Statistik und Verteidigung der Diplomarbeit

Präzisierung der nötigen Mindestkenntnisse

Statistiksupport PH Wallis, Standort Brig

zusammengestellt von

Paul Ruppen

Version vom 4. Dezember 2017

Dieses Dokument wurde mit LaTeX gesetzt.

Inhaltsverzeichnis

In diesem Papier wird angegeben,

- was Sie wissen müssen, wenn sie in Ihrer Diplomarbeit einen statistischen Test oder statistische Tabellen verwenden. Es ist soviel zu wissen, dass man die Tabellen und gelieferten Resultate sinnvoll interpretieren kann. Vorbehalten bleiben andere Regelungen mit der Betreuerin oder dem Betreuer der Diplomarbeit. Aus dem folgenden Text müssen Sie nur das wissen, was in Ihrer Diplomarbeit auftaucht. Der Statistik-Support gibt bei der Zusendung der Auswertungen die entsprechenden Kapitel an.
- Unter „Angabe der Ergebnisse in der Diplomarbeit" wird für jeden Test erläutert, welche Auswertungsergebnisse in den Text einer Diplomarbeit aufzunehmen sind.
- Zusätzlich zum unabdingbaren Wissen werden weitere Erläuterungen zu den Tests gegeben, die Sie aber nicht kennen oder studieren müssen (jeweils als „fakultativ" gekennzeichet). Diese Erläuterungen setzen eine gewisse wahrscheinlichkeitstheoretische Ausbildung voraus. Diese kann man z.B. im Skript unter http://math.logik.ch/.3bb672ea/cmd.14/audience.D (Teil „Einführung in die Wahrscheinlichkeitstheorie") aufgearbeitet werden, was allerdings zeitintensiv ist.
- Es wird vorausgesetzt, dass Sie wissen, was eine Variable und ihrer Ausprägungen (valeurs de la variable) sind sowie was nominale, ordinale und metrische Skalen sind (échelles nominales, ordinales et métriques) (s. skalen.pdf auf der Home-page http://sozio.logik.ch; http://edit.logik.ch/.3bb67306/cmd.14/audience.D).
- Falls nichts anderes vermerkt, werden Tabellen des Statistikprogramms SPSS zugeschickt.
- Sie dürfen vor der Abgabe der Diplomarbeit diese dem Statistiksupport zuschicken unter Angabe der Stellen, wo Statistik vorkommt. Der Statistiksupport überprüft ihre Ausführungen und korrigiert diese falls nötig.
- Zitiervorschlag: Ruppen, P. (2017). *Statistik und Verteidigung der Diplomarbeit: Präzisierung der nötigen Mindestkenntnisse.* Brig: PH Wallis.

1 Häufigkeitsverteilung (Distribution de fréquences)

Falls in Ihrere Diplomarbeit Tabellen mit Häufigkeitsverteilungen auftauchen, sollten Sie folgendes wissen: Eine Häufigkeitsverteilung (distribution de fréquences) gibt die Anzahl der Objekte pro Ausprägung (Kategorie, Wert) einer Variablen an. Das Wort „Verteilung" deutet an, dass es darum geht, wie die Objekte auf die Ausprägungen der Variable verteilt sind.

Beispiel 1. *Bezüglich der Variable „Wohnort" mit den Ausprägungen „Unterwallis", „Mittelwallis" und „Unterwallis" wurde eine Stichprobe von 94 Personen im Wallis erhoben. Es ergeben sich die folgenden Häufigkeiten (effectifs = fréquences). Bei den gültigen Prozente (pourcentage valide) sind* 100% *die Anzahl der antwortenden Personen, während die Prozente die Anzahl der Personen sind inklusive fehlende Angaben. Gibt es keine fehlende Daten, so sind die beiden Spalten identisch. Kumulierte Prozente (pourcentage cumulé) ist die Summe der gültigen Prozente der darüberstehenden Kategorien inklusive der Kategorie, bei der die Prozentzahl steht. So sind die Kumulierten Prozente des Mittelwallis* $17.0 + 37.2 = 54.3$ *(Die Summenbildung ist wegen Rundungen nicht exakt).*

Wohnort					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Unterwallis	16	17	17	17
	Mittelwallis	35	37.2	37.2	54.3
	Oberwallis	43	45.7	45.7	100
	Gesamt	94	100	100	

Tabelle 1: Häufigkeitsverteilung der Variable „Wohnort“ (deutsche Beschriftung)

Wohnort					
		Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Unterwallis	16	17	17	17
	Mittelwallis	35	37.2	37.2	54.3
	Oberwallis	43	45.7	45.7	100
	Gesamt	94	100	100	

Tabelle 2: Häufigkeitsverteilung der Variable „Wohnort“ (französische Beschriftung)

Abgesehen von zwei fehlenden Daten wird in den folgenden zwei Tabellen ein zu oben identischer Datensatz dargestellt. Die fehlenden Daten werden hinter „fehlend System“ (Manquante Système manquant) angegeben, die gültigen Daten hinter „gültig“ (valide). Die Werte unter „Prozente“ und „gültige Prozente“ sind nicht mehr identisch.

Wohnort					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
	Unterwallis	14	14.9	15.2	15.2
Gültig	Mittelwallis	35	37.2	38	53.3
	Oberwallis	43	45.7	46.7	100
	Gesamt	92	97.9	100	
Fehlend	System	2	2.1		
Total		94	100		

Tabelle 3: Häufigkeitsverteilung der Variable „Wohnort“ bei zwei fehlenden Daten (deutsche Beschriftung)

Wohnort		Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
	Unterwallis	14	14.9	15.2	15.2
Valide	Mittelwallis	35	37.2	38	53.3
	Oberwallis	43	45.7	46.7	100
	Total	92	97.9	100	
Manquante	Système manquant	2	2.1		
Total		94	100		

Tabelle 4: Häufigkeitsverteilung der Variable „Wohnort“ bei zwei fehlenden Daten (französische Beschriftung)

2 Kreuztabelle (Tableau croisée)

Falls in Ihrer Diplomarbeit Kreuztabellen eingefügt sind, sollten Sie folgendes Wissen: In einer Kreuztabelle werden die gemeinsamen Häufigkeiten (distribution de fréquences bivariée, distribution de fréquences commune) zweier Variablen dargestellt.

Beispiel 2. *Variable „Wohnort“ mit den Ausprägungen „Unterwallis“, „Mittelwallis“ und „Unterwallis“ und Variable „Zufriedenheit mit der Schule“ mit den Ausprägungen „sehr zufrieden“, „zufrieden“, „mässig zufrieden“ und „unzufrieden“. So sind z.B. 5 Personen im Unterwallis sehr zufrieden, während 12 Personen im Mittelwallis mässig zufrieden sind. Bei „Gesamt“ (französische Tabelle: Total) stehen die jeweiligen Summen der Spalten respektive der Zeilen. So gaben im Unterwallis 16 Personen zu Thema eine Antwort und 26 der antwortenden Personen waren in der gesamten Stichprobe zufrieden. Der Vektor der Summen der Spalten oder der Summen der Zeilen wird „Randverteilung“ genannt (distribution marginale). So ist z.B.* $(16, 35, 43)$ *die Randverteilung bezüglich „Wohnort“ und* $(28, 26, 22, 18)$ *die Randverteilung bezüglich „Zufriedenheit mit Schule“. Die Randverteilung bezüglich einer Variable ist dasselbe wie die Häufigkeitsverteilung der entsprechenden Variable (s. Kapitel „Häufigkeitsverteilung“). Unten rechts gibt die letzte Zahl die Anzahl der gültigen und gezählten Objekte an (im Beispiel 94). Überprüfen Sie, wo Sie in der Tabelle die Variablennamen und die Namen der Ausprägungen der Variablen finden.*

Wohnort * Zufriedenheit mit Schule Kreuztabelle

Anzahl

		Zufriedenheit mit Schule				Gesamt
		sehr zufrieden	zufrieden	mässig zufrieden	unzufrieden	
Wohnort	Unterwallis	5	6	2	3	16
	Mittelwallis	10	2	12	11	35
	Oberwallis	13	18	8	4	43
Gesamt		28	26	22	18	94

Tabelle 5: Kreuztabelle der Variable „Wohnort“ und der Variable „Zufriedenheit mit Schule“ (deutsche Beschriftung)

Tableau croisé Wohnort * Zufriedenheit mit Schule

Effectif

		Zufriedenheit mit Schule				Total
		sehr zufrieden	zufrieden	mässig zufrieden	unzufrieden	
Wohnort	Unterwallis	5	6	2	3	16
	Mittelwallis	10	2	12	11	35
	Oberwallis	13	18	8	4	43
Total		28	26	22	18	94

Tabelle 6: Kreuztabelle der Variable „Wohnort" und der Variable „Zufriedenheit mit Schule" (französische Beschriftung)

3 Exakter Test von Fisher (Test exact de Fisher)

Liegt eine Kreuztabelle vor bei mindestens einer nominalskalierten Variablen so ist der Fisher-Test zu verwenden. Im zugesandten Ausdruck kommen verschiedene Tests vor (s. Tabelle 11). Es ist nur der Fisher-Test anzuschauen. Folgendes müssen Sie wissen:

- Sie müssen Kreuztabellen (tableau croisé) interpretieren können (s. entsprechendes Kapitel).
- Für den Test werden drei Tabellen zugeschickt, die Sie interpretieren können müssen, wobei die erste oft die Angaben für mehrere Tests enthält. Sie enthält die Angaben über die gültigen (valide) und die fehlenden (manquante) Fälle (observations). Fehlend ist ein Fall, falls bezüglich der Variable keine Information über die Ausprägung des Objektes vorliegt = fehlendes Datum. Im Beispiel gibt es keine fehlende Daten.

Verarbeitete Fälle

	Fälle					
	Gültig		Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
Beruf * Rauchverhalten	94	100%	0	0%	94	100%

Tabelle 7: Tabelle bezüglich gültiger und fehlender Daten (deutsche Beschriftung)

Récapitulatif du traitement des observations

	Observations					
	Valide		Manquante		Total	
	N	Pourcent	N	Pourcent	N	Pourcent
Beruf * Rauchverhalten	94	100%	0	0%	94	100%

Tabelle 8: Tabelle bezüglich gültiger und fehlender Daten (französische Beschriftung)

- Die zweite Tabelle gibt die Kreuztabelle an, zusätzlich zur gemeinsamen Häufigkeitsverteilung (Anzahl, effectif) aber noch die erwarteten Häufigkeiten (= erwartete Anzahlen = fréquences anticipées = effectifs théoriques). Diese sind im Beispiel wie folgt zu verstehen: der Anteil der Raucher an der Gesamtstichprobe ist $\frac{41}{94}$. Gibt es keinen Zusammenhang zwischen dem Beruf und dem Rauchverhalten, müssen bei den Lehrpersonen $\frac{41}{94}$ Raucher sein, d.h. der Anteil der Raucher bei den Lehrpersonen muss mit dem Anteil der Raucher in der gesamten Stichprobe übereinstimmen. Es gibt 29 Lehrpersonen in der Stichprobe. Also müssten $\frac{41}{94} \cdot 29 = 12.649$ (= erwartete Anzahl - effectif théorique) Lehrer Raucher sein. Analog müssten bei den Nicht Lehrpersonen $\frac{41}{94} \cdot 65 = 28.351$ Personen Raucher sein. $\frac{53}{94}$ ist der Anteil der Nichtraucher an der Stichprobe. Entsprechend müssten bei Unabhängigkeit (= kein Zusammenhang zwischen den Variablen) $\frac{53}{94} \cdot 29 = 16.351$ Personen Nichtraucher sein, bei den Nicht Lehrpersonen $\frac{53}{94} \cdot 65 = 36.649$. Die Tabelle erlaubt es festzustellen, in welche Richtung Abweichungen der faktischen Häufigkeiten von den erwarteten Häufigkeiten vorliegen. Hat es mehr Lehrpersonen als erwartet, die Raucher sind - oder weniger Lehrer? Manchmal werden in solchen Tabellen noch die Differenzen zwischen Anzahl und Erwarteter Anzahl geliefert (Residuum - résidu - genannt; z.B. $14 - 12.6 = 1.4$).

Beruf * Rauchverhalten Kreuztabelle

			Rauchverhalten		Gesamt
			Raucher	Nichtraucher	
Beruf	Lehrperson	Anzahl	14	15	29
		Erwartete Anzahl	12.6	16.4	29
	Nicht Lehrperson	Anzahl	27	38	65
		Erwartete Anzahl	28.4	36.6	65
Gesamt		Anzahl	41	53	94
		Erwartete Anzahl	41	53	94

Tabelle 9: Kreuztabelle mit erwarteten Häufigkeiten (deutsche Beschriftung)

Tableau croisé Beruf * Rauchverhalten

			Rauchverhalten		Total
			Raucher	Nichtraucher	
Beruf	Lehrperson	Effectif	14	15	29
		Effectif théorique	12.6	16.4	29
	Nicht Lehrperson	Effectif	27	38	65
		Effectif théorique	28.4	36.6	65
Total		Effectif	41	53	94
		Effectif théorique	41	53	94

Tabelle 10: Kreuztabelle mit erwarteten Häufigkeiten (französische Beschriftung)

- Die dritte Tabelle gibt die Testresultate an. Der Fisher-Test überprüft bei Vorliegen von dichotomen Variablen (d.h. jede der Variablen hat nur zwei Ausprägungen, wir sprechen in der Folge vom 2X2-Fall),

  - ob die Abweichung der Häufigkeit in der Zelle $(1, 1)$ (im Beispiel die Zelle Raucher X Lehrperson) von den erwarteten Häufigkeiten gut oder schlecht *durch Zufall erklärbar* ist (zweiseitiger Test, test bilatéral),
  - ob die Häufigkeit der Zelle $(1, 1)$ grösser ist als erwartet und ob diese Abweichung gut oder schlecht durch Zufall erklärbar ist (rechtsseitiger Test, test unilatéral à droite) ,
  - ob in der Zelle $(1, 1)$ die Häufigkeit kleiner ist als erwartet (linksseitiger Test, test unilatéral à gauche) und ob diese Abweichung *gut oder schlecht durch Zufall erklärbar* ist.

  Hat es mehr als zwei Ausprägungen in einer der Variablen, so gibt der Fisher-Test an, ob es Abweichungen der erwarteten von den faktische Häufigkeiten gibt, die gut oder schlecht durch Zufall erklärbar sind.

- Diese Abweichungen sind *schlecht durch Zufall erklärbar*, wenn der p-Wert (valeur p) kleiner ist als 0.05. Der p-Wert ist im hinter „Exakter Test nach Fisher“ (Test exact de Fisher) zu finden in der Spalte „Exakte Signifikanz (2-seitig)“ (Signification exacte (bilatérale)) oder im 2X2-Fall zusätzlich in der Spalte „Exakte Singifikanz (1-seitig)“ (Signification exacte (unilatérale)). Ob Sie einen zwei- oder einseitigen Test verwenden, hängt von Ihrer Problemstellung ab:
  Im Falle des 2X2-Falles:

  - Haben Sie die Hypothese aufgestellt, dass in der Zelle $(1, 1)$ mehr Objekte liegen als zu erwarten sind, liegt ein rechtseitiger Test vor (im Beispiel: mehr Lehrer als zu erwarten rauchen)
  - haben Sie die Hypothese aufgestellt, dass in der Zelle $(1, 1)$ weniger Objekte liegen als zu erwarten sind, liegt ein linksseitiger Test vor (Im Beispiel: Lehrer rauchern weniger als zu erwarten),
  - haben Sie die Hypothese aufgestellt, dass in der Zelle $(1, 1)$ mehr oder weniger Personen liegen als zu erwarten ist, liegt ein zweiseitiger Test vor (im Beispiel: das Rauchverhalten von Lehrpersonen verschieden ist vom Rauchverhalten von Nicht Lehrpersonen).

  Verwenden Sie auf Grund der Problemlage einen zweiseitigen Test, wählen Sie als p-Wert den Wert unter „Exakte Signifikanz (2-seitig)“.
  Verwenden Sie hingegen einen rechstseitigen Test, brauchen Sie als p-Wert die Zahl unter „Exakte Signifikanz (1-seitig)“, sofern die Häufigkeit in der Zelle $(1, 1)$ grösser als die erwartete Häufigkeit ist. Sonst entfällt ein Test, das sie ja in der Zelle $(1, 1)$ eh weniger Personen als erwartet haben und somit die Hyothese bereits falsifiziert ist.
  Verwenden Sie einen linksseitigen Test, brauchen Sie als p-Wert die Zahl unter „Exakte Signifikanz (1-seitig)“, sofern die Häufigkeit in der Zelle $(1, 1)$ kleiner als die erwartete Häufigkeit ist. Sonst entfällt ein Test, das sie ja in der Zelle $(1, 1)$ eh mehr Personen als erwartet haben und somit die Hyothese bereits falsifiziert ist.
  Ist der p-Wert kleiner als 0.05, sprechen wir von einer signifikanten Abweichung der faktischen von den erwarteten Häufigkeiten.
  Von manchen Programmen wird für den exakten Test von Fisher noch das Quotenverhältnis (rapport des cotes, odds ratio) berechnet. Diese wird berechnet durch $\frac{a_{11}/a_{12}}{a_{21}/a_{22}}$, gewöhnlich allerdings leicht korrigiert (unter anderem, um eine Division durch 0 zu verhindern). Im Beispiel Rauchverhalten drückt es das Verhältnis der rauchenden Lehrer zu den nichtrauchenden Lehrern aus - gemessen am Verhältnis der rauchenden Nicht-Lehrer zu den nichtrauchenden Nicht-Lehrern.

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	.370[a]	1	0.543		
Chi-Quadrat nach Pearson[b]	0.147	1	0.702		
Likelihood-Quotient	0.369	1	0.544		
Exakter Test nach Fisher				0.653	0.35
Zusammenhang linear-mit-linear	0.366	1	0.545		
Anzahl der gültigen Fälle	94				

a. 0 Zellen (0.0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete minimale erwarteteHäufigkeit ist 12.65.
b. Wird nur für eine 2x2-Tabelle berechnet

Tabelle 11: Testergebnisse des exakten Testes nach Fisher (deutsche Beschriftung)

Tests du Khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)	Signification exacte (bilatérale)	Signification exacte (unilatérale)
Khi-deux de Pearson	.370[a]	1	0.543		
Correction pour la continuité[b]	0.147	1	0.702		
Rapport de vraisemblance	0.369	1	0.544		
Test exact de Fisher				0.653	0.35
Association linéaire par linéaire	0.366	1	0.545		
Nombre d'observations valides	94				

a. 0 cellules (0.0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 12.65.
b. Calculé uniquement pour un tableau 2x2

Tabelle 12: Testergebnisse des exakten Testes nach Fisher (französische Beschriftung)

Im Falle von mehr als zwei Ausprägungen: Hier ist die Unterscheidung von rechts- und linksseitig nicht besonders sinnvoll, da der Test die Gesamte Abweichung der erwarteten von der faktischen Kreuztabelle misst.

- Angabe der Ergebnisse in der Diplomarbeit beim 2X2-Fall: Liefern Sie die Kreuztabelle und dann im Text für einen
  - *rechtsseitigen Test* fürs Beispiel „Die Hypothese, dass Lehrpersonen vermehrt Raucher sind als andere Personen hat sich nicht bestätigt (n = 94; gültige Werte: 94 ; Exakter Test nach Fisher; rechtseitiger Test, p-Wert = 0.350)“. „L'hypothèse qu'il y a plus de fumeurs parmis les enseignants n'est pas confirmée (n = 94, observations valides: 94 ; Test exact de Fisher; test unilatéral à droite, valeur p = 0.350)“. Wäre der Wert unter exakte Signifikanz (1-seitig) kleiner als 0.05, z.B. 0.024 und hätte es in der Zelle $(1, 1)$ mehr rauchende Lehrpersonen als erwartet, würden wir schliessen: „Die Hypothese, dass Lehrpersonen vermehrt Raucher sind als andere Personen hat sich bestätigt (n = 94; gültige Werte: 94 ; Exakter Test nach Fisher; rechtseitiger Test, p-Wert = 0.024)“. „L'hypothèse qu'il y a plus de fumeurs parmis les enseignants est confirmée (n = 94, observations valides: 94 ; Test exact de Fisher; test unilatéral à droite, valeur p = 0.024)“.

– *linkseitigen Test* fürs Beispiel „Die Hypothese, dass Lehrpersonen weniger oft Raucher sind als andere Personen hat sich nicht bestätigt (n = 94; gültige Werte: 94 ; es hat mehr rauchende Lehrpersonen als erwartet)“. „L'hypothèse qu'il y a moins de fumeurs parmis les enseignants n'est pas confirmée (n = 94, observations valides: 94 ; il y a plus d'enseignants qui fument que l'effectif théorique)“. Wäre der Wert unter exakte Signifikanz (1 - seitig) kleiner als 0.05 (z.B. 0.017) und hätte es weniger rauchende Lehrpersonen als erwartet, würden wir schliessen: „Die Hypothese, dass Lehrpersonen weniger oft Raucher sind als andere Personen hat sich bestätigt (n = 94; gültige Werte: 94 ; Exakter Test nach Fisher; linksseitiger Test, p-Wert = 0.017)“. „L'hypothèse qu'il y a moins de fumeurs parmis les enseignants est confirmée (n = 94, observations valides: 94 ; Test exact de Fisher; test unilatéral à gauche, valeur p = 0.017)“.

– *zweiseitigen Test* fürs Beispiel: „Die Hypothese, dass Lehrpersonen weniger oft oder häufiger Raucher sind als andere Personen hat sich nicht bestätigt (n = 94; gültige Werte: 94 ; Exakter Test nach Fisher; zweiseitiger Test, p-Wert = 0.653)“. „L'hypothèse qu'il y a moins ou plus de fumeurs parmis les enseignants n'est pas confirmée (n = 94, observations valides: 94 ; Test exact de Fisher; test bilatéral, valeur p = 0.653)“. Wäre der Wert unter exakte Signifikanz (2-seitig) kleiner als 0.05, z.B. 0.022, würden wir schliessen: „Die Hypothese, dass Lehrpersonen weniger oft oder häufiger Raucher sind als andere Personen hat sich bestätigt (n = 94; gültige Werte: 94 ; Exakter Test nach Fisher; zweiseitiger Test, p-Wert = 0.022)“. „L'hypothèse qu'il y a moins ou plus de fumeurs parmis les enseignants est confirmée (n = 94, observations valides: 94 ; Test exact de Fisher; test bilatéral, valeur p = 0.022)“.
Angabe der Ergebnisse in der Diplomarbeit bei mehr als zwei Ausprägungen: Liefern Sie die Kreuztabelle und dann im Text z.B. für eine Kreuztabelle mit den Variablen „Wohnortskategorie“ (drei Auspräungen: Talgemeinde, Tourismusgemeinde, nicht touristisches Bergdorf) und der ordinalskalierten Variable „Zufriedenheit“ (mit vier Ausprägungen)

– „Die Hypothese, dass es Unterschiede zwischen der Wohnortskategorie bezüglich Zufriedenheit gibt, wird bestätigt (n = 94; gültige Werte: 94 ; Exakter Test nach Fisher; p-Wert = 0.035)“. „L'hypothèse qu'il y a des différences entre les catégories de domicile par rapport à la satisfaction est confirmée (n = 94, observations valides: 94 ; Test exact de Fisher; valeur p = 0.035)“. Wenn der Wert unter exakte Signifikanz grösser als 0.05, z.B. 0.124, würden wir schliessen: „Die Hypothese ... wird nicht bestätigt (n = 94; gültige Werte: 94 ; Exakter Test nach Fisher; p-Wert = 0.124)“. „L'hypothèse.... n'est pas confirmée (n = 94, observations valides: 94 ; Test exact de Fisher; test unilatéral à droite, valeur p = 0.124)“.

– Manchmal wird zusätzlich der Output des Programm R geliefert:

```
Fisher's Exact Test for Count Data
data: a
p-value = 0.6534
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.4949345 3.4644266
sample estimates:
odds ratio
1.309724
```

Das Quotenverhältnis ist 1.309724 (Das Verhältnis der Rauchenden zu den nicht-rauchenden Lehrern ist um 1.31 mal grösser als bei den Nicht-Lehrern). Das Verhältnis ist aber nicht signifikant grösser als 1 (p-Wert 0.6534, zweiseitiger Test). Das wirklich Quotenverhältnis liegt mit einer Wahrscheinlichkeit von 0.95 im Intervall]0.4949345 3.4644266[. Ist eine der Zellen 0, wird das Quotenverhältnis nicht geliefert.

- Bibliographische Angaben. Der Fisher-Test wurde von Fisher (1966) (Sammlung von Artikeln von R. A. Fisher) entwickelt und wird z.B. in Agresti (2002) ausführlich diskutiert.

Der Exakte Test von Fisher ist ein Klassiker. Es ist nicht unbedingt nötig, eine Referenz anzugeben. Wenn Sie eine solche angeben wollen, würde ich Agresti (2002) empfehlen.

Fakultative Erläuterungen zum Fisher-Test Der Fisher-Test geht für den 2X2-Fall von der Hypergeometrischen Verteilung aus. In einer Grundgesamtheit mit n Objekten gibt es n_{1+} Objekte mit der Ausprägung 1 der Variable 1.

	Variable 2		
Variable 1	Ausprägung 1	Ausprägung 2	Summen
Ausprägung 1	a_{11}	a_{12}	n_{1+}
Ausprägung 2	a_{21}	a_{22}	n_{2+}
Summen	n_{+1}	n_{+2}	n

Man zieht aus dieser Grundgesamtheit eine Stichprobe der Grösse n_{+1}, was der Randhäufigkeit der Variable 2 für die erste Ausprägung entspricht. Der $p-$Wert ist die Wahrscheinlichkeit $P(X \geq a_{11})$, dass mindestens eine Häufigeit von a_{11} (rechsseitiger Test) vorliegt , die Wahrscheinlichkeit $P(X \leq a_{11})$, dass die Häufigkeiten höchstens a_{11} (linksseitiger Test) beträgt (mit der Zufallsvariable X, welche Häufigkeiten in der Zelle $(1,1)$ annimmt). Die zugesandten Tabellen liefern dabei den kleineren der beiden Werte, d.h. $\min\{P(X \geq a_{11}), P(X \leq a_{11})\}$ (=exakte Signifikanz (1-seitig)). Für zweiseitige Tests ist der p-Wert $2 \cdot \min\{P(X \geq a_{11}), P(X \leq a_{11})\}$. Für den rechtsseitigen Test rechnet man also

$$P(X \geq a_{11}) = \sum_{i=a_{11}}^{\min(n_{+1},n_{1+})} \frac{\binom{n_{1+}}{i}\binom{n-n_{1+}}{n_{+1}-i}}{\binom{n}{n_{+1}}}$$

für den linksseitigen

$$P(X \leq a_{11}) = \sum_{i=0}^{a_{11}} \frac{\binom{n_{1+}}{i}\binom{n-n_{1+}}{n_{+1}-i}}{\binom{n}{n_{+1}}}$$

Die obige Abweichung 0.6530 von 0.650 ergibt sich durch Rundungen. 0.350 ist ein gerundeter Wert. Im zweiseitigen Test wird der Wert ungerundete Wert von 0.350 mit zwei multipliziert und erst dann gerundet, was im Beispiel zu 0.653 führt.

Für den Fall mit mehr als zwei Ausprägungen wird die verallgemeinerte hypergeometrische Verteilung verwendet. Die Berechnung des p-Wertes erfolgt mit kombinatorischen Methoden, die sehr rechenintensiv sind.

Es gibt übrigens in der Literatur und in manchen Statistikpaketen alternative nicht-äquivalente Berechungen des p-Wertes. Oft wird auch das Chancen-Verhältnis (Odds-Ratio) mit exaktem oder asymptotischem Vertrauensintervall von Statistikprogrammen geliefert (z.B. das Statistikprogramm R unter dem Befehl fisher.test). Mehr dazu finden Sie bei Agresti (2002).

4 Chi-Quadrat-Test von Pearson (Test du Khi-deux de Pearson)

Der Chi-Quadrat-Test wird bei Kreuztabellen (tableau croisé) anwendet, wenn beide Variablen nominalskaliert sind und mindestens eine der Variablen mehr als zwei Ausprägungen aufweist. Zudem wird er bei ordinalskalierten Variablen mit wenigen Ausprägungen gebraucht, falls die Bedingungen für den Gamma-Test nicht erfüllt sind (Konzentration der Häufigkeiten in der Kreuztablle auf der Haupt- oder der Nebendiagonale). Die erwarteten Häufigkeiten (s. für Erklärung in Kürze) dürfen nicht in mehr als 20% der Zellen kleiner als 5 sein. Folgendes müssen Sie wissen, wenn Sie in Ihrer Diplomarbeit Chi-Quadrat-Tests verwenden.

- Sie müssen Kreuztabellen (tableau croisé) interpretieren können (s. entsprechendes Kapitel).
- Für den Test werden drei Tabellen zugeschickt, die Sie interpretieren können müssen, wobei die erste oft die Angaben für mehrere Tests enthält. Sie enthält die Angaben über die gültigen (valide) und die fehlenden (manquante) Fälle (observations). Fehlend ist ein Fall, falls bezüglich der Variable keine Information über die Ausprägung des Objektes vorliegt = fehlendes Datum. Im Beispiel gibt es zwei fehlende Daten. Es wird nicht angegeben, bei welcher Variablen. Wenn man das wissen will, kann man univariate Häufikgeitsverteilung der Variablen anschauen.

Verarbeitete Fälle

	Fälle					
	Gültig		Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
Beruf * Rauchverhalten	92	97.9%	2	2.5%	94	100.0%

Tabelle 13: Tabelle bezüglich gültiger und fehlender Daten (deutsche Beschriftung)

Récapitulatif du traitement des observations

	Observations					
	Valide		Manquante		Total	
	N	Pourcent	N	Pourcent	N	Pourcent
Beruf * Rauchverhalten	92	97.9%	2	2.5%	94	100.0%

Tabelle 14: Tabelle bezüglich gültiger und fehlender Daten (französische Beschriftung)

- Die zweite Tabelle gibt die Kreuztabelle an, zusätzlich zur gemeinsamen Häufigkeitsverteilung (Anzahl, effectif) aber noch die erwarteten Häufigkeiten (= erwartete Anzahlen = fréquences anticipées = effectifs théoriques). Diese sind im Beispiel wie folgt zu verstehen: der Anteil der sehr zufriedenen Personen an der Gesamtstichprobe ist $\frac{27}{92}$ (s. Randverteilung Zufriedenheit mit Schule). Gibt es keinen Zusammenhang zwischen dem Wohnort und der Tatsache, sehr zufrieden zu sein, müssen bei den Unterwallisern $\frac{27}{92}$ sehr zufrieden sein, d.h. der Anteil der sehr zufriedenen bei den Unterwallisern muss mit dem Anteil der sehr zufriedenen Personen in der gesamten Stichprobe übereinstimmen. Es gibt 14 Unterwalliser in der Stichprobe. Also müssten $\frac{27}{92} \cdot 14 = 4.1087$ (= erwartete Anzahl - effectif théorique) Unterwalliser sehr zufrieden sein. Analog müssten bei den Unterwallisern $\frac{25}{92} \cdot 14 = 3.8043$ zufrieden sein, $\frac{22}{92} \cdot 35 = 8.3696$ der Mittelwalliser mässig zufrieden sein, etc. Zudem werden in solchen Tabellen gewöhnlich noch die Differenzen zwischen Anzahl und Erwarteter Anzahl geliefert (Residuum - résidu - genannt; z.B. $13 - 12.6 = 0.4$ in der Zelle $(3, 1)$), damit man sofort sieht, wo und wie stark die faktischen von den erwarteten Häufigkeiten abweichen.

Wohnort * Zufriedenheit mit Schule Kreuztabelle

			Zufriedenheit mit Schule				Gesamt
			sehr zufrieden	zufrieden zufrieden	mässig zufrieden	unzufrieden unzufrieden	
	Unterwallis	Anzahl	4	5	2	3	14
		Erwartete Anzahl	4.1	3.8	3.3	2.7	14
		Residuen	-0.1	1.2	-1.3	0.3	0
Wohnort	Mittelwallis	Anzahl	10	2	12	11	35
		Erwartete Anzahl	10.3	9.5	8.4	6.8	35
		Residuen	-0.3	-7.5	3.6	4.2	
	Oberwallis	Anzahl	13	18	8	4	43
		Erwartete Anzahl	12.6	11.7	10.3	8.4	43
		Residuen	0.4	6.3	-2.3	-4.4	
Gesamt		Anzahl	27	25	22	18	92

Tabelle 15: Kreuztabelle mit faktischen und erwarteten Häufigkeiten und deren Differenz, den Residuen (deutsche Beschriftung)

Tableau croisé Wohnort * Zufriedenheit mit Schule

			Zufriedenheit mit Schule				Total
			sehr zufrieden	zufrieden zufrieden	mässig zufrieden	unzufrieden unzufrieden	
	Unterwallis	Effectif	4	5	2	3	14
		Effectif théorique	4.1	3.8	3.3	2.7	14
		Résidu	-0.1	1.2	-1.3	0.3	0
Wohnort	Mittelwallis	Effectif	10	2	12	11	35
		Effectif théorique	10.3	9.5	8.4	6.8	35
		Résidu	-0.3	-7.5	3.6	4.2	
	Oberwallis	Effectif	13	18	8	4	43
		Effectif théorique	12.6	11.7	10.3	8.4	43
		Résidu	0.4	6.3	-2.3	-4.4	
Total		Effectif	27	25	22	18	92

Tabelle 16: Kreuztabelle mit faktischen und erwarteten Häufigkeiten und deren Differenz, den Residuen (französische Beschriftung)

- Die dritte Tabelle gibt die Testresultate an. Der Chi-Quadrat-Test von Pearson überprüft, ob die erwarteten und faktischen Häufigkeiten so von einander abweichen, dass die Abweichungen gut oder schlecht durch Zufall erklärbar sind.

  Diese Abweichungen sind *schlecht durch Zufall erklärbar*, wenn der p-Wert (valeur p) kleiner als 0.05 ist. Der p-Wert ist hinter „Chi-Quadrat nach Pearson" zu finden in der Spalte „Asymptotische Signifikanz (2-seitig)". Ist der p-Wert kleiner als 0.05, sprechen wir von einer signifikanten Abweichung der faktischen von den erwarteten Häufigkeiten.

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	17.223[a]	6	0.008
Likelihood-Quotient	19.668	6	0.003
Zusammenhang linear-mit-linear	1.996	1	0.158
Anzahl der gültigen Fälle	92		

a. 4 Zellen (33.3%) haben eine erwartete Häufigkeit kleiner 5.
Die minimale erwartete Häufigkeit ist 2.74.

Tabelle 17: Ergebnisse des Chi-Quadrat-Testes nach Pearson (deutsche Beschriftung)

Tests du Khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)
Khi-deux de Pearson	17.223[a]	6	0.008
Rapport de vraisemblance	19.668	6	0.003
Association linéaire par linéaire	1.996	1	0.158
Nombre d'observations valides	92		

a. 4 cellules (33.3%) ont un effectif théorique inférieur à 5.
L'effectif théorique minimum est de 2.74.

Tabelle 18: Ergebnisse des Chi-Quadrat-Testes nach Pearson (französische Beschriftung)

Im Beispiel sind in einem Drittel der Zellen die erwarteten Häufigkeiten kleiner als 5. Entsprechend ist der in der obigen Tabelle gelieferte p-Wert unzuverlässig und es wird noch entweder ein exakter Test oder eine Monte-Carlo-Simulation geliefert. Werden die Ergebnisse eines exakten Testes oder einer Monte-Carlo-Simulation geliefert, sind diese zu verwenden (statt der asymtotischen des Chi-Quadrat-Tests von Pearson).

```
> chisq.test(table(a),simulate=T,B=10000)

        Pearson's Chi-squared test with simulated p-value (based on 10000
        replicates)

data:  table(a)
X-squared = 17.2234, df = NA, p-value = 0.007399
```

Der p-Wert ist $0.007 < 0.05$. Also unterscheiden sich die faktischen signifikant von den erwarteten Häufigkeiten der Tabelle (Im Beispiel unterscheidet sich der asymptotische p-Wert nur unbedeutend von dem aus der Simulation berechneten p-Wert).

- Angabe der Ergebnisse in der Diplomarbeit: Der Chi-Quadrat-Test gibt nicht an, wo oder in welche Richtungen Abweichungen der faktischen von den erwarteten Häufigkeiten vorliegen. Entsprechend sollten in der Diplomarbeit die Kreuztabelle mit den Häufigkeiten angegeben werden. Der Leser kann dann überprüfen sehen, wo Abweichungen vorliegen. Der Chi-Quadrat-Test erlaubt es nicht, von einzelnen Abweichungen zu sagen, dass sie signifikant sind oder nicht. Allerdings ist davon auszugehen, dass vor allem die grössten Abweichungen für die signifikante Abweichung der gesamten Tabelle verantwortlich sind. Sagt man von der grössten Abweichung, sie sei signifikant (= schlecht durch Zufall erklärbar), ist das sicher nicht falsch.
  Zitierbeispiel: „Die Hypothese, dass die Zufriedenheit mit der Schule vom Wohnort unabhängig ist, kann verworfen werden (n = 94; gültige Werte: 92 ; Chi-Quadrat-Test nach Pearson; p-Wert = 0.007)". „L'hypothèse que la satisfaction avec l'école est indépendante du domicile peut être rejetée (n = 94, observations valides: 92 ; Test du Khi-deux de Pearson; valeur p = 0.007)".
  Wäre der Wert unter „Asymptotische Signifikanz (2-seitig)" grösser als 0.05, z.B. 0.452, würden wir schliessen: „Die Hypothese, dass die Zufriedenheit mit der Schule vom Wohnort unabhängig ist, kann nicht verworfen werden (n = 94; gültige Werte: 92 ; Chi-Quadrat-Test nach Pearson; p-Wert = 0.452)". „L'hypothèse que la satisfaction avec l'école est indépendante du domicile ne peut pas être rejetée (n = 94, observations valides: 94 ; Test exact de Fisher; valeur p = 0.452)".
- Bibliographische Angaben. Der Chi-Quadrat-Test nach Pearson wurde von Pearson (1900) entwickelt und wird z.B. in Agresti (2002) ausführlich diskutiert. Die korrekte Berechnungsmethode der Freiheitgrade wurde allerdings erst von Fisher (1922) entwickelt. Der Chi-Quadrat-Test nach Pearson ist ein Klassiker, der fast in jedem Lehrbuch zur Statistik zu finden ist. Es ist nicht unbedingt nötig, eine Referenz anzugeben. Wenn Sie eine solche angeben wollen, würde ich Agresti (2002) empfehlen.

Fakultative Erläuterungen zum Chi-Quadrat-Test nach Pearson Beim Chi-Quadrat-Test nach Pearson interpretiert man jede einzelne Häufigkeit einer Zelle der Kreuztabelle als Wert einer poisson-verteilten Zufallsvariable, wobei der Erwartungswert dieser Zuvallsvariable mit der erwarteten Häufigkeit identifiziert wird. Die Varianz einer poisson-verteilten Zufallsvariable ist mit deren Erwartungswert (espérance) identisch. Gemäss zentralem Grenzwertsatz ist damit

$$\frac{K_{ij} - k_{ij}^e}{\sqrt{k_{ij}^e}}$$

asymptotisch standardnormalverteilt (K_{ij} ist die Zufallsvariable, welche konkrete Häufigkeiten als Wert annimmt, k_{ij}^e ist die erwartete Häufigkeit). Entsprechend ist gemäss Definition der Chi-Quadrat-Verteilung

$$\frac{\left(K_{ij} - k_{ij}^e\right)^2}{k_{ij}^e}$$

asymptotisch chi-quadrat-verteilt mit einem Freiheitsgrad. Deshalb ist

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(K_{ij} - k_{ij}^e\right)^2}{k_{ij}^e}$$

asymptotisch chi-quadrat-verteilt mit $(I-1)(J-1)$ Freiheitsgraden (I ist die Anzahl der Zeilen der Kreuztabelle und J die Anzahl der Spalten der Kreuztabelle). Freiheitsgrade sind definiert als die Anzahl der unabhängigen Zufallsvariablen, welche in die Summe χ^2 einfliessen. Da die erwarteten Häufigkeiten aus den Randverteilungen berechnet werden, ist pro Zeile und pro Spalte jeweils eine Zellenhäufigkeit fixiert und damit nicht unabhängig, sobald die übrigen Zellenhäufigkeiten festgelegt sind. Entsprechend verlieren wir in jeder Zeile und in jeder Spalte je einen Freiheitsgrad.

5 Rang-Test nach Mann-Whitney (Test de Mann-Whitney)

Falls in Ihrer Diplomarbeit Ergebnisse des Rang-Testes nach Mann-Whithney verwendet werden, sollten Sie folgendes Wissen:

- Der Rang-Test nach Mann-Whithney wird bei einer nominalskalierten Variable mit zwei Ausprägungen (hier „Gruppen" gennannt) und einer ordinalskalierten Variable mit vielen Ausprägungen angewendet. Ist die zweite Variable metrisch skaliert, wird der Rang-Test verwendet, wenn die Voraussetzung der Normalverteiltheit der Daten für den t-Test nicht gegeben sind. Die Daten werden rangiert (je grösser der Wert der Variable, desto grösser der Rang). Es wird getestet, ob sich die Rangmittel (= arithmetisches Mittel der Ränge, moyenne arithmetique des rangs) der beiden Gruppen unterscheiden (zweiseitiger Test, test bilatéral), ob das Rangmittel einer bestimmten Gruppe grösser ist als das Rangmittel der anderen Gruppe (einseitiger Test, test unilatéral).
- Für den Test werden zwei Tabellen zugesandt. Im Beispiel stellt jemand die Hypothese auf, Lehrpersonen hätten in der Mittelschule bessere Mathematiknoten gehabt als Nicht Lehrpersonen. Es liegt also ein einseitiger Test vor. Ein zweiseitiger Test würde vorliegen, falls die Hypothese aufgestellt wird, die Mathematiknoten der beiden Gruppen würden sich unterscheiden. Die erste Tabelle gibt die Rangsummen (= Summe der Ränge) und Rangmittel pro Gruppe an - im Beispiel für Lehrpersonen und Nicht Lehrpersonen). Es zeigt sich, dass das Rangmittel der Lehrpersonen etwas tiefer liegt als das der anderen Personen.

Ränge

	Beruf	N	Mittlerer Rang	Rangsumme
Noten Mathematik	Lehrperson	29	45.9	1331
	Nicht Lehrperson	65	48.22	3134
	Gesamt	94		

Tabelle 19: Kennzahlen für den Mann-Whitney-Test (deutsche Beschriftung)

Rangs

	Beruf	N	Rang moyen	Somme des rangs
Noten Mathematik	Lehrperson	29	45.9	1331
	Nicht Lehrperson	65	48.22	3134
	Total	94		

Tabelle 20: Kennzahlen für den Mann-Whitney-Test (französische Beschriftung)

- Um zu überprüfen, ob diese Differenz schlecht oder gut durch Zufall erklärbar ist, betrachten wir die zweite Tabelle:

Statistik für Test[a]

	Noten Mathematik
Mann-Whitney U	896
Wilcoxon W	1331
Z	-0.381
Asymptotische Signifikanz (2-seitig)	0.703
a. Gruppenvariable: Beruf	

Tabelle 21: p-Wert für den Mann-Whitney-Test (deutsche Beschriftung)

Test[a]

	Noten Mathematik
U de Mann-Whitney	896
W de Wilcoxon	1331
Z	-0.381
Signification asymptotique	0.703
(bilatérale)	0.703
a. Critère de regroupement : Beruf	

Tabelle 22: p-Wert für den Mann-Whitney-Test (französische Beschriftung)

Bei „Asymtptotische Signfikanz (zweiseitig)“ (Signfication asymptotique (bilatérale)) wird der asymptotische zweiseitige p-Wert (valeur p) angegeben. Bei kleineren Fallzahlen wird noch ein exakter zweiseitiger p-Wert (valeur p exact bilatéral) angegeben. Ist dies der Fall, sollte der exakte verwendet werden.

- Angabe der Ergebnisse in der Diplomarbeit: Bei der Verwendung von Ergebnissen mit dem Mann-Whitney-Test sollten Sie die Tabelle mit den Rangmitteln in der Diplomarbeit liefern - alternativ kann man aber auch die Rangmittel der beiden Gruppen im Text angeben. Im Text würde man im Falle eines *zweiseitigen Testes* fürs Beispiel schreiben: „Die Hypothese, dass die Mathematiknoten von Lehrpersonen sich von denen anderer Personen unterscheiden, ist nicht bestätigt (n = 94; gültige Werte: 94 ; Mann-Whitney-Test; p-Wert = 0.703)“. „L'hypothèse que les enseignants ont des notes de mathématiques différentes des autres personnes n'est pas confirmée (n = 94, observations valides: 94 ; Test de Mann-Whithney; valeur p = 0.703)“. Wenn Sie nicht die Tabelle mit den Rangmitteln liefern, kann man die Rangmittel in dieser Klammer angeben: „Die Hypothese, dass die Mathematiknoten von Lehrpersonen sich von denen anderer Personen unterscheiden, ist nicht bestätigt (n = 94; gültige Werte: 94 ; Mann-Whitney-Test; p-Wert = 0.703, Rangmittel Lehrpersonen 45.9, Rangmittel Nicht Lehrpersonen 48.22)“.
  Wenn der p-Wert kleiner-gleich 0.05 ist, z.B. 0.021, würde man die Hyothesen nicht verneinen: „Die Hypothese, dass die Mathematiknoten von Lehrpersonen sich von denen anderer Personen unterscheiden, ist bestätigt (n = 94; gültige Werte: 94 ; Mann-Whitney-Test; p-Wert = 0.021)“. „L'hypothèse que les enseignants ont des notes de mathématiques différentes des autres personnes est confirmée (n = 94, observations valides: 94 ; Test de Mann-Whithney; valeur p = 0.021)“.

Im Falle eines *einseitigen Testes* überprüft man, ob die Rangmittel die gemäss Hypothese erforderliche Ordnung aufweisen. Weist die Gruppe, die gemäss Hypothese ein höheres Rangmittel aufweisen müsste ein kleineres Rangmittel auf, würde man schreiben: „Da die Lehrpersonen ein kleineres Rangmittel aufweisen als die übrigen Personen, ist die Hypothese widerlegt. Ein statistischer Test erübrigt sich“. „Comme les enseignants ont une moyenne de rangs inférieure à celle des autres personnes, l'hypothèse ist infirmée. Il n'est pas nécessaire de faire un test statistique“.
Weist die Gruppe, die gemäss Hypothese ein höheres Rangmittel aufweisen müsste, tatsächlich ein solches auf, so wird der gelieferte p-Wert halbiert. Würden wir die Hypothese „Nicht Lehrpersonen haben höhere Mathematiknoten als Lehrpersonen“ testen, würde es sich um einen einseitigen Test handeln. Wir schreiben: "Die Hypothese, dass Nicht Lehrpersonen höhere Mathematiknoten haben als Lehrpersonen, wird nicht bestätigt (n=94; gültige Werte: 94; Mann-Whitney-Test, p-Wert: 0.352)“. „L'hypothèse que les autres personnes ont des notes de mathématiques supérieure aux enseignants n'est pas confirmée (n=94; valeurs valides: 94; Test de Mann-Whitney, valeur p = 0.352)“.

- Bibliographische Angaben: Der Mann-Whitney-Test geht auf Mann und Whitney (1947) zurück. Es ist nicht nötig, diese Referenz anzugeben. Wenn Sie eine allgemeine Referenz für Rangtests liefern wollen, könnte man Hettmansperger (1984) oder Büning und Trenkler (1994) angeben.

Fakultative Erläuterungen zum Mann-Whitney-Test Exakte p-Werte werden mit Hilfe der Kombinatorik berechnet. Man geht davon aus, dass jede Rangfolge dieselbe Wahrscheinlichkeit hat und zählt die Rangfolgen mit kleiner oder grösseren Rangmitteln. Für den asymtotischen p-Wert werden die folgenden Grössen verwendet:

$U_1 := \left\lvert\{(x_i, x_j) \mid r_{x_i} > r_{x_j} \text{ und } x_i \in S_1, x_j \in S_2\}\right\rvert$
$S_i :=$ Menge der Objekte der Stichprobe i
$r_{x_j} :=$ Rang des Objektes x_j.

$Z = \frac{U_i - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}} \sim N(0,1)$ (asymptotisch)
U_i : Zufallsvariable $= n_1 n_2 + \frac{n_i(n_i+1)}{2} - R_i \quad (i \in \{1,2\})$.
R_i : Zufallsvariable (Wert von $R_i = \sum_{j=1}^{n_i} r_i(x_j)$)
$r_i(x_j) =$ Rang des Datums x_j der Stichprobe i
n_1 : Umfang der ersten Stichprobe
n_2 : Umfang der zweiten Stichprobe

Oft wird für die Berechnung von Z bei verbunden Rängen, d.h. wenn verschiedene Daten dieselbe Ausprägung haben, eine Korrektur am Zähler (= Standardabweichung von U_i) vorgenommen.

6 Rangtest nach Kruskal-Wallis (Test de Kruskal-Wallis)

Falls in Ihrer Diplomarbeit Ergebnisse des Rang-Testes nach Mann-Whithney verwendet werden, sollten Sie folgendes Wissen:

- Der Kruskal-Wallis-Test wird verwendet, wenn eine nominalskalierte Variable (hier Gruppenvariable genannt) und eine abhängige ordinalskalierte Variable mit vielen Ausprägungen vorliegen. Er wird auch statt der Einfaktoriellen Varianzanalyse (ANOVA) verwendet, wenn die metrische Variable die Bedingungen für die Anwendung der ANOVA nicht erfüllt (Normalverteiltheit der Residuen; in allen Gruppen identische Varianzen). Es wird getestet, ob sich die Rangmittel der verschiedenen Gruppen unterscheiden oder nicht.

- Für den Test werden zwei Tabellen zugesandt. Im Beispiel stellt jemand die Hypothese auf, je nach Lehrmethode (es werden drei angewendet) hätten Schüler unterschiedliche Noten. Die erste Tabelle gibt die Rangmittel pro Gruppe an - im Beispiel für die drei Methoden). Es zeigt sich, dass die erste Gruppe zwischen den beiden anderen Gruppen liegt und dass die dritte (freies Arbeiten) am schlechtesten abschneidet.

Ränge

	Methode	N	Rangmittel
Noten	frontal	19	49.921
	frontal mit Gruppenarbeit	21	73.714
	freies Arbeiten	54	36.454
	Gesamt	94	

Tabelle 23: Rangmittel für den Kruskal-Wallis-Test (deutsche Beschriftung)

Rangs

	Methode	N	Rang moyen
Noten	frontal	19	49.921
	frontal mit Gruppenarbeit	21	73.714
	freies Arbeiten	54	36.454
	Total	94	

Tabelle 24: Rangmittel für den Kruskal-Wallis-Test (französische Beschriftung)

- Um die Frage zu überprüfen, ob diese Unterschiede gut oder schlecht durch Zufall erklärbar sind, wird der Kruskal-Wallis-Test druchgeführt, dessen Ergebnis in der folgenden Tabelle erscheint - der p-Wert kann bei „Asymptotische Signifikanz“ (Signification asymptotique) abgelesen werden.Werden neben der asymptotischen Signifikanz auch die Ergebnisse eines exakten Tests geliefert, so sind diese zu verwenden.

- Angabe der Ergebnisse in der Diplomarbeit: Es sollte die Tabelle mit den Rangmittel geliefert werden, damit der Leser sich ein Bild über die Gruppenrangmittel verschaffen kann. Da der p-Wert kleiner ist als 0.05, wird die Hypothese, dass es Unterschiede gibt, im Beispiel angenommen. Man könnte z.B. schreiben „Die Hypothese, dass die drei Methoden zu unterschiedlichen Resultaten führt, wird bestätigt (n = 94; gültige Werte: 94; Kruskal-Wallis-Test, p-Wert = 0.000)“. „L'hypothèse, que les trois méthodes mènent à des résultats différents, est confirmée (n = 94; valeurs valides: 94; Test de Kruskal-Wallis, valeur p = 0.000)“. Der

Test^{a,b}

	Noten
Chi-Quadrat	28.525
df	2
Asymptotische Signifikanz	6.397E-07

a. Kruskal-Wallis-Test
b. Gruppenvariable: Methode

Tabelle 25: p-Wert für den Kruskal-Wallis-Test (deutsche Beschriftung)

Test^{a,b}

	Noten
Khi-deux	28.525
ddl	2
Signification asymptotique	6.397E-07

a. Test de Kruskal Wallis
b. Critère de regroupement : Methode

Tabelle 26: p-Wert für den Kruskal-Wallis-Test (französische Beschriftung)

Kruskal-Wallis-Test erlaubt es nicht, zu sagen, welche Gruppe sich von welcher Gruppe signifikant unterscheidet. Wo und in welche Richtung Unterschiede bestehen, kann man von den Rangmitteln ablesen.

- Bibliographische Angaben: Der Kruskal-Wallis-Test geht auf Kruskal und Wallis (1952) zurück. Es ist nicht unbedingt nötig, diese Referenz anzugeben. Will man allgemein eine Referenz zu Rangtests angeben, kann man Hettmansperger (1984) oder Büning und Trenkler (1994) angeben.

Fakultative Erläuterungen zum Kruskal-Wallis-Test Die Daten werden durchrangiert und dann die Rangmittel pro Gruppe berechnet. Der Test wird wie folgt berechnet:

$K = \frac{12}{n(n+1)} \sum_{j=1}^{m} k_j \left(\bar{R}_j - \frac{n+1}{2}\right)^2$ ist asymptotisch χ^2_{m-1}−verteilt.
$\bar{R}_j = \frac{1}{k_j} \sum_{i=1}^{k_j} r_{ij}$ (Durchschnittlicher Rang der Gruppe j)
r_{ij} = Rang des Datums x_{ij}
n: Anzahl der Daten
k_j : Anzahl der Daten der Gruppe j
m: Anzahl der Gruppen

7 Jonckheere-Terpstra-Test (Test de Jonckheere-Terpstra)

Wenn Sie in Ihrer Diplomarbeit einen Jonckheere-Terpstra-Test verwenden, so sollten Sie wissen,

- dass es darum geht, bei zwei ordinalskalierten Variablen, von denen die unabhängige Variable relativ undifferenziert ist (z.B. nur 4 oder 5 Ausprägungen) und die abhängige Variable differenziert ist (viele Ausprägungen, z.B. 20), zu entscheiden, ob bei den differenzierten Daten bezüglich aufsteigender Ordnung des Faktors ein Trend vorliegt oder nicht. So möchte man z.B. untersuchen, ob 4 Jahrgänge einer Primarschule bezüglich Konzentrationsfähigkeit einen Trend aufweisen (je älter, desto konzentrationsfähiger), wobei wir voraussetzen, dass Konzentratiosfähigkeit relativ differenziert gemessen werden kann (sonst würde man Gamma, s.u., verwenden).

- Wird ein Jonckheere-Terpstra-Test zugesandt, so werden zwei Tests geliefert: der Kruskal-Wallis-Test (s. oben). In diesem müssen Sie nur die Tabelle „Ränge“ anschauen. Sie liefert die Rangmittel. Es ist zu prüfen, ob die Tendenz der Rangmittel der Hypothese entspricht. Im folgenden Beispiel (je älter, desto konzentrierter) stimmt die Ordnung der Rangmittel mit der Hypothese überein (s. 27 und 28).

Ränge			
	Alter	H	Mittlerer Rang
Konzentrationleistung	9.0	10	15.90
	10.0	9	16.67
	11.0	18	25.58
	12.0	5	26.70
	Gesamtsumme	42	

Tabelle 27: Beschreibende Statistik zum Jonckheere-Terpstra-Test. Bei H stehen die Häufigkeiten, es hat z.B. 10 Personen in der Altersklasse 9. Die Rangmittel weisen in Übereinstimmung mit der Hypothese eine aufsteigende Tendenz auf.

Rangs			
	Alter	N	Rang moyen
Konzentrationleistung	9.0	10	15.90
	10.0	9	16.67
	11.0	18	25.58
	12.0	5	26.70
	Total	42	

Tabelle 28: Statistiques descriptives pour le test de Jonckheere-Terpstra. N indique les fréquences, il y a p.ex. 10 personnes dans la classe d'âge 9. Les moyennes de rang sont en concordance avec l'hypothèse d'une tendance croissante avec l'âge.

- Die zweite Tabelle, die Ergebnisse des Kruskal-Wallis-Tests, können Sie übergehen. Als dritte Tabelle werden die Ergebnisse des Jonckheere-Terpstra-Tests geliefert. Im Beispiel (s. Tabelle 29 und 30):

Jonckheere-Terpstra-Test[a]			
			KL (Konzentrationleistung)
Anzahl der Ebenen in Alter			4
H			42
Beobachtete J-T-Statistik			411.000
Mittelwert-J-T-Statistik			308.500
Standardabweichung der J-T-Statistik			43.498
Standard-J-T-Statistik			2.356
Asymp. Sig. (2-seitig)			.018
Monte-Carlo-Sig. (2-seitig)	Sig.		.018[b]
	99 % Konfidenzintervall	Untergrenze	.014
		Obergrenze	.021
Monte-Carlo-Sig. (1-seitig)	Sig.		.009[b]
	99 % Konfidenzintervall	Untergrenze	.006
		Obergrenze	.011

a. Gruppierungsvariable: Alter
b. Basierend auf 10000 Stichprobentabellen mit dem Startwert 624387341.

Tabelle 29: Ergebnisse des Jondkheere-Terpstra-Testes. Der p-Wert beträgt 0.008. Damit ist die steigende Tendenz signifikant von einer Konstanten verschieden.

Test de Jonckheere-Terpstra[a]			
			KL (Konzentrationleistung)
Nombre de niveaux dans Alter			4
N			42
Test de Jonckheere-Terpstra observé			411.000
Moyenne du test de J-T			308.500
Ecart type du test de Jonckheere-Terpstra			43.498
Test de Jonckheere-Terpstra standard			2.356
Sig. asymptotique (bilatérale)			.018
Sig. Monte Carlo	Sig.		.018[b]
(bilatérale)	Intervalle de confiance	Borne inférieure	.014
	à 99 %	Borne supérieure	.021
Sig. Monte Carlo	Sig.		.009[b]
(unilatérale)	Intervalle de confiance	Borne inférieure	.006
	à 99 %	Borne supérieure	.011

a. Variable de regroupement : Alter
b. Basée sur 10000 tables échantillonnées avec valeur de départ 299883525.

Tabelle 30: Résultats du test de Jondkheere-Terpstra. La valeur p se monte à 0.008. Ainsi la tendance croissante des moyennes de rang est significativement différente d'une constante.

Wird ein Monte-Carlo-Test geliefert wie im Beispiel, sind die entsprechenden Ergebnisse zu verwenden. Wird ein exakter Test geliefert, sind die Ergebnisses dieses Testes zu verwenden.

- Angabe der Ergebnisse in der Diplomarbeit: Bei der Verwendung von J-T-Testes sollte in der Diplomarbeit die Tabelle mit den Rangmitteln geliefert werden, damit der Leser sieht, wie die Variablen zusammenhängen. Er kann dann die Interpretation des Testes kontrollieren und die Wichtigkeit der Unterschiede zwischen den Gruppen beurteilen.
  Im Text wird bei einer rechtsseitigen Hypothese (steigender Zusammenhang) im obigen Beispiel geschrieben: „Die Hyothese, dass die Konzentrationsfähigkeit mit dem Alter zunimmt, wird bestätigt (n = 42; gültige Wert: 42; einseitiger Test; p-Wert = 0.008)“. „L'hypothèse que la concentration augmente avec l'âge est confirmée (n = 42; observations valides: 42; test unilatéral, p-valeur = 0.009)“ (il y a différence entre la version francophone et germanophone, parce la valeur p est le résultat d'un test monte carlo).
  Für eine linksseitigen Hypothese (sinkende Tendenz) würde man im obigen Beispiel schreiben: „Die Hyothese, dass je älter jemand ist, desto tiefer ist die Konzentrationsfähigkeit, wird nicht bestätigt (n = 42; gültige Wert: 42; linkseitiger Test, die Tendenz ist wachsend und nicht sinkend)“. „L'hypothèse que plus âgé l'élève, moins grande est la concentration est infirmée (n = 42; observations valides: 42; test unilatéral à gauche, la tendance est croissante)“. Man gibt hier keinen p-Wert an, da die Richtung des Zusammenhangs (positive Steigung) nicht mit der Hypothese übereinstimmt, gemäss derer sich eine negative Steigung ergeben müsste.
  Bei einer zweiseitigen Hypothese würde man postulieren, dass es eine Tendenz zwischen Alter und Konzentrationsfähigkeit gibt, die Richtung wird aber nicht bestimmt (also steigend oder sinkend). Man würde schreiben: „Die Hyothese, dass es einen steigenden oder sinkenden Zusammenhangs zwischen Alter und Konzentrationsfähigkeit gibt, wird bestätigt (n = 42; gültige Werte: 42; zweiseitiger Test; p-Wert = 0.016)“. „L'hypothèse qu'il y a une tendance croissante ou décroissante entre l'âge et la concentration est confirmée (n = 42; observations valides: 42; test bilatéral; p-valeur = 0.019)“.

- Bibliographische Angaben: Der J-T-Test wird in Artikeln von Jonckheere (1954a), Jonckheere (1954b) und Terpstra (o. J.) vorgestellt. Es ist nicht nötig, eine Referenz für den Test anzugeben. Wenn Sie eine solche angeben wollen, würde ich Büning und Trenkler (1994) empfehlen.

Fakultative Erläuterungen zum Jonckheere-Terpstra-Test Berechnet wird die folgende Teststatistik, wobei die Kategorien des Faktors von links nach rechts aufsteigend nummeriert sind:

$$J := \sum_{i=1}^{c-1} \sum_{j=i+1}^{c} U_{ij}$$

mit der Anzahl Klassen c des Faktors und mit

$$U_{ij} := \sum_{s=1}^{n_i} \sum_{t=1}^{n_j} \Psi\left(X_{jt} - X_{is}\right),$$

so dass

$$\Psi\left(X_{jt} - X_{is}\right) := \begin{cases} 1 & \text{für } X_{jt} - X_{is} > 0 \\ 0 & \text{sonst} \end{cases},$$

wobei n_i sowie n_j die Anzahl der Messungen X_{jt} , X_{is} der Klassen i und j des Faktors sind. Man zählt also die Anzahl der Paare, so dass $X_{jt} - X_{is} > 0$ für Klassen j, die rechts der Klassen i liegen. Man kann zeigen, dass die Statistik

$$Z = \frac{J - E(J)}{\sqrt{Var(J)}}$$

näherungsweise standardnormalverteilt ist, wobei mit $N = \sum_i^c n_i$ (= Anzahl Daten)

$$E(J) = \frac{N^2 - \sum_i^c n_i^2}{4}$$
$$Var(J) = \frac{N^2(2N+3) - \sum_i^c n_i^2(2n_i+3)}{72}.$$

Beim Monte-Carlo-Test werden ein paar tausend Stichproben gezogen und die Verteilung der Teststatistik J empirisch bestimmt. Beim exakten Test wird die Verteilung der Teststatistik J mit Hilfe von Kombinatorik bestimmt.

8 Gamma

Wenn Sie in Ihre Diplomarbeit einen Test von Gamma verwenden, sollten Sie folgendes wissen.

- Sie sollten Kreuztabellen (tableaux croisés) interpretieren können (s. entsprechendes Kapitel).
- Sie sollten wissen, dass Gamma eine Art Korrelation für ordinalskalierte Variablen ist mit jeweils wenigen Ausprägungen (z.B. die Variable „Ausbildungsgrad“ mit den Ausprägungen „Sek I“, „Sek II“, „Tertiäre Ausbildung“ und die Variable „Zufriedenheit mit Schule“ mit den Ausprägungen „sehr zufrieden“, „zufrieden“, „mässig zufrieden“, „unzufrieden“). Ist eine der beiden Variablen nominalskaliert mit nur zwei Ausprägungen und die andere ordinalskaliert mit wenig Ausprägungen kann Gamma ebenfalls sinnvoll verwendet werden (z. B. die Variable „Beruf“ mit den Ausprägungen „Lehrperson“ und „Nicht Lehrperson“ und die Variable „Zufriedenheit mit Schule“ mit obigen Ausprägungen).
- Wird ein Gamma-Test zugesandt, so werden drei Tabellen geliefert, die Sie interpretieren können müssen, wobei die erste oft die Angaben für mehrere Tests enthält. Sie enthält die Angaben über die gültigen (valide) und die fehlenden (manquante) Fälle (observations). Fehlend ist ein Fall, falls bezüglich der Variable keine Information über die Ausprägung des Objektes vorliegt = fehlendes Datum. Im Beispiel gibt es keine fehlende Daten.

Verarbeitete Fälle

	Fälle					
	Gültig		Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
Zufriedenheit mit Schule * Ausbildung	94	100%	0	0%	94	100%

Tabelle 31: Tabelle bezüglich gültiger und fehlender Daten (deutsche Beschriftung)

Récapitulatif du traitement des observations

	Observations					
	Valide		Manquante		Total	
	N	Pourcent	N	Pourcent	N	Pourcent
Zufriedenheit mit Schule * Ausbildung	94	100%	0	0%	94	100%

Tabelle 32: Tabelle bezüglich gültiger und fehlender Daten (französische Beschriftung)

- Zweitens wird pro Test eine Kreuztabelle zugeschickt. Für Gamma sollten die Häufigkeiten auf der Hauptdiagonalen (von oben links nach unten rechts) oder auf der Nebendiagonalen (von oben links nach unten rechts) der Kreuztabelle liegen. Es ist auch möglich, dass die Häufigkeiten U-förmig angeordnet sind - dann ist die Berechnung von Gamma nicht angesagt. Die Kreutzabelle dient dazu, Gamma zu interpretieren: liegen die Häufigkeiten z.B. auf der Hauptdiagonalen, so gilt : je weiter rechts man in der Kreuztabelle liegt, desto weiter unten liegt man tendenzmässig in der Kreutztabelle (oder je weiter unten, desto weiter rechts). Was das inhaltlich bedeutet, kann an den Ausprägungen der Variablen in der Kreuztabelle abgelesen werden. In der folgenden Tabelle heisst das: je höher das Ausbildungniveau, desto zuzufriedener ist jemand. Liegen die Häufigkeiten auf der Nebendiagonale, würde das heissen: je weiter unten, desto weniger weit links (z.B. je höher das Ausbildungsniveau, desto zufriedener mit der Schule).

Zufriedenheit mit Schule * Ausbildung Kreuztabelle

Anzahl

		Ausbildung			Gesamt
		Volksschule Volksschule	Berufslehre oder Sek II	Tertiäre Ausbildung	
Zufriedenheit mit Schule	sehr zufrieden	14	9	5	28
	zufrieden	1	11	14	26
	mässig zufrieden	0	5	17	22
	unzufrieden	0	3	15	18
Gesamt		15	28	51	94

Tabelle 33: Kreuztabelle für Gamma-Test. Gamma ist positiv, wenn sich die Häufigkeiten auf der Hauptdiagonale massieren (wie im Beispiel, d.h. von oben links nach unten rechts), negativ, wenn sie sich auf der Nebendiagonale massieren (d.h. von oben rechts nach unten links). (deutsche Beschriftung)

Tableau croisé Zufriedenheit mit Schule * Ausbildung

Effectif

		Ausbildung			Total
		Volksschule Volksschule	Berufslehre oder Sek II	Tertiäre Ausbildung	
Zufriedenheit mit Schule	sehr zufrieden	14	9	5	28
	zufrieden	1	11	14	26
	mässig zufrieden	0	5	17	22
	unzufrieden	0	3	15	18
Total		15	28	51	94

Tabelle 34: Kreuztabelle für Gamma-Test. Gamma ist positiv, wenn sich die Häufigkeiten auf der Hauptdiagonale massieren (wie im Beispiel, d.h. von oben links nach unten rechts), negativ, wenn sie sich auf der Nebendiagonale massieren (d.h. von oben rechts nach unten links). (französische Beschriftung)

- Drittens wird eine Tabelle mit Auswertungsergebnissen geliefert. Gamma, eine Kennzahl des ordinalen Zusammenhangs, die wie die Pearson-Korrelation zwischen -1 und 1 liegt, wird unter „Wert“ (valeur) und hinter „Gamma“ angegeben. Zudem wird der zweiseitige p-Wert geliefert - unter „Näherungsweise Signifikanz“ (Signification approximée) sowie die exakte Signifikanz - verwenden Sie die exakte Signifikanz! Der zweiseitige p-Wert entspricht der Hypothese, dass Gamma von 0 verschieden ist, d.h. dass es einen negativen oder positiven Zusammenhang zwischen den beiden Variablen hat. Ein rechtsseitger Test entspricht der Hypothese, dass Gamma grösser als 0 ist. Ein linksseitiger Test entspricht der Hypothese, dass Gamma kleiner als 0 ist. Einseitige p-Werte erhält man, indem man den zweiwertigen Wert halbiert. Zu beachten ist dabei, ob die Richtung des Zusammenhangs mit der einseitigen Hypothese übereinstimmt. Hat man eine Hypothese, welche gemäss Anordnung der Ausprägungen der Variablen dann bestätigt wird, wenn die Häufigkeiten auf der Hauptdiagonalen liegen, muss Gamma positiv sein und der halbierte zweiseitige p-Wert kleiner-gleich 0.05, damit die Hypothese bestätigt ist. Hat man eine Hypothese, welche gemäss Anordnung der Ausprägungen der Variablen dann bestätigt wird, wenn die Häufigkeiten auf der Nebendiagonalen liegen, muss Gamma negativ sein und der halbierte zweiseitige p-Wert kleiner-gleich 0.05, damit die Hypothese bestätigt ist. Im Beispiel wird die Hypothese, dass die Zuzufriedenheit mit der Schule mit dem Ausbildungsniveau steigt, bestätigt (Gamma = 0.736, p-Wert = 0.000). Werden in der Tabelle mit den Testergebnisse auch Ergebnisse eines exakten Tests geliefert, so sind diese zu verwenden.

Symmetrische Maße

		Wert	Asymptotischer Standardfehler [a]	Näherungsweises T[b]	Näherungsweise	Exakte Signifikanz
Ordinal- bzgl. Ordinalmaß	Gamma	0.73630635	0.08062502	7.018399515	2.2442E-12	0.000
	Anzahl der gültigen Fälle	94				

a. Die Null-Hyphothese wird nicht angenommen.
b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

Tabelle 35: Gamma steht unter „Wert“ (Valeur), der zweiseitige p-Wert unter „Näherungsweise Signifikanz“(deutsche Beschriftung)

Mesures symétriques

		Valeur	Erreur standard asymptotique[a]	T approximé[b]	Signification approximée	Signification exacte
Ordinal par Ordinal	Gamma	0.73630635	0.08062502	7.018399515	2.2442E-12	0.000
	Nombre d'observations valides	94				

a. L'hypothèse nulle n'est pas considérée.
b. Utilisation de l'erreur standard asymptotique dans l'hypothèse nulle.

Tabelle 36: Gamma steht unter „Wert“ (Valeur), der zweiseitige p-Wert unter „Näherungsweise Signifikanz“(französische Beschriftung)

- Angabe der Ergebnisse in der Diplomarbeit: Bei der Verwendung von Gamma sollte in der Diplomarbeit die Kreuztabelle geliefert werden, damit der Leser sieht, wie die Variablen zusammenhängen. Er kann dann Gamma sinnvoll interpretieren.
  Im Text wird bei einer zweiseitigen Hypothese im obigen Beispiel geschrieben: „Die Hyothe-

se, dass es einen Zusammenhang zwischen der Variable Ausbildungsgrad und Zufriedenheit mit der Schule gibt, wird bestätigt (n = 94; gültige Wert: 94; Gamma = 0.736, zweiseitiger Test; p-Wert = 0.000)". „L'hypothèse qu'il y a une corrélation entre la variable „Ausbildungsgrad" et la variable „Zufriedenheit mit der Schule" est confirmée (n = 94; observations valides: 94; Gamma = 0.736, p-valeur = 0.000)"
Im Text wird bei einer rechtsseitigen Hypothese im obigen Beispiel geschrieben: „Die Hypothese, dass je höher der Ausbildungsgrad ist, desto kleiner ist die Zufriedenheit mit der Schule, wird bestätigt (n = 94; gültige Wert: 94; Gamma = 0.736, rechtsseitiger Test; p-Wert = 0.000)". „L'hypothèse que meilleure est la formation, moins grande est la satisfaction avec l'école est confirmée (n = 94; observations valides: 94; Gamma = 0.736, test unilatéral à droite; p-valeur = 0.000)".
Im Text wird bei einer linksseitigen Hypothese im obigen Beispiel geschrieben: „Die Hypothese, dass es je höher der Ausbildungrad ist, desto grösser ist die Zufriedenheit mit der Schule, wird nicht bestätigt (n = 94; gültige Wert: 94; rechtsseitiger Test; Gamma = 0.736)". „L'hypothèse que meilleure est la formation, plus grande est la satisfaction avec l'école est infirmée (n = 94; observations valides: 94; Gamma = 0.736, test unilatéral à gauche)". Man gibt hier keinen p-Wert an, da die Richtung des Zusammenhangs (positives Gamma) nicht mit der Hypothese übereinstimmt, gemäss derer sich ein negatives Gamma ergeben müsste.

- Bibliographische Angaben: Gamma wird in einem Artikel von Goodman und Kruskal (1954) vorgestellt. Es ist nicht nötig, eine Referenz für den Test anzugeben. Wenn Sie eine solche angeben wollen, würde ich Agresti (2002) empfehlen.

Fakultative Erläuterunge zu Gamma Gamma ist wie folgt definiert, wobei k_{ij} die Häufigkeit der $i-$ten Zeile und der j-ten Spalte der Kreuztabelle ist.

$$\gamma = \frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d}$$
$$\Pi_c = \sum_{i=1}^{I}\sum_{j=1}^{J} k_{ij}\left(\sum_{I\geq h>i}\ \sum_{J\geq k>j} k_{hk}\right)$$
$$\Pi_d = \sum_{i=1}^{I}\sum_{j=1}^{J} k_{ij}\left(\sum_{I\geq h>i}\ \sum_{1\leq k<j} k_{hk}\right)$$

Wir erklären die Berechnung am obigen Beispiel:

Wir berechnen zuerst die Anzahl der „konkordanten" Paare. Es handelt sich um die Anzahl von jeweils zwei Objekten a und b für die gilt: b hat einen höheren Ausbildungsgrad und ist weniger zufrieden mit der Schule als als a. In der Zelle a_{11} haben wir 14 Personen. Für jede dieser 14 Personen gibt es $(11 + 14 + 5 + 17 + 3 + 15)$ Personen, die sowohl eine höhere Ausbildung haben und weniger zufriedener sind. Derart gehen wir alle Zellen durch und bilden am Schluss die Summe aller konkordanten Paare. Dies ergibt im Beispiel:

Zufriedenheit mit Schule * Ausbildung Kreuztabelle

Anzahl

		Ausbildung			Gesamt
		Volksschule Volksschule	Berufslehre oder Sek II	Tertiäre Ausbildung	
Zufriedenheit mit Schule	sehr zufrieden	14	9	5	28
	zufrieden	1	11	14	26
	mässig zufrieden	0	5	17	22
	unzufrieden	0	3	15	18
Gesamt		15	28	51	94

Tabelle 37: Kreuztabelle für Gamma-Test. Gamma ist positiv, wenn sich die Häufigkeiten auf der Hauptdiagonale massieren (wie im Beispiel, d.h. von oben links nach unten rechts), negativ, wenn sie sich auf der Nebendiagonale massieren (d.h. von oben rechts nach unten links). (deutsche Beschriftung)

$\Pi_c =$
$14\,(11+14+5+17+3+15)+9\,(14+17+15)+1\,(5+17+3+15)+11\,(17+15)+5\,(15) = 1791$

Analog berechnen wir die „diskordanten“ Paare. Es handelt sich um Paare, die sich nicht konkordant verhalten, d.h. b hat eine höhere Ausbildung und ist zufriedener. In der Zelle a_{13} mit 5 Personen erhalten wir $5(1+11+0+5+0+3)$ Personen, die mehr verdienen und unzufriedener sind. Insgesamt:
$\Pi_d = 5(1+11+0+5+0+3) + 9\,(1) + 14\,(5+0+3+0) + 17\,(3) = 272$

Offenbar besteht ein „positiver“ Zusammenhang zwischen den beiden Variablen, wenn es mehr konkordante als diskordante Paare hat. Es besteht ein umgekehrter Zusammenhang zwischen den beiden Variablen, wenn es mehr diskordante Paare als konkordante Paare hat. Es besteht kein Zusammenhang, wenn es gleichviele konkordante wie diskordante Paare hat. Wir messen das Grössenverhältnis zwischen diskordanten und konkordanten Paaren, indem wir die Differenz dieser zwei Grössen an der Summe der beiden Grösse messen. Im Beispiel erhalten wir.

$$\gamma = \frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d} = \frac{1791-272}{1791+272} = 0.736\,31$$

Gamma variiert zwischen -1 und 1. Da $\Pi_c, \Pi_d \geq 0$, gilt nämlich: $|\Pi_c - \Pi_d| \leq \Pi_c + \Pi_d$. Zudem gilt: Wenn $\Pi_d = 0$, dann gilt: $\frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d} = \frac{\Pi_c}{\Pi_c} = 1$.Wenn $\Pi_c = 0$, dann gilt: $\frac{-\Pi_d}{\Pi_d} = -1$.

In ? (?) wird gezeigt, dass die folgende Grösse näherungsweise standardnormalverteilt ist (in SPSS wird etwas vorsichtiger mit einer t-Verteilung gearbeitet):

$\frac{\Gamma}{\frac{S}{\sqrt{n}}} \sim N(0,1)$ (asymptotisch; für n genügend gross)
S^2 ist die Zufallsvariable, die den Wert $s^2 = \frac{16}{(\pi_c+\pi_d)^4} \cdot \sum_{i=1}^{n} \sum_{j=1}^{m} h_{ij}(\pi_d h_{ij}^c - \pi_c h_{ij}^d)^2$ annimt:
$h_{ij} = \frac{k_{ij}}{n}$
k_{ij} = absolute Häufigkeit der Zelle i, j
n = Stichprobengrösse
$h_{ij}^c = \sum_{a<i} \sum_{b<j} h_{ab} + \sum_{a>i} \sum_{b>j} h_{ab}$
$h_{ij}^d = \sum_{a<i} \sum_{b>j} h_{ab} + \sum_{a>i} \sum_{b<j} h_{ab}$
$\pi_c = \sum_{i=1}^{n} \sum_{j=1}^{m} h_{ij} h_{ij}^c$
$\pi_d = \sum_{i=1}^{n} \sum_{j=1}^{m} h_{ij} h_{ij}^d$
Γ = Zufallsvariable, die konkrete γ als Werte annimmt.

Für den exakten p-Wert werden die kombinatorischen Möglichkeiten durchgerechnet, bei gegebenen Randverteilungen Kreuztabellen zu erhalten. Unter der Hypothese, dass jede Tabelle gleichwahrscheinlich ist, wird der Anteil der Tabellen bestimmt, die ein Gamma grösser-gleich das vorliegende Gamma haben (bei einem rechtsseitigen Test).

9 Spearman-Korrelation (corrélation de Spearman)

Wenn Sie in Ihrer Diplomarbeit einen Test bezüglich der Spearman-Korrelation verwenden, sollten Sie folgendes wissen:

- Die Spearman-Korrelation ist eine Rangkorrelation: Statt auf die Daten wird die Pearson-Korrelation (Produkte-Momenten-Korrelation) auf die Ränge berechnet. Die Spearman-Korrelation wird verwendet, wenn beide Variablen mindestens ordinalskaliert sind und wenn die Daten genügend differenziert sind (d.h. die Gruppen von Objekten mit identischer Ausprägung sind nicht zu gross und es hat nicht zu viele solcher Gruppen. Anders formuliert: es hat fast soviele Ausprägungen wie Daten. Gleiche Ausprägungen werden „Bindungen" genannt). Bei metrisch skalierten Daten wird die Spearman-Korrelation statt der Pearson-Korrelation verwendet, wenn die Bedingungen für die Verwendung der Pearson-Korrelation nicht gegeben sind (ungefähre Punktsymmetrie bezüglich dem Punkt $(\bar{x}, \bar{y})$ - $\bar{x}$ = Mittelwert der $x-$Daten; $\bar{y}$ = Mittelwert der $y-$Daten; Punktwolke der (x_i, y_i) liegt ungefähr auf einer Geraden. Es liegen keine Ausreisser vor).

- Wird ein Spearman-Korrelation-Test zugesandt, so werden zwei Tabellen geliefert, die Sie interpretieren können müssen, wobei die erste oft die Angaben für mehrere Tests enthält. Sie enthält die Angaben über die gültigen (valide) und die fehlenden (manquante) Fälle (observations). Fehlend ist ein Fall, falls bezüglich der Variable keine Information über die Ausprägung des Objektes vorliegt = fehlendes Datum.

- Zweitens wird eine Tabelle mit den Korrelationskoeffizienten (coéfficient de corrélation) und dem p-Wert (valeur p) zugesandt. Der Korrelationskoeffizient ist eine Zahl zwischen -1 und 1. Liegt der Korrelationskoeffizient in der Nähe von 0, so liegt keine Korrelation vor. Liegt der Wert in der Nähe von 1, liegt eine starke positive Korrelation vor (je mehr, desto mehr, z.B. „Je reicher jemand ist, desto mehr Geld für Konsumgüter gibt er aus"). Liegt der Wert in der Nähe von -1, liegt eine starke negative Korrelation vor (je mehr, desto weniger, z.B. „Je mehr jemand raucht, desto weniger lange lebt er"). Bei der Interpretation der Korrelation muss man allerdings beachten, wie die Variablen kodiert sind (was drückt eine tiefe, was drückt eine hohe Zahl aus?): eine positive Korrelation liegt vor, wenn in den Zahlenpaaren hohen Zahlen der Variable x im allgemeinen hohen Zahlen der Variable y entsprechen, eine negative Korrelation liegt vor, wenn in den Zahlenpaaren im Allgemeinen hohen Zahlen der Variable x tiefe Zahlen der Variable y entsprechen.

- Bei einem zweiseitigen Test drückt ein p-Wert von weniger als 0.05 aus, dass die Abweichung des Korrelationskoeffizienten von 0 schlecht durch Zufall erklärbar ist. Bei einem rechtsseitigen Test drückt ein p-Wert von weniger als 0.05 aus, dass die Korrelation schlecht durch Zufall erklärbar nach oben von 0 abweicht (also mehr als zufällig stark positiv ist, der Korrelationskoeffizient muss positiv sein!). Bei einem linksseitigen Test drückt ein p-Wert von weniger als 0.05 aus, dass die Korrelation schlecht durch Zufall ercläbar nach unten von 0 abweicht (also mehr als zufällig stark negativ ist. Der Korrelationskoeffizient muss negativ sein!). Im folgenden Beispiel wurde die Hypothese aufgestellt, dass Studierende mit mehr Selbstachtung weniger stressanfällig sind. Bei der Umfrage mussten die Studierenden sich selber auf einer Skala von 0 bis 9 einschätzen (Selbstachtungsniveau; je grösser die Zahl, desto mehr Selbstachtung) und die Stressanfälligkeit (je mehr Stressgefühle, desto höher die Zahl). Es ergaben sich die folgenden Resultate (s. Tabelle 38 und 39):

Korrelationen

			Estime de soi (De 0 = peu à 9 = beaucoup)	Taux de stress (De 0 = peu à 9 = beaucoup)
Spearman-Rho	Estime de soi (De 0 = peu à 9 = beaucoup)	Korrelationskoeffizient	1.000	-.379**
		Sig. (1-seitig)		.001
		N	62	62
	Taux de stress (De 0 = peu à) 9 = beaucoup	Korrelationskoeffizient	-.379**	1.000
		Sig. (1-seitig)		.001
		N	62	62

Tabelle 38: Hinter „Korrelationskoeffzient“ steht der Korrelationskoeffizient der ersten Variable mit sich selber (immer 1), dann der Korrelationskoeffizient mit der andere Variable. Die Sterne bedeuten, dass die Abweichung von 0 signifikant ist. Der p-Wert steht nach Sig. Es wurde ein einseitiger Test geliefert (1-seitig). N = Anzahl gültiger Daten.

Corrélations

			Estime de soi (De 0 = peu à 9 = beaucoup)	Taux de stress (De 0 = peu à 9 = beaucoup)
Rho de Spearman	Estime de soi (De 0 = peu à 9 = beaucoup)	Coefficient de corrélation	1.000	-.379**
		Sig. (unilatérale)		.001
		N	62	62
	Taux de stress (De 0 = peu à 9 = beaucoup)	Coefficient de corrélation	-.379**	1.000
		Sig. (unilatérale)		.001
		N	62	62

Tabelle 39: Hinter „coéfficient de corrélation“ steht der Korrelationskoeffizient der ersten Variable mit sich selber (immer 1), dann der Korrelationskoeffizient mit der andere Variable. Die Sterne bedeuten, dass die Abweichung von 0 signifikant ist. Der p-Wert steht nach Sig. Es wurde ein einseitiger Test geliefert (unilatéral). N = Anzahl gültiger Daten.

Im Beispiel wird die Hypothese, dass mehr Selbstachtung mit weniger Stressgefühlen einhergehen bestätigt (negativer Korrelationskoeffizient, p-Wert $0.001 < 0.05$). Ob man eine Korrelation von -0.379 inhaltlich als wichtig betrachten kann, ist allerdings eine andere Frage. Werden bei den Testergebnissen auch diejenigen eines exakten Tests angegeben, so sind diese zu verwenden.
Wird ein Bootstrap-Test geliefert (dies etwa, wenn es zuviele Bindungen hat oder andere Bedingungen des Tests nicht erfüllt sind), so steht z.B.

		variable 1	variable 2
variable 1	Coefficient de corrélation		.118
	Sig. (unilatéral)		.193
	N		56
Bootstrap	Biais		-.003
	Résidu standard		.135
Intervalle de confiance à 95 %	Inférieur		-.153
	Supérieur		.379

On indique le coefficient de corrélation (Coefficient de corrlation 0.118) et la valeur p du test normal (Sig. (unilatéral); 0.193). Sous Bootstrap il y a un supplément: Le Biais (Verzerrung) indique la différence entre le Coefficient de corrélation 0.118 et la moyenne de tous les coefficients de corrélation calculés pour les échantillons du bootstrap (p.ex.1000 échantillons). Intervalle de confiance à 95 % avec le bord inférieur et le bord supérieur de cet interval (95% Konfidenzintervall mit unterer Grenze und oberer Grenze des Intervalls, hier ünterer und oberer Wert"genannt): Avec une probabilité de 95 % le véritable coefficient de corrélation se trouve entre -0.153 et 0.379. Le coefficient de corrélation se distingue avec une assez grande probabilité de 0, si l'intervalle de confiance ne contient pas 0.

- Angabe der Ergebnisse in der Diplomarbeit: Im Text wird bei der obigen Hypothese z.B. geschrieben: „Die Hypothese, dass bei mehr Selbstachtung weniger Stressgefühle vorliegen wird bestätigt (n =62, gültige Werte = 62, Spearman-Rho = -0.379, einseitiger Test, p-Wert = 0.001)". „L'hypothèse que plus d'estime de soi est accompagné de moins de sentiments de stress est confirmée (n =62, observations valides = 62, Rho de Spearman = -0.379, test unilatéral, valeur p = 0.001)"
  Hätte die Hypothese gelautet: Selbstachtung ist negativ oder positiv mit Stressgefühlen korreliert, würde man mit den obigen Ergebnissen z.B. schreiben: „Die Hypothese, dass es eine positive oder negative Korrelation der Selbstachtung mit Stressgefühlen gibt, wird bestätigt (n = 62, gültige Werte n = 62; Spearman-Rho = -0.379, zweiseitiger Test, p-Wert = 0.002)". „L'hypothèse qu'il y a une corrélation positive ou négative entre l'estime de soi et les sentiments de stress est confirmée (n = 62, observations valides = 62; Rho de Spearman = -0.379, test bilatéral, valeur p = 0.002)"
  Hätte die Hypothese gelautet: Je mehr Selbstachtung desto mehr Stressgefühle, würde man mit den obigen Ergebnissen z.B. schreiben „Die Hypothese, dass mehr Selbstachtung mit mehr Stressgefühlen einhergeht, wird nicht bestätigt (n = 62, gültige Werte n = 62; Spearman-Rho = -0.379; einseitiger Test)". Da der Korrelationskoeffizent negativ ist und die Hypothese einen positiven behauptet, erübrigt sich die Angabe des p-Wertes. „L'hypothèse que plus d'estime de soi est accompagné de plus de sentiments de stress n'est pas confirmée (n = 62, observations valides = 62; Rho de Spearman = -0.379; test unilatéral)".
  Wird z.B. ein Bootstrap für einen rechtsseitigen Test geliefert , würde man (für obiges Beispiel) schreiben: Die Hypothese einer positiven Korrelation zwischen der Variable 1 und 2 wird nicht bestätigt (n=56, gültige Beobachtungen = 56, Rho von Spearman = 0.18. Bootstrap: Verzerrung: -0.003, 95%-Vertrauensintervall: [-1.53, 0.379]). L'hypothèse d'une corrélation positive entre la Variable 1 et 2 n'est pas confirmée (n=56, observations valides = 56, Rho de Spearman = 0.18. Bootstrap: biais: -0.003, intervalle de confiance à 0.95 %: [-1.53, 0.379]).

- Bibliographische Angaben: Spearmans-Rho wurde von Charles Edward Spearman (1863 - 1945, britischer Psychologe) entwickelt ((Spearman, 1904) und (Spearman, 1906)). Es ist nicht nötig, eine Referenz für den Test anzugeben. Wenn Sie eine solche angeben wollen, würde ich auf (Büning & Trenkler, 1994, S. 232 ff.) referieren.

Fakultative Erläuterungen zum Spearman-Korrelations-Test Die Pearson-Korrelation wird definiert als

$$r = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$$

mit den Standardabweichungen s_x der $x-$Daten und s_y der $y-$Daten. Bei der Spearman-Korrelation wird die Pearson-Korrelation auf die Ränge statt auf die Daten berechnet: dazu werden die Daten nach der $x-$Variable sortiert. Dann werden den Daten aufsteigend Ränge zugeordnet (dem kleinsten Datum den ersten Rang, etc.). Bei gleichen Ausprägungen werden Ränge so zugeordnet, dass sich die Rangsumme nicht verändert. Dies kann man dadurch erreichen, dass jeweils Rangmittel zugeordnet werden: haben wir drei Daten mit gleicher Ausprägung, die sich an 7. 8. und 9. Stelle befinden, ordnen wir den drei Daten den Rang 8 zu. Anschliessend wird nach der $y-$Variable geordnet und es werden analog Ränge zugeordnet. Am Schluss wird die Pearson-Korrelation auf die Ränge berechnet. Beim Testen wird die Hotelling-Pabst-Statistik

$$D = \sum_{i=1}^{n} \left(r\left(x_i\right) - r\left(y_i\right)\right)^2$$

verwendet - mit den Rängen $r\left(x_i\right)$ des Datums x_i und $r\left(y_i\right)$ des Datums y_i. Für kleine Datensätze ($n \leq 30$) werden mit Hilfe der Kombinatorik Tabellen berechnet, die die statistische Beurteilung von Abweichungen von 0 erlauben. Bei grösseren Datensätzen kann gezeigt werden, dass D näherungsweise normalverteilt ist und es gilt näherungsweise (beim Fehlen von Bindungen)

$$\frac{D - E\left(D\right)}{\sqrt{Var\left(D\right)}} = \frac{D - \frac{1}{6}(n^3 - n)}{\sqrt{\frac{(n-1)(n+1)^2 n^2}{36}}} \sim N\left(0, 1\right)$$

Bei Bindungen werden in der Literatur sogenannte Bindungskorrekturen offeriert.

10 Vorzeichentest für verbundene Stichproben (Test des signes)

Wenn Sie in Ihre Diplomarbeit einen Vorzeichentest für verbundene Stichproben verwenden, sollten Sie folgendes wissen:

- Verbundene Stichproben liegen vor, wenn man dieselben Untersuchungsobjekte zweimal misst (z.B. vor und nach einer Behandlung)
- Sie sollten wissen, dass der Vorzeichentest positive Differenzen mit negativen Differenzen pro Untersuchungsobjekt vergleicht. Bei einer zweiseitigen Hypothese testet man, ob es zwischen den beiden Messungen einen Unterschied gibt (z.B. „Durch die Anwendung von Gruppenarbeiten ergibt sich bei den Leistungen von Schülern ein Unterschied"). Bei einem rechtseitigen Test testen wir, ob es mehr positive als negative Differenzen gibt (z.B. „Durch die Anwendung von Gruppenarbeiten ergibt sich bei den Leistungen von Schülern eine Verbesserung"). Bei einem linksseitigen Test testen wir, ob es mehr negative als positive Differenzen gibt (z.B. „Durch die Anwendung von Gruppenarbeiten ergibt sich bei den Leistungen von Schüleren eine Verschlechterung"). Ob ein rechtsseitiger oder ein linkseitiger Test vorliegt, hängt damit auch von der Differenzbildung ab. Kehren wir die Differenzbildung um, wird ein linksseitiger Test zu einem rechtseitigen und umgekehrt. Im Beispiel: die obigen Seiten der Tests ergeben sich, wenn man „Mit Gruppenarbeiten" minus „Vor Gruppenarbeiten" rechnet. Die Seiten der Tests kehren sich um, wenn man „Vor Gruppenarbeiten" minus „Mit Gruppenarbeiten" rechnet.
- Wird ein Vorzeichen-Test zugesandt, so werden zwei Tabellen geliefert, die Sie interpretieren können müssen. Es wird eine Tabelle zugeschickt, welche die Anzahl der positiven und die Anzahl der negativen Differenzen angibt. Bei „Bindungen" (Ex aequo) wird angegeben, wieviele Differenzen 0 sind.

Häufigkeiten

		N
2. Stichprobe Anzahl Punkte - 1. Stichprobe Anzahl Punkte.	Negative Differenzen[a]	5
	Positive Differenzen[b]	15
	Bindungen[c]	6
	Gesamt	26

a. 2. Stichprobe Anzahl Punkte < 1. Stichprobe Anzahl Punkte.
b. 2. Stichprobe Anzahl Punkte > 1. Stichprobe Anzahl Punkte.
c. 2. Stichprobe Anzahl Punkte = 1. Stichprobe Anzahl Punkte.

Tabelle 40: Angaben zum verbundenen Vorzeichtest. Links steht, welche Variable von welcher Variable abgezählt wurde. Dann wird die Anzahl der positiven und die Anzahl der negativen Differenzen angeben. Die Anzahl der 0-Differenzen wird bei Bindungen angegeben. (deutsche Beschriftung)

Fréquences

		N
2. Stichprobe Anzahl Punkte - 1. Stichprobe Anzahl Punkte.	Différences négatives[a]	5
	Différences positives[b]	15
	Ex aequo[c]	6
	Total	26

a. 2. Stichprobe Anzahl Punkte < 1. Stichprobe Anzahl Punkte.
b. 2. Stichprobe Anzahl Punkte > 1. Stichprobe Anzahl Punkte.
c. 2. Stichprobe Anzahl Punkte = 1. Stichprobe Anzahl Punkte.

Tabelle 41: Angaben zum verbundenen Vorzeichtest. Links steht, welche Variable von welcher Variable abgezählt wurde. Dann wird die Anzahl der positiven und die Anzahl der negativen Differenzen angeben. Die Anzahl der 0-Differenzen wird bei Ex aequo angegeben. (französische Beschriftung)

- In der zweiten Tabelle wird angegeben, ob die Unterschiede zwischen negativen und positiven Differenzen gut oder schlecht durch Zufall erklärbar sind. Die Wahrscheinlichkeit, dass man zufälliger Weise die vorliegende oder eine extremere Abweichung der beiden erhält, ist der zweiseitige p-Wert (2-seitige exakte Signifikanz). Bei einem einseitigen Test ist der p-Wert die Hälfte davon, wobei zu überprüfen ist, ob die Tendenz in der ersten Tabelle (Häufigkeiten der positiven oder der negativen Abweichungen) mit der Hypothese übereinstimmt.

Statistik für Test[a]

	2. Stichprobe Anzahl Punkte - 1. Stichprobe Anzahl Punkte
Exakte Signifikanz (2-seitig)	0.041[b]

a. Vorzeichentest.
b. Verwendetete Binomialverteilung.

Tabelle 42: Testergebnisse des Vorzeichentests. Angabe des zweiseitigen p-Wertes. (deutsche Beschriftung)

Test[a]

	2. Stichprobe Anzahl Punkte - 1. Stichprobe Anzahl Punkte
Signification exacte (bilatérale)	0.041[b]

a. Test des signes.
b. VDistribution binomiale utilisée.

Tabelle 43: Testergebnisse des Vorzeichentests. Angabe des zweiseitigen p-Wertes. (französische Beschriftung)

- Angabe der Ergebnisse in der Diplomarbeit: Bei der Verwendung des Vorzeichentests sollte in der Diplomarbeit im Text bei einer zweiseitigen Hypothese im obigen Beispiel geschrieben werden: „Die Hypothese, dass es einen Leistungsunterschied bezüglich Verwendung von Gruppenarbeiten gibt, wird bestätigt (n = 27; gültige Werte: 26; Anzahl positiver Differenzen 15, negative Differenzen 5, Bindungen 6, 2. Stichprobe (mit Gruppenarbeiten) - 1. Stichprobe, verbundener zweiseitiger Vorzeichentest, p-Wert = 0.041)“. „L'hypothèse qu'il y a une différence par rapport à l'utilisation des groupes de travail, est confirmée (n = 27;

observations valides: 26; nombre des différences positives 15, des différences négatives 5, Ex aequo 6, 2. Stichprobe (mit Gruppenarbeiten) - 1. Stichprobe, test des signes bilatéral, valeur p = 0.041)“. „2. Stichprobe (mit Gruppenarbeiten)“ und „1. Stichprobe“ sind die Namen der verwendeten Variablen (wie von Ihnen zugeschickt).
Im Text wird bei einer rechtsseitigen Hypothese im obigen Beispiel geschrieben: „Die Hypothese, dass bei Gruppenarbeiten die Leistung besser ist, wird bestätigt (n = 27; gültige Werte 26; Anzahl positiver Differenzen 15, negative Differenzen 5, Bindungen 6, 2. Stichprobe (mit Gruppenarbeit) - 1. Stichprobe, rechtsseitiger verbundener Vorzeichentest, p-Wert = 0.02)“. „L'hypothèse que les groupes de travail améliorent les connaissances est confirmée (n = 27; observations valides: 26; nombre des différences positives 15, des différences négatives 5, Ex aequo 6, 2. Stichprobe (mit Gruppenarbeiten) - 1. Stichprobe, test des signes unilatéral à droite, valeur p = 0.02)“.
Im Text wird bei einer linksseitigen Hypothese im obigen Beispiel geschrieben: „Die Hypothese, dass bei Gruppenarbeiten die Leistung schlechter wird, wird nicht bestätigt (n = 27; gültige Werte 26; Anzahl positiver Differenzen 15, negative Differenzen 5, Bindungen 6, 2. Stichprobe - 1. Stichprobe, linksseitiger verbundener Vorzeichentest)“. „L'hypothèse que les groupes de travail dégradent les connaissances est infirmée (n = 27; observations valides: 26; nombre des différences positives 15, des différences négatives 5, Ex aequo 6, 2. Stichprobe (mit Grupenarbeiten) - 1. Stichprobe, test des signes unilatéral à gauche)“. Man gibt hier keinen p-Wert an, da die Richtung des Zusammenhangs (Anzahl negativer Differenzen im Verhältnis zur Anzahl der positiven Differenzen) nicht mit der Hypothese übereinstimmt, gemäss derer sich mehr negative als positive Differenzen ergeben müssten.

- Bibliographische Angaben: Der Vorzeichentest ist ein Klassiker. Es ist nicht nötig, eine Referenz für den Test anzugeben. Wenn Sie eine solche angeben wollen, würde ich Büning und Trenkler (1994, S. 167), empfehlen.

Fakultative Erläuterungen zum Vorzeichentest Der Vorzeichentest beruht auf der Binomialverteilung. Man zählt die Anzahl der interessierenden Differenzen und geht davon aus, dass beim Fehlen von Unterschieden zwischen den beiden Stichproben die Wahrscheinlichkeit, dass eine Differenz positiv ist, 0.5 beträgt. Entsprechend ist die Wahrscheinlichkeit einer negativen Differenz ebenfalls 0.5. Bindungen (d.h. 0-Differenzen) lässt man dabei weg. Es gilt für den rechtseitigen p-Wert entsprechend:

$$p - Wert = P(X \geq x) = \sum_{i=x}^{n} \binom{n}{i} 0.5^i 0.5^{n-i}$$

wobei n die Stichprobengrösse ist (Anzahl gültiger $-$ Anzahl Bindungen) und x die Anzahl der positiven Differenzen. Für den linksseitgen p-Wert rechnet man

$$p - Wert = P(X \leq x) = \sum_{i=0}^{x} \binom{n}{i} 0.5^i 0.5^{n-i},$$

und für den zweiseitigen Test

$$2 \min\{P(X \geq x), P(X \leq x)\}.$$

11 Binomialtest (Test binomial)

Wenn Sie in Ihrer Diplomarbeit einen Binomialtest verwenden, sollten Sie folgendes wissen:

- Es wird getestet, ob die Verteilung der Beobachtungen auf die zwei Ausprägungen gleichmässig ist oder nicht (zweiseitiger Test) oder ob in einer Kategorie mehr Beobachtungen vorkommen als in der anderen (einseitiger Test). So möchte man z.B. wissen, ob gleiche viele Leute zufrieden sind wie unzufrieden (zweiseitiger Test) oder ob es mehr zufriedene als unzufriedene Personen gibt (oder dann andere einseitige Hypothese: ob es mehr unzufriedene als zufriedene Personen gibt).

- Wird ein Binomialtest zugesandt, so wird eine Tabellen geliefert, die Sie interpretieren können müssen (s. unten). Sie enthält die Angaben über die gültigen (valide) und die fehlenden (manquante) Fälle (observations), Fehlend ist ein Fall, falls bezüglich der Variable keine Information über die Ausprägung des Objektes vorliegt = fehlendes Datum. Unter „N“ wird die Häufigkeitsverteilung auf die beiden Ausprägungen (Gruppen), unter „Beobachteter Anteil“ die entsprechende relative Häufigkeitsverteilung geliefert. Unter Testanteil wird angeben, wie die Verteilung unter der Nullhypothese aussieht (bei 0.5 also halb-halb).

- Zuletzt wird ein zweiseitiger p-Wert geliefert (unter „Exakte Signifikanz (2-seitig)“). Dieser kann bei einseitigen Hypothesen halbiert werden, sofern die Richtung der Verteilung mit der Hypothese übereinstimmt. Bei einem zweiseitigen Test drückt ein p-Wert von weniger als 0.05 aus, dass die Abweichung der Verteilung von einer Gleichverteilung auf beide Gruppen schlecht durch Zufall erklärbar ist. Bei einer Hypothese, dass eine der Gruppen überrepräsentiert ist, drückt der halbierte p-Wert, wenn er kleiner als 0.05 ist aus, dass die Überrepräsentation der Gruppe schlecht durch Zufall erklärbar ist (sofern die relativen oder absoluten Häufigkeiten in diese Richtung zeigen!). Im beiliegenden Beispiel würde die zweiseitige Hypothese lauten: „Es hat bei den positiv und den negativ eingestellten Personen nicht gleich viele Personen“. Die Hypothse wird bestätigt, da der p-Wert kleiner als 0.05 ist. Die einseitige Hypothese „Es hat mehr positiv als negativ eingestellt Personen“ wird ebenfalls bestätigt, da es erstens mehr positive als negativ eingestellte Personen hat und zweitens der p-Wert $0.009/2 = 0.0045 < 0.05$.
  Die einseitige Hypothese „Es hat mehr negativ als positiv eingestellte Personen“ wird verworfen, da es mehr positiv als negativ eingestellt Personen hat. (s. Tabelle 44 und 45):

Test auf Binomialverteilung

		Kategorie	N	Beobachteter Anteil	Testanteil	Exakte Signifikanz (2-seitig)
Gesamtbeurteilung	Gruppe 1	positiv	25	0.74	0.5	0.009
dichotom	Gruppe 2	negativ	9	0.26	0	0
	Gesamt		34	1	0	0

Tabelle 44: Tabelle zum Einstichproben-Binomialtest

		Test binomial				
		Modalité	N	Proportion observée	Test de proportion	Signification exacte (bilatérale)
Gesamtbeurteilung	Groupe 1	positiv	25	0.74	0.5	0.009
dichotom	Groupe 2	negativ	9	0.26	0	0
	Total		34	1	0	0

Tabelle 45: Tableau pour le test binomial univarié

- Angabe der Ergebnisse in der Diplomarbeit: Im Text wird bei der zweiseitigen Hypothese z.B. geschrieben: „Die Hypothese, dass es Unterschiede in der Einstellung gibt, wird bestätigt (n =34, gültige Werte = 34, beobachteter Anteil 0.74, zweiseitiger Test, p-Wert = 0.009)“. „L'hypothèse qu'il y a des différences d'attitude est confirmée (n =34, observations valides = 34, proportion observée = 0.74, test bilatéral, valeur p = 0.009)“
  Für die Hypothese „Es hat mehr Personen mit positiver Einstellung“ würde man schreiben: „Die Hypthese wird bestätigt (n = 34, gültige Werte = 43; beobachteter Anteil 0.74, einseitiger Test, p-Wert = 0.005)“. „L'hypothèse qu'il y a plus de personne avec une attitude positive est confirmée (n = 34, observations valides = 34; proportion observée = 0.74, test unilatéral, valeur p = 0.005)“
  Für die Hypothese: Es hat mehr Personen mit negativer Einstellung, würde man mit den obigen Ergebnissen z.B. schreiben „Die Hypothese, dass es mehr Personen mit negativer Einstellung hat, wird nicht bestätigt (n = 34, gültige Werte 34; beobachteter Anteil 0.26; einseitiger Test)“. Da der beobachtete Anteil kleiner als 0.5 ist und die Hypothese einen Anteil behauptet, der grösser ist als 0.5, erübrigt sich die Angabe des p-Wertes. „L'hypothèse qu'il y a plus de personnes avec une attitude négative n'est pas confirmée (n = 34, observations valides = 34; proportion observée = 0.26; test unilatéral)“.
- Bibliographische Angaben: der Binomialtest ist ein Klassiker, den man kaum auf eine klare Quelle zurückführen kann. Wesentliche Beiträge zu Binomialverteilung erfolgten durch Jakob Bernoulli (Jakob I. Bernoulli, 1654 - 1705, Schweizer Mathematiker und Physiker). Es ist nicht nötig, eine Referenz für den Test anzugeben. Wenn Sie eine solche angeben wollen, würde ich auf (Büning & Trenkler, 1994, S. 232 ff.) referieren.

Fakultative Erläuterungen zum Binomialtest Der Binomialtest beruht auf der Binomialverteilung: Wird auf Gleichverteilung in zwei Kategorien getestet, so wird $p = 0.5$ gesetzt. Ist x die absolute Häufigkeit der untersuchten Klasse, so ist für den linksseitigen Test der p-Wert definiert durch:

$$P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k \left(1-p\right)^{n-k}$$

für den rechtseitigen Test durch:

$$P(X \geq x) = \sum_{k=x}^{n} \binom{n}{k} p^k \left(1-p\right)^{n-k}$$

Im zweiseitigen Fall durch $2 \cdot \min \left\{P(X \leq x), P(X \geq x)\right\}$

12 Regression logistique ordinale (ordinale logistische Regression)

On calcule un modèle, qui exprime une variable dépendante ordinale par d'autres variables. Exemple: on voudrait savoir s'il y a une relation entre la satisfaction avec une prestation (très content, content, peu content, mécontent) et l'âge (en années), le genre et l'état civil (célibataire, marié, divorcé/veuf). On obtiendrait les résultats suivant:

Récapitulatif de traitement des observations

		N	Pourcentage marginal
satisfaction	mécontent	150	25.0%
	peu content	150	25.0%
	content	150	25.0%
	très content	150	25.0%
etat_civil	célibataire	163	27.2%
	marié	288	48.0%
	divorcé ou veuf	149	24.8%
genre	féminin	296	49.3%
	masculin	304	50.7%
Valide		600	100.0%
Manquant		0	
Total		600	

Ce premier tableau présente les fréquences univariées des variables intégrées dans le modèle.

Informations sur l'ajustement du modèle

Modèle	Log de vraisemblance -2	Khi-deux	ddl	Sig.
Constante uniquement	1663.553			
Final	611.641	1051.912	4	.000

Fonction de lien : Logit.

Ce tableau compare le modèle sans variables au modèle avec les variables „âge“, „genre“ et „état civil“. La „Log de vraisemblance - 2“ devrait se réduire - par l'insertion des variables dans le modèle - d'une manière significative. En l'occurence la réduction est importante (Sig. 0.000 < 0.05).

Qualité d'ajustement

	Khi-deux	ddl	Sig.
Pearson	1072.719	1793	1.000
Déviance	611.641	1793	1.000

Fonction de lien : Logit.

On compare les prédictions du modèle avec les données. Les deux ne devraient pas se distinguer trop: la signification doit être suppérieure à 0.05. Dans l'exemple, le modèle s'adapte très bien aux données ($1 > 0.05$). On pourrait (et devrait en général) faire des analyses plus poussées par rapport à l'adaptation du modèle - ce qui nous mènerait trop loins dans ce cadre.

Le logiciel SPSS offre de plus quelques Pseudo R-deux (dans la littérature on en offre une pluralité). Ces chiffres entre 0 et 1 sont en général beaucoup plus bas que dans l'exemple - même si le modèle s'adapte bien. Les personnes qui connaissent le R^2 de la régression linéaire classique doivent se rendre compte, qu'on ne peut pas comparer sans autre le R^2 et les différents Pseudo R-deux. Les Pseudo R-deux oscillent souvent autour de 0.1. Un pareil nombre pour le R^2 de la regression traditionnelle serait tout à fait insuffisant. Il n'est pas nécessaire de tenir compte de ces Pseudo R-deux.

Estimations des paramètres

							Intervalle de confiance à 95 %	
		Estimation	Erreur standard	Wald	ddl	Sig.	Borne inférieure	Borne supérieure
Seuil	[satisfaction = 1]	20.404	1.287	251.482	1	.000	17.882	22.926
	[satisfaction = 2]	24.755	1.507	269.895	1	.000	21.802	27.709
	[satisfaction = 3]	29.064	1.729	282.697	1	.000	25.676	32.452
Emplacement	age	.591	.035	280.120	1	.000	.522	.660
	[etat_civil=0]	-1.991	.316	39.689	1	.000	-2.611	-1.372
	[etat_civil=1]	-.872	.262	11.094	1	.001	-1.385	-.359
	[etat_civil=2]	0[a]	.	.	0	.	.	.
	[genre=0]	-1.589	.229	47.930	1	.000	-2.039	-1.139
	[genre=1]	0[a]	.	.	0	.	.	.

Fonction de lien : Logit.

a. Ce paramètre est défini sur 0, car il est redondant.

Seuil (dt: Schwelle, en: threshold): il faut contrôler, si les seuils augmentent ou descendent d'une manière monotone. Ici ils augmentent.

Emplacement (dt: Lâge; en:Location): $\exp(Estimation) = e^{Estimation}$ indique p.ex. pour l'état civil le rapport des cotes (en: odds ratio; dt: Quotenverhältnis)

$$\frac{\frac{satisfaction=i+1}{satisfaction=i}}{\frac{etat\_civil=j}{etat\_civil=2}}$$

Remark 3. *Der Begriff „Quote" (fr: cote; en: odds) entstammt dem Wettgeschäft. Es handelt sich um ein Verhältnis von Wahrscheinlichkeiten, nämlich der Wahrscheinlichkeit zu gewinnen geteilt durch dieWahrscheinlichkeit zu verlieren. Ein Quotenverhältnis ist das Verhältnis zweier Quoten: man bildet z.B. das Verhältnis der Quoten von rauchenden zu nichtrauchenden Frauen* *und von rauchenden zu nichtrauchenden Männern. Rauchen z.B. 10 Frauen und 15 nicht und 12 Männer und 10 nicht, ergibt sich das Quotenverhältnis* $\frac{10/15}{12/10} = 0.555\,56$*. Dasselbe Quotenverhältnis ergibt sich, wenn man mit relativen Häufigkeiten (= geschätzte Wahrscheinlichkeiten) arbeitet:* $\frac{(10/35)/(15/35)}{(12/22)/(10/22)} = 0.555\,56$*).* $\frac{1}{0.555\,56} = 1.800$*. Die Raucherquote der Frauen ist im Beispiel fast zweimal kleiner als die Raucherquote der Männer).*

Pour l'exemple en question (état civil et satisfation):

$$\exp(-0.872) = 0.418\,11$$

$\frac{1}{0.418\,11} = 2.391\,7$. La cote d'être dans la prochaine catégorie plus élévée de satisfaction (= d'être d'un degré plus content) est d'un peu plus que le double plus bas pour une personne mariée que pour une personne divorcée/veuve (selon le modèle). sig.=0.001 < 0.05 dit, que l'écart du rapport des cotes 0.41811 de 1 s'explique mal par le hasard.

$$\exp(-1.991) = 0.136\,56$$

$\frac{1}{0.136\,56} = 7.322\,8$. La cote d'être d'un degré plus content est d'un peu plus que sept fois plus bas pour une personne célibataire que pour une personne divorcée/veuve.

$$\exp(0.872 - 1.991) = 0.326\,61$$

$\frac{1}{0.326\,61} = 3.061\,8$ La cote d'être d'un degré plus content est trois fois plus bas pour une personne célibataire que pour une personne marié.

$$\exp(-1.589) = 0.204\,13$$

La cote d'être d'un degré plus content est de 5 fois plus bas pour une femme que pour un homme.

$$\exp(0.591) = 1.805\,8$$

La cote d'être d'un degré plus content est presque le double pour une personne plus âgée d'une année.

Souvent on ne s'intéresse pas aux rapports des cotes mais uniquement à la tendance de l'influence. Sans calculer on peut alors immédiatement conclure des résultats du tableau ci-dessus que les femmes sont moins contentes que les hommes, que les personnes plus âgées sont plus contentes et que les personnes divorcées/veuves sont les plus contentes et que les mariées sont plus contentes que les célibataires (données fictives).

- Si l'hypothèse est dirigée (p.ex. on a l'hypothèse que l'âge a une influence positive sur la satisfaction, et si le coéfficient va dans le sens de l'hypothèse on peut diviser par deux le chiffre indiqué sous „Sig." Dans l'exemple cette hypthèse est confirmée). Les chiffres indiqués sous „Sig." correspondent à des hypothèses bilatérales (il y a influence, mais on n'a pas d'hypothèse sur la direction de cette influence).

- Citer les résultats dans le travail de diplôme: dans le texte on écrit pour une hypothèse bilatérale p.ex.: „L'hypotèse que l'âge a une influence sur la satisfaction est confirmée (n = 600, rapport des cotes 1.8058 par rapport à une année supplémentaire, valeur p = 0.000)". „Die Hypothese, dass das Alter einen Einfluss auf die Zufriedenheit hat, wird bestätigt (n =600, Quotenverhältnis bezüglich eines zusätzlichen Jahres 1.8058, p-Wert = 0.000)"
  Pour l'hypothèse que l'âge a une influence positive sur la satisfaction on écrirait: „L'hypothèse que l'âge a une influence positive sur la satisfaction est confirmée (n = 600, rapport des cotes 1.8058 par rapport à une année supplémentaire, valeur p = 0.000)" (dans l'exemple $\frac{0.000}{2} = 0.000$).
  Für die Hypothese, dass das Alter einen positiven Einfluss auf die Zufriedenheit hat, würde man schreiben: „Die Hypothese, dass das Alter einen positiven Einfluss auf die Zufriedenheit hat, wird bestätigt (n =600, Quotenverhältnis bezüglich eines zusätzlichen Jahres 1.8058,

p-Wert = 0.000)“ (im Beispiel ist $\frac{0.000}{2} = 0.000$).
Pour l'hypothèse que l'âge a une influence negative sur la satisfaction on écrirait: „L'hypotèse que l'âge a une influence négative sur la satisfaction est infirmée (n = 600, rapport des cotes 1. 8058 par rapport à une année supplémentaire, valeur p = 1)“ (1-0=1)
Für die Hypothese, dass das Alter einen negativen Einfluss auf die Zufriedenheit hat, würde man schreiben: „Die Hypothese, dass das Alter einen negativen Einfluss auf die Zufriedenheit hat, wird verworfen (n =600, Quotenverhältnis bezüglich eines zusätzlichen Jahres 1.8058, p-Wert = 1)“ (1-0=1).

- Bibliographische Angaben: die logistische Regression gehört zum Grundstock moderner Statistik. Es ist nicht nötig, eine Referenz für das Verfahren anzugeben. Wenn Sie eine solche angeben wollen, würde ich auf (Hosmer & Lemeshow, 2000) referieren. Das Buch stellt auch einen relativ einfachen Einstieg in die Thematik dar.

Fakultative Erläuterungen zur ordinalen logistischen Regression Zuerst ein paar Bemekrungen zur logistische, zweiwertigen Regression, von der die ordinale logistische Regression eine Verallgemeinerung ist. Die logistische Regression ist ein sogenanntes Verallgemeinertes Lineares Modell, das im diesem Fall wie folgt spezifiziert ist (der rechte Teil der Gleichung ist ein linearer (genauer: affiner) Ausdruck):

$$\ln\left(\frac{P\left(Y=1|\mathbf{x}\right)}{P\left(Y=0|\mathbf{x}\right)}\right)=\sum_{i=0}^{p}\beta_i x_i$$

mit $x_0 = 1$, den erklärenden Variablen $x_1, ..., x_p$, $\mathbf{x} = (1, x_1, ..., x_p)$ und der Zielvariable Y (zu erklärende Variable). $P(Y = 1|\mathbf{x})$ ist die Wahrscheinlichkeit, dass die Zufallsvariable Y den Wert 1 annimmt, unter der Bedingung de Vorliegens eines Vektors $\mathbf{x}$ von erklärenden Variablen. Einfaches Umrechnen ergibt:

$$P\left(Y=1|\mathbf{x}\right)=\frac{\exp\left(\sum\limits_{i=0}^{p}\beta_i x_i\right)}{1+\exp\left(\sum\limits_{i=0}^{p}\beta_i x_i\right)}$$

Zudem sieht man unmittelbar, dass

$$\begin{aligned}&\beta_0+\beta_1 x_1+...+\beta_k\left(x_k+1\right)+...+\beta_p x_p-\left(\beta_0+\beta_1 x_1+...+\beta_k x_k+...+\beta_p x_p\right)\\&=\beta_k\left(x_k+1\right)-\beta_k x_k=\beta_k\end{aligned}$$

Mit der ersten Gleichung gilt mit $\mathbf{x}_{k+1} := (1, x_1, ...x_k + 1, ..., x_p)$ und $\sum\limits_{i=0}^{p}\beta_i x_{i_{k+1}} := \beta_0 + \beta_1 x_1 + ... + \beta_k (x_k + 1) + ... + \beta_p x_p$

$$\begin{aligned}\ln\left(\frac{P\left(Y=1|\mathbf{x}_{k+1}\right)}{P\left(Y=0|\mathbf{x}_{k+1}\right)}\right)-\ln\left(\frac{P\left(Y=1|\mathbf{x}\right)}{P\left(Y=0|\mathbf{x}\right)}\right)&=\ln\left(\frac{\frac{P(Y=1|\mathbf{x}_{k+1})}{P(Y=0|\mathbf{x}_{k+1})}}{\frac{P(Y=1|\mathbf{x})}{P(Y=0|\mathbf{x})}}\right)=\sum_{i=0}^{p}\beta_i x_{i_{k+1}}-\sum_{i=0}^{p}\beta_i x_i\\&=\beta_k\end{aligned}$$

Entsprechend gilt:

$$\exp\left(\beta_k\right)=\frac{\frac{P(Y=1|\mathbf{x}_{k+1})}{P(Y=0|\mathbf{x}_{k+1})}}{\frac{P(Y=1|\mathbf{x})}{P(Y=0|\mathbf{x})}}$$

Der exponentierte Koeffizient drückt also ein Quotenverhältnis aus.

Die ordinale logistische Regression stellt eine Verallgemeinerung der logistischen Regression dar. Es gibt verschiedene Modelle. In SPSS wird die Methode der kontinuierlichen Quoten (englisch: continuation-ratio model) berechnet ($k = 0, ..., K - 1$) mit der Anzahl K der Werte der

Variable Y. Bei den erklärenden Variablen wird jeweils die mit der grössten Zahl kodierte Ausprägung als Referenz gewählt.

$$\ln\left(\frac{P\left(Y > k\right)}{P\left(Y \leq k\right)}\right) = \delta_k + \sum_{i=1}^{p} \beta_i x_i$$

δ_k sind die „Schwellen" (*Seuil)*. Die exponentierten Koeffizienten drücken aus, um wieviel die Quote, in einer höheren Kategorie der Zielvariable zu sein, grösser oder kleiner ist, wenn man in der entsprechenden erklärenden Variable um eine Einheit absteigt und die übrigen Variablen konstant lässt (kontrolliert).

Literatur

Agresti, A. (2002). *Categorical Data Analysis* (2. Aufl.). New York: John Wiley.

Büning, H. & Trenkler, G. (1994). *Nichtparametrische statistische Methoden.* Berlin: Walter de Gruyter.

Fisher, R. A. (1922). On the interpretation of chi-square from contingency tables, and the calculation of P. *J. Roy. Statist. Doc.* (85), 442-450.

Fisher, R. A. (1966). *The Design of Experiments* (8. Aufl.). Edinburgh: Oliver and Boyd.

Goodman, L. A. & Kruskal, W. H. (1954). Measures of Association for Cross Classifications. *Journal of the American Statistical Association* (49), 732–764.

Hettmansperger, T. P. (1984). *Statistical inference based on ranks.* New York: Wiley.

Hosmer, D. W. & Lemeshow, S. (2000). *Applied Logistic Regression.* New York: Wiley.

Jonckheere, A. R. (1954a). A distribution-free k-sample test against ordered alternatives. *Biometrica*, *41*, 133-145.

Jonckheere, A. R. (1954b). A test of significance for the relation between m rankings and k ranked categories. *British Journal of Statistical Psychology*, *7*, 93-100.

Kruskal, W. H. & Wallis, W. (1952). Use of ranks in one criterion variance analysis. *J. Am. Stat. Assoc.* (57), 583-621.

Mann, H. & Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.*, *18*, 50-60.

Pearson, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably suposed to have arisen from random sampling. *Philos. Mag. Ser.*, *50* (5), 157-175.

Spearman, C. (1904). The proof and measurment of association between two things. *American Journal of Psychology*, *15*, 72-101.

Spearman, C. (1906). A footnote for measuring correlation. *British Journal of Psychology*, *2*, 89-108.

Terpstra, T. J. (o. J.). The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking], journal = Indagationes Mathematicae, year = 1952, pages = 327-333.