7. AUFGABEN MIT LÖSUNGEN

In diesem Kapitel werden Übungsaufgaben (mit Lösungen) dargestellt. Die Übungsaufgaben sind unterteilt in „Statistik Einführung“ und „Statistik Vertiefung“.

7.1 AUFGABEN ZU „STATISTIK EINFÜHRUNG“

Aufgabe 1: (Wahrscheinlichkeit)

Wir haben eine Trommel mit insgesamt 200 Bällen. Die Bälle können unterschiedliche Farben haben, sowie zusätzlich noch einen aufgedruckten Stern. Von jeder Kategorie gibt es soundsoviele Bälle (siehe Tabelle)

Bälle	rot	blau	grün	
mit Stern	60	40	30	130
ohne Stern	20	40	10	70
	80	80	40	200

Wie groß ist die Wahrscheinlichkeit,

a) einen blauen Ball zu ziehen?

b) einen roten oder einen grünen Ball zu ziehen?

c) einen grünen Ball mit Stern zu ziehen?

d) einen grünen und einen blauen Ball zu ziehen?

e) zuerst einen grünen und dann einen blauen Ball zu ziehen?

f) entweder einen grünen Ball ohne Stern oder einen blauen Ball mit Stern zu ziehen?

g) entweder einen roten Ball oder einen Ball mit Stern zu ziehen?

h) daß ein blauer Ball einen Stern hat?

i) daß ein Ball ohne Stern rot ist?

Aufgabe 2: (Wahrscheinlichkeit)

Wir befinden uns an der Universität in XYZ. Über diese Universität wissen wir einiges:

1. Es gibt hier gleich viele Studentinnen wie Studenten.

2. Im Fachbereich Psychologie gibt es insgesamt 330 Studentinnen und 170 Studenten.

3. Von den Studentinnen des Fachbereichs Psychologie haben 40% gute statistische Kenntnisse.

4. Von den Studenten des Fachbereichs Psychologie haben 35% gute statistische Kenntnisse.

5. An der Universität studieren insgesamt 5000 Studenten.

a) Wie groß ist die Wahrscheinlichkeit dafür, daß, wenn man sich zufällig eine Person auf dem Campus greift, diese Person weiblich ist?

b) Wie groß ist die Wahrscheinlichkeit dafür, daß diese Person weiblich ist und Psychologie studiert?

c) Wie groß ist die Wahrscheinlichkeit dafür, daß diese Person weiblich ist, Psychologie studiert, aber nur über geringe statistische Kenntnisse verfügt?

Aufgabe 3: (Wahrscheinlichkeit)

Jetzt schauen wir uns die Uni Frankfurt an. Was wissen wir alles?

1. Frauen und Männer sind insgesamt gleichverteilt.

2. Im Fachbereich Psycholgie ist das Verhältnis 0,33(mask.)/0,67(fem.).

3. 70% aller PsychologiestudentInnen haben gute statistische Kenntnisse.

4. 5% aller Studierenden haben sich im Fachbereich Psychologie eingeschrieben.

a) Wie groß ist die Wahrscheinlichkeit dafür, zufällig auf dem Campus einen Psychologiestudenten zu treffen, der auch über gute statistische Kenntnisse verfügt?

b) Nehmen wir einmal an, es gäbe insgesamt 20.000 StudentInnen an der Uni Frankfurt. Wieviele davon sind weiblich, studieren Psychologie und haben gute statistische Kenntnisse?

(Werte siehe Frage 1)

c) Wenn wir wissen, daß insgesamt an der Uni gleichviele Frauen wie Männer eingeschrieben sind, wieviele Frauen hätten wir im Fachbereich Psychologie eigentlich erwartet?

Aufgabe 4: (U-Test, Chi-Quadrat, t-Test, PM-Korr.)

Wir befinden uns in der Bronzezeit, viele, viele Jahre früher. Zwei benachbarte Amazonenstämme, die Penthes und die Sileas, wollen sich nicht mehr jedesmal gegenseitig die Schädel einschlagen, wenn es um die Vorherrschaft um die einzige Wasserstelle geht. Diesmal wollen sie die Frage durch einen sportlichen Wettkampf klären.

a) Als erstes steht Speerweitwurf auf dem Programm. Je fünf Amazonen treten gegeneinander an. Am weitesten wirft eine Penthe. Am zweit- und drittweitesten je eine Silea. An vierter Stelle liegt wieder eine Penthe. Den fünften, sechsten und siebten Platz belegen Frauen vom Stamm der Sileas. Die restlichen Plätze werden von Penthen eingenommen.

Tja, welcher Stamm war nun besser? Der weise Schamane Kaibus behauptete, daß die Rangplätze gleichverteilt seien. Würden Sie dem zustimmen?

b) Als nächster Wettkampf stand Steineschleudern auf dem Programm. Je zwanzig Werferinnen pro Stamm traten an, die aus fünfzig Schritt Entfernung ihre Steine auf ein Ziel schleuderten. Hier die Ergebnisse:

	Penthes	Sileas	
Treffer	12	9	21
kein Treffer	8	11	19
	20	20	40 Würfe insgesamt

Unser weiser und freundlicher Schamane behauptete wieder, es gäbe keinen signifikanten Unterschied. Wie sehen Sie das?

c) Das nächste war kein eigentlicher Wettkampf. Es ging lediglich darum, daß die Penthes behaupteten, im Schnitt mehr Männer zu besitzen; nämlich 4 Stück, bei einer Standardabweichung von 0,75. Die Sileas hatten im Schnitt 2,5 Männer, bei einer Standardabweichung von 0,5. An diesem Vergleich nahmen 30 Penthes und 30 Sileas teil. Würden Sie sagen, daß die Penthes wirklich mehr Männer im Schnitt hatten?

d) Nach dem letzten Vergleich schlugen die Wogen hoch.

Die Sileas hatten zwar weniger Männer, behaupteten jedoch, mit ihren Männern glücklicher zu sein, als es die Penthes mit ihren Männern jemals sein würden.

Unser weiser, kluger, intelligenter, guter, netter und freundlicher Schamane Kaibus betrachtete sich die jeweiligen Männer. Das wirklich auffallende an den sileischen Männern waren ihre großen Nasen. Flugs erstellte Kaibus einen Fragebogen, der das Glücklichsein der sileischen Frauen erfaßte, und er maß noch die Nasengröße der anwesenden sileischen Männer.

Hier die Werte:

Nasengröße	10	7	8	10	6	13	11	9
Glücklichsein	27	18	23	22	14	28	26	22

Kaibus behauptete, daß 1. beide Merkmale normalverteilt seien und daß 2. die Nasengröße der Männer wohl mit dem Glücklichsein der Frauen in Beziehung stehe. Was meinen Sie dazu?

Anmerkung: Der Mann mit der Nasengröße 13 soll angeblich Johannes geheißen haben.

Als Kaibus diese Vermutung geäußert hatte und die Penthen mit Entsetzen feststellen mußten, daß ihre Männer alle über kleine Nasen verfügten, wurden sie dermaßen zornig, daß sie Kaibus mit einer seltsam geformten Keule erschlugen. Alten Überlieferungen zufolge soll die Keule ungefähr so ausgesehen haben:

$$\Sigma$$

Aufgabe 5: (Wahrscheinlichkeit, t-Test)

Irgendwann nach der Bronzezeit kam die Antike. Richtig interessant war es hauptsächlich in Griechenland. Wen gab es da nicht alles: Aristoteles, Plato, Pythagoras, Demosthenes, Sophokles und und und. Einer, den die Geschichtsschreibung völlig übersehen hat, war ein kleiner, eher dicklicher Statistiker und Philosoph namens Kaios, der -rein zufällig- ein Nachfahre des oben erwähnten Schamanen war, was sich unter anderem darin äußerte, daß er manchmal von wilden, kreischenden Amazonen träumte, die Keulenschwingend hinter ihm herliefen. Als er eines Tages in seiner Toga über den Marktplatz von Athen schlenderte, stellte er sich folgende Fragen:

a) Ich könnte mir einen dieser Allobroger kaufen, der mir meinen Acker umgräbt. Leider sind von 10 gekauften Allobrogern höchstens 3 zu gebrauchen. Wenn ich mir jetzt 6 Allobroger kaufe, wie groß ist denn dann die Wahrscheinlichkeit dafür, wenigstens 1 brauchbaren dabei zu haben?

b) Andererseits, wenn ich mir jetzt sechs Allobroger hole, habe ich nicht mehr genug Geld, noch eine Tänzerin zu besorgen. Wenn ich mir dagegen nur vier Allobroger kaufen würde, wie groß ist denn dann die Wahrscheinlichkeit dafür, mindestens 1 guten dabei zu haben?

c) In Sparta haben sie angeblich im Schnitt vier Esel (bei einer Standardabweichung von S=1.5) pro Haushalt. Hier haben die Leute im Schnitt höchstens drei Esel (bei einer Standardabweichung von S=1.2). Angeblich wurde die Befragung bei 50 Spartanern und bei 42 Athenern durchgeführt. Ich frage mich, ob die Spartaner wirklich mehr Esel haben?

Aufgabe 6: (Chi-Quadrat)

Die international renommierte Keksfirma Palzen hat den national renommierten Psychologen Kabudis um Rat gefragt. Palzen verkauft in der BuReDeu drei verschiedene Kekssorten in je drei verschiedenen Geschmacksrichtungen. Die Prozentzahlen sind wie folgt:

KEKSE:

	Form			
Geschmack	rund	quadratisch	dreieckig	
Schoko	10%	15%	5%	30%
Kokos	25%	12%	3%	40%
Butter	5%	13%	12%	30%
	40%	40%	20%	100%

Im Statistiker-Testdorf Holtzhausen am Errl verkaufen sich die Kekssorten folgendermaßen:

KEKSE:

	Form			
Geschmack	rund	quadratisch	dreieckig	
Schoko	50	100	30	180
Kokos	80	30	10	120
Butter	70	20	110	200
	200	150	150	500

Tja, werden in Holtzhausen am Errl die verschiedenen Kekssorten nun entsprechend dem Verkauf in der BuReDeu gekauft oder nicht?

(Achtung bei den Freiheitsgraden!)

Aufgabe 7: (Chi-Quadrat, Phi-Korr., Produkt-Moment-Korr., t-Test)

Irgendwo, in den unendlichen Weiten des Universums, liegt der Planet Stastik Eins. Wir schreiben das Jahr 069624717 a.u.c.

Auf Stastik Eins leben die Efftests und die Fiekors, gar wunderliche Wesen, im Schnitt circa 1.50 m groß; die Efftests mit einem eher orangefarbenen, die Fiekors mit einem eher grünlichen Pelz, der jedoch sorgsam bedeckt gehalten wird.

Der Sternenforscher und Ethno-, Physio-, Psycho- und Logologe Commander McKay, der vor zwanzig Jahren diesen Planeten besuchte, berichtet unter anderem folgende Daten:

	Efftests	Fiekors	
Gumbatse	35	55	90
keine Gumbatse	40	20	60
	75	75	150

*Gumbatse sind kleine, weißlich-gelbe Wesen von rundlicher Gestalt und zäher Konsistenz, die beim Nationalsport Dänniz zum Einsatz kommen.

a) Würden Sie sagen, daß es bei der Verteilung der Gumbatse zwischen Efftests und Fiekors einen Unterschied gibt?

b) Wie hängen denn die Merkmale Efftests-Fiekors/Gumbatse-keine Gumbatse zusammen?

c) Commander McKay berichtet auch von einem zufälligen Besuch in einem öffentlichen Gebäude auf Stastik Eins. 25mal betrat er -ohne anzuklopfen- zufällig ein Zimmer; 13mal traf er auf einen fiekorischen Beamten und 12mal auf einen efftestischen. Mit seiner super-senso-dynamo-Stopuhr (made in Suitzaländ) maß er jedesmal die Zeit, die der betreffende Beamte benötigte, um sein Essen wegzupacken und sich gerade hinzusetzen. Folgende Werte wurden von ihm übermittelt. Die Fiekors benötigten im Schnitt 32 Sekunden (mit einer Standardabweichung S=4 sec); die Efftests 31 Sekunden (mit einer Standardabweichung S=8 sec). Tja, kann man behaupten, daß die Efftests die aufmerksameren Beamten waren?

Bei einem Besuch einer Revuegala zu seinen Ehren stellte er fest, daß die Menge unterschiedlich lange klatschte, je nachdem wieviel Pelz der Tänzerinnen zu sehen war. McKay übermittelte folgende Werte:

Dauer des Klatschens	4	6	1	2	4	3	1	5	4
Menge an sichtbarem Pelz	20	25	12	13	23	18	4	20	19

Dauer des Klatschens wurde in normalverteilten Zeiteinheiten gemessen, Menge an sichtbarem Pelz in ebenfalls normalverteilten Flächeneinheiten.

d) Was für ein Zusammenhang besteht denn nun zwischen den beiden Variablen?

e) Kann dieser Zusammenhang auch auf Populationsebene übertragen werden?

Eine Anekdote erzählt über den späten McKay, als er schon lange seine Tätigkeit an der Universität auf Alpha Fehltauri aufgenommen hatte, daß die Wahrscheinlichkeit dafür, daß er mit zwei verschiedenen Socken an die Uni kam, p=0.5 gewesen sein soll.

Die Wahrscheinlichkeit dafür, daß er den Hörsaal verwechselte, soll p=0.8 gewesen sein und -zu guter Letzt- die Wahrscheinlichekit dafür, daß er mitten im Satz den Faden verlor, p=0.5.

f) Wie groß ist denn die Wahrscheinlichkeit dafür, daß er seine Socken verwechselte, den Hörsaal verpaßte, aber nicht den Faden verlor?

g) Und wie groß ist die Wahrscheinlichkeit dafür, daß bei ihm mal alles klappte?

h) Und wie groß ist die Wahrscheinlichkeit dafür, daß an fünf hintereinander folgenden Tagen mindestens dreimal alles bei ihm schiefging?

Aufgabe 8: (Wahrscheinlichkeit)

Jeder kennt Robinson Crusoe, aber kaum jemand weiß, was ihm wirklich für Gedanken durch den Kopf gingen, als er das erste Mal Fußspuren im Sand sah:

"Oh, nein! Fußspuren! Ein Mensch! Ein Mann.....oder eine Frau! Mal überlegen: Die Wahrscheinlichkeit dafür, daß die Fußspuren von einem Kannibalen stammen, beträgt

Kai Budischewski: Statistik

p(K)=0.8; daß sie nicht von einem Kannibalen stammen, p(n.K.)=0.2. Generell gilt die Wahrscheinlichkeit p(f)=0.1, daß mir jemand freundlich gesinnt ist und p(n.f.)=0.9, daß mir jemand nicht freundlich gesinnt ist. Ein unfreundlicher Kannibale wird mich sofort fressen; ein freundlicher Kannibale wird mich vorher noch fragen, ob ich lieber gekocht oder gebraten werden möchte. Ein unfreundlicher Nicht-Kannibale wird mir einfach nur den Schädel einschlagen, ein freundlicher Nicht-Kannibale wird mich entweder höflich ignorieren (p(i)=0.6) oder mit mir die Flasche Rum köpfen (p(R)=0.4). Zum Glück habe ich ja meinen kleinen Statistikführer dabei, so daß ich die einzelnen Wahrscheinlichkeiten schnell ausrechnen kann."

"Tja, wie groß ist denn die Wahrscheinlichkeit, daß bei sechs verschiedenen Fußspuren höchstens einmal ein freundlicher, Rum trinkender, Nichtkannibale dabei ist?"

"Nun ja, und wie groß ist die Wahrscheinlichkeit dafür, daß höchstens einmal ein Typ dabei ist, der **nicht** freundlich, Rum trinkend und Nichtkannibale ist?"

Aufgabe 13: (Wahrscheinlichkeit)

Ein Dozent entwickelt einen (zugegebenermaßen recht seltsamen) Multiple-Choice-Test. Dieser Test besteht aus 5 Fragen. Zu jeder Frage gibt es 20 Antwortmöglichkeiten, von denen jeweils nur eine einzige richtig ist.

Bitte rechnen Sie bei dieser Aufgabe mit 4 Stellen hinter dem Komma genau!

a) Wie groß ist bei diesem Multiple-Choice-Test die Wahrscheinlichkeit dafür, genau zwei Fragen (egal welche) per Zufall richtig anzukreuzen?

b) Wie groß ist bei diesem Multiple-Choice-Test die Wahrscheinlichkeit dafür, höchstens zwei Fragen (egal welche) per Zufall richtig anzukreuzen?

7.3 AUFGABEN ZU „STATISTIK VERTIEFUNG“

Aufgabe XX: (Matrixalgebra)

Gegeben ist eine Matrix X vom Typ (5*2).

$$X = \begin{bmatrix} 6 & 2 \\ 8 & 4 \\ 5 & 1 \\ 7 & 2 \\ 6 & 3 \end{bmatrix}$$

a) Bilden Sie die transponierte Matrix X'.

b) Berechnen Sie bitte eine Matrix Y als Produkt von X' und X: Y=X'X

c) Berechnen Sie bitte eine Matrix Z als Produkt von X und X': Z = XX'

Lösung:

a)
$$X = \begin{bmatrix} 6 & 2 \\ 8 & 4 \\ 5 & 1 \\ 7 & 2 \\ 6 & 3 \end{bmatrix} \rightarrow transponieren \rightarrow X' = \begin{bmatrix} 6 & 8 & 5 & 7 & 6 \\ 2 & 4 & 1 & 2 & 3 \end{bmatrix}$$

b) Y=X'X

$$\begin{array}{c|c} & \begin{bmatrix} 6 & 2 \\ 8 & 4 \\ 5 & 1 \\ 7 & 2 \\ 6 & 3 \end{bmatrix} \quad X \\ \hline \underset{X'}{\begin{bmatrix} 6 & 8 & 5 & 7 & 6 \\ 2 & 4 & 1 & 2 & 3 \end{bmatrix}} & \underset{Y}{\begin{bmatrix} 210 & 81 \\ 81 & 34 \end{bmatrix}} \end{array}$$

c) Z=XX'

$$\begin{array}{c|c} & \begin{bmatrix}6&8&5&7&6\\2&4&1&2&3\end{bmatrix} X' \\ \hline \underset{X}{\begin{bmatrix}6&2\\8&4\\5&1\\7&2\\6&3\end{bmatrix}} & \underset{Z}{\begin{bmatrix}40&56&32&46&42\\56&80&44&64&60\\32&44&26&37&33\\46&64&37&53&48\\42&60&33&48&45\end{bmatrix}}\end{array}$$

<u>Aufgabe 9:</u> (Regressionsanalyse)

Steinzeit...

Mammute und Säbelzahntiger bevölkern die weiten Steppen Eurasiens.

Die ersten menschenähnlichen Wesen tapsen durch die Gegend, immer auf der Suche nach Nahrung. Dort, dort hinten! Mitten in einer furchtbaren Einöde kann man einen kleinen Stamm Neanderthaler sehen. Es ist der Stamm der Ugluk, wieder ohne Häuptling. Der letzte Häuptling wurde zu Geschnetzeltem verarbeitet, weil er den Stamm in diese furchtbare Einöde geführt hatte. Nun berät der Ältestenrat darüber, wie man endlich einmal einen Häuptling finden könnte, der etwas besser ist. Mit zu diesem Ältestenrat gehört der Zahlenmagier Gorkai, der Vorfahre einer ganzen Generation fähiger Statistiker (erinnert sei hier nur an: Kaibus, Kaios, Kabudis, McKay sowie Kai 'Das ist alles ganz einfach' B.). Und Gorkai sprach: "Woran erkennt man einen guten Häuptling, bzw. was könnte wichtig sein? Zum einen, wie schnell er Tierherden und damit Nahrung findet (Variable Schnelligkeit, in Tagen), zum anderen, wie weise er spricht (Variable Weisheit, in Weisheitspunkten). Ich weiß eine Magie, mit der ich -wenn ihr mir ein wenig helft- vorhersagen kann, wie gut jemand als Häuptling sein wird. Sagt mir, wie war das bei unseren letzten Häuptlingen mit diesen Fähigkeiten?"

Und sie stellten eine Tabelle auf (Oh Tabellenmagie!).

Häuptlingsqualität (y)	Schnelligkeit (x1)	Weisheit (x2)
1	14	6
2	14	4
4	10	6
2	8	8
8	4	8
3	12	5

Tja, nun. Wunderbar, Gorkai, jetzt haben wir eine *Tabelle* erstellt. Jetzt zaubere! Wir wollen wissen,

a) wie gut kann die Häuptlingsqualität anhand der Variablen Schnelligkeit und Weisheit erklärt werden (multiple Bestimmtheit)?

b) wie gut ist die Vorhersage, wenn nur die Variable Schnelligkeit berücksichtigt wird?

c) wie gut ist die Vorhersage, wenn nur die Variable Weisheit berücksichtigt wird?

d) bringt die Hinzunahme der Variablen Weisheit eine signifikante Verbesserung?

e) bringt die Hinzunahme der Variablen Schnelligkeit eine signifikante Verbesserung?

f) welche Häuptlingsqualität erzielt ein Kandidat mit einem Schnelligkeitswert von 4 und einem Weisheitswert von 9? Und wie groß ist das dazugehörige Konfidenzintervall (95%)?

<u>Aufgabe 10:</u> (Multiple Regressionsanalyse)

Wir befinden uns im Jahr 2204. Aufgrund ihrer andauernden Quängeleien nach einem Psychotherapeutengesetz sind alle Psychologen von der Mediziner/Heilpraktikervereinigung auf den kleinen, düsteren Planeten Stastik I verbannt worden.

Nur über einige wenige Spione können unsere wackeren Psychologen den Kontakt zur

Erde aufrechterhalten.

"Wenn wir nur zeigen könnten, daß unsere Anwesenheit auf der Erde den Menschen Gutes bringt!"

"Ja, dann könnten wir sie vielleicht überzeugen, daß wir wieder zurück dürfen."

"Sagt 'mal, da gab es doch diese eine Untersuchung von Prof.Dr.K.Budis, was ist eigentlich daraus geworden?"

"Ich glaube, diese Untersuchung wurde nie ausgewertet, aber die Daten müßten sich noch in meinem Computer befinden. Ich sehe nach."

Und tatsächlich, die Daten waren noch vorhanden:

Untersucht wurde damals der Zusammenhang zwischen allgemeiner Lebenszufriedenheit (Kriterium) und Einstellung gegenüber Psychologen (x_1) sowie Medikamentenkonsum (x_2). Je höher die Werte, desto größer die allgemeine Lebenszufriedenheit, desto positiver die Einstellung gegenüber Psychologen und desto höher der Medikamentenkonsum.

Allgemeine Lebenszufriedenheit	Einstellung gegenüber Psychologen	Medikamentenkonsum
15	19	2
16	14	6
9	11	4
5	9	16
2	3	18
4	5	19
6	8	7
9	10	3
11	13	2
17	17	1

a) Berechnen Sie die optimalen Schätzwerte der Einflußgewichte!

b) Berechnen Sie den Determinationskoeffizienten!

c) Wie gut kann die Variable x_1 alleine die Variation des Kriteriums erklären?

d) Bringt die Hinzunahme des Prädiktors 'Medikamentenkonsum' eine signifikante

Verbesserung der Vorhersageleistung?

e) Berechnen Sie den Schätzwert für die Allgemeine Lebenszufriedenheit für eine Person mit der Einstellung 12 und dem Medikamentenkonsum 8.

Geben Sie auch ein 95%-Konfidenzintervall an.

Aufgabe 11: (Varianzanalyse)

Statistik-Champ Kejbee interessierte sich eines Tages dafür, ob sich die Punktzahlen einer Methodenklausur danach unterschieden, ob die Leute den Taschenrechner STASTIK 5000, STASTIK 2000 oder STASTIK 0815 verwendeten (Faktor B) und ob sich dabei um Jungspunde (bis 30 Jahre) oder um Greise (ab 30 Jahre) (Faktor A) handelte.

Und Kejbee stellte sich eine Tabelle auf:

	Stastik 5000	Stastik 2000	Stastik 0815
Greise	10 12 14 $\boxed{1}$	15 18 14 $\boxed{3}$	18 20 19 $\boxed{5}$
Jungspunde	16 12 10 $\boxed{2}$	17 15 13 $\boxed{4}$	23 12 24 $\boxed{6}$

Gemessen wurden die Punkte in einer Methodenklausur.

a) Globale Effekte?

b) Haupteffekt Faktor A (Alterskategorie)?

c) Haupteffekt Faktor B (Taschenrechnermodell)?

d) Wechselwirkungseffekte?

Aufgabe 12: (Varianzanalyse)

Sehr geehrte Damen und Herren,

wir übertragen heute live die Abschlußfeierlichkeiten und Prämierungen aus dem Museum alter Theorien. 18 Schiedsrichter durften je einmal eine Benotung abgeben. Die alten Theorien stammten aus den Bereichen Psychologie und Psychoanalyse und sie wurden eingeteilt nach dem Komplexitätsgrad: Leicht, mittel oder hoch.

Zur Auswahl standen Theorien von Seligman ('gelernte Hilflosigkeit'), Freud ('Ich-Es-ÜberIch'), James-Lange ('Ich bin traurig, weil ich weine'), Maslow ('Bedürfnishierarchie'), Budis (Einstellung gegenüber Psychologen und die allgemeine Lebenszufriedenheit) und vielen anderen.

Die folgende Tabelle listet die Einzelurteile auf:

	Komplexitätsgrad		
Fachrichtung	leicht	mittel	hoch
Psychologie	40 24 26 $\boxed{1}$	10 8 12 $\boxed{3}$	10 26 24 $\boxed{5}$
Psychoanalyse	16 14 15 $\boxed{2}$	20 22 18 $\boxed{4}$	4 4 7 $\boxed{6}$

a) Globale Effekte?

b) Haupteffekt Faktor A (Fachrichtung)?

c) Haupteffekt Faktor B (Komplexitätsgrad)?

d) Wechselwirkung?

LÖSUNGEN

Aufgabe 1:

zu a) 80 Bälle von 200 sind blau $\Rightarrow$ p(blau)= $\frac{80}{200} = \frac{40}{100} = \mathbf{0.4}$

zu b) 80 Bälle von 200 sind rot $\Rightarrow$ p(rot)= $\frac{80}{200} = \frac{40}{100} = \mathbf{0.4}$;

40 Bälle von 200 sind grün $\Rightarrow$ p(grün)= $\frac{40}{200} = \frac{20}{100} = \mathbf{0.2}$

p(rot **oder** grün)= 0.4 + 0.2 = **0.6**

zu c) 30 Bälle von 200 sind grün und haben einen Stern $\Rightarrow$ p(grün mit Stern)= $\frac{30}{200} = \frac{15}{100} = \mathbf{0.15}$

zu d) p(grün)=0.2 p(blau)=0.4 $\Rightarrow$ p(grün **und** blau)= $0.2 \times 0.4 \times 2$ = **0.16**

Warum wurde hier mit 2 multipliziert? Ganz einfach: Es gibt zwei Möglichkeiten, wie man ziehen kann. Entweder zuerst grün und dann blau **oder** zuerst blau und dann grün. Deswegen muß mit 2 multipliziert werden.

zu e) p(zuerst grün **und** dann blau)= 0.2×0.4 = **0.08**

zu f) p(grün ohne Stern)= $\frac{10}{200} = 0.05$ p(blau mit Stern)= $\frac{40}{200} = 0.2 \Rightarrow$

p(grün ohne Stern **oder** blau mit Stern)=0.2 + 0.05 = **0.25**

zu g) p(rot)=0.4 p(mit Stern)= $\frac{130}{200} = 0.65$

p(rot **oder** mit Stern)= 0.4 + 0.65 - 0.3 = **0.75**

Warum wurde hier 0.3 abgezogen? Weil man sonst die sechzig roten Bälle mit Stern doppelt gezählt hätte, einmal zu den 80 roten Bällen und einmal zu den 130 Bällen mit Stern. Also mußten die sechzig Bälle einmal abgezogen werden. 60 Bälle entsprechen einem Wert von 0.3.

zu h) 80 blaue Bälle gibt es insgesamt, 40 davon haben einen Stern $\Rightarrow$

p(Stern bei blau)=$\frac{40}{80} =$ **0.5**

zu i) 70 Bälle ohne Stern gibt es insgesamt, 20 davon sind rot $\Rightarrow$

p(rot bei ohne Stern)=$\frac{20}{70} =$ **0.286**

Aufgabe 2:

zu a) Da es an der Uni insgesamt gleichviele Studentinnen wie Studenten gibt, ist die Wahrscheinlichkeit p(weiblich)=0.5

zu b) Da es insgesamt 330 Psychologiestudentinnen unter den 5000 Studenten gibt, ist die Wahrscheinlichkeit p(weiblich und Psychologie)=$\frac{330}{5000} =$ **0.066**

zu c) Von den 330 Psychologiestudentinnen haben 40% ($\Rightarrow$p=0.4) gute statistische Kenntnisse; die Wahrscheinlichkeit für weiblich und Psychologie ist p=0.066; Daraus folgt, die Wahrscheinlichkeit für weiblich und Psychologie und gute statistische Kenntnisse ist

p(weiblich und Psychologie und gut stat.) = 0.066×0.4 = 0.0264.

Aufgabe 3:

zu a) p(psych) = 0.05

p(männlich bei psych) = 0.33

p(gut stat.) = 0.7

p(psych **und** männlich **und** gut stat.) = 0.05×0.33×0.7 = **0.01155** $\cong$ 1.155% aller Studierenden

zu b) p(psych) = 0.05

p(weiblich bei psych) = 0.67

p(gut stat.) = 0.7

p(psych **und** weiblich **und** gut stat.) = 0.05×0.67×0.7 = **0.02345** $\cong$ 2.345% aller Studierenden

n×p = Anzahl 20000×0.02345 = **469**

d.h. es gibt 469 weibliche Psychologiestudenten mit guten statistischen Kenntnissen, zumindest würde man soviele erwarten.

zu c) p(psych) = 0.05

p(weiblich erwartet) = 0.50

p(psych **und** weiblich erwartet) = 0.05×0.50 = **0.025** $\cong$ 2.5% aller Studierenden

n×p = Anzahl 20000×0.025 = **500** weibliche Psychologiestudenten hätte man erwartet.

Aufgabe 4:

zu a) Vergleich von Rangplätzen $\Rightarrow$ ordinalskaliert $\Rightarrow$ U-Test (Mann/Whitney), da unabhängige Stichproben

H_0: Die Rangplätze sind gleich verteilt.

H_1: Die Rangplätze sind nicht gleich verteilt.

Wenn man mathematische Hypothesen formulieren will, sollte man dazu den T-Wert (hat nichts mit dem t-Test zu tun!!!) verwenden, etwa folgendermaßen:

H_0: $T_1 = T_2$

H_1: $T_1 \neq T_2$

Normalerweise werden beim U-Test jedoch nur verbale Hypothesen formuliert.

Zuerst werden für jede Stichprobe die Rangplatzsummen (T) gebildet.

$T_{Penthes} = 1+4+8+9+10 = 32 = T_1$ $n_{Penthes} = 5 = n_1$

$T_{Sileas} = 2+3+5+6+7 = 23 = T_2$ $n_{Sileas} = 5 = n_2$

Als nächstes wird der U-Wert errechnet:

$$U = n_1 \times n_2 + \frac{n_1 \times (n_1+1)}{2} - T_1 = 5 \times 5 + \frac{5 \times (5+1)}{2} - 32 = \mathbf{8}$$

Nun wird der U´-Wert errechnet:

$$U' = n_1 \times n_2 - U = 5 \times 5 - 8 = \mathbf{17}$$

Mit dem kleineren der beiden U-Werte geht man in die U-Werte-Tabelle (E) und schlägt dort unter den jeweiligen n_1 und n_2 die Auftretenswahrscheinlichkeit des kleineren U-Wertes nach. Ist diese Auftretenswahrscheinlichkeit kleiner als p = 0.05, dann sind die Rangplätze nicht gleich verteilt, man hat ein signifikantes Ergebnis.

Hier ist der kleinere U-Wert = 8, $n_1 = 5$ und $n_2 = 5$.

Als Auftretenswahrscheinlichkeit findet man für den Wert 8 p(U=8) = 0.210. Dieser Wert ist größer als 0.05, daher liegt hier kein signifikantes Ergebnis vor, die Rangplätze sind nicht überzufällig unterschiedlich verteilt, oder anders gesagt, die unterschiedliche Verteilung der Rangplätze ist wohl eher aus Zufall geschehen und unterliegt keiner systematischen Ursache.

zu b) Ein geradezu klassischer Vier-Felder-Test (Chi-Quadrat)

H_0: Die beobachteten Werte sind gleich (entsprechen) den erwarteten.

H_1: Die beobachteten Werte sind ungleich (unterscheiden sich von) den erwarteten.

Beim Chi^2-Test werden in der Regel keine mathematischen Hypothesen formuliert. Wenn man unbedingt auch mathematische Hypothesen formulieren will, sollte dies ungefähr so aussehen:

H_0: $f_{b1} - f_{e1} = 0$ und $f_{b2} - f_{e2} = 0$ und $f_{b3} - f_{e3} = 0$ und $f_{b4} - f_{e4} = 0$

H_1: $f_{b1} - f_{e1} \neq 0$ und $f_{b2} - f_{e2} \neq 0$ und $f_{b3} - f_{e3} \neq 0$ und $f_{b4} - f_{e4} \neq 0$

Die Formel für den Chi^2-Test (Vierfelder) lautet:

Kai Budischewski: Statistik

$$\chi^2_{emp} = \frac{n \times (b \times c - a \times d)^2}{(a+b) \times (c+d) \times (a+c) \times (b+d)} = \frac{40 \times (8 \times 9 - 12 \times 11)^2}{21 \times 19 \times 20 \times 20}$$

$$\chi^2_{emp} = \frac{40 \times (-60)^2}{159600} = \frac{40 \times 3600}{159600} = \frac{144000}{159600} = \mathbf{0.902}$$

Als nächstes sucht man sich das kritische Chi^2. Dazu benötigt man die Freiheitsgrade. Bei einem Vierfelder-Schema kann nur eine einzige Zahl frei eingesetzt werden, wenn die Randsummen berücksichtigt werden. Die Randsummen wurden hier berücksichtigt, daher hat man ein df=1.

In Tabelle D liest man unter df=1 und α=5% das kritische $\chi^2_{krit(\alpha=0.05, df=1)} =$ **3.841**

Das empirische Chi^2 ist kleiner als das kritische Chi^2, daher nicht signifikant, man bleibt bei der H_0.

$\chi^2_{emp} < \chi^2_{krit} \Rightarrow$ *nicht signifikant* $\Rightarrow H_0$.

Die beobachteten Werte unterscheiden sich nicht von den erwarteten, auch bei diesem Wettkampf gibt es keinen signifikanten Unterschied zwischen den Penthes und den Sileas.

zu c) Die Penthes hatten als Mittelwert 4 Männer, die Sileas als Mittelwert 2,5 Männer. Um Mittelwerte zu vergleichen, wird der t-Test verwendet. Hierbei handelt es sich um unabhängige Stichproben (schließlich kann man nicht von einer Penthe auf eine bestimmte Silea schließen).

Zuerst sollte man aufschreiben, was einem alles gegeben wurde:

$n_{Sileas} = 30$	$n_{Penthes} = 30$
$\bar{x}_{Sileas} = 2{,}5$	$\bar{x}_{Penthes} = 4$
$S_{Sileas} = 0{,}5$	$S_{Penthes} = 0{,}75$

$S^2_{Sileas} = 0{,}25$ $\quad$ $S^2_{Penthes} = 0{,}5625$

Bevor ein t-Test durchgeführt werden kann, muß immer erst ein F-Test durchgeführt werden. Einzige Ausnahme: Wenn ausdrücklich etwas über die Varianzenhomogenität bzw. -heterogenität in der Aufgabe ausgesagt wurde!

F-Test (Varianzenvergleich)

H_0: Die Varianzen sind gleich $\quad$ H_0: $\sigma_1^2 = \sigma_2^2$

H_1: Die Varianzen sind unterschiedlich $\quad$ H_1: $\sigma_1^2 \neq \sigma_2^2$

Die Formeln:

$$F_{emp} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \qquad \hat{\sigma}^2 = S^2 \times \frac{n}{n-1}$$

$$\hat{\sigma}^2_{Penthes} = 0{,}5625 \times \frac{30}{29} = 0{,}582$$

$$\hat{\sigma}^2_{Sileas} = 0{,}25 \times \frac{30}{29} = 0{,}259$$

Beim F-Test wird immer die größere Varianz auf den Bruchstrich (in den Zähler) gesetzt, die kleinere Varianz unter den Bruchstrich (in den Nenner).

$$F_{emp} = \frac{0{,}582}{0{,}259} = 2{,}247$$

Diesen empirischen F-Wert muß man nun mit dem kritischen F-Wert vergleichen. Dazu benötigt man Zähler und Nenner-Freiheitsgrade. Im Zähler steht die Varianz der Penthes mit n = 30 $\Rightarrow$ $df_{Zähler}$ = 29;

im Nenner steht die Varianz der Sileas mit n = 30 $\Rightarrow$ df_{Nenner} = 29. Mit diesen beiden Freiheitsgraden geht man nun in Tabelle C und schlägt dort für α=0.05 den kritischen F-

Wert nach:

$$F_{krit(\alpha=0.05;df_1=29,df_2=29)} = 1{,}85$$

Da der empirische F-Wert größer als der kritische F-Wert ist, handelt es sich um ein signifikantes Ergebnis, man entscheidet sich also für H_1, d.h. die Varianzen sind nicht gleich (homogen) sondern unterschiedlich (heterogen).

Da die Varianzen heterogen sind, muß der t-Test für heterogene Varianzen herangezogen werden:

t-Test für heterogene Varianzen (Mittelwertevergleich)

Die Frage war, ob die Penthes tatsächlich **mehr** Männer haben als die Sileas. Es handelt sich hierbei also um eine einseitige Hypothese (es wird nur nach einer Richtung gefragt, nämlich nach mehr).

H_0: Die Penthes haben nicht mehr Männer als die Sileas

$$\mu_{Penthes} \leq \mu_{Sileas}$$

H_1: Die Penthes haben mehr Männer als die Sileas

$$\mu_{Penthes} > \mu_{Sileas}$$

Die Formeln:

$$t_{emp} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_d} \qquad \hat{\sigma}_d = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} \qquad df = \frac{n_1 + n_2}{2}$$

$$\hat{\sigma}_d = \sqrt{\frac{0{,}582}{30} + \frac{0{,}259}{30}} = \sqrt{0{,}0194 + 0{,}0086} = \sqrt{0{,}028} = 0{,}167$$

$$t_{emp} = \frac{4-2{,}5}{0{,}167} = \frac{1{,}5}{0{,}167} = 8{,}982$$

Dieser empirische t-Wert muß nun mit einem kritischen t-Wert verglichen werden. Um den kritischen t-Wert aus Tabelle B ablesen zu können, benötigt man die Freiheitsgrade. Beim t-Test für heterogene Varianzen berechnen sich die Freiheitsgrade nach folgender Formel:

$$df = \frac{n_1+n_2}{2} = \frac{30+30}{2} = 30$$

Als kritischen t-Wert liest man bei α=0,05, einseitig, folgenden Wert ab:

$$t_{krit(\alpha=0.05;einseitig;df=30)} = 1{,}697$$

Da der empirische t-Wert größer als der kritische t-Wert ist, handelt es sich um ein signifikantes Ergebnis, man entscheidet sich für H_1.

$t_{emp} > t_{krit} \Rightarrow$ signifikant $\Rightarrow H_1$

Die Penthes haben - hochgerechnet (bezogen) auf die Population - überzufällig mehr Männer als die Sileas.

zu d) Hier ist nach einem Zusammenhang gefragt, nämlich ob das eine mit dem anderen in Beziehung stehe. Also: Korrelation. Da beide Merkmale normalverteilt sind, kann für beide Variablen Intervallskalenniveau angenommen werden. Daraus folgt, hier soll eine Produkt-Moment-Korrelation gerechnet werden.

Dazu berechnet man am besten erst einmal den Mittelwert pro Variable und erstellt dann eine Tabelle.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{10+7+8+10+6+13+11+9}{8} = \frac{74}{8} = 9{,}25$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{27+18+23+22+14+28+26+22}{8} = \frac{180}{8} = 22{,}5$$

Pb-Nr.	x	x- $\bar{x}$	(x- $\bar{x}$)^2	y	y- $\bar{y}$	(y- $\bar{y}$)^2	(x- $\bar{x}$)×(y- $\bar{y}$)
1	10	0,75	0,5625	27	4,5	20,25	3,375
2	7	-2,25	5,0625	18	-4,5	20,25	10,125
3	8	-1,25	1,5625	23	0,5	0,25	-0,625
4	10	0,75	0,5625	22	-0,5	0,25	-0,375
5	6	-3,25	10,5625	14	-6,5	42,25	21,125
6	13	3,75	14,0625	28	5,5	30,25	20,625
7	11	1,75	3,0625	26	3,5	12,25	6,125
8	9	-0,25	0,0625	22	-0,5	0,25	0,125
			Σ = 35,5			Σ = 126	Σ = 60,5

$$S_x^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} = \frac{35{,}5}{8} = 4{,}4375 \Rightarrow S_x = 2{,}107$$

$$S_y^2 = \frac{\sum_{i=1}^{n}(y_i\bar{y})^2}{n} = \frac{126}{8} = 15{,}75 \Rightarrow S_y = 3{,}969$$

$$COV_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\times(y_i-\bar{y})}{n} = \frac{60{,}5}{8} = 7{,}5625$$

$$r_{xy} = \frac{COV_{xy}}{S_x \times S_y} = \frac{7{,}5625}{2{,}107\times 3{,}969} = \frac{7{,}5625}{8{,}363} = 0{,}904$$

$$r_{xy}^2 = r \times r = 0{,}904 \times 0{,}904 = 0{,}817$$

$r_{xy}^2 \times 100\% = 81{,}7\%$ erklärte Variation/Varianz

Es besteht ein hoher (deutlicher) positiver Zusammenhang von r = 0,904 zwischen der Variable ´Nasengröße´ und der Variablen ´Glücklichsein´.

Ob dieser Zusammenhang so hoch ist, daß man nicht mehr von Zufall reden kann, er also signifikant wäre, könnte mit dem t-Test für Korrelationen überprüft werden. Die Formeln hierzu lauten:

$$t_{emp} = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}} \qquad df = n - 2$$

Hier angewendet würde man errechnen:

$$t_{emp} = \frac{0{,}904 \times \sqrt{8-2}}{\sqrt{1-0{,}817}} = \frac{0{,}904 \times \sqrt{6}}{\sqrt{0{,}183}} = \frac{0{,}904 \times 2{,}449}{0{,}428} = 5{,}173$$

Als Freiheitsgrade errechnet man: df = n - 2 = 8 - 2 = 6

Als kritischen t-Wert liest man aus Tabelle B für df = 6 und α=0,05 zweiseitig (0,025 einseitig) den Wert:

$$t_{krit(df=6;\alpha=0{,}05,zweiseitig)} = 2{,}447$$

Da der empirische t-Wert größer als der kritische t-Wert ist, hätte man ein signifikantes Ergebnis, die Korrelation ist überzufällig von Null verschieden.

Als Hypothesen hätte man schreiben müssen:

H_0: $\rho = 0$

H_1: $\rho \neq 0$

Aufgabe 5:

zu a)

p(brauchbar) = 3/10 = 0,3

p(unbrauchbar) = 7/10 = 0,7

n = 6

k = wenigstens 1 = 1 oder 2 oder 3 oder 4 oder 5 oder 6

Zuerst sollte für jedes k die mögliche Anzahl an Kombinationen errechnet werden; dies geschieht über die Formel:

$$\binom{n}{k} = \frac{n!}{k! \times (n-k)!}$$

Der Fall k = 1:

$$\binom{n}{k} = \binom{6}{1} = \frac{6!}{1! \times (6-1)!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{1 \times 5 \times 4 \times 3 \times 2 \times 1} = 6$$

Die Wahrscheinlichkeit für **eine** dieser Kombinationen errechnet sich zu:

p = 0,3×0,7×0,7×0,7×0,7×0,7 = 0,05

Die Gesamtwahrscheinlichkeit für einen unter sechs, egal welcher, berechnet sich jetzt als p×Anzahl Kombinationen = 0,05×6 = 0,3

Der Fall k = 2:

$$\binom{n}{k} = \binom{6}{2} = \frac{6!}{2! \times (6-2)!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{2 \times 1 \times 4 \times 3 \times 2 \times 1} = 15$$

Die Wahrscheinlichkeit für **eine** dieser Kombinationen errechnet sich zu:

p = 0,3×0,3×0,7×0,7×0,7×0,7 = 0,022

Die Gesamtwahrscheinlichkeit für zwei unter sechs, egal welche, berechnet sich jetzt als p×Anzahl Kombinationen = 0,022×15 = 0,33

Der Fall k = 3:

$$\binom{n}{k} = \binom{6}{3} = \frac{6!}{3! \times (6-3)!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 3 \times 2 \times 1} = 20$$

Die Wahrscheinlichkeit für **eine** dieser Kombinationen errechnet sich zu:

p = 0,3×0,3×0,3×0,7×0,7×0,7 = 0,009

Die Gesamtwahrscheinlichkeit für drei unter sechs, egal welche, berechnet sich jetzt als p ×Anzahl Kombinationen = 0,009×20 = 0,18

Der Fall k = 4:

$$\binom{n}{k} = \binom{6}{4} = \frac{6!}{4! \times (6-4)!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1 \times 2 \times 1} = 15$$

Die Wahrscheinlichkeit für **eine** dieser Kombinationen errechnet sich zu:

p = 0,3×0,3×0,3×0,3×0,7×0,7 = 0,004

Die Gesamtwahrscheinlichkeit für vier unter sechs, egal welche, berechnet sich jetzt als p ×Anzahl Kombinationen = 0,004×15 = 0,06

Der Fall k = 5:

$$\binom{n}{k} = \binom{6}{5} = \frac{6!}{5! \times (6-5)!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{5 \times 4 \times 3 \times 2 \times 1 \times 1} = 6$$

Die Wahrscheinlichkeit für **eine** dieser Kombinationen errechnet sich zu:

p = 0,3×0,3×0,3×0,3×0,3×0,7 = 0,002

Die Gesamtwahrscheinlichkeit für fünf unter sechs, egal welche, berechnet sich jetzt als

p×Anzahl Kombinationen = 0,002×6 = 0,012

Der Fall k = 6:

$$\binom{n}{k} = \binom{6}{6} = \frac{6!}{6! \times (6-6)!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{6 \times 5 \times 4 \times 3 \times 2 \times 1 \times 1} = 1$$

Die Wahrscheinlichkeit für **eine** dieser Kombinationen errechnet sich zu:

Kai Budischewski: Statistik

p = 0,3×0,3×0,3×0,3×0,3×0,3 = 0,001
Die Gesamtwahrscheinlichkeit für sechs unter sechs berechnet sich jetzt als

p×Anzahl Kombinationen = 0,001×1 = 0,001

Diese Wahrscheinlichkeiten müssen nun noch addiert werden (es kann ja sein 1 oder 2 oder 3 oder...), um die Antwort für die Frage zu errechnen:

p(mindestens 1) = p(1)+p(2)+p(3)+p(4)+p(5)+p(6)

= 0,3 + 0,33 + 0,18 + 0,06 + 0,012 + 0,001 = **0,883**

Diese Aufgabe hätte man auch schneller und einfacher ausrechnen können, wenn man folgende Überlegung angestellt hätte: Jeder Fall ist möglich außer, keinen einzigen brauchbaren dabeizuhaben. Wie groß ist denn die Wahrscheinlichkeit dafür, keinen einzigen brauchbaren dabeizuhaben?

p(0) = 0,7×0,7×0,7×0,7×0,7×0,7 = 0,118

Wieviele Kombinationsmöglichkeiten gibt es hierfür? Nur eine einzige, d.h. die Wahrscheinlichkeit dafür, keinen einzigen dabeizuhaben ist p=0,118.

Da sich Wahrscheinlichkeiten zu 1 addieren, muß die Wahrscheinlichkeit für alle anderen Fälle bis auf diesen 1-p(0) sein, also

p(alles außer 0) = 1- 0,118 = 0,882.

Dies ist - bis auf Rundungsungenauigkeiten - dasselbe Ergebnis.

zu b) Mindestens einen guten heißt 1 oder 2 oder 3 oder 4. D.h. alle Fälle bis auf ´keinen´ (0) brauchbare.

Wahrscheinlichkeiten addieren sich zu 1, also muß die Wahrscheinlichkeit für jeden Fall außer keinen Brauchbaren 1- p(0) sein.

p(0) = 0,7×0,7×0,7×0,7 = 0,240 (es gibt nur eine Kombination)

Daraus folgt: p(1 oder 2 oder 3 oder 4) = 1 - 0,240 = **0,76** zu c) Hier sollen Mittelwerte verglichen werden. Das dafür zuständige Testverfahren ist der t-Test. Vor einem t-Test muß jedoch immer ein F-Test gerechnet werden.

Für den F-Test benötigt man die Varianzen, die man aus den Standardabweichungen errechnen kann:

$$S^2_{Sparta} = S_{Sparta} \times S_{Sparta} = 1{,}5 \times 1{,}5 = 2{,}25$$

$$S^2_{Athen} = S_{Athen} \times S_{Athen} = 1{,}2 \times 1{,}2 = 1{,}44$$

F-Test (Varianzenvergleich)

H_0: Die Varianzen sind gleich $\qquad H_0\colon\ \sigma_1^2 = \sigma_2^2$

H_1: Die Varianzen sind unterschiedlich $\qquad H_1\colon\ \sigma_1^2 \neq \sigma_2^2$

Die Formeln:

$$F_{emp} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \qquad \hat{\sigma}^2 = S^2 \times \frac{n}{n-1}$$

$$\hat{\sigma}^2_{Sparta} = 2{,}25 \times \frac{50}{49} = 2{,}296$$

$$\hat{\sigma}^2_{Athen} = 1{,}44 \times \frac{42}{41} = 1{,}475$$

Beim F-Test wird immer die größere Varianz auf den Bruchstrich (in den Zähler) gesetzt, die kleinere Varianz unter den Bruchstrich (in den Nenner).

$$F_{emp} = \frac{2{,}296}{1{,}475} = 1{,}557$$

Diesen empirischen F-Wert muß man nun mit dem kritischen F-Wert vergleichen. Dazu benötigt man Zähler und Nenner-Freiheitsgrade. Im Zähler steht die Varianz der Spartaner mit n = 50 $\Rightarrow df_{Zähler} = 49$;

im Nenner steht die Varianz der Athener mit n = 42 $\Rightarrow df_{Nenner} = 41$. Mit diesen beiden Freiheitsgraden geht man nun in Tabelle B und schlägt dort für α=0.05 den kritischen F-Wert nach:

$$F_{krit(\alpha=0.05;df_1=49,df_2=41)} = 1{,}66$$

Da der kritische F-Wert größer als der empirische F-Wert ist, handelt es sich um ein nicht signifikantes Ergebnis, man entscheidet sich also für H_0, d.h. die Varianzen sind gleich (homogen) und nicht unterschiedlich (heterogen).

Da die Varianzen homogen sind, muß der t-Test für homogene Varianzen herangezogen werden:

t-Test für homogene Varianzen (Mittelwertevergleich)

Die Frage war, ob die Spartaner tatsächlich **mehr** Esel haben als die Athener. Es handelt sich hierbei also um eine einseitige Hypothese (es wird nur nach einer Richtung gefragt, nämlich nach mehr).

H_0: Die Spartaner haben nicht mehr Esel als die Athener

$$\mu_{Sparta} \leq \mu_{Athen}$$

H_1: Die Spartaner haben mehr Esel als die Athener

$$\mu_{Sparta} > \mu_{Athen}$$

Die Formeln:

$$t_{emp} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)}} \qquad \hat{\sigma}_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{(n_1-1)\times\hat{\sigma}_1^2 + (n_2-1)\times\hat{\sigma}_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$df = n_1 + n_2 - 2$$

$$\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{(50-1)\times 2{,}296+(42-1)\times 1{,}475}{50+42-2}\times\left(\frac{1}{50}+\frac{1}{42}\right)}$$

$$\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{112{,}504+60{,}475}{90}\times(0{,}044)} = \sqrt{\frac{172{,}979}{90}\times 0{,}044} = \sqrt{0{,}085} = 0{,}292$$

$$t_{emp} = \frac{4-3}{0{,}292} = 3{,}425$$

Dieser empirische t-Wert muß nun mit einem kritischen t-Wert verglichen werden. Um den kritischen t-Wert aus TabelleC ablesen zu können, benötigt man die Freiheitsgrade. Beim t-Test für homogene Varianzen berechnen sich die Freiheitsgrade nach folgender Formel:

$df = n_1 + n_2 - 2 = 50 + 42 - 2 = 90$

Als kritischen t-Wert liest man bei α=0,05, einseitig, folgenden Wert ab:

$$t_{krit(\alpha=0.05;einseitig;df=90)} = 1{,}671$$

Da der empirische t-Wert größer als der kritische t-Wert ist, handelt es sich um ein signifikantes Ergebnis, man entscheidet sich für H_1.

$t_{emp} > t_{krit} \Rightarrow$ signifikant $\Rightarrow H_1$

Die Spartaner haben - hochgerechnet (bezogen) auf die Population - überzufällig mehr Esel als die Athener.

Aufgabe 6:

Hier sollen beobachtete Werte (tatsächlicher Verkauf in Holtzhausen a.E.) mit erwarteten/theoretischen Werten (Was hätte in Holtzhausen verkauft werden sollen) verglichen werden. Das hierfür zuständige Testverfahren ist der Chi^2-Test.

H_0: Der Verkauf in Holtzhausen a.E. entspricht dem Verkauf in der BuReDeu

H_1: Der Verkauf in Holtzhausen a.E. entspricht dem Verkauf in der BuReDeu nicht

Die Formel für den Chi^2-Test lautet:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_{bi}-f_{ei})^2}{f_{ei}}$$

Kai Budischewski: Statistik

Hierbei ist:

f_{bi} : beobachteter Wert der Kategorie i

f_{ei} : erwarteter Wert der Kategorie i

k : Anzahl der Kategorien

Die Schwierigkeit liegt eigentlich nur darin, wie man die erwarteten Werte ausrechnet. Gewöhnlich geschieht dies über die Randsummen. Hier liegen jedoch weitere Informationen vor, nämlich die Prozentzahlen des Verkaufs in der BuReDeu. Diese Prozentzahlen sollte man verwenden, um die erwarteten Werte für Holtzhausen a.E. auszurechnen. Das heißt, die Randsummen werden nicht benötigt und sollten daher auch nicht in der Tabelle aufgelistet werden.

KEKSE: beobachtete Werte

	Form		
Geschmack	rund	quadratisch	dreieckig
Schoko	50	100	30
Kokos	80	30	10
Butter	70	20	110

500 Päckchen Kekse wurden insgesamt verkauft. Für die Kategorie Kokos-quadratisch hätte man nach den Angaben für die BuReDeu 12% erwartet. 12% von 500 sind 60 Päckchen (1% = 5 Päckchen).

Die folgende Tabelle listet die erwarteten Werte auf.

KEKSE: erwartete Werte

	Form		
Geschmack	rund	quadratisch	dreieckig
Schoko	50	75	25
Kokos	125	60	15

Butter	25	65	60

Nun wird ganz normal in die Formel eingesetzt:

$$\chi^2_{emp} = \frac{(50-50)^2}{50} + \frac{(100-75)^2}{75} + \frac{(30-25)^2}{25} + \frac{(80-125)^2}{125}$$
$$+ \frac{(30-60)^2}{60} + \frac{(10-15)^2}{15} + \frac{(70-25)^2}{25} + \frac{(20-65)^2}{65} + \frac{(110-60)^2}{60}$$

$$\chi^2_{emp} = \frac{0}{50} + \frac{625}{75} + \frac{25}{25} + \frac{2025}{125} + \frac{900}{60} + \frac{25}{15} + \frac{2025}{25} + \frac{2025}{65} + \frac{2500}{60}$$

$$\chi^2_{emp} = 0 + 8{,}333 + 1 + 16{,}2 + 15 + 1{,}667 + 81 + 31{,}154 + 41{,}667$$

$$\chi^2_{emp} = 196{,}021$$

Dieses empirische Chi^2 muß nun mit einem kritischen Chi^2 verglichen werden. Um das kritische Chi^2 aus der Tabelle lesen zu können, benötigte man die Freiheitsgrade. Da hier nicht über die Randsummen geschätzt wurde, berechnen sich die Freiheitsgrade als df = Anzahl Kategorien - 1 = 9 - 1 = 8.

Als kritischen Chi^2-Wert liest man nun für α = 0,05 und df = 8 aus Tabelle D den Wert 15,507 ab.

Da der empirische Chi^2-Wert größer als der kritische Chi^2-Wert ist, hat man ein signifikantes Ergebnis, d.h. die Verkaufszahlen in Holtzhausen a.E. entsprechen nicht den Verkaufszahlen in der BuReDeu.

Aufgabe 7:

zu a) Hier wird nach einem Unterschied gefragt. Beide Merkmale (Gumbatse/keine Gumbatse und Efftests/Fiekors) sind nominalskaliert. Tja, außerdem liegt hier ein klassisches Vierfelder-Schema vor. Also: Chi^2-Test.

H_0: Es gibt keinen Unterschied in der Verteilung der Gumbatse zwischen Efftests und

Fiekors

H_1: Es gibt einen Unterschied in der Verteilung der Gumbatse zwischen Efftests und Fiekors

Die Formel:

$$\chi^2_{emp} = \frac{n \times (b \times c - a \times d)^2}{(a+b)\times(c+d)\times(a+c)\times(b+d)} = \frac{150 \times (55 \times 40 - 35 \times 20)^2}{90 \times 60 \times 75 \times 75}$$

$$\chi^2_{emp} = \frac{150 \times (1500)^2}{30375000} = \frac{150 \times 2250000}{30375000} = \frac{337500000}{30375000} = 11{,}111$$

Hier wurden die Randsummen in die Berechnung miteinbezogen, daher ergibt sich als Freiheitsgrad df = 1. Hiermit kann man nun aus Tabelle D den kritischen Chi^2-Wert für α =0,05 und df = 1 herauslesen:

$$\chi^2_{krit(\alpha=0{,}05;df=1)} = 3{,}841$$

Da der empirische Chi^2-Wert größer als der kritische Chi^2-Wert ist, hat man ein signifikantes Ergebnis, d.h. man entscheidet sich für H_1. Es gibt einen Unterschied in der Verteilung der Gumbatse zwischen Efftests und Fiekors. Wenn man sich jetzt noch die Tabelle ansieht, kann man auch sagen: Fiekors haben signfikant mehr Gumbatse als Efftests bzw. Efftests haben signifikant weniger Gumbatse.

zu b) Hier wird nach einem Zusammenhang gefragt, nach einer Korrelation. Für nominalskalierte Variablen gibt es die Vierfelder-(Phi-)Korrelation.

Die Formel:

$$Phi = \sqrt{\frac{\chi^2}{n}} = \frac{(b \times c - a \times d)}{\sqrt{(a+b)\times(c+d)\times(a+c)\times(b+d)}}$$

Da man schon den Chi^2-Wert berechnet hat, empfiehlt sich die erste Formel:

$$Phi = \sqrt{\frac{11{,}111}{150}} = \sqrt{0{,}074} = 0{,}272$$

Um den Phi-Koeffizienten interpretieren zu können, benötigt man noch den maximalen Phi-Wert unter Berücksichtigung der Randsummen. D.h. wie müßte die Tabelle aussehen, damit ein maximaler Wert herauskommt? So:

	Efftests	Fiekors	
Gumbatse	15	75	90
keine Gumbatse	60	0	60
	75	75	150

Diese Werte werden nun ganz einfach in die Formel eingesetzt:

$$Phi_{\max} = \frac{(b \times c - a \times d)}{\sqrt{(a+b)\times(c+d)\times(a+c)\times(b+d)}} = \frac{(75\times 60 - 15\times 0)}{\sqrt{90\times 60\times 75\times 75}}$$

$$Phi_{\max} = \frac{4500}{\sqrt{30375000}} = \frac{4500}{5511{,}352} = 0{,}816$$

Nun muß man nur noch den Phi-Wert durch Phi_{max} teilen, dann erhält man den sogenannten Äquivalenzkoeffizienten, den man ähnlich wie den Produkt-Moment-Korrelationskoeffizienten interpretieren kann:

$$Phi_{korr} = r_{Äqui} = \frac{Phi}{Phi_{\max}} = \frac{0{,}272}{0{,}816} = 0{,}333$$

Es besteht also eine mäßige Korrelation zwischen Efftests/Fiekorrs und Gumbatse/keine Gumbatse.

zu c) Hier sollen Mittelwerte verglichen werden. Das dafür zuständige Testverfahren ist der t-Test. Vor einem t-Test muß jedoch immer ein F-Test gerechnet werden.

Für den F-Test benötigt man die Varianzen, die man aus den Standardabweichungen errechnen kann:

$$S^2_{Fiekors} = S_{Fiekors} \times S_{Fiekors} = 4\times 4 = 16$$

$$S^2_{Efftests} = S_{Efftests} \times S_{Efftests} = 8\times 8 = 64$$

F-Test (Varianzenvergleich)

H_0: Die Varianzen sind gleich $\qquad H_0$: $\sigma_1^2 = \sigma_2^2$

H_1: Die Varianzen sind unterschiedlich $\qquad H_1$: $\sigma_1^2 \neq \sigma_2^2$

Die Formeln:

$$F_{emp} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \qquad \hat{\sigma}^2 = S^2 \times \frac{n}{n-1}$$

$$\hat{\sigma}^2_{Fiekors} = 16 \times \frac{13}{12} = 17{,}333$$

$$\hat{\sigma}^2_{Efftests} = 64 \times \frac{12}{11} = 69{,}818$$

Beim F-Test wird immer die größere Varianz auf den Bruchstrich (in den Zähler) gesetzt, die kleinere Varianz unter den Bruchstrich (in den Nenner).

$$F_{emp} = \frac{69{,}818}{17{,}333} = 4{,}028$$

Diesen empirischen F-Wert muß man nun mit dem kritischen F-Wert vergleichen. Dazu benötigt man Zähler und Nenner-Freiheitsgrade. Im Zähler steht die Varianz der Efftests mit n = 12 $\Rightarrow$ $df_{Zähler}$ = 11;

im Nenner steht die Varianz der Fiekors mit n = 13 $\Rightarrow$ df_{Nenner} = 12. Mit diesen beiden Freiheitsgraden geht man nun in Tabelle B und schlägt dort für α=0.05 den kritischen F-Wert nach:

$$F_{krit(\alpha=0.05; df_1=11, df_2=12)} = 2{,}72$$

Da der empirische F-Wert größer als der kritische F-Wert ist, handelt es sich um ein signifikantes Ergebnis, man entscheidet sich also für H_1, d.h. die Varianzen sind nicht gleich (homogen) sondern unterschiedlich (heterogen).

Da die Varianzen heterogen sind, muß der t-Test für heterogene Varianzen herangezogen werden:

t-Test für heterogene Varianzen (Mittelwertevergleich)

Die Frage war, ob die Efftests die **aufmerksameren** Beamten waren. Es handelt sich hierbei also um eine einseitige Hypothese (es wird nur nach einer Richtung gefragt, nämlich nach mehr). Allerdings, aufmerksamer heißt, daß sie schneller (in weniger Zeit) für den Besucher bereit waren.

H_0: Die Efftests sind nicht aufmerksamer als die Fiekors.

$$\mu_{Efftests} \geq \mu_{Fiekors}$$

H_1: Die Efftests sind aufmerksamer als die Fiekors.

$$\mu_{Efftests} < \mu_{Fiekors}$$

Die Formeln:

$$t_{emp} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_d} \qquad \hat{\sigma}_d = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} \qquad df = \frac{n_1 + n_2}{2}$$

$$\hat{\sigma}_d = \sqrt{\frac{17{,}333}{13} + \frac{69{,}818}{12}} = \sqrt{1{,}333 + 5{,}818} = \sqrt{7{,}151} = 2{,}674$$

$$t_{emp} = \frac{32 - 31}{2{,}674} = \frac{1}{2{,}674} = 0{,}374$$

Dieser empirische t-Wert muß nun mit einem kritischen t-Wert verglichen werden. Um den kritischen t-Wert aus Tabelle C ablesen zu können, benötigt man die Freiheitsgrade. Beim t-Test für heterogene Varianzen berechnen sich die Freiheitsgrade nach folgender

Formel:

$$df = \frac{n_1 + n_2}{2} = \frac{13 + 12}{2} = 12{,}5 \approx 13$$

Als kritischen t-Wert liest man bei α=0,05, einseitig, folgenden Wert ab:

$$t_{krit(\alpha=0.05;einseitig;df=13)} = 1{,}771$$

Da der kritische t-Wert größer als der empirische t-Wert ist, handelt es sich um ein nicht signifikantes Ergebnis, man entscheidet sich für H_0.

$t_{emp} < t_{krit} \Rightarrow$ nicht signifikant $\Rightarrow H_0$

Die Efftests sind - bezogen auf die Population - nicht aufmerksamer als die Fiekors.

zu d) Hier wird nach einem Zusammenhang gefragt. Beide Merkmale sind normalverteilt, d.h. sie sind (mindestens) intervallskaliert. Das hierfür zuständige Zusammenhangsmaß ist der Produkt-Moment-Korrelations-Koeffizient.

Dazu berechnet man am besten erst einmal den Mittelwert pro Variable und erstellt dann eine Tabelle.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{4+6+1+2+4+3+1+5+4}{9} = \frac{30}{9} = 3{,}333$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{20+25+12+13+23+18+4+20+19}{9} = \frac{154}{9} = 17{,}111$$

Pb-Nr.	x	x- $\bar{x}$	(x- $\bar{x}$)^2	y	y- $\bar{y}$	(y- $\bar{y}$)^2	(x- $\bar{x}$)×(y- $\bar{y}$)
1	4	0,667	0,445	20	2,889	8,346	1,927
2	6	2,667	7,113	25	7,889	62,236	21,040
3	1	-2,333	5,443	12	-5,111	26,122	11,924
4	2	-1,333	1,777	13	-4,111	16,900	5,480
5	4	0,667	0,445	23	5,889	34,680	3,928
6	3	-0,333	0,111	18	0,889	0,790	-0,296
7	1	-2,333	5,443	4	-13,111	171,898	30,588

8	5	1,667	2,779	20	2,889	8,346	4,816
9	4	0,667	0,445	19	1,889	3,568	1,260
			Σ = 24,001			Σ = 332,886	Σ = 80,667

$$S_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{24,001}{9} = 2,667 \Rightarrow S_x = 1,633$$

$$S_y^2 = \frac{\sum_{i=1}^{n}(y_i \bar{y})^2}{n} = \frac{332,886}{9} = 36,987 \Rightarrow S_y = 6,082$$

$$COV_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \times (y_i - \bar{y})}{n} = \frac{80,667}{9} = 8,963$$

$$r_{xy} = \frac{COV_{xy}}{S_x \times S_y} = \frac{8,963}{1,633 \times 6,082} = \frac{8,963}{9,932} = 0,902$$

$$r_{xy}^2 = r \times r = 0,902 \times 0,902 = 0,814$$

$r_{xy}^2 \times 100\% = 81,4\%$ erklärte Variation/Varianz

Es besteht ein hoher positiver Zusammenhang von r_{xy} = 0,902 zwischen der Dauer des Klatschens und der Menge an sichtbarem Pelz. Es bestehen 81,4% erklärte Varianz.

zu e) Hier ist danach gefragt, ob sich diese Korrelation auch auf die Gesamtbevölkerung hochrechnen läßt, d.h. ob diese Korrelation so groß ist, daß auch auf eine Korrelation in der Gesamtbevölkerung geschlossen werden kann. Der dafür zuständige Test ist der t-Test für Korrelationen.

Die Formel:

$$t_{emp} = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}} \qquad \text{df = n - 2}$$

Als Hypothesen müßte man schreiben:

H_0: In der Population besteht kein Zusammenhang H_0: $\rho = 0$

H_1: In der Population besteht ein Zusammenhang H_1: $\rho \neq 0$

Hier würde man errechnen:

$$t_{emp} = \frac{0,902 \times \sqrt{9-2}}{\sqrt{1-0,814}} = \frac{0,902 \times \sqrt{7}}{\sqrt{0,186}} = \frac{0,902 \times 2,646}{0,431} = 5,538$$

Als Freiheitsgrade errechnet man: df = n - 2 = 9 - 2 = 7

Als kritischen t-Wert liest man aus Tabelle C für df = 7 und α=0,05 zweiseitig (0,025 einseitig) den Wert:

$$t_{krit(df=7;\alpha=0,05,zweiseitig)} = 2,365$$

Da der empirische t-Wert größer als der kritische t-Wert ist, hätte man ein signifikantes Ergebnis, die Korrelation ist überzufällig von Null verschieden, man entscheidet sich für die Alternativhypothese H_1.

zu f) p(Socken verwechseln **und** Hörsaal verpassen **und** Faden nicht verlieren)= 0,5×0,8 ×0,5= **0,2**

zu g) p(Socken nicht verwechseln **und** Hörsaal nicht verpassen **und** Faden nicht verlieren) = 0,5×0,2×0,5= **0,05**

zu h) Als erstes muß man sich fragen: Was heißt mindestens dreimal? Mindestens dreimal heißt: dreimal oder viermal oder fünfmal. Dann muß man sich fragen: Wieviele Kombinationsmöglichkeiten gibt es jeweils für dreimal, viermal und fünfmal. Die Kombinationsmöglichkeiten errechnen sich über die Formel:

Anzahl Kombinationen = $\binom{n}{k} = \frac{n!}{k! \times (n-k)!}$

Das n bleibt immer gleich, nämlich fünf Tage, also n=5. Das k jedoch kann 3 oder 4 oder 5 sein.

Die Wahrscheinlichkeit dafür, daß alles schiefgeht, ist:

p(Socken verwechseln **und** Hörsaal verpassen **und** Faden verlieren)= 0,5×0,8×0,5 = **0,2**. Daraus folgt, die Wahrscheinlichkeit dafür, daß nicht *alles* schiefgeht, beträgt p=0,8. Nun aber zu den Kombinationen.

Der Fall k = 3:

$$\binom{n}{k} = \binom{5}{3} = \frac{5!}{3! \times (5-3)!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1} = 10$$

Die Wahrscheinlichkeit für **eine** dieser Kombinationen errechnet sich zu:

p = 0,2×0,2×0,2×0,8×0,8 = 0,00512

Die Gesamtwahrscheinlichkeit für drei unter fünf, egal welche Tage, berechnet sich jetzt als

p×Anzahl Kombinationen = 0,00512×10 = 0,0512

Der Fall k = 4:

$$\binom{n}{k} = \binom{5}{4} = \frac{5!}{4! \times (5-4)!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1 \times 1} = 5$$

Die Wahrscheinlichkeit für **eine** dieser Kombinationen errechnet sich zu:

p = 0,2×0,2×0,2×0,2×0,8 = 0,00128

Die Gesamtwahrscheinlichkeit für vier unter fünf, egal welche Tage, berechnet sich jetzt als

p×Anzahl Kombinationen = 0,00128×5 = 0,0064

Der Fall k = 5:

$$\binom{n}{k} = \binom{5}{5} = \frac{5!}{5! \times (5-5)!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{5 \times 4 \times 3 \times 2 \times 1 \times 1} = 1$$ (Achtung: 0! = 1 und nicht 0!!!!!)

Die Wahrscheinlichkeit für diese Kombinationen errechnet sich zu:

p = 0,2×0,2×0,2×0,2×0,2 = 0,00032

Diese Wahrscheinlichkeiten müssen nun noch addiert werden (es kann ja sein 3 oder 4 oder 5), um die Antwort für die Frage zu errechnen:

p(mindestens 3) = p(3)+p(4)+p(5) = 0,0512+0,0064+0,00032 = **0,05792**

Aufgabe 8:

Die Schwierigkeit bei dieser Aufgabe liegt eigentlich darin, daß man sich nicht verwirren lassen darf, sondern erst einmal aufschreibt, was alles gegeben wurde:

p(Kannibale)=0,8

p(nicht Kannibale)=0,2

p(freundlich)=0,1

p(unfreundlich)=0,9

p(ignorieren)=0,6

p(nicht ignorieren sondern Rum trinken)=0,4

p(freundlich und Rum trinkend und Nichtkannibale)=0,1×0,4×0,2 = 0,008. Hieraus folgt, daß die Wahrscheinlichkeit dafür, daß die Gegenwahrscheinlichkeit (entweder nicht freundlich oder nicht Rum trinkend oder Kannibale) p=0,992 ist.

Was heißt nun: Bei sechs höchstens einmal? Höchstens einmal heißt: Entweder einmal **oder** keinmal.

Die Kombinationen errechnen sich wieder einmal über die Formel:

Anzahl Kombinationen = $\binom{n}{k} = \frac{n!}{k! \times (n-k)!}$

Das n bleibt immer gleich, nämlich n=6. Das k kann entweder 1 oder 0 sein.

Der Fall k=1:

$$\binom{n}{k} = \binom{6}{1} = \frac{6!}{1! \times (6-1)!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{1 \times 5 \times 4 \times 3 \times 2 \times 1} = 6$$

Die Wahrscheinlichkeit für **eine** dieser Kombinationen errechnet sich zu:

p = 0,008×0,992×0,992×0,992×0,992×0,992 = 0,0053

Die Gesamtwahrscheinlichkeit für einen unter sechs, egal welcher, berechnet sich jetzt als

p×Anzahl Kombinationen = 0,0053×6 = 0,0318

Der Fall k = 0:

$$\binom{n}{k} = \binom{6}{0} = \frac{6!}{0! \times (6-0)!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{1 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} = 1$$ (Achtung: 0! = 1 und nicht 0!!!!!)

Die Wahrscheinlichkeit für diese Kombinationen errechnet sich zu:

p = 0,992×0,992×0,992×0,992×0,992×0,992 = 0,9529

Die einzelnen Wahrscheinlichkeiten müssen nun noch addiert werden:

p(höchstens 1)=p(1)+p(0)= 0,0318+0,9529= **0,9847**

D.h. die Wahrscheinlichkeit dafür, daß höchstens einmal ein freundlicher, Rum trinkender Nichtkannibale dabei ist, ist sehr hoch!

Die Wahrscheinlichkeit dafür, daß ein Typ nicht freundlich, nicht Rum trinkend und nicht

Nichtkannibal=Kannibale ist, beträgt p=0,992 (siehe oben). Die Gegenwahrscheinlichkeit hierfür beträgt also p=0,008.

Was heißt nun: Bei sechs höchstens einmal? Höchstens einmal heißt: Entweder einmal **oder** keinmal.

Die Kombinationen errechnen sich wieder einmal über die Formel:

Anzahl Kombinationen = $\binom{n}{k} = \frac{n!}{k! \times (n-k)!}$

Das n bleibt immer gleich, nämlich n=6. Das k kann entweder 1 oder 0 sein.

Der Fall k=1:

$$\binom{n}{k} = \binom{6}{1} = \frac{6!}{1! \times (6-1)!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{1 \times 5 \times 4 \times 3 \times 2 \times 1} = 6$$

Die Wahrscheinlichkeit für **eine** dieser Kombinationen errechnet sich zu:

p = 0,992×0,008×0,008×0,008×0,008×0,008 = 0,0000

Diese 0,0000 stimmen natürlich nicht ganz. Nur ist die Zahl so klein, daß sie schon sehr nahe an Null herangeht. Ebenso gilt dies für den Fall k=0 und für die Gesamtwahrscheinlichkeit.

Die Gesamtwahrscheinlichkeit für einen unter sechs, egal welcher, berechnet sich jetzt als

p×Anzahl Kombinationen = 0,0000×6 = 0,0000

Der Fall k = 0:

$$\binom{n}{k} = \binom{6}{0} = \frac{6!}{0! \times (6-0)!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{1 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} = 1$$ (Achtung: 0! = 1 und nicht 0!!!!!)

Die Wahrscheinlichkeit für diese Kombinationen errechnet sich zu:

p = 0,008×0,008×0,008×0,008×0,008×0,008 = 0,0000

Die einzelnen Wahrscheinlichkeiten müssen nun noch addiert werden:

p(höchstens 1)=p(1)+p(0)= 0,0000+0,0000= **0,0000**

D.h. die Wahrscheinlichkeit dafür, daß höchstens einmal ein nicht freundlicher, nicht Rum trinkender Kannibale dabei ist, ist extrem gering!

<u>Aufgabe 9:</u> Multiple Regressionsanalyse

a) Berechnung nach ALM:

$$b = (X'X)^{-1} \times X'y\,,\quad R^2 = \frac{b'(X'y) - n\bar{y}^2}{y'y - n\bar{y}^2}\,,\quad n = 6$$

<u>1. Schritt:</u> Berechnung von X'X

$$\begin{array}{c|c} & \begin{bmatrix} 1 & 14 & 6 \\ 1 & 14 & 4 \\ 1 & 10 & 6 \\ 1 & 8 & 8 \\ 1 & 4 & 8 \\ 1 & 12 & 5 \end{bmatrix} \\ \hline X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 14 & 14 & 10 & 8 & 4 & 12 \\ 6 & 4 & 6 & 8 & 8 & 5 \end{bmatrix} & \begin{bmatrix} 6 & 62 & 37 \\ 62 & 716 & 356 \\ 37 & 356 & 241 \end{bmatrix} \end{array}$$

$$\begin{bmatrix} 6 & 62 & 37 \\ 62 & 716 & 356 \\ 37 & 356 & 241 \end{bmatrix} = X'X = \begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 \end{bmatrix}$$

<u>2. Schritt:</u> Berechnung von X'y

Kai Budischewski: Statistik

$$\begin{array}{c|c} & \begin{vmatrix} 1 \\ 2 \\ 4 \\ 2 \\ 8 \\ 3 \end{vmatrix} = y \\ \hline X' = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 14 & 14 & 10 & 8 & 4 & 12 \\ 6 & 4 & 6 & 8 & 8 & 5 \end{vmatrix} & \begin{vmatrix} 20 \\ 166 \\ 133 \end{vmatrix} = X'y = \begin{vmatrix} \Sigma y \\ \Sigma x_1 y \\ \Sigma x_2 y \end{vmatrix} \end{array}$$

<u>3. Schritt:</u> Berechnung der Determinanten von X'X

$$\begin{array}{ccc|l} 6 & 62 & 37 & = 37 \cdot 716 \cdot 37 = 980204 \cdot (-1) \\ 62 & 716 & 356 & = 6 \cdot 356 \cdot 356 = 760416 \cdot (-1) \\ 37 & 356 & 241 & = 62 \cdot 62 \cdot 241 = 926404 \cdot (-1) \\ 6 & 62 & 37 & = 6 \cdot 716 \cdot 241 = 1035336 \\ 62 & 716 & 356 & = 62 \cdot 356 \cdot 37 = 816664 \\ & & & = 37 \cdot 62 \cdot 356 = 816664 \end{array}$$

= -980204 + (-760416) + (-926404) + 1035336 + 816664 + 816664

= <u>1640</u> = Determinante von X´X

<u>4. Schritt:</u> Berechnung der Kofaktorenmatrix:

$$K = \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix}$$

Erster Index: Zeile, zweiter Index: Spalte

$$K = \begin{vmatrix} \begin{vmatrix} 6 & 62 & 37 \\ 62 & 716 & 356 \\ 37 & 356 & 241 \end{vmatrix} & \begin{vmatrix} 6 & 62 & 37 \\ 62 & 716 & 356 \\ 37 & 356 & 241 \end{vmatrix} & |..| \\ |..| & |..| & |..| \\ |..| & |..| & |..| \end{vmatrix}$$

$$K = \begin{vmatrix} \begin{vmatrix} 716 & 356 \\ 356 & 241 \end{vmatrix} & -1\cdot \begin{vmatrix} 62 & 356 \\ 37 & 241 \end{vmatrix} & \begin{vmatrix} 62 & 716 \\ 37 & 356 \end{vmatrix} \\ -1\cdot \begin{vmatrix} 62 & 37 \\ 356 & 241 \end{vmatrix} & \begin{vmatrix} 6 & 37 \\ 37 & 241 \end{vmatrix} & -1\cdot \begin{vmatrix} 6 & 62 \\ 37 & 356 \end{vmatrix} \\ \begin{vmatrix} 62 & 37 \\ 716 & 356 \end{vmatrix} & -1\cdot \begin{vmatrix} 6 & 37 \\ 62 & 356 \end{vmatrix} & \begin{vmatrix} 6 & 62 \\ 62 & 716 \end{vmatrix} \end{vmatrix}$$

$$K = \begin{vmatrix} (716\cdot 241 - 356\cdot 356) & ((-1)\cdot(62\cdot 241 - 37\cdot 356)) & (62\cdot 356 - 37\cdot 716) \\ ((-1)\cdot(62\cdot 241 - 356\cdot 37)) & (6\cdot 241 - 37\cdot 37) & ((-1)\cdot(6\cdot 356 - 37\cdot 62)) \\ (62\cdot 356 - 716\cdot 37) & ((-1)\cdot(6\cdot 356 - 62\cdot 37)) & (6\cdot 716 - 62\cdot 62) \end{vmatrix}$$

Kofaktorenmatrix $K = \begin{vmatrix} 45820 & -1770 & -4420 \\ -1770 & 77 & 158 \\ -4420 & 158 & 452 \end{vmatrix}$

Inverse = $(X'X)^{-1}$

Inverse von X´X: $\frac{1}{\text{Determinante}} \cdot \text{Kofaktorenmatrix} = \frac{1}{1640} \cdot \begin{vmatrix} 45820 & -1770 & -4420 \\ -1770 & 77 & 158 \\ -4420 & 158 & 452 \end{vmatrix}$ Noch **nicht** ausrechnen!!!

5. Schritt: Berechnen der b-Gewichte: $b = (X'X)^{-1} \cdot X'y$

$$\begin{array}{l|l} & \begin{vmatrix} 20 \\ 166 \\ 133 \end{vmatrix} = X'y \\ \hline (X'X)^{-1} = \frac{1}{1640}\cdot \begin{vmatrix} 45820 & -1770 & -4420 \\ -1770 & 77 & 158 \\ -4420 & 158 & 452 \end{vmatrix} & \begin{vmatrix} 21170 \\ -0.978 \\ -1253 \end{vmatrix} \mathbf{= b\text{ - }Gewichte} \end{array}$$

6. Schritt: Berechnung von b'(X'y)

$$\begin{array}{ccc|l} & & & 20 \\ & & & 166 \\ & & & 133 \\ \hline 21170 & -0.978 & -1253 & 94.403 = b'(X'y) \end{array}$$

7. Schritt: Berechnung von y'y

$$\begin{array}{cccccc|l} & & & & & & 1 \\ & & & & & & 2 \\ & & & & & & 4 \\ & & & & & & 2 \\ & & & & & & 8 \\ & & & & & & 3 \\ \hline 1 & 2 & 4 & 2 & 8 & 3 & 98 = y'y \end{array}$$

8. Schritt: Berechnung von $\bar{y}^2$

$$\begin{array}{cccccc|l} & & & & & & 1 \\ & & & & & & 2 \\ & & & & & & 4 \\ & & & & & & 2 \\ & & & & & & 8 \\ & & & & & & 3 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 20 \; :6 = 3{,}333 = \bar{y} \Rightarrow \bar{y}^2 = 11{,}111 \end{array}$$

9. Schritt: Berechnung von R^2

$$R^2 = \frac{94{,}403 - 6 * 11{,}111}{98 - 6 * 11{,}111} = \frac{27{,}736}{31{,}333} = \underline{0{,}8852}$$

Ob dieser Wert signifikant ist, kann nach der Formel

$$F_{emp} = \frac{\left(R_U^2 - R_E^2\right) / df_h}{\left(1 - R_U^2\right) / df_e}$$

berechnet werden.

R_U^2: Determinatiosnkoeffizient des uneingeschränkten Modells

R_E^2: Determinationskoeffizient des eingeschränkten Modells

df_h: Hypothesenfreiheitsgrade; Anzahl der Null-gesetzten b-Gewichte; Zählerfreiheitsgrade df_1

df_e: Fehlerfreiheitsgrade; n - Anzahl der b-Gewichte inklusive b_0; Nennerfreiheitsgrade df_2

Die Hypothesen lauten:

$H_0 : b_1 = 0 und b_2 = 0$

Kai Budischewski: Statistik

$H_1 : b_1 \neq 0 und / oder b_2 \neq 0$

$R_U^2 = 0.8852$

$R_E^2 = 0$

$df_h = 2$

$df_e = 6 - 3 = 3$

$$F_{emp} = \frac{(0.8852 - 0)/2}{(1 - 0.8852)/3} = \frac{0.4426}{0.0383} = 11.56$$

Dieses F_{emp} wird mit einem F_{krit} verglichen:

$F_{krit(\alpha=0.05; df1=2; df2=3)} = 9.55$

$F_{krit} < F_{emp} \Rightarrow signifikant \Rightarrow H_1$, d.h. mindestens einer der Prädiktoren bietet eine signifikante Erklärung des Kriteriums

b) Wie gut ist das Modell, wenn nur die Variable Schnelligkeit berücksichtigt wird?

1. Schritt: Verkleinern der X'X-Matrix auf die interessierenden Variablen:

$$X'X = \begin{array}{c} \\ x_0 \\ x_1 \\ x_2 \end{array} \begin{array}{c} \begin{array}{ccc} x_0 & x_1 & x_2 \end{array} \\ \left| \begin{array}{ccc} 6 & 62 & \cancel{37} \\ 62 & 716 & \cancel{356} \\ \cancel{37} & \cancel{356} & \cancel{241} \end{array} \right| \end{array}$$

2. Schritt: Verkleinern der X'y-Matrix auf die interessierenden Variablen:

$$X'y = \begin{array}{c} \\ x_0 \\ x_1 \\ x_2 \end{array} \begin{array}{c} y \\ \left| \begin{array}{c} 20 \\ 166 \\ \cancel{133} \end{array} \right| \end{array}$$

3. Schritt: Berechnen der b-Gewichte

Berechnung nach ALM:

$$b = \left(X'X\right)^{-1} \times X'y$$

$$X'X = \begin{vmatrix} 6 & 62 \\ 62 & 716 \end{vmatrix}, \quad X'y = \begin{vmatrix} 20 \\ 166 \end{vmatrix}$$

3a) Berechnen der Determinanten von X'X

Determinante = 6*716 - 62*62 = 4296 - 3844 = 452

3b) Berechnen der Inversen von X'X

$$\text{Kofaktorenmatrix} = \begin{vmatrix} 716 & -62 \\ -62 & 6 \end{vmatrix}$$

$$\left(X'X\right)^{-1} = \frac{1}{452} \times \begin{vmatrix} 716 & -62 \\ -62 & 6 \end{vmatrix}$$

$$b = \frac{1}{452} \times \begin{vmatrix} 716 & -62 \\ -62 & 6 \end{vmatrix} \times \begin{vmatrix} 20 \\ 166 \end{vmatrix} = \begin{vmatrix} 8.913 \\ -0.540 \end{vmatrix} = b - \text{Gewichte}$$

Berechnung der b-Gewichte nach Gauß'schem Algorithmus:

1. Schritt: Bilden von zwei Gleichungen anhand der verkleinerten X'X und X'y -Matrix.

$$\begin{array}{lrcrcr} \text{I.} & 6b_0 & + & 62b_1 & = & 20 \\ \text{II.} & 62b_0 & + & 716b_1 & = & 166 \end{array}$$

2. Schritt: Schrittweises Erweitern und Subtrahieren:

$$\begin{array}{lrcrcr|l} \text{I.} & 6b_0 & + & 62b_1 & = & 20 & \times 62 \\ \text{II.} & 62b_0 & + & 716b_1 & = & 166 & \times 6 \end{array}$$

$$\begin{array}{lrcrcr|l} \text{I.} & 372b_0 & + & 3844b_1 & = & 1240 & \\ \text{II.} & 372b_0 & + & 4296b_1 & = & 996 & - \end{array}$$

Kai Budischewski: Statistik

$$\begin{array}{lrcrcr|l} \text{III.} & 0b_0 & - & 452b_1 & = & 244 & \end{array}$$

3.1. Schritt: Berechnung von b_1

$$\begin{array}{rcr|l} -452b_1 & = & 244 & :(-452) \\ b_1 & = & -0.540 & - \end{array}$$

3.2. Schritt: Berechnung von b_0 durch Einsetzen von b_1 in Gleichung 1

$$\begin{array}{lrcrcr|l} \text{I.} & 6b_0 & + & 62\times(-0.540) & = & 20 & \\ \text{I.} & 6b_0 & - & 33.48 & = & 20 & +33.48 \\ & 6b_0 & & & = & 53.48 & :6 \\ & b_0 & & & = & 8.913 & \end{array}$$

4. Schritt: Berechnung von b'(X'y)

$$\begin{vmatrix} 8.912 & -0.540 \end{vmatrix} \times \begin{vmatrix} 20 \\ 166 \end{vmatrix} = 88.6 = b'\left(X'y\right)$$

<u>5. Schritt:</u> Berechnung von y'y

y'y = 98

Berechnung siehe oben

<u>6. Schritt:</u> Berechnung von $\bar{y}^2$

$\bar{y}^2 = 11.111 \Rightarrow n\bar{y}^2 = 66.667$

Berechnung siehe oben

<u>7. Schritt:</u> Berechnung von R^2

$$R^2 = \frac{b'\left(X'y\right) - n\bar{y}^2}{y'y - n\bar{y}^2} = \frac{88.6 - 66.667}{98 - 66.667} = \frac{21933}{31333} = 0.700$$

Nun stellt sich die Frage, ob dieser Determinationskoeffizient signifikant ist. Zur Prüfung auf Signifikanz wird der F-Test herangezogen:

Die Hypothesen lauten:

$H_0 : b_1 = 0$ Die Variable Schnelligkeit leistet keinen Beitrag zur Vorhersage der Häuptlingsqualität

$H_1 : b_1 \neq 0$ Die Variable Schnelligkeit leistet einen von Null verschiedenen Beitrag zur Vorhersage der Häuptlingsqualität.

$R_U^2 = 0.700$

$R_E^2 = 0$

$df_h = 1$

$df_e = 6 - 2 = 4$

$$F_{emp} = \frac{(0.700 - 0) / 1}{(1 - 0.700) / 4} = \frac{0.700}{0.075} = 9.333$$

Dieses F_{emp} wird mit einem F_{krit} verglichen:

$F_{krit(\alpha=0.05;df1=1;df2=4)} = 7.71$

$F_{krit} < F_{emp} \Rightarrow$ signifikant $\Rightarrow H_1$, d.h. der Prädiktor Schnelligkeit alleine erklärt signifikant das Kriterium

c) Wie gut ist das Modell, wenn nur die Variable Weisheit berücksichtigt wird?

1. Schritt: Verkleinern der X'X-Matrix auf die interessierenden Variablen:

$$X'X = \begin{array}{c c} & \begin{matrix} x_0 & x_1 & x_2 \end{matrix} \\ \begin{matrix} x_0 \\ x_1 \\ x_2 \end{matrix} & \begin{vmatrix} 6 & \cancel{62} & 37 \\ \cancel{62} & \cancel{716} & \cancel{356} \\ 37 & \cancel{356} & 241 \end{vmatrix} \end{array}$$

Kai Budischewski: Statistik

2. Schritt: Verkleinern der X'y-Matrix auf die interessierenden Variablen:

$$X'y = \begin{array}{c c} & y \\ \begin{matrix} x_0 \\ x_1 \\ x_2 \end{matrix} & \begin{vmatrix} 20 \\ \cancel{166} \\ 133 \end{vmatrix} \end{array}$$

3. Schritt: Berechnen der b-Gewichte

Berechnung nach ALM:

$$b = (X'X)^{-1} \times X'y$$

$$X'X = \begin{vmatrix} 6 & 37 \\ 37 & 241 \end{vmatrix}, \quad X'y = \begin{vmatrix} 20 \\ 133 \end{vmatrix}$$

3a) Berechnen der Determinanten von X'X

Determinante = 6×241 - 37×37 = 1446 - 1369 = 77

3b) Berechnen der Inversen von X'X

$$\text{Kofaktorenmatrix} = \begin{vmatrix} 241 & -37 \\ -37 & 6 \end{vmatrix}$$

$$(X'X)^{-1} = \frac{1}{77} \times \begin{vmatrix} 241 & -37 \\ -37 & 6 \end{vmatrix}$$

$$b = \frac{1}{77} \times \begin{vmatrix} 241 & -37 \\ -37 & 6 \end{vmatrix} \times \begin{vmatrix} 20 \\ 133 \end{vmatrix} = \begin{vmatrix} -1312 \\ .753 \end{vmatrix} = b - \text{Gewichte}$$

4. Schritt: Berechnung von b'(X'y)

$$\begin{array}{cc|l} & & 20 \\ & & 133 \\ \hline -1312 & .753 & 73.948 = b'(X'y) \end{array}$$

<u>5. Schritt:</u> Berechnung von y'y

y'y = 98

Berechnung siehe oben

<u>6. Schritt:</u> Berechnung von $\bar{y}^2$

$\bar{y}^2 = 11.111 \Rightarrow n\,\bar{y}^2 = 66.667$

Berechnung siehe oben

<u>7. Schritt:</u> Berechnung von R^2

$$R^2 = \frac{b'(X'y) - n\bar{y}^2}{y'y - n\bar{y}^2} = \frac{73.948 - 66.667}{98 - 66.667} = \frac{7.281}{31.333} = 0.232$$

Nun stellt sich die Frage, ob dieser Determinationskoeffizient signifikant ist. Zur Prüfung auf Signifikanz wird der F-Test herangezogen:

Die Hypothesen lauten:

$H_0: b_2 = 0$

$H_1: b_2 \neq 0$

$R^2_U = 0.232$

$R^2_E = 0$

$df_h = 1$

$df_e = 6 - 2 = 4$

$$F_{emp} = \frac{(0.232 - 0)/1}{(1 - 0.232)/4} = \frac{0.232}{0.192} = 1208$$

Dieses F_{emp} wird mit einem F_{krit} verglichen:

$F_{krit(\alpha=0.05;df1=1;df2=4)} = 7.71$

$F_{emp} < F_{krit} \Rightarrow \text{nicht signifikant} \Rightarrow H_0$, d.h. der Prädiktor Weisheit alleine erklärt das Kriterium nicht signifikant.

d) Bringt die Hinzunahme der Variablen Weisheit eine signifikante Verbesserung?

Zur Abklärung dieser Frage wird wieder der F-Test verwendet, allerdings mit folgenden Hypothesen und Werten:

$H_0: b_2 = 0; b_1 = \text{beliebig}$

$H_1: b_2 \neq 0; b_1 = \text{beliebig}$

$R^2_U = 0.8852$

$R^2_E = 0.700$

$df_h = 1$

$df_e = 6 - 3 = 3$

$$F_{emp} = \frac{(0.8852 - 0.700)/1}{(1 - 0.8852)/3} = 4.839$$

$F_{krit(\alpha=0.05;df1=1;df2=3)} = 10.1$

$F_{emp} < F_{krit} \Rightarrow \text{nicht signifikant} \Rightarrow H_0$, d.h. die Hinzunahme der Variablen Weisheit bringt keine signifikante Verbesserung des Modells.

Kai Budischewski: Statistik

e) Bringt die Hinzunahme der Variablen Schnelligkeit eine signifikante Verbesserung?

Zur Abklärung dieser Frage wird wieder der F-Test verwendet, allerdings mit folgenden Hypothesen und Werten:

H_0: $b_1 = 0$; $b_2 =$ beliebig

H_1: $b_1 \neq 0$; $b_2 =$ beliebig

$R_U^2 = 0.8852$

$R_E^2 = 0.232$

$df_h = 1$

$df_e = 6 - 3 = 3$

$$F_{emp} = \frac{(0.8852 - 0.232)/1}{(1 - 0.8852)/3} = 17.05$$

$F_{krit(\alpha=0.05;df1=1;df2=3)} = 10.1$

$F_{krit} < F_{emp} \Rightarrow$ signifikant $\Rightarrow H_1$, d.h. die Hinzunahme der Variablen Schnelligkeit bringt eine signifikante Verbesserung des Modells.

f) Welche Häuptlingsqualität erzielt ein Bewerber mit einem Schnelligkeitswert von 4 Tage und einem Weisheitswert von 9? Bitte noch ein 95%-Konfidenzintervall berechnen.

Ausgehend von den b-Gewichten ergibt sich folgende allgemeine Gleichung:

$y = 21170 + (-0.978) \times x_1 + (-1253) \times x_2$

$x_1 = 4$, $x_2 = 9$

$y = 21170 + (-0.978) \times 4 + (-1253) \times 9$

$y = 21170 - 3.912 - 11277$

y = 5.981

Ein Bewerber mit den oben angeführten Werten hätte eine Häuptlingsqualität von 5.981 Punkten.

Zur Berechnung des Konfidenzintervalles:

Allgemeine Formel: $y - t_{\alpha/2} \times s_e \leq y \leq y + t_{\alpha/2} \times s_e$

<u>1. Schritt:</u> Berechnung von s_e

$$s_e = \sqrt{\frac{Q_e}{df_e}}$$

$Q_e = Q_t - Q_d$

$Q_t = y'y - n\bar{y}^2 = 31.333$

$Q_d = b'(X'y) - n\bar{y}^2 = 27.736$

$Q_e = 31.333 - 27.736 = 3.597$

$df_e = n - Anzahl\ Parameterrestriktionen = 6 - 3 = 3$

$$s_e = \sqrt{\frac{3.597}{3}} = 1.095$$

<u>2. Schritt:</u> Heraussuchen des t-Wertes aus Tabelle

$\alpha = 0.05$, $t_{(\alpha/2=0.025;\,df=3)} = 3.182$

<u>3. Schritt:</u> Berechnen der Grenzen des Konfidenzintervalles

$5.981 - 3.182 \times 1095 \leq y \leq 5.981 + 3.182 \times 1095$

2.497 ≤ y ≤ 9.465

Das Konfidenzintervall umfaßt den Bereich von 2.497 bis 9.465. Innerhalb dieser Grenzen kommt der wahre Wert des Bewerbes mit einer gewissen Wahrscheinlichkeit zum liegen.

Aufgabe 10: Multiple Regressionsanalyse

$$X = \begin{vmatrix} 1 & 19 & 2 \\ 1 & 14 & 6 \\ 1 & 11 & 4 \\ 1 & 9 & 16 \\ 1 & 3 & 18 \\ 1 & 5 & 19 \\ 1 & 8 & 7 \\ 1 & 10 & 3 \\ 1 & 13 & 2 \\ 1 & 17 & 1 \end{vmatrix}, \quad X' = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 19 & 14 & 11 & 9 & 3 & 5 & 8 & 10 & 13 & 17 \\ 2 & 6 & 4 & 16 & 18 & 19 & 7 & 3 & 2 & 1 \end{vmatrix} \quad y = \begin{vmatrix} 15 \\ 16 \\ 9 \\ 5 \\ 2 \\ 4 \\ 6 \\ 9 \\ 11 \\ 17 \end{vmatrix}$$

$$X'X = \begin{vmatrix} 10 & 109 & 78 \\ 109 & 1415 & 588 \\ 78 & 588 & 1060 \end{vmatrix}, \quad X'y = \begin{vmatrix} 94 \\ 1249 \\ 462 \end{vmatrix} \quad K = \begin{vmatrix} 1154156 & -69676 & -46278 \\ -69676 & 4516 & 2622 \\ -46278 & 2622 & 2269 \end{vmatrix}$$

Det.(X'X) = 337192, $\quad (X'X)^{-1} = \frac{1}{337192} \times \begin{vmatrix} 1154156 & -69676 & -46278 \\ -69676 & 4516 & 2622 \\ -46278 & 2622 & 2269 \end{vmatrix}$

a) $b = (X'X)^{-1}X'y = \begin{vmatrix} 0.252 \\ 0.897 \\ -0.080 \end{vmatrix}$

b) $b'X'y = 1107.081$, $\quad y'y = y^2 = 1134$, $\quad \bar{y} = 9.4 \to \bar{y}^2 = 88.36 \to n\bar{y}^2 = 883.6$

$$R^2 = \frac{1107.081 - 883.6}{1134 - 883.6} = 0.892$$

$F_{emp} = \frac{0.892/2}{(1-0.892)/7} = \frac{0.446}{0.015} = 29.733$, $\quad F_{krit} = 4.74 \to$ signifikant

c) $R^2 = 0.885$

$F_{emp} = \frac{0.885/1}{(1-0.885)/8} = 61565$, $\quad F_{krit} = 5.32 \to$ signifikant

d) $R^2 = 0.649$

$F_{emp} = \frac{0.649/1}{(1-0.649)/8} = 14.792$, $\quad F_{krit} = 5.32 \to$ signifikant

e) $F_{emp} = \frac{(0.892-0.885)/1}{(1-0.892)/7} = 0.453$, $\quad F_{krit} = 5.59 \to$ nicht signifikant

f) $F_{emp} = \frac{(0.892-0.649)/1}{(1-0.892)/7} = 15.75$, $\quad F_{krit} = 5.59 \to$ signifikant

g) $y = 0.252 + 0.897 \times 12 - 0.080 \times 8 = 10.376$

$Q_e = Q_t - Q_d = 250.4 - 223.481 = 26.919$

$df_e = 7$

$s_e = \sqrt{\frac{26.919}{7}} = 1961$, $\quad t = 2.365$

$10.376 - 1.961 \times 2.365 \le y \le 10.376 + 1.961 \times 2.365$

$5.738 \le y \le 15.013$

Aufgabe 11: Varianzanalyse

a) Globale Effekte?

$$X'X = 3 \times \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}, \quad (X'X)^{-1} = \frac{1}{3} \times \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

Kai Budischewski: Statistik

$$y=\begin{vmatrix}10\\12\\14\\16\\12\\10\\15\\18\\14\\17\\15\\13\\18\\20\\19\\23\\12\\24\end{vmatrix},\qquad X'y=\begin{vmatrix}36\\38\\47\\45\\57\\59\end{vmatrix},\qquad b=\begin{vmatrix}12\\12.67\\15.67\\15\\19\\19.67\end{vmatrix},\qquad y'y = 4702,\qquad \overline{y}=15.67$$

$\overline{y}^2 = 245.5489 \rightarrow n\overline{y}^2 = 18 \times 245.5489 = 4419.8802$

$b'X'y = 3\times(12^2+12.67^2+15.67^2+15^2+19^2+19.67^2)=4568.9601$

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

$Q_t = 4702 - 4419.8802 = 282.1198$

$Q_d = 4568.9601 - 4419.8802 = 149.0799 = Q_d$

$Q_e = Q_t - Q_d = 282.1198 - 149.0799 = 133.04$

df_h = Anzahl Parameterrestriktionen (Gleichzeichen in der Nullhypothese) = 5

df_e = Anzahl Vpn minus Anzahl Zellen = 18 - 6 = 12

$$F_{emp}=\frac{Q_h\,/\,df_h}{Q_e\,/\,df_e}=\frac{149.0799\,/\,5}{133.04\,/\,12}=2.689$$

$F_{krit(\alpha=0.05;df1=5;df2=12)} = 3.11 \rightarrow$ nicht signifikant, H_0, es gibt keinen globalen Effekt.

b) Haupteffekt Faktor A (Alterskategorie)?

Kontrastmatrix $C=\begin{vmatrix}1 & -1 & 1 & -1 & 1 & -1\end{vmatrix}$

$$Cb = -0.67,\qquad C(X'X)^{-1}=\begin{vmatrix}1 & -1 & 1 & -1 & 1 & -1\end{vmatrix}\begin{vmatrix}\frac{1}{3} & 0 & 0 & 0 & 0 & 0\\ 0 & \frac{1}{3} & 0 & 0 & 0 & 0\\ 0 & 0 & \frac{1}{3} & 0 & 0 & 0\\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0\\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0\\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3}\end{vmatrix}=\begin{vmatrix}\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} & -\frac{1}{3}\end{vmatrix}$$

$$C(X'X)^{-1}C'=\begin{vmatrix}\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} & -\frac{1}{3}\end{vmatrix}\begin{vmatrix}1\\-1\\1\\-1\\1\\-1\end{vmatrix}=\frac{6}{3}=2,\qquad \left(C(X'X)^{-1}C'\right)^{-1}=\frac{1}{2}$$

$Cb(C(X'X)^{-1}C')^{-1}Cb = 0.67\times0.5\times0.67 = 0.056 = Q_h$

H_0: $\mu_1 + \mu_3 + \mu_5 = \mu_2 + \mu_4 + \mu_6$

df_h = Anzahl Parameterrestriktionen (Gleichzeichen in der Nullhypothese) = 1

df_e = Anzahl Vpn minus Anzahl Zellen = 18 - 6 = 12

$$F_{emp}=\frac{Q_h\,/\,df_h}{Q_e\,/\,df_e}=\frac{0.056\,/\,1}{133.04\,/\,12}=0.005$$

$F_{krit(\alpha=0.05;df1=1;df2=12)} = 4.75 \rightarrow$ nicht signifikant, H_0, es gibt keinen Haupteffekt des Faktors Alterskategorien.

c) Haupteffekt Faktor B (Taschenrechnermodell)?

Kontrastmatrix $C = \begin{vmatrix} 2 & 2 & -1 & -1 & -1 & -1 \\ 0 & 0 & 1 & 1 & -1 & -1 \end{vmatrix}$

$Cb = \begin{vmatrix} -20 \\ -8 \end{vmatrix}$, $C(X'X)^{-1} = \begin{vmatrix} 2 & 2 & -1 & -1 & -1 & -1 \\ 0 & 0 & 1 & 1 & -1 & -1 \end{vmatrix} \cdot \begin{vmatrix} \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} \end{vmatrix} = \begin{vmatrix} \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{vmatrix}$

$C(X'X)^{-1}C' = \begin{vmatrix} \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{vmatrix} \cdot \begin{vmatrix} 2 & 0 \\ 2 & 0 \\ -1 & 1 \\ -1 & 1 \\ -1 & -1 \\ -1 & -1 \end{vmatrix} = \begin{vmatrix} \frac{12}{3} & 0 \\ 0 & \frac{4}{3} \end{vmatrix}$, $\left(C(X'X)^{-1}C'\right)^{-1} = \begin{vmatrix} \frac{3}{12} & 0 \\ 0 & \frac{3}{4} \end{vmatrix}$

$Cb'\left(C(X'X)^{-1}C'\right)^{-1} = \begin{vmatrix} -20 & -8 \end{vmatrix} \cdot \begin{vmatrix} \frac{3}{12} & 0 \\ 0 & \frac{3}{4} \end{vmatrix} = \begin{vmatrix} -5 & -6 \end{vmatrix}$ $Cb'\left(C(X'X)^{-1}C'\right)^{-1}Cb = \begin{vmatrix} -5 & -6 \end{vmatrix} \cdot \begin{vmatrix} -20 \\ -8 \end{vmatrix} = 148$

$Cb(C(X'X)^{-1}C')^{-1}Cb = 148 = Q_h$

H_0: $2\times\mu_1 + 2\times\mu_2 = \mu_3 + \mu_4 + \mu_5 + \mu_6$ und $\mu_3 + \mu_4 = \mu_5 + \mu_6$

Kai Budischewski: Statistik

df_h = Anzahl Parameterrestriktionen (Gleichzeichen in der Nullhypothese) = 2

df_e = Anzahl Vpn minus Anzahl Zellen = 18 - 6 = 15

$$F_{emp} = \frac{Q_h / df_h}{Q_e / df_e} = \frac{148 / 2}{133.04 / 12} = 6.675$$

$F_{krit(\alpha=0.05;df1=2;df2=12)} = 3.89 \rightarrow$ signifikant, H_1, es gibt einen Haupteffekt des Faktors Taschenrechnermodell.

d) Wechselwirkungseffekte?

Das Erstellen dieser Kontrastmatrix ist wirklich etwas kompliziert. Was soll eigentlich getestet werden? Am besten man malt in das Versuchsdesign Striche ein:

	Stastik 5000	Stastik 2000	Stastik 0815
Greise	10 12 14 $\boxed{1}$	15 18 14 $\boxed{3}$	18 20 19 $\boxed{5}$
Jungspunde	16 12 10 $\boxed{2}$	17 15 13 $\boxed{4}$	23 12 24 $\boxed{6}$

Zelle 1 gegen Zelle 4 und Zelle 5 sowie Zelle 2 gegen Zelle 3 und Zelle 6.

Außerdem Zelle 3 gegen Zelle 6 und Zelle 4 gegen Zelle 5. Die Minuszeichen werden nun gegenläufig verteilt. 2×Zelle 1 - Zelle 4 - Zelle 5 sowie -2×Zelle 2 + Zelle 3 + Zelle 6

Und weiter: Zelle 3 - Zelle 6 und -Zelle 4 + Zelle 5

Wie gestaltet sich das nun in der Kontrastmatrix? So:

$$C = \begin{vmatrix} 2 & -2 & 1 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{vmatrix}$$

Die Gestaltung der Kontrastmatrix ist das Schwierigste an der Berechnung der Wechselwirkung.

$$Cb = \begin{vmatrix} 0 \\ 134 \end{vmatrix}, \qquad C(X'X)^{-1} = \begin{array}{c|c} & \begin{vmatrix} \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} \end{vmatrix} \\ \hline \begin{vmatrix} 2 & -2 & 1 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{vmatrix} & \begin{vmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{vmatrix} \end{array}$$

$$C(X'X)^{-1}C' = \begin{array}{c|c} & \begin{vmatrix} 2 & 0 \\ -2 & 0 \\ 1 & 1 \\ -1 & -1 \\ -1 & -1 \\ 1 & 1 \end{vmatrix} \\ \hline \begin{vmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{vmatrix} & \begin{vmatrix} \frac{12}{3} & 0 \\ 0 & \frac{4}{3} \end{vmatrix} \end{array}, \qquad \left(C(X'X)^{-1}C'\right)^{-1} = \begin{vmatrix} \frac{3}{12} & 0 \\ 0 & \frac{3}{4} \end{vmatrix}$$

$$Cb'\left(C(X'X)^{-1}C'\right)^{-1} = \begin{array}{c|c} & \begin{vmatrix} \frac{3}{12} & 0 \\ 0 & \frac{3}{4} \end{vmatrix} \\ \hline \begin{vmatrix} 0 & 134 \end{vmatrix} & \begin{vmatrix} 0 & 1 \end{vmatrix} \end{array} \qquad Cb'\left(C(X'X)^{-1}C'\right)^{-1}Cb = \begin{array}{c|c} & \begin{vmatrix} 0 \\ 134 \end{vmatrix} \\ \hline \begin{vmatrix} 0 & 1 \end{vmatrix} & 134 \end{array}$$

$Cb(C(X'X)^{-1}C')^{-1}Cb = 1.34 = Q_h$

df_h = Anzahl Parameterrestriktionen (Gleichzeichen in der Nullhypothese) = 2

df_e = Anzahl Vpn minus Anzahl Zellen = 18 - 6 = 12

$$F_{emp} = \frac{Q_h / df_h}{Q_e / df_e} = \frac{134 / 2}{133.04 / 12} = 0.060$$

$F_{krit(\alpha=0.05;df1=2;df2=12)} = 3.89 \rightarrow$ nicht signifikant, H_0, es gibt keine Wechselwirkung zwischen

Kai Budischewski: Statistik

den Faktoren.

Aufgabe 12: Varianzanalyse

a) Globale Effekte?

$$X'y = \begin{vmatrix} 90 \\ 45 \\ 30 \\ 60 \\ 60 \\ 15 \end{vmatrix}, \qquad b = \begin{vmatrix} 30 \\ 15 \\ 10 \\ 20 \\ 20 \\ 5 \end{vmatrix}, \qquad y'y = 6478, \qquad \overline{y} = 16.67$$

$b'X'y = 6150$, $\quad \overline{y}^2 = 277.890 \rightarrow n \times \overline{y}^2 = 5002.02$

$Q_t = 6478 - 5002.02 = 1475.98$

$Q_d = 6150 - 5002.02 = 1147.98$

$Q_e = 1475.98 - 1147.98 = 328$

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

df_h = Anzahl Parameterrestriktionen (Gleichzeichen in der Nullhypothese) = 5

df_e = Anzahl Vpn minus Anzahl Zellen = 18 - 6 = 12

$$F_{emp} = \frac{Q_h / df_h}{Q_e / df_e} = \frac{1147.98 / 5}{328 / 12} = 8.340$$

$F_{krit(\alpha=0.05;df1=5;df2=12)} = 3.11 \rightarrow$ signifikant, H_1, es gibt einen globalen Effekt.

b) Haupteffekt Faktor A, Theoretische Ausrichtung

Kontrastmatrix $C = \begin{vmatrix} 1 & -1 & 1 & -1 & 1 & -1 \end{vmatrix}$

$Cb = 20$, $\qquad \left(C(X'X)^{-1}C'\right)^{-1} = \frac{1}{2}$

$Q_h = Cb'(C(X'X)^{-1}C')^{-1})Cb = 20 \times 0.5 \times 20 = 200$

H_0: $\mu_1 + \mu_3 + \mu_5 = \mu_2 + \mu_4 + \mu_6$

df_h = Anzahl Parameterrestriktionen (Gleichzeichen in der Nullhypothese) = 1

df_e = Anzahl Vpn minus Anzahl Zellen = 18 - 6 = 12

$$F_{emp} = \frac{Q_h / df_h}{Q_e / df_e} = \frac{200 / 1}{328 / 12} = 7.317$$

$F_{krit(\alpha=0.05;df1=1;df2=12)} = 4.75 \rightarrow$ signifikant, H_1, es gibt einen Haupteffekt des Faktors Theoretische Ausrichtung.

c) Haupteffekt Faktor B, Theoretische Ausrichtung

Kontrastmatrix $C = \begin{vmatrix} 2 & 2 & -1 & -1 & -1 & -1 \\ 0 & 0 & 1 & 1 & -1 & -1 \end{vmatrix}$

$$Cb = \begin{vmatrix} 35 \\ 5 \end{vmatrix}, \qquad \left(C(X'X)^{-1}C'\right)^{-1} = \begin{vmatrix} \frac{3}{12} & 0 \\ 0 & \frac{3}{4} \end{vmatrix}$$

$Q_h = Cb'(C(X'X)^{-1}C')^{-1})Cb = 325$

H_0: $2\times\mu_1 + 2\times\mu_2 = \mu_3 + \mu_4 + \mu_5 + \mu_6$ und $\mu_3 + \mu_4 = \mu_5 + \mu_6$

df_h = Anzahl Parameterrestriktionen (Gleichzeichen in der Nullhypothese) = 2

df_e = Anzahl Vpn minus Anzahl Zellen = 18 - 6 = 15

$$F_{emp} = \frac{Q_h / df_h}{Q_e / df_e} = \frac{325 / 2}{328 / 12} = 5.945$$

$F_{krit(\alpha=0.05;df1=2;df2=12)} = 3.89 \rightarrow$ signifikant, H_1, es gibt einen Haupteffekt des Faktors Komplexitätsgrad.

d) Wechselwirkungseffekte?

$$C = \begin{vmatrix} 2 & -2 & 1 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{vmatrix}$$

$$Cb = \begin{vmatrix} 5 \\ -25 \end{vmatrix}, \qquad \left(C(X'X)^{-1}C'\right)^{-1} = \begin{vmatrix} \frac{3}{12} & 0 \\ 0 & \frac{3}{4} \end{vmatrix}$$

$Cb(C(X'X)^{-1}C')^{-1}Cb = 475 = Q_h$

df_h = Anzahl Parameterrestriktionen (Gleichzeichen in der Nullhypothese) = 2

df_e = Anzahl Vpn minus Anzahl Zellen = 18 - 6 = 12

$$F_{emp} = \frac{Q_h / df_h}{Q_e / df_e} = \frac{475 / 2}{328 / 12} = 8.689$$

$F_{krit(\alpha=0.05;df1=2;df2=12)} = 3.89 \rightarrow$ signifikant, H_1, es gibt eine Wechselwirkung zwischen den Faktoren.