

Multivariates nichtparametrisches Behrens-Fisher-Problem mit Kovariablen

Dissertation
zur Erlangung des mathematisch-
naturwissenschaftlichen Doktorgrades
“Doctor rerum naturalium”
der Georg-August-Universität Göttingen

vorgelegt von
Antonia Zapf
aus München

Göttingen 2009

D7

Referent: Prof. Dr. Edgar Brunner

Korreferent: Prof. Dr. Axel Munk

Tag der mündlichen Prüfung: 23.10.2009

Danksagung

Die vorliegende Arbeit geht zurück auf einen Themenvorschlag von Herrn Prof. Dr. Edgar Brunner, der mich beim Fortgang der Untersuchung betreut und mir darüber hinaus die Möglichkeit gegeben hat, aktuelle Forschungsergebnisse auf Konferenzen vorzustellen sowie in der Abteilung für Medizinische Statistik wertvolle praktische Erfahrung zu sammeln – dafür sei ihm hiermit herzlich gedankt.

Weiterhin danke ich Herrn Prof. Dr. Axel Munk für die Übernahme des Korreferates und Herrn Prof. Dr. Martin Schlather für die Unterstützung im Rahmen des Promotionsstudiengangs 'Applied Statistics and Empirical Methods'.

Außerdem möchte ich mich ganz herzlich bei Herrn Dr. Frank Konietschke und Frau Katharina Lange für das sorgfältige Korrekturlesen und die stete Bereitschaft zur Diskussion bedanken. Weiterhin danke ich den Kollegen der Abteilung für Medizinische Statistik und der Abteilung für Genetische Epidemiologie für die interessante Zeit in Göttingen.

Mein besonderer Dank gilt meinen Eltern für die stetige Unterstützung und vor allem meinem Mann und meinen Töchtern für die unendliche Geduld.

Inhaltsverzeichnis

Danksagung	v
1 Einleitung	1
1.1 Motivierende Beispiele	1
1.2 Parametrisches Behrens-Fisher-Problem (B-F-P)	3
1.3 Nichtparametrisches B-F-P	6
1.4 Multivariates nichtpar. B-F-P	8
1.5 Multivar. nichtpar. B-F-P mit Kovariablen	9
1.6 Aufbau der Arbeit	10
2 Design und Modell	11
2.1 Anforderungen an das Design	12
2.2 Beispiele	14
3 Relative Effekte	19
3.1 Allgemeines	19
3.2 Schätzung	21
3.3 Skalenalternativen	22
3.4 Adjustierter relativer Effekt	23
4 Punkt- und Intervallschätzer	27
4.1 Schätzer der einzelnen relativen Effekte	27
4.2 Geschätzter adjustierter relativer Effekt	27
4.3 Asymptotische Äquivalenz und multivariate Normalverteilung	29
4.4 Schätzung der Kovarianzmatrix	32
4.5 Schätzung der Regressionskoeffizienten	33
4.6 Beobachtbarer Punktschätzer	35
4.7 Beobachtbare Schätzung der Kovarianzmatrix	36
4.8 Konfidenzintervall	37
4.8.1 Normalapproximation	37
4.8.2 t-Approximation	37
4.9 Beispiele	38

5	Transformationsmethode	43
5.1	Transformation des Punktschätzers	43
5.2	Transformation der Konfidenzintervalle	45
5.3	Logit-Transformation	48
5.4	Beispiele	50
6	Hypothesentests	53
6.1	Teststatistiken	53
6.2	Test auf Vorliegen eines Effekts	54
6.3	Tests auf gleiche Effekte	54
6.4	Beispiele	56
7	Simulationen	59
8	Diskussion	65
9	Ausblick	67
9.1	Fehlende Werte	67
9.2	Multiple Kontrasttests und simultane Konfidenzintervalle	70
A	Makro	73
B	Allgemeine Bezeichnungen	75
B.1	Vordefinierte Matrizen	75
B.2	Zählfunktion	76
C	Beweise	77
C.1	Beweis zu Theorem 4.2 auf Seite 28	77
C.2	Beweis zu Satz 4.5 auf Seite 29	78
C.3	Beweis zu Theorem 4.9 auf Seite 32	80
	Literaturverzeichnis	83

Abbildungsverzeichnis

1.1	<i>Meningitis-Studie, Boxplots des Alters.</i>	2
1.2	<i>CAD-Studie, Boxplots der Befundungsdauer für die drei Untersucher und für die beiden Statusgruppen.</i>	3
1.3	<i>CAD-Studie, Boxplots des Alters.</i>	4
1.4	<i>ROC-Kurve, optimaler Fall.</i>	4
1.5	<i>ROC-Kurve, schlechtester Fall.</i>	5
1.6	<i>ROC-Kurve, realistischer Fall.</i>	5
1.7	<i>CAD-Studie, Verteilung der Zielgröße in den beiden Gruppen.</i>	7
1.8	<i>Meningitis, Verteilung des Laborparameters A in den beiden Gruppen.</i>	8
3.1	<i>Stochastische Tendenz.</i>	20
3.2	<i>Skalenalternative.</i>	23
4.1	<i>Meningitis-Studie, Konfidenzintervalle.</i>	39
4.2	<i>CAD-Studie, Konfidenzintervalle.</i>	41
5.1	<i>Meningitis-Studie, transformierte Konfidenzintervalle.</i>	50
5.2	<i>CAD-Studie, transformierte Konfidenzintervalle.</i>	52
7.1	<i>Vergleich der Teststatistiken.</i>	60
7.2	<i>Variation der Stichprobenumfänge.</i>	61
7.3	<i>Variation der Bedingungen.</i>	62
7.4	<i>Variation der Anzahl der Kovariablen.</i>	63
7.5	<i>Vergleich von Shift der Zielvariable und Shift der Kovariable.</i>	64

Tabellenverzeichnis

1.1	<i>CAD-Studie, Häufigkeitstabelle des Scores.</i>	9
2.1	<i>Kombination von zwei Bedingungen.</i>	11
2.2	<i>Design für eine Bedingung ($d = 1$).</i>	13
2.3	<i>Design für mehrere Bedingungen ($d > 1$).</i>	13
2.4	<i>Design des Meningitis-Beispiels.</i>	16
2.5	<i>Design des CAD-Beispiels.</i>	17
4.1	<i>Meningitis-Beispiel, geschätzte relative Effekte.</i>	38
4.2	<i>CAD-Studie, Vergleich der geschätzten relativen Effekte</i>	40
5.1	<i>Meningitis-Studie, Breite der transformierten Konfidenzintervalle.</i>	51
5.2	<i>CAD-Studie, Vergleich der untransformierten und transformierten geschätzten relativen Effekte.</i>	51
5.3	<i>CAD-Studie, Breite der transformierten Konfidenzintervalle.</i>	52
6.1	<i>CAD-Studie, p-Werte der Tests auf Gleichheit der Effekte.</i>	57
7.1	<i>Variation der Stichprobenumfänge, balanciert.</i>	59
7.2	<i>Variation des Stichprobenumfangs n_2 bei $n_1 = 20$.</i>	61
7.3	<i>Variation der Anzahl der Bedingungen.</i>	62
7.4	<i>Variation der Anzahl der Kovariablen.</i>	63
7.5	<i>Modifikation der Verteilung der Kovariablen.</i>	64
9.1	<i>Einfluss rein zufällig fehlender Werte auf den geschätzten relativen Effekt anhand des Meningitis-Beispiels.</i>	69

1 Einleitung

Im ersten Abschnitt werden zwei motivierende Beispiele vorgestellt, anschließend wird der Leser in den Abschnitten 1.2 bis 1.5 schrittweise an das multivariate nichtparametrische Behrens-Fisher-Problem mit Kovariablen herangeführt. Im Abschnitt 1.6 wird dann der Aufbau der Arbeit angegeben.

1.1 Motivierende Beispiele

Im Folgenden werden zwei Beispiele vorgestellt. In den jeweiligen Kapiteln wird die Methodik dann auf die beiden Datensätze angewendet und die Ergebnisse werden diskutiert.

Meningitis-Studie¹

Bei dieser Studie geht es um die Unterscheidung zwischen viraler und bakterieller Meningitis, die klinisch die gleiche Symptomatik haben. Die beiden Formen benötigen jedoch eine unterschiedliche Behandlung, vor allem ist bei der hoch letalen bakteriellen Meningitis eine sehr rasche antibiotische Behandlung nötig. Der Goldstandard, der später mikrobiologisch festgestellt wurde, ist hier entsprechend virale bzw. bakterielle Meningitis. Verschiedene Parameter wurden auf ihre Fähigkeit hin untersucht, die beiden Meningitis-Varianten zu differenzieren. Da die Ergebnisse noch nicht publiziert sind, können die Parameter nicht benannt werden, wir werden sie deshalb mit Laborparameter A, B und C bezeichnen. Eine Randomisierung nach Erkrankung war hier natürlich nicht möglich und bei der Auswertung der Daten stellte sich heraus, dass sich die Altersstruktur in den beiden Gruppen stark unterscheidet (siehe Abbildung 1.1). Das Durchschnittsalter in der viralen Gruppe lag bei 38 Jahren, in der bakteriellen Gruppe bei 63 Jahren. Es kann also beispielsweise sein, dass ein Laborparameter, der mit dem Alter hoch korreliert ist, fälschlicherweise als zur Differenzierung geeignet beurteilt wird.

¹Die Daten dieser Studie wurden uns freundlicherweise von Herrn PD Dr. med. Holger Schmidt aus der Abteilung Neurologie der Universität Göttingen zur Verfügung gestellt.

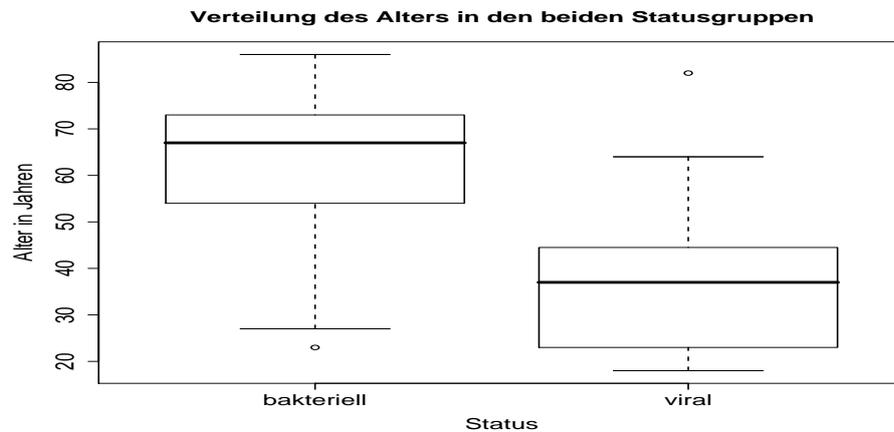


Abbildung 1.1: Meningitis-Studie, Boxplots des Alters.

CAD-Studie²

In dieser Studie wurden 198 Frauen auf Brustkrebs hin untersucht. Davon hatten 95 Frauen Brustkrebs und 103 Frauen nicht, der Goldstandard ist histologisch gesichert. Der Befund wurde von drei Untersuchern mit einer Mammographie erhoben, einmal mit und einmal ohne Unterstützung durch das CAD-System (Computer Aided Design). Dabei war die Zielgröße ein Score gemäß der BI-RADS³ Klassifikation mit folgender Aufschlüsselung:

1. Kategorie: kein Tumor
2. Kategorie: gutartiger Tumor
3. Kategorie: wahrscheinlich gutartiger Tumor
4. Kategorie: vermutlich bösartiger Tumor
5. Kategorie: sehr wahrscheinlich bösartiger Tumor

Das Problem bei dieser Studie war, dass die drei Untersucher unterschiedlich viel Erfahrung hatten, es handelte sich nämlich um einen Studenten ohne Mammographie-Erfahrung, einen Assistenzarzt mit etwa sechs Monaten Mammographie-Erfahrung und einen Facharzt mit mehrjähriger Mammographie-Erfahrung. Zusätzlich zu dem

²Diese Studie stammt aus der Abteilung Diagnostische Radiologie der Universität Göttingen, die Daten wurden uns freundlicherweise von Frau PD Dr. med. Silvia Obenauer zur Verfügung gestellt.

³Breast-Imaging-Reporting-and-Data-System.

Score wurde auch die Befundungszeit pro Untersucher pro Methode pro Bild festgehalten. Wie man in Abbildung 1.2 sehen kann, unterscheidet sich die benötigte Zeit bei den Untersuchern deutlich, bei den Statusgruppen im Schnitt um 10 Minuten. Nun

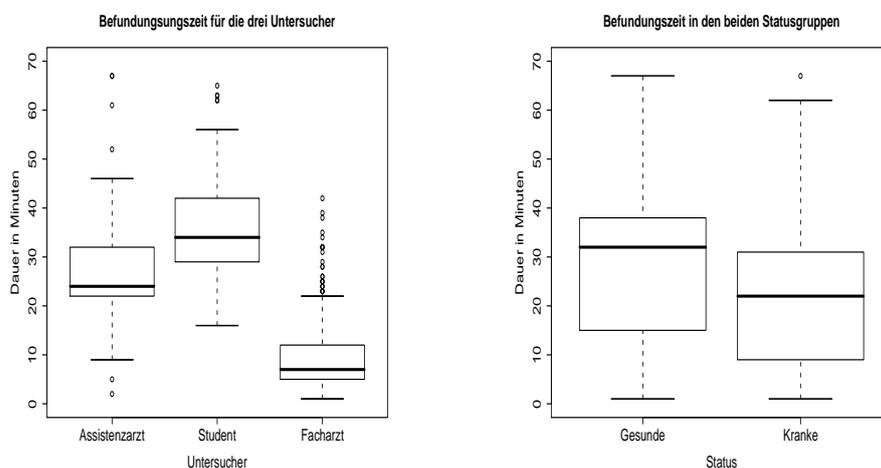


Abbildung 1.2: CAD-Studie, Boxplots der Befundungsdauer für die drei Untersucher und für die beiden Statusgruppen.

stellt sich die Frage, ob die Zeit, die der Untersucher zur Befundung benötigt, einen Einfluss auf die Sicherheit der Ergebnisse hat, bzw. ob das CAD unerfahrenen Ärzten die Diagnose erleichtert.

Zusätzlich war auch noch das Alter der Frauen bekannt, allerdings war hier der Unterschied in den Verteilungen der beiden Statusgruppen gering (siehe Abbildung 1.3).

1.2 Parametrisches Behrens-Fisher-Problem

Das *parametrische* Behrens-Fisher-Problem beschreibt die Schwierigkeit, die Erwartungswerte zweier Stichproben zu vergleichen, die normalverteilt $N(\mu_i, \sigma_i^2)$ sind, wobei die Varianzen unbekannt sind und unterschiedlich sein können. Man testet also die Hypothese $H_0^\mu : \mu_1 = \mu_2$ unter der Annahme σ_1^2, σ_2^2 beliebig. In der Medizin tritt dieses Problem häufig auf, da beispielsweise ein Laborparameter bei kranken Personen im Grenzbereich oftmals weiter gestreut ist als bei Gesunden, bei denen er im Normbereich liegt. Der t-Test ist in diesen Fällen liberal, da Varianzeffekte fälschlicherweise als Verschiebungseffekte interpretiert werden können.

Von Behrens (1929) [5] und Fisher (1936) [21] stammten die ersten Lösungsansätze dieser Problematik, die weiter modifiziert wurden. Heute wird in diesen Fällen hauptsächlich der approximative t-Test verwendet, der auf den Arbeiten von Smith (1936)

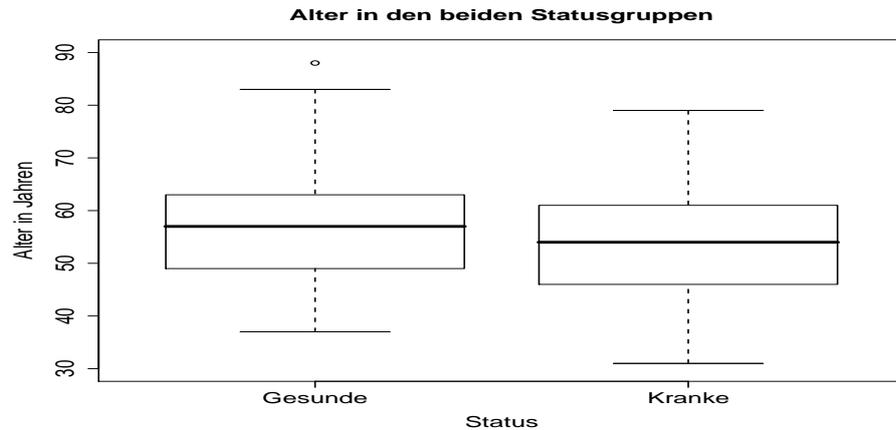


Abbildung 1.3: CAD-Studie, Boxplots des Alters.

[37], Welch (1937) [40] und Satterthwaite (1946) [34] basiert.

Eine Methode, die Trennschärfe zu beurteilen, also wie gut mit einem Verfahren zwei Gruppen unterschieden werden können, ist die Receiver Operating Characteristic Curve (ROC-Kurve). Hierbei wird für jeden möglichen cutpoint Sensitivität und 1–Spezifität berechnet und gegeneinander aufgetragen. Die Fläche unter dieser Kurve, die sogenannte Area Under the Curve ($AUC \in [0,1]$), ist ein Maß für die Trennschärfe und entspricht dem relativen Effekt p (siehe Abschnitt 3.1). Zu beachten ist, dass der relative Effekt nur dann geschätzt werden kann, wenn neben dem Ergebnis des Tests auch der wahre Zustand, der sogenannte Goldstandard, vorliegt.

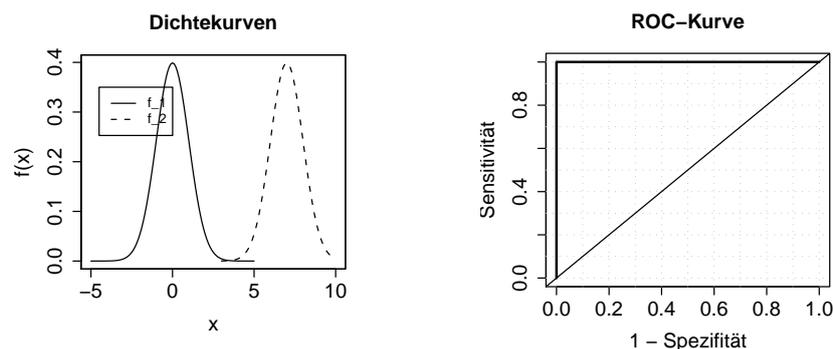


Abbildung 1.4: ROC-Kurve, optimaler Fall.

In den Grafiken 1.4 bis 1.6 sind für drei Beispiele die Dichtefunktionen und zugehörige

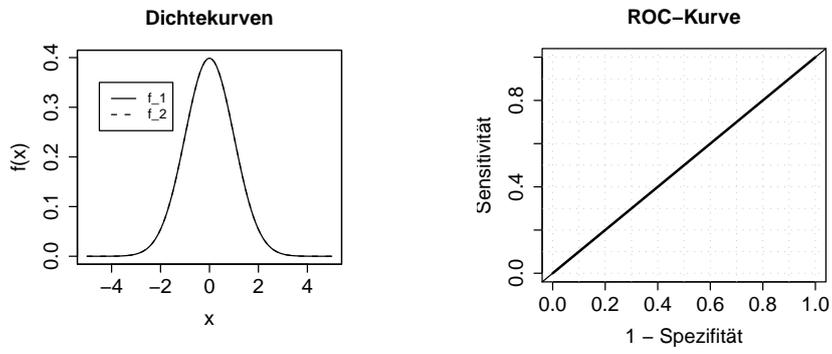


Abbildung 1.5: ROC-Kurve, schlechtester Fall.

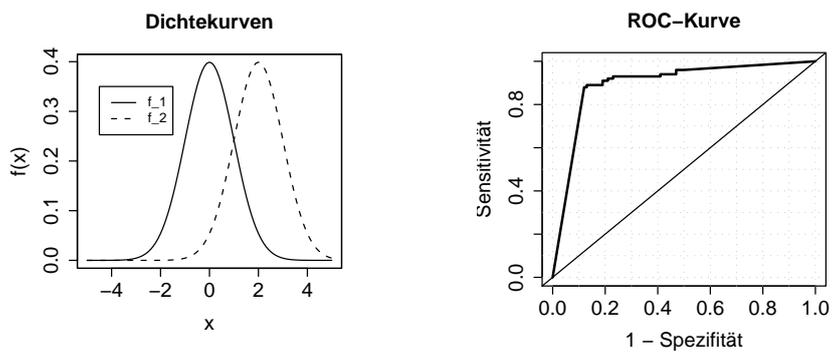


Abbildung 1.6: ROC-Kurve, realistischer Fall.

gen ROC-Kurven gezeichnet. Der optimale Fall, das bedeutet die 100-prozentig sichere Diagnose ist in Grafik 1.4 dargestellt, die zwei Dichtekurven sind komplett getrennt und die AUC ist gleich 1. Dieser Fall ist in der Realität allerdings kaum zu beobachten. Im schlechtesten, ebenfalls eher unrealistischen Fall (Abbildung 1.5), sind die Dichtefunktionen der zwei Gruppen exakt gleich. Das Testergebnis ist dann so unzuverlässig wie eine rein zufällige Entscheidung, die ROC-Kurve verläuft entlang der Winkelhalbierenden und die AUC ist gleich 0.5. Grafik 1.6 gibt realistische Ergebnisse eines guten diagnostischen Verfahrens wieder. Die AUC liegt hier bei 0.89, d.h. die zwei Stichproben lassen sich gut unterscheiden.

Ein typisches Beispiel für den Zwei-Stichproben-Fall sind Diagnosestudien. Bei dieser Art klinischer Studien geht es darum, ein diagnostisches Verfahren daraufhin zu überprüfen, wie gut es gesunde von kranken Personen unterscheiden kann. Ein Beispiel hierfür ist die vorgestellte CAD-Studie, bei der die Diagnose von Brustkrebs mit und ohne Computerunterstützung verglichen wurde. Genauso können die beiden Stichproben aber auch zwei Krankheiten mit dem selben Symptom-Komplex sein, die miteinander verglichen werden sollen. Hierfür ist die Meningitis-Studie ein Beispiel, bei der anhand verschiedener Laborparameter zwischen viraler und bakterieller Meningitis unterschieden werden soll. Hier sind die Kollektive also nicht, wie bei der CAD-Studie, gesunde und kranke Personen, sondern Patienten mit bakterieller und viraler Meningitis. Und so gibt es viele Studien, bei denen es um Unterschiede zwischen zwei Stichproben geht. In dem Meningitis-Beispiel sind die Zielgrößen Parameter, die metrisch und unter Umständen auch normalverteilt sind, allerdings wahrscheinlich in den beiden Gruppen nicht die gleiche Varianz aufweisen. Es tritt dann das parametrische Behrens-Fisher-Problem auf.

1.3 Nichtparametrisches Behrens-Fisher-Problem

Oftmals jedoch ist die Annahme einer Normalverteilung nicht gerechtfertigt, oder aber es kann gar keine parametrische Verteilung angenommen werden. Das nichtparametrische Analogon zum t-Test ist der Wilcoxon-Mann-Whitney-Test (Wilcoxon (1945) [41] und Mann und Whitney (1947) [29]). Bei diesem Test wird auf Gleichheit der Verteilungsfunktionen und damit auf gleiche Varianzen getestet. Die Nullhypothese lautet also $H_0^F : F_1 = F_2$. Im Fall gleicher Varianzen entspricht die nichtparametrische Hypothese H_0^F der parametrischen H_0^μ , im heteroskedastischen Normalverteilungsmodell gilt diese Entsprechung dagegen nicht.

Fligner und Policello (1981) [22] und Brunner und Neumann (1996) [12] entwickelten Tests, die heteroskedastische Verteilungen erlauben, allerdings wurde die Stetigkeit der Daten vorausgesetzt, insbesondere durften keine Bindungen auftreten. Tests

zur Lösung des nichtparametrischen Behrens-Fisher-Problems für große Stichproben stammen unter anderem von Brunner und Puri (1996) [15]. Bei kleinen Stichproben sind Approximationen notwendig, wie sie beispielsweise Brunner und Munzel (2000) [10] angegeben haben. Eine Adjustierung bezüglich Kovariablen wurde hier aber nicht ausgearbeitet.

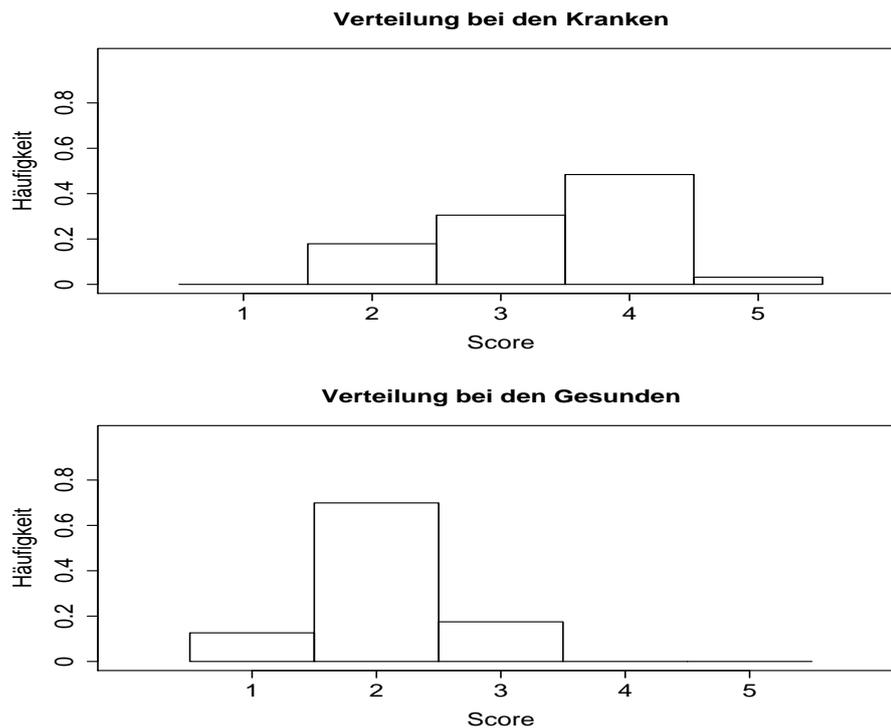


Abbildung 1.7: CAD-Studie, Verteilung der Zielgröße in den beiden Gruppen.

Ein Beispiel für den nichtparametrischen Fall ist wieder die CAD-Studie, bei der die Zielgröße ein Score von 1 bis 5 ist, der die Wahrscheinlichkeit für ein Karzinom wiedergibt. Dessen Verteilung in den beiden Gruppen ist in Abbildung 1.7 dargestellt. Da der Score rein ordinal ist, sind parametrische Verfahren nicht mehr anwendbar.

In Abbildung 1.8 ist zur Veranschaulichung des nichtparametrischen Behrens-Fisher-Problems die Verteilung des Laborparameters A der Meningitis-Studie dargestellt.⁴

⁴Dabei wurde die x-Achse bei 100 abgebrochen, der höchste Wert in der bakteriellen Gruppe war 481.

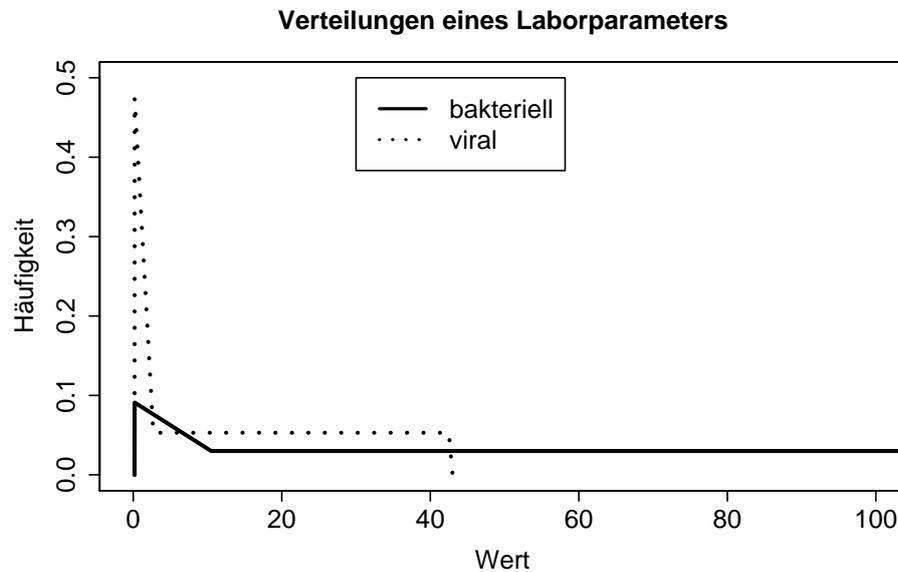


Abbildung 1.8: *Meningitis, Verteilung des Laborparameters A in den beiden Gruppen.*

1.4 Multivariates nichtparametrisches Behrens-Fisher-Problem

Gerade bei der Bewertung uneindeutiger Befunde, wie z.B. bei bildgebenden Verfahren, motorischen Tests oder individuellen Schwellenwerten, entscheidet man sich oftmals für ein Studiendesign mit mehreren Untersuchern (Multi-Reader-Design). Bei Zulassungsverfahren wird von den Regulierungsbehörden im Allgemeinen auch gefordert, dass mindestens zwei, besser drei Untersucher die Ergebnisse beurteilen. Dieses Multi-Reader-Design soll zu genaueren und valideren Ergebnissen führen. Auch kann es vorkommen, dass verschiedene Diagnoseverfahren verglichen werden. Bei dem Beispiel der Meningitis-Studie sind es die verschiedenen Laborparameter, die von Interesse sind (Multi-Methoden-Design). Wenn man die verschiedenen Methoden zusätzlich von mehreren Untersuchern beurteilen lässt, hat man das häufig verwendete Multi-Methoden-Multi-Reader-Design. Für jeden Patienten liegt die Zielgröße also für unterschiedliche Bedingungen vor. Bei der CAD-Studie z.B. wurden die Ergebnisse der zwei Methoden jeweils von drei Untersuchern beurteilt (siehe Tabelle 1.1). Es liegt also jeweils das multivariate nichtparametrische Behrens-Fisher-Problem vor.

Tabelle 1.1: CAD-Studie, Häufigkeitstabelle des Scores.

Kombination		Score der Kranken					Score der Gesunden				
Methode	Untersucher	1	2	3	4	5	1	2	3	4	5
1	1	0	17	29	46	3	13	72	18	0	0
	2	0	10	31	49	5	4	55	35	9	0
	3	0	2	13	33	47	32	62	9	0	0
2	1	0	22	35	34	4	13	72	18	0	0
	2	0	22	46	27	0	8	60	28	7	0
	3	0	7	9	32	47	33	64	5	1	0

1.5 Multivariates nichtparametrisches Behrens-Fisher-Problem mit Kovariablen

In jeglichen Studien sollte nach Kovariablen, von denen ein Einfluss auf die Zielgröße bekannt ist oder vermutet wird, geschichtet werden. Innerhalb dieser Schichten sollte anschließend randomisiert werden. Bei medizinischen Studien sind typische Kovariablen z.B. Alter und Geschlecht. Es kann jedoch vorkommen, dass die Randomisierung aus verschiedenen Gründen unterlassen wird, dass sie ungenügend ist (z.B. aufgrund fehlenden Wissens über den Effekt verschiedener Variablen), oder dass es gar nicht möglich ist zu randomisieren, weil beispielsweise zwei Krankheitsgruppen mit sehr geringer Prävalenz unterschieden werden sollen. In all diesen Fällen kann der unberücksichtigte Einfluss von Kovariablen dazu führen, dass die Ergebnisse verfälscht sind, siehe hierfür z.B. Senn (2004) [35]. Ein Spezialfall ist die Berücksichtigung von Baselinewerten, hier wird oftmals vorgeschlagen, die Differenz oder den Quotienten von Ausgangs- und Endwert zu verwenden. Allerdings ist leicht ersichtlich, dass beispielsweise ein Temperaturanstieg von $2\text{ }C^{\circ}$ bei einer anfänglichen Temperatur von $36.5\text{ }C^{\circ}$ eine andere Bedeutung hat als bei einem Startwert von $39\text{ }C^{\circ}$. Wegen dieser Problematik wird z.B. von Senn eine Adjustierung angeraten. Wenn die Zielgröße nicht normalverteilt bzw. nicht einmal metrisch ist, wie z.B. der Schweregrad einer Krankheit, verbietet sich jedoch die Verwendung linearer Modelle und es müssen nichtparametrische Methoden verwendet werden. Unter anderem haben Bathke und Brunner (2003) [3] und Brumback et al. [8] für bestimmte Modelle nichtparametrische Verfahren um den Einfluss von Kovariablen erweitert.

Das Ziel dieser Arbeit ist, die Lösungsansätze, die es für das multivariate nichtparametrische Behrens-Fisher-Problem bereits gibt (siehe beispielsweise Brunner et al.

(2002) [11]), um die Berücksichtigung von Kovariablen zu erweitern. Dabei sollen sowohl Punkt- und Intervallschätzer als auch Hypothesentests entwickelt werden.

1.6 Aufbau der Arbeit

Zunächst wird im zweiten Kapitel das untersuchte Design und das dazugehörige statistische Modell vorgestellt. Anschließend wird, nach einer kurzen Einführung in die Grundlagen der relativen Effekte, der adjustierte relative Effekt hergeleitet. Das vierte Kapitel widmet sich dem Punkt- und Intervallschätzer des bezüglich Kovariablen adjustierten relativen Effekts. Da diese Schätzer allerdings nicht mehr bereichserhaltend sind, wird im fünften Kapitel die Transformationsmethode erläutert. Im sechsten Kapitel werden dann zwei verschiedene Teststatistiken dargestellt und auf Globaltests angewendet. Nachdem im siebten Kapitel die beiden Testverfahren bei Niveau- und Powersimulationen verglichen und verschiedene Einflüsse untersucht werden, folgt im achten Kapitel eine Zusammenfassung und Diskussion der Ergebnisse. Im neunten Kapitel wird abschließend noch ein Ausblick bezüglich fehlender Werte und multipler Kontrasttests bzw. simultaner Konfidenzintervalle gegeben.

Die Anwendung der Verfahren auf die beiden vorgestellten Beispiele erfolgt jeweils am Ende eines Kapitels. Das Computerprogramm und seine Anwendung werden im Anhang A erläutert. Im Anhang B sind allgemeine Bezeichnungen und vordefinierte Matrizen angegeben. Für einen besseren Lesefluss sind längere Beweise in den Anhang C gestellt worden, im Text wird dann auf die entsprechende Seite verwiesen.

2 Design und Modell

Vom Design her haben wir zwei unabhängige Statusgruppen $i = 1, 2$ mit jeweils n_i Individuen, wobei jedes Individuum unter d Bedingungen beobachtet wird. Die Gesamt-Fallzahl ist dann $N = n_1 + n_2$.

Wie in der Einleitung beschrieben, kann es zum Beispiel von Interesse sein, verschiedene Untersuchungsmethoden zu vergleichen und/oder die Ergebnisse von mehreren Untersuchern beurteilen zu lassen. Es sind auch mehr als zwei Faktoren möglich, die dann nur sinnvoll zusammengefasst werden müssen. Man erhält dann für jede Bedingung ℓ eine ROC-Kurve und als Schätzer für die AUC einen relativen Effekt $p^{(\ell)}$ (siehe Abschnitt 1.2). Voraussetzung ist jedoch, dass jedes Individuum unter denselben Bedingungen beobachtet wird. Ist dem nicht so, stellt sich das Problem fehlender Werte, das in Kapitel 9.1 behandelt wird.

In dieser Arbeit gehen wir beispielhaft von $m = 1, \dots, M$ Methoden und $r = 1, \dots, R$ Readern (Untersuchern) aus, deren Kombination zu einem Faktor $\ell = 1, \dots, d$ zusammengefasst wird, um die Anzahl der Indizes überschaubar zu halten (siehe Tabelle 2.1). Erst im Abschnitt 6.3 (Globaltests auf gleiche Effekte) wird der Faktor wieder aufgetrennt, um die Einzeleffekte Methode und Reader untersuchen zu können.

Tabelle 2.1: *Kombination von zwei Bedingungen.*

Methode m	1	1	1	...	M	M	M
Untersucher r	1	...	R	...	1	...	R
Kombination ℓ	1	...	R	...	$(M - 1) \cdot R + 1$...	$M \cdot R = d$

In diesem Modell werden auch die Kovariablen unter d Bedingungen beobachtet. So ist beispielsweise in der CAD-Studie für jeden Untersucher und für jede Methode die Befundungszeit pro Bild erhoben worden. Bei Kovariablen, die unabhängig von der Bedingung sind, z.B. Alter, gilt das Modell genauso, dann sind die Werte entsprechend für jede Bedingung gleich.

Als statistisches Modell ist also für jedes Individuum s der Statusgruppe i folgender

Beobachtungsvektor gegeben:

$$\mathbf{w}_{is} = \left(\mathbf{w}_{is}^{(1)'}, \dots, \mathbf{w}_{is}^{(d)'} \right)' \quad \text{mit } i = 1, 2 \text{ und } s = 1, \dots, n_i,$$

dabei bezeichnet $\ell = 1, \dots, d$ die Bedingung. Für jede Bedingung ℓ setzt sich der Vektor wiederum aus den Beobachtungen der Zielvariable und der Kovariable(n) zusammen. Es gilt also:

$$\mathbf{w}_{is}^{(\ell)} = \left(X_{0,is}^{(\ell)}, X_{1,is}^{(\ell)}, \dots, X_{k,is}^{(\ell)} \right)',$$

wobei $X_{j,is}^{(\ell)}$ für $j = 0$ der Wert der Zielgröße und für $j = 1, \dots, k$ der Wert der j -ten Kovariable ist. Die Komponenten der Vektoren $\mathbf{w}_{is}^{(\ell)}$ bzw. \mathbf{w}_{is} sind abhängig, da sie zu einem Individuum gehören. Die einzelnen Individuen und die beiden Statusgruppen sind dagegen voneinander unabhängig. Für die Verteilung der einzelnen Beobachtungen gilt:

$$X_{j,is}^{(\ell)} \sim F_{j,i}^{(\ell)}(x) \quad \text{mit } j = 0, \dots, k.$$

Mit der marginalen Verteilung $F_{j,i}^{(\ell)}(x)$ ist die normalisierte Verteilungsfunktion (Ruymgaart, 1980, [33]) gemeint, die folgendermaßen definiert ist:

Definition 2.1 (Normalisierte Verteilungsfunktion)

$$F_{j,i}^{(\ell)}(x) = P \left(X_{j,is}^{(\ell)} < x \right) + \frac{1}{2} P \left(X_{j,is}^{(\ell)} = x \right).$$

Es sind alle stetigen und unstetigen Verteilungen zugelassen. Der einzige Fall, der ausgeschlossen wird, ist der, dass alle $F_{j,i}^{(\ell)}(x)$ Ein-Punkt-Verteilungen sind. Es sind also auch ordinale und binäre Variablen erlaubt, die dabei auftretenden Bindungen stellen kein Problem dar.

Der einfachste Fall des Designs mit nur einer Bedingung ist in Tabelle 2.2 zu sehen, das allgemeine Design ist in Tabelle 2.3 dargestellt.

2.1 Anforderungen an das Design

Für die Ergebnisse der folgenden Kapitel sind sowohl Bedingungen an die Verteilungen als auch an die Stichprobenumfänge nötig. Unter anderem gelten die Eigenschaften von Schätzern und Testverfahren nur asymptotisch, benötigen also hinreichend große

Tabelle 2.2: Design für eine Bedingung ($d = 1$).

Status	Individuum	Zielvariable	Kovariablen		
1	1	$X_{0,11}$	$X_{1,11}$	\dots	$X_{k,11}$
	\vdots		\vdots		
	n_1	$X_{0,1n_1}$	$X_{1,1n_1}$	\dots	$X_{k,1n_1}$
2	1	$X_{0,21}$	$X_{1,21}$	\dots	$X_{k,21}$
	\vdots		\vdots		
	n_2	$X_{0,2n_2}$	$X_{1,2n_2}$	\dots	$X_{k,2n_2}$

Tabelle 2.3: Design für mehrere Bedingungen ($d > 1$).

Bedingung		1	\dots	d
Status	Individuum			
1	1	$\mathbf{w}_{11}^{(1)}$	\dots	$\mathbf{w}_{11}^{(d)}$
	\vdots		\vdots	
	n_1	$\mathbf{w}_{1n_1}^{(1)}$	\dots	$\mathbf{w}_{1n_1}^{(d)}$
2	1	$\mathbf{w}_{21}^{(1)}$	\dots	$\mathbf{w}_{21}^{(d)}$
	\vdots		\vdots	
	n_2	$\mathbf{w}_{2n_2}^{(1)}$	\dots	$\mathbf{w}_{2n_2}^{(d)}$

Stichprobenumfänge. Dementsprechend wird an verschiedenen Stellen im Text auf folgende Annahmen verwiesen:

(A1) $X_{j,is}^{(\ell)}, X_{j,is'}^{(\ell)} \sim F_{j,i}^{(\ell)}(x)$ mit $i = 1, 2$ und $s, s' = 1, \dots, n_i$,

(A2) $N \rightarrow \infty$, so dass $\frac{N}{n_i} \leq N_0 < \infty$ und

(A3) $\text{Var}(F_{j,1}^{(\ell)}(X_{j,2s}^{(\ell)})) > 0$ und $\text{Var}(F_{j,2}^{(\ell)}(X_{j,1s}^{(\ell)})) > 0$ für $j = 0, \dots, k$ und $\ell = 1, \dots, d$.

Das bedeutet, dass bei Studien mit dem entsprechenden Versuchsplan folgende Annahmen erfüllt sein müssen:

(A1) Die Verteilung darf nur von der Statusgruppe, der Ziel- bzw. Kovariable sowie der Bedingung, nicht aber vom Individuum abhängen.

(A2) Der gesamte Stichprobenumfang muss hinreichend groß sein. Dabei hängt die konkrete Größe von der Anzahl der Bedingungen und der Balanciertheit ab. D.h. je mehr Bedingungen vorliegen und je unbalancierter die Stichprobenumfänge der zwei Statusgruppen sind, desto größer muss die Fallzahl sein (siehe Kapitel 7). Die Stichprobenumfänge der beiden Statusgruppen dürfen auch insgesamt nicht zu unbalanciert sein, auch hier ist jedoch kein in jeder Konstellation gültiges Verhältnis anzugeben. In Kapitel 7 werden die Simulationsergebnisse hierzu diskutiert.

2.2 Beispiele

Meningitis-Studie

Das Design der Meningitis-Studie ist in Tabelle 2.4 dargestellt. Das statistische Modell ist für $i = (\text{bakteriell, viral})$ und $s = 1, \dots, n_i$ mit $n_{\text{bakteriell}} = 33$ und $n_{\text{viral}} = 19$:

$$\mathbf{w}_{is} = \left(\mathbf{w}_{is}^{(1)'}, \mathbf{w}_{is}^{(2)'}, \mathbf{w}_{is}^{(3)'} \right)' .$$

Dabei bezeichnet $\ell = (1,2,3)$ die Laborparameter (A,B,C). Weiterhin gilt für jeden Laborparameter ℓ :

$$\mathbf{w}_{is}^{(\ell)} = \left(X_{0,is}^{(\ell)}, X_{1,is}^{(\ell)} \right)'$$

wobei $k = 0$ für die Zielgröße und $k = 1$ für das Alter als Kovariable steht. Die Kovariable ist in diesem Beispiel also unabhängig von der Bedingung.

CAD-Studie

Für die CAD-Studie ist das Design auszugsweise in Tabelle 2.5 abgebildet. Das dazugehörige statistische Modell lautet:

$$\mathbf{w}_{is} = \left(\mathbf{w}_{is}^{(1)'}, \dots, \mathbf{w}_{is}^{(6)'} \right)' .$$

Dabei ist $i = (\text{krank, gesund})$ und $s = 1, \dots, n_i$ mit $n_{\text{k}} = 95$ und $n_{\text{g}} = 103$. Weiterhin setzen sich die Methoden $m = 1,2$ und Untersucher $r = 1,2,3$ zu der Kombination $\ell = (1,1), (1,2), \dots, (2,3) = 1, \dots, 6$ zusammen, und es gilt für jede Kombination ℓ :

$$\mathbf{w}_{is}^{(\ell)} = \left(X_{0,is}^{(\ell)}, X_{1,is}^{(\ell)}, X_{2,is}^{(\ell)} \right)'$$

wobei $k = 0$ für die Zielgröße, $k = 1$ für das Alter als Kovariable und $k = 2$ für die Befundungsdauer als Kovariable steht. Damit ist das Alter für alle Bedingungen gleich, die Befundungsdauer dagegen für alle unterschiedlich.

Tabelle 2.4: Design des Meningitis-Beispiels.

bakteriell					viral				
Patient	A	B	C	Alter	Patient	A	B	C	Alter
1	0.19	37.20	60.00	27	1	42.50	41.00	60.00	19
2	0.19	36.10	60.00	39	2	0.19	37.80	204.08	34
3	329.80	38.00	60.00	37	3	0.19	36.40	60.00	18
5	132.50	40.00	1395.00	48	5	0.19	38.20	163.00	38
6	481.00	36.50	127.00	23	6	0.19	36.60	89.00	44
7	80.00	40.00	60.00	61	7	0.19	37.00	60.00	24
8	265.10	38.70	60.00	80	8	6.10	37.30	100.00	39
9	138.70	38.50	60.00	43	9	10.50	36.40	189.00	82
10	140.70	40.00	60.00	60	10	0.19	40.00	116.00	33
11	214.70	37.20	316.00	67	11	2.70	38.70	76.00	22
12	182.50	40.40	171.00	78	12	5.70	38.10	230.00	20
13	303.60	38.20	510.00	62	13	5.80	37.00	16.65	45
14	28.56	38.40	60.00	86	14	0.19	36.80	290.00	49
15	181.70	37.20	60.00	61	15	0.19	38.50	60.00	41
16	277.40	38.90	117.00	54	16	11.90	36.40	50.24	37
17	14.00	38.60	60.00	76	17	3.50	37.20	55.52	34
18	151.20	36.90	60.00	80	18	11.80	39.80	151.00	64
19	126.00	40.60	60.00	69	19	3.00	36.80	60.00	64
20	388.70	38.20	60.00	55					
21	125.20	38.10	121.00	76					
22	112.60	39.40	154.00	76					
23	290.30	37.00	60.00	61					
24	213.00	38.80	100.00	67					
25	321.10	36.60	60.00	77					
26	168.60	38.00	60.00	69					
27	311.00	36.70	109.00	71					
28	10.50	39.50	2791.00	37					
29	122.90	40.40	158.00	67					
30	198.30	38.80	98.00	73					
31	179.90	40.00	60.00	56					
32	161.00	39.00	60.00	42					
33	0.19	38.00	465.00	71					

Tabelle 2.5: *Design des CAD-Beispiels.*

Status	ID	$m = 1, r = 1$		$m = 1, r = 2$...	$m = 2, r = 3$		Alter
		Score	Zeit	Score	Zeit		Score	Zeit	
gesund	1	2	32	4	42	...	3	9	44
	2	2	15	3	38		2	6	43
	3	2	17	3	43		1	4	63
	4	3	32	3	53		2	7	55
	⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮
	100	1	22	3	32		2	3	62
	101	2	19	2	26		2	2	59
	102	1	13	2	32		2	5	59
	103	2	15	2	26		3	6	43
krank	1	4	61	4	34	...	5	5	46
	2	3	15	4	65		4	7	53
	3	4	34	4	62		5	2	58
	4	4	67	4	52		5	4	52
	⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮
	92	2	22	3	41		5	9	54
	93	3	24	4	44		4	9	55
	94	3	23	4	44		4	19	49
	95	4	42	3	36		5	7	62

3 Relative Effekte

3.1 Allgemeines

In diesem Abschnitt soll ein Überblick über die relativen Effekte gegeben werden, für eine genauere Einführung siehe z.B. Brunner und Munzel (2002) [11].

Der relative Effekt ist, wie schon in der Einleitung erwähnt, ein Maß für den Unterschied zwischen Verteilungen. Im hier untersuchten Fall von zwei Stichproben, also z.B. bei Diagnosestudien, ist der relative Effekt p äquivalent zur Area Under the Curve (AUC, siehe Abschnitt 1.2).

Definition 3.1 (Relativer Effekt)

Für zwei unabhängige Zufallsvariablen $X_1 \sim F_1$ und $X_2 \sim F_2$ heißt

$$p = P(X_1 < X_2) + \frac{1}{2}P(X_1 = X_2)$$

der relative Effekt von X_2 zu X_1 , und damit auch von F_2 zu F_1 . Da X_1 und X_2 unabhängige Zufallsvariablen sind, folgt mit dem Satz von Fubini die Darstellung als Lebesgue-Stieltjes-Integral (Mann und Whitney, 1947 [29]):

Proposition 3.2 (Integraldarstellung des relativen Effektes)

$$p = \int F_1 dF_2 .$$

Beweis:

$$\begin{aligned} p &= P(X_1 < X_2) + \frac{1}{2}P(X_1 = X_2) \\ &= \int P(X_1 < x) dF_2(x) + \int \frac{1}{2}P(X_1 = x) dF_2(x) \\ &= \int \left(P(X_1 < x) + \frac{1}{2}P(X_1 = x) \right) dF_2(x) \\ &= \int F_1 dF_2 . \end{aligned}$$

□

Die Interpretation ist im 2-Stichprobenfall folgende:

Definition 3.3 (Stochastische Tendenz)

X_1 tendiert im Vergleich zu X_2 (stochastisch):

- zu größeren Werten, wenn $p < \frac{1}{2}$ gilt,
- zu kleineren Werten, wenn $p > \frac{1}{2}$ gilt und
- tendenziell gleichen Werten, wenn $p = \frac{1}{2}$ gilt.

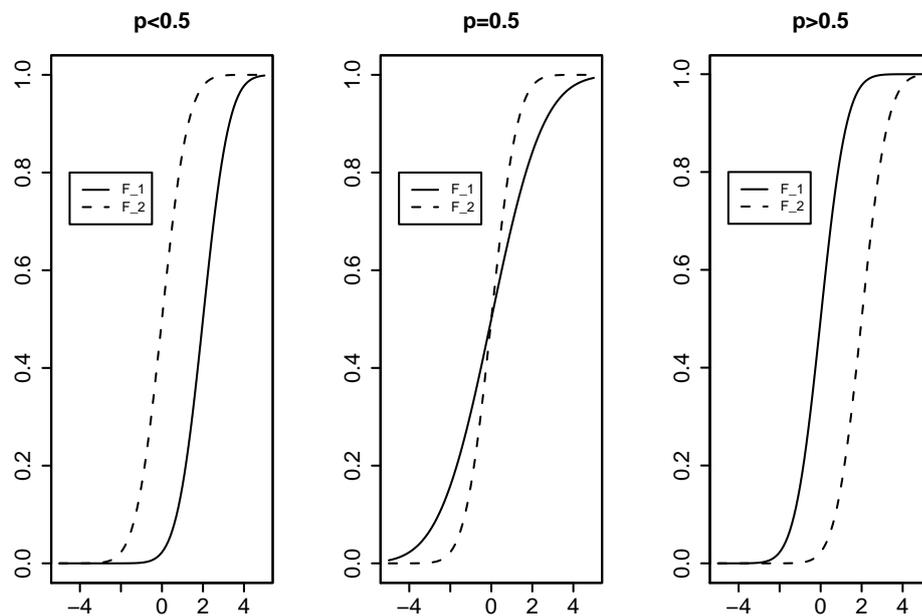


Abbildung 3.1: *Stochastische Tendenz.*

Zur Veranschaulichung der stochastischen Tendenz dient die Grafik 3.1.

Proposition 3.4 (Eigenschaften des relativen Effektes)

Der relative Effekt p hat weiterhin folgende Eigenschaften:

- (1) $p = \frac{1}{2}$, wenn die Zufallsvariablen X_1 und X_2 unabhängig und identisch verteilt sind.
- (2) p ist invariant unter beliebigen ordnungserhaltenden und messbaren Transformationen.

Beweis:

Siehe auch Brunner und Munzel (2002 [11], Proposition 1.6)

Zu (1) Es gilt immer: $P(X_1 < X_2) + P(X_1 = X_2) + P(X_1 > X_2) = 1$. Wenn X_1 und X_2 die gleiche Verteilung besitzen, gilt außerdem: $P(X_1 < X_2) = P(X_1 > X_2)$. Daraus folgt:

$$p = P(X_1 < X_2) + \frac{1}{2}P(X_1 = X_2) = \frac{1}{2}.$$

Zu (2) Sei $m(\cdot)$ eine ordnungserhaltende und messbare Transformation, so gilt in jedem Fall:

$$P(X_1 < X_2) + \frac{1}{2}P(X_1 = X_2) = P(m(X_1) < m(X_2)) + \frac{1}{2}P(m(X_1) = m(X_2)).$$

□

3.2 Schätzung

In Proposition 3.2 auf Seite 19 haben wir den relativen Effekt mit Hilfe der Verteilungsfunktionen F_1 und F_2 angegeben:

$$p = \int F_1 dF_2.$$

Da diese Verteilungsfunktionen leicht geschätzt werden können, verwenden wir für die Schätzung des relativen Effekts das Einsetzungsprinzip. Dabei werden die wahren Verteilungsfunktionen durch ihre Schätzer ersetzt, und wir erhalten als Schätzer für p :

$$\hat{p} = \int \hat{F}_1 d\hat{F}_2.$$

Der erwartungstreue Schätzer für $F_i(x)$ ist dabei folgende empirische Verteilungsfunktion $\hat{F}_i(x)$.

Definition 3.5 (Empirische Verteilungsfunktion)

Für eine Stichprobe X_{i1}, \dots, X_{in} mit der Verteilung $X_{is} \sim F_i(x)$, $s = 1, \dots, n_i$, $i = 1, 2$ heißt

$$\hat{F}_i(x) = \frac{1}{n_i} \sum_{s=1}^{n_i} c(x - X_{is})$$

die normalisierte Version der empirischen Verteilungsfunktion von X_{i1}, \dots, X_{in_i} .

Dabei beschreibt $c(x)$ die normalisierte Version der Zählfunktion (siehe Anhang B).

Berechnen lässt sich der Schätzer ganz leicht mit Hilfe der Mittelränge der Beobachtungen. Diese sind in allen üblichen Statistik-Programmpaketen implementiert und folgendermaßen definiert:

Definition 3.6 (Mittelrang)

$$R_{is} = n_1 \widehat{F}_1(X_{is}) + n_2 \widehat{F}_2(X_{is}) + \frac{1}{2}.$$

Hat man die Mittelränge über die Beobachtungen gebildet, lässt sich der relative Effekt sofort schätzen.

Definition 3.7 (Rangdarstellung von \widehat{p})

Die Rangdarstellung des relativen Effekts lautet:

$$\widehat{p} = \int \widehat{F}_1 d\widehat{F}_2 = \frac{1}{n_1} \left(\overline{R}_2 - \frac{n_2 + 1}{2} \right) = \frac{1}{N} (\overline{R}_2 - \overline{R}_1) + \frac{1}{2},$$

wobei $\overline{R}_i = 1/n_i \sum_{s=1}^{n_i} R_{is}$ der Mittelwert der Ränge in der i -ten Stichprobe ist ($i = 1, 2$). Für den somit erhaltenen geschätzten relativen Effekt gelten bestimmte Eigenschaften.

Proposition 3.8 Eigenschaften von \widehat{p}

- (1) \widehat{p} ist erwartungstreu für p , d.h. $E(\widehat{p}) = p$,
- (2) \widehat{p} ist konsistent für p , d.h. $P(|\widehat{p} - p| > \epsilon) \rightarrow 0$.

Beweis:

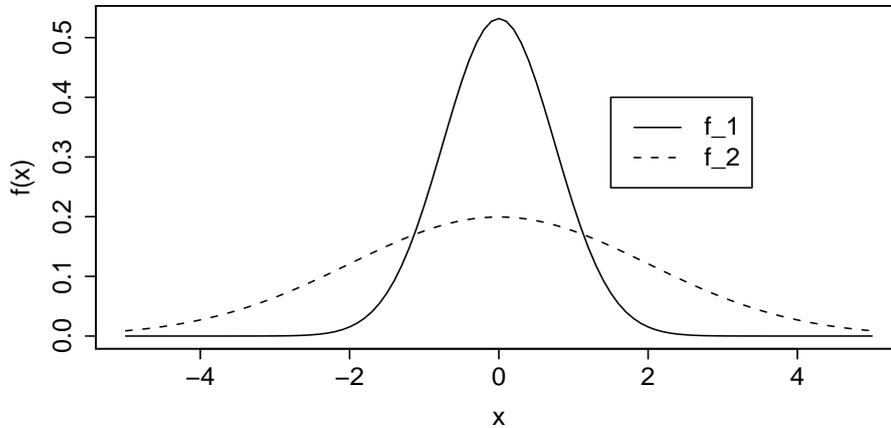
Siehe Brunner und Munzel (2002) [14]), Proposition 4.7.

□

3.3 Skalenalternativen

Der Vorteil der relativen Effekte ist, dass sie lediglich ordinales Skalenniveau voraussetzen, dadurch allgemein verwendbar sind und auch bei Heteroskedastizität zu keinen falschen Testergebnissen führen (siehe Abschnitt 1.2). Allerdings gehen alle Informationen über die Ordnungsstruktur hinaus verloren, was z.B. im Falle reiner Skalenalternativen nachteilig sein kann. Wenn also die zu Grunde liegenden Verteilungen symmetrisch sind und ein identisches Lagezentrum haben, gilt für den relativen Effekt: $p = \int F_1 dF_2 = 0.5$. Die hierüber formulierte Hypothese $H_0^p : p = 0.5$ ist bei Stetigkeit der Verteilungen und Invertierbarkeit in μ_i ($i = 1, 2$) äquivalent zu $H_0^\mu : \mu_1 = \mu_2$.

An der beispielhaften Abbildung 3.2 kann man jedoch sehen, dass hier gilt: $F_1 \neq F_2$. Die Hypothesen H_0^μ und H_0^p würden also nicht abgelehnt werden, H_0^F dagegen bei hinreichend großem Stichprobenumfang schon. Es kann also $p = 0.5$ auch dann gelten, wenn $F_1 \neq F_2$. Deshalb darf kein Testverfahren verwendet werden, das die Gleichheit der Verteilungsfunktionen voraussetzt.

Abbildung 3.2: *Skalernalternative*.

3.4 Adjustierter relativer Effekt

In unserem Modell nehmen wir an, dass die Zielvariable unter $\ell = 1, \dots, d$ verschiedenen Kombinationen beobachtet wird. Dann lautet der relative Effekt der Zielvariable in jeder Kombination ℓ (siehe Tabelle 2.1, S. 11):

$$p_0^{(\ell)} = \int F_{0,1}^{(\ell)} dF_{0,2}^{(\ell)}.$$

Die 0 im Index steht dabei für die Zielvariable. Um nun diese relativen Effekte für den Einfluss von Kovariablen zu adjustieren, verwenden wir den regressionsanalytischen Ansatz. Dieser wurde auch von Brumback et al (2006 [8]) für eine Kovariable in folgender, in unsere Notation übertragener, Form verwendet: $p_0^{(\ell)} = \gamma_1^{(\ell)} p_1^{(\ell)} + p^{(\ell)*}$. In diesem Modell lautet der für Kovariablen adjustierte relative Effekt also

$$p^{(\ell)*} = p_0^{(\ell)} - \gamma_1^{(\ell)} p_1^{(\ell)}.$$

Dabei bezeichnet der Stern bei $p^{(\ell)*}$ in dieser Arbeit immer den adjustierten Effekt. Weiterhin ist $\gamma_1^{(\ell)}$ der Regressionskoeffizient, der den Einfluss der einen Kovariablen auf die Zielgröße beschreibt und $p_1^{(\ell)}$ ist der relative Effekt der einen Kovariable. Ein neuerer Ansatz von Janes et al. (2008 [25] und [24]) zur nichtparametrischen Schätzung der adjustierten AUC ist, dass erst die adjustierte ROC-Kurve mit dem Regressionsansatz geschätzt wird und anschließend die adjustierte AUC als das Integral über die adjustierte ROC-Kurve berechnet wird. Allerdings ist das Verfahren nach Aussage der Autoren bei sehr großen Effekten instabil,

außerdem kann dabei nur ein Faktor berücksichtigt werden.

Dagegen hat das hier vorgestellte Verfahren den Vorteil, dass mehrere Kovariablen und Faktoren berücksichtigt werden können. Außerdem zentrieren wir die relativen Effekte der Kovariablen um den Punkt 0, da der Einfluss bei gleicher Verteilung gleich Null sein soll. Wenn beispielsweise in einer randomisierten Studie die Altersverteilung in beiden Statusgruppen ungefähr gleich ist, für das Alter aber trotzdem adjustiert wird, sollte der adjustierte relative Effekt auch ungefähr gleich dem unadjustierten sein.

Damit erhalten wir also das Regressionsmodell

$$p_0^{(\ell)} = \sum_{j=1}^k \gamma_j^{(\ell)} (p_j^{(\ell)} - 0.5) + p^{(\ell)*}, \quad (3.1)$$

und entsprechend den adjustierten relativen Effekt.

Defintion 3.9 (Adjustierter relativer Effekt)

Der für die Kovariablen $j = 1, \dots, k$ adjustierte relative Effekt ist dann also

$$p^{(\ell)*} = p_0^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} (p_j^{(\ell)} - 0.5),$$

mit $\gamma_j^{(\ell)}$ als Regressionskoeffizient für die j -te Kovariable in der ℓ -ten Kombination.

Durch die Berücksichtigung der Kovariablen soll die Streuung des Messfehlers reduziert und dadurch die Macht des Tests erhöht werden. Außerdem soll so der Bias verringert werden, da die unterschiedliche Verteilung der Kovariablen nicht fälschlicherweise als Effekt in der Zielvariablen interpretiert wird. Der relative Effekt für die Kovariable wird genauso wie der für die Zielvariable gebildet. Wir erhalten also für die Kovariable $j = 1, \dots, k$ in jeder Kombination ℓ den relativen Effekt:

$$p_j^{(\ell)} = \int F_{j,1}^{(\ell)} dF_{j,2}^{(\ell)}.$$

Die Eigenschaften und Interpretation von Abschnitt 3.1. gelten entsprechend für die verschiedenen $p_j^{(\ell)}$ mit $j = 0, \dots, k$.

Für die einfachere Programmierbarkeit wollen wir die relativen Effekte folgendermaßen in einem Vektor darstellen:

$$\begin{aligned} \mathbf{p} &= (p_0^{(1)}, \dots, p_0^{(d)}, (p_1^{(1)} - 0.5), \dots, (p_k^{(d)} - 0.5)) \\ &= (\mathbf{p}'_0, \dots, \mathbf{p}'_k)' \end{aligned}$$

mit $\mathbf{p}_0 = (p_0^{(1)}, \dots, p_0^{(d)})'$ für die Zielvariable und $\mathbf{p}_j = (p_j^{(1)} - 0.5, \dots, p_j^{(d)} - 0.5)'$, $j = 1, \dots, k$ für die Kovariablen. Weiterhin wollen wir die Regressionskoeffizienten in Matrixform darstellen:

$$\mathbf{\Gamma} = (\mathbf{I}, \mathbf{\Gamma}_1, \dots, \mathbf{\Gamma}_k) , \quad (3.2)$$

dabei sei \mathbf{I} die d -dimensionale Einheitsmatrix (siehe Anhang B) und

$$\mathbf{\Gamma}_j = \begin{pmatrix} -\gamma_j^{(1)} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & -\gamma_j^{(d)} \end{pmatrix}$$

die Matrix mit den Regressionskoeffizienten für die j -te Kovariable unter den $\ell = 1, \dots, d$ Bedingungen.

Proposition 3.10 (Matrixschreibweise von \mathbf{p}^*)

Dann gilt für den Vektor der adjustierten relativen Effekte $\mathbf{p}^* = (p^{(1)*}, \dots, p^{(d)*})'$:

$$\mathbf{p}^* = \mathbf{\Gamma} \mathbf{p} .$$

Im nächsten Kapitel werden wir Punkt- und Intervallschätzer für den adjustierten relativen Effekt herleiten und die vorliegenden Eigenschaften zeigen und beweisen.

4 Punkt- und Intervallschätzer

In diesem Kapitel sollen die Punkt- und Intervallschätzer für die adjustierten relativen Effekte hergeleitet werden. Außerdem wird die Methodik im Abschnitt 4.9 auf die beiden Beispiele angewendet und die Ergebnisse werden diskutiert.

4.1 Schätzer der einzelnen relativen Effekte

Vorerst werden die relativen Effekte für Ziel- und Kovariablen einzeln geschätzt. Hierbei werden also nach dem Einsetzungsprinzip die wahren durch die geschätzten Verteilungsfunktionen ersetzt:

$$\hat{p}_j^{(\ell)} = \int \hat{F}_{j,1}^{(\ell)} d\hat{F}_{j,2}^{(\ell)}.$$

Dabei steht $j = 0$ für die Zielvariable und $j = 1, \dots, k$ für die Kovariablen. Die empirische Verteilungsfunktion

$$\hat{F}_{j,i}^{(\ell)}(x) = \frac{1}{N} \sum_{s=1}^N c(x - X_{j,is}^{(\ell)})$$

ist ein erwartungstreuer Schätzer für $F_{j,i}^{(\ell)}(x)$ (siehe hierzu auch Abschnitt 3.2). Die Schätzer $\hat{p}_j^{(\ell)}$ sind mit Hilfe der Ränge exakt und einfach zu berechnen. Dabei werden die Mittelränge für jede Kombination ℓ und für jede Ziel- bzw. Kovariable j einzeln gebildet:

$$\hat{p}_j^{(\ell)} = \frac{1}{N} \left(\bar{R}_{j,2}^{(\ell)} - \bar{R}_{j,1}^{(\ell)} \right) + \frac{1}{2}.$$

Im folgenden Abschnitt soll nun der Punktschätzer für den adjustierten relativen Effekt hergeleitet werden.

4.2 Geschätzter adjustierter relativer Effekt

Vorerst nehmen wir an, dass die $\gamma_j^{(\ell)}$ ($\ell = 1, \dots, d$ und $j = 1, \dots, k$) bekannt sind (siehe hierzu z.B. Langer, 1998 [28]). Erst in Abschnitt 4.5 werden die Regressionskoeffizienten $\gamma_j^{(\ell)}$ geschätzt.

Das Regressionsmodell für den geschätzten relativen Effekt lautet analog zu (3.1) auf S.24:

$$\widehat{p}_0^{(\ell)} = \sum_{j=1}^k \gamma_j^{(\ell)} \left(\widehat{p}_j^{(\ell)} - 0.5 \right) + \widehat{p}^{(\ell)*} . \quad (4.1)$$

Dann ist der geschätzte adjustierte relative Effekt entsprechend folgendermaßen definiert:

Definition 4.1 (Schätzer für $p^{(\ell)*}$)

$$\widehat{p}^{(\ell)*} = \widehat{p}_0^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \left(\widehat{p}_j^{(\ell)} - 0.5 \right) .$$

Dabei nehmen wir vorerst an, dass die $\gamma_j^{(\ell)}$ ($j = 1, \dots, k$) bekannt sind (siehe oben). Somit erhalten wir den Vektor der geschätzten adjustierten relativen Effekte $\widehat{\mathbf{p}}^* = (\widehat{p}^{(1)*}, \dots, \widehat{p}^{(d)*})$. Wir können zeigen, dass der geschätzte adjustierte relative Effekt $\widehat{p}^{(\ell)*}$ folgende Eigenschaften besitzt:

Theorem 4.2 (Eigenschaften von $\widehat{p}^{(\ell)*}$)

(1) $\widehat{p}^{(\ell)*}$ ist konsistent für $p^{(\ell)*}$ in L_2 :

$$E \left[(\widehat{p}^{(\ell)*} - p^{(\ell)*})^2 \right] \rightarrow 0.$$

(2) $\widehat{p}^{(\ell)*}$ ist erwartungstreu für $p^{(\ell)*}$, d.h.

$$E(\widehat{p}^{(\ell)*}) = p^{(\ell)*}.$$

(3) $\widehat{p}^{(\ell)*}$ ist Skalierungs-invariant.

(4) $\widehat{p}^{(\ell)*}$ ist robust gegenüber

- Ausreißern, da die Beobachtungen nicht mit ihren eigentlichen Werten, sondern nur mit ihren Rängen eingehen.
- Verletzungen von Verteilungsannahmen, da alle Verteilungen für $F_{j,i}^{(\ell)}(x)$ erlaubt sind. Der einzige Fall, der ausgeschlossen werden muss, ist, dass alle $F_{j,i}^{(\ell)}(x)$ Ein-Punkt-Verteilungen sind.

Beweis:

Siehe Anhang C.1, Seite 77. □

Anhand des Punktschätzers des relativen Effekts kann noch keine Aussage über die Variabilität des Ergebnisses getroffen werden. Mit einem Bereichsschätzer kann dagegen angegeben werden, in welchem Intervall der wahre Wert mit einer vorgegebenen Wahrscheinlichkeit liegt. Für die Herleitung des Konfidenzintervalls benötigen wir vorerst die asymptotische Verteilung der adjustierten relativen Effekte.

4.3 Asymptotische Äquivalenz und multivariate Normalverteilung

Die Komponenten der adjustierten relativen Effekte sind nicht mehr unabhängig. Um trotzdem die zentralen Grenzwertsätze anwenden und die Verteilung des Schätzers herleiten zu können, benötigen wir die Darstellung der relativen Effekte als Summe unabhängiger Zufallsvariablen, die asymptotisch die gleiche Verteilung haben. Dabei sei $B_N^{(\ell)}$ eine Summe von unabhängigen Zufallsvariablen und folgendermaßen definiert:

Definition 4.3

$$\begin{aligned}
 \sqrt{N}B_N^{(\ell)} &= \sqrt{N} \left(\frac{1}{n_2} \sum_{s=1}^{n_2} F_{0,1}^{(\ell)}(X_{0,2s}^{(\ell)}) - \frac{1}{n_1} \sum_{s=1}^{n_1} F_{0,2}^{(\ell)}(X_{0,1s}^{(\ell)}) + 1 - 2p_0^{(\ell)} - \right. \\
 &\quad \left. - \sum_{j=1}^k \gamma_j^{(\ell)} \left(\frac{1}{n_2} \sum_{s=1}^{n_2} F_{j,1}^{(\ell)}(X_{j,2s}^{(\ell)}) - \frac{1}{n_1} \sum_{s=1}^{n_1} F_{j,2}^{(\ell)}(X_{j,1s}^{(\ell)}) + 1 - 2p_j^{(\ell)} \right) \right) \\
 &= \sqrt{N} \left(\frac{1}{n_2} \sum_{s=1}^{n_2} Y_{0,2s}^{(\ell)} - \frac{1}{n_1} \sum_{s=1}^{n_1} Y_{0,1s}^{(\ell)} + 1 - 2p_0^{(\ell)} - \right. \\
 &\quad \left. - \sum_{j=1}^k \gamma_j^{(\ell)} \left(\frac{1}{n_2} \sum_{s=1}^{n_2} Y_{j,2s}^{(\ell)} - \frac{1}{n_1} \sum_{s=1}^{n_1} Y_{j,1s}^{(\ell)} + 1 - 2p_j^{(\ell)} \right) \right) \\
 &= \sqrt{N} \left(\bar{Y}_{0,2}^{(\ell)} - \bar{Y}_{0,1}^{(\ell)} + 1 - 2p_0^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \left(\bar{Y}_{j,2}^{(\ell)} - \bar{Y}_{j,1}^{(\ell)} + 1 - 2p_j^{(\ell)} \right) \right).
 \end{aligned}$$

Dann zeigen wir also im Folgenden die asymptotische Äquivalenz des adjustierten relativen Effekts in der Komponente ℓ .

Satz 4.4 (Komponentenweise asymptotische Äquivalenz)

Unter den Annahmen (A1) und (A2) auf S. 13 gilt:

$$\begin{aligned}
 \sqrt{N} \left(\hat{p}^{(\ell)*} - p^{(\ell)*} \right) &\doteq \sqrt{N} \left(\int F_{0,1}^{(\ell)} d\hat{F}_{0,2}^{(\ell)} - \int F_{0,2}^{(\ell)} d\hat{F}_{0,1}^{(\ell)} + 1 - 2p_0^{(\ell)} - \right. \\
 &\quad \left. - \sum_{j=1}^k \gamma_j^{(\ell)} \left(\int F_{j,1}^{(\ell)} d\hat{F}_{j,2}^{(\ell)} - \int F_{j,2}^{(\ell)} d\hat{F}_{j,1}^{(\ell)} + 1 - 2p_j^{(\ell)} \right) \right) \\
 &= \sqrt{N}B_N^{(\ell)}.
 \end{aligned}$$

Beweis:

Siehe Anhang C.2, Seite 78.

Da diese Äquivalenz in jeder Komponente ℓ gilt und die Anzahl der Dimensionen beschränkt ist, gilt die Äquivalenz genauso für den Gesamtvektor der geschätzten relativen Effekte.

Satz 4.5 (Asymptotischer Äquivalenzsatz)

Unter den Annahmen (A1) und (A2) auf S. 13 gilt:

$$\sqrt{N}(\widehat{\mathbf{p}}^* - \mathbf{p}^*) \doteq \sqrt{N}\mathbf{B}_N,$$

mit $\mathbf{B}_N = \overline{\mathbf{Y}}_{0,2} - \overline{\mathbf{Y}}_{0,1} + \mathbf{1} - 2\mathbf{p}_0 - \sum_{j=1}^k \gamma_j (\overline{\mathbf{Y}}_{j,2} - \overline{\mathbf{Y}}_{j,1} + \mathbf{1} - 2\mathbf{p}_j)$.

Um die asymptotische Verteilung der relativen Effekte herzuleiten genügt es also, die asymptotische Verteilung des Zufallsvektors $\sqrt{N}\mathbf{B}_N$ zu überprüfen. Dabei bezeichnen $\lambda_1, \dots, \lambda_d$ die Eigenwerte der asymptotischen Kovarianzmatrix $\mathbf{V}_N^* = \text{Cov}(\sqrt{N}\mathbf{B}_N)$, und $\lambda_{\min} = \min(\lambda_1, \dots, \lambda_d)$ sei der kleinste Eigenwert. Dann benötigen wir für die Herleitung der asymptotischen Verteilung von $\sqrt{N}\mathbf{B}_N$ noch die Annahme, dass $\lambda_{\min} > 0$ ist.

(A4) Es gelte $\lambda_{\min} \geq \lambda_0 > 0$, dabei sei λ_0 eine Konstante.

Jetzt können wir für den Vektor der adjustierten relativen Effekte zeigen, dass er asymptotisch multivariat normalverteilt ist.

Satz 4.6 (Multivariate Normalverteilung)

Unter den Annahmen (A1), (A2), (A3) und (A4) ist die Statistik $\sqrt{N}(\widehat{\mathbf{p}}^* - \mathbf{p}^*)$ asymptotisch multivariat normalverteilt mit Erwartungswertvektor $\mathbf{0}$ und Kovarianzmatrix $\mathbf{V}_N^* = \text{Cov}(\sqrt{N}\mathbf{B}_N)$.

Beweis:

Es genügt, die asymptotische Verteilung des Zufallsvektors $\sqrt{N}\mathbf{B}_N$ mit den Komponenten $B_N^{(\ell)}$ $\ell = 1, \dots, d$ zu zeigen. Nach dem Satz von Cramer/Wold genügt es weiterhin die Normalverteilung für jede beliebige Linearkombination der einzelnen Komponenten $B_N^{(\ell)}$ zu zeigen. Dabei setzen sich die Komponenten aus Konstanten und den unabhängigen Zufallsvariablen $Y_{j,i}^{(\ell)}$ zusammen (siehe Definition 4.4 auf S. 29). Die asymptotische Normalität folgt schließlich durch Anwendung des zentralen Grenzwertsatzes von Lindeberg-Feller, weil die Varianzen nach Annahme (A3) und (A4) ungleich 0 sind. Die Lindeberg-Bedingung ist erfüllt, weil die Zufallsvariablen $Y_{j,i}^{(\ell)}$ gleichmäßig beschränkt sind. □

Die Kovarianzmatrix $\mathbf{V}_N^* = \text{Cov}(\sqrt{N}\mathbf{B}_N)$ setzt sich dabei aus den gewichteten Kovarianzmatrizen der beiden Statusgruppen zusammen (siehe unten). Weiterhin sind die Zufallsvektoren $\overline{\mathbf{Y}}_{j,i} = (\overline{Y}_{j,i}^{(1)}, \dots, \overline{Y}_{j,i}^{(d)})$ mit $i = 1, 2$ und $j = 0, \dots, k$ unabhängig voneinander. Dabei sei $\overline{Y}_{j,i}^{(\ell)} = \frac{1}{n_i} \sum_{s=1}^{n_i} Y_{j,is}^{(\ell)}$ und $Y_{j,1s}^{(\ell)} = F_{j,2}^{(\ell)}(X_{j,1s}^{(\ell)})$ bzw. $Y_{j,2s}^{(\ell)} = F_{j,1}^{(\ell)}(X_{j,2s}^{(\ell)})$. Dann ist die asymptotische Kovarianzmatrix folgendermaßen definiert:

Definition 4.7 (Asymptotische Kovarianzmatrix)

$$\begin{aligned}
 \mathbf{V}_N^* &= \text{Var} \left(\sqrt{N} \mathbf{B}_N \right) \\
 &= N \cdot \text{Var} \left(\bar{\mathbf{Y}}_{0,2} - \sum_{j=1}^k \gamma_j \bar{\mathbf{Y}}_{j,2} - \bar{\mathbf{Y}}_{0,1} + \sum_{j=1}^k \gamma_j \bar{\mathbf{Y}}_{j,1} \right) \\
 &= N \left(\text{Var} \left(\bar{\mathbf{Y}}_{0,2} - \sum_{j=1}^k \gamma_j \bar{\mathbf{Y}}_{j,2} \right) + \text{Var} \left(\bar{\mathbf{Y}}_{0,1} - \sum_{j=1}^k \gamma_j \bar{\mathbf{Y}}_{j,1} \right) \right) \\
 &= \frac{N}{n_1} \mathbf{V}_1^* + \frac{N}{n_2} \mathbf{V}_2^* .
 \end{aligned}$$

Proposition 4.8 (Matrixschreibweise der Kovarianzmatrix)

Die asymptotische Kovarianzmatrix \mathbf{V}_i^* für die Statusgruppe $i = 1, 2$ lautet in Matrixschreibweise:

$$\mathbf{V}_i^* = \mathbf{\Gamma} \mathbf{\Sigma}_i^* \mathbf{\Gamma}' .$$

Dabei ist $\mathbf{\Gamma}$ definiert wie in (3.2) auf S. 25, und $\mathbf{\Sigma}_i^*$ ist:

$$\mathbf{\Sigma}_i^* = \begin{pmatrix} \Sigma_{00,i}^* & \cdots & \Sigma_{0k,i}^* \\ \vdots & & \vdots \\ \Sigma_{k0,i}^* & \cdots & \Sigma_{kk,i}^* \end{pmatrix} .$$

Die einzelnen Matrizen $\Sigma_{jj',i}^*$ ($j, j' = 0, \dots, k$) beinhalten die Kovarianzen in der i -ten Statusgruppe zwischen Ziel- und Kovariable bzw. zwischen zwei Kovariablen:

$$\Sigma_{jj',i}^* = \begin{pmatrix} \sigma_{jj',i}^{(1,1)*} & \cdots & \sigma_{jj',i}^{(1,d)*} \\ \vdots & & \vdots \\ \sigma_{jj',i}^{(d,1)*} & \cdots & \sigma_{jj',i}^{(d,d)*} \end{pmatrix} .$$

Dabei ist $\sigma_{jj',i}^{(\ell,\ell')*} = \text{Cov} \left(\mathbf{Y}_{j,i}^{(\ell)}, \mathbf{Y}_{j',i}^{(\ell')} \right)$ die Kovarianz zwischen der (Ziel-/Ko-) Variable j unter der Bedingung ℓ und der (Ziel-/Ko-) Variable j' unter der Bedingung ℓ' .

Dann bestehen die Kovarianzmatrizen \mathbf{V}_i^* der einzelnen Statusgruppen $i = 1, 2$ aus den Elementen $\left(v_i^{(\ell,r)} \right)_{\ell,r}$ mit $\ell, r = 1, \dots, d$. Die Kovarianz zwischen zwei Bedingungen ℓ und r lautet dann

$$v_i^{(\ell,r)*} = \text{Cov} \left(Y_{0,is}^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} Y_{j,is}^{(\ell)}, Y_{0,is}^{(r)} - \sum_{j=1}^k \gamma_j^{(r)} Y_{j,is}^{(r)} \right) .$$

Im folgenden Abschnitt soll ein Schätzer für den Regressionskoeffizienten $\gamma_j^{(\ell)}$ mit $\ell = 1, \dots, d$

und $j = 1, \dots, k$ gefunden werden, damit beobachtbare Punkt- und Intervallschätzer definiert werden können.

4.4 Schätzung der Kovarianzmatrix

Im letzten Abschnitt haben wir die wahre Kovarianz definiert. Da die $\mathbf{Y}_{j,i}^{(\ell)}$ nicht beobachtbar sind, ersetzen wir sie für die Schätzung der asymptotischen Kovarianzmatrix durch ihre empirischen Analoga. Dabei nehmen wir weiterhin die Regressionskoeffizienten $\gamma_j^{(\ell)}$ als Konstanten an. Die geschätzte Kovarianzmatrix $\widehat{\mathbf{V}}_N^*$ lautet dann:

$$\widehat{\mathbf{V}}_N^* = \frac{N}{n_1} \widehat{\mathbf{V}}_1^* + \frac{N}{n_2} \widehat{\mathbf{V}}_2^*,$$

mit den Elementen für $i = 1, 2$:

$$\widehat{\mathbf{V}}_i^* = \left(\widehat{v}_i^{(\ell,r)*} \right)_{\ell,r}.$$

Dabei gilt: $\widehat{v}_i^{(\ell,r)*} = \frac{N}{n_1} \widehat{v}_1^{(\ell,r)*} + \frac{N}{n_2} \widehat{v}_2^{(\ell,r)*}$ mit:

$$\widehat{v}_i^{(\ell,r)*} = \text{Cov} \left(\widehat{Y}_{0,is}^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \widehat{Y}_{j,is}^{(\ell)}, \widehat{Y}_{0,is}^{(r)} - \sum_{j=1}^k \gamma_j^{(r)} \widehat{Y}_{j,is}^{(r)} \right).$$

Zur numerischen Berechnung definieren wir die Vektoren $\widehat{\mathbf{Y}}_{j,i}^{(\ell)} = \left(\widehat{Y}_{j,i1}^{(\ell)}, \dots, \widehat{Y}_{j,in_i}^{(\ell)} \right)$ mit den Komponenten

$$\widehat{Y}_{j,is}^{(\ell)} = \frac{1}{N - n_i} \left(R_{j,is}^{(\ell)} - R_{j,is}^{(\ell)(i)} \right).$$

Der Internrang $R_{j,is}^{(\ell)(i)}$ bezeichnet den Mittelrang der s -ten Beobachtung innerhalb der i -ten Statusgruppe in der ℓ -ten Komponente. Die asymptotische Kovarianz kann also mit Hilfe der Gesamt- und Internränge geschätzt werden. Für diesen einfachen Schätzer gelten folgende Eigenschaften:

Theorem 4.9 (Eigenschaften des Kovarianzschätzers)

Der Schätzer der asymptotischen Kovarianzmatrix hat folgende Eigenschaften:

- (1) $\widehat{\mathbf{V}}_N^*$ ist ein asymptotisch erwartungstreuer Schätzer für \mathbf{V}_N^* .
- (2) $\widehat{\mathbf{V}}_N^*$ ist L_2 -konsistent.
- (3) $\widehat{\mathbf{V}}_N^*$ ist positiv semidefinit.

Beweis:

Siehe Anhang C.3, Seite 80. □

4.5 Schätzung der Regressionskoeffizienten

Alle bisherigen Ergebnisse basieren auf der Annahme, dass die Regressionskoeffizienten $\gamma_j^{(\ell)}$ bekannt sind. In diesem Abschnitt soll nun ein erwartungstreuer Schätzer der Koeffizienten hergeleitet werden.

In der Gleichung (3.1) auf S.24 haben wir die Regressionsgleichung mit den relativen Effekten $p_j^{(\ell)}$ aufgestellt. Da die Komponenten des Vektors $\mathbf{p}_j = (p_j^{(1)}, \dots, p_j^{(d)})$ mit $j = 0, \dots, k$ nicht unabhängig sind (siehe auch Abschnitt 4.3), wollen wir die Zufallsvariablen so transformieren, dass das berechnete Funktional asymptotisch äquivalent zu $p_j^{(\ell)}$ ist. Somit erhalten wir $Y_{j,1s}^{(\ell)} = F_{j,2}^{(\ell)}(X_{j,1s}^{(\ell)})$ bzw. $Y_{j,2s}^{(\ell)} = F_{j,1}^{(\ell)}(X_{j,2s}^{(\ell)})$ als so genannte asymptotische Rangtransformation (ART)¹. Also nehmen wir eine Zerlegung der Zielvariablen in der ART in folgender Form an:

$$Y_{0,is}^{(\ell)} = \sum_{j=1}^k \gamma_j^{(\ell)} (Y_{j,is}^{(\ell)} - 0.5) + Y_{is}^{(\ell)*}.$$

Dabei sei $\ell = 1, \dots, d$, $i = 1, 2$ und $s = 1, \dots, n_i$. Nach Mittelung in den Faktorstufenkombinationen und Zentrierung um diesen Mittelwert $\bar{Y}_{j,i}^{(\ell)} = \frac{1}{N} \sum_{s=1}^{n_i} Y_{j,is}^{(\ell)}$ erhalten wir für jede Ziel- bzw. Kovariable $j = 0, \dots, k$ den Vektor

$$\begin{aligned} \mathbf{Y}_{j,iz}^{(\ell)} &= \left(Y_{j,i1}^{(\ell)} - 0.5 - \left(\bar{Y}_{j,i}^{(\ell)} - 0.5 \right), \dots, Y_{j,in_i}^{(\ell)} - 0.5 - \left(\bar{Y}_{j,i}^{(\ell)} - 0.5 \right) \right)' \\ &= \left(Y_{j,i1}^{(\ell)} - \bar{Y}_{j,i}^{(\ell)}, \dots, Y_{j,in_i}^{(\ell)} - \bar{Y}_{j,i}^{(\ell)} \right)' \end{aligned}$$

und $\mathbf{Y}_{j,z}^{(\ell)} = (\mathbf{Y}_{j,1z}^{(\ell)}, \mathbf{Y}_{j,2z}^{(\ell)})'$. Das Regressionsmodell lautet dann:

$$\mathbf{Y}_{0,z}^{(\ell)} = \sum_{j=1}^k \gamma_j^{(\ell)} \left(\mathbf{Y}_{j,z}^{(\ell)} - 0.5 \right) + \mathbf{Y}_z^{(\ell)*}.$$

Die Vektoren der Kovariablen werden zusammengefasst als $\mathbf{Y}_{1:k,z}^{(\ell)} = \left(\mathbf{Y}_{1,z}^{(\ell)}, \dots, \mathbf{Y}_{k,z}^{(\ell)} \right)'$. Da $F_{j,i}^{(\ell)}$ als Mittel von Verteilungsfunktionen unbekannt ist, sind die $Y_{j,is}^{(\ell)}$ unbeobachtbar. Wird allerdings $F_{j,i}^{(\ell)}$ durch den Schätzer $\hat{F}_{j,i}^{(\ell)}$ ersetzt, erhalten wir beobachtbare Zufallsvariablen. Dann lässt sich der Regressionskoeffizient $\gamma_j^{(\ell)}$ mit der Kleinsten-Quadrate-Methode schätzen.

¹Vergleiche hierzu auch Siemer (1999 [36]).

Definition 4.10 (Geschätzte Regressionskoeffizienten)

Der Vektor der Kleinste-Quadrate-Schätzer für die Bedingung ℓ , $\widehat{\boldsymbol{\gamma}}^{(\ell)} = (\widehat{\gamma}_1^{(\ell)}, \dots, \widehat{\gamma}_k^{(\ell)})$, lautet:

$$\widehat{\boldsymbol{\gamma}}^{(\ell)} = \left(\widehat{\mathbf{Y}}_{1:k,z}^{(\ell)'} \widehat{\mathbf{Y}}_{1:k,z}^{(\ell)} \right)^{-1} \widehat{\mathbf{Y}}_{1:k,z}^{(\ell)'} \widehat{\mathbf{Y}}_{0,z}^{(\ell)}$$

Man benötigt für die Schätzung der Regressionskoeffizienten die Prädiktoren $\widehat{Y}_{j,is}^{(\ell)}$ mit $j = 0, \dots, k$, die aus der Differenz des bereits bekannten Mittelrangs und des Internrangs berechnet werden. Die Rangdarstellung der Prädiktoren lautet dann:

$$\widehat{Y}_{j,1s}^{(\ell)} = \widehat{F}_{j,2}^{(\ell)} \left(X_{j,1s}^{(\ell)} \right) = \frac{1}{n_2} \left(R_{j,1s}^{(\ell)} - R_{j,1s}^{(\ell)(1)} \right)$$

bzw.

$$\widehat{Y}_{j,2s}^{(\ell)} = \widehat{F}_{j,1}^{(\ell)} \left(X_{j,2s}^{(\ell)} \right) = \frac{1}{n_1} \left(R_{j,2s}^{(\ell)} - R_{j,2s}^{(\ell)(2)} \right).$$

Nun wird das arithmetische Mittel gebildet:

$$\widehat{Y}_{j,i}^{(\ell)} = \frac{1}{n_i} \sum_{s=1}^{n_i} \widehat{Y}_{j,is}^{(\ell)}, j = 0, \dots, k.$$

Anschließend werden die zentrierten Prädiktoren $\widehat{\mathbf{Y}}_{j,iz}^{(\ell)}$ berechnet:

$$\widehat{\mathbf{Y}}_{j,iz}^{(\ell)} = \left(\widehat{Y}_{j,i1}^{(\ell)} - \widehat{Y}_{j,i}^{(\ell)}, \dots, \widehat{Y}_{j,is}^{(\ell)} - \widehat{Y}_{j,i}^{(\ell)} \right)'.$$

Dann erhält man unter Verwendung von $\widehat{\mathbf{Y}}_{j,z}^{(\ell)} = \left(\widehat{\mathbf{Y}}_{j,1z}^{(\ell)'}, \widehat{\mathbf{Y}}_{j,2z}^{(\ell)'} \right)'$ und mit der Matrix für die Kovariablen $\widehat{\mathbf{Y}}_{1:k,z}^{(\ell)} = \left(\widehat{\mathbf{Y}}_{1,z}^{(\ell)'}, \dots, \widehat{\mathbf{Y}}_{k,z}^{(\ell)'} \right)'$ die oben angegebenen geschätzten Regressionskoeffizienten.

Theorem 4.11 (Eigenschaften von $\widehat{\boldsymbol{\gamma}}^{(\ell)}$)

Der Vektor der Kleinste-Quadrate-Schätzer $\widehat{\boldsymbol{\gamma}}^{(\ell)}$ ist asymptotisch erwartungstreu für den Vektor der Regressionskoeffizienten $\boldsymbol{\gamma}^{(\ell)}$. Das heißt, es gilt:

$$E \left(\widehat{\boldsymbol{\gamma}}^{(\ell)} - \boldsymbol{\gamma}^{(\ell)} \right) \rightarrow 0$$

Beweis:

Diese Eigenschaft ergibt sich aus dem Satz von Gauss-Markov-Aitken.

□

Da die $\widehat{\mathbf{Y}}_{j,z}^{(\ell)}$ auf der Differenz zwischen Gesamtrang und Internrang basieren, werden die Komponenten Null, wenn die beiden Statusgruppen bei der Kovariable j komplett getrennt sind. Das wiederum hat zur Folge, dass die Inverse nicht existiert. Auf diesen Punkt werden wir in der Diskussion in Kapitel 8 zurückkommen.

Unter Verwendung der angegebenen Kleinste-Quadrate-Schätzer kann nun ein beobachtbarer Punktschätzer für den adjustierten relativen Effekt hergeleitet werden.

4.6 Beobachtbarer Punktschätzer

Im Abschnitt 4.2 wurde der Punktschätzer für den adjustierten relativen Effekt unter Voraussetzung eines bekannten $\gamma_j^{(\ell)}$ definiert. Nun wollen wir beobachtbare Punktschätzer, bedingt auf die geschätzten Regressionskoeffizienten, angeben. Diese lauten dann:

$$\widehat{\mathbf{p}}^{(\ell)*} = \widehat{\mathbf{p}}_0^{(\ell)} - \sum_{j=1}^k \widehat{\gamma}_j^{(\ell)} \left(\widehat{\mathbf{p}}_j^{(\ell)} - 0.5 \right) .$$

Die bewiesenen Eigenschaften aus Theorem 4.2 gelten entsprechend für die beobachtbaren Punktschätzer $\widehat{\mathbf{p}}^{(\ell)*}$, da auf die Konstante $\widehat{\gamma}_j^{(\ell)}$ bedingt wird.

Wir wollen die geschätzten adjustierten relativen Effekte analog zu den wahren in Abschnitt 3.4 darstellen. Das bedeutet, dass die geschätzten Einzeleffekte folgendermaßen aussehen:

$$\widehat{\mathbf{p}} = (\widehat{\mathbf{p}}'_0, \dots, \widehat{\mathbf{p}}'_k)'$$

mit $\widehat{\mathbf{p}}_0 = (\widehat{p}_0^{(1)}, \dots, \widehat{p}_0^{(d)})'$ für die Zielvariable und $\widehat{\mathbf{p}}_j = (\widehat{p}_j^{(1)} - 0.5, \dots, \widehat{p}_j^{(d)} - 0.5)'$, $j = 1, \dots, k$ für die Kovariablen. Die geschätzten Regressionskoeffizienten werden wie auf S.25 in Matrixform dargestellt:

$$\widehat{\mathbf{\Gamma}} = (\mathbf{I}, \widehat{\mathbf{\Gamma}}_1, \dots, \widehat{\mathbf{\Gamma}}_k) , \tag{4.2}$$

dabei gelte:

$$\widehat{\mathbf{\Gamma}}_j = \begin{pmatrix} -\widehat{\gamma}_j^{(1)} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & -\widehat{\gamma}_j^{(d)} \end{pmatrix}$$

als Matrix mit den geschätzten Regressionskoeffizienten pro Kovariable j für die verschiedenen Bedingungen.

Dann gilt für den Vektor der geschätzten adjustierten relativen Effekte

$$\widehat{\mathbf{p}}^* = (\widehat{p}^{(1)*}, \dots, \widehat{p}^{(d)*})'$$

Definition 4.12 (Matrixschreibweise von $\hat{\mathbf{p}}^*$)

$$\hat{\mathbf{p}}^* = \hat{\Gamma} \hat{\mathbf{p}}.$$

Für die Berechnung der Intervallschätzer wird noch die beobachtbare geschätzte Kovarianzmatrix benötigt, die im nächsten Abschnitt hergeleitet werden soll.

4.7 Beobachtbare Schätzung der Kovarianzmatrix

In Abschnitt 4.4 auf S.32 haben wir die geschätzte Kovarianz mit als bekannt angenommenen Regressionskoeffizienten $\gamma_j^{(\ell)}$ definiert. Setzt man nun die geschätzten Regressionskoeffizienten $\hat{\gamma}_j^{(\ell)}$ ein, erhält man die geschätzten Kovarianzen $\hat{v}^{(\ell,r)*} = \frac{N}{n_1} \hat{v}_1^{(\ell,r)*} + \frac{N}{n_2} \hat{v}_2^{(\ell,r)*}$ mit:

$$\hat{v}_i^{(\ell,r)*} = \text{Cov} \left(\hat{Y}_{0,is}^{(\ell)} - \sum_{j=1}^k \hat{\gamma}_j^{(\ell)} \hat{Y}_{j,is}^{(\ell)}, \hat{Y}_{0,is}^{(r)} - \sum_{j=1}^k \hat{\gamma}_j^{(r)} \hat{Y}_{j,is}^{(r)} \right).$$

Für die einfachere Programmierbarkeit wollen wir die beobachtbare geschätzte Kovarianzmatrix auch wieder in der Matrixschreibweise angeben.

Definition 4.13 (Matrixschreibweise der geschätzten Kovarianzmatrix)

$$\hat{\mathbf{V}}_i^* = \hat{\Gamma} \hat{\Sigma}_i^* \hat{\Gamma}'.$$

Dabei ist $\hat{\Gamma}$ definiert wie oben in Abschnitt 4.6, und $\hat{\Sigma}_i^*$ ist:

$$\hat{\Sigma}_i^* = \begin{pmatrix} \hat{\Sigma}_{00,i}^* & \cdots & \hat{\Sigma}_{0k,i}^* \\ \vdots & & \vdots \\ \hat{\Sigma}_{k0,i}^* & \cdots & \hat{\Sigma}_{kk,i}^* \end{pmatrix}.$$

Die Matrizen $\hat{\Sigma}_{jj',i}^*$ ($j, j' = 0, \dots, k$) beinhalten die geschätzten Kovarianzen zwischen Ziel- und Kovariable bzw. zwischen zwei Kovariablen in der Statusgruppe i :

$$\hat{\Sigma}_{jj',i}^* = \begin{pmatrix} \hat{\sigma}_{jj',i}^{(1,1)*} & \cdots & \hat{\sigma}_{jj',i}^{(1,d)*} \\ \vdots & & \vdots \\ \hat{\sigma}_{jj',i}^{(d,1)*} & \cdots & \hat{\sigma}_{jj',i}^{(d,d)*} \end{pmatrix}.$$

Dann bezeichnet $\hat{\sigma}_{jj',i}^{(\ell,\ell')}$ die geschätzte Kovarianz zwischen Ziel- bzw. Kovariable j unter der Bedingung ℓ und Ziel- bzw. Kovariable j' unter der Bedingung ℓ' . Die Eigenschaften aus Theorem 4.9 gelten bei Verwendung der geschätzten Regressionskoeffizienten genauso.

Im nächsten Abschnitt sollen nun Konfidenzintervalle, mit Normal- und mit t-Approximation, hergeleitet werden.

4.8 Konfidenzintervall

4.8.1 Normalapproximation

In Abschnitt 4.3 haben wir die asymptotische multivariate Normalverteilung gezeigt. Daraus resultiert unter den Annahmen (A1) und (A2) auf S. 13 für jede Kombination ℓ :

$$\sqrt{N} \frac{\widehat{p}^{(\ell)*} - p^{(\ell)*}}{\sqrt{\widehat{v}^{(\ell, \ell)*}}} \rightarrow N(0,1) \text{ in Verteilung.}$$

Dabei gilt: $\widehat{v}^{(\ell, \ell)*} = \frac{N}{n_1} \widehat{v}_1^{(\ell, \ell)*} + \frac{N}{n_2} \widehat{v}_2^{(\ell, \ell)*}$ (siehe Seite 36). Mit dieser Annahme und der geschätzten Kovarianzmatrix $\widehat{\mathbf{V}}_N^*$ kommt man zu dem Konfidenzintervall mit Normalapproximation.

Korollar 4.14 (Konfidenzintervall mit Normal-Approximation)

Es seien die Annahmen (A1) und (A2) erfüllt. Dann ist für jedes $\alpha \in (0,1)$:

$$\left[\widehat{p}_{\text{No},u}^{(\ell)*}, \widehat{p}_{\text{No},o}^{(\ell)*} \right] = \left[\widehat{p}^{(\ell)*} \pm \frac{\sqrt{\widehat{v}^{(\ell, \ell)*}} \cdot z_{1-\alpha/2}}{\sqrt{N}} \right]$$

ein asymptotisches $(1 - \alpha)$ -Konfidenzintervall für den adjustierten relativen Effekt $p^{(\ell)*}$. Dabei steht u für die untere, o für die obere Grenze und No für die Normal-Approximation.

4.8.2 t-Approximation

Bei kleinen Stichproben kann es passieren, dass die Konfidenzintervalle mit Normalapproximation das vorgegebene Niveau nicht einhalten. Allerdings kann in diesen Fällen das Quantil $z_{1-\alpha/2}$ der Normalverteilung durch das Quantil $t_{(N-1), 1-\alpha/2}$ der t-Verteilung mit $N - 1$ Freiheitsgraden ersetzt werden. Damit erhält man ein Konfidenzintervall mit t-Approximation. Nähere Erläuterungen zur Herleitung sind beispielsweise bei Domhof (2001 [18], Abschnitt 5.5) zu finden.

Korollar 4.15 (Konfidenzintervall mit t-Approximation)

Für jedes $\alpha \in (0,1)$ ist

$$\left[\widehat{p}_{t,u}^{(\ell)*}, \widehat{p}_{t,o}^{(\ell)*} \right] = \left[\widehat{p}^{(\ell)*} \pm \frac{\sqrt{\widehat{v}^{(\ell, \ell)*}} \cdot t_{f, 1-\alpha/2}}{\sqrt{N}} \right]$$

ein $(1 - \alpha)$ -Konfidenzintervall für den adjustierten relativen Effekt $p^{(\ell)*}$. Dabei steht t für die t-Approximation, und die Varianz $\widehat{v}^{(\ell, \ell)*}$ ist wie oben definiert. Die Freiheitsgrade des

Quantils der t-Verteilung $t_{\widehat{f}^{(\ell)*}, 1-\alpha/2}$ werden folgendermaßen geschätzt:

$$\widehat{f}^{(\ell)*} = \max \left(1, \frac{\left(n_1 \widehat{v}_1^{(\ell, \ell)*} + n_2 \widehat{v}_2^{(\ell, \ell)*} \right)^2}{\left(n_1 \widehat{v}_1^{(\ell, \ell)*} \right)^2 / (n_2 - 1) + \left(n_2 \widehat{v}_2^{(\ell, \ell)*} \right)^2 / (n_1 - 1)} \right).$$

Wenn die Stichproben hinreichend groß sind, gilt:

$$P \left(\widehat{p}^{(\ell)*} \in \left[\widehat{p}_{t,u}^{(\ell)*}, \widehat{p}_{t,o}^{(\ell)*} \right] \right) = 1 - \alpha.$$

Im folgenden Abschnitt sollen die hergeleiteten Punkt- und Intervallschätzer nun auf die zwei Beispiele angewendet werden.

Durch die Adjustierung bezüglich der Kovariablen sind die Punkt- und Intervallschätzer der relativen Effekte nicht mehr bereichserhaltend. Um die Ergebnisse analog zu Definition 3.3 auf Seite 20 interpretieren zu können, muss ein Transformationsverfahren angewendet werden, das dann im fünften Kapitel beschrieben wird.

4.9 Beispiele

Meningitis-Studie

Die unadjustierten relativen Effekte in Tabelle 4.1 zeigen, dass anhand des Laborparameters A die beiden Krankheiten unterschieden werden können. Die Trennschärfe des Laborparameters B ist wesentlich geringer, und der Laborparameter C besitzt für die Diagnosefindung gar keine Aussagekraft. Werden die relativen Effekte nun Alters-adjustiert, wird für die Laborparameter A und C der Effekt kleiner, für den Laborparameter B dagegen größer, der Unterschied unadjustiert zu adjustiert liegt zwischen -0.05 und 0.01 . Die Richtung des Zu-

Tabelle 4.1: *Meningitis-Beispiel, geschätzte relative Effekte.*

Laborparameter	$\gamma_1^{(\ell)}$	Adjustierung	
		ohne	für Alter
A	0.14	0.9219	0.8759
B	-0.07	0.6842	0.7067
C	0.30	0.5319	0.4354

sammenhangs entspricht dem Vorzeichen, und die Stärke der Veränderung entspricht der Größe des geschätzten Regressionskoeffizienten (siehe Tabelle 4.1). Der Laborparameter A ist also, unabhängig vom Alter, ein sehr guter Diagnoseparameter.

In Abbildung 4.1 sind für die drei Laborparameter die Schätzer und die 95%-Konfidenzintervalle dargestellt, und zwar jeweils mit Normal- und mit t-Approximation und sowohl mit als

auch ohne Adjustierung für Kovariablen. Die Konfidenzintervalle geben zusätzlich zu den Punktschätzern eine Information über die Variabilität der Ergebnisse. So ist das Intervall, in dem der wahre Laborparameter A mit einer Wahrscheinlichkeit von 95 Prozent liegt, recht schmal und liegt ungefähr zwischen 0.8 und 1. Das Konfidenzintervall des Laborparameters

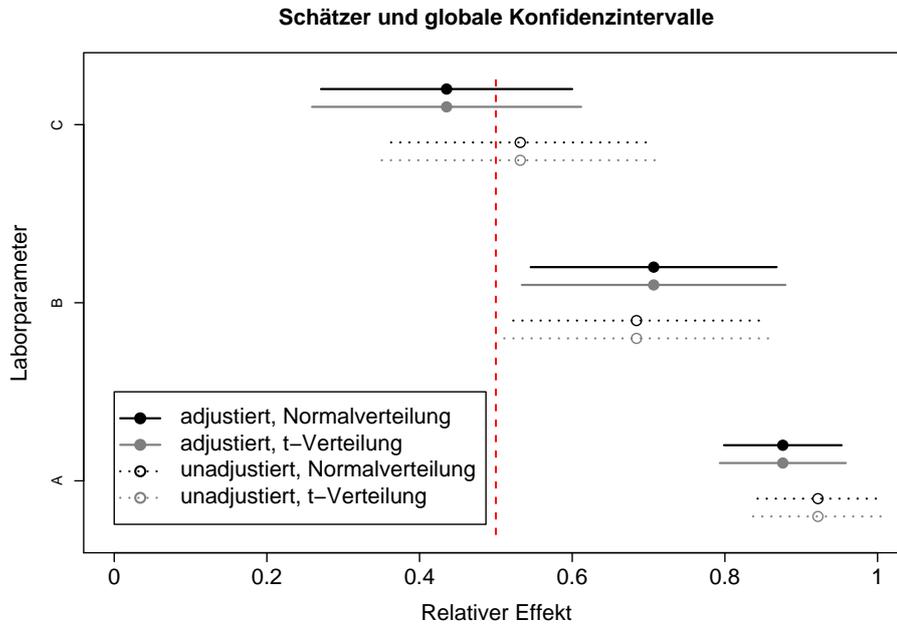


Abbildung 4.1: Meningitis-Studie, Konfidenzintervalle.

C ist dagegen wesentlich breiter und schließt auch den Wert 0.5 mit ein.

Weiterhin sieht man, dass die Konfidenzintervalle mit t-Approximation geringfügig breiter sind als mit Normal-Approximation. Das erklärt sich durch die größeren Quantile der t-Verteilung.

CAD-Studie

Bei der CAD-Studie haben wir sowohl für Befundungszeit und Alter einzeln als auch für beide gemeinsam adjustiert. Die Ergebnisse sind in Tabelle 4.2 zu sehen. Allgemein lässt sich sagen, dass mit CAD-System etwas größere relative Effekte erzielt wurden als ohne. Weiterhin wird die Diagnose, wie es zu erwarten war, mit zunehmender Erfahrung sicherer. Ohne Berücksichtigung der Kovariablen erzielt der Facharzt annähernd 100 Prozent, mit Verwendung des CAD-Systems kann er sich kaum noch verbessern (0.97 vs. 0.99), ähnlich ist es bei dem etwas erfahrenen Assistenzarzt (0.85 vs. 0.88). Bei dem absolut unerfahrenen Studenten erhöht das CAD-System dagegen die AUC von 0.75 auf 0.82, der Nutzen der Computerunterstützung scheint hier größer zu sein. Allerdings kann dieser Eindruck täuschen, da beispielsweise der Facharzt schon ohne CAD eine so sichere Diagnose stellt, dass er sich bei Verwendung des CAD-Systems nicht mehr viel verbessern kann. Dieses Problem besteht bei

allen Effektmaßen, die eine Begrenzung nach unten und/oder oben haben. Umso wichtiger ist in diesen Fällen eine sorgsame Interpretation.

Bei Altersadjustierung ändern sich die Ergebnisse so gut wie gar nicht. Das war auch zu erwarten, da die Verteilung des Alters in den beiden Statusgruppen praktisch gleich ist (siehe Abbildung 1.3 in Kapitel 1.1). Nun wird für die Befundungszeit adjustiert, die zwischen

Tabelle 4.2: CAD-Studie, Vergleich der geschätzten relativen Effekte

		Adjustierung			
		ohne	für Alter	für Zeit	für Alter und Zeit
mit CAD	Assistenzarzt	0.8795	0.8812	0.8267	0.8284
	Student	0.8204	0.8216	0.6567	0.6579
	Facharzt	0.9858	0.9854	0.9864	0.986
ohne CAD	Assistenzarzt	0.8464	0.8495	0.8172	0.8202
	Student	0.7455	0.746	0.6284	0.6289
	Facharzt	0.968	0.9683	0.9685	0.9688

den Statusgruppen etwas und zwischen den Untersuchern deutlich unterschiedlich ist (siehe Abbildung 1.2 in Kapitel 1.1). Beim Facharzt sind die Ergebnisse immer noch fast unverändert, die Sicherheit der Ergebnisse hängt also auch nicht von der benötigten Befundungszeit ab. Bei dem etwas erfahrenen Assistenzarzt werden nach der Adjustierung für die benötigte Zeit die Ergebnisse etwas schlechter, besonders deutlich ist die Veränderung aber bei dem unerfahrenen Studenten. Die relativen Effekte fallen hier um 0.08 bei Verwendung des CAD-Systems und um 0.06 bei der Diagnose ohne Computerunterstützung. Die Interpretation könnte hier sein, dass eine begrenzte Zeit für die Diagnose bei einem unerfahrenen Arzt zu schlechteren Ergebnissen führt. Bei Adjustierung für Alter und benötigter Zeit gemeinsam wird der Effekt der Befundungszeit-Adjustierung noch etwas verstärkt. Es gilt für alle Adjustierungsszenarien und unabhängig davon, ob das CAD-System genutzt wurde, dass der Student am ungenauesten und der Facharzt am genauesten diagnostiziert. Die Unterschiede zwischen den Untersuchern sind allerdings bei Verwendung des CAD-Systems etwas weniger deutlich. Es lässt sich aber keine deutliche Interaktion zwischen Untersucher und Verwendung des CAD-Systems erkennen. Sicher beurteilen lässt sich das aber letztendlich erst mit einem Test auf Interaktion.

Die Punktschätzer und 95%-Konfidenzintervalle der CAD-Studie sind in Abbildung 4.2 dargestellt. Dabei bezeichnet U1 den Assistenzarzt, U2 den Studenten und U3 den Facharzt. Weiterhin sind in der oberen Hälfte der Grafik die Ergebnisse für die Diagnose ohne CAD, und in der unteren Hälfte für die Diagnose mit CAD dargestellt. Um die Ergebnisse übersichtlich zu halten, und da die Altersverteilung in den beiden Statusgruppen sehr ähnlich war, haben wir im Folgenden nur die für die Befundungszeit adjustierten Ergebnisse angegeben.

Es ist zu beobachten, dass die Konfidenzintervalle umso breiter werden, je näher sie an der

Signifikanzgrenze 0.5 liegen. Außerdem sind die Intervalle etwas schmaler, wenn das CAD-System verwendet wurde. Kein Konfidenzintervall schließt die 0.5 mit ein. Besonders bei dem Studenten verdeutlichen die Konfidenzintervalle die Unterschiede zwischen den relativen Effekten mit und ohne Adjustierung für die Befundungszeit.

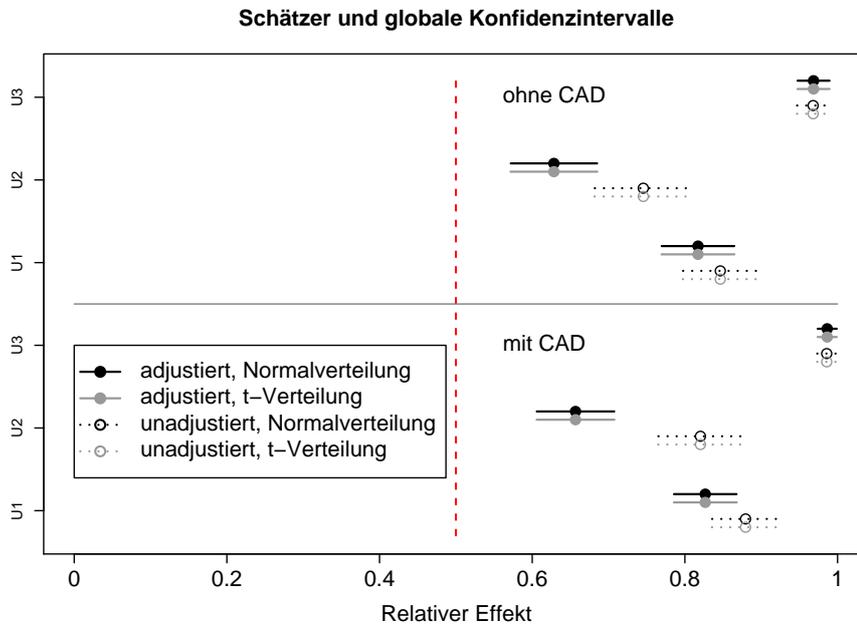


Abbildung 4.2: CAD-Studie, Konfidenzintervalle.

5 Transformationsmethode

In Kapitel 1.2 und in Definition 3.3 auf S. 20 wurde die Interpretation des relativen Effekts angegeben. Für den adjustierten relativen Effekt und das dazugehörige Konfidenzintervall sind diese allerdings so nicht anwendbar, da sie nicht mehr gezwungenermaßen bereichserhaltend sind, also $\notin [0,1]$. Wie auch am Ende von Abschnitt 4.8 bemerkt sind so die Ergebnisse nur eingeschränkt interpretierbar. Deshalb soll in diesem Kapitel die Transformationsmethode für den Punktschätzer und für das Konfidenzintervall vorgestellt werden. Damit wird die Bereichserhaltung erzwungen und die Schätzer sind wieder wie üblich interpretierbar (siehe hierzu u.a. Christophliemk (2001) [16] und Domhof (2001) [18]).

Das Prinzip bei der Transformation ist, dass das offene Einheitsintervall durch eine umkehrbare Funktion g auf den Raum der reellen Zahlen abgebildet wird. Die streng monotone und bijektive Transformation $g : (0,1) \rightarrow \mathbb{R}$ soll drei Eigenschaften erfüllen. Sie sollte stetig differenzierbar und symmetrisch um den Punkt 0.5 sein, und sie sollte eine nullstellenfreie erste Ableitung besitzen. Dann ist die Statistik der transformierten Effekte wieder asymptotisch normalverteilt, so dass auch das dazugehörige Konfidenzintervall konstruiert werden kann. Durch die Umkehrfunktion g^{-1} erhält man dann wieder bereichserhaltende Schätzer und Konfidenzintervalle.

5.1 Transformation des Punktschätzers

Für die adjustierten relativen Effekte gilt nicht mehr $p^{(\ell)*} \in [0,1]$. Um dieses Problem zu beheben, wird vorerst der relative Effekt der Zielgröße und der Kovariablen auf $p_{j,T}^{(\ell)} = g(p_j^{(\ell)})$ mit $\ell = 1, \dots, d$ und $j = 0, \dots, k$ transformiert. Zur Adjustierung verwenden wir wieder den regressionsanalytischen Ansatz. Damit erhalten wir

$$p_T^{(\ell)*} = p_{0,T}^{(\ell)} - \sum_{j=1}^k \gamma_{j,T}^{(\ell)} p_{j,T}^{(\ell)}.$$

Dabei sollen die Regressionskoeffizienten $\gamma_{j,T}^{(\ell)}$ so gewählt werden, dass die Varianz von $p_T^{(\ell)*}$ minimal wird. Der rücktransformierte adjustierte relative Effekt $\tilde{p}^{(\ell)*} = g^{-1}(p_T^{(\ell)*})$ ist dann wieder bereichserhaltend.

Entsprechend wird auch bei den bereichserhaltenen Schätzern vorgegangen. Mit $\widehat{p}_{j,T}^{(\ell)} = g\left(\widehat{p}_j^{(\ell)}\right)$ ($\ell = 1, \dots, d$ und $j = 0, \dots, k$) und dem regressionsanalytischen Ansatz erhält man

$$\widehat{p}_T^{(\ell)*} = \widehat{p}_{0,T}^{(\ell)} - \sum_{j=1}^k \widehat{\gamma}_{j,T}^{(\ell)} \widehat{p}_{j,T}^{(\ell)}.$$

(Die Varianz-minimierenden Schätzer für die Regressionskoeffizienten $\widehat{\gamma}_{j,T}^{(\ell)}$ werden am Ende von Abschnitt 5.2 angegeben.) Durch Anwendung der Umkehrfunktion $g^{-1}(\widehat{p}_T^{(\ell)*}) = \widetilde{p}^{(\ell)*}$ erhält man dann einen bereichserhaltenden Schätzer.

Das Vorgehen der Transformation soll in folgendem Diagramm noch einmal veranschaulicht werden:

$$\begin{aligned} p_j^{(\ell)} &\in [0,1], j = 0, \dots, k \\ &\downarrow g \\ p_{j,T}^{(\ell)} &= g\left(p_j^{(\ell)}\right) \in \mathbb{R}, j = 0, \dots, k \\ &\downarrow \\ p_T^{(\ell)*} &= p_{0,T}^{(\ell)} - \sum_{j=1}^k \gamma_{j,T}^{(\ell)} p_{j,T}^{(\ell)} \in \mathbb{R} \\ &\downarrow g^{-1} \\ \widetilde{p}^{(\ell)*} &= g^{-1}(p_T^{(\ell)*}) \in (0,1) \end{aligned}$$

Definition 5.1 (Rücktransformierter adjustierter relativer Effekt)

Der wahre rücktransformierte adjustierte relative Effekt ist folgendermaßen definiert:

$$\widetilde{p}^{(\ell)*} = g^{-1}\left(p_{0,T}^{(\ell)} - \sum_{j=1}^k \gamma_{j,T}^{(\ell)} p_{j,T}^{(\ell)}\right),$$

und es gelte $\tilde{p}^{(\ell)*} \in (0,1)$. Der dazugehörige Schätzer ist dann

$$\widehat{p}^{(\ell)*} = g^{-1} \left(\widehat{p}_{0,T}^{(\ell)} - \sum_{j=1}^k \widehat{\gamma}_{j,T}^{(\ell)} \widehat{p}_{j,T}^{(\ell)} \right). \quad (5.1)$$

Der Vektor der transformierten Einzeleffekte (analog zu Abschnitt 3.4 auf S. 25) lautet:

$$\mathbf{p}_T = \left(\mathbf{p}'_{0,T}, \dots, \mathbf{p}'_{k,T} \right)',$$

mit $\mathbf{p}_{j,T} = (p_{j,T}^{(1)}, \dots, p_{j,T}^{(d)})'$ für $j = 0, \dots, k$. Entsprechend ist der Vektor der geschätzten transformierten Einzeleffekte:

$$\widehat{\mathbf{p}}_T = \left(\widehat{\mathbf{p}}'_{0,T}, \dots, \widehat{\mathbf{p}}'_{k,T} \right)',$$

und der Vektor der transformierten relativen Effekte lautet entsprechend:

$$\mathbf{p}_T^* = \mathbf{\Gamma}_T \mathbf{p}_T,$$

geschätzt durch

$$\widehat{\mathbf{p}}_T^* = \widehat{\mathbf{\Gamma}}_T \widehat{\mathbf{p}}_T.$$

Proposition 5.2 (Rücktransformierter adjustierter relativer Effekt in Matrixschreibweise)

Der wahre rücktransformierte adjustierte relative Effekt lautet in Matrixschreibweise:

$$\tilde{\mathbf{p}}^* = g^{-1} (\mathbf{\Gamma}_T \mathbf{p}_T^*),$$

und der dazugehörige Schätzer

$$\widehat{\tilde{\mathbf{p}}}^* = g^{-1} (\widehat{\mathbf{\Gamma}}_T \widehat{\mathbf{p}}_T^*),$$

Im folgenden Abschnitt sollen nun die transformierten Konfidenzintervalle definiert werden.

5.2 Transformation der Konfidenzintervalle

Zur Definition der transformierten Konfidenzintervalle muss vorerst die Verteilung der transformierten Effekte hergeleitet werden. Unter zweimaliger Verwendung des multivariaten

Delta-Satzes mit den zwei Jakobi-Matrizen

$$d\phi_{\mathbf{p}} = \begin{pmatrix} dg(p_0^{(1)}) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & dg(p_k^{(d)}) \end{pmatrix} \text{ und}$$

$$d\phi_{\mathbf{p}_T^*} = \begin{pmatrix} dg^{-1}(p_T^{(1)*}) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & dg^{-1}(p_T^{(d)*}) \end{pmatrix}$$

gelangt man zu folgendem Ergebnis:

Satz 5.3 (Asymptotische multivariate Normalverteilung)

Es gelten die Annahmen (A1) und (A2) auf S. 13, $g : (0,1) \rightarrow \mathbb{R}$ und $g^{-1} : \mathbb{R} \rightarrow (0,1)$ seien stetig differenzierbar. Weiterhin sei \mathbf{p} definiert wie in Abschnitt 3.4, mit $p_j^{(\ell)} \in (0,1)$. Dann gilt für den Vektor $\tilde{\mathbf{p}}^* = (\tilde{p}^{(1)*}, \dots, \tilde{p}^{(d)*})$ folgende Verteilung:

$$\sqrt{N} (\hat{\mathbf{p}}^* - \tilde{\mathbf{p}}^*) \sim \text{MVN}(\mathbf{0}, \tilde{\mathbf{V}}^*)$$

mit der Kovarianzmatrix $\tilde{\mathbf{V}}^* = d\phi_{\mathbf{p}_T^*}^{-1} \Gamma_T d\phi_{\mathbf{p}} \Sigma^* d\phi_{\mathbf{p}}' \Gamma_T' d\phi_{\mathbf{p}_T^*}^{-1'}$, dem Vektor der adjustierten Effekte $\tilde{\mathbf{p}}^* = g^{-1}(\hat{\mathbf{p}}_T^*)$ und dem dazugehörigen entsprechenden Schätzer.

Beweis:

Der Vektor der adjustierten relativen Effekte ist definiert als $\mathbf{p}^* = \Gamma \mathbf{p}$ (siehe Proposition 3.10) und es gilt $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}) \xrightarrow{\mathcal{D}} \text{MVN}(\mathbf{0}, \Sigma^*)$ mit $\Sigma^* = \text{Cov}(\hat{\mathbf{p}})$ (analog zum Beweis zu Satz 4.6). Weiterhin ist die Abbildung

$$\phi : (0,1)^{d*(k+1)} \rightarrow \mathbb{R}^{d*(k+1)}, \begin{pmatrix} p_0^{(1)} \\ \vdots \\ p_k^{(d)} \end{pmatrix} \mapsto \begin{pmatrix} g(p_0^{(1)}) \\ \vdots \\ g(p_k^{(d)}) \end{pmatrix}$$

differenzierbar und besitzt die Jacobische Funktionalmatrix $d\phi_{\mathbf{p}}$. Wegen Satz 4.6, und weil außerdem gilt:

$$\sqrt{N} \begin{pmatrix} g(\hat{p}_0^{(1)}) - g(p_0^{(1)}) \\ \vdots \\ g(\hat{p}_k^{(d)}) - g(p_k^{(d)}) \end{pmatrix} = \sqrt{N} \left(d\phi \begin{pmatrix} \hat{p}_0^{(1)} \\ \vdots \\ \hat{p}_k^{(d)} \end{pmatrix} - \phi \begin{pmatrix} p_0^{(1)} \\ \vdots \\ p_k^{(d)} \end{pmatrix} \right),$$

gelangt man mit dem multivariaten Delta-Satz (siehe van der Vaart, 1998 [39]) zu folgender

Verteilungskonvergenz:

$$\begin{aligned} \sqrt{N} (g(\hat{\mathbf{p}}) - g(\mathbf{p})) &\xrightarrow{\mathcal{D}} \text{MVN}(\mathbf{0}, \mathbf{d}\phi_{\mathbf{p}} \boldsymbol{\Sigma}^* \mathbf{d}\phi'_{\mathbf{p}}) \\ \Rightarrow \sqrt{N} (\boldsymbol{\Gamma}_T g(\hat{\mathbf{p}}) - \boldsymbol{\Gamma}_T g(\mathbf{p})) &\xrightarrow{\mathcal{D}} \text{MVN}(\mathbf{0}, \boldsymbol{\Gamma}_T \mathbf{d}\phi_{\mathbf{p}} \boldsymbol{\Sigma}^* \mathbf{d}\phi'_{\mathbf{p}} \boldsymbol{\Gamma}'_T) \end{aligned}$$

mit $\boldsymbol{\Gamma}_T \mathbf{d}\phi_{\mathbf{p}} \boldsymbol{\Sigma}^* \mathbf{d}\phi'_{\mathbf{p}} \boldsymbol{\Gamma}'_T = \mathbf{V}_T^*$. Das heißt

$$\sqrt{N} (\hat{\mathbf{p}}_T^* - \mathbf{p}_T^*) \xrightarrow{\mathcal{D}} \text{MVN}(\mathbf{0}, \mathbf{V}_T^*).$$

Werden die Schätzer nun rücktransformiert, erhält man, unter nochmaliger Verwendung des Deltasatzes, als Verteilung:

$$\begin{aligned} \sqrt{N} (g^{-1}(\hat{\mathbf{p}}_T^*) - g^{-1}(\mathbf{p}_T^*)) &\xrightarrow{\mathcal{D}} \text{MVN}\left(\mathbf{0}, \mathbf{d}\phi_{\mathbf{p}_T^*}^{-1} \boldsymbol{\Gamma}_T \mathbf{d}\phi_{\mathbf{p}} \boldsymbol{\Sigma}^* \mathbf{d}\phi'_{\mathbf{p}} \boldsymbol{\Gamma}'_T \mathbf{d}\phi_{\mathbf{p}_T^*}^{-1'}\right) \\ \Rightarrow \sqrt{N} (\hat{\mathbf{p}}^* - \tilde{\mathbf{p}}^*) &\xrightarrow{\mathcal{D}} \text{MVN}(\mathbf{0}, \tilde{\mathbf{V}}^*). \end{aligned}$$

□

Die einzelnen Elemente der Kovarianzmatrix werden bezeichnet mit $\tilde{v}^{(\ell,r)*}$.

Definition 5.4 (Transformiertes Konfidenzintervall)

Das transformierte Konfidenzintervall ist wegen der asymptotischen Normalverteilung:

$$\left[\hat{p}_{\mathbf{u}}^{(\ell)*}, \hat{p}_{\mathbf{o}}^{(\ell)*} \right] = \left(\left[\hat{p}^{(\ell)*} \pm \frac{\sqrt{\hat{v}^{(\ell,\ell)*}} \cdot q_{1-\alpha/2}}{\sqrt{N}} \right] \right).$$

Dabei ist $\hat{v}^{(\ell,\ell)*}$ definiert als empirischer Schätzer von $\tilde{v}^{(\ell,\ell)*}$ und $q_{1-\alpha/2}$ ist das entsprechende Quantil. Das heißt, für die Normalapproximation ist $q_{1-\alpha/2} = z_{1-\alpha/2}$ und für die t-Approximation ist $q_{1-\alpha/2} = t_{\hat{f}^{(\ell)*}, 1-\alpha/2}$ mit $\hat{f}^{(\ell)*}$ wie in Abschnitt 4.8.2 auf S. 38

Korollar 5.5 (Eigenschaften der transformierten Punkt- und Intervallschätzer)

Unter den in Theorem 5.3 genannten Voraussetzungen gilt:

- (1) Der geschätzte transformierte Effekt $\hat{p}^{(\ell)*}$ (siehe (5.1)) ist asymptotisch erwartungstreu für $p^{(\ell)*}$.
- (2) Unter der zusätzlichen Bedingung, dass $\tilde{v}^{(\ell,r)*} > 0$ ist, ist für alle $\alpha \in (0,1)$

$$\left[\hat{p}_{\mathbf{u}}^{(\ell)*}, \hat{p}_{\mathbf{o}}^{(\ell)*} \right] \subset [0,1]$$

ein asymptotisches transformiertes $(1 - \alpha)$ -Konfidenzintervall für $p^{(\ell)*}$.

Beweis:

Zu (1): Aus der Voraussetzung, dass g^{-1} stetig ist und aus Satz 5.3 ergibt sich die asymptotische Erwartungstreue.

Zu (2): Die Behauptung folgt aus

$$P\left(p^{(\ell)*} \in \left[\widehat{p}_{\mathbf{u}}^{(\ell)*}, \widehat{p}_{\mathbf{o}}^{(\ell)*}\right]\right) = P\left(q_{\alpha/2} \leq \sqrt{N} \frac{\widehat{p}^{(\ell)*} - \widetilde{p}^{(\ell)*}}{\widehat{v}^{(\ell,r)*}} \leq q_{1-\alpha/2}\right) \rightarrow 1 - \alpha.$$

□

Die asymptotische Varianz von $p_T^{(\ell)*}$ wird minimal, wenn der Vektor $\boldsymbol{\gamma}_T^{(\ell)} = (\gamma_{1,T}^{(\ell)}, \dots, \gamma_{k,T}^{(\ell)})'$ folgendermaßen gewählt wird:

$$\boldsymbol{\gamma}_T^{(\ell)} = d\boldsymbol{\phi}_{\mathbf{p}^{(\ell)}[1:k]} \boldsymbol{\gamma}^{(\ell)} d\boldsymbol{\phi}_{\mathbf{p}^{(\ell)}[0]}.$$

Dabei sei $\boldsymbol{\gamma}^{(\ell)}$ wie in Definition 4.10, $d\boldsymbol{\phi}_{\mathbf{p}^{(\ell)}[0]} = dg(p_0^{(\ell)})$ und

$$d\boldsymbol{\phi}_{\mathbf{p}^{(\ell)}[1:k]} = \begin{pmatrix} dg(p_1^{(\ell)}) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & dg(p_k^{(\ell)}) \end{pmatrix}.$$

Der Schätzer lautet entsprechend:

$$\widehat{\boldsymbol{\gamma}}_T^{(\ell)} = d\boldsymbol{\phi}_{\mathbf{p}^{(\ell)}[1:k]} \widehat{\boldsymbol{\gamma}}^{(\ell)} d\boldsymbol{\phi}_{\mathbf{p}^{(\ell)}[0]}.$$

Übliche Transformationsmethoden sind die Logit- und die Probit-Transformation. Im nächsten Abschnitt wollen wir speziell die Anwendung der Logit-Transformation vorstellen.

5.3 Logit-Transformation

Christophliemk ([16] 2001) kam bei einem Vergleich der Logit- und der Probit-Transformation zu dem Schluss, dass die Logit-Transformation insgesamt das Niveau besser einhält. Außerdem wird häufig die Logit-Transformation verwendet, da sie aus der logistischen Regression bekannt ist, bei Simulationen zu guten Ergebnissen führt und einfach anzuwenden ist. Dabei

gilt:

$$g(p_j^{(\ell)}) = \log \left(\frac{p_j^{(\ell)}}{1 - p_j^{(\ell)}} \right)$$

$$dg(p_j^{(\ell)}) = \frac{1}{p_j^{(\ell)}(1 - p_j^{(\ell)})}$$

$$g^{-1}(p^{(\ell)*}) = \frac{\exp(p^{(\ell)*})}{\exp(p^{(\ell)*}) + 1} .$$

Die Jacobi-Matrizen hierbei sind dann:

$$d\phi_{\mathbf{p}} = \begin{pmatrix} \frac{1}{p_0^{(1)}(1-p_0^{(1)})} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{p_k^{(d)}(1-p_k^{(d)})} \end{pmatrix} \text{ und}$$

$$d\phi_{\mathbf{p}_T^*} = \begin{pmatrix} \frac{\exp(p_T^{(1)*})}{(\exp(p_T^{(1)*})+1)^2} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{\exp(p_T^{(d)*})}{(\exp(p_T^{(d)*})+1)^2} \end{pmatrix} .$$

Zu beachten ist, dass die Transformation im Fall vollständig getrennter Statusgruppen nicht definiert ist, das heißt, immer dann, wenn gilt: $\hat{p}_j^{(\ell)} = 0$ oder $\hat{p}_j^{(\ell)} = 1$, mit $j = 0, \dots, k$. Zu beheben ist das Problem, indem in diesen Fällen ein $\varepsilon > 0$ subtrahiert bzw. addiert wird¹. Das bedeutet:

$$\hat{p}_j^{(\ell)} = \begin{cases} \varepsilon & \text{wenn } \hat{p}_j^{(\ell)} = 0 \\ 1 - \varepsilon & \text{wenn } \hat{p}_j^{(\ell)} = 1 \\ \hat{p}_j^{(\ell)} & \text{sonst} \end{cases}$$

Für das Konfidenzintervall gilt:

$$\begin{bmatrix} \hat{p}_U^{(\ell)*} & \hat{p}_O^{(\ell)*} \end{bmatrix} = \hat{p}^{(\ell)*} \pm \frac{\sqrt{\hat{v}^{(\ell, \ell)*}} \cdot q_{1-\alpha/2}}{\sqrt{N}} ,$$

mit $\hat{v}^{(\ell, \ell)*}$ und $q_{1-\alpha/2}$ wie oben.

Im folgenden Abschnitt sollen noch für die zwei Beispiele die transformierten Schätzer und Intervalle angegeben und interpretiert werden.

¹Im Makro (siehe Anhang A) ist $\varepsilon = 1e^{-5}$ voreingestellt.

5.4 Beispiele

Meningitis-Studie

Wenn man die Abbildung 5.1 mit der aus Abschnitt 4.8 vergleicht, fällt auf, dass der Unterschied zwischen den adjustierten und unadjustierten Effekten nach der Transformation geringer ist. Außerdem zeigt sich jetzt auch die erwartete geringere Varianz. Die Konfi-

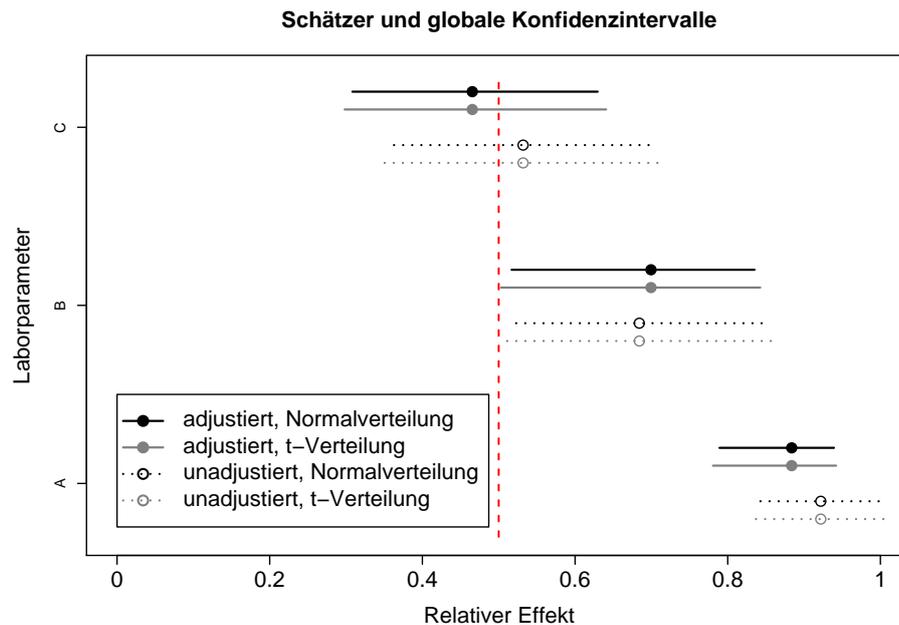


Abbildung 5.1: *Meningitis-Studie, transformierte Konfidenzintervalle.*

denzintervalle sind hier nicht für die Multiplizität der Fragestellungen adjustiert, nur bei den in Kapitel 9.2 angesprochenen simultanen Konfidenzintervallen wird diesem Problem Rechnung getragen. Obwohl aufgrund dieser Problematik keine endgültigen Aussagen über Signifikanz getroffen werden können, ist zu erwarten, dass der Effekt des Laborparameters C nicht signifikant von 0.5 verschieden ist, der des Laborparameters A dagegen schon. Bei dem Laborparameter B liegen die Konfidenzintervalle sehr nah an der Signifikanzgrenze oder berühren sie. Hier wird erst der Test Gewissheit bringen.

Zur Verdeutlichung der Reduzierung der Streuung sind in Tabelle 5.1 für die einzelnen Parameter die jeweiligen Breiten der Konfidenzintervalle zum Vergleich angegeben. Man sieht, dass bei dem Laborparameter C die Varianz am deutlichsten verringert wurde.

Tabelle 5.1: *Meningitis-Studie, Breite der transformierten Konfidenzintervalle.*

	CRP	Temp.	τ -Spiegel
unadjustiert	0.1581	0.3228	0.3382
adjustiert	0.1500	0.3182	0.3213

CAD-Studie

In Tabelle 5.2 sind die für die Untersuchungszeit adjustierten geschätzten relativen Effekte angegeben: untransformiert, analog zu Tabelle 4.2 in Abschnitt 4.8, und transformiert. Es fällt auf, dass die transformierten Effekte tendenziell größer sind, und je näher der Effekt an 0.5 ist, desto stärker wirkt sich die Transformation aus. In der Tabelle 5.3 sind die Breiten

Tabelle 5.2: *CAD-Studie, Vergleich der untransformierten und transformierten geschätzten relativen Effekte.*

		untransformiert	transformiert
mit CAD	Assistenzarzt	0.8267	0.8275
	Student	0.6567	0.7069
	Facharzt	0.9864	0.9864
ohne CAD	Assistenzarzt	0.8172	0.8185
	Student	0.6284	0.6427
	Facharzt	0.9685	0.9685

der Konfidenzintervalle der relativen Effekte bei den verschiedenen Bedingungen angegeben. Man sieht, dass für den Facharzt die Kovariablen-Adjustierung sogar zu breiteren Konfidenzintervallen führt. Bei dem Studenten sind die Intervalle der adjustierten Effekte allerdings wesentlich, und bei dem Assistenzarzt etwas schmaler als die der unadjustierten. Außerdem wird die Streuung der Ergebnisse bei Verwendung des CAD-Systems geringer. In Abbildung 5.2 sind die transformierten Punkt- und Intervallschätzer noch grafisch dargestellt. Wieder unter Berücksichtigung des Multiplizität-Problems scheint für alle Bedingungen der Effekt signifikant von 0.5 verschieden zu sein. Außerdem sieht es so aus, als ob die Sicherheit der Diagnose des Facharztes signifikant besser ist als die des Assistenzarztes oder des Studenten. Die hier gezeigten Punkt- und Intervallschätzer vermitteln optisch einen ersten Eindruck, für exakte Aussagen über Signifikanzen benötigen wir allerdings die Hypothesentests, die im nächsten Kapitel besprochen werden.

Tabelle 5.3: CAD-Studie, Breite der transformierten Konfidenzintervalle.

	mit CAD			ohne CAD		
	Ass.arzt	Student	Facharzt	Ass.arzt	Student	Facharzt
unadj.	0.0876	0.1102	0.0249	0.0976	0.1270	0.0419
adj.	0.0818	0.0930	0.0285	0.0947	0.1109	0.0446

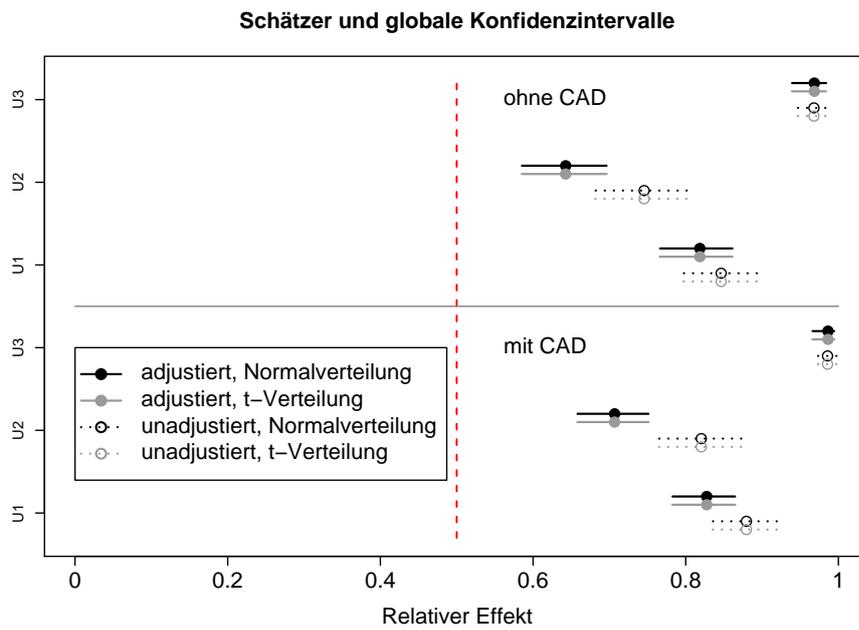


Abbildung 5.2: CAD-Studie, transformierte Konfidenzintervalle.

6 Hypothesentests

In diesem Kapitel beschäftigen wir uns mit Hypothesentests. Im ersten Abschnitt sollen die zwei hier verwendeten Teststatistiken vorgestellt werden. In Abschnitt 6.2 werden diese dann auf die Hypothese angewendet, dass unter mindestens einer Bedingung der bezüglich Kovariablen adjustierte Effekt signifikant von 0.5 verschieden ist und in Abschnitt 6.3 dann auf die Hypothese, dass sich der adjustierte Effekt von mindestens einer Bedingung signifikant von den anderen unterscheidet. Die Tests werden immer auf die rücktransformierten adjustierten relativen Effekte angewendet.

6.1 Teststatistiken

Um die verschiedenen Hypothesen zu überprüfen, betrachten wir in dieser Arbeit zwei Teststatistiken: die Wald- und die ANOVA-Teststatistik. Die beiden Tests sollen im Folgenden einzeln vorgestellt werden. Die Eigenschaften der Tests werden dann bei den Simulationen in Kapitel 7 untersucht.

1. Wald-Test

Bereits 1988 wurde von DeLong et al. (1988, [17]) die Wald-Typ-Statistik (WTS) für den Vergleich unabhängiger Beobachtungen vorgeschlagen. Unter der Voraussetzung, dass für die Kovarianzmatrix $\tilde{\mathbf{V}}_N^*$ gilt: $\tilde{\mathbf{V}}_N^* \rightarrow \tilde{\mathbf{V}}^* \neq 0$ für $N \rightarrow \infty$, sodass $\text{Rg}(\mathbf{C}\tilde{\mathbf{V}}_N^*) = \text{Rg}(\mathbf{C}\tilde{\mathbf{V}}^*)$ für alle $N \geq N_0$, ist die WTS Q_N unter H_0 asymptotisch χ^2 -verteilt. Da der Wald-Test allerdings bei kleinen Stichproben und/oder vielen Faktorstufen zu liberalen Ergebnissen führt und eine nicht-singuläre Kovarianzmatrix Voraussetzung ist, wollen wir als nächstes den ANOVA-Test vorstellen.

2. ANOVA-Test

Die ANOVA-Typ-Statistik (ATS) erhält man, wenn man bei der WTS den Varianzschätzer durch eine bekannte Matrix (hier die Einheitsmatrix \mathbf{I}) ersetzt. Die ATS F_N ist unter der Nullhypothese asymptotisch χ^2 -verteilt. Die Approximation der Verteilung stammt von Box (1954, [6]), Brunner et al. ([9]) haben die ATS 1997 erstmals in der Nichtparametrik verwendet. Viele Aufsätze haben gezeigt, dass die ATS bei kleinen Stichproben bzw. vielen Faktorstufen das vorgegebene Niveau besser einhält als die WTS und dabei eine vergleichbare oder höhere Power erzielt (siehe z.B. Brunner et al., 2002 [14]).

6.2 Test auf Vorliegen eines Effekts

Man kann testen, ob ein Effekt vorliegt, das bedeutet ob unter mindestens einer Bedingung der Effekt signifikant von 0.5 verschieden ist. Die Hypothese lautet dann:

$$H_0 : \tilde{\mathbf{p}}^* = \frac{1}{2} \mathbf{1}_d ,$$

mit $\tilde{\mathbf{p}}^* = (\tilde{p}^{(1)*}, \dots, \tilde{p}^{(d)*})$. Dann sind die Teststatistiken und ihre Verteilungen der Tests:

1. Wald-Test

Die Wald-Teststatistik lautet bei dieser Hypothese:

$$Q_N = N \left(\hat{\tilde{\mathbf{p}}}^* - \frac{1}{2} \mathbf{1}_d \right)' \mathbf{I}_d \left(\mathbf{I}_d \hat{\tilde{\mathbf{V}}}_N \mathbf{I}_d \right)^+ \mathbf{I}_d \left(\hat{\tilde{\mathbf{p}}}^* - \frac{1}{2} \right)' .$$

Dabei bezeichne $()^+$ die Moore-Penrose-Inverse. Es wäre auch jede andere verallgemeinerte g-Inverse verwendbar, mit der Moore-Penrose-Inverse kann man jedoch, insbesondere bei asymptotischer Betrachtung, viele Schwierigkeiten umgehen; unter anderem ist sie numerisch stabil. Unter der Nullhypothese H_0 gilt dann:

$$Q_N \sim \chi_f^2 ,$$

mit $\hat{f} = \text{Rg} \left(\mathbf{I}_d \hat{\tilde{\mathbf{V}}}_N \mathbf{I}_d \right)$.

2. ANOVA-Test

Beim ANOVA-Test lautet in diesem Fall die Teststatistik:

$$F_N = \frac{N}{\text{Sp} \left(\hat{\tilde{\mathbf{V}}}_N \right)} \left(\hat{\tilde{\mathbf{p}}}^* - \frac{1}{2} \right)' \left(\hat{\tilde{\mathbf{p}}}^* - \frac{1}{2} \right) .$$

Unter Hypothese H_0 gilt:

$$F_N \sim \chi_f^2 / f \quad , \quad \text{mit} \\ \hat{f} = \frac{\left[\text{Sp} \left(\hat{\tilde{\mathbf{V}}}_N \right) \right]^2}{\text{Sp} \left(\hat{\tilde{\mathbf{V}}}_N \hat{\tilde{\mathbf{V}}}_N \right)} .$$

6.3 Tests auf gleiche Effekte

Wenn die Zielgröße unter mehr als einer Bedingung festgestellt wurde, kann es außerdem von Interesse sein, zu testen, ob sich die Effekte unter den verschiedenen Bedingungen signifikant voneinander unterscheiden. Dabei können die Kombinationen $\ell = 1, \dots, d$ in die in dieser

Arbeit angenommenen Einzeleffekte Methode $m = 1, \dots, M$ und Reader $r = 1, \dots, R$ mit $d = M \cdot R$ aufgelöst werden. Es gilt also $\ell = (11), (12), \dots, (MR)$ und $\tilde{\mathbf{p}}^* = (\tilde{p}^{(11)*}, \dots, \tilde{p}^{(MR)*})$. Dann gibt es analog zu Kaufmann et al. ([26], 2005) drei Effekte, die getestet werden können:

I) Methodeneffekt:

Differenzieren die Methoden die zwei Statusgruppen unterschiedlich gut? Dafür lautet die Nullhypothese: $H_0^m : \bar{p}^{*(1.)} = \dots = \bar{p}^{*(M.)}$ mit $\bar{p}^{*(m.)} = \frac{1}{R} \sum_{r=1}^R \tilde{p}^{*(mr)}$. Dann ist die entsprechende Kontrastmatrix: $\mathbf{C}_M = \mathbf{P}_M \otimes \frac{1}{R} \mathbf{1}'_R$ mit $\mathbf{P}_M = \mathbf{I}_M - \frac{1}{M} \mathbf{J}_M$. Die Nullhypothese kann damit geschrieben werden als: $H_0^m : \mathbf{C}_M \tilde{\mathbf{p}}^* = \mathbf{0}$.

II) Readereffekt:

Differenzieren die Untersucher unterschiedlich gut? Dabei lautet die Nullhypothese: $H_0^r : \bar{p}^{*(.1)} = \dots = \bar{p}^{*(.R)}$ mit $\bar{p}^{*(.r)} = \frac{1}{M} \sum_{m=1}^M \tilde{p}^{*(mr)}$. Die Kontrastmatrix ist dann: $\mathbf{C}_R = \frac{1}{M} \mathbf{1}'_M \otimes \mathbf{P}_R$ mit $\mathbf{P}_R = \mathbf{I}_R - \frac{1}{R} \mathbf{J}_R$. Die Nullhypothese kann dann also zu $H_0^r : \mathbf{C}_R \tilde{\mathbf{p}}^* = \mathbf{0}$ umformuliert werden.

III) Interaktion Methode-Reader:

Sind die Beurteilungen der Reader homogen bei den unterschiedlichen Methoden?

Dabei lautet dann die Nullhypothese: $H_0^{mr} : \tilde{p}_{mr}^* = \bar{p}^{*(m.)} + \bar{p}^{*(.r)} - \bar{p}^{*(.)}$ mit $\bar{p}^{*(.)} = \frac{1}{M} \frac{1}{R} \sum_{m=1}^M \sum_{r=1}^R \tilde{p}^{*(mr)}$. Anders dargestellt als $H_0^{mr} : \mathbf{C}_{MR} \tilde{\mathbf{p}}^* = \mathbf{0}$ mit $\mathbf{C}_{MR} = \mathbf{P}_M \otimes \mathbf{P}_R$.

Nun sollen für die verschiedenen Tests wieder die Teststatistiken und ihre Verteilungen angegeben werden. Dabei sei \mathbf{C} je nach Hypothese \mathbf{C}_M , \mathbf{C}_R oder \mathbf{C}_{MR} .

1. Wald-Test

Die Teststatistik des Wald-Typs sieht hier folgendermaßen aus:

$$Q_N = N \hat{\tilde{\mathbf{p}}}^{*'} \mathbf{C}' \left(\mathbf{C} \hat{\tilde{\mathbf{V}}}_N \mathbf{C}' \right)^+ \mathbf{C} \hat{\tilde{\mathbf{p}}}^* .$$

Unter Hypothese H_0 gilt dann für große Stichproben:

$$Q_N \sim \chi_{\hat{f}}^2 ,$$

mit $\hat{f} = \text{Rg}(\mathbf{C} \hat{\tilde{\mathbf{V}}}_N \mathbf{C}')$.

2. ANOVA-Typ-Statistik

Die Teststatistik des ANOVA-Typs ist:

$$F_N = \frac{N}{\text{Sp}(\mathbf{T} \hat{\tilde{\mathbf{V}}})} \hat{\tilde{\mathbf{p}}}^{*'} \mathbf{T} \hat{\tilde{\mathbf{p}}}^* ,$$

dabei ist $\mathbf{T} = \mathbf{C}' (\mathbf{C} \mathbf{C}')^{-} \mathbf{C}$. Unter Hypothese H_0 gilt:

$$F_N \sim \chi_{\hat{f}}^2 / f ,$$

$$\text{mit } \hat{f} = \frac{(\text{Sp}(\mathbf{T}\hat{\mathbf{V}}^*))^2}{\text{Sp}(\mathbf{T}\hat{\mathbf{V}}^* \mathbf{T}\hat{\mathbf{V}}^*)}.$$

Die hier vorgestellten Tests auf Vorliegen eines Effektes und auf Gleichheit der Effekte sind allgemein und können zur Überprüfung konkreter Hypothesen nicht ausreichend sein. Wenn in der Medizin beispielsweise untersucht werden soll, mit welchem Test eine Krankheit am sichersten diagnostiziert werden kann, bieten sich multiple Kontrasttests und dazugehörige simultane Konfidenzintervalle an. Dies ist ein separates Projekt, ein Ausblick hierauf wird in Kapitel 9.2 auf Seite 70 gegeben.

6.4 Beispiele

Meningitis-Studie

In dieser Studie war es maßgeblich herauszufinden, ob einer der Laborparameter (A, B und C) zur Differenzierung der beiden Formen der Meningitis geeignet ist. Hierfür verwenden wir den Globaltest auf Vorliegen eines Effektes aus Abschnitt 6.2, α wird hier und bei allen folgenden Tests auf 0.05 festgesetzt. Der p-Wert ist bei beiden Teststatistiken und unabhängig von der Alters-Adjustierung kleiner 0.0001. Der Effekt mindestens eines Parameters ist also von 0.5 verschieden. Das Ergebnis ist konsistent mit dem der Konfidenzintervalle in Abschnitt 5.4 (siehe Abbildung 5.1).

CAD-Studie

Hierbei standen für die Kliniker folgende Fragen im Vordergrund:

1. Kommen die Untersucher zu unterschiedlichen Ergebnissen?
2. Verbessert das CAD-System die Diagnose?
3. Besteht eine Wechselwirkung zwischen Untersucher und CAD-System?

Diese Fragen lassen sich mit den Tests auf Gleichheit der Effekte aus Abschnitt 6.3 beantworten. In Tabelle 6.1 sind die Ergebnisse der Tests angegeben, sowohl mit als auch ohne Adjustierung bezüglich der Befundungszeit. Der Einfluss der Faktoren Untersucher und Methode ist hochsignifikant, die Unterschiede zwischen adjustierten und unadjustierten sowie zwischen den verschiedenen Teststatistiken sind marginal. Es kann zu $\alpha = 5\%$ keine signifikante Wechselwirkung zwischen Untersucher und Methode gezeigt werden, allerdings ist die Interaktion nur knapp nicht signifikant.

Tabelle 6.1: *CAD-Studie, p-Werte der Tests auf Gleichheit der Effekte.*

	Test	Methode	Untersucher	Interaktion
unadjustiert	WTS	< 0.0001	< 0.0001	0.068
	ATS	< 0.0001	< 0.0001	0.059
adjustiert	WTS	0.004	< 0.0001	0.131
	ATS	0.004	< 0.0001	0.056

7 Simulationen

Um die Eigenschaften der entwickelten Verfahren zu untersuchen wurden Niveau- und Powersimulationen durchgeführt. Bei den Powersimulationen wurde das Verhalten bei einer Verschiebung der Zielvariable und bei einer Verschiebung der Kovariable untersucht. Die Anzahl der Simulationen pro Situation ist 10 000.

Grundsätzlich waren die Einstellungen: Korrelation zwischen Ziel- und Kovariable $\rho = 0.8$, Stichprobenumfänge balanciert $n_i = 15$, eine Kovariable, Normalverteilung $N(0,1)$ für Ziel- und Kovariablen, sechs Bedingungen ($d = 6$) und $\alpha = 0.05$. Davon können alle Parameter beliebig variiert werden, wir haben uns hier aber auf die Variation der Stichprobenumfänge im balancierten und im unbalancierten Fall, der Anzahl der Kovariablen und der Anzahl der Bedingungen beschränkt. Wenn nichts anderes angegeben ist, gelten für die nicht variierten Parameter die Grundeinstellungen.

Als erstes werden die beiden verschiedenen Teststatistiken hinsichtlich der Einhaltung des Niveaus und der erzielten Power verglichen.

Tabelle 7.1: *Variation der Stichprobenumfänge, balanciert.*

n_i	adjustiert		unadjustiert	
	WTS	ATS	WTS	ATS
10	80.53	93.05	81.08	93.35
20	90.50	95.02	88.97	95.18
30	92.89	95.28	91.62	94.96
50	93.81	95.49	92.50	94.82

Dafür wurde ein balanciertes Design mit verschiedenen Stichprobenumfängen simuliert. Für die anderen Variablen gelten die Grundeinstellungen (siehe oben). In Tabelle 7.1 sieht man, dass der ANOVA-Test das vorgegebene Niveau besser einhält als der Wald-Test. In den verschiedenen untersuchten Szenarien wiederholt sich dieses Ergebnis. Weiterhin sieht man in Tabelle 7.1, dass bei sechs Bedingungen bereits für $n_i = 20$ das Niveau sicher eingehalten wird, bei $n_i = 10$ sind die Ergebnisse tendenziell liberal. Außerdem halten die adjustierten relativen Effekte das Niveau etwas besser ein als die unadjustierten.

Bei dem Vergleich der Teststatistiken hinsichtlich der Power sieht man in Abbildung 7.1, dass der ANOVA-Test eine größere Power erzielt als der Wald-Test¹. Aus diesem Grund werden wir für die verschiedenen Situationen nur die Powerergebnisse dieses Tests vergleichen.

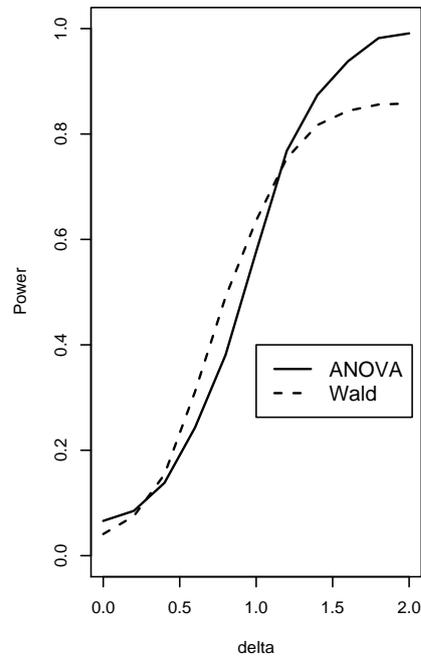


Abbildung 7.1: Vergleich der Teststatistiken.

¹Die Eingaben waren hierbei abweichend von den Grundeinstellungen: $d = 2$ und $n_1 = n_2 = 30$.
Delta ist hier der Shift der Zielvariable.

Anhand der Abbildung 7.2 wird deutlich, wie sehr die Power von der Fallzahl abhängt.

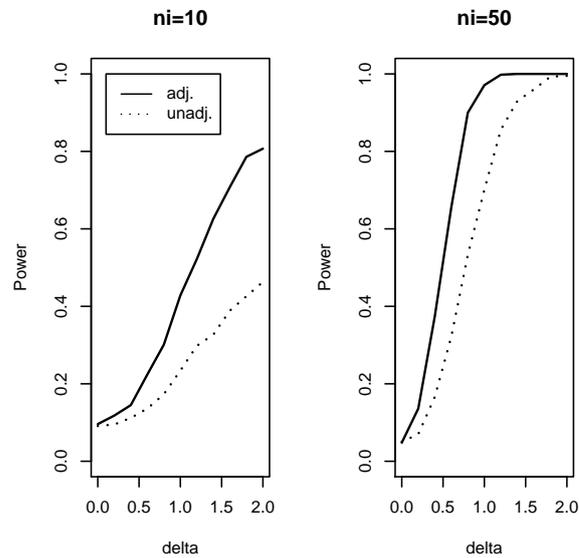


Abbildung 7.2: Variation der Stichprobenumfänge.

Die balancierten Stichprobenumfänge waren hier $n_i = (10,50)$ mit $i = 1,2$, abgebildet ist die Power für den ANOVA-Test. Wie man sehen kann, führt neben der Vergrößerung der Stichprobe auch die Kovariablen-Adjustierung zu einer größeren Power.

Aus Tabelle 7.2 geht hervor, dass die Tests robust gegenüber Unbalanciertheit sind. Hierbei ist $n_1 = 20$ und $n_2 = (20,40,60)$, für die anderen Parameter gelten wieder die Grundeinstellungen von oben. Auch hier halten die adjustierten Effekte das vorgegebene Niveau besser ein als die unadjustierten Effekte.

Tabelle 7.2: Variation des Stichprobenumfangs n_2 bei $n_1 = 20$.

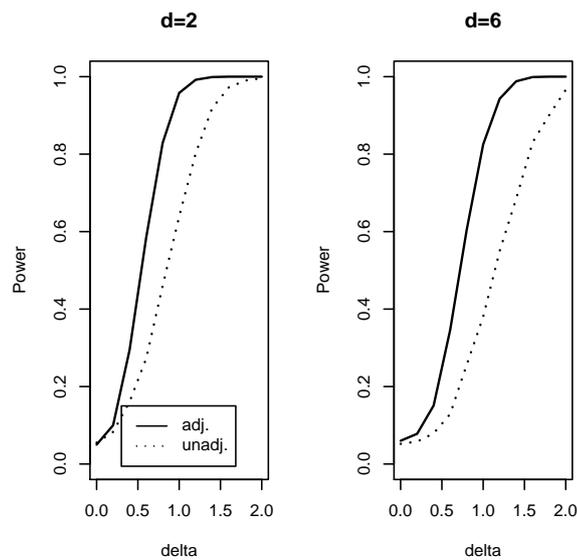
n_2	adjustiert		unadjustiert	
	ATS	Wald	ATS	Wald
20	95.05	94.87	94.92	94.73
40	94.86	94.75	93.93	93.54
60	94.76	94.64	93.87	93.39

Nun soll untersucht werden, inwieweit die Anzahl der Bedingungen einen Einfluss auf die Testergebnisse hat. In Tabelle 7.3 zeigt sich, dass bei $n_i = 15$ im Fall von sechs Bedingungen das Niveau genauso gut eingehalten wird wie bei zwei Bedingungen. Die adjustierten Effekte liegen wieder etwas näher am vorgegebenen α -Level als die unadjustierten.

Tabelle 7.3: *Variation der Anzahl der Bedingungen.*

d	adjustiert		unadjustiert	
	ATS	Wald	ATS	Wald
2	94.97	94.67	94.59	94.54
4	94.30	95.02	94.60	94.13
6	95.25	94.89	95.29	94.86

Die Power wird mit Zunahme der Bedingungen etwas geringer (siehe Abbildung 7.3), allerdings ist der Unterschied zwischen zwei und sechs Bedingungen marginal. Die Powerkurve der adjustierten Effekte ist etwas steiler als die der unadjustierten Effekte.

Abbildung 7.3: *Variation der Bedingungen.*

Im Folgenden wollen wir die Anzahl der Kovariablen variieren. In Tabelle 7.4 wurden zwei bzw. fünf Kovariablen simuliert, die alle in den beiden Statusgruppen die gleiche Verteilung besitzen. Wie man sehen kann, wird für die adjustierten Effekte die Einhaltung des Niveaus mit steigendem k schlechter, während sich bei den unadjustierten Effekten per definitionem nichts ändert.

Tabelle 7.4: *Variation der Anzahl der Kovariablen.*

k	adjustiert		unadjustiert	
	ATS	Wald	ATS	Wald
2	94.90	94.54	94.94	94.76
5	93.84	93.23	94.88	94.82

Bei der Poweranalyse wurde die Auswirkung von zwei bzw. fünf Kovariablen untersucht, einmal mit und einmal ohne Effekt der Kovariable auf die Zielgröße. Dabei wurden nur die adjustierten Effekte untersucht. In Abbildung 7.4 sieht man, dass bei Kovariablen ohne Effekt die Power abnimmt, je mehr Kovariablen berücksichtigt werden.

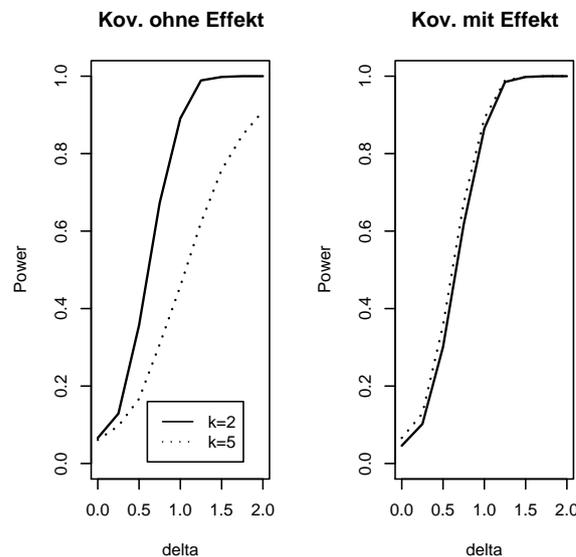


Abbildung 7.4: *Variation der Anzahl der Kovariablen.*

Sind die Kovariablen dagegen in den Statusgruppen unterschiedlich verteilt, ist die Power bei $k = 5$ noch etwas größer als bei $k = 2$.

Nun wollen wir den größten Vorteil der Kovariablen-Adjustierung anhand der Abbildung 7.5 verdeutlichen. Hier wurde die Power bei einem Shift der Zielvariable mit der Power bei einem Shift der Kovariable verglichen. Dabei gelten wieder die Grundeinstellungen wie oben. Bei einem Shift der Zielvariable ist die Power der adjustierten Effekte etwas höher als die der unadjustierten, aber auch die Power der unadjustierten Effekte ist hoch. Bei einem Shift der Kovariable dagegen bleibt die Power der unadjustierten Effekte per definitionem bei $\alpha = 5\%$, während die Power der adjustierten Effekte zügig ansteigt.

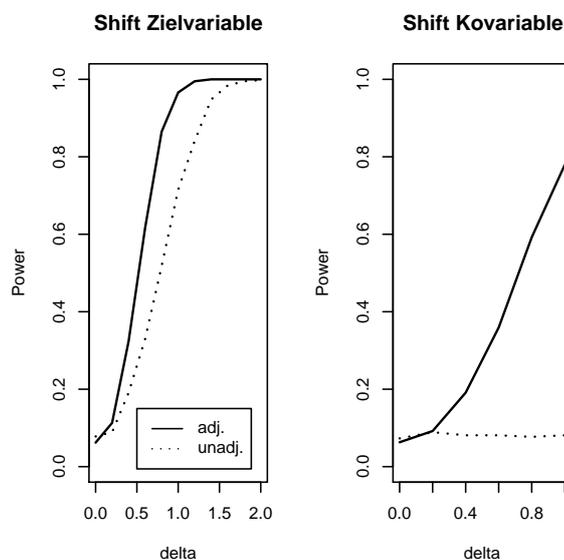


Abbildung 7.5: Vergleich von Shift der Zielvariable und Shift der Kovariable.

In dieser Arbeit wurde anders als z.B. bei Domhof (2001 [18]) nicht angenommen, dass die Kovariablen unter den verschiedenen Bedingungen die gleiche Verteilung besitzen. Deshalb sollen an dieser Stelle die Auswirkungen unterschiedlicher Verteilungen überprüft werden. In Tabelle 7.5 sind beispielhaft die Ergebnisse des Tests auf Gleichheit für die ANOVA-Typ-Statistik angegeben. Dabei galt $n_i = 30$, $r = 3$, $m = 1$ und $k = 1$, und es wurde entweder der Mittelwert verschoben oder die Varianz verändert. Die Ergebnisse entsprechen genau den Ergebnissen der Theorie. Die Kleinste-Quadrate-Schätzer sind erwartungstreu, entsprechend bleibt das empirische Niveau auch bei starker Verschiebung der Erwartungswerte sehr nah an dem vorgegebenen Niveau. Dagegen ist der Kleinste-Quadrate-Schätzer bei unterschiedlichen Varianzen nach dem Satz von Gauss-Markov nicht mehr ein optimaler Schätzer, was sich auch bei den Simulationen zeigt. Allerdings ist selbst bei starken Unterschieden in der Varianz das Ergebnis nur leicht konservativ.

Tabelle 7.5: Modifikation der Verteilung der Kovariablen.

Veränderung der Varianz	empirisches Niveau
$N(0,1.5)$, $N(0,2)$, $N(0,2.5)$	95.99
$N(0,2)$, $N(0,3)$, $N(0,4)$	96.13
Verschiebung des Erwartungswertes	empirisches Niveau
$N(0.5,1)$, $N(1,1)$, $N(1.5,1)$	95.38
$N(1,1)$, $N(2,1)$, $N(3,1)$	95.19

8 Diskussion

In dieser Arbeit wurden Punkt- und Intervallschätzer sowie Hypothesentests für das multivariate nichtparametrische Behrens-Fisher-Problem um eine Adjustierung bezüglich Kovariablen erweitert. Wir konnten für die Konfidenzintervalle zeigen, dass durch die Kovariablen-Adjustierung die Varianz im Allgemeinen verringert wird und die Aussagen dadurch exakter werden. Bei den Hypothesentests führt die Berücksichtigung von Kovariablen zu einer höheren Power bei gleich guter bzw. besserer Einhaltung des Niveaus.

Die Simulationen haben gezeigt, dass die adjustierten relativen Effekte das Niveau im Allgemeinen mindestens genauso gut einhalten wie die unadjustierten. Auch die Power ist bei den adjustierten Effekten höher, vor allem bei Kovariablen mit einem großen Effekt auf die Zielgröße. Lediglich bei vielen Kovariablen mit gleicher Verteilung in den beiden Statusgruppen halten die adjustierten Effekte das Niveau nicht mehr so gut ein. Da Tests auf einen Effekt der Kovariable nicht Bestandteil dieser Arbeit sind, muss bereits bei der Studienplanung entschieden werden, welche Kovariablen berücksichtigt werden sollen. Für das Beispiel der CAD-Studie würde das z.B. bedeuten, dass man bezüglich der Befundungszeit adjustiert, da zu erwarten ist, dass sie zwischen den Untersuchern stark und den Statusgruppen etwas variiert. Tests auf einen signifikanten Einfluss der Kovariablen (siehe z.B. Bathke[2]) wären hier sicherlich eine sinnvolle Erweiterung.

Das Problem bei der Schätzmethode der Regressionskoeffizienten ist, dass die Inverse unter Umständen nicht existiert und damit der Schätzer nicht berechnet werden kann. Das ist dann der Fall, wenn die Verteilung der Kovariablen in den beiden Statusgruppen komplett getrennt ist, es also keinen Überlappungsbereich gibt. Es wäre deshalb denkbar, eine andere Schätzmethode zu verwenden, wovon wir allerdings Abstand genommen haben. Denn wenn die Verteilungen der Kovariablen in den beiden Statusgruppen so unterschiedlich sind, wenn also beispielsweise in der einen Statusgruppe nur Frauen und in der anderen nur Männer sind, ist die Sinnhaftigkeit eines Vergleichs ohnehin fraglich.

Eine andere mögliche Erweiterung wäre noch die Übertragung der Verfahren auf andere Designs. So kann es sein, dass der Faktor Untersucher nicht unter dem Faktor Methode verschachtelt ist, sondern dass die beiden Faktoren gekreuzt sind. Das wäre z.B. der Fall wenn ein CT¹-Spezialist CT-Bilder, und ein MRT²-Spezialist MRT-Bilder begutachtet. Dementsprechend sind verschiedene Designs denkbar, siehe hierzu u.a. Lange (2008, [27]). Auch wäre es sicher von Interesse, die Verfahren auf den Fall von mehr als zwei Statusgruppen zu

¹Computer-Tomographie

²Magnet-Resonanz-Tomographie

übertragen.

Da die entwickelten Verfahren auch für die Praxis anwendbar sein sollen, wird der nächste Schritt sein, das Computerprogramm, das hierfür geschrieben wurde (siehe Anhang A), in Form eines R- und eines SAS-Makros der Öffentlichkeit zugänglich zu machen.

Mit dieser Arbeit ist insbesondere ein wichtiger Beitrag für die Auswertung von Diagnosestudien geleistet worden. Diese Studien gewinnen in der medizinischen Forschung zunehmend an Gewicht, da eine richtige Diagnose die Voraussetzung für eine effektive Therapie ist. Immer wieder kommt es vor, dass erst bei der Auswertung der Daten Kovariablen entdeckt werden, die einen Einfluss auf den diagnostischen Test haben und für die nicht randomisiert wurde. Dieses Problem kann mit den hier entwickelten Verfahren behoben werden. Angesichts der besprochenen möglichen Erweiterungen und Verallgemeinerungen wird außerdem klar, dass dieses Gebiet noch ein großes Entwicklungspotenzial besitzt, das in weiterer Forschungsarbeit ausgeschöpft werden soll.

9 Ausblick

Die in dieser Arbeit vorgestellten Verfahren sind die elementaren, die für Auswertungen klinischer Studien im Allgemeinen nötig sind. Darüber hinaus gibt es unserer Meinung nach noch zwei Themen mit großer Praxisrelevanz. Das sind erstens fehlende Werte und zweitens Kontrasttests und die dazugehörigen simultanen Konfidenzintervalle. In diesem Kapitel soll zu diesen beiden Themen ein kleiner Ausblick gegeben werden.

9.1 Fehlende Werte

Da in unserem Modell keine Messwiederholungen angenommen werden, bedeutet ein fehlender Wert, dass für ein Subjekt unter einer oder mehreren Bedingungen kein Response vorliegt. Speziell für den Fall, dass dieses Fehlen rein zufällig ist (sogenanntes missing completely at random), wollen wir das Modell dahingehend erweitern, dass fehlende Werte zugelassen werden können. Am Ende dieses Abschnitts werden Nutzen und Risiken der Berücksichtigung von rein zufällig fehlenden Werten diskutiert.

Rein zufälliges Fehlen liegt zum Beispiel dann vor, wenn aus labortechnischen Gründen ein Ergebnis fehlt, oder wenn prinzipiell jedes Subjekt nur unter einem Teil der Bedingungen untersucht wird und die Zuordnung zufällig erfolgt. Wir wollen hier weder das sogenannte missing at random berücksichtigen, noch das sogenannte informative missing, wenn also der Grund für das Fehlen mit dem Outcome zusammenhängt. Das wäre z.B. der Fall, wenn man Patienten, denen es besonders schlecht geht, nicht so viele verschiedene Verfahren zumuten will. Mit der Lösung dieses Problems beschäftigten sich unter anderem Domhof et al. (2002) [19] und Akritas et al. (2002) [1].

Um den Fall fehlender Werte zu berücksichtigen, führen Brunner et al. (1999) [13] folgenden Indikator ein:

$$\lambda_{j, is}^{(\ell)} = \begin{cases} 1 & \text{falls } X_{j, is}^{(\ell)} \text{ vorliegt} \\ 0 & \text{falls } X_{j, is}^{(\ell)} \text{ fehlt} \end{cases}$$

mit $j = 0, \dots, k$. Dann ist $n_{j, i}^{(\ell)} = \sum_{s=1}^{n_i} \lambda_{j, is}^{(\ell)}$ die Anzahl der vorhandenen Beobachtungen bei Response bzw. Kovariable j unter Bedingung ℓ in Statusgruppe i . Die Gesamtanzahl der Beobachtungen in Kombination ℓ für $j = 0, \dots, k$ ist entsprechend $N_j^{(\ell)} = n_{j, 1}^{(\ell)} + n_{j, 2}^{(\ell)}$. Dann sind die empirischen Verteilungsfunktionen $\hat{F}_{j, i}^{(\ell)}$ in der Kombination ℓ folgendermaßen

definiert:

$$\widehat{F}_{j,i}^{(\ell)}(x) = \frac{1}{n_{j,i}^{(\ell)}} \sum_{s=1}^{n_i} \lambda_{j,is}^{(\ell)} c\left(x - X_{j,is}^{(\ell)}\right).$$

Damit kommt man zu folgender Definition der einzelnen geschätzten relativen Effekte:

$$\widehat{p}_j^{(\ell)} = \frac{1}{N_j^{(\ell)}} \left(\overline{R}_{j,2}^{(\ell)} - \overline{R}_{j,1}^{(\ell)} \right) + \frac{1}{2}, \quad j = 0, \dots, k.$$

Dabei gilt $\overline{R}_{j,i}^{(\ell)} = \frac{1}{n_{j,i}^{(\ell)}} \sum_{s=1}^{n_i} \lambda_{j,is}^{(\ell)} R_{j,is}^{(\ell)}$, und $R_{j,i}^{(\ell)}$ ist der Rang von $X_{j,i}^{(\ell)}$ unter allen $N_j^{(\ell)}$ beobachteten Werten. Dann ist der Schätzer des adjustierten relativen Effekts:

$$\begin{aligned} \widehat{p}^{(\ell)*} &= \widehat{p}_0^{(\ell)} - \sum_{j=1}^k \widehat{\gamma}_j^{(\ell)} \left(\widehat{p}_j^{(\ell)} - 0.5 \right) \\ &= \frac{1}{N_0^{(\ell)}} \left(\overline{R}_{0,2}^{(\ell)} - \overline{R}_{0,1}^{(\ell)} \right) + \frac{1}{2} - \sum_{j=1}^k \widehat{\gamma}_j^{(\ell)} \left(\frac{1}{N_j^{(\ell)}} \left(\overline{R}_{j,2}^{(\ell)} - \overline{R}_{j,1}^{(\ell)} \right) \right). \end{aligned}$$

Bei der Schätzung der Regressionskoeffizienten (siehe Definition 4.2, S. 29) ändern sich nur die Definitionen von $\widehat{Y}_{j,is}^{(\ell)}$ und $\widehat{Y}_{j,i}^{(\ell)}$:

$$\widehat{Y}_{j,is}^{(\ell)} = \frac{1}{N_j^{(\ell)} - n_{j,i}^{(\ell)}} \lambda_{j,is}^{(\ell)} \left(R_{j,is}^{(\ell)} - R_{j,i}^{(\ell)} \right)$$

und

$$\widehat{Y}_{j,i}^{(\ell)} = \frac{1}{n_{j,i}^{(\ell)}} \sum_{s=1}^{n_i} \widehat{Y}_{j,is}^{(\ell)}, \quad j = 0, \dots, k.$$

Es gelten alle bisherigen Ergebnisse, die für einen kompletten Datensatz hergeleitet wurden, wenn die kleinste Anzahl beobachteter Werte für die Variable j in der Kombination ℓ groß genug ist. Das heißt, wenn gilt:

$$\min n_{j,i}^{(\ell)} \rightarrow \infty, \text{ für alle } \ell = 1, \dots, d, \quad j = 1, \dots, k, \text{ und } i = 1, 2,$$

so dass $\frac{N}{n_{j,i}^{(\ell)}} \leq N_0 < \infty$.

Für die Schätzung der einzelnen Kovarianzen können nur die Subjekte verwendet werden, für die unter *beiden* Bedingungen Werte vorliegen. Das kann unter Umständen zu einer starken

Reduktion der Fallzahl führen. Wir definieren hierfür:

$$\lambda_{j, is}^{(\ell, r)} = \begin{cases} 1 & \text{falls } X_{j, is}^{(\ell)} \text{ und } X_{j, is}^{(r)} \text{ vorliegt} \\ 0 & \text{falls } X_{j, is}^{(\ell)} \text{ und/oder } X_{j, is}^{(r)} \text{ fehlt} \end{cases} .$$

Dann ist $n_{j, i}^{(\ell, r)} = \sum_{s=1}^{n_i} \lambda_{j, is}^{(\ell, r)}$ die Anzahl der entsprechenden Beobachtungen unter den Bedingungen ℓ und r . Die Kovarianzmatrix ist definiert als:

$$\widehat{\mathbf{V}}_i^* = \left(\widehat{v}_i^{(\ell, r)*} \right)_{\ell, r=1, \dots, d} ,$$

mit den einzelnen Komponenten für $i = 1, 2$ wie in Abschnitt 4.4 auf S.29. Dabei ändern sich nur die Komponenten $\widehat{Y}_{j, 1s}^{(\ell)}$ folgendermaßen:

$$\widehat{Y}_{j, 1s}^{(\ell)} = \frac{1}{n_{j, 2}^{(\ell, r)}} \left(R_{j, is}^{(\ell)} - R_{j, is}^{(\ell)(i)} \right) ,$$

und für Statusgruppe $i = 2$ entsprechend. Für diese modifizierten relativen Effekte mit dazugehörigen Varianzen können dann wieder die Transformationsmethoden aus Kapitel 5 angewendet werden.

Simulation fehlender Werte am Beispiel der Meningitis-Studie

An der Meningitis-Studie haben wir den Einfluss fehlender Werte untersucht. Dabei wurden rein zufällig 1, 2, 5 und 10 Prozent der Daten durch fehlende Werte ersetzt. In Tabelle 9.1 sind die Ergebnisse angegeben. Wie man sieht, führen bei großen Effekten (hier der Laborpa-

Tabelle 9.1: *Einfluss rein zufällig fehlender Werte auf den geschätzten relativen Effekt anhand des Meningitis-Beispiels.*

Laborparameter	Geschätzter adjustierter relativer Effekt bei einem Anteil fehlender Werte von				
	0 %	1 %	2 %	5 %	10 %
A	0.8759	0.7328	0.6256	0.735	0.6459
B	0.7067	0.6749	0.6749	0.6844	0.6336
C	0.4354	0.5921	0.547	0.5913	0.5729

rameter A) bereits 1 % fehlende Werte zu stark verfälschten Ergebnissen. Bei einem weiteren Anstieg des Anteils fehlender Werte ist die Zunahme der Verfälschung dann nicht mehr so stark. Je näher der Effekt an 0.5 ist, desto geringer ist der Einfluss fehlender Werte.

Zu beachten ist, dass die Kovarianzmatrix bei fehlenden Werten gegebenenfalls nicht mehr positiv semidefinit ist, und dass die Schätzer für die Kovarianzmatrix durch fehlende Werte

stark verfälscht sein kann. Dennoch wurde das Computerprogramm so geschrieben, dass fehlende Werte berücksichtigt werden können. Denn in der Praxis kommt es häufig vor, dass einzelne Werte rein zufällig fehlen, und in diesen Fällen soll das Makro trotzdem anwendbar bleiben. Zur Kontrolle sollten aber auch Imputationsverfahren¹ verwendet und die Ergebnisse verglichen werden.

9.2 Multiple Kontrasttests und simultane Konfidenzintervalle

Im Falle mehrerer Bedingungen ist die Aussage, dass unter mindestens einer Bedingung der Effekt ungleich 0.5 ist, oder dass sich die Effekte von mindestens zwei Bedingungen unterscheiden (siehe Kapitel 6), oft nicht zufriedenstellend. Für den Anwender ist meistens eine genauere Aussage von Interesse, z.B. unter welchen Bedingungen ein Effekt vorliegt, bzw. unter welcher Bedingung der Effekt am größten ist. Um aber genauere Aussagen treffen zu können sind mehrere Tests nötig, wie z.B. der Vergleich aller Gruppen mit einer Kontrollgruppe oder der paarweise Vergleich aller Gruppen. Hierbei darf aber das globale Niveau das vorgegebene α -Niveau nicht überschreiten, weshalb entsprechend bezüglich der multiplen Vergleiche adjustiert werden muss. Hierfür gibt es verschiedene Möglichkeiten, die einfachste ist wohl die Bonferroni-Adjustierung. Hierbei ergibt sich das lokale aus dem globalen α geteilt durch die Anzahl der Tests. Allerdings ist die Bonferroni-Adjustierung konservativ, es wird also überkorrigiert. Auch viele andere Korrekturverfahren haben Nachteile oder benötigen spezielle Voraussetzungen. Ganz anders hingegen ist die Herangehensweise bei den multiplen Kontrasttests. Dabei wird eine Kontrastmatrix $\mathbf{C} = (\mathbf{c}_1, \dots, \mathbf{c}_Q)'$ definiert, deren Zeilen $\mathbf{c}_q = (c_{q1}, \dots, c_{qd})$ die einzelnen Alternativen formulieren (mit $q = 1, \dots, Q$). Dabei muss gelten, dass $\sum_{\ell=1}^d c_{q\ell} = 0$. Im parametrischen Ansatz lauten die einzelnen Hypothesen dann $H_0 : t_q = 0$ mit dem Kontrast $t_q = \sum_{\ell=1}^d c_{q\ell} \mu_{\ell}$, siehe hierzu Bretz et al. (2001, [7]). Munzel und Hothorn (2001, [31]) und Munzel und Tamhane (2002, [32]) haben die Ansätze für die Anwendung in nichtparametrischen Modellen übertragen, hierbei wurden mehrere unabhängige Gruppen verglichen. Übertragen auf unser Design, also auf abhängige Messungen zweier unabhängiger Stichproben, lautet die Globalhypothese im nichtparametrischen Fall:

$$H_0 : \mathbf{p}^c = \mathbf{0} ,$$

mit $\mathbf{Cp}^* = \mathbf{p}^c = (p_1^c, \dots, p_Q^c)'$, und die jeweiligen Einzelhypothesen

$$H_{0,q} : p_q^c = 0 .$$

Es existieren einige vordefinierte Kontrastmatrizen, von denen hier eine Auswahl vorgestellt wird.

¹Z.B. best case, worst case, last observation carried forward.

- Tukey-Kontraste: Hierbei werden alle Paarvergleiche durchgeführt, ohne eine Gewichtung für die Stichprobenumfänge (Tukey (1953) [38]).
- Dunnett-Kontraste: Hierbei werden alle Gruppen mit der Kontrollgruppe verglichen, wir nehmen hier die erste Gruppe als Kontrollgruppe an. Auch bei der Dunnett-Methode bleiben die Stichprobenumfänge unberücksichtigt (Dunnett (1955) [20]).
- Sequenzielle Kontraste: Hierbei werden alle benachbarten Gruppen miteinander verglichen, auch wieder ungewichtet für die Stichprobenumfänge (Bauer und Budde (1992) [4]).
- Average-Kontraste: Hierbei wird jede Gruppe mit den anderen gemeinsam verglichen, dabei wird für die Stichprobenumfänge gewichtet.
- Changepoint-Kontraste: Hierbei werden die Gruppen schrittweise in zwei Obergruppen aufgeteilt und diese zwei Obergruppen miteinander verglichen, auch hier wird für die Stichprobenumfänge gewichtet (Hirotzu (1979) [23]).
- Williams-Kontraste: Hierbei wird die erste Stichprobe erst mit der letzten verglichen, dann mit den letzten beiden usw., auch wieder mit Gewichtung für die Stichprobenumfänge (Williams (1971) [42]).
- Marcus-Kontraste: Diese Methode entspricht der Kombination aus Changepoint- und Williams-Kontrasten, auch wieder mit Gewichtung (Marcus (1976) [30]).

Die im vierten Kapitel vorgestellten lokalen Konfidenzintervalle halten jedes für sich das lokale Niveau ein, insgesamt wird das globale Niveau jedoch überschritten. Damit das globale α eingehalten wird, müssen auch die Konfidenzintervalle für die multiplen Vergleiche adjustiert werden (siehe oben). Eine wesentlich bessere Methode ist die Anwendung simultaner Konfidenzintervalle, die das globale Niveau einhalten und zugleich zu schmaleren Intervallen führen. Dabei werden die Konfidenzintervalle für die verschiedenen Alternativen simultan gebildet. Die Alternativen werden, wie bei den multiplen Kontrasttests über die Matrix $\mathbf{C} = (\mathbf{c}_1, \dots, \mathbf{c}_Q)'$ formuliert.

Die von Munzel und Hothorn (2001, [31]) erarbeiteten Verfahren sind Gegenstand aktueller Forschung. Die dabei entwickelte Methodik soll dann in einem nächsten Schritt auch für die Berücksichtigung von Kovariablen angepasst werden.

A Makro

Alle erarbeiteten Verfahren sind für das statistische Programmpaket R in einem Makro namens NMBFP (für Nichtparametrisches Multivariates Behrens-Fisher-Problem) implementiert.

Das Makro wird aufgerufen durch:

NMBFP(DATA	= Datensatz,
	STATE	= Status,
	RESPONSE	= Zielvariable,
	COVARIATES	= Kovariablen,
	FACTOR1	= 1. Faktor,
	FACTOR2	= 2. Faktor,
	ALPHA	= α -Niveau,
	PLOT	= Grafiken ja/nein)

Der Goldstandard (STATE) muss zweistufig vorliegen, ansonsten wird das Makro abgebrochen und eine Fehlermeldung zurückgegeben. Die Zielvariable kann in beliebigem Skalenniveau vorliegen und auch fehlende Werte können behandelt werden (sie sollten mit NA gekennzeichnet sein). Wenn eine Kovariable bzw. Baseline-Variable vorliegt, wird sie direkt über den Variablennamen angegeben. Wenn mehrere Kovariablen berücksichtigt werden sollen, werden sie in der Form $c(\text{Kovariable1}, \text{Kovariable2}, \dots)$ angegeben; es gibt keine Beschränkung der Anzahl der Kovariablen. Der erste Faktor bezeichnet die Variable eines Subplot-Faktors (z.B. Methode), der zweite Faktor die des zweiten Subplot-Faktors (z.B. Untersucher). Beide Angaben sind optional, wird der Faktor 2 nicht angegeben, werden die Daten unter der Annahme nur eines Faktors ausgewertet, ist auch für den Faktor 1 keine Angabe gemacht worden, werden die Daten unter der Annahme einer einzigen Bedingung ausgewertet. Entsprechend werden nur die Tests durchgeführt, die bei dem jeweiligen Modell möglich sind. Das α -Niveau kann beliebig definiert werden, wenn die Angabe fehlt wird die Voreinstellung 5% verwendet. Die Variable PLOT erfordert einen logischen Wert (TRUE/FALSE) und gibt an, ob Grafiken erstellt werden sollen, voreingestellt ist hier TRUE.

Der Output des Programms beinhaltet als Kenngrößen die Anzahl der Faktorstufen der beiden Faktoren und die Gesamtzahl der Bedingungen d , desweiteren die Anzahl der Kovariablen, die Stichprobenumfänge (in den beiden Status-Gruppen und gesamt) und die Anzahl der fehlenden Werte unter den einzelnen Bedingungen.

Als Maßzahlen werden für die verschiedenen Bedingungen die unadjustierten und die transformierten adjustierten relativen Effekte inklusive Konfidenzintervalle ausgegeben. Von den durchgeführten Tests werden die Testgrößen und dazugehörigen p-Werte ausgegeben. Wenn `Plot=TRUE` gilt, werden die Schätzer der adjustierten und unadjustierten Effekte und die dazugehörigen Konfidenzintervalle gezeichnet.

B Allgemeine Bezeichnungen

Vektoren werden in dieser Arbeit immer mit fett gedruckten kleinen Buchstaben bezeichnet, Matrizen mit fett gedruckten großen Buchstaben oder Symbolen.

B.1 Vordefinierte Matrizen

- der Einservektor:

$$\mathbf{1}_d = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^{d \times 1}$$

- die Einsermatrix:

$$\mathbf{J}_d = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{d \times d}$$

- die Einheitsmatrix:

$$\mathbf{I}_d = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{d \times d}.$$

Für zwei beliebige Matrizen

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pq} \end{pmatrix} \in \mathbb{R}^{p \times q} \quad \text{und}$$

$$\mathbf{B} = \begin{pmatrix} b_{11} & \cdots & b_{1s} \\ \vdots & & \vdots \\ b_{r1} & \cdots & b_{rs} \end{pmatrix} \in \mathbb{R}^{r \times s}$$

ist das Kronecker-Produkt folgendermaßen definiert:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1q}\mathbf{B} \\ \vdots & & \vdots \\ a_{p1}\mathbf{B} & \cdots & a_{pq}\mathbf{B} \end{pmatrix} \in \mathbb{R}^{pr \times qs}.$$

B.2 Zählfunktion

- links-stetige Version der Zählfunktion

$$c^-(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

- rechts-stetige Version der Zählfunktion

$$c^+(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

- normalisierte Version der Zählfunktion

$$c(x) = \frac{1}{2} [c^+(x) + c^-(x)]$$

C Beweise

In diesem Kapitel sind längere Beweise aufgeführt, die den Lesefluss stören würden. An der Überschrift ist jeweils abzulesen auf welches Theorem sich der Beweis bezieht.

C.1 Beweis zu Theorem 4.2 auf Seite 28

Zu (1) Bekannt ist folgende Eigenschaft:

$$E(\widehat{p}_j^{(\ell)} - p_j^{(\ell)})^2 = O\left(\frac{1}{N}\right), j = 0, \dots, k$$

(Beweis siehe Brunner et al. (2002) [14]). Damit gilt:

$$\begin{aligned} & E\left(\widehat{p}^{(\ell)*} - p^{(\ell)*}\right)^2 \\ &= E\left(\widehat{p}_0^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \left(\widehat{p}_j^{(\ell)} - 1/2\right) - \left(p_0^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \left(p_j^{(\ell)} - 1/2\right)\right)\right)^2 \\ &= E\left(\widehat{p}_0^{(\ell)} - p_0^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \left(\widehat{p}_j^{(\ell)} - p_j^{(\ell)}\right)\right)^2 \end{aligned}$$

(wegen der c_r -Ungleichung)

$$\leq 2E\left(\widehat{p}_0^{(\ell)} - p_0^{(\ell)}\right)^2 + 2E\left(\sum_{j=1}^k \gamma_j^{(\ell)} \left(\widehat{p}_j^{(\ell)} - p_j^{(\ell)}\right)\right)^2.$$

Für den ersten Teil gilt: $E\left(\widehat{p}_0^{(\ell)} - p_0^{(\ell)}\right)^2 = O\left(\frac{1}{N}\right)$ (siehe oben). Der zweite Teil lässt sich mit der Jensen-Ungleichung folgendermaßen schreiben:

$$E\left(\sum_{j=1}^k \gamma_j^{(\ell)} \left(\widehat{p}_j^{(\ell)} - p_j^{(\ell)}\right)\right)^2 \leq k \sum_{j=1}^k \gamma_j^{(\ell)2} E\left(\widehat{p}_j^{(\ell)} - p_j^{(\ell)}\right)^2.$$

Da gilt: $E\left(\widehat{p}_j^{(\ell)} - p_j^{(\ell)}\right)^2 = O\left(\frac{1}{N}\right)$ (siehe oben) und k beschränkt ist, gilt:

$$E\left(\widehat{p}^{(\ell)*} - p^{(\ell)*}\right)^2 \rightarrow 0 \text{ für } N \rightarrow \infty.$$

Zu (2) Aus der L_2 -Konsistenz folgt automatisch auch die asymptotische Erwartungstreue.

Zu (3) Die Beobachtungen $\mathbf{X}^{(\ell)} = (X_{j, is}^{(\ell)})_{i=1,2; s=1, \dots, n_i; j=0, \dots, k}$ gehen in die Berechnung von $\hat{p}^{(\ell)*}$ in Form ihrer Ränge ein. Wenn nun die Beobachtungen transformiert werden, so dass gilt:

$\mathbf{X}^{(\ell)} = (\text{const} \cdot X_{j, is}^{(\ell)})_{i=1,2; s=1, \dots, n_i; j=0, \dots, k}$ (const sei eine beliebige Konstante), so hat dies keine Auswirkung auf die Ränge. Das heißt:

$$\begin{aligned} R_{j, is}^{(\ell)} &= n_1 \widehat{F}_{j,1}^{(\ell)}(X_{j, is}^{(\ell)}) + n_2 \widehat{F}_{j,2}^{(\ell)}(X_{j, is}^{(\ell)}) + \frac{1}{2} \\ &= n_1 \widehat{F}_{j,1}^{(\ell)}(\text{const} \cdot X_{j, is}^{(\ell)}) + n_2 \widehat{F}_{j,2}^{(\ell)}(\text{const} \cdot X_{j, is}^{(\ell)}) + \frac{1}{2}. \end{aligned}$$

□

C.2 Beweis zu Satz 4.5 auf Seite 29

Unter den Annahmen (A1) und (A2) auf S. 13 und für bekannt angenommenes $\gamma_j^{(\ell)}$ gilt folgende asymptotische Äquivalenz:

$$\begin{aligned} &\sqrt{N} \left(\hat{p}^{(\ell)*} - p^{(\ell)*} \right) \\ &= \sqrt{N} \left(\hat{p}_0^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \left(\hat{p}_j^{(\ell)} - 0.5 \right) - \left(p_0^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \left(p_j^{(\ell)} - 0.5 \right) \right) \right) \\ &= \sqrt{N} \left(\hat{p}_0^{(\ell)} - p_0^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \left(\hat{p}_j^{(\ell)} - p_j^{(\ell)} \right) \right) \\ &= \sqrt{N} \left(\int \widehat{F}_{0,1}^{(\ell)} d\widehat{F}_{0,2}^{(\ell)} - \int F_{0,1}^{(\ell)} dF_{0,2}^{(\ell)} - \sum_{j=1}^k \gamma_j \left(\int \widehat{F}_{j,1}^{(\ell)} d\widehat{F}_{j,2}^{(\ell)} - \int F_{j,1}^{(\ell)} dF_{j,2}^{(\ell)} \right) \right) \\ &= \sqrt{N} \left(\int F_{0,1}^{(\ell)} d\widehat{F}_{0,2}^{(\ell)} - \int F_{0,2}^{(\ell)} d\widehat{F}_{0,2}^{(\ell)} + \int \left(\widehat{F}_{0,1}^{(\ell)} - F_{0,1}^{(\ell)} \right) d \left(\widehat{F}_{0,2}^{(\ell)} - F_{0,2}^{(\ell)} \right) + 1 - 2p_0^{(\ell)} \right) \end{aligned}$$

$$\begin{aligned}
& - \sum_{j=1}^k \gamma_j^{(\ell)} \left(\int F_{j,1}^{(\ell)} d\widehat{F}_{j,2}^{(\ell)} - \int F_{j,1}^{(\ell)} d\widehat{F}_{j,2}^{(\ell)} + \int \left(\widehat{F}_{j,1}^{(\ell)} - F_{j,1}^{(\ell)} \right) d \left(\widehat{F}_{j,2}^{(\ell)} - F_{j,2}^{(\ell)} \right) + 1 - 2p_j^{(\ell)} \right) \\
&= \sqrt{N} \left(\int F_{0,1}^{(\ell)} d\widehat{F}_{0,2}^{(\ell)} - \int F_{0,2}^{(\ell)} d\widehat{F}_{0,2}^{(\ell)} + 1 - 2p_0^{(\ell)} \right) \\
& \quad - \sum_{j=1}^k \gamma_j^{(\ell)} \left(\int F_{j,1}^{(\ell)} d\widehat{F}_{j,2}^{(\ell)} - \int F_{j,1}^{(\ell)} d\widehat{F}_{j,2}^{(\ell)} + 1 - 2p_j^{(\ell)} \right) \\
& \quad + \sqrt{N} \left(\int \left(\widehat{F}_{0,1}^{(\ell)} - F_{0,1}^{(\ell)} \right) d \left(\widehat{F}_{0,2}^{(\ell)} - F_{0,2}^{(\ell)} \right) - \sum_{j=1}^k \gamma_j^{(\ell)} \int \left(\widehat{F}_{j,1}^{(\ell)} - F_{j,1}^{(\ell)} \right) d \left(\widehat{F}_{j,2}^{(\ell)} - F_{j,2}^{(\ell)} \right) \right) \\
&= \sqrt{N} B_N^{(\ell)} + \sqrt{N} A_N^{(\ell)} \\
&\doteq \sqrt{N} B_N^{(\ell)}
\end{aligned}$$

Dabei ist $B_N^{(\ell)}$ wie in Definition 4.4, und für

$$\begin{aligned}
A_N^{(\ell)} &= \int \left(\widehat{F}_{0,1}^{(\ell)} - F_{0,1}^{(\ell)} \right) d \left(\widehat{F}_{0,2}^{(\ell)} - F_{0,2}^{(\ell)} \right) - \sum_{j=1}^k \gamma_j^{(\ell)} \int \left(\widehat{F}_{j,1}^{(\ell)} - F_{j,1}^{(\ell)} \right) d \left(\widehat{F}_{j,2}^{(\ell)} - F_{j,2}^{(\ell)} \right) \text{ bzw.} \\
A_j^{(\ell)} &= \int \left(\widehat{F}_{j,1}^{(\ell)} - F_{j,1}^{(\ell)} \right) d \left(\widehat{F}_{j,2}^{(\ell)} - F_{j,2}^{(\ell)} \right) \text{ mit } j = 0, \dots, k \text{ gilt:}
\end{aligned}$$

$$\begin{aligned}
& E \left(\sqrt{N} A_N^{(\ell)} \right)^2 \\
&= E \left(\sqrt{N} A_0^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \sqrt{N} A_j^{(\ell)} \right)^2
\end{aligned}$$

(wegen der c_r -Ungleichung)

$$\leq 2E \left(\sqrt{N} A_0^{(\ell)} \right)^2 + 2E \left(\sum_{j=1}^k \gamma_j^{(\ell)} \sqrt{N} A_j^{(\ell)} \right)^2 .$$

Für den ersten Teil gilt $E \left(\sqrt{N} A_0^{(\ell)} \right)^2 = O \left(\frac{1}{N} \right)$ (siehe Brunner et al. 2002 [11], Theorem 3.3). Für den zweiten Teil gilt mit der Jensen-Ungleichung:

$$E \left(\sum_{j=1}^k \gamma_j^{(\ell)} \sqrt{N} A_j^{(\ell)} \right)^2 \leq k \sum_{j=1}^k \gamma_j^{(\ell)2} E \left(\sqrt{N} A_j^{(\ell)} \right)^2 .$$

Da bekannt ist, dass $E \left(\sqrt{N} A_j^{(\ell)} \right)^2 = O \left(\frac{1}{N} \right)$ (siehe oben) und k beschränkt ist, gilt: $E \left(\sqrt{N} A_N^{(\ell)} \right)^2 \rightarrow 0$ für $N \rightarrow \infty$.

□

C.3 Beweis zu Theorem 4.9 auf Seite 32

Zu (1): Da $\widehat{\mathbf{V}}_N^*$ die Stichprobenkovarianzmatrix ist, ist sie damit auch (asymptotisch) erwartungstreu.

Zu (2): Es genügt die L_2 -Konsistenz für jedes Element von $\widehat{\mathbf{V}}_N^* - \mathbf{V}_N^*$ zu zeigen. Das heißt, es muss bewiesen werden: $\|\widehat{v}_i^{(\ell,r)*} - v_i^{(\ell,r)*}\|_2 \rightarrow 0$, wobei gilt: $\widehat{v}_i^{(\ell,r)*} = \frac{N}{n_1} \widehat{v}_1^{(\ell,r)*} + \frac{N}{n_2} \widehat{v}_2^{(\ell,r)*}$.

Wir definieren $Y_{j,isz}^{(\ell)} = Y_{j,is}^{(\ell)} - \bar{Y}_{j,i}^{(\ell)}$ und $\widehat{Y}_{j,isz}^{(\ell)} = \widehat{Y}_{j,is}^{(\ell)} - \widehat{\bar{Y}}_{j,i}^{(\ell)}$ mit $j = 0, \dots, k$, $i = 1, 2$ und $s = 1, \dots, n_i$. Dann schreiben wir die Kovarianz $v_i^{(\ell,r)*}$ als

$$v_i^{(\ell,r)*} = \frac{N}{n_i(n_i - 1)} \sum_{s=1}^{n_i} \left(Y_{0,isz}^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} Y_{j,isz}^{(\ell)} \right) \left(Y_{0,isz}^{(r)} - \sum_{j=1}^k \gamma_j^{(r)} Y_{j,isz}^{(r)} \right),$$

und die geschätzte Kovarianz $\widehat{v}_i^{(\ell,r)*}$ entsprechend als

$$\widehat{v}_i^{(\ell,r)*} = \frac{N}{n_i(n_i - 1)} \sum_{s=1}^{n_i} \left(\widehat{Y}_{0,isz}^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \widehat{Y}_{j,isz}^{(\ell)} \right) \left(\widehat{Y}_{0,isz}^{(r)} - \sum_{j=1}^k \gamma_j^{(r)} \widehat{Y}_{j,isz}^{(r)} \right).$$

Analog zu Brunner et al. (2002 [14], Theorem 4.1) genügt es mit der Annahme (A2) zu zeigen, dass $E(A_i^2) \rightarrow 0$, wobei $A_i = 1/n_i \sum_{s=1}^{n_i} \Delta_{is}$ und

$$\begin{aligned} \Delta_{is} &= \left(Y_{0,isz}^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} Y_{j,isz}^{(\ell)} \right) \left(Y_{0,isz}^{(r)} - \sum_{j=1}^k \gamma_j^{(r)} Y_{j,isz}^{(r)} \right) \\ &\quad - \left(\widehat{Y}_{0,isz}^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \widehat{Y}_{j,isz}^{(\ell)} \right) \left(\widehat{Y}_{0,isz}^{(r)} - \sum_{j=1}^k \gamma_j^{(r)} \widehat{Y}_{j,isz}^{(r)} \right). \end{aligned}$$

Für den weiteren Beweis wird verwendet, dass für $i = 1, 2$ gilt: $E\left(\widehat{Y}_{j,i1}^{(\ell)} - Y_{j,i1}^{(\ell)}\right)^2 \rightarrow 0$ (siehe Brunner et al., 2002 [14], Theorem 4.1)). Außerdem ist zu beachten, dass die Zufallsvariablen $Y_{j,isz}^{(\ell)}$ und $Y_{j,isz'}^{(\ell)}$ mit $j = 1, \dots, k$ unabhängig sind für $s \neq s'$. Dann zeigen wir die L_2 -Konsistenz beispielhaft für die Varianz $\widehat{v}_i^{(\ell,\ell)*}$, für die Kovarianzen gilt es analog. Dann gilt für $i = 1$:

$$\begin{aligned} &E \left[\widehat{Y}_{0,11z}^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \widehat{Y}_{j,11z}^{(\ell)} - \left(Y_{0,11z}^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} Y_{j,11z}^{(\ell)} \right) \right]^2 \\ &= E \left[\widehat{Y}_{0,11z}^{(\ell)} - Y_{0,11z}^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \left(\widehat{Y}_{j,11z}^{(\ell)} - Y_{j,11z}^{(\ell)} \right) \right]^2 \end{aligned}$$

(wegen der c_r -Ungleichung)

$$\leq 2E \left(\widehat{Y}_{0,11z}^{(\ell)} - Y_{0,11z}^{(\ell)} \right)^2 + 2E \left(\sum_{j=1}^k \gamma_j^{(\ell)} \left(\widehat{Y}_{j,11z}^{(\ell)} - Y_{j,11z}^{(\ell)} \right) \right)^2 .$$

Für den ersten Teil gilt:

$$\begin{aligned} & E \left(\widehat{Y}_{0,11z}^{(\ell)} - Y_{0,11z}^{(\ell)} \right)^2 \\ &= E \left(\widehat{Y}_{0,11}^{(\ell)} - \widehat{Y}_{0,1}^{(\ell)} - \left(Y_{0,11}^{(\ell)} - \bar{Y}_{0,1}^{(\ell)} \right) \right)^2 \end{aligned}$$

(wegen der c_r -Ungleichung)

$$\leq 2E \left(\widehat{Y}_{0,11}^{(\ell)} - Y_{0,11}^{(\ell)} \right)^2 + 2E \left(\widehat{Y}_{0,1}^{(\ell)} - \bar{Y}_{0,1}^{(\ell)} \right)^2 .$$

Dabei wiederum gilt für den ersten Teil: $E \left(\widehat{Y}_{0,11}^{(\ell)} - Y_{0,11}^{(\ell)} \right)^2 \rightarrow 0$ (siehe oben) und für den zweiten Teil:

$$E \left(\widehat{Y}_{0,1}^{(\ell)} - \bar{Y}_{0,1}^{(\ell)} \right)^2$$

(mit der Jensen-Ungleichung)

$$\leq 2E \left(\widehat{Y}_{0,11}^{(\ell)} - Y_{0,11}^{(\ell)} \right)^2 \rightarrow 0 .$$

Für den zweiten Teil von oben gilt:

$$E \left(\sum_{j=1}^k \gamma_j^{(\ell)} \left(\widehat{Y}_{j,11z}^{(\ell)} - Y_{j,11z}^{(\ell)} \right) \right)^2 \leq k \sum_{j=1}^k \gamma_j^{(\ell)2} E \left(\widehat{Y}_{j,11z}^{(\ell)} - Y_{j,11z}^{(\ell)} \right)^2 .$$

Dabei gilt wieder $E \left(\widehat{Y}_{j,11z}^{(\ell)} - Y_{j,11z}^{(\ell)} \right)^2 \rightarrow 0$ (siehe oben), und da k beschränkt ist und der Beweis genauso für $i = 2$ geht, gilt insgesamt für $i = 1, 2$:

$$E \left[\widehat{Y}_{0,i1z}^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \widehat{Y}_{j,i1z}^{(\ell)} - \left(Y_{0,i1z}^{(\ell)} - \sum_{j=1}^k \gamma_j^{(\ell)} \left(Y_{j,i1z}^{(\ell)} \right) \right) \right]^2 \rightarrow 0 .$$

Wird auf die Schätzer $\widehat{\gamma}_j^{(r)}$ bedingt, gilt der Beweis genauso.

Zu (3): Die geschätzte Kovarianzmatrix $\widehat{\mathbf{V}}_N^{(*)}$ ist eine Stichprobenkovarianzmatrix und damit auch positiv semidefinit.

□

Literaturverzeichnis

- [1] AKRITAS M., KUHA J., OSGOOD D.: *A Nonparametric Approach to Matched Pairs With Missing Data*. Sociological Methods and Research 30, 425-454, 2002.
- [2] BATHKE, A.: *Testing monotone effects of covariates in nonparametric mixed models*. Nonparametric Statistics 17, 423-439, 2005.
- [3] BATHKE A., BRUNNER E.: *A Nonparametric Alternative to Analysis of Covariance*. Akritas M., Politis D. (Hrsg.), Recent Advances and Trends in Nonparametric Statistics, Elsevier, 2003.
- [4] BAUER P., BUDDE M.: *Multiple Testing for detecting efficient dose steps*. Biometrical Journal 36, 3-15, 1994.
- [5] BEHRENS, W. U.: *Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen*. Landwirtschaftliche Jahrbücher 68, 807-837, 1929.
- [6] BOX, G.E.P.: *Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification*. The Annals of Mathematical Statistics, 25, 290-302, 1954.
- [7] BRETZ F., GENZ A., HOTHORN L. A.: *On the Numerical Availability of Multiple Comparison Procedures*. Biometrical Journal 43, 645-656, 2001.
- [8] BRUMBACK L., PEPE M., ALONZO T.: *Using the ROC curve for gauging treatment effect in clinical trials*. Statistics in Medicine, 583-588, 2006.
- [9] BRUNNER E., DETTE H., MUNK A.: *Box-Type Approximations in Nonparametric Factorial Designs*. Journal of the American Statistical Association 92, 1494-1502, 1997.
- [10] BRUNNER E., MUNZEL U.: *The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation*. Biometrical Journal, 42, 17-25, 2000.
- [11] BRUNNER E., MUNZEL U.: *Nichtparametrische Datenanalyse*. Springer Verlag, Berlin, 2002.
- [12] BRUNNER E., NEUMANN N.: *Two-Sample Rank Tests in General Models*. Biometrical Journal 4, 395-402, 1986.
- [13] BRUNNER E., MUNZEL U., PURI M.: *Rank-Score Tests in Factorial Designs with Repeated Measures*. Journal of Multivariate Analysis 70, 286-317, 1999.

- [14] BRUNNER E., MUNZEL U., PURI M.: *The multivariate nonparametric Behrens-Fisher problem*. Journal of Statistical Planning and Inference 108, 37-53, 2002.
- [15] BRUNNER E., PURI M. L.: *Nonparametric methods in design and analysis of experiments*. Handbook of Statistics, Vol. 13, 631-703, 1996.
- [16] CHRISTOPHLIEMK, O.: *Die Transformationsmethode im gemischten nichtparametrischen Modell mit Kovariablen*. Diplomarbeit, Göttingen, 2001.
- [17] DELONG E.R., DELONG D.M., CLARKE-PEARSON D.L.: *Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach*. Biometrics 44, 837-45, 1988.
- [18] DOMHOF: *Nichtparametrische relative Effekte*. Dissertation, Göttingen, 2001.
- [19] DOMHOF S., BRUNNER E., OSGOOD D.: *Rank Procedures for Repeated Measures With Missing Values*. Sociological Methods and Research 30, 367-393, 2002.
- [20] DUNNETT, C. W.: *A multiple comparison procedure for comparing several treatments with a control*. Journal of the American Statistical Association 50, 1096-1121, 1955.
- [21] FISHER, R. A.: *The fiducial argument in statistical inference*. Annals of Eugenics 6, 391-398, 1936.
- [22] FLIGNER M. A., POLICELLO G. E.: *Robust Rank Procedures for the Behrens-Fisher Problem*. Journal of the American Statistical Association 76, 162-168, 1981.
- [23] HIROTSU, C.: *The cumulative chi-squares method and a studentized maximal contrast method for testing an ordered alternative in a one-way analysis of variance model*. Reports of Statistical Application Research 26, 12-21, 1979.
- [24] JANES H., PEPE M.: *Adjusting for Covariates in Studies of Diagnostic, Screening, or Prognostic Markers: An Old Concept in a New Setting*. American Journal of Epidemiology, 168, 1, 89-97, 2008.
- [25] JANES H., LONGTON G., PEPE M.: *Accommodating Covariates in ROC Analysis*. UW Biostatistics Working Paper Series, 2008.
- [26] KAUFMANN J., WERNER C., BRUNNER E.: *Nonparametric methods for analyzing the accuracy of diagnostic tests with multiple readers*. Statistical Methods in Medical Research 14, 129-146, 2005.
- [27] LANGE, K.: *Nichtparametrische Designs für faktorielle Diagnosestudien*. Diplomarbeit, Göttingen, 2008.
- [28] LANGER, F.: *Berücksichtigung von Kovariablen im nichtparametrischen gemischten Modell*. Doktorarbeit, Göttingen, 1998.

-
- [29] MANN H. B., WHITNEY D. R.: *On a test of whether one of two random variables is stochastically larger than the other*. Annals of Mathematical Statistics 18, 50-60, 1947.
- [30] MARCUS, R.: *The power of some tests of the equality of normal means against an ordered alternative*. Biometrika 63, 177-183, 1976.
- [31] MUNZEL U., HOTHORN L.: *A Unified Approach to Simultaneous Rank Test Procedures in the Unbalanced One-way Layout*. Biometrical Journal, 43, 5, 553-569, 2001.
- [32] MUNZEL U., TAMHANE A.: *Nonparametric multiple comparisons in repeated measures designs for data with ties*. Biometrical Journal 44, 762-779, 2002.
- [33] RUYMGAART, F.H.: *A unified approach to the asymptotic distribution theory of certain midrank statistics*. In: *Statistique non Parametrique Asymptotique, 1-18*, J.P. Raoult (Ed.), *Lecture Notes on Mathematics, No 821*. Springer, Berlin, 1980.
- [34] SATTERTHWAITE, F. E.: *An Approximate Distribution of Estimates of Variance Components*. Biometrics Bulletin 2, 110-114, 1946.
- [35] SENN, S.: *Statistical Issues in Drug Development*. John Wiley, Chichester, 2004.
- [36] SIEMER, A.: *Die Berücksichtigung von heterogen verteilten Kovariablen in einem nicht-parametrischen gemischten Modell*. Diplomarbeit, Göttingen, 1999.
- [37] SMITH, H. F.: *The Problem of Comparing the Results of Two Experiments With Unequal Errors*. Journal of the Council for Scientific and Industrial Research 9, 211-212, 1936.
- [38] TUKEY, J. W.: *The problem of multiple comparisons*. Unpublished manuscript reprinted in: *The Collected Works of John W. Tukey, Vol. 8, 807-837*. Chapman and Hall, New York, 1994.
- [39] VAART, A.W. VAN DER: *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, 1998.
- [40] WELCH, B. L.: *The Significance of the Difference Between Two Means When the Population Variances are Unequal*. Biometrika 29, 350-362, 1947.
- [41] WILCOXON, F.: *Individual comparisons by ranking methods*. Biometrics 1, 80-83, 1945.
- [42] WILLIAMS, D. A.: *A test for differences between treatment means when several dose levels are compared with a zero dose control*. Biometrics 27, 103-117, 1971.

Lebenslauf

Name: Antonia Zapf
Anschrift: Gellertstr. 46, 30175 Hannover
Geburtsdatum: 18. August 1981
Geburtsort: München
Familienstand: verheiratet, 4 Kinder

Ausbildung

Juni 2001
10/2001 bis 11/2006
Allgemeine Hochschulreife
Studium: Statistik an der Ludwig-Maximilians-Universität in München, Anwendungsgebiet Medizinische Biometrie

April 2004
November 2006
Vordiplom Statistik
Diplom Statistik
Titel der Diplomarbeit: „Probabilistische Kreuzvalidierung in linearen Modellen“

Seit April 2007
Studium in dem Promotionsstudiengang „Applied Statistics and Empirical Methods“ der Georg-August-Universität Göttingen

Berufstätigkeit

08/2002 bis 11/2006
Studentische Hilfskraft am Institut für Statistik an der Ludwig-Maximilians-Universität in München

12/2006 bis 02/2009
Wissenschaftliche Mitarbeiterin in der Abteilung für Medizinische Statistik an der Georg-August-Universität Göttingen

Seit März 2009
Wissenschaftliche Mitarbeiterin am Institut für Biometrie an der Medizinischen Hochschule Hannover

Hannover, 15.09.2009