

**Biometrische Methoden zur Planung und Auswertung
von Sicherheitsstudien**

Habilitationsschrift
zur Erlangung der Lehrbefugnis
für das Lehrgebiet
Biometrie

Dem Fachbereich Statistik der Universität Dortmund
vorgelegt von
Dr. Dieter Hauschke
aus Konstanz

Dortmund 1999

Meiner verstorbenen Mutter Josefina Hauschke in dankbarer Erinnerung gewidmet.

Mein Dank gilt Herrn Prof. Dr. Hothorn für die Anregung und Betreuung der vorliegenden Arbeit.

Ebenso danke ich Herrn Prof. Dr. Schach, Herrn Prof. Dr. Brunner und Herrn Prof. Dr. Kunert für ihr Interesse, das sie der Arbeit entgegengebracht haben. Für anregende Diskussionen danke ich meinen Kollegen, Herrn Burke, Herrn Diletti und Herrn Dr. Neuhäuser.

**Biometrische Methoden zur Planung und Auswertung
von Sicherheitsstudien**

Habilitationsschrift
zur Erlangung der Lehrbefugnis
für das Lehrgebiet
Biometrie

Dem Fachbereich Statistik der Universität Dortmund
vorgelegt von
Dr. Dieter Hauschke
aus Konstanz

Dortmund 1999

Inhaltsverzeichnis

1	Einleitung und Übersicht	1
2	Einführung in die Problemstellung.....	4
3	Zweistichprobensituation.....	12
	3.1 Sicherheitsnachweis für die Differenz von Lokationsparametern.....	12
	3.1.1 Einseitiges Testproblem.....	12
	3.1.2 Zweiseitiges Testproblem	15
	3.1.3 Gütebestimmung und Fallzahlplanung	17
	3.2 Sicherheitsnachweis für den Quotienten von Lokationsparametern....	25
	3.2.1 Einseitiges Testproblem.....	25
	3.2.2 Zweiseitiges Testproblem	27
	3.2.3 Gütebestimmung und Fallzahlplanung	28
4	Crossover-Situation.....	38
	4.1 Sicherheitsnachweis für die Differenz von Lokationsparametern.....	39
	4.1.1 Einseitiges Testproblem.....	40
	4.1.2 Zweiseitiges Testproblem	44
	4.1.3 Gütebestimmung und Fallzahlplanung	45
	4.2 Sicherheitsnachweis für den Quotienten von Lokationsparametern....	49
	4.2.1 Einseitiges Testproblem.....	49
	4.2.2 Zweiseitiges Testproblem	54
	4.2.3 Gütebestimmung und Fallzahlplanung	56
5	K-Stichproben <i>many-to-one</i>-Situation.....	64
	5.1 <i>Simple-tree</i> Alternative	64
	5.1.1 Globaler Sicherheitsnachweis	64
	5.1.1.1 Einseitiges Testproblem.....	64
	5.1.1.2 Zweiseitiges Testproblem.....	66
	5.1.2 Partieller Sicherheitsnachweis.....	68
	5.1.2.1 Einseitiges Testproblem.....	68
	5.1.2.2 Zweiseitiges Testproblem.....	74
	5.2 <i>Simple-order</i> Alternative	76
	5.2.1 Bestimmung der maximal sicheren Dosis.....	77
	5.2.1 Alternative Konzepte	79

6	Konzept der Bioäquivalenz	81
	6.1 Modell, Testproblem und Entscheidungsverfahren.....	81
	6.2 Anwendungsbeispiel.....	83
7	Zusammenfassung und Ausblick	87
8	Literaturverzeichnis	89
9	Anhang: SAS-Programme zur Gütebestimmung.....	97

1 Einleitung und Übersicht

Absence of evidence is not evidence of absence (Altman und Bland, 1994)

In experimentellen Untersuchungen von Pharmaka muß die Wirksamkeit und die Unbedenklichkeit einer neuen Substanz im Vergleich zu einer Kontrolle gezeigt werden. Während für den Wirksamkeitsnachweis ein umfangreiches Schrifttum vorliegt, hat die adäquate biometrische Auswertung von Sicherheitsstudien erst in den letzten Jahren immer größere Bedeutung gewonnen. Ein Grund dafür ist die Vorgehensweise, daß in der Praxis bislang ein *proof of hazard* durchgeführt wird, wobei die Nullhypothese H_0 : *Substanz unbedenklich* gegen die Alternative H_1 : *Substanz bedenklich* getestet wird. Kann die Nullhypothese nicht abgelehnt werden, so wird häufig die Unbedenklichkeit der Substanz gefolgert. Der gravierende Nachteil des *proof of hazard* besteht darin, daß nur die Wahrscheinlichkeit der irrtümlichen Verwerfung der Nullhypothese (Produzentenrisiko) direkt durch den vorgegebenen Fehler 1. Art **a** begrenzt ist.

Weitaus wichtiger ist jedoch die Kontrolle des Konsumentenrisikos, nämlich der irrtümlichen Annahme, daß die Substanz unbedenklich ist. Es stellt sich unmittelbar die Frage, ob ohne Berücksichtigung des Fehlers 2. Art **b** dieses Entscheidungsverfahren korrekt sein kann, zumal in Sicherheitsstudien oftmals eine Begründung der gewählten Fallzahlen nicht auf einer vorgegebenen Güte (*Power*) von $1 - \mathbf{b}$ basiert, sondern sich statt dessen an behördlichen Mindestanforderungen orientiert. Dabei wird übersehen, daß ein relevanter Unterschied bei zum Beispiel kleinen Fallzahlen und/oder großer Versuchsstreuung zu einem falsch-negativen Testergebnis führen kann; statistische Nichtsignifikanz wäre dann nicht kongruent mit vorliegender Unbedenklichkeit.

Vom ethischen, wissenschaftlichen und regulatorischen Standpunkt her gesehen, muß eine sicherheitsrelevante Studie eine ausreichende Güte aufweisen (Healey, 1987). Dieser Tatsache wurde erstmals in dem von der U.S. Food and Drug Administration (1993) erstellten Entwurf der *Guidance for Toxicity Tests* folgendermaßen Rechnung getragen: *A power calculation should be provided for tests that failed to reject the null hypothesis, particularly to justify the adequacy of the sample size.*

Eine vor Versuchsbeginn vorgenommene Gütebetrachtung dient generell dazu sicherzustellen, daß vorhandene Effekte mit ausreichender Wahrscheinlichkeit entdeckt werden können. Wird jedoch die entsprechende Nullhypothese nicht verworfen, so kann eine *a-posteriori* vorgenommene Güteermittlung nur als eine indirekte Methode zum Nachweis

von nicht vorhandenen Unterschieden betrachtet werden (Hayes, 1987; Muller und Benignus, 1992).

Ein weiterer Nachteil des *proof of hazard* besteht ferner darin, daß ein irrelevanter Unterschied zu einem signifikanten Testergebnis führen kann; statistische Signifikanz wäre dann inkongruent mit Unbedenklichkeit der Substanz. Dieses Phänomen tritt zum Beispiel oftmals bei der Analyse von Mutagenitätsuntersuchungen auf, da in einzelnen Mutagenitätsassays die Zelle als Versuchseinheit vorliegt und daher große Fallzahlen bzw. wegen der diskreten Verteilung der Endpunkte eine Vielzahl von Bindungen und somit geringe Varianzschätzungen resultieren (Hauschke *et al.*, 1997).

Insgesamt gesehen kann daher der *proof of hazard* sowohl behördlichen (Kontrolle des Konsumentenrisikos) als auch producentenspezifischen (Ausschluß irrelevanter Unterschiede) Anforderungen nicht gerecht werden. Der wesentliche Grund besteht darin, daß auf der Neyman-Pearson Theorie basierende statistische Testverfahren das Vorliegen der Alternativhypothese anhand der empirischen Evidenz gegen die Nullhypothese demonstrieren. Klar beschrieben wurde dies bereits von Fisher (1935): *The null hypothesis is never proved or established, but is possibly disproved in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis.*

Es ist daher verwunderlich, daß erst Mitte der achtziger Jahre von Bross (1985) für den direkten Sicherheitsnachweis das Konzept des *proof of safety* vorgestellt wurde. Hierbei wird als Nullhypothese H_0 : *Substanz bedenklich* und als Alternative H_1 : *Substanz unbedenklich* gewählt und desweiteren wird ein Schwellenwert (nicht mehr akzeptables Risiko) mit in die Formulierung des Testproblems einbezogen. Bross (1985) und Millard (1987) benutzten den *proof of safety* für die Frage, ob gegen die Wohnbarkeit einer ehemals belasteten Umgebung Sicherheitsbedenken bestehen. Beide Autoren kamen zum Schluß, daß für den direkten Sicherheitsnachweis umfangreichere Fallzahlen als bisher zu Grunde gelegt werden müssen. Basierend auf dieser Erkenntnis, empfahlen Holland und Ordoukhani (1990) den klassischen Nachweis des *proof of hazard* beizubehalten und dabei für den Fehler 1. Art höhere Niveaus zu wählen und zusätzlich gleich dem Fehler 2. Art zu setzen, d.h. $\mathbf{a} = \mathbf{b} = 0.10$ oder sogar in bestimmten Fällen $\mathbf{a} = \mathbf{b} = 0.20$.

Für den Bereich toxikologischer Substanzuntersuchungen wurden von Erickson und McDonald (1995), Stallard und Whitehead (1996), Hauschke (1995, 1997), Hansson (1997), Neuhäuser und Hothorn (1995, 1997) und Hauschke und Hothorn (1998) weitergehende Arbeiten für den *proof of safety* publiziert. Erickson und McDonald legten in ihrer Arbeit den Schwerpunkt auf die Bestimmung von Schwellenwerten bei vorgegebener Anzahl von

Behandlungsgruppen, unterschiedlichen Fallzahlen und Variationskoeffizienten. Stallard und Whitehead schlugen als Alternative zur klassischen Auswertung von Kanzerogenitätsstudien mittels Petoanalyse (Peto *et al.*, 1980) einen modellabhängigen Ansatz zur adäquaten statistischen Beurteilung vor, wobei sie sich jedoch nur auf den Vergleich einer Behandlungsgruppe mit einer Kontrollgruppe konzentrierten. Hansson diskutierte nochmals die Durchführbarkeit und die resultierenden Konsequenzen des direkten Sicherheitsnachweises aus wissenschaftstheoretischer Sicht. Neuhäuser und Hothorn verwendeten den direkten Ansatz zur Kontrolle des Konsumentenrisikos bei der Auswertung der Dosis-Wirkungs-Abhängigkeit des Ames Mutagenitätsassay. In den Arbeiten von Hauschke und Hothorn wurden die den Testverfahren entsprechenden Konfidenzintervalle für die Zwei- und Mehrstichprobensituation im Falle partiell bzw. vollständig geordneter Alternativen dargestellt.

Das Ziel der vorliegenden Arbeit ist eine systematische Darstellung statistischer Verfahren für den direkten Nachweis der Unbedenklichkeit von Substanzen. Die Problematik des falsch-negativen bzw. falsch-positiven Ergebnisses bei der klassischen Hypothesenformulierung wird im zweiten Kapitel zunächst anhand von Datensätzen aus der Toxikologie verdeutlicht. Im folgenden dritten Kapitel wird die statistische Methodik für die Zweistichprobensituation ausführlich dargestellt. Ein wesentlicher Punkt ist dabei die Herleitung der exakten Güte- und Fallzahlbestimmung mit zusätzlicher Angabe geeigneter Approximationsformeln.

Das *Crossover*-Design kann durch eine elementare Transformation der zugrundeliegenden Zufallsvariablen auf eine Zweistichprobensituation zurückgeführt werden. Daher werden die im dritten Kapitel hergeleiteten Verfahren im vierten Kapitel auf die *Crossover*-Situation übertragen.

Im fünften Kapitel werden die biometrischen Methoden für den *proof of safety* in der Mehrstichprobensituation für partiell bzw. vollständig geordnete Alternativen hergeleitet. Das zentrale Anliegen besteht dabei in der Ermittlung sicherer Substanzen bzw. in der Bestimmung der maximal sicheren Dosis.

Nach der Darstellung des Konzeptes der Bioäquivalenz im sechsten Kapitel, werden im siebten Kapitel die wesentlichen Ergebnisse der Arbeit zusammengefaßt und es wird im Ausblick auf noch offene Probleme hingewiesen.

2 Einführung in die Problematik

Zur Veranschaulichung der Fragestellung wird die Zweistichprobensituation mit n_0 bzw. n_1 Beobachtungen pro Gruppe betrachtet. Die zur Sicherheitsbeurteilung relevante Zielgröße sei beschrieben durch die Zufallsvariable X_{ij} , die unabhängig identisch verteilt ist (u.i.v. \sim) mit der stetigen Verteilungsfunktion $F_i(x) = F(x - \mathbf{m}_i)$, $i = 0, 1$ und $j = 1, \dots, n_i$. Dabei kennzeichnen \mathbf{m}_i die Lageparameter, der Index $i = 0$ die Kontrolle (Vehikel- oder Negativkontrolle) und der Index $i = 1$ die Behandlungsgruppe. Desweiteren wird angenommen, daß sich ein möglicher sicherheitsrelevanter Substanzeffekt durch höhere Ausprägungen der Zielgröße auszeichnet, d.h. $\mathbf{m}_1 > \mathbf{m}_0$. Das traditionelle Testproblem des *proof of hazard* läßt sich dann folgendermaßen formulieren:

$$H_0: \mathbf{m}_1 - \mathbf{m}_0 \leq 0 \quad (\text{Substanz in der vorliegenden Untersuchung unbedenklich})$$

versus

$$H_1: \mathbf{m}_1 - \mathbf{m}_0 > 0 \quad (\text{Substanz in der vorliegenden Untersuchung bedenklich}).$$

Die einseitige Formulierung des Testproblems in Sicherheitsstudien ist meist aus inhaltlichen Überlegungen gerechtfertigt (Dubey, 1991) und sollte ferner zur Minimierung des Fehlers 2. Art verwendet werden (Hothorn, 1991, 1992; Hothorn und Hauschke, 1998).

Trifft man nun die parametrische Verteilungsannahme $F_i(x) = \mathbf{F}\left(\frac{x - \mathbf{m}_i}{\mathbf{s}}\right)$, wobei $\mathbf{F}(\cdot)$ die Verteilungsfunktion der Standardnormalverteilung bezeichnet, und nimmt man ferner an, daß die Varianz \mathbf{s}^2 zwar unbekannt aber in beiden Gruppen gleich ist, so kann obiges Testproblem mittels des Zweistichproben- t -Tests überprüft werden. Die Nullhypothese H_0 wird genau dann verworfen, falls

$$t_1 = \frac{\bar{X}_1 - \bar{X}_0}{S \sqrt{\frac{1}{n_0} + \frac{1}{n_1}}} \geq t_{\alpha, n_0 + n_1 - 2},$$

wobei

$$\bar{X}_0 = \frac{1}{n_0} \sum_{j=1}^{n_0} X_{0j}, \quad \bar{X}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} X_{1j}$$

die empirischen Mittelwerte der beiden Stichproben bezeichnen,

$$S^2 = \frac{1}{n_0 + n_1 - 2} \sum_{i=0}^1 \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

die gepoolte empirische Varianz ist und der kritische Wert

$t_{a,n}$ dem $(1 - \alpha)$ -Quantil der zentralen t -Verteilung mit n Freiheitsgraden entspricht.

In Abbildung 1 wird die Abhängigkeit der Güte des Zweistichproben- t -Tests von dem relativ zur Kontrollgruppe dargestellten Variationskoeffizienten ($CV_0 = \frac{S}{m_0}$) und der Alternativen dargestellt, wobei die Differenz $m_1 - m_0$ ebenfalls als Anteil von m_0 ausgedrückt wird. Mit $\alpha = 0.05$ und jeweils 10 Versuchseinheiten pro Gruppe wurden für subchronische Toxizitätsuntersuchungen übliche Größen gewählt. In Tabelle 1 werden exemplarisch die entsprechenden Gütwerte an der Stelle $m_1 - m_0 = 0.1m_0$ für ausgewählte haematologische Endpunkte an Ratten angegeben. Es ist offensichtlich, daß bei den in der Praxis vorkommenden Variabilitäten nur dramatische Effekte mit hinreichend hoher Wahrscheinlichkeit entdeckt werden können: Nimmt der Variationskoeffizient zu bzw. ist ein möglicher Unterschied zwischen den Lokationsparametern nicht sehr ausgeprägt, so verringert sich die Güte des Entscheidungsverfahrens dramatisch.

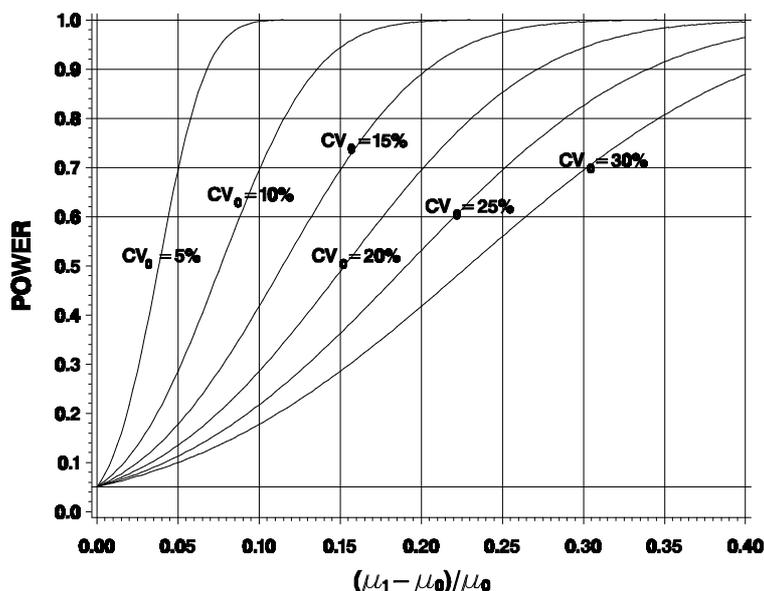


Abbildung 1: Die Gütefunktion des einseitigen Zweistichproben- t -Tests t_1 für $\alpha = 0.05$, $n_0 = n_1 = 10$ und verschiedene Variationskoeffizienten CV_0

Tabelle 1: Gütewerte für ausgewählte haematologische Endpunkte von Ratten (Chanter *et al.*, 1987) an der Stelle $m_1 - m_0 = 0.1m_0$, $\alpha = 0.05$, $n_0 = n_1 = 10$

Endpunkt	CV_0 (%)	Güte
Haemoglobin	4	0.9999
Prothrombin Zeit	9	0.7717
Thrombozyten	15	0.4167
Leukozyten	28	0.1905
Neutrophile	66	0.0936

Zur praktischen Illustration diene die folgende chronische Toxizitätsprüfung an weiblichen Wistar-Ratten, in der unter anderem die Serum-ASAT-Werte nach einer Applikationsdauer von sechs Monaten erhoben wurden. In Tabelle 2 sind die Einzelwerte und in Abbildung 2 die entsprechenden Box-plots für die Kontrolle und für die mit der Substanz behandelten Gruppe dargestellt. Geplant waren 19 Tiere pro Gruppe; infolge mortalitätsbedingter Ausfälle innerhalb der Behandlungsgruppe ergaben sich jedoch unterschiedliche Fallzahlen.

Tabelle 2: Serum-ASAT-Werte ($mmol/L$) weiblicher Wistar-Ratten

	Kontrollgruppe	Behandlungsgruppe
	1.33	1.53
	1.78	1.75
	1.53	2.12
	1.95	2.83
	1.83	2.58
	1.47	2.37
	1.87	2.92
	1.55	2.08
	2.58	3.01
	2.17	2.67
	1.97	2.06
	1.62	2.25
	2.25	3.08
	3.53	2.01
	2.92	1.58
	1.78	
	2.22	
	2.55	
	2.75	
Anzahl Tiere	19	15
Mittelwert	2.09	2.32
Standardabweichung	0.57	0.51

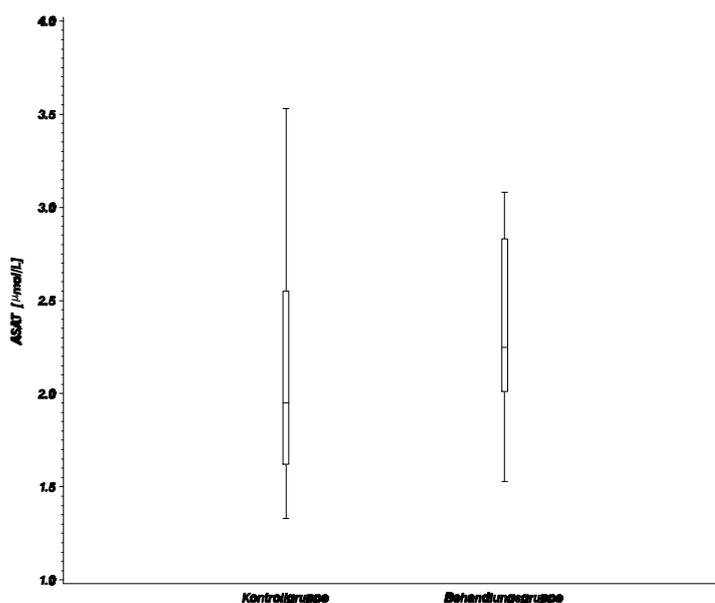


Abbildung 2: Box-Plots für die Serum-ASAT-Werte (mmol/L) weiblicher Wistar-Ratten

Die zugrundeliegende Fragestellung beinhaltet eine mögliche sicherheitsrelevante Erhöhung der ASAT-Werte, dabei wurde eine Erhöhung von $d=0.45$ als minimal biologisch relevant angesehen (Hothorn, 1992). Führt man den klassischen Test auf

$$H_0: m_1 - m_0 \leq 0 \quad (\text{keine Erhöhung in den ASAT-Werten})$$

versus

$$H_1: m_1 - m_0 > 0 \quad (\text{erhöhte ASAT-Werte}),$$

mittels des t -Tests durch, so erhält man einen p -Wert von 0.108. In der Praxis folgert man nun aus der Nichtverwerfung der Nullhypothese das fehlende Vorliegen der Erhöhung der ASAT-Werte und übersieht dabei die folgende von Altman und Bland (1994) treffend formulierte Tatsache: *Absence of evidence is not evidence of absence*. Vertauscht man Null- und Alternativhypothese und führt man zusätzlich mit $d=0.45$ einen Schwellenwert (nicht mehr akzeptable Erhöhung) ein, so erhält man das Testproblem:

$$H_0: m_1 - m_0 \geq 0.45 \quad (\text{relevant erhöhte ASAT-Werte})$$

versus

$$H_1: m_1 - m_0 < 0.45 \quad (\text{keine relevante Erhöhung in den ASAT-Werten}).$$

Die Auswertung dieser verschobenen Nullhypothese ergibt nun einen p -Wert von 0.132, somit kann nicht auf eine unbedenkliche Erhöhung der ASAT-Werte geschlossen werden.

Es sei darauf hingewiesen, daß das Testen verschobener Nullhypothesen erstmals von Victor (1987) zur Beurteilung der klinischen Relevanz für den Nachweis der Wirksamkeit in klinischen Studien verwendet wurde (siehe dazu auch Hauschke *et al.*, 1999a).

Die oben gewählte einseitige Formulierung des Testproblems ist in der überwiegenden Anzahl von Sicherheitsstudien angezeigt. Es existieren jedoch auch Fragestellungen bei denen eher ein zweiseitiges Testproblem relevant ist. Ein Beispiel dafür ist der Aspekt der Reversibilität möglicher negativer Effekte bei der Beurteilung von Substanzen. Üblicherweise werden in derartigen Versuchen zwei zusätzliche Behandlungsgruppen mitgeführt. Es handelt sich dabei um eine unbehandelte Kontrolle (Negativgruppe) und um eine weitere Gruppe, die mit der höchsten Dosis der Substanz behandelt wird. Die Versuchseinheiten werden nach Abschluß des eigentlichen Experimentes nicht direkt untersucht, sondern erst nach einer ausreichend langen Erholungsphase. In Tabelle 3 sind die terminalen Körpergewichte männlicher Ratten der beiden Erholungsgruppen aufgelistet (Hauschke und Hothorn, 1998) und in Abbildung 3 sind die zugehörigen Box-Plots dargestellt.

Tabelle 3: Terminale Körpergewichte (g) männlicher Ratten in den Erholungsgruppen

	Kontrollgruppe	Hohe Dosis
	519.8	465.0
	535.7	492.2
	491.7	466.3
	508.7	555.5
	565.5	528.1
	485.8	483.8
	520.7	502.8
	507.5	468.1
	428.9	519.2
	466.0	500.2
Anzahl Tiere	10	10
Mittelwert	503.0	498.1
Median	508.1	496.2
Standardabweichung	37.9	29.7

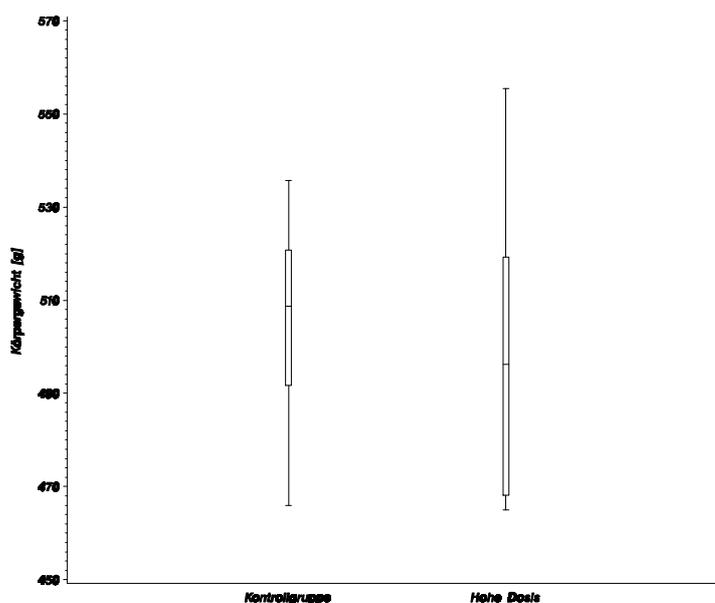


Abbildung 3: Box-Plots für die terminalen Körpergewichte (g) männlicher Ratten

Die inhaltliche Fragestellung beinhaltet die Vergleichbarkeit der terminalen Körpergewichte beider Gruppen. Das klassische Testproblem lautet somit

$$H_0: \mathbf{m}_1 - \mathbf{m}_0 = 0 \quad \text{versus} \quad H_1: \mathbf{m}_1 - \mathbf{m}_0 \neq 0$$

und H_0 wird genau dann verworfen, falls $|t_1| \geq t_{\frac{\alpha}{2}, n_0 + n_1 - 2}$. Führt man den Test nun mit den

obigen Daten durch, so erhält man einen p -Wert von 0.75. In der Praxis folgert man nun aus der Nichtverwerfung der Nullhypothese irrtümlicherweise die Vergleichbarkeit der Körpergewichte. Analog zur einseitigen Fragestellung führt die korrekte Formulierung des Testproblems auf das zweiseitige Äquivalenzproblem:

$$H_0: \mathbf{m}_1 - \mathbf{m}_0 \leq \mathbf{d}_1 \quad \text{oder} \quad \mathbf{m}_1 - \mathbf{m}_0 \geq \mathbf{d}_2$$

versus

$$H_1: \mathbf{d}_1 < \mathbf{m}_1 - \mathbf{m}_0 < \mathbf{d}_2,$$

wobei das Intervall $(\mathbf{d}_1, \mathbf{d}_2)$, $\mathbf{d}_1 < 0 < \mathbf{d}_2$, den Sicherheitsbereich definiert. Die entsprechende Auswertung dieses Testproblems wird im dritten Kapitel detailliert beschrieben.

Eine weitere Fragestellung in Sicherheitsuntersuchungen ist die Bestimmung einer maximal sicheren Dosis (Hauschke, 1995; Hothorn und Hauschke, 2000; Tamhane *et al.*, 1999). Dazu wird ein Experiment mit einer Vehikelkontrolle (Versuchseinheiten erhalten nur

das Trägermedium, zum Beispiel das Lösungsmittel) sowie meist drei Dosierungen der Substanz durchgeführt. Zur Veranschaulichung diene die folgende Untersuchung (Tabelle 4), in der Tieren die Substanz in unterschiedlichen Dosierungen verabreicht wird. Die Zielvariable ist der Anteil erkrankter Tiere; Experiment I und II unterscheiden sich nur durch die Fallzahlen (Gaylor, 1983) von 20 bzw. 60 pro Gruppe.

Eine in der Praxis noch übliche Auswertungsstrategie basiert auf dem Abschlußtestprinzip (Marcus *et al.*, 1976) und besteht darin, daß im ersten Schritt der Cochran-Armitage-Test mit allen Behandlungsgruppen durchgeführt wird (Lüdin, 1985). Ist der p -Wert kleiner oder gleich 0.05, so wird der Test ohne die höchste Dosis wiederholt. Im Falle einer erneuten Signifikanz zum Niveau von 0.05 erfolgt der entsprechende Vergleich zwischen Kontrolle und kleinster Dosis. Liegt ein nichtsignifikantes Ergebnis ($p > 0.05$) vor, so wird die in dem Vergleich vorliegende höchste Dosis als maximal sichere Dosis bezeichnet. Es sei darauf hingewiesen, daß diese Dosis in der Literatur auch mit *NOED* (*no observed effect dose*) oder *NOEL* (*no observed effect level*) bezeichnet wird (Williams, 1971, 1972; Shirley, 1977; Tukey *et al.*, 1985; Rom *et al.*, 1994).

Tabelle 4: Dosis-Wirkungsanalyse in zwei Experimenten (Gaylor, 1983)

Behandlungsgruppe	Anteil erkrankter Tiere	
	Experiment 1	Experiment 2
Kontrolle	0/20	0/60
Kleine Dosis D_1	1/20	3/60
Mittlere Dosis D_2	2/20	6/60
Hohe Dosis D_3	10/20	30/60

Die exakten, konditionalen p -Werte sind in der Tabelle 5 aufgelistet; zusätzlich wurde die asymptotische Güte nach Nam (1987) berechnet. Während im ersten Experiment mit jeweils 20 Tieren pro Gruppe die mittlere Dosis D_2 als maximal sichere Dosis resultiert, ist in dem zweiten Experiment mit Verdreifachung der Fallzahlen, aber ansonsten gleichbleibenden Anteilen, die kleinste Dosis D_1 die maximal sichere Dosis. Betrachtet man die jeweiligen Güten, so ist es offensichtlich, daß sich durch entsprechende Wahl der Stichprobengrößen unterschiedliche, maximal sichere Dosierungen ergeben. Je kleiner die Stichprobe, desto größer ist die maximal sichere Dosis; dieses stellt jedoch genau das Entgegengesetzte dar, was

experimentell beabsichtigt ist. Daher wird im fünften Kapitel die probate Auswertungsstrategie zur Bestimmung der maximal sicheren Dosis hergeleitet.

Tabelle 5: Exakte p -Werte des Cochran-Armitage-Tests

Behandlungsgruppe	Anteil	Experiment 1	Experiment 2
Kleine Dosis D_1	0.05	–	0.121
Mittlere Dosis D_2	0.10	0.144 (0.28 [§])	0.009 (0.75 [§])
Hohe Dosis D_3	0.50	< 0.001	< 0.001

[§] Asymptotische Güte

Es sei darauf hingewiesen, daß in der vorliegenden Arbeit das Lokationsmodell mit stetigen Verteilungen betrachtet wird, d.h. es wird angenommen, daß ein Behandlungseffekt nur eine Verschiebung der zugrundeliegenden Verteilungsfunktion bewirkt. Die in den jeweiligen Modellen vorgestellten Methoden beziehen sich somit ausschließlich auf Lokationsvergleiche; hinsichtlich entsprechender Dispersionsvergleiche wird auf die Arbeiten von Guilbaud (1993) und Bauer und Bauer (1994) verwiesen.

In der Arzneimittelforschung, insbesondere bei der Entwicklung von Psychopharmaka, liegen oftmals Ordinalskalen vor, anhand derer eine Beurteilung hinsichtlich Wirksamkeit und Sicherheit erfolgt. Die Herleitung entsprechender Testverfahren für den Äquivalenznachweis erweist sich schon im Falle dichotomer Zufallsvariablen als nicht trivial und erst in jüngster Zeit sind exakte Methoden publiziert worden (Chan, 1998; Kopit und Berger 1998). Die Verallgemeinerung dieser Verfahren auf ordinal skalierte Daten wird ein Forschungsschwerpunkt der kommenden Jahre sein.

3 Zweistichprobensituation

Nach den vorhergehenden Ausführungen ist es evident, daß in Sicherheitsstudien die Kontrolle der irrtümlichen Annahme der Unbedenklichkeit von übergeordneter Bedeutung ist. Daher wird in diesem Kapitel der direkte Sicherheitsnachweis für die Zweistichprobensituation vorgestellt, wobei der benötigte Schwellenwert \mathbf{d} zunächst als bekannt vorausgesetzt wird und die Formulierung des Testproblems mittels der Differenz der Lokationsparameter erfolgt. Diese Vorgehensweise wird durch den Index \mathbf{d} in der Formulierung der Hypothesen und Teststatistiken gekennzeichnet. Die Angabe eines Schwellenwertes in absoluten Einheiten erweist sich jedoch in der Praxis häufig als sehr schwierig. Daher werden im folgenden zusätzlich Verfahren dargestellt, die es ermöglichen, diesen Wert auch relativ zum Erwartungswert der Kontrollgruppe zu definieren. Als Konsequenz ergibt sich, daß der Schwellenwert - absolut gesehen - unbekannt ist und das Testproblem nur durch geeignete Verfahren hinsichtlich des Quotienten von Lokationsparametern adäquat zu lösen ist; diese werden im folgenden durch den Index \mathbf{q} kenntlich gemacht.

3.1 Sicherheitsnachweis für die Differenz von Lokationsparametern

3.1.1 Einseitiges Testproblem

Für den Vergleich einer Behandlungsgruppe mit der Kontrollgruppe stellt sich das einseitige Testproblem wie folgt dar:

$$H_0^{\mathbf{d}}: \mathbf{m}_1 - \mathbf{m}_0 \geq \mathbf{d} \quad (\text{Substanz in der vorliegenden Untersuchung bedenklich})$$

versus

$$H_1^{\mathbf{d}}: \mathbf{m}_1 - \mathbf{m}_0 < \mathbf{d} \quad (\text{Substanz in der vorliegenden Untersuchung unbedenklich}),$$

wobei $(-\infty, \mathbf{c})$, $\mathbf{c} > 0$, den Bereich des nicht relevanten Unterschiedes festlegt. Auf die folgenden zwei Aspekte sei nochmals hingewiesen: Zum einen wird ein irrelevanter Unterschied zugelassen, und zum anderen wird die zu zeigende Vermutung als Alternative $H_1^{\mathbf{d}}$ formuliert. Letzteres hat zur Folge, daß die Wahrscheinlichkeit für die irrtümliche Annahme der Unbedenklichkeit direkt durch den Fehler 1. Art \mathbf{a} , welcher *a-priori* vorgebar ist, begrenzt wird. Im Falle der Normalverteilung mit homogenen, aber unbekanntem Varianzen, d.h. $F_i(x) = F\left(\frac{x - \mathbf{m}_i}{\mathbf{s}}\right)$, $i = 0, 1$, wird die Nullhypothese $H_0^{\mathbf{d}}$ abgelehnt, falls

$$t_1^d = \frac{\bar{X}_1 - \bar{X}_0 - \mathbf{d}}{S \sqrt{\frac{1}{n_0} + \frac{1}{n_1}}} \leq -t_{\mathbf{a}, n_0+n_1-2}.$$

Die Durchführung des Tests verlangt nun im Gegensatz zum klassischen *proof of hazard* die explizite Einbeziehung des Schwellenwertes \mathbf{d} in den Zähler der Teststatistik. Sollte es jedoch nicht möglich sein, diesen Wert *a-priori* zu definieren, so kann die Berechnung von entsprechenden Konfidenzintervallen eine Alternative darstellen: Kann \mathbf{d} somit nur *a-posteriori* definiert werden, so ist der Schluß auf Unbedenklichkeit zum Niveau \mathbf{a} möglich, falls die obere Grenze \mathbf{d}_\ddagger des einseitigen $100(1-\mathbf{a})\%$ -Konfidenzintervalles für die Differenz $\mathbf{m}_1 - \mathbf{m}_0$ unterhalb von \mathbf{d} liegt:

$$(-\infty, \mathbf{d}_\ddagger] = \left(-\infty, \bar{X}_1 - \bar{X}_0 + t_{\mathbf{a}, n_0+n_1-2} S \sqrt{\frac{1}{n_0} + \frac{1}{n_1}} \right] \subset (-\infty, \mathbf{d}).$$

Verzichtet man auf die Annahme der Normalverteilung und postuliert man nur das Lokationsmodell X_{ij} u.i.v. $\sim F_i(x) = F(x - \mathbf{m}_i)$, stetig, $i = 0, 1$ und $j = 1, \dots, n_i$, so gestatten der Zweistichproben-Wilcoxon-Test und das entsprechende verteilungsfreie Konfidenzintervall nach Moses (Hollander und Wolfe, 1973; Lehmann, 1975) ein analoges, nichtparametrisches Entscheidungsverfahren zur Beurteilung der Sicherheit. Sei $R_{1j}(\mathbf{d})$ der Rang von $X_{1j} - \mathbf{d}$ in der kombinierten Stichprobe

$$X_{01}, \dots, X_{0n_0}, X_{11} - \mathbf{d}, \dots, X_{1n_1} - \mathbf{d},$$

und $R_1(\mathbf{d}) = \sum_{j=1}^{n_1} R_{1j}(\mathbf{d})$ die Rangsumme bezüglich der um \mathbf{d} verschobenen Beobachtungen der Behandlungsgruppe. Die Nullhypothese H_0^d wird verworfen, falls

$$R_1(\mathbf{d}) \leq n_1(n_0 + n_1 + 1) - r_{\mathbf{a}, n_0, n_1}^d,$$

wobei $r_{\mathbf{a}, n_0, n_1}^d$ das $(1-\mathbf{a})$ -Quantil der Wilcoxonstatistik $R_1(\mathbf{d})$ bezeichne. Analog zur parametrischen Auswertung wird H_0^d abgelehnt, falls das verteilungsfreie einseitige $100(1-\mathbf{a})\%$ -Konfidenzintervall für die Differenz der Lokationsparameter $\mathbf{m}_1 - \mathbf{m}_0$ vollständig im Sicherheitsbereich $(-\infty, \mathbf{d})$ enthalten ist. Die Dualität zwischen Teststatistik und

Konfidenzintervall läßt sich mittels des Theorems 4 von Lehmann (1975, Seite 87) zeigen. Wesentlich ist jedoch dabei, daß **keine Bindungen** vorliegen; diese Prämisse wird im folgenden für alle verteilungsfreien Verfahren vorausgesetzt.

Die Konstruktion des Konfidenzintervalles basiert auf den geordneten Werten $D_1 \leq \dots \leq D_{n_0 n_1}$ der insgesamt $n_0 n_1$ Differenzen $X_{1j} - X_{0j^*}$, $j = 1, \dots, n_1$ und $j^* = 1, \dots, n_0$.

Damit ergibt sich als einseitiges $100(1-\alpha)\%$ -Konfidenzintervall:

$$(-\infty, U_1^d], \text{ wobei } U_1^d = D_{u_1}^d = D_{n_0 n_1 + 1 - C_a^1}^d, \quad C_a^1 = \frac{n_1(2n_0 + n_1 + 1)}{2} + 1 - r_{\alpha, n_0, n_1}^d.$$

Zur Vereinfachung sind in Tabelle 6 die zur Konstruktion von einseitigen 95%-Konfidenzintervallen benötigten Werte u_1 für eine Gesamtstichprobe vom Umfang $n_0 + n_1 = 8, \dots, 30$ angegeben. Für größere n_0 und n_1 läßt sich C_a^1 mittels

$$C_a^1 = \frac{n_0 n_1}{2} - z_\alpha \sqrt{\frac{n_0 n_1 (n_0 + n_1 + 1)}{12}}$$

approximieren (Hollander und Wolfe, 1973); dabei bezeichnet z_α das $(1-\alpha)$ -Quantil der Standardnormalverteilung.

Tabelle 6: Zweistichprobensituation: Indizes u_1 zur Konstruktion des einseitigen nichtparametrischen 95%-Konfidenzintervalles für $m_1 - m_0$

$n_0 =$	$n_1 = 4$	5	6	7	8	9	10	11	12	13	14	15
4	15	18	21	24	27	30	33	36	39	42	45	48
5	18	21	25	29	32	36	39	43	47	50	54	57
6	21	25	29	34	38	42	46	50	55	59	63	67
7	24	29	34	38	43	48	53	58	63	67	72	77
8	27	32	38	43	49	54	60	65	70	76	81	87
9	30	36	42	48	54	60	66	72	78	84	90	96
10	33	39	46	53	60	66	73	79	86	93	99	106
11	36	43	50	58	65	72	79	87	94	101	108	115
12	39	47	55	63	70	78	86	94	102	109	117	125
13	42	50	59	67	76	84	93	101	109	118	126	134
14	45	54	63	72	81	90	99	108	117	126	135	144
15	48	57	67	77	87	96	106	115	125	134	144	153

3.1.2 Zweiseitiges Testproblem

Die einseitige Formulierung des Testproblems ist in der überwiegenden Anzahl von Sicherheitsstudien angezeigt. Es existieren jedoch auch Fragestellungen (Reversibilität von substanzbedingten Effekten), bei denen eher ein zweiseitiges Testproblem relevant ist. Das entsprechende Testproblem lautet

$$H_0^d: \mathbf{m}_1 - \mathbf{m}_0 \leq \mathbf{d}_1 \quad \text{oder} \quad \mathbf{m}_1 - \mathbf{m}_0 \geq \mathbf{d}_2$$

versus

$$H_1^d: \mathbf{d}_1 < \mathbf{m}_1 - \mathbf{m}_0 < \mathbf{d}_2,$$

wobei das Intervall $(\mathbf{d}_1, \mathbf{d}_2)$, $\mathbf{d}_1 < 0 < \mathbf{d}_2$, den Sicherheitsbereich oder, anders ausgedrückt, den Bereich des nicht relevanten Unterschiedes definiert. Eine Zerlegung in zwei einseitige Testprobleme ergibt:

$$H_{01}^d: \mathbf{m}_1 - \mathbf{m}_0 \leq \mathbf{d}_1 \quad \text{versus} \quad H_{11}^d: \mathbf{m}_1 - \mathbf{m}_0 > \mathbf{d}_1$$

und

$$H_{02}^d: \mathbf{m}_1 - \mathbf{m}_0 \geq \mathbf{d}_2 \quad \text{versus} \quad H_{12}^d: \mathbf{m}_1 - \mathbf{m}_0 < \mathbf{d}_2.$$

Die Nullhypothese H_0^d ist die Vereinigung von H_{01}^d und H_{02}^d , und die Alternative H_1^d läßt sich als Durchschnitt von H_{11}^d und H_{12}^d darstellen:

$$H_0^d = \bigcup_{r=1}^2 H_{0r}^d \quad \text{versus} \quad H_1^d = \bigcap_{r=1}^2 H_{1r}^d.$$

Entsprechend des *intersection-union*-Prinzips (Berger, 1982) wird H_0^d zum Niveau \mathbf{a} genau dann abgelehnt, falls **beide** Hypothesen H_{01}^d und H_{02}^d jeweils zum Niveau \mathbf{a} verworfen werden können. Bei der parametrischen Annahme $F_i(x) = F\left(\frac{x - \mathbf{m}_i}{\mathbf{s}}\right)$, $i = 0, 1$, mit unbekannter, aber homogener Varianz \mathbf{s}^2 führt die Anwendung des Zweistichproben-*t*-Tests zur Ablehnung von H_{01}^d bzw. von H_{02}^d , falls

$$t_{11}^d = \frac{\bar{X}_1 - \bar{X}_0 - \mathbf{d}_1}{S \sqrt{\frac{1}{n_0} + \frac{1}{n_1}}} \geq t_{\mathbf{a}, n_0 + n_1 - 2} \quad \text{und} \quad t_{12}^d = \frac{\bar{X}_1 - \bar{X}_0 - \mathbf{d}_2}{S \sqrt{\frac{1}{n_0} + \frac{1}{n_1}}} \leq -t_{\mathbf{a}, n_0 + n_1 - 2}.$$

Die Ablehnung von H_0^d mittels der zwei einseitigen Zweistichproben- t -Tests (Schuirmann, 1987) ist äquivalent zur Inklusion des zweiseitigen $100(1-2\alpha)\%$ -Konfidenzintervalles $[\mathbf{d}_-, \mathbf{d}_+]$ für die Differenz $\mathbf{m}_1 - \mathbf{m}_0$ im Sicherheitsbereich $(\mathbf{c}_1, \mathbf{c}_2)$:

$$[\mathbf{d}_-, \mathbf{d}_+] = \left[\bar{X}_1 - \bar{X}_0 \pm t_{\alpha, n_0+n_1-2} S \sqrt{\frac{1}{n_0} + \frac{1}{n_1}} \right] \subset (\mathbf{d}_-, \mathbf{d}_+).$$

Das analoge verteilungsfreie Verfahren wurde von Hauschke *et al.* (1990) publiziert, wobei anstelle des Zweistichproben- t -Tests der Zweistichproben-Wilcoxon-Test verwendet wird.

$$H_{01}^d \text{ wird abgelehnt, falls } R_1(\mathbf{d}_1) \geq r_{\alpha, n_0, n_1}^d$$

$$\text{und } H_{02}^d \text{ wird abgelehnt, falls } R_1(\mathbf{d}_2) \leq n_1(n_0 + n_1 + 1) - r_{\alpha, n_0, n_1}^d.$$

Die Verwerfung beider Einzelhypothesen ist äquivalent zur Inklusion des verteilungsfreien zweiseitigen $100(1-2\alpha)\%$ -Konfidenzintervalles $[L_1^d, U_1^d]$ für $\mathbf{m}_1 - \mathbf{m}_0$ im Sicherheitsbereich $(\mathbf{c}_1, \mathbf{c}_2)$ (Hauschke und Steinijs, 1991; Chow und Liu, 1992), wobei $L_1^d = D_{l_1}^d = D_{c_a^1}^d$ und $U_1^d = D_{u_1}^d = D_{n_0 n_1 + 1 - c_a^1}^d$. In Tabelle 7 sind die zur Konstruktion von zweiseitigen 90%-Konfidenzintervallen benötigten Werte der Indizes l_1 und u_1 für eine Gesamtstichprobe vom Umfang $n_0 + n_1 = 8, \dots, 30$ angegeben.

Tabelle 7: Zweistichprobensituation: Indizes l_1/u_1 zur Konstruktion des zweiseitigen nichtparametrischen 90%-Konfidenzintervalles für $\mathbf{m}_1 - \mathbf{m}_0$

$n_0 =$	$n_1 = 4$	5	6	7	8	9	10	11	12	13	14	15
4	2/15	3/18	4/21	5/24	6/27	7/30	8/33	9/36	10/39	11/42	12/45	13/48
5	3/18	5/21	6/25	7/29	9/32	10/36	12/39	13/43	14/47	16/50	17/54	19/57
6	4/21	6/25	8/29	9/34	11/38	13/42	15/46	17/50	18/55	20/59	22/63	24/67
7	5/24	7/29	9/34	12/38	14/43	16/48	18/53	20/58	22/63	25/67	27/72	29/77
8	6/27	9/32	11/38	14/43	16/49	19/54	21/60	24/65	27/70	29/76	32/81	34/87
9	7/30	10/36	13/42	16/48	19/54	22/60	25/66	28/72	31/78	34/84	37/90	40/96
10	8/33	12/39	15/46	18/53	21/60	25/66	28/73	32/79	35/86	38/93	42/99	45/106
11	9/36	13/43	17/50	20/58	24/65	28/72	32/79	35/87	39/94	43/101	47/108	51/115
12	10/39	14/47	18/55	22/63	27/70	31/78	35/86	39/94	43/102	48/109	52/117	56/125
13	11/42	16/50	20/59	25/67	29/76	34/84	38/93	43/101	48/109	52/118	57/126	62/134
14	12/45	17/54	22/63	27/72	32/81	37/90	42/99	47/108	52/117	57/126	62/135	67/144
15	13/48	19/57	24/67	29/77	34/87	40/96	45/106	51/115	56/125	62/134	67/144	73/153

Zur Veranschaulichung werden für den Rattenversuch hinsichtlich der Reversibilität (Tabelle 3) die 90%-Konfidenzintervalle für $m_1 - m_0$ entsprechend des parametrischen bzw. nichtparametrischen Ansatzes in Tabelle 8 dargestellt. Bei einem Sicherheitsbereich (c_1, c_2) von $\pm 40g$ führen beide Methoden zur Verwerfung der Nullhypothese und somit zur Annahme der Reversibilität der durch die Substanz bedingten Körpergewichtsveränderung.

Tabelle 8: Zweiseitige 90%-Konfidenzintervalle für $m_1 - m_0$ für den Beispieldatensatz aus Tabelle 3

	90%-Konfidenzintervall
Parametrisches Verfahren	[-31.3, 21.5]
Nichtparametrisches Verfahren	[-36.0, 19.8]

3.1.3 Gütebestimmung und Fallzahlplanung

Zum Zwecke der Güteberechnung oder der Fallzahlermittlung wird in dieser Arbeit generell ein balanciertes Design angenommen, d.h. $n_0 = n_1 = n$. Die Güte des einseitigen Zweistichproben- t -Tests t_1^d ist die Wahrscheinlichkeit für die korrekte Annahme von H_1^d und läßt sich über die nichtzentrale t -Verteilung ermitteln:

$$P\left[t_1^d \leq -t_{\alpha, 2n-2} \mid m_1 - m_0 < d \mid s\right] = \frac{\sqrt{2p}}{G(n-1)2^{n-2}} \int_0^{\infty} F\left(\frac{-t_{\alpha, 2n-2} x}{\sqrt{2n-2}} - D\right) x^{2n-3} F'(x) dx,$$

wobei $F'(x) = \frac{1}{\sqrt{2p}} e^{-\frac{x^2}{2}}$ die Dichte der Standardnormalverteilung bezeichnet und der Nichtzentralitätsparameter sich wie folgt darstellt:

$$D = \frac{m_1 - m_0 - d}{s \sqrt{\frac{2}{n}}}.$$

Die nichtzentrale t -Verteilung ist in den gängigen Statistikprogrammen als Funktion enthalten, so daß die entsprechende Berechnung der Güte ohne Probleme erfolgen kann. In verschiedenen Publikationen, zum Beispiel Phillips (1990) und Liu und Chow (1992), wird zur Vereinfachung der Darstellung des Nichtzentralitätsparameters der Schwellenwert d der Wert $m_1 - m_0$ aus der Alternative und die zugrundeliegende Standardabweichung s relativ zum Erwartungswert der Kontrollgruppe definiert:

$$\mathbf{d} = \mathbf{J}^* \mathbf{m}_0, \quad \mathbf{m}_1 - \mathbf{m}_0 = \mathbf{J} \mathbf{m}_0, \quad \mathbf{s} = CV_0 \mathbf{m}_0$$

und somit gilt für den Nichtzentralitätsparameter $D = \frac{J - J^*}{CV_0 \sqrt{\frac{2}{n}}}$.

In Abbildung 4 wird die Gütefunktion für verschiedene Variationskoeffizienten CV_0 dargestellt; dabei werden ein Sicherheitsbereich von $(-\infty, \mathbf{d}) = (-\infty, 0.2\mathbf{m}_0)$ und 10 Versuchseinheiten pro Gruppe gewählt.

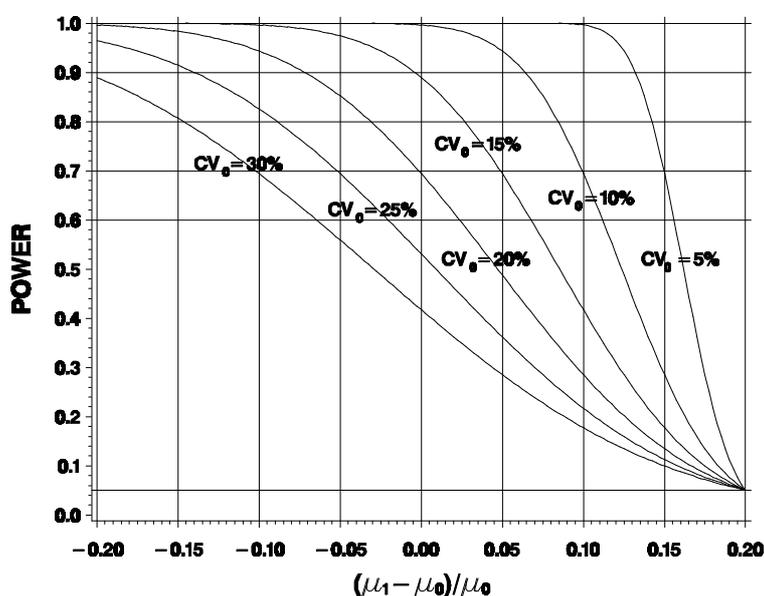


Abbildung 4: Zweistichprobensituation: Die Gütefunktion des einseitigen Tests t_1^d bei einem Sicherheitsbereich von $(-\infty, \mathbf{d}) = (-\infty, 0.2\mathbf{m}_0)$ für $\alpha = 0.05$, $n_0 = n_1 = 10$ und verschiedene CV_0

Den benötigten Stichprobenumfang n (pro Gruppe), um beim Testen von H_0^d versus H_1^d einen Fehler 1. Art α und einen Fehler 2. Art β an einer festgewählten Stelle $\mathbf{J}, \mathbf{J} < \mathbf{J}^*$, nicht zu überschreiten, kann iterativ nach der folgenden bekannten Ungleichung (Mace, 1964) bestimmt werden:

$$n \geq 2 \left(\frac{CV_0}{\mathbf{J}^* - \mathbf{J}} \right)^2 (t_{\alpha, 2n-2} + t_{\beta, 2n-2})^2.$$

In Tabelle 9 sind die benötigten Fallzahlen angegeben, um eine Güte von mindestens 0.80 zu erreichen. Dabei werden die weiteren Einflußgrößen wie folgt vorgegeben: $\alpha = 0.05$, $\mathbf{J}^* = 0.20$, $\mathbf{J} = 0.00, 0.05, 0.10, 0.15$ und $CV_0 = 5.0, 7.5, 10.0, \dots, 30.0\%$. Die exakten Werte

stehen in der ersten Zeile, und die approximativen werden, falls sie von den exakten abweichen, in der zweiten Zeile angegeben; der Bindestrich bedeutet somit Übereinstimmung der Werte.

Tabelle 9: Zweistichprobensituation: Exakter und approximativer Stichprobenumfang pro Gruppe für eine Güte von mindestens 0.80 bei $\alpha = 0.05$, $(-\infty, \mathbf{d}) = (-\infty, 0.2\mathbf{m})$ und verschiedene CV_0

CV_0 (%)	J			
	0.00	0.05	0.10	0.15
5.0	2	3	4	14
	3	-	5	-
7.5	3	4	8	29
	-	5	-	-
10.0	4	7	14	51
	5	-	-	-
12.5	6	10	21	78
	-	-	-	79
15.0	8	14	29	112
	-	-	-	113
17.5	11	18	39	153
	-	-	-	-
20.0	14	23	51	199
	-	-	-	-
22.5	17	29	64	252
	-	-	-	-
25.0	21	36	78	310
	-	-	79	-
27.5	25	43	95	375
	-	-	-	-
30.0	29	51	112	446
	-	-	113	-

Die statistische Methodik zur Bestimmung der Güte für das zweiseitige Testproblem wurde von Owen (1965) hergeleitet. Dabei wird die folgende Eigenschaft benötigt:

Lemma 1:

Die Korrelation zwischen den Teststatistiken t_{11}^d und t_{12}^d entspricht dem Wert 1.

Beweis:

Es gilt

$$\text{Var}(\bar{X}_1 - \bar{X}_0 - \mathbf{d}) = \text{Var}(\bar{X}_1 - \bar{X}_0 - \mathbf{d}_2),$$

$$\text{Kov}(\bar{X}_1 - \bar{X}_0 - \mathbf{d}, \bar{X}_1 - \bar{X}_0 - \mathbf{d}_2) = \text{Var}(\bar{X}_1 - \bar{X}_0 - \mathbf{d})$$

und daher

$$\text{Korr}(t_{11}^d, t_{12}^d) = \mathbf{r}(t_{11}^d, t_{12}^d) = \mathbf{r}^d = 1.$$

▷

Unter Ausnutzung dieser Eigenschaft kann nach Owen (1965) die zugrundeliegende nichtzentrale bivariate t -Verteilung durch die entsprechende Differenz zweier Integrale bestimmt werden, und somit erhält man die Wahrscheinlichkeit für die korrekte Verwerfung von H_0^d :

$$\begin{aligned} P[t_{11}^d \geq t_{a,2n-2} \wedge t_{12}^d \leq -t_{a,2n-2} | \mathbf{d}_1 < \mathbf{m}_1 - \mathbf{m}_0 < \mathbf{d}_2, \mathbf{s}] \\ = Q(-t_{a,2n-2}, \mathbf{D}_2, 0, \mathbf{y}) - Q(t_{a,2n-2}, \mathbf{D}_1, 0, \mathbf{y}), \end{aligned}$$

wobei

$$Q(t, \mathbf{D}, u, \mathbf{y}) = \frac{\sqrt{2\mathbf{p}}}{\mathbf{G}(n-1)2^{n-2}} \int_u^y \mathbf{F} \left(\frac{tx}{\sqrt{2n-2}} - \mathbf{D} \right) x^{2n-3} \mathbf{F}'(x) dx,$$

$$\mathbf{y} = \frac{\mathbf{D}_1 - \mathbf{D}_2}{A_1 - A_2}, \quad A_1 = \frac{t_{a,2n-2}}{\sqrt{2n-2}}, \quad A_2 = \frac{-t_{a,2n-2}}{\sqrt{2n-2}}$$

und

$$\mathbf{D}_1 = \frac{\mathbf{m}_1 - \mathbf{m}_0 - \mathbf{c}_1}{\mathbf{s} \sqrt{\frac{2}{n}}}, \quad \mathbf{D}_2 = \frac{\mathbf{m}_1 - \mathbf{m}_0 - \mathbf{c}_2}{\mathbf{s} \sqrt{\frac{2}{n}}}.$$

Definiert man wieder entsprechend

$$\mathbf{c}_1 = \mathbf{J}_1 \mathbf{m}_0, \quad \mathbf{c}_2 = \mathbf{J}_2 \mathbf{m}_0, \quad \mathbf{m}_1 - \mathbf{m}_0 = \mathbf{J} \mathbf{m}_0, \quad \mathbf{s} = \text{CV}_0 \mathbf{m}_0,$$

so folgt für die Nichtzentralitätsparameter:

$$\mathbf{D}_1 = \frac{\mathbf{J} - \mathbf{J}_1}{\text{CV}_0 \sqrt{\frac{2}{n}}} \quad \text{und} \quad \mathbf{D}_2 = \frac{\mathbf{J} - \mathbf{J}_2}{\text{CV}_0 \sqrt{\frac{2}{n}}}.$$

Die von Owen angegebenen Formeln zur Bestimmung von $Q(t, \mathbf{D}, u, \mathbf{y})$ wurden für die vorliegende Arbeit mittels SAS/IML (siehe Anhang) realisiert und dienen als Basis zur Berechnung der entsprechenden Gütefunktionen. In der Abbildung 5 werden die Gütefunktionen für einen symmetrischen Sicherheitsbereich von $(-0.2\mathbf{m}_0, 0.2\mathbf{m}_0)$, für einen Variationskoeffizienten von 15.0% und für die Stichprobenumfänge von jeweils 10 bzw. 20 pro Gruppe dargestellt. Die Gütefunktionen sind bei dieser Wahl des Äquivalenzbereiches symmetrisch zu 0 und erreichen auch dort ihr Maximum. Man sieht, daß für den Stichprobenumfang von 10 pro Gruppe eine maximale Güte von etwa 0.80 erreicht wird.

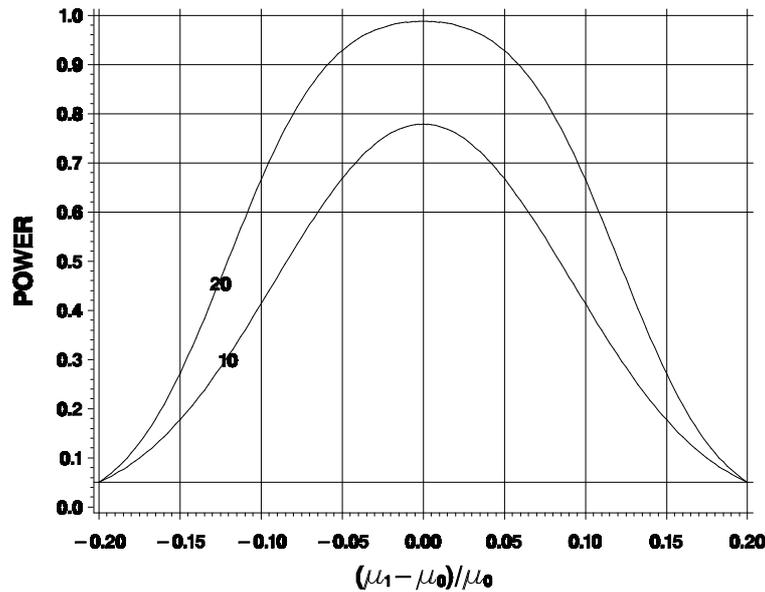


Abbildung 5: Zweistichprobensituation: Die Gütefunktion des Entscheidungsverfahrens, basierend auf den zwei einseitigen Tests t_{11}^d, t_{12}^d bei einem Sicherheitsbereich von $(\underline{d}, \bar{d}) = (-0.2m, 0.2m)$ für $\alpha = 0.05, n_0 = n_1 = 10, 20$ und $CV_0 = 15.0\%$

Aufgrund der komplexen Berechnung der Gütefunktionen haben Hauschke und Hothorn (1998) zur Berechnung der Fallzahlen die folgenden einfachen Approximationsformeln angegeben.

Satz 1:

Die Approximationsformeln für die Fallzahlen (pro Gruppe) für eine Güte von mindestens $1 - b$ bei vorgegebenem a, CV_0 und einem zu 0 symmetrisch gewählten Sicherheitsbereich, d.h. $J_1 = -J_2$, lauten:

$$\text{i) falls } J = 0 \quad n \geq 2 \left(\frac{CV_0}{J_2} \right)^2 \left(t_{a, 2n-2} + t_{\frac{b}{2}, 2n-2} \right)^2,$$

$$\text{ii) } 0 < J < J_2 \quad n \geq 2 \left(\frac{CV_0}{J_2 - J} \right)^2 \left(t_{a, 2n-2} + t_{b, 2n-2} \right)^2$$

und

$$\text{iii) } -J_2 < J < 0 \quad n \geq 2 \left(\frac{CV_0}{-J_2 - J} \right)^2 \left(t_{a, 2n-2} + t_{b, 2n-2} \right)^2.$$

Beweis:

i) Aus $\mathbf{J} = 0$ folgt $\mathbf{m}_1 = \mathbf{m}_0$, und die Güte lautet

$$P \left[\frac{\bar{X}_1 - \bar{X}_0 - \mathbf{d}_1}{S \sqrt{\frac{2}{n}}} \geq t_{\mathbf{a}, 2n-2} \wedge \frac{\bar{X}_1 - \bar{X}_0 - \mathbf{d}_2}{S \sqrt{\frac{2}{n}}} \leq -t_{\mathbf{a}, 2n-2} \right] =$$

$$P \left[\mathbf{d}_1 + t_{\mathbf{a}, 2n-2} S \sqrt{\frac{2}{n}} \leq \bar{X}_1 - \bar{X}_0 \leq \mathbf{d}_2 - t_{\mathbf{a}, 2n-2} S \sqrt{\frac{2}{n}} \right] =$$

$$P \left[\frac{\mathbf{d}_1}{S \sqrt{\frac{2}{n}}} + t_{\mathbf{a}, 2n-2} \leq \frac{\bar{X}_1 - \bar{X}_0}{S \sqrt{\frac{2}{n}}} \leq \frac{\mathbf{d}_2}{S \sqrt{\frac{2}{n}}} - t_{\mathbf{a}, 2n-2} \right].$$

Für $\mathbf{m}_1 = \mathbf{m}_0$ ist die Statistik $\frac{\bar{X}_1 - \bar{X}_0}{S \sqrt{\frac{2}{n}}}$ zentral t -verteilt. Zudem gilt mit $\mathbf{d}_1 = -\mathbf{d}_2$ die folgende

Symmetrieeigenschaft

$$\frac{\mathbf{d}_2}{S \sqrt{\frac{2}{n}}} - t_{\mathbf{a}, 2n-2} = - \left(\frac{-\mathbf{d}_2}{S \sqrt{\frac{2}{n}}} + t_{\mathbf{a}, 2n-2} \right).$$

Verlangt man nun, daß die Power mindestens $1 - \mathbf{b}$ beträgt, und behandelt man die empirische Fehlervarianz S^2 wie eine deterministische Größe und ersetzt sie durch \mathbf{s}^2 (Chow und Liu, 1992), so folgt

$$\left| \frac{\mathbf{d}_2}{\mathbf{s} \sqrt{\frac{2}{n}}} - t_{\mathbf{a}, 2n-2} \right| \geq t_{\frac{\mathbf{b}}{2}, 2n-2}$$

und dies ist äquivalent zu

$$n \geq 2 \left(\frac{\mathbf{s}}{\mathbf{d}_2} \right)^2 \left(t_{a,2n-2} + t_{\frac{b}{2},2n-2} \right)^2.$$

Mit $\hat{\mathbf{c}}_2 = \mathbf{J}_2 \mathbf{m}_0$ und $\mathbf{s} = \mathbf{C}V_0 \mathbf{m}_0$ folgt

$$n \geq 2 \left(\frac{\mathbf{C}V_0}{\mathbf{J}_2} \right)^2 \left(t_{a,2n-2} + t_{\frac{b}{2},2n-2} \right)^2.$$

ii) Für $0 < \mathbf{m}_1 - \mathbf{m}_0 < \mathbf{d}_2$ lautet die Güte

$$P \left[\frac{\mathbf{d}_1 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} + t_{a,2n-2} \leq \frac{\bar{X}_1 - \bar{X}_0 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} \leq \frac{\mathbf{d}_2 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} - t_{a,2n-2} \right] =$$

$$P \left[\frac{-\mathbf{d}_2 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} + t_{a,2n-2} \leq \frac{\bar{X}_1 - \bar{X}_0 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} \leq \frac{\mathbf{d}_2 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} - t_{a,2n-2} \right],$$

da $\mathbf{d}_1 = -\mathbf{d}_2$. Die Symmetrieeigenschaft liegt nicht mehr vor:

$$- \left(\frac{\mathbf{d}_2 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} - t_{a,2n-2} \right) = \frac{-\mathbf{d}_2 + (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} + t_{a,2n-2} > \frac{-\mathbf{d}_2 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} + t_{a,2n-2}.$$

Verlangt man nun, daß die Power mindestens $1 - \mathbf{b}$ beträgt und ersetzt man S^2 durch \mathbf{s}^2 , so können bei Verwendung der Ungleichung

$$\frac{\hat{\mathbf{c}}_2 - (\mathbf{m}_1 - \mathbf{m}_0)}{\mathbf{s} \sqrt{\frac{2}{n}}} - t_{a,2n-2} \geq t_{\frac{b}{2},2n-2}$$

sehr große Stichprobengrößen resultieren. Daher wird die folgende Beziehung verwendet:

$$P \left[\frac{\mathbf{d}_1 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} + t_{\mathbf{a}, 2n-2} \leq \frac{\bar{X}_1 - \bar{X}_0 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} \leq \frac{\mathbf{d}_2 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} - t_{\mathbf{a}, 2n-2} \right]$$

$$\leq P \left[\frac{\bar{X}_1 - \bar{X}_0 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} \leq \frac{\mathbf{d}_2 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} - t_{\mathbf{a}, 2n-2} \right].$$

Dies führt zu

$$\frac{\mathbf{d}_2 - (\mathbf{m}_1 - \mathbf{m}_0)}{S \sqrt{\frac{2}{n}}} - t_{\mathbf{a}, 2n-2} \geq t_{\mathbf{b}, 2n-2}$$

oder äquivalent dazu

$$n \geq 2 \left(\frac{\mathbf{s}}{\mathbf{d}_2 - (\mathbf{m}_1 - \mathbf{m}_0)} \right)^2 (t_{\mathbf{a}, 2n-2} + t_{\mathbf{b}, 2n-2})^2.$$

Mit $\mathbf{d}_2 = \mathbf{J}_2 \mathbf{m}_0$, $\mathbf{m}_1 - \mathbf{m}_0 = \mathbf{J} \mathbf{m}_0$ und $\mathbf{s} = CV_0 \mathbf{m}_0$ folgt

$$n \geq 2 \left(\frac{CV_0}{\mathbf{J}_2 - \mathbf{J}} \right)^2 \left(t_{\mathbf{a}, 2n-2} + t_{\frac{\mathbf{b}}{2}, 2n-2} \right)^2.$$

iii) Erfolgt analog zu ii)

▷

In der Tabelle 10 sind die exakten und approximativen Fallzahlen für $\mathbf{a} = 0.05$, $\mathbf{b} = 0.20$, $\mathbf{J}_1 = -\mathbf{J}_2 = -0.20$, $CV_0 = 5.0, 10.0, \dots, 30.0\%$ und für $|\mathbf{J}| = 0.00, 0.02, 0.05, 0.10, 0.15$ angegeben. Es sei darauf hingewiesen, daß die Approximationsformeln für die Werte $|\mathbf{J}| < 0.05$ und $\mathbf{J} \neq 0$ zu unlogischen Resultaten führen können: Die ermittelten Werte sind an der Stelle 0.02 geringer als an der für die Fallzahlbestimmung optimalen Stelle 0. Dieses

Phänomen beruht auf dem diskreten Schritt von $t_{b,2n-2}$ für $J \neq 0$ nach $t_{\frac{b}{2},2n-2}$ für $J = 0$.

Daher sollte in diesen Situationen unbedingt die exakte Methode verwendet werden.

Tabelle 10: Zweistichprobensituation: Exakter und approximativer Stichprobenumfang pro Gruppe für eine Güte von mindestens 0.80 bei $\alpha = 0.05$, von $(\underline{c}, \underline{d}) = (-0.2\mathbf{m}_0, 0.2\mathbf{m}_0)$ und verschiedene CV_0

CV_0 (%)	$ J $					
	0.00	0.02	0.05	0.10	0.15	
5.0	3	3	3	4	14	
	-	-	-	5	-	
7.5	4	4	5	8	29	
	-	-	5	-	-	
10.0	6	6	7	14	51	
	-	5	-	-	-	
12.5	8	8	10	21	78	
	-	7	-	-	79	
15.0	11	11	14	29	112	
	-	10	-	-	113	
17.5	14	15	18	39	153	
	-	13	-	-	-	
20.0	18	19	24	51	199	
	-	17	23	-	-	
22.5	23	24	29	64	252	
	-	21	-	-	-	
25.0	28	29	36	78	310	
	-	25	-	79	-	
27.5	34	35	43	95	375	
	-	30	-	-	-	
30.0	40	41	51	112	446	
	-	36	-	113	-	

3.2 Sicherheitsnachweis für den Quotienten von Lokationsparametern

Wie in den einleitenden Bemerkungen zu Kapitel 3 bereits erwähnt, ist die Festlegung des Sicherheitsbereiches in absoluten Einheiten in der Praxis oftmals sehr schwierig. Aus diesem Grunde werden im folgenden Methoden dargestellt, die es ermöglichen, diesen Wert relativ zum Erwartungswert der Kontrollgruppe zu definieren. Daraus resultiert die Notwendigkeit, entsprechende Testprobleme durch geeignete Verfahren bezüglich des Quotienten $\frac{\mathbf{m}_1}{\mathbf{m}_0}$,

$\mathbf{m}_0 \neq 0$, adäquat zu lösen.

3.2.1 Einseitiges Testproblem

Das einseitige Testproblem mit einem relativ zum Erwartungswert der Kontrolle definierten Schwellenwert, d.h. $\underline{c} = f\mathbf{m}_0$, $0 < f$, stellt sich wie folgt dar:

$$H_0^d: \mathbf{m}_1 - \mathbf{m}_0 \geq \mathbf{d} = f\mathbf{m}_0 \Leftrightarrow H_0^q: \frac{\mathbf{m}_1}{\mathbf{m}_0} \geq 1 + f = \mathbf{q}$$

versus

$$H_1^d: \mathbf{m}_1 - \mathbf{m}_0 < \mathbf{d} = f\mathbf{m}_0 \Leftrightarrow H_1^q: \frac{\mathbf{m}_1}{\mathbf{m}_0} < 1 + f = \mathbf{q},$$

wobei $(-\infty, \mathbf{q})$, $\mathbf{q} > 1$, den auf den Quotienten der Erwartungswerte $\frac{\mathbf{m}_1}{\mathbf{m}_0}$ bezogenen Bereich des nicht relevanten Unterschiedes definiert.

Sasabuchi (1988) zeigte, daß unter der parametrischen Annahme $F_i(x) = \mathbf{F}\left(\frac{x - \mathbf{m}_i}{\mathbf{s}}\right)$,

$i = 0, 1$, das folgende Testverfahren ein Likelihood-Ratio-Test zum Niveau \mathbf{a} ist: H_0^q wird genau dann ablehnt, falls

$$t_1^q = \frac{\bar{X}_1 - \mathbf{q}\bar{X}_0}{S \sqrt{\frac{\mathbf{q}^2}{n_0} + \frac{1}{n_1}}} \leq -t_{\mathbf{a}, n_0 + n_1 - 2}.$$

Hauschke *et al.* (1999b) zeigten, daß die Bedingung $t_1^q \leq -t_{\mathbf{a}, n_0 + n_1 - 2}$ äquivalent ist zu

$$(-\infty, \mathbf{q}_+^1] \subset (-\infty, \mathbf{q}) \quad \text{und} \quad \bar{X}_0^2 > a_0,$$

mit

$$\mathbf{q}_+^1 = \frac{\bar{X}_0 \bar{X}_1 + \sqrt{a_0 \bar{X}_1^2 + a_1 \bar{X}_0^2 - a_0 a_1}}{\bar{X}_0^2 - a_0}, \quad a_0 = \frac{S^2}{n_0} t_{\mathbf{a}, n_0 + n_1 - 2}^2 \quad \text{und} \quad a_1 = \frac{S^2}{n_1} t_{\mathbf{a}, n_0 + n_1 - 2}^2.$$

Der Wert \mathbf{q}_+^1 ist die obere Grenze des einseitigen $100(1-\mathbf{a})\%$ -Konfidenzintervalles nach Fieller (1954) für $\frac{\mathbf{m}_1}{\mathbf{m}_0}$, daher gestattet die Verwendung dieses Konfidenzintervalles ebenfalls eine *a-posteriori* zu treffende Sicherheitsbeurteilung. Die Bedingung $\bar{X}_0^2 > a_0$ bedeutet, daß der Lokationsparameter der Kontrolle \mathbf{m}_0 statistisch signifikant ungleich Null ist und stellt eine ausreichende Bedingung dafür dar, daß sich die Fiellersche Konfidenzmenge als ein reguläres Konfidenzintervall darstellt.

Für das Beispiel der chronischen Toxizitätsprüfung an weiblichen Wistar-Ratten (Tabelle 2) ergibt die Berechnung der oberen Grenze des einseitigen 95%-Konfidenzintervalles für

$\frac{m_1}{m_0}$ den Wert $q_+^1 = 1.29$. Postuliert man eine relevante Erhöhung von $d = 0.25m_0$, so folgt für

die obere Schranke von $\frac{m_1}{m_0}$ der Wert $q = 1.25$. Die obere Grenze des Konfidenzintervalles

liegt klar oberhalb von q und daher kann eine sicherheitsrelevante Erhöhung der ASAT-Werte nicht ausgeschlossen werden.

3.2.2 Zweiseitiges Testproblem

Das zur Formulierung

$$H_0^d: m_1 - m_0 \leq d_1 \quad \text{oder} \quad m_1 - m_0 \geq d_2$$

versus

$$H_1^d: d_1 < m_1 - m_0 < d_2$$

äquivalente Testproblem für den Quotienten der Erwartungswerte lautet

$$H_0^q: \frac{m_1}{m_0} \leq q_1 \quad \text{oder} \quad \frac{m_1}{m_0} \geq q_2$$

versus

$$H_1^q: q_1 < \frac{m_1}{m_0} < q_2,$$

mit dem Sicherheitsbereich (q_1, q_2) , $q_1 = 1 + f_1$, $q_2 = 1 + f_2$ und $-1 < f_1 < 0 < f_2$.

Das Testen der zweiseitigen Fragestellung ist äquivalent zur simultanen Überprüfung der beiden einseitigen Testprobleme:

$$H_{01}^q: \frac{m_1}{m_0} \leq q_1 \quad \text{versus} \quad H_{11}^q: \frac{m_1}{m_0} > q_1$$

und

$$H_{02}^q: \frac{m_1}{m_0} \geq q_2 \quad \text{versus} \quad H_{12}^q: \frac{m_1}{m_0} < q_2.$$

Entsprechend des *intersection-union*-Prinzips wird $H_0^q = \bigcup_{r=1}^2 H_{0r}^q$ zum Niveau α zugunsten

von $H_1^q = \bigcap_{r=1}^2 H_{1r}^q$ abgelehnt, falls **beide** Einzelhypothesen H_{01}^q und H_{02}^q jeweils zum Niveau

\mathbf{a} verworfen werden können. Unter der Voraussetzung $F_i(x) = \mathbf{F}\left(\frac{x - \mathbf{m}_i}{\mathbf{s}}\right)$, $i = 0, 1$, führt die Anwendung der Sasabuchi-Teststatistiken zur Ablehnung von H_{01}^q bzw. von H_{02}^q , falls

$$t_{11}^q = \frac{\bar{X}_1 - \mathbf{q}_1 \bar{X}_0}{S \sqrt{\frac{\mathbf{q}_1^2}{n_0} + \frac{1}{n_1}}} \geq t_{\mathbf{a}, n_0+n_1-2} \quad \text{und} \quad t_{12}^q = \frac{\bar{X}_1 - \mathbf{q}_2 \bar{X}_0}{S \sqrt{\frac{\mathbf{q}_2^2}{n_0} + \frac{1}{n_1}}} \leq -t_{\mathbf{a}, n_0+n_1-2}.$$

Die Bedingungen $t_{11}^q \geq t_{\mathbf{a}, n_0+n_1-2}$ und $t_{12}^q \leq -t_{\mathbf{a}, n_0+n_1-2}$ sind äquivalent zu

$$[\mathbf{q}_-^1, \mathbf{q}_+^1] \subset (\mathbf{q}_1, \mathbf{q}_2) \quad \text{und} \quad \bar{X}_0^2 > a_0,$$

wobei

$$\mathbf{q}_\pm^1 = \frac{\bar{X}_0 \bar{X}_1 \pm \sqrt{a_0 \bar{X}_1^2 + a_1 \bar{X}_0^2 - a_0 a_1}}{\bar{X}_0^2 - a_0}.$$

Der Schluß auf Sicherheit zum Niveau \mathbf{a} ist somit möglich, falls das Fiellersche $100(1-2\mathbf{a})\%$ -Konfidenzintervall $[\mathbf{q}_-^1, \mathbf{q}_+^1]$ regulär und vollständig im Sicherheitsbereich enthalten $(\mathbf{q}_1, \mathbf{q}_2)$ ist (Hauschke *et al.*, 1999c).

Berechnet man für den Reversibilitätsversuch (Tabelle 3) nach der obigen Methode das zweiseitige 90%-Konfidenzintervall für $\frac{\mathbf{m}_1}{\mathbf{m}_0}$, so erhält man [0.94, 1.04]. Bei einem Sicherheitsbereich von $(\mathbf{d}_1, \mathbf{d}_2) = (-0.1\mathbf{m}_0, +0.1\mathbf{m}_0)$ für die Differenz $\mathbf{m}_1 - \mathbf{m}_0$ folgt ein Bereich von $(\mathbf{q}_1, \mathbf{q}_2) = (0.9, 1.1)$ für $\frac{\mathbf{m}_1}{\mathbf{m}_0}$. Das Konfidenzintervall ist im Sicherheitsbereich enthalten, und somit ist die Annahme der Reversibilität der durch die Substanz bedingten Körpergewichtsveränderung bei diesen Vorgaben gerechtfertigt.

3.2.3 Gütebestimmung und Fallzahlplanung

Die Güte des einseitigen Sasabuchi-Tests t_1^q ist die Wahrscheinlichkeit für die korrekte Annahme von H_1^q , und läßt sich über die nichtzentrale t -Verteilung ermitteln:

$$P\left[t_1^q \leq -t_{\mathbf{a}, 2n-2} \left| \frac{\mathbf{m}_1}{\mathbf{m}_0} < \mathbf{q}, \mathbf{s} \right. \right] = \frac{\sqrt{2\mathbf{p}}}{\mathbf{G}(n-1)2^{n-2}} \int_0^\infty \mathbf{F}\left(\frac{-t_{\mathbf{a}, 2n-2} x}{\sqrt{2n-2}} - \mathbf{q}\right) x^{2n-3} \mathbf{F}'(x) dx,$$

wobei der Nichtzentralitätsparameter lautet:

$$q = \frac{m_1 - q m_0}{s \sqrt{\frac{1+q^2}{n}}} = \frac{\frac{m_1}{m_0} - q}{CV_0 \sqrt{\frac{1+q^2}{n}}}$$

In Abbildung 6 wird die Gütefunktion für verschiedene Variationskoeffizienten dargestellt; dabei werden ein Sicherheitsbereich von $(-\infty, q) = (-\infty, 1.2)$ und 10 Versuchseinheiten pro Gruppe gewählt.

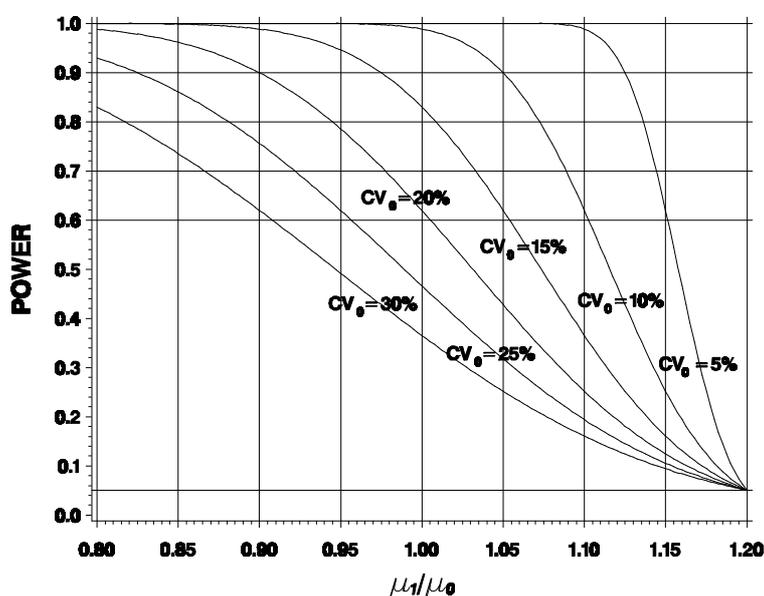


Abbildung 6: Zweistichprobensituation: Die Gütefunktion des einseitigen Tests t_1^q bei einem Sicherheitsbereich von $(-\infty, q) = (-\infty, 1.2)$ für $\alpha = 0.05$, $n_0 = n_1 = 10$ und verschiedene CV_0

Den notwendigen Stichprobenumfang n (pro Gruppe), um beim Testen von H_0^q versus H_1^q die Fehler \mathbf{a} und \mathbf{b} an der Stelle $\frac{m_1}{m_0}, \frac{m_1}{m_0} < q$, nicht zu überschreiten, kann man iterativ nach der folgenden Ungleichung berechnen (Hauschke *et al.*, 1999b):

Satz 2:

Die Approximationsformel für die Fallzahlen (pro Gruppe) für eine Güte von mindestens $1 - \mathbf{b}$ bei vorgegebenem \mathbf{a} , CV_0 und einem Sicherheitsbereich $(-\infty, q)$ lautet:

$$n \geq (1 + \mathbf{q}^2) \left(\frac{CV_0}{\mathbf{q} - \frac{\mathbf{m}_1}{\mathbf{m}_0}} \right)^2 (t_{\mathbf{a}, 2n-2} + t_{\mathbf{b}, 2n-2})^2.$$

Beweis:

Die Wahrscheinlichkeit H_0^q zu verwerfen ist

$$P \left[t_1^q \leq -t_{\mathbf{a}, 2n-2} \left| \frac{\mathbf{m}_1}{\mathbf{m}_0} < \mathbf{q}, \mathbf{s} \right. \right] = P \left[\frac{\bar{X}_1 - \mathbf{q}\bar{X}_0}{S\sqrt{\frac{1+\mathbf{q}^2}{n}}} \leq -t_{\mathbf{a}, 2n-2} \left| \frac{\mathbf{m}_1}{\mathbf{m}_0} < \mathbf{q}, \mathbf{s} \right. \right] =$$

$$P \left[\frac{\bar{X}_1 - \mathbf{q}\bar{X}_0 - (\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0)}{S\sqrt{\frac{1+\mathbf{q}^2}{n}}} \leq -t_{\mathbf{a}, 2n-2} - \frac{\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0}{S\sqrt{\frac{1+\mathbf{q}^2}{n}}} \left| \frac{\mathbf{m}_1}{\mathbf{m}_0} < \mathbf{q}, \mathbf{s} \right. \right].$$

Verlangt man nun, daß die Power mindestens $1 - \mathbf{b}$ beträgt und ersetzt man S^2 durch \mathbf{s}^2 , so folgt analog zu Satz 1 die Behauptung:

$$n \geq (1 + \mathbf{q}^2) \left(\frac{CV_0}{\mathbf{q} - \frac{\mathbf{m}_1}{\mathbf{m}_0}} \right)^2 (t_{\mathbf{a}, 2n-2} + t_{\mathbf{b}, 2n-2})^2$$

▷

In Tabelle 11 sind die Fallzahlen angegeben, um eine Güte von mindestens 0.80 zu erreichen. Dabei werden die weiteren Einflußgrößen wie folgt vorgegeben: $\mathbf{a} = 0.05$, $\mathbf{q} = 1.20$, $\frac{\mathbf{m}_1}{\mathbf{m}_0} = 1.00, 1.05, 1.10, 1.15$ und $CV_0 = 5.0, 7.5, 10.0, \dots, 30.0\%$.

Tabelle 11: Zweistichprobensituation: Exakter und approximativer Stichprobenumfang pro Gruppe für eine Güte von mindestens 0.80 bei $\alpha = 0.05$, $(-\infty, \mathbf{q}) = (-\infty, 1.2)$ und verschiedene CV_0

CV_0 (%)	m_1/m_0			
	1.00	1.05	1.10	1.15
5.0	3	3	5	16
	-	-	-	-
7.5	4	5	10	35
	-	-	-	-
10.0	5	8	16	62
	-	-	-	-
12.5	7	12	25	95
	-	-	-	96
15.0	10	16	35	137
	-	-	-	-
17.5	13	22	47	186
	-	-	-	-
20.0	16	28	62	243
	-	-	-	-
22.5	20	35	78	307
	-	-	-	-
25.0	25	43	95	378
	-	-	96	-
27.5	30	52	115	458
	-	-	-	-
30.0	35	62	137	544
	-	-	-	-

Die exakte Gütebestimmung für das zweiseitige Testproblem erfolgt direkt über die bivariate nichtzentrale t -Verteilung, wobei das folgende Ergebnis benötigt wird:

Lemma 2:

Die Korrelation zwischen den beiden Teststatistiken t_{11}^q und t_{12}^q lautet

$$\text{Korr}(t_{11}^q, t_{12}^q) = \mathbf{r}(t_{11}^q, t_{12}^q) = \mathbf{r}^q = \frac{1 + \mathbf{q}_1 \mathbf{q}_2}{\sqrt{(1 + \mathbf{q}_1^2)(1 + \mathbf{q}_2^2)}}.$$

Beweis:

Mit $\text{Var}(\bar{X}_1) = \text{Var}(\bar{X}_0) = \frac{\mathbf{s}^2}{n}$ folgt

$$\text{Var}(\bar{X}_1 - \mathbf{q}_r \bar{X}_0) = \frac{\mathbf{s}^2(1 + \mathbf{q}_r^2)}{n}, \quad r = 1, 2,$$

$$\text{Kov}(\bar{X}_1 - \mathbf{q}_1 \bar{X}_0, \bar{X}_1 - \mathbf{q}_2 \bar{X}_0) = \text{Var}(\bar{X}_1) + \mathbf{q}_1 \mathbf{q}_2 \text{Var}(\bar{X}_0) = \frac{\mathbf{s}^2(1 + \mathbf{q}_1 \mathbf{q}_2)}{n}$$

und damit

$$\mathbf{r}^q = \frac{\frac{\mathbf{s}^2(1+\mathbf{q}_1\mathbf{q}_2)}{n}}{\sqrt{\frac{\mathbf{s}^2(1+\mathbf{q}_1^2)}{n} \frac{\mathbf{s}^2(1+\mathbf{q}_2^2)}{n}}} = \frac{1+\mathbf{q}_1\mathbf{q}_2}{\sqrt{(1+\mathbf{q}_1^2)(1+\mathbf{q}_2^2)}}.$$

▷

Die Wahrscheinlichkeit der korrekten Annahme von H_1^q ist

$$P\left[t_{11}^q \geq t_{a,2n-2} \wedge t_{12}^q \leq -t_{a,2n-2} \mid \mathbf{q}_1 < \frac{\mathbf{m}_1}{\mathbf{m}_0} < \mathbf{q}_2, \mathbf{s}\right] =$$

$$Q(\infty, -t_{a,2n-2}, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{r}^q) - Q(t_{a,2n-2}, -t_{a,2n-2}, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{r}^q),$$

wobei

$$Q(t_1, t_2, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{r}) = P\left[t_{11}^q \leq t_1 \wedge t_{12}^q \leq t_2 \mid \mathbf{q}_1 < \frac{\mathbf{m}_1}{\mathbf{m}_0} < \mathbf{q}_2, \mathbf{s}\right] =$$

$$\frac{\sqrt{2\mathbf{p}}}{\mathbf{G}(n-1)2^{n-2}} \int_0^\infty F_2\left(\frac{t_1 x}{\sqrt{2n-2}} - \mathbf{Q}_1, \frac{t_2 x}{\sqrt{2n-2}} - \mathbf{Q}_2, \mathbf{r}\right) x^{2n-3} F'(x) dx,$$

$$F_2(x, y, \mathbf{r}) = \frac{1}{2\mathbf{p}\sqrt{1-\mathbf{r}^2}} \int_{-\infty}^x \int_{-\infty}^y \exp\left(-\frac{u^2 - 2\mathbf{r}uv + v^2}{2(1-\mathbf{r}^2)}\right) dv du,$$

und für die entsprechenden Nichtzentralitätsparameter gilt

$$\mathbf{Q}_1 = \frac{\mathbf{m}_1 - \mathbf{q}_1 \mathbf{m}_0}{\mathbf{s} \sqrt{\frac{1+\mathbf{q}_1^2}{n}}} = \frac{\frac{\mathbf{m}_1}{\mathbf{m}_0} - \mathbf{q}_1}{CV_0 \sqrt{\frac{1+\mathbf{q}_1^2}{n}}} \quad \text{und} \quad \mathbf{Q}_2 = \frac{\mathbf{m}_1 - \mathbf{q}_2 \mathbf{m}_0}{\mathbf{s} \sqrt{\frac{1+\mathbf{q}_2^2}{n}}} = \frac{\frac{\mathbf{m}_1}{\mathbf{m}_0} - \mathbf{q}_2}{CV_0 \sqrt{\frac{1+\mathbf{q}_2^2}{n}}}.$$

Die Berechnung der obigen Integrale erfolgt mittels entsprechender SAS-Algorithmen (Genz und Bretz, 1999, siehe Anhang). In der Abbildung 7 werden die Gütefunktionen für einen Sicherheitsbereich von $(\mathbf{q}_1, \mathbf{q}_2) = (0.8, 1.25)$, einen Variationskoeffizienten von 15.0% und für Stichprobenumfänge von jeweils 10 und 20 pro Gruppe dargestellt. Die

Gütefunktionen sind nicht symmetrisch zu 1 und erreichen ihr Maximum ebenfalls für kleine Fallzahlen und/oder große Variationskoeffizienten nicht an der Stelle 1.

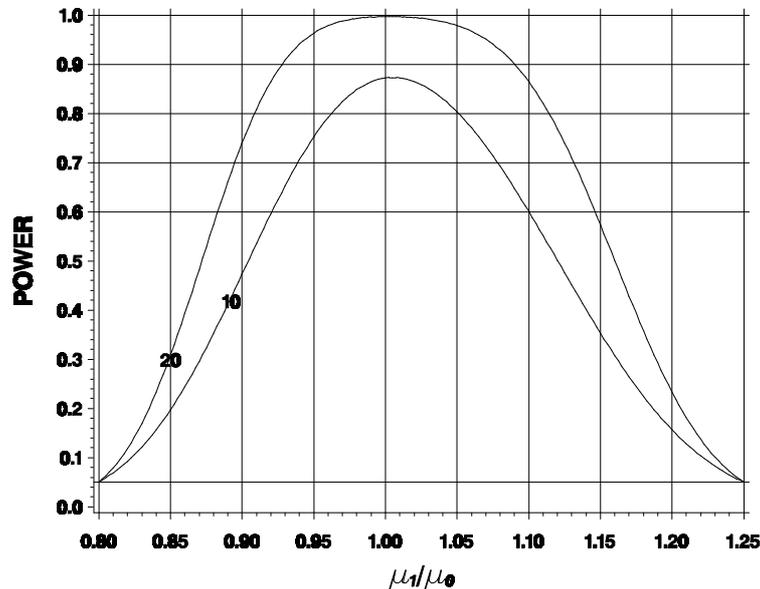
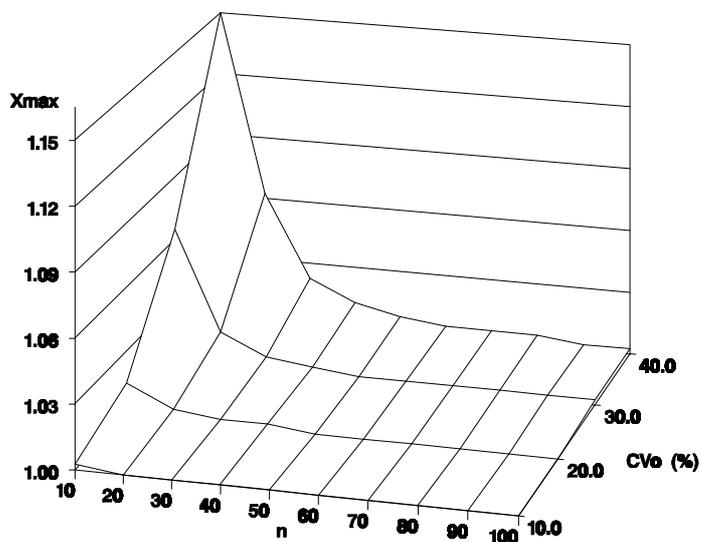
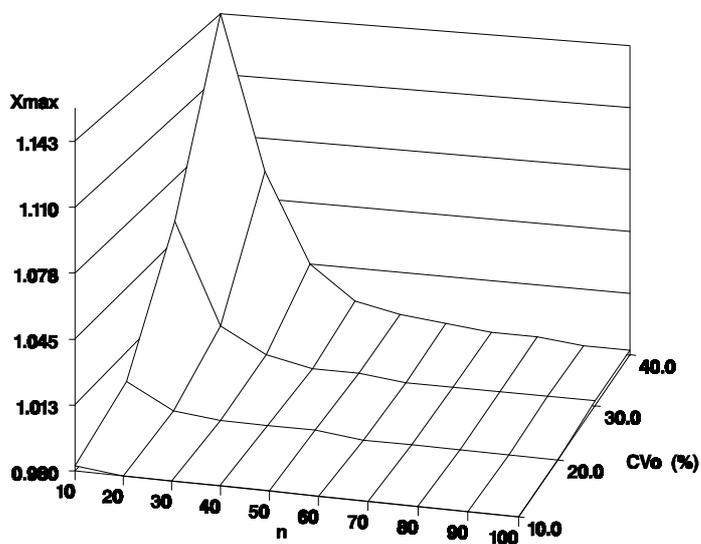


Abbildung 7: Zweistichprobensituation: Die Gütefunktion des Entscheidungsverfahrens, basierend auf den zwei einseitigen Tests t_{11}^q, t_{12}^q bei einem Sicherheitsbereich von $(q_1, q_2) = (0.8, 1.25)$ für $\alpha = 0.05, n_0 = n_1 = 10, 20$ und $CV_0 = 15.0\%$

In den Abbildungen 8 und 9 ist die Stelle der maximalen Güte (X_{max}) in Abhängigkeit von der Fallzahl, vom Variationskoeffizienten und für verschiedene Sicherheitsbereiche dargestellt. Man erkennt, daß für hinreichend große Fallzahlen das Maximum an der Stelle 1 genau dann angenommen wird, wenn ein Sicherheitsbereich der Gestalt $(q_1, q_2) = \left(q_1, \frac{1}{q_1} \right)$ gewählt wird (Kieser *et al.*, 1997).

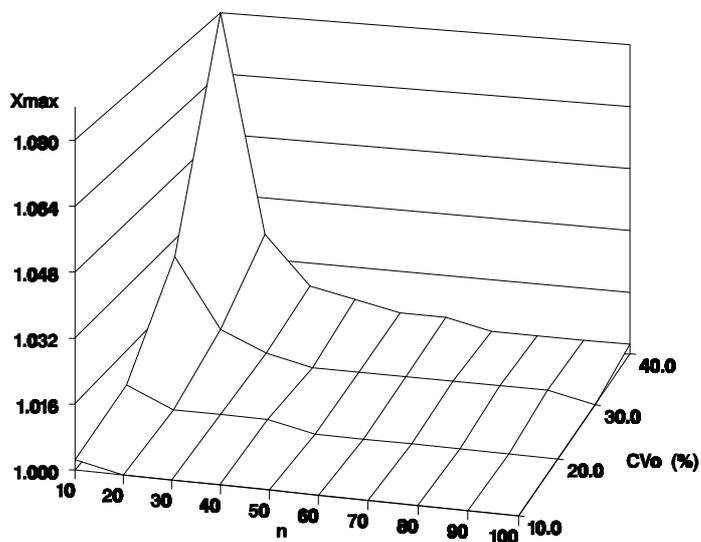


(a)

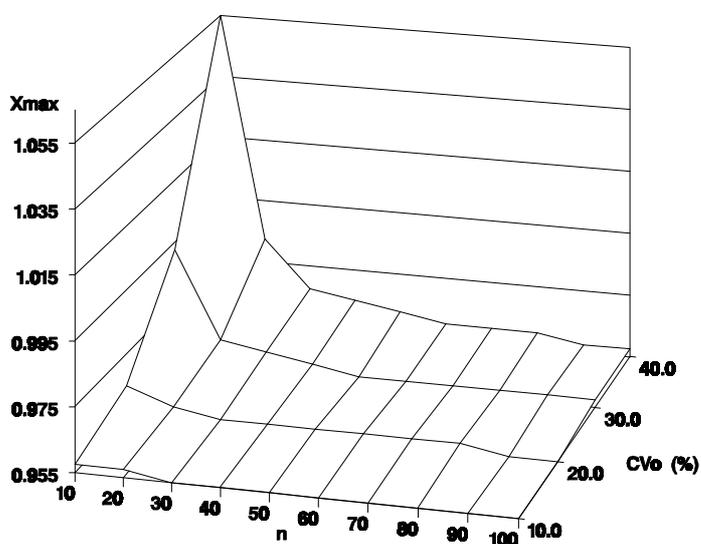


(b)

Abbildung 8: Zweistichprobensituation: Stelle der maximalen Güte X_{max} in Abhängigkeit von der Fallzahl und vom Variationskoeffizienten, basierend auf den zwei einseitigen Tests t_{11}^q, t_{12}^q für $\alpha = 0.05$ bei einem Sicherheitsbereich von (a) $(q_1, q_2) = (0.8, 1.25)$, (b) $(q_1, q_2) = (0.8, 1.20)$



(a)



(b)

Abbildung 9: Zweistichprobensituation: Stelle der maximalen Güte X_{max} in Abhängigkeit von der Fallzahl und vom Variationskoeffizienten, basierend auf den zwei einseitigen Tests t_{11}^q, t_{12}^q für $\alpha = 0.05$ bei einem Sicherheitsbereich von (a) $(q_1, q_2) = (0.7, 1.43)$, (b) $(q_1, q_2) = (0.7, 1.30)$

Zur vereinfachten Berechnung der Fallzahlen haben Kieser und Hauschke (1999a) die folgenden Approximationsformeln hergeleitet.

Satz 3:

Die Approximationsformeln für die Fallzahlen (pro Gruppe) für eine Güte von mindestens $1 - \mathbf{b}$ bei vorgegebenem \mathbf{a} , CV_0 für einen Sicherheitsbereich der Gestalt $(\mathbf{q}_1, \mathbf{q}_2) = \left(\frac{1}{\mathbf{q}_2}, \mathbf{q}_2\right)$

lauten:

$$\text{i) falls } \frac{\mathbf{m}_1}{\mathbf{m}_0} = 1 \quad n \geq \left(1 + \mathbf{q}_2^2\right) \left(\frac{CV_0}{\mathbf{q}_2 - 1}\right)^2 \left(t_{\mathbf{a}, 2n-2} + t_{\frac{\mathbf{b}}{2}, 2n-2}\right)^2,$$

$$\text{ii) } 1 < \frac{\mathbf{m}_1}{\mathbf{m}_0} < \mathbf{q}_2 \quad n \geq \left(1 + \mathbf{q}_2^2\right) \left(\frac{CV_0}{\mathbf{q}_2 - \frac{\mathbf{m}_1}{\mathbf{m}_0}}\right)^2 \left(t_{\mathbf{a}, 2n-2} + t_{\mathbf{b}, 2n-2}\right)^2$$

und

$$\text{iii) } \frac{1}{\mathbf{q}_2} < \frac{\mathbf{m}_1}{\mathbf{m}_0} < 1 \quad n \geq \left(1 + \frac{1}{\mathbf{q}_2^2}\right) \left(\frac{CV_0}{\frac{1}{\mathbf{q}_2} - \frac{\mathbf{m}_1}{\mathbf{m}_0}}\right)^2 \left(t_{\mathbf{a}, 2n-2} + t_{\mathbf{b}, 2n-2}\right)^2.$$

Beweis:

Die Herleitung erfolgt analog zu Satz 1.

▷

In der Tabelle 12 sind die exakten (Hauschke *et al.*, 1999c) und die approximativen (Kieser und Hauschke, 1999a) Fallzahlen für $\mathbf{a} = 0.05$, $\mathbf{b} = 0.20$, $(\mathbf{q}_1, \mathbf{q}_2) = (0.8, 1.25)$, $CV_0 = 5.0$, 10.0, ..., 30.0% und für verschiedene Quotienten $\frac{\mathbf{m}_1}{\mathbf{m}_0}$ angegeben.

Tabelle 12: Zweistichproben-Situation: Exakter und approximativer Stichprobenumfang pro Gruppe für eine Güte von mindestens 0.80 bei $\alpha = 0.05$, $(\mathbf{q}_1, \mathbf{q}_2) = (0.8, 1.25)$ und verschiedene CV_0

CV_0 (%)	$\mathbf{m}_1, \mathbf{m}_0$							
	0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20
5.0	11	4	3	3	3	3	5	17
	-	-	-	-	-	-	-	-
7.5	24	7	4	3	4	5	10	37
	-	-	-	4	-	-	-	-
10.0	42	11	6	5	5	8	17	65
	-	-	-	-	-	-	-	-
12.5	65	17	8	7	8	12	26	100
	-	-	-	-	-	-	-	-
15.0	92	24	12	9	10	17	37	144
	93	-	11	-	-	-	-	-
17.5	125	32	15	12	14	23	50	195
	-	-	-	-	13	-	-	-
20.0	163	42	20	15	18	29	65	255
	-	-	19	-	17	-	-	-
22.5	207	53	25	19	22	37	81	322
	-	-	24	-	21	-	-	-
25.0	255	65	30	23	27	45	100	397
	-	-	29	-	26	-	-	-
27.5	308	78	36	28	32	54	121	480
	-	-	35	-	31	55	-	481
30.0	366	92	43	33	38	65	144	572
	-	93	42	-	37	-	-	-

In diesem dritten Kapitel wurden die Entscheidungsverfahren für den direkten Sicherheitsnachweis dargestellt. Desweiteren erfolgte eine exakte Gütebestimmung, und neben der korrespondierenden exakten Fallzahlbestimmung wurden zusätzlich Approximationsformeln zur Verfügung gestellt. Vergleicht man in den jeweiligen Tabellen die approximativen mit den exakt ermittelten Werten, so läßt sich eine weitgehende Übereinstimmung feststellen.

Es sei darauf hingewiesen, daß für den Quotienten im Parallelgruppenvergleich bislang nur parametrische Methoden existieren. Eine finite Adaption auf nichtparametrische Verfahren ist nicht trivial, da schon unter der Normalverteilungsannahme $F_i(x) = F\left(\frac{x - \mathbf{m}_i}{\mathbf{s}}\right)$, $i = 0, 1$, wegen $Var(\bar{X}_1) \neq Var(\bar{X}_0)$ die Voraussetzung des Lokationsmodells verletzt ist.

4 Crossover-Situation

Ein ursprünglich aus der Landwirtschaft stammender Versuchsplan (Jones und Kenward, 1989) ist das *Crossover*-Design mit zwei Perioden und zwei Behandlungen. Jede Versuchseinheit erhält beide Behandlungen, wobei in der ersten Sequenz die Behandlungen in der Reihenfolge (0, 1) und in der zweiten Sequenz in der umgekehrten Reihenfolge (1, 0) verabreicht werden. Die Zuordnung der Versuchseinheiten zu den beiden Sequenzen erfolgt zufällig. Ein wesentlicher Vorteil im Vergleich zum Parallelgruppenvergleich besteht darin, daß die intraindividuelle Variabilität, d.h. Variabilität innerhalb der Versuchseinheit, meist wesentlich geringer ist als die interindividuelle Variabilität und somit geringere Stichprobenumfänge als im Parallelgruppenvergleich resultieren. Jedoch ist es auch offensichtlich, daß dieses Design nur dann zur Anwendung kommen kann, falls eine vollständige Reversibilität der zu untersuchenden Effekte vorliegt und daher nicht die Applikation der jeweiligen Behandlung in der zweiten Periode in Frage stellt. Zwischen den beiden Perioden ist somit eine hinreichend lange *Washout*-Phase zu wählen, so daß mögliche *Carryover*-Effekte auszuschließen sind. Die Zufallsvariable X_{ijk} bezeichnet die Beobachtung der Zielvariablen bei der j -ten Versuchseinheit, $j = 1, \dots, n_i$, in der i -ten Sequenz, $i = 0, 1$ aus der k -ten Periode, $k = 1, 2$.

Das allgemeine bivariate Modell, welches im folgenden als Modell I bezeichnet wird, lautet:

$$X_{ijk} = \mathbf{m}_h + \mathbf{p}_k + \mathbf{e}_{ijk},$$

wobei \mathbf{m}_h den Effekt der Behandlung h , $h = 0$ für $i + 1 = k$ und $h = 1$ für $i + 1 \neq k$ bezeichnet und \mathbf{p}_k die Perioden-Effekte darstellen mit $\mathbf{p}_1 + \mathbf{p}_2 = 0$. Weiterhin wird angenommen, daß für die bivariaten Zufallsvektoren gilt: $(\mathbf{e}_{ij1}, \mathbf{e}_{ij2})'$ u.i.v. $\sim F_i(x_1, x_2)$, stetig, $i = 0, 1$. In der Tabelle 13 ist das entsprechende Schema der Beobachtungen angegeben.

Tabelle 13: Schema der Zufallsvariablen im Zwei-Perioden-*Crossover*-Design mit zwei Behandlungen

Behandlungssequenz	Periode 1	Periode 2
(0, 1)	$X_{0j1} = \mathbf{m}_0 + \mathbf{p}_1 + \mathbf{e}_{0j1}$	$X_{0j2} = \mathbf{m}_1 + \mathbf{p}_2 + \mathbf{e}_{0j2}$
(1,0)	$X_{1j1} = \mathbf{m}_1 + \mathbf{p}_1 + \mathbf{e}_{1j1}$	$X_{1j2} = \mathbf{m}_0 + \mathbf{p}_2 + \mathbf{e}_{1j2}$

Das klassische univariate Modell (Grizzle, 1965):

$$X_{ijk} = \mathbf{m}_h + \mathbf{p}_k + b_{ij} + e_{ijk},$$

bezeichnet im folgenden als Modell II, ist mit $\mathbf{e}_{ijk} = b_{ij} + e_{ijk}$ als Spezialfall im allgemeinen Modell I enthalten. Dabei bezeichnet b_{ij} den zufälligen Effekt der j -ten Versuchseinheit mit b_{ij} u.i.v. $\sim F_b(x) = F\left(\frac{x}{\mathbf{s}_b}\right)$ und e_{ijk} den zufälligen Fehlerterm mit e_{ijk} u.i.v. $\sim F_e(x) = F\left(\frac{x}{\mathbf{s}_e}\right)$; die Zufallsvariablen b_{ij} und e_{ijk} werden als unabhängig vorausgesetzt. Neben der Annahme identischer Varianzen setzt Modell II auch stets positive Kovarianzen für die Beobachtungen der ersten und zweiten Periode voraus:

$$\text{Kov}(X_{ijk}, X_{ijl}) = \begin{cases} \mathbf{s}_b^2 & k \neq l \\ \mathbf{s}_b^2 + \mathbf{s}_e^2 & k = l \end{cases} \quad k, l = 1, 2.$$

Da diese Voraussetzungen sicherlich sehr restriktiv sein können (Lehmacher 1987, 1991; Hauschke *et al.*, 1996), werden statistische Entscheidungsverfahren sowohl für Modell I als auch für Modell II hergeleitet.

4.1 Sicherheitsnachweis für die Differenz von Lokationsparametern

Seien Y_{ij}^d die j -ten intraindividuellen Periodendifferenzen der Beobachtungen in der 1. und 2. Periode innerhalb der i -ten Sequenz: $Y_{ij}^d = X_{ij1} - X_{ij2}$, und ferner seien $H_i(y)$ die stetigen Verteilungsfunktionen der Differenzen der zugehörigen Fehlerterme:

$$H_i(y) = P(\mathbf{e}_{ij1} - \mathbf{e}_{ij2} \leq y), \quad i = 0, 1, \quad j = 1, \dots, n_i.$$

Verlangt man $H_0(y) = H_1(y) = H(y)$, so folgt unmittelbar:

$$Y_{0j}^d \text{ u.i.v. } \sim H(y + \mathbf{m}_1 - \mathbf{m}_0 + \mathbf{p}_2 - \mathbf{p}_1), \quad j = 1, \dots, n_0,$$

$$Y_{1j}^d \text{ u.i.v. } \sim H(y + \mathbf{m}_0 - \mathbf{m}_1 + \mathbf{p}_2 - \mathbf{p}_1), \quad j = 1, \dots, n_1,$$

d.h. die Verteilungen der Differenzen der beiden Sequenzen unterscheiden sich nur durch den Lokationsunterschied von $2(\mathbf{m}_1 - \mathbf{m}_0)$. Daher können Hypothesen über $\mathbf{m}_1 - \mathbf{m}_0$ durch Verfahren für zwei unverbundene Stichproben untersucht werden.

4.1.1 Einseitiges Testproblem

Für die parametrische Analyse im Modell I wird eine bivariate Normalverteilung mit Erwartungswertvektor $(0, 0)'$ und Varianz-Kovarianzmatrizen \mathbf{S}_i angenommen, d.h. $(\mathbf{e}_{ij1}, \mathbf{e}_{ij2})'$ u.i.v. $\sim F_i(x_1, x_2) = N_2(0, \mathbf{S}_i)$, wobei

$$\mathbf{S}_0 = \begin{pmatrix} \mathbf{s}_0^2 & \mathbf{s}_{01} \\ \mathbf{s}_{01} & \mathbf{s}_1^2 \end{pmatrix} \quad \text{und} \quad \mathbf{S}_1 = \begin{pmatrix} \mathbf{s}_1^2 & \mathbf{s}_{01} \\ \mathbf{s}_{01} & \mathbf{s}_0^2 \end{pmatrix}.$$

Zur Konstruktion eines Testverfahren benötigt man das folgende Lemma.

Lemma 3:

Der Schätzer $\frac{\bar{X}_1 - \bar{X}_0}{2}$ ist normalverteilt mit Erwartungswert $\mathbf{m}_1 - \mathbf{m}_0$ und Varianz

$$\mathbf{s}_d^2 \frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right), \quad \text{wobei}$$

$$\mathbf{s}_d^2 = \mathbf{s}_0^2 + \mathbf{s}_1^2 - 2\mathbf{s}_{01}$$

und

$$\bar{X}_0 = \left(\frac{1}{n_0} \sum_{j=1}^{n_0} X_{0j1} + \frac{1}{n_1} \sum_{j=1}^{n_1} X_{1j2} \right) \quad \text{und} \quad \bar{X}_1 = \left(\frac{1}{n_0} \sum_{j=1}^{n_0} X_{0j2} + \frac{1}{n_1} \sum_{j=1}^{n_1} X_{1j1} \right).$$

Beweis:

Die Verteilungsfunktionen der Periodendifferenzen sind:

$$Y_{ij}^d \text{ u.i.v. } \sim H(y - E(Y_{ij}^d)) = \mathbf{F} \left(\frac{y - E(Y_{ij}^d)}{\text{Var}(Y_{ij}^d)} \right),$$

mit

$$E(Y_{0j}^d) = \mathbf{m}_0 - \mathbf{m}_1 + \mathbf{p}_1 - \mathbf{p}_2, \quad E(Y_{1j}^d) = \mathbf{m}_1 - \mathbf{m}_0 + \mathbf{p}_1 - \mathbf{p}_2,$$

$$\text{Var}(Y_{ij}^d) = \mathbf{s}_d^2 = \mathbf{s}_0^2 + \mathbf{s}_1^2 - 2\mathbf{s}_{01}, \quad i = 0, 1 \quad \text{und} \quad j = 1, \dots, n_i.$$

Der Erwartungswert für die Differenz der transformierten Zufallsvariablen aus den beiden Stichproben beträgt $E(Y_{1j}^d - Y_{0j}^d) = 2(\mathbf{m}_1 - \mathbf{m}_0)$; ein erwartungstreuer Schätzer für die Differenz $\mathbf{m}_1 - \mathbf{m}_0$ ist somit:

$$\frac{\bar{Y}_1^d - \bar{Y}_0^d}{2} = \frac{1}{2} \left(\frac{1}{n_1} \sum_{j=1}^{n_1} (X_{1j1} - X_{1j2}) - \frac{1}{n_0} \sum_{j=1}^{n_0} (X_{0j1} - X_{0j2}) \right) = \frac{\bar{X}_1 - \bar{X}_0}{2}.$$

Die Varianz des Schätzers ist

$$\text{Var} \left(\frac{\bar{X}_1 - \bar{X}_0}{2} \right) = \text{Var} \left(\frac{1}{2} \left(\frac{1}{n_1} \sum_{j=1}^{n_1} Y_{1j}^d - \frac{1}{n_0} \sum_{j=1}^{n_0} Y_{0j}^d \right) \right) = \mathbf{s}_d^2 \frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right),$$

da

$$\text{Var}(Y_{ij}^d) = \mathbf{s}_d^2, \quad i = 0, 1 \quad \text{und} \quad j = 1, \dots, n_i.$$

▷

Mittels der Periodendifferenzen wurde das *Crossover*-Design auf die Zweistichprobensituation reduziert

$$Y_{ij}^d \text{ u.i.v. } \sim \mathbf{F} \left(\frac{y - E(Y_{ij}^d)}{\text{Var}(Y_{ij}^d)} \right),$$

$$E(Y_{ij}^d) = (-1)^{i+1} (\mathbf{m}_1 - \mathbf{m}_0) + \mathbf{p}_1 - \mathbf{p}_2, \quad \text{Var}(Y_{ij}^d) = \mathbf{s}_d^2,$$

$$i = 0, 1 \quad \text{und} \quad j = 1, \dots, n_i.$$

Die Varianz \mathbf{s}_d^2 kann somit erwartungstreu durch S_d^2 geschätzt werden, mit

$$S_d^2 = \frac{1}{n_0 + n_1 - 2} \sum_{i=0}^1 \sum_{j=1}^{n_i} (Y_{ij}^d - \bar{Y}_i^d)^2.$$

Das einseitig formulierte Testproblem

$$H_0^d: \mathbf{m}_1 - \mathbf{m}_0 \geq \mathbf{d} \quad \text{versus} \quad H_1^d: \mathbf{m}_1 - \mathbf{m}_0 < \mathbf{d}$$

wird somit abgelehnt, falls

$$t_1^d = \frac{\bar{X}_1 - \bar{X}_0 - \mathbf{d}}{S_d \sqrt{\frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)}} \leq -t_{\mathbf{a}, n_0 + n_1 - 2}.$$

Bei einer *a-posteriori* vorgenommenen Festlegung von \mathbf{d} kann mit einer Irrtumswahrscheinlichkeit von \mathbf{a} auf Unbedenklichkeit geschlossen werden, falls die obere Grenze \mathbf{d}_\ddagger des einseitigen $100(1-\mathbf{a})\%$ -Konfidenzintervalles für die Differenz $\mathbf{m}_1 - \mathbf{m}_0$ unterhalb von \mathbf{d} liegt:

$$(-\infty, \mathbf{d}_\ddagger] = \left[-\infty, \frac{\bar{X}_1 - \bar{X}_0}{2} + t_{\mathbf{a}, n_0 + n_1 - 2} S_d \sqrt{\frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)} \right] \subset (-\infty, \mathbf{d}).$$

Im Modell II ergibt sich als Unterschied zu Modell I nur ein anderer Varianzschätzer, da

$$\text{Var}(Y_{ij}^d) = \mathbf{s}_d^2 = \mathbf{s}_0^2 + \mathbf{s}_1^2 - 2\mathbf{s}_{01} = 2(\mathbf{s}_b^2 + \mathbf{s}_e^2) - 2\mathbf{s}_b^2 = 2\mathbf{s}_e^2,$$

$$i = 0, 1 \quad \text{und} \quad j = 1, \dots, n_i.$$

Die Varianz des Schätzers $\frac{\bar{X}_1 - \bar{X}_0}{2}$ beträgt nun $\mathbf{s}_e^2 \frac{1}{2} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)$, und ein erwartungstreuer

Schätzer (Jones und Kenward, 1989) für \mathbf{s}_e^2 ist

$$S_e^2 = \frac{1}{n_0 + n_1 - 2} \left(\sum_{i=0}^1 \sum_{j=1}^{n_i} \sum_{k=1}^2 X_{ijk}^2 - \sum_{i=0}^1 \sum_{j=1}^{n_i} \frac{X_{ij.}^2}{2} - \sum_{i=0}^1 \frac{X_{i.1}^2 + X_{i.2}^2}{n_i} + \sum_{i=0}^1 \frac{X_{i..}^2}{2n_i} \right),$$

mit

$$X_{ij.}^2 = (X_{ij1} + X_{ij2})^2, \quad X_{i.k}^2 = \left(\sum_{j=1}^{n_i} X_{ijk} \right)^2 \quad \text{und} \quad X_{i..}^2 = \left(\sum_{j=1}^{n_i} \sum_{k=1}^2 X_{ijk} \right)^2.$$

Aufgrund der Reduktion auf die Zweistichprobensituation, gilt zwischen den beiden Varianzschätzern die folgende Beziehung (Hauschke und Steinijans, 1991):

$$S_e^2 = \frac{S_d^2}{2}.$$

H_0^d wird im Modell II abgelehnt, falls

$$t_1^d = \frac{\bar{X}_1 - \bar{X}_0 - \mathbf{d}}{S_e \sqrt{\frac{1}{2} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)}} \leq -t_{\mathbf{a}, n_0 + n_1 - 2}$$

$$\Leftrightarrow$$

$$(-\infty, \mathbf{d}_+] = \left(-\infty, \frac{\bar{X}_1 - \bar{X}_0}{2} + t_{\mathbf{a}, n_0 + n_1 - 2} S_e \sqrt{\frac{1}{2} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)} \right] \subset (-\infty, \mathbf{d}).$$

Verzichtet man auf die Normalverteilungsannahme, so kann wegen des vorliegenden Lokationsmodells

$$Y_{0j}^d \text{ u.i.v. } \sim H(y + \mathbf{m}_1 - \mathbf{m}_0 + \mathbf{p}_2 - \mathbf{p}_1), \quad j = 1, \dots, n_0,$$

$$Y_{1j}^d \text{ u.i.v. } \sim H(y + \mathbf{m}_0 - \mathbf{m}_1 + \mathbf{p}_2 - \mathbf{p}_1), \quad j = 1, \dots, n_1,$$

für beide Modelle der Wilcoxon-Test und das verteilungsfreie Konfidenzintervall nach Moses als Entscheidungsverfahren herangezogen werden. Dabei muß jedoch beachtet werden, daß die Rangsumme der um $2\mathbf{d}$ verschobenen j -ten intraindividuellen Periodendifferenzen verwendet wird. Sei $R_{1j}(2\mathbf{d})$ der Rang von $Y_{1j}^d - 2\mathbf{d}$ in der kombinierten Stichprobe

$$Y_{0j}^d, \dots, Y_{0n_0}^d, Y_{11}^d - 2\mathbf{d}, \dots, Y_{1n_1}^d - 2\mathbf{d},$$

und $R_1(2\mathbf{d}) = \sum_{j=1}^{n_1} R_{1j}(2\mathbf{d})$. H_0^d wird verworfen, falls $R_1(2\mathbf{d}) \leq n_1(n_0 + n_1 + 1) - r_{\mathbf{a}, n_0, n_1}^{2\mathbf{d}}$, wobei $r_{\mathbf{a}, n_0, n_1}^{2\mathbf{d}}$ das $(1-\mathbf{a})$ -Quantil der Wilcoxonstatistik $R_1(2\mathbf{d})$ ist. Die Konstruktion des Konfidenzintervalles basiert wiederum auf den geordneten Werten $D_1^d \leq \dots \leq D_{n_0 n_1}^d$ der insgesamt $n_0 n_1$ Differenzen $Y_{1j}^d - Y_{0j^*}^d$, $j = 1, \dots, n_1$ und $j^* = 1, \dots, n_0$. Damit ergibt sich als einseitiges $100(1-\mathbf{a})\%$ -Konfidenzintervall für $\mathbf{m}_1 - \mathbf{m}_0$:

$$\left(-\infty, \frac{U_1^d}{2} \right], \quad \text{wobei } U_1^d = D_{u_1}^d = D_{n_0 n_1 + 1 - C_a^1}^d, \quad C_a^1 = \frac{n_1(2n_0 + n_1 + 1)}{2} + 1 - r_{\mathbf{a}, n_0, n_1}^{2\mathbf{d}}.$$

4.1.2 Zweiseitiges Testproblem

Die zweiseitige Formulierung des Entscheidungsproblems

$$H_0^d: \mathbf{m}_1 - \mathbf{m}_0 \leq \mathbf{d}_1 \text{ oder } \mathbf{m}_1 - \mathbf{m}_0 \geq \mathbf{d}_2$$

versus

$$H_1^d: \mathbf{d}_1 < \mathbf{m}_1 - \mathbf{m}_0 < \mathbf{d}_2$$

wird wieder in zwei einseitige Testprobleme zerlegt

$$H_{01}^d: \mathbf{m}_1 - \mathbf{m}_0 \leq \mathbf{d}_1 \text{ versus } H_{11}^d: \mathbf{m}_1 - \mathbf{m}_0 > \mathbf{d}_1$$

und

$$H_{02}^d: \mathbf{m}_1 - \mathbf{m}_0 \geq \mathbf{d}_2 \text{ versus } H_{12}^d: \mathbf{m}_1 - \mathbf{m}_0 < \mathbf{d}_2 .$$

Die Nullhypothese $H_0^d = \bigcup_{r=1}^2 H_{0r}^d$ wird im Modell I zum Niveau \mathbf{a} genau dann abgelehnt, falls

$$t_{11}^d = \frac{\bar{X}_1 - \bar{X}_0 - \mathbf{d}_1}{S_d \sqrt{\frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)}} \geq t_{\mathbf{a}, n_0 + n_1 - 2} \quad \text{und} \quad t_{12}^d = \frac{\bar{X}_1 - \bar{X}_0 - \mathbf{d}_2}{S_d \sqrt{\frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)}} \leq -t_{\mathbf{a}, n_0 + n_1 - 2},$$

was wiederum äquivalent ist zur Inklusion des $100(1-2\mathbf{a})\%$ -Konfidenzintervalles $[\mathbf{d}_-, \mathbf{d}_+]$ für $\mathbf{m}_1 - \mathbf{m}_0$ im Sicherheitsbereich:

$$[\mathbf{d}_-, \mathbf{d}_+] = \left[\frac{\bar{X}_1 - \bar{X}_0}{2} \pm t_{\mathbf{a}, n_0 + n_1 - 2} S_d \sqrt{\frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)} \right] \subset (\mathbf{d}_1, \mathbf{d}_2).$$

Analog gilt im Modell II, daß H_0^d zum Niveau \mathbf{a} genau dann abgelehnt wird, falls

$$t_{11}^d = \frac{\bar{X}_1 - \bar{X}_0 - \mathbf{d}_1}{S_e \sqrt{\frac{1}{2} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)}} \geq t_{\mathbf{a}, n_0 + n_1 - 2} \quad \text{und} \quad t_{12}^d = \frac{\bar{X}_1 - \bar{X}_0 - \mathbf{d}_2}{S_e \sqrt{\frac{1}{2} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)}} \leq -t_{\mathbf{a}, n_0 + n_1 - 2}$$

\Leftrightarrow

$$[\mathbf{d}_-, \mathbf{d}_+] = \left[\frac{\bar{X}_1 - \bar{X}_0}{2} \pm t_{\mathbf{a}, n_0+n_1-2} S_e \sqrt{\frac{1}{2} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)} \right] \subset (\mathbf{d}_1, \mathbf{d}_2).$$

Für das entsprechende verteilungsfreie Verfahren in beiden Modellen mit Zweistichproben-Wilcoxon-Tests gilt

$$H_{01}^d \text{ wird abgelehnt, falls } R_1(2\mathbf{d}) \geq r_{\mathbf{a}, n_0, n_1}^{2\mathbf{d}}$$

$$\text{und } H_{02}^d \text{ wird abgelehnt, falls } R_1(2\mathbf{d}_2) \leq n_1(n_0 + n_1 + 1) - r_{\mathbf{a}, n_0, n_1}^{2\mathbf{d}_2}.$$

Die Verwerfung beider Einzelhypothesen ist äquivalent zur Inklusion des verteilungsfreien zweiseitigen $100(1-2\mathbf{a})\%$ -Konfidenzintervalles $\left[\frac{L_1^d}{2}, \frac{U_1^d}{2} \right]$ für $\mathbf{m}_1 - \mathbf{m}_0$ im Sicherheitsbereich $(\mathbf{c}_1, \mathbf{c}_2)$ wobei $L_1^d = D_{l_1}^d = D_{c_a^1}^d$ und $U_1^d = D_{u_1}^d = D_{n_0 n_1 + 1 - c_a^1}^d$.

4.1.3 Gütebestimmung und Fallzahlbestimmung

Unter Annahme des Modells II kann die Güte des einseitigen Zweistichproben- t -Tests t_1^d mittels der nichtzentralen t -Verteilung analog zur Zweistichprobensituation bestimmt werden:

$$P[t_1^d \leq -t_{\mathbf{a}, 2n-2} | \mathbf{m}_1 - \mathbf{m}_0 < \mathbf{d}, \mathbf{s}_e] = \frac{\sqrt{2\mathbf{p}}}{\mathbf{G}(n-1)2^{n-2}} \int_0^\infty \mathbf{F} \left(\frac{-t_{\mathbf{a}, 2n-2} x}{\sqrt{2n-2}} - \mathbf{D} \right) x^{2n-3} \mathbf{F}'(x) dx,$$

mit Nichtzentralitätsparameter

$$\mathbf{D} = \frac{\mathbf{J} - \mathbf{J}^*}{\text{CV}_e \sqrt{\frac{1}{n}}}, \text{ wobei } \mathbf{d} = \mathbf{J}^* \mathbf{m}_0, \mathbf{m}_1 - \mathbf{m}_0 = \mathbf{J} \mathbf{m}_0, \text{CV}_e = \frac{\mathbf{s}_e}{\mathbf{m}_0}.$$

In Abbildung 10 wird die Gütefunktion für verschiedene Variationskoeffizienten bei einem Sicherheitsbereich von $(-\infty, \mathbf{c}) = (-\infty, 0.2\mathbf{m}_0)$ und jeweils 10 Versuchseinheiten pro Sequenz dargestellt.

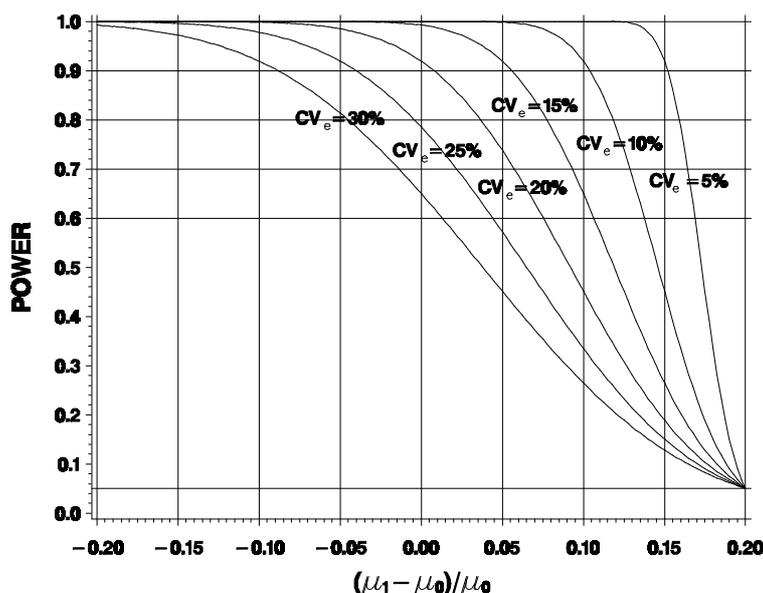


Abbildung 10: Crossover-Situation: Die Gütefunktion des einseitigen Tests t_1^d bei einem Sicherheitsbereich von $(-\infty, \mathbf{d}] = (-\infty, 0.2\mathbf{m}_0)$ für $\mathbf{a} = 0.05$, $n_0 = n_1 = 10$ und verschiedene CV_e

Den benötigten Stichprobenumfang n (pro Sequenz) kann man iterativ nach der folgenden Ungleichung bestimmen:

$$n \geq \left(\frac{CV_e}{J^* - J} \right)^2 \left(t_{\mathbf{a}, 2n-2} + t_{\mathbf{b}, 2n-2} \right)^2,$$

wobei die Herleitung analog zur entsprechenden Ungleichung in der Zweistichprobensituation erfolgt.

In Tabelle 14 sind die benötigten Fallzahlen angegeben, um eine Güte von mindestens 0.80 zu erreichen. Dabei werden die weiteren Einflußgrößen wie folgt vorgegeben: $\mathbf{a} = 0.05$, $J^* = 0.20$, $J = 0.00, 0.05, 0.10, 0.15$ und $CV_e = 5.0, 7.5, 10.0, \dots, 30.0\%$.

Da die Korrelation \mathbf{r}^d zwischen den Teststatistiken t_{11}^d und t_{12}^d genau dem Wert 1 entspricht, wird die Wahrscheinlichkeit für die korrekte Verwerfung für das zweiseitige Testproblem unter Ausnutzung der Owen-Technik berechnet:

$$P[t_{11}^d \geq t_{\mathbf{a}, 2n-2} \wedge t_{12}^d \leq -t_{\mathbf{a}, 2n-2} \mid \mathbf{d} < \mathbf{m}_1 - \mathbf{m}_0 < \mathbf{d}_2, \mathbf{s}_e] =$$

$$Q(-t_{\mathbf{a}, 2n-2}, \mathbf{D}_2, 0, \mathbf{y}) - Q(t_{\mathbf{a}, 2n-2}, \mathbf{D}_1, 0, \mathbf{y}),$$

wobei

$$D_1 = \frac{J - J_1}{CV_e \sqrt{\frac{1}{n}}} \quad \text{und} \quad D_2 = \frac{J - J_2}{CV_e \sqrt{\frac{1}{n}}}.$$

Tabelle 14: Crossover-Situation: Exakter und approximativer Stichprobenumfang pro Sequenz für eine Güte von mindestens 0.80 bei $\alpha = 0.05$, $(-\infty, \mathbf{d}) = (-\infty, 0.2\mathbf{m})$ und verschiedene CV_e

CV_e (%)	J			
	0.00	0.05	0.10	0.15
5.0	2	2	3	7
	3	3	-	8
7.5	3	3	5	15
	-	-	-	-
10.0	3	4	7	26
	-	-	8	-
12.5	4	6	11	40
	-	-	-	-
15.0	5	7	15	57
	-	8	-	-
17.5	6	10	20	77
	-	-	-	-
20.0	7	12	26	100
	8	-	-	-
22.5	9	15	32	126
	-	-	33	-
25.0	11	18	40	156
	-	-	-	-
27.5	13	22	48	188
	-	-	-	-
30.0	15	26	57	224
	-	-	-	-

In der Abbildung 11 werden die Gütefunktionen für einen symmetrischen Sicherheitsbereich von $(-0.2\mathbf{m}_0, 0.2\mathbf{m}_0)$, einen Variationskoeffizienten von 15.0% und für die Stichprobenumfänge von jeweils 10 bzw. 20 pro Sequenz dargestellt.

Die vereinfachte Berechnung der Fallzahlen erfolgt analog zur Zweistichprobensituation nach den folgenden Approximationsformeln, wobei ebenfalls ein zu 0 symmetrischer Sicherheitsbereich gewählt wurde, d.h. $J_1 = -J_2$:

$$\text{falls } J = 0 \quad n \geq \left(\frac{CV_e}{J_2} \right)^2 \left(t_{a,2n-2} + t_{\frac{b}{2},2n-2} \right)^2,$$

$$0 < J < J_2 \quad n \geq \left(\frac{CV_e}{J_2 - J} \right)^2 \left(t_{a,2n-2} + t_{b,2n-2} \right)^2$$

$$\text{und } -J_2 < J < 0 \quad n \geq \left(\frac{CV_e}{-J_2 - J} \right)^2 \left(t_{a,2n-2} + t_{b,2n-2} \right)^2.$$

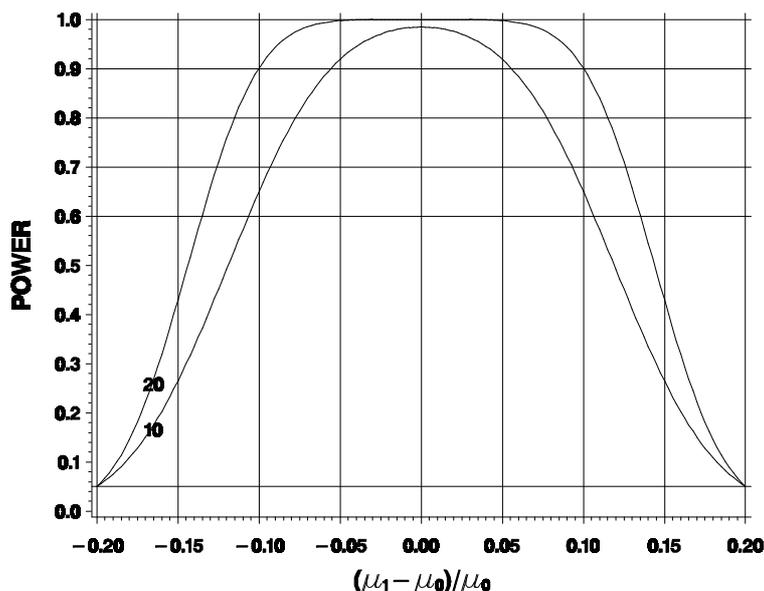


Abbildung 11: Crossover-Situation: Die Gütefunktion des Entscheidungsverfahrens, basierend auf den zwei einseitigen Tests t_{11}^d, t_{12}^d bei einem Sicherheitsbereich von $(\underline{d}, \underline{d}) = (-0.2m, 0.2m)$ für $\alpha = 0.05, n_0 = n_1 = 10, 20$ und $CV_e = 15.0\%$

In der Tabelle 15 sind die exakten und approximativen Fallzahlen für $\alpha = 0.05, b = 0.20, J_1 = -J_2 = -0.20, CV_e = 5.0, 10.0, \dots, 30.0\%$ und für $|J| = 0.00, 0.02, 0.05, 0.10, 0.15$ angegeben.

Tabelle 15: Crossover-Situation: Exakter und approximativer Stichprobenumfang pro Sequenz für eine Güte von mindestens 0.80 bei $\alpha = 0.05, (\underline{d}, \underline{d}) = (-0.2m, 0.2m)$ und verschiedene CV_e

CV_e (%)	$ J $			
	0.00	0.05	0.10	0.15
5.0	2	2	3	7
	3	3	-	8
7.5	3	3	5	15
	-	-	-	-
10.0	4	4	7	26
	-	-	8	-
12.5	5	6	11	40
	-	-	-	-
15.0	6	7	15	57
	-	8	-	-
17.5	8	10	20	77
	-	-	-	-
20.0	10	12	26	100
	-	-	-	-
22.5	12	15	32	126
	-	-	33	-
25.0	15	19	40	156
	-	18	-	-
27.5	17	22	48	188
	18	-	-	-
30.0	20	26	57	224
	21	-	-	-

4.2 Sicherheitsnachweis für den Quotienten von Lokationsparametern

Seien

$$Y_{ij}^q = \begin{cases} \mathbf{q}X_{ij1} - X_{ij2} & i = 0 \\ X_{ij1} - \mathbf{q}X_{ij2} & i = 1 \end{cases}$$

die um \mathbf{q} korrigierten j -ten intraindividuellen Periodendifferenzen, $j = 1, \dots, n_i$, und seien ferner $G_i(y)$ die Verteilungsfunktionen der Differenzen der zugehörigen Fehlerterme:

$$G_i(y) = \begin{cases} P(\mathbf{q}\mathbf{e}_{ij1} - \mathbf{e}_{ij2} \leq y) & i = 0 \\ P(\mathbf{e}_{ij1} - \mathbf{q}\mathbf{e}_{ij2} \leq y) & i = 1 \end{cases} .$$

Mit der Voraussetzung $G_0(y) = G_1(y) = G(y)$ folgt unmittelbar:

$$Y_{0j}^q \text{ u.i.v. } \sim G(y + \mathbf{m}_1 - \mathbf{q}\mathbf{m}_0 + \mathbf{p}_2 - \mathbf{q}\mathbf{p}_1), \quad j = 1, \dots, n_0,$$

$$Y_{1j}^q \text{ u.i.v. } \sim G(y + \mathbf{q}\mathbf{m}_0 - \mathbf{m}_1 + \mathbf{q}\mathbf{p}_2 - \mathbf{p}_1), \quad j = 1, \dots, n_1,$$

d.h. die Verteilungen der Differenzen der beiden Sequenzen unterscheiden sich nur durch den Lokationsunterschied von

$$2(\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0) + \mathbf{p}_1 - \mathbf{q}\mathbf{p}_2 + \mathbf{p}_2 - \mathbf{q}\mathbf{p}_1 = 2(\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0) + (\mathbf{p}_1 + \mathbf{p}_2)(1 - \mathbf{q}) = 2(\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0),$$

da $\mathbf{p}_1 + \mathbf{p}_2 = 0$. Somit können Hypothesen über $\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0 \leq 0 \Leftrightarrow \frac{\mathbf{m}_1}{\mathbf{m}_0} \leq \mathbf{q}$ durch Verfahren für zwei unverbundene Stichproben untersucht werden.

4.2.1 Einseitiges Testproblem

Im Model I mit der bivariaten Normalverteilungsannahme $(\mathbf{e}_{ij1}, \mathbf{e}_{ij2})' \text{ u.i.v. } \sim N_2(0, \mathbf{S}_i)$,

$$\mathbf{S}_0 = \begin{pmatrix} \mathbf{s}_0^2 & \mathbf{s}_{01} \\ \mathbf{s}_{01} & \mathbf{s}_1^2 \end{pmatrix} \quad \text{und} \quad \mathbf{S}_1 = \begin{pmatrix} \mathbf{s}_1^2 & \mathbf{s}_{01} \\ \mathbf{s}_{01} & \mathbf{s}_0^2 \end{pmatrix},$$

benötigt man zur Konstruktion von Entscheidungsverfahren das folgende Lemma.

Lemma 4:

Der Schätzer $\frac{\bar{X}_1 - \mathbf{q}\bar{X}_0}{2}$ ist normalverteilt mit Erwartungswert $\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0$ und Varianz

$$\mathbf{s}_q^2 \frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right), \text{ wobei}$$

$$\mathbf{s}_q^2 = \mathbf{q}^2 \mathbf{s}_0^2 + \mathbf{s}_1^2 - 2\mathbf{q}\mathbf{s}_{01}.$$

Beweis:

Die Verteilungsfunktionen der um \mathbf{q} korrigierten Periodendifferenzen sind:

$$Y_{ij}^q \text{ u.i.v. } \sim G(y - E(Y_{ij}^q)) = \mathbf{F} \left(\frac{y - E(Y_{ij}^q)}{\text{Var}(Y_{ij}^q)} \right),$$

mit

$$E(Y_{0j}^q) = \mathbf{q}\mathbf{m}_0 - \mathbf{m}_1 + \mathbf{q}\mathbf{p}_1 - \mathbf{p}_2, \quad E(Y_{1j}^q) = \mathbf{m}_1 - \mathbf{q}\mathbf{m}_0 + \mathbf{p}_1 - \mathbf{q}\mathbf{p}_2,$$

$$\text{Var}(Y_{ij}^q) = \mathbf{s}_q^2 = \mathbf{q}^2 \mathbf{s}_0^2 + \mathbf{s}_1^2 - 2\mathbf{q}\mathbf{s}_{01}, \quad i = 0, 1 \quad \text{und} \quad j = 1, \dots, n_i.$$

Der Erwartungswert für die Differenz der transformierten Zufallsvariablen aus den beiden Stichproben beträgt

$$E(Y_{1j}^q - Y_{0j}^q) = 2(\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0) + (\mathbf{p}_1 + \mathbf{p}_2)(1 - \mathbf{q}) = 2(\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0),$$

und ein erwartungstreuer Schätzer für die Differenz $\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0$ ist

$$\frac{\bar{Y}_1^q - \bar{Y}_0^q}{2} = \frac{1}{2} \left(\frac{1}{n_1} \sum_{j=1}^{n_1} (X_{1j1} - \mathbf{q}X_{1j2}) - \frac{1}{n_0} \sum_{j=1}^{n_0} (\mathbf{q}X_{0j1} - X_{0j2}) \right) = \frac{\bar{X}_1 - \mathbf{q}\bar{X}_0}{2}.$$

Die Varianz des Schätzers ist

$$\text{Var} \left(\frac{\bar{X}_1 - \mathbf{q}\bar{X}_0}{2} \right) = \text{Var} \left(\frac{\bar{Y}_1^q - \bar{Y}_0^q}{2} \right) = \mathbf{s}_q^2 \frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)$$

da

$$\text{Var}(Y_{ij}^d) = \mathbf{s}_q^2, \quad i = 0, 1 \quad \text{und} \quad j = 1, \dots, n_i.$$

▷

Die Varianz \mathbf{s}_q^2 kann erwartungstreu durch $S_q^2 = \mathbf{q}^2 S_0^2 + S_1^2 - 2\mathbf{q}S_{01}$ geschätzt werden, wobei

$$S_0^2 = \frac{1}{n_0 + n_1 - 2} \left(\sum_{j=1}^{n_0} (X_{0j1} - \bar{X}_{0.1})^2 + \sum_{j=1}^{n_1} (X_{1j2} - \bar{X}_{1.2})^2 \right),$$

$$S_1^2 = \frac{1}{n_0 + n_1 - 2} \left(\sum_{j=1}^{n_0} (X_{0j2} - \bar{X}_{0.2})^2 + \sum_{j=1}^{n_1} (X_{1j1} - \bar{X}_{1.1})^2 \right),$$

$$S_{01} = \frac{1}{n_0 + n_1 - 2} \left(\sum_{j=1}^{n_0} (X_{0j1} - \bar{X}_{1.1})(X_{0j2} - \bar{X}_{0.2}) + \sum_{j=1}^{n_1} (X_{1j1} - \bar{X}_{1.1})(X_{1j2} - \bar{X}_{1.2}) \right)$$

$$\text{und } \bar{X}_{i.k} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ijk}.$$

Die Verteilung von $\frac{S_q^2}{\mathbf{s}_q^2} (n_0 + n_1 - 2)$ ist \mathbf{c}^2 -verteilt mit $n_0 + n_1 - 2$ Freiheitsgraden (Locke, 1984). Das auf den Quotienten der Erwartungswerte bezogene einseitige Testproblem H_0^q mit einem irrelevanten Bereich von $(-\infty, \mathbf{q})$, $\mathbf{q} > 1$,

$$H_0^q: \frac{\mathbf{m}_1}{\mathbf{m}_0} \geq \mathbf{q} \quad \text{versus} \quad H_1^q: \frac{\mathbf{m}_1}{\mathbf{m}_0} < \mathbf{q},$$

wird genau dann abgelehnt, falls

$$t_1^q = \frac{\bar{X}_1 - \mathbf{q}\bar{X}_0}{S_q \sqrt{\frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)}} \leq -t_{\mathbf{a}, n_0 + n_1 - 2}.$$

Die Bedingung $t_1^q \leq -t_{\mathbf{a}, n_0 + n_1 - 2}$ ist äquivalent zu

$$(-\infty, \mathbf{q}_+^1) \subset (-\infty, \mathbf{q}) \quad \text{und} \quad \bar{X}_0^2 > a_0,$$

wobei

$$\mathbf{q}_+^1 = \frac{(\bar{X}_1 \bar{X}_0 - a_{01}) + \sqrt{(\bar{X}_1 \bar{X}_0 - a_{01})^2 - (\bar{X}_1^2 - a_1)(\bar{X}_0^2 - a_0)}}{(\bar{X}_0^2 - a_0)},$$

$$a_1 = \frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right) S_1^2 t_{\mathbf{a}, n_0+n_1-2}^2, \quad a_0 = \frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right) S_0^2 t_{\mathbf{a}, n_0+n_1-2}^2, \quad a_{01} = \frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right) S_{01}^2 t_{\mathbf{a}, n_0+n_1-2}^2.$$

Ist die obere Grenze \mathbf{q}_+^1 des einseitigen $100(1-\mathbf{a})\%$ -Konfidenzintervalles für $\frac{\mathbf{m}_1}{\mathbf{m}_0}$ unterhalb von \mathbf{q} , so kann zum Niveau \mathbf{a} auf vorliegende Sicherheit geschlossen werden.

Im Modell II wird H_0^q genau dann abgelehnt, falls

$$t_1^q = \frac{\bar{X}_1 - \mathbf{q}\bar{X}_0}{2 \sqrt{S_e^2(1+\mathbf{q}^2) + S_b^2(1-\mathbf{q})^2}} \sqrt{\frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)} \leq -t_{\mathbf{a}, n_0+n_1-2},$$

mit

$$S_b^2 = \frac{1}{2} \left(\sum_{i=0}^1 \sum_{j=1}^{n_i} \frac{X_{ij}^2}{2} - \sum_{i=0}^1 \frac{X_{i..}^2}{2n_i} - S_e^2 \right).$$

Kieser und Hauschke (1999b) haben gezeigt, daß die Bedingung $t_1^q \leq -t_{\mathbf{a}, n_0+n_1-2}$ äquivalent ist zu

$$(-\infty, \mathbf{q}_+^1) \subset (-\infty, \mathbf{q}) \quad \text{und} \quad \bar{X}_0^2 > a_{e+tb},$$

wobei

$$\mathbf{q}_+^1 = \frac{(\bar{X}_1 \bar{X}_0 - a_b) + \sqrt{(\bar{X}_1 \bar{X}_0 - a_b)^2 - (\bar{X}_1^2 - a_{e+tb})(\bar{X}_0^2 - a_{e+tb})}}{(\bar{X}_0^2 - a_{e+tb})},$$

$$a_b = \frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right) S_b^2 t_{\mathbf{a}, n_0+n_1-2}^2, \quad a_{e+tb} = \frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right) (S_e^2 + S_b^2) t_{\mathbf{a}, n_0+n_1-2}^2.$$

Man beachte, daß der Schätzer S_b^2 aufgrund der folgenden Beziehung nach der Satterthwaites Methode berechnet wurde:

$$E \left(\frac{1}{n_0 + n_1 - 2} \left(\sum_{i=0}^1 \sum_{j=1}^{n_i} \frac{X_{ij}^2}{2} - \sum_{i=0}^1 \frac{X_{i..}^2}{2n_i} \right) \right) = 2\mathbf{s}_b^2 + \mathbf{s}_e^2.$$

Es sei daher darauf hingewiesen, daß im Vorliegen der Varianzhomogenität (Modell II) die Teststatistik t_1^q nur approximativ t -verteilt ist (Vuorinen und Tuominen, 1994), wobei für die nach der Satterthwaites Approximation berechneten Freiheitsgrade \mathbf{n} gilt

$$n_0 + n_1 - 2 \leq \mathbf{n} \leq 2(n_0 + n_1 - 2).$$

Nimmt man nun die untere Grenze, so resultiert eine konservative Vorgehensweise, welche auch von Vuorinen und Tuominen (1994) und Vuorinen und Turunen (1996) aus praktischen Erwägungen vorgeschlagen wird.

Verzichtet man auf die Normalverteilungsannahme, so kann wegen des vorliegenden Lokationsmodells

$$Y_{0j}^q \text{ u.i.v. } \sim G(y + \mathbf{m}_1 - \mathbf{q}\mathbf{m}_0 + \mathbf{p}_2 - \mathbf{q}\mathbf{p}_1), \quad j = 1, \dots, n_0,$$

$$Y_{1j}^q \text{ u.i.v. } \sim G(y + \mathbf{q}\mathbf{m}_0 - \mathbf{m}_1 + \mathbf{q}\mathbf{p}_2 - \mathbf{p}_1), \quad j = 1, \dots, n_1,$$

sowohl für Model I als auch für Modell II der Wilcoxon-Test (Vuorinen und Turunen, 1997) und das verteilungsfreie Konfidenzintervall nach Moses als Entscheidungsverfahren herangezogen werden. Unter H_0^q gilt

$$Y_{0j}^q \text{ u.i.v. } \sim G(y + \mathbf{p}_2 - \mathbf{q}\mathbf{p}_1) = G(z), \quad j = 1, \dots, n_0$$

$$Y_{1j}^q \text{ u.i.v. } \sim G(y + \mathbf{q}\mathbf{p}_2 - \mathbf{p}_1) = G(z), \quad j = 1, \dots, n_1,$$

da aus $\mathbf{p}_1 + \mathbf{p}_2 = 0$ folgt, daß $\mathbf{p}_1 - \mathbf{q}\mathbf{p}_2 = \mathbf{q}\mathbf{p}_1 - \mathbf{p}_2$. Dabei muß jedoch nun beachtet werden, daß die Rangsumme der um \mathbf{q} korrigierten j -ten intraindividuellen Periodendifferenzen Y_{1j}^q verwendet wird. Sei $R_{1j}(\mathbf{q})$ der Rang von $Y_{1j}^q = X_{1j1} - \mathbf{q}X_{1j2}$ in der kombinierten Stichprobe

$$Y_{01}^q, \dots, Y_{0n_0}^q, Y_{11}^q, \dots, Y_{1n_1}^q =$$

$$\mathbf{q}X_{011} - X_{012}, \dots, \mathbf{q}X_{0n_01} - X_{0n_02}, X_{111} - \mathbf{q}X_{112}, \dots, X_{1n_11} - \mathbf{q}X_{1n_12}$$

und $R_1(\mathbf{q}) = \sum_{j=1}^{n_1} R_{1j}(\mathbf{q})$. Die Nullhypothese H_0^q wird verworfen, falls

$$R_1(\mathbf{q}) \leq n_1(n_0 + n_1 + 1) - r_{\mathbf{a}, n_0, n_1}^q,$$

wobei $r_{\mathbf{a},n_0,n_1}^q$ das $(1-\mathbf{a})$ -Quantil der Wilcoxonstatistik $R_1(\mathbf{q})$ ist. Die Konstruktion des Konfidenzintervalles basiert wiederum auf den geordneten Werten $D_1^q \leq \dots \leq D_{n_0 n_1}^q$ der insgesamt $n_0 n_1$ Differenzen $Y_{1j}^q - Y_{0j^*}^q, j = 1, \dots, n_1$ und $j^* = 1, \dots, n_0$. Damit ergibt sich als einseitiges $100(1-\mathbf{a})\%$ -Konfidenzintervall für $\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0$:

$$\left(-\infty, \frac{U_1^q}{2} \right], \text{ wobei } U_1^q = D_{u_1}^q = D_{n_0 n_1 + 1 - C_a^1}^q, \quad C_a^1 = \frac{n_1(2n_0 + n_1 + 1)}{2} + 1 - r_{\mathbf{a},n_0,n_1}^q.$$

Liegt die obere Grenze des einseitigen Konfidenzintervalles unterhalb von 0, so kann die Nullhypothese verworfen und auf Sicherheit geschlossen werden.

4.2.2 Zweiseitiges Testproblem

Das Testen der zweiseitigen Fragestellung

$$H_0^q: \frac{\mathbf{m}_1}{\mathbf{m}_0} \leq \mathbf{q}_1 \quad \text{oder} \quad \frac{\mathbf{m}_1}{\mathbf{m}_0} \geq \mathbf{q}_2$$

versus

$$H_1^q: \mathbf{q}_1 < \frac{\mathbf{m}_1}{\mathbf{m}_0} < \mathbf{q}_2,$$

mit $\mathbf{q}_1 = 1 + f_1, \mathbf{q}_2 = 1 + f_2$ und $-1 < f_1 < 0 < f_2$ ist äquivalent zur simultanen Überprüfung der beiden einseitigen Testprobleme:

$$H_{01}^q: \frac{\mathbf{m}_1}{\mathbf{m}_0} \leq \mathbf{q}_1 \quad \text{versus} \quad H_{11}^q: \frac{\mathbf{m}_1}{\mathbf{m}_0} > \mathbf{q}_1$$

und

$$H_{02}^q: \frac{\mathbf{m}_1}{\mathbf{m}_0} \geq \mathbf{q}_2 \quad \text{versus} \quad H_{12}^q: \frac{\mathbf{m}_1}{\mathbf{m}_0} < \mathbf{q}_2.$$

$H_0^q = \bigcup_{r=1}^2 H_{0r}^q$ wird bei der bivariaten Normalverteilungsannahme (Modell I) verworfen, falls

$$t_{11}^q = \frac{\bar{X}_1 - \mathbf{q}_1 \bar{X}_0}{S_{q_1} \sqrt{\frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)}} \geq t_{\mathbf{a}, n_0 + n_1 - 2} \quad \text{und} \quad t_{12}^q = \frac{\bar{X}_1 - \mathbf{q}_2 \bar{X}_0}{S_{q_2} \sqrt{\frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)}} \leq -t_{\mathbf{a}, n_0 + n_1 - 2}.$$

Die Verwerfung beider einseitigen Hypothesen ist äquivalent zu

$$[\mathbf{q}_-^1, \mathbf{q}_+^1] \subset (\mathbf{q}_1, \mathbf{q}_2) \quad \text{und} \quad \bar{X}_0^2 > a_0, \quad \text{wobei}$$

$$\mathbf{q}_\pm^1 = \frac{(\bar{X}_1 \bar{X}_0 - a_{01}) \pm \sqrt{(\bar{X}_1 \bar{X}_0 - a_{01})^2 - (\bar{X}_1^2 - a_1)(\bar{X}_0^2 - a_0)}}{(\bar{X}_0^2 - a_0)}.$$

Der Schluß auf Sicherheit zum Niveau \mathbf{a} ist somit möglich, falls das $100(1-2\mathbf{a})\%$ -Konfidenzintervall für $\frac{\mathbf{m}_1}{\mathbf{m}_0}$ regulär und vollständig im Sicherheitsbereich enthalten $(\mathbf{q}_1, \mathbf{q}_2)$ ist (Hauschke *et al.*, 1999c).

Im Modell II (Kieser und Hauschke, 1999b) wird $H_0^q = \bigcup_{r=1}^2 H_{0r}^q$ genau dann abgelehnt, falls

$$t_{11}^q = \frac{\bar{X}_1 - \mathbf{q}_1 \bar{X}_0}{\sqrt{S_e^2(1+\mathbf{q}_1^2) + S_b^2(1-\mathbf{q}_1)^2} \sqrt{\frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)}} \leq -t_{\mathbf{a}, n_0 + n_1 - 2}$$

und

$$t_{12}^q = \frac{\bar{X}_1 - \mathbf{q}_2 \bar{X}_0}{\sqrt{S_e^2(1+\mathbf{q}_2^2) + S_b^2(1-\mathbf{q}_2)^2} \sqrt{\frac{1}{4} \left(\frac{1}{n_0} + \frac{1}{n_1} \right)}} \geq t_{\mathbf{a}, n_0 + n_1 - 2}.$$

Die Verwerfung beider Hypothesen ist äquivalent zu

$$[\mathbf{q}_-^1, \mathbf{q}_+^1] \subset (\mathbf{q}_1, \mathbf{q}_2) \quad \text{und} \quad \bar{X}_0^2 > a_{e+b},$$

wobei

$$\mathbf{q}_\pm^1 = \frac{(\bar{X}_1 \bar{X}_0 - a_b) \pm \sqrt{(\bar{X}_1 \bar{X}_0 - a_b)^2 - (\bar{X}_1^2 - a_{e+b})(\bar{X}_0^2 - a_{e+b})}}{(\bar{X}_0^2 - a_{e+b})}.$$

Für das entsprechende verteilungsfreie Verfahren mit Zweistichproben-Wilcoxon-Tests gilt

$$H_{01}^q \text{ wird abgelehnt, falls } R_1(\mathbf{q}_1) \geq r_{\mathbf{a},n_0,n_1}^{q_1}$$

$$\text{und } H_{02}^q \text{ wird abgelehnt, falls } R_1(\mathbf{q}_2) \leq n_1(n_0 + n_1 + 1) - r_{\mathbf{a},n_0,n_1}^{q_2}.$$

Die Verwerfung beider Einzelhypothesen ist äquivalent zur Bedingung, daß die untere Grenze des einseitigen $100(1-\mathbf{a})\%$ -Konfidenzintervalles für $\mathbf{m}_1 - \mathbf{q}_1 \mathbf{m}_0$ oberhalb von 0 liegt, und ferner, daß die obere Grenze des einseitigen $100(1-\mathbf{a})\%$ -Konfidenzintervalles für $\mathbf{m}_1 - \mathbf{q}_2 \mathbf{m}_0$ unterhalb von 0 liegt:

$$\left[\frac{L_1^{q_1}}{2}, \infty \right) \subset (0, \infty) \quad \text{und} \quad \left(-\infty, \frac{U_1^{q_2}}{2} \right] \subset (-\infty, 0),$$

$$\text{wobei } L_1^{q_1} = D_{l_1}^{q_1} = D_{C_a^1}^{q_1} \quad \text{und} \quad U_1^{q_2} = D_{u_1}^{q_2} = D_{n_0 n_1 + 1 - C_a^1}^{q_2}.$$

4.2.3 Gütebestimmung und Fallzahlplanung

Die Güte des einseitigen Sasabuchi-Tests t_1^q im *Crossover*-Design unter Annahme des Modells II läßt sich über die nichtzentrale t -Verteilung ermitteln:

$$P \left[t_1^q \leq -t_{\mathbf{a},2n-2} \left| \frac{\mathbf{m}_1}{\mathbf{m}_0} < \mathbf{q}, \mathbf{s}_q \right. \right] = \frac{\sqrt{2\mathbf{p}}}{\mathbf{G}(n-1)2^{n-2}} \int_0^\infty \mathbf{F} \left(\frac{-t_{\mathbf{a},2n-2} x}{\sqrt{2n-2}} - \mathbf{q} \right) x^{2n-3} \mathbf{F}'(x) dx,$$

mit dem Nichtzentralitätsparameter:

$$\mathbf{q} = \frac{\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0}{\mathbf{s}_q \sqrt{\frac{1}{2n}}} = \frac{\mathbf{m}_1 - \mathbf{q}\mathbf{m}_0}{\sqrt{\frac{\mathbf{s}_e^2(1+\mathbf{q}^2) + \mathbf{s}_b^2(1-\mathbf{q})^2}{2n}}} = \frac{\frac{\mathbf{m}_1}{\mathbf{m}_0} - \mathbf{q}}{\sqrt{\frac{CV_e^2(1+\mathbf{q}^2) + CV_b^2(1-\mathbf{q})^2}{2n}}},$$

mit $\mathbf{s}_e = CV_e \mathbf{m}_0$ und $\mathbf{s}_b = CV_b \mathbf{m}_0$.

In Abbildung 12 wird die Gütefunktion für verschiedene Variationskoeffizienten CV_e bei einem festen $CV_b = 30\%$, einem Sicherheitsbereich von $(-\infty, \mathbf{q}) = (-\infty, 1.2)$ und jeweils 10 Versuchseinheiten pro Sequenz dargestellt.

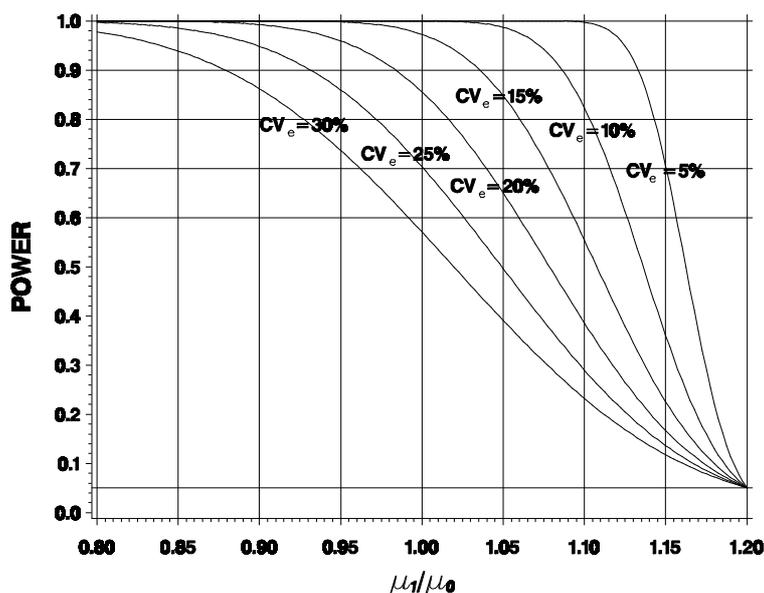


Abbildung 12: Crossover-Situation: Die Gütefunktion des einseitigen Tests t_1^q bei einem Sicherheitsbereich von $(-\infty, q) = (-\infty, 1.2)$ für $\alpha = 0.05$, $n_0 = n_1 = 10$, $CV_b = 0.30$ und verschiedene CV_e

Den notwendigen Stichprobenumfang n (pro Sequenz), um beim Testen von H_0^q versus H_1^q die Fehler α und β an der Stelle $\frac{m_1}{m_0}, \frac{m_1}{m_0} < q$, nicht zu überschreiten, kann man iterativ nach der folgenden Ungleichung berechnen:

$$n \geq \frac{1}{2} \frac{CV_e^2(1+q^2) + CV_b^2(1-q)^2}{\left(q - \frac{m_1}{m_0}\right)^2} (t_{\alpha, 2n-2} + t_{\beta, 2n-2})^2.$$

Die Herleitung erfolgt analog zur Zweistichprobensituation.

In den Tabellen 16 und 17 sind die exakten und approximativen Fallzahlen angegeben, um eine Güte von mindestens 0.80 zu erreichen. Dabei sind die weiteren Einflußgrößen wie folgt vorgegeben: $\alpha = 0.05$, $q = 1.20$, $\frac{m_1}{m_0} = 1.00, 1.05, 1.10, 1.15$ und $CV_0 = 5.0, 7.5, 10.0, \dots, 30.0\%$.

Tabelle 16: Crossover-Situation: Exakter Stichprobenumfang pro Sequenz für eine Güte von mindestens 0.80 bei $\alpha = 0.05$, $(-\infty, \mathbf{q}) = (-\infty, 1.2)$ und verschiedene CV_e , CV_b .

CV_e (%)	CV_b (%)	m_1/m_0			
		1.00	1.05	1.10	1.15
5.0	20.0	2	3	4	11
	40.0	3	3	5	17
	60.0	3	4	8	27
	80.0	4	6	11	40
	100.0	5	8	15	58
10.0	20.0	4	5	9	13
	40.0	4	6	11	39
	60.0	4	7	13	49
	80.0	5	8	17	63
	100.0	6	10	21	81
15.0	20.0	6	9	19	71
	40.0	6	10	20	77
	60.0	7	11	23	87
	80.0	8	12	26	101
	100.0	9	14	31	119
20.0	20.0	9	15	32	124
	40.0	9	16	33	130
	60.0	10	17	36	140
	80.0	11	18	39	154
	100.0	12	20	44	171
25.0	20.0	13	22	49	192
	40.0	14	23	50	198
	60.0	14	24	53	208
	80.0	15	26	56	221
	100.0	16	28	61	239
30.0	20.0	18	32	70	275
	40.0	19	32	71	281
	60.0	19	33	74	291
	80.0	20	35	77	304
	100.0	21	37	81	322

Für das zweiseitige Testproblem erfolgt die Gütebestimmung nach

$$P \left[t_{11}^q \geq t_{\alpha, 2n-2} \wedge t_{12}^q \leq t_{\alpha, 2n-2} \mid \mathbf{q}_1 < \frac{m_1}{m_0} < \mathbf{q}_2, \mathbf{s}_{q_1}, \mathbf{s}_{q_2} \right] =$$

$$Q(\infty, -t_{\alpha, 2n-2}, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{r}^q) - Q(t_{\alpha, 2n-2}, -t_{\alpha, 2n-2}, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{r}^q),$$

wobei

$$Q(t_1, t_2, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{r}) = P \left[t_{11}^q \leq t_1 \wedge t_{12}^q \leq t_2 \mid \mathbf{q}_1 < \frac{m_1}{m_0} < \mathbf{q}_2, \mathbf{s} \right] =$$

$$\frac{\sqrt{2p}}{G(n-1)2^{n-2}} \int_0^\infty F_2 \left(\frac{t_1 x}{\sqrt{2n-2}} - \mathbf{q}_1, \frac{t_2 x}{\sqrt{2n-2}} - \mathbf{q}_2, \mathbf{r} \right) x^{2n-3} F(x) dx$$

und

$$\mathbf{q}_1 = \frac{m_1 - q_1 m_0}{s_{q_1} \sqrt{\frac{1}{2n}}} \quad \text{und} \quad \mathbf{q}_2 = \frac{m_1 - q_2 m_0}{s_{q_2} \sqrt{\frac{1}{2n}}}.$$

Tabelle 17: Crossover-Situation: Approximativer Stichprobenumfang pro Sequenz für eine Güte von mindestens 0.80 bei $\alpha = 0.05$, $(-\infty, \mathbf{q}) = (-\infty, 1.2)$ und verschiedene CV_e , CV_b .

CV_e (%)	CV_b (%)	m_1/m_0			
		1.00	1.05	1.10	1.15
5.0	20.0	3	3	4	11
	40.0	3	3	5	17
	60.0	3	4	8	27
	80.0	4	6	11	40
	100.0	5	8	16	58
10.0	20.0	4	5	9	13
	40.0	4	6	11	39
	60.0	4	7	13	49
	80.0	5	8	17	63
	100.0	6	10	21	81
15.0	20.0	6	9	19	71
	40.0	6	10	20	77
	60.0	7	11	23	87
	80.0	8	12	26	101
	100.0	9	14	31	119
20.0	20.0	9	15	32	124
	40.0	9	16	33	130
	60.0	10	17	36	140
	80.0	11	18	39	154
	100.0	12	20	44	171
25.0	20.0	13	22	49	192
	40.0	14	23	50	198
	60.0	14	24	53	208
	80.0	15	26	56	221
	100.0	16	28	61	239
30.0	20.0	18	32	70	275
	40.0	19	32	71	281
	60.0	19	33	74	291
	80.0	20	35	77	304
	100.0	21	37	82	322

Im Modell II gilt

$$\mathbf{q}_1 = \frac{\frac{\mathbf{m}_1}{\mathbf{m}_0} - \mathbf{q}_1}{\sqrt{\frac{CV_e^2(1+\mathbf{q}_1^2) + CV_b^2(1-\mathbf{q}_1)^2}{2n}}} \quad \text{und} \quad \mathbf{q}_2 = \frac{\frac{\mathbf{m}_1}{\mathbf{m}_0} - \mathbf{q}_2}{\sqrt{\frac{CV_e^2(1+\mathbf{q}_2^2) + CV_b^2(1-\mathbf{q}_2)^2}{2n}}}.$$

Der benötigte Korrelationskoeffizient \mathbf{r}^q folgt mit Lemma 5.

Lemma 5:

Es gilt

$$\mathbf{r}^q = \frac{CV_e^2(1+\mathbf{q}_1\mathbf{q}_2) + CV_b^2(1+\mathbf{q}_1\mathbf{q}_2 - \mathbf{q}_1 - \mathbf{q}_2)}{\sqrt{(CV_e^2(1+\mathbf{q}_1^2) + CV_b^2(1-\mathbf{q}_1)^2)(CV_e^2(1+\mathbf{q}_2^2) + CV_b^2(1-\mathbf{q}_2)^2)}}$$

Beweis:

$$Var(\bar{X}_1 - \mathbf{q}_r \bar{X}_0) = \frac{1}{2n} (\mathbf{s}_e^2(1+\mathbf{q}_r^2) + \mathbf{s}_b^2(1-\mathbf{q}_r)^2) = \frac{\mathbf{m}_0^2}{2n} (CV_e^2(1+\mathbf{q}_r^2) + CV_b^2(1-\mathbf{q}_r)^2), \quad r = 1, 2.$$

$$Kov(\bar{X}_1 - \mathbf{q}_1 \bar{X}_0, \bar{X}_1 - \mathbf{q}_2 \bar{X}_0) = Var(\bar{X}_1) + Var(\bar{X}_0)\mathbf{q}_1\mathbf{q}_2 - Kov(\bar{X}_1, \bar{X}_0)(\mathbf{q}_1 + \mathbf{q}_2) =$$

$$\frac{1}{2n} (\mathbf{s}_e^2 + \mathbf{s}_b^2 + (\mathbf{s}_e^2 + \mathbf{s}_b^2)\mathbf{q}_1\mathbf{q}_2 - \mathbf{s}_b^2(\mathbf{q}_1 + \mathbf{q}_2)) = \frac{1}{2n} (\mathbf{s}_e^2(1+\mathbf{q}_1\mathbf{q}_2) + \mathbf{s}_b^2(1+\mathbf{q}_1\mathbf{q}_2 - \mathbf{q}_1 - \mathbf{q}_2)) =$$

$$\frac{\mathbf{m}_0^2}{2n} (CV_e^2(1+\mathbf{q}_1\mathbf{q}_2) + CV_b^2(1+\mathbf{q}_1\mathbf{q}_2 - \mathbf{q}_1 - \mathbf{q}_2))$$

und daher

$$\mathbf{r}^q = \frac{CV_e^2(1+\mathbf{q}_1\mathbf{q}_2) + CV_b^2(1+\mathbf{q}_1\mathbf{q}_2 - \mathbf{q}_1 - \mathbf{q}_2)}{\sqrt{(CV_e^2(1+\mathbf{q}_1^2) + CV_b^2(1-\mathbf{q}_1)^2)(CV_e^2(1+\mathbf{q}_2^2) + CV_b^2(1-\mathbf{q}_2)^2)}}.$$

▷

In Abbildung 13 werden die Gütefunktionen für einen Sicherheitsbereich von $(\mathbf{q}_1, \mathbf{q}_2) = (0.8, 1.25)$, für die Variationskoeffizienten $CV_e = 32.5\%$, $CV_b = 30.0\%$ und für verschiedene Fallzahlen dargestellt. Anhand der Grafik erkennt man, daß bei kleinen Stichproben das aktuelle Niveau an der Stelle 0.8 kleiner ist als das nominelle Niveau. Bei größer werdenden Stichprobenanzahlen konvergiert es jedoch gegen das nominelle Niveau (Schuirmann, 1989).

Yee (1986) leitete bereits untere und obere Grenzen für die Güte her, die jedoch nur in der Nähe von $\frac{m_1}{m_0} = 1$ nahezu übereinstimmten und somit exakte Gütwerte erbrachten. Vuorinen und Tuominen (1994) ermittelten entsprechende Gütwerte mittels Simulationen für verschiedene Modellannahmen. Dabei untersuchten sie auch das von Mandallaz and Mau (1981) vorgeschlagene restriktive Modell mit festen Effekten der Versuchseinheiten.

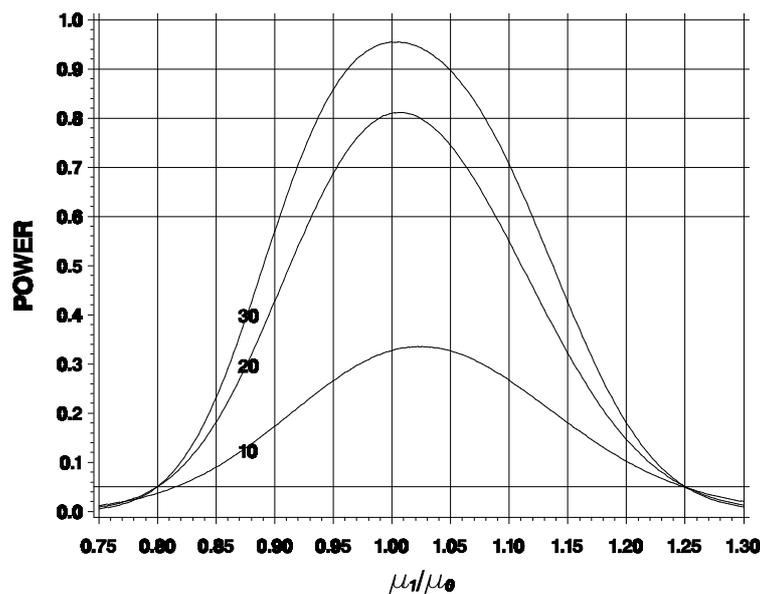


Abbildung 13: Crossover-Situation: Die Gütefunktion des Entscheidungsverfahrens, basierend auf den zwei einseitigen Tests t_{11}^q, t_{12}^q bei einem Sicherheitsbereich von $(q_1, q_2) = (0.8, 1.25)$ für $\alpha = 0.05, n_0 = n_1 = 10, 20, 30, CV_e = 32.5$ und $CV_b = 30.0\%$

Entsprechende Fallzahlen, um eine Güte von mindestens 0.80 zu erreichen, werden in Tabelle 18 angegeben. Die vereinfachte Berechnung der Fallzahlen erfolgt nach den folgenden Approximationsformeln. Dabei wurde folgender Sicherheitsbereich gewählt, d.h. $q_1 = \frac{1}{q_2}$:

$$\text{falls } \frac{m_1}{m_0} = 1 \quad n \geq \frac{1}{2} \frac{CV_e^2(1+q_2^2) + CV_b^2(1-q_2)^2}{(q_2 - 1)^2} \left(t_{a, 2n-2} + t_{\frac{b}{2}, 2n-2} \right)^2,$$

$$1 < \frac{m_1}{m_0} < q_2 \quad n \geq \frac{1}{2} \frac{CV_e^2(1+q_2^2) + CV_b^2(1-q_2)^2}{\left(q_2 - \frac{m_1}{m_0} \right)^2} \left(t_{a, 2n-2} + t_{b, 2n-2} \right)^2$$

$$\text{und } \frac{1}{q_2} < \frac{m_1}{m_0} < 1 \quad n \geq \frac{1}{2} \frac{CV_e^2 \left(1 + \frac{1}{q_2}\right) + CV_b^2 \left(1 - \frac{1}{q_2}\right)^2}{\left(\frac{1}{q_2} - \frac{m_1}{m_0}\right)^2} (t_{a,2n-2} + t_{b,2n-2})^2.$$

Die Herleitung dieser Formeln erfolgt analog zur Zweistichprobensituation.

Die der Tabelle 18 entsprechenden approximativen Fallzahlen sind in Tabelle 19 aufgelistet.

Tabelle 18: Crossover-Situation: Exakter Stichprobenumfang pro Sequenz für eine Güte von mindestens 0.80 bei $\alpha = 0.05$, $(q_1, q_2) = (0.8, 1.25)$, und verschiedene CV_e , CV_b

CV_e (%)	CV_b (%)	m_1/m_0							
		0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20
5.0	20.0	8	3	3	2	2	3	4	12
	40.0	14	5	3	3	3	4	6	21
	60.0	24	7	4	3	4	5	10	37
	80.0	38	10	5	4	5	8	16	59
	100.0	56	15	7	5	7	11	23	86
10.0	20.0	23	7	4	3	4	5	10	36
	40.0	29	8	5	4	4	6	12	45
	60.0	39	11	6	5	5	8	16	61
	80.0	53	14	7	6	6	10	21	82
	100.0	71	19	9	7	8	13	28	110
15.0	20.0	49	13	7	5	6	10	20	76
	40.0	55	15	7	6	7	11	22	85
	60.0	65	17	8	7	8	12	26	100
	80.0	78	21	10	8	9	15	31	122
	100.0	96	25	12	9	11	18	38	150
20.0	20.0	84	22	11	8	10	16	34	131
	40.0	90	23	11	9	10	17	36	140
	60.0	100	26	13	10	11	18	40	156
	80.0	114	29	14	11	13	21	45	177
	100.0	132	34	16	12	14	24	52	205
25.0	20.0	130	33	16	12	14	24	51	202
	40.0	136	35	17	13	15	25	54	212
	60.0	146	37	18	14	16	26	58	227
	80.0	160	41	19	15	17	29	63	249
	100.0	177	45	21	16	19	32	70	276
30.0	20.0	186	47	22	17	20	33	73	289
	40.0	192	49	23	18	20	34	76	299
	60.0	201	51	24	19	21	36	79	314
	80.0	215	55	26	20	23	38	85	336
	100.0	233	59	28	21	25	41	92	364

Tabelle 19: *Crossover*-Situation: Approximativer Stichprobenumfang pro Sequenz für eine Güte von mindestens 0.80 bei $\alpha = 0.05$, $(\mathbf{q}_1, \mathbf{q}_2) = (0.8, 1.25)$, und verschiedene CV_e , CV_b

CV_e (%)	CV_b (%)	$\mathbf{m}_1 / \mathbf{m}_2$							
		0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20
5.0	20.0	8	3	3	3	3	3	4	12
	40.0	14	5	3	3	3	4	6	22
	60.0	24	7	4	4	4	5	10	37
	80.0	38	11	5	5	5	8	16	59
	100.0	56	15	7	6	7	11	23	86
10.0	20.0	24	7	4	4	4	5	10	36
	40.0	29	8	5	4	4	6	12	45
	60.0	39	11	6	5	5	8	16	61
	80.0	53	14	7	6	6	10	22	82
	100.0	71	19	9	7	8	13	29	110
15.0	20.0	49	13	7	6	6	10	20	76
	40.0	55	15	7	6	7	11	22	85
	60.0	65	17	8	7	8	12	26	100
	80.0	79	21	10	8	9	15	31	122
	100.0	96	25	12	10	11	18	38	150
20.0	20.0	84	22	11	9	9	16	34	131
	40.0	90	24	11	9	10	17	36	140
	60.0	100	26	12	10	11	18	40	156
	80.0	114	29	14	11	12	21	45	177
	100.0	132	34	16	13	14	24	52	205
25.0	20.0	130	33	16	13	14	24	52	202
	40.0	136	35	16	13	14	25	54	212
	60.0	146	37	17	14	15	26	58	227
	80.0	160	41	19	15	17	29	63	249
	100.0	177	45	21	17	19	32	70	277
30.0	20.0	186	47	22	17	19	33	73	290
	40.0	192	49	22	18	20	34	76	299
	60.0	202	51	24	19	21	36	80	314
	80.0	215	55	25	20	22	38	85	336
	100.0	233	59	27	21	24	42	92	364

Mittels der Periodendifferenzen bzw. der um \mathbf{q} korrigierten Periodendifferenzen kann das *Crossover*-Design auf die Zweistichprobensituation reduziert werden. Daher konnten die im dritten Kapitel hergeleiteten biometrischen Methoden auf die *Crossover*-Situation übertragen werden. Bemerkenswert ist wiederum die gute Übereinstimmung der mittels Approximationsformeln ermittelten Fallzahlen mit den exakt ermittelten Werten.

Im Kapitel 6 wird die entsprechende Auswertung des *Crossover*-Designs an einem realen Datensatz veranschaulicht.

5 K-Stichproben *many-to-one*-Situation

5.1 *Simple-tree* Alternative

Ein für Sicherheitsuntersuchungen typisches Design ist die folgende Versuchsanordnung:

$$\{\text{Kontrolle, Behandlung}_1, \text{Behandlung}_2, \dots, \text{Behandlung}_k\},$$

wobei die k Behandlungen verschiedene Substanzen bzw. unterschiedliche Kombinationen von Substanzen sein können. Zur Validierung des Experimentes, d.h. zur Überprüfung der Sensitivität der verwendeten Versuchseinheiten gegenüber der Behandlung, wird in der Toxikologie zusätzlich eine Positivgruppe mit bekanntem toxischem, mutagenem, teratogenem oder kanzerogenem Potential mitgeführt. Die adäquate Auswertung für die Einbeziehung der Positivkontrolle in die statistische Analyse wurde von Hothorn (1995) und Hothorn *et al.* (1999) beschrieben.

Es bezeichnen X_{ij} die Zielgrößen, die unabhängig identisch verteilt sind nach den stetigen Verteilungsfunktionen $F_i(x) = F(x - \mathbf{m}_i)$, $i = 0, 1, \dots, k$ und $j = 1, \dots, n_i$, wobei \mathbf{m}_i die Lageparameter bezeichnen, der Index $i \geq 1$ die Behandlungsgruppen und der Index $i = 0$ die Kontrollgruppe. Die zur Sicherheitsuntersuchung relevanten Parameter sind entweder die insgesamt k Differenzen $\mathbf{m}_i - \mathbf{m}_0$ oder die Quotienten $\frac{\mathbf{m}_i}{\mathbf{m}_0}$, $i = 1, \dots, k$, der Lokationsparameter, je nachdem, ob der Schwellenwert absolut oder nur relativ zur Kontrolle angegeben werden kann.

5.1.1 Globaler Sicherheitsnachweis

5.1.1.1 Einseitiges Testproblem

Die einseitig formulierten Einzelhypothesen für die Differenzen und die Quotienten lauten:

$$H_{0i}^d: \mathbf{m}_i - \mathbf{m}_0 \geq \mathbf{d} \quad \text{versus} \quad H_{li}^d: \mathbf{m}_i - \mathbf{m}_0 < \mathbf{d}$$

bzw.

$$H_{0i}^q: \frac{\mathbf{m}_i}{\mathbf{m}_0} \geq \mathbf{q} \quad \text{versus} \quad H_{li}^q: \frac{\mathbf{m}_i}{\mathbf{m}_0} < \mathbf{q}.$$

In der Situation des globalen Sicherheitsnachweises muß die Unbedenklichkeit **aller** k Behandlungen nachgewiesen werden. Die einseitige Formulierung des Testproblems ist:

$$H_0^d : \bigcup_{i=1}^k H_{0i}^d \quad \text{versus} \quad H_1^d : \bigcap_{i=1}^k H_{1i}^d$$

bzw.

$$H_0^q : \bigcup_{i=1}^k H_{0i}^q \quad \text{versus} \quad H_1^q : \bigcap_{i=1}^k H_{1i}^q.$$

Unter der Normalverteilungsannahme $F_i(x) = F\left(\frac{x - \mathbf{m}_i}{\mathbf{s}}\right)$, $i = 0, \dots, k$, wird H_0^d unter Ausnutzung des *intersection-union*-Prinzips genau dann zum Niveau \mathbf{a} abgelehnt, falls **alle** Einzelhypothesen H_{0i}^d jeweils zum Niveau \mathbf{a} abgelehnt werden können:

$$t_i^d = \frac{\bar{X}_i - \bar{X}_0 - \mathbf{d}}{S \sqrt{\frac{1}{n_0} + \frac{1}{n_i}}} \leq -t_{\mathbf{a}, \mathbf{u}} \quad \forall i \in (1, \dots, k),$$

mit

$$\bar{X}_0 = \frac{1}{n_0} \sum_{j=1}^{n_0} X_{0j}, \quad \bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, \quad S^2 = \frac{1}{\mathbf{u}} \sum_{i=0}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \quad \text{und} \quad \mathbf{u} = \sum_{i=0}^k n_i - (k+1).$$

Äquivalent dazu ist die Inklusion **aller** einseitigen $100(1-\mathbf{a})\%$ -Konfidenzintervalle für die Differenzen $\mathbf{m}_i - \mathbf{m}_0$, $i = 1, \dots, k$, im Sicherheitsbereich:

$$(-\infty, \mathbf{d}_+] = \left[-\infty, \bar{X}_i - \bar{X}_0 + t_{\mathbf{a}, \mathbf{u}} S \sqrt{\frac{1}{n_0} + \frac{1}{n_i}} \right] \subset (-\infty, \mathbf{d}) \quad \forall i \in (1, \dots, k).$$

Das nichtparametrische Verfahren basiert auf der paarweise vorgenommenen Rangvergabe (*separate ranking*), d.h. $R_{ij}(\mathbf{d})$ bezeichnet den Rang von $X_{ij} - \mathbf{d}$ in der kombinierten Stichprobe vom Umfang $n_0 + n_i$:

$$X_{01}, \dots, X_{0n_0}, X_{i1} - \mathbf{d}, \dots, X_{in_i} - \mathbf{d}.$$

H_0^d wird abgelehnt, falls $R_i(\mathbf{d}) = \sum_{j=1}^{n_i} R_{ij}(\mathbf{d}) \leq n_i(n_0 + n_i + 1) - r_{\mathbf{a}, n_0, n_i}^d$ für alle $i \in (1, \dots, k)$,

dabei bezeichnet $r_{\mathbf{a}, n_0, n_i}^d$ das $(1-\mathbf{a})$ -Quantil der Wilcoxonstatistik $R_i(\mathbf{d})$. Die einseitigen $100(1-\mathbf{a})\%$ -Konfidenzintervalle für $\mathbf{m}_i - \mathbf{m}_0$ ergeben sich analog zur Zweistichprobensituation:

$$(-\infty, U_i^d], \text{ wobei } U_i^d = D_{u_i}^d = D_{n_0 n_i + 1 - C_a^i}^d, \quad C_a^i = \frac{n_i(2n_0 + n_i + 1)}{2} + 1 - r_{a, n_0, n_i}^d$$

und globale Sicherheit liegt zum Niveau \mathbf{a} vor, falls $(-\infty, U_i^d] \subset (-\infty, \mathbf{d}) \quad \forall i \in (1, \dots, k)$.

Die auf den Quotienten basierende Globalhypothese H_0^q wird verworfen, falls

$$t_i^q = \frac{\bar{X}_i - \mathbf{q}\bar{X}_0}{S \sqrt{\frac{\mathbf{q}^2}{n_0} + \frac{1}{n_i}}} \leq -t_{a,u} \quad \forall i \in (1, \dots, k),$$

was wiederum äquivalent ist zu

$$(-\infty, \mathbf{q}_+^i] \subset (-\infty, \mathbf{q}) \quad \forall i \in (1, \dots, k) \quad \text{und} \quad \bar{X}_0^2 > a_0,$$

mit

$$\mathbf{q}_+^i = \frac{\bar{X}_0 \bar{X}_i + \sqrt{a_0 \bar{X}_i^2 + a_i \bar{X}_0^2 - a_0 a_i}}{\bar{X}_0^2 - a_0}, \quad a_0 = \frac{S^2}{n_0} t_{a,u}^2 \quad \text{und} \quad a_i = \frac{S^2}{n_i} t_{a,u}^2.$$

Liegen **alle** oberen Grenzen \mathbf{q}_+^i der einseitigen $100(1-\mathbf{a})\%$ -Konfidenzintervalle für die Quotienten $\frac{\mathbf{m}_i}{\mathbf{m}_0}$, $i = 1, \dots, k$ unterhalb von \mathbf{q} , kann zum Niveau \mathbf{a} auf globale Sicherheit geschlossen werden.

5.1.1.2 Zweiseitiges Testproblem

Die k zweiseitig formulierten Einzelhypothesen

$$H_{0i}^d: \mathbf{m}_i - \mathbf{m}_0 \leq \mathbf{d}_1 \quad \text{oder} \quad \mathbf{m}_i - \mathbf{m}_0 \geq \mathbf{d}_2 \quad \text{versus} \quad H_{0i}^d: \mathbf{d}_1 < \mathbf{m}_i - \mathbf{m}_0 < \mathbf{d}_2$$

bzw.

$$H_{0i}^q: \frac{\mathbf{m}_i}{\mathbf{m}_0} \leq \mathbf{q}_1 \quad \text{oder} \quad \frac{\mathbf{m}_i}{\mathbf{m}_0} \geq \mathbf{q}_2 \quad \text{versus} \quad H_{0i}^q: \mathbf{q}_1 < \frac{\mathbf{m}_i}{\mathbf{m}_0} < \mathbf{q}_2$$

werden in jeweils zwei einseitige Testprobleme zerlegt:

$$H_{0i1}^d: \mathbf{m}_i - \mathbf{m}_0 \leq \mathbf{d}_1 \quad \text{versus} \quad H_{1i1}^d: \mathbf{m}_i - \mathbf{m}_0 > \mathbf{d}_1$$

und

$$H_{0i2}^d: \mathbf{m}_i - \mathbf{m}_0 \geq \mathbf{d}_2 \quad \text{versus} \quad H_{1i2}^d: \mathbf{m}_i - \mathbf{m}_0 < \mathbf{d}_2$$

bzw.

$$H_{0i1}^q : \frac{m_i}{m_0} \leq q_1 \quad \text{versus} \quad H_{1i1}^q : \frac{m_i}{m_0} > q_1$$

und

$$H_{0i2}^q : \frac{m_i}{m_0} \geq q_2 \quad \text{versus} \quad H_{1i2}^q : \frac{m_i}{m_0} < q_2.$$

Die Globalhypothesen lauten dann:

$$H_0^d : \bigcup_{i=1}^k \bigcup_{r=1}^2 H_{0ir}^d \quad \text{versus} \quad H_1^d : \bigcap_{i=1}^k \bigcap_{r=1}^2 H_{1ir}^d$$

bzw.

$$H_0^q : \bigcup_{i=1}^k \bigcup_{r=1}^2 H_{0ir}^q \quad \text{versus} \quad H_1^q : \bigcap_{i=1}^k \bigcap_{r=1}^2 H_{1ir}^q.$$

Bei Normalverteilungsannahme wird H_0^d verworfen, falls jeweils **alle** Einzelhypothesen mittels zweier einseitiger t -Tests abgelehnt werden, oder äquivalent dazu, falls **alle** korrespondierenden zweiseitigen $100(1-\alpha)\%$ -Konfidenzintervalle im Sicherheitsbereich enthalten sind:

$$t_{i1}^d = \frac{\bar{X}_i - \bar{X}_0 - \mathbf{d}_1}{S \sqrt{\frac{1}{n_0} + \frac{1}{n_i}}} \geq t_{\alpha,u} \quad \text{und} \quad t_{i2}^d = \frac{\bar{X}_i - \bar{X}_0 - \mathbf{d}_2}{S \sqrt{\frac{1}{n_0} + \frac{1}{n_i}}} \leq -t_{\alpha,u} \quad \forall i \in (1, \dots, k)$$

\Leftrightarrow

$$[\mathbf{d}_-, \mathbf{d}_+] = \left[\bar{X}_i - \bar{X}_0 \pm t_{\alpha,u} S \sqrt{\frac{1}{n_0} + \frac{1}{n_i}} \right] \subset (\mathbf{d}_1, \mathbf{d}_2) \quad \forall i \in (1, \dots, k).$$

Es sei darauf hingewiesen, daß Bofinger und Bofinger (1993) für den globalen Sicherheitsnachweis den maximalen Abstand $\mathbf{m}^* = \max_{i=1}^k |\mathbf{m}_i - \mathbf{m}_0|$ betrachteten. Dabei definierten die Autoren als Nullhypothese $H_0^d : \mathbf{m}^* \leq \mathbf{d}$, welche verworfen werden kann, falls die obere Grenze des $100(1-\alpha)\%$ -Konfidenzintervalles für \mathbf{m}^* unterhalb von \mathbf{d} liegt. Giani und Straßburger (1994) hingegen definierten die globale Sicherheit in der standardisierten Form $\frac{\mathbf{m}^*}{\mathbf{S}}$ und leiteten ein entsprechendes Testverfahren her. Der Nachteil beider Methoden besteht in der komplexen Bestimmung der kritischen Werte, welche nur mit

aufwendigen Algorithmen zu lösen sind. Unter Ausnutzung des *intersection-union*-Prinzips ergibt sich jedoch, wie oben dargestellt, ein wesentlich einfacheres Verfahren zum direkten Nachweis der globalen Sicherheit.

Für das verteilungsfreie Verfahren gilt, daß H_0^d abgelehnt wird, falls

$$R_i(\mathbf{d}) \geq r_{\mathbf{a}, n_0, n_i}^d \quad \text{und} \quad R_i(\mathbf{d}_2) \leq n_i(n_0 + n_i + 1) - r_{\mathbf{a}, n_0, n_i}^d \quad \forall i \in (1, \dots, k)$$

$$\Leftrightarrow$$

$$[L_i^d, U_i^d] \subset (\mathbf{d}_1, \mathbf{d}_2) \quad \forall i \in (1, \dots, k), \quad L_i^d = D_{l_i}^d = D_{C_a^i}^d \quad \text{und} \quad U_i^d = D_{u_i}^d = D_{n_0 n_i + 1 - C_a^i}^d.$$

Die Entscheidungsregel hinsichtlich der auf den Quotienten basierenden Nullhypothese H_0^q erfolgt ebenfalls mittels der entsprechenden Testverfahren bzw. mit der dazu äquivalenten Inklusionsregel:

$$t_{i1}^q = \frac{\bar{X}_i - \mathbf{q}_1 \bar{X}_0}{S \sqrt{\frac{\mathbf{q}_1^2}{n_0} + \frac{1}{n_i}}} \geq t_{\mathbf{a}, \mathbf{u}} \quad \text{und} \quad t_{i2}^q = \frac{\bar{X}_i - \mathbf{q}_2 \bar{X}_0}{S \sqrt{\frac{\mathbf{q}_2^2}{n_0} + \frac{1}{n_i}}} \leq -t_{\mathbf{a}, \mathbf{u}} \quad \forall i \in (1, \dots, k)$$

$$\Leftrightarrow$$

$$[\mathbf{q}_-^i, \mathbf{q}_+^i] \subset (\mathbf{q}_1, \mathbf{q}_2) \quad \forall i \in (1, \dots, k) \quad \text{und} \quad \bar{X}_0^2 > a_0,$$

mit

$$\mathbf{q}_\pm^i = \frac{\bar{X}_0 \bar{X}_i \pm \sqrt{a_0 \bar{X}_i^2 + a_i \bar{X}_0^2 - a_0 a_i}}{\bar{X}_0^2 - a_0}, \quad a_0 = \frac{S^2}{n_0} t_{\mathbf{a}, \mathbf{u}}^2 \quad \text{und} \quad a_i = \frac{S^2}{n_i} t_{\mathbf{a}, \mathbf{u}}^2.$$

5.1.2 Partieller Sicherheitsnachweis

5.1.2.1 Einseitiges Testproblem

Der Nachweis der globalen Sicherheit bedingt, daß für alle einzelnen Behandlungen der Nachweis der Sicherheit geführt werden muß. Diese Anforderung ist sicherlich ein sehr restriktives Kriterium und wird daher im folgenden bei der Untersuchung der partiellen Sicherheit abgeschwächt, welche im Nachweis **mindestens einer** sicheren Behandlung besteht. Die entsprechenden Einzelhypothesen lauten:

$$H_0^d : \bigcap_{i=1}^k H_{0i}^d \quad \text{versus} \quad H_1^d : \bigcup_{i=1}^k H_{1i}^d$$

bzw.

$$H_0^q : \bigcap_{i=1}^k H_{0i}^q \quad \text{versus} \quad H_1^q : \bigcup_{i=1}^k H_{1i}^q .$$

Diese Formulierung entspricht dem *union-intersection*-Problem (Roy, 1953). Im Gegensatz zum *intersection-union*-Problem, bei dem **alle** Einzelhypothesen abgelehnt werden müssen, wird nun die Globalhypothese verworfen, falls **mindestens eine** der k Einzelhypothesen zum entsprechend adjustierten Niveau abgelehnt werden kann.

Basierend auf der Normalverteilungsannahme $F_i(x) = F\left(\frac{x - \mathbf{m}_i}{\mathbf{s}}\right)$, $i = 0, 1, \dots, k$, kann mittels des Dunnett-Verfahrens (Dunnett und Tamhane, 1991) der Nachweis der partiellen Sicherheit durchgeführt werden. Wesentlich ist dabei, daß dieses Verfahren den multiplen Fehler 1. Art kontrolliert, d.h. die Wahrscheinlichkeit, mindestens eine wahre Einzelhypothese irrtümlicherweise abzulehnen. Die i -te Einzelhypothese H_{0i}^d für die Differenz der Erwartungswerte oder H_{0i}^q für den Quotienten der Erwartungswerte wird verworfen, falls

$$t_i^d = \frac{\bar{X}_i - \bar{X}_0 - \mathbf{d}}{S \sqrt{\frac{1}{n_0} + \frac{1}{n_i}}} \leq -t_{\mathbf{a}, \mathbf{u}, R_k^d}$$

bzw.

$$t_i^q = \frac{\bar{X}_i - \mathbf{q}\bar{X}_0}{S \sqrt{\frac{\mathbf{q}^2}{n_0} + \frac{1}{n_i}}} \leq -t_{\mathbf{a}, \mathbf{u}, R_k^q} ,$$

wobei die kritischen Werte $t_{\mathbf{a}, \mathbf{u}, R_k^d}$ und $t_{\mathbf{a}, \mathbf{u}, R_k^q}$ folgendermaßen hergeleitet werden.

Lemma 6:

Unter der jeweiligen Globalhypothese H_0^d bzw. H_0^q sind die Teststatistiken $(t_1^d, t_2^d, \dots, t_k^d)'$ und $(t_1^q, t_2^q, \dots, t_k^q)'$ k -dimensional zentral t -verteilt mit \mathbf{u} Freiheitsgraden und Korrelationsmatrix $R_k^d = (\mathbf{r}_{ij}^d)_{i,j=1,\dots,k}$ bzw. $R_k^q = (\mathbf{r}_{ij}^q)_{i,j=1,\dots,k}$, wobei

i) $\mathbf{r}_{ij}^d = \mathbf{I}_i^d \mathbf{I}_j^d$, $1 \leq i \neq j \leq k$, wobei

$$\mathbf{I}_i^d = \frac{1}{\sqrt{1 + \frac{n_0}{n_i}}}, \quad i = 1, \dots, k.$$

ii) $\mathbf{r}_{ij}^q = \mathbf{I}_i^q \mathbf{I}_j^q$, $1 \leq i \neq j \leq k$, wobei

$$\mathbf{I}_i^q = \frac{1}{\sqrt{\mathbf{q}^2 + \frac{n_0}{n_i}}}, \quad i = 1, \dots, k.$$

Beweis:

i)

$$\mathbf{r}_{ij}^d = \frac{\text{Kov}(\bar{X}_i - \bar{X}_0 - \mathbf{d}, \bar{X}_j - \bar{X}_0 - \mathbf{d})}{\sqrt{\text{Var}(\bar{X}_i - \bar{X}_0 - \mathbf{d}) \text{Var}(\bar{X}_j - \bar{X}_0 - \mathbf{d})}} = \frac{\frac{\mathbf{s}^2}{n_0}}{\mathbf{s}^2 \sqrt{\left(\frac{1}{n_0} + \frac{1}{n_i}\right) \left(\frac{1}{n_0} + \frac{1}{n_j}\right)}} =$$

$$\frac{1}{\sqrt{\left(1 + \frac{n_0}{n_i}\right)}} \cdot \frac{1}{\sqrt{\left(1 + \frac{n_0}{n_j}\right)}} = \mathbf{I}_i^d \mathbf{I}_j^d.$$

ii)

$$\mathbf{r}_{ij}^q = \frac{\text{Kov}(\bar{X}_i - \mathbf{q}\bar{X}_0, \bar{X}_j - \mathbf{q}\bar{X}_0)}{\sqrt{\text{Var}(\bar{X}_i - \mathbf{q}\bar{X}_0) \text{Var}(\bar{X}_j - \mathbf{q}\bar{X}_0)}} = \frac{\frac{\mathbf{s}^2}{n_0} \mathbf{q}^2}{\mathbf{s}^2 \sqrt{\left(\frac{\mathbf{q}^2}{n_0} + \frac{1}{n_i}\right) \left(\frac{\mathbf{q}^2}{n_0} + \frac{1}{n_j}\right)}} =$$

$$\frac{\mathbf{q}}{\sqrt{\left(\mathbf{q}^2 + \frac{n_0}{n_i}\right)}} \cdot \frac{\mathbf{q}}{\sqrt{\left(\mathbf{q}^2 + \frac{n_0}{n_j}\right)}} = \mathbf{I}_i^q \mathbf{I}_j^q.$$

▷

$t_{\mathbf{a}, n, R_k^d}$ bezeichnet das $(1-\mathbf{a})$ -Quantil der Statistik $\max_{i=1}^k t_i^d$ und $t_{\mathbf{a}, n, R_k^q}$ das $(1-\mathbf{a})$ -Quantil der Statistik $\max_{i=1}^k t_i^q$. Diese Quantile können wegen der Produkteigenschaft der Korrelations-

koeffizienten mittels der SAS-Funktion PROBMC berechnet werden (Hauschke *et al.*, 1999d).

Die äquivalente Beurteilung ergibt sich unter Verwendung der simultanen $100(1-\mathbf{a})\%$ -Konfidenzintervalle. Die i -te Einzelhypothese H_{0i}^d oder H_{0i}^q , $i = 1, \dots, k$, wird abgelehnt, falls

$$(-\infty, \mathbf{d}_+^i] = \left(-\infty, \bar{X}_i - \bar{X}_0 + t_{\mathbf{a}, \mathbf{u}, R_k^d} S \sqrt{\frac{1}{n_0} + \frac{1}{n_i}} \right) \subset (-\infty, \mathbf{d})$$

bzw.

$$(-\infty, \mathbf{q}_+^i] \subset (-\infty, \mathbf{q}) \quad \text{und} \quad \bar{X}_0^2 > a_0,$$

mit

$$\mathbf{q}_+^i = \frac{\bar{X}_0 \bar{X}_i + \sqrt{a_0 \bar{X}_i^2 + a_i \bar{X}_0^2 - a_0 a_i}}{\bar{X}_0^2 - a_0}, \quad a_0 = \frac{S^2}{n_0} t_{\mathbf{a}, \mathbf{u}, R_k^q}^2 \quad \text{und} \quad a_i = \frac{S^2}{n_i} t_{\mathbf{a}, \mathbf{u}, R_k^q}^2.$$

Es sei darauf hingewiesen, daß die obigen Konfidenzintervalle $(-\infty, \mathbf{q}_+^i]$ für die jeweiligen

Quotienten $\frac{\mathbf{m}_i}{\mathbf{m}_0}$, $i = 1, \dots, k$, nicht mehr für eine *a-posteriori* vorgenommene Sicherheits-

beurteilung verwendet werden können. Dies liegt an der Tatsache, daß zur Konstruktion dieser Konfidenzintervalle der Wert \mathbf{q} zur Berechnung der Korrelation und somit zur Bestimmung der kritischen Werte $t_{\mathbf{a}, \mathbf{u}, R_k^q}$ benötigt wird. Auf Kosten eines Güteverlustes ist dies jedoch unter Ausnutzung der \mathbf{a} -Korrektur nach Bonferroni leicht zu beheben, d.h. man ersetzt die kritische Grenze $t_{\mathbf{a}, \mathbf{u}, R_k^q}$ durch den Wert der univariaten t -Verteilung $t_{\frac{\mathbf{a}}{k}, \mathbf{u}}$.

Zur Veranschaulichung diene die von Bedotto *et al.* (1989) veröffentlichte Studie *Cardiac hypertrophy induced by thyroid hormone is independent of loading conditions and beta-adrenoceptor blockade*. In diesem pharmakologischen Experiment an Ratten untersuchten die Autoren in einem Telexperiment, ob die durch die Applikation des Schilddrüsenhormons T_4 induzierte links-ventrikuläre Hypertrophie durch gleichzeitige Behandlung mit Captopril, Hydralazin, Propranolol bzw. einer Kombination von Captopril und Propranolol gehemmt werden kann. Die getroffenen Schlußfolgerungen basierten auf dem indirekten Ansatz des fehlenden Unterschiedes zwischen jeweiliger Behandlung und der Kontrolle (T_4). Postuliert man jedoch die Hypothese, daß bei gleichzeitiger Verabreichung dieser Substanzen die durch T_4 induzierte Hypertrophie nur irrelevant erhöht wird, so folgt der direkte Sicherheitsnachweis als probate Methode (Hauschke und Hothorn, 1998). In der folgenden

Tabelle 20 sind die empirischen Daten und die oberen Grenzen q_+^i der simultanen 95%-Konfidenzintervalle für die Quotienten $\frac{m_i}{m_0}$, $i = 1, \dots, 4$, dargestellt. Dabei wurden die simultanen 95%-Konfidenzintervalle bei *a-priori* Vorgabe von $q = 1.07$ (fiktive Vorgabe) und zusätzlich unter Verwendung der Bonferroni-Korrektur ermittelt; dazu wurde die für das Gesamtexperiment vorliegende Varianzschätzung $S = 0.1538$ mit 93 Freiheitsgraden benutzt (Dunnnett und Tamhane, 1991). Mit *a-priori* vorgegebenem $q = 1.07$ folgt eine irrelevante Erhöhung bei gleichzeitiger Gabe von Captopril bzw. Hydralazin. Verwendet man die konservativere Bonferroni-Adjustierung, so kann die Unbedenklichkeit nur bei gleichzeitiger Gabe von Captopril gefolgert werden.

Tabelle 20: Mittelwerte, Fallzahlen für die Variable Linksherz-Gewicht normiert auf das Körpergewicht (mg/g) und obere simultane 95%-Konfidenzgrenzen (bei *a-priori* Vorgabe von $q=1.07$ und nach Bonferroni-Adjustierung) für den Vergleich mit der Kontrolle T₄

Behandlungsgruppe	Mittelwert	Fallzahl	Obere Konfidenzgrenze für $q=1.07$	Obere Konfidenzgrenze Bonferroni-adjustiert
T ₄	2.52	10	-	-
T ₄ + Captopril	2.49	9	1.051	1.054
T ₄ + Propranolol	2.60	12	1.092	1.094
T ₄ + Hydralazin	2.54	10	1.069	1.072
T ₄ + Captopril + Propranolol	2.56	10	1.078	1.081

Basierend auf der Abschlußtestprozedur (Marcus *et al.*, 1976), haben Dunnnett und Tamhane (1991) ein *step-down* Verfahren hergeleitet, welches im Vergleich zur obigen *single-step* Methode eine größere Güte aufweist. Dabei werden die Teststatistiken der Größe nach geordnet:

$$t_{(1)}^d \leq t_{(2)}^d \leq \dots \leq t_{(k)}^d \quad \text{bzw.} \quad t_{(1)}^q \leq t_{(2)}^q \leq \dots \leq t_{(k)}^q.$$

Die jeweils korrespondierenden Hypothesen seien mit $H_{0(1)}^d, H_{0(2)}^d, \dots, H_{0(k)}^d$ bzw. $H_{0(1)}^q, H_{0(2)}^q, \dots, H_{0(k)}^q$ bezeichnet. Die Hypothesen $H_{0(i)}^d$ bzw. $H_{0(i)}^q$, $i = 1, \dots, k$, werden abgelehnt, falls

$$t_{(i)}^d \leq -t_{a,u,R_{k-i+1}}^d \quad \text{und alle Hypothesen } H_{0(1)}^d, H_{0(2)}^d, \dots, H_{0(i-1)}^d \text{ verworfen wurden}$$

bzw.

$t_{(i)}^q \leq -t_{\mathbf{a}, \mathbf{u}, R_{k-i+1}^q}$ und alle Hypothesen $H_{0(1)}^q, H_{0(2)}^q, \dots, H_{0(i-1)}^q$ verworfen wurden.

Das *step-down* Verfahren wird für den Beispieldatensatz von Bedotto *et al.* (1989) verwendet. In der Tabelle 21 sind die geordneten Teststatistiken $t_{(i)}^q$ und die kritischen Werte $-t_{0.05, 93, R_{4-i+1}}^q, i = 1, \dots, 4$, dargestellt. Mit $\mathbf{q} = 1.07$ folgt neben der irrelevanten Erhöhung bei gleichzeitiger Gabe von Captopril ebenfalls eine nicht relevante Erhöhung für die gleichzeitige Verabreichung von Hydralazin.

Tabelle 21: Geordnete Teststatistiken und kritische Werte für die Vergleiche mit der Kontrolle T₄

Behandlungsgruppe	$t_{(i)}^q$	$-t_{0.05, 93, R_{4-i+1}}^q$
T ₄ + Captopril	-2.825	-2.178
T ₄ + Hydralazin	-2.196	-2.077
T ₄ + Captopril + Propranolol	-1.915	-1.930
T ₄ + Propranolol	-1.409	-1.661

Eine analoge *step-up* Prozedur wurde von Dunnett und Tamhane (1992, 1995) vorgeschlagen, wobei jedoch der Nachweis der Kontrolle des multiplen Fehlers 1. Art für ungleiche Korrelationen bislang noch nicht erbracht wurde. Wesentlich einfacher gestaltet sich die Verwendung der \mathbf{a} -Korrektur nach Bonferroni; dabei werden die kritischen Werte $t_{\mathbf{a}, \mathbf{u}, R_k^d}$ und $t_{\mathbf{a}, \mathbf{u}, R_k^q}$ der k -dimensionalen t -Verteilung durch den Wert der univariaten t -Verteilung $t_{\frac{\mathbf{a}}{k}, \mathbf{u}}$ ersetzt. Insbesondere für großes k ist die Vorgehensweise konservativ, daher wurde diese durch die *step-down* Methode nach Holm (1979) verbessert. Dazu werden nun die p -Werte der Teststatistiken der Größe nach geordnet:

$$p_{(1)}^d \leq p_{(2)}^d \leq \dots \leq p_{(k)}^d \quad \text{bzw.} \quad p_{(1)}^q \leq p_{(2)}^q \leq \dots \leq p_{(k)}^q,$$

die korrespondierenden Hypothesen $H_{0(i)}^d$ bzw. $H_{0(i)}^q, i = 1, \dots, k$, werden abgelehnt, falls

$$p_{(i)}^d \leq \frac{\mathbf{a}}{k-i+1} \quad \text{und alle Hypothesen } H_{0(1)}^d, H_{0(2)}^d, \dots, H_{0(i-1)}^d \text{ verworfen wurden}$$

bzw.

$p_{(i)}^q \leq \frac{\mathbf{a}}{k-i+1}$ und alle Hypothesen $H_{0(1)}^q, H_{0(2)}^q, \dots, H_{0(i-1)}^q$ verworfen wurden.

Es sei darauf hingewiesen, daß in Spezialfällen, zum Beispiel bei Unabhängigkeit der Teststatistiken, die Bonferroni-Prozedur und die Holm-Modifikation weiter verbessert werden können (Simes, 1986; Hochberg, 1988; Hommel, 1988; Rom, 1990; Samuel-Cahn, 1996).

Eine verteilungsfreie Prozedur zum multiplen Niveau \mathbf{a} hat Steel (1959) für den *many-to-one* Vergleich vorgeschlagen: Für identische Stichprobenumfänge der Behandlungsgruppen $n_1 = n_2 = \dots = n_k = m$ und m nicht notwendigerweise identisch mit n_0 , wird die Hypothese

H_{0i}^d abgelehnt, falls $R_i(\mathbf{d}) = \sum_{j=1}^{n_i} R_{ij}(\mathbf{d}) \leq m(n_0 + m + 1) - r_{\mathbf{a},k,n_0,m}^d$, dabei bezeichnet $r_{\mathbf{a},k,n_0,m}^d$ das

$(1-\mathbf{a})$ -Quantil der Statistik $\max_{i=1}^k R_k(\mathbf{d})$. Exakte kritische Werte hat Steel für $k = 2, 3$ und $n_0 = m = 3, 4, 5$ tabelliert; einfache Approximationsformeln wurden von Hochberg und Tamhane (1987) und Hsu (1996) angegeben. Die einseitigen $100(1-\mathbf{a})\%$ -Konfidenzintervalle für $\mathbf{m}_i - \mathbf{m}_0$ sind

$$(-\infty, U_i^d], \text{ wobei } U_i^d = D_{u_i}^d = D_{n_0 m + 1 - C_a^i}^d, \quad C_a^i = \frac{m(2n_0 + m + 1)}{2} + 1 - r_{\mathbf{a},k,n_0,m}^d.$$

5.1.2.2 Zweiseitiges Testproblem

Analog zum globalen Sicherheitsnachweis lautet die zweiseitige Hypothese für den partiellen Nachweis der Unbedenklichkeit:

$$H_0^d: \bigcap_{i=1}^k \bigcup_{r=1}^2 H_{0ir}^d \quad \text{versus} \quad H_1^d: \bigcup_{i=1}^k \bigcap_{r=1}^2 H_{1ir}^d$$

bzw.

$$H_0^q: \bigcap_{i=1}^k \bigcup_{r=1}^2 H_{0ir}^q \quad \text{versus} \quad H_1^q: \bigcup_{i=1}^k \bigcap_{r=1}^2 H_{1ir}^q.$$

Bei Normalverteilungsannahme kann H_0^d unter Verwendung der Bonferroni-Korrektur überprüft werden, wobei die k Einzelhypothesen mittels der zwei einseitigen t -Tests getestet werden oder äquivalent dazu, mittels der Inklusionsregel:

$$t_{i1}^d = \frac{\bar{X}_i - \bar{X}_0 - \mathbf{d}_1}{S \sqrt{\frac{1}{n_0} + \frac{1}{n_i}}} \geq t_{\frac{\mathbf{a}}{k}, u} \quad \text{und} \quad t_{i2}^d = \frac{\bar{X}_i - \bar{X}_0 - \mathbf{d}_2}{S \sqrt{\frac{1}{n_0} + \frac{1}{n_i}}} \leq -t_{\frac{\mathbf{a}}{k}, u}$$

\Leftrightarrow

$$[\mathbf{d}_-, \mathbf{d}_+] = \left[\bar{X}_i - \bar{X}_0 \pm t_{\frac{\mathbf{a}}{k}, u} S \sqrt{\frac{1}{n_0} + \frac{1}{n_i}} \right] \subset (\mathbf{d}_1, \mathbf{d}_2).$$

Die Bonferroni-Adjustierung kann wiederum durch die *step-down* Prozedur nach Holm verbessert werden: es seien $p_i^d = \max(p_{i1}^d, p_{i2}^d)$ und $p_i^q = \max(p_{i1}^q, p_{i2}^q)$, wobei p_{ir}^d, p_{ir}^q die p -Werte der Teststatistiken t_{ir}^d und t_{ir}^q bezeichnen, $r = 1, 2$. Die p -Werte werden der Größe nach geordnet:

$$p_{(1)}^d \leq p_{(2)}^d \leq \dots \leq p_{(k)}^d \quad \text{bzw.} \quad p_{(1)}^q \leq p_{(2)}^q \leq \dots \leq p_{(k)}^q,$$

und die korrespondierenden Hypothesen $H_{0(i)}^d$ bzw. $H_{0(i)}^q, i = 1, \dots, k$, werden abgelehnt, falls

$$p_{(i)}^d \leq \frac{\mathbf{a}}{k-i+1} \quad \text{und alle Hypothesen } H_{0(1)}^d, H_{0(2)}^d, \dots, H_{0(i-1)}^d \text{ verworfen wurden}$$

bzw.

$$p_{(i)}^q \leq \frac{\mathbf{a}}{k-i+1} \quad \text{und alle Hypothesen } H_{0(1)}^q, H_{0(2)}^q, \dots, H_{0(i-1)}^q \text{ verworfen wurden.}$$

Unter der Annahme identischer Stichprobenumfänge der Behandlungsgruppen, d.h. $n_1 = n_2 = \dots = n_k = m$, haben Bofinger und Bofinger (1995) die folgenden simultanen $100(1-\mathbf{a})\%$ -Konfidenzintervalle für $\mathbf{m}_i - \mathbf{m}_0$ hergeleitet:

$$\left(\left(\bar{X}_i - \bar{X}_0 - c_k S \sqrt{\frac{1}{n_0} + \frac{1}{m}} \right)^-, \left(\bar{X}_i - \bar{X}_0 + c_k S \sqrt{\frac{1}{n_0} + \frac{1}{m}} \right)^+ \right),$$

wobei $x^- = \min(0, x)$, $x^+ = \max(0, x)$ und c_k spezielle, aus der multivariaten t -Verteilung abgeleitete kritische Werte darstellen. Die Einzelhypothesen lassen sich dann mit der Inklusionsregel testen:

$$\left[\left(\bar{X}_i - \bar{X}_0 - c_k S \sqrt{\frac{1}{n_0} + \frac{1}{m}} \right)^-, \left(\bar{X}_i - \bar{X}_0 + c_k S \sqrt{\frac{1}{n_0} + \frac{1}{m}} \right)^+ \right] \subset (\mathbf{d}, \mathbf{d}_2).$$

Für das analoge verteilungsfreie Verfahren gilt, daß H_0^d abgelehnt wird, falls

$$R_i(\mathbf{d}_1) \geq r_{\frac{a}{k}, n_0, n_i}^d \quad \text{und} \quad R_i(\mathbf{d}_2) \leq n_i(n_0 + n_i + 1) - r_{\frac{a}{k}, n_0, n_i}^d$$

$$\Leftrightarrow$$

$$[L_i^d, U_i^d] \subset (\mathbf{d}, \mathbf{d}_2), \quad L_i^d = D_{l_i}^d = D_{C_{\frac{a}{k}}^i}^d \quad \text{und} \quad U_i^d = D_{u_i}^d = D_{n_0 n_i + 1 - C_{\frac{a}{k}}^i}^d.$$

Die Überprüfung von H_0^q und der Einzelhypothesen erfolgt analog:

$$t_{i1}^q = \frac{\bar{X}_i - \mathbf{q}_1 \bar{X}_0}{S \sqrt{\frac{\mathbf{q}_1^2}{n_0} + \frac{1}{n_i}}} \geq t_{\frac{a}{k}, u} \quad \text{und} \quad t_{i2}^q = \frac{\bar{X}_i - \mathbf{q}_2 \bar{X}_0}{S \sqrt{\frac{\mathbf{q}_2^2}{n_0} + \frac{1}{n_i}}} \leq -t_{\frac{a}{k}, u}$$

$$\Leftrightarrow$$

$$[\mathbf{q}_-, \mathbf{q}_+] \subset (\mathbf{q}_1, \mathbf{q}_2) \quad \text{und} \quad \bar{X}_0^2 > a_0,$$

mit

$$\mathbf{q}_{\pm}^i = \frac{\bar{X}_0 \bar{X}_i \pm \sqrt{a_0 \bar{X}_i^2 + a_i \bar{X}_0^2 - a_0 a_i}}{\bar{X}_0^2 - a_0}, \quad a_0 = \frac{S^2}{n_0} t_{\frac{a}{k}, u}^2 \quad \text{und} \quad a_i = \frac{S^2}{n_i} t_{\frac{a}{k}, u}^2.$$

5.2 Simple-order Alternative

Statt der verschiedenen Substanzen oder entsprechender Kombinationen liegen nun k aufsteigende Dosierungen vor:

$$\{\text{Kontrolle, Dosis}_1, \text{Dosis}_2, \dots, \text{Dosis}_k\}.$$

Die Zielgrößen X_{ij} sind unabhängig identisch verteilt nach den stetigen Verteilungsfunktionen $F_i(x) = F(x - \mathbf{m}_i)$, $i = 0, 1, \dots, k$ und $j = 1, \dots, n_i$; zusätzlich wird für die Lokationsparameter meist eine monotone Ordnung postuliert: $\mathbf{m}_0 \leq \mathbf{m}_1 \leq \dots \leq \mathbf{m}_k$. Das Ziel besteht in der Bestimmung der maximalen Dosierung, die im Vergleich zur Kontrolle nur irrelevant erhöht ist.

5.2.1 Bestimmung der maximal sicheren Dosis

Kann der Schwellenwert d absolut angegeben werden, so ist maximal sicherere Dosis wie folgt definiert:

$$MAXSD^d = D_{i^*} \quad \text{wobei} \quad i^* = \max(i : m_i - m_0 < d)$$

und zusätzlich

$$\forall j, j = 1, \dots, i^* - 1 : m_j - m_0 < d.$$

Unter der Monotonieannahme $m_0 \leq m_1 \leq \dots \leq m_k$ ist die zusätzliche Bedingung automatisch erfüllt, daß alle Dosierungen, die kleiner als die maximal sichere Dosis sind, auch sicher sind (Hothorn und Hauschke, 1999). Ausgehend von dem Konzept der *a-priori* geordneten Hypothesen (Hothorn und Lehmacher, 1991; Maurer *et al.*, 1995; Hsu und Berger, 1998) kann eine Prozedur abgeleitet werden, die einerseits den multiplen Fehler 1. Art auf α begrenzt und andererseits die Annahme der monotonen Ordnung nicht benötigt (Neuhäuser und Hothorn, 1995, 1997; Hauschke, 1995; Hauschke und Hothorn, 1998). Beginnend mit dem Testproblem:

$$H_{01}^d : m_1 - m_0 \geq d$$

versus

$$H_{11}^d : m_1 - m_0 < d$$

wird ein Test zum Niveau α durchgeführt. Falls H_{01}^d abgelehnt werden kann, wird

$$H_{02}^d : m_2 - m_0 \geq d$$

versus

$$H_{12}^d : m_2 - m_0 < d$$

getestet. Wird H_{01}^d nicht abgelehnt, so ist die Aussage zu treffen, daß im vorliegenden Design die $MAXSD^d$ nicht bestimmt werden kann; inhaltlich hat dies zur Folge, daß ein neues Experiment mit kleiner gewählten Dosierungen erfolgen muß. Allgemein erfolgt der Übergang von

$$H_{0i}^d : m_i - m_0 \geq d$$

versus

$$H_{1i}^d : m_i - m_0 < d$$

zu

$$H_{0i+1}^d : m_{i+1} - m_0 \geq d$$

versus

$$H_{1i+1}^d : m_{i+1} - m_0 < d$$

genau dann, falls H_{0i}^d zum Niveau α verworfen werden konnte. Als Schätzer für die $MAXSD^d$ dient die maximale Dosis, die im Vergleich zur Kontrolle keine relevante Erhöhung aufweist. Es sei darauf hingewiesen, daß bei der Bestimmung der $MAXSD^d$ die Auswahl der Dosierungen von eminenter Bedeutung ist. Ohne Angabe des funktionalen Zusammenhanges zwischen Dosis und Effekt kann - falls überhaupt möglich - nur eine der gewählten Dosierungen als $MAXSD^d$ abgeleitet werden.

Kann der Schwellenwert d nur relativ zum Erwartungswert der Kontrolle angegeben werden, so lautet die entsprechende Definition der maximal sichereren Dosis:

$$MAXSD^q = D_{i^*} \quad \text{wobei} \quad i^* = \max(i : \frac{m_i}{m_0} < q)$$

und zusätzlich

$$\forall j, j = 1, \dots, i^* - 1 : \frac{m_j}{m_0} < d.$$

Die Bestimmung der maximal sicheren Dosis $MAXSD^q$ erfolgt analog zur Ermittlung von $MAXSD^d$ (Tamhane *et al.*, 1999) und wird im folgenden an einem Beispiel mit drei Dosierungen mit der normalverteilten Zielvariablen Serum-ASAT-Werte erläutert. Hierbei wird die Varianzschätzung $S = 0.5694$ mit 59 Freiheitsgraden benutzt und als irrelevante Erhöhung sei $q = 1.25$ gewählt. Die Teststatistiken lauten:

$$t_i^q = \frac{\bar{X}_i - 1.25\bar{X}_0}{S \sqrt{\frac{1.25^2}{n_0} + \frac{1}{n_i}}}, \quad i = 1, 2, 3.$$

In der Tabelle 22 sind die Ergebnisse für den Vergleich der Dosisgruppen mit der Kontrollgruppe dargestellt. Bei einem Niveau von $\alpha = 0.05$ bricht das Verfahren der *a-priori* geordneten Hypothesen beim Vergleich der Dosisgruppe 200mg/kg, da der zugehörige *p*-Wert größer ist als 0.05. Als maximal sichere Dosis folgt somit die Dosis 20mg/kg.

Tabelle 22: Mittelwerte, Fallzahlen für die Zielvariable ASAT-Werte (mmol/L), Werte der Teststatistiken und entsprechende p -Werte

Dosisgruppe (mg/kg)	Mittelwert	Fallzahl	t_i^q	p -Wert
0	2.087	19	-	-
20	2.064	16	-2.5147	0.0073
200	2.323	15	-1.3005	0.0992
1000	2.753	13	0.6350	-

5.2.2 Alternative Konzepte

Eine übliche Auswertungsstrategie basiert auf dem Abschlußtestprinzip (Marcus *et al.*, 1976) und besteht darin, daß im ersten Schritt ein Trendtest (Jonckheere, 1954; Bartholomew, 1961) mit allen Behandlungsgruppen durchgeführt wird:

$$H_0 : \mathbf{m}_0 = \mathbf{m}_1 = \dots = \mathbf{m}_k$$

versus

$$H_1 : \mathbf{m}_0 \leq \mathbf{m}_1 \leq \dots \leq \mathbf{m}_k, \quad \mathbf{m}_0 < \mathbf{m}_k.$$

Ist der p -Wert kleiner oder gleich \mathbf{a} , so wird der Test ohne die höchste Dosis D_k wiederholt (Hothorn *et al.*, 1997). Im Falle einer erneuten Signifikanz zum Niveau \mathbf{a} erfolgt der entsprechende Vergleich ohne die Dosierungen D_k und D_{k-1} . Diese Verfahrensweise des Ausschlusses der jeweils höchsten Dosis wird solange fortgesetzt, bis erstmalig ein nichtsignifikantes Ergebnis ($p > \mathbf{a}$) vorliegt. Die in diesem letzten Vergleich vorliegende höchste Dosis wird dann als maximal sichere Dosis *MAXSD* oder *NOED* (*no observed effect dose*) oder *NOEL* (*no observed effect level*) bezeichnet (Williams, 1971, 1972; Shirley, 1977; Tukey *et al.*, 1985; Rom *et al.*, 1994). Bei Verletzung der Monotonieannahme, zum Beispiel bei Vorliegen einer Umbrella-Alternativen sind entsprechend modifizierte Verfahren zu benutzen (Mack und Wolfe, 1981; Hettmansperger und Norton, 1987).

Brown und Erdreich (1989) und Yanagawa *et al.* (1994) wiesen darauf hin, daß bei dieser klassischen Bestimmung der *MAXSD* die Stichprobengröße in nicht akzeptabler Weise eingeht: je geringer die Fallzahl, desto größer ist die maximal sichere Dosis. Daher wurde in der letzten Zeit (Barnes *et al.*, 1995; Gaylor, 1996; Gaylor und Chen, 1996) das von Crump (1984) entwickelte Konzept der *benchmark dose* (*BMD*) eingehend diskutiert. Zur Berechnung dieser Dosis benötigt man jedoch eine Annahme über den funktionalen Zusammenhang zwischen Dosis und Effekt und kann dann mittels Extrapolation eine Dosis

schätzen, die im Vergleich zur Kontrolle eine nur irrelevante Erhöhung aufweist. Um der vorliegenden Variabilität Rechnung zu tragen, wird auch die untere Grenze des entsprechenden Konfidenzintervalles als Schätzer verwendet.

Bewertend zu dem Konzept der *BMD* ist zu sehen, daß sich die Angabe der funktionalen Beziehung zwischen Effekt und Dosis in der Praxis als äußerst schwierig erweist, zumal eine Verifizierung der getroffenen Modellannahmen nicht möglich ist. Hinzu kommt wegen der meist geringen Anzahlen von eingesetzten Versuchseinheiten, daß der Schätzer der *BMD* oftmals sehr nahe bei Null liegt (Murrell *et al.*, 1998).

6 Konzept der Bioäquivalenz

Im Rahmen humanpharmakologischer Bioäquivalenzstudien wird die Ähnlichkeit der Wirkstoffkonzentrationen verschiedener Formulierungen eines Wirkstoffs im Plasma anhand vorgegebener Grenzen untersucht. Wegen der wesentlichen Annahme, daß der Bioäquivalenznachweis die Gleichwertigkeit bezüglich Wirksamkeit und Sicherheit impliziert (Metzler, 1974; Röhmel, 1998), sind derartige Studien insbesondere bei der Entwicklung von Generika von großer Bedeutung. Unter Bezugnahme auf das wirkstoffidentische Referenzprodukt kann dann eine Arzneimittelzulassung ohne weitere aufwendige klinische Prüfungen erfolgen, so daß zusätzliche Entwicklungskosten und -zeiten für den Nachahmerproduzenten entfallen. Bioäquivalenzstudien sind ferner von Bedeutung bei der Umstellung der Entwicklungsform, mit der die klinischen Versuche gemacht wurden, auf die definitive Handelsform bzw. bei Änderung der Hilfsstoffzusammensetzung. Die Anwendung und aktuelle Erweiterungen werden am Beispiel einer Dosisäquivalenzstudie illustriert.

6.1 Modell, Testproblem und Entscheidungsverfahren

Die Wirksamkeit eines Arzneimittels ist in der Regel an eine bestimmte minimale Plasmakonzentration gebunden, damit der erwünschte Effekt eintritt. Zu hohe Plasmaspiegel sind dagegen oft mit unerwünschten Nebenwirkungen verbunden. Deshalb besteht die Notwendigkeit, zum Nachweis der Bioäquivalenz die Gleichwertigkeit sowohl für das Ausmaß als auch für die Geschwindigkeit der Resorption zu bestimmen (CPMP, 1991; FDA, 1992). Statt der Plasmaspiegelkurve selbst werden daraus abgeleitete pharmakokinetische Kenngrößen berechnet. Als charakterisierende Größen dienen für das Ausmaß die Fläche unter der Plasmaspiegelkurve ($AUC = \text{area under the curve}$) und für die Geschwindigkeit bei schnell freisetzenden Formulierungen Zeitpunkt (t_{max}) und Höhe des maximalen Serumspiegels (C_{max}). Bei kontrolliert freisetzenden Arzneimittelformen hingegen sind zur Charakterisierung der Resorptionsgeschwindigkeit andere Zielgrößen aussagefähiger (Hauschke und Steinijans, 1993).

Für viele Substanzen ist eine große interindividuelle Variabilität bekannt. Während zum Beispiel für die AUC zwischen den Individuen eines Kollektivs Variationskoeffizienten von 30% üblich sind, beträgt in einem *Crossover*-Design der intraindividuelle Variationskoeffizient oft nur noch 15% (Hauschke *et al.*, 1994). Aus diesem Grunde werden Bioäquivalenzstudien als *Crossover* angelegt, wobei *Carryover*-Effekte durch eine hinreichend lang gewählte *Washout*-Phase vernachlässigt werden können. Beim einfachen

Zwei-Perioden-*Crossover*-Design werden Test (T) und Referenz (R) somit intraindividuell verglichen, wobei die Probanden den Behandlungssequenzen RT bzw. TR zufällig zugeordnet werden.

Für die meisten Kenngrößen wird angenommen, daß sie logarithmisch normalverteilt sind. Diese Annahme ist für die AUC durch die pharmakokinetische Grundgleichung $AUC = \frac{f \cdot Dosis}{Clearance}$ gerechtfertigt, wobei f , $0 < f \leq 1$, den bioverfügbaren Dosisanteil und damit den Formulierungseffekt darstellt, während die $Clearance$ (Maß für die Eliminationsgeschwindigkeit) dem Individuumseffekt entspricht (Steinijans und Hauschke, 1990). Eine analoge Argumentation gilt für alle konzentrationsabhängigen Kenngrößen, insbesondere für C_{max} . Das entsprechende multiplikative Modell lautet:

$$X_{ijk} = \mathbf{m}_h \cdot \mathbf{p}_k \cdot b_{ij} \cdot e_{ijk},$$

dabei bezeichnet die Zufallsvariable X_{ijk} die Beobachtung der pharmakokinetischen Kenngröße bei dem j -ten Probanden, $j = 1, \dots, n_i$, in der i -ten Sequenz, $i = 0$ (RT), 1 (TR) aus der k -ten Periode, $k = 1, 2$. Der Effekt der Formulierung h wird durch \mathbf{m}_h gekennzeichnet, $h = R$ für $i + 1 = k$ und $h = T$ für $i + 1 \neq k$, für die Periodeneffekte gilt $\mathbf{p}_1 \cdot \mathbf{p}_2 = 0$. Für den zufälligen

Probandeneffekt b_{ij} gelte $\ln b_{ij}$ u.i.v $\sim \mathbf{F}\left(\frac{x}{\mathbf{s}_b}\right)$ und für den zufälligen Fehlerterm e_{ijk} gelte $\ln e_{ijk}$ u.i.v $\sim \mathbf{F}\left(\frac{x}{\mathbf{s}_e}\right)$.

Es sei darauf hingewiesen, daß für einige Kenngrößen der Resorptionsgeschwindigkeit, zum Beispiel t_{max} , ein additives Modell angenommen wird (Hauschke und Steinijans, 1993). Die Auswertung erfolgt üblicherweise mittels des verteilungsfreien Verfahrens nach Hauschke *et al.* (1990).

Aus Behördensicht ist die irrtümliche Ablehnung der Bioäquivalenz (Produzentenrisiko) von sekundärer Bedeutung; wesentlich ist die Kontrolle der irrtümlichen Annahme der Bioäquivalenz (Konsumentenrisiko). Aus diesem Grund wird das Entscheidungsproblem mit einer Intervallalternativen formuliert:

$$H_0^q : \frac{\mathbf{m}_T}{\mathbf{m}_R} \leq \mathbf{q}_1 \quad \text{oder} \quad \frac{\mathbf{m}_T}{\mathbf{m}_R} \geq \mathbf{q}_2$$

versus

$$H_1^q: \mathbf{q}_1 < \frac{\mathbf{m}_T}{\mathbf{m}_R} < \mathbf{q}_2,$$

wobei $(\mathbf{q}_1, \mathbf{q}_2)$, $0 < \mathbf{q}_1 < 1 < \mathbf{q}_2$, $\mathbf{q}_1 = \frac{1}{\mathbf{q}_2}$ den Äquivalenzbereich bezeichne. Im Falle der AUC und C_{max} werden üblicherweise $\mathbf{q}_1 = 0.80$ und $\mathbf{q}_2 = 1.25$ gewählt (FDA, 1992), wobei in Europa für C_{max} auch die Grenzen $\mathbf{q}_1 = 0.70$ und $\mathbf{q}_2 = 1.43$ zulässig sind (CPMP, 1991).

Durch logarithmische Transformation der Kenngröße wird das multiplikative Modell in das additive Modell übergeführt und die äquivalente Formulierung des Testproblems lautet

$$H_0^d: \ln \mathbf{m}_T - \ln \mathbf{m}_R \leq \ln \mathbf{q}_1 \quad \text{oder} \quad \ln \mathbf{m}_T - \ln \mathbf{m}_R \geq \ln \mathbf{q}_2$$

versus

$$H_1^d: \ln \mathbf{q}_1 < \ln \mathbf{m}_T - \ln \mathbf{m}_R < \ln \mathbf{q}_2.$$

Daher können die im Kapitel 4.1 dargestellten Methoden für die Differenz von Lokationsparametern direkt verwendet werden. Man berechnet zum Beispiel zunächst das $100(1-2\alpha)\%$ -Konfidenzintervall für $\ln \mathbf{m}_T - \ln \mathbf{m}_R$ und erhält durch anschließende inverse Transformation ($\exp(\cdot)$) der Intervallgrenzen ein $100(1-2\alpha)\%$ -Konfidenzintervall für $\frac{\mathbf{m}_T}{\mathbf{m}_R}$. Ist

die Annahme der Log-Normalverteilung für die entsprechenden Kenngrößen zweifelhaft, so kann das nichtparametrische Verfahren nach Hauschke *et al.* (1990) verwendet werden.

6.2 Anwendungsbeispiel

Das Beispiel bezieht sich auf eine Dosisäquivalenzstudie, bei der im randomisierten Zwei-Perioden-*Crossover* Versuch mit 7tägiger *Washout*-Phase an 18 gesunden Probanden 1000mg Theophyllin als Pellets in unterschiedlicher Kapselgrößen verglichen wurden (Steinijans *et al.*, 1989). Je zwei Kapseln mit 200 und 300mg Theophyllin dienten als Referenz, zwei Kapseln mit 500mg dienten als Test. Im Akutteil der Studie, d.h. nach Gabe der ersten Dosis, erfolgte die Blutentnahmen zur Theophyllinbestimmung im Plasma vor und 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 28, 36, 40, 44 und 48 Stunden nach der abendlichen Applikation um 20 Uhr. Die Berechnung der AUC (Tabelle 23) erfolgte bis 48 Stunden mit der linearen Trapezregel und wurde aufgrund des terminalen monoexponentiellen

Konzentrationsverlaufes bis Unendlich extrapoliert (Sauter *et al.*, 1992). Der Anteil der bis zum letzten Meßzeitpunkt bestimmten *AUC* zu der Gesamt-*AUC* betrug 94 (91-99)% in der Testsituation und 95 (84-99)% in der Referenzsituation; Angaben als Median (Minimum – Maximum).

Tabelle 23: Behandlungssequenz und *AUC*-Werte (*mg/L·h*)

Sequenz	Periode 1	Periode 2
<i>TR</i>	228.04	288.79
<i>RT</i>	339.03	329.76
<i>TR</i>	288.21	343.37
<i>RT</i>	242.64	258.19
<i>RT</i>	249.94	201.56
<i>TR</i>	217.97	225.77
<i>TR</i>	133.13	235.89
<i>RT</i>	184.32	249.64
<i>TR</i>	213.78	215.14
<i>TR</i>	248.98	245.48
<i>TR</i>	163.93	134.89
<i>RT</i>	209.30	231.98
<i>RT</i>	207.40	234.19
<i>TR</i>	245.92	223.39
<i>RT</i>	239.84	241.25
<i>RT</i>	211.24	255.60
<i>TR</i>	188.05	169.70
<i>RT</i>	230.36	256.55

In Tabelle 24 ist das parametrische und das verteilungsfreie 90%-Konfidenzintervall für $\frac{m_T}{m_R}$ aufgeführt. Beide Konfidenzintervalle sind vollständig im Äquivalenzbereich (0.80, 1.25) enthalten, so daß mit einer maximalen Irrtumswahrscheinlichkeit (Konsumentenrisiko) von 0.05 auf Äquivalenz hinsichtlich des Ausmaßes der Resorption geschlossen werden kann.

Tabelle 24: Zweiseitige 90%-Konfidenzintervalle für m_T / m_R für den Beispieldatensatz aus Tabelle 23

	90%-Konfidenzintervall
Parametrisches Verfahren	[0.93, 1.08]
Nichtparametrisches Verfahren	[0.94, 1.10]

Die klassische Beurteilung der Bioäquivalenz (*average bioequivalence*) basiert bislang nur auf den Lokationsparametern. Zur Beurteilung der Ähnlichkeit hinsichtlich der Variabilität wird routinemäßig ein *Groups-by-periods-Plot* erstellt; für die Dosisäquivalenzstudie erkennt

man anhand der Abbildung 14, daß die Referenzformulierung eine größere Variabilität aufweist.

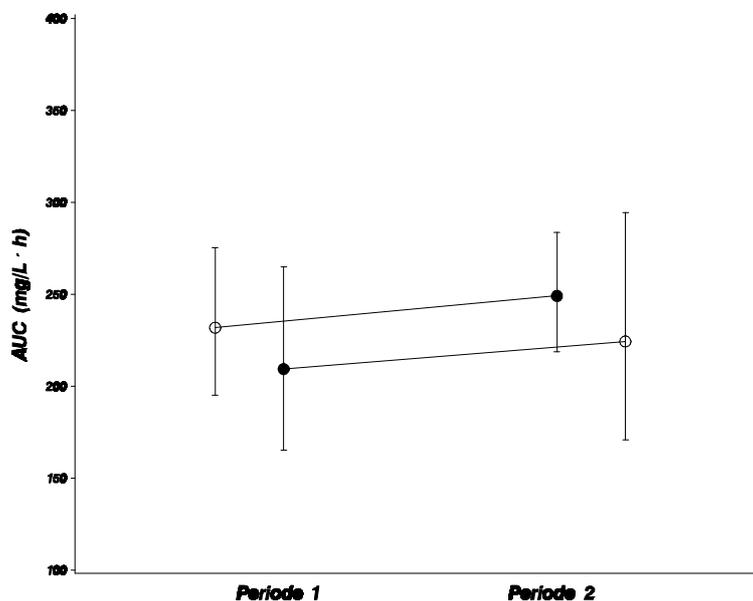


Abbildung 14: *Groups-by-Periods-Plot* für den Beispieldatensatz aus Tabelle 23 (● : Test, ○ : Referenz):
Geometrische Mittel und 68%-Bereiche

Neben der Forderung der Vergleichbarkeit hinsichtlich der Variabilität ist das Nichtvorhandensein einer ausgeprägten *subject-by-formulation* Interaktion ein wesentliches Maß für die Austauschbarkeit von Referenz- und Testformulierung; in der Abbildung 15 wird dies grafisch dargestellt. Man erkennt, daß - bezogen auf die individuellen Bioverfügbarkeits-

quotienten $\frac{X_{0j2}}{X_{0j1}}$ bzw. $\frac{X_{1j1}}{X_{1j2}}$, $j = 1, \dots, n_i$, $i = 0, 1$ - die Mehrzahl der Werte zwar sehr nahe bei

1 ist, aber auch extreme Werten oberhalb und unterhalb von 1 vorliegen.

Diese beiden zusätzlichen Anforderungen für die Austauschbarkeit von Formulierungen wurden in dem von der Food and Drug Administration (1997) erstellten Entwurf *In vivo bioequivalence studies based on population and individual bioequivalence approaches* gestellt.

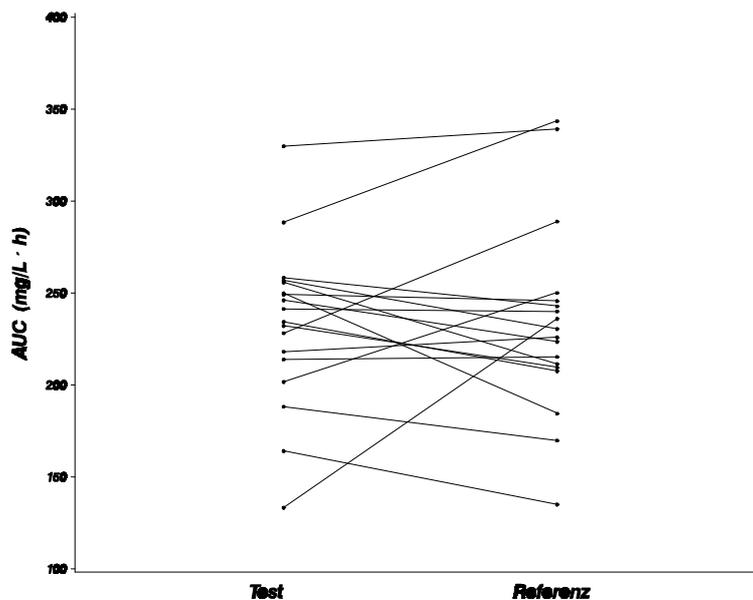


Abbildung 15: *Subject-by-treatment* Interaktion für den Beispieldatensatz aus Tabelle 23

Es sei jedoch darauf hingewiesen, daß die in dem Entwurf geforderte statistische Methodik noch Unzulänglichkeiten aufweist (Hauschke *et al.*, 1999e), zum Beispiel die fehlende Interpretierbarkeit des gewählten Parameters (*aggregate criterion*) oder die nicht adäquate Ermittlung der Stichprobengröße.

7 Zusammenfassung und Ausblick

Bislang ist es in der Praxis üblich, bei der Beurteilung von Substanzen hinsichtlich ihrer Sicherheit den indirekten Ansatz des *proof of hazard* anzuwenden. Dabei wird die Nullhypothese H_0 : *Substanz unbedenklich* gegen die Alternative H_1 : *Substanz bedenklich* getestet, und man folgert die Unbedenklichkeit der Substanz, falls die Nullhypothese zum Niveau α nicht verworfen werden kann. Der entscheidende Nachteil dieses Ansatzes ist, daß nur die Wahrscheinlichkeit der irrtümlichen Verwerfung der Nullhypothese (Produzentenrisiko) direkt durch den vorgegebenen Fehler 1. Art α begrenzt ist. Weitaus wichtiger ist jedoch die Kontrolle des Konsumentenrisikos, nämlich der irrtümlichen Annahme, daß die Substanz unbedenklich ist.

Daher wurde in dieser Arbeit der direkte Sicherheitsnachweis (*proof of safety*) vorgestellt. Hierbei wird als Nullhypothese H_0 : *Substanz bedenklich* und als Alternative H_1 : *Substanz unbedenklich* gewählt und desweiteren wird ein Schwellenwert mit in die Formulierung des Testproblems einbezogen. Für die Zweistichprobensituation wurde zunächst die statistische Methodik für die Differenz von Lokationsparametern dargestellt. Eine wesentliche Voraussetzung bei der Durchführung entsprechender Testverfahren ist dabei, daß der Schwellenwert *a-priori* in absoluten Einheiten angegeben werden kann. Die Festlegung des nicht mehr akzeptablen Risikos kann sich – insbesondere in *Screening*-Studien, bei denen eine Vielzahl toxikologischer Zielgrößen untersucht werden – als schwierig erweisen. Dennoch sollte dies zumindest bei konfirmatorischen Sicherheitsstudien für die primären Zielvariablen erfolgen. Es sei darauf hingewiesen, daß auch beim klassischen Ansatz des *proof of hazard* die Angabe des Schwellenwertes verlangt wird, da diese Größe für eine adäquate Fallzahlplanung benötigt wird.

Häufig wird der Schwellenwert auch relativ zum Erwartungswert einer Kontrollgruppe definiert, so daß sich das daraus resultierende Testproblem auf den Quotienten der Lokationsparameter bezieht. Daher wurden für diese Fragestellung die adäquaten biometrischen Methoden vorgestellt. Es sei darauf hingewiesen, daß die Methoden sich ausschließlich auf Lokationsvergleiche beziehen. Für einen simultanen Lokations- und Dispersionsvergleich kann das von Bauer und Bauer (1994) vorgestellte Konzept übertragen werden. Bauer und Bauer publizierten ein Verfahren zum simultanen Äquivalenznachweis für die Differenz von Lokationsparametern und für die Skalenparameter. Eine analoge Vorgehensweise für die Situation des simultanen Äquivalenznachweises für den Quotienten von Lokationsparametern und für die Skalenparameter erfordert jedoch zunächst, daß die Teststatistik für den Quotienten der Lokationsparameter auf den Fall ungleicher Varianzen verallgemeinert wird.

In der Zweistichprobensituation existieren bislang für die Beurteilung des Quotienten nur parametrische Methoden, da schon unter der Normalverteilungsannahme die Voraussetzung des Lokationsmodells verletzt ist. Es sollte daher untersucht werden, ob in dieser Situation sich alternative Methoden, zum Beispiel Bootstrap-Verfahren, als geeignet erweisen.

Das *Crossover*-Design kann durch die Betrachtung der Periodendifferenzen bzw. der korrigierten Periodendifferenzen auf eine Zweistichprobensituation zurückgeführt werden. Daher wurden die für die Zweistichprobensituation hergeleiteten Verfahren auf die *Crossover*-Situation übertragen. Bemerkenswert ist die Tatsache, daß die Transformation zu einer Varianzhomogenität innerhalb der beiden resultierenden Stichproben führt. Dies ist unabhängig davon, ob die den jeweiligen Behandlungen zugrundeliegenden Varianzen homogen sind oder nicht.

In der Mehrstichprobensituation bestand das zentrale Anliegen in der Ermittlung sicherer Substanzen bzw. in der Bestimmung der maximal sicheren Dosis. In der *many-to-one*-Situation wurden zum Nachweis der globalen Sicherheit das *intersection-union*-Prinzip verwendet, so daß wiederum die für die Zweistichprobensituation hergeleiteten Verfahren zur Anwendung kamen. Für den partiellen Sicherheitsnachweis wurde in der einseitigen Fragestellung die Dunnett-Prozedur verwendet bzw. zur Beurteilung der Quotienten der Lokationsparameter entsprechend verallgemeinert. Bei der zweiseitigen Fragestellung wurde zur Konstruktion der simultanen Konfidenzintervalle die Bonferroni-Korrektur verwendet. Daher sollte untersucht werden, ob diese konservative Vorgehensweise nicht durch ein Verfahren ersetzt werden kann, das das Niveau vollständiger ausschöpfen kann.

Ausgehend von dem Konzept der *a-priori* geordneten Hypothesen wurde zur Bestimmung der maximal sicheren Dosis eine wiederum auf Zweistichprobenverfahren basierende Prozedur zum multiplen Niveau α hergeleitet.

Abschließend wurde das Konzept der Bioäquivalenz dargestellt und anhand eines realen Datensatzes die zusätzliche Anforderung der individuellen Bioäquivalenz illustriert.

In der vorliegenden Arbeit wurden generell stetige Verteilungsfunktionen vorausgesetzt. Häufig liegen jedoch Ordinalskalen vor, anhand derer eine Beurteilung hinsichtlich der Sicherheit erfolgt. Die Herleitung exakter Testverfahren für den Äquivalenznachweis im Falle dichotomer Zufallsvariablen wurde erst in jüngster Zeit erbracht und die Verallgemeinerung dieser Verfahren auf ordinal skalierte Daten wird ein Forschungsschwerpunkt der kommenden Jahre sein.

8 Literaturverzeichnis

- Altman, D., Bland, M. (1994). Regression towards the mean. *British Medical Journal*, **308**: 1499.
- Bartholomew, D. J. (1961). Ordered tests in the analysis of variance. *Biometrika*, **48**: 325-332.
- Barnes, D. G., Daston, G. P., Evans, J. S., Jarabek, A. M., Kavlock, R. J., Kimmel, C. A., Park, C., Spitzer, H. L. (1995). Benchmark dose workshop: criteria for use of a benchmark dose to estimate a reference dose. *Regulatory Toxicology and Pharmacology*, **21**: 296-306.
- Bauer, P., Bauer, M. M. (1994). Testing equivalence simultaneously for location and dispersion of two normally distributed populations. *Biometrical Journal*, **36**: 643-660.
- Bedotto, J. E., Gay, R. G., Graham, S. D., Morkin, E., Goldman, S. (1989). Cardiac hypertrophy induced by thyroid hormone is independent of loading conditions and beta-adrenoceptor blockade. *Journal of Pharmacology and Experimental Therapeutics*, **248**: 632-636.
- Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics*, **24**: 295-300.
- Bofinger, E., Bofinger, M. (1993). Equivalence of normal means compared with a control. *Communications in Statistics - Theory and Methods*, **22**: 3117-3141.
- Bofinger, E., Bofinger, M. (1995). Equivalence with respect to a control: stepwise tests. *Journal of the Royal Statistical Society B*, **57**: 721-733.
- Brown, I. D., Erdreich, L. S. (1989). Statistical uncertainty in the no-observed-effect-level. *Fundamental and Applied Toxicology*, **13**: 235-244.
- Bross, I. D. (1985). Why proof of safety is much more difficult than proof of hazard. *Biometrics*, **41**: 785-793.
- Brown, K. G., Erdreich, L. S. (1989). Statistical uncertainty in the no-observed-adverse-effect level. *Fundamental and Applied Toxicology*, **13**: 235-244.
- Chan, I. S. F. (1998). Exact tests of equivalence and efficacy with a non-zero lower bound for comparative studies. *Statistics in Medicine*, **17**: 1403-1413.
- Chanter, D. O., Tuck, M. G., Coombs, D. W. (1987). The chances of false negative results in conventional toxicology studies with rats. *Toxicology*, **43**: 65-74.
- Chow, S.-C., Liu, J.-P. *Design and Analysis of Bioavailability and Bioequivalence Studies*, Marcel Dekker, New York, 1992.
- CPMP Working Party on Efficacy and Medicinal Products (1991). Investigation of bioavailability and bioequivalence, III/54/89-EN.
- Crump, K. S. (1984). A new method for determining allowable daily intakes. *Fundamental and Applied Toxicology*, **4**: 854-871.

- Diletti, E., Hauschke, D., Steinijans, V. W. (1991). Sample size determination for bioequivalence assessment by means of confidence intervals. *International Journal of Clinical Pharmacology, Therapy and Toxicology*, **29**: 1-8.
- Dunnett, C. W., Tamhane, A. C. (1991). Step-down multiple tests for comparing treatments with a control in unbalanced one-way layout. *Statistics in Medicine*, **10**: 939-947.
- Dunnett, C. W., Tamhane, A. C. (1992). A step-up multiple test procedure. *Journal of the American Statistical Association*, **87**: 162-170.
- Dunnett, C. W., Tamhane, A. C. (1995). Step-up multiple testing of parameters with unequally correlated estimates. *Biometrics*, **51**: 217-227.
- Erickson, W. P., McDonald, L. L. (1995). Tests for bioequivalence of control media and test media in studies of toxicity. *Environmental Toxicology and Chemistry*, **14**: 1247-1256.
- Dubey, S. D. (1991). Some thoughts on the one-sided and two-sided tests. *Journal of Biopharmaceutical Statistics*, **1**: 139-150.
- Fieller, E. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society B*, **16**, 175-185.
- Fisher, R. A. *The design of experiments*, Oliver and Boyd, London, 1935.
- Food and Drug Administration. Guidance on statistical procedures for bioequivalence studies using a standard two-treatment crossover design. Informal communication by the Division of Bioequivalence, Office of Generic Drugs, Rockville, MD, 1992.
- Food and Drug Administration. Toxicological principles for the safety assessment of direct food additives and color additives used in food, *Draft*, 1993.
- Food and Drug Administration. In vivo bioequivalence studies based on population and individual bioequivalence approaches, *Draft Guidance for Industry*, Center for Drug Evaluation and Research, Rockville, 1997.
- Gaylor, D. W. (1983). The use of safety factors for controlling risk. *Journal of Toxicology and Environmental Health*, **11**: 329-336.
- Gaylor, D. W. (1996). Quantalization of continuous data for benchmark dose estimation. *Regulatory Toxicology and Pharmacology*, **24**: 246-250.
- Gaylor, D. W., Chen, J. J. (1996). Precision of benchmark dose estimates for continuous (nonquantal) measurements of toxic effects. *Regulatory Toxicology and Pharmacology*, **24**: 19-23.
- Genz, A., Bretz F. (1999). Numerical computation of multivariate *t*-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation*, (zur Veröffentlichung eingereicht).
- Giani, G., Straßburger, K. (1994). Testing and selecting for equivalence with respect to a control.

- Journal of the American Statistical Association*, **89**: 320-329.
- Grizzle, J. E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics*, **21**: 467-480.
- Guilbaud, O. (1993). Exact inference about the within-subject variability in 2x2 crossover trials. *Journal of the American Statistical Association*, **88**: 939-946.
- Hansson, S. O. (1997). Can we reverse the burden of proof? *Toxicology Letters*, **90**: 223-228.
- Hauschke, D., Steinijans, V. W., Diletti, E. (1990). A distribution-free procedure for the statistical analysis of bioequivalence studies. *International Journal of Clinical Pharmacology, Therapy and Toxicology*, **28**: 72-78.
- Hauschke, D., Steinijans, V. W. (1991). Ein verteilungsfreies Verfahren zur statistischen Auswertung von Bioäquivalenzstudien. In: *Biometrie in der chemisch-pharmazeutischen Industrie* (Vollmar, J.), Gustav Fischer, Stuttgart, 69-89.
- Hauschke, D., Steinijans, V. W., Diletti, E., Burke, M. (1992). Sample size determination for bioequivalence assessment using a multiplicative model. *Journal of Pharmacokinetics and Biopharmaceutics*, **20**: 557-561.
- Hauschke, D., Steinijans, V. W. (1993). Modified-release dosage forms: acceptance criteria and statistics for bioequivalence. *Drug Information Journal*, **27**: 903-909.
- Hauschke, D., Steinijans, V. W., Böhm, A. (1994). Bioäquivalenzstudien – Methoden und Anwendungen. In: *Medizinische Dokumentation. Entwicklung, heutiger Stand, Perspektiven*, (Gaus, W., Kugelmann, M.), Universitätsverlag Ulm, Ulm, 382-387.
- Hauschke, D. (1995). Testprinzipien bei Safety-Studien in der Toxikologie. In: *Medizinische Forschung-Ärztliches Handeln*, (Trampisch H. J., Lange, S.), MMV Medizin Verlag, München, 104-107.
- Hauschke, D., Steinijans, V. W., Hothorn, L. A. (1996). A note on Welch's approximate *t*-solution to bioequivalence assessment. *Biometrika*, **83**: 236-237.
- Hauschke, D. (1997). Statistical proof of safety in toxicological studies. *Drug Information Journal*, **31**: 357-361.
- Hauschke, D., Hayashi, M., Lin, K. K., Lovell, D. P., Robinson, W. D., Yoshimura, I. (1997). Recommendations for biostatistics of mutagenicity studies. *Drug Information Journal*, **31**: 323-326.
- Hauschke, D., Hothorn, L. A. (1998). Safety assessment in toxicological studies: proof of safety versus proof of hazard. In: *Design and Analysis of Animals Studies in Pharmaceutical Development*, (Chow, S.-C., Liu, J.-P.), Marcel Dekker, New York, 197-225.
- Hauschke, D., Schall, R., Luus, H. G. (1999a). Statistical significance versus clinical significance. In:

- Encyclopedia of Biopharmaceutical Statistics*, (Chow, S.-C.), Marcel Dekker, New York, (zur Veröffentlichung angenommen).
- Hauschke, D., Kieser, M., Hothorn, L. A. (1999b). Proof of safety in toxicology based on the ratio of two means for normally distributed data. *Biometrical Journal*, **41**: 295-304.
- Hauschke, D., Kieser, M., Diletti, E., Burke, M. (1999c). Sample size determination for proving equivalence based on the ratio of two means for normally distributed data. *Statistics in Medicine*, **18**: 93-105.
- Hauschke, D., Kieser, M., Hothorn, L. A., Bretz, F. (1999d). Statistical methods for multiple comparisons with a control based on the ratio of two means for normally distributed data, (in Vorbereitung).
- Hauschke, D., Steinijans, V. W. (1999e). The U.S. draft guidance regarding population and individual bioequivalence approaches: comments by a research-based pharmaceutical company. *Statistics in Medicine*, (zur Veröffentlichung angenommen).
- Hayes, J. P. (1987). The positive approach to negative results in toxicology studies. *Ecotoxicology and Environmental Safety*, **14**: 73-77.
- Healy, G. F. (1987). Power calculations in toxicology. *Alternatives to Laboratory Animals*, **15**: 132-139.
- Hettmansperger, T. P., Norton, R. M. (1987). Tests for patterned alternatives in k -sample problems. *Journal of the American Statistical Association*, **82**: 292-299.
- Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, **75**: 800-802.
- Hochberg, Y., Tamhane, A. C. *Multiple comparison procedures*, Wiley, New York, 1987.
- Holland, B., Ordoukhani, N. K. (1990). Balancing type I and type II error probabilities: further comments on proof of safety vs proof of hazard. *Communications in Statistics - Theory and Methods*, **19**: 3557-3570.
- Hollander, M., Wolfe, D. A. *Nonparametric Statistical Methods*. Wiley, New York, 1973.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**: 65-70.
- Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, **75**: 383-386.
- Hothorn, L. A. (1991). General statistical principles in testing of toxicological studies. In *Statistical Methods in Toxicology*, (Hothorn, L. A.), Springer Verlag, Berlin, 111-131.
- Hothorn, L. A., Lehmacher, W. (1991). A simple testing procedure control versus k treatments for one-sided ordered alternatives, with application in toxicology. *Biometrical Journal*, **33**: 179-189.

- Hothorn, L. A. (1992). Biometrische Analyse toxikologischer Untersuchungen. In: *Statistisches Know how in der medizinischen Forschung*, (Adams, J.), Ullstein Mosby, Berlin, 475-590.
- Hothorn, L. A. (1995). Biostatistical analysis of the control vs k treatments design including a positive control group. In: *Biometrie in der chemisch-pharmazeutischen Industrie*, (Vollmar, J.), Gustav Fischer Verlag, Stuttgart, 19-26.
- Hothorn, L. A., Neuhäuser, M., Koch, H.-F. (1997). Analysis of randomized dose-finding studies: closure test modification based on multiple contrast test. *Biometrical Journal*, 39: 467-479.
- Hothorn, L. A., Hauschke, D. (1998). Principles in statistical testing in randomized toxicological studies. In: *Design and Analysis of Animal Studies in Pharmaceutical Development*, (Chow, S.-C., Liu, J.-P.), Marcel Dekker, New York, 79-133.
- Hothorn, L. A., Hauschke, D. (2000). Identifying the maximum safe dose: a multiple testing approach. *Journal of Biopharmaceutical Statistics*, (zur Veröffentlichung angenommen).
- Hothorn, L. A., Hayashi, M., Seidel, D. (1999). Dose-response relationship in mutagenicity assays including an appropriate positive control group: a multiple testing approach. *Environmental and Ecological Statistics*, (zur Veröffentlichung angenommen).
- Hsu, J. C. *Multiple comparisons*, Chapman and Hall, London, 1996.
- Hsu, J. C., Berger, R. L. (1998). Stepwise confidence intervals without multiplicity adjustment for dose response, multiple endpoints and toxicity studies. *Technical Report*.
- Jonckheere, A. R. (1954). A distribution-free k -sample test against ordered alternatives. *Biometrika*, **41**: 133-145.
- Jones, B., Kenward, M. G. *Design and Analysis of Cross-over Trials*, Chapman and Hall, London, 1989.
- Kieser, M., Hauschke, D., Wassmer, G. (1997). Some results on the power function of Sasabuchi's test, *Technical Report*.
- Kieser, M., Hauschke, D. (1999a). Approximate sample sizes for testing hypotheses about the ratio and difference of two means. *Journal of Biopharmaceutical Statistics*, (zur Veröffentlichung angenommen).
- Kieser, M., Hauschke, D. (1999b). Statistical methods for testing the ratio of two location parameters in the crossover design, (in Vorbereitung).
- Kopit, J. S., Berger, R. L. (1998). A more powerful test for a practical difference between binomial proportions. *Proceedings of the Biopharmaceutical Section of the American Statistical Association*, (zur Veröffentlichung angenommen).
- Lehmacher, W. (1987). Verlaufskurven und Crossover: Statistische Analyse von Verlaufs-kurven im

- Zwei-Stichproben-Vergleich und von Crossover-Versuchen. In: Medizinische Informatik und Statistik (Überla, K., Reichertz, P. L., Victor, N.), Springer Verlag, Berlin.
- Lehmacher, W. (1991). Analysis of the crossover design in the presence of residual effects. *Statistics in Medicine*, **10**: 891-899.
- Lehmann, E. L. *Nonparametrics: statistical methods based on ranks*, Holden-Day, San Francisco, 1975.
- Liu, J.-P., Chow, S.-C. (1992). Sample size determination for the two one-sided tests procedure in bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics*, **20**: 101-104.
- Locke, C. S. (1984). An exact confidence interval from untransformed data for the ratio of two formulation means. *Journal of Pharmacokinetics and Biopharmaceutics*, **12**, 649-655.
- Lüdin, E. (1985). A test procedure based on ranks for the statistical evaluation of toxicological studies. *Archives of Toxicology*, **58**: 57-58.
- Mace, A. E. *Sample-size determination*, Reinhold Publishing Corporation, New York, 1964.
- Mack, G. A., Wolfe, D. A. (1981). *K*-sample rank tests for umbrella alternatives. *Journal of the American Statistical Association*, **76**: 175-181.
- Mandallaz, D. and Mau, J. (1981). Comparison of different methods for decision-making in bioequivalence assessment. *Biometrics*, **37**: 213-222.
- Marcus, R., Peritz, E., Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, **63**: 655-660.
- Maurer, W., Hothorn, L. A., Lehmacher, W. (1995). Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypotheses. In: *Biometrie in der chemisch-pharmazeutischen Industrie* (Vollmar, J.), Gustav Fischer, Stuttgart, 3-18.
- Metzler, C. M. (1974). Bioavailability – a problem in equivalence. *Biometrics*, **30**: 300-317.
- Millard, S. P. (1987). Proof of safety vs proof of hazard. *Biometrics*, **43**: 719-725.
- Muller, K. E., Benignus, V. A. (1992). Increasing scientific power with statistical power. *Neurotoxicology and Teratology*, **14**: 211-219.
- Nam, J. (1987). A simple approximation for calculating sample sizes for detecting linear trends in proportions. *Biometrics*, **43**: 701-705.
- Neuhäuser, M., Hothorn, L. A. (1995). Auswertung der Dosis-Wirkungs-Abhängigkeit des Ames Mutagenitätsassay bei direkter Kontrolle des Konsumentenrisikos. In: *Medizinische Forschung-Ärztliches Handeln*, (Trampisch H. J., Lange, S.), MMV Medizin Verlag, München, 113-116.
- Neuhäuser, M., Hothorn, L. A. (1997). The control of the consumer risk in the Ames assay. *Drug Information Journal*, **31**: 363-367.
- Owen, D. B. (1965): A special case of a bivariate non-central *t*-distribution. *Biometrika*, **52**: 437-446.

- Peace, K. E. (1993). Design and analysis considerations for safety data, particularly adverse events. In: *Drug Safety Assessment in Clinical Trials*, (Gilbert, G. S.), Marcel Dekker, New York, 305-316.
- Peto, R., Pike, M. C., Day, N. E., Gray, R. G., Lee, P. N., Parish, S., Peto, J., Richards, S., Wahrendorf, J. (1980). Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. In: *Long-term and Short-term Screening Assays for Carcinogens: A Critical Appraisal*, IARC Monographs Supplement 2, Lyon, 311-426.
- Phillips, K. F. (1990). Power of the two one-sided tests procedure in bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics*, **18**: 137-144.
- Rom, D. M. (1990). A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika*, **77**: 663-665.
- Rom, D. M., Costello, R. J., Connell L. T. (1994). On closed test procedures for dose-response analysis. *Statistics in Medicine*, **13**: 1583-1596.
- Röhmel, J. (1998). Therapeutic equivalence investigations: statistical considerations. *Statistics in Medicine*, **17**: 1703-1714.
- Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Annals of Mathematical Statistics*, **24**: 220-238.
- Samuel-Cahn, E. (1996). Is the Simes improved Bonferroni procedure conservative? *Biometrika*, **83**: 928-933.
- Sasabuchi, S. (1988). A multivariate one-sided test with composite hypotheses determined by linear inequalities when the covariance matrix has an unknown scale factor, *Memoirs of the Faculty of Science, Kyushu University, Series A, Mathematics*, **42**, 9-19
- Sauter, R., Steijnmans, V. W., Diletti, E., Böhm, A., Schulz, H. U. (1992). Presentation of results from bioequivalence studies. *International Journal of Clinical Pharmacology, Therapy and Toxicology*, **30**: 233-256.
- Schuurmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, **15**: 657-680.
- Schuurmann, D. J. (1989). Confidence intervals for the ratio of two means from a crossover study. *American Statistical Association, Proceedings of the Biopharmaceutical Section*, 121-126.
- Shirley, E. (1977). A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. *Biometrics*, **33**: 386-389.
- Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**: 751-754.
- Stallard, N., Whitehead, A. (1996). An alternative approach to the analysis of animal carcinogenicity studies. *Regulatory Toxicology and Pharmacology*, **23**: 244-248.

- Steel, R. G. D. (1959). A multiple rank sum test: treatments versus control. *Biometrics*, **15**: 560-572.
- Steinijans, V. W., Sauter, R., Jonkman, J. H. G., Schulz, H. U., Stricker, H., Blume, H. (1989): Bioequivalence studies: single vs multiple dose. *International Journal of Clinical Pharmacology, Therapy and Toxicology*, **27**: 261-266.
- Steinijans, V. W., Hauschke, D. (1990). Update on the statistical analysis of bioequivalence studies. *International Journal of Clinical Pharmacology, Therapy and Toxicology*, **28**: 105-110.
- Tamhane, A. C., Hochberg, Y., Dunnett, C. W. (1996). Multiple test procedures for dose finding. *Biometrics*, **52**: 21-37.
- Tamhane, A. C., Dunnett, C. W., Wetherington, J. D., Green, J. W. (1999). Multiple test procedures for identifying a safe dose. *Biometrics*, (zur Veröffentlichung eingereicht).
- Tukey, J. W., Ciminera, J. L., Heyse, J. F. (1985). Testing the statistical certainty of a response to increasing doses of a drug. *Biometrics*, **41**: 295-301.
- Victor, N. (1987). On clinically relevant differences and shifted nullhypotheses. *Methods of Information in Medicine*, **26**: 109-116.
- Vuorinen, J., Tuominen, J. (1994). Fieller's confidence intervals for the ratio of two means in the assessment of average bioequivalence from crossover data. *Statistics in Medicine*, **13**: 2531-2545.
- Vuorinen, J., Turunen, J. (1996). A three-step procedure for assessing bioequivalence in the general mixed model framework. *Statistics in Medicine*, **15**: 2635-2655.
- Vuorinen, J., Turunen, J. (1997). A simple three-step procedure for parametric and nonparametric assessment of bioequivalence. *Drug Information Journal*, **32**: 167-180.
- Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics*, **27**: 103-117.
- Williams, D. A. (1972). The comparison of several dose levels with a zero dose control. *Biometrics*, **28**: 519-531.
- Yanagawa, T., Kikuchi, Y., Brown, K. G. (1994). Statistical issues on the no-observed-adverse-effect level in categorical response. *Environmental Health Perspectives*, **102**, **Supplement 1**: 95-101
- Yee, K. F. (1986). The calculation of probabilities in rejecting bioequivalence. *Biometrics*, **42**: 961-965.

Anhang: SAS-Programme zur Gütebestimmung

Kursiv geschriebene Zeilen beinhalten z.B. Plausibilitätsprüfungen zu den jeweiligen Funktionsparametern und haben mit den Algorithmen nichts zu tun. Deswegen werden auch Funktionen, die in diesen Zeilen vorzufinden sind, nicht im Listing aufgeführt.

Zeilen wie:

```
onerr='if fa00002=0 then do;ok=0;resume;end;';
call push(onerr);
```

werden als Vorbereitung zur Laufzeitfehler-Erkennung benutzt.

Modul PSA_ERR gibt Fehlermeldungen aus, hier die Fehlerliste:

```
PSA (-1) Parameter(s) not of appropriate type
PSA (-2) BVAL not 0 or 1
PSA (-3) Check parameter(s): No of rows,columns
PSA (-4) Integer parameter(s) are not integer
PSA (-5) NF < 2
PSA (-6) Call to SAS-function PROBT failed
PSA (-7) R < 0.001
PSA (-8) Error number -8
PSA (-9) PS_POW2R: Problems with integration
SSC (-401) Errorreturn from a NCT-module
```

```
/* SSC.INC
```

```
Start      : 28.10.94
Last edit: <13-NOV-1998 08:35:31.88>
```

```
Version   : V1.1
```

```
Power calculation
```

```
*/
```

```
start ssc_pow(design      /* Design Crossover, 2-unpaired groups */
             ,model       /* Additive Multiplicative Sasabuchi */
             ,sideval     /* 1-2-Sided */
             ,sidedir     /* Direction if 1-Sided */
             ,alpha       /* Alpha */
             ,n           /* Samplesize */
             ,x           /* X-value (Vector) */
             ,cv          /* Coefficient of variation */
             ,cv2         /* 2nd CV (Sasabuchi Crossover) */
             ,th1        /* Lower limit */
             ,th2        /* Upper limit */
             ,powerval    /* Returned Powervalue */
             ) global(ps_dp1,ps_dp2,ps_cnt,ps_tech);
```

```
run i_psa001(dummy);
```

```

modnam='SSC_POW';

ps_dp1=0;
ps_dp2=0;

tt=type(design)+type(model)+type(sideval)+type(sidedir)+type(alpha)
+type(n)+type(x)+type(cv)+type(th1)+type(th2)+type(powerval);

if tt^='NNNNNNNNNN' then do;
  ierr= -1;
  run psa_err(modnam,ierr);
  return(ierr);
end;
t1=nrow(design)+nrow(model)+nrow(sideval)+nrow(sidedir)+nrow(alpha)
+nrow(n)+nrow(cv)+nrow(th1)+nrow(th2);
t2=ncol(design)+ncol(model)+ncol(sideval)+ncol(sidedir)+ncol(alpha)
+ncol(n)+ncol(x)+ncol(cv)+ncol(th1)+ncol(th2)+ncol(powerval);
if t1+t2 ^= 20 then do;
  ierr= -3;
  run psa_err(modnam,ierr);
  return(ierr);
end;

if (ceil(n)-floor(n))^=0 then do;
  ierr= -4;
  run psa_err(modnam,ierr);
  return(ierr);
end;

powerval = -1;

pp=0;
run i_ssc001(1,design,model,sideval,sidedir,alpha,n,pp,x,
            cv,cv2,th1,th2,ierr);
if ierr ^= 0 then do;
  run psa_err(modnam,ierr);
  return(ierr);
end;

nx=nrow(x);
powerval=j(nx,1,-1);

do m=1 to nx;
  if x[m]=. then do;
    powerval[m]=.;
  end; else do;

    if model=3 then do;
      nf=2#n-2;          /* Sasabuchi Crossover and Unpaired 2 groups */
    end; else do;
      if design=1 then do;
        nf = n-2;
        c=2;             /* Crossover */
      end; else do;
        nf = n-2;
        c=4;             /* Unpaired two groups */
      end;
    end;

    tv = tinv(1-alpha,nf);

    if model = 1 then do;
      u = cv # sqrt(c/n);          /* Add. */
      d1= (x[m]-th1)/u;           /* Left or twosided */
      d2= (x[m]-th2)/u;           /* Right or twosided */
    end; else do;
      if model=2 then do;
        u = sqrt( log(1+cv#cv)#(c/n) ); /* Mult. */
        v = log(x[m]);
        d1= (v-log(th1))/u;        /* Left or twosided */
      end;
    end;
  end;
end;

```

```

    d2= (v-log(th2))/u;                                /* Right or twosided */
end; else do;
/*
    if design=1 then do;                                MODEL = 3      SASABUCHI  */
/* Crossover */
    ut1= (1-th1)##2;
    ut2= (1-th2)##2;
    wt1= cv#cv#(1+th1#th1)+cv2#cv2#ut1;
    wt2= cv#cv#(1+th2#th2)+cv2#cv2#ut2;
    u1= sqrt(wt1/2/n);
    u2= sqrt(wt2/2/n);
    rx1= cv#cv#(1+th1#th2)+cv2#cv2#(1+th1#th2-th1-th2);
    corr= rx1/sqrt(wt1#wt2);
end; else do;                                /* Unpaired two groups */
    ut1= 1+th1#th1;
    ut2= 1+th2#th2;
    u1= cv # sqrt(ut1/n);
    u2= cv # sqrt(ut2/n);
    corr= (1+th1#th2)/sqrt(ut1#ut2);
end;
d1= (x[m]-th1)/u1;
d2= (x[m]-th2)/u2;
end;
end;

ierr = -401;
kerr = 0;

temp=0;
if sideval = 1 then do;
    if sidedir < 0 then do;                                /* Onesided */
        kerr = ps_nct(tv,nf,d1,temp);
        if kerr ^=0 then return(ierr);
        power = 1 - temp;
    end; else do;
        kerr = ps_nct(-tv,nf,d2,temp);
        if kerr ^=0 then return(ierr);
        power = temp;
    end;
end; else do;
    if model=3 then do;                                /* SASABUCHI */
        part1=0;part2=0;
        kerr=ps_pow2r(1e6,-tv,nf,d1,d2,corr,part1);
        if kerr ^=0 then return(ierr);
        ps_dp1=ps_cnt;
        ps_dp2=ps_tech#10;
        kerr=ps_pow2r( tv,-tv,nf,d1,d2,corr,part2);
        if kerr ^=0 then return(ierr);
        ps_dp1=ps_dp1+ps_cnt;
        ps_dp2=ps_dp2+ps_tech;
        power = part1 - part2;
    end; else do;
        fakt = sqrt(nf)/(2#tv);
        r = (d1-d2) # fakt;                                /* Twosided */
        q1=0;q2=0;
        kerr=ps_nctfr(-tv,nf,d2,r,q1);
        if kerr ^=0 then return(ierr);
        kerr=ps_nctfr( tv,nf,d1,r,q2);
        if kerr ^=0 then return(ierr);
        power = q1 - q2;
    end;
end;

powerval[m]=power;
end;
end;

return(0);
finish;

```

```

/* NCT.INC

Start      : 28.10.94
Last edit: <13-NOV-1998 08:35:31.88>

Version   : V1.0

Noncentral t-distribution
*/

start ps_nct(tv      /* T-value */
             ,nf     /* Degrees of freedom */
             ,d      /* Noncentral parameter */
             ,nct    /* nct(tv,nf,d) */
             ) global(fa00002);

run i_psa001(dummy);

modnam='PS_NCT';

if type(tv)+type(nf)+type(d)+type(nct) ^= 'NNNN' then do;
  ierr= -1;
  run psa_err(modnam,ierr);
  return(ierr);
end;
if nrow(tv)+nrow(nf)+nrow(d)+nrow(nct)+
  ncol(tv)+ncol(nf)+ncol(d)+ncol(nct) ^= 8 then do;
  ierr= -3;
  run psa_err(modnam,ierr);
  return(ierr);
end;

if (ceil(nf)-floor(nf))^=0 then do;
  ierr= -4;
  run psa_err(modnam,ierr);
  return(ierr);
end;

nct= -1;

if nf < 2 then do;
  ierr= -5;
  run psa_err(modnam,ierr);
  return(ierr);
end;

ok=1;
onerr='if fa00002=0 then do;ok=0;resume;end;';
call push(onerr);

if d < 0 then do;
  fa00002=0;
  nct = 1 - probt(-tv,nf,-d);
  fa00002=1;
end; else do;
  fa00002=0;
  nct = probt(tv,nf,d);
  fa00002=1;
end;
if ok=0 then do;
  ierr= -6;
  run psa_err(modnam,ierr);
  return(ierr);
end;
return(0);
finish;

start ps_nctfr(tv      /* T-value */
              ,nf     /* Degrees of freedom */
              ,d      /* Noncentral parameter */
              ,r      /* Integration-limit */

```

```

                ,nctfr /* NCT-fraction */
                );
/*
Fraction of the UNIVARIATE NON-CENTRAL-t-DISTRIBUTION
(-> Bivariate )
Fraction of Cumulative distribution function  $Q(t,d;0,R)$ 
f
The integral in the noncentral-t-distribution INTEGRAL  $\int_0^{\infty}$ 
0
is now reduced to  $\int_0^R$ 
0
NCTFR :  $Q(t,d;0,R)$ 
f
Formulas from D.B.Owen p.439
*/

run i_psa001(dummy);

modnam='PS_NCTFR';

if type(tv)+type(nf)+type(d)+type(r)+type(nctfr) ^= 'NNNNN' then do;
    ierr= -1;
    run psa_err(modnam,ierr);
    return(ierr);
end;
if nrow(tv)+nrow(nf)+nrow(d)+nrow(r)+nrow(nctfr)+
ncol(tv)+ncol(nf)+ncol(d)+ncol(r)+ncol(nctfr) ^= 10 then do;
    ierr= -3;
    run psa_err(modnam,ierr);
    return(ierr);
end;

if (ceil(nf)-floor(nf))^=0 then do;
    ierr= -4;
    run psa_err(modnam,ierr);
    return(ierr);
end;

nctfr = -1;

if nf < 2 then do;
    ierr= -5;
    run psa_err(modnam,ierr);
    return(ierr);
end;
if r < 1e-3 then do;
    ierr= -7;
    run psa_err(modnam,ierr);
    return(ierr);
end;

sq2phi= 2.506628274631; /* SQRT(2*PHI) */

a=tv/sqrt(nf);
b=nf/(nf+tv#tv);

sqb=sqrt(b);

par1=d#a#sqb;
par2=(d#a#b-r)/sqb;

fak= 0.3989422804014; /* 1/sqrt(2*PHI) */

x=d#sqb; dens =fak#exp(-x#x/2); /* Density of normal distribution */
x=par1; dens1 =fak#exp(-x#x/2);
x=par2; dens2 =fak#exp(-x#x/2);
x=r; densr =fak#exp(-x#x/2);

```

```

x=a#r-d;      densa =fak#exp(-x#x/2);

mk0=a#sqb#dens#(probnorm(par1)-probnorm(par2));      /* M0 */
hk0=-densr#probnorm(a#r-d);                          /* H0 */
meven=mk0 + hk0;                                     /* Even sum */

if nf>=3 then do;
  mk=b#(d#a#mk0+a#dens#(dens1-dens2));              /* M1 */
  hk=r#hk0;                                          /* A3 = 1 */ /* H1 */
  modd =mk + hk;                                     /* Odd sum */
end;

if nf>=4 then do;

  /* Usage of scalars, because they are faster as matrices and we must
  use (sure?) a loop to calculate our result!

  At K=2 (Loopbegin) ak2 : A4   (Loopend) ak : A3   ak1 : A4

  At K=3 (Loopbegin) ak2 : A5 = 1/3/A4 = 1/3/ak1

  General : ak2 = 1/k/ak1   with ak2==A[k+2] , ak1==A[k+1]

  At K=2 L1 = L[k-1]
  At K=3 L2 = L[k-1] = ak2#r#L[k-2]

  */

  lv=a#b#r#densr#densa/2.;

  ak =1;
  ak1=1;

  iflag=1;
  do k=2 to nf-2;
    ak2=1/k/ak1;
    if k>2 then lv = ak2#r#lv;
    mk1=(k-1)/k # b # (ak#d#a#mk + mk0) - lv;
    hk1=ak2#r#hk;
    if iflag=1 then do;
      meven=meven+mk1+hk1;
    end; else do;
      modd =modd +mk1+hk1;
    end;
    iflag = -iflag;
    ak =ak1;  ak1=ak2;
    mk0=mk;   mk =mk1;
    hk =hk1;
  end;
end;

if int(mod(nf,2)) = 0 then do;
  nctfr=probnorm(-d)+sq2phi#meven;
end; else do;
  tfunc1=0;tfunc2=0;tfunc3=0;
  idummy=ps_t_fct(d#sqb,a      ,tfunc1);
  idummy=ps_t_fct(r,(a#r-d)/r,tfunc2);
  if abs(d) < 1e-10 then do;
    dval = 1e-10;
    if d<0 then dval = -dval;
  end; else dval=d;
  aval= a - r/b/dval;
  idummy=ps_t_fct(d#sqb,aval,tfunc3);
  nctfr=probnorm(r) + 2#(tfunc1-tfunc2-tfunc3) + 2#modd;
  if d >= 0 then nctfr= nctfr-1;
end;

return(0);
finish;

start ps_t_fct(h      /* H-value */

```

```

        ,a      /* A-value */
        ,tfunc /* T-function returnvalue */
    );
/*
-- T-function, see OWEN "Tables for computing bivariate normal
probabilities"
Annals of Mathematical Statistics, 27 (1956)
Page 1075..1090
*/

run i_psa001(dummy);

modnam='PS_T_FCT';

if type(h)+type(a)+type(tfunc) ^= 'NNN' then do;
    ierr= -1;
    run psa_err(modnam,ierr);
    return(ierr);
end;
if nrow(h)+nrow(a)+nrow(tfunc)+
    ncol(h)+ncol(a)+ncol(tfunc) ^= 6 then do;
    ierr= -3;
    run psa_err(modnam,ierr);
    return(ierr);
end;

phi=3.14159265359;
/*
-- Integral for A= 0..1
*/
if a < 0 then do;
    aa=-a;
    ineg=1;
end; else do;
    aa=a;
    ineg=0;
end;
if aa > 1 then do;
    mode=2;
    g1=probnorm(abs(h)); /* If A<0 , use T(h,-a) = -T(h,a) */
    g2=probnorm(abs(aa#h));
    func=(g1+g2)/2-g1#g2;
    hh=aa#h;
    aa=1/aa;
end; else do;
    mode=1;
    hh=h;
end;

if abs(hh) > 6 then do;
    tf=0; /* Convergence to zero */
end; else do;
    hh=hh##2;
    hf=exp(-hh/2);

    sf=1; /* i=0 : h^0/(2^0*0!) */
    sumf=sf; /* SUM i=0 */
    vfak=1; /* (-1)^0 */

    sum=(1-hf)#aa; /* c0*A^(2*0+1) */

    jx=1;

    c=1;
    do while(abs(c)>1e-10);
        sf=sf#hh/2/jx; /* i=j : H^(2*i)/2^i/i! */
        sumf=sumf+sf; /* Sum i=0..j */
        vfak=-vfak; /* (-1)^j */
        ind = 2#jx+1;
        c=vfak#(1-hf#sumf)/ind # aa##ind; /* Cj * a^(2j+1) */
        sum=sum+c;
        jx=jx+1;
    end;
end;

```

```

end;

tf=(atan(aa)-sum)/2/phi;
end;

if mode = 1 then tfunc=tf; else tfunc=func-tf;
if ineg = 1 then tfunc = -tfunc;

return(0);
finish;

start ps_pow2r(t1      /* Part of 1st integration limit to PROBBNRM */
              ,t2      /* Part of 2nd integration limit to PROBBNRM */
              ,nf      /* Degrees of freedom */
              ,d1      /* 1st noncentrality value */
              ,d2      /* 2nd noncentrality value */
              ,r       /* Correlation */
              ,pow2r) /* Result: POWER */
  global(ps_cnt,ps_tech,ps_zt1,ps_zt2,ps_znf,ps_zd1,ps_zd2,ps_zr);

run i_psa001(dummy);

modnam='PS_POW2R';

pow2r = -1; /* Returnvalue if error */
ps_cnt=0;
ps_tech=0;

if type(t1)+type(t2)+type(nf)+type(d1)+type(d2)+type(r) ^= 'NNNNNN' then do;
  ierr= -1;
  run psa_err(modnam,ierr);
  return(ierr);
end;
if nrow(t1)+nrow(t2)+nrow(nf)+nrow(d1)+nrow(d2)+nrow(r) ^= 6 then do;
  ierr= -3;
  run psa_err(modnam,ierr);
  return(ierr);
end;
if (ceil(nf)-floor(nf))^=0 then do;
  ierr= -4;
  run psa_err(modnam,ierr);
  return(ierr);
end;
if nf < 2 then do;
  ierr= -5;
  run psa_err(modnam,ierr);
  return(ierr);
end;

ierr=0; /* Errorreturn of function */

xmax = sqrt(nf-1); /* Approximative maximum */

/* 1st try with QUAD */

ps_zt1=t1;
ps_zt2=t2;
ps_znf=nf;
ps_zd1=d1;
ps_zd2=d2;
ps_zr =r;

call quad(pret,"PS_FPW2Q",{1e-6 .p}) eps=1e-9 peak=xmax scale=0.1 msg="NO";

if pret^=. then do;
  pow2r=pret;
  ps_tech=1; /* Technique used */
  return(ierr);
end;

```

```

/* QUAD has problems, calculate integral with own technique */
ps_tech=2;                /* Technique used now */

delta=0.1;
if 2#xmax/delta<200 then delta=xmax/100;

flag=0;
count=0;
do until(flag=1);
  count=count+1;
  xvec= do(1e-5,2#xmax,delta);
  yvec= ps_fpw2r(xvec,t1,t2,nf,d1,d2,r);
  nxy=ncol(xvec);

  ymaxind=loc(yvec=max(yvec));
  ymaxind=ymaxind[1];      /* If more than 1 maximum */

  if ymaxind<0.1#nxy | ymaxind>0.9#nxy then do;
    if count>=2 then do;
      if sum(yvec)#delta < 1e-15 then do;      /* Practically no area */
        pow2r=sum(yvec)#delta;                /* 'pseudo' trapeziodal */
      end; else do;
        ierr=-9;                               /* Could not calculate */
        run psa_err(modnam,ierr);
      end;
      return(ierr);
    end;
    if ymaxind<0.1#nxy then xmax=xmax/2; else xmax=2#xmax;
  end; else do;
    flag=1;
  end;
end;

ymax=yvec[ymaxind];

ycrit = ymax/1e8;

xrange=loc(yvec>ycrit);    /* Get range of Y > YCRIT */

goon=1;
if type(xrange)="U" then do;
  goon=0;
end; else do;
  xindlow=xrange[1];
  xindup=xrange[ncol(xrange)];
  if (xindup-xindlow)<5 then goon=0;
end;

if goon=0 then do;
  ierr=-9;                               /* Could not calculate */
  run psa_err(modnam,ierr);
  return(ierr);
end;

xbeg=xvec[xindlow];
xend=xvec[xindup];

if xindup>=(nxy-1) then do;
  flag=0;
  count=0;
  do until(flag=1);
    count=count+1;
    if count>=10 then do;
      ierr=-9;
      run psa_err(modnam,ierr);
      return(ierr);
    end;
    xend=xend#1.2;
    ylast= ps_fpw2r(xend,t1,t2,nf,d1,d2,r);
    if ylast<ycrit then flag=1;
  end;
end;

```

```

ninter=1000;          /* Must be EVEN */
delta=(xend-xbeg)/ninter;
xvec=do(xbeg,xend,delta);
yvec= ps_fpw2r(xvec,t1,t2,nf,d1,d2,r);
nxy=ncol(xvec);

/* Simpsons formula

      delta
INT : ----- [ y0 + yn + 4 (y1 + y3 + ... + yn-1) + 2 (y2 + y4 + ... + yn-2) ]
      3
N even integer: NINTER == N   NXY = NINTER + 1

Index: YVEC[1]==Y   YVEC[nxy]=Y
      0             ninter
*/

iodd=t(do(1,ninter-1,2)); /* odd index Start 1 == y1 == YVEC[2] */
ieven=t(do(2,ninter-2,2)); /* even index Start 2 == y2 == YVEC[3] */

psum =yvec[1]+yvec[ninter+1]+4#sum(yvec[iodd+1])+2#sum(yvec[ieven+1]);
pow2r=psum#delta/3;
return(ierr);
finish;

start ps_fpw2r(xvec,t1,t2,nf,d1,d2,r);
g1=(xvec#t1/sqrt(nf)-d1);
g2=(xvec#t2/sqrt(nf)-d2);
qwer=probbnrm(g1,g2,r); /* 1 row NX columns */

lfak=(nf-1)#log(xvec) - (nf/2-1)#log(2) - lgamma(nf/2) - xvec#xvec/2;
yvec=qwer#exp(lfak);
return(yvec);
finish;

/* Interface module to QUAD */

start ps_fpw2q(xvec) global(ps_cnt,ps_zt1,ps_zt2,ps_znf,ps_zd1,ps_zd2,ps_zr);
ps_cnt=ps_cnt+1;
yvec=ps_fpw2r(xvec,ps_zt1,ps_zt2,ps_znf,ps_zd1,ps_zd2,ps_zr);
return(yvec);
finish;

/*
About PS_FPW2R

Basic procedure received from Frank Bretz (Uni Hannover) (Feb 1998)


$$z = \int_0^{\infty} \frac{BI(x*t1/sf-d1; x*t2/sf-d2) * x^{f-1} * e^{-x*x/2}}{\Gamma(f/2) * 2} dx$$


t1,t2 : Limits to Bi-Normal-Integral
dl,d2 : Noncentrality parameters
f      : Degrees of freedom
sf     = sqrt(f)
r      : Correlation

```

$$BI(a,b) = \frac{1}{2*PI * SQRT(1-r*r)} \int_{-\infty}^a \int_{-\infty}^b \frac{u*u-2*r*u*v+v*v}{2*(1-r*r)} dv du$$

Due to numeric problems we take the logarithm :

$$LFAK = (f-1)*\text{LOG}(x) - (f/2-1)*\text{LOG}(2) - \text{LGAMMA}(f/2) - x*x/2$$

(LGAMMA: Log of Gammafunction which is usable with great f)

 Shape of curve:

```

      *
    *  **
  **      *
****      *****
-----^-----
      Max
  
```

To know the maximum is a great help. Let us determine an approximate maximum.

Consider only following part:

$$y = x^a * e^{-x*x/2} \quad \text{witht } a=f-1$$

$$y' = x^a (-x) e^{-x*x/2} + a x^{a-1} e^{-x*x/2}$$

$$y' = 0 \rightarrow x^{a+1} = a x^{a-1} \rightarrow x^2 = a \rightarrow x = \text{SQRT}(a) = \text{SQRT}(f-1)$$

*/