

BIAS. für Windows

Handbuch Version 11

Hanns Ackermann



Copyright 1989-2020

Das Programm "BiAS. für Windows" und dessen Handbücher sind gemäß UrhG urheberrechtlich geschützt. Sowohl unbefugte Vervielfältigungen als auch unbefugter Vertrieb des Programms bzw. der Handbücher oder Teilen davon sind somit illegal und werden straf- und zivilrechtlich geahndet. Alle Rechte an dem Softwareprodukt **BiAS.** inklusive Handbücher liegen ausschließlich beim Autor.

Der rechtmäßige Erwerb der Programm-CD bzw. des Original-Downloads und des Handbuches gestattet die Nutzung des Programms **BiAS.** analog zur Benutzung eines Buches. Entsprechend der Unmöglichkeit, dass ein Buch gleichzeitig von verschiedenen Personen an verschiedenen Orten gelesen wird, so darf eine lizenzierte Kopie des Softwareprodukts **BiAS.** nicht von mehreren Personen an unterschiedlichen Orten bzw. auf unterschiedlichen Geräten gleichzeitig benutzt werden. Kopien der Original-CD bzw. des Original-Downloads dürfen ausschließlich nur zum Zweck der Datensicherung angefertigt werden.

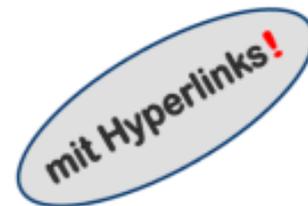
Gewährleistung

Für die Richtigkeit des Inhaltes dieses Handbuches und des Softwareprodukts **BiAS.** wird keine Garantie übernommen. Der Autor, die unter „Vorab“ genannten Beitragenden und der Verlag haften deshalb nicht für eventuell enthaltene Fehler, noch für Neben- und Folgeschäden, die in Verbindung mit der Benutzung dieses Materials entstehen.

Sollten sich einzelne Text- oder Programmteile als fehlerhaft herausstellen, so wird dies nach Maßgabe des Autors auf der Homepage des Programms bekannt gegeben. Gegebenenfalls modifizierte Updates stehen lizenzierten Benutzern per Download kostenfrei zur Verfügung.

Autor

Dr. Hanns Ackermann
Goethe-Univ. Frankfurt
Theodor Stern - Kai 7
60590 Frankfurt am Main



Homepage

<https://www.bias-online.de>

Verlag

Version 1.00: 1989
Version 11.12: 2020

Alle weiteren Aktualisierungen im Programm und auf der Homepage !

epsilon-Verlag GbR Hochheim Darmstadt 1989-2020 • ISBN 3-9803214-8-7

Inhalt



Vorab ...	4
0. Wichtige Informationen zum ersten Start	6
1. Struktur des Programms	8
2. Daten, Dateneditor und das Datenbankmodul XTab	15
3. Import und Export von Dateien und Microsoft's Excel	25
4. Exemplarischer Aufruf einer Funktion	28
5. Graphische Ausgabe	31
6. Installation des Programms	35
7. Validierung des Programms	36
8. Das Leistungsspektrum im Überblick	43
A. Anhang: Validierung und Beispiele	49
A.1 Deskriptive Statistik	49
A.2 Konfidenzintervalle	57
A.3 Vierfelder-Tafeln, Binomial- und Poisson-Tests	60
A.4 Kontingenztafeln	69
A.5 Konfigurationsfrequenzanalyse (KFA)	74
A.6 Tests auf Gauß-Verteilung	78
A.7 Regression und Korrelation	83
A.8 Parametrische Statistik	106
A.9 Nicht-parametrische Statistik	122
A.10 Überlebenszeitanalyse („Survival-Analyse“)	141
A.11 Clusteranalyse	154
A.12 Diskriminanzanalyse	158
A.13 Zeitreihenanalyse	166
A.14 Faktorenanalyse	173
A.15 Reliabilitätsanalyse	176
A.16 Toleranzbereiche und Qualitätssicherung	182
A.17 Pharmakokinetische Parameter und Bioäquivalenz	187
A.18 Methodenvergleiche und Diagnostische Tests	192
A.19 Epidemiologie und “Evidence Based Medicine”	202
A.20 Multiples Testen von Nullhypothesen	204
A.21 Meta-Analyse mit p-Werten	205
A.22 Interims-Analyse: Zwischenauswertungen	208
A.23 Fallzahl- und Powerberechnungen	210
A.24 Randomisierung	226
A.25 Dichte- und Verteilungsfunktionen	228
Literatur	232
Übersicht	246

Dieses Handbuch steht einschließlich Anhang (“Beispiele zur Validierung“) bis Windows XP als Help-“Hypertext-Handbuch“ zur Verfügung; Zum Aufruf (und zur optionalen Nachinstallation ab Vista!) vgl. Abbildung 1. Abbildung 28 in Kapitel 4 zeigt für Windows-Versionen bis XP den Programm-Aufruf der in **BiAS**. verfügbaren Methoden. Ein Überblick über **BiAS**.’ Leistungsspektrum findet sich auf der letzten Seite dieses eBooks, detaillierter im 8. Kapitel und im Internet unter <https://www.bias-online.de>.

Vorab ...

BiAS. ist ein biometrisches Programmpaket, das an der Goethe-Universität Frankfurt am Main speziell für Mediziner und für Humanwissenschaftler aus dem Umfeld der Medizin entwickelt wurde. **BiAS.** bietet ein umfassendes, an den konkreten Fragestellungen der aktuellen medizinisch-biologischen Forschung orientiertes Spektrum von statistischen und biometrischen Methoden. **BiAS.** ist ein Akronym für "Biometrische Analyse von Stichproben".

BiAS.' Methodenspektrum beinhaltet neben "konventionellen" statistischen Methoden auch viele weitere Verfahren, die in der Medizin und deren assoziierten Fachgebieten zwar von grundlegender Bedeutung, aber in anderen Programmen nicht unbedingt verfügbar sind. Das Programm ist somit besonders für Mediziner und Pharmakologen, aber auch für Biologen und Psychologen im Studium und in der wissenschaftlichen Praxis von Interesse: Das Anwenderspektrum reicht von Studenten, Diplomanden und Doktoranden bis hin zu professionellen universitären und industriellen Wissenschaftlern.

Statistik-Programme werden vielfach benutzt wie Programme zur Textverarbeitung. Textverarbeitungsprogramme setzen bei ihren Benutzern eine gewisse natürliche Intelligenz und gelegentlich Grundkenntnisse der Orthographie voraus, Statistik-Programme dagegen beruhen einerseits auf einem ausgedehnten Teilgebiet der Mathematik (Maß- und Wahrscheinlichkeitstheorie, Mathematische Statistik) und andererseits auf eher anwendungsorientierten Fachgebieten wie zum Beispiel der Biometrie. Der Autor rät deshalb davon ab, **BiAS.** ohne adäquate Vorkenntnisse oder vor Lektüre wenigstens eines adäquaten Lehrbuches zu starten. Die Verfügbarkeit über ein wissenschaftliches Programmpaket bedeutet nicht die Verfügbarkeit über dessen wissenschaftliche Inhalte.

Erfahrungsgemäß erlernen die meisten Benutzer die Bedienung eines für sie neuen Programms exemplarisch anhand von speziellen Beispielen, so dass in diesem Handbuch lediglich - aber dafür sehr ausführlich! - die Anwendung des bekannten U-Tests von Wilcoxon, Mann und Whitney beschrieben wird. Die Bedienung der übrigen Module ergibt sich damit: Zur Unterstützung der "informatischen Intuition" der Benutzer/innen dient der Validierungsteil des Handbuches, womit alle Eingaben und Datenformate am praktischen Beispiel ersichtlich sind. Der Autor hofft, dass dieser vielleicht etwas ungewöhnliche Weg einen leichteren Zugang zu der Materie verschafft als eine sich mehrfach wiederholende Aneinanderreihung von technischen Details. Das Handbuch liegt mit gleichen Inhalten sowohl als Hypertext- als auch als eBook-Dokument vor und ist, wie Abbildung 1 unter XP zeigt, im Hauptmenü des Programms per "Buch-Icon" verfügbar. Falls die Hypertext-Version unter Ihrer aktuellern Installation nicht vorhanden ist, kann diese optional nachträglich eingerichtet werden: Mehr dazu finden Sie am Ende dieses Abschnitts.

Der zweite Teil des Handbuchs befasst sich mit der bereits erwähnten Validierung des Programms (Näheres dazu in Kapitel 7!). Alle zur Validierung des Programms verwendeten Datensätze sind im Programm-Unterverzeichnis "..\Valide" enthalten.

Beide Teile dieses Handbuchs werden bis Windows XP vom Programm wahlweise als Hypertext-Handbuch oder als eBook aufgerufen. Falls das Hypertext-Handbuch in Ihrem OS nicht verfügbar ist, wird ohne Abfrage direkt das eBook aufgerufen.

BiAS.' erste Programmversion erschien 1989 unter DOS. Eine vorläufige Portierung nach Windows 95/NT unter C++ und PB/DLL wurde im Jahre 1996 von Herrn Dipl.-Ing. Michael Stotz und mit selbstloser Unterstützung durch Professor Gerd Küveler vorgenommen, Herr Dipl.-Math. Georg Zelmer übernahm dabei die Programmierung des Dateneditors. Bei den späteren Implementierungen haben mir viele Mitglieder des PowerBasic-Forums, insbesondere Chris Boss, Lance Edmonds, Ian McAllister, Semen Matusovski, Dave Navarro und Eric Pearson unter anderem bei Fragen zur Windows-API und -GDI sehr geholfen. Don Dickinson schrieb die Druckroutinen für die Graphik-Ausgabe, Edwin Knoppert den Import aus Excel und Børje Hagsten das Pause-Spiel. James Klutho ist Autor des aktuellen Dateneditors, Jørgen Ibsen hat seinen im Ausgabe-Editor optimal lesbaren Font „Dina“ zur freien Verfügung gestellt.

Frau Professor Eva Herrmann hat mit großem Engagement, vielen konstruktiven Vorschlägen und mit eigenem Quellcode (Valz-Algorithmus und Kerndichteschätzer!) zu der aktuellen Version beigetragen. Herr Dr. Yusuke Asai, Frau Dr. Dimitra Bon, Frau Dr. Tje Lin Chung, Frau Dr. Natalie Filmann, Frau Aksana Käfer B.Sc., Herr Prof. Felix Kawerin und Frau Dr. Julia Rey haben die aktuelle Programmversion einer erneuten kritischen Überprüfung unterzogen, ihrerseits viele Vorschläge eingebracht und auf die eine oder andere „Ecke und Kante“ des Programms hingewiesen.

Allen genannten, aber auch vielen hier namentlich nicht genannten Beteiligten schulde ich für ihre stets bereitwillige und immer kompetente Hilfe und für ihre zahlreichen wertvollen Beiträge meinen ganz herzlichen Dank!

1989 – 2020

Hanns Ackermann

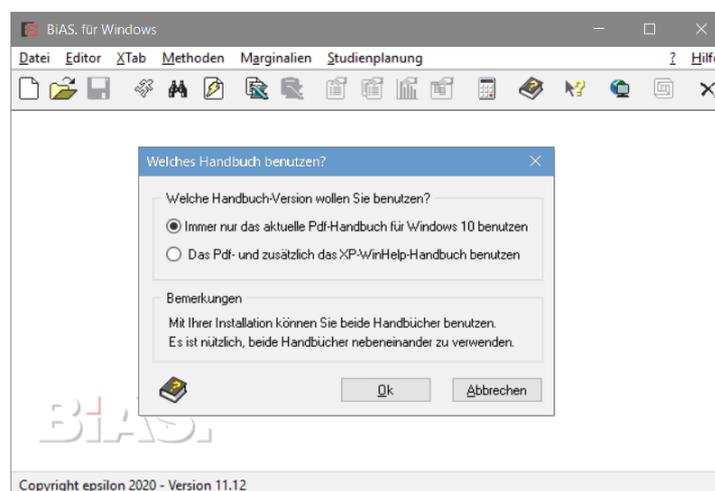


Abbildung 1: XP-Handbücher optional ab Windows Vista, hier unter Windows 10

Das bis Windows XP verwendete WinHelp-Handbuch können Sie optional auch unter allen aktuelleren Windows-Versionen zusätzlich zu diesem eBook verwenden, wenn Sie auf Microsoft's Homepage die Datei WinHlp32.exe downloaden und installieren. (Links dazu unten, ein Kopieren von WinHlp32.exe aus XP in **BiAS.**' Programmverzeichnis (!!)) ist ebenfalls möglich, von MS aber nicht erwünscht.) **BiAS.** erkennt jede Installation und stellt die XP-Handbücher gemäß Abbildung 1 zu Verfügung. Aufschluss über Windows' Installationsstatus und Wahlmöglichkeiten gibt das Menü „Hilfe: Welches Handbuch?“

- ▶ www.microsoft.com/downloads/details.aspx?displaylang=de&FamilyID=6ebcfad9-d3f5-4365-8070-334cd175d4bb
- ▶ www.microsoft.com/downloads/details.aspx?displaylang=de&FamilyID=258aa5ec-e3d9-4228-8844-008e02b32a2c

0. Wichtige Informationen zum 1. Start ▲

Bitte nehmen Sie sich jetzt zwei Minuten Zeit, die folgenden Informationen aufmerksam durchzulesen. Diese ganz kurze Einführung in die Bedienung des Programms erleichtert Ihnen die ersten "Gehversuche" mit **BiAS**. sicher ganz erheblich. Eine Lektüre der folgenden Kapitel kann aber vielleicht auch ganz nützlich sein!

Das vermutlich schwierigste Problem bei der ersten Benutzung eines neuen Programms ist die korrekte Eingabe von Daten:

Als Benutzer von Word, Excel oder auch von anderen Programmen unter Windows ist Ihnen geläufig, dass die unterschiedlichen Programme in der Regel ganz unterschiedliche Datenformate verwenden: **BiAS**. verwendet seinerseits ein spezielles Dateiformat mit der Datei-Extension "DAT". Im zweiten Kapitel finden Sie exemplarisch die Standard-Testdatei Test.DAT, die in vielen Beispielen des Handbuches verwendet wird.

Wichtig: BiAS. verwendet zwecks Kompatibilität mit anderen Programmen für *alle* Berechnungen ausschließlich *numerische* Daten, niemals *Textangaben*! Dies betrifft speziell die Bezeichnungen von beispielsweise Diagnosen, Therapien etc., wie ebenfalls in Kapitel 2 erläutert wird.

Ihre eigenen Daten können Sie über **BiAS**.' Daten-Editor eingeben. Der Editor wird in **BiAS**.' Toolbar per "Leeres Blatt" bzw. „Datei öffnen“ aufgerufen, Einzelheiten finden Sie im 2. Kapitel.

Wichtig: Nach *Öffnen* einer bereits vorhandenen Datei erscheint das Datengrid des Editors und im Status-Fenster (ganz unten!) wird der Name der geöffneten Datei angegeben. Beachten Sie bitte das Handbuch zur Bedienung des Editors, speziell das Rechtsclick-Menü.

Eine Alternative zur Dateneingabe per Editor ist die *Schnelleingabe* von Daten, die ebenfalls in **BiAS**.' Toolbar aufgerufen werden kann ("Blitz-Icon", unten).

Ein Import und Export von Dateien unterschiedlicher Formate ist ebenfalls vorgesehen, insbesondere eine Übertragung von Daten aus Excel über die Zwischenablage. Einzelheiten dazu finden Sie in Kapitel 3.

Nachdem Ihnen der Umgang mit Dateien etwas vertraut geworden ist, sollten Sie einen ersten Aufruf einer Funktion versuchen.

Die zweite wesentliche Hürde bei der ersten Benutzung eines Programms ist der Aufruf einer der angebotenen Funktionen:

Für erste Versuche - auch ohne eine eigene Dateneingabe - können Sie **BiAS**.' Test-Datei Test.DAT verwenden, die bereits oben erwähnt wurde. Die Datei ist in **BiAS**.' Programm-Verzeichnis enthalten und wird in Kapitel 2 kommentiert. Vor Verwendung muss die Datei natürlich via "Datei öffnen" geöffnet werden.

Wichtig: Nach *Datei Öffnen* wird der Name der "aktiven Datei" in der Statuszeile angezeigt (ganz unten im Fenster!). Eine Modifikation der Daten erfolgt über das Rechtsclick-Menü des Editors.

Nützlich: Das beim ersten *Öffnen* verwendete Verzeichnis kann per Icon auf unterschiedliche Weisen voreingestellt werden (Icon "Startverzeichnis", dazu weiter unten): Bitte lesen Sie dazu auch die entsprechenden Erläuterungen zum Arbeitsverzeichnis.

Sicher ist Ihnen der bekannte Zwei-Stichproben-t-Test oder auch der U-Test von Wilcoxon, Mann und Whitney geläufig:

Eine exemplarische Anleitung zum Aufruf dieser statistischen Tests finden Sie in Kapitel 4. Versuche mit anderen Tests oder Graphiken werden Ihnen zeigen, dass die Bedienung des Programms grundsätzlich *immer* nach dem gleichen Muster der genannten beiden Tests verläuft.

Nach Lektüre dieser kurzen Einführung sind Sie sicher in der Lage, mit eigenen Daten die ersten Auswertungen mit BiAS. vorzunehmen. Bitte beachten Sie dazu noch die abschließenden Hinweise:

BiAS. wurde speziell für die Anwendung im medizinischen Bereich entwickelt, und deshalb finden Sie hier sehr viele Methoden, die in anderen Programmpaketen nicht oder nicht ohne weiteres verfügbar sind; einen ersten Eindruck können Sie sich mit **BiAS.**' Menü oder mit Hilfe der Kapitel 7 und 8 verschaffen. Das "Pfeil- und Fragezeichen-Icon" (Symbol unten) in **BiAS.**' Toolbar zeigt Ihnen zu allen Methoden überschaubare Anwendungsbeispiele aus Medizin und Biologie zur Auffrischung bereits vorhandenen Wissens.

In diesem Handbuch (speziell im Anhang: "Beispiele zur Validierung") finden sich ebenfalls zahlreiche Beispiele zu den verfügbaren statistischen Verfahren. Die Help-Hypertext-Version des Handbuchs als Windows-Hilfe-Datei kann bis WinXP auch per "Buch-und-Fragezeichen-Icon" (unten) aufgerufen werden. Ab Windows Vista liegt das Handbuch ausschließlich im vorliegenden eBook-Format vor, es sei denn:

Vielleicht noch interessant: Die ab Windows-Vista verfügbare Handbuchversion kann optional durch die - inhaltlich äquivalente - frühere XP-WinHelp-Version ergänzt werden, wozu in der Regel eine einfache Nachinstallation der Windows-Hilfe erforderlich ist. Einzelheiten zur Installation finden Sie gegebenenfalls im Abschnitt "Vorab" auf Seite 5 dieses Handbuchs, Aufschluss über Windows' Installationsstatus gibt - vgl. Abbildung 2 - das Menü "Hilfe: Welches Handbuch benutzen?".

BiAS. kann - wie alle anderen Statistik-Programme auch - kein Lehrbuch ersetzen: Auf in anderen Programmen gelegentlich vorhandene, wissenschaftlichen Fragestellungen jedoch unangemessene Handlungsanleitungen im Sinne von *was tun wenn* wurde somit explizit verzichtet. Bitte greifen Sie bei allen fachspezifischen Fragen auf die angegebene Literatur zurück: Hinweise dazu finden Sie bitte im Literaturverzeichnis.

Angesprochene Toolbar-Icons:

-  Schnelleingabe von Daten
-  Neue Datei öffnen
-  Vorhandene Datei öffnen
-  Startverzeichnis/Arbeitsverzeichnis einstellen
-  "Pfeil- und Fragezeichen": Anwendungsbeispiele
-  "Buch und Fragezeichen": Handbuch ▲

1. Struktur des Programms

Die ersten Versionen des Programms wurden in den Jahren 1989-1997 unter dem Betriebssystem MS-DOS mit Hilfe der Sprachen *PowerBasic3*, *C* und *Assembler* entwickelt (und sind unverändert lauffähig unter Windows 7/8!). Die Nachfolgerversionen - als erste die Version 5 für Windows 3.x - waren ebenfalls ohne besondere Anforderungen an die Hardware ab 8086/88 mit CGA-Graphik und 460 Kb RAM lauffähig, ebenfalls eine Programmversion für das Betriebssystem OS/2 von IBM.

BiAS.' aktuelle Windows-Version 11 erfordert minimal Windows 95/NT oder optimal nachfolgende Versionen, einen Prozessor ab 486 und VGA-Graphik und ist unter allen Windows-32/64bit-Betriebssystemen lauffähig. Das Programm wurde fast vollständig mit Hilfe von *PowerBASIC*, der nativen Windows-API/GDI und marginal mit *Delphi*, *Fortran* und *Assembler* implementiert. Dank des exzellenten *PowerBASIC*-Compilers beansprucht das Programm trotz seines Leistungsumfangs nur kaum mehr als 20Mb der Festplatte und bietet optimale Laufzeiten: *PowerBASIC* ist ein Produkt von Bob Zale's PowerBASIC Inc., mehr dazu unter www.powerbasic.com.

Nach Start des Programms per Windows' Explorer, dort "Ausführen" oder natürlich optimal per Icon erscheint als erstes der Eingangsbildschirm des Programms bzw. das Hauptmenü, das einen Überblick über die implementierten Verfahren gibt: Beispiele dazu finden sich in Abbildungen 2 und 3. Jeder Punkt des Haupt-Menüs führt zu einem oder zu mehreren Untermenüs, wie im Folgenden exemplarisch erläutert wird. Einige wichtige Informationen zu **BiAS.** - zum Beispiel bezüglich der verwendeten Dateistruktur - gibt das Menü *Hilfe* (aber auch dieses Handbuch!):

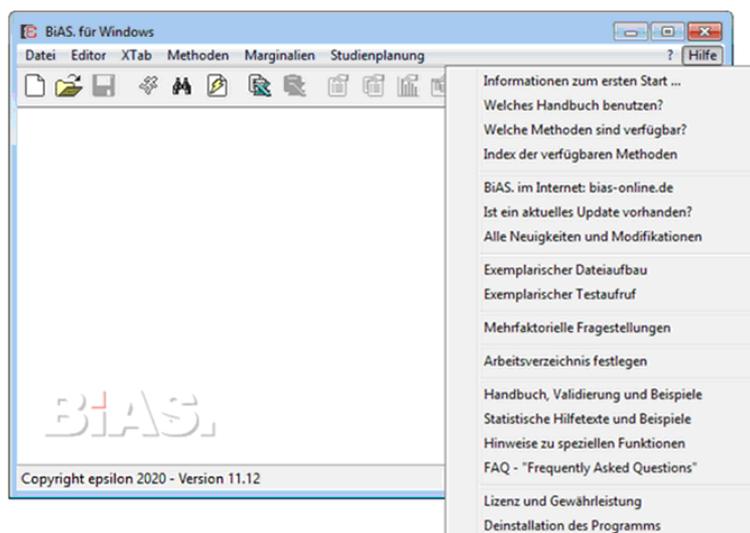


Abbildung 2: BiAS.' Hauptmenü und das Pulldown-Menü "Hilfe"

Alle Menüs des Programms können optimal mit Hilfe der Maus oder alternativ per Shortcuts bedient werden; wie üblich, erfolgt die gewünschte Menü-Auswahl mit der linken Maustaste. Alle Eingabedialoge, Auswahlfenster, Buttons etc. sind nach Windows-Standard ebenfalls per Maus und gegebenenfalls per Shortcuts bedienbar.

Alle statistischen Methoden - mit Ausnahme der Fallzahl- und Power-Berechnungen, der Randomisierungen und des Zufallszahlengenerators, diese finden sich im Menü *Studienplanung* - sind unter den beiden Menüpunkten *Methoden* und *Marginalien* zusammengefasst. Die am häufigsten verwendeten Methoden der Teststatistik (t-Tests, Varianzanalyse, Regression, Wilcoxon-Tests, Kruskal-Wallis-Test, Tests auf Gauß-Verteilung, Survival-Analyse etc.) finden sich im ersten Untermenü *Methoden*, die Cluster- und Diskriminanzanalyse und viele andere Verfahren mehr im zweiten Untermenü *Marginalien*. Die nächste Abbildung 3 zeigt die sukzessive Auswahl einer Methode - hier des Grubbs-Ausreißer-Tests - durch die hierarchischen Menüs:

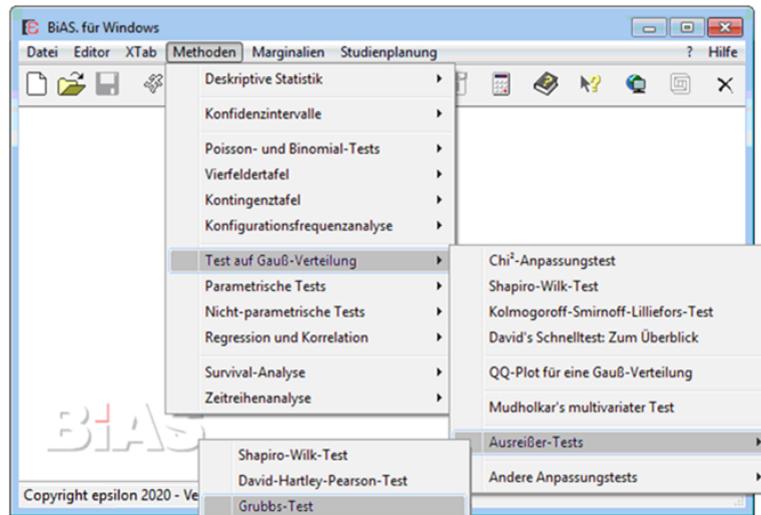


Abbildung 3: Beispiel für BiAS.' Menüführung

Am Ende jeder Programmeinheit wird ein Ausgabe-Fenster aufgerufen, das alle Rechenergebnisse und Literaturhinweise enthält. Abbildung 4 zeigt einen Auszug dieses Fenster nach Durchführung des Fisher-Tests:

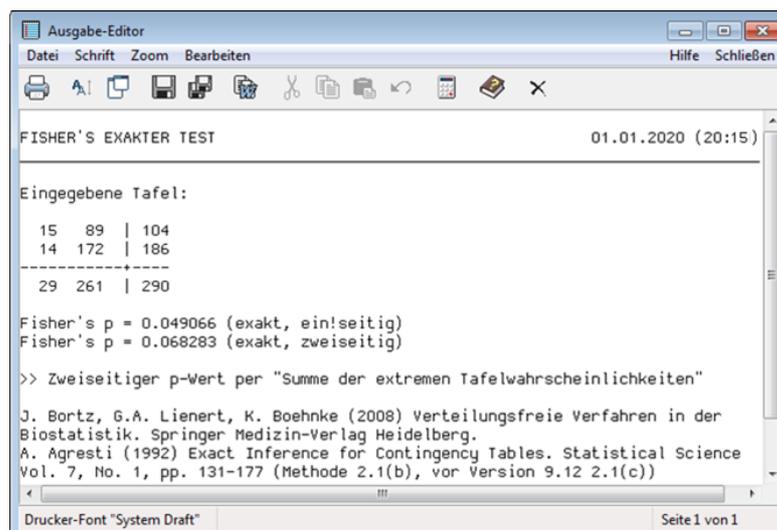


Abbildung 4: BiAS.' Ausgabe-Editor am Beispiel von Fisher's Exact Test

Das Drucker-Icon im Menü des Ausgabe-Fensters eröffnet den üblichen Drucker-Dialog. Eine Einstellung des Druckers ist in der Menü-Leiste des Ausgabe-Editors unter "Datei: Drucker einrichten" vorgesehen. Alle anderen Funktionen entnehmen Sie bitte ebenfalls der Menü-Leiste in Abbildung 4:

Die Einstellung der Schrift (im Menü und per Icon "AA") bezieht sich stets auf die Bildschirm- und auf die Druckausgabe. Es kann die Monotype-Schrift „Dina“ von Jørgen Ibsen als draft/bold (im Druck: Courier-New draft/bold) und alternativ die zwar kompaktere, aber schlechter lesbare Windows-Systemschrift gewählt werden.

Mit Hilfe der *einfachen* Diskette speichert man den kompletten aktuellen Inhalt des Ausgabe-Editors in eine Datei mit frei zu wählendem Namen. Mit Hilfe der *Doppel-*Diskette wird - nach Eingabe eines frei wählbaren Dateinamens - die aktuelle Ausgabe zunächst wie bei der einfachen Diskette gespeichert. Bei einem erneuten Aufruf dieser Funktion wird jedoch kein neuer Datei-Name mehr erfragt, sondern die aktuelle Ausgabe den früheren Ausgaben hinzugefügt. Durch dieses kumulative Speichern können alle Ausgaben eines Laufes in einem Protokoll "gesammelt" und später bearbeitet und/oder ausgedruckt werden.

Den kompletten Text des Ausgabefensters kann man mit Hilfe des MS-Word-Symbols "W" zur weiteren Verarbeitung (zum Beispiel mit MS-Word oder mit anderen Textprogrammen) in die Zwischenablage kopieren. Dies ist - ohne Markierung! - nur für die komplette Ausgabe vorgesehen, markierte Teile der Ausgabe kopiert man via *Kopieren*:

Die Funktionen *Ausschneiden*, *Kopieren*, *Einfügen* und *Wiederherstellen* entsprechen den üblichen, zum Beispiel in MS-Word verwendeten Windows-Funktionen zum Editieren und behandeln immer nur *Teile* der Ausgabe. Bei einer Ausgabe auf den Drucker wird bei markierten Textteilen ungeachtet anderer Einstellungen nicht der gesamte Text, sondern nur die markierte Textpassage ausgedruckt.

Das Buch-und-Fragezeichen-Icon beantwortet eventuelle Fragen zur Druckausgabe. Das Kreuz-Icon beendet den Ausgabe-Editor, worauf wieder **BiAS.**' zuletzt aktiver Eingabe-Dialog erscheint.

Von einigem Interesse ist vielleicht auch **BiAS.**' Hilfefunktion, die explizit mit Hilfe des Pfeil-und-Fragezeichen-Icons aufgerufen werden kann:

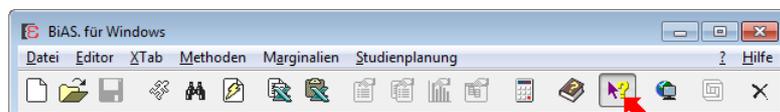


Abbildung 5: Aufruf der Hilfe-Funktion: Biometrische Beispiele

Die Hilfefunktion ist Bestandteil eines zweiteiligen Hilfesystems: Die Hilfefunktion erläutert an Hand von Beispielen die biometrischen Inhalte aller Verfahren, der zweite Teil des Hilfesystems ist in den Menüs über die Wahlmöglichkeit *Hinweise* verfügbar und beschäftigt sich in der Hauptsache mit programmtechnischen Details, weiterführenden Erläuterungen und Hinweisen.

Nach Aufruf meldet sich **BiAS.**' Hilfesystem mit einem Eröffnungsbildschirm, die Bedienung erfolgt mit Auswahlbuttons:

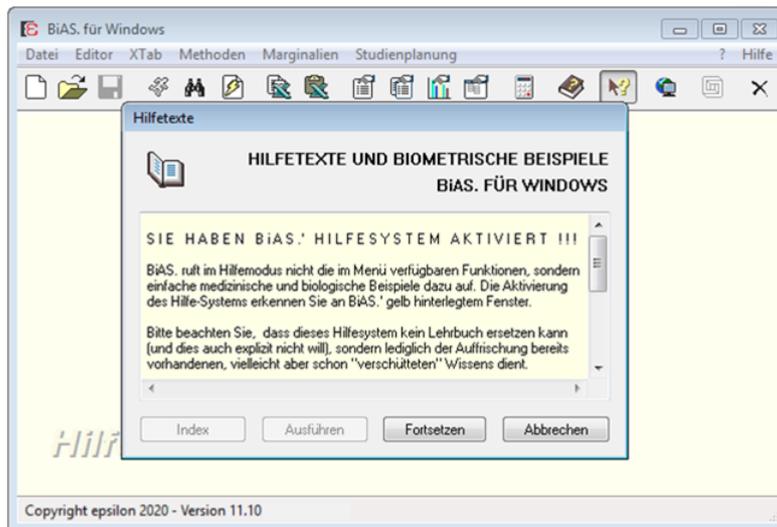


Abbildung 6: Eröffnungsbildschirm der Hilfefunktion

Benutzt man jetzt nach <Weiter> das Hauptmenü in gewohnter Weise, so erscheinen in einem Fenster an Stelle der gewählten Funktion Hinweise und speziell einfache Beispiele zu der angewählten Methode. Abbildung 7 zeigt dazu ein Beispiel anhand des Kaplan-Meier-Schätzers:



Abbildung 7: Beispiel für BiAS.' Hilfetext

Der Autor verfolgt mit diesen Hilfetexten ausdrücklich *nicht* die Absicht, ein statistisches Lehrbuch zu ersetzen, vielmehr sind die Hilfe-Texte lediglich dazu geeignet, vielleicht irgendwann vorhandenes Wissen wieder zu aktivieren: In der Einleitung wurde bereits darauf hingewiesen, dass **BiAS.** - wie alle anderen statistischen Programmpakete auch - erst nach Lektüre eines einschlägigen Lehrbuches verwendet werden sollte. Im Abschnitt "Literatur" finden sich dazu einige grundlegende Lehrbücher und weitere Literaturhinweise zu allen im Programm angebotenen Verfahren. Zum raschen Nachschlagen empfiehlt sich Sachs (1968-2003) bzw. die Nachfolgerwerke von Sachs und Hedderich (2006,2009) und von Hedderich und Sachs (2015).

Mit Hilfe des Buch-und-Fragezeichen-Icons kann man bis Windows XP die Validierungs- und Bedienungsbeispiele des 7. Handbuch-Kapitels aufrufen (bitte dazu die Bemerkungen in der Einleitung beachten!!). Auf dem gleichen Wege kann auch, ebenfalls bis Windows XP, die im Programm implementierte Hypertext-Version dieses Handbuches aufgerufen werden. Für Windows ab Vista bzw. für aktuellere Windows-Versionen steht primär dieses Pdf-Handbuch zur Verfügung.

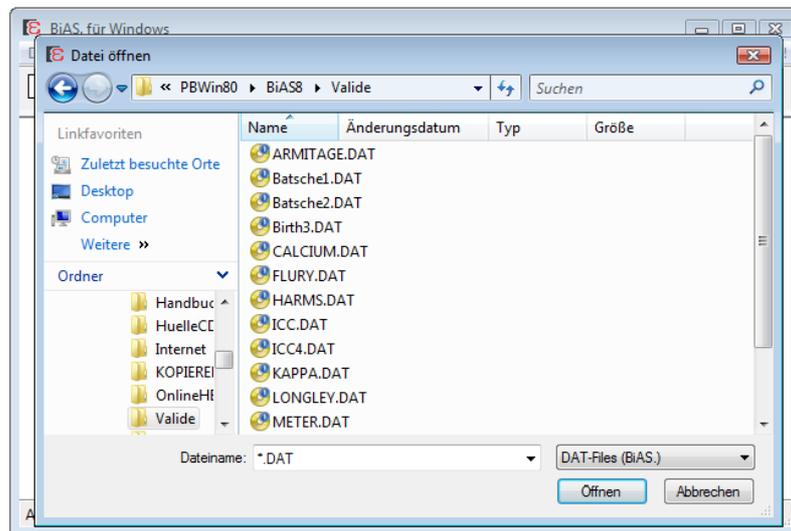


Abbildung 8: Eingabe einer aktiven Datei im Menü “Datei“ bzw. per Icon

Per *Datei öffnen* wird, wie üblich, eine Eingabedatei zur weiteren Verarbeitung ausgewählt. Die Struktur einer Eingabedatei wird im nächsten Kapitel oder auch im Menü *Infos* beschrieben und ergibt sich zwangsläufig bei Verwendung des Editors (*Neue Datei eingeben*). **BIAS.** Dateien entsprechen den SDF-Standard-Dateien und können in dBasell- und in Text-Dateien (Windows, Excel) exportiert werden; vgl. dazu den Import und Export von Dateien im zweiten Kapitel. **BIAS.** merkt sich einen eingegebenen Dateinamen für den Bereich des gesamten Programms, wobei spätere Modifikationen natürlich möglich sind. Abbildung 8 zeigt die Eingabe einer “Aktiven Datei”, die im Anschluss für die weitere Verwendung zur Verfügung steht.

Die nächste Abbildung zeigt den Aufruf für die Einstellung des Standard-Arbeitsverzeichnisses. Nach der ersten Installation des Programms wird das Verzeichnis, in dem Sie **BIAS.** installiert haben, als Standard verwendet: Modifikationen sind hier möglich.

BIAS. liest Daten mit einer einfachen Genauigkeit (d.h. auf 6 Dezimalstellen genau); sodass möglichst nur Eingabewerte zum Beispiel $|X| < 999999$ (ohne Nachkommastellen) oder zum Beispiel $|X| < 99.9999$ (2 Vor- und 4 Nachkommastellen) verwendet werden sollten. Andere Eingaben können zu entsprechenden Rundungsfehlern etwa bei der Berechnung von arithmetischem Mittelwert oder von Streuungskomponenten führen. Ungeachtet dieser Festlegung führt **BIAS.** fast alle Berechnungen intern in doppelt- oder dreifachgenauer Darstellung aus (d.h. auf 16- bzw. 18 Dezimalstellen genau), worauf weiter unten noch ausführlicher eingegangen wird. Vorsicht: **BIAS.** fragt die Beträge von Eingabedaten nicht explizit ab, eine Überprüfung ist lediglich im Modul *Import-Export* möglich („*Format einer importierten Datei prüfen*“).

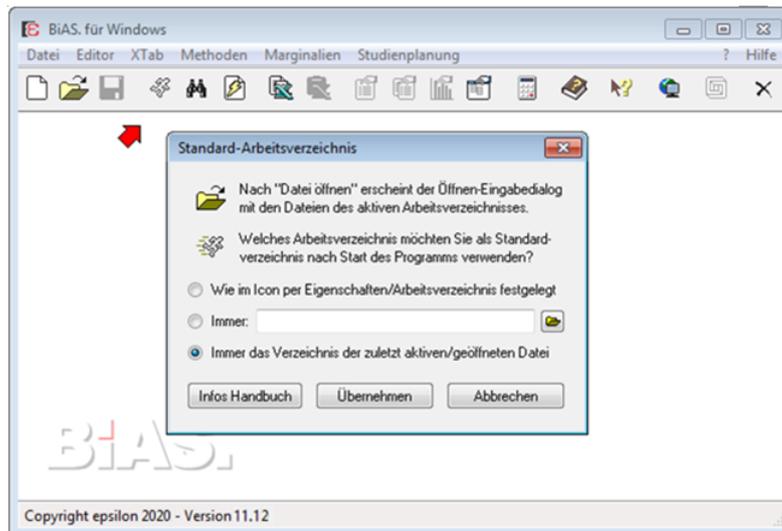


Abbildung 9: Festlegung des Standard-Arbeitsverzeichnisses

Bei der Eingabe von negativen Zahlen müssen negative Vorzeichen *ohne* Zwischenraum vor dem Betrag der einzugebenden Zahl stehen, da ansonsten das Programm den Wert "0" als Eingabe erkennt. (Also z.B. nicht "- 1", sondern "-1" eingeben!) Die oben erwähnte Import-Prüfung von Dateien stellt diesen Fehler ebenfalls fest.

BiAS. verwendet aus systematischen Gründen grundsätzlich nur natürliche Zahlen, also ganze Zahlen größer als Null (die obere Grenze beträgt 1000) als Gruppennummern; andere Werte können unter Umständen zu Fehlern führen. Nicht zulässige Gruppennummern - zum Beispiel aus Fremddateien - sollten deshalb gegebenenfalls umbenannt werden. Das Procedere zeigt Abbildung 12 im 2. Kapitel.

BiAS. arbeitet mit formatierten numerischen ASCII-Dateien, die im nächsten Kapitel ("Editor") näher beschrieben werden; vgl. auch Menü Infos, "Wie ist eine Datei aufgebaut?". **BiAS.**' Dateien entsprechen im Wesentlichen dem Standard-Dateiformat SDF und können auch als dBasell-, SPSS- und Windows- bzw. Excel-Text-Dateien (*.DBF, *.SAV, *.TXT, *.CSV) importiert und exportiert werden.

Eine Datenmatrix bzw. eine Eingabe-Datei enthält *zeilenweise* die Beobachtungseinheiten (das sind Probanden, Patienten etc.) und *spaltenweise* die Werte von Merkmalen (Messwerte, z.B. Laborwerte eines Patienten) resp. die Werte einer Variablen an verschiedenen Zeitpunkten. Als weitere Variablen kommen noch eventuell Gruppierungsvariablen hinzu, die, z.B. für eine Varianzanalyse, die Gruppenzugehörigkeit der Daten zu Behandlungsgruppen bzw. Bedingungen kennzeichnen. **BiAS.** erlaubt grundsätzlich eine beliebige Anzahl von Zeilen (also Beobachtungseinheiten wie Patienten, Probanden o.ä.), jedoch maximal 200 Variablen pro Zeile. Die verschiedenen Verfahren gestatten jeweils eine Auswahl von Variablen/Spalten bzw. gewisser Teilstichproben/Gruppen. Der Stichprobenumfang ist, von einigen Ausnahmen abgesehen, auf maximal $n=2000$ Werte pro Gruppe bzw. in Abhängigkeit von der Gruppenzahl durch den vorgesehenen Speicherplatz beschränkt. Eine Ausnahme bilden z.B. die Datenbankbefehle in XTab, die diesen Beschränkungen nicht unterworfen sind oder einige regressions- und korrelationsanalytische Methoden, die mit $n \leq 10000$ Werten durchgeführt werden können. Auf alle eventuellen speziellen Beschränkungen wird in den entsprechenden Moduln hingewiesen.

Merkmalsträger mit fehlenden Werten in mindestens einer Komponente, die für eine konkrete uni- oder multivariate Fragestellung relevant ist, werden von **BiAS**. bei der Auswertung ignoriert. Fehlende Werte in einer Datenstruktur werden - wie üblich - einfach nicht eingetragen und die Felder sind somit *leer*: Keinesfalls darf ein Dezimalpunkt o.ä. verwendet werden, da Letzteres unweigerlich zu Fehlern führt!

“Kritische” Berechnungen werden von **BiAS**. in dreifachgenauer Darstellung (18 Stellen) vorgenommen: Dies betrifft die Parametrische Statistik, speziell etwa die Berechnung von Streuungskomponenten und den Bereich der Regressions- und Korrelationsrechnung. **BiAS**. genügt somit den aktuellen Anforderungen an die Rechengenauigkeit der verwendeten Algorithmen, die sich wie Two-Pass-Algorithmen verhalten; vgl. dazu etwa Berk (Amer. Statist. 41, 3, 1987, pp. 222-8) oder Dallal (Amer. Statist. 42, 3, 1988, pp. 212-6). Von wenigen Ausnahmen abgesehen, werden die Berechnungen in allen anderen Fällen mit doppelter Genauigkeit (16 Stellen) durchgeführt.

Alle Rangordnungen in **BiAS**. werden mit einem “schnellen”, das heißt $O(n \cdot \log n)$ -Assembler-Algorithmus vorgenommen, wodurch auch bei nicht-parametrischen Verfahren zufriedenstellende Rechenzeiten erzielt werden können.

Ein eventuell vorhandener Coprozessor wird von **BiAS**. erkannt und gegebenenfalls auch verwendet. Eine besondere Einstellung ist dazu nicht erforderlich.

Zu allen nicht-parametrischen Testverfahren gibt **BiAS**. sowohl approximative als auch exakte Prüfgrößen einschließlich deren “Überschreitungswahrscheinlichkeiten” p („ p -Werte“) an; dies gilt natürlich auch für alle parametrischen Verfahren. Speziell im relevanten Fall von kleinen Stichprobenumfängen kann die Prüfung einer Nullhypothese somit ohne Verwendung von Tabellen nicht nur approximativ, sondern auch anhand der exakten Nullverteilung eines statistischen Testes erfolgen. (Unter einer “Überschreitungswahrscheinlichkeit” versteht man die Wahrscheinlichkeit, dass - bei Gültigkeit der Nullhypothese - die vorliegende oder eine “extremere” Stichprobensituation eintritt. Bei exakt errechneten p -Werten speziell in der parametrischen Statistik, aber auch in der nicht-parametrischen Statistik etwa zum U-Test oder zu Fisher’s exaktem Test ist die Schreibweise “ $p = p_{\text{errechnet}}$ ” angemessen, wogegen bei nicht-parametrischen Tests mit tabellenorientierten p -Werten in der Regel “ $p \leq p_{\text{Tabelle}}$ ” verwendet werden sollte. **BiAS**. verwendet in Abhängigkeit von der Sachlage das Gleichheitszeichen oder die \leq -Relation.)

BiAS. bestimmt zu allen Graphiken die Skalierungen der Koordinatenachsen durch einen speziellen Algorithmus, so dass sich stets sinnvolle und sinnvoll gerundete Werte ergeben. Umständliche Eingaben seitens des Benutzers bzw. unhandliche Intervallskalierungen durch das Programm entfallen damit. Trotzdem gibt **BiAS**. optional Gelegenheit zur Modifikation der vorgegebenen Skalierung, zum Beispiel um verschiedene Grafiken mit der gleichen Skalierung zu versehen. Die Kopfzeile einer Graphik und deren Achsenbeschriftung kann optional überschrieben und die Größe der Achsenbeschriftungen kann verändert werden. Hardcopies von Graphiken sind im Graphik-Editor (vgl. Kapitel 5) oder nach Abschluss der Graphik via *Letzte Graphik zeigen* vom Hauptmenü aus möglich. Im Abschnitt “Graphische Ausgabe” findet sich dazu ein Beispiel; vgl. Sie dazu das Dialogfeld *Graphik-Einstellungen* in Abbildung 31 des 5. Kapitels.



2. Daten, Editor und XTab

BiAS.' Editor wird vom Hauptmenü aus aufgerufen und arbeitet mit formatierten numerischen ASCII-Dateien. Jeder Datei - in diesem Handbuch gelegentlich auch als "Datenmatrix" oder als "Tabelle" bezeichnet - sind die Namen der Variablen bzw. Spalten vorangestellt; die Namen sollten aus Kompatibilitätsgründen nicht länger als 8 Zeichen sein. **BiAS.**' Dateneditor ist in Abbildung 10 mit einem Popup-Menü dargestellt, das per Rechtsklick auf die Namenzeile geöffnet wird und, von wenigen Zusatzfunktionen im Menü „Editor“ abgesehen, alle Elemente des Editors enthält:

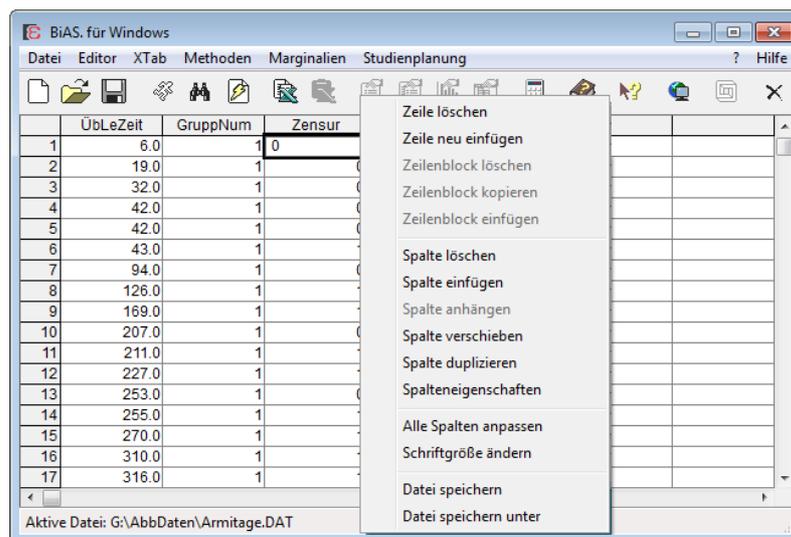


Abbildung 10: **BiAS.**' Editor mit Popup-Menü

Bei Rechtsklick auf ein *aktiviertes* Edit-Fenster (fetter Rand!) erscheint das übliche Menü mit „Cut-Copy-Paste-Undo“. Eine Datei-Übertragung z.B. aus Excel erfolgt *ausschließlich* über den rechten Excel-Button der Toolbar, mehr dazu weiter unten.

BiAS.' Editor gestattet die Neugenerierung von Dateien, verschiedene Korrektur- und Ergänzungsmöglichkeiten bereits vorhandener Dateien, beliebige Datentransformationen (u.a. via Formelinterpreter) und natürlich eine klartextliche Bezeichnung von Variablen; Namen und Formate können zu einem späteren Zeitpunkt geändert werden. Alle Funktionen erklären sich nach Aufruf selbst und werden - mit Ausnahme des Formelinterpreters - deshalb nicht im Einzelnen erläutert.

Eine Datei kann prinzipiell beliebig viele Zeilen ("Merkmalsträger", also Patienten, Probanden etc.) und maximal 200 Spalten ("Variablen" bzw. "Parameter") umfassen.

Alle Werte, die einem Merkmalsträger (d.h., Patient/Proband o.ä.) zugeordnet sind, werden in **BiAS.** stets *in genau einer Zeile* angeordnet. Somit entspricht eine Zeile *genau einem* Merkmalsträger und eine Spalte *genau einem* Merkmal (Variable).

Merkmalsträger können zu *Gruppen* ("Stichproben", z.B. Therapien oder Diagnosen) zusammengefasst werden. Variablen, die Gruppen kennzeichnen, werden in **BiAS.** stets als *Gruppenvariablen* bezeichnet, im Beispiel etwa ist dies die dritte Variable

“GruppNum”. Gruppen werden immer mit Zahlen 1,2,...,1000 bezeichnet, Gruppennummer 0 ist nicht zulässig. Dazu ein Beispiel im geöffneten Editor:

	VarEins	VarZwei	GruppNum	Gruppe
1		2.0	1	Gruppe A
2	2.0	3.0	1	Gruppe A
3	6.0	5.0	1	Gruppe A
4	4.0	5.0	1	Gruppe A
5	5.0	4.0	1	Gruppe A
6	10.0	20.0	2	Gruppe B
7	15.0	25.0	2	Gruppe B
8	20.0	30.0	2	Gruppe B
9	30.0	35.0	3	Gruppe C
10	32.0	33.0	3	Gruppe C
11	36.0	32.0	3	Gruppe C
12	39.0	39.0	3	Gruppe C
13				
14				
15				
16				
17				

Aktive Datei: G:\AbbDaten\Test.DAT

Abbildung 11: BiAS.’ Test-Datei Test.DAT

Die fehlende Angabe in der ersten Zeile/Spalte bedeutet einen fehlenden Wert, eine spezielle Symbolik ist in solchen Fällen also nicht erforderlich. Alle eventuell vorhandenen Textangaben werden ebenfalls als “fehlend” interpretiert.

Im Beispiel existiert nur die Gruppenvariable *GruppNum*. Prinzipiell dürfen beliebig viele Gruppenvariablen verwendet werden, die ganzzahlige Werte von 1 bis 1000 annehmen können. Die Verwendung von Gruppenvariablen ist nicht obligat.

Gruppennummern können im Programm mit der Funktion *Gruppennummern ändern* (im Editor-Menü) sehr einfach verändert werden:

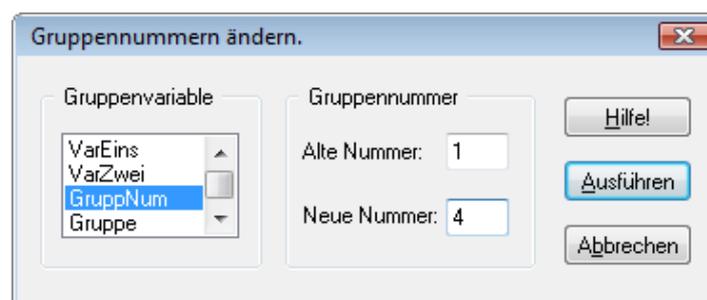


Abbildung 12: Gruppennummern ändern

Für jeden Merkmalsträger, also pro Zeile der Datenmatrix, können bis zu 200 Variablen eingegeben werden. Eine Zeile darf inclusive Trennzeichen (blanks, “Leerzeichen”) bis zu 32768 Zeichen lang sein.

Bei Neueingabe einer Datei (nächste Abbildung!) und analog bei späterem Ergänzen von Spalten (also von Variablen, nach Rechtsklick) ist für jede Spalte eine *klartextliche Bezeichnung* (Namen optimal nicht länger als 8 Zeichen!), die *Typ-Angabe* und eine Angabe des *Spalten-Formats* erforderlich:

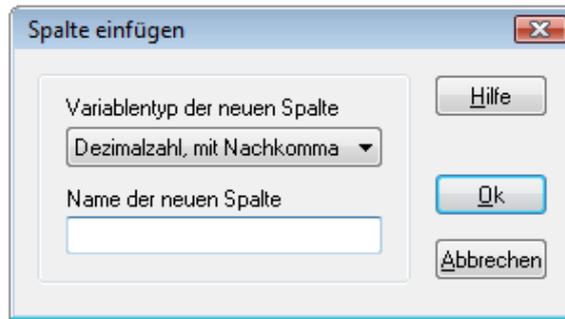


Abbildung 13a: Neue Daten eingeben bzw. Spalte einfügen

In **BiAS.**' Editor können vier verschiedene Datentypen verwendet werden:

Typ Dezimalzahl: Dies sind Zahlen mit Dezimalstellen („Numerisch mit Nachkomma“). Bei der Definition einer Spalte muss dieser oder ein anderer Typ (dazu weiter unten) angegeben werden.

Typ Ganzzahl: Dies sind ganzzahlige Variablen („Numerisch ohne Nachkomma“). Bei Definition einer Spalte sollte dieser Variablentyp zum Beispiel für „Gruppenvariablen“ verwendet werden (die immer ganzzahlig und positiv sein müssen).

Typ Datum: Datumsvariablen werden immer im Standard-Format TT.MM.JJJJ - z.B. also 31.01.2004 – eingegeben. Datumsvariablen können nur in Transformationen, nicht aber für allgemeine Berechnungen verwendet werden.

Typ Text: Textvariablen dienen nur zur Information und können niemals für Berechnungen o.ä. verwendet werden!

BiAS. verwendet für alle Berechnungen ausschließlich *numerische* Daten, keine *Textvariablen*! Die Verwendung von Textvariablen dient nur zur Information, die Werte insbesondere von Gruppenvariablen müssen also immer *Zahlenwerte* sein, keine *textlichen* Bezeichnungen. Die Darstellung von Zahlen beträgt 8 Stellen.

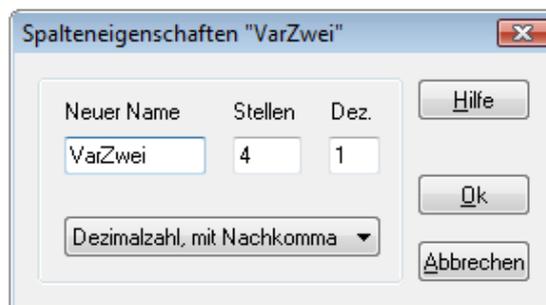


Abbildung 13b: Spalteneigenschaften ändern

Die Spaltenbezeichnungen, der Datentyp und das zunächst intern festgelegte Datenformat können im Rechtsklick-Menü des Dateneditors per „*Spalteneigenschaften*“ jederzeit geändert werden. Letzteres ist in praxi jedoch nur bei eventuell gewünschter Reduktion der Nachkommastellen einer Dezimalzahl von Belang, ansonsten eher gegenstandslos. Falls größere als die vorläufig vorgesehenen Werte eingegeben werden, ändert der Editor selbständig die Formatierung, Entsprechendes gilt auch für eine eventuell größere Anzahl von Nachkommastellen von Dezimalzahlen.

Bitte beachten Sie bei der Dateneingabe und -korrektur die übliche “<Enter>-Konvention” zur fortschreitenden Eingabe und Korrektur der Zellen der Datenmatrix. Bei Eingabe von Daten wird via <Enter> in der letzten Zelle einer Zeile automatisch eine neue Zeile erzeugt und der Eingabefocus auf die erste Zelle der neuen Zeile gesetzt. <Tab> führt die Eingabe nur in der aktuellen Zeile fort.

Bei Erstellung einer Datei werden die Eigenschaften (d.h. Namen und Typ) jeder Spalte gemäß Konvention separat abgefragt (Abbildung 13). Nach der Definition einer Spalte wählt man per Rechts-Click die Funktion “ Spalte einfügen“, um der Datei eine weitere Spalte hinzuzufügen (dazu die Abbildungen 10 und 13a).

Nach Abschluss der Dateneingabe, Korrektur oder einer Modifikation wählen Sie - in Abweichung von vielen Microsoft-Programmen - immer *Datei speichern* (per R-Click oder per Toolbar), um Ihre Eingaben zu sichern. Damit wird einem unbeabsichtigten Verlust Ihrer Daten vorgebeugt. Ihre neue bzw. modifizierte Datei können Sie wie gewohnt vom Hauptmenü aus via *Datei öffnen* für die Benutzung in **BiAS**. öffnen.

In **BiAS**.’ Editor stehen einige Funktionen zum Ergänzen und Löschen von Zeilen und Spalten zur Verfügung (vgl. Rechtsclick-Menü in Abbildung 10: Zeilen ergänzen oder löschen, Spalten löschen, einfügen und anhängen). Bezug ist jeweils diejenige Zeile bzw. Spalte, die den Eingabe-Focus besitzt: Ein Feld dieser Zeile bzw. Spalte ist mit einem etwas dickeren Rand versehen.

Das *Kopieren* von Spalten erfolgt im Rechtsclick-Menü per „*Spalte kopieren*“. Die Spalte mit dem Edit-Focus wird - nach Abfrage eines neuen Namens - identisch vor der aktuellen Spalte eingefügt, eine *Verschiebung* von Spalten erfolgt im Editor per Rechtsclick und Wahl der Zielposition der Spalte.

Analoge Operationen mit *Zeilenblöcken* sind ebenfalls möglich: Ein Zeilenblock wird per linker Maustaste in der Spalte der Zeilenheader markiert (und dadurch farblich hinterlegt) und per Rechtsclick-Menü die gewünschte Operation ausgeführt, *mehrere* Leerzeilen können nach Markierung leerer Zeilen am Ende des Datengrids und Markierung der Zielposition per R-Click-Menü eingefügt werden.

In **BiAS**.’ Editor wird die Einstellung des Arbeitsverzeichnisses grundsätzlich nicht geändert. Öffnen Sie dagegen eine neu erzeugte oder mit *Speichern unter* gespeicherte Datei per Icon vom Hauptmenü aus, so wird dort gleichzeitig das Arbeitsverzeichnis als Default auf das Verzeichnis der eben geöffneten Datei eingestellt.

Eine Datei kann prinzipiell beliebig viele Zeilen (“Merkmalsträger”, also Patienten, Probanden) und maximal 200 Spalten (“Variablen”, zum Beispiel medizinische Parameter) umfassen.

Fehlende Werte in einer Datenstruktur werden, wie auch an anderer Stelle erwähnt, ihrer Natur folgend einfach nicht eingetragen, das entsprechende Feld bleibt also ohne einen Eintrag *leer*.

Erwähnenswert ist die Backup-Funktion des Editor-Moduls (nur in **BiAS**.’ Hauptmenü): Einerseits kann man, wie üblich, vollständige Kopien einer vorhandenen Datei herstellen, andererseits können auch “eingeschränkte” Kopien einer Datei erzeugt werden, wobei die erste und die letzte Zeilennummer angegeben wird.

BiAS. akzeptiert auch lange Dateinamen. Aus Kompatibilitätsgründen ist es jedoch günstig, trotzdem nur Dateinamen mit maximal 8 Zeichen zu verwenden. Dies wird im Editor bei Eingabe abgefragt.

Neben der Dateneingabe über den Editor sind zwei weitere Varianten denkbar:

Die erste Variante besteht in einem Import (Export ist auch möglich: Kapitel 3!) von Dateien. Dieser erfolgt nicht im Rahmen des Editors, sondern in **BiAS.**' Hauptmenü. Speziell Excel-Dateien können auch über die Zwischenablage (per Shortcuts bzw. via Strg-<C> und Strg-<V>) importiert bzw. exportiert werden.

Die zweite Variante besteht in der Schnelleingabe von Daten (im Editor-Modul via Hauptmenü erreichbar, Blitz-Icon), womit **BiAS.**' Editor und dessen Formatspezifikationen umgangen werden kann. Eine Modifikation der "auf die Schnelle" eingelesenen Daten ist nur über **BiAS.**' Editor möglich.

Häufig genügt es, Daten ohne eine präzise Definition der Datenmatrix (vgl. dazu **BiAS.**' Editor) einzugeben. **BiAS.**' Schnelleingabe verwendet Windows' Zwischenablage und wird per "Blitz-Icon" in der Toolbar oder im Editor-Modul des Hauptmenüs aufgerufen; eine Sicherung der eingegebenen Daten in **BiAS.**' Dat-Format ist vorgesehen.

Zur Trennung der Daten wird im Fenster der "Schnelleingabe" ein Semikolon verwendet; äquivalent dazu kann stattdessen auch ein (nicht sichtbares!) <Tab> eingegeben werden:

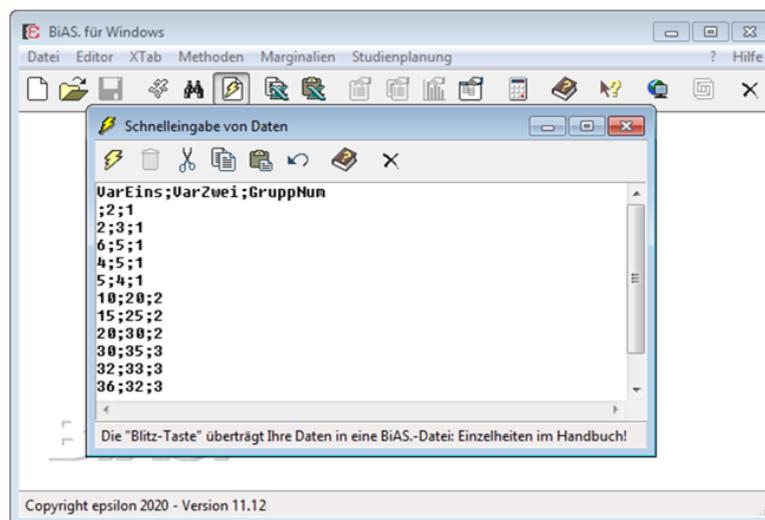


Abbildung 14: "Schnelleingabe" von Daten

BiAS.' Test-Daten (Abbildung 11: "Test.DAT", steht in **BiAS.**' Verzeichnis!) können via Schnelleingabe auch gemäß der Abbildung 14 eingegeben werden, die im Daten-Editor obligatorische Definition der Spalten ist hier nicht erforderlich. Bitte beachten Sie den fehlenden Wert in der zweiten Zeile in Abbildung 14: Vor dem ersten Semikolon wurde kein Wert angegeben! Einige Tips zur Dateneingabe - zum Beispiel bezüglich der Nachkommastellen - finden Sie weiter unten.

Die Kopfzeile mit den Spaltennamen wird obligat als erste Zeile eingegeben, in den nächsten Zeilen folgen die numerischen (!) Daten. Als Trennzeichen zur Abgrenzung von Spaltennamen in der ersten bzw. von Daten in den folgenden Zeilen wird das Semikolon oder äquivalent ein "<Tab>" verwendet.

Nach Abschluss der Dateneingabe übertragen Sie die Daten mit Hilfe der *Blitz-Taste* in eine **BiAS**.-Datei, deren Datei-Name abgefragt wird. Das Fenster in Abbildung 15 dient der Bestätigung, ob die erste Zeile aus Spaltennamen oder - entgegen der Empfehlung - bereits aus Daten besteht:

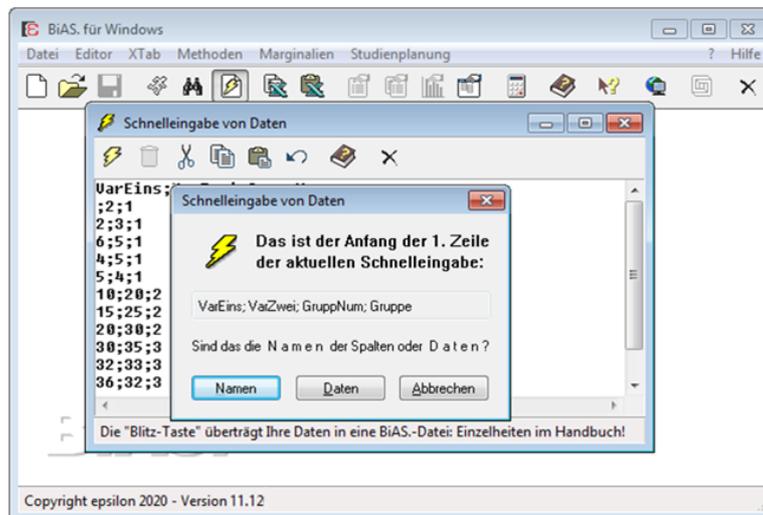


Abbildung 15: "Schnelleingabe" von Daten: Abspeichern

BiAS. erlaubt beliebige Transformationen von Daten: Über **BiAS.** Hauptmenü (dort: *Editor*, dann *Transformationen*) können beliebige Transformationen von Dateien durchgeführt werden. (Für Kenner der DOS-Version: Der Interpreter "**BiAS.** BASIC" der früheren DOS-Version ist unter Windows nicht mehr verfügbar!)

Fehlende Werte in einer Datenstruktur werden, wie bereits erwähnt, ihrer Natur folgend einfach nicht eingetragen, das entsprechende Feld bleibt also leer. Das Transformationsmodul (siehe Abbildung 16) interpretiert diese Daten in entsprechender Weise als fehlend.

BiAS. erfragt nach Wahl des Moduls eine Ein- und eine Ausgabedatei. Falls die Ausgabedatei bereits existiert, so wird dies reklamiert und **BiAS.** gibt Gelegenheit zur Korrektur bzw. zur Bestätigung des gewählten Programmnamens. Grundsätzlich können die Ein- und Ausgabedatei identisch sein, auch wenn sich ein Überschreiben der Eingabedatei nicht unbedingt empfiehlt. Abbildung 16 gibt einen Überblick über die Eingabemöglichkeiten.

Die Eingabe erfolgt grundsätzlich in *genau einer* Zeile, die in konventioneller BASIC-Notation zu formulieren ist. Stößt der Formelinterpreter bei der Berechnung auf einen fehlenden Wert, so wird auch die Zielvariable (links vom Gleichheitszeichen) undefiniert. Prinzipiell wählt man Variablen in der List-Box aus, drückt *In die Formel* (oder schreibt gleich z.B. X1 für die erste Variable) und definiert gemäß der Hinweise in Abbildung 16 eine Transformation: Zum Beispiel erreicht man mit "X6=X1+X2", dass die 6. Spalte einer Datenmatrix durch die Summe der ersten beiden Spalten ersetzt wird. Weitere Einzelheiten und Beispiele entnehmen Sie bitte den im Programm verfügbaren "Hinweisen" zum Formelinterpreter und auch der Abbildung 16:

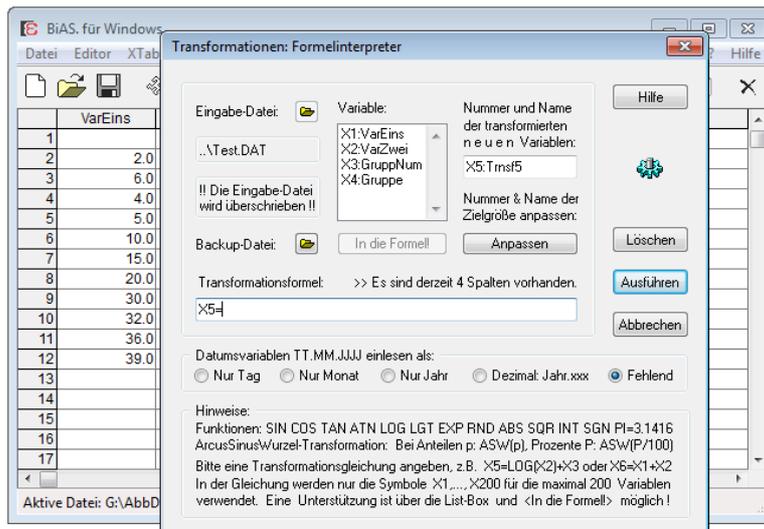


Abbildung 16: Dialogfeld "Transformationen"

Als Ergänzung zum Dateneditor besitzt **BiAS.** ein Datenbankmodul ("XTab") zur Bearbeitung und Verknüpfung von Dateien:

BiAS. bietet im Modul XTab ein eingeschränktes Datenbankmodul zur Unterstützung des Editors. XTab sieht Selektionen von Zeilen einer Datei gemäß benutzerdefinierter Kriterien vor und erlaubt neben vielfältigen Verknüpfungen von Dateien auch das Auszählen von Daten nach Definition von Bedingungen bzw. gemäß logischer Konstruktionen. Wie in vielen anderen Modulen des Programms gibt darüber auch **BiAS.** technisches Hilfesystem Aufschluss (Hinweise). Es empfiehlt sich, die Funktionen des Editors und des Menüs XTab anhand des Test-Datensatzes "Test.Dat" (Abbildung 11, steht auch in **BiAS.** Verzeichnis!) ausführlich auszuprobieren.

Nach Aufruf des Datenbankmoduls XTab erscheint das Menü in Abbildung 17, dessen Funktionen anschließend kurz erläutert werden:

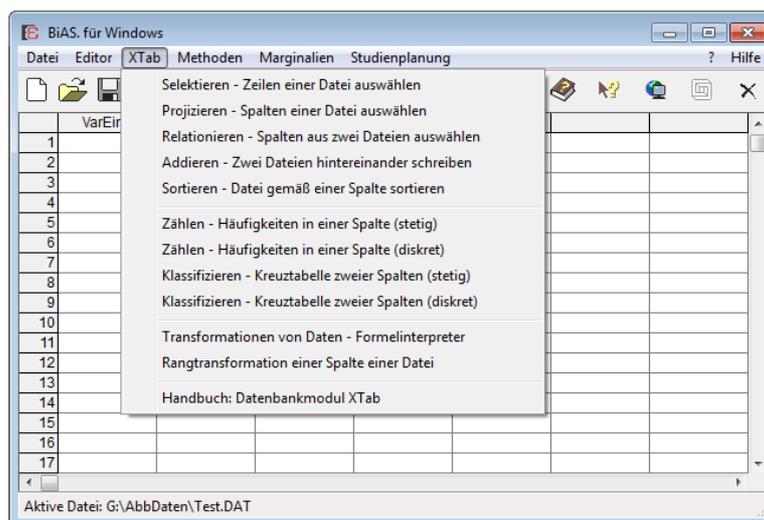


Abbildung 17: Menü des Datenbankmoduls "XTab"

Selektieren: Per frei definierbarem Selektionskriterium (vgl. auch die ausführliche Beschreibung am Ende dieses Kapitels) oder per Zeilennummern können Zeilen (die Daten von "Merkmalsträgern", zum Beispiel von Patienten) einer Datei ausgewählt und in eine neue Datei geschrieben werden. Beispiel: "Wähle alle Zeilen aus, die in Variable 3 größer sind als 100".

Projizieren: Eine "Projektion" bedeutet eine Auswahl von Variablen (Spalten bzw. "Merkmale") einer Datei, die in eine neue Datei geschrieben werden.

Relationieren: Im Gegensatz zur Projektion werden hier zwei Eingabedateien verwendet, aus denen ebenfalls Spalten ausgewählt und in eine gemeinsame Ausgabedatei geschrieben werden. **BIAS.** Relationierung setzt - im Gegensatz zu der üblichen Definition bei Relationalen Datenbanksystemen - Dateien voraus, die zeilenweise identische Merkmalsträger beinhalten. Vorsicht: Das Programm reklamiert lediglich Dateien mit unterschiedlicher Zeilenzahl, prüft die Merkmalsträger jedoch nicht auf Identität!

Addieren: Zwei Dateien werden "hintereinander" in eine neue Ausgabedatei geschrieben. Vorausgesetzt sind dabei zwei Dateien, in denen jeder Merkmalsträger in nur genau einer Datei enthalten ist und beide Dateien jeweils spaltenweise identische und identisch formatierte Variablen (Spalten) enthalten. **BIAS.** überprüft dies lediglich via Namen und Anzahl, prüft jedoch nicht, ob in den beiden Dateien Merkmalsträger mehrfach vorkommen!

Sortieren: Das Modul sortiert Dateien gemäß einer zu wählenden Variablen (das heißt, Spalte). Die in früheren Versionen vorhandene Beschränkung auf Dateigrößen von 64Kb wurde aufgehoben, dies unter Umständen mit dem Nachteil eines ungünstigeren Laufzeitverhaltens.

Zählen: Das Modul dient der Ermittlung von Klassenhäufigkeiten für eine Variable (Spalte), dies wahlweise für diskrete oder stetige Daten. Es kann ein Selektionskriterium verwendet werden, so dass bei der Auszählung gegebenenfalls nur diejenigen Zeilen berücksichtigt werden, für die das Selektionskriterium zutrifft.

Klassifizieren: Das Modul ermittelt eine bivariate Häufigkeitstafel für zwei Variablen, das heißt, es werden die Häufigkeiten der gewünschten "Felder" ausgezählt, wozu bei stetigen Variablen Klassengrenzen vorgegeben werden können. Weiterhin kann ein Selektionskriterium angewendet werden, um die Auszählung zum Beispiel auf eine bestimmte Teilstichprobe (vielleicht auf eine Gruppe wie Geschlecht o.ä.) zu beschränken. Die Ausgabe des Moduls besteht aus einer Häufigkeitstabelle.

Das Selektionskriterium wird, wie oben beschrieben, in verschiedenen Funktionen des Programms verwendet. Die folgende Beschreibung bezieht sich auf das Selektieren von Zeilen in Selektieren, die Verwendung in den Modulen Klassifizieren und Auszählen erfolgt analog.

Der Benutzer kann entscheiden, ob die Ausgabe dieser Programmeinheit nicht nur in das Ausgabefenster, sondern auch in eine neu zu generierende Ausgabe-Datei erfolgen soll. Letzteres kann in vielfältiger Weise von Nutzen sein, zum Beispiel um Teilstichproben (Strata, Subgruppen) zu "selektieren".

Das Dialogfeld des Selektionskriteriums fordert Sie zur Definition eines logischen Ausdrucks auf, der algebraische Bedingungen für eine oder mehrere Variablen verknüpft; dazu die nächste Abbildung 18:

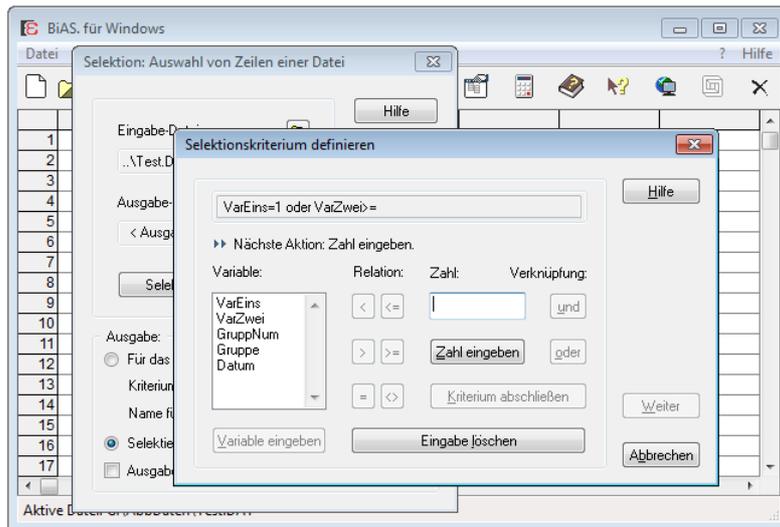


Abbildung 18: Dialogfeld zum Selektionskriterium

Die erste Eingabeaufforderung "Nächste Aktion" zur Eingabe einer ersten Variablen auf (in der Listbox auswählen und <Variable eingeben>). Anschließend erscheint die nächste Kommentarzeile "Nächste Aktion: Relation eingeben", wonach mit Hilfe der Funktionstasten unter "Relation" eine der sechs Relationen gewählt werden kann. Die nächste Kommentarzeile "Nächste Aktion: Zahlenwert eingeben" schließt die Bedingung für die erste gewählte Selektionsvariable ab. Zur Fortsetzung der Definition des Selektionskriteriums muss eine logische Verknüpfung angegeben werden, die nach der Kommentarzeile "Nächste Aktion: Verknüpfung angeben" gewählt wird. An dieser Stelle kann auch die Definition per <Ende> abgeschlossen werden. Ein Selektionskriterium könnte entsprechend Abbildung 18 beispielsweise lauten: "VarEins = 1 oder VarZwei >= 5".

Es ist zu beachten, dass die Verknüpfung *und* im Sinne einer Klammerung bei algebraischen Operationen stärker bindet als *oder*. Bei Eingabe von *und* bzw. *oder* erscheint wieder der Text "Nächste Aktion: Variable auswählen" und die Definition des Selektionskriteriums ist wie beschrieben fortzusetzen.

Beispielsweise möchte man alle Zeilen der Datei Test.Dat (steht in **BiAS.** Verzeichnis!) ermitteln, die in Gruppe 1 enthalten sind und für die die erste Variable größer als 2, aber kleiner als 5 ist. Dazu definiert das Selektionskriterium mit "GruppNum=1 und VarEins>2 und VarEins<5". **BiAS.** schreibt die vierte Zeile der Datei Test.Dat in eine Ausgabe-Datei (z.B. Test.Sel) und in das Ausgabefenster; die Variablennamen aus Test.Dat werden in die Ausgabe-Datei Test.Sel übernommen. Die Ausgabe wird durch den Text "Die Bedingung wird von 1 Zeile erfüllt. Wegen fehlender Werte wurde 1 Zeile nicht berücksichtigt" abgeschlossen: Nur die zweite Zeile (2.0 3.0 1) der Datei Test.Dat erfüllt das Kriterium, der letzte Kommentar ist durch den fehlenden Wert der ersten Variablen der ersten Zeile bedingt.

Das Selektionskriterium spielt - optional - zum Beispiel auch im Modul *Klassifizieren* eine Rolle, da man sich unter Umständen auf das Auszählen von Teilkollektiven beschränken möchte. Die Eingabe von Selektionskriterien ist im Modul *Klassifizieren* die gleiche wie oben beschrieben:

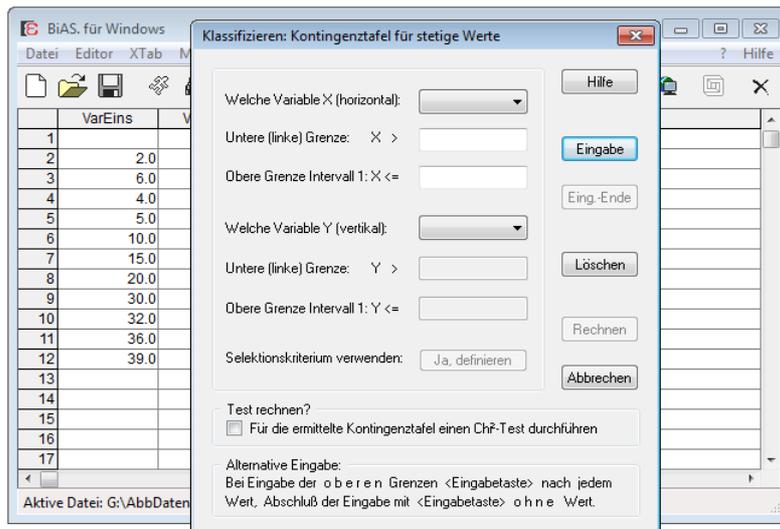


Abbildung 19: Dialogfeld zur “Kreuzklassifikation”

Man wählt in den Auswahlfeldern (Combo-Boxen) in offensichtlicher Weise die beiden Variablen X und Y (als “horizontale” und “vertikale” Variable der Kreuzklassifikation) und definiert für beide Variablen Intervallgrenzen. Nach Eingabe der jeweils unteren Grenze gibt man alle relevanten oberen Intervallgrenzen ein: Abschluss dieser Eingabe ist die <Eingabetaste> *ohne* Werteingabe! Wie erwähnt, kann das optionale Selektionskriterium das Auszählen auf Schichten bzw. Subgruppen beschränken.

Viele Anwender benutzen das Tabellenkalkulationsprogramm Excel von Microsoft zur Eingabe und Verwaltung ihrer Daten. Excel ist dafür sicher hervorragend geeignet und gestattet eine außerordentlich flexible Behandlung der Datenstrukturen, zum Beispiel können auch mehrere Dateien “nebeneinander” bearbeitet werden. Abbildung 21 zeigt die Test-Datei Test.DAT im Excel-Fenster.

Vor Start eines Projektes sollten Sie sich über das für Sie günstigste Vorgehen einigen und klären, ob Sie Ihre Daten in **BiAS**. mit Hilfe des Editors etc. eingeben oder ob Sie den flexiblen Excel-Editor verwenden wollen. Im zweiten Fall können Sie Ihre Daten per Zwischenablage in **BiAS**. importieren, wie im nächsten Kapitel ausführlich dargestellt wird.

Unabhängig von Ihrer Entscheidung sollten Sie Ihre Daten stets in nur *genau einer* Datei verwalten: Ein “paralleles” Ergänzen und Modifizieren in einer **BiAS**.-Datei einerseits und in einer Excel-Datei andererseits ist nicht nur mühsam, sondern führt früher oder später mit Sicherheit zu Problemen!

Bitte beachten Sie, dass Excel zwar als Tabellenkalkulationsprogramm hervorragende Dienste leistet, aber kein Statistik-Programm darstellt. Die wenigen, zum Teil statistisch unbefriedigend implementierten Methoden sind für wissenschaftliche Auswertungen in der Regel nicht ausreichend, wenn man sich nicht nur auf triviale Methoden beschränken möchte. Einige Ansichten dazu finden sich im Internet, entsprechende Urls finden Sie auf **BiAS**.’ Homepage <https://www.bias-online.de>.



3. Import, Export und Excel

Fast alle unter Windows lauffähigen Programme verwenden interne Datenformate, die in aller Regel anderen Programmen nicht zugänglich sind. Zum Datenaustausch besteht aber praktisch immer die Möglichkeit des Exports und des Imports von Text-Dateien (*.TXT und *.CSV), sehr viele Programme können auch dBaseIII-DBF- oder Standard-Dateien (*.SDF) lesen und schreiben oder sind wenigstens über die Zwischenablage zugänglich.

BiAS. gestattet den Import- und Export von Excel-XLS-, Windows' CSV- und TXT-Dateien, dBaseIII-Dateien (DBF), "Plain-Text-Dateien" (SDF, formatiert!), SPSS und **BiAS.**-Dateien früherer DOS-Versionen, außerdem ist ein Transfer via Zwischenablage möglich. Über die vorhandenen Import-Varianten (Export ist analog) gibt die nächste Abbildung Aufschluss. Bitte beachten Sie, dass der Import und Export nicht vom Editor, sondern immer über das Menü *Datei:Import/Export* vorgenommen wird.

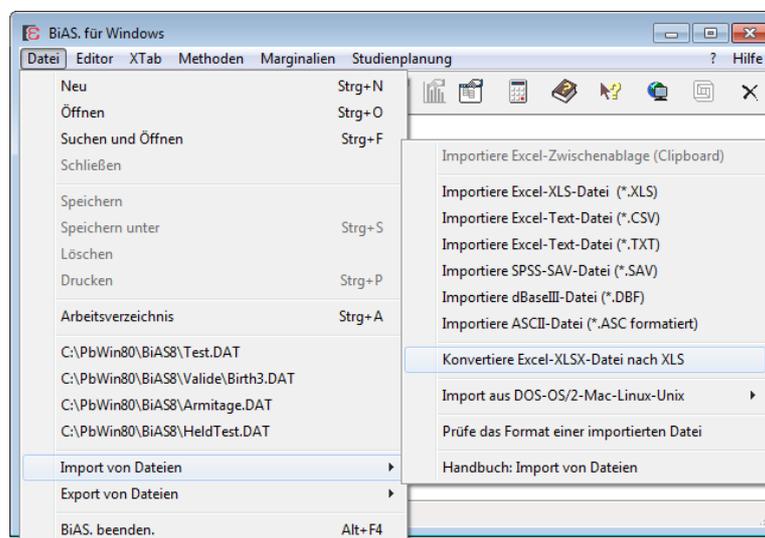


Abbildung 20: Untermenü "Import" im Pulldown-Menü "Datei"

Grundsätzlich muss man nur den Namen incl. Pfad der zu importierenden bzw. zu exportierenden Datei und wie in Abb. 8 den Namen incl. Pfad der Zieldatei angeben.

Vorsicht: Die Extensionen TXT, CSV etc. sind üblich, aber nicht verbindlich und werden von manchen Autoren auch vertauscht (TXT ist konventionell TAB-delimited, CSV ist (semi)colon-delimited!). Bei Transferproblemen hilft sehr oft eine einfache Anpassung an die Namenskonvention des Zielprogramms.

Vorsicht: Der Import und Export erfolgt *nicht* über **BiAS.**' Editor, sondern *stets* über die Funktionen der oben in Abbildung 20 dargestellten Module "Import" und "Export"!

Vorsicht: Natürlich muss man auch beim Import **BiAS.**' Datenformat beachten!

Der Import aus Excel ist auch für Excel-Dateien mit mehreren Tabellen möglich, da gegebenenfalls im Programm über ein Dialogfeld eine der Tabellen zum Import ausgewählt werden kann. Bitte beachten Sie in Excel unbedingt, dass die Namen der

Spalten nur in der *ersten* Zeile der Datei stehen dürfen und die Zellen unterhalb der ersten Zeile (“Datenmatrix”, pro Merkmalsträger/Patient *eine* Zeile!) formatiert sein müssen: Dazu werden in Excel die gewünschten Zellen markiert und (vgl. Excel-Menü!) via “Format-Zellen” mit dem korrekten Format versehen. Die Einstellung “Standardformat” führt unter Umständen zu Fehlern!

Eine unproblematische Alternative ist der Import über die Zwischenablage: **BiAS.** verarbeitet Excel-XLS-Dateien - wie oben beschrieben - direkt, über den Import und Export anderer Standard-Dateiformate wie zum Beispiel CSV/Excel oder DBF/dBaselll oder ganz einfach über Windows’ Zwischenablage (“Clipboard”). Natürlich muss man auch beim Import über die Zwischenablage **BiAS.**’ Datenformat beachten.

Zum Clipboard-Import von Daten markiert man zunächst in Excel die relevanten Daten und kopiert diese mit <Strg>-C in die Zwischenablage (nächste Abbildung). Bitte markieren Sie in Excel nach Möglichkeit auch die Spaltenbezeichnungen (Variablenamen), um diese für **BiAS.** zur Verfügung zu stellen:

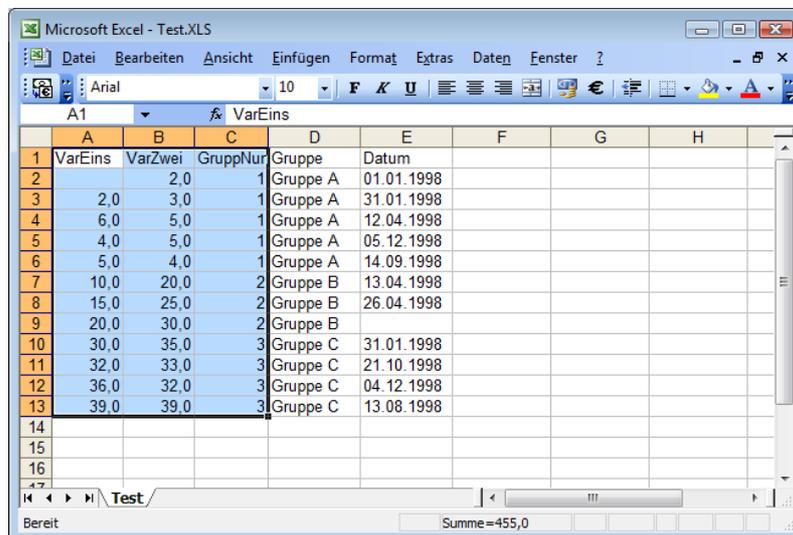


Abbildung 21: Datei Test.Dat in Excel

Bitte beachten Sie, dass Spalten-Namen nur in der *ersten* Zeile der Datei stehen dürfen und die Zellen unterhalb der ersten Zeile (also die “Datenmatrix”) formatiert sein sollten! Nach Wechsel zu **BiAS.** betätigt man in der Toolbar das rechte Excel-Icon (alternativ <Strg>-V) für den Import der in Excel beschriebenen Zwischenablage:

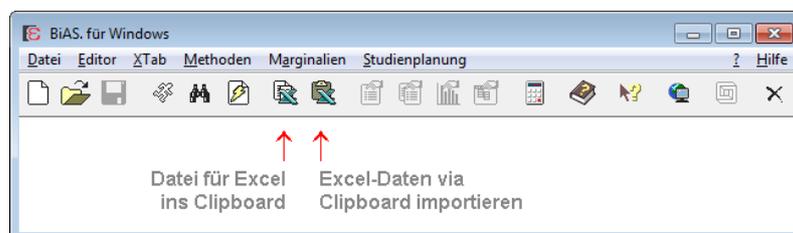


Abbildung 22: BiAS.’ Toolbar-Funktionen “Excel-Import und -Export”

Beim Daten-Import aus Excel fragt **BiAS.** - analog zur Schnelleingabe in Kapitel 2, Abbildung 15 - ab, ob die in Excel markierten Daten in der *ersten* Zeile (wie in Abbildung 21!) Namen resp. Bezeichnungen der Spalten (d.h. "Variablenamen") enthalten: Im Beispiel von Abbildung 21 ist dies mit "Namen" zu beantworten:

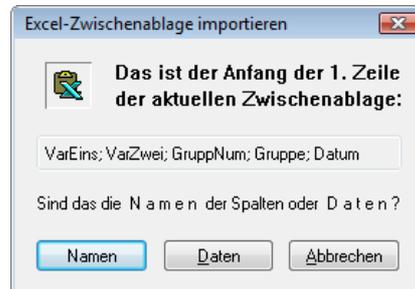


Abbildung 23: Excel-Zwischenablage importieren

Die danach importierte Datei steht jetzt für **BiAS.** zur Verfügung. Optional können die importierten Daten als **BiAS.-DAT**-Datei abgespeichert (<Speichern> nach Wahl eines Datei-Namens) oder aber temporär als Clipboard-Datei (<Abbruch>) verwendet werden.

Die importierte Excel-Datei (als DAT-Datei oder in der Zwischenablage) ist automatisch die "Aktive Datei" und wird somit beim nächsten Aufruf zum Beispiel eines statistischen Tests verwendet.

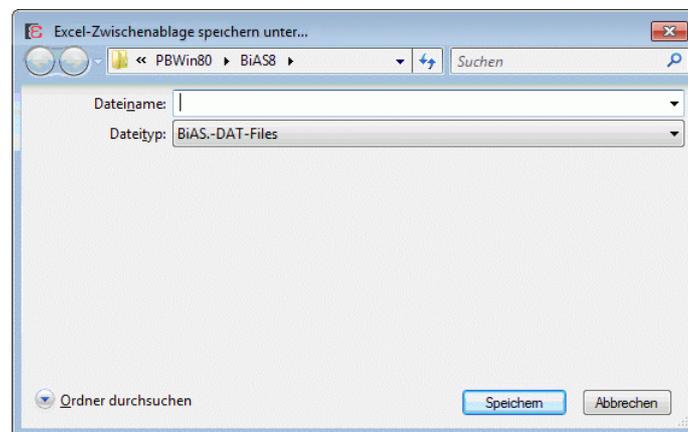


Abbildung 24: Abfrage eines Dateinamens zur optionalen Speicherung

Der *Export* für Excel erfolgt analog, vergleichen Sie bitte das oben abgebildete Icon in Abbildung 22 bzw. die Toolbar-Funktion "Datei für Excel ins Clipboard". Nach Betätigen des *linken* Icons (Export!) wechselt man zu Excel und kann dort die Daten via <Strg>-V in Excel einlesen.

Zum Nachvollziehen des Imports und Exports stehen die verwendeten Daten als Dateien in **BiAS.'** Verzeichnis zur Verfügung: Dies sind die beiden Dateien Test.XLS und Test.DAT.



4. Aufruf einer Funktion

Die Datei "Test.DAT" (Abbildung 11, steht auch in **BiAS.**' Verzeichnis) beinhaltet Daten, die hier zur exemplarischen Demonstration eines Aufrufs einer Funktion - im aufgeführten Beispiel des Wilcoxon-Mann-Whitney-U-Tests - benutzt werden.

Im Hauptmenü (obere Menüleiste) wählt man *Methoden*, im anschließend erscheinenden Menü *Nicht-parametrische Verfahren* und zuletzt endlich den *Wilcoxon-Mann-Whitney-U-Test*. Der Bildschirm zeigt vor dem Maus-Klick bzw. vor <Eingabetaste> das Bild der nächsten Abbildung 25:

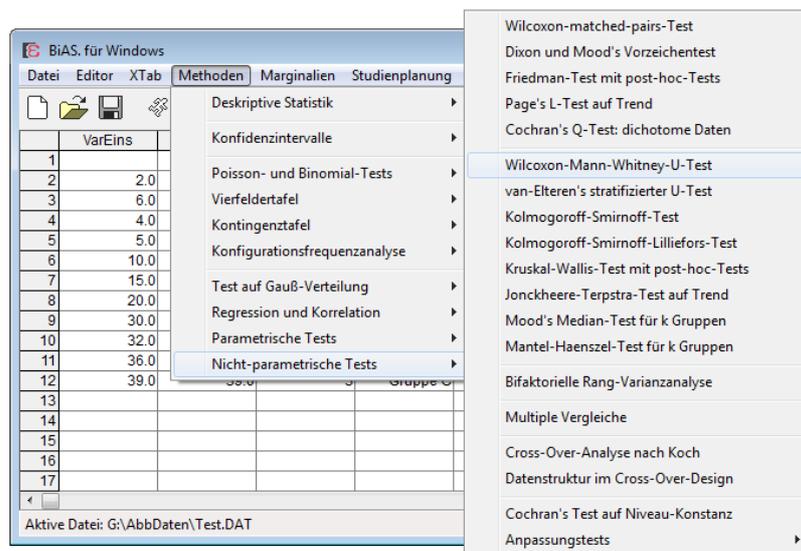


Abbildung 25: Menü-Auswahl des Wilcoxon-Mann-Whitney-Tests

Die im Beispiel dieses Kapitels verwendete Test-Datei umfasst 12 Zeilen ("Patienten") bzw. 3 Spalten ("Merkmale"). Das erste und das zweite Merkmal erhielten bereits bei der Dateneingabe die klartextlichen Namen "VarEins" und "VarZwei". Die dritte Variable=Spalte erhielt den Namen "GruppNum" und kennzeichnet die Gruppenzugehörigkeit der Zeilen bzw. der Merkmalsträger; GruppNum nimmt im Beispiel die Werte 1, 2 und 3 an (drei Gruppen). Diese Gruppenzugehörigkeit kann beispielsweise durch verschiedene Behandlungen, Diagnosen oder auch durch andere Kriterien 1, 2 und 3 definiert sein.

Die erste Eingabe ist die Gruppierungsvariable (die hier obligat angegeben werden muss, denn es handelt sich bei dem gewählten U-Test um den Vergleich zweier Gruppen!); diese wird im Beispiel, wie Abbildung 26 zeigt, im ersten Auswahlfeld (Combo-Box, in der Abbildung nicht "aufgeklappt") mit "GruppNum" definiert.

Die Frage des Programms "Welche Variable testen?" beantwortet man mit dem Namen der Zielgröße, auszuwählen in der zweiten Combo-Box (in der Abbildung 26 ebenfalls nicht "aufgeklappt") mit z.B. "VarEins", also der ersten Spalte der Eingabedatei. (Betrachtet man drei Medikationen zur Behandlung der Hypertonie, so könnte die Variable "VarEins" die Blutdrucksenkung nach Medikation darstellen.)

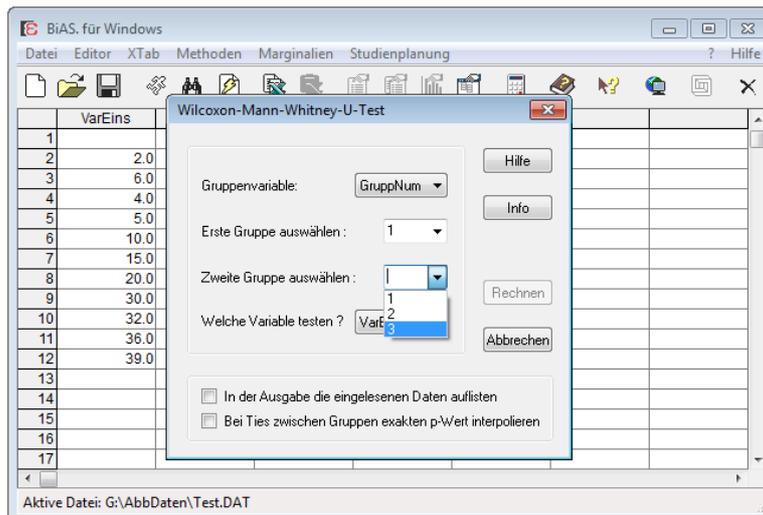


Abbildung 26: Dialogfeld zum Wilcoxon-Mann-Whitney-Test

Zum Abschluss gibt man in dem Dialogfeld die beiden testrelevanten Gruppen (Behandlungen, im Beispiel: Gruppennummern "1" und "3") bekannt, die man gemäß der Nullhypothese vergleichen möchte. Diese Gruppennummern sind im Beispiel Werte der Gruppierungsvariablen "GruppNum" der 3. Spalte (im Hypertoniebeispiel z.B. die Medikationen "1" und "3"; Vorsicht: Gruppennummern sind *immer* >0!).

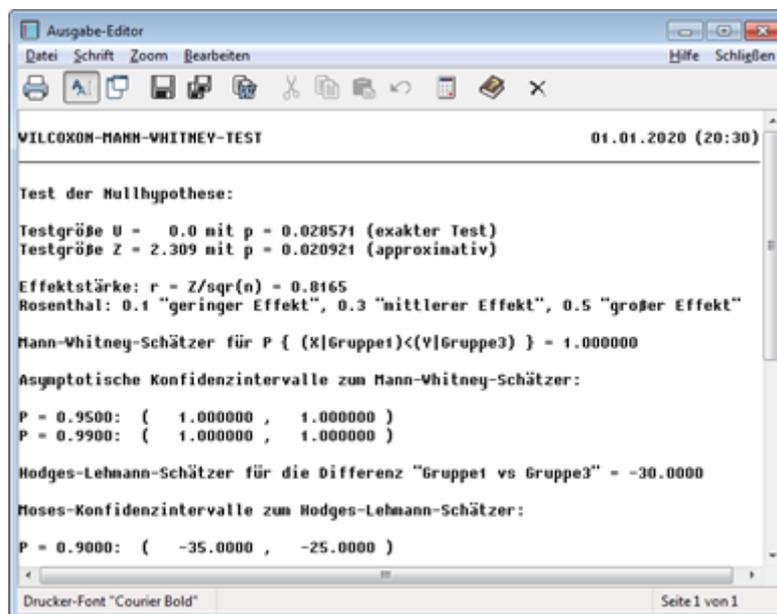


Abbildung 27: Ergebnis der Rechnung im Ausgabe-Editor

Die Ausgabe - als Ergebnis des Wilcoxon-Mann-Whitney-U-Tests - erfolgt in ein Ausgabefenster, das in Abbildung 27 wiedergegeben wird. Im Ausgabefenster kann beliebig "geblättert" werden, bei Bedarf kann das Ergebnis der Berechnung auf den Drucker ausgegeben werden oder aber zwecks Dokumentation oder späterer Verwendung gespeichert werden (Disketten-Icons, kumulativ oder pro Test).

Eine biometrisch-statistische Bewertung des Ergebnisses wird hier nicht vorgenommen. Bitte bedenken Sie, dass eine inhaltliche Analyse von Programmresultaten nur mit Vorkenntnissen bzw. nach Lektüre einschlägiger statistischer Lehrbücher möglich ist; Hinweise finden sich reichlich in den Literaturangaben. Am Ende jeder Ausgabe gibt **BiAS**. Literaturhinweise, die eine Lektüre oder ein Zitieren der verwendeten Methoden erleichtern.

Nach dem ersten "Ausprobieren" des Programms anhand des Wilcoxon-Mann-Whitney-U-Tests ist Ihnen **BiAS**. Bedienungssystematik bereits (fast) vertraut und es sollte Ihnen - wie dem Autor zahlreiche Anwender des Programms bestätigt haben! - möglich sein, auch andere Testverfahren oder auch deskriptive Methoden des Programms erfolgreich zu benutzen. Als Hilfestellung für weitere Versuche dient Ihnen Kapitel 7, das bis WinXP nur im Programm als "Validierung" verfügbar ist:



Abbildung 28: Hilfethemen in der Validierung (Programmaufruf unter WinXP)

Unter XP gelangt man über das Buch-Icon (zum Beispiel Abbildung 5) zum Validierungsteil (dieser entspricht Kapitel 7 dieses eBooks!), dessen "Hilfethemen" ebenfalls einen Überblick über die implementierten Methoden geben. (Bei Bedarf bitte die Bemerkungen in der Einleitung zur Verfügbarkeit der Hypertexte beachten!!)

BiAS. Methodenspektrum beinhaltet neben den "üblichen" statistischen Methoden auch zahlreiche weitere Verfahren, die im medizinischen Bereich zwar von grundlegender Bedeutung, aber in anderen Programmen nicht oder nur selten verfügbar sind. Einen Überblick über den Leistungsumfang gibt die letzte Seite, **BiAS**. Homepage <https://www.bias-online.de> oder in vollem Umfang die Demoversion des Programms, die unter der genannten Adresse verfügbar ist. Insgesamt stehen mehr als 400 implementierte Verfahren zur Verfügung.

BiAS. entstand im medizinischen Bereich und ist, wie das Methodenspektrum zeigt, somit besonders für Mediziner und Pharmakologen, aber auch für Biologen und Psychologen im Studium und in der wissenschaftlichen Praxis von Interesse: Das Anwenderspektrum reicht von Studenten, Diplomanden und Doktoranden bis hin zu professionellen universitären und industriellen Wissenschaftlern.



5. Graphische Ausgabe

BiAS. sieht zahlreiche graphische Ausgaben vor; einen Überblick dazu gibt die letzte Seite. Als Beispiel zeigt Abbildung 30 eine graphische Darstellung einer Kaplan-Meier-Überlebenszeitschätzung zusammen mit Konfidenzintervallen nach Hall und Wellner (Gruppe 2 der Datei "Armitage.Dat", steht in **BiAS.**' Verzeichnis). Die erforderlichen Eingaben werden im Dialogfeld der Abbildung 29 dargestellt.

BiAS. enthält mehr als 60 verschiedene graphische Darstellungsmöglichkeiten, die alle (Kreisdiagramme ausgeschlossen) in analoger Weise behandelt werden können. Nach Speicherung einer Graphik kann diese im Hauptmenü wieder eingelesen werden (Tool-Icon "Letzte Graphik"), um via MS-Paint weitere Modifikationen (z.B. Änderungen der Beschriftung, zusätzliche Beschriftungen o.ä.) vornehmen zu können. Vorsicht: MS-Paint liest je nach Version nur BMP-Graphiken ("Bitmaps").

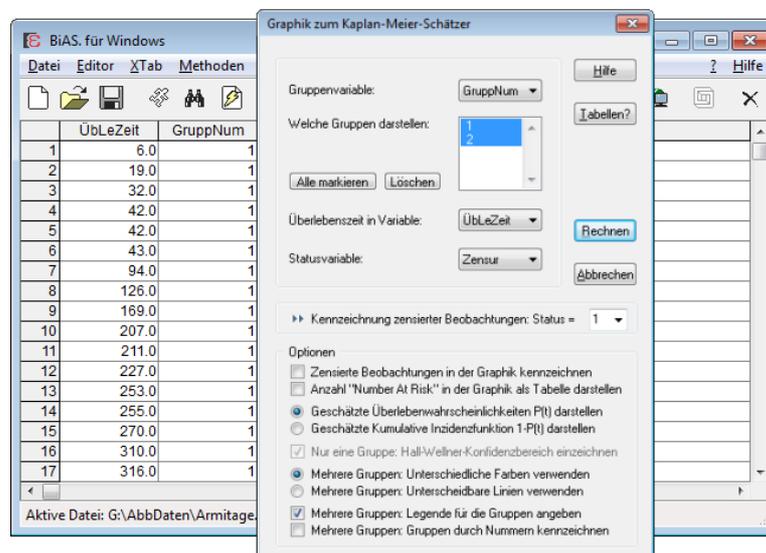


Abbildung 29: Dialogfeld zum Kaplan-Meier-Schätzer

Die Funktion wird unter *Marginalien/Survivalanalyse* aufgerufen und man erhält im Beispiel nach Wahl der Gruppe 2 die Graphik in Abbildung 30. Diese Graphik kann nach Abfrage in eine BMP, GIF, JPG oder PCX-Graphik-Datei geschrieben werden, die wiederum in ein beliebiges Textprogramm wie zum Beispiel MS-Word eingelesen werden kann. Auf diesem Wege können **BiAS.**' Graphiken auch an anderer Stelle weiterverarbeitet werden, z. B. in Microsoft's Paintbrush oder in PowerPoint.

Das Menü des Graphik-Editors (nächste Abbildung 30) sieht einige Graphik-Optionen vor: *Bearbeiten* beinhaltet zahlreiche Graphikeinstellungen (dazu Abbildung 31) und ist auch per Rechtsklick aufrufbar, an Stelle von *Drucken* und *Kopieren* (ins Clipboard) kann man auch kürzer <Strg>-P und <Strg>-C verwenden. *Speichern* speichert, wie erwähnt, die Graphik alternativ in eine Datei. Mit *Kopieren* wird die Graphik - zum Beispiel für MS-Word - in die Zwischenablage kopiert. Bitte beachten Sie, dass BMP-Color-Dateien ca. 600Kb umfassen, die anderen Formate - speziell GIF-Dateien - dagegen deutlich weniger. Die Option *Größe* gestattet eine pixelgenaue Modifikation der Graphikgröße, *Zoom* verändert die Größe in nur zwei Stufen.

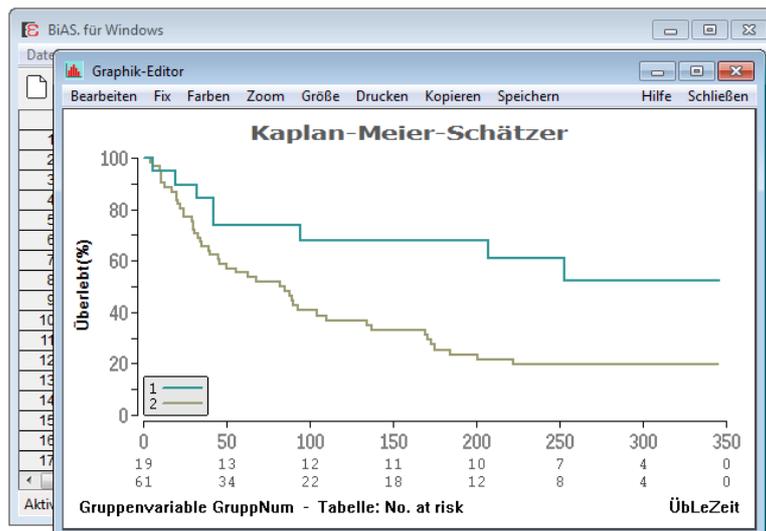


Abbildung 30: Überlebenszeitschätzung nach Kaplan und Meier

Im Kapitel “Struktur” wurde bereits auf die automatische, aber modifizierbare Skalierung und Beschriftung der Graphiken hingewiesen. Weitere Optionen sind vorgesehen (z.B. Ein- und Ausblenden des Koordinatengitters per *Netzgitter*, Modifizierung der Skalierung per *Skalierung*), wie das Dialogfeld des Graphik-Editors (aufgerufen via *Bearbeiten* bzw. per Rechtsklick) in Abbildung 31 zeigt:

Wichtig: Alle Modifikationen der Voreinstellung können *dauerhaft* (also auch für Neustarts!) im Menü *Fix* des Graphik-Menüs (2. Menüpunkt von links) festgelegt werden, um eine individuell fixierte Gestaltung der Graphiken zu ermöglichen.

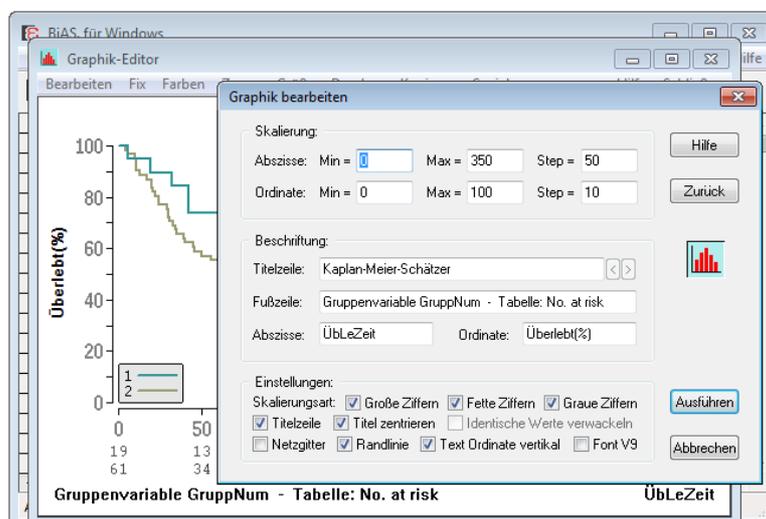


Abbildung 31: Dialogfeld “Graphik-Einstellungen”

Bei *nominaler* Abszisse (also bei Gruppen, Box-Plots, Zeitverläufen etc.) wird diese *textlich* skaliert, so dass das Dialogfeld zwecks beliebiger Veränderung der Beschriftung in modifizierter Darstellung angezeigt wird: Weitere Einzelheiten dazu finden sich unter *Hilfe* im Menü.

Die im Programm verwendeten „Standardfarben“ sind nach Installation vorgegeben und aus Abbildung 33 ersichtlich. Die Farben sind auf zweierlei Weise modifizierbar:

Nach Doppelklick auf z.B. die Farbe der jeweils ersten Säule erscheint der bekannte Windows-Dialog zur Farbwahl, im Beispiel der Abbildung 32 mit der Farbe „grün“. Nach Anpassung und „Ok“ wird die Farbe in der Graphik geändert. Jede Farbwahl bleibt während der Bearbeitung einer Graphik erhalten, bei einem erneuten Aufruf oder bei Aufruf einer anderen Graphik wird wieder die Standardpalette verwendet. Alle neuen Farben können aber im Menü per „Fix“ *permanent* gespeichert werden.

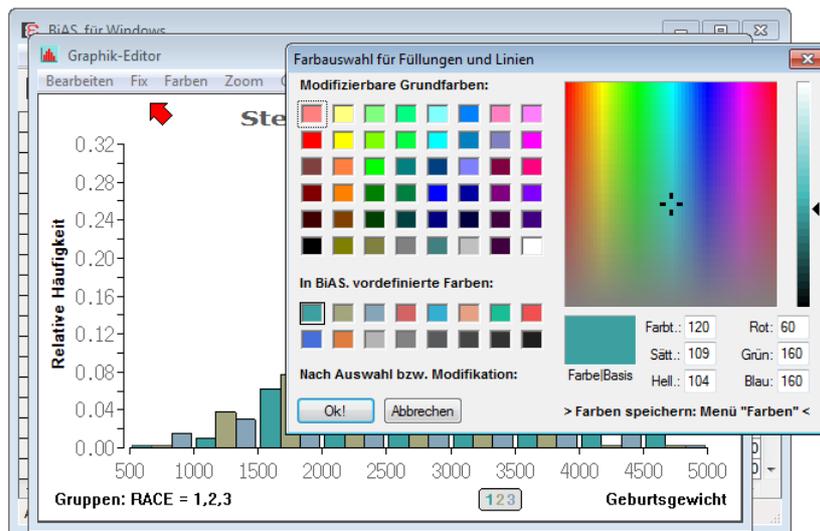


Abbildung 32: Dialogfeld „Farbauswahl“ nach Doppelklick auf eine Farbe

Eine zweite Variante der Farbwahl ist über das Graphikmenü per „Farben“ verfügbar. Der entsprechende Dialog findet sich in der nächsten Abbildung 33:

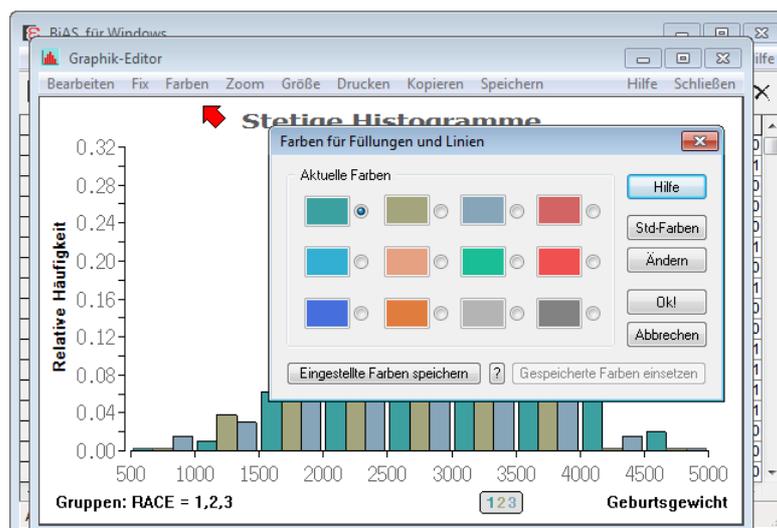


Abbildung 33: Dialogfeld „Farbauswahl“ nach Wahl von „Farben“ im Menü

Per „Fix“ oder auch im Dialog kann die modifizierte Farbpalette wie oben *permanent* – und wie oben auch für erneute Programmaufrufe! – fixiert und gespeichert werden.

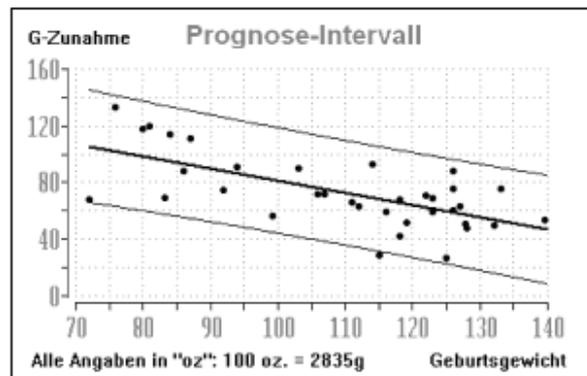
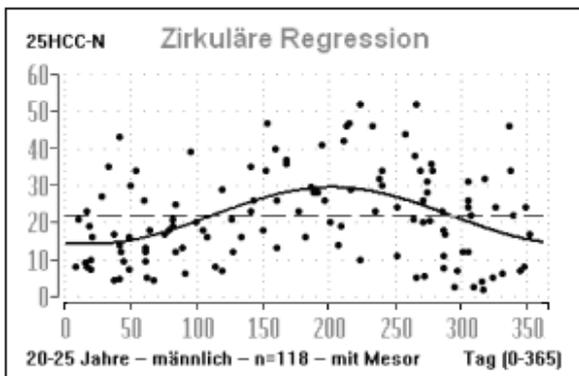
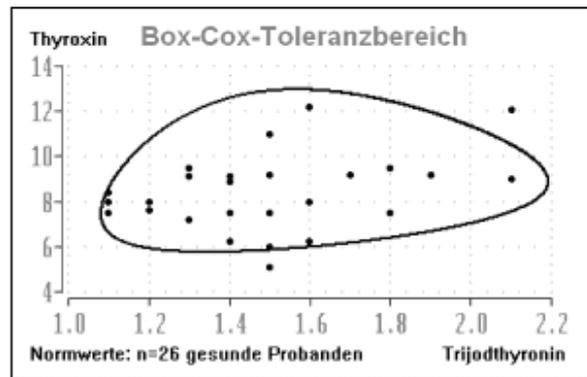
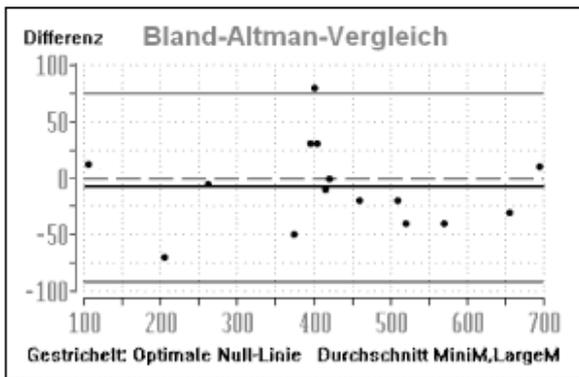
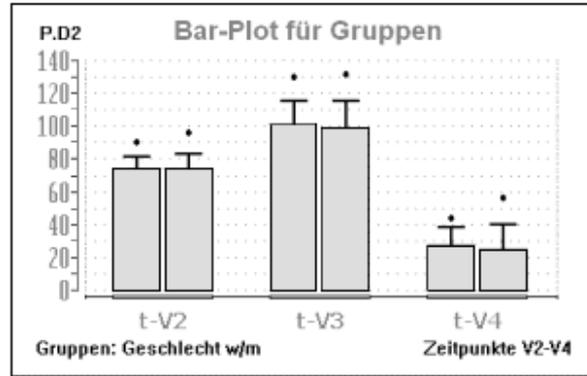
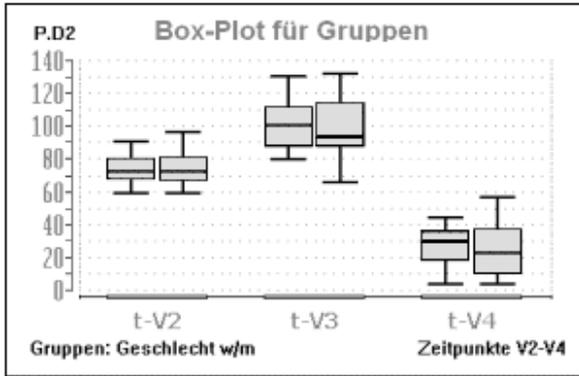
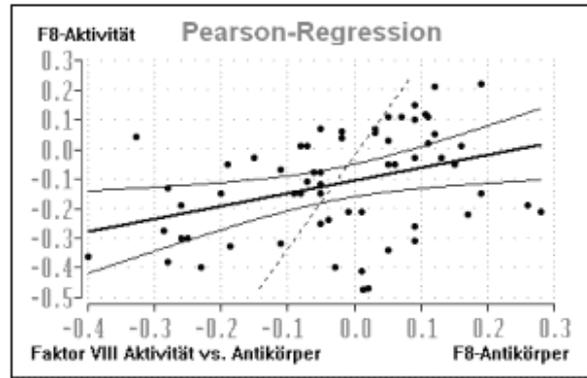
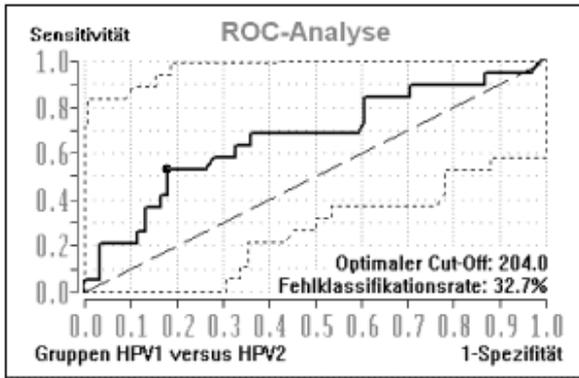


Abbildung 34: Beispiele einiger Graphiken in optionaler Grauwert-Darstellung

Im Anhang finden sich im Zusammenhang mit speziellen Verfahren (zum Beispiel Lin's CC und viele andere) circa 60 weitere, hier nicht aufgeführte Graphiken. ▲

6. Installation

BiAS.' Installation erfolgt durch Starten entweder der Setup-Datei DemoBiAS.EXE oder der Datei BiAS11.EXE, die Sie von **BiAS.**' Homepage *bias-online.de* heruntergeladen haben (bei Erhalt einer Original-CD: *Setup.exe* starten!). Prinzipiell ist nur der Pfad des Zielverzeichnisses anzugeben, in dem **BiAS.** installiert werden soll:



Abbildung 35: Installation des Programms

Abbildung 36 zeigt die drei in **BiAS.**' verfügbaren Varianten zum Festlegen des Verzeichnisses, das jeweils nach Programmstart und bei erneutem „Datei öffnen“ als Standardverzeichnis verwendet werden soll. Voreinstellung nach Installation ist das Verzeichnis (vgl. Abb. 36!), das im jeweils vorgängigen Öffnen-Dialog gewählt wurde:

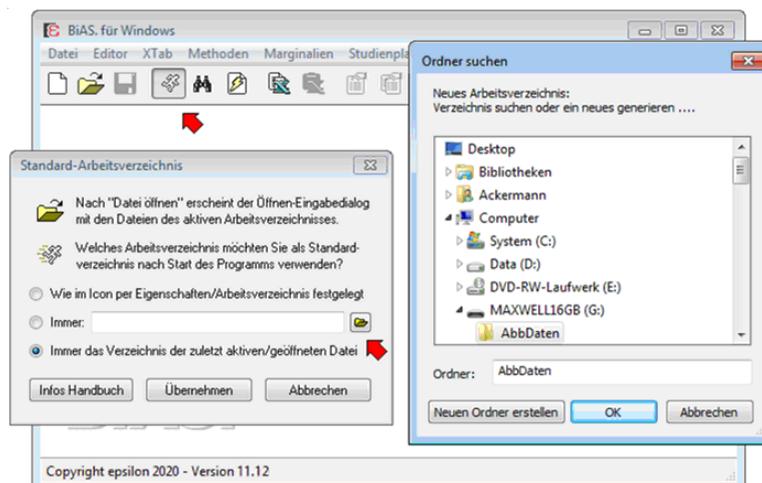


Abbildung 36: Auswahl des Standard-Startverzeichnisses bei „Datei öffnen“

Handbuch: Bis WinXP werden **BiAS.**' „WinHelp-Handbücher“ während des Setups installiert und sind im Programm aufrufbar. Ab Vista wird das vorliegende eBook installiert, das ebenfalls in **BiAS.**' Toolbar per „Buch-Icon“ verfügbar ist. Die XP-Handbücher können bei vorhandenem *WinHlp32* optional zusätzlich verwendet werden: Einzelheiten ergeben sich aus Abbildung 1 und dem Kommentar dazu auf Seite 5.

7. Validierung

BiAS. wurde nach bestem Wissen und Gewissen geplant, entworfen, programmiert und ausführlich getestet. Zur Validierung der Software wurden alle Programmteile, soweit in anderen Programmen verfügbar, anhand der Ergebnisse der vorhandenen Standardsoftware überprüft. In allen anderen Fällen wurden durchgerechnete Beispiele aus der Literatur mit den von **BiAS.** errechneten Ergebnissen verglichen oder mit "Bleistift und Papier" errechnete Ergebnisse herangezogen. Erfahrungsgemäß lassen sich trotzdem Fehler nur selten vermeiden. Der Autor übernimmt deshalb weder eine Garantie für die Korrektheit des Programms **BiAS.** und des Handbuchs noch eine Haftung für mögliche "Folgeschäden", ist aber für Hinweise und Verbesserungsvorschläge stets dankbar.



Abbildung 37: Validierungsteil des Handbuchs: XP-Hypertext-Version

In der vorliegenden Zusammenstellung der Validierung (siehe Anhang, unter XP auch als "Hypertext-Handbuch" gemäß Abbildung 37 verfügbar) finden sich fast ausschließlich Beispiele und Ergebnisse aus bekannten und allgemein verfügbaren Lehrbüchern:

Mit der "Validierung" des Programms wurde versucht, alle Rechenergebnisse überprüfbar und nachvollziehbar zu machen, so dass hier weitgehend auf Lehrbuchbeispiele hingewiesen wird; wenige Ausnahmen beziehen sich auf Beispiele aus Originalarbeiten und auf Ergebnisse der Standardprogramme SPSS, SAS, Testimate und SYSTAT. Dabei ist leider als Nachteil in Kauf zu nehmen, dass die durchgerechneten Beispiele der Lehrbücher häufig nur mit geringer Rechengenauigkeit bzw. mit Rundungsfehlern angegeben werden, so dass mitunter Abweichungen bereits in der 3. oder 4. Stelle der Ergebnisse vorhanden sind. Auf solche Abweichungen wird im Folgenden hingewiesen, auffällige Abweichungen von den Ergebnissen der Lehrbuchbeispiele wurden nach Möglichkeit durch Per-Hand-Berechnungen überprüft und können in allen Fällen erklärt werden.

Mit **BiAS.**' Validierung und den Beispielen zur Bedienung erübrigt sich nach Ansicht des Autors eine Aufstellung der verwendeten Formeln: Einerseits besteht nicht die

Absicht, mit **BiAS.**' Handbuch einschlägige Lehrbücher oder eine einschlägige Vorbildung zu ersetzen, andererseits kann das erforderliche biometrische Basiswissen bzw. die zugrunde gelegten Formeln etc. ohne weiteres der zitierten [Literatur](#) entnommen werden.

Die durchgerechneten Beispiele des Validierungsteils sind durch die Verwendung von Hyperlinks als "Hypertext-Handbuch" leichter zu lesen und einfacher zu finden als in einem gedruckten Handbuch: Zum Aufruf per Icon vergleichen Sie bitte die Abbildung 36 im vorhergehenden [Kapitel 6](#). Der Text des Hypertext-Handbuchs entspricht vollständig dem Text des Anhangs. Falls Sie auch ab Windows Vista das Hypertext-Handbuch verwenden möchten, installieren Sie – wie auf Seite 5 ausführlich beschrieben – nachträglich die Datei *WinHlp32.exe*. Das Programm erkennt die Installation und stellt zusätzlich die vor Vista verwendete Hypertext-Version des Handbuchs bereit!

Für alle in **BiAS.** implementierten Verfahren findet sich jeweils eine vollständige Ein- und Ausgabe des Programms. Die Eingabedaten zu den gerechneten Beispielen werden ebenfalls angegeben; alle Eingabedateien wurden mit den Autorennamen bezeichnet, die Extension einer Datei entspricht der Seitennummer des zitierten Lehrbuches bzw. der Originalarbeit (vgl. Literaturangaben).

Die "Beispiele" dienen einerseits der Validierung des Programms, andererseits kann man aber auch in Zweifelsfällen an den Beispielen die Bedienung des Programms einfach nachvollziehen. Aus diesem Grund wurde in den ersten Kapiteln eine nur exemplarische Einweisung in die Bedienung vorgenommen und auf detailliertere Darstellung verzichtet.

Die Reihenfolge der Validierungsbeispiele ist die gleiche wie die der Literaturzitate im letzten Kapitel. Zu jedem dieser Beispiele findet sich eine kurze Erläuterung und, etwas eingerückt, die vollständige Ausgabe des Programms einschließlich Literaturhinweisen.

In [Kapitel 1](#) ("Struktur des Programms") wurde **BiAS.**' Hilfesystem beschrieben (vergleichen Sie dazu auch bitte die Abbildungen 5 und 6). In **BiAS.**' Hilfemodus werden die angewählten Funktionen nicht ausgeführt, sondern es werden zu allen Menüpunkten einfache medizinische und biologische Beispiele in einem Fenster gezeigt. Dieser biometrisch orientierte Teil des Hilfesystems kann somit als eine "schnelle Ergänzung" zu den Validierungsbeispielen dieses Kapitels verwendet werden, soll aber keinesfalls die Lektüre eines Lehrbuches ersetzen.

Alle graphischen Methoden wurden im Anhang exemplarisch reproduziert; aus Platzgründen können jedoch nicht alle Optionen dargestellt werden. In Kapitel 4 findet sich eine Graphik aus dem Bereich der Survival-Analyse; diese kann mit Hilfe der Datei "Armitage.Dat" (steht in **BiAS.**' Verzeichnis!) nachvollzogen werden. **BiAS.** sieht mehr als 80 verschiedene Graphiken vor, die sachlich den jeweiligen Moduln des Menüs zugeordnet sind. Zum Verständnis der Bedienung empfiehlt es sich, das Beispiel der Überlebenskurve nachzuvollziehen ([5. Kapitel](#), Abbildungen 29-31!). Alle anderen verfügbaren Graphiken, auch die in Abbildung 34, folgen im Wesentlichen diesem Beispiel.

Eine explizite Validierung der Graphiken wurde somit *nicht* vorgenommen und erscheint einerseits nicht sinnvoll und andererseits auch nicht möglich.

Im Folgenden finden Sie alle Details zur Validierung der Rechenergebnisse des Programms. Zu jedem Test bzw. zu jeder Methode wird jeweils die entsprechende Quelle zur Überprüfung der Resultate angegeben. Falls eine entsprechende Referenz nicht existierte, wurden ausführliche per-Hand-Berechnungen durchgeführt. Eine Überprüfung der Graphiken ergab sich vollständig durch ausführliche per-Hand-Berechnungen und, sofern möglich, mit anschließenden Vergleichen mit anderen Programmpaketen. Details werden in der weiteren Aufstellung nicht aufgeführt.

Deskriptive Statistik

Schätzgrößen mit den Daten Graf et al. (1987, p.379): Das Beispiel wurde gerechnet mit den Faktoren 1, 10000 und 0.0001, bis auf Darstellungsgenauigkeit mit identischen Ergebnissen. Weitere Überprüfung durch Addition einer Konstanten ($c=1000$) und mit großen Stichproben ($n \leq 1000$) unterschiedlicher Größenordnung. • Box-Plot, Bar-Plot, Quartilen, Perzentilen, Histogramm, Scattergramm: Daten aus Graf et al. (1987, p.379, interpolierte Werte der Quartilen wurden per Hand gerechnet: $P(0.25)=1.4275$, $P(0.75)=1.5850$, Median=1.52. • Die interpolierten Perzentilen und Quartilen können mit Lienert (1973, p. 42) überprüft werden. BiAS. benutzt bei der Berechnung Tukey's "statistisch äquivalente Blöcke", vgl. dazu auch Beran und Hall (1993).

Konfidenzintervalle

Konfidenzintervall für μ : vgl. t-Test. • Konfidenzintervall für Median: Gemäß Tabellen Sachs (1992, pp.415-6 • Poisson-Konfidenzintervall: Sachs (1992, p.292) • Binomial-Konfidenzintervall: Sachs (1992, p.436) • Konfidenzintervall für mittleren Korrelationskoeffizienten: Sachs (1999, Kap. 552) • Die übrigen Konfidenzintervalle werden (Hinweise!) im Zusammenhang mit analogen Testverfahren überprüft.

Vierfelder-Tafel

χ^2 -Test (Yates-korrigiert): Walter (1980, p.122, pp.140-1), Ackermann (1977, p. 75), Zahlreiche Per-Hand-Berechnungen • Fisher-Test: Büning und Trenkler (1978, p. 247), Sachs (1992, p. 478), Hartung (1982, p.414) • McNemar-Test: Approximativ: Fleiss (1973, p.74), Walter (1980, p.124). Exakt: Binomial-Tabellen • Median-Test: Sachs (1992, p.390), auch mit χ^2 - und Fisher-Test überprüfbar. • Cohen's Kappa: Fleiss (1973, pp.143ff), Bortz-Lienert-Boehnke (1990, pp. 450-453) (Vorsicht Rechengenauigkeit!) • Pearson's Kontingenzkoeffizient: Sachs (1992, p.476), CC ohne Yates-Korrektur. • Gart's Cross-Over-Analyse: Gart (1969, pp.75ff), Bortz et al.(1990, p.165-5) • Konfidenzintervall für $\mu_1-\mu_2$: Armitage (1988, p.123ff) • Äquivalenztest: R. Johnson (1988, pp.907ff) • Beurteilung diagnostischer Test: Per-Hand-Berechnungen • Epidemiologie: Sachs (1992, pp.311-319).

Kontingenztafel

χ^2 -Test: Armitage (1988, p.373), Sachs (1992, p.591), Sachs (1992, p.604/6 (Abweichung in Gesamt- χ^2 , Korrekter Wert Sachs (1992, p.595!). Sachs (1992, p.595), Sachs (1992, p.588) • Yates und Cochran's Trend-Test: Armitage (1988, p.373), Sachs (1992, p.604/6 (Gesamt- χ^2 vgl. χ^2 -Test), Sachs (1992, p.591) • Mantel-Haenszel-Test: De Jonge (1983, pp.155ff): $p=0.006863$, $\chi^2=7.3085$. Cohen's Kappa: Fleiss (1973, p.143ff), Bortz (1984/89, p.459) • Pearson's Kontingenzkoeffizient: Sachs (1992, p.476), CC ohne Yates-Korrektur.

Konfigurationsfrequenzanalyse

Einstichproben-KFA: Bortz (1984/89, p.200), Bortz et al. (1990, p.156), Krauth-Lienert (1973, p.34) • Lehmacher (1981) • Mehrstichproben-KFA: Bortz et al. (1990, p.158), Krauth-Lienert (1973), Haf et al. (1985, pp.70-1) • Clusterung binärer und/oder kategorialer Daten: Lautsch, E., G.A. Lienert (1993)

Test auf Gauß-Verteilung

χ^2 -Anpassungstest: Per-Hand-Berechnungen; BiAS. berechnet die Intervallbreiten gemäß Sturges, so dass Lehrbuchbeispiele i.a. nicht reproduziert werden. • Shapiro-Wilk-Test: Graf et al. (1987, p.393) • Kolmogoroff-Smirnoff-Lilliefors-Test: Graf et al. (1987, p.393) (Präzision!) • David-Test: Sachs (1992, p.424), Per-Hand-Berechnungen • Grubbs-Test: Graf et al. (1987, p.391,393)

Regression und Korrelation

Pearson-Regression: Armitage (1988, p.148), Sachs (1992, p.521, p.522), Ackermann (1977, p.91) Vertrauensbereich Sachs (1992, pp.552-3) • Polynomiale Regression: Sachs (1992, p.561), Armitage (1988, pp.148,266), Hartung (1982, pp.592-4), Per-Hand-Berechnungen auch mit Perfect-Fit-Daten • Multiple Regression: Sachs (1992, p.561), Armitage (1982, pp.148,266), Hartung (1982, pp.599-600 (es ist $t=\sqrt{F}$ für $df=1!$), Longley-Daten • Vergleich von Regressionskoeffizienten: Sachs (1992, p.554), Linder/Berchtold (1982, p.191ff) • Hahn-Prognoseintervalle: Per-Hand-Berechnungen gemäß Sachs (1992, p.557) auf Grund der (überprüften) Resultate • Pearson-Korrelation: Sachs (1992, p.521), Sachs (1992, p.522), Hartung (1992, p.547), Ackermann (1977, p.94) • Partielle Korrelation: Per-Hand-Berechnungen gemäß Sachs (1992, pp.570-3) auf Grund der verifizierten Korrelationskoeffizienten, analog Hartung (1982, p. 562) • Multiple Korrelation: Per-Hand-Berechnungen gemäß Sachs (1992, pp.570-3) auf Grund der verifizierten Korrelationskoeffizienten, analog Hartung (1982, p.564), Bortz (1984/89, pp.559-560) (F wg. Rundung von R ungenau, $F=24.24$, vgl. Multiple Regression) • Rang-Korrelation: Lienert (1973, p.685-7) • Spearman's Rangkorrelation: Sachs (1992, p.513, p.522) • Abbauverfahren Multiple Regression: Flury-Riedwyl (1983) Fisher-Verlag. • Vergleich von Korrelationskoeffizienten Sachs (1999, Kap. 552) • Logistische Regression: Hosmer/Lemeshow (2000) • Zirkuläre Regression: Batschelet (1981) Kapitel 8 und 9.

Parametrische Statistik

Einstichproben-t-Test und Konfidenzintervalle: Graf et al. (1987, p.399ff), Walter (1980, pp.127ff), Ackermann (1977, p.84) • Zweistichproben-t-Test und Konfidenzintervalle: Graf et al. (1987, p.397ff), Lorenz (1992, pp.170ff), Walter (1980, p.126) • Kolmogoroff-Smirnoff-Test (Lilliefors-Variante) in t-Tests & Varianzanalyse: Identisch mit Test in "Prüfung auf Gauß-Verteilung" • F-Test (Varianzhomogenität) in t-Tests und Varianzanalyse: Identisch mit Test in "Prüfung auf Gauß-Verteilung" • Welch-Test: Sachs (1990, p.356), Lorenz (1992, p.175) • F-Test: Sachs (1992, p.348), Lorenz (1992, p.178) (Exakte Berechnung von s^2 mit z.B. Zweistichproben-t-Test, Lorenz gibt nur 4 Dezimalstellen genau an!) • Zur Varianz- und Kovarianzanalyse: BiAS. rechnet grundsätzlich mit a posteriori-Vergleichen, bitte beim Vergleich der Ergebnisse beachten! • Einweg-Varianzanalyse: Freund/Littel bzw. SAS, incl. Scheffe-Tests, Datensatz aus Freund/Littel, Bortz (1984/89, pp.458ff, 300ff, 325ff, 329ff), Ackermann (1977, p.96) • Zweiweg-Varianzanalyse: Bortz (1984/89, 355ff), Weitere Prüfung der Scheffe-Vergleiche: Vergleich mit Einweg-Varianzanalyse mit vernachlässigtem zweiten Faktor • Lineare Kontraste: Bortz (1984/89, 325ff) Vorsicht: Als a posteriori-Vergleich besitzt der Zähler $df=2$, also $F=58.17(9)$ • Orthogonale Polynome: Prüfung Koeffizienten

Ackermann (1979, p.604) • Einweg-Kovarianzanalyse (reduziertes Modell): Freund/ Littel bzw. SAS, inclusive Scheffe-Tests, Datensatz aus Freund/Littel, Bortz (1984, p.458ff) • Zweiweg-Kovarianzanalyse: Bortz (1984/89, 479ff), Vorsicht: Fehler in "Einzelvergleiche 2) und 3)": $n=6$, nicht $n=3$! Weitere Validierungen der Scheffe-Vergleiche: Einweg-Kovarianzanalyse mit vernachlässigtem zweiten Faktor, Vergleich mit Zweiweg-Varianzanalyse mit vernachlässigter Kovariablen. • Bartlett-Test: Sachs (1992, p.617), Bortz (1984/89, pp.329,345) (Rechengenauigkeit!) • Einstichproben-Hotelling's Test: Kramer (1972, p.51, p.54) (Vorsicht: Kramer prüft in diesen Beispielen $H_0(\mu_1=15, \mu_2=6, \mu_3=2.85)$ für p.51, für p.54 $H_0(\mu_1=553, \mu_2=635, \mu_3=532, \mu_4=613)$; die Daten der Beispiele wurden deshalb vor Berechnung via BiAS.' BASIC transformiert ($x'_i=x_i-\mu_i$). Die Ergebnisse bei Kramer sind - vermutlich wegen der Matrizeninversion - relativ ungenau: Beispiel p.51: $T_2=24.9692$, p.54: $T_2=1.1602$.) • Zweistichproben-Hotelling's Test: Kramer (1972, p.73 (dort $T_2=23.891$, vgl. Bemerkung zu Einstichproben-Hotelling-Test) • Cross-Over-Analyse: Daten Grizzle (1965), Abt (1985).

Nicht-parametrische Statistik

Die Berechnung der exakten p-Werte für die beiden Wilcoxon-Tests wurde mit allgemein verfügbarer Tabellen überprüft (u.a. exakte Werte bei Siegel (1956, pp.271, Tab. J), ebenfalls die exakten p-Werte für Binomial-basierte Tests (z.B. McNemar-Test, Vergleich Häufigkeiten etc.) Alle Tests werden bei Bindungen ausnahmslos mit Tie-Korrektur berechnet! • Wilcoxon-matched-pairs-Test: Hollander-Wolfe (1973, p.30) (Daten p.29), Graf et al. (1987, p.399ff), Hodges-Lehmann-Schätzer, Tukey-Konfidenzintervall: Hollander-Wolfe (1973, pp.50-52, pp.29-36) • Wilcoxon-Mann-Whitney-Test: Graf et al. (1987, p.399), Ackermann (1977, p.87), Hodges-Lehmann-Schätzer und Moses-Konfidenzintervall: Hollander-Wolfe (1973, pp.69,75,79) • Kolmogoroff-Smirnoff-Test: Hartung (1982, pp.519, 523) • Kruskal-Wallis-Test: Daten Sachs (1992, p.394, Vorsicht: Sachs berechnet die Prüfgröße H^{\wedge} ohne Tie-Korrektur! Büning et al. (1978, p.201,206) • Dunn-Test: Dunn (1964, pp.241ff) • Friedman-Test: Daten Klotz (1980), zwei fehlende Werte ersetzt durch Zeilenmedian, viele Bindungen (!), Ergebnis ist identisch mit Klotz, Sachs (1992, pp.669,671) • Test nach Rijkooort & Wise (Vorsicht: 2. Freiheitsgrad bei Sachs (1992, p.667) muss korrekt lauten $(c-1)(r-1)$), Conover (1980, p.301) • Wilcoxon-Wilcox-Test: Sachs (1992, p.675) • Dixon und Mood's Vorzeichentest: Tabellen der Binomial-Verteilung mit $p=0.5$. (Es wird nur die exakte Version berechnet.) • Cochran's Q-Test: Büning et al. (1978, pp.227-8) • Vergleich zweier Häufigkeiten/Poisson-Verteilungen: Approximativ: Sachs (1992, p.297, kontinuieritätskorrigiert, für $df=1$ ist $z^2=\chi^2$. Exakt: Tabellen der Binomial-Verteilung mit $p=0.5$ • Cross-Over-Analyse nach Koch: Daten Grizzle (1965), Abt (1985), Daten Koch (1972): Prüfung der Rangsummen R_1, R_2 , Prüfgröße $U=\text{Min}(U_1, U_2)$ mit per-Hand-Berechnungen.

Survival-Analyse

Kaplan-Meier-Schätzungen: Armitage et al. (1988, pp.431-2), Harms (1988, p.182) • Gehan-Wilcoxon-Test: Daten Peto-Pike (1977, p.31, Stratum "Normale Nierenfunktion", identische Ergebnisse mit "Testimate" (IDV München) • Log-Rank-Tests Peto-Pike und Cox-Mantel: Daten Peto und Pike (1977, p.31), Stratum "Normale Nierenfunktion", identische Ergebnisse mit Testimate" (IDV München), Harms (1988, p.185), Armitage (1988, pp.430-2) • Stratifizierter Test: Daten Peto und Pike (1977, p.31), identische Ergebnisse mit "Testimate". Zwei Gruppen: Programm WinStat 3.1 (Vorsicht: WinStat 3.1 ist in bestimmten Situationen problematisch!!) • Hall-Wellner-Konfidenzintervalle: Harris und Albert (1991, pp.17,41) • Hazard-Raten: Harris und Albert (1991, pp.17,41) und dazu ausführliche per-Hand-Berechnung.

Cluster-Analyse

Mc-Queen-Rubin: Deichsel und Trampisch (1985, pp.3-4) • Single Linkage: Alle Verfahren (Distanzen) wurden durch per-Hand-Berechnungen verifiziert. Deichsel und Trampisch (1985, pp.3-4). • KFA-Clusterung binärer und/oder kategorialer Daten: Lautsch, E., G.A. Lienert (1993)

Diskriminanzanalyse

Lineare DA: Deichsel und Trampisch (1985, pp.86-7), Graham (19xx, pp.115ff), geringe Abweichungen der Trenngeraden evtl. durch zwei Wertepaare fraglicher Korrektheit erklärbar. Verifizierung durch Per-Hand-Berechnungen. • Quadratische DA: Programm nach Inst. of Math. Stat. (ed., 1989), Discriminant Analysis and Clustering. Stat. Science Vol. 4, No. 1, p.34-69. Verifiziert mit Linearer DA. • Box-Test in Quadratischer DA: Bortz (1984/89, pp.440-2) • Stoller's nicht-parametrische DA: Stoller (1954, pp.770ff), per-Hand-gerechnete Beispiele. • KFA-Diskrimination binärer/kategorialer Daten: Lautsch, E., G.A. Lienert (1993), Krauth (1993).

Zeitreihenanalyse

Varianzanalyse mit Messwiederholungen: Bortz (1984/89, p.412) (Rechengenauigkeit des F-Wertes, beide MQ stimmen überein!) Box-Test in ANOVA: Bortz (1984/89, pp.440-2) • Mann-Test: Hartung (1982, p.249-51), Koziol et al. (1978, pp.753ff) • Box-Plots: Daten Graf et al. (1987, p.379), vgl. Bemerkungen zur Deskriptiven Statistik.

Faktorenanalyse

Korrelationsmatrizen nach Pearson und Spearman: Bortz p.638, Modul Regression und Korrelation • Explorative FA: Haf et al. (1985, pp.90-96), Bortz (1984/89, p.638) (Vorsicht: in Bortz (1979) wird eine Hauptkomponentenanalyse durchgeführt, deren Ergebnisse hier gut mit BiAS.' Centroid-Methode übereinstimmen.) • Konfirmatorische Faktorenanalyse.: Haf et al. (1985, pp.101/2), Bortz (1984/89, p.638) (Vgl. Bemerkung zur Explorativen Faktorenanalyse.)

Toleranzbereiche

Parametrisch: Per-Hand-Berechnungen gemäß Walter (1980, p.117) • Nicht-parametrisch: Überprüfung an Hand Tabellen aus Ackermann (1985), zahlreiche Per-Hand-Berechnungen (vgl. auch Perzentilen) • Perzentilen, Quartilen: Daten Graf et al. (1987, p.379 (vgl. "Deskriptive Statistik"), zahlreiche Per-Hand-Berechnungen.

Pharmakokinetische Parameter und Bioäquivalenz

Berechnung von u.a. AUC, Cmax, Tmax, T1/2: per Hand, T1/2 gemäß TopFit (Fa. Gödecke et al.), Algorithmus 2.1. • Parametrisches Intervall für A/B: Wie Konfidenzintervall für μ • Parametrisches Intervall für $\ln(A/B)$: Wie Konfidenzintervall für μ • Nicht-parametrisches Intervall für A/B, $\ln(A/B)$: Wie nicht-parametrisches Konfidenzintervall für Median • Nicht-parametrisches Tukey-Intervall für $\ln(A/B)$: Hauschke (1990) • Nicht-parametrisches Moses-Intervall für A/B, $\ln(A/B)$: Hauschke (1990).

Methodenvergleiche und Diagnostischer Test

Bland- und Altman's Methodenvergleich: Bland und Altman, Lancet 1986, Ackermann 1997
• Cohen's Kappa wie zur Vierfeldertafel • Beurteilung diagnostischer Tests: R.S. Galen, S.R. Gambino (1979) Fischer Stuttgart, Binomial-Konfidenzintervalle mit BiAS.' Methoden, Abel (1993), Hippokrates Stuttgart. • ROC-Kurven: Per-Hand-Berechnungen, Abel (1993) und Campbell (1994).

Mehrere Nullhypothesen und Interimsanalyse

Multiple Tests: Zahlreiche per-Hand-Berechnungen, Holm (1979, pp.65ff), Rüger (1978, pp.591ff), Simes (1986, pp.751ff), Hommel (1988, pp.383ff) • Meta-Analyse, Einseitige Alternative: Zahlreiche Per-Hand-Berechnungen nach den zitierten Arbeiten. Fisher (1971, pp.99-101), Koziol et al. (1978, pp.753ff), Winer (1971) • Meta-Analyse, Zweiseitige Alternative: Per-Hand-Berechnungen nach den zitierten Arbeiten. Winer (1971), Koziol et al. (1978, pp.753ff) • Zwischenauswertungen: Pocock (1977), Bauer und Köhne (1994).

Fallzahlen und Power

BiAS. rundet bei Fallzahlschätzungen grundsätzlich auf. Dadurch ergeben sich gelegentlich Abweichungen von der Referenzliteratur. • Konfidenzintervall für μ : Per-Hand-Berechnungen • Schätzung von Wahrscheinlichkeiten: Mace (1964, p.63) • Vergleich binomialverteilter Größen: Schouten (1982) und Scheidermann (1964) • Äquivalenz binomialverteilter Größen: Blackwelder (1982), Machin und Campbell (1987) Tab. 4.1 • χ^2 -Test mit Yates-Korrektur: Casagrande (1978) • Länge eines Konfidenzintervalls: Bristol (1989), Per-Hand-Berechnung • Korrelationskoeffizient: Machin und Campbell (1987, pp.92ff) Tab. 8.1 • Einstichproben-t-Test: Machin und Campbell (1987) Tab. 7.1 (Vorsicht: BiAS. rechnet nicht mit z-Approximation, sondern via t-Verteilung, wodurch gelegentlich größere (exaktere) Werte resultieren) • Zweistichproben-t-Test: Werner (1984, p.247-9), Bratcher et al.(1970), Machin und Campbell (1987) Tab. 7.2 (vgl. auch Bemerkung zum Einstichproben-Test) • Varianzanalyse: Werner (1984, p.247-9), Bratcher et al. (1970). Nachtests Abt (1985/87), handgerechnete Beispiele. • Cross-Over-Analyse: Per-Hand-Berechnungen nach Zimmermann und Rahlfs (1980) • Log-Rank-Test: Machin und Campbell (1987, pp.98ff,115ff)+Tab. 9.2, Freedman (1982) • Vorzeichen- und Wilcoxon-Test: Noether (1987, p.646) • Epidemiologie: Lemeshow (1988, pp.759ff) • Many-one-Vergleiche: Horn und Vollandt (1995). • Bauer-Köhne-Design: Bauer und Köhne (1994).

Verteilungen

Alle Verteilungen wurden an Hand einschlägiger Tabellen überprüft. In der Regel ergaben sich Übereinstimmungen in mind. 4 Stellen (soweit vorhanden). Gelegentlich verwendete Tabellenwerte (z.B. Shapiro-Wilk-Test) wurden ebenfalls mehrfach und unabhängig überprüft. Eine Überprüfung der exakten p-Werte zum Wilcoxon-Test ist mit Hilfe der Tabelle J aus Siegel (1956) möglich. Des Weiteren wurden zum Vergleich einschlägige Programme herangezogen.

Die verwendeten Validierungsdaten finden Sie im Unterverzeichnis ..\Valide des Verzeichnisses, in dem das Programm installiert wurde.



8. Leistungsspektrum



- **Studienplanung und -design**
 - Vollständige Randomisierung
 - Randomisierung in Blöcken
 - Randomisierung im Cross-Over-Design
 - Uni- und bivariate Zufallszahlen
 - Lateinische Quadrate
 - Gruppensequentielle und Adaptive Designs
- **Fallzahl-Berechnungen**
 - Konfidenzintervall für μ
 - Ein- und Zweistichproben-t-Test
 - Varianzanalyse (Anova)
 - Studentized Range (Anova)
 - Einstichproben-t-Test auf Äquivalenz
 - Zweistichproben-t-Test auf Äquivalenz
 - Test auf Therapeutische Äquivalenz
 - Einfache lineare Regressions- und Korrelationskoeffizienten
 - Multiple Regressions- und Korrelationskoeffizienten, Omnibus-Tests
 - t-Test: Many-One-Vergleiche
 - Cross-Over-Analyse im Grizzle-Modell
 - Bland-Altman-Vergleich (Unterschied und Äquivalenz)
 - Wahrscheinlichkeit θ schätzen
 - Binomialen Anteil θ testen
 - Vergleich binomialer Grössen
 - χ^2 -Vierfeldertafel-Test
 - Äquivalenz binomial-verteilter Grössen
 - Dixon-Mood's Vorzeichen-Test
 - McNemar-Test
 - Cohen's Kappa-Koeffizient
 - Wilcoxon-matched-pairs-Test
 - Wilcoxon-Mann-Whitney-U-Test
 - Rangkorrelation: Spearman's rho und Kendall's tau
 - U-Test: Many-One-Vergleiche
 - Log-Rank-Test
 - Relatives Risiko und Odds-Ratio
 - Fallzahlen zum Bauer-Köhne-Design
 - Fallzahlen für gruppensequentielle Designs
 - Fallzahlen für clusterrandomisierte Studien
- **Power-Berechnungen**
 - Ein- und Zweistichproben-t-Test
 - Varianzanalyse (Anova)
 - Einfache lineare Regressions- und Korrelationskoeffizienten
 - Multiple Regressions- und Korrelationskoeffizienten, Omnibus-Tests
 - Cross-Over-Analyse im Grizzle-Modell
 - Binomialen Anteil testen und Vergleich binomialer Grössen
 - Poisson-Häufigkeit testen und Vergleich von Poisson-Häufigkeiten

χ^2 -Vierfeldertafel-Test
 Äquivalenz binomial-verteilter Größen
 Einstichproben- und Zweistichproben-t-Test auf Äquivalenz
 Test auf Therapeutische Äquivalenz
 Dixon-Mood's Vorzeichen-Test
 McNemar-Test
 Cohen's Kappa-Koeffizient
 Wilcoxon-matched-pairs-Test
 Wilcoxon-Mann-Whitney-U-Test
 Rangkorrelation: Spearman's rho und Kendall's tau
 Log-Rank-Test (Survival-Analyse)
 Bland-Altman's Methodenvergleich

- **Deskriptive Statistik: Numerische Methoden**

Durchschnitt und Standardabweichung
 Median und Quartilen Q_1 und Q_3
 Perzentilen und Quantilen
 Variationskoeffizienten CV und CV[%]
 Alle Parameter auch für Zeitreihen
 Uni- und bivariate Häufigkeitstabellen

- **Deskriptive Statistik: Graphische Methoden**

Diverse stetige und diskrete Histogramme
 Histogramme für Nominaldaten
 Vergleichende Histogramme
 Kumulative Häufigkeiten
 Kreisdiagramm
 Pareto-Diagramm
 Box- und Bar-Plots (auch gruppiert)
 Scattergram
 Ladder-Plots
 Gruppierte und nicht-gruppierte Box- und Bar-Plots für Zeitverläufe
 Gruppierte und nicht-gruppierte Einzeldarstellung für Zeitverläufe
 ROC-Kurven einschließlich Konfidenzband und Test
 Kaplan-Meier-Kurven
 Bland-Altman-, Passing-Bablok- und Lin-Plots
 Weitere spezielle Graphiken, zum Beispiel in der Regressionsrechnung

- **Konfidenzintervalle** finden sich zum Teil explizit, aber auch zu fast allen Testverfahren und werden hier nicht im Einzelnen aufgeführt.

- **Tests auf Gauß-Verteilung**

χ^2 -Anpassungstest mit Graphik
 Kolmogoroff-Smirnoff-Lilliefors-Test (auch für bekannte μ und σ), mit Graphik
 Wilk-Shapiro-Test
 Wilk-Shapiro-Royston-Test
 Grubbs' Ausreißer-Test
 David-Hartley-Pearson's Ausreißer-Test
 David's multipler Test auf Ausreißer
 Mudholkar's multivariater Test auf Gauß-Verteilung

- **Vierfelder-Tafeln**
 - χ^2 -Vierfelder-Test mit/ohne Yates-Korrektur
 - χ^2 -Test nach Klassifikation
 - Fisher's Exact Test
 - McNemar- χ^2 -Test
 - McNemar-Binomial-Test
 - Cohen's Kappa
 - Mantel-Haenszel-Test
 - Median-Test
 - Äquivalenztest
 - Diagnostische Tests
 - Gart's Cross-Over-Analyse
- **Kontingenztafeln**
 - χ^2 -Kontingenztafel-Test
 - Haldane-Dawson- und Craddock-Flood-Test
 - Yates-Cochran-Test auf Trend
 - Mantel-Haenszel-Zimmermann-Test
 - Pearson's Kontingenzkoeffizient
 - Terwilliger-Ott's Symmetrie-Test
 - Bowker's Felder-Symmetrie-Test
 - Cohen's Kappa für k Ratern und s Stufen, auch gewichtet
 - Vergleich von $k \geq 2$ Kontingenztafeln
 - χ^2 -Test nach 2-Weg-Klassifikation
 - Mehrdimensionale Kontingenztafeln
 - Analyse einzelner Felder per KFA
- **Regressions- und Korrelation**
 - Lineare Regression mit Konfidenzintervall und Graphik
 - Polynomiale Regression mit Abbau und Graphik
 - Multiple Regression mit Abbau
 - Logistische Regression mit Abbau
 - Vergleich von Regressionsgeraden
 - Hahn-Prognose-Intervall mit Graphik
 - Lineare Korrelation
 - Multiple und Partielle Korrelation
 - Zirkuläre Regression und Korrelation
 - Vergleich von Korrelationskoeffizienten
 - Korrelationsmatrizen
 - Rang-Regression
 - Spearman's Rang-Korrelation
 - Intra-Cluster-Korrelation
 - Intraclass-Korrelation
- **Parametrische Methoden**
 - Einstichproben-t-Test
 - Zweistichproben-t-Test
 - Welch's modifizierter t-Test
 - Varianzgleichheit: F-Test, Bartlett's Test und Box-Test
 - Varianzanalyse (ANOVA)

Kovarianzanalyse (COVAR)
Bifaktorielle ANOVA und COVAR
ANOVA für Messwiederholungen
Orthogonale Polynome in der ANOVA
Lineare Kontraste in der ANOVA
ANOVA-Trend-Tests
Multiple Vergleiche nach Scheffé
Grizzle's Cross-Over-Analyse
Multivariate Ein- und Zweistichproben-Hotelling's T^2 -Tests
Simultane Konfidenzintervalle zu beiden Hotelling-Tests

- **Äquivalenz-Tests**

Einstichproben-t-Test auf Äquivalenz
Zweistichproben-t-Test auf Äquivalenz
Test auf Therapeutische Äquivalenz ("Nicht-Unterlegenheit")
Bland-Altman-Vergleich auf Äquivalenz
Äquivalenz binomial-verteilter Grössen
Diverse Bioäquivalenz-Prüfungen im Cross-Over
Äquivalenz-Tests mit Konfidenzintervallen sind nicht explizit aufgeführt,
zum Beispiel zum U-Test u.v.a.

- **Survival-Analyse**

Kaplan-Meier-Schätzer mit Graphik
Hall-Wellner-Konfidenzintervalle
Log-Rank-Test, auch für $k > 2$ Gruppen
Stratifizierter Log-Rank-Test
Peto-Pike-Test
Cox-Mantel-Test
Gehan-Wilcoxon-Test
Relative Hazards für $k \geq 2$ Gruppen mit Konfidenzintervallen

- **Nicht-parametrische Methoden**

Wilcoxon-matched-pairs-Test
Wilcoxon-Mann-Whitney-U-Test
Hodges-Lehmann-Schätzer mit Konfidenzintervall
Mann-Whitney-Schätzer mit Konfidenzintervall
Kolmogoroff-Smirnoff-Test
Mantel-Haenszel-Test für $k \geq 2$ Gruppen
Koch's Nonparametrische Cross-Over-Analyse
Gart's Dichotome Cross-Over-Analyse
Page-Test auf Trend
Friedman-Test mit Wilcoxon-Wilcox-Test als multipler post-hoc-Test
Multiple Schaich-Hamerle- und Conover-Tests
Kendall und Babington-Smith's W-Koeffizient
Kruskal-Wallis-Test mit u.a. Dunn-Holm-Vergleichen als post-hoc-Tests
Nemenyi's multiple Vergleiche
Multiple Conover-Iman-Tests
Cochrans Q-Test
Anpassungstests, u.a. Test auf Gleichverteilung
Vergleich von Häufigkeiten

Diverse Äquivalenztests, auch mit Konfidenzintervallen
Exakte nonparametrische p-Werte zu (fast) allen Verfahren

- **Binomial- und Poisson-Tests**

Diverse Binomial-Tests

Test für den Binomial-Parameter θ

Test und Binomial-Konfidenzintervalle für θ , $\theta_1 - \theta_2$ und θ_1 / θ_2

Poisson-Konfidenzintervall für λ

Vergleich von Poisson-Häufigkeiten

- **Cluster-Analyse**

Cluster-Analysen mit Datenstandardisierungen und mehreren Metriken:

L1-Norm, City-Block, Canberra-Metrik, Tanimoto-Distanz, Simple-Matching

Single-Linkage-Clusterung mit diesen Metriken und Euklidischer Distanz

Kmeans-Verfahren mit Euklidischer Distanz

Dichotome Cluster-Analyse gemäß KFA

- **Diskriminanzanalyse**

Lineare Diskriminanzanalyse

Quadratische Diskriminanzanalyse

Abbaumodell der Diskriminanzanalyse

Dichotome Diskriminanzanalyse gemäß KFA

Stoller's nonparametrische Diskriminanzanalyse

- **Faktorenanalyse**

Explorative Faktorenanalyse

Konfirmatorische Faktorenanalyse

- **Reliabilitätsanalyse**

Intraclass-Korrelation incl. Konfidenzintervall

Cronbach's alpha incl. Konfidenzintervall

Cohen's kappa incl. Konfidenzintervall, auch als "weighted kappa"

- **Zeitreihen-Analyse**

Deskriptive Statistiken im Zeitverlauf

Box- & Bar-Plots im Zeitverlauf, auch gruppiert

Graphik der Einzelverläufe, auch gruppiert

ANOVA für Messwiederholungen

Mann-Test

Simultane Konfidenzintervalle

Hotelling's Test

Page's Trend-Test

Yates-Cochran-Test auf Trend

Wallis-Moore's Phasentest

Wald-Wolfowitz' Iterationstest

Cochran's Niveau-Konstanz-Test

- **Konfigurationsfrequenzanalyse (KFA)**

χ^2 -, Binomial- und Lehmann-Test

KFA für mehrere Stichproben

Dichotome KFA-Cluster- & Diskriminanzanalyse

- **Norm- und Toleranzbereiche**
 Uni- und bivariate Toleranzbereiche
 Parametrische und Nichtparametrische Toleranzbereiche
 Box-Cox-transformierte Toleranzbereiche
- **Pharmakokinetische Methoden**
 Parameter: AUC, $T_{1/2}$, λ , C_{min} , C_{max} , T_{max} , C_{av} , PTF, PTS, HVD, MRT und AUCM
 Diverse Bioäquivalenz-Prüfungen im Cross-Over
 Moses- und Tukey-Konfidenzintervalle für Bioäquivalenz
- **Epidemiologische Methoden**
 Odds-Ratio und Relatives Risiko mit Konfidenzintervallen
 Mantel-Haenszel-Test
- **Diagnostische Tests**
 Sensitivität, Spezifität, Effizienz, Prädiktive Werte
 Youden-Index
 Likelihood-Ratios mit Konfidenzintervallen
 Vergleich von Likelihood-Ratios, "paired" und "unpaired"
 Vergleich von Prädiktiven Werten, "paired" und "unpaired"
 Cohen's Kappa (gewichtet und ungewichtet)
 Konfidenzintervalle zu allen Parametern
 Diverse Symmetrietests
 ROC-Analyse und ROC-Vergleiche, "paired" und "unpaired"
- **Methodenvergleiche**
 Bland-Altman-Verfahren mit Graphik
 Passing-Bablok-Regression mit Graphik
 Lin's Konkordanz-Korrelation
 Mc-Nemar's Symmetrietest
 Bowker's Symmetrietest
 Terwilliger und Ott's Test
 Cohen's Kappa-Koeffizient, auch als "weighted kappa"
- **Interimsanalyse: Zwischenauswertungen**
 Analyse nach Pocock und nach O'Brien & Fleming
 Adaptives Bauer-Köhne-Design
- **Mehrere Nullhypothesen – Multiple Tests**
 α -Adjustierungen (Bonferroni, Holm, Hommel, Simes und Rüger)
 Abschlusstests bei hierarchisch geordneten Hypothesen
 Benjamini-Hochberg's False Discovery Rate FDR
- **Meta-Analyse mit p-Werten**
 Diverse ein- und zweiseitige Methoden, Verfahren nach
 Fisher, Tippet, Mudholkar-George und mit inverser Gauß-Verteilung
- **Evidence Based Medicine (EBM)**
 Parameter CER, EER, ARR, RRR und NNT mit Konfidenzintervallen
 Test zum Risikovergleich Likelihood-Ratio positiv und negativ mit KI's



Anhang: Validierung und Beispiele — ▲

Wegen Überschneidung der Themenbereiche wurden in vielen Abschnitten [Links](#) eingefügt, die jeweils mit blau unterlegt und unterstrichen gekennzeichnet wurden. Die blauen Pfeile am Seitenrand, ebenfalls Links, führen zurück zum Abschnittsanfang.

A.1 Deskriptive Statistik ▲

Schätzgrößen Durchschnitt \bar{x} , Median, SD, Quartilen, CV etc.

Deskriptive Größen in der Zeitreihenanalyse ([A.13](#))

Graphiken (Hier nur teilweise aufgeführt, mehr dazu im Programm):

Histogramm für stetige Größen

Kerndichteschätzer

Histogramm für diskrete Größen

Histogramm für nominale Größen

Empirische Verteilungsfunktion - Kumulative Häufigkeiten

Kumulative Klassenhäufigkeiten

Kreisdiagramm

Pareto-Diagramm

Ladder-Plots

Dot-Plots (Einzelwerte)

Scattergram (Streudiagramm, Punktwolke)

Box- und Bar-Plots

Box- und Bar-Plots für Zeitverläufe ([A.13](#))

Einzelverläufe bei Zeitreihen ([A.13](#))

Angepasste Gauß-Verteilung (Histogramm und kumulativ) ([A.6](#))

Regressionsgerade ([A.7](#))

Prognoseintervall ([A.7](#))

Polynomiale Regression ([A.7](#))

Cluster-Analyse ([A.11](#))

Diskriminanzanalyse ([A.12](#))

Schätzgrößen: Graf et al. (1987, p. 379) geben ein Beispiel zur Berechnung von Durchschnitt \bar{x} , Standardabweichung s , Streuung s^2 , Median, Spannweite, Variationskoeffizient und Quartilen an. Die erste Spalte der hier verwendeten Daten entstammt Graf et al., die beiden nächsten Spalten entstanden durch Multiplikation der ersten Spalte mit 10000 und 1/10000, die letzte durch Addition der Konstanten $x=1000$. **Vorsicht: BiAS.** verwendet für die Ein- und Ausgabe die einfache Darstellungsgenauigkeit, also 6 Stellen.

Die Funktion "Schätzgrößen" besitzt keine Ausgabe eingelesener Daten - hier per Editor:

Editor: Ausdrucken

Die Datei Graf.379 umfasst 4 Spalten bzw. 10 Zeilen (0.3Kb)

Var:	1	2	3	4
1:	1.62	16200.0	0.000162	1001.62
2:	1.42	14200.0	0.000142	1001.42

```

3:  1.58 15800.0 0.000158 1001.58
4:  1.39 13900.0 0.000139 1001.39
5:  1.54 15400.0 0.000154 1001.54
6:  1.35 13500.0 0.000135 1001.35
7:  1.50 15000.0 0.000150 1001.50
8:  1.48 14800.0 0.000148 1001.48
9:  1.58 15800.0 0.000158 1001.58
10: 1.60 16000.0 0.000160 1001.60

```

Die nächste Ausgabe erfolgte via "Schätzgrößen" im Menü "Methoden":

Schätzgrößen

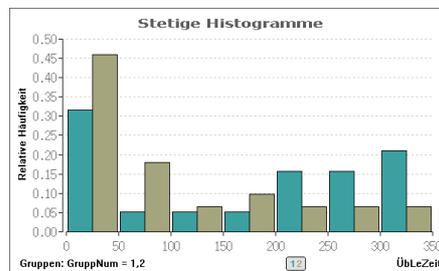
Eingabedatei ist Graf379.DAT

Nr.	Variable	n	miss	Xquer	s	Median	Minimum	Maximum
			SEM	CV	CV(%)	Range	1.Quartil	3.Quartil
1	Asche	10	0	1.5060	0.0942	1.5200	1.3500	1.6200
			0.0298	0.0625	1.9777	0.2700	1.4125	1.5850
2	Asche*X	10	0	15060.00	941.8658	15200.00	13500.00	16200.00
			297.8441	0.0625	1.9777	2700.00	14125.00	15850.00
3	Asche/X	10	0	0.0002	9.419E-06	0.0002	0.0001	0.0002
			2.98E-06	0.0625	1.9777	2.70E-05	0.0001	0.0002
4	Asche+X	10	0	1001.51	0.0942	1001.52	1001.35	1001.62
			0.0298	9.41E-05	0.0030	0.2700	1001.41	1001.59

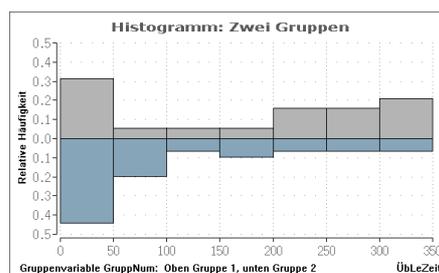
Literatur:

L. Sachs (1992) *Angewandte Statistik*, Springer Kap.314
 B.L. van der Waerden (1957) *Math. Statistik*, Springer S.345

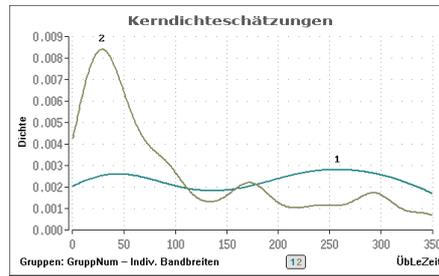
Stetiges Histogramm (eine oder mehrere Gruppen): Zur einfachen Nachvollziehbarkeit werden die Daten aus BiAS.' Testdatei Armitage.DAT verwendet (alle Überlebenszeiten ungeachtet der Gruppeneinteilung). Ein diskretes oder kumuliertes Histogramm dazu ist ebenfalls möglich, außerdem eine vereinigte Darstellung des stetigen und kumulativen Histogramms.



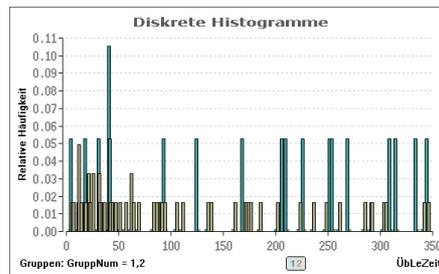
Stetiges Histogramm (Graphik für zwei Gruppen): Zum Nachvollziehen werden die Daten aus BiAS.' Testdatei Armitage.DAT verwendet. Nach oben ist Gruppe 1, nach unten ist Gruppe 2 des Datensatzes aufgetragen. Vgl. auch Graphik zu Kaplan-Meier-Schätzer.



Kerndichteschätzer: Diese umfasst eine alternative Darstellung für eine oder auch für mehrere Gruppen. Die Darstellung erfolgt hier – ungeachtet des inhaltlichen Kontextes - exemplarisch für die Daten aus Armitage.DAT.

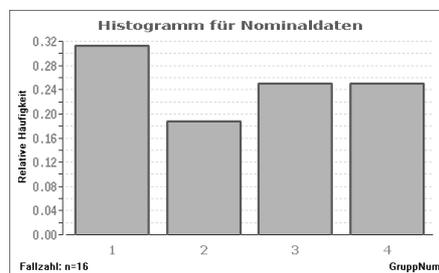


Diskretes Histogramm (eine oder mehrere Gruppen): Zur einfachen Nachvollziehbarkeit werden die Daten aus **BiAS.** Testdatei Armitage.DAT verwendet (alle Überlebenszeiten). Ein stetiges Histogramm oder ein kumuliertes Histogramm dazu ist ebenfalls möglich, außerdem eine vereinigte Darstellung des stetigen und kumulativen Histogramms. Analog zum vergleichenden Histogramm für zwei Gruppen bei stetiger Skala ist auch für diskrete Skalen eine vergleichende Darstellung von Gruppen möglich, in diesem Anhang aber nicht enthalten.

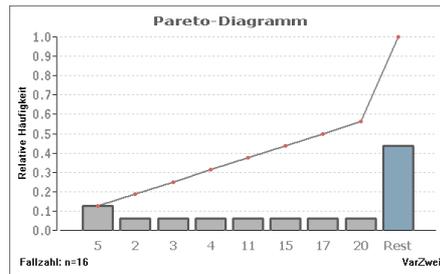


Bitte beachten Sie die unterschiedlichen Größen der Skalierungen: In Kapitel 5 wurde **BiAS.** Graphik-Menü (Abbildung 31) beschrieben, das viele Möglichkeiten zur Modifikation der Graphiken vorsieht. Diese Modifikationsmöglichkeiten betreffen speziell die Skalierung, die Beschriftung, eine Verwendung des Netzgitters (hier in allen Graphiken vorhanden) und die Darstellung einer Randlinie.

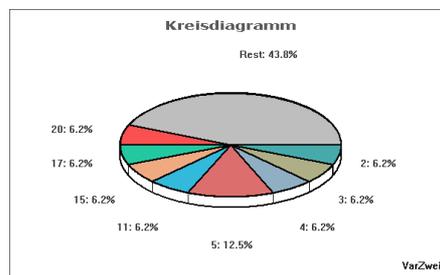
Nominales Histogramm: Zur einfachen Nachvollziehbarkeit werden die Daten aus **BiAS.** Standard-Testdatei Test.DAT verwendet, davon GruppNum. Nominale Histogramme können auch als Kreisdiagramme ("Torten") dargestellt werden. Vgl. Sie bitte auch die Pareto-Diagramme. Die Skalierung kann "freitextlich" via Dialogfeld "Bearbeiten" modifiziert werden.



Pareto-Diagramm: Welche (möglichst wenigen!) Substanz-Klassen kommen mit einer gewissen (möglichst großen!) Häufigkeit vor? Die nominalen Kategorien sind im Pareto-Diagramm ihrer Häufigkeit nach geordnet: Bitte vergleichen Sie dazu auch die Nominal-Histogramme. Auch hier kann die Skalierung "freitextlich" bearbeitet werden.

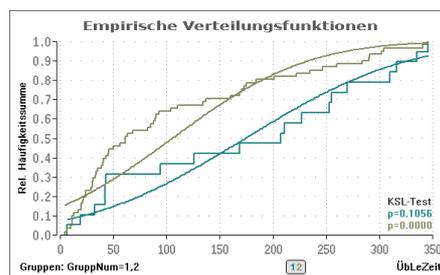


Kreisdiagramm: Zur einfachen Nachvollziehbarkeit werden die Daten aus **BiAS.**' Standard-Testdatei Test.DAT verwendet, davon Variable VarZwei. Kreisdiagramme (sog. "Torten") können auch als Nominale Histogramme dargestellt werden. Optional können die Kategorien geordnet werden. Vgl. Sie bitte dazu auch die Pareto-Diagramme.

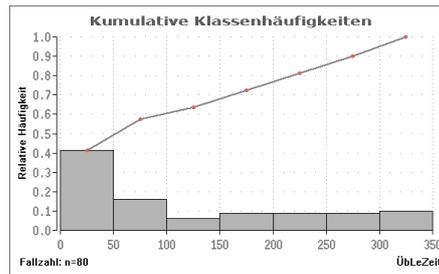


Alle Graphiken mit *nominaler* Skalierung wie zum Beispiel das Pareto-Diagramm oder das Histogramm für Nominal-Klassen (oben) oder weiter unten gruppierte und ungruppierte Box- und Bar-Plots und einige andere gestatten auch eine *textliche* Skalierung/Beschriftung auf der Abszisse.

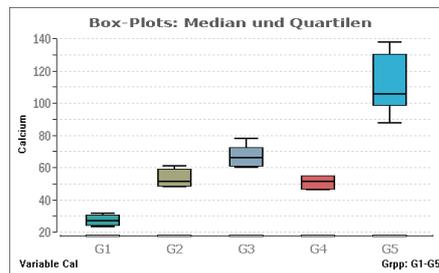
Verteilungsfunktion - Kumulative Häufigkeiten: Zur einfachen Nachvollziehbarkeit werden – ungeachtet des Kontextes - die Daten aus **BiAS.**' Standard-Testdatei Armitage.DAT verwendet, davon Variable ÜbLeZeit. Vergleichen Sie dazu auch stetige Histogramme und diskrete Histogramme, außerdem die vereinigte Darstellung des stetigen und kumulativen Histogramms. Optional ist ein Kolmogoroff-Smirnoff-Test (Tabelle rechts unten) möglich.



Kumulative Klassenhäufigkeiten: Zur einfachen Nachvollziehbarkeit werden die Daten aus **BiAS.**' Standard-Testdatei Armitage.DAT verwendet, davon Variable ÜbLeZeit. Das Diagramm vereinigt ein stetiges Histogramm und ein kumulatives Histogramm in einer gemeinsamen Abbildung. Bitte beachten Sie, dass die Darstellung - im Gegensatz zu nominalen Pareto-Diagrammen - die Häufigkeiten gemäß ihrer quantitativ gegebenen Ordnung wiedergibt.

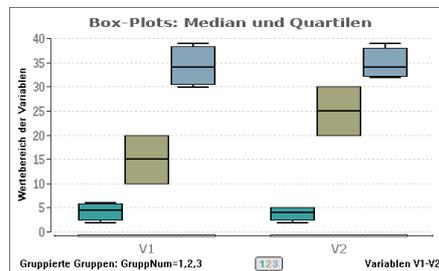


Box-Plot: Box-Plots können mit Median & Quartilen oder mit Durchschnitt & Standardabweichung dargestellt werden. Hier wird die Variable "Cal" aus "Calcium.DAT" verwendet, optional auch inclusive der Einzelwerte darstellbar. Eine Alternative zu Box-Plots sind Bar-Plots.



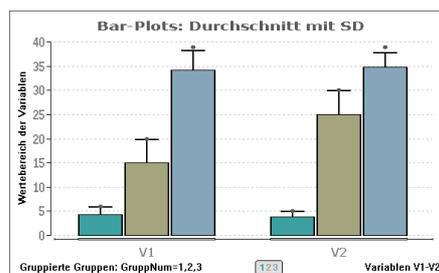
Box: Quartilen Q_1 und Q_3 - Dicker Strich: Median - Antennen: Extremwerte

Gruppierte Darstellungen (hier mit anderen Daten: ..\Valide\Test.DAT) sind ebenfalls möglich:

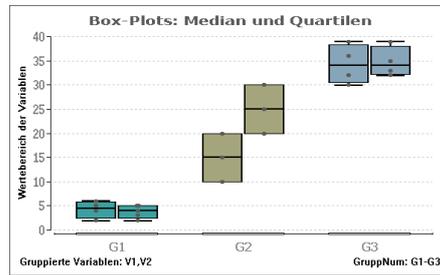


Alle Graphiken werden hier nur mit einer geringen Auflösung von ca. 400x200 Pixel dargestellt, die Auflösung im Programm hängt von Ihrer Bildschirm-Einstellung ab.

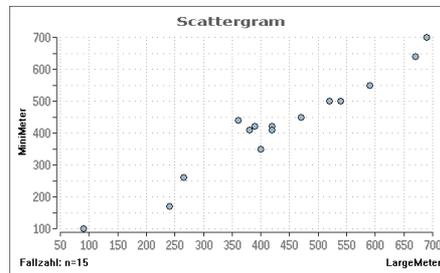
Bar-Plot: Bar-Plots stellen den Durchschnitt \bar{x} als Säule mit aufgesetzter Standardabweichung s dar. Die Darstellung verwendet - wie in der letzten Abbildung - die Datei ..\Valide\Test.



Box- und Bar-Plots: Beide Varianten können nicht nur – wie oben – mit nebeneinander angeordneten Variablen dargestellt werden, sondern die Variablen können auch innerhalb der Gruppen angeordnet werden. Die folgende Darstellung verwendet wieder die Datei ..\Valide\Test.



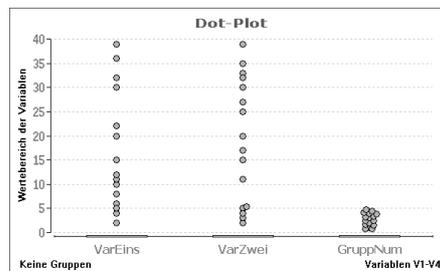
Scattergram: Vergleich zweier Messgeräte; es wurden die Daten in ..\Valide\Meter.DAT verwendet. Bei identischen Werten kann **BIAS**, diese - optional - "verwackeln", um alle Werte sichtbar zu machen. Vgl. auch das Bland-Altman-Verfahren und Passing-Bablok's Regression.



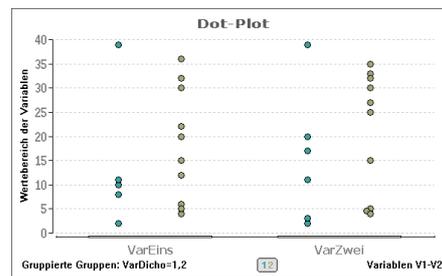
„Scattergrams“ bzw. „Punktwolken“ werden auch in der Regressionsrechnung verwendet, dies im Zusammenhang mit der Darstellung von Regressionsgeraden oder angepasster Polynome.

Bitte beachten Sie die unten aufgeführten Optionen zur Darstellung, mit denen u.a. die Beschriftung geändert werden kann oder z.B. das Netzgitter ein- und ausgeblendet werden kann.

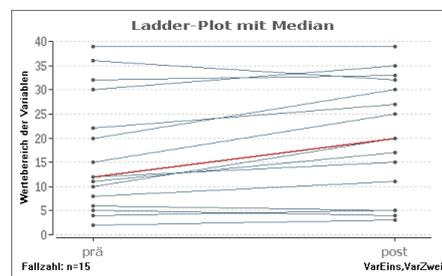
Dot-Plot: Gruppierete oder ungruppierete Darstellung von Einzelwerten, dies analog zu Box- und Barplots. Bei identischen Werten kann **BIAS**, diese - optional - "verwackeln", um alle Werte sichtbar zu machen. Es können vier verschiedene „Dots“ ausgewählt werden. Zur Darstellung wird wieder die Testdatei „..\Valide\Test.DAT“ verwendet:



Die Daten können wie oben auch „gruppiert“ dargestellt werden. In der nächsten Abbildung wurden die Gruppen „gruppiert“, alternativ können auch die Variablen „gruppiert“ werden:



Ladder-Plot: Darstellung von „gepaarten“ Einzelwerten beispielsweise zum Vergleich „prä vs. post“ zum der Einstichproben-t-Test. Zur Darstellung wird die Testdatei „Test.DAT“ verwendet. Die rote Linie kennzeichnet den Verlauf der Mediane:



Weitere im Programm verfügbare Graphiken:

Die oben dargestellten Graphiken sind alle über das Menü „Deskriptive Statistik“ zu erhalten. Viele weitere Graphiken sind nur im Zusammenhang mit Methoden anderer Menüs verfügbar: Dies sind insbesondere die Menüs

- Regression und Korrelation
- Test auf Gauß-Verteilung
- Überlebenszeitanalyse
- Zeitreihenanalyse
- Toleranzbereiche
- Methodenvergleiche
- Diagnostische Tests
- Clusteranalyse
- Diskriminanzanalyse

Die insgesamt etwa 60 Graphiken stehen alle im Zusammenhang mit den Beispielen der entsprechenden Menüs und können wie die oben abgebildeten Graphiken im Graphik-Editor bearbeitet werden.

Viele Optionen zu den Graphiken stehen bereits in den entsprechenden Dialog-Boxes zur Verfügung, andere sind im Graphik-Manü verfügbar. Näheres dazu findet sich auf der nächsten Seite.

Optionen in BiAS.' Graphiken:

Hier wie an vielen anderen Stellen dieser Dokumentation werden nicht alle im Programm vorgesehenen Graphik-Optionen explizit aufgeführt:

Zum Beispiel können Box- und Bar-Plots optional in unterschiedlicher Weise, gegebenenfalls auch mit Einzelwerten, dargestellt werden. Bei identischen Werten kann **BiAS.** diese - ebenfalls optional - "verwackeln", um alle Werte sichtbar zu machen; diese Möglichkeit besteht auch in vielen anderen Graphiken. Für Bland-und Altman's Verfahren etwa kann man optional die Regressionsgerade incl. Konfidenzgrenzen eintragen, ähnliches gilt auch in der Regressionsrechnung. Die Liste der Beispiele lässt sich fortsetzen:

Da hier natürlich nicht alle Graphik-Varianten und -Optionen aufgeführt werden können, wurden alle relevanten Validierungsdateien in einem Unterverzeichnis von **BiAS.** ("..\Valide", in der Download-Version in der Datei Valide.Zip) zugänglich gemacht. Mit Hilfe dieser Daten können Sie alle Optionen ausführlich ausprobieren!

In **BiAS.**' Graphik-Editor gibt das Menü "Einstellungen" einen Überblick über die Standard-Optionen (Gitter, Skalierung, Beschriftung, Schriftart, identische Punkte verwackeln, Dateiausgabe in Bitmap- oder PCX-Datei u.v.a.m.). Bitte vergleichen Sie dazu auch **BiAS.**' Handbuch (Kapitel 5, "Graphische Ausgabe")!

Eine Darstellung des Dialogfeldes "Einstellungen" im Graphik-Editor finden Sie im unteren Bereich des Dialogs in Abbildung 38 und analog in Abbildung 31, in deren Zusammenhang alle Einstellungen etc. ausführlicher beschrieben werden; weiteres dazu in [Kapitel 5](#).



Abbildung 38: Dialogfeld "Graphik bearbeiten"

Das Dialogfeld "Einstellungen" ist "kontextsensitiv": Es werden bei allen Graphiken nur die relevanten, modifizierbaren Einstellungen angezeigt und die oben dargestellten Positionen gegebenenfalls verändert, wie dies zum Beispiel bei Box-Plots der Fall ist. Näheres dazu, wie erwähnt, in [Kapitel 5](#).

Bitte beachten Sie, dass Farben nur im Graphik-Menü verändert werden können, äquivalent kann der aus Windows bekannte Dialog zur Farbmodifikation durch *Doppelklick* auf die entsprechende Farbe aufgerufen werden: Vergleichen Sie dazu bitte den Abschnitt 5 („Graphische Ausgabe“).

Die Auswahl der Farbe(n) kann für spätere Verwendungen „fixiert“ werden: Klicken Sie dazu im Menü des Graphik-Editors auf die Auswahl „Fix“ im Graphikmenü (Abbildungen 30 bis 33); Näheres dazu wieder in Abschnitt 5.

Falls eine Farbe nicht veränderbar ist – zum Beispiel bei Koordinatenachsen – erscheint der Hinweis „Keine Farbe“.

A.2 Konfidenzintervalle



Konfidenzintervall für den Erwartungswert μ

Konfidenzintervall für den Median

Konfidenzintervall für den Binomialparameter θ

Konfidenzintervall für $\theta_1 - \theta_2$ ("paired" und "unpaired")

Konfidenzintervall für θ_1 / θ_2 ("paired" und "unpaired")

Poisson-Konfidenzintervall für den Parameter λ

Poisson-Konfidenzintervall für die Differenz $\lambda_1 - \lambda_2$

Sonstige, in anderen Moduln enthaltene Konfidenzintervalle:

Tukey-Konfidenzintervalle für den Median ([A.9](#))

Moses-Konfidenzintervalle für den Median ([A.9](#))

Konfidenzintervall für Paardifferenzen ([A.8](#))

Konfidenzintervall für Mittelwertsdifferenzen ([A.8](#))

Konfidenzintervalle für Binomial-Differenzen ([A.3](#))

Konfidenzintervall für die Steigung einer Regressionsgeraden ([A.7](#))

Konfidenzintervall für Differenzen von Steigungen ([A.7](#))

Konfidenzintervall für Differenzen von Achsenabschnitten ([A.7](#))

Graphische Konfidenzintervalle für die Regression ([A.7](#))

Konfidenzintervall für den Korrelationskoeffizienten ([A.7](#))

Konfidenzintervall für mittleren Korrelationskoeffizienten ([A.7](#))

Simultane Konfidenzintervalle (eine Stichprobe) ([A.8](#))

Simultane Konfidenzintervalle (zwei Stichproben) ([A.8](#))

Konfidenzintervalle für Kaplan-Meier-Überlebenszeitkurven ([A.13](#))

Graphisches Konfidenzintervall für eine Kaplan-Meier-Kurve ([A.13](#))

Konfidenzintervalle für das Relative Risiko und das Odds-Ratio ([A.19](#))

Konfidenzintervalle zur Beurteilung diagnostischer Test ([A.18](#))

Konfidenzintervall für den Youden-Index ([A.18](#))

Konfidenzintervalle für Likelihood-Ratios ([A.18](#) und [A.19](#))

Konfidenzintervalle zum Vergleich von Likelihood-Ratios ([A.18](#))

Konfidenzintervalle zum Vergleich von Prädiktiven Werten ([A.18](#))

Konfidenzintervalle für Epidemiologische und EbM-Parameter ([A.19](#))

Es werden nicht zu allen Konfidenzintervallen Validierungsbeispiele angegeben, da die oben aufgeführten Methoden weitgehend nicht explizit, sondern im Zusammenhang mit den entsprechenden Testverfahren aufgeführt werden.

Konfidenzintervall für den Erwartungswert μ : Mit den gleichen Daten wie in "Schätzgrößen" (Datei Graf.379) wird für jede Variable ein Konfidenzintervall für den Erwartungswert μ berechnet (Einstichproben-Situation), eine Kontrolle ist wieder durch die 2., 3. und 4. Spalte gegeben:

Konfidenzintervalle für den Erwartungswert μ



Eingabedatei ist Graf379.DAT

Var.	n	P	t	X-quer	s	Li.Grenze	Re.Grenze
1	10	0.9500	2.2622	1.5060	0.0942	1.4386	1.5734
2	10	0.9500	2.2622	15060.0000	941.8658	14386.2295	15733.7705
3	10	0.9500	2.2622	0.0002	9.419E-06	0.0001	0.0002
4	10	0.9500	2.2622	1001.5060	0.0942	1001.4386	1001.5734

Literatur: L. Sachs (1992) Angewandte Statistik, Springer. Kap.31ff

Konfidenzintervall für den Median: Die Konfidenzintervalle werden wieder für die gleichen Daten wie in "Schätzgrößen" berechnet (Datei Graf.379, Einstichproben-Situation).

Konfidenzintervalle für den Median



Eingabedatei ist Graf379.DAT

Var.	n	P	Min	Max	Median	Li.Grenze	Re.Grenze
1	10	0.9500	1.3500	1.6200	1.5200	1.3900	1.6000
2	10	0.9500	13500.0000	16200.0000	15200.0000	13900.0000	16000.0000
3	10	0.9500	0.0001	0.0002	0.0002	0.0001	0.0002
4	10	0.9500	1001.3500	1001.6200	1001.5200	1001.3900	1001.6000

Literatur: L. Sachs (1992) Angew. Statistik Kap. 314ff, Springer.

Binomial-Konfidenzintervall für θ : In 20 gleichartigen Versuchen mit den möglichen Ergebnissen A und B wird 7 mal A bzw. 13 mal B beobachtet. Gewünscht ist ein 95%-Vertrauensbereich für den Parameter θ einer Binomialverteilung aufgrund von $p^{\wedge}=7/20=0.35$ (Sachs (1996, p. 436)):

Konfidenzintervall für theta einer Binomial-Verteilung



Konfidenz P: 0.950000
Realisationen n: 20
Erfolge k: 7
Parameter θ^{\wedge} : 0.350000
Untere Grenze: 0.153909
Obere Grenze: 0.592189

Literatur: L. Sachs (1992) Angew. Statistik Kap. 45, Springer.

Konfidenzintervall für $\theta_1-\theta_2$ ("unpaired"): In einem klinischen Versuch wird eine alte mit einer neuen Behandlung verglichen (Zeilen). Von 257 Patienten mit Behandlung A sterben 41, von 244 Patienten unter B sterben 64. Berechnet wird ein 95%-Konfidenzintervall für die Differenz der Überlebensraten (Armitage und Berry (1988, p.124)):

Px100%-Konfidenzintervalle und Test für $\delta = \theta_1 - \theta_2$



Eingegebene Tafel:	Erwartete Häufigkeiten $\Phi(i,j)$:
41 216 257	53.9 203.1 257.0
64 180 244	51.1 192.9 244.0
-----+-----	-----+-----
105 396 501	105.0 396.0 501.0

Schätzungen für Theta1 und Theta2 der Zeilen: P1 = 0.1595 ; P2 = 0.2623

Konfidenzintervalle für Delta=Theta1-Theta2

	ohne Yates-Korrektur:	mit Yates-Lorrektur:
P=90%:	(-0.1624 , -0.0431)	(-0.1664 , -0.0391)
P=95%:	(-0.1738 , -0.0317)	(-0.1778 , -0.0277)
P=99%:	(-0.1962 , -0.0094)	(-0.2002 , -0.0054)

BEMERKUNG: Für $n < 60$ nur die Kontinuitätskorrigierten Intervalle verwenden!

P. Armitage, G. Berry (1987) *Statistical Methods*, Blackwell Sci.Pub. S.123
J.L. Fleiss (1981) *Stat. Methods for Rates and Proportions*. Wiley NY, p.29

Konfidenzintervall für $\theta_1 - \theta_2$ ("paired"): Konfidenzintervall für die Differenz $\theta_1 - \theta_2$ zweier abhängiger binomialer Größen (beide Merkmale am gleichen Patienten!). Ein Beispiel findet sich im Zusammenhang mit dem McNemar-Test in [Anhang 3](#).

Konfidenzintervall für θ_1 / θ_2 ("unpaired" bzw. "paired"): Die Fragestellung ist analog zu der der Differenz $\theta_1 - \theta_2$, so dass auf die entsprechenden Beispiele hingewiesen wird. Die Validierung für "paired" erfolgte gemäß Bonnett und Price (2006), für "unpaired" gemäß Koopman (1984).

Poisson-Konfidenzintervall für λ : Auf Grund von $x=10$ Beobachtungen in einem bestimmten Zeitraum soll ein 95%-Vertrauensbereich für den Parameter λ der Poisson-Verteilung berechnet werden (Sachs 1992, p. 292). In [Anhang 3](#) kann auch zusätzlich eine Nullhypothese geprüft werden:

Konfidenzintervall für λ einer Poisson-Verteilung ▲

Eingegebene Häufigkeit: $n = 10$

95%-Konfidenzintervall für λ : [4.7954 ; 18.3904]

← Konfidenz nach Wahl!

Literatur: L. Sachs (1992) *Angew. Statistik*, Kap. 164. Springer.

Poisson-Konfidenzintervall für die Differenz $\lambda_1 - \lambda_2$: Bei zwei Buchhaltern werden beim Schreiben von je 250 Rechnungen bei einem 6, beim anderen 16 Fehler entdeckt (Sachs 1992, p.297). In [Anhang 3](#) kann auch zusätzlich eine Nullhypothese geprüft werden:

Vergleich zweier Häufigkeiten ▲

Erste Häufigkeit: $n_1 = 6$; Zweite Häufigkeit: $n_2 = 16$

95%-Konfidenzintervall für $\lambda_1 - \lambda_2$: [-20.19 ; 0.19]

← Konfidenz nach Wahl!

Literatur: L. Sachs (1992) *Angewandte Statistik*. Springer-Verlag. Beispiel S. 297.

Sonstige Konfidenzintervalle:

In anderen Modulen finden sich weitere Konfidenzintervalle. Dies sind zum Beispiel Konfidenzintervalle für $\mu_1 - \mu_2$, Konfidenzintervalle für die Regressions- und Korrelationsrechnung, für den Median (Tukey- und Moses-Intervalle), in der Survival-Analyse (Hall-Wellner-Konfidenzbereiche), Simultane Konfidenzintervalle in der multivariaten Situation und Konfidenzintervalle für das Relative Risiko und Odds-Ratio. Diese Konfidenzintervalle werden im Zusammenhang mit den korrespondierenden Testverfahren behandelt und finden sich einschließlich Beispielen in den entsprechenden Modulen.

Bitte beachten Sie, dass alle Konfidenzintervalle auch für teststatistische Fragestellungen, insbesondere auch für Äquivalenztests verwendet werden können, worauf im Einzelnen hingewiesen wird.

▲

A.3 Vierfeldertafeln, Binomial- und Poisson-Tests



Pearson's χ^2 -Vierfeldertafel-Test

Woolf's G-Test (Log-Likelihood)

Fisher's exakter Test

McNemar-Test

Mood's Median-Test

Cohen's Kappa

Pearson's Kontingenzkoeffizient

Gart's dichotome Cross-Over-Analyse

Mantel-Haenszel-Test

Stratifizierter Mantel-Haenszel-Test

Test und Konfidenzintervall für $\theta_1 - \theta_2$ ("paired" und "unpaired")

Test und Konfidenzintervall für θ_1 / θ_2 ("paired" und "unpaired")

Äquivalenztest für θ_1 und θ_2

Binomial-Wahrscheinlichkeit θ testen, mit Konfidenzintervall

Poisson-Erwartungswert λ testen, mit Konfidenzintervall

Vergleich zweier Häufigkeiten / Poisson-Verteilungen

Beurteilung diagnostischer Tests (Spezifität, Sensitivität etc.) (A.18)

Test und Konfidenzintervall zum Vergleich zweier Likelihood-Ratios (A.18)

Test und Konfidenzintervall zum Vergleich zweier Prädiktiver Werte (A.18)

Epidemiologie (Relatives Risiko und Odds-Ratio)

χ^2 -Vierfelder-Test: Es wird untersucht, ob der Rhesusfaktor (Spalten) in irgendeiner Form vom Geschlecht (Zeilen) der Merkmalsträger abhängt Ackermann (1977, p.75). Die Nullhypothese lautet: "Es besteht kein Zusammenhang zwischen dem Rhesusfaktor und dem Geschlecht des Merkmalsträgers." Eine Modifikation des Tests *ohne* Yates-Korrektur ist über die Kontingenztafelanalyse möglich, aber nicht empfehlenswert.

Die Validierung des G-Tests erfolgte mit elem.com/~btilly/effective-ab-testing/g-test-calculator.html.

χ^2 -Vierfeldertafel-Test und Woolf's G-Test



Eingegebene Tafel aus Sachs (2003)

16	4		20
25	15		40
-----+-----			
41	19		60

Erwartete Häufigkeiten ϕ_{ij} :

13.7	6.3		20.0
27.3	12.7		40.0
-----+-----			
41.0	19.0		60.0

Chi²-Vierfeldertafel-Test (Pearson):

Chi² = 1.1650 (p = 0.280440, zweiseitig, Yates-korrigiert, stets für n<=60)

Chi² = 1.8870 (p = 0.169537, zweiseitig, nicht korrigiert, auch falls n>60)

G-Test (Woolf, Log-Likelihood-Test):

Chi² = 1.2058 (p = 0.272174, zweiseitig, Yates-korrigiert, stets für n<=30)

Chi² = 1.9786 (p = 0.159538, zweiseitig, nicht korrigiert, auch falls n>30)

95%-Konfidenzintervall für $\Delta = \Theta_1 - \Theta_2$ (Anteile "positiv" der Zeilen):
 Yates-korrigiert: [-0.0932 , 0.4432] ::: unkorrigiert: [-0.0557 , 0.4057]
 Schätzwerte: $\Theta_1^{\wedge} = 16/20 = 0.8000$, $\Theta_2^{\wedge} = 25/40 = 0.6250$, $\Delta^{\wedge} = 0.1750$

Pearson's Assoziationskoeffizient $\phi = 0.1773$ mit $p = 0.169537$ (zweiseitig)

>> ϕ liegt analog zu Pearson's Korrelationskoeffizient zwischen -1 und 1.
 >> Cohen's Empfehlung: $\phi = 0.1$ "small", $\phi = 0.3$ "medium", $\phi = 0.5$ "large"

Literatur:

- L. Sachs (2003) *Angewandte Statistik*, Springer, Kap.62ff
- B. Woolf (1957) The log likelihood ratio test. *Ann. Hum. Gen.* 21, pp.397-409.
- J.H. McDonald (2014) *Handbook of Biol. Statistics*. Sparky House Publ. Baltimore

Fisher-Test: Der exakte Test von Fisher behandelt speziell bei kleinen Fallzahlen analoge Fragestellungen wie der χ^2 -Vierfelder-Test. Die *zweiseitige* Berechnung des p-Wertes erfolgt ab Version 9.12 konventionell gemäß „Summiere die Wahrscheinlichkeit p_0 der vorliegenden und die Wahrscheinlichkeiten $p_i; p_i \leq p_0$ aller bei den gegebenen Randhäufigkeiten möglichen Tafeln“ (Agresti (1992, 2.1b)) und optional wie bis 9.11 gemäß 2.1c per Erwartungswerte, womit gelegentliche Abweichungen von Programmen bei zweiseitigen p-Werten erklärbar sind. Ein Beispiel aus Sachs (1992), Seite 478:

Fisher-Test ▲

Eingegebene Tafel aus Sachs p. 478:

2	8		10
10	4		14
12	12		24

Fisher's p = 0.018037 (exakt, ein!seitig)
 Fisher's p = 0.036075 (exakt, zweiseitig)

Literatur:

- J. Bortz, G.A. Lienert, K. Boehnke (2008) *Verteilungsfreie Verfahren in der Biostatistik*. Springer Medizin-Verlag Heidelberg.
- A. Agresti (1992) *Exact Inference for Contingency Tables*. *Statistical Science* Vol. 7, No. 1, pp. 131-77 (Methode 2.1(a), 2.1(b) (konventionell) und 2.1(c)).

McNemar-Test: Ein Arzt behandelt 38 Patienten zweimal im Abstand von einem Monat mit Aktiv und mit Placebo (Walter (1980, p.124)). Die Beurteilung der Wirkungen erfolgt anhand der Skala "stark/ gering", Eingänge sind die Behandlungen. Vgl. auch Gart's Cross-Over-Analyse und Bowker's Test. Bitte beachten Sie, dass es sich hier, im Gegensatz zum χ^2 -Test, um eine "paired" Situation handelt. Ausführliche Konfidenzintervalle werden weiter unten angegeben!

Mc-Nemar-Test ▲

Eingegebene Tafel:

9	15		24
4	10		14
13	25		38

Diskordante Beurteilungen:

Beobachtete absolute Häufigkeiten $n_{21} = 4$ und $n_{12} = 15$, Differenz = -11
 Geschätzte Wahrscheinlichkeiten $h_{21} = 0.1053$ und $h_{12} = 0.3947$, Differenz = -0.2895

Test der Nullhypothese H_0 ("Differenz diskordanter Paare ist Null"):

Approx. Test mit! Edwards-Korrektur: $\chi^2=5.26$ mit $p=0.021781$ (zweiseitig)
Approx. Test ohne Edwards-Korrektur: $\chi^2=6.37$ mit $p=0.011617$ (zweiseitig)
Exakter Binomial-Test: $p=0.009605$ (einseitig) und $p=0.019211$ (zweiseitig)

>> Vorzugsweise sollte der exakte Binomial-Test verwendet werden.

Zimmermann's Effektstärke $w^2 = \chi^2/n = 0.168$ (ohne Edwards-Korrektur), die Gesamtfallzahl $n = 38$ wird dabei berücksichtigt:

>> Bewertung der "praktischen Relevanz" der Effektstärke w^2 (nach Cohen):
>> $w^2=0.01$ "geringe", $w^2=0.10$ "mittlere" und $w^2=0.25$ "große prakt. Relevanz"

Chinn's Effektstärke f und Cohen's d nur bzgl. der Wechselfelder, die Gesamtfallzahl von $n = 38$ wird dabei **n i c h t** berücksichtigt:

Chinn's $f=\ln(OR)=\ln(15/4)=1.322$, ergibt Cohen's $d=f/\sqrt{\pi^2/3}=0.729$

>> Bewertung der Bedeutung der Effektstärken f und d nach Cohen:
>> Cohen: $d=0.2$ "geringer", $d=0.5$ "mittlerer" und $d=0.8$ "großer Effekt"

Approximative Konfidenzintervalle für die Differenz diskordanter Paare:

$P=0.90$: [-0.4782 , -0.1008] $P=0.95$: [-0.5143 , -0.0647]
 $P=0.98$: [-0.5563 , -0.0226] $P=0.99$: [-0.5849 , 0.0060]

Die Konfidenzintervalle sind aus dem χ^2 -Test **o h n e** Edwards-Korrektur abgeleitet: Der unkorrigierte Test ergibt einen p -Wert von $p=0.011617$.

Literatur:

J.L. Fleiss (2003) *Statistical Methods for Rates and Proportions*, Wiley NY
S. Chinn (2000) A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine* 19, pp.3127-31
J. Cohen (1977,1988) *Statistical power analysis for the behavioral sciences*. 1st ed. Academic Press New York, 2nd ed. Hillsdale New York, Erlbaum Assoc.
H. Zimmermann (1984) Die praktische Relevanz des McNemar-Tests. *Biometrical Journal* 26, 2, pp.219-21

Mood's Median-Test: Zwei Parallelgruppen werden in Bezug auf Leistung, Heilungserfolg o.ä. mit Hilfe eines Scores beurteilt (Sachs 1992, p. 390):

Median-Test

Eingabedatei ist ... \Valide\Sachs390.DAT
davon Variable 1 (Tab109)
und Gruppen-Var. 3 (Gruppe) = 1 2

Eingelesene Werte der Gruppe 1: 7 14 22 36 40 48 49 52
Eingelesene Werte der Gruppe 2: 3 5 6 10 17 18 20 39

Resultierende Tafel ($M =$ Median der gepoolten Stichproben = 19)

	< M	> M		
-----+-----				
Stichprobe 1:	2	6		8
Stichprobe 2:	6	2		8
-----+-----				
Gesamt:	8	8		16

Es sind keine Werte vorhanden, die gleich dem gepoolten Median M sind.

>> Cave: Die Voraussetzungen für den asymptotischen χ^2 -Test sind nicht erfüllt,
>> eine Verifizierung ist per Fisher-Freeman-Halton's Test in "Tafeln" möglich:

>> 100.0% der Erwartungshäufigkeiten sind < 5, vertretbar sind maximal 20%.

Prüfgröße: Brandt und Snedecor's $\chi^2=2.250$ mit $df=1$ und $p=0.133614$
Die Prüfgröße verwendet für 2 Gruppen Yates' Kontinuitätskorrektur.

Alternativ: Test mit Fisher's exaktem Test (zweiseitig): $p=0.131935$

Literatur:

A.M. Mood (1950) Introduction to the Theory of Statistics, McGraw-Hill
G.W. Brown, A.M. Mood (1951) On Median Tests for Linear Hypotheses. Proc. of 2nd Berkeley Symposium on Math., Statist. and Prob., Univ. California.
H. Büning, G. Trenkler (1978,1994) Nicht-parametrische Methoden, deGruyter
L. Sachs (2003) Angewandte Statistik, Springer-Verlag Kap. 392, 533.

Cohen's Kappa: Zwei Beurteiler sind an Hand einer zweistufigen Skala in Bezug auf ihre Übereinstimmung zu überprüfen (Bortz et al. (1990, p.451)). (In BiAS. steht Cohen's Kappa auch für mehr als 2 Beurteiler und mehr als 2 Stufen zur Verfügung, vgl. Sie bitte BiAS.' Menü.) Der horizontale Eingang entspricht Rater 1, der vertikale Rater 2; in 9 bzw. 4 Fälle stimmen die Beurteilungen überein:

Cohen's Kappa-Koeffizient

Bezeichnung: Bortz-Lienert-Boehnke p.451

Eingegebene Tafel:

9	3		12
4	4		8
-----+			
13	7		20

Cohen's Kappa = 0.2553

Prüfgröße Z = 1.1483 (p = 0.250829)

Literatur:

W.G.Cohen (1960) Errors of measurements. Technometrics 10, pp. 637-66
J.L.Fleiss (1973) Stat. Meth. for Rates & Propor. Wiley NY, pp. 143-47

Eine allgemeinere Variante von Cohen's Kappa findet sich im Modul "Kontingenztafeln" für den Fall von $k>2$ Ratern und/oder $s>2$ Stufen des Merkmals: Näheres in [Anhang 4](#).

Pearson's Kontingenzkoeffizient: Zwei Behandlungen (Zeilen) sind bezüglich des Therapieeffektes (Spalten) zu vergleichen (Sachs (1992, p.450)). Pearson's Kontingenzkoeffizient ist ein Maß für die Enge des Zusammenhanges der beiden kategorialen Merkmale:

Pearson's Kontingenzkoeffizient

Bezeichnung: Sachs pp. 450/476

Eingegebene Tafel:

15	85		100
4	77		81
-----+			
19	162		181

Erwartete Häufigkeiten ϕ_{ij} :

10.5	89.5		100.0
8.5	72.5		81.0
-----+			
19.0	162.0		181.0

$\chi^2 = 3.8107$ (p = 0.050927, zweiseitig)

Pearson's Kontingenzkoeffizient C = 0.2031 (mit Yates-Korrektur)
 Pearson's Kontingenzkoeffizient C = 0.2278 (ohne Yates-Korrektur)

Literatur: L. Sachs (1992) Angew. Statistik, Springer, Kap.466,621ff.

Gart's Cross-Over-Analyse: Bortz et al. (1990, p.164): In einer Stichprobe von Patienten reagieren n=19 unterschiedlich auf zwei Präparate, davon Nvp=8 Patienten in der Reihenfolge VP, Npv=11 in der Reihenfolge PV. In VP ist V achtmal, in PV ist V siebenmal und P viermal wirksam. Patienten, die auf beide Präparate gleich reagieren, gehen nicht in die Auswertung ein:

Cross-Over-Analyse (Gart) ▲

Bezeichnung: Bortz-Lienert-Boehnke pp.164-5

		Reihenfolge 'AB'		Reihenfolge 'BA'		Zusammenfassung			
		A		A		A			
		+	-	+	-	+	-		
B	+	0	0	+	0	4	+	0	4
	-	8	0	-	7	0	-	15	0

	Exakter Test (Gart)	Approximativer Test (Z (Gauß))
Haupteffekt:	p = 0.006549 (ein!s.)	p = 0.010878 (ein!s.) Z = 2.29
	p = 0.012821 (zweis.)	p = 0.021756 (zweis.) Z = 2.29
Periodeneffekt:	p = 0.085139 (ein!s.)	p = 0.094471 (ein!s.) Z = 1.31
	p = 0.103199 (zweis.)	p = 0.188941 (zweis.) Z = 1.31

Literatur:

J.J. Gart (1969) An exact test for comparing matched proportions in crossover designs. Biometrika Vol. 56, pp. 75-80.

Mantel-Haenszel-Test: Schuhmacher und Schulgen (2002, Seite 60) beschreiben den Vergleich zweier Behandlungen (Zeilen) bei Hinterwandinfarkt bzgl. "Gestorben" und "Überlebt" (Spalten). Die Daten für den Hinterwandinfarkt finden sich in der eingegebenen Tafel, die Ergebnisse können mit dem stratifizierten MH-Test bzw. mit SPSS verifiziert werden.

Mantel-Haenszel-Test ▲

Eingegebene Tafel:

3	86		89
11	69		80
-----+			
14	155		169

Prüfgröße:

Chi² = 5.9382 mit fg = 1 und p = 0.014816

Literatur:

N. Mantel (1963) Chi²-tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. JASA 58, pp. 690-700

Stratifizierter Mantel-Haenszel-Test: Schuhmacher und Schulgen (2002) beschreiben den Vergleich zweier Behandlungen, dies stratifiziert nach Hinterwandinfarkt und Vorderwandinfarkt. Die Daten können über eine Datei eingelesen werden und werden wegen des Umfangs (n=313 Patienten) hier nicht angegeben. Dieses Modul wurde anhand der Ergebnisse von Schuhmacher und Schulgen verifiziert (dort wird jedoch keine Stetigkeitskorrektur verwendet, die Resultate beim Konfidenzintervall weichen - vermutlich wegen Rundungsfehlern - geringfügig ab), ebenfalls mit SPSS, dessen Ergebnisse mit denen von BiAS. exakt übereinstimmen.

Stratifizierter Mantel-Haenszel-Test

Eingabedatei ist ..\Valide\Schuhmacher.DAT
davon Variable 1 (Therapie) und Variable 2 (Erfolg)

Stratifizierungsvariable Stratum = 1,2 (Infarktlokalisierung)

>> Der Wert x1=1 entspricht der ersten Zeile bzw. Spalte der Vierfeldertafel
>> eines Stratums, der Wert x2=2 entspricht der zweiten Zeile bzw. Spalte.

Strata-spezifische Vierfeldertafeln (Zeilen: Erfolg, Spalten: Therapie):

Tafel 1:	3	86		89	OR=0.2188
	11	69		80	p=0.014816
	-----+				
	14	155		169	

Tafel 2:	6	67		73	OR=0.7052
	8	63		71	p=0.538439
	-----+				
	14	130		144	

DerSimonian-Laird-Test auf Homogenität der ORs: $\text{Chi}^2=1.77$ (df=1): p=0.183117

Breslow-Day-Test auf Homogenität der ORs: $\text{Chi}^2=1.83$ (df=1): p=0.176622

Tarone-Test auf Homogenität der ORs: $\text{Chi}^2=1.83$ (df=1): p=0.176665

Gemeinsames gewichtetes Odds-Ratio der Strata-spezifischen ORs: OR=0.4131

95%-Konfidenzintervall für das gemeinsame OR: CI = [0.1813 ; 0.9414]

Prüfgröße $\text{Chi}^2=3.86$ (df=1): p=0.049486 (mit Kontinuitätskorrektur, empfohlen)

Prüfgröße $\text{Chi}^2=4.67$ (df=1): p=0.030618 (ohne (!) Kontinuitätskorrektur)

Literatur:

N Mantel, W Haenszel (1959) Statistical aspects of the analysis of data from retrospective studies of disease. J. National Cancer Inst. 22, pp. 719-48
J Robins, S Greenland, NE Breslow (1986) A general estimator for the variance of the Mantel-Haenszel odds ratio. Amer. J. Epidemiol. 124, 5, pp. 719-23
R DerSimonian, N Laird (1986) Meta-analysis in clinical trials. Controlled Clinical Trials, 7, pp. 177-88.
M Schuhmacher, G Schulgen (2002) Methodik klinischer Studien, Springer-Verlag

Test und Konfidenzintervall für $\theta_1-\theta_2$ ("unpaired"): In einem klinischen Versuch wird eine alte mit einer neuen Behandlung verglichen (Zeilen). Von 257 Patienten mit Behandlung A sterben 41, von 244 Patienten unter B sterben 64. Berechnet wird ein 95%-Konfidenzintervall für die Differenz der beiden Überlebensraten (Armitage und Berry (1988, p.124)):

P-100%-Konfidenzintervalle und Test für $\delta = \theta_1 - \theta_2$



Bezeichnung: Armitage und Berry pp. 124-5

Eingegebene Tafel:

41	216		257
64	180		244
-----+			
105	396		501

Erwartete Häufigkeiten $\Phi(i,j)$:

53.9	203.1		257.0
51.1	192.9		244.0
-----+			
105.0	396.0		501.0

Prüfgrößen:

Gauss' Z1 = 2.8247 (p = 0.004733) (ohne Yates-Korrektur)

Gauss' Z2 = 2.7149 (p = 0.006630) (mit! Yates-Korrektur)

Schätzungen für θ_1 und θ_2 der Zeilen:

P1 = 0.1595

P2 = 0.2623

Konfidenzintervalle für $\delta = \theta_1 - \theta_2$:

	ohne Yates-Korrektur:	mit Yates-Lorrektur:
P=90%:	(-0.1624 , -0.0431)	(-0.1664 , -0.0391)
P=95%:	(-0.1738 , -0.0317)	(-0.1778 , -0.0277)
P=99%:	(-0.1962 , -0.0094)	(-0.2002 , -0.0054)

BEMERKUNG: Die Tests können sich wegen abweichender Varianzen vom Test-Ergebnis per korrespondierendem Konfidenzintervall unterscheiden. Für die H_0 -Prüfung bei $n < 60$ immer nur den Test Z2 resp. das korrigierte Konfidenzintervall verwenden!!

Literatur: L. Sachs (1997) Angewandte Statistik, Springer, Kap.62ff.

P. Armitage, G. Berry (1987) Statistical Methods, Blackwell Sci.Pub. S.123

J.L. Fleiss (1981) Stat. Methods for Rates and Proportions. Wiley NY, p.29

Test für θ_1/θ_2 ("unpaired"): Die Fragestellung ist analog zu der der Differenz $\theta_1 - \theta_2$ des obigen Beispiels und betrachtet die Relation der beiden "ungepaarten" Binomial-Parameter. Die Validerung erfolgte mit Hilfe eines Beispiels aus dem Artikel von Koopman (1984), beachten Sie dazu bitte auch das analoge Konfidenzintervall, das im Modul "Konfidenzintervalle" verfügbar ist.

Tests für $\theta_1 - \theta_2$ und θ_1/θ_2 ("paired"): Analog wie bei der Differenz und dem Quotienten für "unpaired", jedoch mit jeweils *zwei* Binomial-Merkmalen am *gleichen* Patienten bzw. Merkmalsträger. Ein Validierungsbeispiel für die Differenz $\theta_1 - \theta_2$ findet sich weiter oben im Zusammenhang mit dem McNemar-Test, ein Validierungsbeispiel für die Relation θ_1/θ_2 bei Bonett und Price (2006).

Äquivalenztests können bei allen Fragestellungen mit Hilfe der Konfidenzintervalle nach der "Intervall-Inklusions-Regel" durchgeführt werden: Nach Definition eines "Äquivalenzbereichs" wird geprüft, ob das 90%(!)-Konfidenzintervall *gänzlich* im Äquivalenzbereich enthalten ist. Ist dies der Fall, so schließt man auf Äquivalenz zum Signifikanzniveau $\alpha = 5\%$ (!).

Ein Äquivalenztest von θ_1 und θ_2 kann auch explizit durchgeführt werden:

Äquivalenztest für θ_1 und θ_2 ("unpaired"): Vergleich von "Regular care" mit "Nurse-Practitioner" (Zeilen) bezüglich der Spalten "Adequate" und "Not adequate". Gewünscht ist ein Test auf Äquivalenz der Raten der Zeilen (Johnson (1988, p. 907)). Beim Vergleich der Ergebnisse ist die geringe Rechengenauigkeit bei Johnson zu beachten.

Äquivalenztest für θ_1 und θ_2

Bezeichnung: Johnson p.907 (cave Rechengenauigkeit!)

Eingegebene Tafel:			Erwartete Häufigkeiten ϕ_{ij} :		
148	77	225	151.0	74.0	225.0
115	52	167	112.0	55.0	167.0
-----+-----			-----+-----		
263	129	392	263.0	129.0	392.0

Äquivalenzintervall $(-\phi, +\phi)$: $\phi = 0.100000$
Differenz $\delta = P_1 - P_2 = 0.657778 - 0.688623$: $\delta = -0.030845$

Äquivalenztest mit $(1-2\alpha)$ -Intervall:

Ohne Kontinuitätskorrektur: $p = 0.073964$ (zweiseitig, nur $n > 60$!)

Mit! Kontinuitätskorrektur: $p = 0.090489$ (zweiseitig, auch $n < 60$)

Literatur: R.Johnson (1988) p-values in 2x2-tables. Biometrics Vol.44,3,pp.907-10

Epidemiologie (Relatives Risiko und Odds-Ratio): Zur Beurteilung eines Risikofaktors sind zu berechnen: Inzidenzraten, Odds Ratio, Relatives Risiko, zugehörige Konfidenzintervalle, zuschreibbares Risiko und der Mantel-Haenszel-Test. Das folgende Beispiel entstammt Sachs (1992, p.312). Bemerkung: Ein in der Epidemiologie gelegentlich erforderlicher Test auf Äquivalenz kann einfach mit Hilfe der errechneten Konfidenzintervalle vor-genommen werden: Zum Niveau z.B. $\alpha = 0.05$ verwendet man ein $P = (1-2\alpha) = 90\%$ -Konfidenzintervall und entscheidet auf Äquivalenz, wenn das Intervall gänzlich im vorgegebenen Äquivalenzbereich (z.B. $\pm 10\%$) liegt.

Epidemiologie

Bezeichnung: Sachs p. 314

	Krankheit +	Krankheit -
Faktor +	24	96
Faktor -	48	592

Pearson-Mantel-Haenszel's $\chi^2 = 18.3866$ ($p < 0.000018$)

Inzidenzrate positiv = 0.2000
Inzidenzrate negativ = 0.0750
Zuschreibbares Risiko = 0.1250
Relatives Risiko = 2.6667 'Starker Schaden'
Odds Ratio = 3.0833 'Starker Schaden'

Konfidenzintervalle:	-----90%-----	-----95%-----	-----99%-----
Relatives Risiko:	1.8305 3.8848	1.7032 4.1752	1.4794 4.8068
Odds Ratio:	2.0019 4.7491	1.8429 5.1587	1.5677 6.0643

Literatur:

L. Sachs (1992) Angewandte Statistik Kap.2, Springer-Verlag

Binominal-Wahrscheinlichkeit θ testen: Die Wahrscheinlichkeit akuter Leukämien bei Kindern bis zu 4 Jahren wird in der BRD mit $p = 0.00104$ angegeben. Bei einer neueren epidemiologischen Untersuchung im Umkreis eines Atomkraftwerks stellte man fest, dass innerhalb eines Jahres dort $k=3$ von $n=4000$ exponierten Kindern im Alter von bis zu 4 Jahren erkrankten. Ist die Inzidenzrate $p=0.00104$ damit zu vereinbaren? Da im Test $P(k \geq 3) = 0.008810$ ist, muss die Nullhypothese im einseitigen Test verworfen werden; vgl. dazu Ackermann (1995-2010):

Binomial-Wahrscheinlichkeit θ testen

Stichwort: Atomkraftwerke - Leukämieinzidenz

Anzahl "Versuche" $n = 4000$

Anzahl "Erfolge" $k = 3$

Hypothetischer Wert aus H_0 : $\theta_0 = 0.000104$

Errechnete relative Häufigkeit $k/n = 0.000750$

Einseitige Binomial-Tests der Nullhypothese $H_0(\theta=\theta_0)$:

Zur Alternative $H_a(\theta<\theta_0)$: $P(k\leq 3) = p = 0.999104$

Zur Alternative $H_a(\theta>\theta_0)$: $P(k\geq 3) = p = 0.008810$

Zweiseitiger Test der Nullhypothese $H_0(\theta=\theta_0)$:

Per Fisher's F mit Ablehnungsbereich $p=2*(p/2)$: $p = 0.017619$

$P*100\%$ -Konfidenzintervalle für θ :

$P = 0.90$: (0.000204 , 0.001937)

$P = 0.95$: (0.000155 , 0.002190)

$P = 0.98$: (0.000109 , 0.002509)

$P = 0.99$: (0.000084 , 0.002742)

Die 90%- und 98%-Konfidenzintervalle sind für Äquivalenztests mit den Irrtumswahrscheinlichkeiten $\alpha=0.05=5\%$ und $\alpha=0.01=1\%$ verwendbar!

Literatur: L.Sachs (1999) Angewandte Statistik, 3. Auflage Springer.

P.Armitage, G.Berry (1988) Stat. Methods in Medical Research, Blackwell.

Poisson-Test für den Parameter λ : In einem Zeitraum wurden $x=10$ Ereignisse beobachtet. Die Hypothese lautet, dass im Mittel $\lambda=15$ Ereignisse eintreten. Zusätzlich zum Test wird ein 95%-Konfidenzintervall berechnet, beliebige Konfidenzintervalle in [Anhang 2](#) (Sachs 1992, p. 292):

Test des Parameters λ einer Poisson-Verteilung

Nullhypothese $H_0(\lambda_0=15)$

Beobachtete absolute Häufigkeit $\lambda=10$

Poisson-Test der H_0 : $p=0.118464$ ein!seitig und $p=0.236929$ zweiseitig

95%-Konfidenzintervall: [4.7954 ; 18.3904]

Literatur: L. Sachs (1992) Angew. Statistik, Kap. 164. Springer.

Vergleich zweier Häufigkeiten/Poisson-Verteilungen: Bei zwei Buchhaltern werden beim Schreiben von je 250 Rechnungen bei einem 6, beim anderen 16 Fehler entdeckt. Unterscheiden sich die beiden Buchhalter "signifikant"? Sachs (1992, p.297):

Vergleich zweier Häufigkeiten

Erste Häufigkeit: $n_1 = 6$; Zweite Häufigkeit: $n_2 = 16$

95%-Konfidenzintervall für die Differenz n_1-n_2 : [-20.19 ; 0.19]

Prüfgröße $\chi^2 = 3.682$ ($p = 0.055009$ zweiseitig)

Exakter Binomial-Test: ($p = 0.026239$ ein!seitig)

Exakter Binomial-Test: ($p = 0.052479$ zweiseitig)

Literatur: L. Sachs (1992) Angewandte Statistik. Springer-Verlag. Beispiel S. 297.

A.4 Kontingenztafeln



χ^2 -Test für eine cxr-Kontingenztafel
Exakter Fisher-Freeman-Halton-Test
Haldane-Dawson-Test
Craddock-Flood-Test
Yates-Cochran-Test auf Trend
Mantel-Haenszel-Test
Cohen's Kappa-Koeffizient (mehrere Stufen, 2 Rater)
Cohen's Kappa-Koeffizient (mehrere Stufen, mehrere Rater)
Pearson's Kontingenzkoeffizient
Terwilliger und Ott's Marginal-Symmetrietest
Bowker's Kontingenztafel-Symmetrietest

χ^2 -Test und Fisher-Freeman-Halton-Test für eine cxr-Kontingenztafel: Vergleich von "Carriers" und "Non- Carriers" (Zeilen) von Streptococcus pyogenes, klassifiziert nach Größe der Tonsillen ("Vorhanden, nicht vergrößert", "vergrößert", "stark vergrößert" (Spalten)). Das Beispiel entstammt Armitage und Berry (1988, p.373):

cxr-Kontingenztafel mit χ^2 -Test

Eingegebene Tafel: Armitage und Berry Seite 373

19	29	24		72
497	560	269		1326
-----+-----				
516	589	293		1398

Prüfgrößen:

$\chi^2 = 7.8848$ mit $fg = 2$ und $p = 0.019402$

Exakter Test und p-Wert mit Valz und Thompson's Algorithmus: $p = 0.023721$

Literatur: L. Sachs (1992) Angew. Statistik, Springer 1992, Kap.62.

Haldane-Dawson-Test: Der Test dient der Untersuchung von schwach besetzten Tafeln bei mehr als 5 Zeilen bzw. Spalten. Bortz, Lienert und Boehnke (1990, p.138) geben dazu ein Beispiel an: In einer Bibliothek für 6 verschiedene Studienrichtungen (Zeilen) befinden sich 8 Statistiklehrbücher (Spalten). Leihen bestimmte Fachrichtungen bestimmte Bücher häufiger aus als andere?

Haldane und Dawson's Test

Eingegebene Tafel: Bortz-Lienert-Boehnke Seite 138

1	0	0	0	4	0	0	0		5
0	1	0	3	4	0	1	2		11
0	2	0	2	0	2	1	2		9
2	5	4	0	0	0	1	4		16
1	0	0	0	4	0	3	0		8
0	1	1	0	3	1	0	0		6
-----+-----									
4	9	5	5	15	3	6	8		55

Prüfgröße:
Haldane-Dawson's U = 3.2171
p = 0.000647 (stets einseitig!)

Literatur:

J.S.B. Haldane (1940) The mean and variance of χ^2 , when used as a test of homogeneity, when samples are small. Biometrika 31, pp.346-55.

Craddock-Flood-Test: Der Test dient der Untersuchung von schwach besetzten Tafeln bei Vorliegen von bis zu 5 Zeilen/Spalten (vgl. auch Haldane-Dawson-Test). Bortz, Lienert und Boehnke (1990, p.139) geben dazu ein Beispiel: Es soll überprüft werden, welche von 3 Futtermischungen von 4 untersuchten Hunderassen bevorzugt wird. (Hier wird nur H_0 ("kein Zusammenhang") geprüft. Einzelne Felder können mit Hilfe der Konfigurationsfrequenzanalyse untersucht werden!) Die eingegebene Tafel weist zu geringe Besetzungszahlen auf und ist deshalb für eine Analyse mit dem χ^2 -Test ungeeignet.

Craddock und Flood's Test ▲

Bezeichnung: Bortz-Lienert-Boehnke p.139

Eingegebene Tafel:

0	5	3		8
3	1	3		7
4	0	6		10
0	0	4		4
-----+-----				
7	6	16		29

Prüfgröße:

Craddock-Flood's $\chi^2 = 16.8942$
mit fg = 6 , n = 29 und p(konservativ) = 0.009680

Falls nicht signifikant, bitte folgende Tabelle verwenden:

n=	16	18	20	25	30	35	40	50	60	'∞'
$\alpha=0.050$:	11.9	12.0	12.1	12.2	12.3	12.4	12.4	12.5	12.6	12.6
$\alpha=0.010$:	15.1	15.3	15.5	15.7	15.9	16.1	16.2	16.4	16.5	16.8
$\alpha=0.001$:	20.0	20.2	20.4	20.7	20.8	20.9	21.0	21.2	21.4	22.4

Literatur:

J.M. Craddock, C.R. Flood (1970) The distribution of χ^2 statistic in small contingency tables. Applied Statistics, 19, pp.173-81.

Yates-Cochran-Test auf Trend: Vergleich dreier Therapieformen (Spalten: Symptomatisch, Spezifisch-Normaldosis, Spezifisch-Doppeldosis) bezgl. Therapeutischem Erfolg (Zeilen: Geheilt in x Wochen, geheilt in x+y Wochen, Gestorben). Weiterhin interessiert eine Prüfung auf Trend ("Korrelation"). Das Beispiel entstammt Sachs (1992, p.596):

cxr-Kontingenztafel mit Yates-Cochran-Test auf Trend ▲

Bezeichnung: Sachs pp. 595/606

Eingegebene Tafel:

14	22	32		68
18	16	8		42
8	2	0		10
-----+-----				
40	40	40		120

Zeilenscores: 1 2 3
 Spaltenscores: 1 2 3

Folgende Felder (mehr als 1/5) haben Erwartungshäufigkeit $\varphi < 5$:

Feld	Erwartete Häufigkeit
(3 , 1)	3.33
(3 , 2)	3.33
(3 , 3)	3.33

Variationsursache	Chi-Quadrat	fg	p
Lineare Regression	20.2935	1	0.0000
Abweichung von der Regression	1.2829	3	0.7332
Gesamtvariation	21.5765	4	0.0002

Literatur: L. Sachs (1992) Angew. Statistik, Springer 1992, Kap.62

Mantel-Haenszel-Test: De Jonge (1983, p.155) beschreibt den Vergleich zweier z.B. Behandlungsgruppen mit ordinalskalierten Zielgröße, die nur wenige Stufen besitzt. In der nachfolgend angegebenen Tafel beziehen sich die Zeilen auf die beiden Behandlungsgruppen, die Spalten auf die 5 Stufen der Zielgröße. Die Daten der Tafel stellen die absoluten Häufigkeiten dar, die in der jeweiligen Behandlung-Merkmalstufen-Kombination beobachtet wurden. Die Nullhypothese lautet "Die beiden Behandlungen/Bedingungen unterscheiden sich nicht in Bezug auf den Behandlungserfolg":

Mantel-Haenszel-Test ▲

Bezeichnung: De Jonge p.155ff

Eingegebene Tafel:

1	13	16	15	7	52
5	21	14	9	3	52
6	34	30	24	10	104

Spaltenscores: 1 2 3 4 5

Prüfgröße:

$\chi^2 = 7.3085$ mit $fg=1$ und $p = 0.006863$

Literatur:

N. Mantel (1963) χ^2 -tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. JASA 58, pp. 690-700.

Cohen's Kappa-Koeffizient: Die Fragestellung lautet analog zum Vierfelder-Fall. Zwei Sozialarbeiter A und B beurteilen $N=100$ jugendpsychiatrische Patienten, ob deren Störung als "Verwahrlosung", "Neurose" oder "Psychose" einzustufen ist. Zu beurteilen ist die Übereinstimmung der beiden Sozialarbeiter A und B. (Bortz et al. (1990, p.459)); Cohen's Kappa κ ist ein standardisiertes Maß für die Übereinstimmung der beiden "Rater" und nimmt Werte zwischen 0 und 1 an; "1" bedeutet eine totale Übereinstimmung. (In **BiAS**. steht Cohen's Kappa auch für mehr als 2 Beurteiler zur Verfügung, vgl. Sie bitte **BiAS**.' Menü.) Die Ausgabe des „Weighted Kappa“ wurde hier nicht aufgenommen:

cxr-Kontingenztafel mit Cohen's Kappa-Koeffizient



Bezeichnung: Bortz-Lienert-Boehnke pp.459-60

Eingegebene Tafel:

53	5	2		60
11	14	5		30
1	6	3		10
-----+				
65	25	10		100

Cohen's Kappa = 0.428571
Prüfgröße Z = 5.45906
mit p = 0.000000

90%-Konfidenzintervall für Kappa: [0.2994 ; 0.5577]

95%-Konfidenzintervall für Kappa: [0.2747 ; 0.5824]

99%-Konfidenzintervall für Kappa: [0.2263 ; 0.6308]

Literatur:

W.G. Cohen (1960) Errors of Measurement in Statistics. Technom. 10, pp. 637-666.

J.L. Fleiss (1973) Stat. Methods for Rates and Proportions pp.143-147. Wiley NY.

Pearson's Kontingenzkoeffizient: Analog zum Vierfelder-Fall sind zwei oder mehr Behandlungen (Zeilen) bezüglich des Therapieeffektes (Spalten) zu vergleichen. Pearson's Kontingenzkoeffizient ist ein Maß für die Enge des Zusammenhanges der beiden kategorialen Merkmale. Das Beispiel entstammt Sachs (1992, p.476):

cxr-Kontingenztafel mit Pearson's Kontingenzkoeffizient



Bezeichnung: Sachs p.476

Eingegebene Tafel:

15	85		100
4	77		81
-----+			
19	162		181

Prüfgröße: $\chi^2 = 4.8222$ mit $fg=1$ und $p = 0.028096$

Pearson's Kontingenzkoeffizient C = 0.2278

Literatur:

L.Sachs (1992) Angewandte Statistik, Springer. Kap. 622.

K. Pawlik (1959) Der maximale Kontingenzkoeffizient im Falle nicht-quadratischer Kontingenztafeln. Metrika 2, pp.150-166.

Terwilliger und Ott's Symmetrie-Test: Der Test prüft auf Homogenität bzw. Symmetrie der Marginalverteilungen zweier abhängiger Verteilungen und findet zum Beispiel in der Genetik häufig Anwendung. In Ermangelung eines Beispiels aus der Literatur wird hier ein per Hand gerechnetes Beispiel angegeben, die Daten entstammen Sachs (1999, p.607):

Terwilliger-Ott-Marginalsymmetrie-Test



Bezeichnung: Sachs p.607

Eingegebene Tafel:

0	10	16	15		41
4	2	10	4		20
12	4	3	6		25
8	4	1	1		14
-----+					
24	20	30	26		100

Prüfgröße:

Chi² = 8.5007

mit fg = 3

und p = 0.036721

Literatur: J. Terwilliger, J. Ott (1992) A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. Hum. Hered. 42, pp. 337-346.

Bowker's Kontingenztafel-Symmetrietest: Analog zum McNemar-Test wird eine Kontingenztafel auf Symmetrie untersucht. Die verwendeten Daten entstammen Sachs (1999, p. 607):

Bowker's Kontingenztafel-Symmetrietest



Bezeichnung: Sachs p.607

Eingegebene Tafel:

0	10	16	15		41
4	2	10	4		20
12	4	3	6		25
8	4	1	1		14
-----+					
24	20	30	26		100

Prüfgröße für Diagonale 'links oben nach rechts unten':

Chi² = 11.4161, FG=6 und p = 0.076335

Prüfgröße für Diagonale 'links unten nach rechts oben':

Chi² = 15.2000, FG=6 und p = 0.018757

Literatur: A.H. Bowker (1948) A test for symmetry in contingency tables. J. Amer. Stat. Ass. 43, pp. 572-574.

A.5 Konfigurationsfrequenzanalyse (KFA) ▲

Einstichproben-Konfigurationsfrequenzanalyse
 Mehrstichproben-Konfigurationsfrequenzanalyse
 KFA zur Clusteranalyse binärer und kategorialer Daten
 KFA zur Diskrimination binärer und kategorialer Daten

Einstichproben-Konfigurationsfrequenzanalyse: N=570 Studenten werden nach ihrer Studienrichtung (1) und ihrer Einstellung zur Gentechnologie (2) befragt. Es fragt sich, ob typische Konfigurationen beobachtet werden können: Das Beispiel ist Bortz et al. (1990, p.155/6) entnommen.

Die Eingabe der Daten erfolgt in Form der absoluten Häufigkeiten der Zellen der Tafel, die durch Angabe der Merkmale und deren Stufen definiert wird. Die Indizierung bei Eingabe der "Konfiguration..." bezieht sich auf die Stufen der gewählten Merkmale/Eingänge der Tafel; im Verlauf der Eingabe läuft der letzte Merkmalsstufenindex des zuletzt definierten Merkmals "am schnellsten", der des zuerst definierten Merkmals "am langsamsten". Dies entspricht der Ausgabe der Ergebnisse:

Konfigurationsfrequenzanalyse KFA

Konf 1 2	Absol. Häufigkeiten	Erwart. Häufigkeiten	Prägn. unkorr	Prägn. Betzin	Typ	Chi ²	p Test	Chi ² - Test	p Binom. Test	p Leh- macher
1 1	135	111.7	0.0509	0.1228	T	4.863	0.027444	0.013924	0.000076	
1 2	108	131.3	0.0531	0.1283	A	4.136	0.041970	0.020433	0.000076	
2 1	31	32.6	0.0030	0.0073	A	0.082	0.774710	0.768160	0.677569	
2 2	40	38.4	0.0031	0.0074	T	0.070	0.791794	0.784591	0.677569	
3 1	57	69.9	0.0257	0.0621	A	2.370	0.123724	0.100314	0.014557	
3 2	95	82.1	0.0264	0.0637	T	2.016	0.155686	0.124884	0.014557	
4 1	39	47.8	0.0169	0.0407	A	1.621	0.202917	0.183423	0.055615	
4 2	65	56.2	0.0171	0.0414	T	1.379	0.240251	0.216118	0.055615	
Sum	570	570.0				16.536	0.000880	; fg = 3		

```
>> Prägnanz uncorr: Unkorr. Werte Q der Prägnanz ("determination coefficient")
>> Prägnanz Betzin: Maximum-korrigierte Werte Q'=Q/Qmax<=1 gemäß Betzin et al.
>> Alle p-Werte beziehen sich auf z w e i seitige Tests.
>> Zur konfirmatorischen Analyse ist eine  $\alpha$ -Korrektur (z.B. Bonferroni, hier per
>>  $\alpha^*=\alpha/8$  bzw.  $p'=8*p$ ) der p-Werte durchzuführen!
```

Literatur:

J. Krauth, A. Lienert (1973) Die Konfigurationsfrequenzanalyse. Karl Alber Freiburg
 J. Bortz et al. (2008) Verteilungsfreie Methoden. 3. Aufl. Springer-Verlag
 W. Lehmaner (1981) Simultaneous test procedure in configural frequency analysis.
 Biom. J. 23, 5, pp. 429-436.
 J. Betzin, P. Bollmann-Sdorra (2003) On the determination coefficient in CFA.
 Psychology Science 45, 2, pp.400-20

Vorsicht: Der Freiheitsgrad der Tafel wird bei Bortz et al. nicht korrekt angegeben; richtig ist hier der Freiheitsgrad $df=(4-1)\cdot(2-1)=3$. Zur Validierung des Lehmaner- und Binomial-Tests vgl. Lehmaner (1981) (im Test wurden Lehmaners Prüfgrößen Z_{ijk} reproduziert, im Programm werden p-Werte angegeben) und Krauth-Lienert (1973).

In der angegebenen Tafel bedeutet ein *Typ* eine Zelle mit einer höheren beobachteten als erwarteten Häufigkeit, ein *Antityp* ist umgekehrt definiert als eine unterfrequentierte Zelle. Eine Beurteilung erfolgt optimal mit Hilfe des Lehmaner-Tests entsprechend der p-Werte in der letzten Spalte (bitte aber auch

die Prägnanzmaße beachten!), wobei im Falle einer konfirmatorischen Analyse eine α -Korrektur zum Beispiel nach Bonferroni oder günstiger nach Bonferroni-Holm durchgeführt werden sollte.

Mehrstichproben-Konfigurationsfrequenzanalyse: Zwei Stichproben (hier Männer und Frauen) werden anhand der Merkmale Suizidmotivation (1) (Somatische Erkrankung, Psychische Erkrankung, Alkoholismus) und dem Suiziderfolg (2) klassifiziert. Nullhypothese ist, dass Männer und Frauen bezüglich der sechs Merkmals-kombinationen homogen sind. Das Beispiel von Bortz et al. (1990, p.158) findet sich auch in dem Buch von Krauth und Lienert zur KFA. Die Dateneingabe (absolute Häufigkeiten der Zellen) entspricht der Einstichproben-KFA, wobei die Häufigkeiten der hier 2 Stichproben "nebeneinander" eingegeben werden:

Konfigurationsfrequenzanalyse KFA ▲

Bezeichnung: Bortz-Lienert-Boehnke p. 158

Konf		Absolute Häufigkeiten		
1	2	Stichpr 1	Stichpr 2	Summe
1	1	86.00	64.00	150.00
1	2	16.00	18.00	34.00
2	1	61.00	76.00	137.00
2	2	25.00	47.00	72.00
3	1	47.00	7.00	54.00
3	2	27.00	8.00	35.00
Sum		262.00	220.00	482.00

Konf		Erwartete Häufigkeiten		
1	2	Stichpr 1	Stichpr 2	Summe
1	1	81.54	68.46	150.00
1	2	18.48	15.52	34.00
2	1	74.47	62.53	137.00
2	2	39.14	32.86	72.00
3	1	29.35	24.65	54.00
3	2	19.02	15.98	35.00
Sum		262.00	220.00	482.00

Konf		Prüfgrößen χ^2 mit df=1		
1	2	Stichpr 1	Stichpr 2	Summe
1	1	0.24	0.29	0.54
1	2	0.33	0.40	0.73
2	1	2.44	2.90	5.34
2	2	5.11	6.08	11.19
3	1	10.61	12.64	23.25
3	2	3.34	3.98	7.32
Sum		22.07	26.29	48.36

Prüfgröße $\chi^2 = 48.36$ mit $df = 5$ ($p = 0.000000$)

Konf		p-Werte zu Prüfgrößen χ^2 mit df=1		
1	2	Stichpr 1	Stichpr 2	Summe
1	1	0.620988	0.589482	0.464247
1	2	0.563812	0.528774	0.392918
2	1	0.118574	0.088518	0.020875
2	2	0.023836	0.013661	0.000823
3	1	0.001125	0.000378	0.000001
3	2	0.067487	0.046007	0.006802

Zur konfirmatorischen Analyse ist eine α -Korrektur (z.B. Bonferroni, $\alpha^*=\alpha/6$ für die "Summe") der p-Werte zum Chi²-Test durchzuführen!

Literatur:

J. Bortz et al. (1990) Verteilungsfreie Methoden. Springer.
 J. Krauth, A. Lienert (1973) Die Konfigurationsfrequenzanalyse. Verlag Karl Alber Freiburg, München.

Clustering kategorialer und binärer Daten: Der Datensatz (Lautsch/Lienert 1993) gibt die beobachteten und die unter der Cluster-Nullhypothese erwarteten Häufigkeiten von Mustern aus t=5 Ja-Nein-Symptomen bei n=150 psychosomatisch Depressiven wieder. Die erwarteten Häufigkeiten errechnen sich mit $e=n/2^t$. Z.B. weist der erste p-Wert des Musters "+++++" ein Cluster von 12 Patienten aus, die an allen 5 Symptomen leiden.

Clustering kategorialer Daten ▲

Konf	Absol. Häufigkeiten	Erwart. Häufigkeiten	Chi ² -Test	p	Chi ² -Test
1 2 3 4 5					
+	12	4.69	11.408	0.000731	***
+	4	4.69	0.101	0.750832	
+	7	4.69	1.141	0.285476	
+	1	4.69	2.901	0.088534	
+	7	4.69	1.141	0.285476	
+	2	4.69	1.541	0.214494	
+	7	4.69	1.141	0.285476	
+	1	4.69	2.901	0.088534	
+	0	4.69	4.688	0.030383	*
+	2	4.69	1.541	0.214494	
+	1	4.69	2.901	0.088534	
+	0	4.69	4.688	0.030383	*
+	0	4.69	4.688	0.030383	*
+	2	4.69	1.541	0.214494	
+	0	4.69	4.688	0.030383	*
+	0	4.69	4.688	0.030383	*
+	7	4.69	1.141	0.285476	
+	4	4.69	0.101	0.750832	
+	11	4.69	8.501	0.003550	**
+	7	4.69	1.141	0.285476	
+	7	4.69	1.141	0.285476	
+	8	4.69	2.341	0.126022	
+	9	4.69	3.967	0.046387	*
+	17	4.69	32.341	0.000000	***
+	2	4.69	1.541	0.214494	
+	1	4.69	2.901	0.088534	
+	2	4.69	1.541	0.214494	
+	9	4.69	3.967	0.046387	*
+	0	4.69	4.688	0.030383	*
+	5	4.69	0.021	0.885234	
+	4	4.69	0.101	0.750832	
+	11	4.69	8.501	0.003550	**

Chi²-Werte heuristisch mit df=1 prüfen!

Literatur: E. Lautsch, G.A. Lienert (1993) Binärdatenanalyse. Beltz-Verlag

Diskrimination kategorialer und binärer Daten: Der Datensatz (Krauth (1993, p. 127) gibt die beobachteten Häufigkeiten von Mustern aus t=5 Ja-Nein-Symptomen bei n1=6117 schwangeren und n2=7448 nicht-schwangeren Frauen. Es fragt sich, ob sich die beiden Gruppen bezüglich der 5 Symptome unterscheiden bzw. ob es für die eine und/oder die andere Gruppe typische sog. "Diskriminanztypen" gibt. Da es sich um ein exploratives Verfahren handelt, wurden die Fisher'schen exakten p-Werte *keiner* alpha-Adjustierung (zum Beispiel nach Bonferroni-Holm) unterzogen, die aber - so bei Krauth (1993) - prinzipiell denkbar ist, um keine "Artefakte" von Diskriminanztypen zu erhalten.

KFA: Diskriminanzanalyse binärer und kategorialer Daten ▲

Konfigurat	Absolute Häufigkeiten		Fisher's p	Zuordnung zu
1 2 3 4 5	Stichpr 1	Stichpr 2	(einseitig)	Stichprobe
-----	-----	-----	-----	-----
+ + + + +	0.00	9.00	0.004525	2**
+ + + + -	0.00	43.00	0.000000	2***
+ + + - +	0.00	74.00	0.000000	2***
+ + + - -	8.00	590.00	0.000000	2***
+ + - + +	0.00	0.00	1.000000	2
+ + - + -	2.00	9.00	0.064361	2
+ + - - +	1.00	18.00	0.000186	2***
+ + - - -	64.00	224.00	0.000000	2***
+ - + + +	0.00	5.00	0.049870	2*
+ - + + -	7.00	36.00	0.000065	2***
+ - + - +	0.00	30.00	0.000000	2***
+ - + - -	59.00	517.00	0.000000	2***
+ - - + +	1.00	2.00	0.573359	2
+ - - + -	36.00	23.00	0.009959	1**
+ - - - +	8.00	18.00	0.100503	2
+ - - - -	1144.00	636.00	0.000000	1***
- + + + +	0.00	6.00	0.027373	2*
- + + + -	4.00	57.00	0.000000	2***
- + + - +	0.00	95.00	0.000000	2***
- + + - -	27.00	954.00	0.000000	2***
- + - + +	2.00	0.00	0.203329	1
- + - + -	14.00	32.00	0.030467	2*
- + - - +	4.00	33.00	0.000008	2***
- + - - -	187.00	526.00	0.000000	2***
- - + + +	1.00	4.00	0.254768	2
- - + + -	17.00	59.00	0.000031	2***
- - + - +	0.00	48.00	0.000000	2***
- - + - -	162.00	1247.00	0.000000	2***
- - - + +	6.00	2.00	0.089276	1
- - - + -	82.00	35.00	0.000000	1***
- - - - +	20.00	56.00	0.000567	2***
- - - - -	4261.00	2060.00	0.000000	1***
-----	-----	-----	-----	-----
Sum	6117.00	7448.00		

Fisher's exact Test: *:p<0.05, **:p<0.01, ***:p<0.001
 Es wurde **keine** Alpha-Adjustierung durchgeführt!

Literatur:

- E. Lautsch, S. von Weber (1995) Methoden und Anwendungen der KFA. Beltz.
- J. Krauth (1993) Einführung in die KFA. Beltz.

A.6 Tests auf Gauß-Verteilung



χ^2 -Anpassungstest
Shapiro-Wilk-Test
Kolmogoroff-Smirnoff-Lilliefors-Test
Mudholkar-Test (multivariat)
David-Test ("Schnelltest")
Grubbs-Test (Ausreißer-Test)
Graphik zum χ^2 -Anpassungstest
Graphik zum Kolmogoroff-Smirnoff-Test
Quantil-Quantil-Plot ("QQ-Plot")

χ^2 -Anpassungstest: Es ist zu beurteilen, ob eine Stichprobe denkbarer Weise einer Gauß-Verteilung entstammt; der Test wird im Programm durch eine Graphik der Häufigkeiten ergänzt (Histogramm und angepasste Gauß-Verteilung). Zu den verwendeten Daten von Graf et al. (1987, p.393) ist zu bemerken, dass der χ^2 -Test bei kleinen Stichproben an sich kontraindiziert ist; besser sollte in solchen Fällen Shapiro-Wilk's Test oder Kolmogoroff-Smirnoff's Test verwendet werden.

χ^2 -Anpassungstest

Eingabedatei ist Graf393.DAT

Untersucht wird Variable 1

Eingelesene Daten:

470 490 478 484 484 464 448 473 453 453 461

Intervall	Beobachtete Häufigkeit	Erwartete Häufigkeit
1	3.0000	1.4751
2	2.0000	2.7096
3	2.0000	3.0326
4	3.0000	2.0679
5	1.0000	0.8591

Prüfgröße Chi² = 2.5571
Freiheitsgrad fg = 2
und p = 0.278436

Literatur: L. Sachs (1992) Angew. Statistik, Springer-Verlag, Kap.43

Shapiro-Wilk-Test: Es ist zu beurteilen, ob eine Stichprobe denkbarer Weise einer Gauß-Verteilung entnommen ist. Der Shapiro-Wilk-Test ist der vermutlich günstigste Test für diese Fragestellung. Die Daten zum Test finden sind in Graf et al. (1987, p.393):

Shapiro-Wilk-Test



Eingabedatei ist Graf393.DAT

Untersucht wird Variable 1

Eingelesene Daten:

470 490 478 484 484 464 448 473 453 453 461

Shapiro-Wilk's Prüfgröße $W = 0.9473$ ($p > 0.100000$ exakter Wert ($n \leq 50$))
Royston's Prüfgröße $Z(W) = -0.2198$ ($p = 0.586977$ approximativ ($n \geq 7$))

Literatur:

S.S. Shapiro, M.B. Wilk (1965) An analysis of variance test for normality (complete samples). Biometrika 52, pp.591-611.

J.P. Royston (1982) An extension of Shapiro and Wilk's test for normality to large samples. Appl. Stat. 31, 2, pp. 115-24.

Kolmogoroff-Smirnoff-Lilliefors-Test: Es ist zu beurteilen, ob eine Stichprobe möglicherweise einer Gauß-Verteilung entstammt. Der Test wird im Programm durch eine testanaloge Graphik der kumulativen Häufigkeiten ergänzt. Ein Test für *bekannte* μ und σ^2 (die ursprüngliche Version von Kolmogoroff & Smirnoff) ist ebenfalls möglich. Hier ein Beispiel mit Daten aus Graf et al. (1987, p.393):

Kolmogoroff-Smirnoff-Lilliefors-Test ▲

Eingabedatei ist Graf393.DAT
Untersucht wird Variable 1

Eingelesene Daten:

470 490 478 484 484 464 448 473 453 453 461

Berechnetes Arithmetisches Mittel $X_{\text{quer}} = 468.909088$

Berechnete Standardabweichung $SD = s = 14.250997$

Kolmogoroff-Smirnoff-Lilliefors' delta = 0.140591 : $p > 0.100000$ (tabellarisch)

Dallal-Wilkinson-korrigierter p-Wert zum KSL-Test : $p = 0.978178$ (empfohlen)

Literatur:

A. Kolmogoroff (1941) Confidence limits for an unknown distribution function. Ann. Math. Stat. 12, pp. 461-463.

N. Smirnoff (1948) Tables for estimating the goodness of fit of empirical distributions. Ann. Math. Stat. 19, pp. 279-281.

H.W. Lilliefors (1967) On the Kolmogoroff-Smirnoff test for normality with mean and variance unknown. J.Am.Stat.Ass. 62, pp. 399-402.

G.E. Dallal, L. Wilkinson (1986) An Analytic Approximation to the Distribution of Lilliefors' Test Statistic for Normality. Am. Stat. 40, pp. 294-6.

Mudholkar-Test: Es sollen 15 Studenten bezüglich 4 Variablen beurteilt werden. Zur Prüfung der Voraussetzung „Gauß-Verteilung“ für den intendierten Hotelling's T^2 -Test (vgl. weiter unten) wird ein 4-variater Mudholkar-Test durchgeführt. Die Daten und die Fragestellung ist Kramer (1972, p.52) entnommen; es ist zu bemerken, dass die Zielgrößen in Kramer's Beispiel qualitativer Natur sind, so dass eher an eine Auswertung mit Friedman's Rang-Varianzanalyse gedacht werden sollte.

Mudholkar-Test ▲

Eingabedatei ist Kramer54A.DAT

davon Variablen 5(Verbal_0), 6(MathSAT0), 7(English0), 8(MathAch0)

i	Var. 5	Var. 6	Var. 7	Var. 8
1	-28.0000	-26.0000	-17.0000	18.0000
2	-146.0000	-20.0000	-45.0000	-59.0000
3	-15.0000	20.0000	-143.0000	12.0000
4	3.0000	26.0000	66.0000	-1.0000
5	137.0000	75.0000	79.0000	-38.0000
6	-126.0000	-63.0000	-128.0000	-69.0000
7	-67.0000	-20.0000	-149.0000	-22.0000
8	65.0000	-50.0000	102.0000	34.0000
9	-100.0000	-67.0000	-144.0000	-48.0000
10	51.0000	-3.0000	6.0000	28.0000
11	16.0000	-60.0000	60.0000	47.0000

12	3.0000	26.0000	-67.0000	-30.0000
13	-8.0000	29.0000	-45.0000	66.0000
14	9.0000	7.0000	-15.0000	45.0000
15	86.0000	44.0000	123.0000	-11.0000

Mudholkar-Test auf 4-variate Gauß-Verteilung:
 Prüfgröße N = 1.3472 (p = 0.177909) (asymptotisch)

Marginalverteilungen (univariate Kolmogoroff-Smirnoff-Tests)
 5: Kolmogoroff-Smirnoff's S = 0.1125 (p > 0.20 (Dallal-korrigiert))
 6: Kolmogoroff-Smirnoff's S = 0.1169 (p > 0.20 (Dallal-korrigiert))
 7: Kolmogoroff-Smirnoff's S = 0.1406 (p > 0.20 (Dallal-korrigiert))
 8: Kolmogoroff-Smirnoff's S = 0.0850 (p > 0.20 (Dallal-korrigiert))

Literatur:
 G.S. Mudholkar et al. (1992) A test of p-variate normality. Biometrika 79, 4, pp. 850-4.

Der Originalartikel von Mudholkar et al. enthält kein Beispiel, so dass zur Überprüfung des Programms Per-Hand-Berechnungen durchgeführt wurden.

David-Test: Es ist zu beurteilen, ob eine Stichprobe möglicherweise aus einer Gauß-Verteilung entnommen ist. Der David-Test wird in **BiAS**. zum raschen Überblick über die gesamte Datensituation einer Datei verwendet, bei speziellen Fragestellungen empfehlen sich eher andere Testverfahren, z.B. Shapiro-Wilk's Test oder Kolmogoroff-Smirnoff's Test. Der Datensatz des Beispiels von Graf et al. (1987, p.393) umfasst nur eine Variable, in umfangreicheren Datensätzen werden alle oder eine selektierte Anzahl von Variablen bzw. Spalten - auch unter Berücksichtigung vielleicht vorhandener Gruppierungsvariablen - getestet:

David-Test ▲

Eingabedatei ist Graf393.DAT
 Untersucht wird Variable 1

Eingelesene Daten:
 470 490 478 484 484 464 448 473 453 453 461

Var.	n	David	p
1	11	2.9472	>0.10

Literatur:
 H.A. David, H.O. Hartley, E.S. Pearson (1954) The distribution of the ratio, in a single normal sample, of range to standard deviation. Biometrika pp.482-493
 E.S. Pearson, M.A. Stephens (1964) The ratio of range to standardeviation in the same normal sample. Biometrika 51, pp. 484-487.

Grubbs-Test: Es ist zu beurteilen, ob eine Stichprobe Ausreißerwerte enthält. Der Grubbs-Test kann versagen, wenn nicht nur ein Ausreißer, sondern eine Gruppe von Ausreißern existiert. In den Daten von Graf et al. (1987, p.391) wird der Wert x=527 als Ausreißer entdeckt:

Grubbs-Test ▲

Eingabedatei ist Graf391.DAT
 Untersucht wird Variable 1

Bemerkung: Die Datei Graf391.DAT ist die gleiche Datei wie Graf393:dat (vgl. z.B. Wilk-Shapiro-Test), jedoch ergänzt durch den 4. Wert X=527 als "Ausreißer"!

1: 470 7: 464
 2: 490 8: 448
 3: 478 9: 473
 4: 527 10: 453
 5: 484 11: 453
 6: 484 12: 461

Durchschnitt = 473.7500
 Standardabweichung = 21.5833
 Spannweite = 79.0000
 Fallzahl n = 12

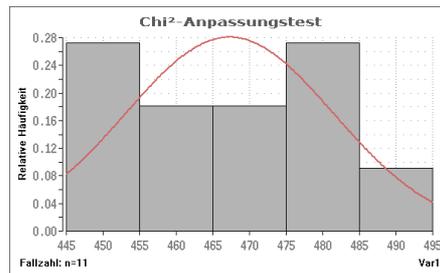
	Nr.	Wert	Prüfgröße	p (1s.)	p (2s.)
Minimum:	8	448.0000	1.1930	1.000000	1.000000
Maximum:	4	527.0000	2.4672	0.017687	0.035374

p(1s.): Einseitiger p-Wert & p(2s.): Zweiseitiger p-Wert
 Zweiseitiger Test: $p = \text{Minimum}\{p(2s.)\} = 0.035374$

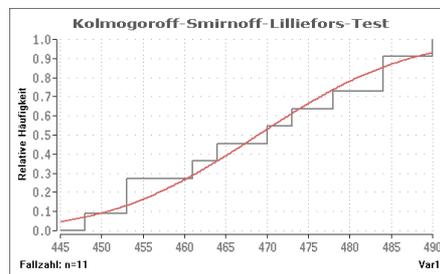
Literatur:

F.E. Grubbs, G. Beck (1972) Extension of sample sizes and percentage points for significance tests of outlying observations. Technom. 14, 4, pp. 847-54

Graphik einer angepassten Gauß-Verteilung: Es wurden die Daten aus der Datei Graf.393 verwendet (Berechnung vgl. Anpassungstest). Die Abweichungen der empirischen von der theoretischen Kurve bilden die Testgrundlage. (**Vorsicht:** möglichst "große" Fallzahlen verwenden!)



Graphik einer angepassten Gauß-Verteilungsfunktion: Es wurden die Daten aus Datei Graf.393 verwendet. Die maximale Abweichung der Kurven entspricht der Prüfgröße des Kolmogoroff - Smirnof - Tests (alle Berechnungsergebnisse in **BiAS**. via Lilliefors-Dallal-Wilkinson-Variante).



Die Skalierungsintervalle speziell in der Darstellung der Dichtefunktion der Gauß-Verteilung können optional frei gewählt werden. In der Voreinstellung benutzt **BiAS**. die Sturges'sche Regel zur Festlegung der Anzahl Intervalle.

QQ-Plot zur Anpassung an eine Gauß-Verteilung: Das QQ-Plot kann zur graphischen Bewertung der Anpassung von Daten an eine Normal- oder - je nach Einstellung im Dialog, Radio-Buttons "Typ" - an eine Log-Normalverteilung verwendet werden. η^2 („Eta²“) ist der Anteil der aufklärten Varianz an der Gesamtvarianz; zur Bewertung der Größenordnung von η^2 kann man die - wenn auch nur heuristisch begründete - Faustregel verwenden: "Mindestens 85% kann als "akzeptabel", mindestens 90% als "gut" und mindestens 95% als "sehr gut" bewertet werden. Optimal ist natürlich ein Wert von η^2 nahe bei dem Maximalwert von 1. Prinzipiell erscheint es sinnvoll zu sein, zur Entscheidungsfindung einen objektiven statistischen Test heranzuziehen: Im Programm ist dazu etwa der Kolmogoroff-Smirnoff-Lilliefors-Test verfügbar.

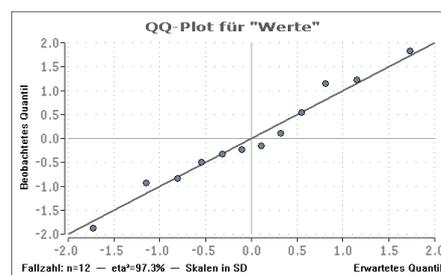
Zur Quantil-Transformation können unterschiedliche Standardverfahren gewählt werden. Gängig sind die beiden Methoden "Rankit" und "Blom".

Zur Verifikation wurde ein Beispiel von Persike (2010) verwendet, die Daten finden sich unten. „Werte“ sind Persikes Testdaten, „TrnWerte“ ist zur Überprüfung der Log-Normalverteilung exponentiell transformiert.

Eine textliche Ausgabe ist nicht vorgesehen, eta² findet sich in der Graphik.

Datei ..\Valide\Persike.DAT ▲

Nr.:	Werte	TrnWerte
1:	44.00	81.4509
2:	61.00	445.8578
3:	48.00	121.5104
4:	68.00	897.8473
5:	25.00	12.1825
6:	60.00	403.4288
7:	41.00	60.3403
8:	37.00	40.4473
9:	45.00	90.0171
10:	36.00	36.5982
11:	43.00	73.6998
12:	53.00	200.3368



Literatur:

Persike, M. (2010) Vorlesung "Mathematische und statistische Methoden II" des Psychologischen Instituts der Universität Mainz.

Gibbons, J., Chakraborti, S. (2010) Nonparametric statistical inference, 5th ed. CRC Press - A Chapman and Hall Book.



Einfache lineare Regression mit Graphik
Vergleich zweier Regressionsgeraden
Hahn-Prognoseintervall mit Graphik
Multiple Regression mit Abbau-Verfahren
Multiple Regression mit additiven und multiplikativen Termen
Logistische Regression mit Abbau-Verfahren
Polynomiale Regression mit Graphik und Abbau-Verfahren
Zirkuläre Regression und Korrelation mit Graphik
Einfache lineare Korrelation
Vergleich zweier unabhängiger Korrelationskoeffizienten
Vergleich zweier abhängiger Korrelationskoeffizienten
Trivariate Multiple Korrelation (quantitativ und ordinal)
Trivariate Partielle Korrelation (quantitativ und ordinal)
Rang-Regression (X:quantitativ, Y:ordinal)
Rang-Korrelation: Spearman's rho und Kendall's tau
Pearson-Korrelationsmatrix
Spearman-Korrelationsmatrix
Intracluster-Korrelation (quantitativ und dichotom)
Intraclass-Korrelation (in [A.15](#))
Bland-Altman-Regression (in [A.18](#))
Passing-Bablok-Regression (in [A.18](#))
Lin's Konkordanz-Korrelation (in [A.18](#))
Cox-Regression (in [A.10](#))
Punktbiseriale Korrelation dichotom vs. quantitativ oder ordinal

Einfache lineare Regression: Hängt die Gewichtszunahme (Y="IncrWeight") von Neugeborenen vom Geburtsgewicht (X="Birthweight") ab? Die Daten entstammen Armitage und Berry (1988, p.148), eine Graphik (weiter unten) ist möglich:

Einfache lineare Regression

Name der Eingabedatei: Armitage.148

Modell: $Y = \beta_0 + \beta_1 X_1$

Zielgröße Y = Variable 2 = IncrWei

Einflussgröße X = Variable 1 = BirthWei

Eingelesene Daten:

i	Y	X
1	68	72
2	63	112
3	66	111
4	72	107
5	52	119
6	75	92
7	76	126
8	118	80

9	120	81
10	114	84
11	29	115
12	42	118
13	48	128
14	50	128
15	69	123
16	59	116
17	27	125
18	60	126
19	71	122
20	88	126
21	63	127
22	88	86
23	53	142
24	50	132
25	111	87
26	59	123
27	76	133
28	72	106
29	90	103
30	68	118
31	93	114
32	91	94

Anzahl fehlender Werte: 0

Regression Y auf X:

Regressionskoeffizient b =	-0.8643
Konstanter Achsenabschnitt c =	167.8701
Testgröße t der Regression t =	4.9198
Überschreitungswahrsch. p =	0.000029
Stichprobenumfang n =	32
Anzahl der Freiheitsgrade n-2 =	30

Konfidenzintervalle für β_1 (Steigung):

P = 0.95:	(-1.2231, -0.5055)
P = 0.99:	(-1.3475, -0.3812)

Konfidenzintervalle für β_0 (Achsenabschnitt):

P = 0.95:	(127.2633, 208.4769)
P = 0.99:	(113.1913, 222.5489)

Regression X auf Y:

Regressionskoeffizient b =	-0.5166
Konstanter Achsenabschnitt c =	148.5760
Testgröße t der Regression t =	4.9198
Überschreitungswahrsch. p =	0.000029
Stichprobenumfang n =	32
Anzahl der Freiheitsgrade n-2 =	30

Durchschnitte, Quadratsummen und Streuungen:

Summe der Quadrate SQxx =	10262.0000
Summe der Quadrate SQyy =	17168.4688
Summe der Produkte SPxy =	-8869.7500
Durchschnitt Xquer =	111.7500
Durchschnitt Yquer =	71.2813
Streuung S ² x =	331.0323
Streuung S ² y =	553.8216
Kovarianz S ² xy =	-286.1210
Reststreuung S ² x.y =	189.3206
Reststreuung S ² y.x =	316.7361

Literatur: L. Sachs (1992) Angewandte Statistik. Springer, Kap. 51

Vergleich zweier Regressionskoeffizienten: Es wurde die Wirkung des Gonadotropingehaltes zweier Harnextrakte auf das Uterusgewicht von Mäusen untersucht (Linder und Berchtold (1982, p.191)). Da bei derartigen Versuchen die Wirkung in aller Regel linear mit dem Logarithmus der Dosis ansteigt, wird bei den folgenden Berechnungen der Logarithmus dualis verwendet. Die hier einzugebenden Kenngrößen wurden mit Hilfe der Regressionrechnung ermittelt; die zugehörigen Daten finden sich im Anschluss.

Vergleich zweier Regressionskoeffizienten ▲

	1. Regression	2. Regression
Stichprobenumfänge n:	15	15
Steigungen b:	259.0000	218.0000
Achsenabschnitte c:	336.9330	112.0670
Durchschnitte Xquer:	0.0000	1.0000
Durchschnitte Yquer:	336.9330	330.0670
Streuungen S ² x:	0.7143	0.7143
Streuungen S ² y:	56304.5000	44603.6016
Reststreuungen S ² (um R.):	9034.8213	11477.7090

Homogenität der Varianzen (zweiseitig):

Prüfgröße F = 1.2704 mit fg = (14, 14) (p = 0.660446)

Test von Ho($\beta_{11}=\beta_{12}$) für die Steigungen β_1 :

Gleiche Restvarianzen:

Prüfgröße t = 0.9053 mit fg = 26 (p = 0.373639)

Ungleiche Restvarianzen, n₁,n₂>20:

Prüfgröße Z = 0.9053 (p = 0.365327)

Ungleiche Restvarianzen, n₁<20 oder n₂<20:

Prüfgröße t = 0.9053 mit fg = 26 (p = 0.373639)

Falls parallel: Test von Ho($\beta_{01}=\beta_{02}$) für die Achsenabschnitte β_0 :

Prüfgröße t = 5.6585 mit fg = 26 (p = 0.000006)

Konfidenzintervalle für $\beta_{11}-\beta_{12}$ (Steigungen):

P = 0.95: (-52.0965, 134.0965)

P = 0.99: (-84.8502, 166.8502)

Falls parallel: Konfidenzintervalle für $\beta_{01}-\beta_{02}$ (Achsenabschnitte):

P = 0.95: (156.2329, 334.4992)

P = 0.99: (124.8736, 365.8584)

Falls nicht parallel: Konfidenzintervalle für Schnittpunkt (Aszisse!):

VORSICHT: Die Fieller-Konfidenzintervalle sind nur gültig, wenn der Schnittpunkt S=-5.4845 der Regressionsgeraden im Bereich des untersuchten Intervalls der Abszisse liegt.

Zur Orientierung (Xquer,SD pooled): S = Xquer -7.09 * SD (p < 1E-6!!)

P = 0.95: (-1.1510, 5.0310)

P = 0.99: (-0.7212, 3.1424)

Literatur:

L. Sachs (1992) Angewandte Statistik. Springer, Kap. 558.

A.Linder, W. Berchtold (1982) Stat. Methoden, Birkhäuser, p.190ff.

Daten zum Vergleich zweier Regressionsgeraden: Daten aus Linder und Berchtold (1982, p.191):

Editor: Ausdrucken ▲

Die Datei Linder.191 umfasst 3 Spalten bzw. 30 Zeilen (.3Kb).

Variablenamen der Datei Linder.191:

1:Dosis 2:Ld(Dosis) 3:Wirkung 4:Extrakt

Ausgabe der Datei Linder.191:

```
Nr.:  1  2  3  4      Nr.:  1  2  3  4
-----
 1: 0.5 -1  83  1      16: 1.0  0 164  2
 2: 0.5 -1  84  1      17: 1.0  0 255  2
 3: 0.5 -1  90  1      18: 1.0  0  64  2
 4: 0.5 -1  84  1      19: 1.0  0  95  2
 5: 0.5 -1  75  1      20: 1.0  0 154  2
 6: 1.0  0 560  1      21: 2.0  1 150  2
 7: 1.0  0 268  1      22: 2.0  1 350  2
 8: 1.0  0 372  1      23: 2.0  1 275  2
 9: 1.0  0 247  1      24: 2.0  1 122  2
10: 1.0  0 185  1      25: 2.0  1 410  2
11: 2.0  1 710  1      26: 4.0  2 487  2
12: 2.0  1 516  1      27: 4.0  2 525  2
13: 2.0  1 620  1      28: 4.0  2 585  2
14: 2.0  1 510  1      29: 4.0  2 600  2
15: 2.0  1 650  1      30: 4.0  2 715  2
```

Polynomiale Regression mit Abbau-Verfahren: In Hartung et al. (1982, p.592) wird die Wirkung einer Stimulanz (X) auf das Leistungsvermögen (Y) untersucht. Es wird im Ansatz ein kubischer, also polynomialer Zusammenhang unterstellt. Die angepasste Funktion kann optional nach Abfrage auch als Graphik dargestellt werden.

Polynomiale Regression mit Abbau



Eingabedatei ist Hartung592.DAT

Zielgröße Y = Variable 2 = Leistung
Einflussgröße X = Variable 1 = DosisSti

Modell: Leistung = b(0) + b(1)*DosisSti^1 + b(2)*DosisSti^2 + b(3)*DosisSti^3

```
  i  Y  X
  2  1
-----
 1 3.2 0.0
 2 5.0 0.2
 3 4.7 0.4
 4 5.8 0.6
 5 7.1 0.8
 6 7.3 1.0
 7 6.5 1.2
 8 7.5 1.4
 9 6.9 1.6
10 6.4 1.8
11 5.3 2.0
12 5.8 2.2
13 3.9 2.4
14 3.5 2.6
15 1.7 2.8
16 0.6 3.0
```

Residuen-Ausreißer-Test (Grubbs/David): p>0.10: OK!

Tabelle der Varianzanalyse (vollständiges Modell):

	SQ	df	S ²	F	p
Auf der Regression:	5.9277E+01	3	1.9759E+01	8.4292E+01	0.000000
Um die Regression:	2.8129E+00	12	2.3441E-01		
Total:	6.2090E+01	15			

Multiples Bestimmtheitsmaß B = 0.9547
 Multipler Korrelationskoeffizient R = 0.9771

Prüfung der partiellen Regressionskoeffizienten b(i):

i	b(i)	s(b(i))	F	df1	df2	p
0	3.30075					
1	6.16048	1.1627E+00	2.8074E+01	1	12	0.000189
2	-2.57428	9.1837E-01	7.8574E+00	1	12	0.015952
3	0.07136	2.0096E-01	1.2611E-01	1	12	0.728664

Abbau des Modells: Vollständiger Abbau für alpha=0 !

Modellabbau: Die Einflussgröße Nr. 3 des Modells wird jetzt eliminiert:

Multiples Bestimmtheitsmaß B = 0.9542
 Multipler Korrelationskoeffizient R = 0.9768

Prüfung der partiellen Regressionskoeffizienten b(i):

i	b(i)	s(b(i))	F	df1	df2	p
0	3.37868					
1	5.78740	4.8104E-01	1.4474E+02	1	13	0.000000
2	-2.25315	1.5468E-01	2.1219E+02	1	13	0.000000

Abbau beendet: Noch vorhandene Variablen signifikant bei alpha=0.01!

Achtung: Die Numerierung der Koeffizienten bezieht sich auf die Reihenfolge der Summanden im Modell, nicht auf die Numerierung in der Datei!!

Literatur:

M.M. Tatsuoka (1971) Multivariate Analysis. John Wiley.
 B. Flury, H. Riedwyl (1983) Angew. multivariate Statistik, Fischer.

Multiple Regression: Hartung et al. (1982, p.599) untersuchen wird die Bearbeitungszeit (Y) eines Werkstückes in Abhängigkeit vom Durchmesser (X1) und der Länge (X2) der Werkstücke:

Multiple Regression

Name der Eingabedatei: Hartung.599

Modell: $Y = X1 \cdot \beta(1) + \beta(0) + \beta(2) \cdot X2 + \beta(3) \cdot X3$
 Zielgröße Y = Variable 1 = BearbZei
 Einflussgrößen X = Variable 2 = Durchmes
 Variable 3 = Länge

i	Y	X1	X2	X3
		1	2	3
1	24	4	31	
2	28	6	3	
3	25	5	7	
4	13	3	29	
5	27	5	27	
6	14	3	18	
7	15	3	28	
8	32	6	17	
9	14	4	8	
10	28	5	12	

Tabelle der Varianzanalyse:

	SQ	df	S ²	F	p
Auf der Regr.:	4.3303E+02	2	2.1651E+02	4.3335E+01	p = 0.000114
Um die Regr.:	3.4974E+01	7	4.9963E+00		
Total:	4.6800E+02	9			

Bestimmtheitsmaß B = 0.9253

Prüfung der Koeffizienten b(i):

i	b(i)	F	df	p
0	-12.37866			
1	6.82892	7.9333E+00	7	0.000046
2	0.24063	7.5785E+00	7	0.028384

Literatur: M.M. Tatsuoka (1971) *Multivariate Analysis*. John Wiley.

Abbaumodell zur Regression: Flury und Riedwyl (Angew. Multivariate Statistik, Fischer 1983) betrachten die Abhängigkeit der entwickelten Wärme beim Härten von Zement in Abhängigkeit von der Materialzusammensetzung in Form von vier Parametern ("Hall-Daten"). Es ist die Frage, ob weniger als eben diese vier Parameter zur Erklärung der Wärmeentwicklung ausreichen. Als Abbruchkriterium wird die Signifikanzschwelle $\alpha=0.05$ für die partiellen Regressionskoeffizienten verwendet:

Multiple Regression: Abbauverfahren

Eingabedatei ist Flury.Dat

Zielgröße Y = Variable 1 = Y
 Einflussgrößen X = Variable 2 = X1
 Variable 3 = X2
 Variable 4 = X3
 Variable 5 = X4

Modell: Y = b(0) + b(1)*X1 + b(2)*X2 + b(3)*X3 + b(4)*X4

Residuen-Ausreißer-Test (Grubbs/David): p>0.10: OK!

Tabelle der Varianzanalyse (vollständiges Modell):

	SQ	df	S ²	F	p
Auf der Regression:	2.6679E+03	4	6.6697E+02	1.1148E+02	0.000000
Um die Regression:	4.7864E+01	8	5.9830E+00		
Total:	2.7158E+03	12			

Multipl. Bestimmtheitsmaß B = 0.9824

Multipler Korrelationskoeffizient R = 0.9911

Prüfung der partiellen Regressionskoeffizienten b(i):

i	b(i)	s(b(i))	F	df1	df2	p
0	62.40535					
1	1.55110	7.4477E-01	4.3375E+00	1	8	0.070822
2	0.51017	7.2379E-01	4.9682E-01	1	8	0.500901
3	0.10191	7.5471E-01	1.8234E-02	1	8	0.895923
4	-0.14406	7.0905E-01	4.1280E-02	1	8	0.844072

Abbau des Modells: Eliminationskriterium ist alpha=0.05 !

Modellabbau: Die Einflussgröße Nr. 3 des Modells wird jetzt eliminiert:

Multipler Bestimmtheitsmaß B = 0.9823
Multipler Korrelationskoeffizient R = 0.9911

Prüfung der partiellen Regressionskoeffizienten b(i):

i	b(i)	s(b(i))	F	df1	df2	p
0	71.64831					
1	1.45194	1.1700E-01	1.5401E+02	1	9	0.000001
2	0.41611	1.8561E-01	5.0259E+00	1	9	0.051687
4	-0.23654	1.7329E-01	1.8633E+00	1	9	0.205395

Modellabbau: Die Einflussgröße Nr. 4 des Modells wird jetzt eliminiert:

Multipler Bestimmtheitsmaß B = 0.9787
Multipler Korrelationskoeffizient R = 0.9893

Prüfung der partiellen Regressionskoeffizienten b(i):

i	b(i)	s(b(i))	F	df1	df2	p
0	52.57735					
1	1.46831	1.2130E-01	1.4652E+02	1	10	0.000000
2	0.66225	4.5855E-02	2.0858E+02	1	10	0.000000

Abbau beendet: Noch vorhandene Variablen signifikant bei alpha=0.05!

Achtung: Die Numerierung der Koeffizienten bezieht sich auf die Reihenfolge der Summanden im Modell, nicht auf die Numerierung in der Datei!!

Literatur:

M.M. Tatsuoka (1971) Multivariate Analysis. John Wiley.
B. Flury, H. Riedwyl (1983) Angew. multivariate Statistik, Fischer.

Hahn-Prognoseintervall: Abhängigkeit der Gewichtszunahme ("IncreWei") vom Geburtsgewicht ("BirthWei") von Neugeborenen. Mit Hilfe der Hahn-Prognose-Intervalle lassen sich mit einer gewünschten Konfidenz P zu gegebenen Geburtsgewichten die zugehörigen Gewichtszunahmen prognostizieren (Armitage und Berry (1988, p.148)). Eine Graphik ist möglich.

Hahn-Prognose-Intervall

Name der Eingabedatei: Armitage.148

Modell: $Y = X_2 = \beta(0) + \beta(1) * X_1$

Zielgröße Y = Variable 2 = IncreWei
Einflussgröße X = Variable 1 = BirthWei

Eingelesene Daten:

i	Y	X
	2	1
1	68	72
2	63	112
3	66	111
4	72	107
5	52	119
6	75	92
7	76	126
8	118	80

9 120 81
 10 114 84
 11 29 115
 12 42 118
 13 48 128
 14 50 128
 15 69 123
 16 59 116
 17 27 125
 18 60 126
 19 71 122
 20 88 126
 21 63 127
 22 88 86
 23 53 142
 24 50 132
 25 111 87
 28 72 106
 29 90 103
 26 59 123
 27 76 133
 30 68 118
 31 93 114
 32 91 94

Regression Y auf X:

Regressionskoeffizient $b = -0.8643$
 Konstanter Achsenabschnitt $c = 167.8701$
 Reststreuung $S^2_{y.x} = 316.7361$
 Testgröße t der Regression $t = 4.9198$
 Überschreitungswahrsch. $p = 0.000029$
 Stichprobenumfang $n = 32$
 Anzahl der Freiheitsgrade $n-2 = 30$

Konfidenzintervalle für β_1 :

$P = 0.95: (-1.2231, -0.5055)$
 $P = 0.99: (-1.3475, -0.3812)$

Konfidenzintervalle für β_0 : ▲

$P = 0.95: (127.2633, 208.4769)$
 $P = 0.99: (113.1913, 222.5489)$

Durchschnitte und Quadratsummen:

Summe der Quadrate $SQ_{xx} = 10262.0000$
 Summe der Quadrate $SQ_{yy} = 17168.4688$
 Summe der Produkte $SP_{xy} = -8869.7500$
 Durchschnitt $X_{\text{quer}} = 111.7500$
 Durchschnitt $Y_{\text{quer}} = 71.2813$
 Streuung $S^2_x = 331.0323$
 Streuung $S^2_y = 553.8216$
 Kovarianz $S^2_{xy} = -286.1210$

Hahn's 95%-Prognose-Intervall:

X	Y^{\wedge}	Untere Grenze	Obere Grenze
70.000000	107.367004	67.533180	147.200836
80.000000	98.723709	60.095810	137.351608
90.000000	90.080406	52.354523	127.806290
100.000000	81.437111	44.287189	118.587036
110.000000	72.793816	35.878536	109.709091
120.000000	64.150520	27.122078	101.178963
130.000000	55.507225	18.020966	92.993484
140.000000	46.863930	8.587559	85.140305
150.000000	38.220634	-1.158146	77.599411

Literatur:

L. Sachs (1992) Angewandte Statistik. Springer, Kap. 51

G.J. Hahn (1972) Simultaneous prediction intervals for a regression model. Technometrics 14, pp. 203-214

Einfache lineare Korrelation: Untersucht wird der Zusammenhang zwischen zwei gleichberechtigten, quantitativen Größen X1 und X2. Sachs (1992, p.522) gibt dazu ein Beispiel an:

Einfache lineare Korrelation ▲

Name der Eingabedatei: Sachs.522

Modell: $Y = X1 = \beta(0) + \beta(1) * X2$

Zielgröße Y = Variable 1

Einflussgröße X = Variable 2

Eingelesene Daten:

```
i  Y  X
   1  2
-----
1 13 12
2 17 17
3 10 11
4 17 13
5 20 16
6 11 14
7 15 15
```

Anzahl fehlender Werte: miss=0

Korrelation:

```
Korrelationskoeffizient r =          0.7087
Bestimmtheitsmaß B = r2 =          0.5023
Testgröße der Korrelation =          2.2464
Überschreitungswahrsch. p =          0.074612
Stichprobenumfang n =                7
Anzahl der Freiheitsgrade n-2 =        5
```

Effektstärke - Bewertung des Korrelationskoeffizienten |r| nach Evans (1996):
<0.2:Poor - 0.2-0.4:weak - 0.4-0.6:moderate - 0.6-0.8:strong - >0.8:optimal

Konfidenzintervalle für rho:

P = 0.95: (-0.0951, 0.9531)

P = 0.99: (-0.3827, 0.9744)

Durchschnitte und Quadratsummen:

```
Summe der Quadrate SQxx =      28.0000
Summe der Quadrate SQyy =      77.4286
Summe der Produkte SPxy =      33.0000
Durchschnitt Xquer =          14.0000
Durchschnitt Yquer =          14.7143
Streuung S2x =                4.6667
Streuung S2y =                12.9048
Kovarianz S2xy =              5.5000
```

Literatur:

L. Sachs (1992) Angewandte Statistik. Springer, Kap. 51

J.D. Evans (1996) Straightforward Statistics for the Behavioral Sciences. Brooks/Cole Publishing, Pacific Grove, Calif.

G.S. Mudholkar, M. McDermott, D.K. Srivastava (1992) A test of p-variate normality. Biometrika 79, 4, pp. 850-4.

Vergleich zweier unabhängiger Korrelationskoeffizienten: Dieses Modul dient dem Vergleich zum Beispiel zweier "Strata" (zum Beispiel Frauen und Männer) bezüglich der Korrelation zweier Größen. Ist die Enge des Zusammenhanges in der einen Gruppe die gleiche wie in der anderen Gruppe? Sachs (1999, Kap. 5) gibt ein gerechnetes Beispiel dazu an.

Falls die verglichenen Koeffizienten nicht signifikant verschieden sind, kann u.a. ein Konfidenzintervall für den mittleren Korrelationskoeffizienten angegeben werden.

Vergleich zweier unabhängiger Korrelationskoeffizienten ▲

Stichwort: Sachs (1999, Kap.5/552)

Korrelationskoeffizient r_1 = 0.6000
Stichprobenumfang n_1 = 28.0000

Korrelationskoeffizient r_2 = 0.8000
Stichprobenumfang n_2 = 23.0000

Prüfgröße $N = 1.351550$

Einseitige Fragestellung: $p = 0.088260$
Zweiseitige Fragestellung: $p = 0.176519$

Falls **k e i n** signifikanter Unterschied zwischen r_1 und r_2 besteht:

Schätzung des gemeinsamen Korrelationskoeffizienten $r = 0.703074$

Prüfgröße $N = 5.858636$

Einseitige Fragestellung: $p = 0.000000$
Zweiseitige Fragestellung: $p = 0.000000$

90%-Konfidenzintervall: (0.556780 ; 0.807065)

95%-Konfidenzintervall: (0.523522 ; 0.822833)

99%-Konfidenzintervall: (0.453718 ; 0.850328)

Literatur: L. Sachs (1999) Angewandte Statistik Kap. 5/552, Springer-Verlag.

Vergleich zweier abhängiger Korrelationskoeffizienten: Zum Vergleich zweier abhängiger Korrelationskoeffizienten sind prinzipiell zwei Situationen zu unterscheiden:

1) „Overlapped“: Ein Korrelationskoeffizient berechnet sich für zwei Variablen A und B, der andere für A und C. Beispiel: Vergleich der Korrelation des IQ von Sohn und Mutter (r_{12}) mit der Korrelation Sohn und Vater (r_{13}) in der gleichen Familie (i.e. Stichprobe!). r_{23} bedeutet den Korrelationskoeffizienten von Mutter und Vater. Zur Verifikation der Berechnungen unten und zur Erläuterung kann das Beispiel aus psych.unl.edu/psycrs/statpage/biv_corr_comp_eg.pdf verwendet werden.

2) „Non-overlapped“: Der erste Korrelationskoeffizient berechnet sich für die Variablen A und B, der andere für die beiden Variablen C und D. Beispiel: Vergleich der Korrelation des IQ von Vater und Mutter (r_{12}) mit der Korrelation Sohn und Tochter (r_{34}) in der gleichen Familie, also in der gleichen Stichprobe. Eine Verifikation ist mit den Daten aus Raghunathan et al. (1996) möglich.

Vergleich zweier abhängiger Korrelationskoeffizienten ▲

Stichwort: Daten aus dem oben genannten Link!

Korrelationskoeffizient $r(1,2) = -0.2050 <<$

Korrelationskoeffizient $r(1,3) = 0.2850 <<$

Korrelationskoeffizient $r(2,3) = -0.4940$

Stichprobenumfang $n = 405$

>> Vergleich der beiden Korrelationskoeffizienten $r(1,2)$ und $r(1,3)$:

Steiger's Prüfgröße $Z = -5.7733$
Zweiseitige Fragestellung: $p = 0.000000$

>> Steiger's Test verwendet Fisher's Z-Transformation, empfohlen.

Vergleich zweier abhängiger Korrelationskoeffizienten

Stichwort: Daten aus Raghunathan (1996)

Korrelationskoeffizient $r(1,2) = 0.3800 <<$
Korrelationskoeffizient $r(3,4) = 0.2500 <<$
Korrelationskoeffizient $r(1,3) = 0.4500$
Korrelationskoeffizient $r(1,4) = 0.5300$
Korrelationskoeffizient $r(2,3) = 0.3100$
Korrelationskoeffizient $r(2,4) = 0.5500$

Stichprobenumfang $n = 603$

>> Vergleich der beiden Korrelationskoeffizienten $r(1,2)$ und $r(3,4)$:

Raghunathan's Prüfgröße $Z = 2.8692$
Zweiseitige Fragestellung: $p = 0.004115$

>> Raghunathan's Test verwendet Fisher's Z-Transformation $r(1,2)$ und $r(3,4)$

J.H. Steiger (1980) Tests for comparing elements of a correlation matrix. Psychological Bulletin 87, 245-51.

T.E. Raghunathan, R. Rosenthal, D.B. Rubin (1996) Comparing correlated but nonoverlapping correlations. Psychological Methods, Vol 1(2), 178-83.

B. Weaver, K.L. Wuensch (2013) SPSS and SAS programs for comparing Pearson correlation coefficients. Behaviour Research Methods Vol. 45, 3, 880-95.

Trivariate Multiple Korrelation: Es wird untersucht, ob die Intelligenz eines Schülers auf Grund der Gedächtnisleistung und der Deutschnote "vorhergesagt" werden kann (Bortz 1989, p.559).
Vorsicht: Da *alle* Parameter eher nicht als quantitativ, sondern als ordinal einzustufen sind, sollte hier die Rang-Korrelation (weiter unten) verwenden werden!

Multiple Korrelation

Name der Eingabedatei: Bortz559.DAT

Zielgröße Y = Variable 3 = Intellig
Einflussgrößen X = Variable 1 = Gedächtn
Variable 2 = NoteDeut

Multiple Korrelation zwischen y =Intellig und den beiden Variablen x =Gedächtn, e =NoteDeut

Eingelesene Daten:

i	Y	X X	i	Y	X X
	3	1 2		3	1 2
1	107	12 2	6	97	13 4
2	105	12 3	7	92	12 4
3	101	13 3	8	118	10 1
4	102	10 4	9	111	14 2
5	114	11 2	10	95	15 3

```

Korrelationskoeffizient R-y.xe =      0.9348
Bestimmtheitsmaß B = R2 =          0.8738
Testgröße F der Korrelation =        24.2377
Überschreitungswahrsch. p =        0.000714
Stichprobenumfang n =                10
Freiheitsgrade (k,n-k-1) =          ( 2 , 7)

```

Vergleich mit den Bivariaten Korrelationen:

```

r-yx: Testgr. F mit df=(1, 7) =      36.4043
      Überschreitungswahrsch. p =    0.000524
r-ye: Testgr. F mit df=(1, 7) =       6.0903
      Überschreitungswahrsch. p =    0.042968

```

Paarweise Korrelationen:

```

r-yx =   -0.4665   ( p = 0.174145 )
r-ye =   -0.8741   ( p = 0.000942 )
r-xe =    0.1594   ( p = 0.659940 )

```

Literatur: L. Sachs (1997) Angewandte Statistik. Springer, Kap. 58.

Trivariate Multiple Rang-Korrelation: Diese erfolgt analog zur multiplen parametrischen Korrelation, arbeitet jedoch mit Rangdaten. Eine Verifikation ist – nach Rangtransformation – mit der genannten Pearson-Korrelation möglich. Vorsicht: Bortz und Lienert (2008) verwenden in ihrem Beispiel *keine* Tie-Korrektur, wodurch sich von BiAS. abweichende Resultate ergeben!

Literatur: Kendall M.G., J.D. Gibbons (1990) Rank correlation methods. Arnold Publ. Ltd. London
 Bortz J., G.A. Lienert (2008) Kurzgefasste Statistik für die klinische Forschung. 3. Auflage Springer

Trivariate Partielle Korrelation: Es wird untersucht, ob die Note in Deutsch und die Intelligenz eines Schülers korreliert sind, wenn man den Einfluss des Gedächtnisse "ausschaltet". Es ergibt sich (mit den Daten von Bortz (1989, p.559) ein bemerkenswertes Ergebnis! Wegen des ordinalen Charakters aller Parameter sollte jedoch die Partielle Rang-Korrelation verwendet werden. Die rechnerische Überprüfung des Moduls wurde mit Daten aus L.Sachs (1997, Kap. 58) vorgenommen.

Partielle Korrelation ▲

Name der Eingabedatei: Bortz559.DAT

```

Variablen X1 = Variable2 = NoteDeut
           X2 = Variable1 = Intellig
Störgröße E = Variable3 = Gedächtn

```

Partielle Korrelation zwischen NoteDeut und Intellig,
 "bereinigt" von Einfluss der Stör-Variablen Gedächtn

Eingelesene Daten:

```

  i  x1   E   x2
-----
  1  2   12  107
  2  3   12  105
  3  3   13  101
  4  4   10  102
  5  2   11  114
  6  4   13   97
  7  4   12   92
  8  1   10  118
  9  2   14  111
 10  3   15   95

```

```

Korrelationskoeffizient r-xy.e =    -0.9158
Bestimmtheitsmaß B = r² =          0.8387
Testgröße t der Korrelation =      6.0336
Überschreitungswahrsch. p =      0.000524
Stichprobenumfang n =              10
Anzahl der Freiheitsgrade n-3 =    7

```

Paarweise Korrelationen:

```

r-yx =    -0.8741    ( p = 0.000942 )
r-ye =     0.1594    ( p = 0.659940 )
r-xe =    -0.4665    ( p = 0.174145 )

```

Literatur: L. Sachs (1997) Angewandte Statistik. Springer, Kap. 58.

Trivariate Partielle Rang-Korrelation: Diese erfolgt analog zur partiellen parametrischen Korrelation, arbeitet jedoch mit Rangdaten. Eine Verifikation ist – nach Rangtransformation, Editor-Modul! – mit der genannten Pearson-Korrelation möglich. Vorsicht: Bortz und Lienert (2008) verwenden in ihrem Beispiel *keine* Tie-Korrektur, wodurch sich von BiAS. abweichende Resultate ergeben!

Literatur: Kendall M.G., J.D. Gibbons (1990) Rank correlation methods. Arnold Publ. Ltd. London
 Bortz J., G.A. Lienert (2008) Kurzgefasste Statistik für die klinische Forschung. 3. Auflage Springer

Rang-Regression: Lienert (1973, p.685) diskutiert in einem Rechenbeispiel die Abhängigkeit der Reaktionszeit (Y) vom gemessenen Blutalkoholgehalt (X): Y ist bei Lienert ein ordinales Merkmal!

Rang-Regression



Name der Eingabedatei: Lienert.685

Modell: $Y = X^2 = \beta(0) + \beta(1) * X1$

Zielgröße Y = Variable 2 = Reaktion

Einflussgröße X = Variable 1 = Promille

Eingelesene Daten:

```

i   Y   X
  2   1
-----
1  2  0.4
2  4  0.5
3  1  0.6
4  5  0.8
5  3  0.9
6  9  0.9
7  8  1.0
8  7  1.2
9 10  1.3
10 6  1.7
11 11 1.7

```

```

Regressionskoeffizient (Absolut) b =    5.4124
                        (Prozent) b =   45.1031
Achsenabschnitt        (Absolut) c =    0.5876
                        (Prozent) c =    4.8969
Testgröße (normalverteilt) N =    2.2730
Überschreitungswahrsch. (2seitig) p =  0.023028
Stichprobenumfang      n =           11

```

Literatur: G.A. Lienert (1973) Verteilungsfreie Methoden in der Biostatistik. Verlag Anton Hain, Meisenheim am Glan. pp. 684-687.

Rangkorrelation: Spearman's rho und Kendall's tau Untersucht wird die Korrelation der beiden ordinalskalierten Größen "Leistungen im Praktikum und Seminar" bei n=10 Studenten (Sachs 1992, p.513). Die Verifikation von rho und tau wurde mit Hilfe von Standardprogrammen durchgeführt. Eine Berechnung exakter p-Werte nach Valz und Thompson (1994) ist optional möglich, vgl. Ausgabe.

Rangkorrelation: Spearman's rho und Kendall's tau ▲

Name der Eingabedatei: Sachs.513

Korreliere: Y = Variable 1 = Praktik.
mit: X = Variable 2 = Seminar

Eingelesene Daten:

i	Y	X
	1	2
1	7	8
2	6	4
3	3	5
4	8	9
5	2	1
6	10	7
7	4	3
8	1	2
9	5	6
10	9	10

Spearman's Korrelationskoeffizient rho = 0.854545

Zweiseitiger Test (tabellenorientiert, n<=30): p < 0.010000
Zweiseitiger Test mit Edgeworth-Approximation: p = 0.003505
Zweiseitiger exakter Test (Valz-Algorithmus): p = 0.002863

Approximative Tests für rho (n>30, zweiseitig, wegen n=10 hier nicht adäquat!):

Student's t = 4.6537 mit df = 8 für n > 30: p = 0.001637
Gauss' Z = 2.5636 für n > 30: p = 0.010358

Approximative Konfidenzintervalle für rho (Student-Test-basiert, n>30):

P=0.90: [0.5131 , 1.0000] P=0.95: [0.4311 , 1.0000]
P=0.98: [0.3227 , 1.0000] P=0.99: [0.2384 , 1.0000]

Kendall's Korrelationskoeffizient tau = 0.688889 ("tau-b": tie-korrigiert)

Es sind 0 Bindungen ("Ties") vorhanden, gemäß "tau-b" 0.0% der Werte

Zweiseitiger exakter Test (Best, n<=20, mit Tie-Korrektur): p = 0.004687
Zweiseitiger exakter Test (Valz-Algorithmus): p = 0.004687

Zweiseitiger approximativer Test (für n>20, Gauß-Approx.): p = 0.005559

Approximative Konfidenzintervalle für tau (Gauß-Test-basiert, n>10):

P=0.90: [0.2802 , 1.0000] P=0.95: [0.2019 , 1.0000]
P=0.98: [0.1109 , 1.0000] P=0.99: [0.0489 , 1.0000]

L. Sachs (2003) Angewandte Statistik. 11. Aufl. Springer-Verlag, Kap. 53
C. Spearman (1904) The proof and measurement of association between two things. Amer. J. Psychol. 15, pp. 72-101
DJ. Best, DE. Roberts (1975) The Upper Tail Probabilities of Spearman's Rho. J. Roy. Stat. Soc. Ser. C, Applied Statistics Vol. 24, No. 3, pp. 377-379
DJ. Best, PG. Gipps (1974) The Upper Tail Probabilities of Kendall's Tau. J. Roy. Stat. Soc. Ser. C, Applied Statistics Vol. 23, No. 1, pp. 98-100
PD. Valz, ME. Thompson (1994) Exact inference for Kendall's S and Spearman's rho with Extensions to Fisher's Exact Test in rxc Contingency Tables. Journal of Computational and Graphical Statistics 3, pp. 459-472.

Korrelationsmatix (Pearson): Die Datei Bortz638.DAT enthält 6 geometrische Größen von Zylindern, die paarweise miteinander korreliert werden (Bortz 1977, p.638). Korrelationsmatrizen werden für bis zu maximal 20 Variablen berechnet.

Pearson-Korrelationsmatix ▲

Eingabedatei ist \Bortz638:DAT
 davon Variablen 1 (Durchmes), 2 (Länge), 3 (Grundfl.),
 4 (Mantelf.), 5 (Volumen), 6 (Diagonal)

1. Zeile : Korrelationskoeffizient r
2. Zeile : Überschreitungswahrscheinlichkeit p

	Durchme	Länge	Grundfl	Mantelf	Volumen	Diagona
Durchmes :	1.0000	0.0000	0.9897	0.8115	0.8953	0.5559
	0.0000	1.0000	0.0000	0.0000	0.0000	0.0026
Länge :	0.0000	1.0000	0.0000	0.5410	0.3482	0.8230
	1.0000	0.0000	1.0000	0.0036	0.0751	0.0000
Grundfl. :	0.9897	0.0000	1.0000	0.8032	0.9045	0.5579
	0.0000	1.0000	0.0000	0.0000	0.0000	0.0025
Mantelf. :	0.8115	0.5410	0.8032	1.0000	0.9687	0.8741
	0.0000	0.0036	0.0000	0.0000	0.0000	0.0000
Volumen :	0.8953	0.3482	0.9045	0.9687	1.0000	0.7665
	0.0000	0.0751	0.0000	0.0000	0.0000	0.0000
Diagonal :	0.5559	0.8230	0.5579	0.8741	0.7665	1.0000
	0.0026	0.0000	0.0025	0.0000	0.0000	0.0000

Literatur: L. Sachs (1997) Angewandte Statistik, 8. Aufl. Springer.

Korrelationsmatix (Spearman): Die Datei Bortz638.DAT enthält 6 geometrische Größen von Zylindern, die paarweise miteinander korreliert werden (Bortz (1977, p.638). Korrelationsmatrizen werden für bis zu maximal 20 Variablen berechnet.

Spearman-Korrelationsmatix ▲

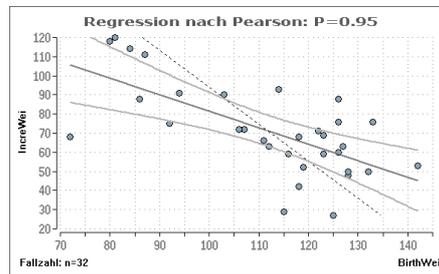
Eingabedatei ist Bortz638.DAT
 davon Variablen 1 (Durchmes), 2 (Länge), 3 (Grundfl.),
 4 (Mantelf.), 5 (Volumen), 6 (Diagonal)

1. Zeile : Korrelationskoeffizient r
2. Zeile : Überschreitungswahrscheinlichkeit p
 (für n<=30 exakt, p>0.2 immer approximativ)

	Durchme	Länge	Grundfl	Mantelf	Volumen	Diagona
Durchmes :	1.0000	0.0000	1.0000	0.8496	0.9474	0.5551
	0.0000	1.0000	0.0020	0.0020	0.0020	0.0100
Länge :	0.0000	1.0000	0.0000	0.5044	0.3158	0.8194
	1.0000	0.0000	1.0000	0.0100	0.2000	0.0020
Grundfl. :	1.0000	0.0000	1.0000	0.8496	0.9474	0.5551
	0.0020	1.0000	0.0000	0.0020	0.0020	0.0100
Mantelf. :	0.8496	0.5044	0.8496	1.0000	0.9664	0.8861
	0.0020	0.0100	0.0020	0.0000	0.0020	0.0020
Volumen :	0.9474	0.3158	0.9474	0.9664	1.0000	0.7866
	0.0020	0.2000	0.0020	0.0020	0.0000	0.0020
Diagonal :	0.5551	0.8194	0.5551	0.8861	0.7866	1.0000
	0.0100	0.0020	0.0100	0.0020	0.0020	0.0000

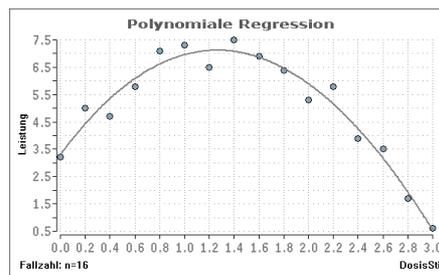
Literatur: L. Sachs (1997) Angewandte Statistik, 8. Aufl. Springer.

Graphik zur Regression: Datei Armitage.148, Gewichtszunahme in Abhängigkeit vom Geburtsgewicht. Zusätzlich zur Geraden "Y auf X" ist die Gerade "X auf Y" (gestrichelt) eingetragen. Die Konfidenzhyperbeln beziehen sich auf die Gerade "Y auf X".

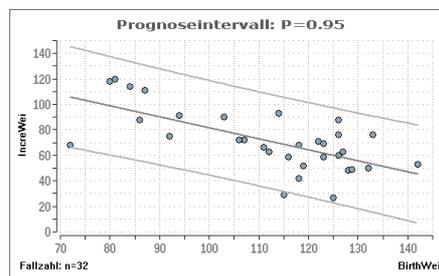


Die Regressionsgerade "X auf Y" (gestrichelt) und die Konfidenzbänder der Regressionsgeraden "Y auf X" werden optional (Regressionsdialog) dargestellt:

Graphik zur polynomialen Regression: Dargestellt wird das angepasste Polynom 3. Grades $Y=a+bx+cx^2+dx^3$ für die Variablen Dosis und Leistung aus der Datei Hartung592.DAT, die Berechnungen finden sich an anderer Stelle:



Graphik zum Prognoseintervall: Daten aus Datei Armitage.148, Gewichtszunahme in Abhängigkeit vom Geburtsgewicht. Die Grenzen wurden mit $P=0.95$ ermittelt, nicht zu verwechseln mit dem Konfidenz - Intervall für die Regressionsgerade! Optionen sind möglich:



Bitte beachten Sie die in Kapitel 5 ("Graphische Ausgabe") beschriebenen Möglichkeiten zur Modifikation von Graphiken: Nach Darstellung einer Abbildung kann im Menü des Graphik-Editors unter "Bearbeiten" ein Eingabedialog aufgerufen werden, der die Modifikation der Beschriftung, der Skalenbeschriftung, der Art und Größe der Beschriftung und des Netzgitters gestattet.

Logistische Regression: Hosmer, D.W. und Lemeshow, S. (2000, pp. 247-252) verwenden Daten aus einer Studie zu Risikofaktoren für geringes Geburtsgewicht von Kindern. Die Daten wurden 1986 am Baystate Medical Center, Springfield, Massachusetts gesammelt.

Der Parameter LOW bedeutet normales (0) bzw. niedriges (1) Geburtsgewicht (es sind nur Werte 0 und 1 zulässig!), die übrigen Parameter sind die möglichen Einflussgrößen. Es wird hier *kein* Abbau der gegebenen Einflussgrößen durchgeführt, im Programm ist dies optional möglich:

Logistische Regression

Eingabedatei ist Birth3.dat

Zielgröße ist Variable 2 (LOW)
mit Einflussvariablen 3 (AGE), 4 (LWT), 6 (RACE_2), 7 (RACE_3), 8 (SMOKE),
9 (PTL), 10 (HT), 11 (UI), 12 (FTV)

Variablen nicht in der Gleichung - Individuelle Tests der Einflussgrößen:

Variable	beta	SD(beta)	Wald's p	Odds-Ratio	95%-Konfidenzintervall
AGE	-0.0512	0.0316	0.105672	0.9501	0.8930 1.0109
LWT	-0.0141	0.0062	0.023657	0.9860	0.9741 0.9981
RACE_2	0.5636	0.4348	0.194931	1.7569	0.7493 4.1198
RACE_3	0.4322	0.3251	0.183657	1.5406	0.8147 2.9133
SMOKE	0.7041	0.3212	0.028367	2.0219	1.0774 3.7944
PTL	0.8018	0.3189	0.011927	2.2296	1.1934 4.1655
HT	1.2135	0.6115	0.047205	3.3654	1.0151 11.1574
UI	0.9469	0.4189	0.023794	2.5778	1.1341 5.8590
FTV	-0.1351	0.1576	0.391171	0.8736	0.6415 1.1897

>> Berechnung des Gesamtmodells mit allen 9 vorgegebenen Einflussgrößen:

Hosmer-Lemeshow's Test auf Güte der Anpassung des Gesamtmodells:
Modellabweichung=Devianz=-2LL=201.285; Chi²=183.095 mit df=179: p=0.401303
Ho(-2LL=0) akzeptiert: Die Modellanpassung mit 9 Variablen ist erfolgreich

Overall-LR-Test des Gesamtmodells: Chi²=33.387 mit df=9: p=0.000114

Pseudo-Bestimmtheitsmaße R² des angepassten Modells:
McFadden's R²=0.142271 - Cox-Snell's R²=0.161928 - Nagelkerke's R²=0.227716

Klassifikation gemäß Eingabe ("Beobachtet") und Gesamtmodell ("Vorhergesagt")

	Vorhergesagt=0	Vorhergesagt=1	Gesamtzahl	Anteil richtig
Beobachtet=0	117	13	130	90.00%
Beobachtet=1	36	23	59	38.98%
Gesamtzahl:			189	74.07%

Variablen in der Gleichung - Partielle Koeffizienten der Einflussgrößen:

Variable	beta	SD(beta)	Wald's p	Odds-Ratio	95%-Konfidenzintervall
Const.	0.4806	1.2105	0.691339	1.6171	0.1508 17.3432
AGE	-0.0295	0.0375	0.430129	0.9709	0.9022 1.0448
LWT	-0.0154	0.0070	0.027519	0.9847	0.9713 0.9983
RACE_2	1.2723	0.5334	0.017062	3.5689	1.2547 10.1516
RACE_3	0.8805	0.4458	0.048257	2.4121	1.0068 5.7791
SMOKE	0.9388	0.4067	0.020983	2.5570	1.1522 5.6747
PTL	0.5433	0.3493	0.119863	1.7217	0.8682 3.4145
HT	1.8633	0.7055	0.008261	6.4450	1.6170 25.6874
UI	0.7676	0.4645	0.098438	2.1547	0.8669 5.3556
FTV	0.0653	0.1744	0.708009	1.0675	0.7585 1.5024

Literatur: D.W. Hosmer, S. Lemeshow (2000) Applied Logistic Regression, Wiley

Intracluster-Korrelation (quantitative Daten): Intracluster-Korrelationen spielen bei Cluster-randomisierten Studien eine Rolle, um die Streuung innerhalb der Cluster mit der Streuung zwischen den Clustern zu vergleichen. Das nachfolgende Beispiel mit 5 Clustern wurde mit fiktiven Daten gerechnet und, soweit möglich, via Anova verifiziert. Die weiteren Angaben wurden mit Bleistift und Papier nachgerechnet.

Eine Fallzahlberechnung für Cluster-randomisierte Studien findet sich unter "Fallzahlberechnungen".

Einweg-Varianzanalyse: Intracluster-Korrelation ▲

Eingabedatei ist ..\Meter3.DAT
davon Variable 2 (PulsVor) und Gruppen-Var. 1 (Dosis) = 6 8 10 12 14

Tabelle der Varianzanalyse:

Streuung	fg	SQ	DQ	F	p
zwischen	4	532.3515	133.0879	1.3627	0.283934
innerhalb	19	1855.6072	97.6635		
total	23	2387.9587			

Bartlett-Test (Homogenität der Varianzen):

Prüfgröße $\chi^2 = 5.5173$ mit $fg = 4$ ($p = 0.238209$)

Test auf Gauß-Verteilung (Dallal-Wilkinson-korrigiert):

Stichprobe 6: Kolmogoroff-Smirnoff's delta = 0.1709 ($p > 0.200000$) Ok!
 Stichprobe 8: Kolmogoroff-Smirnoff's delta = 0.2487 ($p > 0.200000$) Ok!
 Stichprobe 10: Kolmogoroff-Smirnoff's delta = 0.2081 ($p > 0.200000$) Ok!
 Stichprobe 12: Kolmogoroff-Smirnoff's delta = 0.2923 ($p > 0.200000$) Ok!
 Stichprobe 14: Kolmogoroff-Smirnoff's delta = 0.2966 ($p = 0.107624$) Ok!

Mittlere Fallzahl pro Cluster ("Adjusted mean cluster size") $n = 5.25$

Design-Effekt $Deff=1+ICC*(n-1) = 1.170391$

Intracluster-Korrel. $ICC=1-n*SQ_{in}/(n-1)*SQ_{to} = 0.040092$ ($p=0.789833$ 2-s.)

Approximatives 90%-Konfidenzintervall für ICC: [-0.207333 ; 0.287517]

Approximatives 95%-Konfidenzintervall für ICC: [-0.235294 ; 0.334917]

Approximatives 99%-Konfidenzintervall für ICC: [-0.235294 ; 0.427558]

Literatur:

SM. Kerry, JM. Bland (1998) The intracluster correlation coefficient in cluster randomisation. British Medical Journal 316, pp. 1455-60.
 MK. Campbell et al. (2000) Analysis of cluster randomized trials in primary care: a practical approach. Family Care 17,2, pp.192-6
 SA. Knox, P. Chondros (2004) Observed intra-cluster-correlation in a cluster survey sample of patient encounters in general practice in Austria. BMC Medical Research Methodology 2004, 4, 30ff

Intracluster-Korrelation (dichotome Daten): Intracluster-Korrelationen spielen bei Cluster-randomisierten Studien eine Rolle, um die Streuung innerhalb der Cluster mit der Streuung zwischen den Clustern zu vergleichen. Das nachfolgende Beispiel mit 4 Clustern wurde mit fiktiven Daten gerechnet und mit per-Hand-Berechnungen verifiziert.

Eine Fallzahlberechnung für Cluster-randomisierte Studien findet sich unter "Fallzahlberechnungen".

Intracluster-Korrelation (dichotom) ▲

Eingabedatei ist ..\Valide\DiskreterICC.DAT
 davon Variable 2 (Wert)
 und Gruppenvar. 1 (Cluster) = 1,2,3,4

ClusterNr	X=0	X=1	Summe
1	4 36.4%	7 63.6%	11
2	5 55.6%	4 44.4%	9
3	2 50.0%	2 50.0%	4
4	1 11.1%	8 88.9%	9
Summe	12 36.4%	21 63.6%	33

Brandt-Snedecor's Testgröße $\chi^2=4.234$ mit $df=3$ und $p=0.237267$ (zweiseitig)

"Adjusted mean cluster size" $n = 8.2$

Design-Effekt $Deff=1+ICC*(n-1) = 1.058532$

Intracluster-Korrelation ICC = 0.008073

Literatur:

A Donner, N Klar (2000) Design and Analysis of Cluster Randomization Trials in Health Research. Oxford University Press American ed. New York, pp. 112-3

MK Campbell, PM Fayers, JM Grimshaw (2005) Determinants of the intracluster correlation coefficient in cluster randomized trials. Clinical Trials 2:99,107

H Chakraborty, J Moore, WA Carlo, TD Hartwell, LL Wright (2009) A simulation

based technique to estimate intracluster correlation for a binary variable. Contemporary Clinical Trials 30, 71-80

Intraclass-Korrelation: Diese findet sich im [Anhang 15](#): "Reliabilitätsanalyse". ▲

Zirkuläre Regression und Korrelation: In BiAS. werden drei Situationen für kreisverteilte Daten angeboten; alle beruhen auf dem Modell $Y=M+A*\text{COS}(\text{Phi}-V)$ mit M =Mesor ("mean level"), A =Amplitude und V =Acrophase ("Verschiebung").

Zum Testen werden artifizielle Daten aus Batschelet (1981, Kapitel 8 und 9) verwendet. Die Größe Phi ist zum Beispiel ein jahreszeitlicher oder monatlicher Rhythmus.

Zur *Regression* wird die Abhängigkeit einer linearen und einer zirkulären Größe betrachtet, zum Beispiel die Vitamin-D-Abhängigkeit (Y) von der Jahreszeit (Phi als Tag des Jahres). Nach Transformation $Y=y_i$ und $X=x_i=\text{cos}(\text{phi}-V)$ (stehen ebenfalls in der Datei "..\Valide\Batsche.Dat") kann mit den linearisierten x_i ein Test mit der konventionellen linearen Regressionsrechnung erfolgen. Eine Graphik (unten) zur Zirkulären Regression ist vorgesehen.

Zirkuläre Regression ▲

Eingabedatei ist .. \Batsche2.DAT

X1 = Y = Variable 2 = YY
X2 = X = Variable 1 = Zeit

Eingelesene Daten:

i	Y	X
1	12.9	0
2	13.3	30
3	14.4	60
4	20.8	90
5	19.1	120
6	19.6	150
7	18.8	180
8	21.3	210
9	15.0	240
10	13.6	270
11	13.6	300
12	12.6	330

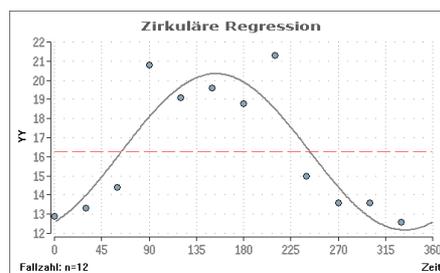
REGRESSION: Y=X1=linear, X=X2=zirkulär; Periode ist der Vollkreis mit 360°

Modellanpassung von $X1=M+A*\text{COS}(X2-V)$
und Test mit linearer Regression für $Y=y_i=X1$ und $X=x_i=\text{cos}(X2-V)$:

Mesor ("mean level"): M = 16.2500
Amplitude: A = 4.0864
Acrophase ("Verschiebung"): V = 153.4907
Periode: P = 360.0000

Angepasstes Modell mit $p = 0.000074$:
 $Y = X1 = 16.2500 + 4.0864 * \text{COS}(X2-153.4907)$

Literatur: E. Batschelet (1981) Circular Statistics in Biology, Acad. Press.



Zur *Korrelation* gibt es 2 verschiedene Methoden: Als nächstes wird der *Zusammenhang* einer linearen und einer zirkulären Größe untersucht. Die Daten sind die gleichen wie oben, der Test und die Berechnung des Korrelationskoeffizienten erfolgt ebenfalls wie oben:

Zirkuläre Regression und Korrelation ▲

Eingabedatei ist .. \Batsche2.DAT

X1 = Y = Variable 2 = YY
 X2 = X = Variable 1 = Zeit

KORRELATION: X1=linear, X2=zirkulär; Periode ist der Vollkreis mit 360°

Modellanpassung von $X1=M+A*\cos(X2-V)$ mit Verschiebung ("Acrophase") $V = 153.4907$ und linearer Korrelation für $y_i=X1$ und $x_i=\cos(X2-V)$:

Korrelationskoeffizient $r = 0.897686$ ($p=0.000074$)

Als letzte Variante findet sich in **BiAS**. die Korrelation von *zwei zirkulären Größen*. Der Test erfolgt mit Hilfe einer Tabelle von Stephens (1979):

Zirkuläre Regression und Korrelation ▲

Eingabedatei ist \VALIDE\ Batsche1.DAT

X1 = Y = Variable 1 = Y
 X2 = X = Variable 2 = X

i	Y	X
1	60	120
2	90	135
3	135	210
4	150	150
5	210	195
6	240	240
7	255	300
8	300	315

KORRELATION: Beide Variablen sind zirkulär; Periode ist der Vollkreis mit 360°

Korrelation positiv: $r(p) = 0.864284$

Korrelation negativ: $r(n) = 0.243207$

Gesamt: $\text{Max}(r(p),r(n)) = 0.864284$

Mardia's Prüfgröße M = 2.444565

Ist $M > S$ (S aus nachfolgender Tabelle), so liegt eine signifikante Korrelation vor (Tabelle nach Stephens (1979), aus Batschelet (1981)):

n	Signifikanzniveau alpha			
	0.10	0.05	0.025	0.01
5	1.62	1.78	1.96	2.10
10	1.67	1.85	2.03	2.20
20	1.70	1.89	2.06	2.25
50	1.71	1.91	2.08	2.28
>>	1.72	1.92	2.10	2.30

Literatur: E. Batschelet (1981) *Circular Statistics in Biology*, Acad. Press.

Bemerkung: die Prüfgröße M ist größer als alle Schwellenwerte in Stephens' Tabelle, mithin ist der Test signifikant bei $\alpha=0.01$.

Punktbiseriale Korrelation dichotom vs. quantitativ: In Bortz (2004) wird eine Korrelation zwischen Familienstatus und der Anzahl der sozialen Kontakte behandelt. Die Verifikation des Programms erfolgt mit diesem Beispiel.

Punktbiseriale Korrelation "dichotom-quantitativ" ▲

Eingabedatei ist ...\\Valide\\Bortz225.DAT

X1 = Y = Variable 1 = Verheiratet
X2 = X = Variable 2 = Kontakte

Eingelesene Daten:

i	Y	X
	1	2
1	1	18.00
2	1	12.00
3	1	16.00
4	1	15.00
5	1	12.00
6	1	14.00
7	1	13.00
8	1	9.00
9	1	12.00
10	1	17.00
11	1	13.00
12	1	11.00
13	2	17.00
14	2	12.00
15	2	16.00
16	2	19.00
17	2	20.00
18	2	16.00
19	2	11.00
20	2	18.00
21	2	12.00
22	2	17.00
23	2	19.00
24	2	20.00
25	2	19.00
26	2	13.00
27	2	18.00

Punktbiserialer Korrelationskoeffizient $r.pb = 0.470183$

Test der Nullhypothese:

Testgröße $t=2.6637$ mit $df=25$ und $p=0.013329$

>> $r.pb$ liegt analog zu Pearson's Korrelationskoeffizienten zwischen -1 und 1.
>> Das Vorzeichen von $r.pb$ ergibt sich aus der Codierung der dichotomen Werte.

Durchschnitte, Standardabweichungen und Streuungen:

	Xquer	Std.Abw.	Streuung	n
X1 = 1.00	13.5000	2.6112	6.8182	12
X1 = 2.00	16.4667	3.0675	9.4095	15
Gesamt:	15.1481	3.1950	10.2080	27

Literatur: J Bortz (2005/2010) Statistik. 6. Auflage Springer-Verlag (pp. 224ff)

Punktbiseriale Korrelation dichotom vs. ordinal: In Bortz (2004) wird eine Korrelation zwischen Geschlecht und der Reihung von Schülerinnen und Schülern nach der Einschätzung eines Lehrers angegeben. Die Verifikation des Programms erfolgt mit diesem Beispiel, die Berechnung der p-Werte mit Hilfe des Wilcoxon-Mann-Whitney-Tests. Analog wurde der punktbiseriale Korrelationskoeffizient r_{pb} anhand der Wilcoxonschen Rangsummen einschließlich der Bindungskorrekturen verifiziert.

Punktbiseriale Korrelation "dichotom-ordinal" ▲

Eingabedatei ist ...\\Valide\\Bortz231.DAT

X1 = Y = Variable 1 = Geschlecht
X2 = X = Variable 2 = Platz

Punktbiserialer Korrelationskoeffizient $r_{pb} = -0.607143$

Test der Nullhypothese:

Testgröße U = 11.0 mit p = 0.054079 (exakter Test)
Testgröße Z = 1.967 mit p = 0.049141 (approximativ)

>> r_{pb} liegt analog zu Pearson's Korrelationskoeffizienten zwischen -1 und 1.
>> Das Vorzeichen von r_{pb} ergibt sich aus der Codierung der dichotomen Werte.

	Median	Minimum	Maximum	n	Rangsumme	Normiert
X1 = 1.00	11.0000	2.0000	15.0000	8	81.0	10.1
X1 = 2.00	5.0000	1.0000	11.0000	7	39.0	5.6

Literatur: J Bortz (2005/2010) Statistik. 6. Auflage Springer-Verlag (pp. 231ff)



Einstichproben-t-Test
Zweistichproben-t-Test
Welch-Test bei ungleichen Varianzen
F-Test auf Gleichheit zweier Varianzen
Bartlett's Test und Levene's Test auf Varianzengleichheit
Vergleich von Variationskoeffizienten
Pitman und Morgan's Vergleich zweier abhängiger Varianzen
Unifaktorielle Varianzanalyse mit Scheffé-Tests
Unifaktorielle Varianzanalyse mit Linearen Kontrasten
Unifaktorielle Varianzanalyse mit Orthogonalen Polynomen
Unifaktorielle Kovarianzanalyse mit Scheffé-Test
Bifaktorielle Varianzanalyse mit Scheffé-Tests
Bifaktorielle Kovarianzanalyse mit Scheffé-Test
Hotelling's Einstichproben-T²-Test mit Simultanen Konfidenzintervallen
Hotelling's Zweistichproben-T²-Test mit Simultanen Konfidenzintervallen
Varianzanalyse mit Messwiederholungen
Cross-Over-Analyse im Grizzle-Modell

Einstichproben-t-Test: Graf et al. (1987, p.400): Untersuchung eines Mittels zur Festigkeit textiler Materialien bei n=10 Materialsorten; pro Materialart wird eine Probe behandelt und eine nicht behandelt. (Der Test ist auch als "echter" Einstichproben-Test gegen ein festes μ durchführbar.)

Einstichproben-t-Test

Eingabedatei ist Graf400.DAT
davon Variable 1 (Behandelt) - Variable 2 (Unbehandelt)

i	Var 1	Var 2	Diff
1:	14.7000	12.1000	2.6000
2:	14.0000	10.9000	3.1000
3:	12.9000	13.1000	-0.2000
4:	16.2000	14.5000	1.7000
5:	10.2000	9.6000	0.6000
6:	12.4000	11.2000	1.2000
7:	12.0000	9.8000	2.2000
8:	14.8000	13.7000	1.1000
9:	11.8000	12.0000	-0.2000
10:	9.7000	9.1000	0.6000

X-quer = 1.2700
s = 1.1265
n = 10
df = 9

Test auf Gauß-Verteilung der Differenzen (Dallal-Wilkinson-korrigiert)
Kolmogoroff-Smirnoff's S = 0.1248 (p > 0.20) OK!
Student's Prüfgröße t = 3.5651 (p = 0.006071)

Konfidenzintervalle:
P = 0.90: (0.6170 , 1.9230)
P = 0.95: (0.4642 , 2.0758)
P = 0.99: (0.1123 , 2.4277)

Literatur: L. Sachs (1992) Angewandte Statistik, Springer. Kap.32ff

Zweistichproben-t-Test: Für zwei photoelektrische Densitometer wird untersucht, ob sie vergleichbare Werte für die optische Dichte eines Filmes liefern, wozu mit beiden Instrumenten jeweils 6 Messungen am gleichen Film durchgeführt wurden (Graf et al. (1987, p.397)):

Zweistichproben-t-Test ▲

Eingabedatei ist Graf397.DAT
 davon Variable 1 (Dichte)
 und Gruppe Var. 2 (Densitom) = 1,2

Gruppe 1		Gruppe 2	
1:	2.6800	1:	2.6400
2:	2.6900	2:	2.6500
3:	2.6700	3:	2.6500
4:	2.6900	4:	2.6400
5:	2.7000	5:	2.6300
6:	2.6800	6:	2.6300

	X-quer	s	s ²	n
Stichprobe 1	2.6850	0.0105	0.0001	6
Stichprobe 2	2.6400	0.0089	8.000E-05	6

Differenz der Mittelwerte: 0.0450

Cohen's Effektstärke d, mit gepoolter Standardabweichung: d = 4.6169
 Cohen: d=0.2 geringer Effekt, d=0.5 mittlerer Effekt, d=0.8 großer Effekt

Test auf Gauß-Verteilung (Dallal-Wilkinson-korrigiert):
 Stichprobe 1: Kolmogoroff-Smirnoff's S = 0.1832 (p > 0.20) OK!
 Stichprobe 2: Kolmogoroff-Smirnoff's S = 0.2016 (p > 0.20) OK!

Homogenität der Varianzen (zweiseitig):
 Prüfgröße F = 1.3750 mit df = (5, 5) (p = 0.735278)

t-Test für gleiche Varianzen:
 Testgröße t = 7.9967 mit df = 10 (p = 0.000012)

t-Test für ungleiche Varianzen per df interpoliert):
 Welch's t = 7.9967 mit df = 9.76 (p = 0.000014)

Konfidenzintervalle für die Differenz $\delta = \mu_1 - \mu_2$:
 P = 0.90: (0.0348 , 0.0552)
 P = 0.95: (0.0325 , 0.0575)
 P = 0.99: (0.0272 , 0.0628)

Literatur: L. Sachs (1992) Angewandte Statistik. Springer, Kap.36,153

Welch-Test: Lorenz (1988, p.175): Vergleich von säugenden und nicht-säugenden Mäusen bezüglich der Trächtigkeitsdauer; die Varianzen in den beiden Gruppen werden als unterschiedlich angenommen (F-Test!). In diesem Beispiel ist der Test auf Gauß-Verteilung ebenfalls positiv, so dass eher ein nicht-parametrischer Test verwendet werden sollte (Wilcoxon-Mann-Whitney-Test.)

Welch-Test ▲

Eingabedatei ist Lorenz175.DAT
 davon Variable 1 (DauerGra)
 und Gruppe(n) Var. 2 (Säugend) = 1, 2

Gruppe 1	Gruppe 2
1: 18.0	1: 25.0
2: 18.0	2: 19.0
3: 19.0	3: 20.0
4: 20.0	4: 24.0

5: 19.0	5: 19.0
6: 19.0	6: 22.0
7: 17.0	7: 27.0
8: 19.0	8: 22.0
9: 19.0	9: 19.0
10: 20.0	10: 18.0
11: 21.0	11: 19.0
12: 19.0	12: 23.0
	13: 24.0
	14: 20.0
	15: 21.0

	X-quer	s	s ²	n
Stichprobe 1	19.0000	1.0445	1.0909	12
Stichprobe 2	21.4667	2.6690	7.1238	15

Differenz der Mittelwerte: -2.4667

Test auf Gauß-Verteilung (Dallal-Wilkinson-korrigiert):

Stichprobe 1: Kolmogoroff-Smirnoff's S = 0.2500 (p = 0.03691) Cave!!

Stichprobe 2: Kolmogoroff-Smirnoff's S = 0.1753 (p > 0.20000) OK!

Homogenität der Varianzen (zweiseitig):

Prüfgröße F = 6.5302 mit df = (14, 11) (p = 0.003504)

t-Test für ungleiche Varianzen (per df interpoliert):

Welch's t = 3.2792 mit df = 18.99 (p = 0.003949)

Literatur: L. Sachs (1992) Angewandte Statistik. Springer. (t-Test: Kap.36, F-Test: Kap.153, Welch-Test: Kap.36/2)

F-Test auf Varianzgleichheit: Vergleich zweier Streuungen, zum Beispiel als Voraussetzung zum t-Test oder zum Vergleich der Messgenauigkeit zweier Analysemethoden; vgl. auch nächstes Beispiel (Sachs 1992, p.348):

F-Test auf Varianzgleichheit

Bezeichnung: Sachs p. 348

Standardabweichung s(1) = 2.8284

Streuung s²(1) = 8.0000

Stichprobenumfang n(1) = 20.0000

Standardabweichung s(2) = 1.7321

Streuung s²(2) = 3.0000

Stichprobenumfang n(2) = 20.0000

Prüfgröße F = 2.6667 mit df = (19 , 19)

Einseitige Fragestellung: p = 0.019220

Zweiseitige Fragestellung: p = 0.038439

Literatur: L. Sachs (1992) Angewandte Statistik, Springer. Kap.153.

Levene's Test auf Varianzgleichheit: Der Levene-Test ist ein robuster Test auf Varianzgleichheit und setzt nicht notwendigerweise Gauß-Verteilungen voraus. Der Test wurde mit Hilfe von SPSS überprüft. Unter „Nichtparametrik“ findet sich eine nicht-parametrische Variante.

Im Programm wurde zur ANOVA der Test von Bartlett bevorzugt, da dieser bei zur ANOVA ohnehin vorausgesetzten Gauß-Verteilungen sensibler ist. Bartlett's Test steht nur implizit im Zusammenhang mit der Einweg-Varianzanalyse zur Verfügung.

Levene's Test auf Varianzgleichheit

Eingabedatei ist Test.DAT
davon Variable 2 (VarZwei)
und Gruppen-Var. 3 (GruppNum) = 1 2 3

	Xquer	s	s ²	n
Stichprobe 1	3.8000	1.3038	1.7000	5
Stichprobe 2	25.0000	5.0000	25.0000	3
Stichprobe 3	34.7500	3.0957	9.5833	4

Levene's Prüfgröße F = 1.726 mit df = (2,9) und p = 0.231978

Literatur: H. Levene (1960) In: Contributions to Probability and Statistics. Essays in Honor of H. Hotelling. Olkin et al. Stanford Univ Press, pp.278-92

Vergleich von Variationskoeffizienten: Dazu wird ein Beispiel von Sachs (2003) verwendet:

Vergleich von Variationskoeffizienten

Variationskoeffizient CV(1) = 0.1000
Stichprobenumfang n(1) = 30.0000

Variationskoeffizient CV(2) = 0.1300
Stichprobenumfang n(2) = 30.0000

Gauß' Prüfgröße zum Test: Z = 1.4168

Einseitige Fragestellung: p = 0.078265
Zweiseitige Fragestellung: p = 0.156530

Literatur: L. Sachs (2003) Angewandte Statistik, Springer. Kap. 36.

Pitman und Morgan's Vergleich zweier abhängiger Varianzen: Hierzu werden die Daten aus der Beispieldatei Test.DAT verwendet (siehe dort). Eine Validierung erfolgt mit den verwendeten Moduln aus BiAS., das sind die einfache lineare Korrelation, der Mudholkar-Test und die Spearman-Rangkorrelation.

Pitman-Morgan-Test auf Gleichheit abhängiger Varianzen

Eingabedatei ist C:\PbWin80\BiAS8\Test.DAT
davon Variable 1 (VarEins) und Variable 2 (VarZwei)

Durchschnitte, Standardabweichungen und Streuungen:

Variable	Xquer	Std.Abw.	Streuung	n	df
VarEins	16.8000	12.3242	151.8857	15	
VarZwei	20.0667	12.4869	155.9238	15	
Differenzen	-3.2667	4.3502	18.9238	15	14

Student's Prüfgröße t = 2.9083 mit df = 14 (p = 0.011452)

Test auf Gauß-Verteilung der Differenzen (Dallal-Wilkinson-korrigiert):
Kolmogoroff-Smirnoff's delta = 0.1655 (p > 0.100000) Ok!

Pitman-Morgan-Test für $H_0(\sigma^2(\text{VarEins}) = \sigma^2(\text{VarZwei}))$ per Pearson-Korrelation von $U=X+Y$ und $V=X-Y$, mit $X=\text{VarEins}$ und $Y=\text{VarZwei}$:

Korrelationskoeffizient $r(U,V) = -0.0380$
 Testgröße der Korrelation $t = -0.1371$
 Überschreitungswahrsch. $p = 0.893042$
 Stichprobenumfang $n = 15$
 Anzahl Freiheitsgrade $fg=n-2 = 13$

Mudholkar-Test auf bivariate Gauß-Verteilung als Voraussetzung für den PM-Test:

Mudholkar's Prüfgröße $N = 1.6752 : p = 0.093889$ (PM-Test nur falls nicht-sign.)

>> Falls keine bivariate Gauß-Verteilung: PM-Modifikation nach McCulloch (1987)
 >> mit Spearman's Korrelation von $U=X+Y$ und $V=X-Y$ (dazu: Bonate (2000 p.176)):

>> Spearman's Korrelationskoeffizient $\rho(U,V) = -0.115632$
 >> Zweiseitiger Test (Edgeworth-Approximation): $p = 0.685768$

Literatur:

R. Wilcox (2015) Comparing the variances of two dependent variables. Journal of Statistical Distributions and Applications 2:7
 P.L. Bonate (2000) Analysis of pretest-posttest designs. Chapman & Hall
 E.J.G. Pitman (1939) A Note on Normal Correlation. Biometrika 31, pp. 9-12
 W.A. Morgan (1939) A test for the significance between two variances in a sample from a normal bivariate population. Biometrika 32, pp. 13-19
 C.E. McCulloch (1987) Tests for equality of variance for paired data. Commun. Stat. Theory Methods 16, pp.1377-91.
 G.S. Mudholkar, M. McDermott, D.K. Srivastava (1992) A test of p-variate normality. Biometrika Vol. 79, 4, pp. 850-4.

Unifaktorielle Varianzanalyse (ANOVA): Vergleich von vier Lehrmethoden für den Englischunterricht. Den Lernerfolg (Zielgröße!) soll ein Englischtest erfassen (Bortz (1977, p.300)). Bemerkung: Die Zielgröße ist von ihrer Qualität her ordinal, weshalb hier eher an eine Analyse mit dem Kruskal-Wallis-Test gedacht werden sollte.

Unifaktorielle Einweg-Varianzanalyse (ANOVA) ▲

Eingabedatei ist Bortz300.DAT
 davon Variable 1 (ScoreTst)
 Alle Gruppen Var. 3 (U_Method)

Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
1: 2.000	1: 3.000	1: 6.000	1: 5.000
2: 1.000	2: 4.000	2: 8.000	2: 5.000
3: 3.000	3: 3.000	3: 7.000	3: 5.000
4: 3.000	4: 5.000	4: 4.000	4: 3.000
5: 1.000	5: 0.000	5:10.000	5: 2.000

	X-quer	s	s ²	n
Stichprobe 1	2.0000	1.0000	1.0000	5
Stichprobe 2	3.0000	1.8708	3.5000	5
Stichprobe 3	7.0000	2.2361	5.0000	5
Stichprobe 4	4.0000	1.4142	2.0000	5

Tabelle der Varianzanalyse:

Streuung	fg	SQ	DQ	F	p
zwischen	3	70.0000	23.3333	8.1159	0.001642
innerhalb	16	46.0000	2.8750		
total	19	116.0000			

Bartlett-Test (Homogenität der Varianzen):
 Prüfgröße $\chi^2 = 2.4231$ mit $fg = 3$ ($p = 0.489356$)

F-Test der Varianzanalyse bei heterogenen Varianzen - Welch-korrigiert:
 Welch's $F=6.708$ mit $fg_1=3$, $fg_2=8.52$ und $p=0.01269$ (per fg_2 interpoliert)

Test auf Gauß-Verteilung (Dallal-Wilkinson-korrigiert):
 Stichprobe 1: Kolmogoroff-Smirnoff's $S = 0.2413$ ($p > 0.20000$) Ok!
 Stichprobe 2: Kolmogoroff-Smirnoff's $S = 0.1456$ ($p = 0.17314$) Ok!
 Stichprobe 3: Kolmogoroff-Smirnoff's $S = 0.1274$ ($p > 0.20000$) Ok!
 Stichprobe 4: Kolmogoroff-Smirnoff's $S = 0.2398$ ($p = 0.03270$) Cave!!

Multiple Vergleiche nach Scheffe:

Stichproben	Differenz	F	df1	df2	p
1 2	1.0000	0.2899	3	16	0.832074
1 3	5.0000	7.2464	3	16	0.002749
1 4	2.0000	1.1594	3	16	0.355850
2 3	4.0000	4.6377	3	16	0.016180
2 4	1.0000	0.2899	3	16	0.832074
3 4	3.0000	2.6087	3	16	0.087381

Literatur: L. Sachs (1992) Angewandte Statistik, Springer-Verlag. (ANOVA Kap.73, Bartlett-Test Kap.723, Scheffe-Test Kap.742)

Unifaktorielle Varianzanalyse mit linearen Kontrasten: Es wird die Wirkung eines neuen Präparates zur Behandlung von Depressionen geprüft, wozu drei Gruppen von Patienten (Placebo, einfache Dosis, doppelte Dosis) gebildet werden. Die Depressivität der Patienten (Zielgröße!) wird mit Hilfe eines Fragebogens eingeschätzt (Bortz (1977, p.325). Bemerkung: Mit der an sich ordinalen Zielgröße sollte eher eine nicht-parametrische Auswertung unternommen werden!

Einweg-Varianzanalyse mit linearen Kontrasten ▲

Eingabedatei ist Bortz325.DAT
 davon Variable 1 (Depress.)
 Alle Gruppen Var. 2 (Behandlg)

Gruppe 1	Gruppe 2	Gruppe 3
1: 18.0000	1: 19.0000	1: 16.0000
2: 22.0000	2: 16.0000	2: 13.0000
3: 25.0000	3: 16.0000	3: 12.0000
4: 19.0000	4: 15.0000	4: 12.0000
5: 22.0000	5: 17.0000	5: 14.0000
6: 19.0000	6: 16.0000	6: 16.0000
7: 21.0000	7: 20.0000	7: 13.0000
8: 17.0000	8: 15.0000	8: 13.0000
9: 21.0000	9: 16.0000	9: 14.0000
10: 22.0000	10: 16.0000	10: 12.0000

	X-quer	s	s ²	n
Stichprobe 1	20.6000	2.3664	5.6000	10
Stichprobe 2	16.6000	1.6465	2.7111	10
Stichprobe 3	13.5000	1.5092	2.2778	10

Tabelle der Varianzanalyse:

Streuung	fg	SQ	DQ	F	p
zwischen	2	253.4000	126.7000	35.8961	0.000000
innerhalb	27	95.3000	3.5296		
total	29	348.7000			

Bartlett-Test (Homogenität der Varianzen):
 Prüfgröße $\chi^2 = 2.0605$ mit $fg = 2$ ($p = 0.356911$)

Test auf Gauß-Verteilung (Dallal-Wilkinson-korrigiert):
 Stichprobe 1: Kolmogoroff-Smirnoff's $S = 0.1771$ ($p > 0.20000$) OK!
 Stichprobe 2: Kolmogoroff-Smirnoff's $S = 0.3422$ ($p = 0.00154$) Cave!!!
 Stichprobe 3: Kolmogoroff-Smirnoff's $S = 0.2298$ ($p > 0.14702$) OK!

Linearer Kontrast nach Scheffe:
 $1.00 \cdot \mu_1 - 0.50 \cdot \mu_2 - 0.50 \cdot \mu_3$
 Prüfgröße $F = 29.0895$ mit $df = (2, 27)$ ($p = 0.000000$)

Literatur:
 L. Sachs (1992) Angewandte Statistik, Springer-Verlag. (ANOVA Kap.73, Bartlett-Test Kap.723, Scheffe-Kontraste Kap.732)

Unifaktorielle Varianzanalyse mit orthogonalen Polynomen: Durchgerechnete Beispiele aus der Literatur wurden nicht gefunden, deshalb hier wieder Verwendung des Datensatzes Test.DAT. Die ANOVA kann mit der üblichen Einweg-Varianzanalyse überprüft werden, die orthogonalen Polynome incl. Test wurden durch per-Hand-Berechnungen nachvollzogen:

Orthogonale Polynome (ANOVA) ▲

Eingabedatei ist Test.DAT
 davon Variable 2 (VarZwei)
 und Gruppen-Var. 3 (GruppNum) = 1 2 3

Gruppe 1	Gruppe 2	Gruppe 3
1: 2.0	1: 20.0	1: 35.0
2: 3.0	2: 25.0	2: 33.0
3: 5.0	3: 30.0	3: 32.0
4: 5.0		4: 39.0
5: 4.0		

	Xquer	s	s ²	n
Stichprobe 1	3.8000	1.3038	1.7000	5
Stichprobe 2	25.0000	5.0000	25.0000	3
Stichprobe 3	34.7500	3.0957	9.5833	4

Tabelle der Varianzanalyse:

Streuung	fg	SQ	DQ	F	p
zwischen	2	2253.3667	1126.6833	118.5289	0.000000
innerhalb	9	85.5500	9.5056		
total	11	2338.9167			

Bartlett-Test (Homogenität der Varianzen):
 Prüfgröße $\chi^2 = 4.2396$ mit $fg = 2$ ($p = 0.120055$)

Test auf Gauß-Verteilung (Dallal-Wilkinson-korrigiert):
 Stichprobe 1: Kolmogoroff-Smirnoff's $\delta = 0.2213$ ($p > 0.200000$) Ok!
 Stichprobe 2: Kolmogoroff-Smirnoff's $\delta = 0.1747$ ($p > 0.200000$) Ok!
 Stichprobe 3: Kolmogoroff-Smirnoff's $\delta = 0.2178$ ($p > 0.200000$) Ok!

Orthogonale Polynome:

Argumente der 3 Stufen (Gruppen):

Stufe	Stichprobe	Argument
1	1	1.0000
2	2	2.0000
3	3	4.0000

Koeffizientenmatrix der Orthogonalkomponenten:

-15.0000	-3.0000	21.0000
1.6000	-4.0000	1.0000

Orthogonalkomponenten p-ter Ordnung:

Linear : Z(1) = 2409.0000
Quadratisch : Z(2) = -130.6000

Summe der Quadrate der ANOVA:

	i	SQ(Zi)	F	Fg1	Fg2	p
Linear:	1	1990.1512	209.3672	1	9	0.000000
Quadratisch:	2	263.2155	27.6907	1	9	0.000519

Literatur:

L. Sachs (1997) Angewandte Statistik, 8. Aufl. Springer-Verlag. (ANOVA Kap.73, Bartlett-Test Kap.723)

H. Ackermann (1979) A concise formula to compute general orthogonal contrast coefficients. Biom. J. 21, Vol. 7, pp. 597-606.



Unifaktorielle Kovarianzanalyse: Es wird geprüft, ob eine geschlechtstherapeutische Behandlung auf verschiedene Verhaltensstörungen (Konzentrationsstörung, Potenzstörung, Hysterisches Verhalten: "Stufen Faktor A", hier identisch mit der Indizierung="Index") Einfluss besitzt. Die Zielgröße ist der von einem Expertengremium eingeschätzte Therapieerfolg. Als Kontrollvariable wird die verbale Intelligenz mit erfasst (Bortz (1977, p.458)). Bemerkung: Mit der an sich ordinalen Zielgröße sollte nach Möglichkeit eine nicht-parametrische Auswertung unternommen werden!

Unifaktorielle Kovarianzanalyse



Eingabedatei ist Bortz458.DAT

davon Variable 2 (Score) mit Kovariable 1 (VerbalA)

Faktor A: Stufe | Variable 3 (Diagnose)

-----+-----	
1	1
2	2
3	3

Variable 2	Kovariable	A: Index	Stufe
5.00000	7.00000	1	1
6.00000	9.00000	1	1
6.00000	8.00000	1	1
4.00000	5.00000	1	1
5.00000	5.00000	1	1
5.00000	11.00000	2	2
4.00000	12.00000	2	2
2.00000	8.00000	2	2
1.00000	7.00000	2	2
3.00000	9.00000	2	2
2.00000	12.00000	3	3
1.00000	10.00000	3	3

1.00000	9.00000	3	3
1.00000	10.00000	3	3
2.00000	13.00000	3	3

Tabelle der Kovarianzanalyse:

Source	fg	SQ	DQ	F	p
Faktor A	2	42.8498	21.4249	65.4827	0.000001
Rest	11	3.5990	0.3272		
Total	13	46.4488			

Voraussetzungen:

Bartlett-Test (Homogenität der Varianzen):
 Prüfgröße $\chi^2 = 3.9845$ mit $df = 2$ ($p = 0.136386$)
 Homogenität der Steigungen innerhalb der Stufen:
 Prüfgröße $F = 2.12$ mit $df = (2, 9)$ ($p = 0.175572$)
 Schätzung der zusammengefassten Regression und Korrelation:
 Regressionskoeffizient $b = 0.50490$
 Korrelationskoeffizient $r = 0.86193$
 Prüfgröße $F = 26.01$ mit $df = (1, 9)$ ($p = 0.000645$)

Linearität Abhängigkeit der Zielvariablen von der Kontrollvariablen:
 Prüfgröße $F = 0.72$ mit $df = (1, 11)$ ($p = 0.413892$)

Durchschnitte:

Stufen	Unkorrigiert	Korrigiert
1	5	5.2000
2	5	3.0000
3	5	1.4000

Multiple Vergleiche nach Scheffe (Korrigierte Differenzen):

Stufen	Faktor A	Differenz	F	df1	df2	p
1	2	3.5127	33.3347	2	11	0.000021
1	3	5.8196	65.3363	2	11	0.000001
2	3	2.3069	18.1512	2	11	0.000328

Literatur: J. Bortz (1989) Lehrbuch der Statistik Kap.7,10. Springer

Bifaktorielle Varianzanalyse: Vergleich dreier Behandlungsmethoden mit der Zielgröße "Depressivität". Da das Geschlecht möglicherweise die Fehlervarianz vergrößert, wird dieses als 2. Faktor erfasst (Quelle: Bortz (1977, p.355)); hier sind die "Stufen" identisch mit den "Indices" der Faktoren! Bemerkung: Mit der offenbar ordinalen Zielgröße sollte besser eine nicht-parametrische Auswertung vorgenommen werden!

Bifaktorielle Varianzanalyse ▲

Eingabedatei ist Bortz355.DAT, davon Variable 1 (Depress.).

Faktor A:	Stufe	Variable 3 (Dosis)	Faktor B:	Stufe	Variable 2 (Geschl.)
	1	1		1	1
	2	2		2	2
	3	3			

Variable 1	A: Index	Stufe	B: Index	Stufe
22.00000	1	1	1	1
25.00000	1	1	1	1
22.00000	1	1	1	1
21.00000	1	1	1	1
22.00000	1	1	1	1
19.00000	2	2	1	1
16.00000	2	2	1	1
16.00000	2	2	1	1

15.00000	2	2	1	1
17.00000	2	2	1	1
13.00000	3	3	1	1
12.00000	3	3	1	1
12.00000	3	3	1	1
13.00000	3	3	1	1
12.00000	3	3	1	1
18.00000	1	1	2	2
19.00000	1	1	2	2
17.00000	1	1	2	2
21.00000	1	1	2	2
19.00000	1	1	2	2
16.00000	2	2	2	2
20.00000	2	2	2	2
15.00000	2	2	2	2
16.00000	2	2	2	2
16.00000	2	2	2	2
16.00000	3	3	2	2
14.00000	3	3	2	2
16.00000	3	3	2	2
13.00000	3	3	2	2
14.00000	3	3	2	2

Stichprobenumfänge n(i,j):

Faktor B	Faktor A:		
	a = 1	a = 2	a = 3
b = 1	5	5	5
b = 2	5	5	5

Durchschnitte:

Faktor B	Faktor A:		
	a = 1	a = 2	a = 3
b = 1	22.400	16.600	12.400
b = 2	18.800	16.600	14.600

Tabelle der Varianzanalyse für Faktoren mit festen Effekten (Modell I):

Source	fg	SQ	DQ	F	p
Faktor A	2	253.4000	126.7000	59.8583	0.000000
Faktor B	1	1.6333	1.6333	0.7717	0.388416
A x B	2	42.8667	21.4333	10.1260	0.000648
Rest	24	50.8000	2.1167		
Total	29	348.7000			

Varianzanalyse für Faktoren mit zufälligen Effekten (Modelle II und III):

	Faktor A		Faktor B		A x B	
	F	p	F	p	F	p
A:fest B:fest	59.8583	0.000000	0.7717	0.388416	10.1260	0.000648
A:fest B:rand	5.9114	0.144689	0.7717	0.388416	10.1260	0.000648
A:rand B:fest	59.8583	0.000000	0.0762	0.808417	10.1260	0.000648
A:rand B:rand	5.9114	0.144689	0.0762	0.808417	10.1260	0.000648

Prüfung der Voraussetzungen:

Bartlett-Test (Homogenität der Varianzen): $\chi^2=4.834$ mit $df=5$ und $p=0.436482$
 Unterschiede der Stichprobenumfänge $\leq 0.0\%$. Gauß-Verteilung nicht vergessen!

Modell I : Multiple Vergleiche nach Scheffé (Korrigierte Differenzen):

Stufen	Faktor A	Differenz	F	df1	df2	p
1	2	4.0000	18.8976	2	24	0.000012
1	3	7.1000	59.5394	2	24	0.000000
2	3	3.1000	11.3504	2	24	0.000339

Stufen	Faktor B	Differenz	F	df1	df2	p
1	2	0.4667	0.7717	1	24	0.388416

Literatur: J. Bortz (1989) Lehrbuch der Statistik. Springer-Verlag. Kap.7,8 ▲

Bifaktorielle Kovarianzanalyse: Untersucht werden drei Programme über einen Lehrgegenstand (Faktor A, dessen Stufen sind hier identisch mit dessen Werten), Motivationsunterschiede sollen durch zwei verschiedene Instruktionen (Faktor B, vgl. A) erreicht werden. Zielgröße ist der Lernerfolg, als Kovariable wird das Merkmal Intelligenz erfasst (Quelle: Bortz (1977, p.479)). Bemerkung: Mit der ordinalen Zielgröße sollte nach Möglichkeit eher eine nicht-parametrische Auswertung unternommen werden!

Bifaktorielle Kovarianzanalyse ▲

Eingabedatei ist Bortz479.DAT
davon Variable 2 (Var_Y) mit Kovariable 1 (Var_X)

Faktor A: Stufe | Variable 3 (Fak_A)

Stufe	Variable 3 (Fak_A)
1	1
2	2
3	3

Faktor B: Stufe | Variable 4 (Fak_B)

Stufe	Variable 4 (Fak_B)
1	1
2	2

Variable 2	Kovariable	A: Index	Stufe	B: Index	Stufe
13.00000	5.00000	1	1	1	1
17.00000	6.00000	1	1	1	1
18.00000	6.00000	1	1	1	1
10.00000	4.00000	1	1	1	1
9.00000	3.00000	1	1	1	1
12.00000	5.00000	1	1	1	1
10.00000	5.00000	2	2	1	1
14.00000	7.00000	2	2	1	1
17.00000	7.00000	2	2	1	1
19.00000	9.00000	2	2	1	1
11.00000	6.00000	2	2	1	1
14.00000	6.00000	2	2	1	1
21.00000	8.00000	3	3	1	1
19.00000	7.00000	3	3	1	1
13.00000	5.00000	3	3	1	1
13.00000	4.00000	3	3	1	1
16.00000	7.00000	3	3	1	1
15.00000	6.00000	3	3	1	1
20.00000	7.00000	1	1	2	2
16.00000	6.00000	1	1	2	2
14.00000	4.00000	1	1	2	2
12.00000	4.00000	1	1	2	2
19.00000	6.00000	1	1	2	2
15.00000	5.00000	1	1	2	2
17.00000	6.00000	2	2	2	2
22.00000	8.00000	2	2	2	2

19.00000	7.00000	2	2	2	2
13.00000	5.00000	2	2	2	2
12.00000	5.00000	2	2	2	2
20.00000	8.00000	2	2	2	2
14.00000	5.00000	3	3	2	2
25.00000	6.00000	3	3	2	2
22.00000	5.00000	3	3	2	2
19.00000	5.00000	3	3	2	2
15.00000	4.00000	3	3	2	2
18.00000	5.00000	3	3	2	2

Stichprobenumfänge n(i,j):

Faktor B		Faktor A:					
		a = 1	a = 2	a = 3			
b = 1	6	6	6	6			
b = 2	6	6	6	6			

Durchschnitte:

Faktor B		Faktor A:					
		a = 1	a = 2	a = 3			
b = 1	13.167	14.167	16.167				
b = 2	16.000	17.167	18.833				

Tabelle der Kovarianzanalyse für Faktoren mit festen Effekten (Modell I):

Source	fg	SQ	DQ	F	p
Faktor A	2	101.9070	50.9535	16.0009	0.000021
Faktor B	1	111.1898	111.1898	34.9169	0.000002
A x B	2	22.9488	11.4744	3.6033	0.040026
Rest	29	92.3479	3.1844		
Total	34	328.3935			

Kovarianzanalyse für Faktoren mit zufälligen Effekten (Modelle II und III):

	Faktor A		Faktor B		A x B	
	F	p	F	p	F	p
A:fest B:fest	16.0009	0.000021	34.9169	0.000002	3.6033	0.040026
A:fest B:rand	4.4406	0.183803	34.9169	0.000002	3.6033	0.040026
A:rand B:fest	16.0009	0.000021	9.6902	0.089551	3.6033	0.040026
A:rand B:rand	4.4406	0.183803	9.6902	0.089551	3.6033	0.040026

Voraussetzungen:

Bartlett-Test (Homogenität der Varianzen):
 Prüfgröße $\chi^2 = 0.6721$ mit $df = 5$ ($p = 0.984467$)

Unterschiede der Stichprobenumfänge: 0.0%.
 Homogenität der Steigungen innerhalb der Stufen:
 Prüfgröße $F = 1.15$ mit $df = (5, 24)$ ($p = 0.364091$)

Schätzung der zusammengefassten Regression und Korrelation:
 Regressionskoeffizient $b = 2.54909$
 Korrelationskoeffizient $r = 0.87368$
 Prüfgröße $F = 77.40$ mit $df = (1, 24)$ ($p = 0.000000$)

Prüfung auf Gauß-Verteilung nicht vergessen!

Modell I : Multiple Vergleiche nach Scheffe (Korrigierte Differenzen):

Stufen	Faktor A	Differenz	F	df1	df2	p
1	2	2.7403	6.1371	2	29	0.005990
1	3	-1.6421	2.2038	2	29	0.128531
2	3	-4.3824	15.6962	2	29	0.000024

Stufen	Faktor B	Differenz	F	df1	df2	p
1	2	-3.5414	34.9169	1	29	0.000002

Literatur: J. Bortz (1989) Lehrbuch der Statistik. Springer-Verlag. Kap. 7 und 8 ▲

Hotelling's Einstichproben-T²-Test und Simultane Konfidenzintervalle: Beurteilung von 15 Studenten bezüglich 4 Variablen (Quelle: Kramer (1972, p. 52)) Kramer prüft die Hypothese $H_0(\mu_1=553, \mu_2=635, \mu_3=532, \mu_4=613)$; da **BiAS.** ausschließlich $H_0(\mu_1=\mu_2=\mu_3=\mu_4=0)$ prüft, wurden Kramers hypothetische Werte μ_i vor der Berechnung von den jeweiligen Zielgrößen subtrahiert (**BiAS.**' Formelinterpretier). Bemerkung: Die Zielgrößen sind eher qualitativer Natur, so dass an eine Auswertung mit Friedman's Test gedacht werden sollte!

Hotelling's Einstichproben-T²-Test & Simultane Konfidenzintervalle ▲

Eingabedatei ist Kramer54A.DAT

davon Variablen 5 (Verbal_0), 6 (MathSAT0), 7 (English0), 8 (MathAch0)

i	Verbal_0	MathSAT0	English0	MathAch0
1	-28.0000	-26.0000	-17.0000	18.0000
2	-146.0000	-20.0000	-45.0000	-59.0000
3	-15.0000	20.0000	-143.0000	12.0000
4	3.0000	26.0000	66.0000	-1.0000
5	137.0000	75.0000	79.0000	-38.0000
6	-126.0000	-63.0000	-128.0000	-69.0000
7	-67.0000	-20.0000	-149.0000	-22.0000
8	65.0000	-50.0000	102.0000	34.0000
9	-100.0000	-67.0000	-144.0000	-48.0000
10	51.0000	-3.0000	6.0000	28.0000
11	16.0000	-60.0000	60.0000	47.0000
12	3.0000	26.0000	-67.0000	-30.0000
13	-8.0000	29.0000	-45.0000	66.0000
14	9.0000	7.0000	-15.0000	45.0000
15	86.0000	44.0000	123.0000	-11.0000

Anzahl Merkmalsträger n = 15

Anzahl Bedingungen k = 4

Prüfgrößen:

Hotelling's T² = 1.2327

Prüfgröße F = 0.2421 mit df = (4, 11) (p = 0.908498)

Simultane Konfidenzintervalle: P = 0.95

i	Variable	X-Quer	Untere Grenze	Obere Grenze
1	Verbal_0	-8.000000	-91.235786	75.235786
2	MathSAT0	-5.466667	-51.284447	40.351112
3	English0	-21.133333	-120.725876	78.459213
4	MathAch0	-1.866667	-46.464039	42.730705

Prüfung auf Gauß-Verteilung mit Mudholkar-Test!

Literatur:

M.M. Tatsuoka (1971) Multivariate Analysis. Wiley. p. 76 ff.

C.Y. Cramer (1972) A first course in methods of multivariate analysis. Virginia Polytech. Inst. and State Univ.. p. 56 ff.

Hotelling's Zweistichproben-T²-Test & Simultane Konfidenzintervalle: In dem Lehrbuch von Kramer (1972, p.73) findet sich als Beispiel ein Vergleich zweier Fertigungstemperaturen bezüglich zweier Festigkeitsparameter von Stahl:

Hotelling's Zweistichproben-T²-Test & Simultane Konfidenzintervalle ▲

Eingabedatei ist Kramer73.DAT
davon Variablen 1 (YieldPoi), 2 (UltStren)
und Gruppen Var. 3 (Temperat) = 1 und 2

Gruppe 1

i	Var. 1	Var. 2
1	33.0000	60.0000
2	36.0000	61.0000
3	35.0000	64.0000
4	38.0000	63.0000
5	40.0000	65.0000

Gruppe 2

i	Var. 1	Var. 2
1	35.0000	57.0000
2	36.0000	59.0000
3	38.0000	59.0000
4	39.0000	61.0000
5	41.0000	63.0000
6	43.0000	65.0000
7	41.0000	59.0000

Durchschnitte, Standardabweichungen und Streuungen:

Gruppe	Variable	Xquer	Std.Abw.	Streuung	Fallzahl

1	YieldPoi	36.4000	2.7019	7.3000	5
	UltStren	62.6000	2.0736	4.3000	5

2	YieldPoi	39.0000	2.8868	8.3333	7
	UltStren	60.4286	2.7603	7.6190	7

Homogenität der Kovarianzmatrizen:

Box' Prüfgröße $\chi^2 = 0.3808$ mit $df = 3$ ($p = 0.944184$)

Prüfgrößen:

Hotelling's T² = 23.9117

Prüfgröße F = 10.7603 mit $df = (2, 9)$ ($p = 0.004106$)

Simultane Konfidenzintervalle: P = 0.95

Variable	X-Quer1	X-Quer2	Diff X-Quer	Untere Grenze	Obere Grenze
YieldPoi	36.4000	39.0000	-2.6000	-7.6680	2.4680
UltStren	62.6000	60.4286	2.1714	-2.3456	6.6884

Prüfung auf Gauß-Verteilung mit Mudholkar-Test!

M.M. Tatsuoka (1971) Multivariate Analysis. Wiley. p. 81 ff.
C.Y. Cramer (1972) A first course in methods of multivariate analysis. Virginia Polytech. Inst. and State Univ.. p. 80 ff.

Varianzanalyse mit Messwiederholungen: Es wird geprüft, ob der Haut-Widerstand Tageschwankungen unterliegt, wozu eine Gruppe von Probanden morgens, mittags und abends untersucht wurde (Bortz (2005, p.412)). Eine graphische Darstellung ist mit Einzelverläufen und Box-Plots möglich, weitere Berechnungen finden sich in "Deskriptives".

Varianzanalyse mit Messwiederholungen ▲

Eingabedatei ist ..\VALIDE\BORTZ412.DAT
davon Variablen 1 (morgens), 2 (mittags), 3 (abends)

Eingelesene Daten:

i	morgens	mittags	abends
1	7.0000	7.0000	6.0000
2	5.0000	6.0000	8.0000
3	8.0000	9.0000	5.0000
4	6.0000	8.0000	6.0000
5	7.0000	7.0000	5.0000
6	7.0000	9.0000	7.0000
7	5.0000	10.0000	6.0000
8	6.0000	7.0000	4.0000
9	7.0000	8.0000	6.0000
10	5.0000	7.0000	5.0000

Tabelle der Varianzanalyse:

Source	SQ	MQ	df	F	p
zwischen	11.6333	1.2926	9	0.9831	0.485678
innerhalb	45.3333	2.2667	20		
- Faktor	21.6667	10.8333	2	8.2394	0.002881
- Rest	23.6667	1.3148	18		
Gesamt	56.9667	1.9644	29		

Zum Vergleich der Faktorstufen wird der p-Wert bzgl. "Faktor" verwendet.

Prüfung der Voraussetzungen - Homogenität der Varianzen und der Korrelationen zwischen den Stufen des Faktors per Box-Test:

Box' Chi² = 0.925 mit df = 4 und p = 0.920965 Ok!

Test auf Gauß-Verteilung der 3 Variablen (Dallal-Wilkinson-korrigiert):

morgens: Kolmogoroff-Smirnoff's delta = 0.2456 (p = 0.088720) Cave!
mittags: Kolmogoroff-Smirnoff's delta = 0.2424 (p = 0.098658) Cave!
abends: Kolmogoroff-Smirnoff's delta = 0.2301 (p > 0.100000) Ok!

Mudholkar's Test auf multivariate Gauß-Verteilung:

Mudholkar's Prüfgröße N = 1.6136 : p = 0.106613 (asymptotisch)

Multiple Vergleiche - Bonferroni-Holm-korrigierte t-Tests und Scheffé-Tests:

Nr	Mult. Paarvergleich	p-unkorr	p-Bonferroni	p-Bonf.Holm	p-Scheffé
1	morgens....mittags.	0.009106	0.027318	0.018212	0.030195
2	morgens....abends..	0.380856	1.000000	0.380856	0.629252
3	mittags....abends..	0.004763	0.014289	0.014289	0.004035

Literatur: J. Bortz (2005) Lehrbuch der Statistik, Springer. pp. 409-414. ▲

Cross-Over-Analyse (Grizzle-Modell): Die Cross-Over-Analyse vergleicht zwei Behandlungen A und B in zwei Gruppen mit den Behandlungssequenzen AB und BA (Grizzle (1969)).

Bitte beachten: Nach Eingabe der Gruppenvariablen sind die beiden Namen der Variablen einzugeben, die den beiden Perioden zugeordnet sind: Die Eingabeaufforderung lautet: Welche Variablen werden getestet: Periode1: und nach Eingabe Periode 2:. Es handelt sich hierbei um Perioden, nicht um Behandlungen: Zum Beispiel erhielt Gruppe 1 in der ersten Periode die Behandlung A, die Gruppe 2 dagegen B, während in der 2. Periode umgekehrt Gruppe 1 B und Gruppe 2 A erhielt!

Cross-Over (Grizzle-Modell)

Eingabedatei ist Grizzle1.DAT
davon Variablen 1 (Periode1), 2 (Periode2)
und Gruppen Var. 3 (GruppNum) = 1 und 2

Gruppe 1

i	Var. 1	Var. 2
1	0.2000	1.0000
2	0.0000	-0.7000
3	-0.8000	0.2000
4	0.6000	1.1000
5	0.3000	0.4000
6	1.5000	1.2000

Gruppe 2

i	Var. 1	Var. 2
1	1.3000	0.9000
2	-2.3000	1.0000
3	0.0000	0.6000
4	-0.8000	-0.3000
5	-0.4000	-1.0000
6	-2.9000	1.7000
7	-1.9000	-0.3000
8	-2.9000	0.9000

Homogenität der Varianzen (zweiseitig):

Prüfgröße F = 1.2363 mit df1 = 7 und df2 = 5 (p = 0.843713)

Nacheffekte:

Prüfgröße F = 4.5708 mit df1 = 1 und df2 = 12 (p = 0.053788)

Prüfung auf Gauß-Verteilung (Wilk-Shapiro-Test) nicht vergessen!

Behandlungseffekte:

Im Cross-Over:

Prüfgröße F = 2.8610 mit df1 = 1 und df2 = 12 (p = 0.116532)

Erste Periode:

Prüfgröße F = 5.1736 mit df1 = 1 und df2 = 12 (p = 0.042090)

Literatur:

J.E. Grizzle (1965) Two-period cross-over design and its use in clinical trials. Biometrics 21, pp. 467-480.

J.E. Grizzle (1974) Correction Note. Biometrics 21, p. 727.

G.G. Koch (1972) The use of nonparam. methods in the stat. Analysis of the two-period change-over design. Biometrics 28, pp. 577-584.

A.9 Nicht-parametrische Statistik



Wilcoxon's matched-pairs-Test
Tukey-Konfidenzintervall
Wilcoxon-Mann-Whitney-Test ("U-Test")
Asymptotisches Konfidenzintervall zum Mann-Whitney-Schätzer
Moses-Konfidenzintervall zum Hodges-Lehmann-Schätzer
Van-Elteren's stratifizierter U-Test
Kolmogoroff-Smirnoff-Test
Mood's Median-Test
Kruskal-Wallis-Test, mit Headrick's exaktem p-Wert
Multiple Vergleiche nach Dunn, Nemenyi und Conover-Iman
Jonckheere-Terpstra-Test auf Trend
Van der Waerden's Normalscore-Test
Nicht-parametrischer Levene-Test auf Varianzgleichheit
Mantel-Haenszel- und Mantel-Haenszel-Zimmermann-Test
Friedman-Test mit Wilcoxon-Wilcox- und Conover-Vergleichen
Page-Test auf Trend
Wilcoxon-Wilcox-Vergleiche
Schaich-Hamerle-Vergleiche
Dixon und Mood's Vorzeichentest
Cochran's Q-Test
Test auf Gleichverteilung
Anpassungstest (χ^2 -Test, Woolf's G-Test)
Koch's nicht-parametrische Cross-Over-Analyse
Bifaktorielle Rang-Varianzanalyse (Kubinger)

Wilcoxon's matched-pairs-Test: In einer Studie ist der Erfolg einer Tranquilizer-Behandlung zu beurteilen. Als Zielgröße wird der Hamilton-Faktor IV verwendet (vgl. Hollander und Wolfe (1973, p.29)):

Wilcoxon-Test ('matched pairs').

Eingabedatei ist Holl129.DAT
davon Variable 1 (HFvor) - Variable 2 (HFnach)

i	Var 1	Var 2	Diff
1:	1.8300	0.8780	0.9520
2:	0.5000	0.6470	-0.1470
3:	1.6200	0.5980	1.0220
4:	2.4800	2.0500	0.4300
5:	1.6800	1.0600	0.6200
6:	1.8800	1.2900	0.5900
7:	1.5500	1.0600	0.4900
8:	3.0600	3.1400	-0.0800
9:	1.3000	1.2900	0.0100

Stichprobenumfang: n = 9
Ungleiche Paare: n' = 9

Prüfgröße R^{\wedge} = 5.00 (p = 0.039063 exakter Wert)
Prüfgröße Z = 2.0732 (p = 0.038152 approximativ)

Hodges-Lehmann-Schätzer = 0.4600

Tukey-Konfidenzintervalle:

P = 0.9000: (0.0100 , 0.7860)
P = 0.9500: (0.0100 , 0.7860)
P = 0.9900: (-0.1135 , 0.9870)

Literatur:

L. Sachs (1992) *Angewandte Statistik*, Springer. Kap.422.

H. Bünung, G. Trenkler (1978) *Nicht-param. Methoden*, Springer. p.192

Tukey-Konfidenzintervall: Konfidenzintervall für den Median von Messungen des Quotienten der Masse der Erde und der Masse des Mondes (aus: Hollander und Wolfe (1973, p.51). Da der matched-pairs-Test als Paardifferenzen-Test konzipiert ist, wird im Beispiel als zweite Variable Null eingegeben: Der WMP-Test macht in dieser Form natürlich keinen Sinn, wesentlich an der Ausgabe sind nur die Konfidenzintervalle!

Wilcoxon-Test ('matched pairs') ▲

Eingabedatei ist Holl151.DAT

davon Variable 1 (QuotE/M)- Variable 2 (Null)

i	Var 1	Var 2	Diff
1:	81.3001	0.0000	81.3001
2:	81.3015	0.0000	81.3015
3:	81.3006	0.0000	81.3006
4:	81.3011	0.0000	81.3011
5:	81.2997	0.0000	81.2997
6:	81.3005	0.0000	81.3005
7:	81.3021	0.0000	81.3021

Stichprobenumfang: n = 7

Ungleiche Paare: n' = 7

Prüfgröße R^{\wedge} = 0.00 (p = 0.015625 exakter Test)

Prüfgröße Z = 2.3664 (p = 0.017960 approximativ)

Hodges-Lehmann-Schätzer = 81.3008

Tukey-Konfidenzintervalle:

P = 0.9000: (81.3001 , 81.3016)
P = 0.9500: (81.3001 , 81.3016)

Literatur:

L. Sachs (1992) *Angewandte Statistik*, Springer. Kap.422

H. Bünung, G. Trenkler (1978) *Nicht-param. Methoden*, Springer. p.192

Wilcoxon-Mann-Whitney-Test ("U-Test"): Vor und nach einer Änderung in einem Prozess werden die Zeiten zur Ausführung einer Verrichtung gemessen (vgl. Graf et al. (1987, p.399)). Es ist zu bemerken, dass hier ein Parallelgruppenversuch vorliegt, mithin kann jeder Proband in nur *genau einer* Gruppe sein. Als Alternative könnte der Vergleich auch im Sinne des matched-pairs-Tests durchgeführt werden, so dass jeder Proband mit *einem* Wertepaar in die Auswertung eingehen würde.

Die Prüfgrößen Z und U werden bei Bindungen („Ties“) auf Grund von mittleren Rängen („Midranks“) berechnet. Der approximative Test via Z enthält eine Tie-Korrektur, der „exakte p-Wert“ wird mit Hilfe des Verfahrens von Zimmermann (1985) berechnet und ist somit nicht Tie-korrigiert. Die p-Werte des exakten Tests können bei Bindungen zwischen den Gruppen heuristisch korrigiert werden. Weitere Validierungen wurden mit Hilfe von SPSS durchgeführt.

Wilcoxon-Mann-Whitney-Test



Eingabedatei ist Graf399.DAT

davon Variable 1 (Zeit) und Gruppen Var. 2 = 1, 2

Gruppe 1	Gruppe 2
1: 11.0000	1: 14.0000
2: 15.0000	2: 19.0000
3: 21.0000	3: 25.0000
4: 23.0000	4: 29.0000
5: 23.0000	5: 36.0000
6: 27.0000	6: 39.0000
7: 28.0000	7: 45.0000
8: 29.0000	8: 49.0000
9: 35.0000	9: 55.0000
10: 48.0000	10: 60.0000

Rangsumme Stichprobe 1 = 82.50, Normiert: 8.25

Rangsumme Stichprobe 2 = 127.50, Normiert: 12.75

Test der Nullhypothese:

Prüfgröße U = 27.5 mit p = 0.105122 (exakter Test)

Prüfgröße Z = 1.702 mit p = 0.088733 (approximativ)

Exakter Test: Unkorrigierter p-Wert bei Bindungen, Zimmermann (1985)

Hodges-Lehmann-Schätzer = -11.5000

Moses-Konfidenzintervalle für Gruppe1-Gruppe2:

P = 0.9000: (-24.0000 , 0.0000)

P = 0.9500: (-26.0000 , 3.0000)

P = 0.9900: (-31.0000 , 9.0000)

	Median	Minimum	Maximum	n
-----+-----				
Gruppe 1	25.0000	11.0000	48.0000	10
Gruppe 2	37.5000	14.0000	60.0000	10

Literatur:

L. Sachs (2003) Angewandte Statistik, 11. Aufl. Springer, Kap. 394.

H. Zimmermann (1985) Exact Calculation of Permutational Distributions for Two Independent Samples. Biom. J. Vol. 27, No. 4, pp. 431-4.

Moses-Konfidenzintervall zum Hodges-Lehmann-Schätzer: Bei der Untersuchung der Membrangängigkeit von Tritium-haltigen Wasser soll ein Konfidenzintervall für den Unterschied zweier Gestationsgruppen errechnet werden (Hollander und Wolfe (1973, p.69)). Das nicht-parametrische Moses-Konfidenzintervall wird im Kontext des U-Tests errechnet:

Wilcoxon-Mann-Whitney-Test



Eingabedatei ist Holl169.DAT

davon Variable 1 (Pd(E-4cm/s))

und Gruppe(n) Var. 2 = 1, 2

Gruppe 1
1: 0.8000
2: 0.8300
3: 1.8900
4: 1.0400
5: 1.4500
6: 1.3800
7: 1.9100
8: 1.6400
9: 0.7300
10: 1.4600

Gruppe 2
 1: 1.1500
 2: 0.8800
 3: 0.9000
 4: 0.7400
 5: 1.2100

Rangsumme Stichprobe 1 = 90.00, Normiert: 9.00
 Rangsumme Stichprobe 2 = 30.00, Normiert: 6.00

Test der Nullhypothese:
 Prüfgröße U = 15.0 mit p = 0.254412 (exakter Test)
 Prüfgröße Z = 1.225 mit p = 0.220671 (approximativ)

Hodges-Lehmann-Schätzer = 0.3050

Moses-Konfidenzintervalle zum Hodges-Lehmann-Schätzer für Gruppe1-Gruppe2:

P = 0.9000: (-0.0800 , 0.7200)
 P = 0.9500: (-0.1500 , 0.7600)
 P = 0.9900: (-0.3800 , 1.0100)

Die Intervalle gelten für $n_1 > 9$ und/oder $n_2 > 9$, sonst näherungsweise.

	Median	Minimum	Maximum	n
Gruppe 1	1.4150	0.7300	1.9100	10
Gruppe 2	0.9000	0.7400	1.2100	5

Literatur: L. Sachs (1992) Angewandte Statistik, Springer, Kap.394.

Konfidenzintervall zum Mann-Whitney-Schätzer: Wellek (1993) gibt ein Beispiel zum Vergleich zweier Schmerzmittel. Diese Daten wurden zur Überprüfung des MW-Schätzers und der Varianzschätzung verwendet. (Zu der Varianzschätzung des MWS vgl. Bamber (1975), ab $n_1, n_2 > 200$ wird aus Laufzeitgründen die Varianz von Sen (1960) verwendet, vgl. dazu auch - im Kontext der AUC-Berechnung - DeLong (1983).) Die Konfidenzintervalle können z.B. auch für Äquivalenztests verwendet werden, wie in dem Buch von Wellek (1993) dargestellt wird:

Wilcoxon-Mann-Whitney-Test ▲

Eingabedatei ist ... \WELLEK.dat
 davon Variable 3 (Diff)
 und Gruppen-Var. 2 (Gruppe) = 1 2

Gruppe 1
 1: -0.0546
 2: 0.0540
 3: -0.0949
 4: 0.1370
 5: 0.0264
 6: -0.0913
 7: -0.3493
 8: 0.0472
 9: 0.0529
 10: -0.2722

Gruppe 2
 1: 0.1072
 2: 0.3516
 3: -0.1889
 4: -0.0319
 5: 0.0264
 6: 0.1480

7: 0.1875
 8: 0.0672
 9: -0.1881
 10: 0.0099

Test der Nullhypothese:

Prüfgröße U = 33.0 mit p = 0.217563 (exakter Test)
 Prüfgröße Z = 1.285 mit p = 0.198765 (approximativ)

Mann-Whitney-Schätzer für $p=P(X1<X2) = 0.670000$

Asymptotische Konfidenzintervalle zum Mann-Whitney-Schätzer:

P = 0.9000: (0.485300 , 0.854700)
 P = 0.9500: (0.449916 , 0.890084)
 P = 0.9800: (0.408775 , 0.931225)
 P = 0.9900: (0.380761 , 0.959239)

Hodges-Lehmann-Schätzer für 'X1-X2' = -0.0890

Moses-Konfidenzintervalle zum Hodges-Lehmann-Schätzer:

P = 0.9000: (-0.2393 , 0.0297)
 P = 0.9500: (-0.2788 , 0.0583)
 P = 0.9800: (-0.2987 , 0.0933)
 P = 0.9900: (-0.3252 , 0.0977)

	Median	Minimum	Maximum	n	Rangsumme	Normiert
Stichprobe 1	-0.0141	-0.3493	0.1370	10	88.0	8.8
Stichprobe 2	0.0468	-0.1889	0.3516	10	122.0	12.2

Literatur: L. Sachs (1999) Angew. Statistik, 9. Aufl. Springer, Kap. 394.
 M. Hollander & D.A. Wolfe (1973) Nonpar. Statistical Methods. Wiley NY.
 St. Wellek (1993) Stat. Methoden zum Nachweis von Äquivalenz. Fischer.

Van Elteren's stratifizierter U-Test: Der van Elteren-Test ist eine stratifizierte Variante des Wilcoxon-Mann-Whitney-U-Tests.

Ein Beispiel findet sich in <http://support.sas.com/publishing/pubcat/chaps/59390.pdf>: „As an illustration, consider a subset of the data collected in a urinary incontinence trial comparing an experimental drug to a placebo over an 8-week period. The primary endpoint in the trial was a percent change from baseline to the end of the study in the number of incontinence episodes per week. Patients were allocated to three strata according to the baseline frequency of incontinence episodes.”

Unten wurden die Daten dieses Beispiels verwendet, aber wegen des Umfangs nicht explizit aufgeführt: Die Daten finden sich in ...\\Valide\vanElteren.DAT“). Die Verifikation des van-Elteren-Tests erfolgte hiermit und mit zahlreichen anderen Beispielen anhand von SAS.

Stratifizierter van-Eltern-Test ▲

Eingabedatei ist ...\\Valide\vanElteren.DAT
 davon Variable 3 (change)
 und Gruppenvar. 1 (Arm_num) = 1,2

Stratifizierungsvariable stratum = 1,2,3

Die Daten zu dem gerechneten Beispiel sind sehr umfangreich (92+91=183 Fälle) und werden deshalb nicht aufgelistet. Details bei support.sas.com bzw. in der Datei ...\\Valide\vanElteren.DAT im Verzeichnis „Valide“.

Zweiseitige Wilcoxon-Mann-Whitney-Tests für alle Strata:

Stratum	n	m	MW-Schätzer	Z	p
1	40	39	0.458974	-0.636	0.524580
2	32	33	0.354167	-2.023	0.043112
3	20	19	0.322368	-1.897	0.057806

Zweiseitiger van-Elteren-Test:

n	m	Z	p	Wichtung der Strata
92	91	-2.505	0.012234	van Elteren (1960)

n = Fallzahl in Gruppe Arm_num=1, m = Fallzahl in Gruppe Arm_num=2
Ein negativer Wert von Z bedeutet höhere Werte in Gruppe Arm_num=1

Literatur:

P.H. van Elteren (1960) On the combination of independent two sample tests of Wilcoxon. Bull. of the Inst. of Internat. Statist. 37, pp. 351-61.
Y.D. Zhao (2006) Sample size estimation for the van Elteren test - a stratified Wilcoxon-Mann-Whitney test. Statist. in Med. 25, pp. 2675-87.

Kolmogoroff-Smirnoff-Test: Hartung (1982, p.519) führt als Beispiel zum KS-Test einen Vergleich zweier Amperemeter an. (Bitte beachten Sie, dass die Auswertung besser mit Hilfe des Verfahrens von Bland und Altman vorgenommen werden sollte!) Der asymptotische Test wurde verifiziert gemäß SAS und SPSS.

Kolmogoroff-Smirnoff-Test

Eingabedatei ist Hartung519.DAT
davon Variable 1 und Gruppe(n) Var. 2 = 1, 2

Gruppe 1

1: 1.2100
2: 1.2200
3: 1.2300
4: 1.2400
5: 1.2500
6: 1.2600
7: 1.2600
8: 1.2700
9: 1.2700
10: 1.2800

Gruppe 2

1: 1.2000
2: 1.2100
3: 1.2300
4: 1.2400
5: 1.2400
6: 1.2400
7: 1.2500
8: 1.2500
9: 1.2600
10: 1.2700
11: 1.2700
12: 1.2800
13: 1.2900
14: 1.2900
15: 1.2900

Mediane und Extremwerte:

	Median	Minimum	Maximum	n
Stichprobe 1	1.2550	1.2100	1.2800	10
Stichprobe 2	1.2500	1.2000	1.2900	15

Kolmogoroff-Smirnoff's delta S = 0.2000 mit p > 0.1000 (exakter Test)

Hodges' asymptotischer Test für delta mit p = 0.970041 (asymptotisch)

Literatur:

L. Sachs (2003) Angewandte Statistik, Kapitel 393, 11. Aufl. Springer-Verlag

P.J. Kim (1969) On the exact and approx. sampling distribution of the two sample Kolmogorov-Smirnov criterion Dmn. JASA 1969, Vol. 64, pp. 1625-1637.

J.L. Hodges Jr (1957) The Significance Probability of the Smirnov Two-Sample Test, Arkiv for Matematik, 3, pp. 469-486.

Kruskal-Wallis-Test mit multiplen Vergleichen: Vergleich von 4 Gruppen bezüglich einer Zielgröße, die nicht-Gauß-verteilt ist bzw. ordinalen Charakter trägt (Sachs (1992, p.394)).

Bitte beachten: Sachs' Prüfgröße ist nicht tie-korrigiert; **BiAS**. rechnet bei Bindungen grundsätzlich mit tie-Korrektur! In Abweichung von Dunn (1964) wird im Test die (günstigere) α -Adjustierung nach Holm (1979) durchgeführt. Der quasi-exakte p-Wert nach Headrick (2003) für "kleine" Fallzahlen wurde durch Beispiele aus der Originalarbeit und per Tabellen (Sachs (2003) o.a.) verifiziert.

Kruskal-Wallis-Test mit multiplen Vergleichen

Eingabedatei ist Sachs394.DAT

davon Variable 1 (Messwert), Alle Gruppen Var. 2 (Gruppe)

Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
1: 121.0	1: 183.0	1: 127.0	1: 73.0
2: 148.0	2: 496.0	2: 251.0	2: 19.0
3: 153.0	3: 101.0	3: 470.0	3: 58.0
4: 114.0	4: 356.0	4: 163.0	4: 101.0
5: 108.0	5: 262.0	5: 304.0	5: 94.0
	6: 89.0		

Kruskal-Wallis' Prüfgröße H für den Globaltest:

H = 11.530: Chi²{df=3}-verteilt mit p = 0.009179

Headrick's quasi-exakter p-Wert: p = 0.002727 mit KI=[0.002543;0.002920]

	Median	Minimum	Maximum	n	Rangsumme	Normiert
Stichprobe 1	121.0000	108.0000	153.0000	5	52.0	10.4
Stichprobe 2	222.5000	89.0000	496.0000	6	82.5	13.8
Stichprobe 3	251.0000	127.0000	470.0000	5	79.0	15.8
Stichprobe 4	73.0000	19.0000	101.0000	5	17.5	3.5

Multiple Dunn-Vergleiche (Bonferroni-Holm-korrigiert):

Nr	Stichproben	Z	p-unkorr	p-Bonferroni	p-Bonf.Holm
1	1 versus 2	0.892	0.372598	1.000000	0.745196
2	1 versus 3	1.376	0.168807	1.000000	0.506421
3	1 versus 4	1.758	0.078699	0.472196	0.314797
4	2 versus 3	0.546	0.585329	1.000000	0.745196
5	2 versus 4	2.729	0.006353	0.038120	0.031766
6	3 versus 4	3.134	0.001722	0.010335	0.010335

Multiple Conover-Iman-Vergleiche (Bonferroni-Holm-korrigiert, $df\{t\}=17$):

Nr	Stichproben	t	p-unkorr	p-Bonferroni	p-Bonf.Holm
1	1 versus 2	1.264	0.223429	1.000000	0.446859
2	1 versus 3	1.950	0.067849	0.407096	0.203548
3	1 versus 4	2.492	0.023335	0.140008	0.093339
4	2 versus 3	0.773	0.449996	1.000000	0.449996
5	2 versus 4	3.866	0.001239	0.007435	0.006196
6	3 versus 4	4.442	0.000358	0.002145	0.002145

Dunn: Signifikanz am Niveau alpha für alle Vergleiche mit Holm's $p\text{-}(BH)\leq\alpha$
 Conover und Iman: Approx. p-Werte für >3 Gruppen, Signifikanz bei $p\text{-}(BH)\leq\alpha$

Literatur:

W.H. Kruskal, W.A. Wallis (1952) Use of ranks in one-criterion variance analysis. J.Amer. Stat. Ass. 47, pp.583-621 und 48 (1953), pp.907-911.

O.J. Dunn (1964) Multiple comparisons using rank sums. Technometrics 6, pp. 241-52.

S. Holm (1979) A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6, pp. 65-70.

W.J. Conover (1999) Practical Nonparametric Statistics (3rd ed.), John Wiley.

Die multiplen Vergleiche des Dunn-Tests können mit Hilfe der Originalarbeit verifiziert werden. Die verwendeten Daten sind sehr umfangreich und können hier nicht wiedergegeben werden. Zur Holm'schen Methode vgl. "Mehrere Nullhypothesen".

Conover-Iman wurde mit Hilfe von StatsDirect verifiziert, das allerdings die Berechnungen ohne Tie-Korrektur ausführt. Weitere Verifikation mit: Viv Bewick, Liz Cheek and Jonathan Ball (2004) Statistics review 10: Further nonparametric methods, Crit Care. 2004; 8(3): 196-199.

Zur Berechnung von exakten p-Werten für "kleine" Fallzahlen finden Sie in der Arbeit von Headrick (2003) Hinweise und Beispiele zur Verifikation.

Jonckheere-Terpstra-Test: Vergleich von mehreren Gruppen bezüglich einer Zielgröße, die nicht-Gauß-verteilt ist bzw. ordinalen Charakter trägt (Sachs (2003), Hollander und Wolfe (1999)).

Bitte beachten Sie, dass die Darstellung in Sachs (2003), ebenso das dort verwendete Beispiel, *keine* Tie-Korrektur verwendet. Im Folgenden ein Beispiel aus Hollander und Wolfe (1999), weitere Verifizierungen mit dem Programm SPSS:

Jonckheere-Terpstra-Test auf Trend ▲

Eingabedatei ist ..\Hollander205.DAT
 davon Variable 1 (Number) und Gruppen-Var. 2 (Group) = 1 2 3

Errechnete J-T-Statistik J: 29.0
 Erwartungswert $\mu(J)$: 54.0
 Standardabweichung $SD(J)$: 12.259

Positiver Trend, da $\mu(J)-J>0$: Alternativ-Hypothese ist $H_a(\mu_1\leq\mu_2\leq\dots\leq\mu_k)$

Prüfgröße $Z = 2.039$ mit $p = 0.020710$ (einseitiger Test mit Tie-Korrektur)
 Quasi-exakter p-Wert: $p = 0.020987$ (per Headrick, $KI=[0.020477;0.021506]$)

Sukzessive Gruppenvergleiche (alle p-Werte sind einseitig gemäß Relation):

i	Stichproben	Rel	Z	p uncorr	p Bonferroni	p Bonf.-Holm
1	1 versus 2	<	0.647	0.258706	0.517413	0.258706
2	2 versus 3	<	1.373	0.084854	0.169707	0.169707

L. Sachs (2003) Angewandte Statistik, Kapitel 396, 11. Aufl. Springer-Verlag
 M. Hollander, D.A. Wolfe (1999) Nonparametric Statistical Methods, Wiley Inc

Van der Waerden's Normalscores-Test: Der Van-der-Waerden-Test vergleicht wie der Kruskal-Wallis-Test k Gruppen. Die Ränge aus dem KW-Test werden in Quantilen der Standard-Normalverteilung transformiert ("Normalscores") und damit alle weiteren Berechnungen ausgeführt.

Zur Validierung wird ein Beispiel aus Conover (1999, Seite 399) verwendet. Vorsicht beim Vergleich: Conover verwendet tabellarische Quantilen mit einer Genauigkeit von nur drei Nachkommastellen.

Van der Waerden's Normalscores-Test ▲

Eingabedatei ist ... \VanDerWaerden.DAT
davon Variable 1 (Obs) und Gruppenvar. 2 (Group) = 1,2,3,4

Eingelesene Daten:

Nr.:	Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
1:	83.0	91.0	101.0	78.0
2:	91.0	90.0	100.0	82.0
3:	94.0	81.0	91.0	81.0
4:	89.0	83.0	93.0	77.0
5:	89.0	84.0	96.0	79.0
6:	96.0	83.0	95.0	81.0
7:	91.0	88.0	94.0	80.0
8:	92.0	91.0		81.0
9:	90.0	89.0		
10:		84.0		

Prüfgröße T1 für den Globaltest:

T1 = 25.180: Chi²{df=3}-verteilt mit p = 0.000014

	Median	Minimum	Maximum	Range	n
Stichprobe 1	91.0	83.0	96.0	13.0	9
Stichprobe 2	86.0	81.0	91.0	10.0	10
Stichprobe 3	95.0	91.0	101.0	10.0	7
Stichprobe 4	80.5	77.0	82.0	5.0	8

Effektstärke: $\eta^2 = T1/(N-1) = 0.7630$

Rasch: $\eta^2=0.01$ geringer Effekt, 0.06 mittlerer Effekt, 0.14 großer Effekt

Lageskizze der 4 Gruppen mit Extremwerten und Median="x":

```

1:          -----x-----
2:      -----x-----
3:          -----x-----
4: -----x-----

```

Multiple Vergleiche (Bonferroni-Holm-korrigiert, df(t)=30):

Nr	Stichproben	t	p-un CORR	p-Bonferroni	p-Bonf.Holm
1	1 versus 2	2.450	0.020315	0.121893	0.020315
2	1 versus 3	3.237	0.002942	0.017652	0.005884
3	1 versus 4	6.711	0.000000	0.000001	0.000001
4	2 versus 3	5.595	0.000004	0.000026	0.000017
5	2 versus 4	4.501	0.000095	0.000570	0.000285
6	3 versus 4	9.453	0.000000	0.000000	0.000000

Signifikanz am Niveau alpha für alle Vergleiche mit Holm's p-(BH)≤alpha

Literatur:

B.L. van der Waerden (1952,1953) Order tests for the two-sample problem and their power. Indagationes Mathematicae 55, 453-458 und ibid. 56, 80,303-316
W.J. Conover (1999) Practical Nonparameteric Statistics 3rd ed. Wiley
B. Rasch et al. (2010) Quantitative Methoden. Springer-Verlag Heidelberg

Nicht-parametrischer Levene-Test auf Varianzgleichheit: Der Test stellt eine Modifikation des „klassischen“ Levene-Tests zur Überprüfung der Gleichheit der Varianzen von zwei oder mehr Normalverteilungen dar: Die Modifikation von Nordstokke und Zumbo (2010) beruht auf Rangordnungsstatistiken, die Prüfgröße ist analog zu Levene's F definiert. Im Artikel finden sich ausführliche Betrachtungen und Simulationen zur Irrtumswahrscheinlichkeit α und zur Power.

Mangels bereits durchgerechneter Beispiele werden hier exemplarisch die Daten zum Van der Waerden-Test verwendet. Eine Überprüfung bzw. Validierung ist damit einfach möglich: Nach Rangtransformation der Daten (in BiAS. per „Editor, Rangtransformation“, das Resultat ist in der Beipieldatei in der Spalte „Rang_Obs“ enthalten) können die Ränge dem parametrischen Levene-Test zugeführt. Letzterer wurde u.a. mit SPSS und Lehrbuchbeispielen validiert.

Nicht-parametrischer Levene-Test auf Varianzgleichheit ▲

Eingabedatei ist ..\VanDerWaerden.DAT
davon Variable 1 (Obs) und Gruppenvar. 2 (Group) = 1,2,3,4

Eingelesene Daten:

Nr.:	Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
1:	83.0	91.0	101.0	78.0
2:	91.0	90.0	100.0	82.0
3:	94.0	81.0	91.0	81.0
4:	89.0	83.0	93.0	77.0
5:	89.0	84.0	96.0	79.0
6:	96.0	83.0	95.0	81.0
7:	91.0	88.0	94.0	80.0
8:	92.0	91.0		81.0
9:	90.0	89.0		
10:		84.0		

		Median	Minimum	Maximum	Range	n
-----	-----	-----	-----	-----	-----	-----
Stichprobe 1		91.0000	83.0000	96.0000	13.0000	9
Stichprobe 2		86.0000	81.0000	91.0000	10.0000	10
Stichprobe 3		95.0000	91.0000	101.0000	10.0000	7
Stichprobe 4		80.5000	77.0000	82.0000	5.0000	8

Levene's Prüfgröße $F = 1.887$ mit $df = (3,30)$ und $p = 0.153138$

>> Alle Prüfgrößen F sind modifiziert nach Nordstokke und Zumbo (2010).

Paarweise Vergleiche der Streuungen der 4 Gruppen:

Vergleiche		df1	df2	Levene's F	Levene's p	
-----	-----	-----	-----	-----	-----	
1	2		1	17	0.306	0.587356
1	3		1	14	2.093	0.169971
1	4		1	15	4.594	0.048894
2	3		1	15	1.057	0.320284
2	4		1	16	3.054	0.099697
3	4		1	13	0.491	0.495832

>> Cave: Die multiplen Vergleiche beinhalten **k e i n e** alpha-Adjustierung.

Literatur:

H. Levene (1960) Robust Tests for Equality of Variance. In: I. Olkin (ed.) Contributions to Probability and Statistics. Stanford Univ Press, pp. 278-92
D.W. Nordstokke, B.D. Zumbo (2010) A New Nonparametric Levene Test for Equal Variances. *Psicologica: Int. J. Methodol. Experim. Psych.* 31, 2, pp. 401-30.

Mantel-Haenszel-Test und Mantel-Haenszel-Zimmermann-Test: Die Tests dienen dem Vergleich von 2 oder mehr Gruppen bei einer ordinalen Zielgröße mit wenigen Stufen. Im Unterschied zu der Testvariante in "Kontingenztafeln" rechnen beide Funktionen mit den Originalwerten, setzen also keine Kontingenztafel voraus (diese wird vom Programm ermittelt). Die Tests sind in diesem Falle effizienter als der U-Test bzw. der Kruskal-Wallis-Test! Zur Validierung im 2-Gruppen-Fall kann das Beispiel von DeJonge (vgl. Kontingenztafeln) herangezogen werden, in Ermangelung eines Beispiels in Zimmermann (1984) wird der k-Gruppenfall an Hand der Daten von **BIAS**.¹ Testdatensatz Test.DAT verifiziert:

Mantel-Haenszel-Zimmermann-Test ▲

Eingabedatei ist Test.DAT

davon Variable 2 (VarZwei)
und Gruppen-Var. 3 (GruppNum) = 1 2 3

Gruppe 1	Gruppe 2	Gruppe 3
1: 2.0000	1: 20.0000	1: 35.0000
2: 3.0000	2: 25.0000	2: 33.0000
3: 5.0000	3: 30.0000	3: 32.0000
4: 5.0000		4: 39.0000
5: 4.0000		

Scores der Zielgröße (Spalten):

2 3 4 5 20 25 30 32 33 35 39

Ermittelte Tafel (Zeilen=Gruppen):

1	1	1	2	0	0	0	0	0	0	0	0	5
0	0	0	0	1	1	1	0	0	0	0	0	3
0	0	0	0	0	0	0	1	1	1	1	1	4
-----+-----												
1	1	1	2	1	1	1	1	1	1	1	1	12

Prüfgröße: Chi²=10.5977 mit fg=2 und p=0.004997

Literatur:

N. Mantel (1963) Chi²-tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. JASA 58, pp. 690-700.

H. Zimmermann (1984) Erweiterung des Mantel-Haenszel-Testes auf k>2 Gruppen. Biom.J. 26, 2, pp. 223-4.

Mood's Median-Test für k Gruppen: Die Fragestellung ist analog der im Abschnitt „Vierfeldertafeln“, jedoch hier mit einer weiteren Gruppe. Die Validierung folgt den dort gemachten Angaben:

Mood's Median-Test für k Gruppen: ▲

Eingabedatei ist ... \Valide\Sachs390.DAT

davon Variable 1 (Tab109) und Gruppen-Var. 3 (Gruppe) = 1 2 3

Resultierende Tafel (M = Median der gepoolten Stichproben = 18)

	< M	> M		
-----+-----				
Stichprobe 1:	2	6		8
Stichprobe 2:	5	2		7 *)
Stichprobe 3:	4	3		7 *)
-----+-----				
Gesamt:	11	11		22

*) Insgesamt 2 mit dem gepoolten Median identische Werte wurden ignoriert.

Cave: Die Voraussetzungen für den asymptotischen χ^2 -Test sind nicht erfüllt, eine Verifizierung ist per Fisher-Freeman-Halton's Test in "Tafeln" möglich: >> 100.0% der Erwartungshäufigkeiten sind < 5, vertretbar sind maximal 20%.

Prüfgröße: Brandt und Snedecor's $\chi^2=3.429$ mit $df=2$ und $p=0.180092$

Multiple Median-Vergleiche per Fisher-Test (Bonferroni-Holm-korrigiert):

Nr	Stichproben	p-un CORR	p-Bonferroni	p-Bonf.Holm
1	1 versus 2	0.131935	0.395804	0.395804
2	1 versus 3	0.131935	0.395804	0.395804
3	2 versus 3	1.000000	1.000000	1.000000

Für alle Gruppen-Vergleiche immer nur die korrigierten p-Werte verwenden!
Many-one-Vergleiche mit einer fixen Kontrolle: Ist $p\text{-unkorrigiert} \leq \alpha/2$?

Häufigkeiten der 3 Vierfeldertafeln zu den multiplen Vergleichen:

Nr	Stichproben	Median	n11	n12	n21	n22	Sum
1	1 versus 2	19.0000	2	6	6	2	16
2	1 versus 3	21.0000	2	6	6	2	16
3	2 versus 3	16.0000	4	4	4	4	16

Die Mediane sind pro Paarvergleich berechnet. Keine medianidentischen Werte.

Literatur:

A.M. Mood (1950) Introduction to the Theory of Statistics, McGraw-Hill
G.W. Brown, A.M. Mood (1951) On Median Tests for Linear Hypotheses. Proc. of 2nd Berkeley Symposium on Math., Statist. and Prob., Univ. California.
H. Büning, G. Trenkler (1978,1994) Nicht-parametrische Methoden, deGruyter
L. Sachs (2003) Angewandte Statistik, Springer-Verlag Kap. 392, 533.

Friedman-Test: Vergleich von 6 Objekten (i.e. Spalten bzw. "Bedingungen") anhand von 9 Gutachtern (Zeilen): Urteilen die Gutachter gleich? In diesem Beispiel von Sachs (1992, p.671) sind die Testgrößen χ^2 und F von Interesse. Ein weiteres Beispiel findet sich unter "Wilcoxon-Wilcox-Vergleiche".

Friedman-Test mit Wilcoxon-Wilcox-Vergleichen

Eingabedatei ist Sachs671.DAT

davon Variablen 1 (Objekt1), 2 (Objekt2), 3 (Objekt3), 4 (Objekt4)

i	Objekt1	Objekt2	Objekt3	Objekt4
1	3.0000	2.0000	1.0000	4.0000
2	4.0000	2.0000	3.0000	1.0000
3	4.0000	1.0000	2.0000	3.0000
4	4.0000	2.0000	3.0000	1.0000
5	3.0000	2.0000	1.0000	4.0000
6	4.0000	1.0000	2.0000	3.0000
7	4.0000	3.0000	2.0000	1.0000
8	3.0000	1.0000	2.0000	4.0000
9	3.0000	1.0000	4.0000	2.0000

Friedman's Chi² = 10.2000 (fg = 3, p = 0.016940)
Rijkoort-Wise's F = 4.8571 (fg1 = 3, fg2 = 24, p = 0.008842)

Wilcoxon-Wilcox-Test:

Conover-Vergleiche mit Bonferroni:

Spalten 1 und 2: D = 17.00 (p ≤ 0.05)	p = 0.001093	korr: p = 0.006561
Spalten 1 und 3: D = 12.00 (p > 0.05)	p = 0.015056	korr: p = 0.090339
Spalten 1 und 4: D = 9.00 (p > 0.05)	p = 0.061219	korr: p = 0.367313
Spalten 2 und 3: D = 5.00 (p > 0.05)	p = 0.286067	korr: p = 1.000000
Spalten 2 und 4: D = 8.00 (p > 0.05)	p = 0.093647	korr: p = 0.561881
Spalten 3 und 4: D = 3.00 (p > 0.05)	p = 0.518916	korr: p = 1.000000

Literatur:

M. Friedman (1937) The use of ranks to avoid the assumption of normality implicit in the anova. J. Amer. Stat. Assoc. 32, pp.675-701.

F. Wilcoxon, R.A. Wilcox (1964) Some rapid approximate statistical procedure. Lederle Laboratories, Pearl River, New York.

W.J. Conover (1999) Practical Nonparametric Statistics (3rd ed.), Wiley.

L. Sachs (1992) Angewandte Statistik, Springer-Verlag. Kap.76.

Verifikation des Conover-Tests mit: Viv Bewick, Liz Cheek and Jonathan Ball (2004) Statistics review 10: Further nonparametric methods, Crit Care. 2004; 8(3): 196-199. Weitere Berechnungen mit dem Programm StatsDirect.

Page-Test auf Trend: Vergleich von 6 Objekten (Spalten, "Bedingungen") anhand von 9 Gutachtern (Zeilen): Beurteilen die Gutachter gleich, oder besteht ein Trend hinsichtlich der Objekte?

Vorsicht: Die Reihenfolge der Objekte unterscheidet sich in Sachs (1997) vom Friedman-Beispiel, dazu wurde das erste Objekt an das Ende der Spalte verschoben (mit BiAS.' Editor möglich); vgl. dazu auch Sachs Seite 671. Der Test liegt in der approximativen Version vor.

Ein zweites Beispiel: n Patienten mit Prostata-Carcinom werden mit einem Zytostatikum behandelt. Sinkt der PSA-Wert im Zeitverlauf von 0, 3, 6, 12 und 24 Monaten? (Wird hier nicht numerisch ausgeführt.)

Page-Test auf Trend

Eingabedatei ist Sachs671.DAT (Reihenfolge der Objekte gemäß Ha modifiziert!)
davon Variablen 1 (Objekt2), 2 (Objekt3), 3 (Objekt4), 4 (Objekt1)

i	Objekt2	Objekt3	Objekt4	Objekt1
1	2.0000	1.0000	4.0000	3.0000
2	2.0000	3.0000	1.0000	4.0000
3	1.0000	2.0000	3.0000	4.0000
4	2.0000	3.0000	1.0000	4.0000
5	2.0000	1.0000	4.0000	3.0000
6	1.0000	2.0000	3.0000	4.0000
7	3.0000	2.0000	1.0000	4.0000
8	1.0000	2.0000	4.0000	3.0000
9	1.0000	4.0000	2.0000	3.0000

Aufwärts-Trend: Page's Prüfgröße L = 252.0 (p = 0.001107, einseitig)

Abwärts-Trend: Page's Prüfgröße L = 198.0 (p = 0.999252, einseitig)

Literatur:

E.B. Page (1963) Ordered hypothesis for multiple treatments: A significance test for linear ranks. J. Amer. Statist. Assoc. 58, pp. 216-30.

L. Sachs (1997) Angewandte Statistik, Springer-Verlag. Kap.76.

Wilcoxon-Wilcox-Vergleiche: Sechs Personen (Zeilen) erhalten sechs verschiedene Diuretika (Spalten). Gemessen wird die Natriumausscheidung nach 2 Stunden. Von Interesse sind hier multiple Vergleiche nach Wilcoxon-Wilcox (Sachs (1992, p.675)). Die multiplen Vergleiche des Tests erhält man als "Nachtests" zum Friedman-Test:

Friedman-Test mit Wilcoxon-Wilcox-Vergleichen ▲

Eingabedatei ist Sachs675.DAT
davon Variablen 1 , 2 , 3 , 4 , 5 , 6 (Diuretika)

i	Var. 1	Var. 2	Var. 3	Var. 4	Var. 5	Var. 6
1	3.8800	30.5800	25.2400	4.4400	29.4100	38.8700
2	5.6400	30.1400	33.5200	7.9400	30.7200	33.1200
3	5.7600	16.9200	25.4500	4.0400	32.9200	39.1500
4	4.2500	23.1900	18.8500	4.4000	28.2300	28.0600
5	5.9100	26.7400	20.4500	4.2300	23.3500	38.2300
6	4.3300	10.9100	26.6700	4.3600	12.0000	26.6500

Friedman's Chi² = 23.3333 (fg = 5, p = 0.000291)
Rijkoort-Wise's F = 17.5000 (fg1 = 5, fg2 = 25, p = 0.000000)

Wilcoxon-Wilcox-Test:

Spalten 1 und 2: D = 15.00 (p > 0.05)
 Spalten 1 und 3: D = 17.00 (p > 0.05)
 Spalten 1 und 4: D = 2.00 (p > 0.05)
 Spalten 1 und 5: D = 19.00 (p ≤ 0.05)
 Spalten 1 und 6: D = 25.00 (p ≤ 0.01)
 Spalten 2 und 3: D = 2.00 (p > 0.05)
 Spalten 2 und 4: D = 13.00 (p > 0.05)
 Spalten 2 und 5: D = 4.00 (p > 0.05)
 Spalten 2 und 6: D = 10.00 (p > 0.05)
 Spalten 3 und 4: D = 15.00 (p > 0.05)
 Spalten 3 und 5: D = 2.00 (p > 0.05)
 Spalten 3 und 6: D = 8.00 (p > 0.05)
 Spalten 4 und 5: D = 17.00 (p > 0.05)
 Spalten 4 und 6: D = 23.00 (p ≤ 0.01)
 Spalten 5 und 6: D = 6.00 (p > 0.05)

Literatur:

- M. Friedman (1937) The use of ranks to avoid the assumption of normality implicit in the anova. J. Amer. Stat. Assoc. 32, pp.675-701.
- F. Wilcoxon, R.A. Wilcox (1964) Some rapid approximate statistical procedure. Lederle Laboratories, Pearl River, New York.
- L. Sachs (1992) Angewandte Statistik, Springer-Verlag. Kap.76.

Schaich-Hamerle-Vergleiche: 20 Ratten werden nach Spontanaktivität in 5 Gruppen ("Blöcke", Zeilen!) zu je 4 Tieren eingeteilt und 4 Bedingungen zugeteilt. Zielgröße ist die Anzahl erreichter Umdrehungen in einer Lauftrommel (Bortz et al. (1990, pp.269,275). Schaich-Hamerle-Vergleiche werden - in Anschluss an den Friedman-Test - in **BiAS**. ausschließlich für n > 25 und/oder k > 10 verwendet, ansonsten wird der für kleine n und k günstigere Wilcoxon-Wilcox-Test (vgl. oben) eingesetzt. Die hier vorliegende Ausgabe für n=5, k=4 dient nur der Validierung und Dokumentation.

Friedman-Test mit Schaich-Hamerle-Vergleichen ▲

Eingabedatei ist BortzLB269.DAT
davon Variablen 1 (Co), 2 (ED), 3 (DD), 4 (LP)

i	Var. 1	Var. 2	Var. 3	Var. 4
1	14.0000	11.0000	16.0000	13.0000

2	13.0000	12.0000	15.0000	12.0000
3	12.0000	13.0000	14.0000	11.0000
4	11.0000	14.0000	13.0000	10.0000
5	10.0000	15.0000	12.0000	9.0000

Friedman's Chi² = 8.2653 (fg = 3, p = 0.040835)
Rijkoort-Wise's F = 4.9091 (fg1 = 3, fg2 = 12, p = 0.018823)

Schaich-Hamerle-Vergleiche:

Spalten 1 und 2: D = -0.30 p < 0.987329
Spalten 1 und 3: D = -1.20 p < 0.539870
Spalten 1 und 4: D = 1.10 p < 0.611676
Spalten 2 und 3: D = -0.90 p < 0.749409
Spalten 2 und 4: D = 1.40 p < 0.400969
Spalten 3 und 4: D = 2.30 p < 0.047374

Literatur:

M. Friedman (1937) The use of ranks to avoid the assumption of normality implicit in the anova. J. Amer. Stat. Assoc. 32, pp.675-701.

E. Schaich, A. Hamerle (1984) Verteilungsfreie stat. Prüfverfahren. Springer-Verlag

L. Sachs (1992) Angewandte Statistik, Springer-Verlag. Kap.76.

Dixon und Mood's Vorzeichentest: Hierfür wird das gleiche Beispiel wie zum Wilcoxon-matched-pairs-Test verwendet (Hollander und Wolfe (1973, p.29), Erfolg einer Tranquilizer-Behandlung); bitte beachten Sie die unterschiedliche Power der beiden Tests. Die Validierung der Ergebnisse erfolgte mittels Tabellen der Binomial-Verteilung.

Vorzeichen-Test nach Dixon und Mood ▲

Eingabedatei ist Holl129.DAT

davon Variable 1 (HFvor) - Variable 2 (HFnach)

i	Var 1	Var 2	Diff
1:	1.8300	0.8780	0.9520
2:	0.5000	0.6470	-0.1470
3:	1.6200	0.5980	1.0220
4:	2.4800	2.0500	0.4300
5:	1.6800	1.0600	0.6200
6:	1.8800	1.2900	0.5900
7:	1.5500	1.0600	0.4900
8:	3.0600	3.1400	-0.0800
9:	1.3000	1.2900	0.0100

Vorzeichen:

```
-      0      +
-----
2      0      7
```

Binomial-Wahrscheinlichkeit p = 0.089844 (exakt, ein!seitig)

Binomial-Wahrscheinlichkeit p = 0.179688 (exakt, zweiseitig)

Literatur: W.J. Dixon, A.M. Mood (1946) The statistical sign test. J. Amer. Stat. Assoc. 41, pp. 557-566.

Cochran's Q-Test: In den Jahren 1965-69 (Spalten) werden 10 Personen (Zeilen) befragt, ob sie die CDU gegenüber anderen Parteien bevorzugen. Zu testen ist die Hypothese, dass die Jahre den gleichen Effekt auf die Entscheidung der Befragten besitzen (Büning und Trenkler (1978, p.227)):

Cochran's Q-Test

Bezeichnung: Bünning und Trenkler p.227, Datei ..\Valide\Buehning227.DAT

Eingegebene Daten:

1: 1 1 0 1 0
2: 0 1 0 0 1
3: 1 0 1 0 0
4: 1 1 1 1 1
5: 0 1 0 0 0
6: 1 0 1 1 1
7: 0 0 0 0 0
8: 1 1 1 1 0
9: 0 1 0 1 1
10: 1 0 1 0 0

Cochran's Q = 1.3333 : Chi² mit df = 4 (p = 0.855695)

Literatur:

W.G. Cochran (1950) The comparison of percentages in matched samples. Biometrika 37, pp. 256-266.

H. Bünning, G. Trenkler (1978) Nichtparametrische statistische Methoden. Walter de Gruyter Berlin, New York. pp. 226-228.

Test auf Gleichverteilung: Zur Prüfung eines Würfels werden 60 Würfe durchgeführt. Die beobachteten Häufigkeiten der 6 Augenzahlen werden auf Gleichverteilung (d.h. $\pi_i=1/6$ für $i=1,2,\dots,6$) getestet Sachs (1992, p.422):

Test auf Gleichverteilung

Bezeichnung: Sachs p.422 (Würfel 1-6)

Eingelesene Daten:

1: 7 2:16 3: 8 4:17 5: 3 6: 9

Erwartete Häufigkeit = 10.00

Prüfgröße Chi² = 14.80 (fg = 5) : p = 0.011252

Literatur: L. Sachs (1992) Angewandte Statistik. Springer, Kap.43.

Anpassungstest: In einem Spaltungsversuch werden 3 Phänotypen im Verhältnis 1:2:1 erwartet. Entsprechen die gefundenen Häufigkeiten dem erwarteten Verhältnis? (aus Sachs (2003)):

Anpassungstest

Bezeichnung: Sachs p.421 (Spaltungsversuch)

Eingelesene Daten und singuläre Bewertungen der Kategorien:

Kategorie	Beobachtet	?	Erwartet	RelHäuf.	Wahrsch.	Chi ² (df=1)	p(Chi ²)
1	14	<	20.0	0.1750	0.2500	1.800	0.179713
2	50	>	40.0	0.6250	0.5000	2.500	0.113846
3	16	<	20.0	0.2000	0.2500	0.800	0.371093
Sum	80		80.0	1.0000	1.0000	5.100	0.078082

Chi²-Anpassungstest: Prüfgröße Chi²=5.100 mit df=No.Kat-1=2 und p=0.078082
Wolf's G-Test: Prüfgröße Chi²=5.187 mit df=No.Kat-1=2 und p=0.074763

Literatur:

L. Sachs (2003) Angewandte Statistik. 11. Auflage Springer.

Koch's Cross-Over-Analyse des Grizzle-Modells: Verglichen werden zwei Behandlungen A und B in zwei Gruppen mit den Behandlungssequenzen AB und BA (Grizzle (1965)).

Bitte beachten Sie: Nach Eingabe der Gruppenvariablen sind die beiden Nummern bzw. Namen der Variablen einzugeben, die den beiden Perioden zugeordnet sind: Die Eingabeaufforderung lautet: Welche Variablen werden getestet: Periode1: und nach Eingabe Periode 2:. Es handelt sich hierbei um Perioden, nicht um Behandlungen: Zum Beispiel erhielt Gruppe 1 in der ersten Periode die Behandlung A, die Gruppe 2 dagegen B, während in der 2. Periode umgekehrt Gruppe 1 B und Gruppe 2 A erhielt!

Cross-Over: Nicht-parametrische Analyse nach Koch ▲

Eingabedatei ist Grizzel1.DAT
davon Variablen 1 (Periode1), 2 (Periode2)
und Gruppen Var. 3 (GruppNum) = 1 und 2

Gruppe 1			Gruppe 2		
i	Var. 1	Var. 2	i	Var. 1	Var. 2
1	0.2000	1.0000	1	1.3000	0.9000
2	0.0000	-0.7000	2	-2.3000	1.0000
3	-0.8000	0.2000	3	0.0000	0.6000
4	0.6000	1.1000	4	-0.8000	-0.3000
5	0.3000	0.4000	5	-0.4000	-1.0000
6	1.5000	1.2000	6	-2.9000	1.7000
			7	-1.9000	-0.3000
			8	-2.9000	0.9000

Stichprobenumfänge: n1 = 6 und n2 = 8

Nacheffekte: Wilcoxon's U = 7.00 (p = 0.029304 exakter Wert)
Gauss' Z = 2.1947 (p = 0.028186 approximativ)

Haupteffekte: Wilcoxon's U = 14.50 (p = 0.282384 exakter Wert)
Gauss' Z = 1.2278 (p = 0.219524 approximativ)

Erste Periode: Wilcoxon's U = 8.00 (p = 0.042624 exakter Wert)
Gauss' Z = 2.0724 (p = 0.038225 approximativ)

Literatur:

J.E. Grizzle (1965) Two-period cross-over design and its use in clinical trials. Biometrics 21, pp. 467-480.

G.G. Koch (1972) The use of nonparametric methods in the statistical analysis of two-period change-over-design. Biometrics 28, pp.577-584.

L. Sachs (1992) Angewandte Statistik. Springer-Verlag. Kap.394. ▲

Bifaktorielle Rang-Varianzanalyse: Die Fragestellung ist analog zur parametrischen bifaktoriellen Varianzanalyse definiert. Die Methode folgt dem Prinzip „Ranking after Alignment“, ist auch für nicht-orthogonale Datenstrukturen geeignet und kann bei nicht-erfüllten Voraussetzungen zur parametrischen ANOVA sowohl für quantitative als auch für ordinale Daten angewendet werden. Die Wechselwirkung AxB wird mit Kubinger's „unbiased“ Methode geprüft.

Die Ergebnisse der Rang-ANOVA wurden mit Hilfe von Beispielen aus Bortz, Lienert und Boehnke (2008, Kapitel 6) und Kubinger (1986) überprüft, der implizit verwendete Kruskal-Wallis-Test mit Moduln aus SPSS und BiAS. und die multiplen Conover-Iman-Vergleiche für die beiden Faktoren A und B mit einem Modul aus BiAS. Die bindungskorrigierten Prüfgrößen folgen der Darstellung von Conover (1980,1999), die der Interaktion Kubinger (1986).

Bifaktorielle Rang-Varianzanalyse (Kubinger) ▲

Eingabedatei ist ...\\Valide\\BortzLiBo244.DAT, davon Variable 1 (Wert)

```

Faktor A:  Stufe | Variable 2 (FaktorA)
-----+-----
           1 |           1
           2 |           2
           3 |           3

Faktor B:  Stufe | Variable 3 (FaktorB)
-----+-----
           1 |           1
           2 |           2
           3 |           3

```

Eingelesene Daten:

Wert	A: Index	Stufe	B: Index	Stufe
3.00000	1	1	1	1
3.00000	1	1	1	1
5.00000	1	1	1	1
4.00000	1	1	1	1
5.00000	1	1	2	2
6.00000	1	1	2	2
7.00000	1	1	2	2
7.00000	1	1	2	2
9.00000	1	1	3	3
6.00000	1	1	3	3
8.00000	1	1	3	3
8.00000	1	1	3	3
6.00000	2	2	1	1
7.00000	2	2	1	1
6.00000	2	2	1	1
4.00000	2	2	1	1
6.00000	2	2	2	2
9.00000	2	2	2	2
9.00000	2	2	2	2
4.00000	2	2	2	2
7.00000	2	2	3	3
5.00000	2	2	3	3
5.00000	2	2	3	3
5.00000	2	2	3	3
9.00000	3	3	1	1
9.00000	3	3	1	1
9.00000	3	3	1	1
9.00000	3	3	1	1
3.00000	3	3	2	2
4.00000	3	3	2	2

3.00000	3	3	2	2
5.00000	3	3	2	2
3.00000	3	3	3	3
3.00000	3	3	3	3
6.00000	3	3	3	3
7.00000	3	3	3	3

Stichprobenumfänge $n(i,j)$:

Stufen Faktor B	Stufen Faktor A		
	a = 1	a = 2	a = 3
b = 1	4	4	4
b = 2	4	4	4
b = 3	4	4	4

Mediane:

Stufen Faktor B	Stufen Faktor A		
	a = 1	a = 2	a = 3
b = 1	3.500	6.000	9.000
b = 2	6.500	7.500	3.500
b = 3	8.000	5.000	4.500

Haupteffekt Faktor A: Prüfgröße $H=0.440$ mit $\text{Chi}^2(\text{df}=2)$: $p=0.802667$ ▲

Multiple Conover-Iman-Vergleiche (Bonferroni-korrigiert, $\text{df}\{t\}=33$):

Nr	Stufen Faktor A	t	p-unkorr	p-Bonferroni
1	1 versus 2	0.265	0.792345	1.000000
2	1 versus 3	0.379	0.707001	1.000000
3	2 versus 3	0.645	0.523664	1.000000

Haupteffekt Faktor B: Prüfgröße $H=0.313$ mit $\text{Chi}^2(\text{df}=2)$: $p=0.855062$

Multiple Conover-Iman-Vergleiche (Bonferroni-korrigiert, $\text{df}\{t\}=33$):

Nr	Stufen Faktor B	t	p-unkorr	p-Bonferroni
1	1 versus 2	0.530	0.599763	1.000000
2	1 versus 3	0.151	0.880593	1.000000
3	2 versus 3	0.378	0.707509	1.000000

Interaktion (Kubinger): Prüfgröße $H=21.790$ mit $\text{Chi}^2(\text{df}=4)$: $p=0.000221$

Literatur:

KD Kubinger (1986): A Note on Non-Parametric Tests for the Interaction in Two-Way Layouts. *Biometrical Journal*, 28, 67-72
 J Bortz, GA Lienert, K Boehnke (2008) *Verteilungsfreie Methoden in der Biostatistik*. 3. Auflage Springer-Verlag. Kapitel 6.
 WJ Conover (1999) *Practical Nonparametric Statistics* (3rd ed.), Wiley.

A.10 Überlebenszeitanalyse („Survival-Analyse“)

Kaplan-Meier-Schätzer mit Graphik und Tabellen
Hall-Wellner-Konfidenzbänder mit Graphik und Tabellen
Einarmiger Test von $H_0(p_0|t)$
Gehan-Wilcoxon-Test
Log-Rank-Test (Cox-Mantel und Peto-Pike)
Stratifizierter Log-Rank-Test
Peto-Pike-Test für mehrere Gruppen
Relativer Hazard
Cox-Regression mit Abbaumodell
Graphiken zum Cox-Modell
Competing Risks: Aalen-Johansen-Schätzer mit Graphik und Tabellen
Competing Risks: Goetghebeur-Ryan's Log-Rank-Test
Tests zu Proportional-Hazard-Exponentialmodellen

Kaplan-Meier-Schätzer: In einer Studie werden die Effekte zweier präoperativer Bestrahlungsdosen bei Patienten mit Rektalkarzinomen verglichen (Harris und Albert (1991, p.17)). Von Interesse sind die Schätzungen der Überlebenszeiten. Es ist zu beachten, dass eine zensierte Beobachtung (der Patient kann nicht bis Eintritt des „Zielereignisses“ beobachtet werden) in **BiAS**. durch den Wert 1 ("zensiert") der Zensurvariablen gekennzeichnet wird, sonst wird der Wert 0 ("verstorben") angegeben.

Kaplan-Meier-Überlebenszeitschätzungen

Eingabedatei ist Harris17.DAT
davon Variable 1 (Time)
Zensur-Variable 3 (Censor)
und Gruppe(n) Var. 2 (Group) = 2

Gruppe 2

1:	9.0000	13:	27.0000	25:	38.0000*
2:	12.0000	14:	29.0000*	26:	51.0000*
3:	12.0000*	15:	30.0000*	27:	54.0000*
4:	13.0000*	16:	32.0000*	28:	57.0000
5:	14.0000*	17:	33.0000*	29:	60.0000*
6:	16.0000	18:	33.0000*	30:	67.0000
7:	18.0000*	19:	35.0000	31:	70.0000
8:	19.0000	20:	35.0000	32:	87.0000*
9:	23.0000*	21:	35.0000*	33:	89.0000*
10:	24.0000*	22:	35.0000*	34:	98.0000*
11:	25.0000*	23:	35.0000*	35:	120.0000*
12:	26.0000*	24:	36.0000		

Schätzung der Überlebenswahrscheinlichkeiten nach Kaplan-Meier:

Z(i) : Überlebenszeit
P(i) : Wahrscheinlichkeit, Zeitpunkt Z(i) zu überleben
S(i) : Standardabweichung SE(P(i)) (Greenwood's Formel)
Q(i) : Wahrscheinlichkeit, Zeitpunkt Z(i) zu überleben,
vorausgesetzt, Zeitpunkt Z(i-1) wurde überlebt.
L(i) : Linke Grenze des 95%-Konfidenzintervalles für P(i)
R(i) : Rechte Grenze des 95%-Konfidenzintervalles für P(i)
(Konfidenzintervalle nach Hall-Wellner (1980))

Gruppe 2

Beobachtete Zeitpunkte:

i	Z(i)	P(i)	Q(i)	S(i)	L(i)	R(i)
1	9.0000	0.9714	0.9714	0.0282	0.7474	1.0000
2	12.0000	0.9429	0.9706	0.0392	0.7189	1.0000
6	16.0000	0.9114	0.9667	0.0489	0.6867	1.0000
8	19.0000	0.8789	0.9643	0.0570	0.6530	1.0000
13	27.0000	0.8407	0.9565	0.0661	0.6116	1.0000
20	35.0000	0.7418	0.8824	0.0878	0.4941	0.9895
24	36.0000	0.6799	0.9167	0.0999	0.4125	0.9474
28	57.0000	0.5950	0.8750	0.1182	0.2776	0.9123
30	67.0000	0.4958	0.8333	0.1338	0.1018	0.8898
31	70.0000	0.3966	0.8000	0.1390	0.0000	0.8673

Maximale Überlebenszeit durch Abbruch:

35	120.0000	0.3966	1.0000	0.1390	0.0000	0.8673
----	----------	--------	--------	--------	--------	--------

Mittlere Überlebenszeit = 74.4659 (Integral KMS, maximale Zeit als Event)

Mediane Überlebenszeit = 67.0000 (per Schnittpunkt KMS)

Konfidenzintervalle für die mediane Überlebenszeit (nach Collett):

P=95%: KI = [49.3347;84.6653] ; P=99%: KI = [43.7845;90.2155]

Zeitintervalle:

i	Z(i)	P(i)	Q(i)	S(i)	L(i)	R(i)
1	10.0000	0.9714	0.9714	0.0282	0.7474	1.0000
2	20.0000	0.8789	0.9047	0.0570	0.6530	1.0000
3	30.0000	0.8407	0.9565	0.0661	0.6116	1.0000
4	40.0000	0.6799	0.8088	0.0999	0.4125	0.9474
5	50.0000	0.6799	1.0000	0.0999	0.4125	0.9474
6	60.0000	0.5950	0.8750	0.1182	0.2776	0.9123
7	70.0000	0.3966	0.6667	0.1390	0.0000	0.8673
8	80.0000	0.3966	1.0000	0.1390	0.0000	0.8673
9	90.0000	0.3966	1.0000	0.1390	0.0000	0.8673
10	100.0000	0.3966	1.0000	0.1390	0.0000	0.8673
11	110.0000	0.3966	1.0000	0.1390	0.0000	0.8673
12	120.0000	0.3966	1.0000	0.1390	0.0000	0.8673

Literatur:

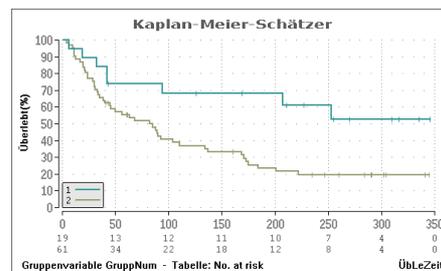
E.L. Kaplan, P. Meier (1958) Nonparametric estimation from incomplete observations. JASA 53, pp. 457-481.

W.J. Hall, J.A. Wellner (1980) Confidence bands for a survival curve from censored data. Biometrika 67, 133-143.

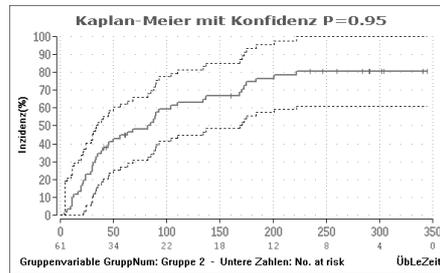
E.K. Harris, A. Albert (1991) Survivorship analysis for clinical studies. Dekker NY

D. Collett (2003) Modelling survival data in medical research. Chapman & Hall

Graphik zum Kaplan-Meier-Schätzer: Vergleich der Effekte zweier präoperativer Bestrahlungsdosen (vgl. die zugehörigen Berechnungen): **BiAS**. sieht einige Optionen zur graphischen Darstellung vor, u.a. auch eine Darstellung als Sterbefunktion. Zensierungen sind hier berücksichtigt!



„**Sterbefunktion**“ für **Kaplan-Meier-Schätzer: BiAS**. berechnet Kaplan-Meier-Schätzer gegebenenfalls mit Hall-Wellner-Konfidenzbändern (weiter unten) als Überlebens- oder Sterbefunktion, hier ein Beispiel (Datei Armitage.DAT) für die Sterbefunktion in Form einer Inzidenzrate:

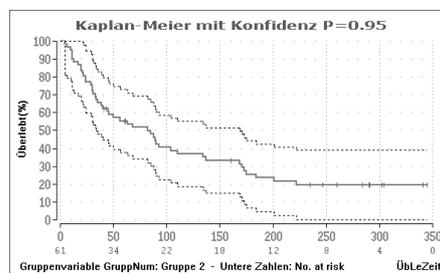


Hall-Wellner-Konfidenzbänder: BiAS. berechnet Hall - Wellner - Konfidenzbänder im Aufruf "Kaplan-Meier-Schätzer" und gibt diese zusammen mit der "Life-Table" als Tabelle aus.

Beim Vorliegen von nur einer Gruppe kann die Survival- bzw. Überlebensfunktion oder - optional - die Sterbefunktion einschließlich Hall-Wellner-Konfidenzband dargestellt werden. Bei Vorliegen von zwei oder mehr Gruppen werden - ohne Hall-Wellner - nur die Survival- bzw. Sterbefunktionen dargestellt.

Beispiel: In einer klinischen Studie werden die Effekte zweier Bestrahlungsdosen bei Patienten mit Rektalkarzinomen verglichen (Harris und Albert (1991, p.17)). Von Interesse sind hier zunächst die Schätzungen der Überlebenszeiten. Es ist zu beachten, dass eine zensierte Beobachtung (z.B. Patient ist verzogen, kann nicht bis Eintritt des Zielereignisses beobachtet werden) in **BiAS** durch den Wert 1 ("zensiert") der Zensurvariablen gekennzeichnet wird, sonst wird der Wert 0 ("verstorben") angegeben.

Die über den Hypertext aufrufbaren Abbildungen werden für beide Gruppen aus diesem Beispiel dargestellt (Armitage.DAT), für die Hall-Wellner-Bereiche wurde die 2. Gruppe verwendet.



Kaplan-Meier-Schätzer – Test der Nullhypothese $H_0(p_0|t)$: Die Nullhypothese des *einarmigen* Tests einer bestimmten Überlebensrate p_0 zum Zeitpunkt t kann einfach per Hall-Wellner-Konfidenzbänder durchgeführt werden: Liegt zum Zeitpunkt t der hypothetische Wert p_0 im Konfidenzintervall, wird die Nullhypothese akzeptiert, ansonsten, wie üblich, zum Testniveau $\alpha=1-P$ abgelehnt.

Im Modul „Studienplanung – Fallzahlberechnung – nicht-parametrische Methoden“ findet sich unter „Kaplan-Meier-Plot: Test von $H_0(p_0|t)$ “ ein Verfahren von Kengo Nagashima (2016) zur Fallzahlberechnung für die *einarmige* Testsituation. Das Verfahren erlaubt die Vorgabe von $\alpha=1-P$ für den Fehler 1. Art und der Power $1-\beta$ des Tests, zusätzlich einige Transformationen der Survival-Funktion $S^{\wedge}(t)$ (log, log-log, logit und Arcsin).

Literatur und Erläuterungen:

Nagashima K. (2016) A sample size estimation tool for one sample non-parametric tests for a survival proportion [Internet]. 2016 Mar 21 Available from: <http://nshi.jp/en/js/onesurvyr/>

Gehan-Wilcoxon-Test: Peto und Pike (1977) beschreiben eine Überlebenszeitstudie aus der Onkologie. Die folgende Auswertung bezieht sich auf Peto-Pike's Gruppe apathologischer Nieren; die Daten sind dem Ausdruck des stratifizierten Log-Rank-Tests zu entnehmen:

Gehan-Wilcoxon-Test

Eingabedatei ist Peto31A.DAT
davon Variable 1 (üLeZeit)
Zensur-Variable 2 (Zensur)
und Gruppe(n) Var. 3 (GruppNum) = 1, 2

Gruppe 1

1: 852.0000*
2: 220.0000
3: 8.0000
4: 1976.0000*
5: 1296.0000*
6: 1460.0000*
7: 1328.0000*
8: 365.0000*

Gruppe 2

1: 180.0000
2: 632.0000
3: 2240.0000*
4: 195.0000
5: 76.0000
6: 70.0000
7: 1990.0000*
8: 700.0000
9: 210.0000
10: 1296.0000 (Zensierte Werte sind durch '*' gekennzeichnet.)

Gehan-Wilcoxon-Test:
Prüfgröße Z = 1.6118 (p = 0.107002)

Literatur:

E.A. Gehan (1965) A generalized Wilcoxon-test for arbitrarily single-censored samples. Biometrika 52, 1 und 2, pp. 203-223.

Log-Rank-Test (Cox-Mantel): Armitage und Berry (1988, p.431) vergleichen in einem Beispiel die Überlebenszeiten von Lymphompatienten verschiedener Stadien. Von Interesse ist der Vergleich beider Gruppen mit dem Log-Rank-Test. (Die Daten zu diesem Beispiel befinden sich in **BiAS.** Verzeichnis. Die Daten der zweiten Gruppe werden auch für die Kaplan-Meier-Schätzungen verwendet.)

Log-Rank-Test

Eingabedatei ist Armitage431.DAT
davon Variable 1 (üLeZeit)
Zensur-Variable 3 (Zensur)
und Gruppe(n) Var. 2 (GruppNum) = 1, 2

Log-Rank-Test:

Peto-Pike's $\chi^2 = 6.5448$ mit $df = 1$ (p = 0.010519)
Cox-Mantel's $\chi^2 = 6.7097$ mit $df = 1$ (p = 0.009589)

Relativer Hazard (Peto-Pike):

RH (1: 2) = 0.3889 95%-Konfidenzintervall: (0.1886 , 0.8081)
RH (2: 1) = 2.5715 95%-Konfidenzintervall: (1.2472 , 5.3020)

Literatur:

P. Armitage, G. Berry (1988) *Statistical Methods in Medical Research*. Blackwell Scientific Publications, 2nd Ed. pp. 421-439.
D.R. Cox, D. Oakes (1984) *Analysis of Survival Data*. Chapman and Hall, London NY.

Log-Rank-Test (Peto-Pike): Peto und Pike (1977) beschreiben eine onkologische Überlebenszeitstudie. Die folgende Auswertung bezieht sich auf Peto-Pike's Gruppe apathologischer Nieren. Die Daten sind dem Daten-Ausdruck des nächsten Beispiels zu entnehmen. Ein Mehrgruppenvergleich findet sich weiter unten!

Log-Rank-Test ▲

Eingabedatei ist Peto31A.DAT
davon Variable 1 (üLeZeit)
Zensur-Variable 2 (Zensur)
und Gruppe(n) Var. 3 (GruppNum) = 1, 2
Cox-Mantel's Chi² = 3.6697 mit df = 1 (p = 0.055410)
Peto-Pike's Chi² = 3.6220 mit df = 1 (p = 0.057020)

Relativer Hazard mit Konfidenzintervallen (nach Peto-Pike):

RH(1:2)=0.2491 95%-Konfidenzintervall: (0.0595 , 1.0424)
99%-Konfidenzintervall: (0.0380 , 1.6344)
RH(2:1)=4.0147 95%-Konfidenzintervall: (0.9594 , 16.8003)
99%-Konfidenzintervall: (0.6118 , 26.3427)

Literatur:

P. Armitage, G. Berry (1988) *Statistical Methods in Medical Research*. Blackwell Scientific Publications, 2nd Ed. pp. 421-439.
D.R. Cox, D. Oakes (1984) *Analysis of Survival Data*. Chapman and Hall, London NY.

Stratifizierter Log-Rank-Test: Peto und Pike (1977) beschreiben eine onkologische Überlebenszeitstudie. Die folgende Auswertung berücksichtigt im Gegensatz zum letzten Beispiel die Schichtung "Nierenfunktion" (Stratum 2 ist mit dem letzten Beispiel identisch).

Editor: Ausdrucken ▲

Die Datei Peto.31 umfasst 4 Spalten bzw. 25 Zeilen (.3Kb).

Variablenamen der Datei Peto.31:
1:übLeZeit 2:Zensiert 3:Behandlg 4:NiereFkt

Ausdruck der Datei Peto.31:

```
Nr.:  1  2  3  4
1:    8  0  1  1
2:   180  0  2  2
3:   632  0  2  2
4:   852  1  1  2
5:    52  0  1  1
6:  2240  1  2  2
7:   220  0  1  2
8:    63  0  1  1
9:   195  0  2  2
10:   76  0  2  2
11:   70  0  2  2
12:    8  0  1  2
13:   13  0  2  1
```

14: 1990 1 2 2
15: 1976 1 1 2
16: 18 0 2 1
17: 700 0 2 2
18: 1296 1 1 2
19: 1460 1 1 2
20: 210 0 2 2
21: 63 0 1 1
22: 1328 1 1 2
23: 1296 0 2 2
24: 365 1 1 2
25: 23 0 2 1

Stratifizierter Log-Rank-Test

Eingabedatei ist Peto31.DAT
davon Variable 1 (üLeZeit)
Zensur-Variable 2 (Zensiert)
und Gruppe(n) Var. 3 (Behandlg) = 1, 2

Stratifizierungsvariablen und Stufen:
NiereFkt: Stufen 1 2

Log-Rank-Tests für alle Strata:

Stratum 1: NiereFkt= 1
Cox-Mantel's Chi² = 2.1904 mit df = 1 (p = 0.138869)
Peto-Pike's Chi² = 1.6526 mit df = 1 (p = 0.198604)
Stratum 2: NiereFkt= 2
Cox-Mantel's Chi² = 3.6697 mit df = 1 (p = 0.055410)
Peto-Pike's Chi² = 3.6220 mit df = 1 (p = 0.057020)

Stratifizierter Log-Rank-Test:
Cox-Mantel's Chi² = 5.7908 mit df = 1 (p = 0.016111)
Peto-Pike's Chi² = 4.8701 mit df = 1 (p = 0.027326)

Literatur:

S.J. Pocock (1983) Clinical Trials. John Wiley.
D.R. Cox, D. Oakes (1984) Analysis of Survival Data. Chapman and Hall, London NY.

Peto-Pike-Test für mehrere Gruppen: Bei der Untersuchung von Überlebenszeiten kann unter Umständen ein Vergleich von mehr als 2 Gruppen von Interesse sein, wozu Peto-Pike's Mehrgruppenvergleich herangezogen werden kann. Daten aus der Literatur können nicht angegeben werden, so dass **BiAS.** Ergebnisse mit den Programmen WinStat und MedCalc verglichen wurden. (Vorsicht in WinStat 3.1 bei Bindungen und zensierten Werten am Ende einer Liste!)

Multiple Vergleiche zu diesem Test sind nicht bekannt, so dass diese mit Hilfe von paarweisen Zweigruppenvergleichen durchgeführt werden. Bitte beachten Sie dabei die Bonferroni-Korrektur bzw. die analoge Korrektur für Many-one-Vergleiche!

Die Konfidenz für die Konfidenzintervalle ist im Eingabedialog des Programms frei wählbar.

Log-Rank-Test (Peto-Pike für k Gruppen)

Eingabedatei ist \Valide\ARMITAGE2.DAT
davon Variable 1 (ÜbLeZeit)
Zensur-Variable 3 (Zensur)
und Gruppe(n) Var. 2 (GruppNum) = 1 2 3

Chi²-Komponenten:

Gruppe 1: Chi² = 3.600 angenähert mit df=1 und p = 0.057783
Gruppe 2: Chi² = 4.280 angenähert mit df=1 und p = 0.038552
Gruppe 3: Chi² = 1.992 angenähert mit df=1 und p = 0.158118

Log-Rank-Test:

Peto-Pike's Chi² = 9.873 mit df = 2 und p = 0.007181 ▲

Multiple Vergleiche und Relative Hazards mit 95%-Konfidenzintervallen:

Vergleich der Paare	Relativer Hazard RH	Konfidenzintervall		p-Werte zu den Paarvergleichen		
		Linke Gr	Rechte Gr	unkorr.	Bonferroni	Bonf.Holm
1 : 2	0.3889	0.1886	0.8018	0.010519	0.031557	0.031557
2 : 1	2.5715	1.2472	5.3020	0.010519	0.031557	0.031557
1 : 3	0.8787	0.3051	2.5305	0.810565	1.000000	0.810565
3 : 1	1.1381	0.3952	3.2776	0.810565	1.000000	0.810565
2 : 3	2.3930	1.0494	5.4570	0.038022	0.114066	0.076044
3 : 2	0.4179	0.1833	0.9529	0.038022	0.114066	0.076044

Zum Multiplen Vergleich die Bonferroni-Holm-korrigierten p-Werte verwenden!
Many-one-Vergleiche mit einer fixen Kontrolle: Ist $p(\text{unkorrigiert}) < \alpha/2$?

Literatur: E.K. Harris, A. Albert (1991) Survivorship Analysis for Clinical Studies, Verlag Marcel Dekker.

Relativer Hazard: Armitage und Berry (1988, p.431) vergleichen in einem Beispiel die Überlebenszeiten von Lymphompatienten verschiedener Stadien; die Daten zu diesem Beispiel befinden sich in BiAS.' Verzeichnis. Die Konfidenz für die Konfidenzintervalle ist im Eingabedialog wählbar.

Relative Hazards sind analog zu Peto-Pike's Mehrgruppenvergleich auch bei Vorliegen von *mehreren* Gruppen berechenbar!

Relativer Hazard ▲

Eingabedatei ist Armitage.Dat
davon Variable 1 (üLeZeit) und Zensur-Variable 3 (Zensur)
Alle Gruppen Var. 2 (GruppNum)

Relative Hazards mit 95%-Konfidenzintervallen (approx. nach Peto-Pike):

Vergl.	RH	LiGr	ReGr	Vergl.	RH	LiGr	ReGr
1 : 2	0.3889	0.1886	0.8018	2 : 1	2.5715	1.2472	5.3020

Im Zwei-Gruppen-Fall ist unter "Log-Rank-Test Peto-Pike/Cox-Mantel" eine ausführlichere Analyse mit Test und mit Konfidenzintervallen möglich.

Literatur: E.K. Harris, A. Albert (1991) Survivorship Analysis for Clinical Studies. Verlag Marcel Dekker.

Cox-Regression: Das Programm berechnet die partiellen Koeffizienten des Cox-Modells, die entsprechenden Odds-Ratios und die zugehörigen p-Werte (Wald). Bei Bindungen der „Events“ wird die Breslow-Methode verwendet. Eine Funktion zum Modellabbau (Wald „backward-elimination“) ist vorhanden. Die Ergebnisse des Programms wurden mit SPSS und „R“ verifiziert, die zugehörigen Daten befinden sich in der Datei CoxNier.DAT und werden hier wegen des Umfangs nicht aufgelistet.

Cox-Regression

Eingabedatei ist ...\\Valide\CoxNier.DAT

Zielgröße ist Variable 8 (survive)

ZensurVar ist Variable 9 (status)

mit Einflussvariablen 2 (diff), 3 (t), 4 (lkb), 5 (m), 6 (uicc), 7 (lkq)

Events: 43 40.6%

Zensiert: 63 59.4%

Gesamt: 106 100.0%

Zensur: Event=0 und Zensiert=1

Events: Bei Bindungen Breslow's Approximation

>> CAVE: Bei 43 Events sollten maximal etwa 4 Prädiktoren verwendet werden,

>> bei Überschreitung ist die Stabilität des Modells möglicherweise unsicher.

Deskriptive Statistik:

Prädiktor	Durchschnitt	Standardabw
diff	2.3491	0.7896
t	2.4623	0.8597
lkb	1.1698	0.8739
m	0.0849	0.2787
uicc	3.6698	1.4584
lkq	0.3715	0.3674

Iterationen mit -2LL-Konvergenzkriterium eps=1E-6:

-2 Log Likelihood = 291.2842 : Start mit dem Null-Modell

-2 Log Likelihood = 285.3767 : Konvergenz nach 19 Iterationen.

>> CAVE: -2LL konvergierte vor dem Koeffizienten des Prädiktors m,

>> der Koeffizient beta weist einen Trend gegen "unendlich" auf!

Overall-Anpassung des Modells:

Chi² = 5.908 mit df = 6 und p = 0.433631

Koeffizienten beta, Standardabweichungen, p-Werte und 95%-Konfidenzintervalle:

Prädiktor	beta	StdAbw	Chi ²	p-Wert	KI-unten	KI-oben
diff	0.1465	0.2157	0.461	0.496930	-0.2763	0.5694
t	-2.9738	1.1355	6.859	0.008821	-5.1994	-0.7482
lkb	-3.2760	1.2179	7.235	0.007150	-5.6631	-0.8888
m	-15.9812	1959.5885	0.000	0.993493	-3856.7748	3824.8123
uicc	3.0757	1.1525	7.122	0.007615	0.8168	5.3347
lkq	0.7104	1.2511	0.322	0.570170	-1.7418	3.1626

Relative Hazards = exp(beta) und 95%-Konfidenzintervalle:

Prädiktor	Rel.Hazard	KI-unten	KI-oben
diff	1.1578	0.7586	1.7672
t	0.0511	0.0055	0.4732
lkb	0.0378	0.0035	0.4111
m	0.0000	0.0000	9.99E+999
uicc	21.6659	2.2632	207.4113
lkq	2.0348	0.1752	23.6313

Baseline Survivor Function an den Prädiktor-Mittelwerten bei "Event": ▲

Zeitpunkt Event	überlebt	nicht überlebt
4.0000	0.997010	0.002990
6.0000	0.993497	0.006503
7.0000	0.982651	0.017349
9.0000	0.967269	0.032731
10.0000	0.956042	0.043958
12.0000	0.926020	0.073980
13.0000	0.919194	0.080806
19.0000	0.910168	0.089832
21.0000	0.900978	0.099022
24.0000	0.891115	0.108885
26.0000	0.871413	0.128587
28.0000	0.859296	0.140704
36.0000	0.846231	0.153769
37.0000	0.820203	0.179797
38.0000	0.805840	0.194160
48.0000	0.743831	0.256169
60.0000	0.000000	1.000000

>> Die maximale Überlebenszeit Tmax = 60.0000 ist ein Event/Ereignis.

>> Mediane Überlebenszeit Tmed = 60.0000 per Schnittpunkt mit BSF-Schätzer.

Baseline Survivor Function per Prädiktor-Mittelwerte in Intervallen:

Intervallgrenze	Letzter Event	überlebt	nicht überlebt
0.0000		1.000000	0.000000
5.0000	4.0000	0.997010	0.002990
10.0000	10.0000	0.956042	0.043958
15.0000	13.0000	0.919194	0.080806
20.0000	19.0000	0.910168	0.089832
25.0000	24.0000	0.891115	0.108885
30.0000	28.0000	0.859296	0.140704
35.0000	28.0000	0.859296	0.140704
40.0000	38.0000	0.805840	0.194160
45.0000	38.0000	0.805840	0.194160
50.0000	48.0000	0.743831	0.256169
55.0000	48.0000	0.743831	0.256169
60.0000	60.0000	0.000000	1.000000

>> Die maximale Überlebenszeit Tmax = 60.0000 ist ein Event/Ereignis.

>> Mediane Überlebenszeit Tmed = 60.0000 per Schnittpunkt mit BSF-Schätzer. ▲

Literatur:

D. Collett (2003) Modelling survival data in medical research. Chapman & Hall
 J.F. Lawless (1982) Statistical Models and Methods for Lifetime Data. Wiley NY
 J.C. Pezzullo (2007) Scripts zum Cox-Modell, statpages.org, Georgetown Univ US
 E. Vittinghoff, C.E. McCulloch (2007) Relaxing the rule of ten events per variable in logistic and Cox regression. Amer. J. Epidemiol. 165:710-8

Graphische Darstellungen der Baseline-Survivor-Funktion und graphische Darstellungen mit individuellen Prädiktor-Werten sind möglich, werden hier aber wegen der Vielfalt im Einzelnen nicht ausgeführt!

Competing Risks - Aalen-Johansen-Schätzer: Aalen und Johansen (1978) beschreiben eine dem Kaplan-Meier-Schätzer analoge Methode, jedoch unter Berücksichtigung von „Competing Risks“. Ein CR-Event verhindert, dass das eigentliche Ziel-Event (zum Beispiel „Tod“) später noch auftreten kann und bedeutet beispielsweise „Tod aus anderen Gründen“. Held (2013) gibt dazu ein Beispiel, das in der nachfolgenden Berechnung verwendet wird. Die Berechnung erfolgt hier zunächst für beide Behandlungen zusammen (Index in der Datenspalte „Behandlung“). Der Index für die Risiken, fortlaufend nummeriert, findet sich in der Spalte „Events“:

Zur Validierung wurden per-Hand-Berechnungen und das Programm „R“ verwendet. Eine graphische Darstellung der Aalen-Johansen-Schätzer findet sich auf der nächsten Seite.

Competing Risks: Aalen-Johansen-Schätzer

Eingabedatei ist C:\PbWin80\BiAS8\Held269.DAT

davon die Zeit-Variable: UebLeZeit

und die Event-Variable: Events

Zensierte Werte sind durch den Wert Events=0 gekennzeichnet.

Es werden alle in Events enthaltenen Risiken dargestellt.

Risiko 1:

ÜbLebzeiten	At Risk	No. Events		Überlebenswahrscheinlichkeiten		
		Ziel	CR	Ziel-Event	CR-Event	Komb-Ev
0.8000	10	1	0	0.9000	1.0000	0.9000
1.2000	9	0	1	0.9000	0.9000	0.8000
1.4000	8	1	0	0.8000	0.9000	0.7000
2.4000	7	1	0	0.7000	0.9000	0.6000
2.9000	6	1	0	0.6000	0.9000	0.5000
3.1000	5	1	0	0.5000	0.9000	0.4000
5.4000	4	0	1	0.5000	0.8000	0.3000
7.9000*	3	0	0	0.5000	0.8000	0.3000
10.5000	2	0	1	0.5000	0.6500	0.1500
15.3000*	1	0	0	0.5000	0.6500	0.1500

Risiko 2:

ÜbLebzeiten	At Risk	No. Events		Überlebenswahrscheinlichkeiten		
		Ziel	CR	Ziel-Event	CR-Event	Komb-Ev
0.8000	10	0	1	1.0000	0.9000	0.9000
1.2000	9	1	0	0.9000	0.9000	0.8000
1.4000	8	0	1	0.9000	0.8000	0.7000
2.4000	7	0	1	0.9000	0.7000	0.6000
2.9000	6	0	1	0.9000	0.6000	0.5000
3.1000	5	0	1	0.9000	0.5000	0.4000
5.4000	4	1	0	0.8000	0.5000	0.3000
7.9000*	3	0	0	0.8000	0.5000	0.3000
10.5000	2	1	0	0.6500	0.5000	0.1500
15.3000*	1	0	0	0.6500	0.5000	0.1500

>> *: Die Überlebenszeit ist bezüglich Ziel-Event und CR-Event(s) zensiert!

>> Komb-Ev: Gesamt-Überleben, ungeachtet der Differenzierung "Ziel" und "CR"

>> Die Graphik bezieht sich auf das CR-adjustierte Überleben bzgl. Ziel-Event

Literatur:

Held L, Rufibach K, Seifert B (2013) Medizinische Statistik. 1st Ed. Pearson Higher Education München.

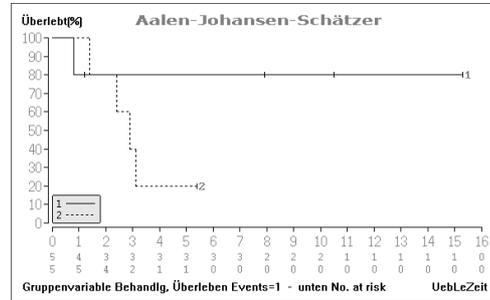
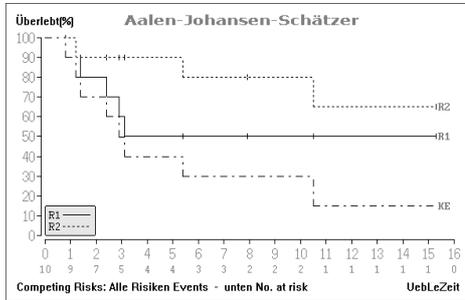
Putter H, Fiocco M, Geskus RB (2007) Tutorial in biostatistics: Competing risks and multi-state models. Statist Med 26, pp. 2389-2430

Aalen O, Johansen S (1978) An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. Scand J Statist 5, pp. 141-50.

Goetghebeur E, Ryan L (1990) A modified log rank test for competing risks with missing failure type. Biometrika 77, 1, pp. 207-11.

Armitage P, Berry G, Matthews JNS (2002) Statistical Methods in Medical Research. 4th Ed. Blackwell Science, Oxford.

Graphik zum Aalen-Johansen-Schätzer: Dazu wurden die Resultate aus den Ausgabe-Tabellen verwendet. Neben dem AJS wird – optional – zusätzlich die Überlebensfunktion für das CR-Event und die für das „kombinierte Event“ berechnet. Eine Angabe der „Number at Risk“ ist möglich, ebenfalls eine Markierung der zensierten Werte; beides ist in der Abbildung nicht enthalten. Links ein Vergleich aller Risiken, rechts die Schätzer für Risiko 1 in den beiden Gruppen:



Competing Risks - Goetghebeur-Ryan's Log-Rank-Test: Zum Gruppenvergleich finden sich anschließend die analogen Berechnung unter Berücksichtigung der beiden Behandlungsgruppen, wozu wieder die Daten aus ..\Valide\Held269.DAT verwendet wurden:

Competing Risks: Aalen-Johansen-Schätzer & Goetghebeur-Ryan's Log-Rank-Test

Eingabedatei ist C:\PbWin80\BiAS8\Held269.DAT

davon die Zeit-Variable: UebLeZeit

und die Event-Variable: Events

Gruppe(n) Var. 2 (Behandlg) = 1 2

Zensierte Werte sind durch den Wert Events=0 gekennzeichnet.

Zielereignis ("Event of Interest"): Events=1

Gruppe Behandlg=1:

ÜbLebZeiten	At Risk	No. Events		Überlebenswahrscheinlichkeiten		
		Ziel	CR	Ziel-Event	CR-Event	Komb-Ev
0.8000	5	1	0	0.8000	1.0000	0.8000
1.2000	4	0	1	0.8000	0.8000	0.6000
7.9000*	3	0	0	0.8000	0.8000	0.6000
10.5000	2	0	1	0.8000	0.5000	0.3000
15.3000*	1	0	0	0.8000	0.5000	0.3000

Gruppe Behandlg=2:

ÜbLebZeiten	At Risk	No. Events		Überlebenswahrscheinlichkeiten		
		Ziel	CR	Ziel-Event	CR-Event	Komb-Ev
1.4000	5	1	0	0.8000	1.0000	0.8000
2.4000	4	1	0	0.6000	1.0000	0.6000
2.9000	3	1	0	0.4000	1.0000	0.4000
3.1000	2	1	0	0.2000	1.0000	0.2000
5.4000	1	0	1	0.2000	0.8000	0.0000

>> *: Die Überlebenszeit ist bezüglich Ziel-Event und CR-Event(s) zensiert!

>> Komb-Ev: Gesamt-Überleben, ungeachtet der Differenzierung "Ziel" und "CR"

>> Die Graphik bezieht sich auf das CR-adjustierte Überleben bzgl. Ziel-Event

Nullhypothesentest per Log-Rank-Test für Competing Risks (Goetghebeur-Ryan):

Prüfgröße Chi² = 1.616 mit df=1: p = 0.203689

Relative Hazards mit 95%-Konfidenzintervallen (Goetghebeur-Ryan & Peto-Pike):

Nr.	Gruppen	RelHaz	LiGr	ReGr	Gruppen	RelHaz	LiGr	ReGr
1	1 : 2	0.2701	0.0359	2.0329	2 : 1	3.7029	0.4919	27.8734

Literatur: Wie zum Aalen-Johansen-Schätzer.

Proportional-Hazard-Exponentialmodell mit individuellen Parametern: Das Modell ist ein parametrisches Proportional-Hazard-Regressionsmodell zur Überlebenszeitanalyse, mit dem getestet wird, ob der Regressionskoeffizient zu einem an anderer Stelle individuell bestimmten Score gleich 1 ist. Ein Beispiel zur Generierung der individuellen Parameter findet sich bei WC Levy et al. (2006) The Seattle Heart Failure Model: Prediction of Survival in Heart Failure. Circulation 113, 1424-33. Entsprechende Daten befinden sich in der Datei ..\Valide\IndivExpMod.DAT. Das Modul wurde mit einem unabhängigen R-Programm verifiziert.

Test zum PH-Exponentialmodell mit individuellen Parametern ▲

Eingabedatei ist ..\Valide\IndivExpMod.DAT
 davon Variable 4 (Zeit)
 Statusvariable 3 (Zensur)
 Surv.-Variable 2 (2yearSur)
 und Gruppenvariable 1 (Gruppe) = 1 2

Gruppe: Gruppe=1				Gruppe: Gruppe=2			
i	2yearSur	Zensur	Zeit	i	2yearSur	Zensur	Zeit
1	69.00	1	7.35	1	87.00	1	8.26
2	71.00	0	6.50	2	85.00	0	8.06
3	96.00	1	3.82	3	80.00	1	8.05
4	92.00	0	6.48	4	94.00	0	5.83
5	84.00	0	5.67	5	87.00	1	5.31
6	82.00	0	1.67	6	88.00	1	6.15
7	87.00	0	5.42	7	93.00	0	8.49
8	91.00	0	4.50	8	85.00	0	5.38
9	93.00	1	2.72	9	86.00	1	5.24
10	90.00	1	9.34	10	94.00	1	5.52
11	93.00	0	3.70	11	87.00	1	4.43
12	89.00	1	4.96	12	90.00	1	4.25
13	92.00	0	8.44	13	83.00	0	1.76
14	75.00	0	0.42	14	91.00	1	4.32
15	74.00	0	0.24	15	85.00	0	9.16
16	81.00	0	2.11	16	90.00	0	0.89
17	95.00	0	6.21	17	81.00	0	5.84
18	87.00	1	5.56	18	83.00	0	0.03

>> Zensierte Werte sind durch den Wert Zensur=1 gekennzeichnet.

>> Die eingelesenen Überlebensraten beziehen sich auf den Zeitpunkt t=2.00

Test auf Abweichung vom individuellen Exponentialmodell - Ho(Lambda0=1): ▲

Nr.	Gruppe	n	#zens	MLS Lambda0	Chi ² (1)	p-Wert
1	1	18	6	1.9031	4.054	0.044055
2	2	18	9	1.3449	0.718	0.396924

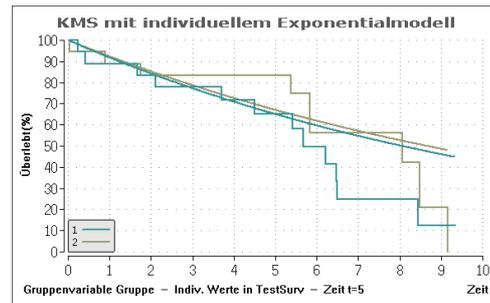
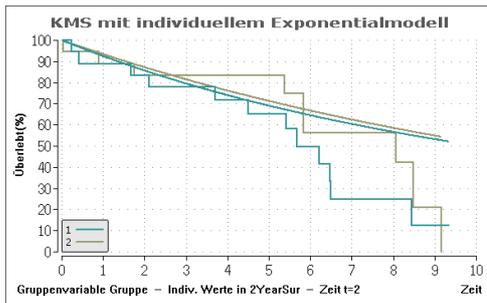
Log-Rank-Test zum Vergleich der Kaplan-Meier-Schätzer:

Peto-Pike's Chi² = 0.833 mit df = 1 und p = 0.361529

>> Detailliertere Auswertungen per Menü unter "Log-Rank-Test"

Literatur:

E. Herrmann (2011) Assessing the predicted survival time in a parametric context. Preprint Univ. Frankfurt am Main.
 D.R. Cox, D. Oakes (1984) Analysis of Survival Data. Chapman & Hall London
 W.C. Levy et al. (2006) The Seattle Heart Failure Model: Prediction of Survival in Heart Failure. Circulation 113: 1424-1433.



Proportional-Hazard-Exponentialmodell mit einem festen Parameter: Das Modell ist wie im Modell für individuelle Parameter ein parametrisches Proportional-Hazard-Regressionsmodell zur Überlebenszeitanalyse, bei dem getestet wird, ob der Regressionskoeffizient β_0 einem festen vorgegebenen Wert entspricht. Vergleichen Sie dazu bitte auch das individualisierte Modell unter "PH-Exponentialmodell mit individuellen Parametern".

Das Procedere erfolgt analog zum individualisierten Modell: Das „feste Modell“ stellt hier lediglich einen Spezialfall des individualisierten Modells dar, zur Überprüfung kann die im Verzeichnis ..\Valide enthaltene Beispieldatei IndivExpMod.DAT verwendet werden: Zur Berechnung mit dem individualisierten Modell enthält die Spalte "TestSurv" die - dort konstanten - individuellen Überlebensraten, womit sich die identischen Resultate wie im festen Modell mit der Rate 0.65=65% für Gruppe 1 bzw. mit der Rate 0.67=67% für Gruppe 2 ergeben. Für alle Rechnungen muss dabei der gleiche Zeitpunkt gewählt werden; im Beispiel wurde „TestSurv“ zusammen mit dem Zeitpunkt t=5 dargestellt (dazu die rechte Abbildung oben).

Alle Daten hierzu finden Sie in der Datei ..\Valide\IndivExpMod.DAT. Das Modul wurde mit einem externen R-Programm verifiziert.

Test zum PH-Exponentialmodell mit einem festen Parameter

Eingabedatei ist ..\Valide\IndivExpMod.DAT
 davon Variable 4 (Zeit)
 Statusvariable 3 (Zensur)
 und Gruppenvariable 1 (Gruppe) = 1

>> Zensierte Werte sind durch den Wert Zensur=1 gekennzeichnet.

>> Eingegebene Überlebensrate zur Zeit t=5.00: P=S(t)=65.00%

>> Daraus errechneter Wert Lambda = 0.086157 gemäß $S(t)=\exp(-\text{Lambda} \cdot t)$

Test auf Abweichung vom gegebenen Exponentialmodell - $H_0(\text{Lambda}=1)$:

Nr.	Gruppe	n	#zens	MLS Lambda0	Chi ² (1)	p-Wert
1	1	18	6	1.6365	2.487	0.114805

Literatur:

- E. Herrmann (2011) Assessing the predicted survival time in a parametric context. Preprint Univ. Frankfurt am Main.
- D.R. Cox, D. Oakes (1984) Analysis of Survival Data. Chapman & Hall London
- W.C. Levy et al. (2006) The Seattle Heart Failure Model: Prediction of Survival in Heart Failure. Circulation 113: 1424-1433.



Clusteranalyse mit Graphik nach McQueen
Single-Linkage-Verfahren (diverse Metriken)
Structogramm zur Clusterung
Alternative Distanzmaße
KFA zur Clusterung binärer und kategorialer Daten (A.5)

Clusteranalyse nach McQueen: Deichsel und Trampisch (1985, p.3) geben einen Datensatz zur Clusteranalyse mit KMeans an; die 20 Wertepaare ergeben sich aus dem folgenden Ausdruck der ermittelten Cluster; eine Graphik ist möglich:

Cluster-Analyse nach McQueen und Rubin

Eingabedatei ist Deichsel003.DAT
davon Variablen 1 (AP[U/l]), 2 (Fe[mg/l])
Die Daten werden untransformiert verarbeitet.

Clusterzentren, Minima und Maxima:

Nr.	n	1	2
1	5	1.3	0.7
		0.8	0.5
		2.0	1.0
2	4	2.5	1.9
		2.1	1.7
		3.0	2.2
3	6	2.2	3.2
		1.7	3.0
		2.8	3.5
4	5	3.7	0.9
		3.2	0.7
		4.5	1.2

Güte der Klassifikation: $F = MQ_{zwi}/MQ_{in} = 52.1119$ mit $df = (3, 16)$

Cluster 1 :

i	Nr.	1	2
4	4	1.5	0.7
6	6	1.1	1.0
9	9	0.8	0.5
11	11	2.0	0.7
12	12	1.2	0.5

Cluster 2 :

i	Nr.	1	2
2	2	3.0	1.7
3	3	2.6	1.8
5	5	2.5	2.2
18	18	2.1	2.0

Cluster 3 :

i	Nr.	1	2
7	7	2.8	3.1
8	8	1.7	3.2

```

10 10  2.1 3.0
14 14  2.5 3.0
16 16  2.2 3.2
17 17  2.1 3.5

```

Cluster 4 :

```

i Nr.  1  2
1  1  4.0 1.0
13 13  4.5 0.7
15 15  3.5 0.7
19 19  3.5 1.2
20 20  3.2 1.1

```

VORSICHT: Die Numerierung (Nr.) der Objekte bezieht sich auf die Reihenfolge der Eingabedatei und berücksichtigt eventuell fehlende Werte/Gruppen!

Literatur:

H. Späth (1977) Cluster-Analyse-Algorithmen. Oldenbourg-Verlag München-Wien, 2. Auflage.
 J. Rubin (1967) Optimal Classification into Groups: An Approach for Solving the Taxonomy Problem. J. Theor. Biol. 15, pp. 103-144.
 J. McQueen (1967) Some Methods for Classification and Analysis of Multivariate Observations. Proc. 5th Berkely Symp. Math. Stat. Prob. 65/66, pp. 281-297.

Single-Linkage-Verfahren: Das Verfahren wird mit den Daten von Deichsel und Trampisch (1985, p.3) unter Verwendung der Euklidischen Distanz durchgeführt. Die 20 verwendeten Wertepaare ergeben sich aus dem folgenden Ausdruck der vier ermittelten Cluster:

Single Linkage: Euklidische Distanz ▲

Eingabedatei ist Deichsel003.DAT
 davon Variablen 1 (AP[U/l]), 2 (Fe[mg/l])
 Die Daten werden untransformiert verarbeitet.

Minimalbaum der Clusterung:

Nr	Nachb	Rang	Distanz		0	1.5556
4	3	20	1.5556	XX		
2	1	19	1.2207	XX		
8	7	18	1.1045	XX		
7	5	16	0.9487	XX		
13	1	15	0.5831	XXXXXXXXXXXXXXXXXXXX		
15	1	14	0.5831	XXXXXXXXXXXXXXXXXXXX		
9	6	13	0.5831	XXXXXXXXXXXXXXXXXXXX		
19	15	12	0.5000	XXXXXXXXXXXXXXXXXXXX		
11	4	11	0.5000	XXXXXXXXXXXXXXXXXXXX		
6	4	10	0.5000	XXXXXXXXXXXXXXXXXXXX		
18	5	9	0.4472	XXXXXXXXXXXXXXXXXXXX		
10	8	8	0.4472	XXXXXXXXXXXXXXXXXXXX		
3	2	7	0.4123	XXXXXXXXXXXXXXXXXXXX		
5	3	6	0.4123	XXXXXXXXXXXXXXXXXXXX		
12	4	5	0.3606	XXXXXXXXXXXX		
17	16	4	0.3162	XXXXXXXXXX		
20	19	3	0.3162	XXXXXXXXXX		
14	7	2	0.3162	XXXXXXXXXX		

Cluster für Schwellendistanz 1.:



Cluster 1

i	Nr.	Werte.....	
1	1	4.0000	1.0000
13	13	4.5000	0.7000
15	15	3.5000	0.7000
19	19	3.5000	1.2000
20	20	3.2000	1.1000

Cluster 2

i	Nr.	Werte.....	
2	2	3.0000	1.7000
3	3	2.6000	1.8000
5	5	2.5000	2.2000
7	7	2.8000	3.1000
14	14	2.5000	3.0000
18	18	2.1000	2.0000

Cluster 3

i	Nr.	Werte.....	
4	4	1.5000	0.7000
6	6	1.1000	1.0000
9	9	0.8000	0.5000
11	11	2.0000	0.7000
12	12	1.2000	0.5000

Cluster 4

i	Nr.	Werte.....	
8	8	1.7000	3.2000
10	10	2.1000	3.0000
16	16	2.2000	3.2000
17	17	2.1000	3.5000

VORSICHT: Die Numerierung (Nr.) der Objekte bezieht sich auf die Reihenfolge der Eingabedatei und berücksichtigt eventuell fehlende Werte/Gruppen!

Literatur:

H. Späth (1977) Cluster-Analyse-Algorithmen. Oldenbourg-Verlag München-Wien.
G. Deichsel, H.J. Trampisch (1985) Cluster- und Diskriminanzanalyse. Fischer-Verl.

Alternative Distanzmaße: Neben der Euklidischen Distanz kann man in **BiAS**. noch einige andere Distanzmaße verwenden: Die Euklidische Distanz gilt für quantitative Merkmale und ist definiert durch die Beziehung $d_{ij} = \sqrt{(\sum (x_{ik} - x_{jk})^2)}$. Die L1-Norm (City-Block-Metrik, Manhattan-Metrik) ist auf quantitative (speziell diskrete) und, falls äquidistante (?) Stufen vorliegen, auch auf ordinale Merkmale anwendbar: $d_{ij} = \sum |x_{ik} - x_{jk}|$. Die Canberra-Metrik $d_{ij} = \sum (|x_{ik} - x_{jk}| / |x_{ik} + x_{jk}|)$ kann auf diskrete, quantitative und ordinale Merkmale angewendet werden. Die Tanimoto-Distanz setzt Merkmale mit den Ausprägungen "0" und "1", also binäre Daten voraus; die Distanz ist definiert durch $d_{ij} = 1 - (\text{Anzahl Komponenten mit } "x_{ik}=1 \text{ und } x_{jk}=1") / (\text{Anzahl Komponenten mit } "x_{ik}=1 \text{ oder } x_{jk}=1")$. Die Tanimoto-Distanz wichtet "1" (z.B. "vorhanden") stärker als "0" ("nicht vorhanden"); Alternative ist die Simple-Matching-Distanz, die diese Eigenschaft nicht besitzt. Die Simple-Matching-Distanz ist auf diskret-quantitative, ordinale und nominale (resp. binäre) Merkmale anwendbar und errechnet sich aus $d_{ij} = \sum (x_{ik} \neq x_{jk}) / N$. Es sollte darauf geachtet werden, dass alle Skalen möglichst die gleiche Anzahl von Stufen besitzen, um eine Dominanz von Skalen mit geringer Stufenzahl zu verhindern. (Literatur: Sokal und Sneath (1973))

Das Ergebnis einer Analyse kann von der unterschiedlichen Variabilität der untersuchten Größen abhängig sein. **BiAS**. sieht deshalb in einigen Fällen eine Standardisierung mit der z-Transformation $x' = (x - \bar{x}) / s$ vor. Andere Standardisierungen sind möglich (vgl. auch **BiAS**.! Formelinterpreter).

Nach Berechnung mehrerer Cluster-Läufe kann ein einfaches Struktogramm der F-Werte in Abhängigkeit von der Cluster-Zahl gezeigt werden. Analog zum Scree-Test der Faktorenanalyse



Lineare Diskriminanzanalyse

2D-Graphik zur Linearen Diskriminanzanalyse

Abbaumodell zur Linearen Diskriminanzanalyse

Quadratische Diskriminanzanalyse

Stoller's nicht-parametrische Diskriminanzanalyse

Univariate parametrische Diskriminanzanalyse

ROC-Kurven zur Diskrimination zweier Gruppen (A.18)

Graphik zu ROC-Kurven (A.18)

Propensity-Score per Diskriminanzanalyse

KFA zur Diskrimination binärer und kategorialer Daten (A.5)

Lineare Diskriminanzanalyse: Eine Gruppe Hämophilie-Überträgerinnen soll anhand der Parameter "Faktor VIII Aktivator" und "Faktor VIII Antikörper" von einer zweiten Gruppe von Nicht-Überträgerinnen optimal getrennt werden (Deichsel und Trampisch (1985, p.86)). Bemerkung: Im Vergleich zur Originalarbeit sind Abweichungen in den Daten festzustellen, so dass sich möglicherweise aus diesem Grund auch geringfügig abweichende Resultate ergeben. Im Zweidimensionalen ist auch eine Graphik möglich.

Die Zuordnungswahrscheinlichkeiten "Prob" in der Ausgabe entsprechen dem "Propensity-Score", der hier nicht explizit aufgeführt wird.

Lineare Diskriminanzanalyse

Eingabedatei ist Herman000.DAT
davon Variablen 1 (F8_Aktiv), 2 (F8_Antik)
und Gruppen Var. 3 (Übertr.) = 1 und 2

Durchschnitte der Variablen:

Var	Gruppe 1	Gruppe 2
1	-0.2474	0.0140
2	0.0400	-0.0383

Kovarianzmatrix (pooled):

0.0184	0.0121
0.0121	0.0199

Homogenität der Kovarianzmatrizen:

Box' Prüfgröße $\chi^2 = 1.9926$ mit $df = 3$ ($p = 0.573940$)

Fisher's Lineare Diskriminanzfunktion $h = (b)*y - c = \text{Summe}(b_i*y_i) - c$:

(Mit den a-priori-Wahrscheinlichkeiten π_1 und π_2 für die beiden Gruppen: $h > \ln(\pi_2/\pi_1)$: zu Gruppe 1, $h < \ln(\pi_2/\pi_1)$: zu Gruppe 2.)

Konstante c = 3.2893
Koeffizient b(1) = -28.0376
Koeffizient b(2) = 20.9792

Mahalanobis' $D^2 = 8.9721$
Hotelling's $T^2 = 116.8072$
Fisher's Prüfgröße F = 57.2584 mit $df = (2, 50)$ ($p = 0.000000$)

Gruppenmittelwerte der Fisher'schen Diskriminanzfunktion:

Gruppe 1: 4.4861 und Gruppe 2: -4.4861

Fehlklassifikationswahrscheinlichkeit bei *g l e i c h e n*
a-priori-Wahrscheinlichkeiten der Gruppen ist $P = 0.067109$

Zuordnung der eingelesenen Wertetupel bei *gleichen* a-priori-Wahrscheinlichkeiten
 π_1 und π_2 für die Gruppen; * bedeutet eine *falsche* Zuordnung des Wertetupels!
Prob= $p(1)$ ist die Wahrscheinlichkeit für Gruppe 1 (Modell) und der *Propensity-Score*

Gruppe	Zuord	Prob	Wertetupel...	
1	1	0.999967	-0.47	0.02
1	1	0.816193	-0.38	-0.28
1	1	0.956912	-0.40	-0.23
1	1	0.966913	-0.32	-0.11
1	1	0.999322	-0.40	-0.03
1	1	0.999779	-0.41	0.01
1	2*	0.466882	-0.15	-0.05
1	1	0.935299	-0.25	-0.05
1	1	0.930894	-0.24	-0.04
1	1	0.915966	-0.21	-0.01
1	1	0.943125	-0.21	0.01
1	1	0.999320	-0.34	0.05
1	1	0.999967	-0.47	0.02
1	2*	0.363526	-0.03	0.09
1	1	0.997237	-0.26	0.09
1	1	0.999318	-0.31	0.09
1	1	0.569319	-0.03	0.13
1	2*	0.446937	0.01	0.16
1	1	0.778924	-0.05	0.15
1	1	0.998415	-0.22	0.17
1	1	0.992626	-0.15	0.19
1	1	0.999443	-0.19	0.26
1	1	0.999791	-0.21	0.28
2	2	0.169738	-0.36	-0.40
2	2	0.031775	-0.19	-0.26
2	2	0.000000	0.24	-0.26
2	2	0.036280	-0.15	-0.20
2	2	0.002805	-0.05	-0.19
2	2	0.003702	-0.03	-0.15
2	2	0.025720	-0.07	-0.11
2	2	0.274125	-0.12	-0.05
2	2	0.157924	-0.11	-0.07
2	2	0.001831	0.07	-0.05
2	2	0.006443	0.01	-0.07
2	2	0.005230	0.01	-0.08
2	2	0.109555	-0.08	-0.05
2	2	0.090702	-0.08	-0.06
2	2	0.003995	-0.13	-0.28
2	2	0.007920	0.04	-0.02
2	2	0.004536	0.06	-0.02
2	2	0.009732	0.07	0.03
2	2	0.012841	0.06	0.03
2	2	0.004847	0.11	0.05
2	2	0.043876	0.03	0.05
2	2	0.007355	0.11	0.07
2	2	0.003659	0.15	0.09
2	2	0.014702	0.10	0.09
2	2	0.102135	0.05	0.12
2	2	0.001280	0.21	0.12
2	2	0.176188	0.02	0.11
2	2	0.016861	0.11	0.11
2	2	0.012791	0.12	0.11
2	2	0.004187	0.22	0.19

Lineare Diskriminanzanalyse: Abbauverfahren

Eingabedatei ist ...\\Test.DAT

davon Variablen 1 (VarEins), 2 (VarZwei)
und Gruppen Var. 3 (GruppNum) = 1 und 3

Abbau des Modells: Eliminationskriterium ist alpha=0.1 !

Prüfung des vollständigen Modells mit 2 Variablen:

Multiples Bestimmtheitsmaß B = 0.9843

Prüfung der partiellen Koeffizienten b(i):

Variable	b(i)	s(b(i))	F	df1	df2	p
VarEins	-0.00649	1.1863E-02	2.9897E-01	1	5	0.608036
VarZwei	-0.02589	1.1754E-02	4.8524E+00	1	5	0.078814

Modellabbau: Die Einflussgröße VarEins wird jetzt eliminiert:

Multiples Bestimmtheitsmaß B = 0.9834

Prüfung der partiellen Koeffizienten b(i):

Variable	b(i)	s(b(i))	F	df1	df2	p
VarZwei	-0.03224	1.7127E-03	3.5438E+02	1	6	0.000001

Abbau beendet: Noch vorhandene Variable/n signifikant bei alpha=0.1!

Zur Ermittlung der Klassifikationsraten und der Zuordnungen der Werte auf Basis des minimalen Modells kann die Lineare Diskriminanzanalyse ('volles Modell') verwendet werden!

Literatur:

M.M. Tatsuoka (1971) Multivariate Analysis. John Wiley.

B. Flury, H. Riedwyl (1983) Angew. multivariate Statistik, Fischer.

Quadratische Diskriminanzanalyse: In der Literatur wurde kein komplett durchgerechnetes Beispiel zur QDA gefunden, so dass diese anhand der Publikation des Inst. of Math. Stat. (1989) und der LDA verifiziert wurde.

Quadratische Diskriminanzanalyse

Eingabedatei ist Herman000.DAT

davon Variablen 1 (F8_Aktiv), 2 (F8_Antik)
und Gruppen Var. 3 (Übertr.) = 1 und 2

Eingabedatei/daten wie zur Linearen Diskriminanzanalyse!

Durchschnitte der Variablen:

Var	Gruppe 1	Gruppe 2
1	-0.2474	0.0140
2	0.0400	-0.0383

Kovarianzmatrix 1 :

0.0198	0.0102
0.0102	0.0190

Kovarianzmatrix 2 :

0.0173	0.0135
0.0135	0.0206

Homogenität der Kovarianzmatrizen:

Box' Prüfgröße $\chi^2 = 1.9926$ mit $df = 3$ ($p = 0.573940$)

Die Prüfung auf p-variate Gauß-Verteilung erfolgt im Modul 'Gauß-Verteilung'!

Quadratische Diskriminanzfunktion $h = y*(MatB)*y-y*(b)+c1$ <?> $c2+2*\ln(\pi2/\pi1)$
(Mit den a-priori-Wahrscheinlichkeiten $\pi1$ und $\pi2$ für die beiden Gruppen:
 $h > c2+2*\ln(\pi2/\pi1)$: zu Gruppe 1, $h < c2+2*\ln(\pi2/\pi1)$: zu Gruppe 2)

Matrix MatB:

48.1448	-39.9215
-39.9215	26.5970

Vektor b:

Koeffizient b(1) = 47.0089
Koeffizient b(2) = -34.3332

Konstante c1 = -4.9059
Konstante c2 = -0.4416

Gruppenmittelwerte der Quadratischen Diskriminanzfunktion:

Gruppe 1	Gruppe 2
11.8763	-6.7887

Zuordnung der eingelesenen Wertetupel bei g l e i c h e n
a-priori-Wahrscheinlichkeiten $\pi1$ und $\pi2$ für die Gruppen;
* bedeutet eine f a l s c h e Zuordnung des Wertetupels!

Gruppe	Zuordnung	Wertetupel...
1	1	-0.47 0.02
1	1	-0.38 -0.28
1	1	-0.40 -0.23
1	1	-0.32 -0.11
1	1	-0.40 -0.03
1	1	-0.41 0.01
1	1	-0.15 -0.05
1	1	-0.25 -0.05
1	1	-0.24 -0.04
1	1	-0.21 -0.01
1	1	-0.21 0.01
1	1	-0.34 0.05
1	1	-0.47 0.02
1	1	-0.03 0.09
1	1	-0.26 0.09
1	1	-0.31 0.09
1	1	-0.03 0.13
1	1	0.01 0.16
1	1	-0.05 0.15
1	1	-0.22 0.17
1	1	-0.15 0.19
1	1	-0.19 0.26
1	1	-0.21 0.28
2	2	-0.36 -0.40
2	2	-0.19 -0.26
2	2	0.24 -0.26
2	2	-0.15 -0.20
2	2	-0.05 -0.19
2	2	-0.03 -0.15
2	2	-0.07 -0.11
2	2	-0.12 -0.05
2	2	-0.11 -0.07
2	2	0.07 -0.05
2	2	0.01 -0.07
2	2	0.01 -0.08
2	2	-0.08 -0.05
2	2	-0.08 -0.06

Stoller's nicht-parametrische Diskriminanzanalyse: Stoller (1954) versucht, zwei a-posteriori klassifizierte Gruppen von Studenten auf Grund eines Vortests zu trennen:

STOLLER's nicht-parametrische Diskriminanzanalyse ▲

Eingabedatei ist Stoller772.DAT
davon Variablen 1 (StudScor)
und Gruppen Var. 2 (Gruppe) = 1 und 2

Lfd Nr.	Gru ppe	Sym bol	Eingabe wert X	Trennpkt Xtrenn	Ungl. Regel 1	a-priori-Ws Regel 2	Gleiche Regel 1	a-priori-Ws Regel 2
1	1	x	50.000	50.500	0.3600*	0.6400	0.5294	0.4706
2	1	x	51.000	54.250	0.4000	0.6000	0.5588	0.4412
3	1	x	57.500	60.250	0.4400	0.5600	0.5882	0.4118
4	1	x	63.000	63.500	0.4800	0.5200	0.6176	0.3824
5	1	x	64.000	66.000	0.5200	0.4800	0.6471	0.3529
6	1	x	68.000	70.000	0.5600	0.4400	0.6765	0.3235
7	1	x	72.000	72.500	0.6000	0.4000	0.7059	0.2941
8	1	x	73.000	73.500	0.6400	0.3600	0.7353	0.2647
9	2	x	74.000	74.250	0.6000	0.4000	0.6728	0.3272
10	1	x	74.500	74.750	0.6400	0.3600	0.7022	0.2978
11	2	x	75.000	75.250	0.6000	0.4000	0.6397	0.3603
12	1	x	75.500	75.750	0.6400	0.3600	0.6691	0.3309
13	2	x	76.000	76.250	0.6000	0.4000	0.6066	0.3934
14	1	x	76.500	76.750	0.6400	0.3600	0.6360	0.3640
15	1	x	77.000	77.500	0.6800	0.3200	0.6654	0.3346
16	1	x	78.000	79.500	0.7200	0.2800	0.6949	0.3051
17	1	x	81.000	81.500	0.7600	0.2400	0.7243	0.2757
18	1	x	82.000	83.000	0.8000	0.2000*	0.7537	0.2463*
19	2	x	84.000	84.500	0.7600	0.2400	0.6912	0.3088
20	2	x	85.000	85.500	0.7200	0.2800	0.6287	0.3713
21	2	x	86.000	86.750	0.6800	0.3200	0.5662	0.4338
22	1	x	87.500	88.250	0.7200	0.2800	0.5956	0.4044
23	2	x	89.000	90.000	0.6800	0.3200	0.5331	0.4669
24	2	x	91.000	91.500	0.6400	0.3600	0.4706*	0.5294
25	1	x	92.000	92.000	0.6800	0.3200	0.5000	0.5000

Bei ungleichen a priori-Wahrscheinlichkeiten: Regel 2
X < Xtrenn : Zuordnung zu Gruppe 1
X > Xtrenn : Zuordnung zu Gruppe 2

Die minimale Fehlklassifikationsrate p = 0.2000 ist durch '*' markiert. Xtrenn aus der Tabelle ablesen !

Bei gleichen a priori-Wahrscheinlichkeiten: Regel 2
X < Xtrenn : Zuordnung zu Gruppe 1
X > Xtrenn : Zuordnung zu Gruppe 2

Die minimale Fehlklassifikationsrate p = 0.2463 ist durch '*' markiert. Xtrenn aus der Tabelle ablesen !

Literatur:

D.S. Stoller (1954) Univariate two-population distribution-free discrimination. J. Amer. Statist. Assoc. Vol. 49, pp. 770 - 777.

Univariate parametrische Diskriminanzanalyse: Trennung zweier Gruppen anhand *eines* Parameters (Berechnungen per Hand überprüft via Test.DAT):

Univariate parametrische Diskriminanzanalyse ▲

Eingabedatei ist Test.DAT
 davon Variablen 2 (VarZwei)
 und Gruppen Var. 3 (GruppNum) = 2 und 3

Test auf Gauß-Verteilung (Dallal-Wilkinson-korrigiert):
 Gruppe 2 : Kolmogoroff-Smirnoff's delta = 0.1747 (p>0.200000) Ok!
 Gruppe 3 : Kolmogoroff-Smirnoff's delta = 0.2178 (p>0.200000) Ok!

Vergleich der Varianzen (zweiseitiger F-Test):
 Prüfgröße F = 2.61 mit df=(2,3) und p=0.441175 Ok!

	Xquer	s	s ²	n
Stichprobe 2	25.0000	5.0000	25.0000	3
Stichprobe 3	34.7500	3.0957	9.5833	4

Bei gleichen a-priori-Wahrscheinlichkeiten der beiden Gruppen:

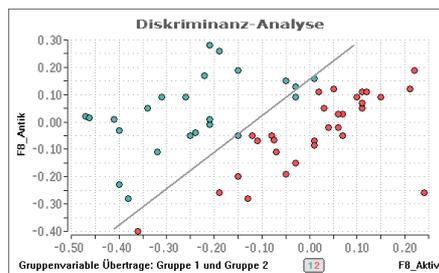
Optimaler Trennpunkt (gewichtet): 29.8750
 X < Trennpunkt: Zuordnung zu Gruppe 2
 X > Trennpunkt: Zuordnung zu Gruppe 3

Zuordnungsraten (Resubstitution):

	richtig	falsch	gesamt
Gruppe 2: n	2	1	3
%	66.67	33.33	100.00
Gruppe 3: n	4	0	4
%	100.00	0.00	100.00
gesamt n	6	1	7
%	85.71	14.29	100.00

Literatur: M.M. Tatsuoaka (1971) Multivariate Analysis. John Wiley.
 B. Flury, H. Riedwyl (1983) Angew. multivariate Statistik, Fischer.

Graphik 2D-Diskriminanzanalyse: Trennung von zwei Gruppen (Überträgerinnen und Nicht-Überträgerinnen der Haemophilie) anhand der beiden Parameter FaktorVIII-Aktivität und -Antikörper (Datei Herman.000). Die Berechnungen und Optionen finden sich an anderer Stelle. Die Darstellung erfolgt hier optional mit farbig unterscheidbaren Punkte (Legende), eine Darstellung mit Gruppennummern ist ebenfalls möglich.



A.13 Zeitreihenanalyse



Varianzanalyse mit Messwiederholungen

Mann-Test auf Trend

Iterationstest

Wallis-Moore's Phasenhäufigkeitstest

Cochran's Niveauekonstanz-Test

Neumann's Trend-Test

Page-Test auf Trend

Durbin-Watson-Test

Ljung-Box-Test

Deskriptives (\bar{X} quer, s etc.)

Graphik Box-Plot Zeitverläufe

Graphik Bar-Plot Zeitverläufe

Graphik Einzelverläufe

Cusum-Analyse für dichotome und quantitative Daten

Varianzanalyse mit Messwiederholungen: Es wird geprüft, ob der Haut-Widerstand Tageschwankungen unterliegt, wozu eine Gruppe von Probanden morgens, mittags und abends untersucht wurde (Bortz (1977,2005, p.412)). Eine graphische Darstellung ist mit Einzelverläufen und Box-Plots möglich, weitere Berechnungen finden sich in "Deskriptives".

Unifaktorielle Varianzanalyse mit Messwiederholungen

Eingabedatei ist Bortz412.DAT

davon Variablen 1 (morgens), 2 (mittags), 3 (abends)

i	morgens	mittags	abends
1	7.0000	7.0000	6.0000
2	5.0000	6.0000	8.0000
3	8.0000	9.0000	5.0000
4	6.0000	8.0000	6.0000
5	7.0000	7.0000	5.0000
6	7.0000	9.0000	7.0000
7	5.0000	10.0000	6.0000
8	6.0000	7.0000	4.0000
9	7.0000	8.0000	6.0000
10	5.0000	7.0000	5.0000

Tabelle der Varianzanalyse:

Source	SQ	MQ	DF	F	p
zwischen	11.6333	1.2926	9	0.9831	0.485678
innerhalb	45.3333	2.2667	20		
Zeit	21.6667	10.8333	2	8.2394	0.002881
Rest	23.6667	1.3148	18		
Gesamt	56.9667	1.9644	29		

Zum Vergleich der Faktorstufen wird der p-Wert bzgl. "Zeit" verwendet.

Prüfung der Voraussetzungen (Homogenität der Varianzen und der Korrelationen zwischen den Stufen des Faktors: Box-Test) :

Chi² = 0.9249 mit df = 4 (p = 0.920965) Ok!

Test auf Gauß-Verteilung der 3 Variablen (Dallal-Wilkinson-korrigiert):

morgens: Kolmogoroff-Smirnoff's delta = 0.2456 (p = 0.088720) Cave!

mittags: Kolmogoroff-Smirnoff's delta = 0.2424 (p = 0.098658) Cave!

abends: Kolmogoroff-Smirnoff's delta = 0.2301 (p > 0.100000) Ok!

Multiple Vergleiche - Bonferroni-Holm-korrigierte t-Tests und Scheffé-Tests:

Nr	Mult. Paarvergleich	p-unkorr	p-Bonferroni	p-Bonf.Holm	p-Scheffé
1	morgens....mittags.	0.009106	0.027318	0.018212	0.030195
2	morgens....abends..	0.380856	1.000000	0.380856	0.629252
3	mittags....abends..	0.004763	0.014289	0.014289	0.004035

Literatur: J. Bortz (2005) Lehrbuch der Statistik, Springer. pp. 423ff.

Mann-Test auf Trend: Besteht in den Jahren 1940-1955 ein Trend in der Anzahl unehelicher Geburten im Staate Kansas (Hartung (1982, p.250))?

Mann-Test auf Trend ▲

Eingabedatei ist Hartung250.DAT

davon Variablen 1 (1940), 2 (1945), 3 (1947), 4 (1950), 5 (1955)

i	Var. 1	Var. 2	Var. 3	Var. 4	Var. 5
1	318.0000	593.0000	564.0000	530.0000	694.0000

Durchschnitte der Variablen:

Nr	Var	Xquer
1	1	318.0000
2	2	593.0000
3	3	564.0000
4	4	530.0000
5	5	694.0000

Prüfung auf Trend für die Zeitreihe der Durchschnitte (zweiseitig):

Prüfgröße (Gauß) N = 0.9798 (p = 0.327187)

Prüfung auf Trend für die 1 unabhängige Zeitreihe mit der 'Inversen Normalverteilung' (zweiseitig):

Prüfgröße (Gauß) N = 0.9798 (p = 0.327187)

Literatur:

H.B. Mann (1945) Nonparametric Tests against Trend. Econometrika 13, pp. 245-9.

J. Hartung (1982) Statistik. Oldenburg München Wien.

B.J. Winer (1971) Statistical Principles in Experimental Design. McGraw Hill NY.

Cochran's Niveauekonstanz-Test: Es werden zum Beispiel Hormon-Werte einer Patientin untersucht. Es stellt sich die Frage, ob die Hormonwerte vor und während der Schwangerschaft konstant bleiben oder ob man von einer Niveaueverschiebung ausgehen muss:

Cochran's Niveauekonstanz-Test ▲

Durchschnitt X_{q1} : 12.1
Anzahl Messwert n_1 : 12
Durchschnitt X_{q2} : 15.8
Anzahl Messwert n_2 : 14

$\text{Chi}^2 = 6.2771$
mit $df=1$
und $p=0.012231$

Literatur:

Cochran (1954) Some methods for strengthening the common Chi^2 -tests. Biometrics 10, pp. 417-51

Neumann's Trend -Test: Der Test wird häufig im Zusammenhang mit der Qualitätskontrolle im Labor eingesetzt. In Anhang A.16 (Qualitätssicherung im Labor, RiLiBÄK) findet sich ein durchgerechnetes Beispiel, weitere Überprüfung durch per-Hand-Berechnungen. Im Programm wird die exakte Methode nach Hart verwendet, ab $n=60$ eine asymptotische Methode (Sachs 2003).

Neumann's Trend-Test ▲

Eingabedatei ist ... \VALIDE\Systemex.DAT
davon Variable 2 (Geraet1)

Eingelesene Daten:

1: 4.54	2: 4.56	3: 4.52	4: 4.51	5: 4.53
6: 4.51	7: 4.58	8: 4.54	9: 4.53	10: 4.50
11: 4.54	12: 4.48	13: 4.56	14: 4.52	15: 4.53
16: 4.54	17: 4.52	18: 4.51	19: 4.58	20: 4.51
21: 4.53	22: 4.54	23: 4.49	24: 4.56	25: 4.50

$X_{\text{quer}} = 4.5292$
 $SD=s = 0.0258$
 $s^2 = 0.0007$
 $n = 25$

Test auf Gauß-Verteilung der Werte (Dallal-Wilkinson-korrigiert):

Kolmogoroff-Smirnoff's delta = 0.1378 ($p > 0.100000$) Ok!

Trendtest nach Neumann et al.:

Prüfgröße = 2.8529: $p > 0.050$ (einseitig exakt) > kein Trend.

Literatur: J Neumann, RH Kent, HB Bellinson, BI Hart (1941) The mean square successive difference. Annals of Mathematical Statistics Vol.12, pp.153-62.

Iterationstest: Bildet die Folge von 8 Münzwürfen W W W Z W W Z Z eine zufällige Folge oder existieren "Gruppierungen"? (Aus: Sachs (1992, Kap. 472)) Als Eingabe in das Programm wird die Folge 1 1 1 2 1 1 2 2 verwendet, da BiAS. keine alphanumerischen, sondern nur numerische Eingaben akzeptiert.

Iterationstest (Wald-Wolfowitz) ▲

Bezeichnung: Sachs (1992, Kap. 472)

i	Var.1	Var.2	Var.3	Var.4	Var.5	Var.6	Var.7	Var.8
1	1.00	1.00	1.00	2.00	1.00	1.00	2.00	2.00

Durchschnitte der Variablen:

Nr	Var	Xquer
1	1	1.0000
2	2	1.0000
3	3	1.0000
4	4	2.0000
5	5	1.0000
6	6	1.0000
7	7	2.0000
8	8	2.0000

Iterationstest für die Zeitreihe der Durchschnitte (zweiseitig):
 Prüfgröße (Gauß) N = -0.6179 (p = 0.536632) (Wenig Iterationen!)

Iterationstest für 1 Zeitreihe (zweiseitig):
 Prüfgröße (Gauß) N = -0.6179 (p = 0.536632) (Wenig Iterationen!)

Der Iterationstest ist nur für Alternativdaten und stetige Skalen zulässig. Für 1) ordinale Skalen mit wenigen Stufen und 2) für mittlere Zeitreihen bei ordinalen und nominalen Skalen und kleiner Fallzahl sind Verzerrungen zu erwarten!

Literatur:

- H. Büning, G. Trenkler (1978) Nicht-param. stat. Methoden. De Gruyter.
- L. Sachs (1992) Angewandte Statistik (Kap. 472), Springer.

Wallis-Moore's Phasenhäufigkeitstest: Die Aufeinanderfolge gleicher Vorzeichen von Differenzen $X(i+1)-X(i)$ wird von Wallis und Moore (1941) als "Phase" bezeichnet; der Test basiert auf der Häufigkeit von Plus- und Minusphasen. Ist die Gesamtzahl der Phasen klein, so weist diese auf eine "Trendbeharrlichkeit" hin. (Sachs errechnet die Prüfgröße von $N=2.56$)

Wallis-Moore's Phasenhäufigkeitstest ▲

Eingabedatei ist Sachs486.DAT
 davon Variablen 1 (M1) bis 22 (M22)

VarNr:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Werte:	5	6	2	3	5	6	4	3	7	8	9	7	5	3	4	7	3	5	6	7	8	9

Phasenhäufigkeitstest für 1 Folge/Zeitreihe (zweiseitig)
 bei insgesamt 7 relevanten Phasen (erwartet: 12.3).
 Prüfgröße (Gauß) N = -2.5513 (p = 0.010731) ('Wenig Phasen')

Literatur:

- W.A. Wallis, G.H. Moore (1941) A significance test for time series. J. Amer. Statist. Assoc. 36, pp. 401-409.
- L. Sachs (1997) Angewandte Statistik (Kap. 473), 8. Aufl. Springer.

Durbin-Watson-Test: Beurteilung der zeitlichen Konstanz eines Zielparameters in Abhängigkeit von der Zeit. Diese Tests testen auf Autokorrelation erster Ordnung, d.h., ob die Werte einer Reihe von den jeweiligen unmittelbaren Vorgängern geprägt sind.

Der Ljung-Box-Test wird hier ebenfalls nur für Autokorrelationen *erster* Ordnung verwendet, höhere Ordnungen als hier können im Menü unter "Ljung-Box-Test" berechnet werden.

Der Test erfolgt nach linearer Regression auf Basis der Residuen des Modells oder alternativ mit den "Residuen" bezüglich des Mittelwertes der Zielvariablen.

Die Prüfgröße zum DW-Test wird zunächst „klassisch-konservativ“ anhand von Tabellen aus Savin und White (1977) geprüft. Als alternative Testgröße wird eine t-verteilte Größe verwendet, dazu SPSS20. Eine Überprüfung der Ergebnisse wurde in Teilen mit SPSS20 vorgenommen.

Durbin-Watson-Test

Eingabedatei ist ..\Valide\Test.DAT

Zielgröße Y = Variable 1 = VarEins

Einflussgröße X = Variable 2 = VarZwei

Anzahl fehlender Werte: 1

Regression und Korrelation "Y auf X":

Regressionskoeffizient b =	0.9218
Konstanter Achsenabschnitt c =	-1.2672
Korrelationskoeffizient r =	0.9362
Bestimmtheitsmaß B =	0.8764
Testgröße der Regression t =	7.9885
Überschreitungswahrsch. p =	0.000022
Stichprobenumfang n =	11
Anzahl der Freiheitsgrade n-2 =	9

Konfidenzintervalle für die Steigung=Regressionskoeffizient:

P = 0.90:	(0.7103,	1.1333)
P = 0.95:	(0.6608,	1.1828)
P = 0.99:	(0.5468,	1.2968)

Konfidenzintervalle für den Achsenabschnitt auf der Y-Achse:

P = 0.90:	(-6.5492,	4.0148)
P = 0.95:	(-7.7854,	5.2511)
P = 0.99:	(-10.6314,	8.0970)

>> Beide Tests untersuchen die Residuen bezüglich der Regressionsgeraden.
>> Die Daten werden gemäß der Werte der Variablen X=Zeit=VarZwei geordnet.

Durbin-Watson-Test mit Prüfgröße D und Durbin's h (approximativ):

Durbin-Watson's Prüfgröße D =	1.564692
Durbin-Watson's Korrelationskoeffizient: r =	0.217654
Tabellarischer Test gemäß Savin & White: p >	0.100000
Test via t(df=8), Testgröße t=0.631: p =	0.545803
Durbin's Gauß-verteilte Prüfgröße h =	0.781364
StdAbw des Regressionskoeffizienten: s =	0.115392
Zweis. Überschreitungswahrscheinlichkeit: p =	0.434588

Ljung-Box-Test mit linearem Lag 1. Ordnung:

Korrelationskoeff. Res mit Lag 1. Ordnung r =	0.185717
Ljung und Box' Prüfgröße (Chi ² (df=1)) Q =	0.493218
Zweis. Überschreitungswahrscheinlichkeit: p =	0.482496

Literatur:

J. Durbin, G.S. Watson (1950,1951) Testing for Serial Correlation in Least Squares Regression. I: Biometrika 37, 409-28, II: Biometrika 38, 159-79.
 G.M. Ljung, G.E.P. Box (1978) On a Measure of a Lack of Fit in Time Series Models. Biometrika 65, 297-303.
 N.E. Savin, K.J. White (1977) The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors. Econometrica 45, 1989-96.

Ljung-Box-Test: Beurteilung der zeitlichen Konstanz eines Zielparameters in Abhängigkeit von der Zeit. Der Test beruht auf Autokorrelation k-ter Ordnung, d.h., ob die Werte einer Reihe von den jeweiligen unmittelbaren Vorgängern geprägt sind.

Der Test erfolgt nach linearer Regression auf Basis der Residuen des Modells oder alternativ mit den "Residuen" bezüglich des Mittelwertes der Zielvariablen.

Eine Überprüfung der Ergebnisse wurde mit anderen, bereits validierten Teilen des Programms und weiterhin mit per-Hand-Berechnungen vorgenommen.

Ljung-Box-Test ▲

Eingabedatei ist ..\Valide\Test.DAT

Zielgröße Y = Variable 1 = VarEins
 Einflussgröße X = Variable 2 = VarZwei

Ausgabe zur Regression wie zum Durbin-Watson-Test

CAVE: Empfehlung für n=11: Lag < eLog(11)=2.4 - hier problematisch.

Ljung-Box-Test mit linearem Lag bis 3. Ordnung und Chi²-Komponenten:

>> Der Test untersucht die Residuen bezüglich der Regressionsgeraden.
 >> Die Daten werden gemäß der Werte der Variablen X=Zeit=VarZwei geordnet.

Lag	n(Lag)	r	Chi ²	df	p	p(<=Lag)
1	10	0.185717	0.493	1	0.482496	0.482496
2	9	-0.163396	0.424	1	0.514846	0.632098
3*	8	-0.343834	2.113	1	0.146032	0.386926
Overall-Ljung-Box-Test:			3.031	3		0.386926

CAVE: Der Overall-Test ist wegen der mit "*" markierten Lags problematisch.

Literatur:

G.M. Ljung, G.E.P. Box (1978) On a Measure of a Lack of Fit in Time Series Models. Biometrika 65, 297-303.
 P. Brockwell, R. Davis (2002) Introduction to Time Series and Forecasting 2nd. ed. Springer.

Deskriptives (Zeitreihen): Berechnung von Zeilendurchschnitten, Streuungen, Medianen, Quartilen und Extrema für Verlaufsreihen. Graphische Darstellungen als Einzelverläufe oder Box-Plots sind möglich. Untersucht wird hier der Hautwiderstand zu 3 Zeitpunkten:

Deskriptives: Zeitreihen

Eingabedatei ist Bortz412.DAT

davon Variablen 1 (morgens), 2 (mittags), 3 (abends)

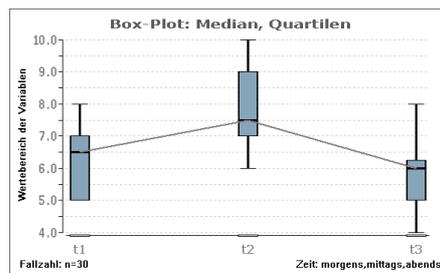
Für alle 3 Zeitpunkte/Spalten mit je n = 10 Versuchseinheiten/Zeilen:

Nr	Xquer	s	Median	Q1	Q3	Minimum	Maximum
1	6.3000	1.0593	6.5000	5.0000	7.0000	5.0000	8.0000
2	7.8000	1.2293	7.5000	7.0000	9.0000	6.0000	10.0000
3	5.8000	1.1353	6.0000	5.0000	6.2500	4.0000	8.0000

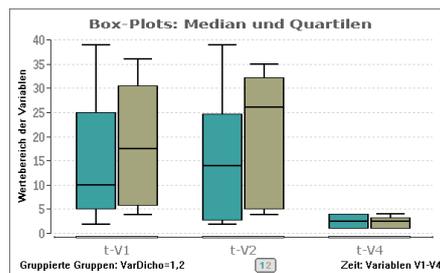
Für alle 10 Versuchseinheiten/Zeilen mit je k = 3 Zeitpunkten/Spalten:

Nr	Xquer	s	Median	Q1	Q3	Minimum	Maximum
1	6.6667	0.5774	7.0000	6.0000	7.0000	6.0000	7.0000
2	6.3333	1.5275	6.0000	5.0000	8.0000	5.0000	8.0000
3	7.3333	2.0817	8.0000	5.0000	9.0000	5.0000	9.0000
4	6.6667	1.1547	6.0000	6.0000	8.0000	6.0000	8.0000
5	6.3333	1.1547	7.0000	5.0000	7.0000	5.0000	7.0000
6	7.6667	1.1547	7.0000	7.0000	9.0000	7.0000	9.0000
7	7.0000	2.6458	6.0000	5.0000	10.0000	5.0000	10.0000
8	5.6667	1.5275	6.0000	4.0000	7.0000	4.0000	7.0000
9	7.0000	1.0000	7.0000	6.0000	8.0000	6.0000	8.0000
10	5.6667	1.1547	5.0000	5.0000	7.0000	5.0000	7.0000

Graphik Box-Plot-Verlaufsreihen: Untersuchung des Haut - Widerstands morgens, mittags und abends (Bortz (1977, p.412)), hier mit Median und Quartilen; als alternative Graphik kommen Bar-Plots und Einzelverläufe in Frage. Berechnungen vgl. ANOVA und "Deskriptives":



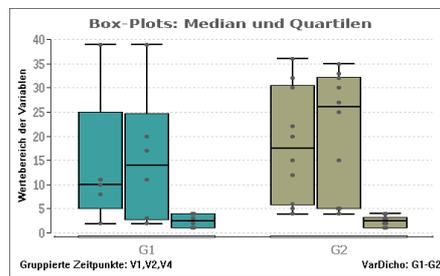
Gruppierte Darstellungen (hier mit anderen Daten) sind ebenfalls möglich:



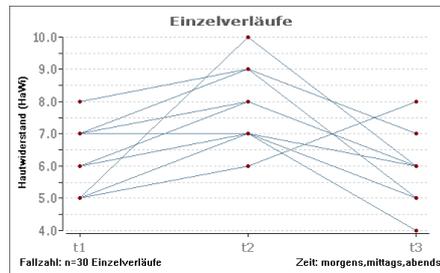
Graphik Bar-Plot-Verlaufsreihen: Untersuchung des Haut - Widerstands morgens, mittags und abends (Bortz (1977, p.412)), hier mit \bar{x} und Standardabweichung s; als alternative Graphik kommen Box-Plots und Einzelverläufe in Frage. Berechnungen vgl. ANOVA und "Deskriptives":



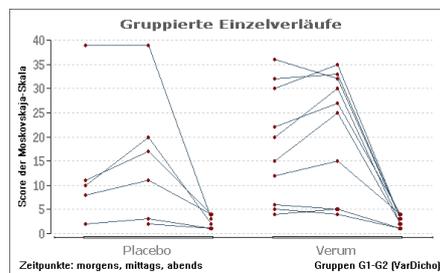
Gruppierte Darstellungen (hier mit anderen Daten und mit optional eingetragenen Werten, aus ..\Valide\Test.DAT) sind ebenfalls möglich:



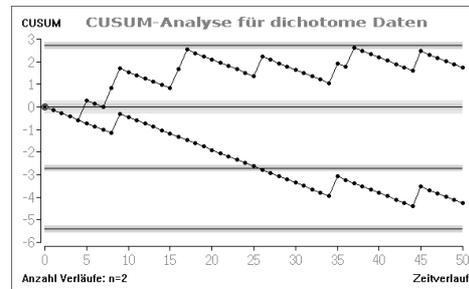
Graphik Einzelverläufe: Untersuchung des Haut-Widerstands morgens, mittags und abends (Bortz (1977, p.412)). Berechnungen finden sich zur ANOVA mit Messwiederholungen und zur Deskriptiven Statistik, eine alternative Graphik ist mit Box-Plots und Bar-Plots möglich:



Gruppierte Darstellungen (hier mit anderen Daten, ohne Referenz) sind ebenfalls möglich: ▲



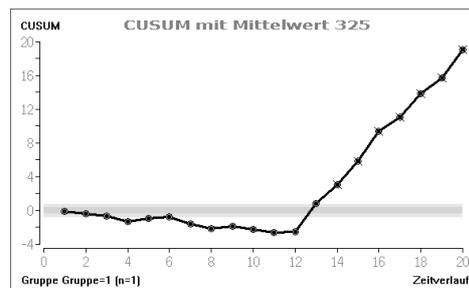
Cusum-Analyse für dichotome Daten: Welche Fehlerraten sind während eines Trainings akzeptabel bzw. unakzeptabel? Zur Verifikation wurden Daten und ein durchgerechnetes Beispiel aus Sivaprakasam und Purva (2010) verwendet. Die Wahrscheinlichkeiten α und β für die fälschliche Ablehnung bzw. Annahme der Nullhypothese können im Eingabedialog gewählt werden. Die grauen Balken bedeuten die Ablehnungsbereiche. Im Eingabedialog sind neben den beiden Irrtumswahrscheinlichkeiten die akzeptable und die nicht-akzeptable Fehlerrate anzugeben. Die verwendeten Daten stehen in der Datei ..\Valide\CusumDicho.DAT:



J. Sivaprakasam, M. Purva (2010) CUSUM analysis to assess competence: what rate is acceptable? Clin. Teach. 7, pp. 257-61

Cusum-Analyse für quantitative Daten: Bei Prins (Editor, 2012) findet sich ein Beispiel zur Kontrolle eines Prozesses. Zur Verifikation wurde dieses Beispiel mit den angegebenen Daten verwendet.

In der Abbildung sind die „out of control“-Werte durch ein „x“ markiert. Die Wahrscheinlichkeiten α und β für die fälschliche Ablehnung bzw. Annahme der Nullhypothese können im Eingabedialog gewählt werden. Die textliche Programmausgabe umfasst eine Tabelle u.a. der Cusum-Werte, wird aber hier nicht angegeben. Die verwendeten Daten stehen in der Datei ..\Valide\CusumQuant.DAT:



J. Prins (editor, 2012) e-Handbook of Statistical Methods. NIST/SEMATECH
<http://www.itl.nist.gov/div898/handbook/pmc/section3/pmc323.htm>

A.14 Faktorenanalyse



Zentroid- und Varimax-Methode
Explorative Faktorenanalyse
Konfirmatorische Faktorenanalyse
Spearman-Korrelationsmatrix
Pearson-Korrelationsmatrix

Explorative Faktorenanalyse: Berechnung einer Faktorenanalyse mittels 6 Maßen an Zylindern: Da die 6 angegebenen geometrischen Größen nicht-linear verknüpft sind, ergeben sich andere als die möglicherweise erwarteten Faktoren Durchmesser und Länge (Bortz (1977, p.638)). (Die Validierung der Methode wurde anhand der Korrelationsmatrix aus Haf/Cheaib p. 90 vorgenommen; da dort die Originaldaten nicht vorliegen, wird hier auf Bortz (1977) zurückgegriffen.)

Faktorenanalyse: Zentroid- und VariMax-Methode

Eingabedatei ist Bortz638.DAT
davon Variablen 1 (Durchmes), 2 (Länge), 3 (Grundfl.), 4 (Mantelf.),
5 (Volumen), 6 (Diagonal)

Pearson-Korrelationsmatrix:

1. Zeile: Korrelationskoeffizient r
2. Zeile: Überschreitungswahrsch. p

	Durchme	Länge	Grundfl	Mantelf	Volumen	Diagona
Durchmes :	1.0000	0.0000	0.9897	0.8115	0.8953	0.5559
	0.0000	1.0000	0.0000	0.0000	0.0000	0.0026
Länge :	0.0000	1.0000	0.0000	0.5410	0.3482	0.8230
	1.0000	0.0000	1.0000	0.0036	0.0751	0.0000
Grundfl. :	0.9897	0.0000	1.0000	0.8032	0.9045	0.5579
	0.0000	1.0000	0.0000	0.0000	0.0000	0.0025
Mantelf. :	0.8115	0.5410	0.8032	1.0000	0.9687	0.8741
	0.0000	0.0036	0.0000	0.0000	0.0000	0.0000
Volumen :	0.8953	0.3482	0.9045	0.9687	1.0000	0.7665
	0.0000	0.0751	0.0000	0.0000	0.0000	0.0000
Diagonal :	0.5559	0.8230	0.5579	0.8741	0.7665	1.0000
	0.0026	0.0000	0.0025	0.0000	0.0000	0.0000

Unrotierte Faktoren (Zentroid-Verfahren):

Durchmes : 0.8435 0.5553
Länge : 0.5041 -0.7163
Grundfl. : 0.8441 0.5618
Mantelf. : 0.9877 -0.0478
Volumen : 0.9647 0.2029
Diagonal : 0.8851 -0.4767

Kommunalitäten:

Durchmes : 1.0000 *)
Länge : 0.7672
Grundfl. : 1.0000 *)
Mantelf. : 0.9777
Volumen : 0.9719
Diagonal : 1.0000 *)

*) >1 wegen Kommunalitätenschätzung per Spaltennorm, cf. Haf/Cheaib.

Rotierte Faktoren (Varimax-Rotation):

Durchmes : 1.0074 -0.0710
Länge : -0.0345 -0.8752
Grundfl. : 1.0118 -0.0662
Mantelf. : 0.7557 -0.6377
Volumen : 0.8897 -0.4246
Diagonal : 0.4137 -0.9163

Varianzanteile (%):

	absol.	kumul.
Faktor 1:	59.56	59.56
Faktor 2:	36.70	96.26

Residuenmatrix nach 2 extrahierten Faktoren:

	Durchme	Länge	Grundfl	Mantelf	Volumen	Diagona
Durchmes :	0.1169	0.0275	-0.0342	-0.0050	-0.0312	-0.0740
Länge :	0.0275	-0.0876	0.0231	0.0089	-0.0072	0.0353
Grundfl. :	-0.0342	0.0231	0.1099	0.0036	-0.0238	-0.0786
Mantelf. :	-0.0050	0.0089	0.0036	0.0409	-0.0255	-0.0229
Volumen :	-0.0312	-0.0072	-0.0238	-0.0255	0.0970	-0.0094
Diagonal :	-0.0740	0.0353	-0.0786	-0.0229	-0.0094	0.1495

Literatur:

J. Bortz (1989) Statistik, Springer-Verlag.
A.L. Comry (1973) A First Course in Factor Analysis, Acad.Press N.Y.
C. Haf, T. Cheaib (1985) Multivariate Statistik, Vieweg-Verlag.
L.Thurstone (1947) Multiple Factor Analysis, Univ. of Chicago
K. Überla (1971) Faktorenanalyse, Springer-Verlag.
E. Weber (1974) Einführung in die Faktorenanalyse, Gustav Fischer Stuttgart.

Konfirmatorische Faktorenanalyse: Zu den 6 geometrischen Zylinder-Maßen von Bortz (1977, p.638) gestattet die konfirmatorische Faktorenanalyse die Untersuchung eines hypothetisierten Ladungsmusters. (Die Validierung der Methode wurde anhand der Korrelationsmatrix aus Haf & Cheaib p. 102 vorgenommen; da dort die Originaldaten nicht vorliegen, wird hier auf Bortz (1977) zurückgegriffen.)

Konfirmatorische Faktorenanalyse

Eingabedatei ist Bortz638.DAT

davon Variablen 1 (Durchmes), 2 (Länge), 3 (Grundfl.),
4 (Mantelf.), 5 (Volumen), 6 (Diagonal)

Pearson-Korrelationsmatrix:

1. Zeile: Korrelationskoeffizient r
2. Zeile: Überschreitungswahrsch. P

	Durchme	Länge	Grundfl	Mantelf	Volumen	Diagona
Durchmes :	1.0000	0.0000	0.9897	0.8115	0.8953	0.5559
	0.0000	1.0000	0.0000	0.0000	0.0000	0.0026
Länge :	0.0000	1.0000	0.0000	0.5410	0.3482	0.8230
	1.0000	0.0000	1.0000	0.0036	0.0751	0.0000
Grundfl. :	0.9897	0.0000	1.0000	0.8032	0.9045	0.5579
	0.0000	1.0000	0.0000	0.0000	0.0000	0.0025
Mantelf. :	0.8115	0.5410	0.8032	1.0000	0.9687	0.8741
	0.0000	0.0036	0.0000	0.0000	0.0000	0.0000
Volumen :	0.8953	0.3482	0.9045	0.9687	1.0000	0.7665
	0.0000	0.0751	0.0000	0.0000	0.0000	0.0000
Diagonal :	0.5559	0.8230	0.5579	0.8741	0.7665	1.0000
	0.0026	0.0000	0.0025	0.0000	0.0000	0.0000

Hypothetisiertes Ladungsmuster:

Durchmes : 1.0000 0.0000
Länge : 0.0000 1.0000
Grundfl. : 0.0000 0.0000
Mantelf. : 0.0000 0.0000
Volumen : 0.0000 0.0000
Diagonal : 0.0000 0.0000

Kommunalitäten:

Durchmes : 1.0000
Länge : 1.0000
Grundfl. : 0.9796
Mantelf. : 0.9512
Volumen : 0.9227
Diagonal : 0.9863

Rotierte Faktoren:

Durchmes : 1.0000 0.0000
Länge : 0.0000 1.0000
Grundfl. : 0.9897 0.0000
Mantelf. : 0.8115 0.5410
Volumen : 0.8953 0.3482
Diagonal : 0.5559 0.8230

Varianzanteile (%):

	absol.	kumul.
Faktor 1:	62.48	62.48
Faktor 2:	34.85	97.33

Residuenmatrix nach 2 extrahierten Faktoren:

	Durchme	Länge	Grundfl	Mantelf	Volumen	Diagona
Durchmes :	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Länge :	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Grundfl. :	0.0000	0.0000	0.0204	-0.0000	0.0185	0.0077
Mantelf. :	0.0000	0.0000	-0.0000	0.0488	0.0538	-0.0223
Volumen :	0.0000	0.0000	0.0185	0.0538	0.0773	-0.0176
Diagonal :	0.0000	0.0000	0.0077	-0.0223	-0.0176	0.0137

Literatur: J. Bortz (1989) Statistik, Springer-Verlag.

A.L. Comrey (1973) A First Course in Factor Analysis, Acad. Press NY

C. Haf, T. Cheaib (1985) Multivariate Statistik, Vieweg-Verlag.

L.L. Thurstone (1947) Multiple Factor Analysis, Univ. Chicago Press.

K. Überla (1971) Faktorenanalyse, Springer-Verlag.

E. Weber (1974) Einführung in die Faktorenanalyse, Fischer Stuttgart.

A.15 Reliabilitätsanalyse

Intraclass-Korrelation

Measurement Error & Repeatability

Cronbach's alpha

Faktorenanalyse (A.14)

Cohen's Kappa-Koeffizient: 2 Rater, dichotom (A.3)

Cohen's Kappa-Koeffizient: 2 Rater, mehrstufig (A.4)

Cohen's Kappa-Koeffizient: k Rater, mehrstufig (A.4)

Pearson's Kontingenzkoeffizient: dichotom (A.3)

Pearson's Kontingenzkoeffizient: mehrstufig (A.4)

Intraclass-Korrelation: Vergleich zweier FEV1-Messungen bzw. Messwiederholungen zu den Zeitpunkten T1 und T2: Die Validierung erfolgt mit Hilfe der Daten und der Resultate aus Bland und Altman (1996), die Überprüfung der Konfidenzintervalle durch per-Hand-Berechnungen.

Intraclass-Korrelation für Messwiederholungen

Eingabedatei ist ...\\Valide\\ICC.DAT, davon Variablen 2 (T1), 3 (T2)

i	Var. 2	Var. 3
1	1.2000	1.2400
2	1.2100	1.1900
3	1.4200	1.8300
4	1.4300	1.3800
5	1.4900	1.6000
6	1.5800	1.3600
7	1.5800	1.6500
8	1.5900	1.6000
9	1.6000	1.5800
10	1.6100	1.6100
11	1.6200	1.6800
12	1.6400	1.6100
13	1.6500	2.0500
14	1.7400	1.8000
15	1.7800	1.7600
16	1.8000	1.7600
17	1.8000	1.8200
18	1.8500	1.7300
19	1.8800	1.8200
20	1.9200	2.0000

Tabelle der Varianzanalyse:

Source	SQ	MQ	df	F	p
zwischen	1.5298	0.0805	19	7.4574	0.000028
innerhalb	0.2167	0.0108	20		
- Items	0.0116	0.0116	1	1.0707	0.313782
- Rest	0.2051	0.0108	19		
Gesamt	1.7465	0.0448	39		

>> Eine gute Reliabilität bedeutet einen signifikanten Unterschied bezüglich
>> "zwischen" und einen nicht-signifikanten Unterschied bezüglich "Items".

>> Beurteilung von wiederholten Messungen nach Bland und Altman (1996), der
 >> ICR ist definiert durch $ICR = (\#Items * SQ_{zwi} - SQ_{ges}) / ((\#Items - 1) * SQ_{ges})$:

Intraclass-Korrelationskoeffizient ICR = 0.751848 (p=0.000000 zweiseitig)

Approximatives 90%-Konfidenzintervall für ICR: [0.591956 ; 0.911740]

Approximatives 95%-Konfidenzintervall für ICR: [0.561325 ; 0.942371]

Approximatives 99%-Konfidenzintervall für ICR: [0.501458 ; 1.000000]

Mittlerer Messfehler ("Measurement error"): SD = Wurzel(MQ_{in}) = 0.1041

Mittlere Differenz zwischen zwei Messungen ("Repeatability") = 0.1472

>> Beurteilung des ICC gemäß Landis und Koch: 0-0.20: slight / 0.21-0.40: fair

>> 0.41-0.60: moderate / 0.61-0.80: substantial / 0.81-1.00: (almost) perfect.

Literatur:

JM Bland, DG Altman (1996) Measurement error and correlation coefficients.

British Medical Journal 313, pp. 41-2.

JM Bland, DG Altman (1996) Measurement error. Brit. Med. J. 313, 744-6.

R Osborne, WSB Paterson (1952) On the sampling variance of heritability estimates derived from variance analyses. Proc.R.Soc.Edinb. B, 64:456-61.

Intraclass-Korrelation: Vergleich von k=4 Ratern, die n=6 Personen beurteilen: Die Validierung erfolgt mit Hilfe der Daten aus Shrout und Fleiss (1979), die Überprüfung der Konfidenzintervalle durch Shrout und Fleiss (1979), per-Hand-Berechnungen und mit den Programmen SPSS und MedCalc. BIAS. rechnet bei ICC(2,1)-Konfidenzintervallen grundsätzlich mit ganzzahlig gerundeten Freiheitsgraden, geringfügige Abweichungen von Ergebnissen aus SPSS sind deshalb möglich.

Intraclass-Korrelation zum Ratervergleich ▲

Eingabedatei ist ...\\Valide\\ShroutFleiss.DAT, davon Variablen R1, R2, R3, R4

i	R1	R2	R3	R4
1	9.0000	2.0000	5.0000	8.0000
2	6.0000	1.0000	3.0000	2.0000
3	8.0000	4.0000	6.0000	8.0000
4	7.0000	1.0000	2.0000	6.0000
5	10.0000	5.0000	6.0000	9.0000
6	6.0000	2.0000	4.0000	7.0000

Tabelle der Varianzanalyse:

Source	SQ	MQ	df	F	p
zwischen	56.2083	11.2417	5	11.0272	0.000135
innerhalb	112.7500	6.2639	18	1.7947	0.164769 *
- Rater	97.4583	32.4861	3	31.8665	0.000001
- Rest	15.2917	1.0194	15		
Gesamt	168.9583	7.3460	23		

*) Zu ICC(1,1) und ICC(1,k): $F = MQ_{zwischen} / MQ_{innerhalb}$ mit $df=(5,18)$

>> Eine gute Reliabilität bedeutet einen signifikanten Unterschied bezüglich

>> "zwischen" und einen nicht-signifikanten Unterschied bezüglich "Rater".

Alle Intraclass-Korrelationskoeffizienten gemäß Shrout und Fleiss (1979):

Subjekte werden von verschiedenen, zufällig ausgewählten Ratern beurteilt:

ICC(1,1) = 0.165742 (p=0.164769) -- Modell "one-way random, single measure"

ICC(1,k) = 0.442797 (p=0.164769) -- Modell "one-way random, average measure"

Subjekte werden von den gleichen, aber zufällig ausgewählten Ratern beurteilt:
ICC(2,1) = 0.289764 (p=0.000135) -- Modell "two-way random, single measure"
ICC(2,k) = 0.620051 (p=0.000135) -- Modell "two-way random, average measure"

Subjekte werden von den gleichen, jedoch fest vorgegebenen Ratern beurteilt:
ICC(3,1) = 0.714841 (p=0.000135) -- Modell "two-way mixed, single measure"
ICC(3,k) = 0.909316 (p=0.000135) -- Modell "two-way mixed, average measure"

>> ICC(1,.), ICC(2,.): Reliabilität i.S. von "absolute rater agreement"
>> ICC(3,.): Reliabilität i.S. von "rater consistency" (justiertes Modell)
>> ICC(.,1): Reliabilität bzgl. des Ratings eines einzelnen, indiv. Raters
>> ICC(.,k): Reliabilität bzgl. des Rating-Mittelwerts von k versch. Ratern

>> Beurteilung der ICCs gemäß Landis & Koch: 0-0.20: slight / 0.21-0.40: fair
>> 0.41-0.60: moderate / 0.61-0.80: substantial / 0.81-1.00: (almost) perfect
>> Negative ICCs indizieren eine Reliabilität von 0: Wirtz und Caspar (2002).

95%-Konfidenzintervalle für die Koeffizienten ICC(1,1), ICC(2,1) und ICC(3,1):

95%-Konfidenzintervall für ICC(1,1) = 0.165742: [-0.132932 ; 0.722560]
95%-Konfidenzintervall für ICC(2,1) = 0.289764: [0.021618 ; 0.759943]
95%-Konfidenzintervall für ICC(3,1) = 0.714841: [0.342465 ; 0.945858]

>> *Zusätzlich im Programm: Multiple Vergleiche der Rater und deren Korrelationen*

Literatur:

PE Shrout, JL Fleiss (1979) Intraclass Correlations: Uses in Assessing Rater Reliability. Psychological Bulletin 86, 2, pp. 420-428.
JL Fleiss, J Cohen (1973) The equivalence of weighted kappa and the intra-class correlation coefficient as measures of reliability. Educational and Psychological Measurement 33, pp. 613-619
JR Landis, GG Koch (1977) The measurement of observer agreement for categorical data. Biometrics Vol. 33(1), pp. 159-174
M Wirtz, F Caspar (2002). Beurteilerübereinstimmung und Beurteilerreliabilität. Göttingen: Hogrefe.

Cohen's Kappa: Bitte vergleichen Sie dazu die beiden Abschnitte [A.3](#) ("Vierfeldertafeln") und [A.4](#) ("Kontingenztafeln"). Kappa ist – auch als "weighted kappa" - im Programm für ≥ 2 Rater und ≥ 2 Merkmalsstufen verfügbar. Hier ein Beispiel für 2 Rater und 2 Stufen (wie in Abschnitt A3):

Cohen's Kappa-Koeffizient

Bezeichnung: Bortz-Lienert-Boehnke p.451

Eingegebene Tafel:

9	3		12
4	4		8
-----+-----			
13	7		20

Cohen's Kappa = 0.2553
Prüfgröße Z = 1.1483 (p = 0.250829)

Literatur:

WG Cohen (1960) Errors of measurements. Technometrics 10, pp. 637-66
JL Fleiss (1973) Stat. Meth. for Rates & Propor. Wiley NY, pp. 143-47

Eine allgemeinere Variante von Cohen's Kappa findet sich im Modul "Kontingenztafeln" für den Fall von $k > 2$ Ratern und/oder $s > 2$ Stufen des Merkmals, wobei auch eine Dateneingabe in der üblichen Form über eine Eingabedatei möglich ist: Einzelheiten dazu finden Sie in [Anhang 4](#) des Handbuchs.

Cronbach's alpha: Ein Fragebogen soll die Patientenzufriedenheit messen. Der Fragebogen umfasst 4 Items und es soll die "interne Validität" geprüft werden, d.h., ob die 4 Fragen/Items prinzipiell das gleiche Konstrukt beurteilen. Die Validierung erfolgt mit Hilfe von SPSS, die der Konfidenzintervalle mit einem Taschenrechner.

Cronbach's alpha ▲

Eingabedatei ist ICC4.DAT, davon Var. 1(Item1), 2(Item2), 3(Item3), 4(Item4)

i	Var. 1	Var. 2	Var. 3	Var. 4
1	21.0000	20.0000	21.5000	23.0000
2	21.0000	21.5000	24.0000	25.5000
3	20.5000	24.0000	24.5000	26.0000
4	23.5000	24.5000	25.0000	26.5000
5	21.5000	23.0000	22.5000	23.5000
6	20.0000	21.0000	21.0000	22.5000
7	21.5000	22.5000	23.0000	25.0000
8	23.0000	23.0000	23.5000	24.0000
9	20.0000	21.0000	22.0000	21.5000
10	16.5000	19.0000	19.0000	19.5000
11	24.5000	25.0000	28.0000	28.0000
12	26.0000	25.0000	29.0000	31.0000
13	21.5000	22.5000	23.0000	26.5000
14	23.0000	22.5000	24.0000	27.5000
15	25.5000	27.5000	26.5000	27.0000
16	20.0000	23.5000	22.5000	26.0000
17	24.5000	25.5000	27.0000	28.5000
18	22.0000	22.0000	24.5000	26.5000
19	24.0000	21.5000	24.5000	25.5000
20	23.0000	20.5000	31.0000	26.0000
21	27.5000	28.0000	31.0000	31.5000
22	23.0000	23.0000	23.5000	25.0000
23	21.5000	23.5000	24.0000	28.0000
24	17.0000	24.5000	26.0000	29.5000
25	22.5000	25.5000	25.5000	26.0000
26	23.0000	24.5000	26.0000	30.0000
27	22.0000	21.5000	23.5000	25.0000

Tabelle der Varianzanalyse:

Source	SQ	MQ	df	F	p
zwischen	518.3796	19.9377	26	9.5925	0.000000
innerhalb	399.3125	4.9298	81		
- Items	237.1921	79.0640	3	38.0396	0.000000
- Rest	162.1204	2.0785	78		
Gesamt	917.6921	8.5766	107		

Mittlere ("Average") Inter-Item-Korrelation = 0.682350

Cronbach's alpha = 0.895752 (p=0.000000 zweiseitig)

Asymptotisches 90%-Konfidenzintervall für alpha: [0.825182;0.937834]

Asymptotisches 95%-Konfidenzintervall für alpha: [0.806990;0.943694]

Asymptotisches 99%-Konfidenzintervall für alpha: [0.765763;0.953604]

Literatur:

JM Bland, DG Altman (1997) Statistics Notes: Cronbach's alpha. British Medical Journal 314, pp. 572-3.

AJ Kroning, PH Franses (2003) Confidence intervals for comparing Cronbach's coefficient alpha values. Preprint Elsevier Science. ▲



Percentilen und Quartilen

Univariater parametrischer Toleranzbereich

Univariater nicht-parametrischer Toleranzbereich

Univariater Box-Cox-transformierter Toleranzbereich

Bivariater parametrischer Toleranzbereich

Bivariater nicht-parametrischer Toleranzbereich

Bivariater Box-Cox-transformierter Toleranzbereich

Qualitätssicherung im Labor: RiLi-BÄK

Percentilen und Quartilen: BiAS. berechnet Percentilen und Quartilen mit Hilfe von Tukey's "Statistisch äquivalenten Blöcken", also in gleicher Weise wie nicht-parametrische univariate Toleranzbereiche. Zur Überprüfung kann das entsprechende Modul der nicht-parametrischen Toleranzbereiche herangezogen werden ("Univariater nicht-parametrischer Normbereich"). Das gleiche Berechnungsverfahren wird in vielen Standardprogrammen wie SPSS, SAS, BMDP etc. verwendet.

Percentilen



Eingabedatei ist \Graf379.DAT, davon Variable 1 (Asche)

1:	1.620000	2:	1.420000
3:	1.580000	4:	1.390000
5:	1.540000	6:	1.350000
7:	1.500000	8:	1.480000
9:	1.580000	10:	1.600000

Median	=	1.5200			
Minimum	=	1.3500	Maximum	=	1.6200
1. Quartil	=	1.4125	3. Quartil	=	1.5850

95.00%-Percentile nicht berechenbar: n ist zu klein.

Literatur:

G.A. Lienert (1973) Verteilungsfreie Methoden in der Biostatistik.

Hain Meisenheim am Glan p. 40ff.

H. Ackermann (1983) Sind $X_{\text{quer}} \pm 2s$ - Bereiche nützliche diagnostische Hilfsmittel? Med. Welt 34, 7, pp. 212-215.

Univariater parametrischer Toleranzbereich: Exemplarische Berechnung eines Toleranzbereiches anhand der Daten von Graf et al. (1987, p.379):

Parametrischer Toleranzbereich

Eingabedatei ist Graf379.DAT, davon Variable 1 (Asche)

Daten wie oben!

$X_{\text{quer}}=1.5060$

Standardabw. $S = 0.0942$

Fallzahl $n = 10$

Prüfung auf Gauß-Verteilung:

Kolmogoroff-Smirnoff' $\delta = 0.184$

Überschreitungsws. $P > 0.2000$

Toleranzbereich:
Untere Grenze: 1.2825
Obere Grenze: 1.7295

Literatur:

F. Proschan (1953) Confidence and Tolerance Intervals for the Normal Distribution. J. Amer. Stat. Ass. 48, pp. 550-564.
H. Ackermann (1983) Sind $x \pm 2s$ - Bereiche nützliche diagnostische Hilfsmittel? Med. Welt 34, 7, pp. 212-215.

Univariater nicht-parametrischer Toleranzbereich: Exemplarische Berechnung eines Toleranzbereiches anhand der Daten von Graf et al. (1987, p.379):

Nicht-parametrische Toleranzbereiche

Eingabedatei ist Graf379.DAT, davon Variable 1

1:	1.620000	2:	1.420000
3:	1.580000	4:	1.390000
5:	1.540000	6:	1.350000
7:	1.500000	8:	1.480000
9:	1.580000	10:	1.600000

Minimum = 1.3500
Streuung s^2 = 0.0089
n = 10

t(P,df) = 1.8331
df = n-1 = 9

Überdeckung P = 0.9000

Untere Grenze = 1.3249
Obere Grenze = 1.6871

Maximum = 1.6200
Median = 1.5200
n = 10

Überdeckung P = 0.9000

P' = 81.82%:
Untere Grenze = 1.3500
Obere Grenze = 1.6200

Literatur:

S.S. Wilks (1941) Determination of Sample Sizes for Setting Tolerance Limits. Ann. Math. Stat. 12, pp. 91-96.
H. Ackermann (1983) Sind $x \pm 2s$ - Bereiche nützliche diagnostische Hilfsmittel? Med. Welt 34, 7, pp. 212-215.

Univariater Box-Cox-transformierter Toleranzbereich: Die Box-Cox-Transformation transformiert nichtnormalverteilte Methoden anwendbar zu machen. Eine Validierung kann deshalb prinzipiell mit parametrischen Methoden erfolgen. Da in der Literatur keine entsprechenden Beispiele gefunden wurden, erfolgte die Überprüfung durch per-Hand-Berechnungen, Sonderfälle etc.. Eine Überprüfung der Box-Cox-Transformation wurde mit Hilfe von anderen Programmmoduln vorgenommen (z.B. Kolmogoroff-Lilliefors-Test), so dass sich eine explizite Überprüfung erübrigt.

Box-Cox-transformierter Toleranzbereich

Eingabedatei ist Graf379.DAT, davon Variable 1 (Asche)

1:	1.620000	2:	1.420000
3:	1.580000	4:	1.390000
5:	1.540000	6:	1.350000
7:	1.500000	8:	1.480000
9:	1.580000	10:	1.600000

Xquer, s und s^2 beziehen sich auf die transformierten Daten!

Xquer = 0.4533
Standardabw. s = 0.0773
Streuung s^2 = 0.0060
n = 10

t(P,df) = 2.2622
df = n-1 = 9
Überdeckung P = 0.9500

Prüfung auf Gauß-Verteilung der Box-Cox-transformierten Daten:
Box-Cox' Transformationskoeffizient tau = 0.50
Kolmogoroff-Smirnoff's delta = 0.184 (p>0.200000)

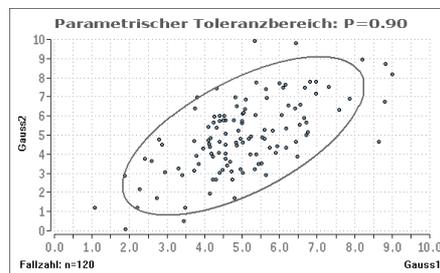
Die Transformation ist nur erfolgreich, wenn delta nicht signifikant ist!

Toleranzbereich (retransformiert):
Untere Grenze = 1.2882
Obere Grenze = 1.7379

Literatur:

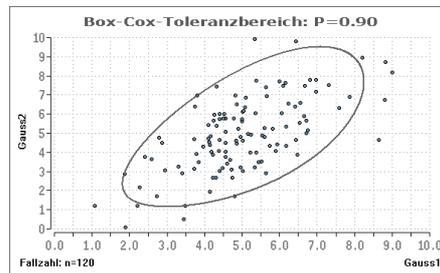
F. Proschan (1953) Confidence and Tolerance Intervals for the Normal Distribution. J. Amer. Stat. Ass. 48, pp. 550-564.
H. Ackermann (1983) Sind Xquer \pm 2s - Bereiche nützliche diagnostische Hilfsmittel? Med. Welt 34, 7, pp. 212-215.
R.A. Rode, V.M. Chinchilli (1988) The Use of Box-Cox-Transformations in the Development of Multivariate Tolerance Regions With Applikationen to Clinical Chemistry. Amer. Statistician 42, 1, pp. 23-30.

Bivariater parametrischer Toleranzbereich: Darstellung mit Gauß-verteiltern Zufallszahlen (Modul Randomisierung/Zufallszahlen, Datei ..\Valide\BivGauss.DAT). Bitte vergleichen Sie dazu auch den bivariaten nicht-parametrischen und Box-Cox-transformierten Toleranzbereich, für die die gleichen Daten verwendet wurden.



Im bivariaten Fall treten bezüglich der Verteilungsformen prinzipiell die gleichen Probleme auf wie im Univariaten. Die Ausgabe des Programms besteht ausschließlich aus Graphiken, die u.a. durch Spezialfälle und Handgerechnetes überprüft wurden. Die Verteilungsformen werden wie im Univariaten mit Hilfe von anderen Programmmoduln überprüft (z.B. Kolmogoroff-Lilliefors-Test), im Bivariaten mit dem Mudholkar-Test (p=0.6213, Marginalverteilungen mit p1=0.0647 und p2=0.6337).

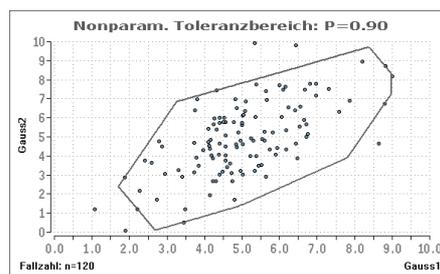
Bivariater Box-Cox-transformierter Toleranzbereich: Die Darstellung erfolgt mit den gleichen Gauß-verteiltern Daten (..\Valide\BivGauss.DAT) wie zum parametrischen Bereich. Bitte vergleichen Sie dazu auch die bivariaten nicht-parametrischen und parametrischen Toleranzbereiche.



Im bivariaten Fall treten bezüglich der Verteilungsformen prinzipiell die gleichen Probleme auf wie im Univariaten. Die Ausgabe des Programms besteht ausschließlich aus Graphiken, die u.a. durch Spezialfälle und Handgerechnetes überprüft wurden. Die Box-Cox-Transformation wurde wie im Univariaten mit Hilfe von anderen Programmmoduln überprüft (z.B. Kolmogoroff-Lilliefors-Test).

Der dargestellte Box-Cox-transformierte Bereich ist mit dem parametrischen Bereich fast (!!) identisch, da die Daten recht gut bivariat Gauß-verteilt sind (p-Werte vgl. parametrischen Bereich oben).

Bivariater nicht-parametrischer Toleranzbereich: Die Darstellung erfolgt mit den gleichen bivariat Gauß-verteiltern Zufallszahlen wie oben. Bitte vergleichen Sie das Resultat auch mit den beiden bivariaten parametrischen und Box-Cox-transformierten Toleranzbereichen.



Diese Normbereiche wurden vom Autor entwickelt, so dass zur Validierung nur auf Ackermann (1985) verwiesen werden kann. Das Buch stellt neben der bivariaten auch Methoden zur multivariaten Situation zur Verfügung. Die Ausgabe des Moduls besteht wie im parametrischen Fall ausschließlich aus Graphiken, die via BMP/GIF zur weiteren Verarbeitung exportiert oder ausgedruckt werden können.

Im univariaten Fall nicht-parametrischer Normbereiche können gegebenenfalls Interpolationen vorgenommen werden, falls bei gewünschtem Anteil π die Anzahl der zu "eliminierender Blöcke" $k=(n+1) \cdot (1-\pi)$ nicht ganzzahlig und/oder nicht geradzahlig resultiert. Bei bivariaten Normbereichen mit π -Erwartung kann jedoch keine Interpolation vorgenommen werden, so dass die Anzahl $= (n+1) \cdot (1-\pi)$ idealerweise durch 4 teilbar sein muss, da in allen anderen Fällen nicht alle n Wertepaare in die Berechnung eingehen können. Das Programm gibt dazu Hinweise.

Qualitätssicherung im Labor gemäß RiLi-BÄK: Berechnung der Beurteilungsparameter QUAMM, VK, SD, weiterhin systematische und zufällige Messabweichungen zur Beurteilung der Richtigkeit und Präzision von Laboranalysen. Zugrundegelegt ist die Neufassung der "RiLi-BÄK" vom 15.2.2008, publiziert im Deutschen Ärzteblatt Jahrgang 195, Heft 7. Zur Validierung wurden Daten aus der Sysmex-Publikation 2007 verwendet (p. 14, unten Literatur; Datei in „..\Valide“), die Ergebnisse wurden anhand der Sysmex-Daten, anderer Module und per-Hand-Berechnungen überprüft:

Qualitätssicherung im Labor gemäß RiLiBÄK

Eingabedatei ist ..\Valide\Systemex.DAT, davon Variable 1 (Geraet1)

Eingelesene Daten:

1:4.54 2:4.56 3:4.52 4:4.51 5:4.53 6:4.51 7:4.58 8:4.54 9:4.53
10:4.50 11:4.54 12:4.48 13:4.56 14:4.52 15:4.53 16:4.54 17:4.52 18:4.51
19:4.58 20:4.51 21:4.53 22:4.54 23:4.49 24:4.56 25:4.50

Deskriptive nicht-parametrische Parameter:

Stichprobenumfang:	n =	25.0000
Minimum:	Min =	4.4800
Maximum:	Max =	4.5800
Median:	Median =	4.5300
1.Quartile:	Q1 =	4.5100
3.Quartile:	Q3 =	4.5400

Kolmogoroff-Smirnoff-Lilliefors-Test auf Gauß-Verteilung:

Kolmogoroff-Smirnoff's delta = 0.138 (p>0.100000): Gauß-Modell Ok.
Der p-Wert des Tests wurde korrigiert nach Dallal und Wilkinson.

Einseitige Grubbs-Tests zur Prüfung der Extremwerte als Ausreißer:

Minimum:	4.4800	p = 0.616290	> Sicher kein Ausreißer.
Maximum:	4.5800	p = 0.521810	> Sicher kein Ausreißer.

Trendtest nach Neumann et al.:

Prüfgröße = 2.8529: p > 0.050 (einseitig exakt) > kein Trend.

Parameter zur Qualitätssicherung gemäß der BÄK-Richtlinien:

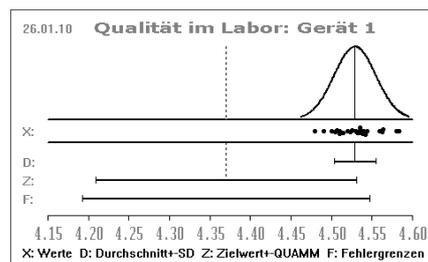
Zielwert nach Benutzervorgabe:	Xo =	4.3700
Arithmetischer Mittelwert (Durchschnitt):	Xquer =	4.5292
Präzision: Zufällige Messabweichung s=StdDev:	SD =	0.0258
Variationskoeffizient VK=CV=SD/Xquer [%]:	VK[%] =	0.5698
Richtigkeit: Systematische Messabweichung:	Delta =	0.1592
Richtigkeit: Relatives Delta=Delta/Xo [%]:	Delta[%] =	3.6430
Quadratischer Mittelwert der Messabweichungen: QUAMM =	QUAMM =	0.1612
Relativer QUAMM=QUAMM/Zielwert [%]:	QUAMM[%] =	3.6887

Laborinterne Fehlergrenzen bezüglich einer Ermittlungsperiode (EP):

Zufällige Messabweichung:	SD(EP) =	0.0258
Systematische Messabweichung:	Delta(EP) =	0.1592
Erweiterungsfaktor für SD(EP):	k =	3.0000
Laborintern selbst ermittelter QUAMM:	QUAMM-Max =	0.1770
Laborinterne untere Fehlergrenze:	UFG =	4.1930
Laborinterne obere Fehlergrenze:	OFG =	4.5470

- Richtlinien der Bundesärztekammer zur Qualitätssicherung laboratoriumsmedizinischer Untersuchungen. Deutsches Ärzteblatt Jg. 105, Heft 7, 15.2.2008
- Die Neufassung der RiLi-BÄK. Publikation des Systemex Xtra 2/2007, pp. 1-16.

Im Programm stehen einige graphische Darstellungen zur Qualitätssicherung zur Verfügung, hier eine Graphik zum oben ausgeführten Beispiel der Systemex-Daten als Illustration der RiLiBÄK-Parameter:



A.17 Pharmakokinetische Parameter und Bioäquivalenz ▲

Berechnung von AUC, Cmax, Tmax und T1/2 etc. (auch Retard-Parameter)
 Berechnung von AUC etc. wie oben, im Cross-Over (auch Retard-Parameter)
 Inklusionsregel mit parametrischem Intervall für A/B, ln(A/B)
 Inklusionsregel mit Tukey-Intervall für ln(A/B)
 Inklusionsregel mit Moses-Intervall für A/B, ln(A/B)
 Mittlere effektive Konzentration EC50 und andere

Berechnung pharmakokinetischer Parameter: Exemplarische Berechnung der Parameter anhand eines Testdatensatzes (Bortz (1977, p.412)), Verifikation per Handberechnungen. Das Programm berechnet folgende Größen:

AUC_1n: Fläche unter der Konzentrations-Zeit-Kurve von t1 bis tn
 AUC_1-unendl: AUC_1n + AUC(tn-unendl) mit $AUC(tn-unendl) = C_n / \lambda$
 AUC_FI: Prozentuale AUC-Fluktuation
 AUMC_1-n Fläche unter der 1. Moment-Kurve von t1 bis tn, zur Berechnung von MRT
 AUMC_1-unendl: Fläche unter der 1. Moment-Kurve von t1-unendlich, zur Berechnung von MRT
 Lambda: Terminale Eliminationskonstante, zur Berechnung von t1/2
 t1/2: Per Lambda aus Log-linearer Regression mit terminalen Zeitpunkten
 Cmin: Konzentration am letzten Zeitpunkt
 Cmax: Maximale Konzentration
 Tmax: Zeitpunkt der maximalen Konzentration
 Cav: Mittlere Konzentration; $Cav = (AUC_1n) / (n\text{-ter Zeitpkt} - 1\text{-ter Zeitpkt})$
 PTF: Prozentuale peak-trough-Fluktuation; $100\% * (Cmax - Cmin) / Cav$
 PTS: Prozentualer Swing; $100\% * (Cmax - Cmin) / Cmin$
 HVD: Half Value Duration, Zeitspanne, in der $C > Cmax/2$
 MRT: Mean Residence Time = $MRT = AUMC_1-unendl / AUC_1-unendl$

Pharmakokinetische Parameter: Eine Stichprobe

Eingabedatei ist ..\Valide\Bortz412.DAT

davon Variablen 1 (morgens) t = 1.0000
 2 (mittags) t = 2.0000
 3 (abends) t = 3.0000

i	Var. 1	Var. 2	Var. 3
1	7.0000	7.0000	6.0000
2	5.0000	6.0000	8.0000
3	8.0000	9.0000	5.0000
4	6.0000	8.0000	6.0000
5	7.0000	7.0000	5.0000
6	7.0000	9.0000	7.0000
7	5.0000	10.0000	6.0000
8	6.0000	7.0000	4.0000
9	7.0000	8.0000	6.0000
10	5.0000	7.0000	5.0000

i	AUC_1n	AUC_1-	Cmin	Cmax	Tmax	T1/2
	Cav	PTF	PTS	AUC_FI	HVD	MRT
	Lambda	AUMC_1n	AUMC_1-			
1	13.5000	52.4230	6.0000	7.0000	1.0000	4.4966
	6.7500	14.81	16.67	4.17	2.0000	7.5495
	0.1542	26.5000	395.7683			

2	12.5000	n.b.	8.0000	8.0000	3.0000	n.b.
	6.2500	0.00	0.00	12.25	2.0000	n.b.
	n.b.	26.5000	n.b.			
3	15.5000	24.0065	5.0000	9.0000	2.0000	1.1792
	7.7500	51.61	80.00	12.20	2.0000	2.8947
	0.5878	29.5000	69.4915			
4	14.0000	34.8564	6.0000	8.0000	2.0000	2.4094
	7.0000	28.57	33.33	7.14	2.0000	4.6783
	0.2877	28.0000	163.0670			
5	13.0000	27.8601	5.0000	7.0000	1.0000	2.0600
	6.5000	30.77	40.00	8.65	2.0000	4.0827
	0.3365	25.0000	113.7445			
6	16.0000	43.8536	7.0000	9.0000	2.0000	2.7581
	8.0000	25.00	28.57	6.25	2.0000	5.1625
	0.2513	32.0000	226.3922			
7	15.5000	27.2457	6.0000	10.0000	2.0000	1.3569
	7.7500	51.61	66.67	14.70	2.0000	3.2934
	0.5108	31.5000	89.7306			
8	12.0000	19.1478	4.0000	7.0000	2.0000	1.2386
	6.0000	50.00	75.00	11.11	2.0000	2.9881
	0.5596	23.0000	57.2159			
9	14.5000	35.3564	6.0000	8.0000	2.0000	2.4094
	7.2500	27.59	33.33	5.82	2.0000	4.6262
	0.2877	28.5000	163.5670			
10	12.0000	26.8601	5.0000	7.0000	2.0000	2.0600
	6.0000	33.33	40.00	8.33	2.0000	4.1975
	0.3365	24.0000	112.7445			

AUC_{1n}: Fläche unter der Konzentrations-Zeit-Kurve von t₁-t_n
 AUC₁₋: AUC_{1n} + AUC(t_n, '-') mit AUC(t_n, '-')=C_n/Lambda
 AUC_{Fl}: Prozentuale AUC-Fluktuation; 100%*(AUC_{über_Cav}+AUC_{unter_Cav})/AUC₁₋
 AUMC_{1n}: Fläche unter der ersten Moment-Kurve
 AUMC₁₋: Fläche unter der ersten Moment-Kurve, zur Berechnung von MRT
 Lambda: Terminale Eliminationskonstante, zur Berechnung von t_{1/2}
 t_{1/2}: Per Lambda aus Log-lin. Regression mit terminalen 2 Zeitpunkten
 C_{min}: Konzentration am letzten Zeitpunkt
 C_{max}: Maximale Konzentration
 T_{max}: Zeitpunkt der maximalen Konzentration
 C_{av}: Mittlere Konzentration; C_{av}=(AUC_{1n})/(n-ter Zeitpkt- 1.Zeitpkt)
 PTF: Prozentuale peak-trough-Fluktuation; 100%*(C_{max}-C_{min})/C_{av}
 PTS: Prozentualer Swing; 100%*(C_{max}-C_{min})/C_{min}
 HVD: Half Value Duration, Zeitspanne, in der C > C_{max}/2
 MRT: Mean Residence Time = MRT = AUMC₁₋ / AUC₁₋

n.b.: Nicht berechenbar (diverse Parameter falls lambda nicht berechenbar).
 "-": Das Symbol bedeutet zB in "AUC₁₋" die Fläche von t₁ bis unendlich.

>> Es werden nur v o l l s t ä n d i g e Zeitverläufe ausgewertet!
 >> Bei lambda<=0 werden lambda und daraus errechnete Werte als n.b. bewertet.

E. Glaser (1985) Pharmakokinetik. pmi-Verlag Frankfurt.
 W.A. Ritschel (1986) Handbook of Basic Pharmacokinetics. Drug Intell. Publ.
 H. Derendorf, Th. Gramatté, H.G. Schäfer, A. Staab (2010) Pharmakokinetik kompakt. Wiss. Verlagsgesellschaft Stuttgart.

Berechnung pharmakokinetischer Parameter im Cross-Over: Zur Verifikation des Moduls wurde eine emplarische Berechnung der Parameter anhand eines Testdatensatzes (Bortz (1977, p.412)) mit Bleistift und Papier vorgenommen (vgl. auch die analoge Berechnung für *eine* Stichprobe). Da das Programm zur Berechnung der Cross-Over-Situation das gleiche Modul impliziert wie im Einstichproben-Fall, wurden die Daten des Einstichproben-Datensatzes mehrfach verwendet und damit auf die Cross-Over-Situation übertragen. Somit erübrigt sich die explizite Berechnung und Ausgabe der pharmakokinetischen Parameter im Cross-Over.

BiAS. berechnet im Cross-Over die gleichen pharmakokinetischen Parameter wie in der obigen Ausgabe für nur eine Gruppe. ▲

Inklusionsregel: Alle nachfolgend aufgeführten Methoden verwenden eine ähnliche Eingabe wie zur Cross-Over-Analyse. Nach Eingabe der Gruppenvariablen sind die beiden Nummern bzw. Namen der Variablen einzugeben, die den beiden Perioden zugeordnet sind: Die Eingabeaufforderung lautet: Welche Variablen werden getestet: Periode1: und nach Eingabe Periode 2:. Es handelt sich hierbei um Perioden, nicht um Behandlungen: Zum Beispiel erhielt Gruppe 1 in der ersten Periode die Behandlung A, die Gruppe 2 dagegen B, während in der 2. Periode umgekehrt Gruppe 1 Behandlung B und Gruppe 2 Behandlung A erhielt!

Die Bezeichnung "Westlake-Regel" wurde aus historischen Gründen beibehalten. Äquivalent dazu könnte auch von einer "Intervall-Inklusionsregel" gesprochen werden: Liegt das Konfidenzintervall für z.B. A/B gänzlich im Äquivalenzbereich, so wird die Nullhypothese beibehalten, liegt es gänzlich außerhalb des Äquivalenzbereichs, so wird die Nullhypothese abgelehnt. In allen anderen Fällen, also bei teilweiser Überschneidung des Konfidenzintervalls und des Äquivalenzbereiches, wird keine Entscheidung getroffen.

Inklusionsregel mit parametrischem Intervall für A/B, ln(A/B): Dies verlangt lediglich die Berechnung von entsprechenden Konfidenzintervallen, wozu Beispiele zur Validierung angegeben werden.

Zur technischen Abfolge vergleichen Sie bitte die "Inklusionsregel mit Tukey-Intervall für ln(A/B)".

Inklusionsregel mit Tukey-Intervall für ln(A/B): Hauschke et al. (1990) beschreiben eine Prüfung zweier Formulierungen A und B auf Äquivalenz im Cross-Over. Zur Prüfung der ln(A/B) werden Tukey-Intervalle verwendet; dabei sollten keine Übertragungseffekte vorhanden sein. Der Test ist verteilungsfrei!

Inklusionsregel mit Tukey-Intervall für ln(A/B) ▲

Eingabedatei ist Hauschkel.DAT
davon Variable 1 (Per_1), 2 (Per_2)
und Gruppen Var. 3 (GruppNum) = 1 und 2

Gruppe 1			Gruppe 2		
i	Var. 1	Var. 2	i	Var. 1	Var. 2
1	550.8500	330.7900	1	423.1800	424.8900
2	378.8300	329.9900	2	448.7600	425.7700
3	333.6500	348.3700	3	548.3300	404.8200
4	420.3200	408.4700	4	551.3900	387.7700
5	386.0500	249.8400	5	548.7100	474.6400
6	412.8200	328.3500	6	372.2100	260.3000

Das Konfidenzintervall und die Äquivalenzbereiche beziehen sich auf den Quotienten $A/B = (\text{Phase 1})/(\text{Phase 2})$ für die erste Gruppe und entsprechend $A/B = (\text{Phase 2})/(\text{Phase 1})$ für die zweite Gruppe. Die Intervalle werden somit auf der Basis von $n = n_1 + n_2 = 6 + 6 = 12$ Werten berechnet:

Hodges-Lehmann-Schätzer: $M = 99.46$
Konfidenz $P = 1 - 2\alpha$ des Konfidenzintervalls: $P = 0.9000$
Signifikanzniveau α für Äquivalenztest: $\alpha = 0.0500$
Konfidenzintervall: (84.83 , 120.02)
Äquivalenzbereich : (80.00 , 120.00) : Keine Entscheidung
Äquivalenzbereich : (70.00 , 142.86) : 'Bioäquivalenz'
Äquivalenzbereich : (80.00 , 125.00) : 'Bioäquivalenz'
Äquivalenzbereich : (90.00 , 111.11) : Keine Entscheidung

Literatur:

W.J. Westlake (1972) Use of Confidence Intervals in Analysis of Comparative Bioavailability Trials. J. Pharm. Sci. 61, pp. 1340-1341.
W. Lehmacher, W. van Eimeren (1986) Zur statistischen Bewertung der Ergebnisse von Bioverfügbarkeitsstudien. Therapiewoche 36, pp.413-420.
D. Hauschke et al. (1990) A distribution-free procedure for the stat. analysis of bioequiv. studies. Int.J.Clin.Pharm.Ther.Tox. 28,2,72-8.

Inklusionsregel mit Moses-Intervall für A/B, ln(A/B): Hauschke et al. (1990): Prüfung zweier Formulierungen A und B auf Äquivalenz im Cross-Over. Zur Prüfung der ln(A/B) werden Moses-Intervalle verwendet. Der Test ist wie der des letzten Beispiels verteilungsfrei!

Inklusionsregel mit Moses-Intervall für A/B, ln(A/B) ▲

Eingabedatei ist Hauschke2.DAT

davon Variable 1 (Periode1), 2 (Periode2)
und Gruppen Var. 3 (Gruppe) = 1 und 2

Gruppe 1			Gruppe 2		
i	Var. 1	Var. 2	i	Var. 1	Var. 2
1	228.0400	288.7900	1	339.0300	329.7600
2	288.2100	343.3700	2	242.6400	258.1900
3	217.9700	225.7700	3	249.9400	201.5600
4	133.1300	235.8900	4	184.3200	249.6400
5	213.7800	215.1400	5	209.3000	231.9800
6	248.9800	245.4800	6	207.4000	234.1900
7	163.9300	134.8900	7	239.8400	241.2500
8	245.9200	223.3900	8	211.2400	255.6000
9	188.0500	169.7000	9	230.3600	256.5500

Das Konfidenzintervall und die Äquivalenzbereiche beziehen sich auf den Quotienten $A/B = (\text{Phase 1})/(\text{Phase 2})$ für die erste Gruppe und entsprechend $A/B = (\text{Phase 2})/(\text{Phase 1})$ für die zweite Gruppe. Die Intervalle werden somit auf der Basis von $n = n_1 + n_2 = 6 + 6 = 12$ Werten berechnet:

Hodges-Lehmann-Schätzer: $M = 103.44$

Konfidenz $P = 1 - 2\alpha$ des Konfidenzintervalls: $P = 0.9000$

Signifikanzniveau α für Äquivalenztest: $\alpha = 0.0500$

Konfidenzintervall: (94.22 , 109.65)

Äquivalenzbereich : (80.00 , 120.00) : 'Bioäquivalenz'

Äquivalenzbereich : (70.00 , 142.86) : 'Bioäquivalenz'

Äquivalenzbereich : (80.00 , 125.00) : 'Bioäquivalenz'

Äquivalenzbereich : (90.00 , 111.11) : 'Bioäquivalenz'

Literatur:

W.J. Westlake (1972) Use of Confidence Intervals in Analysis of Comparative Bioavailability Trials. J. Pharm. Sci. 61, pp. 1340-1341.
W. Lehmacher, W. van Eimeren (1986) Zur statistischen Bewertung der Ergebnisse von Bioverfügbarkeitsstudien. Therapiewoche 36, pp.413-420.
D. Hauschke et al. (1990) A distribution-free procedure for the stat. analysis of bioequiv. studies. Int.J.Clin.Pharm.Ther.Tox. 28,2,72-8.

Mittlere effektive Konzentration EC50 und andere, mit Konfidenzintervall: Die Berechnung der EC50 erfolgt gemäß Alexander et al. (1999) auf Basis der tatsächlich gemessenen Werte, also *ohne* Anpassung eines speziellen Modells. Dazu sind zwei Voraussetzungen erforderlich: Einerseits sollte der maximale Effekt im Bereich der Dosierungen erreicht werden und es sollten die beiden Dosierungen unter- und oberhalb der EC50 relativ "dicht" liegen, da auf den linearen Anteil der "Sigmoidkurve" Bezug genommen wird.

Zur Verifikation des Moduls wurden Beispiele - wie unten aufgeführt - aus Alexander et al. (1999) (cave: Typos in den Ergebnissen zu Experiment 3 und 5!) und aus Chen et al. (2013) verwendet, letztere sind vollständig verifizierbar. In Chen et al. (2013) findet sich ein Vergleich der Ergebnisse der Methode Methode von Alexander et al. mit den Resultaten von unterschiedlichen Modellen mit angepassten sigmoidalen logistischen Funktionen, die zum Teil in anderen Programmen verwendet werden.

Mittlere effektive Konzentration EC50

Eingabedatei ist ...\\Valide\\Alexander1999.DAT
davon Effekt-Variablen 2 (Var3), 3 (Var4), 4 (Var5), 5 (Var6) und 6 (Var7)
und Dosis-Variable 1 (Dosis)

>> Die Dosis-Variable "Dosis" ist in der Datei logarithmisch angegeben.
>> Die Response-Variablen sind in der Datei in absoluten Werten angegeben.
>> Es wird die mittlere effektive Konzentration EC50 berechnet.

Dosis logarithmisch (Log10):

Experiment Var3:	EC50 = -8.7800
Experiment Var4:	EC50 = -9.3589
Experiment Var5:	EC50 = -8.7167
Experiment Var6:	EC50 = -8.5196
Experiment Var7:	EC50 = -8.3300

Gemittelte EC-Werte: EC50 = -8.7410 mit SD = 0.3879

95%-Konfidenzintervall: [-9.2227 , -8.2594]

Nach Rücktransformation per Exp10 untransformiert:

Experiment Var3:	EC50 = 1.659586E-09
Experiment Var4:	EC50 = 4.375938E-10
Experiment Var5:	EC50 = 1.920140E-09
Experiment Var6:	EC50 = 3.022434E-09
Experiment Var7:	EC50 = 4.677349E-09

Gemittelte EC-Werte: EC50 = 1.815315E-09

95%-Konfidenzintervall: [5.988381E-10 , 5.502941E-09]

Literatur:

H Alexander, DJ Browse, SJ Reading, IS Benjamin (1999) A simple and accurate method for calculation of the EC50. J Pharmacol Toxicol 41, pp. 55-8.

Z Chen, R Bertin, G Froidi (2013) EC50 estimation of antioxidant activity in DPPH assay using several statistical programs. Food Chemist 138, pp. 414-20.

Bland-Altman-Vergleich

Graphik zu Bland-Altman's Methodenvergleich

Passing-Bablok-Regression

Graphik zu Passing-Bablok's Methodenvergleich

Lin's Konkordanz-Korrelation

Graphik zu Lin's Konkordanz-Korrelation

ROC-Kurve mit Konfidenzintervall und Graphik

Vergleich von ROC-Kurven („paired“ und „unpaired“)

Mc-Nemar-Test (A.3)

Terwilliger und Ott's Marginal-Symmetrietest (A.4)

Bowker's Kontingenztafel-Symmetrietest (A.4)

Cohen's Kappa (dichotom, 2 Rater) (A.3)

Cohen's Kappa (mehrstufige Zielgröße, 2 Rater) (A.4)

Cohen's Kappa (mehrere Stufen, mehrere Rater) (A.4)

Beurteilung diagnostischer Tests (Sensitivität, Spezifität, Youden etc.)

Likelihood-Ratios und Vergleich von Likelihood-Ratios

Prädiktive Werte und Vergleich von Prädiktiven Werten

Evidence Based Medicine - NNT & Co. (A.19).

Bland-Altman-Vergleich: Vergleich zweier Messgeräte - Messen beide Geräte "dasselbe"? Das Programm sieht verschiedene Optionen zur graphischen Darstellung vor, die an anderer Stelle vorgenommen wird (vgl. auch Scattergramm). Bitte vergleichen Sie auch die Methode von Passing und Bablok, die als nicht-parametrisches Verfahren die gleiche Fragestellung bearbeitet. Die verwendeten Daten wurden Bland und Altman (1986) entnommen:

Bland-Altman's Methodenvergleich

Eingabedatei ist ..\Valide\Meter.DAT

1. Methode: Y=Variable1=MiniM - 2. Methode: X=Variable2=LargeM

Eingelesene Daten:

i	Y	X
	1	2
1	100	90
2	170	240
3	260	265
4	350	400
5	410	380
6	420	390
7	440	360
8	420	420
9	410	420
10	450	470
11	500	520
12	500	540
13	550	590
14	640	670
15	700	690

Korrelation der beiden Methoden (untransformierte Werte):

Korrelationskoeffizient $r = 0.9724$
Bestimmtheitsmaß $B = r^2 = 0.9456$
Testgröße t der Korrelation = 15.0255
Überschreitungswahrsch. $p = 0.000000$
Stichprobenumfang $n = 15$
Anzahl der Freiheitsgrade $n-2 = 13$

Bland und Altman's Transformationen:

Bland-Altman's Abszisse: Durchschnitt $(Y+X)/2$
Bland-Altman's Ordinate: Differenzen $Y-X$

Regression 'Differenzen auf Durchschnitt':

Regressionskoeffizient $b = -0.0117$
Konstanter Achsenabschnitt $c = -3.3548$
Reststreuung $S^2_{y.x} = 1518.0823$
Testgröße t der Regression $t = 0.1783$
Überschreitungswahrsch. $p = 0.861232$
Stichprobenumfang $n = 15$
Anzahl der Freiheitsgrade $n-2 = 13$

Residuen-Ausreißer-Test (Grubbs/David): $p < 0.10$: Cave!

Korrelation der Differenzen und Durchschnitt:

Korrelationskoeffizient $r = -0.0494$
Bestimmtheitsmaß $B = r^2 = 0.0024$
Testgröße t der Korrelation = -0.1783
Überschreitungswahrsch. $p = 0.861232$
Stichprobenumfang $n = 15$
Anzahl der Freiheitsgrade $n-2 = 13$

Durchschnitte, Streuung und Extrema der Differenzen:

Durchschnitt = -8.3333
Streuung $s^2 = 1413.0952$
Standardabw. $s = 37.5912$
Minimum $\min = -70.0000$
Maximum $\max = 80.0000$
Spannweite $R = 150.0000$

Gauß/Ausreißer-Test für die Differenzen (Grubbs/David): $p > 0.10$: OK!

t-Test für die Differenzen:

Student's Prüfgröße $t = 0.8586$ mit $df=14$ ($p = 0.405035$)

Toleranzbereiche für die Differenzen:

$P=0.90$: Untere Grenze = -76.71 Obere Grenze = 60.05
 $P=0.95$: Untere Grenze = -91.60 Obere Grenze = 74.94
 $P=0.99$: Untere Grenze = -123.91 Obere Grenze = 107.24

Relation (Länge des Toleranzbereichs)/(Spannweite der Durchschnitt):

$P=0.90$: Relation = 22.79 %
 $P=0.95$: Relation = 27.76 %
 $P=0.99$: Relation = 38.52 %

Beurteilung: Außer der Korrelation der untransformierten Werte sollen alle Testergebnisse nicht-signifikant ausfallen (möglichst mit $p > 0.10!$). Zu einer Interpretation signifikanter Resultate geben Bland und Altman (Literatur!) einige Hinweise.

Fette durchgezogene Linie: Durchschnitt der Differenzen der Methoden
Dünne durchgezogene Linien: 95%-Toleranzgrenzen für die Differenzen
Langgestrichelte Linie: Optimale Null-Linie bei Übereinstimmung

Literatur:

D.G. Altman, J.M. Bland (1983) *Measurements in Medicine: The Analysis of Method Comparison Studies*. *The Statistician* 32, 3, pp. 307-318.
J.M. Bland, D.G. Altman (1986) *Statistical methods for assessing agreement between two methods of clinical measurements*. *Lancet* 8.2.1986

Passing-Bablok-Regression: Vergleich zweier Messgeräte - Messen beide Geräte "das-selbe"? Die Methode dient der gleichen Fragestellung wie Bland und Altman's Methodenvergleich; dort finden sich auch weitere Aspekte und insbesondere auch teststatistische Verfahren. Es ist zu bemerken, dass die Passing-Bablok-Regression als nicht-parametrisches Verfahren aufzufassen ist. Die verwendeten Daten wurden Bland und Altman (1986) entnommen, die Validierung der Methode erfolgte durch per - Hand - Berechnungen und durch das Programm "Method Validator" von Philippe Marquis. Eine graphische Darstellung ist vorgesehen:

Methodenvergleich: Passing-Bablok-Regression ▲

Eingabedatei ist ..\VALIDE\Meter.dat

Methode 1 (Ordinate): Y = Variable 1 = MiniM

Methode 2 (Abszisse): X = Variable 2 = LargeM

Korrelationskoeffizient r = 0.972396 (p=0.000000)

Passing-Bablok-Regression:

Steigung b = 0.938272

Konfidenzintervall: [0.800000 , 1.085714]

Achsenabschnitt c = 12.098778

Konfidenzintervall: [-49.142895 , 77.999992]

Beide Konfidenzintervalle wurden mit P = 0.95 berechnet.

Zum Test von $H_0(\beta=1)$: H_0 ablehnen, falls $\beta=1$ nicht im Konfidenzintervall

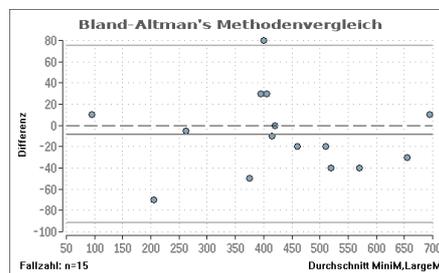
Zum Test von $H_0(c=0)$: H_0 ablehnen, falls $c=0$ nicht im Konfidenzintervall

Literatur:

H. Passing, W. Bablok (1983) *A New Biometrical Procedure for Testing the Equality of Measurements from Two Different Analytical Methods*.

J. Clin. Chem. Clin. Biochem. Vol. 21, 1983, pp. 709-720.

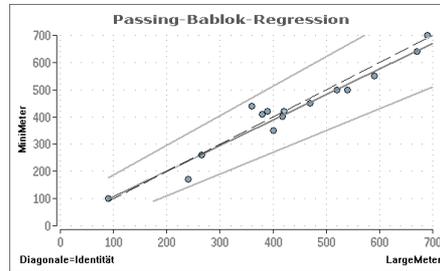
Graphik zum Bland-Altman-Vergleich: Vergleich z.B. zweier Messgeräte - Messen beide Geräte "das-selbe"? Daten aus Bland/Altman (1986), Berechnungen, Scattergramm und eine Legende zur Abbildung (Bedeutung der Geraden!) finden sich weiter unten und an anderer Stelle:



Fette durchgezogene Linie: Durchschnitt der Differenzen der Methoden
Dünne durchgezogene Linien: 95%-Toleranzgrenzen für die Differenzen
Langgestrichelte Linie: Optimale Null-Linie bei Übereinstimmung

Graphik zur Passing-Bablok-Regression: Vergleich z.B. zweier Messgeräte - Messen beide Geräte "dasselbe"? Daten aus Bland/Altman (1986), Berechnungen, Scattergramm und eine Legende zur Abbildung (Bedeutung der Geraden!) finden sich weiter unten:

Modifikationen der Darstellung, insbesondere der Beschriftung, der Skalierung und des Netzgitters sind im Menü "Bearbeiten" des Graphik-Editors vorgesehen.



Gestrichelte Linie: Optimale Null-Linie bei Übereinstimmung der Methoden
 Dicke Linie: Passing-Bablok's Regressionsgerade zum Methodenvergleich
 Dünne Linien: Konfidenzgrenzen zur Regressionsgeraden (nur optional)

Methodenvergleich - Lin's Konkordanz-Korrelation: Vergleich zweier Messgeräte - Messen beide Geräte "dasselbe"? Die Methode dient der gleichen Fragestellung wie Bland-Altman's Methodenvergleich; dort finden sich auch weitere Aspekte und speziell auch teststatistische Verfahren. Die verwendeten Daten wurden Bland und Altman (1986) entnommen, die Validierung der Methode erfolgte durch per - Hand - Berechnungen und durch das Programm "Stata". Eine graphische Darstellung der Punktwolke, des "Schwerpunktes" und der Winkelhalbierenden ist möglich, ebenfalls - im Gegensatz zu anderen Verfahren - ein simultaner Vergleich von *mehreren* Methoden.

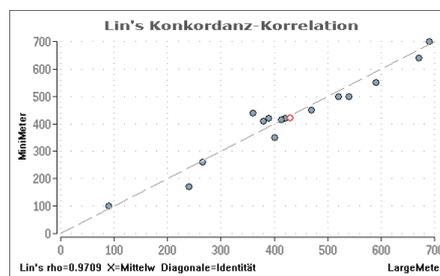
Methodenvergleich: Lin's Konkordanz-Korrelation ▲

Eingabedatei ist ..\VALIDE\Meter.dat

Methode1 = Variable 1 = MiniM
 Methode2 = Variable 2 = LargeM

Eingelesene Daten:

i	Y	X
1	100	90
2	170	240
3	260	265
4	350	400
5	410	380
6	420	390
7	440	360
8	420	420
9	410	420
10	450	470
11	500	520
12	500	540
13	550	590
14	640	670
15	700	690



Pearson's Korrelationskoeffizient $r = 0.972396$ ($p=0.000000$)

Lin's Konkordanz-Korrelationskoeffizient $\rho = 0.970916$ ($p=0.000000$)

Approximatives 90%-Konfidenzintervall für rho: [0.929287 ; 0.988188]
 Approximatives 95%-Konfidenzintervall für rho: [0.916386 ; 0.990068]
 Approximatives 99%-Konfidenzintervall für rho: [0.884412 ; 0.992925]

p-Wert und Konfidenzintervalle werden nach Z-Transformation berechnet!

Literatur: L. I-K. Lin (1989,2000) A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 45, 255-68 Corr. Note: Biometrics 56, 2000, 324-5.

ROC-Kurven: Die Berechnung der ROC-Kurven wurde mit Per-Hand-Berechnungen überprüft. Beispiel auch in Abel (1993). Mit Hilfe der ROC-Kurven kann man die "Güte" eines diagnostischen Tests untersuchen, ob dieser zur Differenzierung z.B. zweier Diagnosen tauglich ist: Der Diagnostische Test ist umso besser, je mehr sich die (diskrete!) Kurve "Sensitivität in Abhängigkeit von 1-Spezifität" nach "links oben" krümmt; eine Graphik incl. Konfidenzintervall ist möglich; letzteres ist ein *simultanes* KI, vgl. Macskarry et al. (2005). Zur Nachvollziehbarkeit werden die Daten aus Test.DAT verwendet:

ROC-Kurven

Eingabedatei ist Test.DAT, davon Variablen 2 (VarZwei) und Gruppen GruppNum=1 und 3

Trennpunkt	Sensitivität	Spezifität	1-Spezifität	fehlklassifiziert
2.5000	1.000000	0.200000	0.800000	0.400000
3.5000	1.000000	0.400000	0.600000	0.300000
4.5000	1.000000	0.600000	0.400000	0.200000
18.5000	1.000000	1.000000	0.000000	0.000000
32.5000	0.750000	1.000000	0.000000	0.125000
34.0000	0.500000	1.000000	0.000000	0.250000
37.0000	0.250000	1.000000	0.000000	0.375000

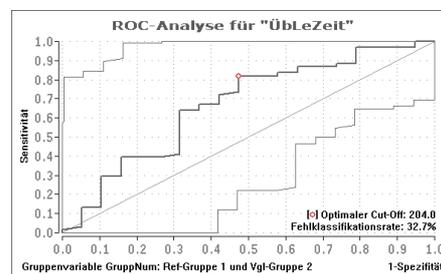
Die Fehlklassifikationsrate gilt für gleiche a-priori-Wahrscheinlichkeiten.

Ein Test auf Abweichung von der Winkelhalbierenden erfolgt äquivalent mit dem Wilcoxon-Mann-Whitney-Test (asymptotisch): Prüfgröße Z = 2.4598 (p=0.013903)

Integral/Fläche unter der ROC-Kurve der interpolierten Trennpunkte = 1.000000

Literatur: Ulrich Abel (1993) Die Bewertung diagnostischer Tests, Hippokrates

Graphik ROC-Kurven: Mit Hilfe der ROC - Kurven kann die "Güte" eines diagnostischen Tests untersucht werden. Die Berechnung der ROC - Kurven wurde mit Per-Hand-Berechnungen überprüft. Zur graphischen Veranschaulichung werden die Daten aus Armitage.DAT verwendet:



Eine Darstellung der Winkelhalbierenden (keine Trennfähigkeit des diagnostischen Tests!) und von Konfidenzintervallen ist möglich (Abbildung!). Wie bei allen Graphiken in **BiAS**. sind Modifikationen der Darstellung, insbesondere der Beschriftung, der Skalierung, des Netzgitters und der Kurvendarstellung bzw. der durch benachbarte Datenpunkte definierte Trennpunkte ebenfalls vorgesehen.

Vergleich von ROC-Kurven (AUCs, "paired" und "unpaired"): "Paired": Für korrelierte AUCs (beide Tests an den gleichen Patienten) wird der nicht-parametrische Test von DeLong et al. (1988) durchgeführt; Verifikation mit alternativen Programmen (MatLab, MedCalc und SAS) und mit Per-Hand-Berechnungen: Dazu im Folgenden ein Beispiel mit artifiziiellen Daten. Der „Globaltest“ zum Vergleich von mehreren AUCs erfolgt ebenfalls mit DeLong et al., Verifikation mit Per-Hand-Berechnungen und SAS. "Unpaired": Die Parameter AUC=Fläche und SD=Standardabweichung der AUC können mit dem ROC-Modul in "Diagnostische Tests" errechnet werden, die Überprüfung erfolgte mit per-Hand-Berechnungen und mit dem Programm "MedCalc". Ein Beispiel dazu erübrigt sich.

Vergleich von ROC-Kurven für "paired" Daten: DeLong-Test ▲

Eingabedatei ist \VALIDE\DeLong.DAT
davon Variablen 1 (VarEins), 2 (VarZwei)
und Gruppen Var. 4 (GruppNum) = 1:gesund und 2:krank

Gruppe 1

i	Var. 1	Var. 2
1	8.0000	2.0000
2	2.0000	3.0000
3	6.0000	5.0000
4	4.0000	6.0000
5	5.0000	4.0000

Gruppe 2

i	Var. 1	Var. 2
1	10.0000	3.0000
2	8.0000	5.0000
3	9.0000	1.0000
4	11.0000	5.0000
5	14.0000	7.0000
6	9.0000	9.0000
7	15.0000	8.0000

Falls "Test positiv" große Werte der Diagnostischen Tests bedeutet (95%-KI):

Test 1 (VarEins): AUC = 0.985714 SD = 0.020203 KI = [0.946117 ; 1.000000]
Test 2 (VarZwei): AUC = 0.671429 SD = 0.166394 KI = [0.345302 ; 0.997555]

Falls "Test positiv" kleine Werte der Diagnostischen Tests bedeutet (95%-KI):

Test 1 (VarEins): AUC = 0.014286 SD = 0.020203 KI = [0.000000 ; 0.053883]
Test 2 (VarZwei): AUC = 0.328571 SD = 0.166394 KI = [0.002445 ; 0.654698]

Differenz der beiden AUCs: 0.314286
Std-Abw. SD der Differenz: 0.171924
Kovarianz der AUCs: Cov = -0.000731

90%-Konfidenzintervall für die Differenz: [0.031496 ; 0.597075]
95%-Konfidenzintervall für die Differenz: [-0.022679 ; 0.651250]
99%-Konfidenzintervall für die Differenz: [-0.128561 ; 0.757132]

Nullhypothese Ho(AUC1=AUC2): Prüfgröße Z = 1.828 mit p = 0.067542 (zweiseitig)

Literatur: E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 44, 3, pp. 837-45. ▲

Beurteilung diagnostischer Tests: Es sind die üblichen Parameter Spezifität, Sensitivität etc. für einen medizinischen Test (zum Beispiel Labortest) zu errechnen. Die Konfidenzintervalle wurden gemäß Sachs (1977, Kap.45) berechnet und überprüft, LR-KI's gemäß Koopman (1984). Das KI zum Youden-Index wurde per Additivität der Varianzen der beiden Parameter approximativ berechnet, vgl. Abel (1993). Einige Definitionen zu den Kenngrößen finden sich bei Galen und Gambino (1979).

Beurteilung diagnostischer Tests ▲

	Testergebnis:		Summe der Zeile
	positiv	negativ	
Anzahl Erkrankte	12	4	16
Anzahl Gesunde	1	32	33
Summe der Spalte	13	36	49

Sensitivität (Richtig positiv) = 75.00% (47.62% , 92.73%)
 Spezifität (Richtig negativ) = 96.97% (84.24% , 99.92%)
 Rate 'Falsch positiv' = 3.03% (0.08% , 15.76%)
 Rate 'Falsch negativ' = 25.00% (7.27% , 52.38%)
 Prädiktiver Wert positiv (ri.pos./pos.ges.)= 92.31% (63.97% , 99.81%)
 Prädiktiver Wert negativ (ri.neg./neg.ges.)= 88.89% (73.94% , 96.89%)
 Effizienz (Rate der richtigen Entscheidungen)= 89.80% (77.77% , 96.60%)
 Youden-Index Y = Sens. + Spezif. - 100% = 71.97% (49.96% , 93.98%)

Likelihood-Ratio positiv = Sens./(1-Spez.) = 24.750 (3.519 ,174.075) "sehr gut"
 Likelihood-Ratio negativ = (1-Sens.)/Spez. = 0.258 (0.110 , 0.604) "gut"

Die Zahlen in Klammern: 95%-Konfidenzintervall für den entsprechenden Parameter.

Literatur: R.S. Galen, S.R. Gambino (1979) Norm und Normabweichung klinischer Daten. Gustav Fischer Stuttgart New York.
 U. Abel (1993) Die Bewertung diagnostischer Tests. Hippokrates Stuttgart.
 P.A.R. Koopman (1984) Confidence Intervals for the Ratio of Two Binomial Proportions. Biometrics 40, pp. 513-517

Likelihood-Ratios und Vergleich von Likelihood-Ratios ("unpaired"): Neben der Berechnung von Likelihood-Ratios ist der Vergleich von zwei Likelihood-Ratios von Interesse, hier mit Hilfe von Konfidenzintervallen und zusätzlich mit einem asymptotischen teststatistischen Vergleich. Eine Überprüfung ist mit dem Beispiel von Koopman (1984) möglich. Die Varianz des unpaired-Konfidenzintervalls für LR1/LR2 folgt Pfanzagl (1974), Überprüfung mit per-Hand-Berechnungen.

Vergleich von Likelihood-Ratios ("unpaired") ▲

TEST 1:	Testergebnis:				Summe der Zeile
	positiv		negativ		
Anzahl Erkrankte	36	90.00%	4	10.00%	40
Anzahl Gesunde	16	20.00%	64	80.00%	80
Summe Spalte	52		68		120

TEST 2:	Testergebnis:				Summe der Zeile
	positiv		negativ		
Anzahl Erkrankte	12	75.00%	4	25.00%	16
Anzahl Gesunde	1	3.03%	32	96.97%	33
Summe Spalte	13		36		49

Alle Konfidenzintervalle KI=[l;r] wurden mit Konfidenz P=0.9500 berechnet:

Berechnungen für Likelihood-Ratios positiv:

LR-Test1 mit KI: 4.5000 [2.8686; 7.0593]
 LR-Test2 mit KI: 24.7500 [3.5190; 174.0752]

LR1/LR2 mit KI: 0.1818 [0.0246; 1.3461] Ho-Test: p=0.095119

Berechnungen für Likelihood-Ratios negativ:

LR-Test1 mit KI: 0.1250 [0.0490; 0.3188]
 LR-Test2 mit KI: 0.2578 [0.1101; 0.6037]

LR1/LR2 mit KI: 0.4848 [0.1368; 1.7179] Ho-Test: p=0.262025

Die Nullhypothese Ho(LR1/LR2=1) bedeutet die Identität der Likelihood-Ratios.

Literatur: .A.R. Koopman (1984) Confidence Intervals for the Ratio of Two Binomial Proportions. Biometrics 40, pp. 513-517

Likelihood-Ratios und Vergleich von Likelihood-Ratios ("paired"): Wie "unpaired"; das paired-Konfidenzintervall wurde mit Hilfe von Bonett & Price (2006) und eigenen Ableitungen berechnet und per Hand überprüft; als Beispiel für "paired" wird das Beispiel von Leisenring (2000) verwendet, vgl. dazu auch den "Vergleich Prädiktiver Werte":

Vergleich von Likelihood-Ratios zweier diagnostischer Tests (paired) ▲

Stichwort: Daten aus Leisenring et al. (2000)

ERKRANKTE	Test 1: positiv	Test 1: negativ	Summe
Test 2: positiv	473 77.80%	81 13.32%	554 91.12%
Test 2: negativ	29 4.77%	25 4.11%	54 8.88%
Summe	502 82.57%	106 17.43%	608 100.00%

GESUNDE	Test 1: positiv	Test 1: negativ	Summe
Test 2: positiv	22 8.37%	44 16.73%	66 25.10%
Test 2: negativ	46 17.49%	151 57.41%	197 74.90%
Summe	68 25.86%	195 74.14%	263 100.00%

Alle Konfidenzintervalle KI=[l;r] wurden mit Konfidenz P=0.9500 berechnet:

Berechnungen für Likelihood-Ratios positiv:

LR-Test1 mit KI: 3.1934 [2.5939; 3.9313]
 LR-Test2 mit KI: 3.6309 [2.9424; 4.4806]

LR1/LR2 mit KI: 0.8795 [0.6645; 1.1640] Test Ho: p=0.369172

Berechnungen für Likelihood-Ratios negativ:

LR-Test1 mit KI: 0.2351 [0.1950; 0.2835]
 LR-Test2 mit KI: 0.1186 [0.0911; 0.1544]
 LR1/LR2 mit KI: 1.9831 [1.4872; 2.6444] Test Ho: p=0.000003

Die Nullhypothese Ho(LR1/LR2=1) bedeutet die Identität der Likelihood-Ratios.

Literatur: D.G. Bonett, R.M. Price (2006) Confidence intervals for a ratio of binomial proportions based on paired data. Statistics in Med. 25, pp.3039-47

Prädiktive Werte und Vergleich von Prädiktiven Werten (unpaired): Neben der Berechnung von Prädiktiven Werten ist der Vergleich von zwei Prädiktiven Werten von Interesse, hier mit Hilfe von Konfidenzintervallen und zusätzlich mit einem asymptotischen teststatistischen Vergleich. Eine Überprüfung ist mit dem Beispiel von Koopman (1984) möglich. Im Beispiel werden die gleichen Daten verwendet wie zum Vergleich von Likelihood-Ratios (unpaired), Überprüfung mit per-Hand-Berechnungen und den entsprechenden Moduln "Beurteilung diagnostischer Tests" und "Quotient von binomialverteilten Größen".

Vergleich von prädiktiven Werten ("unpaired") ▲

TEST 1:	Testergebnis:				Summe der Zeile
	positiv		negativ		
Anzahl Erkrankte	36	90.00%	4	10.00%	40
Anzahl Gesunde	16	20.00%	64	80.00%	80
Summe Spalte	52	43.33%	68	56.67%	120

TEST 2:	Testergebnis:				Summe der Zeile
	positiv		negativ		
Anzahl Erkrankte	12	75.00%	4	25.00%	16
Anzahl Gesunde	1	3.03%	32	96.97%	33
Summe Spalte	13	26.53%	36	73.47%	49

Alle Konfidenzintervalle $KI=[l;r]$ wurden mit Konfidenz $P=0.9500$ berechnet:

Prädiktive Werte positiv (%):

PPV1 Test1 mit KI: 69.2308 [54.8976; 81.2827]
 PPV2 Test2 mit KI: 92.3077 [63.9703; 99.8054]
 PPV1/PPV2 mit KI: 0.7500 [0.5901; 0.9532] Test Ho: $p=0.090377$

Prädiktive Werte negativ (%):

NPV1 Test1 mit KI: 94.1176 [85.6182; 98.3742]
 NPV2 Test2 mit KI: 88.8889 [73.9389; 96.8884]
 NPV1/NPV2 mit KI: 1.0588 [0.9299; 1.2057] Test Ho: $p=0.341093$

Die Tests der $H_0(PV1/PV2=1)$ werden jeweils mit der H_0 -Varianz durchgeführt.

Der Vergleich ist von der Prävalenz abhängig: Nur unstratifiziert verwenden!

Literatur: P.A.R. Koopman, Confidence intervals for the ratio of two binomial proportions. Biometrics 40, 1984, pp. 513-517

Prädiktive Werte und Vergleich von Prädiktiven Werten (paired): Alle Einzelheiten dazu bei Bonett und Price (2006), bei Leisenring et al. (2000) findet sich ein durchgerechnetes Beispiel, das hier zur Validierung verwendet werden kann.

Vergleich von prädiktiven Werten ("paired")

Stichwort: Daten aus Leisenring et al. (2000)

ERKRANKTE	Test 1: positiv	Test 1: negativ	Summe
Test 2: positiv	473 54.31%	81 9.30%	554 63.61%
Test 2: negativ	29 3.33%	25 2.87%	54 6.20%
Summe	502 57.63%	106 12.17%	608 69.80%

GESUNDE	Test 1: positiv	Test 1: negativ	Summe
Test 2: positiv	22 2.53%	44 5.05%	66 7.58%
Test 2: negativ	46 5.28%	151 17.34%	197 22.62%
Summe	68 7.81%	195 22.39%	263 30.20%

Alle Konfidenzintervalle KI=[l;r] wurden mit Konfidenz P=0.9500 berechnet:

Prädiktive Werte POSITIV (%):

PPV1 Test1 mit KI: 88.0702 [85.1223; 90.6159]
PPV2 Test2 mit KI: 89.3548 [86.6563; 91.6710]

PPV1/PPV2 mit KI: 0.9856 [0.9548; 1.0174] Test Ho: p=0.371051

Prädiktive Werte NEGATIV (%):

NPV1 Test1 mit KI: 64.7841 [59.0969; 70.1769]
NPV2 Test2 mit KI: 78.4861 [72.8806; 83.4052]

NPV1/NPV2 mit KI: 0.8254 [0.7624; 0.8936] Test Ho: p=0.000002

Die Nullhypothese Ho(PV1/PV2=1) bedeutet die Identität der jeweiligen Werte.

Der Vergleich ist von der Prävalenz abhängig: Nur unstratifiziert verwenden!

Literatur: C.S. Moskowitz, M.S. Pepe (2006) Comparing the Predictive Values of Diagnostic Tests: Sample Size and Analysis for Paired Study Designs. Clinical Trials Vol. 3, No. 3, pp. 272-279.

A.19 Epidemiologie und "Evidence Based Medicine" ▲

Experimental Event Rate EER mit Konfidenzintervall
Control Event Rate CER mit Konfidenzintervall
Absolute Risiko-Reduktion ARR mit Konfidenzintervall
Relative Risiko-Reduktion RRR mit Konfidenzintervall
Relativer Nutzen ("benefit") bzw. Schaden ("harm")
Number Needed to Treat NNT (benefit/harm) mit Konfidenzintervall
Statistischer Test zum Vergleich zweier Methoden (Gauß-Test)
Likelihood-Ratio positiv/negativ (A.18)
Test zum Vergleich zweier Likelihood-Ratios (A.18)
Relatives Risiko, Odds-Ratio und Inzidenz bei zwei Stichproben (A.3)
Pearson-Mantel-Haenszel-Test (A.3)
Konfidenzintervall für die Inzidenz einer Stichprobe (A.2)

Bitte beachten Sie, dass die aufgeführten Tests bzw. Parameter zum Teil auch in anderen Moduln (Vierfelder-Tafel, Konfidenzintervalle etc.) enthalten sind. An den entsprechenden Stellen finden sich weitere Beispiele zur Validierung.

Evidence Based Medicine: Berechnung der üblichen Größen CER, EER, ARR, RRR, NNT und deren Konfidenzintervalle. Das Modul kann mit der Zielgröße "Nutzen" oder der Zielgröße "Schaden" ausgeführt werden. Das folgende Beispiel entstammt der Literatur (Quelle leider nicht mehr verifizierbar) und betrachtet die Zielgröße "Tod" (also: "Schaden").

Bitte beachten Sie, dass viele im Internet verfügbaren Beispiele zu den hier berechneten Parametern vielfach falsch berechnet sind und nicht mit den mit "BiAS." erzielten Ergebnissen übereinstimmen. Eine Validierung der hier vorliegenden Ergebnisse ist jedoch einfach mit Hilfe des Moduls "Vergleich zweier Binomial-Verteilungen" möglich, da $NNT=1/ARR$. Entsprechendes gilt für die Grenzen der Konfidenzintervalle. Vorsicht: Dabei Vertauschung der Grenzen beachten! Falls die ARR-Intervalle die Null einschließen: Unterscheidung in "harm" (NNH) und "benefit" (NNB) beachten!

Evidence Based Medicine: NNT und Anderes

	Zielereignis "Schaden"				Summe Be- handlungen
	ja		nein		
Novum	228	20.4%	888	79.6%	1116
Standard	275	24.7%	840	75.3%	1115
Summen	503		1728		2231

Test (ohne Yates-Korrektur, 2-seitig): Gauss' $Z=2.3926$ mit $p=0.016729$

Binomial-Konfidenzintervalle für Event-Rate EER(%) "Novum":

--- 90% ---	--- 95% ---	--- 99% ---
18.46 22.52	18.10 22.92	17.41 23.71

Binomial-Konfidenzintervalle für Event-Rate CER(%) "Standard":

--- 90% ---	--- 95% ---	--- 99% ---
22.55 26.88	22.16 27.30	21.41 28.14

Absolute Risiko-Reduktion (Novum-Standard): ARR(%) = 4.23

Konfidenzintervalle für die absolute Risiko-Reduktion ARR(%):

---	90%	---	---	95%	---	---	99%	---
	1.33		7.14		0.77		7.70	
							-0.32	
							8.79	

Relative Risiko-Reduktion (Novum-Standard)/Standard: RRR(%) = 17.17

Konfidenzintervalle für die relative Risiko-Reduktion RRR(%):

---	90%	---	---	95%	---	---	99%	---
	5.38		28.95		3.12		31.21	
							-1.29	
							35.62	

Number needed to treat: NNT("Benefit") = 23.62

Konfidenzintervalle für NNT("Benefit"):

---	90%	---	---	95%	---	---	99%	---
	14.00		75.37		12.99		129.89	
							11.38	
							> 10^6	

Konfidenzintervalle für NNT("Harm"):

---	90%	---	---	95%	---	---	99%	---
	undefiniert			undefiniert			313.97	
							> 10^6	

VORSICHT: Die NNTs beziehen sich immer auf die Therapie "Novum".

Literatur:

R.J. Cook, D.L. Sackett (1995) The number needed to treat: a clinically useful measure of treatment effect. BMJ 1995, 310, pp.452-4

D.G. Altman (1998) Confidence intervals for the NNT. BMJ 317, pp.1309-12



A.20 Multiples Testen von Nullhypothesen



Multiples Testen: Bonferroni-, Holm-, Simes-, Rüger- und Hommel-Tests

Hierarchisches Testen

Benjamini-Hochberg's False-Discovery-Rate

Multiples Testen: Zur Validierung wurden neben Beispielen aus der angegebenen Literatur einige per-Hand-Berechnungen herangezogen.

Bitte beachten Sie, dass bei hierarchisch geordneten Hypothesen die in diesem Abschnitt verwendeten sog. "alpha-Korrekturen" *nicht* anzuwenden sind, da geordnete Hypothesen *lokal* zum Niveau alpha getestet werden, um das *globale* Niveau einzuhalten. Eine Ausnahme bildet gegebenenfalls das Testen von hierarchischen *Gruppen* von Hypothesen. Im Programm und ganz unten auf dieser Seite finden Sie dazu und zu speziellen Techniken ausführliche Hinweise, ebenfalls Literaturangaben aus Vollmar, J. (Hrsg.) Testing principles in clinical and preclinical trials, Gustav Fischer 1995.

Multiple Prüfung von Nullhypothesen.

Testniveau $\alpha = 0.0500$

Ho	P(i)	Bonferroni	Holm (1979)
1	0.0123	sig.	sig.
2	0.0200	n.s.	n.s.
3	0.0314	n.s.	n.s.
4	0.0550	n.s.	n.s.

Global-Test nach Rüger (1978) und Simes (1986):

Ablehnung der Global-Hypothese, da

$P((1)) = 0.0123 < 1 \cdot \alpha/N = 0.0125$

$P((2)) = 0.0200 < 2 \cdot \alpha/N = 0.0250$

$P((3)) = 0.0314 < 3 \cdot \alpha/N = 0.0375$

Multipler Test nach Hommel (1986,1988):

Alle $H_0(I)$ ablehnen, für deren Indexmenge I von Elementar-Hypothesen

gilt: $|I \cap \{1,2\}| \leq 1$

Die Individual-Hypothesen 1, 2 werden abgelehnt.

Bonferroni, C.E. (1936) Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 8:3-62.

G. Hommel (1988) A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika 1988, 75, pp. 383-386

S. Holm (1979) A simple sequential rejective multiple test procedure. Scand. J. Stat. 6, pp. 65-70

B. Rüger (1978) Das max. Signifikanzniveau des Testes 'Lehne H_0 ab, wenn k unter n geg. Tests zur Ablehnung führen'. Biometrika 52, pp.591-611

R.J. Simes (1986) An improved Bonferroni procedure for multiple tests of significance. Biometrika 73, pp. 751-754

A. Victor, A. Elsässer, G. Hommel, M. Blettner (2010) Wie bewältigt man die p-Wert-Flut? Deutsches Ärzteblatt Jg 107, Heft 4, pp. 50-56

Hierarchisches Testen: Bei hierarchisch geordneten Hypothesen ist keine "alpha-Korrektur" erforderlich. Die Hypothesen werden in einer a-priori definierten Reihenfolge jeweils zum Niveau alpha getestet: Wird eine Hypothese in dieser Reihenfolge akzeptiert, so bricht das Testprozedere ab und auch alle "nachfolgenden" Ho müssen akzeptiert werden. Die bereits abgelehnten Nullhypothesen gelten als abgelehnt jeweils zum Niveau alpha.

Beispiel 1: Vergleich Placebo, Standard und Novum: Im ersten Schritt ("sensitivity") $H_0(P=S)$, dann ("efficacy") $H_0(P=N)$ und im dritten Schritt ("superiority") $H_0(S=N)$ prüfen. (Natürlich könnte die dritte H_0 im Sinne einer "therapeutischen Äquivalenz" auch $H_0(N=S-\delta)$ mit der Alternative $H_a(N>S-\delta)$ lauten.) Die - bis zu drei - Tests werden zum nominalen Niveau alpha durchgeführt.

Beispiel 2: Kombination eines Äquivalenz- mit einem Überlegenheitstest: Im ersten Schritt ("Äquivalenz") wird $H_0(N=S-\delta)$ vs. $H_a(N>S-\delta)$ geprüft. Kann die "Nicht-Unterlegenheit" (einseitiger Test!) angenommen werden, so kann auch der Test auf Überlegenheit durchgeführt werden: Im 2. Schritt ist $H_0(N=S)$ gegen $H_a(N>S)$ zu prüfen, dies ebenfalls zum Niveau alpha und natürlich nur genau dann, wenn im ersten Schritt die Nullhypothese abgelehnt wurde.

Beispiel 3: Dosisfindung: Die höchste Dosis wird zuerst, die absteigenden Dosen in absteigender Reihenfolge jeweils zum Niveau alpha geprüft. Wird an einer Stelle die H_0 "kein Effekt" angenommen, bricht das Verfahren ab und auch alle "niedrigeren" H_0 's müssen - ohne diese zu testen!! - beibehalten werden.

Da zum hierarchischen Testen offenbar keine Berechnung, sondern "nur" eine Planung der Testhierarchie durchgeführt werden muss, enthält das Programm lediglich Hinweise auf die Struktur des Abschlusstests.

Das Buch "Testing principles in clinical and preclinical trials" (Hrsg. Joachim Vollmar), Gustav Fischer 1995, bringt einen hervorragenden Überblick über das Thema.

Benjamini-Hochberg's "False Discovery Rate": Zur Validierung wurden neben Beispielen aus der angegebenen Literatur einige per-Hand-Berechnungen herangezogen. Im Folgenden ein Beispiel aus R. Hoffmann et al. "The False Discovery Rate: A new approach to the problem of multiple testing", Datum 2004, Vol. 10, 1, pp. 3-4 (Statistician's Corner). Bitte beachten Sie, dass die Werte geordnet sind, was bei der Eingabe in BiAS. nicht unbedingt der Fall sein muss; ausgegeben wird immer die laufende Nummer der Eingabe zusammen mit den korrespondierenden (geordneten) p-Werten.

Benjamini-Hochberg's "False Discovery Rate" FDR ▲

Das FDR-level wird mit $FDR=0.05$ angenommen:

Lfd.Nr.	p-Wert	BH-Krit.	signif.
1	0.000390	0.003571	*
2	0.001030	0.007143	*
3	0.001590	0.010714	*
4	0.001640	0.014286	*
5	0.007650	0.017857	*
6	0.019600	0.021429	*
7	0.023700	0.025000	*
8	0.031000	0.028571	
9	0.101000	0.032143	
10	0.157000	0.035714	
11	0.284000	0.039286	
12	0.542000	0.042857	
13	0.543000	0.046429	
14	0.793000	0.050000	

Die mit "*" gekennzeichneten Tests (das sind alle mit p-Werten $\leq 0.023700!$) werden bei $FDR=0.050000$ als signifikant aufgefasst.

Literatur:

Y. Benjamini, Y. Hochberg (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 1, pp. 289-300.

R. Hoffmann et al. (2004) "The False Discovery Rate: A new approach to the problem of multiple testing", Datum 2004, Vol. 10, 1, pp. 3-4 (Statistician's Corner)

Bitte beachten Sie, dass Methoden wie die Bonferroni-Adjustierung und die anderen unter "Multiples Testen" behandelten Verfahren, ebenfalls das "Hierarchische Testen" die sogenannte Familywise-error-rate (FWE) kontrollieren, das heißt, die Wahrscheinlichkeit, wenigstens eine falsch-positive Entscheidung zu erhalten bzw. mindestens eine Nullhypothese irrtümlich abzulehnen ist kleiner oder gleich der vorgegebenen Irrtumswahrscheinlichkeit α .

Bei sehr vielen Nullhypothesen werden diese Methoden sehr konservativ, so dass man – wie zum Beispiel bei der Untersuchung von Genexpressionen mit zum Teil hunderten von Nullhypothesen – die Anwendung der "False Discovery Rate" in Erwägung ziehen kann. Die FDR ist der erwartete Anteil von falsch-positiven Ergebnissen und ist immer größer als die FWE. Eine FDR von zum Beispiel 5% bedeutet, dass im Mittel 5% der abgelehnten Nullhypothesen irrtümlich abgelehnt werden, also "falsch positiv" sind.

Ein Beispiel zur "False Discovery Rate" FDR nach Benjamini und Hochberg und weitere Erläuterungen dazu finden sich bei Hoffmann et al. (2004).

Einen lesenswerten Überblick über die Fragestellungen geben A. Victor, A. Elsässer, G. Hommel und M. Blettner in ihrem Artikel „Wie bewältigt man die p-Wert-Flut?“ im Deutschen Ärzteblatt 2010, Jg 107, Heft 4, pp. 50-56.



A.21 Meta-Analyse mit p-Werten

Meta-Analyse bei unabhängigen Versuchen (einseitige Alternativen)
Meta-Analyse bei unabhängigen Versuchen (zweiseitige Alternativen)

Meta-Analyse bei unabhängigen Versuchen (einseitige Alternativen): Dazu wurden neben Beispielen aus der angegebenen Literatur per-Hand-Berechnungen herangezogen:

Zusammenfassen von unabhängigen Versuchen (einseitige Alternative).

Ho(1): p = 0.010000
Ho(2): p = 0.020000
Ho(3): p = 0.030000
Ho(4): p = 0.120000
Ho(5): p = 0.220000

Fisher's Verfahren: $\chi^2 = 31.3163$ (df=10) p = 0.000520
Inverse Normalverteilung: N = 3.6713 p < 0.000121
Tippett's Verfahren: p < 0.049010

Literatur:

R.A. Fisher (1971) *Statistical Methods for Research Workers*. 14. Auflage, Hafner Press. pp. 99-101.
J.A. Koziol, M.D. Perlman (1978) Combining Independent Chi-Squared Tests. *J. Amer. Stat. Ass.* 73, No. 4, pp. 753-763.

Meta-Analyse bei unabhängigen Versuchen (zweiseitige Alternativen): Es werden die gleichen p-Werte verwendet wie oben, jedoch bei unterstellter zweiseitiger Prüfung. Im folgenden Ausdruck gibt VZ=Vorzeichen an, "in welche Richtung" sich in den Individualtests Unterschiede ergaben.

Unabhängige Versuche: zweiseitige Alternative

Ho(1): p = 0.010000, VZ = +
Ho(2): p = 0.020000, VZ = +
Ho(3): p = 0.030000, VZ = +
Ho(4): p = 0.120000, VZ = +
Ho(5): p = 0.220000, VZ = +

Inverse Normalverteilung: N = 4.4067 (p = 0.000010)

Literatur:

B.J. Winer (1971) *Statistical Principles in Experimental Design*. McGraw Hill New York.
J.A. Koziol, M.D. Perlman (1978) Combining Independent Chi2-Tests. *K.Amer, M.D. Stat.. Ass* 73, No. 4, pp. 753-763

A.22 Zwischenauswertungen – Interimsanalyse



Zwischenauswertungen nach Pocock
 Zwischenauswertungen nach Fleming-O'Brien
 Adaptive Interimsanalyse: Das Bauer-Köhne-Design

Zwischenauswertungen nach Pocock: Pocock (1977) beschreibt eine gruppensequentielle Methode zur Planung und Auswertung von Studien:

Zwischenauswertungen nach Pocock

Bei $k-1$ geplanten Zwischenauswertungen (die k -te Auswertung ist die Endauswertung!!) ergibt sich das nominale Signifikanzniveau α -korr= α -korr der $k-1$ Interimsanalysen und der Endauswertung. α ist das Signifikanzniveau der gesamten Studie. Für $\beta=0.10$ und 0.05 (bzw. Power= $1-\beta$) wird die mittlere Anzahl von Auswertungen bis zum Abschluss angegeben.

k	alpha=0.05			alpha=0.01		
	a-korr	$\beta=0.10$	$\beta=0.05$	a-korr	$\beta=0.10$	$\beta=0.05$
2	0.0294	1.41	1.31	0.0056	1.47	1.37
3	0.0221	1.88	1.71	0.0041	2.00	1.83
4	0.0182	2.36	2.12	0.0033	2.53	2.29
5	0.0158	2.84	2.53	0.0028	3.06	2.75
6	0.0142	3.32	2.94	0.0025	3.59	3.22
7	0.0130	3.80	3.35	0.0023	4.12	3.68
8	0.0120	4.28	3.77	0.0021	4.65	4.14
9	0.0112	4.76	4.18	0.0019	5.18	4.61
10	0.0106	5.24	4.59	0.0018	5.71	5.07
15	0.0086	7.63	6.66	0.0015	8.36	7.39
20	0.0075	10.03	8.72	0.0013	11.00	9.70

Literatur: S.J. Pocock (1977) Group sequential methods in the design and analysis of clinical trials. Biometrika 64, pp. 191-199.

Adaptive Interimsanalyse nach Bauer und Köhne: Daten aus: F. Krummenauer (2004) Leitlinien Sektion DOG-Glaukom der Deutschen Ophthalmol. Gesellschaft. Es handelt sich um eine Vergleichstudie zum Niveau $\alpha=0.05$. Die erste Phase wird mit $\alpha_1=0.03$ ausgeführt, daraus errechnet sich ein $\alpha_0=0.30$ ("Stop for futility"). Mit dem (fiktiven) $p_1=0.098310$ der ersten Phase ($\alpha_1 < p_1 < \alpha_0$) errechnet sich $\alpha_2=0.088546$ für die zweite Phase, deren p-Wert wiederum 0.075620 betrage: Damit ergibt sich ein Gesamt-p-Wert von $p=0.043874 < 0.05$, womit die H_0 abgelehnt wird. Falls $p_1 < \alpha_1$, kann bereits in der ersten Phase H_0 abgelehnt werden, im geschilderten Fall $\alpha_1 < p_1 < \alpha_0$ wird die zweite Phase unter Neuberechnung der Fallzahl etc. angegangen.

BAUER-KÖHNE-DESIGN

Irrtumswahrscheinlichkeit alpha = 0.050000

Phase 1:

Irrtumswahrscheinlichkeit alpha1 = 0.030000

Irrtumswahrscheinlichkeit alpha0 = 0.298488

Phase 2:

p-Wert der 1. Phase: = 0.098310

Irrtumswahrscheinlichkeit alpha2 = 0.088546

Abschluss:

p-Wert der 2. Phase: = 0.075620

Gesamt-p-Wert: p = 0.043874

Literatur: P. Bauer, K. Köhne (1994) Evaluation of experiments with adaptive interim analysis. Biometrics 50, pp. 1029-41.

Zwischenauswertungen nach Fleming-O'Brien: Fleming et al. (1979) beschreiben eine gruppensequentielle Methode zur Planung und Auswertung von Studien:

ZWISCHENAUSWERTUNGEN (FLEMING-O'BRIEN)

Bei k-1 geplanten Zwischenauswertungen (die k-te Auswertung ist die Endauswertung!!) ergeben sich die nominalen Signifikanzniveaus alpha-i (i=1,..,k-1) der k-1 einzelnen Interimsanalysen; i=k bedeutet die Endauswertung bzw. deren Signifikanzniveau. Alpha ist das Signifikanzniveau der gesamten Studie.

Alpha = 0.01							
k=2		k=3			k=4		
alpha-1	alpha-2	alpha-1	alpha-2	alpha-3	alpha-1	alpha-2	alpha-3
0.00100	0.00943	0.00050	0.00056	0.00952	0.00033	0.00037	0.00043
0.00200	0.00872	0.00100	0.00114	0.00889	0.00067	0.00075	0.00089
0.00300	0.00791	0.00150	0.00174	0.00817	0.00100	0.00114	0.00137
0.00400	0.00704	0.00200	0.00234	0.00738	0.00133	0.00153	0.00184
0.00500	0.00612	0.00250	0.00296	0.00651	0.00167	0.00194	0.00233
Alpha = 0.05							
k=2		k=3			k=4		
alpha-1	alpha-2	alpha-1	alpha-2	alpha-3	alpha-1	alpha-2	alpha-3
0.00500	0.04806	0.00250	0.00296	0.04831	0.00167	0.00194	0.00233
0.01000	0.04519	0.00500	0.00612	0.04588	0.00333	0.00400	0.00488
0.01500	0.04177	0.00750	0.00936	0.04292	0.00500	0.00612	0.00753
0.02000	0.03788	0.01000	0.01268	0.03949	0.00667	0.00827	0.01026
0.02500	0.03355	0.01250	0.01606	0.03558	0.00833	0.01047	0.01306

Tabellen für k=5,..,10 finden sich bei Fleming et al. (vgl. Literatur).

Literatur: Th.R. Fleming, D.P. Harrington, P.C. O'Brien (1979) Designs for Group Sequential Tests. Controlled Clinical Trials 5, 4, pp. 348-361



Konfidenzintervall für μ
Ein- und Zweistichproben-t-Test auf Unterschied
Ein- und Zweistichproben-t-Test auf Äquivalenz
Einweg-Varianzanalyse (ANOVA)
Studentized Range
Cross-Over-Analyse
Einfache lineare und multiple Korrelationskoeffizienten r
Einfache lineare und multiple Regressionskoeffizienten β
Multiple Regression und Korrelation: Omnibus-Test
Vergleich von Korrelationskoeffizienten
Spearman's rho
Kendall's tau
Logistische Regression
Schätzung einer Wahrscheinlichkeit
Test eines binomialen Anteils
Vergleich binomialverteilter Größen
Äquivalenz binomialverteilter Größen
 χ^2 -Vierfelder-Test
Dixon-Mood's Vorzeichentest
McNemar-Test
Cohen's Kappa
Wilcoxon-matched-pairs-Test
Wilcoxon-Mann-Whitney-U-Test
Kruskal-Wallis-Test
Log-Rank-Test
Kaplan-Meier-Schätzer: Test $H_0(P|t)$
Cox-Regression
Odds-Ratio und Relatives Risiko
Many-one-Vergleiche zum t-Test und zum U-Test
 χ^2 -Kontingenztafeln
Bland-Altman-Methodenvergleich
Intraclass- Korrelationen
Cluster-randomisierte Studien
Konfidenzintervall und Test für eine AUC
ROC-Kurven-Vergleich für „paired“ und „unpaired“

Auf eine explizite Darstellung bzw. Validierung der Power-Berechnungen wurde verzichtet, da diese sich komplementär aus den aufgeführten Fallzahlberechnungen ergeben.

Bei allen parametrischen Fallzahlberechnungen ist zu beachten, dass BiAS. in der Regel nicht nur mit den vielfach üblichen und auch häufig in der Literatur verwendeten Gauß-Approximationen arbeitet, sondern weitgehend iterative Formeln oder nicht-zentrale Verteilungen für zum Beispiel die t-Verteilung verwendet. Dadurch ergeben sich unter Umständen Abweichungen von gängigen Referenztabellen oder von anderen Programmen.

Konfidenzintervall für μ : Zur Berechnung des Stichprobenumfanges für Konfidenzintervalle für μ wurden Per-Hand-Berechnungen herangezogen. **Vorsicht:** Das zitierte Lehrbuch von Werner (1992, p.242) rechnet mit einer Gauß-Approximation!

Konfidenzintervall

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Länge des Intervalls L : 20.0000
Schätzung für Varianz σ^2 : 340.0000
Stichprobenumfang n = 16

Literatur: Werner (1992) Medizinische Statistik. Urban & Schwarzenbeck Mü.

Einstichproben-t-Test: Machin und Campbell (1987) geben in Tabellen Beispiele zur Fallzahlberechnung für den Einstichproben-t-Test an:

Einstichproben-t-Test

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000
Schätzung für Varianz σ^2 : 1.5000
Minimale relevante Differenz δ : 0.7000
Stichprobenumfang n = 35 (Zweiseitige Fragestellung)

Literatur:

Mace (1962) Sample Size Determination. Chapman & Hall London.

Machin & Campbell (1987) Statistical Tables. Blackwell Scien. Publ.

Zweistichproben-t-Test: Machin und Campbell (1987) geben in Tabellen Beispiele zur Fallzahlberechnung für den Einstichproben-t-Test an, weitere Beispiele finden sich in Werner (1984, p. 247):

Zweistichproben-t-Test

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000
Schätzung für die Varianz σ^2 : 4.0000
Minimale relevante Differenz δ : 2.0000

Stichprobenumfang $n_1 = n_2 = 23$ (Zweiseitige Fragestellung)

Literatur:

Mace (1962) Sample Size Determination. Chapman & Hall London.

Machin & Campbell (1987) Statistical Tables. Blackwell Scien. Publ.

Werner (1992) Medizinische Statistik. Urban & Schwarzenbeck München.

Einstichproben-t-Test auf Äquivalenz: Dazu wurden Vergleiche mit dem Programm "N" und Ergebnissen des Tests auf Unterschied angestellt; die Berechnung erfolgt für zweiseitige Äquivalenz.

Delta ist die maximale, medizinisch irrelevante Differenz zwischen den beiden fraglichen Gruppen.

Bitte beachten Sie, dass für die *einseitige* Fragestellung die gleiche Berechnung vorgenommen wird wie für die *einseitige* Fragestellung "Unterschied", das heißt, Sie geben im Fenster für die *zweiseitige* Fragestellung "Unterschied" einfach an Stelle von alpha den Wert $2 \cdot \alpha$ ein.

Einstichproben-t-Test auf Äquivalenz

Gewünschte Irrtumswahrscheinlichkeit alpha : 0.0500
Gewünschte Irrtumswahrscheinlichkeit beta : 0.1000
Schätzung für die Varianz σ^2 : 12.0000
Minimale relevante Differenz δ_0 : 1.0000

Stichprobenumfang $n = 132$ (Zweiseitige Fragestellung)

Literatur:

Mace (1962) Sample Size Determination. Chapman & Hall London.
Machin & Campbell (1987) Statistical Tables. Blackwell Scien. Publ.

Zweistichproben-t-Test auf Äquivalenz: Dazu wurden Vergleiche mit dem Programm "N" und den Ergebnissen des Tests auf Unterschied angestellt; die Berechnung wird für zweiseitige Äquivalenz durchgeführt. Delta ist die maximale, medizinisch irrelevante Differenz zwischen den beiden fraglichen Gruppen.

Bitte beachten Sie, dass für die *einseitige* Fragestellung die gleiche Berechnung vorgenommen wird wie für die *einseitige* Fragestellung "Unterschied", das heißt, Sie geben im Fenster für die *zweiseitige* Fragestellung "Unterschied" einfach an Stelle von alpha den Wert $2 \cdot \alpha$ ein.

Zweistichproben-t-Test auf Äquivalenz

Gewünschte Irrtumswahrscheinlichkeit alpha : 0.0500
Gewünschte Irrtumswahrscheinlichkeit beta : 0.1000
Schätzung für die Varianz σ^2 : 12.0000
Minimale relevante Differenz δ_0 : 2.0000
Stichprobenumfang $n_1 = n_2 = 66$ (Zweiseitige Fragestellung)

Literatur:

Mace (1962) Sample Size Determination. Chapman & Hall London.
Machin & Campbell (1987) Statistical Tables. Blackwell Scien. Publ.
Werner (1992) Medizinische Statistik. Urban & Schwarzenbeck München.

Einweg-Varianzanalyse: Werner (1984, p.247) gibt Beispiele zur Fallzahlschätzung für die Varianzanalyse. Zur Kontrolle der Fallzahlschätzungen für die Nachtests (Studentized Range) wurden Beispiele nach Abt (1987) per Hand gerechnet.

Einweg-Varianzanalyse

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000
Schätzung für die gemeinsame Varianz σ^2 : 4.0000
Anzahl der Gruppen (Stichproben) k : 7
Minimale relevante Spannweite δ der μ_1, \dots, μ_n : 2.0000

F-Test der Varianzanalyse:

Stichprobenumfang $n_1=n_2= \dots =n_k = 36$
Studentized Range (Nachtests!):
Stichprobenumfang $n_1=n_2= \dots =n_k = 52$

Literatur:

Mace (1962) Sample Size Determination. Chapman & Hall London.
Abt (1987) Kurs 'Versuchsplanung' Baseler Biometrische Region.

Cross-Over-Analyse: Für die Cross-Over-Situation wurden Per-Hand-Berechnungen gemäß der Ergebnisse von Julious (2004) bzw. Zimmermann und Rahlfs (1980) durchgeführt:

Cross-Over-Analyse

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000
Schätzung für die Varianz σ^2 : 4.0000
Minimale relevante Differenz δ : 1.0000
Korrelationskoeffizient der Behandlungen ρ : 0.8000

Stichprobenumfang $n_1 = n_2 = 9$ (Zweiseitige Fragestellung)

Literatur:

Mace (1962) Sample Size Determination. Chapman & Hall London.

Julious, S.A. (2004) Sample sizes for clinical trials with normal data. Statistics in Medicine 23, pp. 1921-86.

Pearson's Korrelationskoeffizient r: Machin und Campbell (1987) geben Tabellen zur Fallzahlschätzung für den Korrelationskoeffizienten.

Pearson's Korrelationskoeffizient r

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000

Nachzuweisender Korrelationskoeffizient r : 0.8000

Stichprobenumfang $n = 12$ (Zweiseitige Fragestellung)

Literatur: Machin & Campbell (1987) Stat. Tables. Blackw. Scien. Publ.

Vergleich zweier Korrelationskoeffizienten: Die Ergebnisse wurden durch per-Hand-Berechnungen und Angaben aus Machin und Campbell (1987) verifiziert.

Vergleich zweier Korrelationskoeffizienten

2-seitige Irrtumswahrscheinlichkeit α = 0.0500
Irrtumswahrscheinlichkeit β = 0.1000
Korrelationskoeffizient r_1 = 0.2000
Korrelationskoeffizient r_2 = 0.6000
Stichprobenumfänge $n_1 = n_2$ = 91.0000

Literatur: Machin & Campbell (1987) Statistical Tables. Blackwell Scien. Publ.

Kendall's tau: Die Ergebnisse wurden durch per-Hand-Berechnungen und durch Beispiele aus Hollander und Wolfe (1999) verifiziert.

Kendall's tau

2-seitige Irrtumswahrscheinlichkeit α = 0.0500
Irrtumswahrscheinlichkeit β = 0.1000
"Nachzuweisender" Korrelationskoeffizient τ = 0.4000
Stichprobenumfang n = 30.0000

Literatur:

G.E. Noether (1987) Sample Size Determination for Some Common Nonparametric Tests.

J. Amer. Stat. Ass. Vol. 82, No. 398, pp. 645-7.

Hollander & Wolfe (1999) Nonparametric Statistical Methods, Wiley.

Spearman's rho: Die Ergebnisse wurden durch per-Hand-Berechnungen und gemäß Hollander und Wolfe (1999) verifiziert.

Spearman's rho ▲

2-seitige Irrtumswahrscheinlichkeit alpha = 0.0500
Irrtumswahrscheinlichkeit beta = 0.1000
"Nachzuweisender" Korrelationskoeffizient r = 0.5000
Stichprobenumfang (per Pitman-Effizienz): n = 41.0000
Stichprobenumfang (asympt. per Gauß-Test): n = 44.0000

Literatur: Hollander & Wolfe (1999) Nonparametric Statistical Methods, Wiley.

Regressionskoeffizient β : Die Verifikation erfolgte mit dem Programm PASS. F.Y. Hsieh, D.A. Bloch, M.D. Larsen (1998) geben Beispiele für die Multiple Regression.

Regressionskoeffizient β ▲

Gewünschte Irrtumswahrscheinlichkeit alpha: 0.0500
Gewünschte Irrtumswahrscheinlichkeit beta: 0.1000

Standardabweichung Sigma(X): 0.1000
Standardabweichung Sigma(Y): 0.2000
'Nachzuweisender' Regressionskoeffizient β : 0.5000

Stichprobenumfang n = 160 (Zweiseitige Fragestellung)

Literatur:

F.Y. Hsieh, D.A. Bloch, M.D. Larsen (1998) A Simple Method of Sample Size Calculation for Linear and Logistic Regression. Statistics in Medicine Vol. 17, 14, pp. 1623-1634.

Logistische Regression: Die Berechnung unterscheidet sich in Abhängigkeit von der (zunächst nur einen) Kovariablen:

Bei quantitativen Kovariablen X ist die Eventrate am Mittelwert von X und das Odds-Ratio $OR = \exp(\beta)$ erforderlich, β ist die Änderung des Log-Odds pro Einheit von X. Bei binären Kovariablen ist der Anteil Events bei jeweils X=0 und X=1 erforderlich, außerdem der geschätzte Anteil Merkmalsträger mit X=1 bzw. X=0.

Im Falle von einer Multiplen Logistischen Regression mit p Kovariablen berechnet man für die relevante die Fallzahl für den Test der partiellen Koeffizienten. Zusätzlich ist dazu die Angabe des (geschätzten) Multiplen Korrelationskoeffizienten erforderlich.

Alle Berechnungen wurden verifiziert mit Hilfe von Beispielen aus Hsieh, Bloch und Larsen (1998) Tabelle I und II.

Logistische Regression: Quantitative Kovariable ▲

2-seitige Irrtumswahrscheinlichkeit alpha: 0.0500
Irrtumswahrscheinlichkeit beta: 0.0500
Nachzuweisende Effektgröße Odds-Ratio OR: 1.5000
Äquivalent dazu: Effektgröße $\beta = \text{Log}(OR)$: 0.4055
Eventrate bei Mittelwert der Kovariablen: 0.4000
Stichprobenumfang bei einer Kovariablen: n = 330.0000

2-seitige Irrtumswahrscheinlichkeit alpha: 0.0500
Irrtumswahrscheinlichkeit beta: 0.0500
Nachzuweisende Effektgröße Odds-Ratio OR: 1.5000
Äquivalent dazu: Effektgröße $\beta = \text{Log}(OR)$: 0.4055

Eventrate bei Mittelwert der Kovariablen: 0.4000
Multipler Korrelationskoeffizient $R(X,23_p)$ = 0.5000
Stichprobenumfang bei p Kovariablen: n = 440.0000

Literatur: F.Y. Hsieh, D.A. Bloch, M.D. Larsen (1998) A Simple Method of Sample Size Calculation for Linear and Logistic Regression. Statistics in Medicine Vol. 17, 14, pp. 1623-1634.

Logistische Regression: Binäre Kovariable ▲

2-seitige Irrtumswahrscheinlichkeit alpha: 0.0500
Irrtumswahrscheinlichkeit beta: 0.0500
Anteil Merkmalsträger mit Kovariable X=1: A1 = 0.2000
Eventrate bei X=0: $ER\{X=0\}$ = 0.0500
Eventrate bei X=1: $ER\{X=1\}$ = 0.1000
Stichprobenumfang bei einer Kovariablen: n = 2186.0000

2-seitige Irrtumswahrscheinlichkeit alpha: 0.0500
Irrtumswahrscheinlichkeit beta: 0.0500
Anteil Merkmalsträger mit Kovariable X=1: A1 = 0.2000
Eventrate bei X=0: $ER\{X=0\}$ = 0.0500
Eventrate bei X=1: $ER\{X=1\}$ = 0.1000
Multipler Korrelationskoeffizient $R(X,23_p)$ = 0.1000
Stichprobenumfang bei p Kovariablen: n = 2208.0000

Literatur: F.Y. Hsieh, D.A. Bloch, M.D. Larsen (1998) A Simple Method of Sample Size Calculation for Linear and Logistic Regression. Statistics in Medicine Vol. 17, 14, pp. 1623-1634.

Schätzung einer Wahrscheinlichkeit: Mace (1962, p.63) macht Angaben zur Fallzahl-schätzung für ein Konfidenzintervall für eine Wahrscheinlichkeit (Beispiel Mace p.63: Mace errechnet $n=258.6$)

Schätzung einer Wahrscheinlichkeit ▲

Gewünschte Irrtumswahrscheinlichkeit α : 0.1000
Gewünschte Länge des Intervalls L : 0.1000
Schätzung der Wahrscheinlichkeit p : 0.6000

Stichprobenumfang n = 259

Literatur:

Mace (1962) Sample Size Determination. Chapman & Hall London.

Test eines binomialen Anteils: Im Verlaufe einer Virusepidemie soll geprüft werden, ob sich die Erfolgsrate einer symptomatischen Behandlung von bisher 0.90 (Ho!) auf 0.85 (HA!) verschlechtert hat.

Dazu wählt man $\alpha=0.10$ und $\beta=0.10$. (Daten zur Validierung nach Lemeshow et al. (1990), die früher verwendete Berechnung nach Mace (1962) ergibt $n=285$, Abweichung ergeben sich nur bei sehr kleinen und sehr großen Werten von θ_0 .)

Test eines binomialen Anteils ▲

Gewünschte Irrtumswahrscheinlichkeit α : 0.1000
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000
Angenommene Wahrscheinlichkeit θ_0 : 0.9000

Angenommene Wahrscheinlichkeit θ_a : 0.8500
Minimale relevante Differenz $\delta_o = \theta_a - \theta_o$: -0.0500

Stichprobenumfang $n = 284$ (Einseitige Fragestellung)

Literatur: Lemeshow, S., D.W. Hosmer, J. Klar, S.K. Lwanga (1990) Adequacy of Sample Size in Health Studies. John Wiley Chichester New York

Vergleich binomialverteilter Größen: Schouten (1982) gibt Tabellen zur Fallzahlschätzung $n=n_1=n_2$ für binomialverteilte Größen an. Bei gewünschten Fallzahlen $n_1 < n_2$ wird eine modifizierte Methode von Schouten verwendet (weiter unten):

Vergleich binomialverteilter Größen: $n_1 = n_2$ ▲

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000

Angenommene Wahrscheinlichkeit θ_1 : 0.1000
Angenommene Wahrscheinlichkeit θ_2 : 0.3000
Minimale relevante Differenz $\delta_o = \theta_2 - \theta_1$: 0.2000
Erwartete Länge des Konfidenzintervalls: $L = 0.2532$

Stichprobenumfang $n_1 = n_2 = 87$ (Zweiseitige Fragestellung)

Bei gewünschten Fallzahlen $n_1 < n_2$ und bei gleichen Parametern α , δ wie oben, aber für $n_2 = 2 \cdot n_1$ ergeben sich die beiden Fallzahlen $n_1 = 65$ und $n_2 = 130$, also mit $65 + 130 = 185$ "Fällen" ein um insgesamt 11 größerer Stichprobenumfang als bei gleichen Fallzahlen $n_1 = n_2$:

Vergleich binomialverteilter Größen: $n_1 < n_2$

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000

Angenommene Wahrscheinlichkeit θ_1 : 0.1000
Angenommene Wahrscheinlichkeit θ_2 : 0.3000
Minimale relevante Differenz δ_0 : 0.2000
Gewünschtes Verhältnis $k = n_2/n_1 > 1$: 2.0000

Stichprobenumfänge $n_1 = 65$, $n_2 = 130$ (Zweiseitige Fragestellung)

Literatur:

Schouten H.J.A. (1982) Comparing two independent binomial proportions by a modified chi squared test - required sample sizes. Biom.J. 24, 1, pp. 93-6

Casagrande JT et al. (1978) An improved approximate formula for calculating sample sizes for comparing two binomial distributions. Biometrics 34, pp. 483-486

Bristol DR (1989) Sample sizes for constructing confidence intervals and testing hypothesis. Stat. in Med. 8, 7, pp. 803-11

Äquivalenz binomialverteilter Größen: Blackwelder (1982) gibt dazu ausführliche Tabellen an, die zur Validierung dienen können.

Äquivalenz binomialverteilter Größen ▲

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000

Angenommene Wahrscheinlichkeit θ_1 : 0.1000
Angenommene Wahrscheinlichkeit θ_2 : 0.2000

Maximale zulässige Differenz δ gemäß H_0 : 0.0500

Stichprobenumfang $n_1 = n_2 = 96$ (Einseitige Fragestellung)
Stichprobenumfang $n_1 = n_2 = 117$ (Zweiseitige Fragestellung)

Literatur:

W.C. Blackwelder (1982) 'Proving the Null Hypothesis' in Clinical Trials. Controlled Clinical Trials 3, pp. 345-355.

χ^2 -Vierfeldertafel-Test: Schouten (1982), Casagrande et al. (1978) geben Tabellen und Beispiele zur Fallzahlschätzung an. Zur erwarteten Länge des Konfidenzintervalls vgl. Bristol (1989):

Chi²-Vierfelder-Test ▲

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000

Angenommene Wahrscheinlichkeit θ_1 : 0.1000
Angenommene Wahrscheinlichkeit θ_2 : 0.3000
Minimale relevante Differenz $\delta_0 = \theta_2 - \theta_1$: 0.2000

Erwartete Länge des Konfidenzintervalls: $L = 0.2456$
Stichprobenumfang $n_1 = n_2 = 92$ (Zweis. Fragestellung, mit Yates-Korrektur.)

Literatur:

Schouten H.J.A. (1982) Comparing two independent binomial proportions by a modified chi squared test - required sample sizes. Biom.J. 24, 1, pp. 93-6
Casagrande JT et al. (1978) An improved approximate formula for calculating sample sizes for comparing two binom. distrib.. Biometrics 34, pp. 483-486
Bristol DR (1989) Sample sizes for constructing confidence intervals and testing hypothesis. Stat. in Med. 8, 7, pp. 803-11

Vorzeichen-Test: Bei der Planung einer Studie zum Erfolg einer krankengymnastischen Behandlung wird ein einseitiges Testniveau von $\alpha=0.10$ und ein Risiko $\beta=0.10$ festgelegt. Falls in mehr als 75% aller Fälle eine Verbesserung zu erwarten ist, möchte man dies durch den Vorzeichen-Test zeigen können (Daten aus Noether (1987)). Beachten Sie bitte, dass BiAS. zweiseitige Werte von α erwartet, so dass bei der Eingabe der Wert $\alpha'=2\alpha=0.20$ verwendet werden muss!

Vorzeichen-Test ▲

Gewünschte Irrtumswahrscheinlichkeit α : 0.2000
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000
Testrelevante Wahrscheinlichkeit $P(X>0)$: 0.7500
Stichprobenumfang $n = 27$ (approximativ)

G.E. Noether (1987) Sample Size Determination for Some Common Non-parametric Tests. J. Amer. Stat. Ass. Vol. 82, No. 398, pp. 645-7.

McNemar-Test: In einer ophthalmologisch-operativen Studie werden Katarakt-Patienten auf einem Auge nach Methode A, auf dem anderen Auge nach Methode B operiert. Bei der Planung der Studie wird ein zweiseitiges Testniveau von $\alpha=0.05$ und ein Risiko $\beta=0.10$ festgelegt. Es wird vermutet, dass das Ergebnis "A positiv, B negativ" in 0,03=3% der Fälle auftritt, "A negativ, B positiv" dagegen in 0,16=16% der Fälle. Damit ist eine Fallzahl von $n=100$ (entsprechend $d=19$ diskordanter Fälle) erforderlich. Bei einseitigen Fragestellungen verwendet man $\alpha'=2\cdot\alpha$:

McNemar-Test

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000
P(Bedingung A positiv, B negativ): 0.0300
P(Bedingung A negativ, B positiv): 0.1600

Zweiseitige Fragestellung:

Erforderliche Anzahl diskordanter Paare: $d = 19$
Stichprobenumfang ("average") $n = 100$

Literatur:

JE Connet, JA Smith, RB McHugh (1987) Sample size and power for paired-matched case-control-studies. Stat. in Med. 6, pp. 53-9

Cohen's Kappa: Es soll die $H_0: \text{Kappa}=0.6$ gegen $H_1: \text{Kappa}=0.9$ getestet werden (0.6 bedeutet nach Landis und Koch (1977, Biometrics 33, p. 159ff) eine „substantielle“ Übereinstimmung). Von den insgesamt $2n$ Ratings (bei 2 Ratern!) wird ein Anteil von $\pi=0.30$ als „positiv“ angenommen. Damit ergibt sich eine Fallzahl von $n=66$; Varifikation gemäß Donner und Eliasziw (1992 p. 1518).

Cohen's Kappa-Koeffizient

2-seitige Irrtumswahrscheinlichkeit $\alpha =$ 0.0500
Irrtumswahrscheinlichkeit $\beta =$ 0.2000
Kappa der Nullhypothese $H_0: \text{Kappa}_0 =$ 0.6000
Nachzuweisendes Kappa der $H_a: \text{Kappa}_1 =$ 0.9000
Anteil positiver Beurteilungen gesamt: $\pi =$ 0.3000
Stichprobenumfang $n =$ 66.0000

Literatur:

A. Donner, M. Eliasziw (1992) A goodness-of-fit approach to inference procedures for the kappa statistic. Stat. in Med. Vol. 11, pp. 1511-1519.

Wilcoxon-matched-pairs-Test: Fragestellung wie im Beispiel zum Vorzeichen-Test. Die Angabe der Wahrscheinlichkeit $P(X+X'>0)$ ist oft unklar, so dass näherungsweise konservativ $P(X>0)$ eingesetzt werden kann oder aber günstiger zur Planung der Vorzeichen-Test verwendet wird.

Wilcoxon-matched-pairs-Test

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000
Testrelevante Wahrscheinlichkeit $P(X+X'>0)$: 0.7500

Stichprobenumfang $n = 57$ (approximativ)

Literatur:

G.E. Noether (1987) Sample Size Determination for Some Common Nonparametric Tests. J. Amer. Stat. Ass. Vol. 82, No. 398, pp. 645-7.

Wilcoxon-Mann-Whitney-U-Test: Vergleich zweier Therapien mit Hilfe von zwei unabhängigen Gruppen von Patienten. Es wird erhofft, dass Therapie (X) nur mit einer Wahrscheinlichkeit von $1/3$ ungünstiger ist als Therapie (Y) bzw. dass Therapie (X) mit einer Wahrscheinlichkeit 0.67 günstiger ist als Therapie (Y). Validierung mit Daten aus Noether (1987):

Wilcoxon-Mann-Whitney-Test



Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.2000
Testrelevante Wahrscheinlichkeit $P(X<Y)$: 0.3300
Stichprobenumfang $n1 = n2 = 46$ (approximativ)

Literatur:

G.E. Noether (1987) Sample Size Determination for Some Common Nonparametric Tests. J. Amer. Stat. Ass. Vol. 82, No. 398, pp. 645-7.

Achtung: Für Shift-Alternativen kann Fan's verallgemeinerte Fallzahlberechnung zum Kruskal-Wallis-Test verwendet werden. Bei vergleichenden Berechnungen auf Grundlage des Mann-Whitney-Schätzers können minimale Abweichungen der Noether'schen von Fan's Methode auftreten, da Noether eine Gauß-Approximation und Fan die nichtzentrale χ^2 -Verteilung verwendet. Entsprechendes gilt für die Power-Berechnung.

Kruskal-Wallis-Test: Fan et al. (2010) geben ein konkretes Beispiel zur Berechnung, das unten verwendet wurde. Die erforderlichen paarweisen Mann-Whitney-Schätzer können anhand der Fan-Daten mit Hilfe des U-Tests ermittelt werden; Beispiele zur direkten Eingabe per Pilot-Datei und zu den „Shift-Alternativen“ werden hier nicht angegeben. Da das Beispiel aus Fan et al. auf einer relativ kleinen Pilot-Studie mit $n=40$ beruht, sollte - wie im Programm vorgesehen - eine Fallzahl-Adjustierung erfolgen. Die Daten zu dem folgenden Beispiel stehen in der Datei ..\Valide\Fan.DAT zur Verfügung:

Kruskal-Wallis-Test



Irrtumswahrscheinlichkeit alpha : 0.0500
Irrtumswahrscheinlichkeit beta : 0.2000
Anzahl der Gruppen (Stichproben) k : 3.0000
Mann-Whitney-Schätzer Gruppe 1 versus 2: 0.6788
Mann-Whitney-Schätzer Gruppe 1 versus 3: 0.3929
Mann-Whitney-Schätzer Gruppe 2 versus 3: 0.2952
Nicht-adjustierte Fallzahlen:
Stichprobenumfang $n1=n2=...=nk =$ 29.0000
Stichprobenumfang gesamt: $n=Summe(ni) =$ 87.0000
Adjustierte Fallzahlen für $N_{pilot} = 40$:
Stichprobenumfang $n1=n2=...=nk =$ 36.0000
Stichprobenumfang gesamt: $n=Summe(ni) =$ 108.0000

Literatur:

C. Fan, D. Zhang, C.-H. Zhang (2011) On Sample Size of the Kruskal-Wallis-Test with Application to a Mouse Peritoneal Cavity Study. Biometrics 67, 1, pp. 213-24.

Achtung: Die generischen und die adjustiert-generischen Fallzahlen aus Fan et al. können beide mit BiAS. reproduziert werden. Die Resultate für die Shift-Hypothesen aus Fan et al. können nicht exakt verifiziert werden, vermutlich deshalb, da in dem Paper (Tabelle 5) die Berechnungen mit offensichtlich unzureichender Genauigkeit und der Median der Gruppe 3 fehlerhaft angegeben sind: Die Fallzahlen in BiAS. sind geringfügig größer. Alle Berechnungen zu den Shift-Alternativen wurden mit zwei weiteren Beispielen (beide sind ebenfalls in der Datei ..\Valide\Fan.DAT enthalten) durch per-Hand-Berechnungen, MatLab und durch die generischen Fallzahlen der Fan-MWS-Methode verifiziert. Kommentierte Details dazu finden sich in der Datei ..\Valide\Fan.DAT.

Log-Rank-Test: Machin und Campbell (1987, pp.98ff,115ff, Tab. 9.2) geben Beispiele zur Fallzahlschätzung für Log-Rank-Tests:

Log-Rank-Test

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000
Angenommene Überlebensrate θ_1 : 0.1000
Angenommene Überlebensrate θ_2 : 0.3000
Minimale relevante Differenz $\delta_0 = \theta_2 - \theta_1$: 0.2000
Angenommener Relativer Hazard RH(1:2) : 1.9125
Withdrawal-Rate = 0.0000: Stichprobenumfang n1 = n2 = 67 (65)
Withdrawal-Rate = 0.1000: Stichprobenumfang n1 = n2 = 75 (72)
Erforderliche Mindestzahl 'Ereignisse' (Tod o.ä.) e = 108 (103)
Die Ergebnisse beziehen sich auf eine zweiseitige Fragestellung!
Die Angaben in Klammern beziehen sich auf die exakte Varianz des Log-Rank-Tests und sind stets günstiger als Freedman's Tabellen.

Literatur:

L.S. Freedman (1982) Tables of the number of patients required in clin. trials using the logrank test. Stat. in Med. Vol.1,2, pp.121-30
D. Machin, M.J. Campbell (1987) Statistical tables for the design of clinical trials. Blackwell Sci. Publ.

Kaplan-Meier-Schätzer - Test von $H_0(P|t)$: Die Fallzahlberechnungen wurden anhand der unten zitierten Publikation von Nagashima (2016) validiert. Im Modul sind verschiedene Transformationen (log-log (SAS), log (Kumul. Harard, CRAB), logit, ArcSinWurzel und untransformiert) möglich.

Fallzahlen: Einarmiger Test von $H_0(P|t)$

Irrtumswahrscheinlichkeit alpha = 0.0500
Irrtumswahrscheinlichkeit beta = 0.2000
Nullhypothese: Survival-Rate = 0.4000
Alternativhypothese: Survival-Rate = 0.6000
Rekrutierungsdauer = 2.0000
Follow-up-Dauer = 4.0000
Zeitpunkt t der Analyse = 3.0000
Transformation - Methode 1: log-log (SAS/Lifetest)
Fallzahl einseitiger Test: n1 = 42.0000
Fallzahl zweiseitiger Test: n2 = 54.0000

Literatur und Erläuterungen:

Nagashima K. (2016) A sample size estimation tool for one sample nonparametric tests for a survival proportion [Internet 2016 Mar 21] nshi.jp/en/js/onesurvyr/

Cox-Regression: Durchgerechnetes Beispiel aus Wang und Chow (2007). Die Fallzahlen gelten für randomisierte Studien bzw. nach Kovariablen stratifizierte Studien, bei unbalancierten Kovariablen ist Vorsicht geboten:

Cox-Regression

2-seitige Irrtumswahrscheinlichkeit α : 0.0500
Irrtumswahrscheinlichkeit β : 0.2000
Angenommener Relativer Hazard RH(1:2) = 4.4817
Withdrawal-Rate = 0.0000: Stichprobenumfang n1 = n2 = 7
Withdrawal-Rate = 0.8000: Stichprobenumfang n1 = n2 = 35

Literatur:

H. Wang, S.C. Chow (2007) Sample Size Calculation for Comparing Time-to-Event Data. Wiley Encyclopedia of Clinical Trials, J.Wiley.
D.A. Schoenfeld (1983) Sample Size Formula for the Proportional-Hazards Regression Model. Biometrics 39, pp. 499-503.

Odds-Ratio (Epidemiologie): Sachs (1992, p. 316-8) bringt Beispiele zur Feststellung der Kollektivumfänge bei Odds-Ratios:

Odds-Ratio (Epidemiologie) ▲

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Mutmaßlicher Parameter : 2.0000
Angenommener Anteil Exponierter bei Kontrolle (%) : 30.0000
Gewünschte Genauigkeit Epsilon (%) : 10.0000

Stichprobenumfang für Odds Ratio: $n_1 = n_2 = 2027$

Literatur:

L.Sachs (1992) Angewandte Statistik Kap. 2, Springer.
S. Lemeshow et al. (1988) Sample size requirements for studies estimating odds ratios or relative risks. Stat. in Med. 7, pp. 759-746.

Relatives Risiko (Epidemiologie): Sachs (1992, p. 316-8) bringt Beispiele zur Feststellung der Kollektivumfänge beim Relativen Risiko:

Relatives Risiko (Epidemiologie) ▲

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Mutmaßlicher Parameter 1.7500
Angenommener Anteil Nicht-Exponierter mit Krankheit (%) 20.0000
Gewünschte Genauigkeit Epsilon (%) : 10.0000

Stichprobenumfang für Relatives Risiko: $n_1 = n_2 = 2027$

Literatur:

L.Sachs (1992) Angewandte Statistik Kap. 2, Springer.
S. Lemeshow et al. (1988) Sample size requirements for studies estimating odds ratios or relative risks. Stat. in Med. 7, pp. 759-746.

Many-one-Vergleiche (t-Test): Diese Situation wird bei Horn und Vollandt (1995) behandelt. Es können $k \geq 2$ Gruppen gewählt werden.

Many-One-Vergleiche (t-Test) ▲

Fallzahlberechnung für den Vergleich von $k=3$ Gruppen mit Kontrolle (0) auf Basis des Zweistichproben-t-Tests und Bonferroni-Adjustierung:

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
Gewünschte Irrtumswahrscheinlichkeit β : 0.1000

Schätzung der gemeinsamen Varianz σ^2 : 2.4000
Minimale relevante Differenz δ_0 : 1.1000

Gleiche Stichprobenumfänge (Zweiseitige Fragestellung):

Stichprobenumfang $n_0=n_1=\dots=n_k = 55$
Stichprobenumfang gesamt $n=n_0+\sum n_i = 220$

Optimal ungleiche Stichprobenumfänge (Zweiseitige Fragestellung):
 Stichprobenumfang Kontrolle: no = 76
 Stichprobenumfang Vergleiche: ni = 44
 Stichprobenumfang gesamt $n=no+\sum ni = 208$

VORSICHT: Im Test gilt für alle Vergleiche das adjustierte Niveau $\alpha^*=0.016667!$

Literatur:

Mace (1962) Sample Size Determination. Chapman & Hall London.
 Machin & Campbell (1987) Statistical Tables. Blackwell Scien. Publ.
 Werner (1992) Medizinische Statistik. Urban & Schwarzenbeck München.
 Horn & Vollandt (1995) Multiple Tests und Auswahlverfahren. Fischer.

Many-one-Vergleiche (Wilcoxon-Mann-Whitney-U-Test): Diese Situation wird bei Horn und Vollandt (1995) behandelt. Es können $k \geq 2$ Gruppen gewählt werden.

Many-One-Vergleiche (U-Test) ▲

Fallzahlberechnung für den Vergleich von $k=4$ Gruppen mit Kontrolle (0) auf Basis des Wilcoxon-Mann-Whitney-Tests und Bonferroni-Adjustierung:

Gewünschte Irrtumswahrscheinlichkeit α : 0.0500
 Gewünschte Irrtumswahrscheinlichkeit β : 0.1000

Testrelevante Wahrscheinlichkeit $P(X<Y)$: 0.3000

Gleiche Stichprobenumfänge (Zweiseitige Fragestellung):

Stichprobenumfang $no=n1=\dots=nk = 60$
 Stichprobenumfang gesamt $n=no+\sum ni = 300$

Optimal ungleiche Stichprobenumfänge (Zweiseitige Fragestellung):

Stichprobenumfang Kontrolle: no = 90
 Stichprobenumfang Vergleiche: ni = 45
 Stichprobenumfang gesamt $n=no+\sum ni = 270$

VORSICHT: Im Test gilt für alle Vergleiche das adjustierte Niveau $\alpha^*=0.012500!$

Literatur:

G.E. Noether (1987) Sample Size Determination for Some Common Nonparametric Tests. J. Amer. Stat. Ass. Vol. 82, No. 398, pp. 645-7
 M.Horn & R.-Vollandt (1995) Multiple Tests und Auswahlverf. Fischer

c×r-Kontingenztafel: Zur Verifikation wurde das Programm PASS verwendet (Hintze (2011)). Die Eingabe erfordert alternativ Cohen's Effektstärke W oder Cramér's V:

c×r-Kontingenztafel ▲

Irrtumswahrscheinlichkeit alpha = 0.050000
 Irrtumswahrscheinlichkeit beta = 0.200000
 Power = 0.800000
 Effektstärke: Cohen's W = 0.300000 << Eingabe Cohen's W
 Effektstärke: Cramér's V = 0.173205
 Anzahl Zeilen r = 5
 Anzahl Spalten c = 4
 Errechnete Fallzahl n = 193

```

Irrtumswahrscheinlichkeit alpha = 0.050000
Irrtumswahrscheinlichkeit beta = 0.200000
Power = 0.800000
Effektstärke: Cohen's W = 0.519615
Effektstärke: Cramér's V = 0.300000 << Eingabe Cramér's V
Anzahl Zeilen r = 5
Anzahl Spalten c = 4

Errechnete Fallzahl n = 65

```

Literatur:

J.L. Hintze (2011) PASS User's Guide p.630, NCSS, Kaysville, Utah.
 J. Cohen (1988) Statistical power analysis for the behavioral sciences.
 2nd edition Hillsdale, NJ: Erlbaum.

Bland-Altman-Methodenvergleich: Die Berechnung erfolgt über die maximale Spannweite (Range) der Differenzen, es ist $\text{Range}=2 \times (\text{Maximale Differenz})$. Die Formel sind analog zum Einstichproben-t-Test.

Bland-Altman-Methodenvergleich ▲

```

Gewünschte Irrtumswahrscheinlichkeit alpha : 0.0500
Gewünschte Irrtumswahrscheinlichkeit beta : 0.1000

Schätzung für die maximale Differenz: 6.0000
Minimale relevante mittlere Differenz delta0: 1.0000

```

Stichprobenumfang n = 97 (Zweiseitige Fragestellung)

```

Gewünschte Irrtumswahrscheinlichkeit alpha : 0.0500
Gewünschte Irrtumswahrscheinlichkeit beta : 0.1000

Schätzung für die maximale Differenz: 6.0000
Minimale relevante mittlere Differenz delta0: 1.0000

```

Stichprobenumfang n = 100 (Zweiseitige Fragestellung, Äquivalenztest)

Literatur:

Mace (1962) Sample Size Determination. Chapman & Hall London.
 Machin & Campbell (1987) Statistical Tables. Blackwell Scien. Publ.

Cluster-randomisierte Studien: Campbell et al. (2000) geben einen Korrekturfaktor an ("Design-Effekt"), so dass aus den Fallzahlen einer Patienten-randomisierten Studie, der Intracluster-Korrelation und der Fallzahl pro Cluster die Cluster-korrigierte Fallzahl errechnet werden kann. Die Validierung erfolgt mit Bleistift, Papier und Taschenrechner:

Cluster-randomisierte Studien ▲

```

Fallzahl der Patienten-randomisierten Studie: 340.0000
Intracluster-Korrelationskoeffizient ICC: 0.1000
Patienten/Fälle pro Cluster: 8.0000
-----
Design-Effekt D: 1.7000
Fallzahl der Cluster-randomisierten Studie: 578.0000

```

Literatur:

MK. Campbell, J. Mollison, N. Steen, JM Grimshaw, M. Eccles (2000) Analysis of cluster randomized trials in primary care: a practical approach. Family Care 17,2, pp.192-6

SM. Kerry, JM. Bland (1998) The intraclass correlation coefficient in cluster randomisation. British Medical Journal 316, pp. 1455-60.

Intraclass-Korrelationskoeffizient ICC: Die Fallzahlberechnungen wurden anhand der Arbeiten von Bonett (2002) und Zou (2012) validiert:

Intraclass-Korrelationskoeffizient ▲

Konfidenz P des Konfidenzintervalls (zweiseitig) =	0.9500
Gewünschte Länge des Konfidenzintervalls =	0.1000
Anzahl der Rater =	3.0000
Berechnung für ICC =	0.8000
Stichprobenumfang n =	140.0000

Literatur:

DG Bonett (2002) Sample size requirements for estimation intraclass correlations with desired precision. Stat. in Med. 21, pp. 1331-5.

GY Zou (2012) Sample size formulas for estimation intraclass correlation coefficients with precision and assurance. Stat. in Med. 31, pp. 3972-81.

ROC-Kurven und AUC: Die Fallzahlberechnungen wurden anhand der unten zitierten Publikationen von JA Hanley und BJ McNeil (1982,1983) validiert. Die Berechnungen der „gemeinsamen“ Korrelation bei abhängigen ROC-Kurven („paired“) erfolgen analog zu PASS (Hintze 2011).

Fallzahlen: Konfidenzintervall für eine AUC ▲

Gewünschte Konfidenz P des Intervalls: P =	0.9500
Geschätzter Parameter Theta: AUC =	0.7000
Gewünschte Länge des Konfidenzintervalls: L =	0.1000
Schätzung für die Standardabweichung sigma =	0.0255

Approximative Stichprobenumfänge: ng = nk =	208
Gesamt-Stichprobenumfang N = ng+nk =	416

Literatur:

JA Hanley, BJ McNeil (1982) The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. Radiology 148, 1, pp. 29-36

Fallzahlen: Test der Nullhypothese Ho(AUC=0.5) ▲

Einseitige Irrtumswahrscheinlichkeit alpha =	0.0500
Irrtumswahrscheinlichkeit beta =	0.2000
Geschätzter Wert der AUC: theta =	0.7000
Stichprobenumfänge (approximativ) ng = nk =	26
Gesamt-Stichprobenumfang N = ng+nk =	52

Literatur:

GE Noether (1987) Sample Size Determination for Some Common Nonparametric Tests. J. Amer. Stat. Ass. Vol. 82, No. 398, pp. 645-7.

JA Hanley, BJ McNeil (1982) The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. Radiology 148, 1, pp. 29-36.

Fallzahlen: Vergleich von zwei unabhängigen ROC-Kurven



Irrtumswahrscheinlichkeit alpha = 0.0500
Irrtumswahrscheinlichkeit beta = 0.2000
Angenommene Fläche unter der ROC1-Kurve: AUC1 = 0.7500
Angenommene Fläche unter der ROC2-Kurve: AUC2 = 0.8500
Minimale relevante Differenz delta=AUC2-AUC1 = 0.1000

Fallzahlen für den einseitigen Test: ng=nk = 137
Fallzahlen für den zweiseitigen Test: ng=nk = 175
>> Es sind ng "Gesunde" und nk "Kranke" erforderlich.
>> Die Fallzahlen gelten jeweils für beide ROC-Kurven!

Literatur:

JA Hanley, BJ McNeil (1982) The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. Radiology 148, 1, pp. 29-36

Fallzahlen: Vergleich von zwei abhängigen ROC-Kurven



Irrtumswahrscheinlichkeit alpha = 0.0500
Irrtumswahrscheinlichkeit beta = 0.2000
Angenommene Fläche unter der ROC1-Kurve: AUC1 = 0.7500
Angenommene Fläche unter der ROC2-Kurve: AUC2 = 0.8500
Minimale relevante Differenz delta=AUC2-AUC1 = 0.1000
Korrelation der Methoden bei "Gesunden": r = 0.5000
Korrelation der Methoden bei "Kranken": r = 0.7000

Fallzahlen für den einseitigen Test: ng=nk = 62
Fallzahlen für den zweiseitigen Test: ng=nk = 79
>> Es sind ng "Gesunde" und nk "Kranke" erforderlich.

Literatur:

JA Hanley, BJ McNeil (1982) The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. Radiol. 143 pp. 29-36
JA Hanley, BJ McNeil (1983) A method of comparing the areas under receiver operating charact. curves derived from the same cases. Radiol. 148 pp. 839-43

Auf eine explizite Darstellung bzw. Validierung der Power-Berechnungen wurde verzichtet, da diese sich in aller Regel komplementär aus den aufgeführten Fallzahlberechnungen ergeben.

A.24 Randomisierungen



Zufallsstichprobe (Gleichverteilung)
Vollständige Randomisierung
Randomisierte Blöcke
Cross-Over-Versuchspläne
Lateinische Quadrate
Gauß-verteilte Zufallszahlen (uni- und bivariat)

Zufallsstichproben: Der Stichprobenumfang und die zulässige kleinste bzw. größte Zahl sind im Programm wählbar. Bitte beachten Sie, dass jede Zufallszahl nur genau einmal vorkommen darf, so dass die kleinste und die größte Zahl in geeigneter Weise gewählt werden müssen!

Zufallsstichprobe

Stichprobenumfang n = 10
Kleinste Zahl: 1
Größte Zahl : 100
27 35 32 71 62 67 37 41 43 90

Vollständige Randomisierung: Der Stichprobenumfang (Anzahl Versuchseinheiten) und die Anzahl der Gruppen sind im Programm wählbar.

Vollständige Randomisierung

Versuchseinheiten: n = 10, Anzahl Gruppen = 2

Lfd Nr	Gruppe
1	1
2	1
3	1
4	2
5	2
6	1
7	2
8	1
9	2
10	2

Randomisierte Blöcke: Der Stichprobenumfang pro Block und die Anzahl Blöcke sind im Programm wählbar.

Randomisierte Blöcke

Blockumfang n = 4 ; Anzahl Blöcke = 5

Block 1:	1	3	4	2
Block 2:	4	3	2	1
Block 3:	2	3	1	4
Block 4:	1	4	3	2
Block 5:	4	2	1	3

Cross-Over-Versuchspläne: Der Stichprobenumfang („Versuchseinheiten“) und die Umfänge der Blöcke sind im Programm wählbar.

Cross-Over-Versuchspläne

Versuchseinheiten: n = 10
Umfang Gruppe 1: 5
Umfang Gruppe 2: 5

Lfde Nr.	Gruppe
1	AB
2	BA
3	AB
4	BA
5	BA
6	AB
7	AB
8	AB
9	BA
10	BA

Lateinische Quadrate: Die Ordnung des gewünschten Lateinischen Quadrats ist im Programm wählbar. Es wird das standardisierte lexikographische und ein randomisiertes Lateinisches Quadrat ausgegeben:

Lateinische Quadrate

Ordnung k = 5

Standardisiertes Lateinisches Quadrat:

A B C D E
B C D E A
C D E A B
D E A B C
E A B C D

Randomisiertes Lateinisches Quadrat:

E C D B A
B E A D C
C A B E D
D B C A E
A D E C B

A.25 Dichte- und Verteilungsfunktionen

Gauß-Verteilung

t-Verteilung

χ^2 -Verteilung

F-Verteilung

Binomial-Verteilung

Poisson-Verteilung

Nicht-zentrale t-Verteilung

Nicht-zentrale χ^2 -Verteilung

Nicht-zentrale F-Verteilung

Studentized Range

Es wird die Verteilungsfunktion und deren Inverse berechnet.

Gauß-verteilte Zufallszahlen (uni- und bivariat): Zur Erzeugung Gauß-verteilter Zufallszahlen können für beliebige Gruppennzahlen und beliebige Fallzahlen in den Gruppen jeweils Werte von μ und σ^2 vorgegeben werden. Auf ein wenig informatives Zahlenbeispiel wird aus Platzgründen verzichtet.

Die ermittelten Zufallszahlen werden zur weiteren Verwendung in eine (frei zu wählende) Datei abgespeichert.

Gauß-Verteilung: Zur Überprüfung der Berechnung von Verteilungsfunktionen und deren Inversen können beliebige einschlägige Tabellen herangezogen werden. Zur Validierung und zur Illustration der Eingabe werden im Folgenden nur zwei Beispiele angegeben.

Gauß-Verteilung

Irrtumswahrscheinlichkeit $\alpha = 0.050000$ (zweiseitig)
Kritischer Wert $G = 1.959964$

Gauß-Verteilung

Prüfgröße $N = 1.959964$
Überschreitungswahrscheinlichkeiten:
Einseitig: $p = 0.025000$
Zweiseitig: $2p = 0.050000$

t-Verteilung: Zur Überprüfung der Berechnung von Verteilungsfunktionen und deren Inversen können beliebige einschlägige Tabellen herangezogen werden. Zur Validierung und zur Illustration der Eingabe werden im Folgenden nur zwei Beispiele angegeben.

t-Verteilung

Irrtumswahrscheinlichkeit $\alpha = 0.010000$ (zweiseitig)
Freiheitsgrad $df = 10$
Kritischer Wert $t = 3.169269$

t-Verteilung

Prüfgröße $t = 3.169269$
Freiheitsgrad $df = 10$
Überschreitungswahrscheinlichkeiten:
Einseitig $p = 0.005000$
Zweiseitig: $2p = 0.010000$

Chi-Quadrat-Verteilung: Zur Überprüfung der Berechnung von Verteilungsfunktionen und deren Inversen können beliebige einschlägige Tabellen herangezogen werden. Zur Validierung und zur Illustration der Eingabe werden im Folgenden nur zwei Beispiele angegeben.

χ^2 -Verteilung

Irrtumswahrscheinlichkeit $2\alpha = 0.050000$
Freiheitsgrad = 1
Kritischer Wert $\chi^2 = 3.841459$

χ^2 -Verteilung

Prüfgröße $\chi^2 = 3.841459$
Freiheitsgrad = 1
Überschreitungswahrscheinlichkeit $p = 0.050000$

F-Verteilung: Zur Überprüfung der Berechnung von Verteilungsfunktionen und deren Inversen können beliebige einschlägige Tabellen herangezogen werden. Zur Validierung und zur Illustration der Eingabe werden im Folgenden nur zwei Beispiele angegeben.

F-Verteilung

Irrtumswahrscheinlichkeit $\alpha = 0.010000$ (zweiseitig)
Freiheitsgrade $df1 = 10$ und $df2 = 20$
Kritischer Wert $F = 3.368182$

F-Verteilung

Prüfgröße $F = 3.368182$
Freiheitsgrade $df1 = 10$ und $df2 = 20$
Überschreitungswahrscheinlichkeit $p = 0.010000$

Binomial-Verteilung: Zur Überprüfung der Berechnung von Verteilungsfunktionen und deren Inversen können beliebige einschlägige Tabellen herangezogen werden. Zur Validierung und zur Illustration der Eingabe wird im Folgenden nur ein Beispiel angegeben.

Binomial-Verteilung

Anzahl Versuche $n = 20$
Anzahl Erfolge $k = 5$
Erfolgswahrscheinlichkeit $p = 0.600000$
 $P(k \leq 5) = 0.001612$
 $P(k \geq 5) = 0.999683$
 $P(k = 5) = 0.001294$

Poisson-Verteilung: Zur Überprüfung der Berechnung von Verteilungsfunktionen und deren Inversen können beliebige einschlägige Tabellen herangezogen werden. Zur Validierung und zur Illustration der Eingabe wird im Folgenden nur ein Beispiel angegeben.

Poisson-Verteilung

Erwartungswert $\lambda = 10.00000$

Anzahl Ereignisse $k = 3$

$P(k \leq 3) = 0.010336$

$P(k = 3) = 0.007567$

$P(k \geq 3) = 0.997231$

Nicht-zentrale t-Verteilung: Zur Überprüfung der Berechnung von Verteilungsfunktionen und deren Inversen können beliebige einschlägige Tabellen herangezogen werden. Zur Validierung und zur Illustration der Eingabe werden im Folgenden nur zwei Beispiele angegeben.

Nicht-zentrale t-Verteilung

Irrtumswahrscheinlichkeit $\alpha = 0.050000$ (zweiseitig)

Freiheitsgrad $df = 10$

NC-Parameter $c = 0.5000$

Kritischer Wert $t = 2.482953$

Nicht-zentrale t-Verteilung

Prüfgröße $t = 2.482953$

Freiheitsgrad $df = 10$

NC-Parameter $c = 0.5000$

Überschreitungswahrscheinlichkeiten:

Einseitig: $p = 0.025000$

Zweiseitig: $2p = 0.050000$

Nicht-zentrale Chi-Quadrat-Verteilung: Zur Überprüfung der Berechnung von Verteilungsfunktionen und deren Inversen können beliebige einschlägige Tabellen herangezogen werden. Zur Validierung und zur Illustration der Eingabe werden im Folgenden nur zwei Beispiele angegeben.

Nicht-zentrale χ^2 -Verteilung

Irrtumswahrscheinlichkeit $\alpha = 0.050000$ (zweiseitig)

Freiheitsgrad = 10

NC-Parameter $c = 0.800000$

Kritischer Wert $\chi^2 = 19.742844$

Nicht-zentrale χ^2 -Verteilung

Prüfgröße $\chi^2 = 19.742844$

Freiheitsgrad = 10

NC-Parameter $c = 0.800000$

Überschreitungswahrscheinlichkeit $p = 0.050000$

Nicht-zentrale F-Verteilung: Zur Überprüfung der Berechnung von Verteilungsfunktionen und deren Inversen können beliebige einschlägige Tabellen herangezogen werden. Zur Validierung und zur Illustration der Eingabe werden im Folgenden nur zwei Beispiele angegeben.

Nicht-zentrale F-Verteilung

Irrtumwahrscheinlichkeit $\alpha = 0.100000$ (zweiseitig)
Freiheitsgrade $df1 = 10$ und $df2 = 20$
NC-Parameter $c = 0.100000$
Kritischer Wert $F = 1.956084$

Nicht-zentrale F-Verteilung

Prüfgröße $F = 1.956084$
Freiheitsgrade $df1 = 10$ und $df2 = 20$
NC-Parameter $c = 0.100000$
Überschreitungswahrscheinlichkeit $p = 0.100000$

Studentized Range: Nach Angaben von Lund et al. (1983) stimmen die Resultate für $df > 2$ i.d.R. auf mindestens 4 Stellen mit den Standard-Tabellen aus Harter et al. (1959) überein, für $df < 3$ ist die Genauigkeit gelegentlich geringfügig geringer. Die Berechnung der p-Werte entspricht Lund et al. (1983) JRSA Applied Statistics, Vol. 32 No. 2, pp. 204-10. Eine Referenz zur Validierung ist Harter et al. (1959) Techn. Rep. Dayton WADC.

Studentized Range

Studentized Range: $R = 5.00000$
Anzahl Gruppen: $k = 5$
Freiheitsgrad: $df = 12$
p-Wert: $p = 0.027670$

Literatur:

Harter, H.L., D. Clemm, E.H. Guthrie (1959) The probability integrals of the range and of studentized range. WADC Technical Report 58-484 Vol. II, Dayton, Ohio, Wright Air Craft Development Center
Lund, R.E., J.R. Lund (1983) Probabilities and upper quantiles for the Studentized Range. Applied Statistics, Vol.32 No.2, 1983, pp. 204-210

Literatur

Die Grundzüge der Biometrie können in den einführenden Lehrbüchern

- Hanns Ackermann (2014) Medizinische Biometrie, 7. Auflage epsilon Darmstadt
- Rolf J. Lorenz (1996) Grundbegriffe der Biometrie, 4. Auflage Fischer Stuttgart

nachgelesen werden. Ein umfassender, auch an Mediziner und Sozialwissenschaftler gerichteter Überblick über aktuelle statistische Methoden und zahllose Beispiele finden sich in den anwendungsorientierten Werken von

- Jürgen Bortz (2005) Lehrbuch der Statistik, 6. Auflage Springer-Verlag Heidelberg
- Irene Guggenmoos-Holzmann, K. Wernecke (1995) Med. Statistik, Springer-Verlag
- Joachim Hartung (1982) Statistik, Oldenburg-Verlag München Wien
- Ralf-Dieter Hilgers et al. (2003) Einführung in die Med. Statistik, Springer-Verlag
- Lothar Sachs (2003) Angewandte Statistik, 11. Aufl. Springer-Verlag Heidelberg *)
- Martin Schumacher, Gabi Schulgen (2002) Methodik klin. Studien, Springer-Verlag
- Christel Weiß (2010) Basiswissen Medizinische Statistik, 5. Aufl. Springer-Verlag

Das als "Praxisratgeber für Kliniker und Ärzte" konzipierte Buch

- Andreas J. W. Goldschmidt (Hrsg. 1997) Medizinische Statistik und Klinische Forschung - von der Idee zum Ergebnis, Springer-Verlag Heidelberg

ist ganz der praktischen Seite der Biometrie gewidmet und gibt unter anderem Aufschluss über GCP-Richtlinien, über Prüfpläne und ethische Aspekte und spricht auch juristische Gesichtspunkte von medizinischen Studien an.

BiAS. gibt zum Abschluss jeder Berechnung spezielle Literaturhinweise, die eine weitergehende Lektüre erleichtern und die verwendeten Methoden zitierbar machen.

Auf den nächsten Seiten finden sich die wichtigsten, in **BiAS.**' Version 11 zitierten bzw. verwendeten Lehrbücher und Originalarbeiten. Spezielle Literaturangaben zu den Beispielen des Validierungsteils sind - wie in **BiAS.**' Programmausgabe - jeweils den Beispielen angefügt und finden sich alle in der nachfolgenden Aufstellung.

Die aktuellen *ICH-Guidelines* (Statistical Principles for Clinical Trials) finden Sie unter www.ich.org, die *Guidelines for Good Clinical Practice GCP* unter www.emea.eu.int .

*) Zu L. Sachs (2003): Zitate im Handbuch und im Programm beziehen sich häufig auch auf frühere Auflagen. Nachfolgerwerk ist J. Hedderich und L. Sachs (2015) *Angewandte Statistik: Methodensammlung mit R*, Springer.

Weitere verwendete Lehrbücher und Originalarbeiten:

- Aalen O.O., S. Johansen (1978) An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand J Statist* 5, pp. 141-50.
- Abel U. (1993) *Die Bewertung diagnostischer Tests*, Hippokrates-Verlag Stuttgart
- Abramowitz M., I. Stegun (1972) *Handbook of Mathematical Functions*. 10th edition Dover Publications New York. Download: http://people.math.sfu.ca/~cbm/aands/abramowitz_and_stegun.pdf
- Abt K. (1960) Analyse de covariance et analyse par differences. *Metrika* 3, pp. 26-45, 95-116, 177-211.
- Abt K. (1967) On the identification of the significant independent variables in linear models. *Metrika* 12, pp. 1-15, 81-96.
- Ackermann H. (1979) A concise formula to compute general orthogonal contrast coefficients. *Biom J* 21, 7, 597-606.
- Ackermann H. (1985) *Mehrdimensionale nicht-parametrische Normbereiche*. Springer-Verlag Berlin-HD-NY-Tokyo
- Ackermann H. (1991) A Friedman-type test for circular data and arbitrary cell frequencies. *Biom J* 33, 6, pp. 643-654.
- Ackermann H. (1991) A program package for **Biometrical Analysis of Samples**. *Comp. Stat. & Data Analysis* 11, 2, pp. 223-4.
- Ackermann H. (1997) Box-Cox-transformierte multivariate Normbereiche. *Inform Biom Epidem in Med & Biol* 28, 4, pp. 197-212.
- Ackermann H. (1995-2014) *Grundlagen der Medizinischen Biometrie*. 1.-7. Auflage epsilon-Verlag Darmstadt
- Agresti A. (1992) Exact Inference for Contingency Tables. *Statistical Science* Vol 7, 1, pp. 131-77.
- Alexander H., D.J. Browse, S.J. Reading, I.S. Benjamin (1999) A simple and accurate method for calculation of the EC50. *J Pharmacol Toxicol* 41, pp. 55-8.
- Altman D.G., J.M. Bland (1983) Measurements in Medicine: The Analysis of Method Comparison Studies. *The Statistician* 32, 3, pp. 307-318.
- Altman D.G. (1998) Confidence Intervals for the Number Needed to Treat. *Brit. Med. J.* 317, 1309-12
- Altman D.G. (Editor, 1994-2007) *Statistical Notes*. *British Medical Journal* <http://www.bmj.com> und <http://www.jerrydallal.com/LHSP/bmj.htm>
- Altman D.G., Matthews J.N.S. (1996) Interaction 1: Heterogeneity of effects. *Brit. Med. J.* 1996;313:486
- Andersen P., Borgan, O. Gill, R., Keiding, N. (1993) *Statistical models based on counting processes*. New York: Springer
- Anscombe F.J., J.W. Tukey (1963) Examination and analysis of residuals. *Technom.* 5,2, p.141-60
- APV-Richtlinie (1985) "Untersuchungen zur Bioverfügbarkeit". *Pharm. Ind.* 47, pp. 627-632.
- Armitage P. (1955) Tests for linear trends in proportions and frequencies. *Biometrics* 11:375-386.
- Armitage P., G. Berry (1987/88) *Statistical Methods in Medical Research*. Blackwell Sci Publ.
- Armitage P., G. Berry, J.N.S. Matthews (2002) *Statistical Methods in Medical Research*. 4th edition Blackwell Sci Publ.
- Backhaus K., B. Erichson, W. Plinke, R. Weiber (2011) *Multivariate Analysemethoden - Eine anwendungsorientierte Einführung*. 13. Auflage Springer-Verlag Berlin
- Bamber D. (1975) The area above the ordinal diminace graph and the area below the receiver operating characteristic graph. *J. Math. Psych.* 12, pp. 387-415.
- Barker C. (2009) The mean, median and confidence intervals of the Kaplan-Meier survival estimate - Computations and applications. *The American Statistician* Vol. 63, No. 1, pp. 78-80.

- Batschelet E. (1981) Circular Statistics in Biology. Academic Press.
- Bauer P., K. Köhne (1994) Evaluation of experiments with adaptive interim analysis. *Biometrics* 50, pp. 1029-41.
- Beasley J.D., S.G. Springer (1977) The percentage points of the normal distribution. *Appl. Stat.* 26, 1
- Benjamini Y., Y. Hochberg (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Series B*, 57, 1, pp. 289-300.
- Beran, R. P. Hall (1993) Interpolated nonparametric prediction intervals and confidence intervals. *J. Royal Stat. Soc. Series B*, 55, 3, pp. 643-52.
- Berk K.N. (1987) Effective Microcomputer Statistical Software. *Amer. Stat.* Vol. 41, No. 3, pp. 222-8
- Bernoulli, J. (1713) *Ars coniectandi, opus posthumum*. Thurneysen Brothers Basel
- Best D.J., P.G. Gipps (1974) Algorithm AS 71: The Upper Tail Probabilities of Kendall's Tau. *Applied Statistics* 23, 1, pp. 98-100
- Best D. J., D.E. Roberts (1975), Algorithm AS 89: The Upper Tail Probabilities of Spearman's rho. *Applied Statistics*, Vol 24, No 3, pp. 377-379.
- Betz M.A, Elliott S.D. (1984) The method of unweighted means in multivariate analysis of variance. *J. Educ. Behav. Statist.* 9, 3, pp. 193-213
- Betzin J., P. Bollmann-Sdorra (2003) On the determination coefficient in CFA. *Psychology Sci.* 45, 2, pp.400-20
- Bewick V., L. Cheek, J. Ball (2004) Statistics review 10: Further nonparametric methods, *Critical Care* 8(3) pp. 196-199
- Blackwelder W.C. (1982) 'Proving the Null Hypothesis' in Clinical Trials. *Contr. Clin. Trials* 3, 345-55
- Bland J.M. (2000) *An introduction to medical statistics*, 3rd ed. Oxford: Oxford University Press
- Bland J.M., D.G. Altman (1986) Statistical methods for assessing agreement between two methods of clinical measurements. *Lancet* 8.2.1986
- Bland J.M., D.G. Altman (1996) Statistics Notes: Measurement error and correlation coefficients. *Brit. Med. J.* 313: 41-2.
- Bland J.M., D.G. Altman (1997) Statistics Notes: Cronbach's alpha. *Brit. Med. J.* 314, pp. 572-3.
- Bland J.M., D.G. Altman (1999) Measuring agreement IN method comparison studies. *Stat Methods Med Res* 8: 135-60.
- Bock H.H. (1974) *Automatische Klassifikation*. Vandenhoeck und Ruprecht. Göttingen Zürich.
- Bock J., H. Toutenberg (1991) Sample size determination in clinical research. In: *Handbook of Statistics* 8, North-Holland.
- Bonett D.G. (2002) Sample size requirements for estimation intraclass correlations with desired precision. *Stat. in Med.* 21, pp. 1331-5.
- Bonett D.G., R.M. Price (2006) Confidence intervals for a ratio of binomial proportions based on paired data. *Stat. in Med.* 25, pp.3039-47.
- Bonferroni C.E. (1936) *Teoria statistica delle classi e calcolo delle probabilità*. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 8:3-62.
- Borenstein M., L.V. Hedges, J.P.T.Higgins, H.R. Rothstein (2009) *Introduction to meta-analysis*. Chichester, UK: Wiley
- Bortz J. (1977-2005) *Lehrbuch der Statistik*. 1. bis 6. Auflage, Springer-Verlag.
- Bortz J., C. Schuster (2010) *Statistik für Human- und Sozialwissenschaftler*. 7. Auflage, Springer-Verlag.
- Bortz J., G.A. Lienert, K. Boehnke (2008) *Verteilungsfreie Methoden in der Biostatistik*. 3. Auflage Springer
- Bortz J., G.A. Lienert (2008) *Kurzgefasste Statistik für die klinische Forschung*. 3. Auflage Springer
- Bowker A.H. (1948) A test for symmetry in contingency tables. *J. Amer. Stat. Ass.* 43, pp. 572-574.

- Box G.E.P., W.G. Hunter, J.S. Hunter (1978) *Statistics for experimenters*. John Wiley New York.
- Bratcher T.L., M.A. Moran, W.J. Zimmer (1970) Tables of sample sizes in anova *J. Qual. Tech.* 2,156
- Bristol D.R. (1989) Sample sizes for constructing confidence intervals and testing hypothesis. *Stat. in Med.* 8, 7, pp. 803-11.
- Brockwell P., R. Davis (2002) *Introduction to Time Series and Forecasting* 2nd. ed. Springer.
- Brookmeyer R., J. Crowley (1982) A confidence interval for the median survival time. *Biometrics* Vol. 38, pp. 29-41.
- Brown B.W. (1980) The crossover experiment for clinical trials. *Biometrics* 36, pp. 69-79.
- Brown G.W., A.M. Mood (1951) On Median Tests for Linear Hypotheses. *Proc. of the 2nd Berkeley Symposium on Math., Statist. and Prob., Univ. California.*
- Bühning H., G. Trenkler (1978) *Nichtparametrische statistische Methoden*. de Gruyter Berlin, NY.
- Campbell G. (1994) *Advances in statistical methodology for the evaluation of diagnostic and laboratory tests*. *Statistics in Medicine* 13, pp. 499-508.
- Campbell M.K., J. Mollison, N. Steen, J.M. Grimshaw, M. Eccles (2000) Analysis of cluster randomized trials in primary care: a practical approach. *Family Care* 17,2, pp.192-6
- Casagrande J.T., M.C. Pike, P.G. Smith (1978) An improved approximate formula for Calculating sample sizes for comparing two binomial distributions. *Biometrics* 34, pp. 483-486.
- Chen E.H. (1971) The power of the Shapiro-Wilk W-Test for normality in samples from Contaminated normal distributions. *J. Amer. Stat. Ass.* 66, pp. 760-762.
- Chen Z., R Bertin, G Frolidi (2013) EC50 estimation of antioxidant activity in DPPH assay using several statistical programs. *Food Chemist* 138, pp. 414-20.
- Chinn S. (2000) A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine* 19, pp.3127-31
- Chou Y.-M., G.M. Johnson (1987) Sample sizes for strong two-sided distribution-free tolerance limits. *Statistical Papers* 28, 1, pp. 117-31.
- Cochran W.G. (1950) The comparison of percentages in matched samples. *Biometrika* 37, 256-266.
- Cochran W.G. (1954) Some methods for strengthening the common chi²-tests. *Biometrics* 10, 417-51.
- Cohen J. (1960) *Errors of Measurement in Statistics*. *Technometrics* 10, pp. 637-666.
- Cohen J. (1977,1988) *Statistical power analysis for the behavioral sciences*. 1st ed. Academic Press New York, 2nd ed. Hillsdale New York, Erlbaum Assoc.
- Cohen J. (1992) A power primer. In: *Psychological Bulletin*, Vol. 112, No. 1 pp. 155-159.
- Colditz A., J.N. Miller, F. Mosteller (1988) Measuring gain in the evaluation of medical technology - The probability of a better outcome. *Intl. J. of Technology in Health Care*, 4, pp.637-642.
- Collett D. (2003) *Modelling survival data in medical research*. Chapman and Hall 2nd ed.
- Comrey A.L. (1973) *A First Course in Factor Analysis*. Academic Press New York London.
- Connet J.E., J.A. Smith, R.B. McHugh (1987) Sample size and power for paired-matched case-control-studies. *Stat. in Med.* 6, pp. 53-9
- Conover W.J. (1980,1999) *Practical Nonparametric Statistics*, 1st & 3rd edition John Wiley
- Conover W.J., R.L. Iman (1979) On multiple comparisons procedures. Technical Report Los Alamos Scientific Laboratory
- Conover W.J., R.L. Iman (1981) Rank Transformations as a Bridge between Parametric and Nonparametric Statistics, *The American Statistician* 35, 3, pp. 124-129
- Cook R.J., D.L. Sackett (1995) The number needed to treat: a clinically useful measure of treatment effect. *British Med. J.* 310, pp.452-4
- Cox D.R., D. Oakes (1984) *Analysis of Survival Data*. Chapman and Hall, London New York.
- CPMP Working Party on Efficiency of Medicinal Products (1989) Note for Guidance: Investigation of Bioavailability and Bioequivalence.

- Craddock J.M., C.R. Flood (1970) The distribution of chi-squared statistic in small Contingency tables. *Applied Statistics*, 19, pp.173-81.
- Cramer C.Y. (1972) A first course in methods of multivariate analysis. Virginia Polytech. Inst. pp.56 ff
- Cramér H. (1946) *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press
- D'Agostino R.B. (1998) Tutorial in Biostatistics: Propensity Score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statist. in Medicine* 17, pp. 2265-2281.
- Dallal G.E. (1988) Statistical Microcomputing - Like It Is. *Amer. Stat.* Vol. 42, No. 3, pp. 212-6.
- Dallal G.E. (1985) PC-SIZE. A Program for Sample Size Determinations Version 2.0. USDA Human Nutr. Res. Center on Aging at Tufts University. 711 Washington Street, Boston, MA 02111.
- Dallal G.E., L. Wilkinson (1986) An Analytic Approximation to the Distribution of Lilliefors' Test-Statistic for Normality. *Am. Stat.* 40, pp. 294-6.
- Daniel C. (1976) *Applications of statistics to industrial experimentation*. John Wiley New York.
- Daniel C. (1978) Patterns in residuals in the two-way-layout. *Technometrics* 20, 4, pp. 385-396.
- Daniel C., F.S. Wood (1971) *Fitting equations to data*. John Wiley New York
- David H.A., H.O. Hartley, E.S. Pearson (1954) The distribution of the ratio, in a single normal sample, of range to standard deviation. *Biometrika* 41, pp. 482-493.
- Deichsel G., H.J. Trampisch (1985) *Clusteranalyse und Diskriminanzanalyse*. G. Fischer Stuttgart
- De Jonge H. (1983) Deficiencies in clinical reports for registration of drugs. *Stat. in Med.* 2, 2, pp. 155-66.
- DeLong E.R., D.M. DeLong, D.L. Clarke-Pearson (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44, 3, pp. 837-45.
- Derendorf H., Th. Gramatté, H.G. Schäfer, A. Staab (2010) *Pharmakokinetik kompakt*. Wiss. Verlagsgesellschaft Stuttgart.
- DerSimonian R., N. Laird (1986) Meta-analysis in clinical trials. *Controlled Clin. Trials*, 7, pp. 177-88.
- Diletti E., D. Hauschke, V.W. Steinijans (1991) Sample size determination for Bioequivalence assessment by means of confidence intervals. *Int. J. Clin. Pharm.* 29,1, pp.1-8.
- Dixon W.J., A.M. Mood (1946) The statistical sign test. *J. Amer. Statist. Assoc.* 41, pp. 557-566.
- Donner A., M. Eliasziw (1992) A goodness-of-fit approach to inference procedures for the kappa statistic: confidence interval construction, significance testing and sample size estimation. *Stat. in Med.* Vol. 11, pp. 1511-1519.
- Donner A., N. Klar (2000) *Design and Analysis of Cluster Randomization Trials*, Oxford University Press New York 2000.
- Dunlap W.P., R.S. Powell, T.K. Konnerth (1977) A Fortran IV function for calculating probabilities associated with the Studentized Range statistic. *Behavior Research Methods and Instrumentation* 9, 4, pp. 373-75.
- Dunn O.J. (1964) Multiple comparisons using rank sums. *Technometrics* 6, pp. 241-252.
- Durbin J., G.S. Watson (1950,1951) Testing for Serial Correlation in Least Squares Regression. I: *Biometrika* 37, 409-28, II: *Biometrika* 38, 159-79.
- Edwards A.L. (1948) Note on the "correction for continuity" in testing the significance Between correlated proportions. *Psychometrika* 13, pp. 185-7.
- Elliott S.D. (1989) The method of unweighted means in univariate and multivariate analysis of variance. *J. Educ. Psychol. Measurement* 49, 399-405
- Evans J.D. (1996) *Straightforward Statistics for the Behavioral Sciences*. Brooks/Cole Publishing, Pacific Grove, Calif.
- Fan C., D. Zhang, C.-H. Zhang (2011) On Sample Size of the Kruskal-Wallis-Test with Application to a Mouse Peritoneal Cavity Study. *Biometrics* 67, 1, pp. 213-24.

- Feldman S.E., E. Klinger (1963) Short cut calculation of the Fisher-Yates 'exact test'. *Psychometrika* 28, pp.289-291.
- Fisher R.A. (1971) *Statistical Methods for Research Workers*. 14. Auflage, Hafner Press. pp. 99-101.
- Fleiss J.L. (1973-2003) *Statistical Methods for Rates and Proportions*. John Wiley New York, pp. 143-147. 3. Auflage 2003.
- Fleiss J.L., J. Cohen (1973) The equivalence of weighted kappa and the intra-class correlation coefficient as measures of reliability. *Educational and Psychological Measurement*. 33, pp. 613-619
- Fleiss J.L., B. Levin, M.C. Paik (2003) *Statistical methods for rates and proportions*, 3rd ed. Hoboken: John Wiley & Sons.
- Freedman L.S. (1982) Tables of the numbers of patients required in clinical trials using the Logrank test. *Statistics in Medicine* Vol. 1, No. 2, pp. 121-130.
- Freeman G.H., J.H. Halton (1951) Note on exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika*, 38:141-149, 1951.
- Freeman M.F., J.W. Tukey (1950) Transformations related to the angular and the square root. *The Annals of Mathematical Statistics* 21:607-11.
- Friedman M. (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Stat. Assoc.* 32, pp.675-701.
- Galen R.S., S.R. Gambino (1979) *Norm und Normabweichung klinischer Daten*. G Fischer-Verlag
- Garbin C. (2012) In: http://psych.unl.edu/psycrs/statpage/biv_corr_comp_eg.pdf Univ. Nebraska USA
- Gart J.J. (1969) An exact test for comparing matched proportions in Cross-Over-designs. *Biometrika* 56, pp.75-80
- Gehan E.A. (1965) A generalized Wilcoxon-test for arbitrarily single-censored samples. *Biometrika* 52, 1 und 2, pp. 203-223.
- Gibbons J., Chakraborti, S. (2010) *Nonparametric statistical inference*, 5th ed. CRC Press - A Chapman and Hall Book.
- Glantz S.A. (2012) *Primer of Biostatistics*. McGraw Hill New York.
- Glaser W.R. (1978) *Varianzanalyse*. UTB Gustav Fischer Stuttgart New York.
- Goetghebeur E., Ryan L. (1990) A modified log rank test for competing risks with missing failure type. *Biometrika* 77, 1, pp. 207-11.
- Goldschmidt A. (Hrsg., 1996) *Medizinische Statistik und Klinische Forschung - von der Idee zum Ergebnis*. Springer-Verlag.
- Goldstein R. (1989) Power and Sample Size via DOS Computers. *Amer. Stat.* Vol. 43, No.4, 253-60
- Graf U., H.J. Henning, K. Stange, P-T. Wittrich (1987) *Formeln und Tabellen der angewandten mathematischen Statistik*. Springer-Verlag.
- Gray R. (1988) A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics* 16, pp. 1141-54.
- Grizzle J.E. (1965) Two-period cross-over design and its use in clinical trials. *Biometrics* 21, 467-80.
- Grizzle J.E. (1967) Continuity correction in the chi-squared test for 2x2-tables. *Amer. Stat.* 21, 28-32.
- Grizzle J.E. (1974) Correction Note. *Biometrics* 21, p. 727.
- Grubbs F.E. (1969) Procedures for detecting outlying observations in samples. *Technometrics* 11:1-21.
- Grubbs F.E., G. Beck (1972) Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics* 14, 4, pp. 847-854.
- Guttman I. (1970) *Statist. Tolerance Regions: Classical and Bayesian*. Griffin's Stat. Monographs 26
- Haf C.M., T. Cheaib (1985) *Multivariate Statistik in den Natur- und Verhaltenswiss.* Vieweg
- Hajek J., Z. Sid k (1967) *Theory of Rank Tests*. Academic Press New York, p. 61.

- Haldane J.S.B. (1940) The mean and variance of chi-squared, when used as a test of Homogeneity when samples are small. *Biometrika* 31, pp.346-55.
- Hahn G.J. (1972) Simultaneous prediction intervals for a regression model. *Technom.* 14, 203-14
- Hall W.J., J.A. Wellner (1980) Confidence bands for a survival curve from censored data. *Biometrika* 67, pp. 133-43.
- Hanley J.A., K.O. Hajian-Tilaki (1997) Sampling variability of nonparametric estimates of the areas under receiver operating characteristic curves: an update. *Academic Radiology* 4:49-58.
- Hanley J.A., B.J. McNeil (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29-36.
- Hanley J.A., B.J. McNeil (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148:839-843.
- Harms V. (1988) *Biomathematik, Statistik und Dokumentation*. Harms-Verlag Kiel.
- Harris E.K., A. Albert (1991) *Survivorship Analysis for Clinical Studies*. Marcel Deller Inc. New York
- Hart B.I. (1942) Significance levels for the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* Vol. 13, pp. 445-7.
- Harter H.L., D. Clemm, E.H. Guthrie (1959) The probability integrals of the range and of studentized range. WADC Technical Report 58-484 Vol. II, Dayton, Ohio, Wright Air Craft Development Center.
- Hartung J. (1982) *Statistik*. Oldenburg München Wien.
- Hartung J., B. Elpelt (1986) *Multivariate Statistik*. Oldenburg München Wien.
- Hauschke D., V.W. Steinijans, E. Diletti (1990) A distribution free procedure for the statistical analysis of bioequivalence studies. *J. Clin. Pharm. Therapy & Toxikol.* 28, 2, pp. 72-78.
- Headrick T.C. (2003) An algorithm for generating exact critical values for the Kruskal-Wallis one-way-ANOVA. *J. Mod. Appl. Stat. Meth.* 2, 1, pp. 268-71
- Hedderich J., L. Sachs (2015), *Angewandte Statistik: Methodensammlung mit R*, Springer-Verlag 15. Auflage. Siehe auch Sachs (1968-2003) und Sachs und Hedderich (2009)
- Held L., K. Rufibach, B. Seifert (2013) *Medizinische Statistik*. Pearson Higher Education München
- Herrmann E. (2011) Assessing the predicted survival time in a parametric context. Preprint IBMM Univ. Frankfurt am Main.
- Herrmann E., M. Ahmed, M. Shahrashoub, A. Orth, H. Ackermann (2010) Assessing approximative algorithms for calculating Spearman's rank correlation coefficient for moderate sample size in the case of ties. DAGStat Dortmund März 2010.
- Hilgers R.A. (1991) Distribution-free confidence bounds for ROC curves. *Methods of Information in Medicine* 30:96-101.
- Hintze J.L. (2011) *PASS User's Guide*, NCSS, Kaysville, Utah.
<http://www.ncss.com/software/pass/pass-documentation/> (Kontingenztafel: Chap. 250, ROC-Vergleich: Chap. 262)
- Hodges J.L. Jr. (1957) The Significance Probability of the Smirnov Two-Sample Test, *Arkiv for Matematik*, 3, pp. 469-486.
- Hollander M., D.A. Wolfe (1973,1999) *Nonparametric statistical methodes*. 1st & 2nd ed. John Wiley.
- Holm S. (1979) A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, pp. 65-70.
- Hommel G. (1988) A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* 1988, 75, pp. 383-386.
- Horn M., R. Vollandt (1995) *Multiple Tests und Auswahlverfahren*. Gustav Fischer - Verlag Stuttgart-Jena.
- Hosmer D.W., S.Lemeshow (2000) *Applied Logistic Regression*, John Wiley
- Hotelling, H. (1931) The generalisation of Student's ratio. *Ann. Math. Stat.* 2, pp. 360-78
- Hsieh F., D. Bloch, M. Larsen (1998) A simple method of sample size calculation for linear and logistic regression. *Statist. Med.* 17, pp. 1623-1634.

- Hsu J.C., P. Qiu, L.J. Hin, D.O. Mutti, K. Zadnik (2004) Multiple comparisons with the best ROC curve. Lecture Notes Monograph Series 47, pp. 65-75.
- Inst. of Math. Stat. (ed., 1989) Discriminant Analysis and Clustering. Stat. Science 4, No. 1, 34-69
- Johnson R. (1988) p-values in 2x2-tables. Biometrics 44, 3, pp. 907-10.
- Jonckheere A.R. (1954) A distribution free k sample test against ordered alternatives. Biometrika 41, pp. 131-45
- Jonge H. De (1983) Deficiencies in clin. reports for registration of drugs. Stat. in Med. 2, pp.155-66.
- Julious S.A. (2004) Sample sizes for clinical trials with normal data. Stat. in Med. 23, pp. 1921-86.
- Kaplan E.L., P. Meier (1958) Nonparametric estimation from incomplete observations. JASA 53, pp. 457-81
- Katz D., J. Baptista, S.P. Azen, M.C. Pike (1978) Obtaining confidence intervals for the risk ratio in cohort studies. Biometrics 34, pp. 469-474.
- Kendall M.G., J.D. Gibbons (1990) Rank correlation methods. Edward Arnold Publ. Ltd. London
- Kerry SM, JM Bland (1998) The intraclass correlation coefficient in cluster randomisation. British Medical Journal 316, pp. 1455-60.
- Klein J.P., M.L. Moeschberger (2003) Survival Analysis. Techniques for censored and truncated data, 2nd ed. New York: Springer Publishers
- Klinger H. (1992) Statistische Modelle für das Bioäquivalenzproblem. 38. Biom. Koll. der DR der Int. Biom. Gesellsch. Gießen 16.-20.3.1992.
- Klotz J. (1980) A modified Friedman-Test with missing observations and ordered categorical data. Biometrics 36, pp.665-70.
- Knox SA., P. Chondros (2004) Observed intra-cluster-correlation in a cluster survey sample of patient encounters in general practice in Australia. BMC Medical Research Methodology 2004, 4, 30ff
- Koch G.G. (1972) The use of nonparametric methods in the statistical analysis of the two-period change-over design. Biometrics 28, pp. 577-584.
- Kolmogoroff A. (1933) Sulla determinazione empirica di una legge di distribuzione. Giornale Inst. Ital. Attuari 4, pp. 83-91.
- Kolmogoroff A. (1941) Confidence limits for an unknown distribution function. Ann. Math. Stat. 12, 461-63
- Koopman P.A.R. (1984) Confidence Intervals for the Ratio of Two Binomial Proportions. Biometrics 40, pp. 513-517.
- Koziol J.A., M.D. Perlman (1978) Combining Independent chi squared Tests. J. Amer. Stat. Ass. 73,4, 753-63
- Kramer C.Y. (1972) A first course in methods of multivariate analysis. Virginia Polytech. Inst., Blacksburg, Virginia.
- Krauth J. (1993) Einführung in die Konfigurationsfrequenzanalyse. Beltz-Verlag.
- Krauth J., A. Lienert (1973) Die Konfigurationsfrequenzanalyse. Karl Alber Verlag Freiburg-München
- Kreyszig E. (1970) Introductory Mathematical Statistics. John Wiley.
- Kroning AJ., PH. Franses (2003) Confidence intervals for comparing Cronbach's coefficient alpha values. Preprint Elsevier Science.
- Krouwer J.S. (2008) Why Bland-Altman plots should use X, not $(Y+X)/2$ when X is a reference method. Statistics in Medicine 27:778-780
- Kruskal W.H., W.A. Wallis (1952) Use of ranks in one-criterion variance analysis. J.Amer. Stat. Ass. 47, pp.583-621 und 48 (1953), pp.907-911.
- Kubinger K.D. (1986): A Note on Non-Parametric Tests for the Interaction in Two-Way Layouts. Biometrical Journal, 28, 67-72
- Küffner H., R. Wittenberg (1985) Datenanalysesysteme für statistische Auswertungen. G. Fischer

- Landis J.R., G.G. Koch (1977) The measurement of observer agreement for categorical data. *Biometrics* Vol. 33(1), pp. 159-174
- Lane D.M. et al. (2014) Introduction to Statistics: An Interactive Multimedia Course of Study. onlinestatbook.com, Rice-University.
- Lantz B. (2013) The impact of sample non-normality on ANOVA and alternative methods. *Brit. J of Mathem. and Statist. Psychology* Vol 66(2), pp. 224-44.
<http://onlinelibrary.wiley.com/doi/10.1111/j.2044-8317.2012.02047.x/full>
- Lautsch E., G.A. Lienert (1993) *Binärdatenanalyse*. Beltz-Verlag
- Lautsch E., von Weber (1995) *Methoden und Anwendungen der KFA*. Beltz-Verlag
- Lawless J.F. (1982) *Statistical Models and Methods for Lifetime Data*. Wiley NY
- Lee E.T. (1980) *Statistical Methods for Survival Data Analysis*. Wordsworth, Belmont, Calif.
- Lehmacher W. (1981) A more powerful simultaneous test procedure in configural frequency analysis. *Biom.J.* 23, 5, pp. 429-36.
- Lehmacher W. (1991) Analysis of the Crossover Design in the Presence of Residual Effects. *Statist. in Med.* 10, pp. 891-9.
- Lehmacher W., W. van Eimeren (1986) Zur statistischen Bewertung der Ergebnisse von Bioverfügbarkeitsstudien. *Therapiewoche* 36, pp. 413-420.
- Lehmann E.L. (1975,1998) *Nonparametrics: Statistical Methods Based on Ranks*. 1. und revidierte 1. Auflage Prentice Hall, New Jersey.
- Leisenring W., T. Alonzo, M.S. Pepe (2000) Comparisons of Predictive Values of Binary Medical Diagnostic Tests for Paired Designs. *Biometrics* 56, pp. 345-351
- Lemeshow S., D.W. Hosmer, J. Klar (1988) Sample size requirements for studies estimating odds ratios or relative risks. *Statist. in Med.* 7, pp. 759-64.
- Lemeshow S., D.W. Hosmer, J. Klar, S.K. Lwanga (1990) *Adequacy of Sample Size in Health Studies*. John Wiley Chichester New York
- Lienert G.A. (1973) *Verteilungsfreie Methoden in der Biostatistik*. Anton Hain Meisenheim am Glan.
- Lilliefors H.W. (1967) On the Kolmogoroff-Smirnoff test for normality with mean and Variance unknown. *J.Am.Stat.Ass.* 62, pp. 399-402.
- Lin L.I-K. (1989,2000) A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* 45, pp. 255-68. Correction Note: *Biometrics* 56, 2000, pp. 324-5.
- Linder A., W. Berchtold (1982) *Stat. Methoden*, Birkhäuser.
- Ljung G.M., G.E.P. Box (1978) On a Measure of a Lack of Fit in Time Series Models. *Biometrika* 65, 297-303.
- Lock R.H. (1993) A comparison of five student versions of statist. packages. *Amer. Stat.* 47,2, 136-45
- Lorenz R.J. (1996) *Grundbegriffe der Biometrie*. Gustav-Fischer-Verlag Stuttgart. 4. Auflage
- Long J.S. (1997) *Regression Models for categorical and limited dependent variables*. Thousand Oaks, Sage Publications
- Lund R.E., J.R. Lund (1983) Probabilities and upper quantiles for the Studentized Range. *Applied Statistics*, Vol.32 No.2, 1983, pp. 204-210.
- Mace A.E. (1964) *Sample-Size Determination*. Chapman & Hall, Ltd., London.
- Machin D., M.J. Campbell (1987) *Stat. Tables for Design of Clin.I Trials*. Blackw. Sci. Publ.
- Macskassy S., Provost, F., Rosset, S. (2005) ROC Confidence Bands: An Empirical Evaluation. *Proc. of the 22nd Int. Conference on Machine Learning (ICML)*. Bonn, Germany.
- Mann H.B. (1945) Nonparametric Tests against Trend. *Econometrika* 13, pp. 245-9.
- Mantel N. (1970) Why stepdown procedures in variable selection. *Technometrics* 12, 3, pp. 621-625.
- Mantel N., W. Haenszel (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst.* 22, pp. 719-48.

- Mantel N. (1963) Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel-procedure. *J. Amer. Statist. Assoc.* 58, 690-700.
- Matthews J.N.S., D.G. Altman (1996) Interaction 2: compare effect sizes not P values. *Brit Med J* 1996;313:808
- Maurer W., L.A. Hothorn, W. Lehmacher (1995) Multiple Comparisons in Drug Clinical Trials and Preclinical Assays: A-priori Ordered Hypotheses. In: Vollmar, J. (Hrsg.) *Testing principles in clinical and preclinical trials*, Gustav Fischer 1995
- McBride G.B. (2005) A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. NIWA Client Report: HAM2005-062.
- McClish D.K. (1987) Comparing the areas under more than two independent roc curves. *Medical Decision Making* 7, pp. 149-55.
- McDonald J.H. (2014) *Handbook of Biol. Statistics*. Sparky House Publishing Baltimore, Maryland: <http://udel.edu/~mcdonald/HandbookBioStat.pdf>
- McQueen J. (1967) Some Methods for Classification and Analysis of Multivariate Observations. *Proc. 5th Berkely Symp. Math. Stat. Prob.* 65/66, pp. 281-297.
- Miller A.J. (1992) Algorithm AS 274: Least Squares Routines to Supplement Those of Gentleman. *Applied Statistics*, Vol. 41, No. 2, pp. 458-478
- Miller A.J. (2004) A collection of mathematical and statistical routines. *J. of Stat. Soft.* 9, 3, pp.1-3
- Miller L.H. (1956) Table of percentage points of Kolmogorov statistics. *J. Am. Stat. Ass.* 51, pp.111-121
- Miosga P., T.L. Chung, H. Ackermann, E. Herrmann, W. Helm (2012) A new approach towards simulation of the Rule of Ten for a Cox regression. Poster, Biometrisches Kolloquium 2012 in Berlin.
- Moore P.G. (1955) The properties of the mean square successive difference in samples from various populations. *J. Amer. Statist. Assoc.* Vol. 50, pp. 434-56.
- Morgan W.A. (1939) A test for the significance between two two variances in a sample from a normal bivariate population. *Biometrika* 32, pp. 13-19
- Mosteller F., J.W. Tukey (1977) *Data analysis and regression*. Addison Wesley, Reading, Mass.
- Moskowitz C.S., M.S. Pepe (2006) Comparing the Predictive Values of Diagnostic Tests: Sample size and Analysis for Paired Study Designs. *Clinical Trials* Vol. 3, No. 3, pp. 272-279.
- Mudholkar G.S., M. McDermott, D.K. Srivastava (1992) A test of p-variate normality. *Biometrika* 79, 4, pp. 850-4.
- Murray D.M. (1998) *The Design and Analysis of Group Randomized Trials*, Oxford Univ. Press 1998
- Nagashima K. (2016) A sample size estimation tool for one sample nonparametric tests for a survival proportion [Internet]. 2016 Mar 21 <http://nshi.jp/en/js/onesurvey/>
- Neter J., W. Wasserman (1974) *Applied linear statistical models*. R. D. Irwin, Inc., Homewood, Ill.
- Nemenyi P. (1963) *Distribution-free multiple Comparisons*. Thesis Princeton University.
- Neumann J., R.H. Kent, H.B. Bellinson, B.I. Hart (1941) The mean square successive difference. *Annals of Mathematical Statistics* Vol.12, pp.153-62.
- Noether G.E. (1967) *Elements of Nonparametric Statistics*. Wiley New York.
- Noether G.E. (1987) Sample Size Determination for Some Common Non-parametric Tests. *J. Amer. Stat. Ass.* Vol. 82, No. 398, pp. 645-7.
- Nordstokke D.W., B.D. Zumbo (2010) A New Nonparametric Levene Test for Equal Variances. *Psicologica: Int. J. Methodol. Experim. Psych.* 31, 2, pp. 401-30.
- Oberfeld D., Th. Franke (2013) Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data. *Behavior Research Methods* Vol 45, 3, pp. 792-812
- O'Brien P.C., T.R. Fleming (1979) A multiple testing procedure for clinical trials. *Biometrics* 35, pp. 549-56.
- Ortseifen C., T. Bruckner, M. Burke und M. Kieser (1997) An Overview of Software Tools for Sample Size Determination. *Inform., Biometrie und Epidemiol. in Med. und Biol.* 28, 2, pp. 91-118.

- Osborne R, WSB. Paterson (1952) On the sampling variance of heritability estimates derived from variance analyses. Proc. R. Soc. Edinb. Sect. B 64:456-461.
- Overall J.E., C.J. Klett (1972) Applied Multivariate Analysis. McGraw-Hill New York.
- Owen D.B. (1962) Handbook of Statistical Tables. Addison Wesley.
- Passing H., W. Bablok (1983) A New Biometrical Procedure for Testing the Equality of Measurements from Two Different Analytical Methods. J. Clin. Chem. Clin. Biochem. Vol. 21, pp. 709-720.
- Pawlik K. (1959) Der maximale Kontingenzkoeffizient im Falle nicht-quadratischer Kontingenztafeln. Metrika 2, 150-66
- Pearson E.S., M.A. Stephens (1964) The ratio of range to standard deviation in the same Normal sample. Biometrika 51, pp. 484-487.
- Peduzzi P., J. Concato, E. Kemper, T.R. Holford, A.R. Feinstein (1996) A simulation study of the number of events per variable in logistic regression analysis. J. Clin. Epidemiol. 49, pp. 1373-9.
- Persike M. (2010) Vorlesung "Mathematische und statistische Methoden II" des Psychologischen Instituts der Universität Mainz.
- Peto R. , M.C. Pike et al. (1977) Design and analysis of randomized clinical trials requiring Prolonged observation of each patient. II Analysis and examples. Brit. J. Cancer 35, p.31.
- Pezzullo J.C. (2007) Scripts zum Cox-Modell, statpages.org, Georgetown Univ US
- Pitman E.J.G. (1939) A Note on Normal Correlation. Biometrika 31, pp. 9-12
- Pocock S.J. (1983) Clinical Trials - A Practical Approach. John Wiley & Sons.
- Pocock S.J. (1977) Group sequential methods in the design and analysis of clinical trials. Biometrika 64, pp. 191-199.
- Pratt J.W., J.D. Gibbons (1981) Concepts of Nonparametric Theory. Springer Ser. in Statistics 1981
- Prescott P. (1979) Critical values for a sequential test for many outliers. Appl. Stat. 28, pp. 36-9.
- Prins (editor, 2012) e-Handbook of Statistical Methods. NIST/SEMATECH
<http://www.itl.nist.gov/div898/handbook/pmc/section3/pmc323.htm>
- Proschan F. (1953) Conf. and Tolerance Interv. for the Normal Distribution. JASA 48, pp. 550-64
- Putter H., Fiocco M., Geskus R.B. (2007) Tutorial in biostatistics: Competing risks and multi-state models. Statistics in Medicine 26, pp. 2389-430.
- Raghunathan T.E., R. Rosenthal, D.B. Rubin (1996) Comparing correlated but nonoverlapping correlations. Psychological Methods, Vol 1(2), pp. 178-183
- Rahlf's V. (1987) "N". IDV Gauting/München. <http://www.idv-cro.eu/cms/index.php?id=2&L=1>
- Rasch B., Friese, M., Hofmann, W., Naumann, E. (2010) Quantitative Methoden 2. Einführung in die Statistik für Psychologen und Sozialwissenschaftler. Springer-Verlag Heidelberg.
- Ritschel W.A. (1986) Handbook of Basic Pharmacokinetics. Drug Intelligence Publ., Inc.
- Robins J., S. Greenland, NE. Breslow (1986) A general estimator for the variance of the Mantel-Haenszel odds ratio. Amer. J. Epidemiol. 124, 5, pp. 719-23
- Rode R.A., V.M. Chinchilli (1988) The Use of Box-Cox-Transformations in the Development of Multivariate Tolerance Regions With Applications to Clinical Chemistry. Amer. Stat. 42, 1, pp. 23-30.
- Rosenbaum P.R., D.B. Rubin (1983) The central role of the propensity score in observational studies for causal effects. Biometrika, 70, pp. 41-55.
- Rosenthal R. (1991) Meta-Analytic Procedures for Social Research. Sage Publications, Inc.
- Royston J.P. (1982) An extension of Shapiro and Wilk's test for normality to large samples. Appl. Stat. 31, 2, pp. 115-24.
- Rubin J. (1967) Optimal Classification into Groups: An Approach for Solving the Taxonomy Problem. J. Theor. Biol. 15, pp. 103-144.
- Rüger B. (1978) Das maximale Signifikanzniveau des Testes "Lehne H_0 ab, wenn k unter n gegebenen Tests zur Ablehnung führen." Metrika 25, pp. 171-178.

- Ryan T.P. (2013) Sample Size Determination and Power. John Wiley & Sons, Wiley Series in Probability and Statistics
- Sachs L. (1970-1993) Statistische Methoden I und II. Springer-Verlag 1. bis 7. Auflage.
- Sachs L. (1968-2003) Angewandte Statistik, Springer-Verlag 1. bis 11. Auflage. Siehe auch: Sachs und Hedderich (2006,2009) 12./13. Auflage und Hedderich und Sachs (2012,2015) 14./15. Auflage
- Sachs L., J. Hedderich (2006,2009) Angewandte Statistik: Methodensammlung mit R, Springer-Verlag 12. und 13. Auflage. Siehe auch Sachs (1968-2003) und Hedderich und Sachs (2012/2015)
- Savin N.E., K.J. White (1977) The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors. *Econometrica* 45, 1989-96.
- Schaich E., A. Hamerle (1984) Verteilungsfreie statistische Prüfverfahren. Springer Berlin.
- Schendera C. (2010) Clusteranalyse mit SPSS. Oldenbourg-Verlag München.
- Schmider E., M. Ziegler, E. Danay, L. Beyer, M. Bühner (2010) Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology* 2010 Vol. 6(4), pp. 147-51. <http://psycontent.metapress.com/content/0676589h9h340604/fulltext.pdf>
- Schneiderman M.A. (1964) The proper size of a clinical trial: "Grandman's strudel" method. *The Journal of New Drugs* 4, pp. 3-11. <http://onlinelibrary.wiley.com/doi/10.1002/j.1552-4604.1964.tb00097.x/pdf>
- Schouten H.J.A. (1982) Comparing two independent binomial proportions by a modified chi squared test - required sample sizes. *Biom.J.* 24, 1, pp. 93-6.
- Schuhmacher M., G. Schulgen (2002) Methodik klinischer Studien, Springer-Verlag
- Sedlmeier P., F. Renkewitz (2008) Forschungsmethoden und Statistik in der Psychologie. Pearson Studium München
- Sen P.K. (1960) On some convergence properties of U-statistics. *Calc. Stat. Ass. Bull.* 10, pp. 1-18.
- Senn S., G. D'Angelo, D. Potvin (2004) Carry-over in cross-over trials in bioequivalence: theoretical concerns and empirical evidence. *Pharmaceutical Statistics* 2004, 3, pp. 133-42.
- Shapiro S.S., M.B. Wilk (1965) An analysis of variance test for normality. *Biometrika* 52, pp.591-611
- Shrout PE, JL Fleiss (1979) Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin* 86, 2, pp. 420-428.
- Siegel, S. (1956) Nonparametric statistics. McGraw-Hill.
- Sivaprakasam J., M. Purva (2010) CUSUM analysis to assess competence: what rate is acceptable? *Clin. Teach.* 7, pp. 257-61.
- Simes R.J. (1986) An improved Bonferroni procedure for multiple tests of signif. *Biometrika* 73, 751-4
- Smirnov N.W. (1939) On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Universität Moskov. Ser. Internat. Sec.A* 2, pp. 3-8.
- Smirnov N.W. (1948) Tables for estimating the goodness of fit of empirical distributions. *Ann. Math. Stat.* 19, pp. 279-281.
- Sokal P.H.A., R.R. Sneath (1973) Numerical Taxonomy. Freeman & Comp., San Francisco.
- Späth H. (1983) Cluster-Formation und -Analyse. Oldenbourg-Verlag.
- Späth H. (1977) Cluster-Analyse-Algorithmen. Oldenbourg-Verlag München Wien, 2. Auflage.
- Spearman C. (1904) The proof and measurement of ass. between two things. *Am.J.Psy.* 15, 72-101
- Sprent P. (1969) Models in regression and related topics. Methuen & Co Ltd., London.
- Spilker B. (1984) Guide to clinical studies and developing protocols. Raven Press New York.
- Steiger J.H. (1980) Tests for comparing elements of a correlation matrix. *Psychological Bulletin* 87, 245-51.
- Steinhausen D., K. Langer (1977) Clusteranalyse. deGruyter Berlin.
- Steinijans V.W., E. Diletti (1983) Statistical Analysis of Bioavailability Studies: Parametric And Nonparametric Confidence Intervals. *Eur. J. Clin. Pharmacol.* 24, pp. 127-36.

- Steinijans V.W., D. Hauschke (1990) Update on the statistical analysis of bioequivalence studies. *J. Clin. Pharm.* 28, 3, pp. 105-110.
- Stoller D.S. (1954) Univariate two-population distributionfree discrimination. *JASA* 49, pp. 770-777.
- Sturges H.A. (1926) The choice of a class interval. *JASA* 21, pp. 65-6.
- Tatsuoka M.M. (1971) *Multivariate Analysis*. John Wiley.
- Thurstone L. (1947) *Multiple Factor Analysis*. Univ. of Chicago Press.
- Terwilliger J., J. Ott (1992) A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered.* 42, pp. 337-346.
- Tukey J.W. (1977) *Exploratory data analysis*. Addison-Wesley Publish. Co. Reading, Mass.
- Überla K. (1971) *Faktorenanalyse*. Springer-Verlag Berlin Heidelberg New York.
- Valz P.D., M.E. Thompson (1994) Exact inference for Kendall's S and Spearman's rho with Extensions to Fisher's Exact Test in rxc Contingency Tables. *Journal of Computational and Graphical Statistics* 3, pp. 459-472.
- van Elteren P.H. (1960) On the combination of independent two sample tests of Wilcoxon. *Bull. of the Inst. of International Statistics* 37, pp. 351-61.
- Van der Waerden B.L. (1952) Order tests for the two-sample problem and their power. *Indagationes Mathematicae*, 14, pp. 453-458
- Van der Waerden B.L. (1953) Order tests for the two-sample problem II+III. *Proc. Of the Koninklijke Nederlandse Akad. van Wetenschappen Series A*, pp. 564, 303-316
- Vaz S., T. Falkmer, A.E. Passmore, R. Parsons, P. Andreou (2013) The Case for Using the Repeatability Coefficient When Calculating Test–Retest Reliability. *PLoS ONE* 8(9): e73990. doi:10.1371/journal.pone.0073990
- Vergara I.A., T. Norambuena, E. Ferrada, A.W. Slater, F. Melo (2008) StAR: a simple tool for the statistical comparison of ROC curves. *BMC Bioinformatics* 9, 265-9.
- Victor A., A. Elsässer, G. Hommel, M. Blettner (2010) Wie bewältigt man die p-Wert-Flut? *Deutsches Ärzteblatt* Jg 107, Heft 4, pp. 50-54
- Vittinghoff E., C.E. McCulloch (2007) Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol.* 165, pp. 710-8.
- Vollmar J. (Hrsg., 1995) *Testing principles in clinical and preclinical trials*, Gustav Fischer
- Waerden B.L. van der (1957) *Mathematische Statistik*, Springer-Verlag. S. 345.
- Walter E. (1980) *Biomathematik für Mediziner*. Teubner.
- Wan X., W. Wang, J. Liu, T. Tong (2014) Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Medical Research Methodology* 14:135
- Weaver B., K.L. Wuensch (2013) SPSS and SAS programs for comparing Pearson correlation coefficients. *Behav Res Methods* 45, 3, 880-95
- Weber E. (1974) *Einführung in die Faktorenanalyse*. Gustav Fischer-Verlag Stuttgart.
- Welch B.L. (1937) The significance of the difference between two means when the Population variances are unequal. *Biometrika* 29, pp. 350-361.
- Welch B.L. (1951) On the Comparison of Several Mean Values: An Alternative Approach. *Biometrika* 38, pp. 330-336.
- Wellek S., M. Blettner (2012) On the Proper Use of the Crossover Design in Clinical Trials. *Deutsches Ärzteblatt International* 109 (15), pp. 276-281.
- Werner J. (1984) *Medizinische Statistik*. Urban & Schwarzenberg München Wien Baltimore.
- Westlake W.J. (1972) Use of Confidence Intervals in Analysis of Comparative Bioavailability Trials. *J. Pharm. Sci.* 61, pp. 1340-1341.
- Wilcox R. (2015) Comparing the variances of two dependent variables. *Journal of Stat. Distrib. and Applications*. pp. 2-7

- Wilcoxon F., R.A. Wilcoxon (1964) Some rapid approximate statistical procedure. Lederle Lab. NY.
- Wilks S.S. (1941) Determination of Sample Sizes for Setting Tolerance Limits. *Annals Math. Stat.* 12, pp. 91-96
- Winer B.J. (1971) *Statistical Principles in Experimental Design*. McGraw Hill New York.
- Wirtz M., F. Caspar (2002) *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Wishart D.W. (1984) *CLUSTAN*. G. Fischer - Verlag Stuttgart New York.
- Woolf B. (1957) The log likelihood ratio test. *Ann. Hum. Gen.* 21, pp.397-409.
- Youden W.J. (1950) An index for rating diagnostic tests. *Cancer* 3:32-35.
- Zhao Y.D. (2006) Sample size estimation for the van Elteren test - a stratified Wilcoxon-Mann-Whitney test. *Statist. in Med.* 25, pp. 2675-87.
- Zhou X.H., M.A. Obuchowski, D.K. McClish (2002) *Statistical methods in diagnostic medicine*. Wiley-Interscience.
- Zimmermann H. (1984) Die praktische Relevanz des McNemar-Tests. *Biometrical Journal* 26, 2, pp.219-21
- Zimmermann H. (1984) Erweiterung des Mantel-Haenszel-Testes auf $k > 2$ Gruppen. *Biom. J.* 26, 2, 223-4
- Zimmermann H. (1985) Exact Calculation of Permutational Distributions for Two Dependent Samples. *Biom. J.* 27, 3, pp. 349-52
- Zimmermann H. (1985) Exact Calculation of Permutational Distributions for Two Independent Samples. *Biom. J.* 27, 4, pp. 431-4
- Zimmermann H., V.W. Rahlfs (1980) Model Building and Testing for the Change Over Design. *Biom. J.* 22, pp.197-210.
- Zou G.Y. (2012) Sample size formulas for estimation intraclass correlation coefficients with precision and assurance. *Stat. in Med.* 31, pp. 3972-81.
- Zwiener I., M. Blettner, G. Hommel (2011) Survival Analysis. *Deutsches Arzteblatt International* 108, 10, pp. 163-169.

Datenmanagement und Datenbankmodul

Daten-Editor, Daten-Transformationen, Missing-Data-Code, Formelinterpreter, Import und Export von dBase-, ASCII- und Windows-Dateien (XLS, DBF, CSV, SAV und TXT) Selektion, Projektion, Addition, Relationieren, Sortieren, Klassifizieren (auch mit χ^2 -Test), Zählen.

Fallzahl- und Powerberechnung

Konfidenzintervalle, Ein- und Zweistichproben-Tests, Varianzanalyse, Kruskal-Wallis-Test, Schätzung von Wahrscheinlichkeiten, Vergleich binomialer Größen, Regression, Korrelation, Cross-Over-Analyse, Log-Rank-Test, Äquivalenztests, Many-One-Vergleiche, Epidemiologie (Odds Ratio, Relatives Risiko), Bland-Altman-Methodenvergleich, Bauer-Köhne-Design, Cluster-randomisierte Studien.

Randomisierung

Vollständige Randomisierung, Randomisierung in Blöcken, Cross-Over-Versuchspläne, Lateinische Quadrate, Zufallsstichproben (Gleich- und Normalverteilung).

Deskriptive Statistik

Durchschnitt, Standardabweichung, Extremwerte, Median, Quartilen, Variationskoeffizient etc., viele Graphiken.

Graphik

Histogramme (stetig, diskret, kumulativ, Pareto), Kreisdiagramme, Scattergram, diverse Box- und Bar-Plots, Zeitreihen-Plots, Gauß-Anpassung, Regressionsfunktion mit Konfidenzintervall, Hahn-Prognose-Intervall, Kaplan-Meier-Plot, Cluster- und Diskriminanzanalyse, Bland-Altman, Passing-Bablok- & Lin-Plot, Toleranzbereiche, ROC-Kurven, BMP-, JPG-, GIF- und PCX- Graphik-Export.

Konfidenzintervalle

Konfidenzintervalle für μ und $\mu_1-\mu_2$, Regression und Korrelation, Median (u.a. Moses- und Tukey-Intervalle), Survival-Analyse (Intervalle nach Hall und Wellner), Simultane Konfidenzintervalle (Hotelling), Binomial- und Poisson-Konfidenzintervalle für θ , $\theta_1-\theta_2$, θ_1/θ_2 , λ und $\lambda_1-\lambda_2$, Relatives Risiko, Odds-Ratio und viele andere Verfahren.

Verteilungen

χ^2 , t-, F- und Gauß-Verteilung mit Inversen, Nicht-zentrale Verteilungen, Binomial- und Poisson-Verteilung.

Test auf Gauß-Verteilung

χ^2 -Test, Shapiro-Wilk-Test (n beliebig!), Kolmogoroff-Smirnoff-Tests, David-Test, Ausreißer-Tests (u.a. Grubbs-Test), Mudholkar-Test auf p-variate Gauß-Verteilung.

Vierfelder- und c_{xr}-Kontingenztafel

χ^2 -Test, Fisher-Test (auch zweiseitig-exakt!), McNemar-Test, Median-Test, Cohen's Kappa, Pearson's Kontingenzkoeffizient, Cross-Over-Analyse, Konfidenzintervall für $\theta_1-\theta_2$, Mehrdimensionale Tafeln, Vergleich von Tafeln, Bewertung diagnostischer Tests, Äquivalenztest, Mantel-Haenszel-Test, Tests und Konfidenzintervalle für Odds-Ratio und Relatives Risiko, Diagnostische Tests, Terwilliger-Ott- und Bowker's Test und einige andere.

Regressions- und Korrelationsanalyse

Pearson-Regression und Korrelation, Multiple, Logistische & Polynom. Regression mit Abbau, Vergleich von Regressionen, Multiple & Partielle Korrelation, Rang-Regression, Spearman's Rang-Korrelation, Residuen-Analyse, Korrelationsmatrix, Cronbach's α und IntraClass-Korrelation (Reliabilitätsanalyse), IntraCluster-Korrelation (Cluster-randomisierte Studien), Lin's CC, zahlreiche Graphiken.

Parametrische Testverfahren

Ein- und Zweistichproben-t-Test, Welch-Test, Ein- und Zweifaktorielle Varianz- und Kovarianzanalyse incl. multipler Vergleiche (Scheffé), Lineare Kontraste, Orthogonale Polynome, Hotelling's Tests, Cross-Over-Analyse (Grizzle), Äquivalenztests, Intracluster- und Intraclass-Korrelation.

Nicht-parametrische Testverfahren

Wilcoxon- & Wilcoxon-Mann-U-Whitney-Test, Tukey- und Moses-Konfidenzintervalle, Hodges-Lehmann- und Mann-Whitney-Schätzer mit KIs, exakte p-Werte, Kolmogoroff-Smirnoff-Test, Kruskal-Wallis-Test (post-hoc-Tests: Dunn-Holm, Conover-Iman, Nemenyi), Friedman-Test (post-hoc-Tests: Wilcoxon-Wilcox, Schaich-Hamerle, Conover-Iman), Dixon-Mood's Vorzeichen-Test, Test auf Gleichverteilung, Anpassungstest, Cochran's Q-Test, Vergleich von Häufigkeiten bzw. Poisson-Verteilungen, Cross-Over-Analyse nach Koch, Äquivalenztests, Mantel-Haenszel-Test.

Survival-Analyse

Life-Tables (Kaplan-Meier-Schätzer und Hall-Wellner-Konfidenzbereiche, Graphik), Gehan-Test, Log-Rank-Test (Peto-Pike, Cox-Mantel), Stratifizierter Log-Rank-Test mit Kaplan-Meier-Schätzern, Relativer Hazard, k-Gruppen.

Cluster- und Diskriminanzanalyse

KMeans- & Single-Linkage-Clusteranalyse mit verschiedenen Metriken, Lineare & Quadratische Diskriminanzanalyse, Stoller's nicht-parametrische Diskriminanzanalyse, Diskriminanzanalyse mit Abbaumodell.

Zeitreihen-Analyse

ANOVA, Mann-Test, Iterations- und Phasenhäufigkeitstest, Simultane Tests, Graphik (Einzelverläufe und Box-Plots der Verläufe).

Faktorenanalyse

Explorative und konfirmatorische Faktorenanalyse mit Zentroid- und Varimax-Methode.

Konfigurationsfrequenzanalyse

Ein- und Mehrstichproben-Konfigurationsfrequenzanalyse (χ^2 , Binomial- und Lehman-Test).

Pharmakologie und Bioäquivalenz

Berechnung u.a. der Größen AUC(t_1, t_n), AUC(t_1, ∞), C_{max}, t_{max}, t_{1/2}, Cross-Over-Analyse, Inklusionsregel u.a. mit nichtparametrischen Tukey- und Moses-Intervallen, EC50.

Diagnostische Tests und Methodenvergleiche

Falsch-positiv und -negativ, Sensitivität, Spezifität, Effizienz, Prädiktive Werte, Youden-Index (alle mit Konfidenzintervallen), ROC-Kurven & Vergleiche mit DeLong, Bland-Altman-, Passing-Bablok- und Lin-Verfahren mit Tests.

Toleranzbereiche

Parametrische und nicht-parametrische Toleranzbereiche (uni- und bivariat), Perzentilen, Quartilen, Box-Cox-Transformationen.

Multiple Testverfahren & Meta-Analyse

Multiple Testverfahren (Bonferroni, Holm, Hommel, Simes, Rüger), Meta-Analyse (Tippett, Fisher, Inverse Normalverteilung), Interimsanalysen nach Pocock, O'Brien-Fleming, Bauer-Köhne-Design, Benjamini-Hochberg's FDR.

epsilon-Verlag GbR

<https://www.bias-online.de>



Aktuelle Informationen zu **BiAS.** finden Sie im Internet unter <https://www.bias-online.de>

BiAS. ist als Download und auf CD erhältlich

© epsilon 1989-2020 • ISBN 3-9803214-8-7