Studienbrief

SRH Fernhochschule – The Mobile University

STATISTIK

Titel-Nr. 0699-04

SRH FERNHOCHSCHULE
The Mobile University

1. Auflage Dezember 2009
2. Auflage Mai 2012
3. Auflage Januar 2013
4. Auflage Mai 2016

© 2016 SRH Fernhochschule
The Mobile University
Staatlich anerkannte Hochschule der SRH Hochschulen GmbH

Lange Straße 19, 88499 Riedlingen
Telefon: 07371 9315-0, Telefax: 07371 9315-15

Das Werk einschließlich aller Abbildungen ist urheberrechtlich geschützt.
Jede Verwertung außerhalb der Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Bearbeitung in elektronischen Systemen.

Autoren

Kai Budischewski
Diplom-Psychologe

Ausbildung

- Abgeschlossenes Studium der Psychologie an der Goethe-Universität Frankfurt am Main
- Promotion in theoretischer Medizin am Universitätsklinikum Frankfurt am Main

Berufspraxis

- Wissenschaftlicher Mitarbeiter an der Klinik für Strahlentherapie und Radioonkologie, Frankfurt am Main
- Wissenschaftlicher Mitarbeiter an der Abteilung für Medizinische Psychologie und Medizinische Soziologie der Gutenberg-Universität Mainz
- Wissenschaftlicher Mitarbeiter am Institut für Medizinische Psychologie der Universität Rostock
- Seit 2006 Dozent an der SRH Hochschule Heidelberg

Prof. Dr. Frederik Ornau
Kapitel 2.1, 2.2, 3.2, 3.3, 3.4, 3.6

10/1993 - 09/1995	Studium der Mathematik an der Universität des Saarlandes
09/1995	Vordiplom in Mathematik
10/1995 - 04/1998	Studium der Betriebswirtschaftslehre an der Universität des Saarlandes mit Schwerpunkt in Statistik, Operations Research und Ökonometrie, Diplomarbeit am Lehrstuhl für Statistik und Ökonometrie, Thema: „Diagnose und Modellierung von Strukturbrüchen“
04/1998	Abschluss zum Diplom-Kaufmann
02/2005	Promotion zum Dr. rer. pol. an der Universität des Saarlandes, Thema der Dissertation: „Statistische Eigenschaften von Prozessen mit autoregressiver bedingter Wartezeit“
07/1998 - 02/2005	Wissenschaftlicher Mitarbeiter am Lehrstuhl für Statistik und Ökonometrie, Universität des Saarlandes, Saarbrücken
03/2005 - 03/2006	Risikocontroller bei HSBC Trinkaus & Burkhardt KGaA, Düsseldorf
04/2006 - 08/2007	Portfolioanalyst bei HSBC Global Asset Management (Deutschland) GmbH, Düsseldorf
09/2007 - 09/2011	Portfoliomanager bei HSBC Global Asset Management (Deutschland) GmbH, Düsseldorf
seit 10/2011	Professor an der SRH Fernhochschule
seit 01/2015	Leiter des Studiengangs Betriebswirtschaft und Management (B.A.)

Vorwort

Liebe Leserin, lieber Leser,

oh, oh, ein Studienbrief zu Statistik ...

… und beim ersten Blättern erblickt man viele furchteinflößende Formeln ...

Ich weiß, dass Statistik bei vielen Studierenden ein eher weniger beliebtes Fach ist. Und außerdem studiert man Psychologie und nicht Mathematik. Wozu das Ganze also?

Kurz gesagt: Ohne Statistik keine Forschung! Ohne Forschung kein Fortschritt!

Einige von Ihnen werden im Laufe Ihres Studiums selbst Forschung betreiben. Spätestens dann werden Sie mit Statistik konfrontiert. Vielleicht untersuchen Sie, welchen Einfluss der Führungsstil auf die Mitarbeitermotivation hat. Vielleicht interessieren Sie sich dafür, in welchem Zusammenhang Job-Enrichment und Leistung stehen oder wie sich Coaching bei Burnout gefährdeten Personen auswirkt.

Für alle diese Fragestellungen werden Sie – mal mehr, mal weniger – Statistik benötigen.

Am Ende der Kapitel finden Sie Übungsaufgaben, mit denen Sie Ihren Lernfortschritt überprüfen können. Die dazugehörigen Lösungen finden Sie am Ende des Studienbriefs.

Vielleicht stellen Sie ja im Laufe des Lesens und Durcharbeitens fest, dass manches zwar schlimm aussieht, aber gar nicht so schlimm ist!

Viel Spaß und viel Erfolg beim Lernen,

Darmstadt, November 2009 Kai Budischewski

Inhaltsverzeichnis

1 Einleitung

Auf den amerikanischen Autor und Gesellschaftskritiker Mark Twain geht angeblich folgendes Zitat zurück: "There are three kinds of lies: Lies, damned lies, and statistics."

Statistik hat keinen besonders guten Ruf. Nahezu jeder kennt zumindest einen der beiden folgenden Sätze:

„Mit Statistik kann man alles beweisen" und „Ich glaube nur der Statistik, die ich selber gefälscht habe".

Als Dozent, der sich auf Methodenlehre – unter anderem also Statistik – spezialisiert hat, frage ich die Studierenden zu Anfang immer, ob die beiden letztgenannten Sätze denn zutreffen. Und? Was denken Sie?

Warum Statistik?

Statistik findet immer dann eine Anwendung, wenn entweder aufgrund der Stichprobenergebnisse auf die Grundgesamtheit geschlossen werden soll oder wenn zwei Stichproben bezüglich eines Merkmales miteinander verglichen werden sollen.

Statistik ist nur ein Werkzeug!

Statistik ist ein wichtiges Instrument der Psychologie, ein Werkzeug, ohne das es kaum geht. In der Psychologie gibt es so gut wie keine Gesetze, so dass man sagen könnte: Ich mache das, dann passiert das. In der Psychologie werden Wahrscheinlichkeitsaussagen gemacht, also: Wenn ich dieses mache, wird wahrscheinlich jenes passieren.

Beispiel 1:
Zwei Personen werfen je zehn Mal eine Münze. Person A hat viermal Zahl geworfen, Person B sieben Mal Zahl. Bedeutet das, dass die Münzen etwa unterschiedlich sind, oder würden sie, wenn sie nur oft genug werfen, beide gleich häufig Zahl haben?

Beispiel 2:
Therapieform A hat bei zwanzig Depressiven eine Besserung um 10% bewirkt, Therapieform B hat bei fünfzehn Depressiven eine Besserung um 15% bewirkt. Bedeutet das nun, dass Therapieform B besser ist? Oder ist Therapieform B nur zufällig in dieser Stichprobe besser, in Wirklichkeit aber schlechter als Therapieform A?

Das erste Beispiel war noch einfach, da sich jeder vorstellen kann, dass beim Münzenwerfen – wenn nur oft genug geworfen wird – gleich häufig Zahl auftreten wird. Für das zweite Beispiel werden jedoch schon sogenannte Signifikanztests benötigt.

Allgemein kann man vier große Anwendungsgebiete für statistische Tests ausmachen:

1. Man untersucht eine Stichprobe hinsichtlich eines bestimmten Merkmals, z.B. 'Fähigkeit, Statistik zu begreifen'. Nun will man von der Stichprobe aus auf die Population/Grundgesamtheit/Bevölkerung schließen. In der Stichprobe ermittelt man einen durchschnittlichen Fähigkeitswert von 15.75 Punkten und eine Standardabweichung von 3.2 Punkten. Aus diesen Daten wird dann auf die Populationswerte geschlossen.

2. Man untersucht zwei Stichproben (oder mehr) hinsichtlich eines bestimmten Merkmals, wobei die eine Stichprobe unterschiedliche Voraussetzungen als die andere Stichprobe hat (z.B. Training, andere Lernmethode o. ä.). Hierbei vergleicht man die Ergebnisse der Stichproben und führt etwaige Unterschiede auf die unterschiedlichen Voraussetzungen zurück. Beispiel: Statistik-Tutoriumsgruppe A lernt nach dem Lehrbuch X. Statistik-Tutoriumsgruppe B lernt nach dem Lehrbuch Y. Gemessen werden die Noten in der Statistikklausur. Gruppe A hat als Mittelwert 2.4; Gruppe B einen Mittelwert von 1.8. Kann behauptet werden, dass Gruppe B einen signifikant besseren Notendurchschnitt hat als Gruppe A? Eine Untergruppe hiervon ist, wenn man eine Stichprobe mehrmals misst, z.B. vor und nach einem Training (siehe auch: abhängige Stichproben).

3. Man untersucht den Zusammenhang zwischen zwei Variablen. Zum Beispiel könnte man untersuchen, ob zwischen dem Merkmal: 'Geschlecht' und dem Merkmal: 'Erfolgreich absolviertes Studium' ein Zusammenhang besteht. Zusammenhänge werden über sogenannte Korrelationsverfahren berechnet.

4. Auf der Grundlage bisheriger Erfahrungen wird versucht, eine Vorhersage/Prognose zu treffen: In einem Bewerbungstest hat eine Person mit einer bestimmten Punktzahl abgeschnitten. Wird diese Person ihre Arbeit gut erledigen können bzw. ist diese Person für die Arbeitsstelle geeignet oder nicht?

2 Wahrscheinlichkeitsrechnung

Lernziele

Wenn Sie dieses Kapitel bearbeitet haben, dann wissen Sie,

- was man unter einem Zufallsexperiment versteht;
- was man unter dem Begriff der Wahrscheinlichkeit versteht;
- wann man von Laplace-Wahrscheinlichkeiten sprechen darf;
- was unter dem Additionstheorem und dem Multiplikationstheorem zu verstehen ist;
- wofür der Ausdruck Fakultät steht;
- was unter einer Binomialverteilung zu verstehen ist;
- wie die Zufallsratewahrscheinlichkeit bei Multiple-Choice-Aufgaben berechnet werden kann;
- was unter einer Normalverteilung zu verstehen ist und
- welche besonderen Kennzeichen die sogenannte Standardnormalverteilung besitzt.

Wahrscheinlichkeiten sind das Natürlichste von der Welt. Nahezu jeder benutzt sie in der Umgangssprache: „Wahrscheinlich komme ich so gegen 20:00 Uhr bei Dir vorbei." „Wahrscheinlich habe ich eine 2 bis 3 in der Klausur." „Wenn ich eine Münze 100mal werfe, wird wahrscheinlich so circa 50mal Kopf fallen." „Wahrscheinlich ist es besser, einen Sturzhelm beim Motorradfahren zu tragen." „Wenn ich bei diesem Wetter spazieren gehe, werde ich wahrscheinlich krank."

Bei all diesen Beispielen wägt man unbewusst Wahrscheinlichkeiten gegeneinander ab. Beim ersten Beispiel ist die Wahrscheinlichkeit, zwischen 19:55 und 20:05 vorbei zu kommen (je nach Pünktlichkeitsbewusstsein der Person) wohl recht hoch, es könnte aber auch sein, dass man davon abweichend doch etwas früher oder später kommt. In der Wahrscheinlichkeitsrechnung geschieht das Bestimmen von Wahrscheinlichkeiten auf mathematischem Wege. Wahrscheinlichkeiten liegen dabei immer zwischen 0 und 1, wobei auch die Zahlen 0 und 1 zulässige Werte sind. Abgekürzt werden Wahrscheinlichkeiten üblicherweise mit dem Buchstaben P (in Anlehnung an das englische Wort „probability").

2.1 Grundbegriffe der Wahrscheinlichkeitsrechnung

Im Folgenden werden einige wichtige Begriffe eingeführt, die für das weitere Verständnis relevant sind. Unter einem **Zufallsexperiment** versteht man einen Vorgang, der beliebig oft wiederholbar ist, nach einer bestimmten Vorschrift ausgeführt wird und dessen Ergebnis nicht eindeutig vorherbestimmbar ist. Die möglichen Ausgänge eines Zufallsexperimentes nennt man **Ergebnisse.**

So lässt sich beispielsweise das Werfen eines Würfels als Zufallsexperiment auffassen. Die Ergebnisse dieses Zufallsexperimentes sind die Zahlen 1, 2, 3, 4, 5 und 6. Aber auch das Werfen einer Münze oder das Ziehen einer Kugel aus einer Urne mit verschiedenen Kugeln können als Zufallsexperimente aufgefasst werden. Die Menge aller Ergebnisse eines Zufallsexperimentes wird als **Ergebnisraum** bezeichnet und durch den griechischen Buchstaben Omega symbolisiert. Im Beispiel „Werfen eines Würfels" sähe der Ergebnisraum wie folgt aus:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Von den bis jetzt kennengelernten Ergebnissen eines Zufallsexperimentes sind die sogenannten Ereignisse zu unterscheiden. Ereignisse können einzelne Ergebnisse oder eben auch zusammengesetzte Ergebnisse eines Zufallsexperimentes sein, für die wir uns interessieren und von denen wir Wahrscheinlichkeiten für deren Eintreten bestimmen wollen. Ereignisse lassen sich mathematisch als einfache Mengen beschreiben. Nehmen wir an, wir interessieren uns für das Ereignis A = „Beim einmaligen Werfen eines Würfels fällt eine gerade Zahl". Die Darstellung dieses Ereignisses in Mengenschreibweise sähe dann wie folgt aus:

$$A = \{2, 4, 6\}.$$

Ereignisse werden im Allgemeinen mit großen lateinischen Buchstaben A, B, C, ... bezeichnet. Mathematisch sind Ereignisse nichts anderes als Teilmengen von Ω. Eine Menge A ist übrigens genau dann Teilmenge einer Menge Ω, wenn jedes Element von A auch in Ω enthalten ist. Bei den möglichen Ergebnissen eines Zufallsexperimentes handelt es sich natürlich ebenfalls um Ereignisse, man nennt die Ergebnisse eines Zufallsexperimentes folglich auch **Elementarereignisse**. Die Anzahl der Elemente einer Menge A nennt man auch Mächtigkeit (oder Kardinalität) und bezeichnet diese mit $|A|$. Die Mächtigkeit von Ω (also die Anzahl der Elemente von Ω) beträgt $|\Omega| = 6$. Ist A ein Ereignis, so nennt man das Ereignis $\overline{A}$ das **Komplementärereignis** von A, wenn gilt: $A \cup \overline{A} = \Omega$.[1]

Zur Übung werden in der nachfolgenden Tabelle für das Zufallsexperiment „Einmaliges Werfen eines Würfels" verschiedene Ereignisse betrachtet.

Verbale Beschreibung der Ereignisse	**Mathematische Darstellung**
A = „Werfen einer 1"	$A = \{1\}$
B = „Werfen einer 6"	$B = \{6\}$
C = „Werfen einer geraden Zahl"	$C = \{2, 4, 6\}$
D = „Werfen einer ungeraden Zahl"	$D = \{1, 3, 5\}$
E = „Werfen einer Zahl größer als 2"	$E = \{3, 4, 5, 6\}$
F = „Werfen einer Zahl von 1 bis 6" (sicheres Ereignis)	$F = \{1, 2, 3, 4, 5, 6\} = \Omega$
G = „Werfen keiner Zahl von 1 bis 6" (unmögliches Ereignis)	$G = \{\,\} = \emptyset$ (leere Menge)

Tabelle 1: Verschiedene Ereignisse im Zufallsexperiment „Werfen eines Würfels". (Quelle: Eigene Darstellung)

Man bezeichnet zwei Ereignisse A und B als **disjunkte Ereignisse**, wenn gilt: $A \cap B = \emptyset$, d.h. wenn die Schnittmenge beider Ereignisse leer ist.

Nachdem der Begriff der Ereignisse eingeführt wurde, kann man sich nun Gedanken darüber machen, wie wahrscheinlich denn gewisse Ereignisse sind, also wie wahrscheinlich es zum Beispiel ist, eine ungerade Zahl beim einmaligen Werfen eines Würfels zu würfeln. Nun kommen wir (endlich) dazu, den Begriff der Wahrscheinlichkeit einzuführen.

Die Wahrscheinlichkeit $P(A)$ gibt an, wie wahrscheinlich es ist, dass das Ereignis A bei einem Zufallsexperiment auftritt. Diesem Ereignis wird ein Wert $0 \leq P(A) \leq 1$ zugeordnet. Einem sicheren Ereignis wird dabei die Wahrscheinlichkeit 1 , einem unmöglichen Ereignis die Wahrscheinlichkeit 0 zugeordnet. Je höher die Wahrscheinlichkeit ist, desto wahrscheinlicher wird das Ereignis eintreten.

[1] Auf eine Vertiefung mengentheoretischer Grundlagen und Operationen wird an dieser Stelle verzichtet.

Nach der **klassischen Definition der Wahrscheinlichkeit** (nach **Laplace**[2]) ist die Wahrscheinlichkeit, dass bei einem bestimmten Zufallsexeperiment das Ereignis A eintritt, der Quotient aus der Anzahl der günstigen Fälle und der Anzahl der möglichen Fälle:[3]

$$P(A) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{|A|}{|\Omega|}.$$

Wenden wir uns nun wieder dem Zufallsexperiment „Einmaliges Werfen eines Würfels" zu. Die möglichen Ergebnisse dieses Zufallsexperimentes sind die sechs Augenzahlen des Würfels. Man betrachtet die Ereignisse A = „Werfen einer 1" und B = „Werfen einer gerade Zahl". In Mengenschreibweise lassen sich diese Ereignisse wie folgt notieren:

$$A = \{1\}$$

und

$$B = \{2, 4, 6\}.$$

Unterstellt man, dass jede der sechs Augenzahlen (= Elementarereignisse) des Würfels gleichwahrscheinlich sind (also ein fairer Würfel zum Einsatz kommt), so lassen sich die Wahrscheinlichkeiten der Ereignisse A und B wie folgt bestimmen:

$$P(A) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{|A|}{|\Omega|} = \frac{1}{6}$$

und

$$P(B) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{|B|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}.$$

Kann die Wahrscheinlichkeit von Ereignissen nicht theoretisch bestimmt werden, so behilft man sich in der Wahrscheinlichkeitsrechnung mit einem anderen Ansatz. Möchte man z.B. die Wahrscheinlichkeit des Ereignisses bestimmen, dass bei einer verbogenen Münze Kopf fällt, so hilft die bisherige Betrachtung nicht wirklich weiter. Es kann nicht gesagt werden, dass die Wahrscheinlichkeit für Kopf bei der verbogenen Münze genau wie bei einer „normalen Münze" auch gleich 0.5 beträgt.

Oder wie sieht es aus, wenn wir die Wahrscheinlichkeit dafür bestimmen möchten, dass die Anwendung einer neuen Behandlungsmethode bei psychosomatischen Störungen erfolgreich ist? Wie wahrscheinlich ist es, dass die Therapie zur Heilung geführt hat? Bei der verbogenen Münze könnte man ausprobieren, wie oft denn z.B. Kopf erscheint, wenn die Münze 100 Mal geworfen wird. Fällt dabei nur 10 Mal Kopf, so kann man davon ausgehen, dass die Münze nicht besonders „fair" ist. Im zweiten Fall könnte man (an dieser Stelle recht vereinfachend beschrieben) eine Stichprobe von Personen untersuchen, welche die Therapieform in Anspruch genommen haben um so einen Eindruck davon zu gewinnen, wie wahrscheinlich es ist, dass die Therapieform zur Heilung führt.

---

2 Pierre-Simon (Marquis de) Laplace (1749 - 1827), französischer Mathematiker, Physiker und Astronom.

3 Es wird dabei unterstellt, dass der Ergebnisraum Ω endlich ist und alle Elementarereignisse gleichwahrscheinlich sind.

Beide Betrachtungen haben gemeinsam, dass man aufgrund von aufgetretenen Häufigkeiten auf die (unbekannte) Wahrscheinlichkeit schließen möchte. Für ausreichend große n (= Stichprobengröße bzw. Anzahl aller Fälle) kann man annehmen:[4]

$$P(A) \approx \frac{\text{Anzahl der Fälle, bei denen A eintritt}}{\text{Anzahl aller Fälle}}.$$

In der Statistik spricht man bei einem solchen Vorgehen, also z.B. dem Schluss aus einer Stichprobe auf eine unbekannte Wahrscheinlichkeit (oder einem anderen Parameter) auch von **Schätzung**. Je größer der Stichprobenumfang ist, desto genauer ist im Allgemeinen auch die Schätzung.

2.2 Additions- und Multiplikationstheorem

Gegeben seien zwei beliebige Ereignisse A und B eines Zufallsexperimentes. Die Wahrscheinlichkeit des Ereignisses $A \cup B$ ist dann:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Diese Beziehung wird als **Additionstheorem** bzw. als **Additionssatz für Wahrscheinlichkeiten** bezeichnet.

Schließen sich die Ereignisse A und B gegenseitig aus (d.h. sind A und B disjunkte Ereignisse), dann ist $P(A \cap B) = P(\emptyset) = 0$ und somit gilt **(nur wenn A und B sich gegenseitig ausschließen)**:

$$P(A \cup B) = P(A) + P(B).$$

Beachten Sie, dass obige Formel nur gilt, wenn die Ereignisse A und B disjunkt sind!

Sind A und $\overline{A}$ hingegen Komplementärereignisse, dann gilt (weil A und $\overline{A}$ disjunkt sind) mit vorstehender Formel:

$$P(A \cup \overline{A}) = P(A) + P(\overline{A}).$$

Da $P(A \cup \overline{A}) = P(\Omega) = 1$ folgt damit direkt:

$$P(A) = 1 - P(\overline{A}) \text{ bzw. } P(\overline{A}) = 1 - P(A).$$

Gegeben seien zwei Ereignisse A und B eines Zufallsexperimentes. Unter der **bedingten Wahrscheinlichkeit für A gegeben B** versteht man die Wahrscheinlichkeit für das Eintreten von A unter der Bedingung, dass B eingetreten ist. Man berechnet die bedingte Wahrscheinlichkeit des Ereignisses A gegeben B (die Schreibweise dafür lautet: A|B) wie folgt:

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

falls $P(B) > 0$.

[4] Die theoretische Grundlage dazu bildet das Gesetz der großen Zahlen, nach dem sich die relative Häufigkeit (= Anzahl der Fälle, bei denen A eintritt geteilt durch die Anzahl aller Fälle) mit wachsendem Stichprobenumfang immer mehr an die theoretische Wahrscheinlichkeit annähert.

Die bedingte Wahrscheinlichkeit für A gegeben B ist also nur definiert, falls P(B) nicht gleich Null ist. Wäre P(B) nämlich gleich Null, dann wäre B ein unmögliches Ereignis, kann daher nicht eingetreten sein und somit wäre die Betrachtung P(A|B) in diesem Falle (also bei P(B) = 0) nicht sinnvoll.

Vertauscht man in obiger Formel A und B, so ergibt sich, diesmal falls P(A)>0:

$$P(B|A) = \frac{P(A \cap B)}{P(A)},$$

da $P(A \cap B) = P(B \cap A)$. Durch Umstellen der Formel für P(A|B) erhält man:

$$P(A \cap B) = P(A|B) \cdot P(B)$$

und durch Umstellen der Formel für P(B|A) erhält man:

$$P(A \cap B) = P(B|A) \cdot P(A).$$

Offensichtlich gilt daher

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A).$$

Den vorstehenden wichtigen Satz nennt man auch **Multiplikationstheorem**.

Sind die Ereignisse A und B unabhängig[5] voneinander, dann ist P(A|B)=P(A) und es folgt aus dem Multiplikationstheorem der **Multiplikationssatz für unabhängige Ereignisse**:

$$P(A \cap B) = P(A) \cdot P(B).$$

Wir betrachten das Konzept der bedingten Wahrscheinlichkeit am nachfolgenden Beispiel näher.[6]

Beispiel: „Im Jahr 1997 gibt es in der Bevölkerung 1000 Personen mit einer bestimmten seltenen psychischen Störung. 5 Jahre später haben 600 Personen Therapie gemacht, 350 Personen haben Therapie gemacht und sind geheilt und 50 Personen sind ohne Therapie geheilt. Als Zufallsexperiment betrachten wir die rein zufällige Auswahl einer Person aus dieser Population von 1000 Patienten. Wie wahrscheinlich ist das Ereignis „Heilung“ unter der Bedingung, dass die Person Therapie gemacht hat?“[7]

Wir definieren die folgenden Ereignisse: H = „Heilung“ und T = „Therapie“. Aus der Aufgabenstellung können wir folgende Wahrscheinlichkeiten ermitteln:

$$P(T) = \frac{\text{Anzahl der Personen, bei denen eine Therapie durchgeführt wurde}}{\text{Anzahl aller Personen mit einer bestimmten seltenen psych. Störung}} = \frac{600}{1000} = 0.6,$$

$$P(H \cap T) = \frac{350}{1000} = 0.35, \quad P\left(H \cap \overline{T}\right) = \frac{50}{1000} = 0.05.$$

---

5 A und B sind unabhängig voneinander, wenn die Tatsache, dass Ereignis B eingetreten ist für das Eintreten des Ereignisses A keine Bedeutung hat, d.h. $P(A|B) = P(A)$ und natürlich auch umgekehrt, d.h. die Tatsache, dass Ereignis A eingetreten ist, hat keinerlei Bedeutung für das Eintreten des Ereignisses B, also: $P(B|A) = P(B)$.

6 Zu den Ausführungen des nachfolgenden Beispiels vgl. Nachtigall, C./Wirtz, M.: 2009, S. 71f.

7 Nachtigall, C./Wirtz, M.: 2009, S. 71.

Gesucht ist die Wahrscheinlichkeit des Ereignisses „Heilung“ unter der Bedingung, dass die Person Therapie gemacht hat, also P(H|T). Gemäß der Formel für die bedingte Wahrscheinlichkeit lässt sich diese Wahrscheinlichkeit wie folgt berechnen:

$$P(H|T)=\frac{P(H\cap T)}{P(T)}=\frac{0.35}{0.6}=0{,}583.$$

Die Wahrscheinlichkeit einer Heilung liegt also bei 0.583, falls im Vorfeld eine Therapie gemacht wurde.

Erweitern wir diese Aufgabe noch ein wenig und fragen uns, wie groß allgemein die Wahrscheinlichkeit einer Heilung ist (also unabhängig davon, ob eine Therapie durchgeführt wurde oder nicht). Gesucht ist also die Wahrscheinlichkeit P(H).

Die Wahrscheinlichkeit P(H) lässt sich wie folgt bestimmen:[8]

$$P(H)=P(H\cap T)+P\left(H\cap\overline{T}\right)=0.35+0.05=0.4.$$

Die Wahrscheinlichkeit einer Heilung (unabhängig davon, ob eine Therapie durchgeführt wurde oder nicht) liegt also bei 0.4; die Wahrscheinlichkeit einer Heilung, falls im Vorfeld eine Therapie durchgeführt wurde, liegt bei 0.583. Die Anwendung einer Therapie erhöht also die Wahrscheinlichkeit für eine Heilung.

2.3 Binomialverteilung

Wir betrachten im Folgenden zunächst das mehrfache Werfen einer Münze. Wirft man eine faire Münze genau einmal, so können für die Ereignisse K = „Kopf“ und Z = „Zahl“ leicht die nachstehenden Wahrscheinlichkeiten ermittelt werden:

$$P(K)=0.5 \text{ und } P(Z)=0.5.$$

Wirft man eine Münze genau zweimal (hintereinander), so können die folgenden Kombinationen auftreten:

1. Wurf	2. Wurf
Kopf	Kopf
Kopf	Zahl
Zahl	Kopf
Zahl	Zahl

Tabelle 2: Kombinationsmöglichkeiten beim zweifachen Münzwurf.
(Quelle: Eigene Darstellung)

Der Ergebnisraum wäre in diesem Fall:

$$\Omega=\{(K, K), (K, Z), (Z, K), (Z, Z)\}.$$

[8] Dieser Ausdruck folgt aus einer mengentheoretischen Betrachtung, da gilt: $H = (H \cap T) \cup (H \cap \overline{T})$. Sie können sich dies zum Beispiel durch Zeichnen von sog. Venn-Diagrammen verdeutlichen. Das resultierende Ergebnis lässt sich dann aus dem Additionstheorem ermitteln: da $(H \cap T) \cap (H \cap \overline{T}) = \emptyset$, ergibt sich $P(H) = P(H \cap T) + P(H \cap \overline{T})$.

Die Wahrscheinlichkeit, dass beim ersten Wurf Kopf und beim zweiten Wurf Kopf fällt, beträgt (vgl. die Ausführungen zur Laplace-Wahrscheinlichkeit):

$$P(\text{"erster Wurf Kopf und zweiter Wurf Kopf"}) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{1}{4} = 0.25.$$

Ebenso ergibt sich:

$$P(\text{"erster Wurf Kopf und zweiter Wurf Zahl"}) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{1}{4} = 0.25,$$

$$P(\text{"erster Wurf Zahl und zweiter Wurf Kopf"}) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{1}{4} = 0.25,$$

$$P(\text{"erster Wurf Zahl und zweiter Wurf Zahl"}) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{1}{4} = 0.25.$$

Wie groß ist nun die Wahrscheinlichkeit dafür, dass beim zweimaligen Werfen einer Münze genau einmal Kopf und einmal Zahl erscheint? Man muss dann die beiden Fälle „erster Wurf Kopf und zweiter Wurf Zahl“ sowie „erster Wurf Zahl und zweiter Wurf Kopf“ berücksichtigten, die Wahrscheinlichkeit für dieses Ereignis beträgt also:

P („genau einmal Kopf und genau einmal Zahl“)
= P(„erster Wurf Kopf und zweiter Wurf Zahl“)
+ P(„erster Wurf Zahl und zweiter Wurf Kopf“) = 0.25 + 0.25 = 0.5

Nun steigern wir das Ganze: Es wird dreimal geworfen!
Welche Kombinationen können auftreten?

Kombination	Ereignis (1. Wurf, 2.Wurf, 3. Wurf)	Wahrscheinlichkeit für das Eintreten des Ereignisses
1	(K, K, K)	0.5 x 0.5 x 0.5 = 0.125
2	(K, K, Z)	0.5 x 0.5 x 0.5 = 0.125
3	(K, Z, K)	0.5 x 0.5 x 0.5 = 0.125
4	(K, Z, Z)	0.5 x 0.5 x 0.5 = 0.125
5	(Z, Z, Z)	0.5 x 0.5 x 0.5 = 0.125
6	(Z, Z, K)	0.5 x 0.5 x 0.5 = 0.125
7	(Z, K, Z)	0.5 x 0.5 x 0.5 = 0.125
8	(Z, K, K)	0.5 x 0.5 x 0.5 = 0.125

Tabelle 3: Dreifacher Münzwurf.
(Quelle: Eigene Darstellung)

Wie groß ist denn nun die Wahrscheinlichkeit dafür, dass bei diesen drei Würfen genau einmal Kopf dabei ist? Das können die Kombinationen 4 oder 6 oder 7 sein.

P(“genau einmal Kopf“) = 0.125 + 0.125 + 0.125 = 0.375

Das heißt, im Idealfall wird in 37.5% aller Fälle genau einmal Kopf dabei sein, wenn dreimal geworfen wird.

Und noch einmal wird gesteigert: Es wird 10mal geworfen!!!

Wer jetzt glaubt, hier würde eine Tabelle mit allen möglichen Kombinationen erscheinen, hat sich getäuscht!![9] Stattdessen wird eine verkürzte Tabelle dargestellt, die angibt, wie oft jeweils Kopf oder Zahl gefallen sind, sowie die dazugehörige Wahrscheinlichkeit.

Anzahl Kopf	Anzahl Zahl	Wahrscheinlichkeit
0	10	0.00098 = 0.098%
1	9	0.00980 = 0.980%
2	8	0.04395 = 4.395%
3	7	0.11719 = 11.719%
4	6	0.20508 = 20.508%
5	5	0.24609 = 24.609%
6	4	0.20508 = 20.508%
7	3	0.11719 = 11.719%
8	2	0.04395 = 4.395%
9	1	0.00980 = 0.980%
10	0	0.00098 = 0.098%

Tabelle 4: Zehnfacher Münzwurf.
(Quelle: Eigene Darstellung)

Woher stammen nun die Wahrscheinlichkeiten, bzw. woher weiß man, wie viele Kombinationsmöglichkeiten es z.B. für 2 Mal Kopf und 8 Mal Zahl) gibt?

Mit der Formel

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

kann man berechnen, wie viele Kombinationsmöglichkeiten es gibt, dass bei n Durchführungen genau k-mal ein bestimmtes Ergebnis auftritt. Der linke Teil der Formel wird gesprochen als „n über k". Das Ausrufezeichen auf der rechten Seite nennt man Fakultät! Der Ausdruck 3! heißt also: 3 Fakultät. Gemeint ist damit, dass bei der Berechnung die Zahlen von 1 bis 3 miteinander multipliziert werden:

$3! = 1 \cdot 2 \cdot 3 = 6$.

Beachten Sie: 0! = 1.

Beispiel: Es soll die Wahrscheinlichkeit berechnet werden, dass bei zehn Würfen zweimal Kopf fällt und achtmal Zahl. Also: n = 10, k = 2.

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} = \binom{10}{2} = \frac{10!}{2! \cdot (10-2)!} = \frac{10!}{2! \cdot 8!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{10 \cdot 9}{2 \cdot 1}$$

D.h. es gibt 45 verschiedene Kombinationen.

Die Wahrscheinlichkeit für eine dieser Kombinationen davon ergibt sich als:

$$P(2 \times \text{K UND } 8 \times \text{Z}) = P(\text{K}) \times P(\text{K}) \times P(\text{Z}) \times P(\text{Z}) \times P(\text{Z}) \times P(\text{Z}) \times P(\text{Z}) \times P(\text{Z}) \times P(\text{Z}) \times P(\text{Z})$$

$$P(2 \times \text{K UND } 8 \times \text{Z}) = 0{,}5^2 \times 0{,}5^8 = 0{,}5^{10} = 0{,}000976562$$

[9] Würde man alle diese Kombinationen aufschreiben, käme man auf insgesamt $2^{10} = 1024$ Fälle.

Daraus folgt, die Gesamtwahrscheinlichkeit für zweimal Kopf und achtmal Zahl (egal in welcher Reihenfolge) ist 45×0.000976562 = 0.04395. Und genau dieses Ergebnis steht oben in der Tabelle.

Bälle in einer Trommel und Münzwürfe sind nicht unbedingt das Arbeitsgebiet von Wirtschaftswissenschaftler und Psychologen. Wozu das Ganze also?

Denken wir an einen Multiple-Choice-Test im Rahmen einer Bewerberauswahl. Dieser Multiple-Choice-Test (MC-Test) bestehe aus zehn Aufgaben. Zu jeder Aufgabe seien vier Antwortmöglichkeiten vorgegeben. Nur eine der Antwortmöglichkeiten sei richtig; die drei anderen Antwortmöglichkeiten seien falsch.

Damit haben wir bereits bestimmte Kennzahlen: n = 10, P(richtig) = 0.25; P(falsch) = 0.75

Wie groß ist die Wahrscheinlichkeit dafür, dass eine Person per Zufall (d.h. per Raten, ohne etwas zu wissen), genau k=3 der Aufgaben in diesem MC-Test richtig ankreuzt?

Die Lösung funktioniert genau wie die Berechnungen beim Münzwurf.

Zuerst benötigen wir die Anzahl an (Kombinations-) Möglichkeiten dafür, drei Richtige und sieben Falsche bei zehn Aufgaben zu haben.

$$\text{Anzahl Möglichkeiten} = \binom{n}{k} = \frac{n!}{k!\cdot(n-k)!} = \frac{10!}{3!\cdot 7!} = \frac{10\cdot 9\cdot 8}{3\cdot 2\cdot 1} = 120$$

Als nächstes brauchen wir die Wahrscheinlichkeit für eine dieser Kombinationen:

$$P(\text{z.B. RRRFFFFFFF}) = 0.25\cdot 0.25\cdot 0.25\cdot 0.75\cdot 0.75\cdot 0.75\cdot 0.75\cdot 0.75\cdot 0.75\cdot 0.75$$
$$= 0.25^3\cdot 0.75^7 = 0.015625\cdot,13348 = 0.00209$$

Zum Schluss müssen wir die beiden Zwischenergebnisse kombinieren:
P(k=3) = 120 × 0.00209 = 0.2508

Die Wahrscheinlichkeit dafür, dass eine Person per Zufall genau drei Mal richtig und sieben Mal falsch ankreuzt – ohne irgendetwas zu wissen – beträgt P=0.2504, d.h. 25%.

Wie groß ist die Wahrscheinlichkeit dafür, dass eine Person per Zufall genau k=2 Aufgaben in diesem MC-Test richtig ankreuzt?

$$\text{Anzahl Möglichkeiten} = \binom{n}{k} = \frac{n!}{k!\cdot(n-k)!} = \frac{10!}{2!\cdot 8!} = \frac{10\cdot 9}{2\cdot 1} = 45$$

Als nächstes brauchen wir die Wahrscheinlichkeit für eine dieser Kombinationen:

$$P(\text{z.B. RRFFFFFFFF}) = 0.25\cdot 0.25\cdot 0.75\cdot 0.75\cdot 0.75\cdot 0.75\cdot 0.75\cdot 0.75\cdot 0.75\cdot 0.75$$
$$= 0.25^2\cdot 0.75^8 = 0.0625\cdot 0.10011 = 0.00626.$$

Jetzt müssen die Ergebnisse wieder kombiniert werden:
P(k=2) = 45 · 0.00626 = 0.2817

Die Wahrscheinlichkeit dafür, dass eine Person per Zufall genau zwei Mal richtig ankreuzt, beträgt P=0.2817, d.h. 28%. Diese Berechnungen können nun für verschiedene k's durchgeführt werden.

Sind für ein Ereignis nur zwei Ausprägungen vorhanden (richtig oder falsch; Gewinn oder Niete oder ähnliches) und werden Wahrscheinlichkeiten dafür berechnet, wie oft z.B. Richtig auftaucht, wenn so und so oft per Zufall angekreuzt wird, dann handelt es sich immer um eine sogenannte **Binomialverteilung**.

Die Formel einer Binomialverteilung (mit den Parametern n und p) lautet:

$$P(k)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k},\ k=0,\ 1,\ \ldots,\ n.$$

Manchmal wird die Binomialverteilung auch statt mit (1-p) einfach mit q geschrieben:

$$P(k)=\binom{n}{k}\cdot p^k\cdot q^{n-k},\ k=0,\ 1,\ \ldots,\ n,$$

wobei dann selbstverständlich für q der Ausdruck 1-p eingesetzt werden muss.

2.4 Normalverteilung

Zentrales Grenzwerttheorem
Generell gilt, dass die Verteilung von Mittelwerten aus Stichproben des Umfangs n, die sämtlich derselben Grundgesamtheit entnommen wurden, mit zunehmendem Stichprobenumfang in eine Normalverteilung übergeht.

In anderen Worten: Ziehe ich aus einer sehr großen Stichprobe mehrere kleine Stichproben und berechne ich für jede kleinere Stichprobe den Mittelwert und stelle ich diese Mittelwerte grafisch dar, dann folgt diese Mittelwerteverteilung einer Normalverteilung.

Aus je mehr Einzelwerten diese kleineren Stichproben bestehen, umso deutlicher wird die Nähe zur Normalverteilung. Normalverteilungen sind achsensymmetrische Verteilungen (um den Erwartungswert μ),, die sich links und rechts der x-Achse asymptotisch nähern.

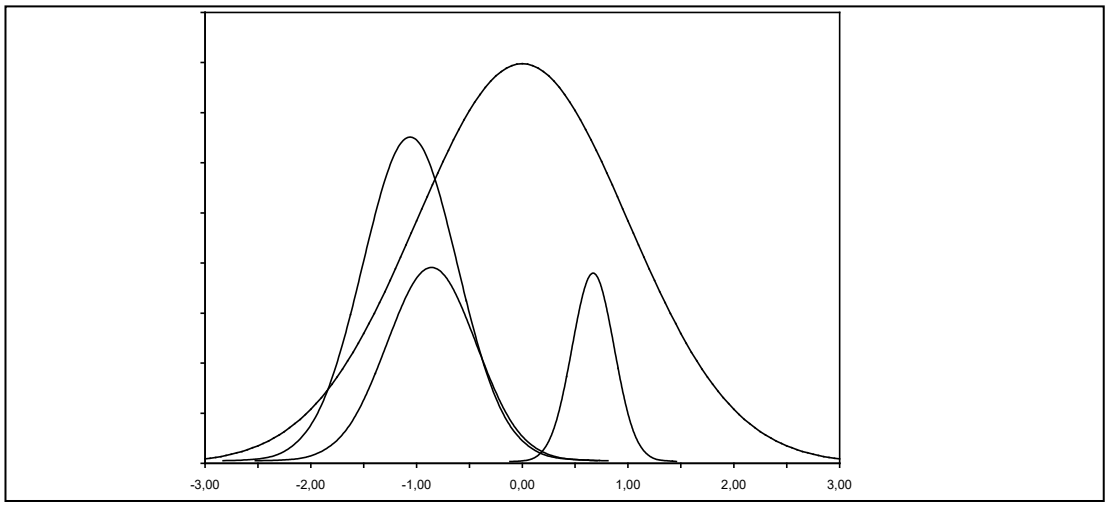

Abbildung 1: Normalverteilungen mit unterschiedlichem Erwartungswert und unterschiedlicher Standardabweichung. (Quelle: Eigene Darstellung)

Eine Standardabweichung beschreibt die durchschnittliche Abweichung der Daten vom Mittelwert. Je größer die Standardabweichung, desto breiter die Normalverteilung; je kleiner die Standardabweichung, desto schmaler ist die Normalverteilung. Normalverteilungen können also anhand ihres Mittelwertes und ihrer Standardabweichung charakterisiert werden.

Eine Normalverteilung mit dem Erwartungswert = 0 und der Standardabweichung = 1 bezeichnet man als **Standardnormalverteilung**. Die Normalverteilungen gehen auf den deutschen Mathematiker C.G. Gauss zurück. Ihre Wichtigkeit kommt daher, weil viele Merkmale und Fähigkeiten in der Bevölkerung/Population/Grundgesamtheit normalverteilt sind. Eines der bekanntesten Beispiele dafür ist der (normierte) Intelligenzquotient. Der IQ hat einen Mittelwert von 100 Punkten und eine Standardabweichung von 15 Punkten.

Wie lässt sich nun z.B. ein IQ von 115 Punkten interpretieren? Naja, besser als der Durchschnitt/ Mittelwert. Man kann aber noch weitergehen, indem man berechnet, wie viel Prozent der Fläche unterhalb der Normalverteilungskurve bis zu diesem Wert liegen. In diesem Beispiel sind das 84,13 Prozent der Fläche. Man könnte also sagen, dass jemand mit einem IQ von 115 Punkten besser als 84 Prozent seiner Grundgesamtheit ist, also zu den besten 16 Prozent gehört. Dieses Umrechnen in Prozent ist der normale Weg der Interpretation. Statt Prozent würde man nur sagen: Die Person hat einen Prozentrang von 84.

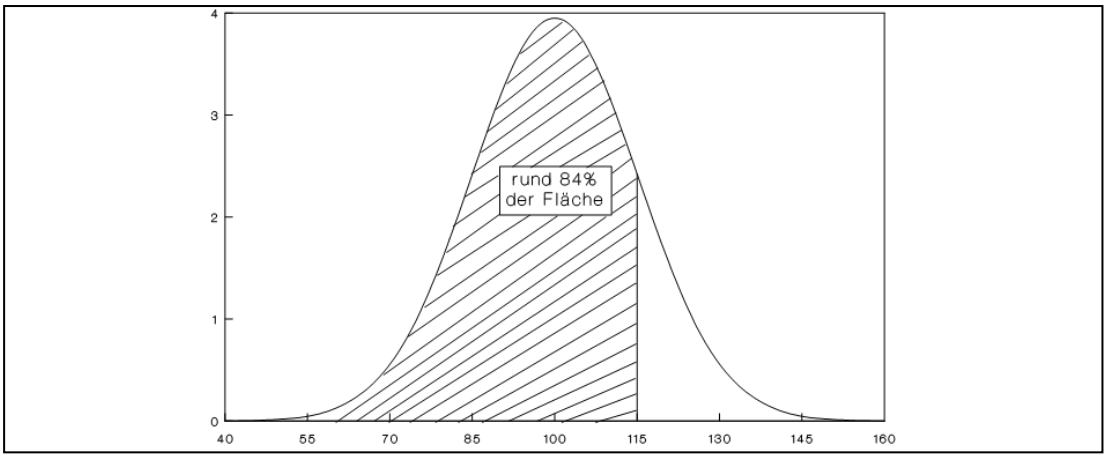

Abbildung 2: Fläche bis IQ=115 Punkte.
(Quelle: Eigene Darstellung)

Die Normalverteilung nennt man auch Zufallsverteilung. Man geht nämlich davon aus, dass – wenn Werte zufällig verteilt werden – sie einer Normalverteilung folgen. Aus diesem Gedankengang entwickelte sich der Begriff 'Signifikanz'.

Beispiel: In einer Schulklasse wird das mathematische Verständnis gemessen. Danach erhält die Klasse eine Woche Intensivtraining zur Steigerung des mathematischen Verständnisses. Nach dieser Woche wird das mathematische Verständnis wieder gemessen. Vorher hatte die Klasse ein durchschnittliches Verständnis von 20 Punkten auf einer Messskala. Nachher lag das durchschnittliche Verständnis bei 27 Punkten. Die Frage ist nun: Ist das durchschnittliche Verständnis bei der zweiten Messung zufällig größer oder hat das Intensivtraining wirklich etwas gebracht? Anders gefragt: Ist die Differenz zwischen dem ersten Mittelwert (20 Punkte) und dem zweiten Mittelwert (27 Punkte) bedeutsam oder rein zufällig? Statt bedeutsam wird in der Statistik der Begriff 'Signifikanz' verwendet.

Wo ist nun die Verbindung zur Normalverteilung? Ganz einfach: Man geht davon aus, dass sich auch die Differenzen normalverteilen; d.h. wenn man zufällig Differenzen bildet, gibt es Differenzen, die häufiger auftreten und solche, die weniger häufig auftreten. Eine Differenz, die in 5% aller Fälle oder weniger rein zufällig auftritt, wird als 'signifikant' bezeichnet. Oder anders ausgedrückt: Ein Ereignis wird dann als signifikant bezeichnet, wenn die Wahrscheinlichkeit dafür, dass dieses Ereignis rein zufällig eintritt, kleiner oder gleich p=0.05 ist.

Beispiel: Man wirft eine Münze zehnmal und zählt aus, wie oft Kopf bzw. Zahl erscheint. Diesen Vorgang wiederholt man zehntausendmal. Am Schluss wird man eine Tabelle ähnlich der nachfolgenden erhalten:

Kopf	Zahl	Häufigkeit des Auftretens
0	10	10
1	9	98
2	8	439
3	7	1172
4	6	2051
5	5	2460
6	4	2051
7	3	1172
8	2	439
9	1	98
10	0	10

Tabelle 5: 10.000mal 10facher Münzwurf.
(Quelle: Eigene Darstellung)

Woher stammen diese Zahlen nun?

1. Die Wahrscheinlichkeit, dass beim einfachen Münzwurf Kopf erscheint ist 0.5.
   Die Wahrscheinlichkeit, dass beim einfachen Münzwurf Zahl erscheint ist 0.5.
2. Mittels der Binomialverteilung lassen sich die Wahrscheinlichkeiten für die verschiedenen Alternativen berechnen.
3. Wahrscheinlichkeit multipliziert mit Anzahl (10.000) ergibt die Häufigkeit des Auftretens.

In der nachstehenden Abbildung geben die Zahlen auf der x-Achse an, wie oft jeweils Kopf und Zahl gefallen sind. 0-10 bedeutet also: keinmal Kopf und zehnmal Zahl.

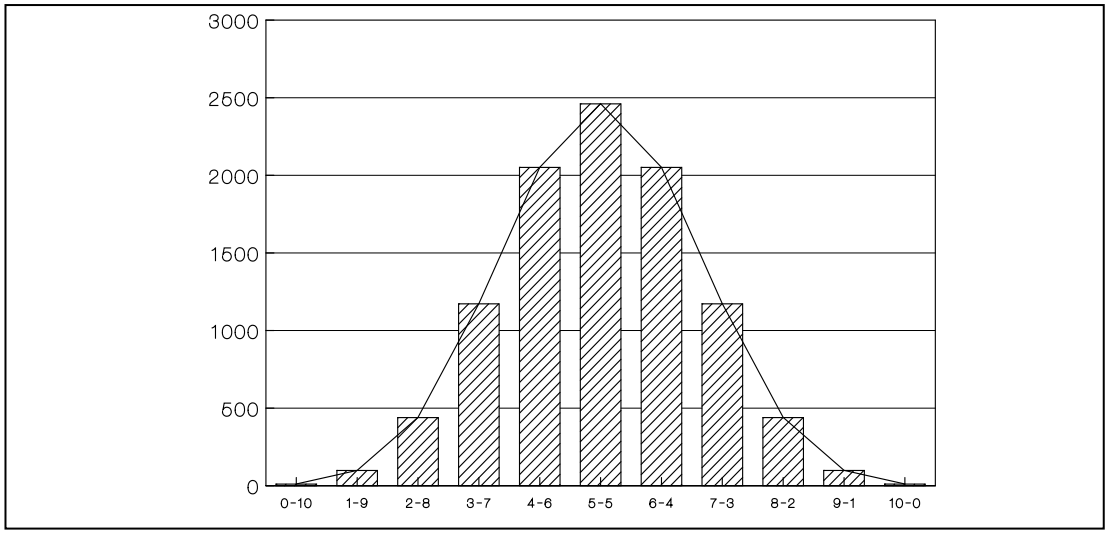

Abbildung 3: 10.000mal 10facher Münzwurf.
(Quelle: Eigene Darstellung)

Wirft jemand nun zehnmal hintereinander Zahl, kann dies natürlich aus Zufall geschehen sein. Da die Wahrscheinlichkeit dafür jedoch kleiner als 5% ist – nämlich p(10×Zahl) = 0.000977 –, würde man dieses Ereignis als überzufällig interpretieren und behaupten, dies sei signifikant, d.h. mit der Münze stimmt irgendetwas nicht. Hier zeigt sich auch gleich ein Dilemma der Statistik: Es kann ja eben sein, dass dieses 10mal-hintereinander-Zahl-werfen wirklich purer Zufall war. In diesem Fall hätte man eine falsche Aussage getroffen, wenn man behaupten würde, mit der Münze stimme irgendetwas nicht.

Übrigens: Wenn man sich die Verteilung ansieht, stellt man fest, dass sie sich doch sehr einer Normalverteilung annähert.

Übungsaufgaben zu Kapitel 2

001 Eine Münze wird dreimal geworfen. Man unterscheidet Kopf (K) und Zahl (Z). Es werden folgende Ereignisse betrachtet: A = „Beim ersten Wurf erscheint Kopf" und B = „Beim dritten Wurf erscheint Zahl".

a) Geben Sie den Ergebnisraum Ω für dieses Zufallsexperiment an!
b) Wie viele Elemente enthält Ω?
c) Stellen Sie die Ereignisse A und B in Mengenschreibweise dar.
d) Beschreiben Sie folgende Ereignisse in Worten und geben Sie die zugehörigen Mengen an: $A \cap B$; $A \cup B$; $\overline{A}$; $A \cap \overline{B}$; $\overline{A} \cap \overline{B}$.
e) Welcher Zusammenhang besteht zwischen $A \cup B$ und $\overline{A} \cap \overline{B}$?

002 Es werde nacheinander mit zwei fairen Würfeln geworfen. Man betrachte die folgenden Ereignisse: A = „Beide Augenzahlen sind gerade", B = „Die Summe aus beiden Augenzahlen beträgt 8" und C = „Beide Augenzahlen sind kleiner oder gleich 4".

a) Geben Sie den Ergebnisraum Ω für dieses Zufallsexperiment an!
b) Wie groß ist die Mächtigkeit von Ω?
c) Stellen Sie die Ereignisse A, B und C in Mengenschreibweise dar.
d) Man bestimme die Wahrscheinlichkeit der folgenden Ereignisse:
A; B;C; $\overline{A}$; $A \cap B$; $A \cup B$; $A \cap B \cap C$; $A \setminus C$ (d.h. A ohne C).

003 Wir befinden uns an einer Universität in einer bestimmten Stadt. Über diese Universität wissen wir:

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.

Berechnen Sie die folgenden Wahrscheinlichkeiten:

a) Wie groß ist die Wahrscheinlichkeit dafür, dass eine zufällig auf dem Campus ausgewählte 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 und über gute statistische Kenntnisse verfügt?

004 Die Wahrscheinlichkeit dafür, dass eine Person den Statistikunterricht versteht, sei p(verstehen)=0.8.

a) Wie hoch ist die Wahrscheinlichkeit dafür, dass von einer Seminargruppe mit n=25 Teilnehmern alle den Statistikunterricht verstehen?

b) Wie hoch ist die Wahrscheinlichkeit dafür, dass in der Seminargruppe mit n=25 Personen niemand den Statistikunterricht versteht?

005 Die Wahrscheinlichkeit, sich zu irgendeiner Zeit am „richtigen Ort" zu befinden, beträgt p=0.3.

Die Wahrscheinlichkeit, sich zur „richtigen Zeit" irgendwo zu befinden, beträgt p=0.5. Gehen Sie davon aus, dass die Ereignisse „richtige Zeit" und „falscher Ort" bzw. „falsche Zeit" und „richtiger Ort" unabhängig sind.

a) Wie groß ist die Wahrscheinlichkeit dafür, zur richtigen Zeit am falschen Ort zu sein?

b) Wie groß ist die Wahrscheinlichkeit dafür, zur falschen Zeit am richtigen Ort zu sein?

c) Wie groß ist die Wahrscheinlichkeit dafür, dass sich von n=5 Personen mindestens 3 zur richtigen Zeit am richtigen Ort befinden?

006 Multiple-Choice, 4 Fragen, 5 Antwortmöglichkeiten

In einem Test hat man vier Multiple-Choice-Fragen mit je fünf Antwortmöglichkeiten, von denen immer nur eine richtig ist.

Beispiel: „Die Durchfallquote in einer Statistik-Klausur liegt bei:
a) 2% ☐
b) 10% ☐
c) 20% ☐
d) 35% ☐
e) 52% ☐"

a) Wie groß ist die Wahrscheinlichkeit dafür, bei einer Frage per Zufall die richtige Antwort anzukreuzen?

b) Wie groß ist die Wahrscheinlichkeit dafür, bei einer Frage per Zufall eine falsche Antwort anzukreuzen?

c) Wie gesagt, der Test besteht aus insgesamt vier solcher Fragen wie das Beispiel oben. Wie groß ist die Wahrscheinlichkeit dafür, die erste richtig und die zweite richtig und die dritte falsch und die vierte falsch zu beantworten?

d) Wie groß ist die Wahrscheinlichkeit dafür, die erste richtig und die zweite falsch und die dritte richtig und die vierte falsch zu beantworten?

e) Wie groß ist die Wahrscheinlichkeit dafür, genau zwei Fragen richtig zu beantworten (egal ob die dritte und vierte oder die zweite und dritte oder...)

007 Multiple-Choice, 5 Fragen, 20 Antwortmöglichkeiten

Ein Dozent entwickelt einen (zugegebenermaßen recht seltsamen) Multiple-Choice-Test. Dieser Test besteht aus 5 Fragen. Zu jeder Frage gibt es jedoch 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?

008 Was sind die Kennzeichen einer Standardnormalverteilung?

3 Grundlagen

Lernziele

Wenn Sie dieses Kapitel durchgearbeitet haben, können Sie,

- die unterschiedlichen Skalenniveaus erläutern und verschiedene Beispiele nennen;
- erläutern, was man unter einer einfachen Zufallsstichprobe, einer geschichteten Stichprobe und einer Klumpenstichprobe versteht;
- den Unterschied zwischen abhängigen und unabhängigen Stichproben erläutern;
- nachvollziehen, was es mit einseitigen und mit zweiseitigen Hypothesen auf sich hat;
- erläutern, was man unter dem alpha- und beta-Fehler versteht und
- erläutern, was Freiheitsgrade sind.

3.1 Skalenniveaus

Daten sind nicht gleich Daten! Das dem in der Tat so ist, verdeutlicht das nachfolgende Beispiel.

Wir untersuchen bei zehn Versuchspersonen das Geschlecht, den Intelligenzquotienten, die Körpergröße und den Schulabschluss. Diese exemplarischen Daten sind in Tabelle 6 aufgeführt.

Versuchsperson Nr.	Geschlecht	Intelligenzquotient (IQ)	Körpergröße in cm	Schulabschluss
1	männlich	96	180	Gymnasium
2	weiblich	102	170	Hauptschule
3	weiblich	105	175	Realschule
4	weiblich	99	180	Gymnasium
5	männlich	103	165	Hauptschule
6	männlich	106	185	Realschule
7	weiblich	100	165	Gymnasium
8	weiblich	104	155	Gymnasium
9	männlich	100	175	Hauptschule
10	weiblich	100	170	Realschule

Tabelle 6: Daten von zehn Versuchspersonen.
(Quelle: Eigene Darstellung)

Was kann man mit diesen Daten nun alles machen? Man könnte z.B. ausrechnen, wie groß die Versuchspersonen durchschnittlich sind; also man könnte den Mittelwert der Körpergröße ermitteln.

Wie wird ein Mittelwert berechnet? Man addiert alle Werte und teilt durch die Anzahl der Personen. Die Anzahl der Personen wird üblicherweise mit n und der Mittelwert (auch arithmetisches Mittel genannt) mit $\bar{x}$ bezeichnet.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{180+170+175+180+165+185+165+155+175+170}{10} = \frac{1720}{10} = 172$$

Das arithmetische Mittel der Körpergröße liegt also bei 172 cm.
Einen Mittelwert aus der Variablen Geschlecht oder aus der Variablen Schulabschluss zu berechnen, macht hingegen keinen wirklichen Sinn.

In der Variablen Körpergröße stecken aber noch weitere Informationen: Man kann z.B. sagen, jemand mit 170cm ist größer als jemand mit 160cm. Eine solche Aussage ist beim Geschlecht (streng genommen) nicht möglich. Bei der Variablen Schulabschluss darf so eine Aussage jedoch getroffen werden. Das Abitur ist im Normalfall höher einzustufen als ein Realschulabschluss, ein Realschulabschluss ist wiederum höherwertiger als ein Hauptschulabschluss. Auch bei der Variablen IQ kann gesagt werden: Jemand mit einem IQ von 105 ist intelligenter als jemand mit einem IQ von 95. Beim IQ kann man auch den Mittelwert berechnen, was wiederum beim Schulabschluss nicht möglich ist.

Variablen unterscheiden sich also darin, welche Aussagen jeweils getroffen werden können.

Bei der Variablen Geschlecht kann lediglich gesagt werden: Weiblich oder männlich, gleich oder ungleich.

Bei der Variablen Schulabschluss kann man eine Rangreihe bilden: Abitur ist besser als Realschulabschluss ist besser als Hauptschulabschluss, also gleich/ungleich und ist besser/schlechter.

Bei den Variablen IQ und Körpergröße kann man außerdem noch angeben, um wie viel jemand intelligenter oder größer ist, man kann auch plus und minus rechnen, also gleich/ungleich und größer/kleiner und plus/minus.

Bei der Variablen Körpergröße kann außerdem noch folgendes gesagt werden: 10 × 10 cm sind 100 cm. So eine Aussage ist beim IQ nicht möglich: 10 × 10 IQ-Punkte sind eben nicht ein IQ von 100!!

In der Statistik unterscheidet man vier verschiedene Skalenniveaus und zwar je nachdem, welche Transformationen:der Messwerte durchgeführt werden dürfen, wird unterschieden in:

- **1. Nominalskalenniveau:**
  Die Daten unterscheiden sich nach gleich und ungleich. (=,≠)
  Beispiel: Geschlecht; Parteizugehörigkeit, Nationalität
- **2. Ordinalskalenniveau:**
  Die Daten unterscheiden sich nach gleich/ungleich, größer und kleiner. (=,≠,<,>)
  Beispiel: Windstärken; Schulabschluss, Schulnoten
- **3. Intervallskalenniveau:**
  Die Daten unterscheiden sich nach gleich/ungleich, größer/kleiner, plus/minus. (=,≠,<,>,+,-)
  Beispiel: Temperatur in Celsius, Intelligenzquotient
  Kennzeichen: Kein natürlicher Nullpunkt
- **4. Verhältnisskalenniveau:** Die Daten unterscheiden sich nach gleich/ungleich, größer/kleiner, plus/minus, mal und geteilt. (=,≠,<,>,+,-,*,:)
  Beispiel: Meter, Zeit, Umsatz
  Kennzeichen: Natürlicher Nullpunkt vorhanden

3.2 Stichprobenarten

Unter einer **Stichprobe** (englisch „sample") versteht man eine (nach bestimmten Kriterien) durchgeführte Auswahl von Objekten (z.B. Personen, Haushalte, Unternehmen etc.) aus einer Grundgesamtheit von Objekten (oft auch Population genannt, englisch „population"). Mit bestimmten Kriterien ist gemeint, dass die Auswahl zum Beispiel durch den Zufall gesteuert sein kann oder auch nach anderen Auswahlkriterien, auf die wir im Folgenden kurz eingehen werden.

Stichproben können wie schon beschrieben klassifiziert werden in Zufallsstichproben und systematische Verfahren.[10] Zu den Zufallsstichproben zählen z.B. die einfache Zufallsstichprobe, die geschichtete Zufallsstichprobe sowie die Klumpenstichprobe. Von den systematischen Verfahren seien an dieser Stelle das Quotenverfahren sowie die systematische Auswahl genannt.

Eine **einfache Zufallsstichprobe** ist eine Stichprobe aus einer Grundgesamtheit von Objekten, bei der jedes Objekt die gleiche Chance hat, in die Stichprobe zu gelangen. Eine **geschichtete Stichprobe** ist eine Stichprobe aus einer Grundgesamtheit von Objekten, bei der aus jeder speziellen Gruppierung (Schicht) der Grundgesamtheit eine einfache Zufallsstichprobe gezogen wird. Geschichtete Zufallsstichproben bieten sich bei recht heterogener Struktur der Grundgesamtheit an. Eine **Klumpenstichprobe** ist eine Stichprobe aus einer Grundgesamtheit, bei der die Grundgesamtheit in Gruppen (sog. Klumpen bzw. engl. Cluster) eingeteilt werden kann. Man wählt dann zufällig (mit einer einfachen Zufallsstichprobe) eine bestimmte Menge von Klumpen aus. Charakteristisch für die Klumpenstichprobe ist, dass in den ausgewählten Clustern dann eine Vollerhebung durchgeführt wird.

Die drei vorstehend beschriebenen Stichprobenarten zählen zu den Zufallsstichproben. Das **Quotenverfahren** wird den systematischen Stichprobenverfahren (auch nicht-probabilistische Stichprobenverfahren genannt) zugeordnet. Das Quotenverfahren kommt in der empirischen Sozialforschung ebenfalls häufig zum Einsatz. Man geht dabei davon aus, dass die Grundgesamtheit stark unterschiedlich verteilt ist. Also teilt man die Grundgesamtheit in verschiedene Quoten auf, z.B. nach Altersklassen, Geschlecht, Bildungsgrad etc. Nun konstruiert man die Stichprobe so, dass die Quotenanteile der Stichprobe mit den Quotenanteilen der Grundgesamtheit übereinstimmen. Wenn z.B. der Anteil der 30 bis 39 jährigen in der Bevölkerung bei 12% liegt, dann wird die Stichprobe so aufgebaut, dass auch 12% der Objekte in der Stichprobe zu dieser Altersklasse gehören usw. Die Auswahl der Objekte obliegt dem Untersuchungsleiter. Es handelt sich bei dem Quotenverfahren um keine Zufallsstichprobe, da keine zufällige Auswahl von Objekten erfolgt.

Die **systematische Auswahl** (auch bewusste Auswahl genannt) ist davon geprägt, dass subjektive Erwägungen bei der Auswahl eine Rolle spielen. Bei einer Mitarbeiterbefragung würde eine systematische Auswahl so aussehen, dass nur männliche Mitarbeiter befragt werden, die in dem Monaten Mai, Juni oder Juli Geburtstag haben.

In der Psychologie und den Wirtschaftswissenschaften hat man oft auch den Fall vorliegen, dass nicht nur eine Stichprobe betrachtet wird, sondern dass zwei (oder mehrere) Stichproben vorliegen, z.B. wenn verschiedene Gruppen von Personen untersucht werden. Zwei Stichproben können zum einen unterschieden werden danach, ob eine Abhängigkeit oder keine Abhängigkeit vorliegt.

Hat man beispielsweise zwei Gruppen vorliegen, die man miteinander vergleichen möchte und sind die Personen in den Gruppen verschieden, so handelt es sich um **unabhängige Stichproben**.

---

[10] Vgl. zu den Ausführungen des nachfolgenden Abschnittes Atteslander, P.: 2008, S. 256ff.

Hat man zwei Gruppen vorliegen und sind die Personen in den Gruppen gleich (also z.B. Test vor einer Coaching-Maßnahme, Test nach einer Coaching-Maßnahme), so handelt es sich um eine **abhängige Stichprobe**. Man weiß also, dass z.B. die erste Person aus der ersten Stichprobe und die erste Person aus der zweiten Stichprobe Herr Müller ist. Es müssen aber nicht zwingend die Personen in den beiden Gruppen identisch sein. Untersucht man z.B. eine spezielle Eigenschaft von Ehepaaren (z.B. Frau Müller ist in der ersten Gruppe die erste Person, Herr Müller ist in der zweiten Gruppe die erste Person), so handelt es sich auch um eine abhängige Stichprobe. In diesem Fall entsteht eine Abhängigkeit durch die Verbundenheit der Ehepartner.

Nach diesen Vorüberlegungen kommen wir zu einer „handfesteren" Definition der beiden Begriffe:

„Zwei Stichproben heißen voneinander **abhängig,** wenn sie den **gleichen** Umfang haben und zu **jedem** Wert der einen Stichprobe **genau ein** Wert der anderen Stichprobe gehört und umgekehrt. Zwischen abhängigen Stichproben besteht somit eine Kopplung. Man spricht daher in diesem Zusammenhang auch von **verbundenen** oder **korrelierten** Stichproben.

Zwei Stichproben, die diese beiden Bedingungen **nicht** zugleich erfüllen, heißen dagegen voneinander **unabhängig (unabhängige** Stichproben). So sind beispielsweise zwei Stichproben von **unterschiedlichen** Umfang stets voneinander **unabhängig**."[11]

Die obige Begriffsdefinition stellt noch einen zusätzlichen interessanten Aspekt heraus, nämlich den des Stichprobenumfangs und dessen Einfluss auf eine Abhängigkeit/Unabhängigkeit der Stichproben. Nur wenn beide Stichproben den gleichen Stichprobenumfang aufweisen, besteht überhaupt die Möglichkeit einer Verbundenheit beider Stichproben; bei ungleichem Umfang beider Stichproben, sind diese quasi per Konstruktion als unabhängige Stichproben anzusehen.

3.3 Hypothesen und Fehlerarten

Hypothesen sind Annahmen oder Vermutungen über die Realität. Vor der Untersuchung einer Stichprobe bildet man Hypothesen, die man dann mit Hilfe der Stichprobe untersucht. In diesem Zusammenhang unterscheidet man zwischen folgenden zwei Typen von Hypothesen:

- Die Nullhypothese H_0: Die Hypothese, die die Sachverhalte beschreibt, die nicht unseren Erwartungen entspricht – die wir also falsifizieren wollen.
- Die Alternativhypothese H_1: Die Hypothese, die die Sachverhalte beschreibt, die unseren Erwartungen entspricht – die wir aber nicht direkt verifizieren können.

K. R. Popper, der den kritischen Rationalismus begründete, legt in seinen Überlegungen dar, dass man Hypothesen nicht beweisen sondern nur widerlegen (falsifizieren) kann.[12] Also ist es naheliegend, den interessierenden Sachverhalt als Alternativhypothese H_1 (auch Forschungshypothese oder Gegenhypothese genannt) zu formulieren. Das entsprechende Gegenteil des Sachverhalts wird als Nullhypothese H_0 formuliert. Wenn wir die Nullhypothese H_0 widerlegen können, so nehmen wir die Forschungshypothese H_1 bis auf Weiteres als gültig an.

---

11 Papula, L.: 2009, S. 455.

12 Vgl. dazu auch die Ausführungen im Studienbrief „Grundlagen der Empirischen Sozialforschung".

Nun lassen wir uns auf ein kleines Gedankenexperiment ein. Nehmen wir an, die Nullhypothese H_0 wäre tatsächlich richtig. Wenn wir im Rahmen unserer statistischen Untersuchung zum Entschluss kommen, H_0 anzunehmen, so ist unsere Entscheidung richtig! Sollten wir H_0 ablehnen (obwohl die Nullhypothese ja eigentlich richtig wäre), so machen wir einen Fehler. In der Statistik nennt man diesen Fehler den "Fehler 1. Art". Überlegen wir weiter: Nehmen wir an, die Alternativhypothese H_1 wäre in Wirklichkeit richtig. Wenn wir nach der statistischen Untersuchung zu dem Schluss kommen, die Nullhypothese abzulehnen, haben wir alles richtig gemacht. Aber was ist, wenn wir die Nullhypothese annehmen würden? Dann hätten wir einen Fehler gemacht. Da dies ein anderer Fehler als eben war, nennt man diesen Fehler den Fehler 2. Art. Die nachfolgende Tabelle fasst diese Überlegungen zusammen.

		Tatsächliche Situation	
		H_0 ist wahr	H_1 ist wahr
Entscheidung:	Nullhypothese H_0 wird angenommen	Entscheidung richtig!	Fehlentscheidung Fehler 2. Art (β-Fehler)
	Nullhypothese H_0 wird abgelehnt	Fehlentscheidung Fehler 1. Art (α-Fehler)	Entscheidung richtig!

Tabelle 7: Hypothesen und Fehlerarten.
(Quelle: Eigene Darstellung)

In der Literatur findet man häufig auch die Begriffe "α-Fehler" und "β-Fehler" in diesem Zusammenhang. Diese korrespondieren mit den Begriffen "Fehler 1. Art" und "Fehler 2. Art". Beim statistischen Hypothesentest ist α eine sehr wichtige Größe um welche wir uns gleich noch weitere Gedanken machen werden.

Doch zuerst zurück zu unserem Gedankenexperiment: Wissen wir nach der Durchführung einer Hypothesenprüfung denn eigentlich, was in Wirklichkeit wahr ist? Nein, dies wissen wir eben nicht. Daher können wir auch nicht sagen, ob unsere Entscheidung richtig war oder ob wir einen Fehler gemacht haben und die falsche Hypothese angenommen haben. Aus diesem Grund legt man bei statistischen Untersuchungen im Vorfeld schon die (maximale) Wahrscheinlichkeit fest, einen Fehler 1. Art zu begehen. Diese (maximale) Wahrscheinlichkeit nennt man dann das Signifikanzniveau α des statistischen Hypothesentests. In der Praxis wird das Signifikanzniveau häufig bei 1% bzw. 5% angesetzt, d.h. $\alpha = 0{,}01$ bzw. $\alpha = 0{,}05$.

3.4 Hypothesentesten

Bei einem **statistischen Hypothesentest** wird im Allgemeinen nach folgendem Schema vorgegangen:

- Formulierung der **Nullhypothese** H_0 und der **Alternativhypothese** H_1.
- Festlegung des **Signifikanzniveaus** α.
- Festlegung einer geeigneten **Teststatistik**.
- Bestimmung des sogenannten **kritischen Bereichs**, der Grundlage für die Ablehnung der Nullhypothese ist.
- Berechnung des **Wertes der Teststatistik** aus einer Stichprobe.
- Entscheidung über die **Ablehnung oder die Nicht-Ablehnung** der Nullhypothese.

Die konkrete Ausgestaltung des Hypothesentests hängt von der Art der zu testenden Parameter (z.B. Mittelwerte, Varianzen, etc.), von der Gerichtetheit der Hypothese (einseitig oder zweiseitig)

und von der Art der Fragestellung (z.B. Vergleich des Stichprobenmittelwerts mit einem erwarteten Wert oder Vergleich zweier oder mehrerer Stichprobenmittelwerte, usw.) ab.

3.5 Hypothesenarten

Beginnen wir mit einem Beispiel.

Beispiel: Die Frage sei, ob sich zwei Mittelwerte unterscheiden.
Diese Frage lässt sich in zwei entgegengesetzte Aussagen umschreiben:
Entweder

a) die Mittelwerte unterscheiden sich nicht oder
b) die Mittelwerte unterscheiden sich.

Eine dieser beiden Aussagen muss zutreffen; die Frage ist nur welche.

Genau nach diesem Schema werden Hypothesen formuliert:

Die sogenannte **Nullhypothese (H0)** behauptet, dass kein Unterschied besteht, die sogenannte **Alternativhypothese (H1)** behauptet, dass ein Unterschied besteht. Mit verschiedenen statistischen Verfahren kann nun nachgeprüft werden, welche dieser beiden Aussagen (Hypothesen) zutrifft.

Nun kann man nicht nur fragen, ob ein Unterschied besteht, sondern man kann auch fragen: Ist der eine Mittelwert (signifikant) größer als der andere Mittelwert? In diesem Fall wird die Frage in folgende zwei Aussagen umgeformt:

a) Der Mittelwert A ist kleiner oder gleich groß wie Mittelwert B;
b) der Mittelwert A ist größer als der Mittelwert B.

Die Formulierung a) entspricht der Nullhypothese – schließlich steht dort auch, dass kein Unterschied besteht -, die Formulierung b) entspricht wieder der Alternativhypothese. Wie schon erwähnt, gibt es in der Statistik keine Gesetze (Wenn ich dies mache, wird jenes passieren). Stattdessen werden Wahrscheinlichkeitsaussagen gemacht (Wenn ich dies mache, wird wahrscheinlich jenes passieren). Wenn man nun prüfen will, ob man die Nullhypothese (H0) oder die Alternativhypothese (H1) annehmen soll, geschieht dies auch über Wahrscheinlichkeitsaussagen. Mit einer Wahrscheinlichkeit von x% trifft die Nullhypothese zu. Man hat sich darauf geeinigt, folgendermaßen zu verfahren:

- Wenn die Wahrscheinlichkeit dafür, dass die Nullhypothese zutrifft, kleiner als 5% ist, entscheidet man sich für die Alternativhypothese.
- Ist die Wahrscheinlichkeit dafür, dass die Nullhypothese zutrifft, größer als 5%, entscheidet man sich für die Nullhypothese.

Die verschiedenen statistischen Testverfahren (in Kapitel 5 kommen genug davon) berechnen nun die Wahrscheinlichkeit dafür, dass die Nullhypothese zutrifft. Prinzipiell unterscheidet man folgende Arten von Hypothesen:

Ungerichtete Hypothesen (auch: unspezifische Hypothesen oder zweiseitige Hypothesen)
Gefragt wird, ob sich die Werte von Gruppe A von den Werten der Gruppe B unterscheiden (=,≠), d. h. die Werte können kleiner oder größer sein.

Mathematisch:[13] H0: $\mu_A=\mu_B$ H1: $\mu_A\neq\mu_B$
Aussage: Die Werte von Gruppe A sind unterschiedlich/nicht unterschiedlich von den Werten der Gruppe B.

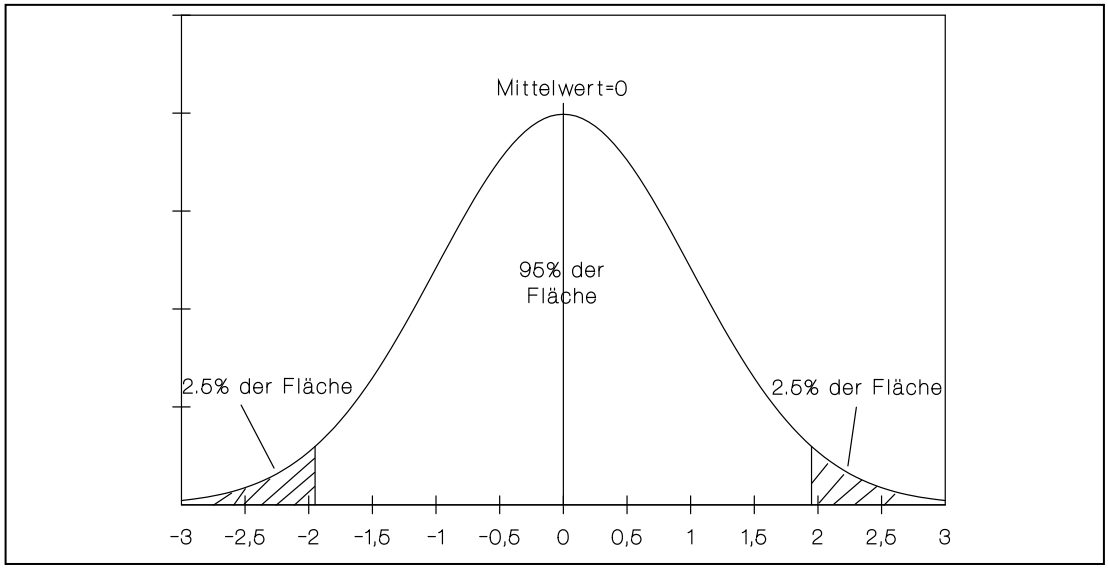

Abbildung 4: Annahme- bzw. Verwerfungsbereich bei zweiseitigen Hypothesen.
(Quelle: Eigene Darstellung)

Gerichtete Hypothesen (auch: spezifische Hypothesen oder einseitige Hypothesen)
Gefragt wird, ob die Werte der Gruppe A besser/schlechter sind als die Werte der Gruppe B, d.h. es wird nur in eine Richtung (nämlich nur nach größer oder nur nach kleiner) geprüft. (≥,< oder >,≤).
Mathematisch: H0: $\mu_A\geq\mu_B$ H1: $\mu_A<\mu_B$ oder H0: $\mu_A\leq\mu_B$ H1: $\mu_A>\mu_B$
Aussage: Die Werte der Gruppe A sind besser/schlechter als die Werte der Gruppe B.

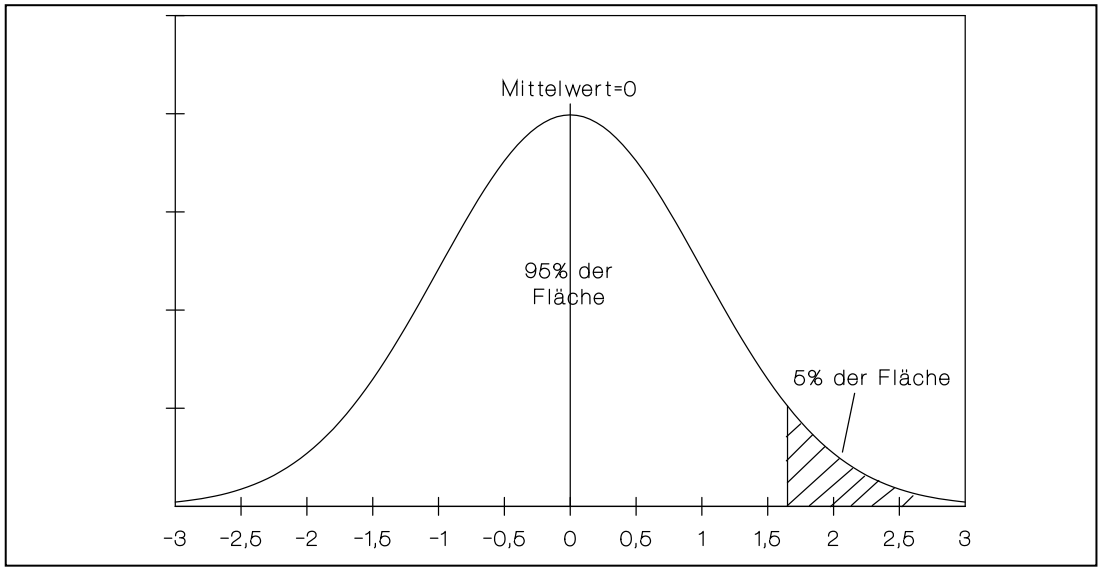

Abbildung 5: Annahme- bzw. Verwerfungsbereich bei einseitigen Hypothesen.
(Quelle: Eigene Darstellung)

[13] Bei mathematischen Hypothesen werden in der Regel griechische Buchstaben verwendet, siehe Kapitel 7.

3.6 Freiheitsgrade

Der Begriff der Freiheitsgrade (englisch degrees of freedom, kurz: df) wird in den nachfolgenden Kapiteln noch häufiger erscheinen. An dieser Stelle soll daher kurz erläutert werden, was man darunter versteht.

Nehmen wir an, wir haben ein arithmetisches Mittel aus einer Stichprobe von n = 3 Beobachtungen berechnet. Beträgt der Mittelwert der drei Beobachtungswerte x_1, x_2 und x_3 genau gleich 3, so können nur zwei der drei Beobachtungswerte variieren! Nimmt man beispielsweise bestimmte Werte für x_1 und x_2 an, so ist der dritte Beobachtungswert (wenn man das arithmetische Mittel kennt) quasi vorgegeben, man sagt dazu auch der dritte Beobachtungswert ist determiniert. Beträgt $\bar{x} = 3$ und sind $x_1 = 2$ und $x_2 = 1$, so ist x_3 vorgegeben und muss demnach $x_3 = 6$ sein. Beträgt $\bar{x} = 3$ und sind $x_1 = 2$ und $x_3 = 6$, so ist x_2 vorgegeben und muss demnach $x_2 = 1$ sein. Beträgt $\bar{x} = 3$ und sind $x_2 = 1$ und $x_3 = 6$, so ist x_1 determiniert und muss demnach $x_1 = 2$ sein.

Allgemein kann man festhalten, falls das arithmetische Mittel einer Stichprobe ($x_1, x_2, \ldots, x_n$) vom Stichprobenumfang n bekannt ist, können nur n -1 Beobachtungen frei variieren, in diesem Fall ergeben sich genau n - 1 Freiheitsgrade (df = n - 1). Ebenso verhält es sich, wenn die Varianz einer Stichprobe vom Stichprobenumfang n bekannt ist. Auch in diesem Fall ergeben sich n - 1 Freiheitsgrade.

Übungsaufgaben zu Kapitel 3

009 Welchen Fehler begeht man, wenn man sich auf Grund der Stichprobenergebnisse für die Nullhypothese H0 entscheidet, obwohl in der Grundgesamtheit die Alternativhypothese H1 gilt?

010 Welchen Fehler begeht man, wenn man sich auf Grund der Stichprobenergebnisse für die Alternativhypothese H1 entscheidet, obwohl in der Grundgesamtheit die Nullhypothese H0 gilt?

011 Erläutern Sie kurz den Begriff „Stichprobe".

012 Geben Sie ein Beispiel für eine Untersuchung, bei der abhängige Stichproben vorliegen!

013 Nennen Sie die vier unterschiedlichen Skalenniveaus und geben Sie jeweils ein Beispiel!

014 Welches Skalenniveau besitzt eine Variable, welche die Merkmalsausprägungen : „sehr zufrieden", „eher zufrieden", „eher unzufrieden", „sehr unzufrieden" aufweist?

015 In einem Fragebogen, der die Einstellung zum Tod erfassen soll, werden verschiedene Sätze dargeboten, auf die man mit ja oder nein antworten kann.
Beispiel: „Ich habe Angst vor dem Tod." ja ☐ nein ☐
Welches Skalenniveau liegt eigentlich vor?

4 Deskriptivstatistik

Lernziele

Wenn Sie dieses Kapitel durchgearbeitet haben, dann wissen Sie,

- was unter den Maßen der zentralen Tendenz zu verstehen ist;
- was ein Median und ein Modalwert sind;
- welche Dispersionsmaße es gibt;
- wie die Varianz und die Standardabweichung berechnet werden;
- welche Korrelationstechniken es gibt;
- wie eine Phi-Korrelation, eine Spearman-Rang-Korrelation und eine Produkt-Moment-Korrelation berechnet werden und
- was z-Werte sind und in welchem Zusammenhang sie mit Prozenträngen stehen.

Die Deskriptivstatistik dient dazu, mit möglichst wenigen Kennwerten möglichst viele Informationen darzustellen. Hat man eine Untersuchung durchgeführt und möchte die erhobenen Daten anderen Personen präsentieren, wird man wohl kaum lange Listen mit Zahlen darstellen, sondern eben versuchen, die Daten in der nötigen Kürze zu beschreiben. Typische Kennwerte in der Deskriptivstatistik sind z.B.: Minimum, Maximum, Durchschnitt (Mittelwert, arithmetisches Mittel), Anzahl an Probanden, Prozentangaben und Ähnliches.

4.1 Maße der zentralen Tendenz

Die Maße der zentralen Tendenz geben Auskunft über das **Zentrum** der Daten. Das bekannteste Maß der zentralen Tendenz ist das arithmetische Mittel (auch als Mittelwert oder Durchschnitt bezeichnet). Das arithmetische Mittel wird häufig mit $\bar{x}$ bezeichnet und berechnet sich als Quotient aus der Summe der beobachteten Werte und der Anzahl dieser Werte. Neben dem arithmetischen Mittel sind noch der Median und der Modus als Lagemaß bedeutend:

- Der Median (MD) stellt die Mitte einer nach Größe sortierten Werteliste dar.
- Der Modalwert (Modus, MO) ist derjenige Wert, der am häufigsten in der Werteliste vorkommt.

Vp-Nummer	Alter	Geschlecht	IQ
1	30	männlich (m)	100
2	40	weiblich (w)	105
3	35	weiblich	95
4	25	weiblich	100
5	20	weiblich	102
6	25	männlich	104
7	27	weiblich	96
8	25	männlich	100
9	30	männlich	98
10	20	weiblich	103

Tabelle 8: Rohdaten der Stichprobe von zehn Personen.
(Quelle: Eigene Darstellung)

Wollte man jemandem diese Stichprobe in wenigen Worten beschreiben, so könnte man beispielsweise das Durchschnittsalter berechnen. Anders ausgedrückt: Man berechnet das arithmetische Mittel des Alters:

$$\bar{x}_{Alter} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{30+40+35+25+20+25+27+25+30+20}{10} = 27.7 \text{ Jahre}$$

Warum ist gerade der Mittelwert so interessant? Nun, ein Mittelwert ist nicht nur eine Zahl, sondern er vermittelt auch ein Bild. Wenn man über eine Stichprobe nur wüsste, dass die Personen im Schnitt 30 Jahre alt wären, denkt man sich die Stichprobe ungefähr folgendermaßen: Ein paar Personen knapp unter 30, ein paar knapp über 30. Nur sehr wenige weit unter 30 Jahren und nur sehr wenige weit über 30 Jahren. D.h., ausgehend vom Mittelwert schätzt man das Alter der Personen, der Mittelwert dient also als Schätzer. Allgemein kann man sagen: Der Mittelwert ist immer der beste Schätzwert, wenn man sonst nichts über die Verteilung der Werte weiß!

Weiterhin könnte man noch angeben, welcher Wert am häufigsten auftaucht. Das ist der Wert 25. Den Wert, der am häufigsten auftaucht, nennt man **Modalwert** (MO).

Des Weiteren könnte es noch interessant sein, von welchem Wert rechts und links jeweils 50% der Stichprobe liegen. Dazu schreibt man am besten erst einmal alle Werte geordnet nebeneinander: 20 20 25 25 25 – 27 30 30 35 40.

Der Wert, neben dem rechts und links jeweils 50% liegen, befindet sich zwischen 25 und 27, es bietet sich daher der Wert 26 an. Rechts und links neben dem Wert 26 liegen jeweils 50% der Stichprobe. Diesen Wert nennt man **Median** (MD).

4.2 Dispersionsmaße

Ebenso ist der kleinste Wert (das **Minimum**) von Interesse; nämlich 20 Jahre, sowie der größte Wert (das **Maximum**), 40 Jahre. Zusammengefasst könnte man den Bereich angeben, innerhalb dessen alle Werte liegen: 20 bis 40 Jahre. Diesen Bereich nennt man **Range**.

Jetzt könnte man noch fragen, wie sich die einzelnen Werte um den Mittelwert verteilen, d.h., liegen die einzelnen Werte alle dicht beim Mittelwert oder sind sie weit verstreut. Solch ein Streuungsmaß ist die **Varianz**. Die Varianz beschreibt die mittlere/durchschnittliche quadratische Abweichung vom Mittelwert. Die Varianz (S^2) berechnet sich nach folgender Formel:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n}$$

$\bar{x}$ = Mittelwert der Stichprobe,
x = Wert der einzelnen Versuchspersonen,
n = Anzahl Versuchspersonen,
i = Laufindex (von der ersten bis zur n-ten Versuchsperson).

$$S^2 = \frac{(30-27.7)^2 + (40-27.7)^2 + (35-27.7)^2 + (25-27.7)^2 + \ldots + (20-27.7)^2}{10}$$

$$S^2 = \frac{(2.3)^2 + (12.3)^2 + (7.3)^2 + (-2.7)^2 + \ldots + (-7.7)^2}{10} = 28.952$$

Die Varianz als solche ist jedoch nicht besonders gut interpretierbar. Die Wurzel aus der Varianz, die sogenannte Standardabweichung (S), lässt sich besser interpretieren.

$$S = \sqrt{S^2} = \sqrt{28.952} = 5.381$$

Wenn ein Merkmal/eine Variable normalverteilt ist, liegen in dem Bereich von rechts und links je eine Standardabweichung vom Mittelwert rund 68% aller Werte!

Folgende Abbildung soll dies verdeutlichen:

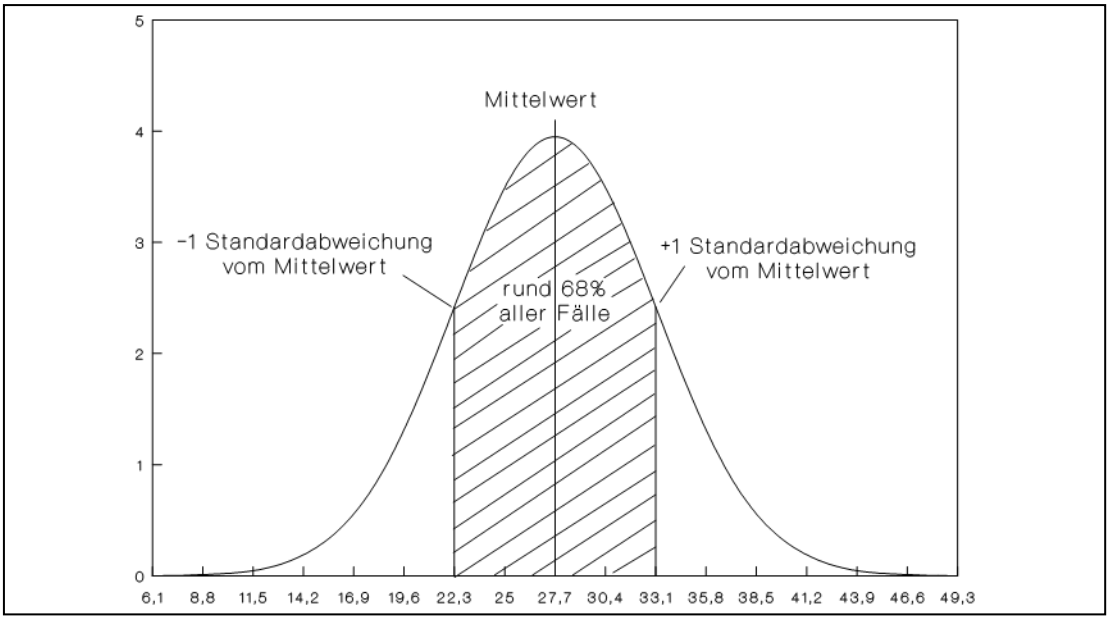

Abbildung 6: Interpretation der Standardabweichung.
(Quelle: Eigene Darstellung)

Je kleiner die Standardabweichung ist, umso enger verteilen sich die Werte um den Mittelwert. Je größer die Standardabweichung ist, umso breiter verteilen sich die Werte um den Mittelwert.

Bezogen auf obiges Beispiel würde man sagen: Im Bereich von 27.7 – 5.381 Jahre bis 27.7 + 5.381 Jahre liegen rund 68% aller Werte, vorausgesetzt, dass die Variable Alter normalverteilt ist.

Noch ein paar Worte zur Varianz: Man könnte sich fragen, warum in der Formel quadriert wird. Nun, gesetzt den Fall, man quadriert nicht, was passiert dann? Unter der Voraussetzung, dass die Werte normalverteilt sind, liegen rechts und links vom Mittelwert gleich viele Werte. Würde man nun einfach immer nur die Differenzen zum Mittelwert bilden und aufsummieren, müsste immer Null herauskommen – wenn die Werte normalverteilt sind. Tja, und mit dem Wert Null kann man nicht viel anfangen! Daher quadriert man die Differenzen. Dadurch werden alle Werte positiv und das Ergebnis ist nicht Null. Genauso gut hätte man auch hoch vier rechnen können; man hat sich allerdings auf hoch zwei geeinigt. Nun ergeben sich aus dem Quadrieren natürlich auch Konsequenzen: Eine große Differenz wird durch das Quadrieren noch größer; d.h. ein Wert, der extrem weit vom Mittelwert entfernt liegt, wirkt sich auf die Varianz stärker aus als ein Wert, der sehr nah beim Mittelwert liegt. Ein interessanter Nebeneffekt ist die oben erwähnte Interpretierbarkeit der Standardabweichung.

Der Vollständigkeit halber (und weil wir diese später brauchen werden) seien hier noch zwei Kennwerte aufgeführt:

Zum einen die so genannte „geschätzte Populationsvarianz" $\hat{\sigma}^2$ sowie die Wurzel daraus, die so genannte „geschätzte Populationsstandardabweichung" $\hat{\sigma}$. Die Berechnungen sind zum Glück relativ einfach. Um die geschätzte Populationsvarianz zu ermitteln, wird folgende Formel verwendet:

$$\hat{\sigma}^2 = S^2 \times \frac{n}{n-1}$$

Wird hiervon nun die Wurzel gezogen, erhält man die geschätzte Populationsstandardabweichung. Im angelsächsischen Sprachraum werden die Begrifflichkeiten (Stichproben-) Standardabweichung bzw. (Stichproben-) Varianz und (geschätzte) Populationsstandardabweichung bzw. Populationsvarianz leider genau entgegengesetzt verwendet. Warum erwähne ich das hier? Naja, die meisten PC-Programme (z.B. Excel) und auch viele Taschenrechner orientieren sich an den angelsächsischen Begrifflichkeiten. Testen Sie es einfach mal aus.

Wenn man eine Verteilung mit dem Mittelwert = 0 und der Standardabweichung = 1 hat, wird dies Standardnormalverteilung und auch z-Werte-Verteilung genannt.

4.3 z-Werte

Die sogenannten z-Werte spielen in der Statistik eine relativ große Rolle. Zum einen können über die z-Werte Probanden aus verschiedenen Stichproben direkt miteinander verglichen werden. Zum anderen werden die z-Werte über die zentrale (Basis) Gleichung der Statistik berechnet. Nach diesem Vorspann nun die Formel:

$$z_i = \frac{x_i - \overline{x}}{S}$$

x_i = Wert eines Probanden
S = Standardabweichung der Stichprobe, der der Proband angehört
$\overline{x}$ = Mittelwert der Stichprobe, der der Proband angehört
i = Index für die Person

Es lohnt, sich diese Formel näher zu betrachten:
Zuerst wird die Differenz des Wertes des Probanden zum Mittelwert gebildet, dann wird diese Differenz an der Standardabweichung relativiert. Dadurch wird der Mittelwert der z-Werte-Verteilung = 0 und die Standardabweichung = 1, unabhängig davon, wie groß der Mittelwert der Stichprobe und die Standardabweichung sind. Hier wird also ein Maß gebildet, das unabhängig von Mittelwert und Standardabweichung ist. Dieses Vorgehen, dass die Differenz eines Wertes zum dazugehörigen Mittelwert an der dazugehörigen Standardabweichung relativiert wird, ist für viele statistische Testverfahren gültig. Nun aber ein Beispiel für die Anwendbarkeit der z-Werte:

Als Chef der Personalabteilung muss man darüber entscheiden, welcher von zwei Bewerbern eingestellt werden soll. Bewerber A hat den Leistungstest LLL (Leisten-Leisten-Leisten) gemacht, Bewerber B den Leistungstest AAA (Arbeit-Arbeit-Arbeit) durchgeführt. Beide Testverfahren sollen die Leistungsfähigkeit bei einer Bürotätigkeit messen. Im Leistungstest LLL kennt man aus einer großen Untersuchung den Mittelwert $\overline{x}$ = 50 und die Standardabweichung S = 8. Für den Leistungstest AAA kennt man den Mittelwert $\overline{x}$ = 80 und die Standardabweichung S = 15.

Bewerber A hat als Punktzahl im LLL den Wert 60 erreicht, Bewerber B im AAA den Wert 90. Vorausgesetzt, eine hohe Punktzahl sei wünschenswert, welcher Bewerber sollte nun eingestellt werden? Ein direkter Vergleich, der eine hat 60 Punkte, der andere 90, verbietet sich, da es ja zwei verschiedene Tests sind. Also: z-Wert berechnen:

$$\text{Bewerber A}: z_A = \frac{60-50}{8} = \frac{10}{8} = 1.25$$

$$\text{Bewerber B}: z_B = \frac{90-80}{15} = \frac{10}{15} = 0.67$$

Nun kann man eine Aussage treffen: Bewerber A hat besser abgeschnitten als Bewerber B, wäre also zu bevorzugen. Ohne das Wissen über den jeweiligen Mittelwert und die jeweilige Standardabweichung hätte man keine Aussage treffen können. Statt zwei Bewerbern um eine Stelle hätte es sich auch um zwei Depressive handeln können, deren Schweregrad der Depression mit unterschiedlichen Verfahren ermittelt wurde.

Die z-Werte-Verteilung hat wiederum einen Mittelwert von 0 und eine Standardabweichung von 1. Die z-Werte-Verteilung ist auch unter dem Namen **Standardnormalverteilung** bekannt.

Für z-Werte im Bereich von –3 bis +3 finden Sie in Tabelle A auch noch die Flächenanteile an der Normalverteilung. Wozu ist das gut?

Auch dazu wieder ein Beispiel: Eine Person hat einen IQ von 115. Wie intelligent ist diese Person eigentlich in Bezug auf die Bevölkerung? Zu den wie viel Prozent Intelligentesten gehört diese Person eigentlich?

Es ist bekannt, dass der Mittelwert des IQ bei 100 liegt. Weiterhin hat die gängige IQ-Skala eine Standardabweichung von S=15..

Mit diesen Angaben (IQ der Person, Mittelwert und Standardabweichung) können wir den z-Wert dieser Person bestimmen:

$$z = \frac{x_i - \overline{x}}{S} = \frac{115-100}{15} = 1$$

Wenn Sie in Tabelle A (siehe Anhang) den z-Wert=1 nachschlagen, finden Sie daneben eine Flächenangabe von 0.8413. Diese Flächenangabe bedeutet: Die Person mit dem IQ=115 bzw. dem $z\text{-Wert}_{IQ}$=1 ist intelligenter/genau so intelligent wie 84% der Bevölkerung bzw. die Person gehört zu den intelligentesten 16% aller Leute.

4.4 Korrelationsmaße

Eine Korrelation gibt Auskunft darüber, ob ein linearer Zusammenhang zwischen Variablen besteht, und wenn ja, wie groß dieser Zusammenhang ist. Was ist damit gemeint?

Beispiel: Schulnoten.
Von sechs Schülern werden die Mathematik-, die Deutsch- und die Physiknoten erhoben.
Die folgende Tabelle listet die Werte auf:

Schüler Nr.	Mathematik	Deutsch	Physik
1	2	3	2
2	1	2	2
3	4	5	3
4	3	1	3
5	6	2	4
6	5	4	4

Tabelle 9: Schulnoten von sechs Schülern.
(Quelle: Eigene Darstellung)

Wenn man sich die Tabelle betrachtet, fällt folgendes auf:

Wenn jemand eine gute Note in Mathematik hat, hat er auch eine gute Note in Physik. Hat jemand eine schlechte Note in Mathematik, hat er auch eine schlechte Note in Physik. Man könnte also vermuten, dass es hier einen Zusammenhang zwischen den Noten in Mathematik und Physik gibt. Weiterhin fällt auf, dass es zwischen den Noten in Mathematik und Deutsch anscheinend keine Beziehung gibt. Zwischen den Noten in Physik und Deutsch ebenfalls nicht. Denken wir uns jetzt einen Idealfall:

Schüler Nr.	Mathematik	Deutsch	Physik
1	1	3	1
2	2	2	2
3	3	5	3
4	4	1	4
5	5	2	5
6	6	4	6

Tabelle 10: Schulnoten ideal.
(Quelle: Eigene Darstellung)

Wenn man nun die Noten in Physik und Mathematik vergleicht, sieht man, dass es immer dieselben Noten sind. Jemand, der in Physik eine „1“ hat, hat auch in Mathematik eine „1“. Wenn das nun immer so wäre, bräuchte man keine Noten mehr in Physik zu erheben, weil die Mathematiknote ausreichen würde und umgekehrt.

Anders ausgedrückt: Wenn man wüsste, welche Note ein Schüler in Mathematik hat, wüsste man automatisch auch, welche Note er in Physik hat bzw. umgekehrt.

Noch anders ausgedrückt: Wenn man wüsste, welche Note jemand in Physik hat, könnte man sehr gut (d.h. ohne Fehler) auf die Mathematiknote schätzen. Abbildung 7 soll dies veranschaulichen.

Wenn ein Zusammenhang nicht perfekt ist, kann man auch nicht so genau schätzen.

- Je geringer ein Zusammenhang ist, umso ungenauer wird eine Schätzung, umso größer wird der Fehler beim Schätzen!
- Je größer ein Zusammenhang ist, umso genauer wird eine Schätzung, umso kleiner wird der Fehler beim Schätzen!

Zusammenhänge lassen sich über Korrelationen berechnen. Dazu wird ein so genannter Korrelationskoeffizient gebildet. Ein Korrelationskoeffizient ist eine genormte Größe, d.h. er kann nur bestimmte Werte annehmen. In diesem Fall nur Werte zwischen –1 und +1. Mit Hilfe des Korrelationskoeffizienten lässt sich die Größe eines Zusammenhangs in Prozent angeben.

Es gilt: Korrelationskoeffizient^2 ×100 = Prozent.

Man spricht allerdings nicht von 'Prozent Zusammenhang', sondern von 'Prozent determinierter Variation/Varianz'. Das Quadrat des Korrelationskoeffizienten ist so wichtig, dass es einen eigenen Namen bekommen hat: Determinationskoeffizient. Man kann also auch sagen: Determinationskoeffizient ×100 = Prozent determinierte Varianz.

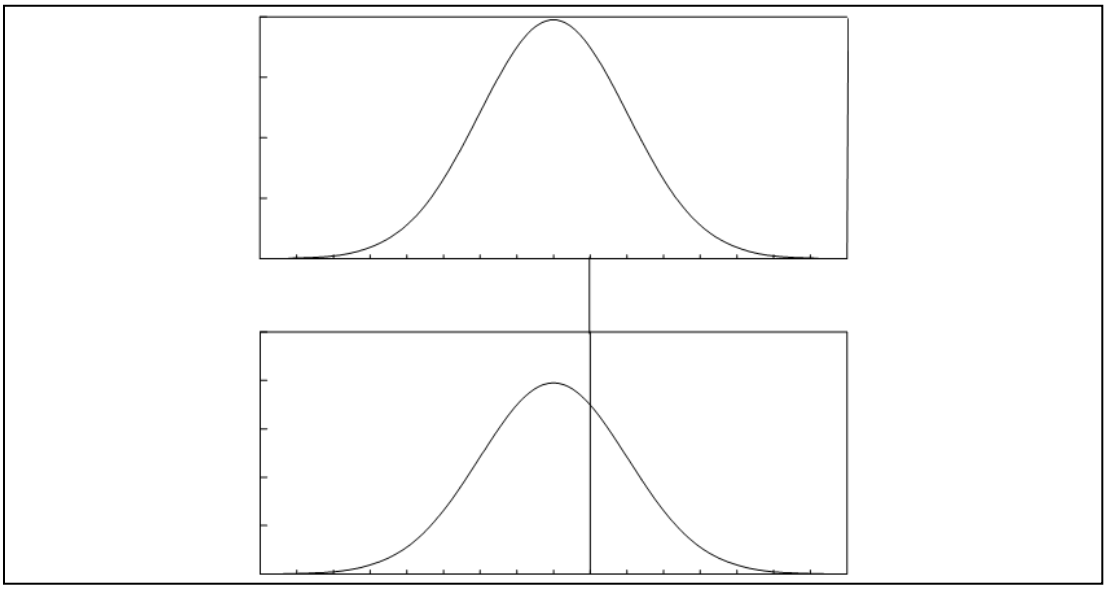

Abbildung 7: Perfekter Zusammenhang zwischen zwei Variablen. Die obere Verteilung stellt die Verteilung der Mathematiknoten dar, die untere die Verteilung der Physiknoten.
(Quelle: Eigene Darstellung)

Wie man nun einen Korrelationskoeffizienten berechnet, hängt davon ab, welches Skalenniveau die Variablen jeweils haben. Tabelle 12 zeigt eine Übersicht über die verschiedenen Korrelationskoeffizienten:

Skalenniveau		**Variable 1**		
		Nominal	**Ordinal**	**Intervall**
	Nominal	Phi, CC, Cramer´s V		r_{Pbis}, r_{bis}
Variable 2	**Ordinal**		Spearman-Rangkorrelation	
	Intervall	r_{Pbis}, r_{bis}		Produkt-Moment-Korrelation

Tabelle 11: Übersicht der Korrelationen.
(Quelle: Eigene Darstellung)

CC ist die Abkürzung für Kontingenzkoeffizient C.
Phi ist die Abkürzung für Phi-Korrelation.
r_{Pbis} ist die Abkürzung für Punkt-biseriale Korrelation.
r_{bis} ist die Abkürzung für Biseriale Korrelation.

Sind also z.B. beide Variablen intervallskaliert, berechnet man den Produkt-Moment-Korrelationskoeffizienten. Sind beide Variablen nominalskaliert, kann man entweder den Kontingenzkoeffizienten C, die Phi-Korrelation oder Cramer´s V berechnen.

Zusammenfassend kann festgehalten werden:

Ein Korrelationskoeffizient von +1 bedeutet: Perfekter positiver Zusammenhang, d.h. ein hoher Wert in der einen Variablen bedeutet einen hohen Wert in der anderen Variablen, bzw. ein niedriger Wert in der einen Variablen bedeutet ein niedriger Wert in der anderen Variablen.

Ein Korrelationskoeffizient von –1 bedeutet: Perfekter negativer Zusammenhang, d.h. ein hoher Wert in der einen Variablen bedeutet einen niedrigen Wert in der anderen Variablen, bzw. ein niedriger Wert in der einen Variablen bedeutet einen hohen Wert in der anderen Variablen.

Ein Korrelationskoeffizient von 0 bedeutet: Die Variablen hängen überhaupt nicht zusammen. Einem hohen Wert in der einen Variablen kann sowohl ein hoher wie ein niedriger Wert in der anderen Variablen entsprechen und umgekehrt.

4.4.1 Phi-Korrelation

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

Die Phi-Korrelation dient zur Berechnung eines Zusammenhangs bei zwei nominalskalierten Variablen, wenn beide jeweils in zwei Stufen vorliegen.

	Frauen	**Männer**	
gute Psychologen	60 a	50 b	110 gute Psych. insgesamt
schlechte Psychologen	40 c	50 d	90 schlechte Psych. insgesamt
	100 Frauen insgesamt	100 Männer insgesamt	200 Personen insgesamt

Tabelle 12: Geschlecht und Qualität von zweihundert Psychologen.
(Quelle: Eigene Darstellung)

Man könnte sich nun fragen, ob es einen Zusammenhang zwischen dem Geschlecht und der Qualität als Psychologe gibt. Setzen wir die Werte doch mal in die Formel ein:

$$\text{Phi} = \frac{50 \cdot 40 - 60 \cdot 50}{\sqrt{100 \cdot 100 \cdot 110 \cdot 90}} = \frac{2000 - 3000}{\sqrt{99000000}} = \frac{-1000}{9949.87} = -0.101$$

Das Vorzeichen spielt bei der Phi-Korrelation keine Rolle! Wenn man die obige Tabelle ein wenig umstellt, erhält man ein positives Vorzeichen. Wie wollte man ein Vorzeichen auch interpretieren? Geht nicht! Und noch einen viel gravierenderen Unterschied zu den anderen Korrelationstechniken gibt es: Ein Phi-Korrelationskoeffizient hat keine festen Grenzen! Das bedeutet, dass man außer dem eigentlichen Phi-Koeffizienten auch noch einen Maximal-Wert errechnen muss, um den Phi-Koeffizienten sinnvoll interpretieren zu können.

Wie berechnet sich nun ein Phi_{max}-Wert? Dazu eine Frage: Wie hätte die obige Tabelle aussehen müssen, damit ein maximaler Zusammenhang bestanden hätte? So:

	Frauen	**Männer**	
gute Psychologen	100 a	0 b	100 gute Psych. insgesamt
schlechte Psychologen	0 c	100 d	100 schlechte Psych. insgesamt
	100 Frauen insgesamt	100 Männer insgesamt	200 Personen insgesamt

Tabelle 13: Maximaler Zusammenhang zwischen Geschlecht und Qualität als Psychologe. (Quelle: Eigene Darstellung)

In dieser Tabelle besteht ganz klar ein Zusammenhang zwischen Geschlecht und Qualität als Psychologe. Alle Frauen sind gute Psychologen, alle Männer sind schlechte Psychologen. Die Tabelle hat nur einen Nachteil: Die Randsummen stimmen nicht mit der Tabelle oben überein. Wie hätte die Tabelle denn aussehen müssen, dass ein maximaler Zusammenhang besteht, die Randsummen aber erhalten bleiben? So:

	Frauen	**Männer**	
gute Psychologen	100 a	10 b	110 gute Psych. insgesamt
schlechte Psychologen	0 c	90 d	90 schlechte Psych. insgesamt
	100 Frauen insgesamt	100 Männer insgesamt	200 Personen insgesamt

Tabelle 14: Maximaler Zusammenhang unter Berücksichtigung der Randsummen. (Quelle: Eigene Darstellung)

In dieser Tabelle ist unter Berücksichtigung der Randsummen ein maximaler Zusammenhang gegeben. Hier kann nun das Phi_{max} ausgerechnet werden, indem einfach wieder die obige Formel angewendet wird.

$$Phi_{max} = \frac{0 \times 10 - 100 \times 90}{\sqrt{100 \times 100 \times 110 \times 90}} = \frac{0 - 9000}{\sqrt{99000000}} = \frac{-9000}{9949.87} = -0.905$$

Jetzt hat man sowohl das Phi als auch das Phi_{max}. Wenn man nun Phi durch Phi_{max} teilt, erhält man das korrigierte Phi_{korr} oder auch den so genannten Äquivalenzkoeffizienten, den man mit den anderen Korrelationskoeffizienten vergleichen kann.

$$\text{Äquivalenzkoeffizient} \quad r_{Äqui} = \frac{Phi}{Phi_{max}} = \frac{-0.101}{-0.905} = 0.112$$

Jetzt kann man sagen, es besteht ein (recht) schwacher Zusammenhang zwischen der Variablen 'Geschlecht' und der Variablen 'Qualität als Psychologe'.

4.4.2 Spearman-Rang-Korrelation

Die Spearman-Rang-Korrelation findet Verwendung, wenn zwei ordinalskalierte Variablen korreliert werden sollen. Ordinalskaliert heißt, dass Ränge vorliegen. Wie vergibt man nun Ränge oder Rangplätze? Angenommen, fünf Leute führen je zwei verschiedene Leistungstests durch und sie erreichen die folgenden Werte:

Proband-Nr.:	Leistungstest A	Leistungstest B
1	14	25
2	16	23
3	18	28
4	11	22
5	13	24

Tabelle 15: Werte von fünf Probanden in zwei unterschiedlichen Leistungstests. (Quelle: Eigene Darstellung)

Gesetzt den Fall, dass ein hoher Wert besser ist als ein niedriger, würde man die Rangplätze so vergeben: Bester Wert = Rangplatz 1, zweitbester Wert = Rangplatz 2 usw. Tabelle 16 listet noch einmal die Werte sowie die Rangplätze auf:

Proband-Nr.:	Leistungstest A	Leistungstest B
1	14 (3.)	25 (2.)
2	16 (2.)	23 (4.)
3	18 (1.)	28 (1.)
4	11 (5.)	22 (5.)
5	13 (4.)	24 (3.)

Tabelle 16: Werte und Rangplätze von fünf Probanden in zwei unterschiedlichen Leistungstests. (Quelle: Eigene Darstellung)

Soweit ganz einfach. Probleme ergeben sich erst dann, wenn mehrere Personen die gleiche Punktzahl haben. Diese müssen sich dann einen Rangplatz teilen. Man spricht dann von verbundenen Rängen. Wie dann die Rangplätze vergeben werden, soll Tabelle 17 darstellen:

Proband-Nr.:	Leistungstest A	Leistungstest B
1	14 (4.)	25 (2.5.)
2	16 (2.)	25 (2.5.)
3	18 (1.)	28 (1.)
4	14 (4.)	22 (5.)
5	14 (4.)	24 (4.)

Tabelle 17: Werte und verbundene Rangplätze von fünf Probanden in zwei unterschiedlichen Leistungstests. (Quelle: Eigene Darstellung)

Als Regel kann man sich merken: Die Summe der verbundenen Rangplätze muss der Summe der getrennten Rangplätze – wenn man solche vergeben hätte – entsprechen. Beim Leistungstest A haben die Personen 1, 4 und 5 als Rangplatz jeweils die 4. (4+4+4=12) bekommen. Hätten für diese Personen getrennte Ränge vergeben werden können, wären das die Rangplätze 3., 4. und 5. (3+4+5=12) gewesen. Beim Leistungstest B haben die Probanden 1 und 2 jeweils den Rangplatz 2.5. (2.5+2.5=5) bekommen. Das entsprach den Rangplätzen 2. und 3. (2+3=5).
Nun aber endlich die Formel:

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot \left(n^2 - 1\right)}$$

r_s = Spearman-Rang-Korrelationskoeffizient
d^2 = Differenz zwischen den Rangplätzen zum Quadrat
n = Anzahl der Personen

Es muss also noch die **Differenz zwischen den Rangplätzen** (nicht zwischen den Werten!) gebildet und ins Quadrat gesetzt werden. Nehmen wir noch einmal die erste Tabelle:

	Leistungstest A		Differenz		Leistungstest B	
Pb-Nr.:	**Wert**	**Rang**	**di**	d_i^2	**Wert**	**Rang**
1	14	3.	1	1	25	2.
2	16	2.	-2	4	23	4.
3	18	1.	0	0	28	1.
4	11	5.	0	0	22	5.
5	13	4.	1	1	24	3.
n = 5				$\sum_{i=1}^{6} d_i^2 = 6$		

Tabelle 18: Werte, Rangplätze und Differenzen von fünf Probanden in zwei unterschiedlichen Leistungstests. (Quelle: Eigene Darstellung)

$$r_s = 1 - \frac{6 \cdot 6}{5 \cdot (25-1)} = 1 - \frac{36}{120} = 1 - 0.3 = 0.7$$

Hier hätte man also eine mittlere bis hohe positive Korrelation zwischen den Rangplätzen. Wer in dem einen Leistungstest einen relativ hohen (niedrigen) Wert hat, hat auch in dem anderen Leistungstest einen relativ hohen (niedrigen) Wert.

Soweit recht einfach. Wenn jedoch viele verbundene Ränge auftauchen (als Regel kann man sich merken: Mehr als 20%), kann diese Formel nicht mehr verwendet werden, sondern es muss eine Korrekturformel angewandt werden:

$$r_s = \frac{2 \cdot \left(\frac{n^3 - n}{12}\right) - T - U - \sum_{i=1}^{n} d_i^2}{2 \cdot \sqrt{\left(\frac{n^3 - n}{12} - T\right) \cdot \left(\frac{n^3 - n}{12} - U\right)}} \qquad T = \sum_{j=1}^{k(x)} \frac{t_j^3 - t_j}{12}, \; U = \sum_{j=1}^{k(y)} \frac{u_j^3 - u_j}{12}$$

t = Anzahl der Leute, die sich einen gemeinsamen Rangplatz in der Variablen x teilen;
u = Anzahl der Leute, die sich einen gemeinsamen Rangplatz in der Variablen y teilen;
k(x) und k(y) = Anzahl der verbundenen Ränge in der Variablen x bzw. y.

Natürlich, ein Beispiel:

	Leistungstest A (x)		Differenz		Leistungstest B (y)	
Pb.-Nr.:	**Wert**	**Rang**	**di**	d_i^2	**Wert**	**Rang**
1	16	2.5.	1.5	2.25	25	4.
2	16	2.5.	-2	4	23	4.5.
3	18	1.	-3.5	12.25	23	4.5.
4	11	5.	0.5	0.25	23	4.5.
5	11	5.	0.5	0.25	23	4.5.
6	11	5.	3	9	24	2.
n = 6				$\sum_{i=1}^{6} d_i^2 = 28$		

Tabelle 19: Werte, Rangplätze und Differenzen von sechs Probanden in zwei unterschiedlichen Leistungstests. (Quelle: Eigene Darstellung)

Im Leistungstest A (der Variablen x) teilen sich 2 Leute den Rangplatz 2.5. (t1 = 2) und 3 Leute den Rangplatz 5. (t2 = 3).

k(x) = 2
(es gibt zwei verschiedene verbundene Ränge, nämlich zum einen 2.5. und zum anderen 5.)

Im Leistungstest B (der Variablen y) teilen sich 4 Leute den Rangplatz 4.5. (u1 = 4).
k(y) = 1
(es gibt als verbundenen Rangplatz nur den Rang 4.5.)

So, mit diesen Werten kann man nun alles berechnen. Zuerst T und U:

$$T = \sum_{j=1}^{2} \frac{t_j^3 - t}{12} = \frac{2^3 - 2}{12} + \frac{3^3 - 3}{12} = \frac{8-2}{12} + \frac{27-3}{12} = \frac{6}{12} + \frac{24}{12} = 2.5$$

$$U = \sum_{j=1}^{1} \frac{u_j^3 - u}{12} = \frac{4^3 - 4}{12} = \frac{64-4}{12} = \frac{60}{12} = 5$$

Und nun berechnet man r_s:

$$r_s = \frac{2 \cdot \left(\frac{6^3-6}{12}\right) - 2.5 - 5 - 28}{2 \cdot \sqrt{\left(\frac{6^3-6}{12} - 2.5\right) \times \left(\frac{6^3-6}{12} - 5\right)}} = \frac{2 \cdot \left(\frac{216-6}{12}\right) - 35.5}{2 \cdot \sqrt{\left(\frac{210}{12} - 2.5\right) \cdot \left(\frac{210}{12} - 5\right)}}$$

$$r_s = \frac{2 \cdot 17.5 - 35.5}{2 \cdot \sqrt{(17.5 - 2.5) \cdot (17.5 - 5)}} \frac{35 - 35.5}{2 \cdot \sqrt{15 \cdot 12.5}} = \frac{-0.5}{2 \cdot \sqrt{187.5}} = \frac{-0.5}{27.386} = -0.018$$

In diesem Beispiel hätte man eine sehr geringe negative Korrelation zwischen den beiden Leistungstestergebnissen errechnet.

Das wäre also der Rechenweg, wenn mehr als 20% verbundene Ränge/Rangplätze vorliegen. Einfacher kann man es sich machen, wenn man (mit einem Taschenrechner!) eine Produkt-Moment-Korrelation über die Rangplätze – nicht über die ursprünglichen Werte!! – berechnet. Das Ergebnis stimmt, bis auf kleinere Rundungsungenauigkeiten, überein.

4.4.3 Produkt-Moment-Korrelation

Covarianz: $$COV_{xy} = \frac{\sum_{i=1}^{n} [(x_i - \bar{x}) \cdot (y_i - \bar{y})]}{n}$$

Korrelationskoeffizient: $$r_{PM} = \frac{COV_{xy}}{S_x \cdot S_y} = \frac{n \cdot \sum_{i=1}^{n} (x_i \cdot y_i) - \left(\sum_{i=1}^{n} x_i\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right] \cdot \left[n \cdot \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}}$$

Die Formeln sehen nicht schlecht aus, oder? Aber, nichts wird so heiß gegessen, wie es gekocht wird. Folgendes Problem: Man habe aus vielen Café-, Kneipen- und Disco-Besuchen die These entwickelt, dass eine Person als umso angesehener und sympathischer gilt, je mehr diese Person redet (unabhängig vom Inhalt). Man will nun prüfen, ob dieser intuitive Zusammenhang auch wirklich besteht. Dazu hat man bei zehn Personen einen Sympathie-Wert und einen Redefluss-

Wert (Worte pro Minute) erhoben. Beide Variablen seien intervallskaliert und normalverteilt. Je größer die Werte, desto sympathischer, bzw. redseliger sind die Personen.

Die folgende Tabelle listet die Werte auf:

Pb-Nummer:	Sympathie-Wert (x)	Redefluss-Wert (y)
1	10	21
2	8	16
3	11	22
4	7	13
5	13	27
6	12	23
7	10	19
8	9	18
9	11	23
10	10	20

Tabelle 20: Sympathie- und Redefluss-Werte bei zehn Probanden.
(Quelle: Eigene Darstellung)

Zuerst berechnet man jeweils den Mittelwert: $\bar{x} = 10.1 \qquad \bar{y} = 20.2$

Danach plustert man die obige Tabelle auf, indem man sie um einige Spalten erweitert:

Pb-Nr.	x	x^2	$(x-\bar{x})^2$	y	y^2	$(y-\bar{y})^2$	x×y	$(x-\bar{x})\times(y-\bar{y})$
1	10	100	$(-0.1)^2$	21	441	$(0.8)^2$	210	-0.08
2	8	64	$(-2.1)^2$	16	256	$(-4.2)^2$	128	8.82
3	11	121	$(0.9)^2$	22	484	$(1.8)^2$	242	1.62
4	7	49	$(-3.1)^2$	13	169	$(-7.2)^2$	91	22.32
5	13	169	$(2.9)^2$	27	729	$(6.8)^2$	351	19.72
6	12	144	$(1.9)^2$	23	529	$(2.8)^2$	276	5.32
7	10	100	$(-0.1)^2$	19	361	$(-1.2)^2$	190	0.12
8	9	81	$(-1.1)^2$	18	324	$(-2.2)^2$	162	2.42
9	11	121	$(0.9)^2$	23	529	$(2.8)^2$	253	2.52
10	10	100	$(-0.1)^2$	20	400	$(-0.2)^2$	200	0.02
Σ=	101	1049	28.9	202	4222	141.6	2103	62.8

Tabelle 21: x- und y-Werte sowie weitere Zwischenergebnisse zur Berechnung der Produkt-Moment-Korrelation.
(Quelle: Eigene Darstellung)

In der Tabelle stehen nun alle Werte, um den Produkt-Moment-Korrelationskoeffizienten r_{PM} zu errechnen. Entweder berechnet man diesen über die Covarianz COV_{xy} und die Standardabweichungen oder über die andere Formel. Hier wird beides dargestellt:

a) Berechnung über Covarianz und Standardabweichungen:

$$S_x^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} = \frac{28.9}{10} = 2.89 \Rightarrow S_x = 1.7; \quad S_y^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n} = \frac{141.6}{10} = 14.16 \Rightarrow S_y = 3.763$$

$$COV_{xy} = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\cdot(y_i-\bar{y})]}{n} = \frac{62.8}{10} = 6.28; \quad r_{PM} = \frac{COV_{xy}}{S_x \cdot S_y} = \frac{6.28}{1.7 \cdot 3.763} = 0.9817$$

b) Berechnen über andere Formel

$$r_{PM} = \frac{n \cdot \sum_{i=1}^{n}(x_i \cdot y_i) - \left(\sum_{i=1}^{n} x_i\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right] \cdot \left[n \cdot \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} = \frac{10 \cdot 2103 - 101 \cdot 202}{\sqrt{\left[10 \cdot 1049 - (101)^2\right] \cdot \left[10 \cdot 4222 - (202)^2\right]}}$$

$$r_{PM} = \frac{21030 - 20402}{\sqrt{[10490 - 10201] \cdot [42220 - 40804]}} = \frac{628}{\sqrt{289 \cdot 1416}} = \frac{628}{639.706} = 0.9817$$

Welche Formel man nimmt, ist Geschmackssache. Wenn man sich für eine Formel entscheidet, reduziert sich natürlich auch die obige Tabelle.

Nun aber zur Interpretation.

$r_{PM} = 0.9817$ positiv

$r_{PM}^2 = 0.9637 \Rightarrow r_{PM}^2 \times 100\% = 96.37\%$ determinierte (erklärte, erklärbare) Varianz.

Es besteht also ein Zusammenhang von 96.37% determinierter Varianz. Der Zusammenhang kann höchstens 100% betragen, d.h. der Zusammenhang hier ist sehr hoch! Die zu 100 fehlenden Prozent (hier 3.63%) nennt man freie Variation/Varianz oder auch Fehlervarianz. Des Weiteren ist der Zusammenhang positiv, d.h. jemand mit einem hohen Sympathie-Wert hat auch einen hohen Redefluss-Wert, bzw. jemand mit einem niedrigen Redefluss-Wert hat einen niedrigen Sympathie-Wert. Kann man hieraus folgern, dass jemand einen hohen Sympathie-Wert hat, weil er viel redet, oder kann man hieraus folgern, dass jemand viel redet, weil er einen hohen Sympathie-Wert hat? Nein! Weder das eine noch das andere. Man kann aus diesem Korrelationskoeffizient nur ablesen, dass ein Zusammenhang besteht. Eine ursächliche (kausale) Interpretation ist nicht möglich! Es kann natürlich eine kausale Beziehung geben, mit der Korrelation kann man dies aber nicht feststellen! Natürlich wird man trotzdem beim nächsten Café, Kneipen- oder Disco-Besuch möglichst viel reden, in der Hoffnung, dadurch sympathischer zu wirken.

Wie wäre es denn zu interpretieren gewesen, hätte man einen negativen Korrelationskoeffizienten errechnet?

$r_{PM} = -0.9817$ negativ

$r_{PM}^2 = 0.9637 \Rightarrow r_{PM}^2 \times 100\% = 96.37\%$ determinierte Varianz

Es wäre immer noch ein Zusammenhang von 96.37% erklärter Varianz gewesen, also ein sehr hoher Zusammenhang. Allerdings wäre dieser Zusammenhang negativ, d.h. jemand mit einem hohen Sympathie-Wert hätte einen niedrigen Redefluss-Wert, bzw. jemand mit einem niedrigen Sympathie-Wert hätte einen hohen Redefluss-Wert gehabt.

Auch hier ist zu beachten, dass eine kausale Interpretation auch hier nicht sinnvoll ist!

Wenn man – ausgehend von r_{PM} = 0.9817 – schätzen wollte, welchen Sympathie-Wert jemand hat, der einen bestimmten Redefluss-Wert XX hat, wie hätte man diesen geschätzt? Da der Zusammenhang sehr groß ist, würde man beim Schätzen nur einen geringen Fehler machen (Fehlervarianz = 3.63%). Die folgende Abbildung soll dies verdeutlichen der der Fehlerbereich ist in der Abbildung mit eingezeichnet.

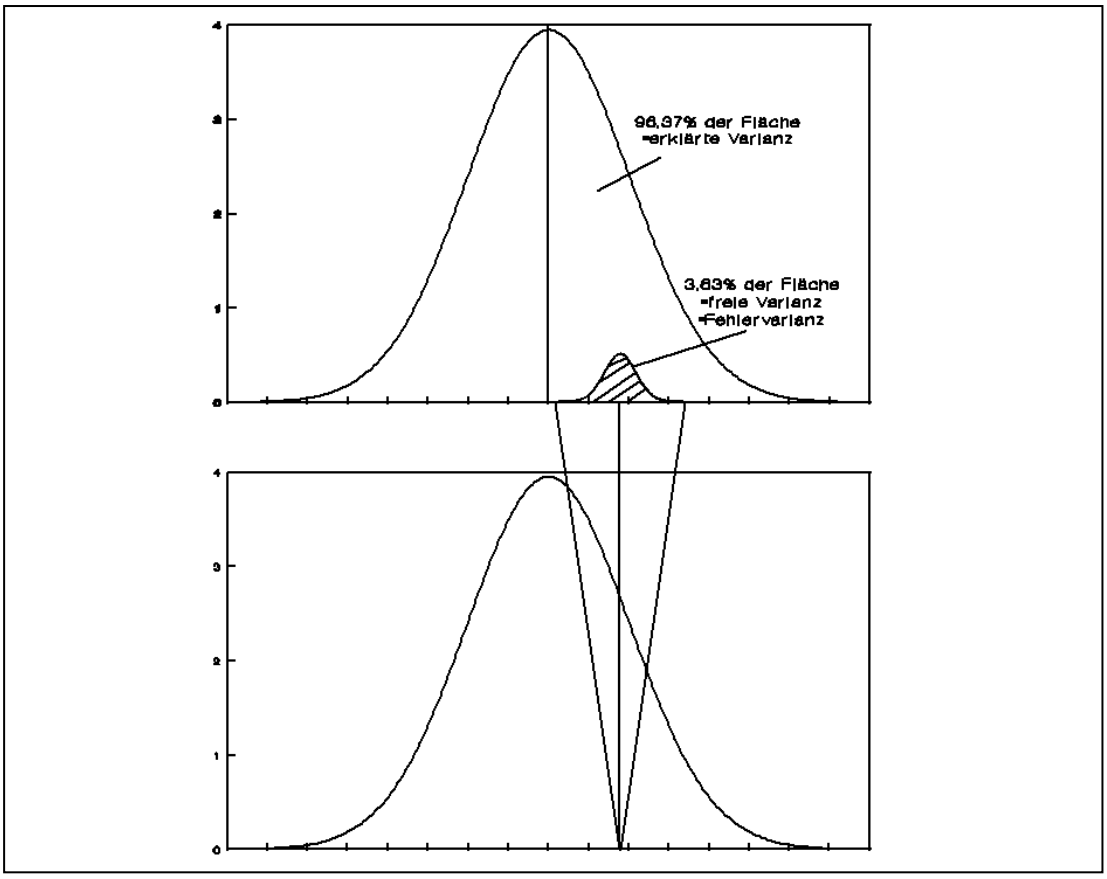

Abbildung 8: Schätzung bei einer Korrelation von r = 0.9817.
(Quelle: Eigene Darstellung)

Die untere Kurve stellt die Verteilung der Redefluss-Werte dar, die obere Kurve die Verteilung der Sympathie-Werte.

Wie hätte es denn ausgesehen, wenn die Korrelation nicht 0.9817 sondern nur 0.8 gewesen wäre?

r = 0.8 $\Rightarrow$ r^2 = 0.64 $\Rightarrow$ 64% determinierte Varianz $\Rightarrow$ 100% – 64% = 36% freie Varianz oder auch Fehlervarianz (Abbildung 9).

Wie hätte es denn ausgesehen, wenn die Korrelation nicht 0.8 sondern nur 0.5 gewesen wäre?

r = 0.5 $\Rightarrow$ r^2 = 0.25 $\Rightarrow$ 25% determinierte Varianz $\Rightarrow$ 100% - 25% = 75% freie Varianz oder auch Fehlervarianz (Abbildung 10).

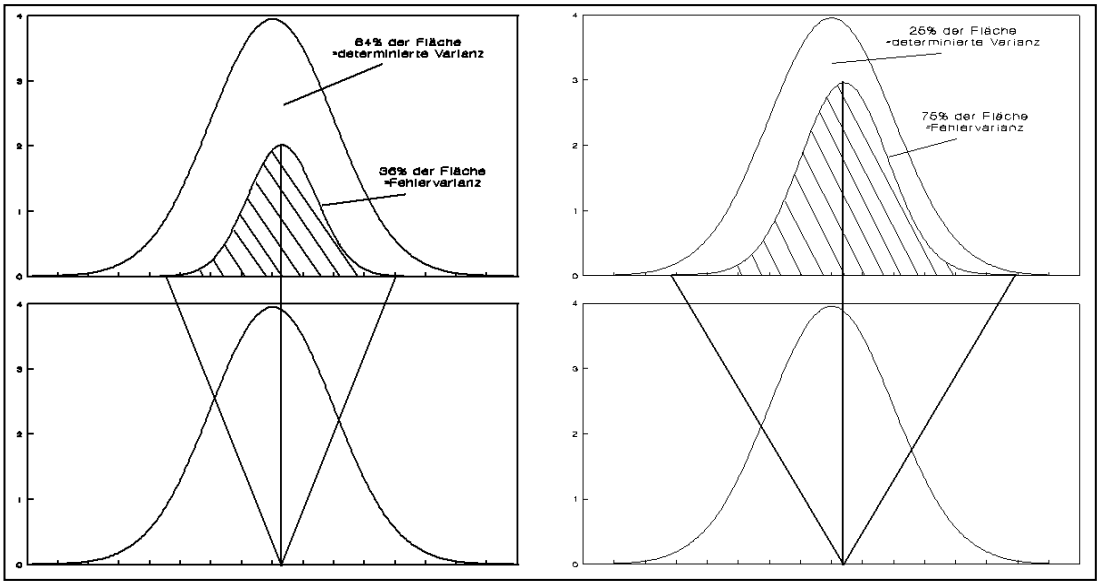

Abbildung 9: Schätzung bei einer Korrelation von r = 0.8 bzw. r = 0.
(Quelle: Eigene Darstellung)

Wie hätte es denn ausgesehen, wenn die Korrelation nicht 0.5 sondern nur 0.3 gewesen wäre?

$r = 0.3 \Rightarrow r^2 = 0.09 \Rightarrow 9\%$ determinierte Varianz $\Rightarrow 100\% - 9\% = 91\%$ freie Varianz oder auch Fehlervarianz (Abb. 11).

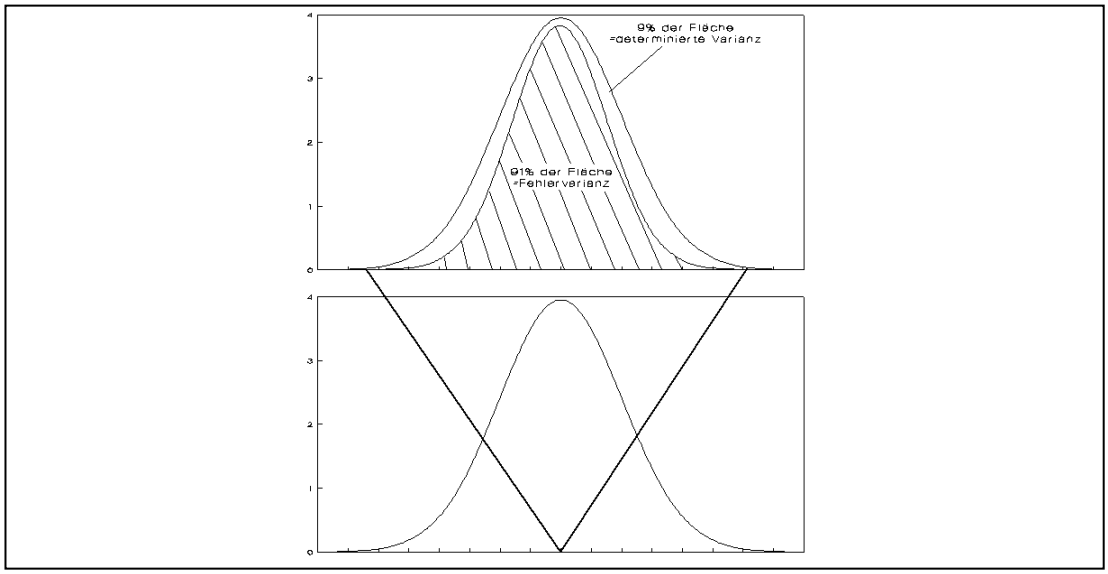

Abbildung 10: Schätzung bei einer Korrelation von r = 0.3.
(Quelle: Eigene Darstellung)

Übungsaufgaben zu Kapitel 4

016 Nennen und beschreiben Sie kurz drei verschiedene Maße der zentralen Tendenz!

017 Nennen Sie drei Dispersionsmaße!

018 Im Rahmen eines Fragebogens werden u.a. die folgenden Variablen abgefragt: Körpergröße, Zufriedenheit mit dem gekauften Produkt (1 = sehr zufrieden, 2 = zufrieden, 3 = mittelmäßig zufrieden, 4 = unzufrieden, 5 = sehr unzufrieden) und Geschlecht. Welches Skalenniveau liegt bei den einzelnen Variablen jeweils vor? Nennen Sie Lagemaße zu den einzelnen Variablen, deren Angabe bzw. Berechnung aufgrund des jeweiligen Skalenniveaus sinnvoll sind.

019 In einer Stichprobe von n = 10 Personen wurde die Körpergröße der Probanden wie folgt festgestellt : x_1=175; x_2=182; x_3=164; x_4=170; x_5=189; x_6=192; x_7=164; x_8=176; x_9=178; x_{10}=160 [cm]. Bestimmen Sie den Modus, den Median und das arithmetische Mittel dieser Stichprobe.

020 Wie würden Sie ein $r^2 = -0.90$ interpretieren?

021 Als Leiter(in) der Personalabteilung haben Sie auch die Aufgaben als Frauenbeauftragte(r) übernommen. Nun interessieren Sie sich dafür, wie es um den Frauenanteil in den verschiedenen Abteilungen der Firma, in der Sie arbeiten, bestellt ist. Als erstes betrachten Sie die Abteilungen „EDV“ und „Personal“ und notieren jeweils das Geschlecht der dort Arbeitenden. Heraus kommt folgende Tabelle:

Mitarbeiter Nr.	Abteilung	Geschlecht
1	EDV	w(eiblich)
2	Personal	m(ännlich)
3	Personal	w
4	EDV	m
5	Personal	w
6	Personal	w
7	EDV	m
8	EDV	m

Besteht ein Zusammenhang zwischen den Variablen „Geschlecht“ und „Abteilungszugehörigkeit“?

Raum für Notizen

5 Einfache Inferenzstatistik

Lernziele

Wenn Sie dieses Kapitel durchgearbeitet haben, dann wissen Sie,

- wie ein Chi²-Test funktioniert;
- was ein F-Test macht und wie er durchgeführt wird;
- wie man einen Mann-Whitney-U-Test rechnet;
- welche Arten eines t-Tests es gibt, wie man die jeweilige t-Test-Variante rechnet und wann sie angewendet wird.

Inferenzstatistik oder auch „schließende Statistik" meint den Schluss von der Stichprobe auf die Population!!! Da es praktisch unmöglich ist, immer alle Menschen zu untersuchen, wenn man wissen will, wie ein bestimmtes Merkmal – z.B. Konzentrationsleistung, Fähigkeit, Statistik zu begreifen oder ähnliches – verteilt ist, also welchen Mittelwert es gibt und welche Standardabweichung, bleibt einem nur die Möglichkeit, mit einer Stichprobe zu arbeiten. Stichproben haben aber den unangenehmen Nachteil, die Population/Bevölkerung/Grundgesamtheit nie völlig genau zu repräsentieren. Die Werte in einer Stichprobe werden praktisch immer anders sein als die Werte der Grundgesamtheit.

Wenn man jetzt die Ergebnisse zweier Stichproben miteinander vergleicht, kann man dies

- direkt tun; dann sind etwaige Unterschiede zwischen den Stichproben aber nicht auf die Population übertragbar. Die Ergebnisse beziehen sich nur auf die Stichproben!
- auf Populationsebene tun, d.h. man rechnet von den Stichprobenergebnissen auf die Population hoch und vergleicht erst dann. Hier sind die Ergebnisse dann auch auf die Population übertragbar.

Im Allgemeinen vergleicht man immer auf Populationsebene. Wenn man wissen will, ob es einen Unterschied zwischen zwei Lehrmethoden gibt, möchte man dies ja schließlich auch auf alle Schüler generalisieren und nicht bloß für die Schüler der Schule XYZ in der Stadt ABC.

Alle Signifikanztests prüfen auf Populationsebene, d.h. von den Kennwerten der Stichprobe (Mittelwert, Varianz etc.) wird immer auf die Kennwerte der Population geschätzt und erst dann wird verglichen. Dazu wird ein inferenzstatistischer Kennwert gebildet. Beim Mittelwertevergleich ist das z.B. der sogenannte t-Wert. Man kann nun aus der Theorie ableiten, wie sich t-Werte verteilen. Die Verteilung eines inferenzstatistischen Kennwertes ist allerdings von der Stichprobengröße abhängig. Allgemein kann man jedoch sagen: Bei sehr großen Stichproben verteilen sich die inferenzstatistischen Kennwerte normal, d.h. sie folgen einer Normalverteilung. Da man aus der Theorie die Verteilung der Werte ableiten kann, kann man auch die Auftretenswahrscheinlichkeit für jeden Wert errechnen. Die Auftretenswahrscheinlichkeit meint die Wahrscheinlichkeit dafür, dass ein Wert rein zufällig auftritt. Das Errechnen der Auftretenswahrscheinlichkeit ist allerdings relativ umständlich. Daher verwendet man Tabellen, die für verschiedene Stichprobengrößen den Wert enthalten, dessen Auftretenswahrscheinlichkeit $p = 0.05$, d.h. 5% beträgt. Diese Werte nennt man kritische Werte. Aus den Stichprobenergebnissen errechnet man nun einen inferenzstatistischen Kennwert. Ist dieser empirische – d.h. aus den Stichprobenergebnissen errechnete – Wert größer als der in der Tabelle stehende kritische Wert, ist die Wahrscheinlichkeit für den empirischen Wert kleiner als $p = 0.05$, d.h. der Wert ist signifikant!

5.1 Zum Begriff der Signifikanz

Signifikanz besteht, wenn man nicht glaubt, dass etwas nur aus Zufall passiert ist. Um noch einmal das Beispiel mit dem Münzenwerfen aufzugreifen: Wenn jemand 10mal hintereinander Kopf wirft, dann kann das zwar aus Zufall passiert sein, aber das müsste schon ein großer Zufall sein. Viel eher stimmt etwas mit der Münze nicht!

Wenn dahingegen jemand 3mal hintereinander Kopf wirft, denkt man noch: Naja, Zufall ... Auf alle Fälle würde man noch nicht vermuten, dass mit der Münze etwas nicht stimmt.

Was hier passiert, ist, dass man die Größe des Zufalls intuitiv bewertet. 3mal hintereinander Kopf zu werfen, ist mehr oder minder normal, bzw. nur ein geringer Zufall. 10mal hintereinander Kopf zu werfen, ist überhaupt nicht mehr normal und wäre schon ein sehr großer Zufall. Und weil es ein sehr großer Zufall wäre, vermutet man, dass es eben kein Zufall mehr ist, sondern dass etwas mit der Münze nicht in Ordnung ist.

Allgemein hat man sich geeinigt, ein Ereignis (wie z.B. 10mal hintereinander Kopf zu werfen) dann als signifikant zu bezeichnen, wenn die (Zufalls-) Wahrscheinlichkeit für dieses Ereignis kleiner als 5% ist.

5.2 Chi²-Test

Ein χ^2-Test vergleicht beobachtete Häufigkeiten mit erwarteten Häufigkeiten!!!

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_b - f_e)^2}{f_e} \qquad \chi^2 = \frac{n \times (a \times d - b \times c)^2}{(a+b) \times (c+d) \times (a+c) \times (b+d)}$$

Dieses Symbol ist kein X sondern ein griechisches χ und heißt Chi! Man spricht also von Chi-Quadrat-Tests.

Beispiel: Man beobachtet einen Tag lang, wie oft in einem Kaufhaus welches Waschmittel wie oft gekauft wird. In diesem Kaufhaus gibt es die Waschmittel A, B und C.

Kategorie	A	B	C
Wie oft gekauft:	100	80	120
Insgesamt wurden 300 Waschmittel verkauft.			

Tabelle 22: Verkaufshäufigkeiten von drei Waschmitteln.
(Quelle: Eigene Darstellung)

Die Frage ist nun, ob die unterschiedlichen Verkaufszahlen der Waschmittel signifikant sind oder ob sie doch im Schnitt gleich häufig verkauft werden.

H0: Die Waschmittel werden gleichhäufig verkauft.
H1: Die Waschmittel werden nicht gleichhäufig verkauft.
(Dies sind ungerichtete, zweiseitige Hypothesen!)

Hat also der hier sichtbare Unterschied auch dann Bestand, wenn man nicht nur einen Tag lang auszählt, sondern z.B. ein Jahr lang auszählen würde?

Wie gesagt, der χ^2 -Test vergleicht beobachtete mit erwarteten Häufigkeiten. Die beobachteten Häufigkeiten stehen in der Tabelle. Wie werden nun die erwarteten Häufigkeiten errechnet?

Ganz einfach: Gesamtzahl (300) geteilt durch die Anzahl an Kategorien (A, B, C = 3) = $\frac{300}{3} = 100$

Erwartet hätte man, dass jeweils 100mal Waschmittel A, B und C verkauft werden.

Die Formel für den χ^2-Test lautet: $\chi^2 = \sum_{i=1}^{k} \frac{(f_b - f_e)^2}{f_e}$ df = k - 1 = 3 - 1 = 2

f_b = Beobachtete Häufigkeit (f = Frequency)
f_e = Erwartete Häufigkeit
k = Anzahl der Kategorien (hier 3)

$$\chi^2_{emp} = \frac{(100-100)^2}{100} + \frac{(80-100)^2}{100} + \frac{(120-100)^2}{100} = 8$$

Als empirischen χ^2-Wert errechnet man also 8.

In Tab. D im tabellarischen Anhang findet man kritische χ^2-Werte.
Der kritische χ^2-Wert ist $\chi^2_{krit(\alpha=0.05;df=2)} = 5.991$.

$\chi^2_{emp} > \chi^2_{krit} \Rightarrow$ signifikant $\Rightarrow$ H1.

Mit einer Irrtumswahrscheinlichkeit von 5% kann behauptet werden, dass die Waschmittel nicht gleich häufig verkauft werden. Bezogen auf die Tabelle würde man also sagen: Waschmittel C wird signifikant häufiger und Waschmittel B signifikant seltener verkauft als man es erwartet hätte.

Ein anderes Beispiel: Man hat bei zweihundert Psychologen das Geschlecht ermittelt, sowie jeweils eingeteilt, ob es sich um gute oder schlechte Psychologen handelt. Die folgende Tabelle listet das Ergebnis auf:

	Frauen	**Männer**	
gute Psychologen	60 a	50 b	110 gute Psych. insgesamt
schlechte Psychologen	40 c	50 d	90 schlechte Psych. insgesamt
	100 Frauen insgesamt	100 Männer insgesamt	200 Personen insgesamt

Tabelle 23: Geschlecht und Qualität von zweihundert Psychologen.
(Quelle: Eigene Darstellung)

Die Frage ist nun, ob es zwischen Frauen und Männern gleich viele gute bzw. schlechte Psychologen gibt oder ob Unterschiede bestehen.

- H0: Die Verteilung von Frauen und Männern bezüglich der Qualität als Psychologen (gut bzw. schlecht) unterscheidet sich nicht voneinander.
- H1: Die Verteilung von Frauen und Männern bezüglich der Qualität als Psychologen (gut bzw. schlecht) unterscheidet sich voneinander.

Man könnte jetzt davon ausgehen, dass es sowohl gleich viele gute wie schlechte Psychologen gibt und dass diese sich gleich auf Frauen und Männer verteilen. Man könnte also für jedes der vier Felder den Wert 50 annehmen.

	Frauen	Männer	
gute Psychologen	50	50	100
schlechte Psychologen	50	50	100
	100	100	200

Tabelle 24: Erwartete Werte (Gleichverteilung).
(Quelle: Eigene Darstellung)

Wenn man so vorgeht, ignoriert man aber die Tatsache, dass es insgesamt mehr gute Psychologen (110) als schlechte Psychologen (90) gab. Stattdessen sollte man folgendermaßen vorgehen:

	Frauen	Männer	
gute Psychologen	a	b	110
schlechte Psychologen	c	d	90
	100	100	200

Tabelle 25: Randsummen.
(Quelle: Eigene Darstellung)

Aus diesen Randsummen bildet man die einzelnen Wahrscheinlichkeiten:

p(Frau) = 100 von 200 = 0.5
p(Mann) = 100 von 200 = 0.5
p(gut) = 110 von 200 = 0.55
p(schlecht) = 90 von 200 = 0.45

Wie groß ist denn p dafür, dass die Ereignisse Frau und gut gleichzeitig eintreffen?
p(Frau und gut) = p(Frau) × p(gut) = 0.5 × 0.55 = 0.275 = 27.5%

Und für die anderen Kombinationen:

p(Frau und schlecht) = 0.5 × 0.45 = 0.225 = 22.5%
p(Mann und gut) = 0.5 × 0.55 = 0.275 = 27.5%
p(Mann und schlecht) = 0.5 × 0.45 = 0.225 = 22.5%

Mit diesen Prozentangaben kann man nun berechnen, wie viele Personen pro Zelle (Feld) es denn sind. Einfacher geht es, wenn man die Wahrscheinlichkeit mit der Gesamtanzahl (hier: 200) multipliziert. Dies liefert die erwarteten Werte:

Feld a (Frau und gut) = 0.275 × 200 = 55
Feld b (Frau und schlecht) = 0.225 × 200 = 45
Feld c (Mann und gut) = 0.275 × 200 = 55
Feld d (Mann und schlecht) = 0.225 × 200 = 45

	Frauen	Männer	
gute Psychologen	55	55	110
schlechte Psychologen	45	45	90
	100	100	200

Tabelle 26: Erwartete Häufigkeiten unter Berücksichtigung der Randsummen.
(Quelle: Eigene Darstellung)

Trägt man diese Werte in die Tabelle ein, sieht man, dass die Randsummen erhalten bleiben. Das Berücksichtigen der Randsummen hat jedoch auch einen kleinen Nachteil: Wie viele Zahlen kann man in dieses Vierfelderschema frei einsetzen, wenn die Randsummen erhalten bleiben sollen? Nur eine einzige: Die anderen ergeben sich hier automatisch!

Würde man in das Feld a (Frau und gut) z.B. den Wert 70 einsetzen, müsste man in das Feld b (Mann und gut) den Wert 40 einsetzen, damit als Randsumme 110 herauskommt. Weiterhin müsste man in das Feld c (Frau und schlecht) den Wert 30 eintragen, damit sich als Randsumme 100 ergibt. In das letzte Feld d (Mann und schlecht) müsste der Wert 60 eingetragen werden.

Allgemein kann man sagen: Werden die erwarteten Werte über die Randsummen errechnet, berechnen sich die Freiheitsgrade als df = (Anzahl Spalten -1) × (Anzahl Zeilen -1)!

Für das Beispiel hier wären die Freiheitsgrade also: df = (2 - 1) × (2 - 1) = 1

Da man jetzt beobachtete und erwartete Werte hat, kann man ganz normal die χ^2-Formel anwenden:

$$\chi^2_{emp} = \sum_{i=1}^{k} \frac{(f_b - f_e)^2}{f_e} = \frac{(60-55)^2}{55} + \frac{(50-55)^2}{55} + \frac{(40-45)^2}{45} + \frac{(50-45)^2}{45} = 2.02$$

Als kritischen χ^2-Wert liest man aus der Tabelle: $\chi^2_{krit(\alpha=0.05;df=1)} = 3.84$

$\chi^2_{emp} < \chi^2_{krit} \Rightarrow$ nicht signifikant $\Rightarrow$ H0.

Es kann nicht behauptet werden, dass die Verteilung von Frauen und Männern bezüglich der Qualität als Psychologen (gut bzw. schlecht) sich voneinander unterscheidet.

Diese Anordnung der Variablen (2×2) ist unter dem Namen Vierfelder-Schema bekannt geworden. Dementsprechend spricht man manchmal auch vom Vierfelder-Test. Für dieses Vierfelder-Schema hat man auch eine eigene Formel entwickelt:

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

a, b, c und d bezeichnen die einzelnen Felder (siehe oben). Auch bei dieser Formel werden die Randsummen berücksichtigt (nichts anderes steht ja im Nenner!). D.h., auch hier ist der Freiheitsgrad df = 1. n beschreibt die Gesamtanzahl der Personen.

Um die Höhe einer Phi-Korrelation auf Überzufälligkeit zu überprüfen, gibt es den Chi²-Test. Die Formeln:

$$\chi^2 = \frac{n \times (a \times d - b \times c)^2}{(a+b) \times (c+d) \times (a+c) \times (b+d)} = n \times Phi^2; \quad df = 1$$

Um diesen Signifikanztest durchführen zu können, benötigt man also entweder eine Vier-Felder-Tafel oder das bereits berechnete Phi sowie die Anzahl der Versuchspersonen (n).

Verwenden wir das Beispiel aus dem Kapitel „Phi-Korrelation". Da bei der Phi-Korrelation mehrere Phi-Werte berechnet wurden (Phi, Phi_{max}, Phi_{korr}) stellt sich als erstes die Frage, welches Phi denn nun benötigt wird.

Es ist das allererste, unkorrigierte Phi! Phi = 0,101, n=200.

Nun denn: Besteht zwischen der Qualität als Psychologe und dem Geschlecht ein Zusammenhang? Zuerst sollen wieder die Hypothesen aufgestellt werden:

H0: Zwischen der Qualität als Psychologe und dem Geschlecht besteht kein Zusammenhang
H1: Zwischen der Qualität als Psychologe und dem Geschlecht besteht ein Zusammenhang

oder mathematisch: $H_o : \varphi = 0$
$H_1 : \varphi \neq 0$

Berechnen wir den empirischen Chi²-Wert:

$\chi^2_{emp} = n \times Phi^2 = 200 \times 0{,}101^2 = 200 \times 0{,}0102 = 2{,}04$ df = 1

Als nächstes benötigen wir den kritischen Chi²-Wert bei alpha = 5%, zweiseitig und df=1 aus Tabelle D: $Chi^2_{krit(alpha=5\%,\ zweiseitig,\ df=1)} = 3.841$.

Der empirische Chi²-Wert ist kleiner als der kritische Chi²-Wert, also nicht signifikant, also H0! Zwischen Geschlecht und Qualität als Psychologe besteht kein statistisch bedeutsamer Zusammenhang.

5.3 U-Test nach Mann-Whitney

Ein U-Test vergleicht Rangplatzsummen!

$$U = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - T_1 \qquad U' = n_1 \cdot n_2 - U \qquad \mu_U = \frac{n_1 \cdot n_2}{2} \qquad z = \frac{U - \mu_U}{\sigma_U}$$

$$\sigma_{Ukorr} = \sqrt{\frac{n_1 \cdot n_2}{N \cdot (N-1)} \cdot \left(\frac{N^3 - N}{12} - \sum_{i=1}^{k} \frac{t_i^3 - t_i}{12} \right)} \qquad \sigma_U = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}$$

Zwei Sprintmannschaften A und B treten mit je zwei Startern (A1, A2, B1, B2) an.

Welche Zieleinläufe kann es nun geben?

Möglichkeit	1. Platz	2. Platz	3. Platz	4. Platz
1	A1	A2	B1	B2
2	A1	A2	B2	B1
3	A2	A1	B1	B2
4	A2	A1	B2	B1
5	A1	B1	A2	B2
6	A1	B2	A2	B1
7	A2	B1	A1	B2
8	A2	B2	A1	B1
9	A1	B1	B2	A2
10	A1	B2	B1	A2
11	A2	B1	B2	A1
12	A2	B2	B1	A1
13	B1	A1	A2	B2
14	B2	A1	A2	B2
15	B1	A2	A1	B2
16	B2	A2	A1	B1
17	B1	B2	A1	A2
18	B2	B1	A1	A2
19	B1	B2	A2	A1
20	B2	B1	A2	A1
21	B1	A1	B2	A2
22	B2	A1	B1	A2
23	B1	A2	B2	A1
24	B2	A2	B1	A1

Tabelle 27: Mögliche Zieleinläufe bei vier Startern.
(Quelle: Eigene Darstellung)

Bei 24 Möglichkeiten beträgt die Wahrscheinlichkeit für eine dieser Möglichkeiten $p = \frac{1}{24}$

Wenn wir jetzt nicht die einzelnen Läufer, sondern die Mannschaften betrachten:

Wie viele Möglichkeiten gibt es, dass Mannschaft A die ersten beiden Plätze und Mannschaft B die letzten beiden Plätze belegt?

Das sind die Möglichkeiten 1 bis 4.

Wie hoch ist die Zufallswahrscheinlichkeit dafür, dass Mannschaft A die ersten beiden Plätze und Mannschaft B die letzten beiden Plätze belegt?

$$p = \frac{4}{24} = \frac{1}{6} = 0{,}167$$

Gibt es nun eine Möglichkeit, diese große Tabelle zu umgehen? Bzw. allein über die Anzahl der Starter der beiden Mannschaften ausrechnen zu können, wie hoch die Wahrscheinlichkeit für einen deutlichen Sieg der Mannschaft A (d.h. alle von Mannschaft A belegen Plätze vor allen der Mannschaft B) ist?

Bezogen auf die Mannschaften geht es momentan um folgenden Zieleinlauf:
1. A 2. A 3. B 4. B 5. B

Wie viele A's stehen vor den B's?

Naja, vor dem B auf dem dritten Platz stehen 2 A's. Vor dem B auf dem vierten Platz stehen 2 A's. Vor dem B auf dem fünften Platz stehen ebenfalls 2 A's. Man könnte sagen, vor den B's stehen insgesamt 6 A's.

Wie viele B's stehen vor den A's? Nun ja, gar keine: = 0 B's.

Nehmen wir den kleineren der beiden Werte (also 0) und nennen ihn U.

Mannschaft B besteht aus $n_1 = 3$ Startern, Mannschaft A aus $n_2 = 2$ Startern.

Gehen Sie mit diesen drei Werten in die Tabelle E.

Unter $n_1 = 2$, $n_2 = 3$ und U = 0 finden Sie als Wahrscheinlichkeit p = 0,100.
Das ist der Wert, den wir ausgerechnet hatten.

Weil es so schön ist, kommt noch eine Beispielrechnung:

Wieder treten zwei Mannschaften gegeneinander an, diesmal jedoch beide Mannschaften mit jeweils 4 Startern, insgesamt befinden sich also acht Läufer im Wettkampf.

Wie groß ist nun die Wahrscheinlichkeit dafür, dass per Zufall die vier Läufer von Mannschaft A alle vor den vier Läufern der Mannschaft B platziert sind?

Betrachten wir den Mannschaftsbezogenen Zieleinlauf:
1. A 2. A 3. A 4. A 5. B 6. B 7. B 8. B

Wie oft steht ein A vor einem B? Naja, vor dem B auf Platz fünf stehen 4 A's, ebenso vor dem B auf Platz 6, 7 und 8. Insgesamt steht sozusagen 16-mal ein A vor einem B.

Wie oft steht ein B vor einem A? Genau 0-mal.

Nehmen wir wieder den kleineren Wert (0) und nennen ihn U.

Die Mannschaftsgrößen waren: $n_1 = 4$, $n_2 = 4$.

Mit diesen Werten finden wir in Tabelle E eine Wahrscheinlichkeit von p = 0,014.

Gibt es nun eine Möglichkeit, diese U-Werte zu berechnen, ohne jedes Mal ausrechnen zu müssen, wie viele des einen Teams vor wie vielen des anderen Teams stehen?

Diese Möglichkeit gibt es. Bekannt geworden ist sie als **Mann-Whitney-U-Test**.

Die Formeln: $U = n_1 \times n_2 + \frac{n_1 \times (n_1 + 1)}{2} - T_1$ $\qquad U' = n_1 \times n_2 - U$

n_1 und n_2 bezeichnen die jeweiligen Mannschaftsstärken bzw. Stichprobengrößen.
T_1 ist einfach die Summe der erzielten Rangplätze einer der beiden Mannschaften.

Hier hatten wir:
Mannschaft A: $n_1 = 4$, Rangplatzsumme $T_1 = 1 + 2 + 3 + 4 = 10$
Mannschaft B: $n_2 = 4$, Rangplatzsumme $T_2 = 5 + 6 + 7 + 8 = 26$

Als U errechnet man:

$$U = n_1 \times n_2 + \frac{n_1 \times (n_1 + 1)}{2} - T_1 = 4 \times 4 + \frac{4 \times (4 + 1)}{2} - 10 = 16 + \frac{4 \times 5}{2} - 10 = 16 + 10 - 10 = 16$$

Das zweite U (U') berechnet sich als: $U' = n_1 \times n_2 - U = 4 \times 4 - 16 = 16 - 16 = 0$

Das sind die beiden U-Werte, die bereits oben ermittelt wurden.

Und noch eine letzte Rechnung:

Der Mannschaftszieleinlauf gestaltete sich folgendermaßen:
1. A 2. A 3. A 4. B 5. A 6. B 7. B 8. B

Wie hoch ist die Wahrscheinlichkeit dafür, dass sich ein solcher Zieleinlauf per Zufall ergibt?
Mannschaft A: $n_1 = 4$, Rangplatzsumme$T_1 = 1 + 2 + 3 + 5$
Mannschaft B: $n_2 = 4$, Rangplatzsumme $T_2 = 4 + 6 + 7 + 8$

Als U errechnet man:

$$U = n_1 \times n_2 + \frac{n_1 \times (n_1 + 1)}{2} - T_1 = 4 \times 4 + \frac{4 \times (4 + 1)}{2} - 11 = 16 + \frac{4 \times 5}{2} - 11 = 16 + 10 - 11 = 15$$

Das zweite U (U') berechnet sich als: $U' = n_1 \times n_2 - U = 4 \times 4 - 15 = 16 - 15 = 1$

Mit dem kleineren U geht es in Tabelle E. Als einseitige Wahrscheinlichkeit kann man ablesen:
p(U=1) = 0,029.

Die Wahrscheinlichkeit dafür, dass sich per Zufall der oben dargestellte Zieleinlauf ergibt, beträgt also 2,9%. Das ist sehr niedrig. Also wird es wohl kein Zufall sein, sondern z. B. an den Trainingsmethoden liegen.

5.4 F-Test

Ein F-Test vergleicht Varianzen!!!

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

Warum könnte es wichtig sein, Varianzen zu vergleichen?
Dazu sollte man sich noch einmal ins Gedächtnis rufen, was Varianzen aussagen. Varianzen geben Auskunft über die Breite einer Verteilung. Folgende Abbildung soll dies verdeutlichen:

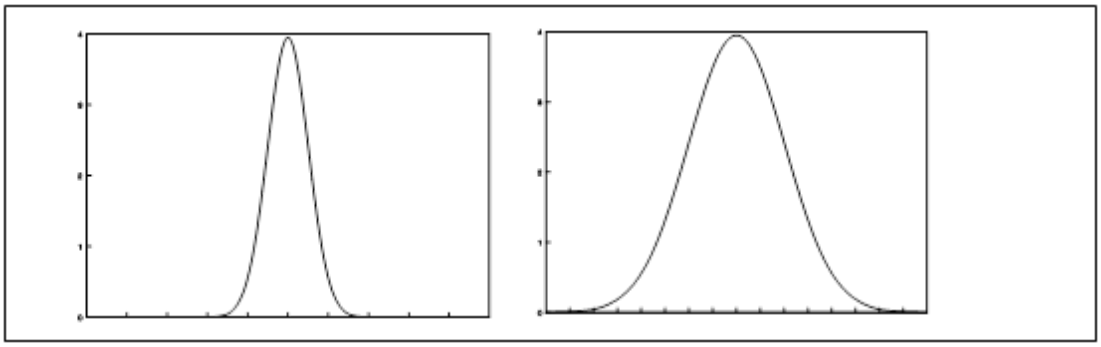

Abbildung 11: Zwei Verteilungen. Die linke Grafik stellt eine Verteilung mit einem Mittelwert von 100 und einer Varianz von 1 dar; bei der rechten Grafik ist ein Mittelwert von 100 und eine Varianz von 100 abgetragen. (Quelle: Eigene Darstellung)

Beide Stichprobenverteilungen haben denselben Mittelwert. Trotzdem würde niemand behaupten, die beiden Verteilungen seien identisch. Die der linken Verteilung zugrunde liegenden Werte sind alle ziemlich ähnlich, während die Werte der rechten Verteilung eine viel größere Streuung haben. Wollte man nun die Mittelwerte miteinander vergleichen, muss man selbstverständlich die Breite oder Streuung der Verteilung mit einbeziehen. Der F-Test prüft nun, ob der Unterschied in der Streuung bzw. Varianz der Verteilungen so groß ist, dass man von Signifikanz reden muss. Diese Prüfung auf Varianzengleichheit (Varianzhomogenität) bzw. -ungleichheit (-heterogenität) ist eine Vorbedingung für den im Anschluss folgenden t-Test. Man könnte auch fragen, ob beide Stichproben zu derselben Population/Grundgesamtheit gehören. Wenn ja, dann dürften sich die Varianzen eigentlich nicht unterscheiden. Unterscheiden sich die Varianzen jedoch, so kann man nicht mehr davon ausgehen, dass beide Stichproben zu derselben Population gehören.

Typischerweise hofft man beim F-Test, dass die Varianzen gleich sind, d.h. man prüft eigentlich auf die Nullhypothese H0. In diesem Fall wählt man häufig einen alpha-Fehler von 10%.

Nun ein Beispiel:
Man hat bei zwei Stichproben die Variable 'Wortflüssigkeit' erhoben. 'Wortflüssigkeit' meint, wie viele Worte einer Person in einer bestimmten Zeit zu einem bestimmten Thema einfallen. Einfachstes Beispiel wäre die Aufgabe, in einer Minute so viele Wörter wie möglich aufzuschreiben, die mit einem 'A' anfangen. Hier nun die Kennwerte:

Stichprobe A: $n_A = 20$ $S_A = 2$
Stichprobe B: $n_B = 15$ $S_B = 4$

Die Fragestellung lautet:
Sind die Varianzen gleich (also homogen) oder unterschiedlich (also heterogen)?

H0: $\sigma_A^2 = \sigma_B^2$ Die Varianzen sind gleich.

H1: $\sigma_A^2 \neq \sigma_B^2$ Die Varianzen sind ungleich.

Zuerst werden aus den Standardabweichungen jeweils die Varianzen gebildet:

$S_A = 2 \Rightarrow S_A^2 = 2 \cdot 2 = 4 \qquad S_B = 4 \Rightarrow S_B^2 = 4 \cdot 4 = 16$

Aus diesen Stichprobenvarianzen werden nun jeweils die geschätzten Populationsvarianzen $\hat{\sigma}^2$ errechnet:

$$\hat{\sigma}_A^2 = S_A^2 \times \frac{n_A}{n_A - 1} = 4 \times \frac{20}{20 - 1} = 4.211$$

$$\hat{\sigma}_B^2 = S_B^2 \times \frac{n_B}{n_B - 1} = 16 \times \frac{15}{15 - 1} = 17.143$$

Aus diesen beiden geschätzten Populationsvarianzen wird nun der empirische F-Wert errechnet. Hierbei gilt, dass die größere der beiden Varianzen immer in den Zähler, d.h. auf den Bruchstrich, gesetzt wird, die kleinere Varianz immer in den Nenner, d.h. unter den Bruchstrich.

$$F_{emp} = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_A^2} = \frac{17.143}{4.211} = 4.071$$

Um den kritischen F-Wert abzulesen, benötigt man zwei Freiheitsgrade, nämlich den Zählerfreiheitsgrad df_1 und den Nennerfreiheitsgrad df_2. Im Zähler steht die Varianz der Stichprobe B, die wiederum aus 15 Personen bestand. Daraus folgt: $df_{Zähler} = 15–1 = 14$. Im Nenner steht die Varianz der Stichprobe A mit 20 Personen. Daraus folgt: $df_{Nenner} = 20–1 = 19$.

Mit diesen beiden Freiheitsgraden geht man nun in Tabelle B und liest dort den kritischen F-Wert ab: $F_{krit} = 2.26$; $F_{emp} > F_{krit} \Rightarrow$ signifikant $\Rightarrow$ H1.

Mit einer Irrtumswahrscheinlichkeit von 10% kann behauptet werden, dass die Varianzen unterschiedlich (heterogen) sind!

5.5 t-Testverfahren

Ein t-Test vergleicht Mittelwerte!!!
Voraussetzungen für den t-Test: Intervallskalenniveau der Variablen und Normalverteilung.

Ein t-Test überprüft, ob ein Unterschied zwischen Mittelwerten bedeutsam (= signifikant) oder nicht bedeutsam (= nicht signifikant) ist. Abbildung 12 soll dies verdeutlichen.

Dazu vermengt der t-Test die Informationen über die Mittelwerte sowie über die jeweilige Standardabweichung und bildet hieraus einen Wert, dessen Auftretenswahrscheinlichkeit nachgeschlagen werden kann. Auftretenswahrscheinlichkeit heißt, mit welcher Wahrscheinlichkeit dieser Wert rein zufällig auftreten kann. Ist diese Auftretenswahrscheinlichkeit kleiner als p = 0.05, dann spricht man von Signifikanz. Das heißt, der vorliegende Mittelwertsunterschied wäre bedeutsam!

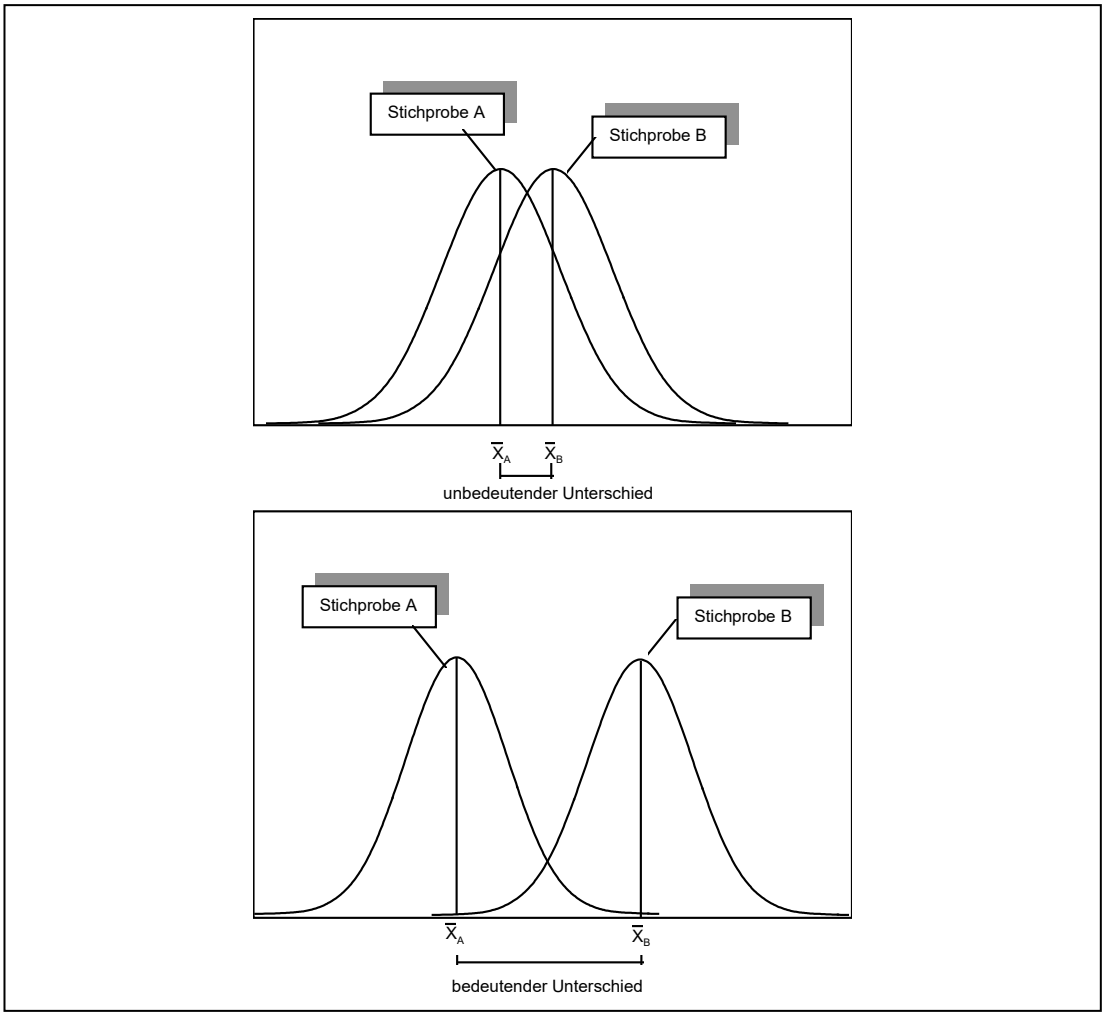

Abbildung 12: Unbedeutende und bedeutende Mittelwertunterschiede.
(Quelle: Eigene Darstellung)

Den t-Test gibt es in mehreren Versionen:

1. Für abhängige Stichproben
2. Für unabhängige Stichproben → bei gleichen (homogenen) Varianzen
   → bei ungleichen (heterogenen) Varianzen

Je nachdem, wie die Ausgangslage ist, müssen für die jeweiligen t-Tests verschiedene Formeln benutzt werden. Die Formeln sind jedoch gar nicht so unterschiedlich:

5.5.1 t-Test für abhängige Stichproben

$$t_{emp} = \frac{\bar{x}_d}{\hat{\sigma}_{\bar{x}d}} \qquad \hat{\sigma}_{\bar{x}d} = \frac{\hat{\sigma}_d}{\sqrt{n}} \qquad \bar{x}_d = \frac{\sum_{i=1}^{n} d_i}{n} \qquad \hat{\sigma}_d = \sqrt{\frac{\sum_{i=1}^{n} d_i^2 - \frac{\left(\sum_{i=1}^{n} d_i\right)^2}{n}}{n-1}}$$

$$df = n - 1$$

Ein Beispiel: Es soll untersucht werden, wie eine neue Therapie bei depressiven Patienten anschlägt. Dazu sollen 10 Depressive vor Beginn der Therapie auf einer Skala von 0 (keine Depression) bis 20 (schwere Depression) den Grad ihrer Depression angeben. Nach der Therapie beurteilen dieselben 10 Patienten noch einmal den Schweregrad ihrer Depression.
Hier sind die Daten:

Patient Nr.	Vor der Therapie	Nach der Therapie	Differenz (d_i)	d_i^-
1	12	10	2	4
2	16	15	1	1
3	17	18	-1	1
4	15	14	1	1
5	17	12	5	25
6	20	14	6	36
7	13	8	5	25
8	16	16	0	0
9	14	13	1	1
10	16	11	5	25
			$\Sigma = 25$	$\Sigma = 119$

Tabelle 28: Grad der Depression vor und nach einer Therapie.
(Quelle: Eigene Darstellung)

Zuerst werden die Hypothesen aufgestellt:

H0: $\Delta = 0$ Zwischen vor und nach der Therapie gibt es keinen Unterschied, keine Differenz (daher das Delta!).

H1: $\Delta \neq 0$ Zwischen vor und nach der Therapie gibt es einen Unterschied.

$$\bar{x}_d = \frac{\sum_{i=1}^{n} d_i}{n} = 2.5$$

$$\hat{\sigma}_d = \sqrt{\frac{\sum_{i=1}^{n} d_i^2 - \frac{\left(\sum_{i=1}^{n} d_i\right)^2}{n}}{n-1}} = \sqrt{\frac{119 - \frac{(25)^2}{10}}{10-1}} = \sqrt{\frac{119-62.5}{9}} = \sqrt{\frac{56.5}{9}} = \sqrt{6.2778} = 2.5056$$

$$\hat{\sigma}_{\bar{x}d} = \frac{\hat{\sigma}_d}{\sqrt{n}} = \frac{2.5056}{\sqrt{10}} = \frac{2.5056}{3.1623} = 0.7923$$

$$t_{emp} = \frac{\bar{x}_d}{\hat{\sigma}_{\bar{x}d}} = \frac{2.5}{0.7923} = 3.16$$

$$t_{krit(\alpha=0.05;df=9)} = 2.262$$

$t_{emp} > t_{krit}$ → signifikant → H1 → Mit einer Irrtumswahrscheinlichkeit von 5% kann behauptet werden, dass die Werte vor der Therapie sich von den Werten nach der Therapie unterscheiden! (Ob es den Patienten besser geht oder schlechter, muss man nun aus der Tabelle ablesen).

5.5.2 t-Test für unabhängige Stichproben

a) Für homogene Varianzen

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

$$\hat{\sigma}^2 = S^2 \cdot \frac{n}{n-1}; \quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}; \quad df = n_1 + n_2 - 2$$

Folgendes Problem:
Man arbeitet an einer Schule. An dieser Schule läuft ein Großversuch, bei dem geprüft werden soll, welche von zwei neuen Lehrmethoden (A, B) den Schülern mehr beibringt. Dazu wurde die Klasse 1 in der Lehrmethode A unterrichtet, Klasse 2 in der Lehrmethode B. Am Ende des Schuljahres bekamen beide Klassen denselben Test vorgelegt, um zu prüfen, wie gut der Schulstoff gelernt wurde. Hier sind die Ergebnisse:

Klasse 1:
Im Schnitt (d.h. als Mittelwert) erreichten die Schüler 15 Punkte in diesem Test,
bei einer Standardabweichung von 2 Punkten.
Klasse 2:
Im Schnitt erreichten die Schüler 17 Punkte in diesem Test, bei einer Standardabweichung von 1.5 Punkten.

In Klasse 1 befinden sich 20 Schüler; in Klasse 2 befinden sich 25 Schüler.

Die Frage ist nun, ob der hier sichtbare Unterschied von zwei Punkten auch noch erhalten bleibt, wenn auf Populationsebene hochgerechnet wird.

Vor einem t-Test muss immer erst ein F-Test gerechnet werden!
(Ausnahme: Wenn Aussagen über die Varianzhomogenität bzw. -heterogenität vorliegen.)

1. Überprüfung auf Varianzhomogenität – F-Test

H0: $\sigma_1^2 = \sigma_2^2$ Die Varianzen sind gleich (homogen).
H1: $\sigma_1^2 \neq \sigma_2^2$ Die Varianzen sind unterschiedlich (heterogen).

$$S_1^2 = 2 \cdot 2 = 4 \rightarrow \hat{\sigma}_1^2 = 4 \cdot \frac{20}{19} = 4.21$$

$$S_2^2 = 1.5 \cdot 1.5 = 2.25 \rightarrow \hat{\sigma}_2^2 = 2.25 \cdot \frac{25}{24} = 2.34$$

$$F_{emp} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{4.21}{2.34} = 1.80 \qquad F_{krit(\alpha=0.10;df_1=19;df_2=24)} = 2.04$$

$F_{emp} < F_{krit}$ → nicht signifikant → H0 → Es kann nicht behauptet werden, dass die Varianzen unterschiedlich sind → t-Test für homogene Varianzen!

2. t-Test für unabhängige Stichproben bei homogenen Varianzen

H0: $\mu_1 = \mu_2$ Die Mittelwerte sind gleich.
H1: $\mu_1 \neq \mu_2$ Die Mittelwerte unterscheiden sich.

$$\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{(n_1-1)\cdot\hat{\sigma}_1^2+(n_2-1)\cdot\hat{\sigma}_2^2}{n_1+n_2-2}\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{\frac{19\cdot 4.21+24\cdot 2.34}{20+25-2}\cdot\left(\frac{1}{20}+\frac{1}{25}\right)} = 0.534$$

$$t_{emp} = \frac{\bar{x}_1-\bar{x}_2}{\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)}} = \frac{17-15}{0.534} = 3.745$$

$$df = n_1+n_2-2 = 20+25-2 = 43$$

$$t_{krit(\alpha=0.05;df=43)} = 2.021$$

$t_{emp} > t_{krit}$ → signifikant → H1 → Die Mittelwerte sind – bezogen auf die Population – mit einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$ unterschiedlich! D.h. es besteht ein Unterschied zwischen den Lehrmethoden; folglich bringt Lehrmethode B den Schülern mehr bei.

b) Für heterogene Varianzen (einfache Variante)

$$t_{emp} = \frac{\bar{x}_1-\bar{x}_2}{\hat{\sigma}_d};\quad \hat{\sigma}_d = \sqrt{\frac{\hat{\sigma}_1^2}{n_1}+\frac{\hat{\sigma}_2^2}{n_2}};\quad \hat{\sigma}^2 = S^2\cdot\frac{n}{n-1};\quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n};\quad df = \frac{n_1+n_2}{2}$$

Gesetzt den Fall, beim obigen Beispiel wäre der F-Test signifikant ausgefallen, die Varianzen wären also heterogen gewesen. In diesem Fall hätte man den t-Test nicht mit den beim obigen Beispiel angegebenen Formeln rechnen dürfen, sondern man müsste entweder den U-Test durchführen oder die hier angegebenen Formeln verwenden. Die Hypothesen zum t-Test wären natürlich gleich geblieben:

H0: $\mu_1 = \mu_2$ Die Mittelwerte sind gleich.
H1: $\mu_1 \neq \mu_2$ Die Mittelwerte unterscheiden sich.

Die Rechnung hätte also folgendermaßen ausgesehen:

$$\hat{\sigma}_d = \sqrt{\frac{4.21}{20}+\frac{2.34}{25}} = \sqrt{0.2105+0.0936} = \sqrt{0.3041} = 0.551$$

$$t_{emp} = \frac{17-15}{0.551} = 3.630$$

$$df = \frac{20+25}{2} = 22.5 \approx 23$$

Als kritischen t-Wert hätte man für $\alpha = 0.05$ und df = 23 abgelesen: $t_{krit(\alpha=0{,}05;df=23)} = 2.069$

Da der empirische t-Wert größer als der kritische t-Wert ist, handelt es sich um ein signifikantes Ergebnis, man entscheidet sich folglich für die Alternativhypothese H1.

Die Mittelwerte sind – bezogen auf die Population – mit einer Irrtumswahrscheinlichkeit von 5% unterschiedlich; Lehrmethode B bringt den Schülern mehr bei.

Außer dieser einfachen Variante gibt es noch eine etwas exaktere, aber ungleich rechenintensivere Variante, den sogenannten Welsh-Test, wie er von Statistikprogrammen verwendet wird.

5.5.3 t-Test für Korrelationen

Um eine Produkt-Moment-Korrelation oder eine Spearman-Rang-Korrelation daraufhin zu überprüfen, ob der Wert dieser Korrelation (noch) im Zufallsbereich liegt, gibt es den sogenannten t-Test für Korrelationen.

Die Formel dazu lautet:

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

Um den empirischen t-Wert berechnen zu können, benötigt man also eine Angabe über die Höhe des Zusammenhangs, z.B. r = 0.98, sowie die Anzahl der Versuchspersonen, die diesem r zu Grunde liegen, z. B. n = 10.

Bevor gerechnet wird, sollte man immer die Fragestellung und damit einhergehend die Hypothesen formulieren.

Z.B.: Besteht ein Zusammenhang zwischen Sympathie und Redefluss?

Hieraus ergeben sich die Hypothesen:

H0: Es besteht kein Zusammenhang zwischen Sympathie und Redefluss (bzw. der Zusammenhang ist Null).
H1: Es besteht ein Zusammenhang zwischen Sympathie und Redefluss (bzw. der Zusammenhang ist nicht Null)

oder mathematisch:
$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

(Wie üblich bei Hypothesen wird auch hier ein griechischer Buchstabe verwendet, in diesem Fall der griechische Buchstabe für r: ρ, gesprochen: rho)

Berechnen wir nun den empirischen t-Wert:

$$t_{emp} = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}98 \times \sqrt{10-2}}{\sqrt{1-0.96}} = \frac{0.98 \times \sqrt{8}}{\sqrt{0.04}} = \frac{0.98 \times 2.83}{0.2} = 13.87; \quad df = 10-2 = 8$$

Als kritischen t-Wert für diese zweiseitige Hypothese bei df = 8 Freiheitsgraden finden wir in Tabelle C: $t_{krit(alpha=5\%,\ zweiseitig,\ df=8)} = 2.306$.

Der empirische t-Wert ist größer als der kritische t-Wert, also signifikant, also H1.
Mit einer Irrtumswahrscheinlichkeit von 5% besteht zwischen Sympathie und Redefluss ein Zusammenhang.

Noch ein Beispiel: Besteht zwischen Sympathie und Redefluss ein positiver Zusammenhang?

Ausgehend von dieser Fragestellung müssen nun gerichtete Hypothesen aufgestellt werden:

H0: Es besteht kein positiver Zusammenhang zwischen Sympathie und Redefluss
H1: Es besteht ein positiver Zusammenhang zwischen Sympathie und Redefluss

oder mathematisch: $H_0 : \rho \leq 0$
$H_1 : \rho > 0$

Die Berechnung des empirischen t-Werts ist die gleiche wie im vorherigen Beispiel.
Was sich ändert, ist nun der kritische t-Wert: $t_{krit(alpha=5\%,\ einseitig,\ df=8)} = 1{,}860$.

Der empirische t-Wert ist größer als der kritische t-Wert, also signifikant, also wird H1 angenommen.
Mit einer Irrtumswahrscheinlichkeit von 5% besteht zwischen Sympathie und Redefluss folglich ein positiver Zusammenhang.

Übungsaufgaben zu Kapitel 5

022 Fähigkeiten von Mitarbeitern

Sie erheben bei acht Mitarbeitern einer Abteilung deren Fähigkeiten im Umgang mit dem PC auf einer Skala von 0 (keine Fähigkeit vorhanden) bis 10 (exzellente Fähigkeiten vorhanden), das Geschlecht sowie die Anzahl bisher besuchter Fortbildungen im PC-Bereich. Die erhobenen Daten lauten:

Geschlecht	**w**	**w**	**m**	**m**	**w**	**m**	**m**	**w**
Fähigkeit	4	7	9	4	5	2	3	1
Anzahl Fortbildungen	0	4	6	2	3	1	1	2

Gehen Sie davon aus, dass die Fähigkeitswerte Intervallskalenniveau besitzen.

a) Berechnen Sie für Frauen und Männer getrennt jeweils Mittelwert, Standardabweichung und Varianz der Fähigkeitswerte.
b) Unterscheiden sich die Varianzen der Fähigkeitswerte von Frauen und Männern?
c) Unterscheiden sich Frauen und Männer hinsichtlich ihrer durchschnittlichen Fähigkeiten im Umgang mit dem PC?
d) Berechnen Sie bitte, ob es einen Zusammenhang zwischen der Fähigkeit im Umgang mit dem PC (x) und der Anzahl bisher besuchter Fortbildungen (y) gibt.
   Prüfen Sie dann, ob der Zusammenhang statistisch signifikant ist.

023 Was meint der Begriff „Inferenzstatistik"?

024 Pädagogischer Ansatz des Kindergartens und Schuleignung

Der große Dienstleistungskonzern, in dem Sie arbeiten, möchte eine/n betriebseigene Kindertagesstätte/Kindergarten für die Mitarbeiter aufbauen. Als Pädagogische Konzepte werden der Montessori-Ansatz und der Situations-Ansatz diskutiert.
Sie sollen nun eine kleine Studie durchführen, um herauszufinden, ob sich diese beiden Ansätze bezüglich der künftigen Schuleignung der Kinder unterscheiden. Dazu führen Sie in einem Kindergarten mit Montessori-Ansatz bei n = 7 Fünfjährigen und in einem Kindergarten mit Situationsansatz bei n = 5 Fünfjährigen einen Schuleignungstest durch. Hier sind die Daten:

Punktzahl Schuleignungstest	**Pädagogischer Ansatz Kindergarten**
17	M(ontessori)
13	M
14	M
9	M
6	M
5	M
12	M
16	S(ituations-Ansatz)
15	S
11	S
8	S
10	S

Für den Schuleignungstest gilt: Je mehr Punkte, desto eher (schon) schulgeeignet.
Nach Angaben des Testmanuals für den Schuleignungstest kann nicht von einer Normalverteilung der Werte ausgegangen werden.
Unterscheiden sich die beiden pädagogischen Konzepte hinsichtlich der Schuleignung statistisch bedeutsam voneinander?

6 Einfachregression und Konfidenzintervall

Lernziele

Nach der Lektüre dieses Kapitels sollten Sie wissen,

- was eine Einfachregression leistet,
- was unter einem b-Gewicht zu verstehen ist,
- in welchem Verhältnis Korrelation und Genauigkeit einer Schätzung stehen,
- warum und wie ein Konfidenzintervall gebildet wird.

6.1 Einfachregression

Regression meint Schätzung!

Wenn man weiß, dass zwei Variablen (z.B. Depression und Angst) mehr oder minder zusammenhängen, könnte man ja versuchen, von der einen auf die andere Variable zu schätzen. Dieses Schätzen kann man nach der Methode „Pi mal Daumen" oder eben auf mathematischem Wege machen. Und wie dieser mathematische Weg aussieht, ist Thema dieses Kapitels.

Eine Vorbemerkung noch: Eine Schätzung ist natürlich umso genauer, je stärker die beiden Variablen zusammenhängen; d.h. der Fehler beim Schätzen ist umso geringer, je höher die Korrelation zwischen den Variablen ist, bzw. die Schätzung wird umso ungenauer (und damit der Fehler umso größer), je geringer die Korrelation ist.

Also: **Korrelation und Regression gehören zusammen, bzw. bauen aufeinander auf**.

Die folgende Abbildung zeigt eine Verteilung der Werte von zwei hoch miteinander korrelierenden Variablen:

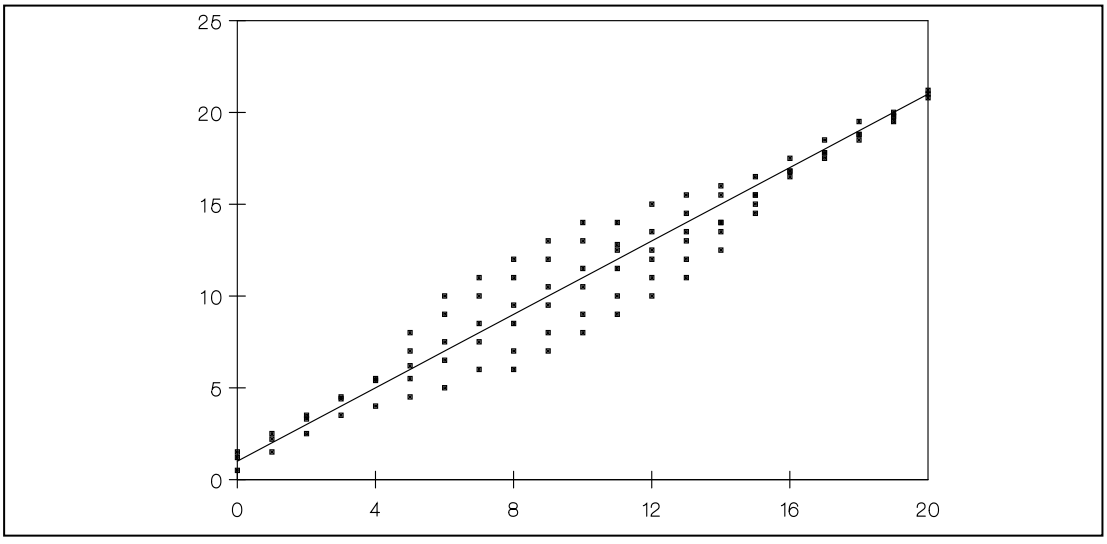

Abbildung 13: Datenwolke von zwei hoch miteinander korrelierenden Variablen.
(Quelle: Eigene Darstellung)

Bei dem Beispiel mit den Schulnoten war es ja noch einfach: Wenn die Noten hoch miteinander korrelieren, heißt dies: Jemand mit einer Eins in Mathematik wird wohl auch eine Eins in Physik haben, bzw. umgekehrt. Wenn die beiden Variablen die gleichen Einheiten haben (Schulnoten zwischen Eins und Sechs) und auch gleich verteilt sind, ist das Schätzen recht einfach. Was aber, wenn die beiden Variablen nicht die gleichen Einheiten haben oder vielleicht anders verteilt sind?

Bei den Abbildungen in Kapitel 4 war das Schätzen immer sehr einfach. Um jetzt hier schätzen zu können, benötigt man eine sogenannte **Schätzgleichung**. Ein anderer Name für Schätzgleichung ist **Regressionsgleichung**. Wie bestimmt man nun diese Regressionsgleichung?

Voraussetzung ist, dass es sich um intervallskalierte Variablen handelt!
Wie üblich, ein paar Formeln:

Fall 1: Schätzen von der Variablen X auf die Variable Y

Die allgemeine Gleichung lautet: $\hat{y} = b \cdot x + a$
(Das Dach auf dem y bedeutet, dass dieser Wert geschätzt wird.)

$$b = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{COV_{xy}}{S_x^2} = \frac{r_{x,y} \cdot S_y}{S_x} \qquad a = \bar{y} - b \cdot \bar{x}$$

Fall 2: Schätzen von der Variablen Y auf die Variable X

Die allgemeine Gleichung lautet: $\hat{x} = b \cdot y + a$

$$b = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n \cdot \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2} = \frac{COV_{xy}}{S_y^2} = \frac{r_{x,y} \cdot S_x}{S_y} \qquad a = \bar{x} - b \cdot \bar{y}$$

Nehmen wir nun einmal die Werte aus Kapitel 4 und schätzen. Zuerst von den Sympathiewerten (x) auf die Redeflusswerte (y):

$\bar{x} = 10.1 \qquad S_x^2 = 2.89 \qquad \bar{y} = 20.2 \qquad S_y^2 = 14.16 \qquad COV_{xy} = 6.28$

Fall 1: Schätzen von X auf Y

$$b = \frac{COV_{xy}}{S_x^2} = \frac{6.28}{2.89} = 2.17 \qquad a = \bar{y} - b \cdot \bar{x} = 20.2 - 2.17 \cdot 10.1 = -1.717$$

Hieraus folgt als erste Schätzgleichung: $\hat{y} = b \cdot x + a = 2.17 \cdot x - 1.717$

Jemand mit einem Sympathiewert von 12 hätte also als geschätzten Redeflusswert von:

$$\hat{y} = 2.17 \times 12 - 1.717 = 24.323$$

Fall 2: Schätzen von Y auf X

$$b = \frac{COV_{xy}}{S_y^2} = \frac{6.28}{14.16} = 0.44 \qquad a = 10.1 - 0.44 \cdot 20.2 = 1.212$$

Als zweite Schätzgleichung resultiert also: $\hat{x} = 0.44 \cdot y + 1.212$

Jemand mit einem Redeflusswert von 23 hätte also als geschätzten Sympathiewert:

$$\hat{x} = 0.44 \cdot 23 - 1.212 = 11.33$$

Die nachstehende Tabelle listet die richtigen und die (je nach Gleichung) geschätzten Werte für Sympathie und Redefluss auf:

Pb-Nr.	x	x	y	y
1	10	10.452	21	19.983
2	8	8.252	16	15.643
3	11	10.892	22	22.153
4	7	6.932	13	13.473
5	13	13.092	27	26.493
6	12	11.332	23	24.323
7	10	9.572	19	19.983
8	9	9.132	18	17.813
9	11	11.332	23	22.153
10	10	10.012	20	19.983

Tabelle 29: Sympathie- und Redeflusswerte, beobachtet und geschätzt.
(Quelle: Eigene Darstellung)

Die Schätzwerte liegen gar nicht so weit neben den beobachteten. In welchem Fall würden die geschätzten Werte denn genau den beobachteten Werten entsprechen? Natürlich bei einer Korrelation von 1 (oder –1). Die Differenz zwischen den geschätzten und den beobachteten Werten ist natürlich umso größer, je kleiner der Zusammenhang ist. Diese Differenz ist der Fehler beim Schätzen, der Schätzfehler.

Diesen Schätzfehler versucht man normalerweise auszugleichen. Wie dies gemacht wird? Indem man nicht einen bestimmten Wert (z.B. 10.012) angibt, sondern einen Bereich.

Folgende Überlegung:
Wenn man sagt: „Mit dem Auto brauche ich von Frankfurt/M. nach München genau 4 Stunden und 25 Minuten" oder: „mit dem Auto brauche ich von Frankfurt/M. nach München 4 bis 5 Stunden". Mit welcher Aussage liegt man wahrscheinlich (schon wieder dieses Wort...) eher richtig?

Mit der zweiten Aussage, dem Bereich. In der Psychologie nennt man so einen Bereich Vertrauensbereich, Vertrauensintervall oder Konfidenzintervall. Und genau darum geht es im nächsten Kapitel.

6.2 Konfidenzintervall

Ein Konfidenzintervall ist ein Bereich, in dem ein Wert mit einer bestimmten Wahrscheinlichkeit liegt!

Angenommen, ein Proband führt einen Test zur Konzentrationsfähigkeit durch und erreicht einen Prozentrang von PR = 60. Nur ist er heute leider etwas unausgeschlafen. Einige Zeit später führt der Proband denselben Test noch einmal durch, diesmal ist er total aufgedreht und erreicht den Prozentrang PR = 80. Aufgrund der unterschiedlichen Bedingungen erreicht dieser Proband zwei unterschiedliche Werte. Welcher von diesen Werten ist jetzt der richtige, der wahre Wert des Probanden? Vielleicht der in der Mitte, also PR = 70? Man weiß es nicht! Aber, man kann eine Schätzung vornehmen, in welchem Bereich der wahre Wert des Probanden mit einer bestimmten Wahrscheinlichkeit liegt. Diesen Bereich nennt man Vertrauensintervall oder Konfidenzintervall. Es gibt drei große Anwendungsbereiche für Konfidenzintervalle.

a) Konfidenzintervall für den wahren Wert eines Probanden, wenn dieser Proband mehrmals ein und denselben Test durchgeführt hat oder wenn bekannt ist, wie zuverlässig (reliabel) der Test an sich ist. Dieser Aspekt wird im vorliegenden Studienbrief nicht behandelt.
b) Konfidenzintervall für den wahren Wert eines Probanden in Test A, indem über den Wert des Probanden in einem anderen Test (Test B) geschätzt wird, wenn bekannt ist, wie beide Tests (A+B) zusammenhängen (korrelieren). Gemeint sind hier die Einfachregression und die multiple Regression.
c) Konfidenzintervall für den Populationsmittelwert μ, wenn ein Stichprobenmittelwert bekannt ist. Auch dieser Aspekt wird im vorliegenden Studienbrief nicht behandelt.

zu b)
Der Einfachheit halber wird hier noch einmal das Beispiel aus Kapitel 4 aufgegriffen.

Proband-Nummer:	Sympathie-Wert (x)	Redefluß-Wert (y)
1	10	21
2	8	16
3	11	22
4	7	13
5	13	27
6	12	23
7	10	19
8	9	18
9	11	23
10	10	20

Tabelle 30: Sympathie- und Redefluss-Werte bei zehn Probanden.
(Quelle: Eigene Darstellung)

Man hatte folgende statistische Kennwerte berechnet:

$\bar{x} = 10{,}1 \qquad S_x^2 = 2{,}89 \qquad S_x = 1{,}7$

$\bar{y} = 20{,}2 \qquad S_y^2 = 14{,}16 \qquad S_y = 3{,}763$

$COV_{xy} = 6{,}28 \qquad r_{xy} = 0{,}9817$

Als Regressionsgleichung für das Schätzen von X auf Y wurde errechnet: $\hat{y} = 2{,}17 \cdot x - 1{,}717$

Und nun folgen die Formeln für das eigentliche Konfidenzintervall, und zwar für den Fall, dass vom Wert in einer Variablen (hier: Sympathie) auf den wahren Wert in einer zweiten Variablen (hier: Redefluss) geschätzt wurde:

Obergrenze: $\hat{y}_i + t_{(\alpha/2)} \cdot \hat{\sigma}_{yx} \cdot \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{n \cdot S_x^2}}$

Untergrenze: $\hat{y}_i - t_{(\alpha/2)} \cdot \hat{\sigma}_{yx} \cdot \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{n \cdot S_x^2}}$

Standardschätzfehler $\hat{\sigma}_{yx} = \sqrt{S_y^2 \cdot \left(1 - r_{xy}^2\right) \cdot \frac{n}{n-2}} = \sqrt{\frac{n \cdot S_y^2 - n \cdot b_{yx}^2 \cdot S_x^2}{n-2}}$

t(α/2) wird verwendet, um die Wahrscheinlichkeit zu bestimmen, dass der wahre Wert innerhalb dieses Bereichs liegt. Die Freiheitsgrade berechnen sich hier als df = n-2.

Angenommen, man möchte den Bereich haben, innerhalb dessen der wahre Wert mit einer Wahrscheinlichkeit von 95% liegt. 95% bedeuten, dass rechts wie links jeweils 2.5% abgeschnitten werden. Welcher t-Wert schneidet nun auf der rechten Seite 2.5% der Fläche bei df = n-2 = 8 ab?

In Tabelle C findet sich als t-Wert: $t_{(\alpha=0.05;zweiseitig,\alpha=0.025;einseitig;df=8)} = 2.306$

So, jetzt wird der Standardschätzfehler berechnet:

$$\hat{\sigma}_{yx} = \sqrt{14.16 \cdot (1 - 0.9817) \cdot \frac{10}{10-2}} = \sqrt{14.16 \cdot 0.0183 \cdot 1.25} = \sqrt{0.324} = 0.569$$

Für welchen Wert soll denn das Konfidenzintervall gebildet werden? Im vorigen Kapitel hatten wir für den x-Wert = 12 den geschätzten y-Wert = 24.323 errechnet.
Und um dieses $\hat{y} = 24.323$ soll nun das Konfidenzintervall gelegt werden.

Als Obergrenze errechnet man:

Obergrenze: $\hat{y}_i + t_{(\alpha/2)} \cdot \hat{\sigma}_{yx} \cdot \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{n \cdot S_x^2}}$

Obergrenze: $24.323 + 2.306 \cdot 0.569 \cdot \sqrt{\frac{1}{10} + \frac{(12 - 10{,}1)^2}{10 \cdot 2.89}} = 24.323 + 1.312 \cdot \sqrt{\frac{1}{10} + \frac{(1.9)^2}{28.9}}$

Obergrenze: $24.323 + 1.312 \cdot \sqrt{\frac{1}{10} + \frac{3.61}{28.9}} = 24.323 + 1.312 \cdot \sqrt{0.1 + 0.125}$

Obergrenze: $24.323 + 1.312 \cdot \sqrt{0.225} = 24.323 + 1.312 \cdot 0.474 = 24.323 + 0.622$

Obergrenze: 24.945

Als Untergrenze errechnet man:
Untergrenze: 24.323 - 0.622 = 23.701

Als Konfidenzintervall hat man also den Bereich von 23.701 – 24.945 errechnet. Innerhalb dieses Bereichs liegt mit einer Wahrscheinlichkeit von 95% der wahre Redeflusswert eines Probanden, der als Sympathiewert x = 12 hat.

Übungsaufgabe zu Kapitel 6

025 Bierausschank

Alex B, der sein Psychologiestudium abgebrochen hat und jetzt Wirt der Bahnhofskneipe „Zur Pfütze" ist, betrachtet seine Bierverkäufe der letzten paar Monate.
Dazu hat er pro Woche die verkauften Liter (in Hektolitern; 1 hl = 100 Liter) in einer Tabelle aufgelistet.

Woche	1	2	3	4	5	6	7	8	9	10
Verkauftes Bier (hl)	3	4	4	5	4	6	5	7	6	8

a) Berechnen Sie eine Regressionsgerade von „Woche" (x) auf „Verkaufte Biere" (y)!
b) Mit wie viel verkauften Bieren sollte er für die 20. Woche rechnen?
Berechnen Sie auch ein 95%-Konfidenzintervall!
c) Berechnen Sie eine Regressionsgerade von „Verkaufte Biere" (x) auf „Woche" (y)!
d) Ab 40 Hektolitern bekommt Alex B von seinem Großhändler günstige Einkaufsoptionen. In welcher Woche kann er damit rechnen?

7 Faktorenanalyse

Lernziele

Wenn Sie dieses Kapitel bearbeitet haben, dann sollten Sie wissen,

- wie das Fundamentaltheorem der Faktorenanalyse lautet und was es bedeutet,
- was unter der Kommunalität zu verstehen ist,
- was unter dem Eigenwert eines Faktors zu verstehen ist,
- wozu eine Faktorenanalyse dient,
- was das Kaiserkriterium aussagt,
- was das Scree-Kriterium aussagt.

Zuerst einmal: Eine Faktorenanalyse ist nichts Besonderes. Sie ist jedoch sehr aufwändig zu berechnen. Als es noch keine Computer gab, haben einige Personen allein mit dem Durchrechnen einer Faktorenanalyse ihren Ruf erlangt. Nun zum Thema: Man stelle sich vor, zwei Variablen würden ziemlich hoch miteinander korrelieren. Man stellt sich nun die Frage, warum man zwei Variablen erhebt bzw. betrachtet und nicht einfach eine neue Variable erstellt, die die beiden anderen Variablen ersetzt.

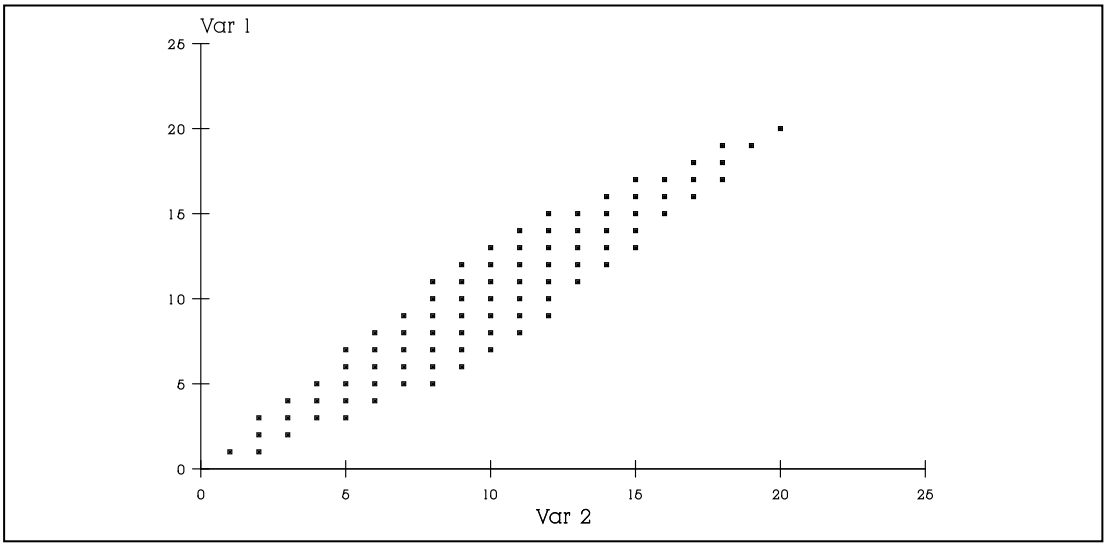

Abbildung 14: Datenwolke bei zwei hoch miteinander korrelierenden Variablen.
(Quelle: Eigene Darstellung)

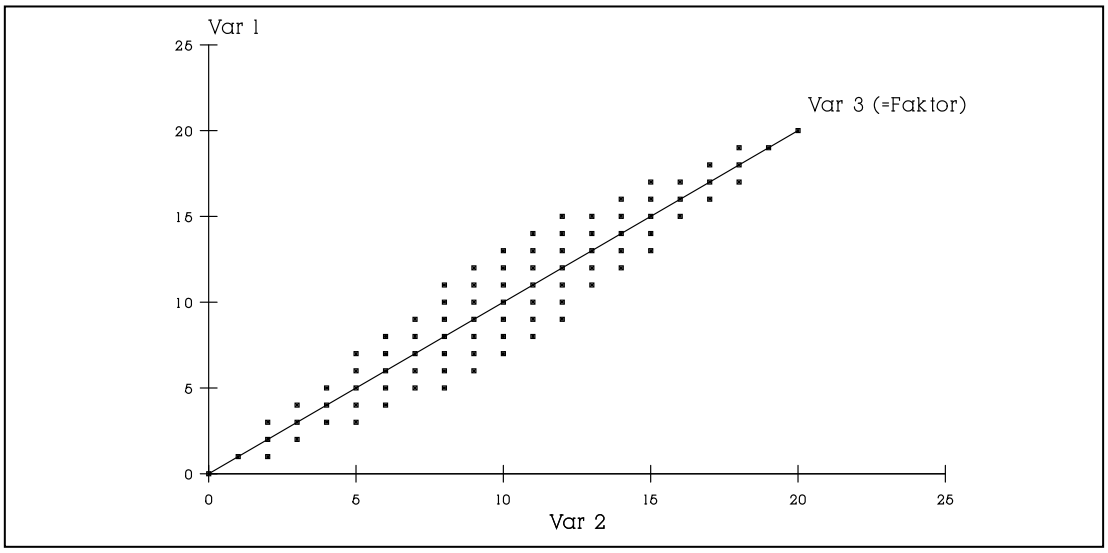

Abbildung 15: Ersetzen zweier hoch korrelierender Variablen durch eine dritte.
(Quelle: Eigene Darstellung)

Genau dies macht eine Faktorenanalyse. Allerdings nicht nur bei zwei Variablen, sondern bei vielen. Anders ausgedrückt: Eine Vielzahl an Variablen wird durch einige wenige Faktoren ersetzt.

Man erhebt beispielsweise Daten für hundert Variablen. Will man nun sehen, wie jede dieser hundert Variablen mit jeder anderen dieser hundert Variablen korreliert, sind das eine ganze Menge Zahlen (um genau zu sein: 4.950). Stattdessen könnte man versuchen, mittels einer Faktorenanalyse den Datenbestand zu reduzieren, auf vielleicht zehn Faktoren.

Als Ergebnis einer Faktorenanalyse erhält man eine sogenannte Faktorladungsmatrix (A) für jede Variable. Diese Faktorladungen sind nichts anderes als die Korrelationen der Variablen mit dem Faktor. Je höher diese Faktorladung, desto besser wird die Variable vom Faktor repräsentiert.

Beispiel:
Man hat fünf Variablen erhoben und die Korrelationen der Variablen errechnet:

	Var1	Var2	Var3	Var4	Var5
Var1	1	0.94	0.91	0.10	0.07
Var2	0.94	1	0.88	0.21	0.14
Var3	0.91	0.88	1	0.09	0.12
Var4	0.10	0.21	0.09	1	0.93
Var5	0.07	0.14	0.12	0.93	1

Tabelle 31: Korrelationsmatrix (R) von fünf Variablen.
(Quelle: Eigene Darstellung)

Wenn man sich diese Korrelationen ansieht, fällt folgendes auf: Variable 1, Variable 2 und Variable 3 korrelieren hoch miteinander, aber nur unbedeutend mit den Variablen 4 und 5. Variable 4 und Variable 5 korrelieren hoch miteinander, aber nur unbedeutend mit den Variablen 1, 2 und 3. Eine Faktorenanalyse könnte (und sollte) jetzt als Ergebnis zwei Faktoren berechnen und eine Faktorladungsmatrix ähnlich der untenstehenden bilden.

	Faktor I	Faktor II
Var1	0.808	0.541
Var2	0.849	0.455
Var3	0.794	0.504
Var4	0.617	0.742
Var5	0.598	0.758

Tabelle 32: Faktorladungsmatrix (Beispiel) für zwei Faktoren und vier Variablen.
(Quelle: Eigene Darstellung)

In dieser Faktorladungsmatrix stehen – wie erwähnt – die Korrelationen der einzelnen Variablen mit den beiden Faktoren. Die Variablen 1, 2 und 3 haben hohe (bzw. die höchsten) Korrelationen mit (oder: laden hoch auf) Faktor I, die Variablen 4 und 5 haben hohe (bzw. die höchsten) Korrelationen mit (laden hoch auf) Faktor II. Anders ausgedrückt: Die Variablen 1, 2 und 3 werden sehr gut von Faktor I repräsentiert, die Variablen 4 und 5 sehr gut von Faktor II.

Erinnern wir uns:
Das Quadrat eines Produkt-Moment-Korrelationskoeffizienten nennt man Determinationskoeffizient. Determinationskoeffizient mit 100 multipliziert ergibt die Prozente an erklärter Varianz. Man könnte sich nun fragen, wie gut, d.h. zu wie viel Prozent erklärter Varianz sich die einzelnen Variablen durch die Faktoren erklären lassen. Dazu bildet man jeweils die Quadrate der Faktorladungen und addiert sie zeilenweise. Die Summe der quadrierten Ladungen nennt man Kommunalität (abgekürzt als h^2). Kommunalität multipliziert mit 100 ergibt die Prozente an erklärter Varianz, gibt also in Prozent an, wie gut die jeweilige Variable durch das Zusammenspiel der Faktoren erklärt wird. Die nachstehende Tabelle gibt ein Beispiel.

Variable	**Faktor I**		**Faktor II**		**Kommunalität** (h^2)
	Faktorladung	**(Faktorladung)^2**	**Faktorladung**	**(Faktorladung)^2**	
Var1	0.808	0.6529	0.541	0.2927	0.9456
Var2	0.849	0.7208	0.455	0.2070	0.9278
Var3	0.794	0.6304	0.504	0.2540	0.8844
Var4	0.617	0.3807	-0.742	0.5506	0.9313
Var5	0.598	0.3576	-0.758	0.5746	0.9322

Tabelle 33: Faktorladungen und Kommunalitäten.
(Quelle: Eigene Darstellung)

Die Kommunalitäten würde man nun so interpretieren: Am besten wird die Variable Var1 durch die Faktoren (Faktorenstruktur) erklärt (repräsentiert), nämlich zu 94.56%. Die Variable Var3 wird am schlechtesten erklärt, nämlich nur zu 88.34%. Insgesamt werden jedoch alle Variablen recht gut erklärt. Als Faustregel kann man sich merken, dass 50% und mehr als gut gelten.

Übrigens: Prinzipiell kann man so viele Faktoren bilden, wie es Variablen gibt! Und wenn man das macht, also beispielsweise hier fünf Faktoren bildet, dann gilt folgendes zentrales Theorem der Faktorenanalyse (Fundamentaltheorem):

$R = A A'$
R = Korrelationsmatrix der Variablen
A = Faktorladungsmatrix

Dieses Theorem heißt nichts anderes, als dass man aus der Faktorladungsmatrix die ursprüngliche Korrelationsmatrix wiederherstellen kann, ohne dass Information verloren geht! Dieses Wiederherstellen ohne Informationsverlust funktioniert leider nur dann, wenn man eben so viele Faktoren wie Variablen bildet. Aber dann hätte man auch keine Faktorenanalyse machen müssen.

Hier taucht allerdings auch ein Problem auf: Wenn man aus der Faktorladungsmatrix die ursprüngliche Korrelationsmatrix wiederherstellen können soll, müssten doch die Variablen durch die Faktorenstruktur jeweils zu 100% erklärt werden. Ansonsten geht doch Information verloren. Dieses Problem nennt man das Kommunalitätenproblem der Faktorenanalyse. Woran dies liegt, folgt weiter unten. Prinzipiell lässt sich jedoch schon hier sagen, dass eine Faktorenstruktur immer etwas Information verliert.

Das Addieren der quadrierten Ladungen kann man natürlich nicht nur zeilenweise, sondern auch spaltenweise durchführen. Dann erhält man den sogenannten Eigenwert der Faktoren. Dieser Eigenwert würde angeben (mit 100 multipliziert), wie viel Prozent der Varianz aller Variablen durch den jeweiligen Faktor erklärt werden. Bei einer Variablen kann dieser Wert natürlich höchstens 100 betragen (mehr als 100 Prozent können für eine Variable nicht aufgeklärt werden), bei fünf Variablen gibt es natürlich dann entsprechend mehr, nämlich insgesamt sozusagen 500 Prozent an erklärter Varianz.

Statt den Eigenwert nun mit 100 zu multiplizieren, interpretiert man den Wert einfach als Anzahl der Variablen, die durch den jeweiligen Faktor repräsentiert/erklärt werden. Folgende Tabelle stellt die Eigenwerte für das obige Beispiel dar:

Variable	Faktor I		Faktor II	
	Faktorladung	(Faktorladung)2	Faktorladung	(Faktorladung)2
Var1	0.808	0.6529	0.541	0.2927
Var2	0.849	0.7208	0.455	0.2070
Var3	0.794	0.6304	0.504	0.2540
Var4	0.617	0.3807	0.742	0.5506
Var5	0.598	0.3576	0.758	0.5746
Eigenwert	2.7424		1.8789	

Tabelle 34: Faktorladungen und Eigenwerte.
(Quelle: Eigene Darstellung)

Man würde also jetzt sagen, der Faktor I erklärt rund 2.7 Variablen und der Faktor II rund 1.8 Variablen. Damit erfüllen sie ein wichtiges Kriterium: Eine Faktorenanalyse soll ja der Datenverminderung (Datenreduktion) dienen, eine große und damit unüberschaubare Anzahl an Variablen auf eine kleinere und damit auch interpretierbare Anzahl reduzieren. Statt der fünf Variablen hätte man hier nun nur zwei Faktoren, die man interpretieren müsste. Wenn ein Faktor nun nur eine einzige oder sogar noch weniger Variablen erklärt (Eigenwert ≤ 1), dient er nicht mehr der Datenreduktion, d.h. er vermindert die Datenmenge nicht mehr!

Dieses Kriterium, dass ein Eigenwert größer als 1 sein soll, damit man von einem Faktor sprechen kann, nennt man **Kaiserkriterium**!

Außer dem Kaiserkriterium gibt es noch ein Extraktionskriterium (d.h. ein Kriterium dafür, wie viele Faktoren gebildet werden sollen) namens **Scree-Test**. Für den Scree-Test werden die Eigenwerte aller Faktoren grafisch dargestellt. Die y-Achse stellt die Höhe des Eigenwerts dar, die x-Achse die einzelnen Faktoren. Die Anweisung für den Scree-Test lautet nun: Ab einer bestimmten Stelle geht die Kurve der Eigenwerte – mehr oder minder – in eine Gerade über. Zu extrahieren sind nur solche Faktoren, die vor dieser Stelle (also vor diesem Knick) liegen. Bezogen auf die nebenstehende Abbildung wären dies die Faktoren 1 bis 3. Die Eigenwerte der Faktoren 4 bis 15 bilden praktisch eine Gerade.

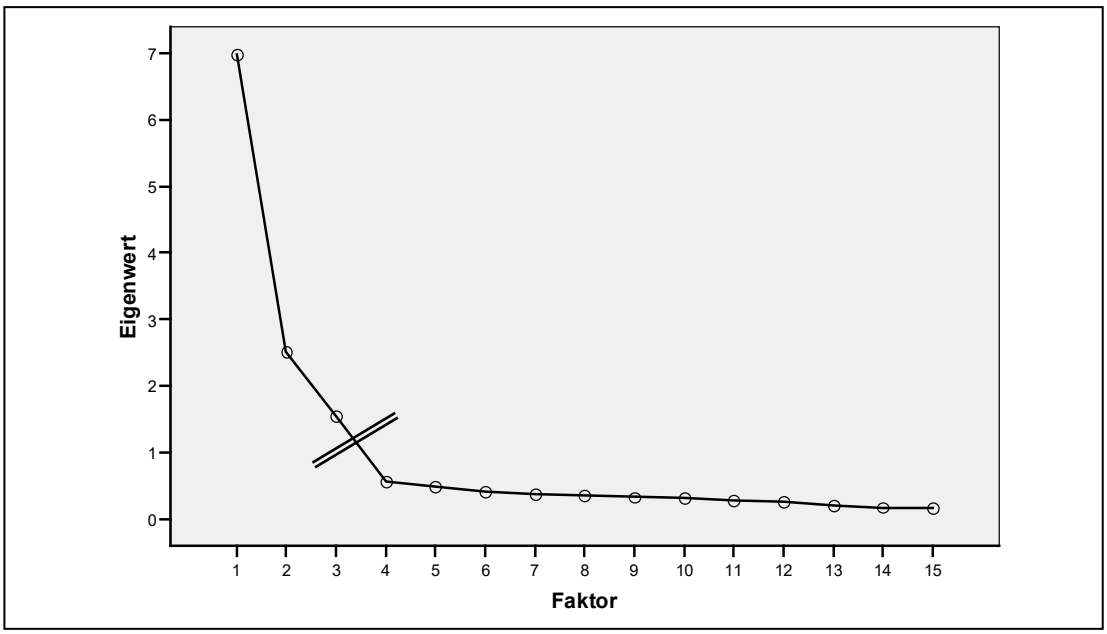

Abbildung 16: Scree-Plot für die Eigenwerte von 15 Faktoren.
(Quelle: Eigene Darstellung)

Bleibt noch ein drittes, inoffizielles Abbruchkriterium, welches meiner Meinung nach jedoch eigentlich das Wichtigste ist: Die Interpretierbarkeit der Faktoren in Verbindung mit der Gesamtgüte der Faktorenanalyse. Kurz zur Erinnerung: Um auszudrücken, wie gut ein Zusammenhang ist, wurde der Begriff des Determinationskoeffizienten r^2 eingeführt. r^2 wurde noch mit 100 multipliziert und dann interpretiert als „Prozent Zusammenhang" oder als „Prozent gemeinsamer Varianz". Für die Faktorenanalyse kann etwas Ähnliches gebildet werden, was dann als „Prozent (durch die Faktoren) erklärte Varianz" genannt wird. Allgemein sollte dieser Wert nicht unter 50% liegen. Gleichzeitig ist noch zu beachten, dass die extrahierten Faktoren gut zu interpretieren sind. Ob jetzt also beispielsweise eine Fünf-Faktorielle Lösung mit 50% erklärter Varianz oder eine Sechs-Faktorielle Lösung mit 58% erklärter Varianz oder eine Sieben-Faktorielle Lösung mit 63% erklärter Varianz genommen wird, entscheidet sich letztlich an der Interpretierbarkeit der gebildeten Faktoren. Hier sind immer Fingerspitzengefühl und Kreativität gefragt!

Vom Prinzip her kann man immer genau so viele Faktoren bilden (extrahieren), wie es Variablen gibt. Dann hätte man jedoch keinerlei Reduktion mehr. Also bricht man das Bilden irgendwann ab. Ein Abbruchkriterium ist beispielsweise das Kaiserkriterium. Dieses lautet: Der Eigenwert eines Faktors muss mindestens gleich 1 sein. Und weil man eben irgendwann abbricht, kann die Faktorenstruktur niemals alle Variablen zu jeweils 100% erklären! Wenn man also alle möglichen Faktoren bilden würde, würden die Variablen jeweils zu 100% erklärt werden, dann hätte man jedoch keine Datenreduktion mehr, sondern genauso viele Faktoren wie vorher Variablen und der Zweck der Faktorenanalyse wäre verfehlt!

Die Faktorenanalyse ist ein Verfahren, welches Variablen gemäß ihrer korrelativen Beziehungen in wenige, voneinander unabhängige Variablengruppen unterteilt. Diese Variablengruppen werden durch einen sogenannten Faktor repräsentiert. Der Faktor liefert sogenannte Faktorladungen für jede Variable. Diese Faktorladung ist nichts anderes als die Korrelation der Variablen mit dem Faktor!

Die Faktorenanalyse ermöglicht es, ohne entscheidenden Informationsverlust viele wechselseitig mehr oder weniger hoch korrelierende Variablen durch wenige, voneinander unabhängige Faktoren zu ersetzen. Die Faktorenanalyse dient hauptsächlich dazu, große Datenmengen zu reduzieren!

Wenn ein komplexes Merkmal, wie z.B. Leistungsmotivation oder Work-Life-Balance untersucht werden soll, denkt man sich viele Indikator-/Prädiktorvariablen aus. Über die Faktorenanalyse kann man berechnen, aus wie vielen Faktoren (Dimensionen) sich das komplexe Merkmal zusammensetzt. Die Faktorenanalyse ist somit ein Verfahren zur Überprüfung der Dimensionalität komplexer Merkmale. Das Wissen um die Dimensionalität eines Merkmales wiederum hat Auswirkungen auf die Testkonstruktion. Wichtig hierbei ist es natürlich, dass eine Faktorenanalyse auch nur solche Dimensionen liefern kann, welche über die Items / Variablen vorgegeben werden. Soll eine Faktorenanalyse der Dimensionierung eines komplexen Konstrukts dienen, ist es anzuraten, möglichst viele Facetten des Konstrukts durch möglichst viele Items abzudecken.

Anfang des 20ten Jahrhunderts entwickelte Spearman (1904) ein Generalfaktorenmodell der Intelligenz. In diesem Modell ging er davon aus, dass alle intellektuellen Leistungen von einem allgemeinen Intelligenzfaktor sowie von mehreren aufgabenspezifischen Intelligenzfaktoren abhängen. Zur Überprüfung dieses Modells entwickelte er die Tetradenmethode, die als Vorläufer der Faktorenanalyse gilt. Thurstone (1931, 1947) war nun maßgeblich an der methodischen Weiterentwicklung der Faktorenanalyse beteiligt und verhalf mit seinem Modell mehrerer gemeinsamer Faktoren der Entwicklung mehrdimensionaler Verhaltens- und Persönlichkeitsmodelle entscheidend zum Durchbruch.

Faktorenanalysen lassen sich auf verschiedene Weisen rechnen.
Die weitverbreitetste ist die Hauptkomponenten-Analyse (principal components analysis = pca) von Hotelling (1933). Weitere sind die Zentroid-Methode von Thurstone (1947) oder die Diagonalmethode von Dwyer (1944).

Übungsaufgaben zu Kapitel 7

026 Wie lautet das Fundamentaltheorem der Faktorenanalyse?

027 Was beschreibt der Eigenwert eines Faktors?

028 Was beschreibt die Kommunalität eines Items?

029 Beschreiben Sie, was unter dem Kaiser- und was unter dem Scree-Kriterium zu verstehen ist.

8 Multiple Regressionsanalyse

Lernziele

Wenn Sie dieses Kapitel bearbeitet haben, dann sollten Sie wissen,

- was eine multiple Regression leistet,
- was unter b-Gewichten zu verstehen ist,
- welche Fragestellungen mit dazugehörigen Hypothesen in einer multiplen Regressionsanalyse mit zwei Prädiktoren üblicherweise aufgestellt werden.

Eine Regressionsanalyse berechnet den Zusammenhang/die Korrelation zwischen mehreren Variablen. In Kapitel 4.4 wurden schon Zusammenhänge berechnet, allerdings immer nur zwischen zwei Variablen. Eine Regressionsanalyse macht nichts anderes, als zwei oder mehr Variablen mit einer anderen zu korrelieren.

Wofür könnte dies wichtig sein? Nun, gesetzt den Fall, man möchte einen aus mehreren Bewerber auswählen und man wüsste, dass die Ergebnisse in den Tests A und B hoch mit der Arbeitsleistung korrelieren, wäre es doch sinnvoll, den Bewerber nach den Ergebnissen in den Tests A und B auszuwählen. Oder anders gesagt, es besteht das Bedürfnis danach, gemachte Erfahrungen transparent zu machen, indem diese Erfahrungen mathematisch dargestellt werden. Durch die Regressionsanalyse wird z.B. eine implizit vorhandene Meinung über den Zusammenhang zwischen Bewerbungstests (oder auch Gesprächen) mit der zukünftigen Arbeitsleistung objektiviert. Kurz gesagt: Ob die Person, die über eine Einstellung mitentscheidet, am Morgen gut oder schlecht aufgestanden ist, verliert (zumindest etwas) an Bedeutung.

Schön wäre es hierbei, wenn die Tests A und B so gut wie gar nicht miteinander korrelierten. Denn dann brächte jedes Testergebnis einen Gewinn. Wenn die Tests A und B hoch miteinander korrelierten, bräuchte man natürlich nicht beide Tests darzubieten.

Wie erkennt man nun, ob zwei oder mehr Tests hoch mit der Arbeitsleistung (oder irgendetwas anderem) korrelieren?

Zuerst muss man mehrere Personen in allen drei Variablen (Test A, Test B und Arbeitsleistung) untersuchen. Aus den Werten der Personen in diesen drei Variablen errechnet man dann eine Regressionsgleichung. Mit Hilfe dieser Regressionsgleichung kann man nun bei zukünftigen Personen aufgrund der Ergebnisse in den Tests A und B errechnen, wie hoch deren Arbeitsleistung wäre.

Die Variablen, mit deren Hilfe eine dritte geschätzt wird, heißen **Prädiktoren** (praedicere = vorhersagen), die Variable, die geschätzt wird, heißt **Kriterium**.

Bei der Einfachregression hatte man lediglich einen Prädiktor (x) und ein Kriterium (y). Hier liegen nun mehrere Prädiktoren (x_1, x_2...) vor und ebenfalls ein Kriterium (y). Bei der Einfachregression erhielt man eine Regressionsgerade. Hier erhält man nun die Kombination zweier Geraden, nämlich eine Ebene (oder Fläche).

Das klingt alles ein wenig kompliziert, deswegen erst einmal ein Beispiel. Und weil es ganz nett ist, wenn Beispiele aufeinander aufbauen, wird hier wieder auf das Beispiel aus der Einfachregression zurückgegriffen. Außer Sympathie und Redeflusswerten werden hier allerdings zusätzlich noch die (intervallskalierten) Attraktivitätswerte dargestellt.

Die Frage soll sein, wie gut von Redefluss- und Attraktivitätswert auf den Sympathiewert geschätzt werden kann.

Pb-Nummer	Sympathie-Wert (y)	Redefluss-Wert (x1)	Attraktivitätswert (x2)
1	10	21	8
2	8	16	2
3	11	22	7
4	7	13	4
5	13	27	7
6	12	23	9
7	10	19	5
8	9	18	2
9	11	23	7
10	10	20	6

Tabelle 35: Sympathie- und Redefluss-Werte bei zehn Probanden.
(Quelle: Eigene Darstellung)

Berechnen wir doch zuerst einmal spaßeshalber die einzelnen Korrelationen.
Wenn man das macht, kommt man zu folgenden Ergebnissen:

	Sympathie	Redefluß	Attraktivität
Sympathie	1.000	0.982	0.755
Redefluß	0.982	1.000	0.729
Attraktivität	0.755	0.729	1.000

Tabelle 36: Korrelationen zwischen den Variablen Sympathie, Redefluß und Attraktivität.
(Quelle: Eigene Darstellung)

Aus diesen Korrelationen lässt sich folgendes ablesen: Redefluss korreliert mit Sympathie sehr hoch, Attraktivität mit Sympathie immer noch hoch. Leider korrelieren Redefluss und Attraktivität auch recht hoch miteinander. Viel besser wäre es, wenn Redefluss und Attraktivität nur schwach miteinander korrelieren würden, denn: Schon hier muss man sich fragen, ob die Variable Attraktivität überhaupt etwas bringt. Redefluss alleine ist schon sehr stark. Na ja, mal sehen.

Wie errechnet man nun solch eine multiple (d.h. mehr als einen Prädiktor enthaltende) Regressionsgleichung?

Zur Berechnung der multiplen Regressionsanalyse ist Matrixalgebra nötig.
Matrixalgebra beschreibt das Rechnen mit Matrizen (Tabellen mit Zahlen).

Unter dem Begriff „Allgemeines lineares Modell" ist nun ein ganzer Katalog an Rechenvorschriften in Matrixalgebra eingeführt worden. Mit Hilfe des Allgemeinen linearen Modells (ALM) lassen sich viele der komplizierteren statistischen Verfahren berechnen.

Es wird hier nur beschrieben, was eine Regressionsanalyse leisten kann und welche Fragestellungen und Hypothesen im Rahmen der multiplen Regressionsanalyse geklärt werden können.

Als multiple Regressionsgleichung für unser Beispiel ergibt sich: $\hat{y} = 1.343 + 0.416 \times x_1 + 0.062 \times x_2$

Was kann man nun mit dieser Formel anfangen?

Mit dieser Formel kann man, wenn man weiß, welchen Attraktivitäts- und welchen Redeflusswert eine Person hat, schätzen, welchen Sympathiewert eine Person haben wird. Nimmt man sich beispielsweise die Werte der Person Nr. 1 aus Tabelle 36 und setzt die Redefluss- und Attraktivitätswerte in die Gleichung ein, ergibt sich:

$$\begin{aligned} \textit{Sympathie} &= 1.343 + 0.416 \times 21 + 0.062 \times 8 \\ &= 1.343 + 8.736 + 0.496 \\ &= 10.575 \end{aligned}$$

Für die Person Nr. 1 würde man also – ausgehend vom Redefluss- und Attraktivitäts-Wert – einen Sympathiewert von 10.575 schätzen.

Ob diese Formel gut ist, kann man beurteilen, wenn man in Tabelle 36 eine weitere Spalte mit den geschätzten Sympathiewerten einfügt:

Pb-Nummer:	Sympathie-Wert (y)	Redefluss-Wert (x1)	Attraktivitätswert (x2)	geschätzter Sympathie-Wert ($\hat{y}$)
1	10	21	8	10.575
2	8	16	2	8.126
3	11	22	7	10.937
4	7	13	4	7.004
5	13	27	7	13.017
6	12	23	9	11.479
7	10	19	5	9.563
8	9	18	2	8.958
9	11	23	7	11.353
10	10	20	6	10.042

Tabelle 37: Sympathie-, Redefluss-, Attraktivitäts- und geschätzte Sympathie-Werte bei zehn Probanden. (Quelle: Eigene Darstellung)

Wie man sieht, lässt sich die Variable Sympathie gut schätzen, wenn man die Attraktivitäts- und Redeflusswerte einer Person kennt. Tabelle 64 listet noch einmal die beobachteten und geschätzten Sympathiewerte auf sowie die Differenz zwischen diesen beiden Werten.

Pb-Nummer:	Sympathie-Wert (y)	geschätzter Sympathie-Wert ($\hat{y}$)	Differenz zwischen y und ($\hat{y}$)
1	10	10.575	+0.575
2	8	8.126	+0.126
3	11	10.937	-0.063
4	7	7.004	+0.004
5	13	13.017	+0.017
6	12	11.479	-0.521
7	10	9.563	-0.437
8	9	8.958	-0.042
9	11	11.353	+0.353
10	10	10.042	+0.042

Tabelle 38: Sympathiewerte, beobachtet und geschätzt, sowie Schätzfehler. (Quelle: Eigene Darstellung)

Noch einmal: Welchen Sinn kann so etwas haben? Gesetzt den Fall, eine Person fragt um Rat, weil sie unter Einsamkeit leidet. Gesetzt den Fall, diese Person sei psychisch normal, könnte man

untersuchen, welchen Redefluss- und welchen Attraktivitätswert diese Person hat. Vielleicht ist diese Person zwar attraktiv, redet aber nur sehr wenig; dann könnte man ihr beibringen, mehr zu reden, bzw. die Ängste nehmen, mit fremden Personen zu reden. Denn wenn es wirklich einen Zusammenhang zwischen dem Redefluss, der Attraktivität und der Sympathie gibt, könnte ein Anheben des Redeflusses auch mit einer Anhebung der Sympathie einhergehen.

Ein anderes Beispiel wäre der Bereich Berufseignung. Gibt es Variablen (Prädiktoren), die eine Voraussage auf den beruflichen Erfolg erlauben? Man könnte sich vorstellen, dass z.B. Studenten bessere Gedächtnisleistungen haben müssen, um erfolgreich durch das Studium zu kommen. Anders ausgedrückt, vielleicht ließe sich von kurz- und langfristigen Gedächtnisleistungen auf den Studienabschluss schließen. Wenn dem so wäre, könnten schon in der Studienberatung die Gedächtnisleistungen eines Studierenden gemessen und dieser Person die Wahrscheinlichkeit für ein erfolgreiches Studium gegeben werden. Dieses Beispiel ist natürlich etwas schwach, da jeder Studiengang spezielle Anforderungen hat und eben nicht nur Gedächtnisleistungen. Vom Prinzip könnte aber auf diese Weise vielleicht vielen StudentInnen ein Studienabbruch erspart werden.

Wie funktioniert eine Regressionsanalyse nun? Zuallererst benötigt man natürlich Daten. Und zwar sowohl die Daten, mit denen geschätzt werden soll (Prädiktoren), als auch die Daten, worauf geschätzt werden soll (Kriterium).

Als nächstes soll nun überlegt werden, was eigentlich alles von Interesse ist.

Im Rahmen der Regressionsanalyse sind folgende Fragestellungen mittlerweile üblich:

a) Wie groß sind die Einflussgewichte der Prädiktorvariablen?
b) Wie groß ist die gemeinsame Korrelation bzw. die multiple Bestimmtheit R^2?
c) Wie gut kann ein Prädiktor alleine das Kriterium erklären?
d) Wie gut kann ein anderer Prädiktor alleine das Kriterium erklären?
e) Wie hoch ist der Zuwachs an Vorhersagegenauigkeit, wenn man zusätzlich zu einem Prädiktor noch einen zweiten Prädiktor in die Berechnung mit aufnimmt?
f) Welchen (Kriteriums-) Wert erzielt eine Person, die im Prädiktor Nr.1 den Wert XX und in Prädiktor Nr.2 den Wert YY hat?
   Wie groß ist das dazugehörige Konfidenz- (Vertrauens-) Intervall?

zu a)
Hier ergeben sich die oben dargestellten b-Gewichte: $b_0 = 1.343$; $b_1 = 0.416$; $b_2 = 0.062$

zu b)
Hier ergibt sich eine multiple Bestimmtheit von: $R^2 = 0.989$

Dieses R^2 muss daraufhin überprüft werden, ob es signifikant ist, d.h. ob das Vorhersage-Modell mit den beiden Prädiktorvariablen überhaupt in der Lage ist, eine geeignete Vorhersage zu treffen.

Die Hypothesen lauten:

$H_0 : b_1 = 0$ *und* $b_2 = 0$	Der Einfluss der beiden Prädiktoren Redefluss und Attraktivität bezüglich der Sympathie ist gleich Null.
$H_1 : b_1 \neq 0$ *und / oder* $b_2 \neq 0$	Der Einfluss mindestens eines Prädiktors bezüglich der Sympathie ist ungleich Null.

Hier ergibt sich, dass mindestens einer der Prädiktoren eine signifikante Erklärung des Kriteriums bietet.

zu c)
Die Hypothesen lauten:

$H_0 : b_1 = 0$ Die Variable Redefluss leistet keinen Beitrag zur Vorhersage der Sympathie

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

Dies entspricht der einfachen Korrelation zwischen der Variablen Redefluss und der Variablen Sympathie bzw. dann den Werten einer Einfachregression (vgl. Kapitel 6).

zu d) siehe c)

zu e)
Wie sehen denn die Hypothesen zu dieser Frage aus?

$H_0 : b_1 = 0; b_2 = \text{beliebig}$ Die Variable Redefluss leistet keinen zusätzlichen Beitrag zur Vorhersage der Sympathie.

$H_1 : b_1 \neq 0; b_2 = \text{beliebig}$ Die Variable Redefluss leistet einen zusätzlichen Beitrag zur Vorhersage der Sympathie.

Hier kann errechnet werden, dass die Hinzunahme des Prädiktors Redefluss eine signifikante Verbesserung des Vorhersagemodells erbringt.

zu f)
Welchen (Kriteriums-) Wert erzielt eine Person, die im Prädiktor Nr.1, Redefluss, den Wert 4 und in Prädiktor Nr.2, Attraktivität, den Wert 20 hat?
Wie groß ist das dazugehörige Konfidenz- (Vertrauens-) Intervall?

Ausgehend von den b-Gewichten ergibt sich folgende Regressionsgleichung:

$$\hat{y} = 1.344 + 0.416 \times x_1 + 0.063 \times x_2$$
$$\hat{y} = 1.344 + 0.416 \times 4 + 0.063 \times 20 = 1.344 + 1.664 + 1.26 = 4.268$$

Für eine Person mit den oben angeführten Werten hätte man einen Sympathiewert von 4.268 Punkten geschätzt.

Die allgemeine Gleichung für ein Konfidenzintervall lautet: $y - t_{\alpha/2} \times s_e \leq y \leq y + t_{\alpha/2} \times s_e$

Hier kann ein S_e berechnet werden von: $s_e = 1.334$

Heraussuchen eines geeigneten t-Wertes aus Tabelle D: $\alpha = 0.05$, $t_{(\alpha/2=0.025;df=7)} = 2.365$

Berechnen der Grenzen des Konfidenzintervalles

$$4.268 - 2.365 \times 1.334 \leq \hat{y} \leq 4.268 + 2.365 \times 1.334$$
$$1.113 \leq \hat{y} \leq 7.423$$

Das Konfidenzintervall umfasst den Bereich von 1.113 bis 7.423. Innerhalb dieser Grenzen liegt der wahre Sympathiewert der Person mit einer gewissen Wahrscheinlichkeit. Damit wären alle Fragestellungen abgearbeitet.

Übungsaufgaben zu Kapitel 8

030 Wozu dient eine multiple Regressionsanalyse?

031 Assessment-Center

Im Rahmen eines Assessmentcenters wurden mit n = 5 Bewerbern „Postkorb-Aufgaben" (x1) und „Sortier-Aufgaben" (x2) durchgeführt. Sechs Monate später wurden diese fünf Bewerber durch Vorgesetzte hinsichtlich ihrer Arbeitsleistung (y) beurteilt.
Mittels multipler Regressionsanalyse wurden die folgenden b-Gewichte bestimmt:
b_0=2, b_1=1,6, b_2=1.
Ein neuer Bewerber erzielt in der Postkorb Aufgabe 5 Punkte, bei der Sortieraufgabe 3 Punkte. Wie viel Punkte wird er bei der Vorgesetztenbeurteilung voraussichtlich erzielen?

9 Univariate Varianzanalyse

Lernziele

Wenn Sie dieses Kapitel bearbeitet haben, dann sollten Sie wissen,

- wozu eine Varianzanalyse dient,
- wie sich die totale Quadratsumme QS_{total} aufteilt,
- wie sich in einer zweifaktoriellen Varianzanalyse die determinierte Quadratsumme QS_{det} aufteilt,
- wie sich eine einfache Varianzanalyse berechnen lässt,
- wie eine Tafel der Varianzanalyse aussieht und wie sie ausgefüllt wird.

Eine Varianzanalyse vergleicht gleichzeitig mehrere Mittelwerte.
Konnte der t-Test immer nur zwei Mittelwerte vergleichen, können mit einer Varianzanalyse praktisch unbegrenzt viele Mittelwerte miteinander verglichen werden.

Die Varianzanalyse untersucht, wie sehr die einzelnen Mittelwerte variieren!

Nun könnte man selbstverständlich auch mit einem t-Test alle diese Mittelwerte – immer im Paarvergleich – gegeneinander testen. Dies hätte nur einen Nachteil: Denken wir an den Begriff der Signifikanz zurück. Wenn die Wahrscheinlichkeit dafür, dass ein Ereignis zufällig eintritt, kleiner als 5% ist, und dieses Ereignis trotzdem eingetreten ist, sprechen wir von Signifikanz. Wenn wir jetzt nicht nur ein Ereignis betrachten, sondern z.B. hundert Ereignisse, dann ist es sehr gut möglich, dass eines von Ihnen tatsächlich per Zufall signifikant ist. Unter tausend Ereignissen ist praktisch immer mindestens eines, das allein per Zufall signifikant ist. Bereits bei nur 14 Ereignissen ist die Wahrscheinlichkeit dafür, dass mindestens eines davon signifikant ist, 50 Prozent! Wie erkennt man nun, ob ein Ereignis wirklich oder nur scheinbar signifikant ist? Gar nicht! Letzten Endes kann man nur hoffen, dass keine Zufallssignifikanz aufgetreten ist. Warum dieser Ausflug zur Wahrscheinlichkeit und zur Signifikanz?

Wie viele Paarvergleiche ergeben sich, wenn man zehn Mittelwerte gegeneinander testen möchte? Bei zehn Mittelwerten ergeben sich bereits 45 Paare. Und die Wahrscheinlichkeit dafür, dass hier bereits ein zufällig signifikantes Ergebnis auftritt, ist gar nicht mehr so klein. Bei einer Varianzanalyse werden die zehn Mittelwerte sozusagen parallel gegeneinander getestet. Die Wahrscheinlichkeit, dass eine Scheinsignifikanz auftritt, wäre immer noch 5%.

Des Weiteren: Wenn ich zehn Mittelwerte habe, interessiert mich zum einen natürlich, ob sich die zehn Mittelwerte voneinander unterscheiden. Stellen wir uns aber einmal vor, diese zehn Mittelwerte ergeben sich, weil Männer und Frauen in je fünf Altersstufen untersucht werden. Man hätte also z.B. folgendes Versuchsdesign:

	Alterskategorie (in Jahren)				
Geschlecht	15 - 17	18 - 20	21 - 25	26 - 35	36 und älter
weiblich	$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$	$\overline{X}_4$	$\overline{X}_5$
männlich	$\overline{X}_6$	$\overline{X}_7$	$\overline{X}_8$	$\overline{X}_9$	$\overline{X}_{10}$

Tabelle 39: Beispiel eines Versuchsdesigns.
(Quelle: Eigene Darstellung)

Hier wäre noch von Interesse, ob sich die einzelnen Alterskategorien -unabhängig vom Geschlecht- voneinander unterscheiden und weiterhin, ob sich Männer und Frauen -unabhängig vom Alter- voneinander unterscheiden. Darüber hinaus könnte man sich vorstellen, dass zwischen Alter und Geschlecht noch eine Beziehung besteht, insofern, dass die jüngeren Frauen und die älteren Frauen bessere Werte als die jüngeren Männer und die älteren Männer erzielen, dass aber die Männer der Alterskategorie 21 bis 25 Jahre bessere Werte erzielen als die Frauen dieser Altersgruppe.

Weil das sehr schwierig klingt, kommt eine Grafik, die das verdeutlichen soll:

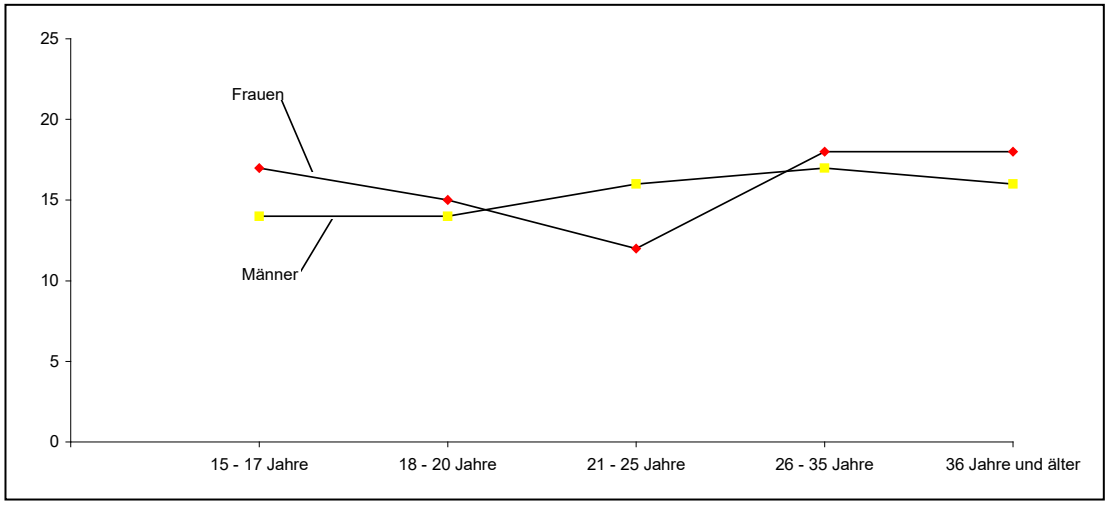

Abbildung 17: Grafische Darstellung einer Wechselwirkung.
(Quelle: Eigene Darstellung)

Was man in dieser Grafik sieht, ist eine sogenannte Wechselwirkung. Und zwar eine Wechselwirkung zwischen den Faktoren Alter und Geschlecht. In der Varianzanalyse spricht man von Faktoren (hier: Alter und Geschlecht) und den dazugehörigen Faktorstufen (hier: fünf Alterskategorien sowie männlich/weiblich). Wenn irgendetwas signifikant wird, spricht man von einem Effekt.

Als Ergebnisse einer Varianzanalyse erhält man nun Aussagen über einen globalen Effekt (die zehn Mittelwerte unterscheiden sich voneinander), über den Haupteffekt des Faktors A (hier: Alter; die Alterskategorien unterscheiden sich voneinander – unabhängig vom Geschlecht), über den Haupteffekt des Faktors B (hier: Geschlecht; Männer und Frauen unterscheiden sich voneinander - unabhängig vom Alter) sowie über eine mögliche Wechselwirkung / einen Wechselwirkungseffekt (die Faktoren A und B hängen in gewisser Weise zusammen). Und das alles auf einmal. Die Varianzanalyse ist selbstverständlich nicht auf zwei Faktoren beschränkt. Sind es drei Faktoren, spricht man vom Haupteffekt Faktor A, Haupteffekt Faktor B, Haupteffekt Faktor C, Wechselwirkung Faktor A × Faktor B, Wechselwirkung A×C, Wechselwirkung B×C sowie der Tripelwechselwirkung A×B×C. Das Ganze lässt sich natürlich noch steigern. In der Regel beschränken sich Varianzanalysen jedoch auf höchstens drei Faktoren.

Die spannende Frage ist nun: Wie wird’s gerechnet? Auch für die Varianzanalyse wird in der Regel Matrixalgebra benötigt. Ein kleines Beispiel lässt sich jedoch auch auf andere Art und Weise berechnen.

Ein Beispiel: Man möchte untersuchen, ob die Dauer der Vorbereitungszeit auf die Statistikklausur und die Regelmäßigkeit des Tutorium-Besuchs einen Einfluss auf die Punktzahl in der Statistikklausur haben. Dazu wurden die Punktzahlen von n = 12 zufällig ausgewählten Studenten erhoben. Der Versuchsplan zu der Untersuchung weist also den Faktor A „Regelmäßigkeit des Tutorium-Besuchs" in zwei Stufen, sowie den Faktor B „Vorbereitungszeit" in zwei Stufen auf. In folgender Tabelle stehen die Punktzahlen der Studenten.

Faktor A: Regelmäßigkeit des Besuchs	Faktor B: Vorbereitungszeit	
	B1: kurzfristige Vorbereitung	B2: langfristige Vorbereitung
A1: regelmäßiger Besuch	12 15 18 (1)	18 21 21 (2)
A2: unregelmäßiger Besuch	13 8 9 (3)	7 14 9 (4)

Tabelle 40: Rohdaten von 12 Personen bei zwei Faktoren mit je zwei Stufen.
(Quelle: Eigene Darstellung)

Um eine Varianzanalyse per Hand rechnen zu können, werden verschiedene Mittelwerte und Quadratsummen – abgekürzt QS – benötigt.

Zuerst berechnen wir einmal den Gesamtmittelwert aller zwölf Werte:

$$\bar{x}_{Gesamt} = \frac{12+15+18+18+21+21+13+8+9+7+14+9}{12} = \frac{165}{12} = 13.75$$

Dann werden noch die Mittelwerte der einzelnen Faktorstufenkombinationen bzw. der einzelnen Zellen benötigt:

$$\bar{x}_{A1,B1\,(Zelle1)} = \frac{12+15+18}{3} = \frac{45}{3} = 15, \quad \bar{x}_{A1,B2\,(Zelle2)} = \frac{18+21+21}{3} = \frac{60}{3} = 20$$

$$\bar{x}_{A2,B1\,(Zelle3)} = \frac{13+8+9}{3} = \frac{30}{3} = 10, \quad \bar{x}_{A2,B2\,(Zelle4)} = \frac{7+14+9}{3} = \frac{30}{3} = 10$$

Jetzt benötigen wir noch die Mittelwerte für die einzelnen Faktorstufen.

$$\bar{x}_{A1} = \frac{12+15+18+18+21+21}{6} = \frac{105}{6} = 17.5, \quad \bar{x}_{A2} = \frac{13+8+9+7+14+9}{6} = \frac{60}{6} = 10$$

$$\bar{x}_{B1} = \frac{12+15+18+13+8+9}{6} = \frac{75}{6} = 12.5, \quad \bar{x}_{B2} = \frac{18+21+21+7+14+9}{6} = \frac{90}{6} = 15$$

Und zum Abschluss brauchen wir noch zwei Mittelwerte, welche für die Berechnung der Wechselwirkung benötigt werden.

$$\bar{x}_{(A1,B1)\,und\,(A2,B2)} = \frac{12+15+18+7+14+9}{6} = \frac{75}{6} = 12.5$$

$$\bar{x}_{(A1,B2)\,und\,(A2,B1)} = \frac{18+21+21+13+8+9}{6} = \frac{90}{6} = 15$$

Nach diesen Vorarbeiten können wir jetzt einige Quadratsummen berechnen. Zuerst die sogenannte totale Quadratsumme QS_{total}. QS_{total} berechnet sich als Summe der quadrierten Differenzen/Abweichungen der einzelnen Rohwerte vom Gesamtmittelwert. Hört sich schlimmer an, als es ist. Der Gesamtmittelwert beträgt 13.75.

Der erste Wert in Zelle 1 ist 12. Die quadrierte Differenz/Abweichung ist dann:

$(12-13.75)^2 = (-1.75)^2 = 3.0625$

Diese Berechnungen müssen nun für alle einzelnen Rohwerte durchgeführt werden. Im Anschluss werden die einzelnen Ergebnisse zusammengezählt. Als Formel könnte man schreiben:

$$QS_{total} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$QS_{total} = (12-13.75)^2 + (15-13.75)^2 + (18-13.75)^2 + (18-13.75)^2 + (21-13.75)^2 + (21-13.75)^2 + (13-13.75)^2 + (8-13.75)^2 + (9-13.75)^2 + (7-13.75)^2 + (14-13.75)^2 + (9-13.75)^2$$

$$QS_{total} = (-1.75)^2 + (1.25)^2 + (4.25)^2 + (4.25)^2 + (7.25)^2 + (7.25)^2 + (-0.75)^2 + (-5.75)^2 + (-4.75)^2 + (-6.75)^2 + (0.25)^2 + (-4.75)^2$$

$$QS_{total} = 3.0625 + 1.5625 + 18.0625 + 18.0625 + 52.5625 + 52.5625 + 0.5625 + 33.0625 + 22.5625 + 45.5625 + 0.0625 + 22.5625$$

$$QS_{total} = 270.25$$

Man könnte nun sagen: Insgesamt (total) variieren die einzelnen Rohwerte um den Gesamtmittelwert mit dem Betrag 270.25.

Wie geht es jetzt weiter? Im nächsten Schritt werden die einzelnen Rohwerte in den Zellen durch den jeweiligen Zellenmittelwert ersetzt. Wir bekommen also folgende Tabelle:

Faktor A: Regelmäßigkeit des Besuchs	Faktor B: Vorbereitungszeit	
	B1: kurzfristige Vorbereitung	B2: langfristige Vorbereitung
A1: regelmäßiger Besuch	15 15 15 [1]	20 20 20 [2]
A2: unregelmäßiger Besuch	10 10 10 [3]	10 10 10 [4]

Tabelle 41: Zellenmittelwerte bei zwei Faktoren mit je zwei Stufen.
(Quelle: Eigene Darstellung)

OK, jetzt kommt ein kleiner, aber äußerst wichtiger Gedankensprung: In den einzelnen Zellen stehen nicht immer die gleichen Werte (z.B. dreimal die 15 in Zelle 1), sondern eben (interindividuell) unterschiedliche Werte (12, 15, 18), d.h. auch innerhalb einer Zelle gibt es eine gewisse Variation der Werte. Durch die Ersetzung der Rohwerte durch den jeweiligen Zellenmittelwert haben wir die Variation innerhalb der Zellen eliminiert bzw. auf Null gesetzt. Jetzt gibt es nur noch eine Variation zwischen den Zellen, aber keine Variation innerhalb der Zellen mehr. Jegliche Variation der Werte geht jetzt nur noch auf die Zellenzugehörigkeit zurück und nicht mehr auf irgendwelche interindividuellen Unterschiede. Mit dieser Tabelle wird nun die sogenannte determinierte Quadratsumme QS_{det} berechnet. Mit dem Begriff „determiniert" ist gemeint, dass die einzelnen Abweichungen vom Gesamtmittelwert durch die Zellenzugehörigkeit begründet (determiniert) sind.

Um QS_{det} zu berechnen, wird jetzt analog zu QS_{total} wieder die Summe der quadrierten Abweichungen zum Gesamtmittelwert gebildet.

$$\begin{aligned}
QS_{\text{det}} &= (15-13.75)^2+(15-13.75)^2+(15-13.75)^2+(20-13.75)^2+(20-13.75)^2 \\
&\quad +(20-13.75)^2+(10-13.75)^2+(10-13.75)^2+(10-13.75)^2+(10-13.75)^2 \\
&\quad +(10-13.75)^2+(10-13.75)^2 \\
QS_{\text{det}} &= (1.25)^2+(1.25)^2+(1.25)^2+(6.25)^2+(6.25)^2+(6.25)^2+(-3.75)^2+(-3.75)^2 \\
&\quad +(-3.75)^2+(-3.75)^2+(-3.75)^2+(-3.75)^2 \\
QS_{\text{det}} &= 1.5625+1.5625+1.5625+39.0625+39.0625+39.0625+14.0625+14.0625 \\
&\quad +14.0625+14.0625+14.0625+14.0625 \\
QS_{\text{det}} &= 206.25
\end{aligned}$$

Man könnte nun sagen: Bedingt (determiniert) durch die Zellenzugehörigkeit variieren die einzelnen Werte um den Gesamtmittelwert mit dem Betrag 206.25.

Wie wäre es nun, wenn nicht die Zellenzugehörigkeit entscheidend wäre, sondern die Zugehörigkeit zu Faktorstufe A1 oder zu Faktorstufe A2?

Ersetzen wir die einzelnen Rohwerte in den Zellen durch den jeweiligen Mittelwert für Faktorstufe A1 bzw. für Faktorstufe A2. Wir bekommen dann folgende Tabelle:

Faktor A: Regelmäßigkeit des Besuchs		
A1: regelmäßiger Besuch	17.5 17.5 17.5 1	17.5 17.5 17.5 2
A2: unregelmäßiger Besuch	10 10 10 3	10 10 10 4

Tabelle 42: Zellenmittelwerte bei Faktor A mit zwei Stufen A1, A2.
(Quelle: Eigene Darstellung)

In dieser Tabelle geht die Variation der Werte nur noch auf die Zugehörigkeit zu Faktorstufe A1 bzw. die Zugehörigkeit zu Faktorstufe A2 zurück. Auch hier kann wieder die Summe der quadrierten Abweichungen vom Gesamtmittelwert (QS_A) berechnet werden.

$$\begin{aligned}
QS_A &= (17.5-13.75)^2+(17.5-13.75)^2+(17.5-13.75)^2+(17.5-13.75)^2+(17.5-13.75)^2 \\
&\quad +(17.5-13.75)^2+(10-13.75)^2+(10-13.75)^2+(10-13.75)^2+(10-13.75)^2 \\
&\quad +(10-13.75)^2+(10-13.75)^2 \\
QS_A &= (3.75)^2+(3.75)^2+(3.75)^2+(3.75)^2+(3.75)^2+(3.75)^2+(-3.75)^2+(-3.75)^2 \\
&\quad +(-3.75)^2+(-3.75)^2+(-3.75)^2+(-3.75)^2 \\
QS_A &= 14.0625+14.0625+14.0625+14.0625+14.0625+14.0625+14.0625+14.0625 \\
&\quad +14.0625+14.0625+14.0625+14.0625 \\
QS_A &= 168.75
\end{aligned}$$

Man könnte nun sagen: Bedingt (determiniert) durch die Zugehörigkeit zur Faktorstufe A1 bzw. zur Faktorstufe A2 variieren die einzelnen Werte um den Gesamtmittelwert mit dem Betrag 168.75.

Wie wäre es nun, wenn nicht die Zellenzugehörigkeit entscheidend wäre, sondern die Zugehörigkeit zu Faktorstufe B1 oder zu Faktorstufe B2? Ersetzen wir die einzelnen Rohwerte in den Zellen durch den jeweiligen Mittelwert für Faktorstufe B1 bzw. für Faktorstufe B2. Wir bekommen also folgende Tabelle:

	Faktor B: Vorbereitungszeit	
	B1: kurzfristige Vorbereitung	B2: langfristige Vorbereitung
	12.5	15
	12.5	15
	12.5 | 1	15 | 2
	12.5	15
	12.5	15
	12.5 | 3	15 | 4

Tabelle 43: Zellenmittelwerte bei Faktor B mit zwei Stufen B1, B2.
(Quelle: Eigene Darstellung)

In dieser Tabelle geht die Variation der Werte nur noch auf die Zugehörigkeit zu Faktorstufe B1 bzw. die Zugehörigkeit zu Faktorstufe B2 zurück. Auch hier kann wieder die Summe der quadrierten Abweichungen vom Gesamtmittelwert (QS_B) berechnet werden.

$$QS_B = (12.5-13.75)^2 + (12.5-13.75)^2 + (12.5-13.75)^2 + (12.5-13.75)^2 + (12.5-13.75)^2 + (12.5-13.75)^2 + (15-13.75)^2 + (15-13.75)^2 + (15-13.75)^2 + (15-13.75)^2 + (15-13.75)^2 + (15-13.75)^2$$

$$QS_B = (-1.25)^2 + (-1.25)^2 + (-1.25)^2 + (-1.25)^2 + (-1.25)^2 + (-1.25)^2 + (1.25)^2 + (1.25)^2 + (1.25)^2 + (1.25)^2 + (1.25)^2 + (1.25)^2$$

$$QS_B = 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625$$

$$QS_B = 18.75$$

Man könnte nun sagen: Bedingt (determiniert) durch die Zugehörigkeit zur Faktorstufe B1 bzw. zur Faktorstufe B2 variieren die einzelnen Werte um den Gesamtmittelwert mit dem Betrag 18.75.

Wie wäre es nun, wenn nicht die Zellenzugehörigkeit entscheidend wäre, sondern die Zugehörigkeit zu den Faktorstufen**kombinationen** A1,B1 und A2,B2 oder zu den Faktorstufenkombinationen A1,B2 und A2,B1? Anders ausgedrückt: Wie wäre es bei einer Wechselwirkung zwischen den Faktoren A und B?

Ersetzen wir die einzelnen Rohwerte in den Zellen durch den jeweiligen Mittelwert für die betreffenden Faktorstufenkombinationen. Wir bekommen also folgende Tabelle:

	Faktor B: Vorbereitungszeit	
	B1: kurzfristige Vorbereitung	B2: langfristige Vorbereitung
	12.5	15
	12.5	15
	12.5 | 1	15 | 2
	15	12.5
	15	12.5
	15 | 3	12.5 | 4

Tabelle 44: Zellenmittelwerte bei Faktor B mit zwei Stufen B1, B2.
(Quelle: Eigene Darstellung)

In dieser Tabelle geht die Variation der Werte nur noch auf die Zugehörigkeit zu den Faktorstufenkombinationen A1,B1 und A2,B2 bzw. die Zugehörigkeit zu den Faktorstufenkombinationen A1,B2 und A2,B1 zurück.

Auch hier kann wieder die Summe der quadrierten Abweichungen vom Gesamtmittelwert (QS_{A*B}) berechnet werden.

$$QS_{A*B} = (12.5-13.75)^2 + (12.5-13.75)^2 + (12.5-13.75)^2 + (15-13.75)^2 + (15-13.75)^2 + (15-13.75)^2 + (15-13.75)^2 + (15-13.75)^2 + (15-13.75)^2 + (12.5-13.75)^2 + (12.5-13.75)^2$$

$$QS_{A*B} = (-1.25)^2 + (-1.25)^2 + (-1.25)^2 + (1.25)^2 + (1.25)^2 + (1.25)^2 + (1.25)^2 + (1.25)^2 + (1.25)^2 + (-1.25)^2 + (-1.25)^2 + (-1.25)^2$$

$$QS_{A*B} = 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625 + 1.5625$$

$$QS_{A*B} = 18.75$$

Man könnte nun sagen: Bedingt (determiniert) durch eine Wechselwirkung zwischen den Faktoren A und B variieren die einzelnen Werte um den Gesamtmittelwert mit dem Betrag 18.75.

Übrigens: Bei einer zweifaktoriellen Varianzanalyse gilt für die determinierte Quadratsumme QS_{det}: $QS_{det} = QS_A + QS_B + QS_{A*B}$.

Daraus folgt: $QS_{A*B} = QS_{det} - QS_A - QS_B$. Man hätte die Quadratsumme der Wechselwirkung also auch wesentlich schneller ermitteln können.

Es bleibt eine letzte Quadratsumme, die berechnet werden muss: Die sogenannte Fehlerquadratsumme QS_{error} bzw. QS_e.

Dazu muss man folgendes wissen: Die totale Quadratsumme QS_{total} setzt sich zusammen aus determinierter Quadratsumme QS_{det} und Fehlerquadratsumme QS_e.

$QS_{total} = QS_{det} + QS_e$

Daraus folgt: $QS_e = QS_{total} - QS_{det} = 270.25 - 206.25 = 64$

Fassen wir die bisherigen Ergebnisse in einer Tabelle zusammen. Man spricht dann von einer Tafel der Varianzanalyse.

Quelle der Variation	Quadratsumme	Freiheitsgrade	F	R^2
Faktor A: Regelmäßigkeit des Besuchs	$QS_A = 168.75$			
Faktor B: Vorbereitungszeit	$QS_B = 18.75$			
Wechselwirkung A×B	$QS_{A\times B} = 18.75$			
determiniert	$QS_{det} = 206.25$			
Fehler	$QS_e = 64$			
total	$QS_{total} = 270.25$			

Tabelle 45: Tafel der Varianzanalyse, erster Teil. (Quelle: Eigene Darstellung)

In dieser Tafel fehlen einige Werte, die bisher noch nicht ausgerechnet wurden.
Beginnen wir mit den Freiheitsgraden. Die Freiheitsgrade für QS_{total} berechnen sich als $df_{total} = n-1$, hier also 12–1 = 11.

Die Freiheitsgrade für QS_A und QS_B berechnen sich jeweils als Anzahl Faktorstufen –1, hier also $df_A = 2–1 = 1$ und $df_B = 2–1 = 1$.

Die Freiheitsgrade für QS_{det} berechnen sich als Anzahl Zellen -1: $df_{det} = 4–1 = 3$.

Die Freiheitsgrade für QS_{A*B} berechnen sich als $df_{A*B} = df_A \times df_B = 1 \times 1 = 1$.

Bleiben noch die Fehlerfreiheitsgrade df_e. Sie berechnen sich als $df_e = df_{total} – df_{det}$.

Analog zu den Quadratsummen gilt für die Freiheitsgrade: $df_{total} = df_{det} + df_e$

und bei einer zweifaktoriellen Varianzanalyse: $df_{det} = df_A + df_B + df_{A*B}$.

Tragen wir die Werte in die Tafel der Varianzanalyse ein.

Quelle der Variation	Quadratsumme	Freiheitsgrade	F	R^2
Faktor A: Regelmäßigkeit des Besuchs	$QS_A = 168.75$	1		
Faktor B: Vorbereitungszeit	$QS_B = 18.75$	1		
Wechselwirkung A×B	$QS_{A×B} = 18.75$	1		
determiniert	$QS_{det} = 206.25$	3		
Fehler	$QS_e = 64$	8		
total	$QS_{total} = 270.25$	11		

Tabelle 46: Tafel der Varianzanalyse, erster Teil.
(Quelle: Eigene Darstellung)

Was uns jetzt noch fehlt, sind Fragestellungen, daraus abgeleitete Hypothesen und eine Anweisung, wie man die Hypothesen auf Signifikanz testet.

Bei einer zweifaktoriellen Varianzanalyse können vier Fragestellungen bearbeitet werden.

1. Unterscheiden sich die vier Zellen im Durchschnitt voneinander oder sind sie gleich?
   Als Nullhypothese könnte man aufstellen:
   H0: Die vier Zellenmittelwerte sind gleich.
   H1: Die vier Zellenmittelwerte sind nicht gleich.

   Mathematisch: H0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
   Zu dieser einen mathematischen Nullhypothese gibt es jetzt leider mehrere Alternativhypothesen....
   H1a: $\mu_1 \neq \frac{\mu_2 + \mu_3 + \mu_4}{3}$ Der erste Mittelwert unterscheidet sich von den drei anderen.
   H1b: $\mu_2 \neq \frac{\mu_3 + \mu_4}{2}$ Der zweite Mittelwert unterscheidet sich von beiden anderen.
   H1c: $\mu_3 \neq \mu_4$ Der dritte Mittelwert unterscheidet sich vom vierten.

   Weil das so ist, werden bei einer Varianzanalyse häufig nur die Nullhypothesen aufgelistet.

2. Gibt es einen Unterschied zwischen den Faktorstufen des Faktors A?
   H0: Die Mittelwerte der Faktorstufen des Faktors A sind gleich
   H1: Die Mittelwerte der Faktorstufen des Faktors A sind nicht gleich.

   H0: $\mu_{A1} = \mu_{A2}$ $\mu_1 + \mu_2 = \mu_3 + \mu_4$
   H1: $\mu_{A1} \neq \mu_{A2}$ $\mu_1 + \mu_2 \neq \mu_3 + \mu_4$

3. Gibt es einen Unterschied zwischen den Faktorstufen des Faktors B?
   H0: Die Mittelwerte der Faktorstufen des Faktors B sind gleich
   H1: Die Mittelwerte der Faktorstufen des Faktors B sind nicht gleich.

   H0: $\mu_{B1} = \mu_{B2}$ $\quad\quad$ $\mu_1 + \mu_3 = \mu_2 + \mu_4$
   H1: $\mu_{B1} \neq \mu_{B2}$ $\quad\quad$ $\mu_1 + \mu_3 \neq \mu_2 + \mu_4$

4. Gibt es eine Wechselwirkung zwischen den Faktoren A und B?
   H0: Der gemeinsame Mittelwert der Faktorstufenkombinationen A1,B1 und A2,B2 unterscheidet sich nicht vom gemeinsamen Mittelwert der Faktorstufenkombinationen A1,B2 und A2,B1.
   H1: Der gemeinsame Mittelwert der Faktorstufenkombinationen A1,B1 und A2,B2 unterscheidet sich vom gemeinsamen Mittelwert der Faktorstufenkombinationen A1,B2 und A2,B1.

   H0: $\mu_{A1,B1 \text{ und } A2,B2} = \mu_{A1,B2 \text{ und } A2,B1}$ $\quad\quad$ $\mu_1 + \mu_4 = \mu_2 + \mu_3$

   Als Alternativhypothese H1 hat man dann
   H1: $\mu_{A1,B1 \text{ und } A2,B2} \neq \mu_{A1,B2 \text{ und } A2,B1}$ $\quad\quad$ $\mu_1 + \mu_4 \neq \mu_2 + \mu_3$

Wir haben die Fragestellungen, wir haben die Hypothesen, es bleibt die Anweisung, wie die Hypothesen getestet werden können.

Es ist zum Glück nur eine, relativ kleine Formel: $F_{emp} = \dfrac{QS_h / df_h}{QS_e / df_e}$

QS_h entspricht der zu testenden Hypothesenquadratsumme, df_h den dazugehörigen Freiheitsgraden.

zu 1) Unterscheiden sich die vier Zellen im Durchschnitt voneinander oder sind sie gleich? Die zu dieser Hypothese gehörende Quadratsumme ist QS_{det}. Zu QS_{det} gehört df_{det}.
Dann rechnen wir doch:

$$F_{emp} = \frac{QS_h / df_h}{QS_e / df_e} = \frac{206.25/3}{64/8} = \frac{68.75}{8} = 8.59$$

Als kritischen F-Wert findet man in Tabelle B: $F_{krit} < F_{emp} \Rightarrow$ signifikant $\Rightarrow$ H1!

$F_{krit(\alpha=0.05;df1=3;df2=8)} = 4.07$

Mit einer Irrtumswahrscheinlichkeit von 5% kann behauptet werden, dass sich die Zellenmittelwerte insgesamt voneinander unterscheiden.

zu 2) Gibt es einen Unterschied zwischen den Faktorstufen des Faktors A?
Die zu dieser Hypothese gehörende Quadratsumme ist QS_A. Zu QS_A gehört df_A.

$$F_{emp} = \frac{QS_h / df_h}{QS_e / df_e} = \frac{168.75/1}{64/8} = \frac{168.75}{8} = 21.09$$

Als kritischen F-Wert findet man in Tabelle B: $F_{krit(\alpha=0.05;df1=1;df2=8)} = 5.32$

$F_{krit} < F_{emp} \Rightarrow$ signifikant $\Rightarrow$ H1!

Mit einer Irrtumswahrscheinlichkeit von 5% kann behauptet werden, dass es einen Haupteffekt des Faktors A gibt; es macht einen signifikanten Unterschied, ob das Tutorium regelmäßig oder unregelmäßig besucht wird. So soll es sein.

zu 3) Gibt es einen Unterschied zwischen den Faktorstufen des Faktors B?
Die zu dieser Hypothese gehörende Quadratsumme ist QS_B. Zu QS_B gehört df_B.

$$F_{emp} = \frac{QS_h / df_h}{QS_e / df_e} = \frac{18.75/1}{64/8} = \frac{18.75}{8} = 2.34$$

Als kritischen F-Wert findet man in Tabelle B: $F_{krit(\alpha=0.05;df1=1;df2=8)} = 5.32$

$F_{emp} < F_{krit} \Rightarrow$ nicht signifikant $\Rightarrow$ H0!

Es kann nicht behauptet werden, dass es einen Haupteffekt des Faktors B gibt; es macht keinen Unterschied, ob kurz- oder langfristig vorbereitet wird. Schade eigentlich!

zu 4) Gibt es eine Wechselwirkung zwischen den Faktoren A und B?
Die zu dieser Hypothese gehörende Quadratsumme ist QS_{A*B}. Zu QS_{A*B} gehört df_{A*B}.

$$F_{emp} = \frac{QS_h / df_h}{QS_e / df_e} = \frac{18.75/1}{64/8} = \frac{18.75}{8} = 2.34$$

Als kritischen F-Wert findet man in Tabelle B: $F_{krit(\alpha=0.05;df1=1;df2=8)} = 5.32$

$F_{emp} < F_{krit} \Rightarrow$ nicht signifikant $\Rightarrow$ H0!

Es kann nicht behauptet werden, dass es eine Wechselwirkung zwischen den Faktoren A und B gibt.

Bezogen auf die Tafel der Varianzanalyse fehlen jetzt lediglich noch die Werte für R^2. Hierzu gibt es folgende Formel:

$$R^2 = \frac{QS_h}{QS_{total}}$$

Dann füllen wir mal die Tafel auf:

Quelle der Variation	Quadratsumme	Freiheitsgrade	F	R^2
Faktor A: Regelmäßigkeit des Besuchs	$QS_A = 168.75$	1	21.09*	=168.75/270.25 =0.624
Faktor B: Vorbereitungszeit	$QS_B = 18.75$	1	2.34 n.s.	0.069
Wechselwirkung A×B	$QS_{A\times B} = 18.75$	1	2.34 n.s.	0.069
determiniert	$QS_{det} = 206.25$	3	8.59*	0.763
Fehler	$QS_e = 64$	8		
total	$QS_{total} = 270.25$	11		

* = signifikant n.s. = nicht signifikant

Tabelle 47: Tafel der Varianzanalyse, zweiter Teil.
(Quelle Eigene Darstellung)

Wie interpretiert man nun das R^2? Praktisch genauso, wie man den Determinationskoeffizienten r^2 interpretieren würde, d.h. mit 100 multiplizieren und dann wie folgt ausdrücken:

Der Faktor A „Regelmäßigkeit des Besuchs des Tutoriums" erklärt 62% der Variation der Ergebnisse in der Statistik-Klausur.

Übungsaufgaben zu Kapitel 9

032 Was macht eine Varianzanalyse?

033 In welche Bestandteile kann die totale Quadratsumme zerlegt werden?

034 In welche Bestandteile kann die determinierte Quadratsumme bei einer zweifaktoriellen Varianzanalyse zerlegt werden

035 Kundenzufriedenheitsstudie

Im Rahmen einer Kundenzufriedenheitsstudie wurde die Kundenzufriedenheit einer Varianzanalytischen Prüfung unterzogen. Geprüft wurde, ob die Kundenzufriedenheit variierte nach einem Faktor A „Haarfarbe Kundenbetreuer" mit den drei Stufen „blond", „braun" und „schwarz" sowie einem Faktor B „Geschlecht Kundenbetreuer" mit den zwei Stufen „weiblich" und „männlich". Die Ergebnisse der Varianzanalyse wurden in eine Tafel der Varianzanalyse übertragen. Unglücklicherweise gerieten Kaffeeflecken auf diese Tafel, so dass verschiedene Zahlen nicht mehr lesbar waren ...

Bitte vervollständigen Sie die Tafel der Varianzanalyse, d.h. ersetzen Sie die Fragezeichen durch Zahlen bzw. durch die Kürzel „sig." für Signifikanz oder „n.s." für nicht signifikant in der Spalte „Signifikanz (Sig.)"!

Gehen Sie bei der Prüfung auf Signifikanz immer von einem alpha=5% bei einseitiger Fragestellung aus! (6 Punkte)

Quelle der Variation	Quadratsumme	Freiheitsgrade	F	Sig.
Faktor A	48	?	?	n.s.
Faktor B	18	?	?	?
Wechselwirkung A*B	?	2	8,15	?
determinierte / Modell	210	?	?	?
Fehler	?	12		
total	306	?		

Raum für Notizen

10 Kruskal-Wallis-H-Test

Lernziele

Wenn Sie dieses Kapitel bearbeitet haben, dann sollten Sie wissen,

- unter welchen Voraussetzungen der Kruskal-Wallis-H-Test angewendet wird,
- wie man ihn berechnet.

Die Einweg-Rangvarianzanalyse nach Kruskal-Wallis kommt zum Einsatz, wenn bei ordinalskalierten Variablen mehr als zwei Gruppen miteinander verglichen werden sollen.

Der H-Test ist somit die Erweiterung des U-Tests nach Mann-Whitney. Analog zum U-Test müssen auch hier zuerst die erhobenen Werte in Rangplätze umgewandelt werden.

Die Formel zur Berechnung des empirischen H-Wertes ist vergleichsweise einfach:

$$H_{emp} = \frac{12}{N \times (N+1)} \times \sum_{j=1}^{k} \frac{T_j^2}{n_j} - 3 \times (N+1)$$

Legende:
N = Gesamtanzahl Probanden
T = Rangplatzsumme einer Faktorstufe
n = Anzahl Probanden einer Faktorstufe
12 = Konstante
3 = Konstante

Anders als in der behandelten Varianzanalyse für intervallskalierte Daten, kann der H-Test jedoch nur mit einem Faktor arbeiten (daher: Einweg ...), d.h. wenn zwei Faktoren (z.B. Geschlecht in den Stufen „weiblich“ vs. „männlich“ und Stimmhöhe in den Stufen „hoch“ vs. „tief“) vorliegen, muss hieraus zuerst ein Faktor gebildet werden, der alle Faktorstufenkombinationen abdeckt. Beispielsweise könnte dieser Faktor Bedingung heißen und vier Faktorstufen umfassen:

Faktor Bedingung: Stufe 1 = weiblich und hohe Stimme
Stufe 2 = weiblich und tiefe Stimme
Stufe 3 = männlich und hohe Stimme
Stufe 4 = männlich und tiefe Stimme.

In dem neuen Faktor 'Bedingung' sind jetzt alle Stufenkombinationen der Faktoren 'Geschlecht' und 'Stimmhöhe' vertreten.

Ein Beispiel: Gemessen wird die ausgestrahlte Autorität (AV) – auf einer Skala von 1 = keine Autorität bis 50 = hohe Autorität – bei weiblichen und männlichen Führungskräften. Bezüglich der ausgestrahlten Autorität sei bekannt, dass diese Werte nicht einer Normalverteilung folgen. Daher wird nicht von einem Intervallskalenniveau ausgegangen, sondern von einem Ordinalskalenniveau. Bei den Führungskräften wird ebenfalls noch ermittelt, ob sie eher eine hohe oder eine tiefe Stimme haben. Insgesamt werden die Führungskräfte also vier Gruppen (oder Faktorstufen) zugeordnet. Zu jeder Gruppe wurden mehrere Mitarbeiter befragt.

Hier sind die Daten:

Faktor Bedingung			
Stufe 1: weiblich, hohe Stimme	**Stufe 2: weiblich, tiefe Stimme**	**Stufe 3: männlich, hohe Stimme**	**Stufe 4: männlich, tiefe Stimme**
48	43	39	18
36	28	35	33
19	30	12	44
45	46	21	29
37	34	27	26
42	38	41	32
	49	16	

Tabelle 48: Rohwerte „Ausgestrahlte Autorität" nach Faktorstufen. (Quelle: Eigene Darstellung)

Im ersten Schritt müssen wieder Rangplätze vergeben werden. Da hohe Werte einer hohen ausgestrahlten Autorität entsprechen, soll hier dem höchsten Wert der Rangplatz 1 zugeordnet werden. Zusätzlich werden dann pro Faktorstufe die Rangplatzsummen T1 bis 4 gebildet.

Faktor Bedingung			
Stufe 1: weiblich, hohe Stimme	**Stufe 2: weiblich, tiefe Stimme**	**Stufe 3: männlich, hohe Stimme**	**Stufe 4: männlich, tiefe Stimme**
2. (48)	6. (43)	9. (39)	24. (18)
12. (36)	19. (28)	13. (35)	15. (33)
23. (19)	17. (30)	26. (12)	5. (44)
4. (45)	3. (46)	22. (21)	18. (29)
11. (37)	14. (34)	20. (27)	21. (26)
7. (42)	10. (38)	8. (41)	16. (32)
	1. (49)	25. (16)	
T1=59	T2=70	T3=123	T4=99
n1=6	n2=7	n3=7	n4=6

Tabelle 49: Rangplätze „Ausgestrahlte Autorität" nach Faktorstufen. (Quelle: Eigene Darstellung)

Insgesamt liegen N=26 Bewertungen vor. Es gibt k = 4 Gruppen / Faktorstufen.

Kommen wir zu den Hypothesen:
H0: Die Gruppen unterscheiden sich hinsichtlich der ausgestrahlten Autorität nicht voneinander.
H1: Die Gruppen unterscheiden sich hinsichtlich der ausgestrahlten Autorität voneinander.

Bevor jetzt der empirische H-Wert berechnet wird, empfiehlt es sich erfahrungsgemäß, die Rangplatzsummen einer kleinen Prüfung zu unterziehen. Notwendig ist das nicht, d.h. wer nicht prüfen will, lässt dies einfach weg.

Die Summe der Rangplatzsummen T1 bis T4 muss der Hälfte des Produkts von Gesamtanzahl mal (Gesamtanzahl +1) entsprechen:
Als Formel:

$$\sum_{i=1}^{k} T_i = \frac{N \times (N+1)}{2}$$

$$59 + 70 + 123 + 99 = \frac{26 \times 27}{2}$$

$$351 = \frac{702}{2} \quad OK!$$

Nach dieser kleinen Prüfung erfolgt nun die Berechnung des empirischen H-Wertes.

$$H_{emp} = \frac{12}{N \times (N+1)} \times \sum_{j=1}^{k} \frac{T_j^2}{n_j} - 3 \times (N+1)$$

$$H_{emp} = \frac{12}{26 \times 27} \times \left(\frac{59^2}{6} + \frac{70^2}{7} + \frac{123^2}{7} + \frac{99^2}{6} \right) - 3 \times 27$$

$$H_{emp} = \frac{12}{702} \times \left(\frac{3481}{6} + \frac{4900}{7} + \frac{15129}{7} + \frac{9801}{6} \right) - 81$$

$$H_{emp} = 0{,}017 \times (580{,}167 + 700 + 2161{,}286 + 1633{,}5) - 81$$

$$H_{emp} = 0{,}017 \times 5074{,}953 - 81 = 86{,}274 - 81 = 5{,}274$$

Jetzt fehlt nur noch ein kritischer H-Wert, mit dem der empirische H-Wert verglichen werden kann. Statt einer neuen Tabelle kann auf die Chi²-Tabelle (mit k-1 Freiheitsgraden) zurückgegriffen werden, wenn es mindestens k = 3 Gruppen sind UND in jeder Gruppe mindestens sechs Werte vorliegen. Dies trifft hier zu (vier Gruppen, in jeder Gruppe mindestens sechs Werte).

Als kritischen H-Wert kann man daher aus Tabelle D bei df = 3 Freiheitsgraden ablesen:
H_{krit}=7.815.
$H_{emp} < H_{krit}$ → nicht signifikant →H0!

Es kann nicht behauptet werden, dass sich die vier Gruppen hinsichtlich der ausgestrahlten Autorität voneinander unterscheiden.

Hat man weniger als sechs Werte pro Gruppe, wird eine spezielle Tabelle kritischer H-Werte benötigt.[14]

Das obige Beispiel war vergleichsweise einfach, da es keine verbundenen Rangplätze gab: Jeder Rohwert tauchte nur ein einziges Mal auf, daher gab es jeden Rangplatz auch nur ein Mal. In der Praxis ist dies leider eher selten der Fall. Für solche Fälle wurde eine Korrekturformel entwickelt. Diese Korrekturformel berechnet einen Korrekturfaktor, an Hand dessen der empirische H-Wert gewichtet werden muss.

14 zu finden beispielsweise bei: Siegel, S.: 2001

Übungsaufgaben zu Kapitel 10

036 Was sind die Voraussetzungen für einen Kruskal-Wallis-H-Test?

037 Bewerberauswahl

Im Rahmen einer Bewerberauswahl haben drei verschiedene Interviewer per Zufall jeweils sechs Bewerber zugewiesen bekommen. Nach dem Interview haben die Interviewer für jeden Bewerber eine Punktzahl zwischen 1 = „gar nicht geeignet" bis 30 „absolut geeignet" vergeben. Für diese Punktevergabe kann kein Intervallskalenniveau angenommen werden. Daher wurden die Punkte bereits in Rangplätze umgewandelt!
Hier sind die Daten (Rangplätze):

Interviewer 1	Interviewer 2	Interviewer 3
5.	17.	7.
15.	13.	8.
10.	12.	2.
9.	11.	6.
4.	14.	16.
18.	3.	1.
Summe: 61	Summe: 70	Summe: 40

Sie stellen sich nun die Frage, ob es vielleicht bei der Punktevergabe Interviewereffekte gibt.

11 Grundlagen

11.1 Rechnen mit dem Summenzeichen

Lautet die Rechenvorschrift: „Addiere alle Werte der Variablen x und teile durch die Anzahl der Werte, dann hast du den Mittelwert!“ wird ein Mathematiker schreiben:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Der Strich über dem x ist das Kürzel für einen Mittelwert: $\bar{x}$ ist also der Mittelwert der Variablen x. Das n steht im Allgemeinen für die Anzahl der Werte. Hat man fünf Versuchspersonen, hat man ein n = 5.

Das Summenzeichen Σ ist die Abkürzung für die Rechenregel: „Addiere die Werte von/bis“. Das 'von' steht unterhalb des Summenzeichens, nämlich von i = 1, d.h. vom ersten Wert, und das 'bis' steht über dem Summenzeichen, nämlich n, d.h. bis zum letzten Wert. Welche Werte addiert werden sollen, steht rechts neben dem Summenzeichen, nämlich die Werte der Variablen x. Das kleine 'i' kennzeichnet, dass die Werte von der ersten (i = 1) bis zur n-ten (n) Versuchsperson addiert werden sollen. Das kleine 'i' nennt man einen Index (Mehrzahl: Indices).

11.2 Griechische Buchstaben

Α,α	alpha	a	Β,β	beta	b
Γ,γ	gamma	g	Δ,δ	delta	d
Ε,ε	epsilon	e	Ζ,ζ	zeta	z
Η,η	eta	ä	Θ,θ	theta	th
Ι,ι	iota	i	Κ,κ	kappa	k
Λ,λ	lambda	l	Μ,μ	my	m
Ν,ν	ny	n	Ξ,ξ	xi	x
Ο,ο	omikron	o (kurz)	Π,π	pi	p
Ρ,ρ	rho	r	Σ,σ	sigma	s
Τ,τ	tau	t	Υ,υ	ypsilon	y
Φ,ϕ	phi	ph	Χ,χ	chi	ch
Ψ,ψ	psi	ps	Ω,ω	omega	o (lang)

Raum für Notizen

Tabellarischer Anhang

Tabelle A: z-Werte (z-Wert, Fläche/Wahrscheinlichkeit; einseitig; 1-α)

-3.00	0.0013	-2.49	0.0064	-1.99	0.0233	-1.49	0.0681	-0.99	0.1611
-2.99	0.0014	-2.48	0.0066	-1.98	0.0239	-1.48	0.0694	-0.98	0.1635
-2.98	0.0014	-2.47	0.0068	-1.97	0.0244	-1.47	0.0708	-0.97	0.1660
-2.97	0.0015	-2.46	0.0069	-1.96	0.0250	-1.46	0.0721	-0.96	0.1685
-2.96	0.0015	-2.45	0.0071	-1.95	0.0256	-1.45	0.0735	-0.95	0.1711
-2.95	0.0016	-2.44	0.0073	-1.94	0.0262	-1.44	0.0749	-0.94	0.1736
-2.94	0.0016	-2.43	0.0075	-1.93	0.0268	-1.43	0.0764	-0.93	0.1762
-2.93	0.0017	-2.42	0.0078	-1.92	0.0274	-1.42	0.0778	-0.92	0.1788
-2.92	0.0018	-2.41	0.0080	-1.91	0.0281	-1.41	0.0793	-0.91	0.1814
-2.91	0.0018	-2.40	0.0082	-1.90	0.0287	-1.40	0.0808	-0.90	0.1841
-2.90	0.0019	-2.39	0.0084	-1.89	0.0294	-1.39	0.0823	-0.89	0.1867
-2.89	0.0019	-2.38	0.0087	-1.88	0.0301	-1.38	0.0838	-0.88	0.1894
-2.88	0.0020	-2.37	0.0089	-1.87	0.0307	-1.37	0.0853	-0.87	0.1922
-2.87	0.0021	-2.36	0.0091	-1.86	0.0314	-1.36	0.0869	-0.86	0.1949
-2.86	0.0021	-2.35	0.0094	-1.85	0.0322	-1.35	0.0885	-0.85	0.1977
-2.85	0.0022	-2.34	0.0096	-1.84	0.0329	-1.34	0.0901	-0.84	0.2005
-2.84	0.0023	-2.33	0.0099	-1.83	0.0336	-1.33	0.0918	-0.83	0.2033
-2.83	0.0023	-2.32	0.0102	-1.82	0.0344	-1.32	0.0934	-0.82	0.2061
-2.82	0.0024	-2.31	0.0104	-1.81	0.0351	-1.31	0.0951	-0.81	0.2090
-2.81	0.0025	-2.30	0.0107	-1.80	0.0359	-1.30	0.0968	-0.80	0.2119
-2.80	0.0026	-2.29	0.0110	-1.79	0.0367	-1.29	0.0985	-0.79	0.2148
-2.79	0.0026	-2.28	0.0113	-1.78	0.0375	-1.28	0.1003	-0.78	0.2177
-2.78	0.0027	-2.27	0.0116	-1.77	0.0384	-1.27	0.1020	-0.77	0.2206
-2.77	0.0028	-2.26	0.0119	-1.76	0.0392	-1.26	0.1038	-0.76	0.2236
-2.76	0.0029	-2.25	0.0122	-1.75	0.0401	-1.25	0.1056	-0.75	0.2266
-2.75	0.0030	-2.24	0.0125	-1.74	0.0409	-1.24	0.1075	-0.74	0.2296
-2.74	0.0031	-2.23	0.0129	-1.73	0.0418	-1.23	0.1093	-0.73	0.2327
-2.73	0.0032	-2.22	0.0132	-1.72	0.0427	-1.22	0.1112	-0.72	0.2358
-2.72	0.0033	-2.21	0.0136	-1.71	0.0436	-1.21	0.1131	-0.71	0.2389
-2.71	0.0034	-2.20	0.0139	-1.70	0.0446	-1.20	0.1151	-0.70	0.2420
-2.70	0.0035	-2.19	0.0143	-1.69	0.0455	-1.19	0.1170	-0.69	0.2451
-2.69	0.0036	-2.18	0.0146	-1.68	0.0465	-1.18	0.1190	-0.68	0.2483
-2.68	0.0037	-2.17	0.0150	-1.67	0.0475	-1.17	0.1210	-0.67	0.2514
-2.67	0.0038	-2.16	0.0154	-1.66	0.0485	-1.16	0.1230	-0.66	0.2546
-2.66	0.0039	-2.15	0.0158	-1.65	0.0495	-1.15	0.1251	-0.65	0.2578
-2.65	0.0040	-2.14	0.0162	-1.64	0.0505	-1.14	0.1271	-0.64	0.2611
-2.64	0.0041	-2.13	0.0166	-1.63	0.0516	-1.13	0.1292	-0.63	0.2643
-2.63	0.0043	-2.12	0.0170	-1.62	0.0526	-1.12	0.1314	-0.62	0.2676
-2.62	0.0044	-2.11	0.0174	-1.61	0.0537	-1.11	0.1335	-0.61	0.2709
-2.61	0.0045	-2.10	0.0179	-1.60	0.0548	-1.10	0.1357	-0.60	0.2749
-2.60	0.0047	-2.09	0.0183	-1.59	0.0559	-1.09	0.1379	-0.59	0.2776
-2.59	0.0048	-2.08	0.0188	-1.58	0.0571	-1.08	0.1401	-0.58	0.2810
-2.58	0.0049	-2.07	0.0192	-1.57	0.0582	-1.07	0.1423	-0.57	0.2843
-2.57	0.0051	-2.06	0.0197	-1.56	0.0594	-1.06	0.1446	-0.56	0.2877
-2.56	0.0052	-2.05	0.0202	-1.55	0.0606	-1.05	0.1469	-0.55	0.2912
-2.55	0.0054	-2.04	0.0207	-1.54	0.0618	-1.04	0.1492	-0.54	0.2946
-2.54	0.0055	-2.03	0.0212	-1.53	0.0630	-1.03	0.1515	-0.53	0.2981
-2.53	0.0057	-2.02	0.0217	-1.52	0.0643	-1.02	0.1539	-0.52	0.3015
-2.52	0.0059	-2.01	0.0222	-1.51	0.0655	-1.01	0.1562	-0.51	0.3050
-2.51	0.0060	-2.00	0.0228	-1.50	0.0668	-1.00	0.1587	-0.50	0.3085
-2.50	0.0062								

Tabelle A: z-Werte (z-Wert, Fläche/Wahrscheinlichkeit; einseitig; 1-α)

-0.49	0.3121	0.01	0.5040	0.51	0.6950	1.01	0.8438	1.51	0.9345
-0.48	0.3156	0.02	0.5080	0.52	0.6985	1.02	0.8461	1.52	0.9357
-0.47	0.3192	0.03	0. 5120	0.53	0.7019	1.03	0.8485	1.53	0.9370
-0.46	0.3228	0.04	0.5160	0.54	0.7054	1.04	0.8508	1.54	0.9382
-0.45	0.3264	0.05	0.5199	0.55	0.7088	1.05	0.8531	1.55	0.9394
-0.44	0.3300	0.06	0.5239	0.56	0.7123	1.06	0.8554	1.56	0.9406
-0.43	0.3336	0.07	0.5279	0.57	0.7157	1.07	0.8577	1.57	0.9418
-0.42	0.3372	0.08	0.5319	0.58	0.7190	1.08	0.8599	1.58	0.9429
-0.41	0.3409	0.09	0.5359	0.59	0.7224	1.09	0.8621	1.59	0.9441
-0.40	0.3446	0.10	0.5398	0.60	0.7257	1.10	0.8643	1.60	0.9452
-0.39	0.3483	0.11	0.5438	0.61	0.7291	1.11	0.8665	1.61	0.9463
-0.38	0.3520	0.12	0.5478	0.62	0.7324	1.12	0.8686	1.62	0.9474
-0.37	0.3557	0.13	0.5517	0.63	0.7357	1.13	0.8708	1.63	0.9484
-0.36	0.3594	0.14	0.5557	0.64	0.7389	1.14	0.8729	1.64	0.9495
-0.35	0.3632	0.15	0.5596	0.65	0.7422	1.15	0.8749	1.65	0.9505
-0.34	0.3669	0.16	0.5636	0.66	0.7454	1.16	0.8770	1.66	0.9515
-0.33	0.3707	0.17	0.5675	0.67	0.7486	1.17	0.8790	1.67	0.9525
-0.32	0.3745	0.18	0.5714	0.68	0.7517	1.18	0.8810	1.68	0.9535
-0.31	0.3783	0.19	0.5753	0.69	0.7549	1.19	0.8830	1.69	0.9545
-0.30	0.3821	0.20	0.5793	0.70	0.7580	1.20	0.8849	1.70	0.9554
-0.29	0.3859	0.21	0.5832	0.71	0.7611	1.21	0.8869	1.71	0.9564
-0.28	0.3897	0.22	0.5871	0.72	0.7642	1.22	0.8888	1.72	0.9573
-0.27	0.3936	0.23	0.5910	0.73	0.7673	1.23	0.8907	1.73	0.9582
-0.26	0.3974	0.24	0.5948	0.74	0.7704	1.24	0.8925	1.74	0.9591
-0.25	0.4013	0.25	0.5987	0.75	0.7734	1.25	0.8944	1.75	0.9599
-0.24	0.4052	0.26	0.6026	0.76	0.7764	1.26	0.8962	1.76	0.9608
-0.23	0.4090	0.27	0.6064	0.77	0.7794	1.27	0.8980	1.77	0.9616
-0.22	0.4129	0.28	0.6103	0.78	0.7823	1.28	0.8997	1.78	0.9625
-0.21	0.4168	0.29	0.6141	0.79	0.7852	1.29	0.9015	1.79	0.9633
-0.20	0.4207	0.30	0.6179	0.80	0.7881	1.30	0.9032	1.80	0.9641
-0.19	0.4247	0.31	0.6217	0.81	0.7910	1.31	0.9049	1.81	0.9649
-0.18	0.4286	0.32	0.6255	0.82	0.7939	1.32	0.9066	1.82	0.9656
-0.17	0.4325	0.33	0.6293	0.83	0.7967	1.33	0.9082	1.83	0.9664
-0.16	0.4364	0.34	0.6331	0.84	0.7995	1.34	0.9099	1.84	0.9671
-0.15	0.4404	0.35	0.6368	0.85	0.8023	1.35	0.9115	1.85	0.9678
-0.14	0.4443	0.36	0.6406	0.86	0.8054	1.36	0.9131	1.86	0.9686
-0.13	0.4483	0.37	0.6443	0.87	0.8078	1.37	0.9147	1.87	0.9693
-0.12	0.4522	0.38	0.6480	0.88	0.8106	1.38	0.9162	1.88	0.9699
-0.11	0.4562	0.39	0.6517	0.89	0.8133	1.39	0.9177	1.89	0.9706
-0.10	0.4602	0.40	0.6554	0.90	0.8159	1.40	0.9192	1.90	0.9713
-0.09	0.4641	0.41	0.6591	0.91	0.8186	1.41	0.9207	1.91	0.9719
-0.08	0.4681	0.42	0.6628	0.92	0.8212	1.42	0.9222	1.92	0.9726
-0.07	0.4721	0.43	0.6664	0.93	0.8238	1.43	0.9236	1.93	0.9732
-0.06	0.4761	0.44	0.6700	0.94	0.8264	1.44	0.9251	1.94	0.9738
-0.05	0.4801	0.45	0.6736	0.95	0.8289	1.45	0.9265	1.95	0.9744
-0.04	0.4840	0.46	0.6772	0.96	0.8315	1.46	0.9279	1.96	0.9750
-0.03	0.4880	0.47	0.6808	0.97	0.8340	1.47	0.9292	1.97	0.9756
-0.02	0.4920	0.48	0.6844	0.98	0.8365	1.48	0.9306	1.98	0.9761
-0.01	0.4960	0.49	0.6879	0.99	0.8389	1.49	0.9319	1.99	0.9767
0.00	0.5000	0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772

Tabelle A: z-Werte (z-Wert, Fläche/Wahrscheinlichkeit; einseitig; 1-α)

2.01	0.9778	2.51	0.9940
2.02	0.9783	2.52	0.9941
2.03	0.9788	2.53	0.9943
2.04	0.9793	2.54	0.9945
2.05	0.9798	2.55	0.9946
2.06	0.9803	2.56	0.9948
2.07	0.9808	2.57	0.9949
2.08	0.9812	2.58	0.9951
2.09	0.9817	2.59	0.9952
2.10	0.9821	2.60	0.9953
2.11	0.9826	2.61	0.9955
2.12	0.9830	2.62	0.9956
2.13	0.9834	2.63	0.9957
2.14	0.9838	2.64	0.9959
2.15	0.9842	2.65	0.9960
2.16	0.9846	2.66	0.9961
2.17	0.9850	2.67	0.9962
2.18	0.9854	2.68	0.9963
2.19	0.9857	2.69	0.9964
2.20	0.9861	2.70	0.9965
2.21	0.9864	2.71	0.9966
2.22	0.9868	2.72	0.9967
2.23	0.9871	2.73	0.9968
2.24	0.9875	2.74	0.9969
2.25	0.9878	2.75	0.9970
2.26	0.9881	2.76	0.9971
2.27	0.9884	2.77	0.9972
2.28	0.9887	2.78	0.9973
2.29	0.9890	2.79	0.9974
2.30	0.9893	2.80	0.9974
2.31	0.9896	2.81	0.9975
2.32	0.9898	2.82	0.9976
2.33	0.9901	2.83	0.9977
2.34	0.9904	2.84	0.9977
2.35	0.9906	2.85	0.9978
2.36	0.9909	2.86	0.9979
2.37	0.9911	2.87	0.9979
2.38	0.9913	2.88	0.9980
2.39	0.9916	2.89	0.9981
2.40	0.9918	2.90	0.9981
2.41	0.9920	2.91	0.9982
2.42	0.9922	2.92	0.9982
2.43	0.9925	2.93	0.9983
2.44	0.9907	2.94	0.9984
2.45	0.9929	2.95	0.9984
2.46	0.9931	2.96	0.9985
2.47	0.9932	2.97	0.9985
2.48	0.9934	2.98	0.9986
2.49	0.9936	2.99	0.9986
2.50	0.9938	3.00	0.9987

Tabelle B: F-Werte, einseitig, Zähler-df 1-10, Nenner-df 1-25

Nen-	Fläche	Zähler-df,									
ner-df,		1	2	3	4	5	6	7	8	9	10
1	0.95	161.00	200.00	216.00	225.00	230.00	234.00	237.00	239.00	241.00	242.00
	0.99	4052.00	4999.00	5403.00	5625.00	5764.00	5859.00	5928.00	5981.00	6022.00	6056.00
2	0.95	18.50	19.00	19.20	19.20	19.30	19.30	19.40	19.40	19.40	19.40
	0.99	98.50	99.00	99.20	99.20	99.30	99.30	99.40	99.40	99.40	99.40
3	0.95	10.10	9.55	9.28	9.12	9.10	8.94	8.89	8.85	8.81	8.79
	0.99	34.10	30.80	29.50	28.70	28.20	27.90	27.70	27.50	27.30	27.20
4	0.95	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
	0.99	21.20	18.00	16.70	16.00	15.50	15.20	15.00	14.80	14.70	14.50
5	0.95	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
	0.99	16.30	13.30	12.10	11.40	11.00	10.70	10.50	10.30	10.20	10.10
6	0.95	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
	0.99	13.70	10.90	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	0.95	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
	0.99	12.20	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	0.95	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
	0.99	11.30	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	0.95	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
	0.99	10.60	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	0.95	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
	0.99	10.00	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	0.95	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
	0.99	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	0.95	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
	0.99	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	0.95	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
	0.99	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	0.95	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
	0.99	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	0.95	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
	0.99	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	0.95	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
	0.99	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	0.95	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
	0.99	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	0.95	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
	0.99	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	0.95	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
	0.99	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	0.95	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
	0.99	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	0.95	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
	0.99	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	0.95	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
	0.99	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	0.95	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
	0.99	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	0.95	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
	0.99	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	0.95	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
	0.99	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13

Tabelle B: F-Werte, einseitig, Zähler-df 1-10, Nenner-df 26-∞

Nen-	Fläche	Zähler-df,									
ner-df,		1	2	3	4	5	6	7	8	9	10
26	0.95	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
	0.99	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	0.95	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
	0.99	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	0.95	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
	0.99	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	0.95	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
	0.99	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	0.95	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
	0.99	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
32	0.95	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14
	0.99	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93
34	0.95	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12
	0.99	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89
36	0.95	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11
	0.99	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95	2.86
38	0.95	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09
	0.99	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92	2.83
40	0.95	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
	0.99	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
42	0.95	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06
	0.99	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86	2.78
44	0.95	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05
	0.99	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84	2.75
46	0.95	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04
	0.99	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82	2.73
48	0.95	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03
	0.99	7.19	5.08	4.22	3.74	3.43	3.20	3.04	2.91	2.80	2.71
50	0.95	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03
	0.99	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70
55	0.95	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01
	0.99	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66
60	0.95	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
	0.99	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
65	0.95	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03	1.98
	0.99	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.80	2.69	2.61
70	0.95	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97
	0.99	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59
80	0.95	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95
	0.99	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55
100	0.95	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93
	0.99	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50
125	0.95	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.96	1.91
	0.99	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.66	2.55	2.47
150	0.95	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89
	0.99	6.81	4.75	3.91	3.45	3.14	2.92	2.76	2.63	2.53	2.44
175	0.95	3.90	3.05	2.66	2.42	2.27	2.15	2.06	1.99	1.93	1.89
	0.99	6.78	4.73	3.90	3.43	3.12	2.91	2.74	2.61	2.51	2.42
200	0.95	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88
	0.99	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41
∞	0.95	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83
	0.99	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Tabelle B: F-Werte, einseitig, Zähler-df 11-20, Nenner-df 1-25 (Fortsetzung)

Nen-	Fläche	Zähler-df,									
ner-df,		11	12	13	14	15	16	17	18	19	20
1	0.95	243.00	244.00	245.00	245.00	246.00	246.00	247.00	247.00	248.00	248.00
	0.99	6083.00	6106.00	6126.00	6143.00	6157.00	6170.00	6181.00	6192.00	6201.00	6209.00
2	0.95	19.40	19.40	19.40	19.40	19.40	19.40	19.40	19.40	19.40	19.40
	0.99	99.40	99.40	99.40	99.40	99.40	99.40	99.40	99.40	99.40	99.40
3	0.95	8.76	8.74	8.73	8.71	8.70	8.69	8.68	8.67	8.67	8.66
	0.99	27.10	27.10	27.00	26.90	26.90	26.80	26.80	26.80	26.70	26.70
4	0.95	5.94	5.91	5.89	5.87	5.86	5.84	5.83	5.82	5.81	5.80
	0.99	14.40	14.40	14.30	14.20	14.20	14.20	14.10	14.10	14.00	14.00
5	0.95	4.71	4.68	4.66	4.64	4.62	4.60	4.59	4.58	4.57	4.56
	0.99	9.96	9.89	9.82	9.77	9.72	9.68	9.64	9.61	9.58	9.55
6	0.95	4.03	4.00	3.98	3.96	3.94	3.92	3.91	3.90	3.88	3.87
	0.99	7.79	7.72	7.66	7.60	7.56	7.52	7.48	7.45	7.42	7.40
7	0.95	3.60	3.57	3.55	3.53	3.51	3.49	3.48	3.47	3.46	3.44
	0.99	6.54	6.47	6.41	6.36	6.31	6.28	6.24	6.21	6.18	6.16
8	0.95	3.31	3.28	3.26	3.24	3.22	3.20	3.19	3.17	3.16	3.15
	0.99	5.73	5.67	5.61	5.56	5.52	5.48	5.44	5.41	5.38	5.36
9	0.95	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96	2.95	2.94
	0.99	5.18	5.11	5.05	5.01	4.96	4.92	4.89	4.86	4.83	4.81
10	0.95	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80	2.79	2.77
	0.99	4.77	4.71	4.65	4.60	4.56	4.52	4.49	4.46	4.43	4.41
11	0.95	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67	2.66	2.65
	0.99	4.46	4.40	4.34	4.29	4.25	4.21	4.18	4.15	4.12	4.10
12	0.95	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57	2.56	2.54
	0.99	4.22	4.16	4.10	4.05	4.01	3.97	3.94	3.91	3.88	3.86
13	0.95	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48	2.47	2.46
	0.99	4.02	3.96	3.91	3.86	3.82	3.78	3.75	3.72	3.69	3.66
14	0.95	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41	2.40	2.39
	0.99	3.86	3.80	3.75	3.70	3.66	3.62	3.59	3.56	3.53	3.51
15	0.95	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35	2.34	2.33
	0.99	3.73	3.67	3.61	3.56	3.52	3.49	3.45	3.42	3.40	3.37
16	0.95	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.30	2.29	2.28
	0.99	3.62	3.55	3.50	3.45	3.41	3.37	3.34	3.31	3.28	3.26
17	0.95	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.26	2.24	2.23
	0.99	3.52	3.46	3.40	3.35	3.31	3.27	3.24	3.21	3.19	3.16
18	0.95	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.22	2.20	2.19
	0.99	3.43	3.37	3.32	3.27	3.23	3.19	3.16	3.13	3.10	3.08
19	0.95	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.18	2.17	2.16
	0.99	3.36	3.30	3.24	3.19	3.15	3.12	3.08	3.05	3.03	3.00
20	0.95	2.31	2.28	2.25	2.22	2.20	2.18	2.17	2.15	2.14	2.12
	0.99	3.29	3.23	3.18	3.13	3.09	3.05	3.02	2.99	2.96	2.94
21	0.95	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12	2.11	2.10
	0.99	3.24	3.17	3.12	3.07	3.03	2.99	2.96	2.93	2.90	2.88
22	0.95	2.26	2.23	2.20	2.17	2.15	2.13	2.11	2.10	2.08	2.07
	0.99	3.18	3.12	3.07	3.02	2.98	2.94	2.91	2.88	2.85	2.83
23	0.95	2.24	2.20	2.18	2.15	2.13	2.11	2.09	2.08	2.06	2.05
	0.99	3.14	3.07	3.02	2.97	2.93	2.89	2.86	2.83	2.80	2.78
24	0.95	2.21	2.18	2.15	2.13	2.11	2.09	2.07	2.05	2.04	2.03
	0.99	3.09	3.03	2.98	2.93	2.89	2.85	2.82	2.79	2.76	2.74
25	0.95	2.20	2.16	2.14	2.11	2.09	2.07	2.05	2.04	2.02	2.01
	0.99	3.06	2.99	2.94	2.89	2.85	2.81	2.78	2.75	2.72	2.70

Tabelle B: F-Werte, einseitig, Zähler-df 11-20, Nenner-df 26-∞ (Fortsetzung)

Nenner-df,	Fläche	Zähler-df, 11	12	13	14	15	16	17	18	19	20
26	0.95	2.18	2.15	2.12	2.09	2.07	2.05	2.03	2.02	2.00	1.99
	0.99	3.02	2.96	2.90	2.86	2.81	2.78	2.75	2.72	2.69	2.66
27	0.95	2.17	2.13	2.10	2.08	2.06	2.04	2.02	2.00	1.99	1.97
	0.99	2.99	2.93	2.87	2.82	2.78	2.75	2.71	2.68	2.66	2.63
28	0.95	2.15	2.12	2.09	2.06	2.04	2.02	2.00	1.99	1.97	1.96
	0.99	2.96	2.90	2.84	2.79	2.75	2.72	2.68	2.65	2.63	2.60
29	0.95	2.14	2.10	2.08	2.05	2.03	2.01	1.99	1.97	1.96	1.94
	0.99	2.93	2.87	2.81	2.77	2.73	2.69	2.66	2.63	2.60	2.57
30	0.95	2.13	2.09	2.06	2.04	2.01	1.99	1.98	1.96	1.95	1.93
	0.99	2.91	2.84	2.79	2.74	2.70	2.66	2.63	2.60	2.57	2.55
32	0.95	2.10	2.07	2.04	2.01	1.99	1.97	1.95	1.94	1.92	1.91
	0.99	2.86	2.80	2.74	2.70	2.65	2.62	2.58	2.55	2.53	2.50
34	0.95	2.08	2.05	2.02	1.99	1.97	1.95	1.93	1.92	1.90	1.89
	0.99	2.82	2.76	2.70	2.66	2.61	2.58	2.54	2.51	2.49	2.46
36	0.95	2.07	2.03	2.00	1.98	1.95	1.93	1.92	1.90	1.88	1.87
	0.99	2.79	2.72	2.67	2.62	2.58	2.54	2.51	2.48	2.45	2.43
38	0.95	2.05	2.02	1.99	1.96	1.94	1.92	1.90	1.88	1.87	1.85
	0.99	2.75	2.69	2.64	2.59	2.55	2.51	2.48	2.45	2.42	2.40
40	0.95	2.04	2.00	1.97	1.95	1.92	1.90	1.89	1.87	1.85	1.84
	0.99	2.73	2.66	2.61	2.56	2.52	2.48	2.45	2.42	2.39	2.37
42	0.95	2.03	1.99	1.96	1.94	1.91	1.89	1.87	1.86	1.84	1.83
	0.99	2.70	2.64	2.59	2.54	2.50	2.46	2.43	2.40	2.37	2.34
44	0.95	2.01	1.98	1.95	1.92	1.90	1.88	1.86	1.84	1.83	1.81
	0.99	2.68	2.62	2.56	2.52	2.47	2.44	2.40	2.37	2.35	2.32
46	0.95	2.00	1.97	1.94	1.91	1.89	1.87	1.85	1.83	1.82	1.80
	0.99	2.66	2.60	2.54	2.50	2.45	2.42	2.38	2.35	2.33	2.30
48	0.95	1.99	1.96	1.93	1.90	1.88	1.86	1.84	1.82	1.81	1.79
	0.99	2.64	2.58	2.53	2.48	2.44	2.40	2.37	2.33	2.31	2.28
50	0.95	1.99	1.95	1.92	1.89	1.87	1.85	1.83	1.81	1.80	1.78
	0.99	2.63	2.56	2.51	2.46	2.42	2.38	2.35	2.32	2.29	2.27
55	0.95	1.97	1.93	1.90	1.88	1.85	1.83	1.81	1.79	1.78	1.76
	0.99	2.59	2.53	2.47	2.42	2.38	2.34	2.31	2.28	2.25	2.23
60	0.95	1.95	1.92	1.89	1.86	1.84	1.82	1.80	1.78	1.76	1.75
	0.99	2.56	2.50	2.44	2.39	2.35	2.31	2.28	2.25	2.22	2.20
65	0.95	1.94	1.90	1.87	1.85	1.82	1.80	1.78	1.76	1.75	1.73
	0.99	2.53	2.47	2.42	2.37	2.33	2.29	2.26	2.23	2.20	2.17
70	0.95	1.93	1.89	1.86	1.84	1.81	1.79	1.77	1.75	1.74	1.72
	0.99	2.51	2.45	2.40	2.35	2.31	2.27	2.23	2.20	2.18	2.15
80	0.95	1.91	1.88	1.84	1.82	1.79	1.77	1.75	1.73	1.72	1.70
	0.99	2.48	2.42	2.36	2.31	2.27	2.23	2.20	2.17	2.14	2.12
100	0.95	1.89	1.85	1.82	1.79	1.77	1.75	1.73	1.71	1.69	1.68
	0.99	2.43	2.37	2.31	2.27	2.22	2.19	2.15	2.12	2.09	2.07
125	0.95	1.87	1.83	1.80	1.77	1.75	1.73	1.71	1.69	1.67	1.66
	0.99	2.39	2.33	2.28	2.23	2.19	2.15	2.11	2.08	2.05	2.03
150	0.95	1.85	1.82	1.79	1.76	1.73	1.71	1.69	1.67	1.66	1.64
	0.99	2.37	2.31	2.25	2.20	2.16	2.12	2.09	2.06	2.03	2.00
175	0.95	1.84	1.81	1.78	1.75	1.72	1.70	1.68	1.66	1.65	1.63
	0.99	2.35	2.29	2.23	2.19	2.14	2.10	2.07	2.04	2.01	1.98
200	0.95	1.84	1.80	1.77	1.74	1.72	1.69	1.67	1.66	1.64	1.62
	0.99	2.34	2.27	2.22	2.17	2.13	2.09	2.06	2.03	2.00	1.97
∞	0.95	1.79	1.75	1.72	1.69	1.67	1.64	1.62	1.60	1.59	1.57
	0.99	2.25	2.18	2.13	2.08	2.04	2.00	1.97	1.93	1.90	1.88

Tabelle B: F-Werte, einseitig, Zähler-df 17-50, Nenner-df 1-25 (Fortsetzung)

Nen-	Fläche	Zähler-df,									
ner-df,		25	30	40	50	75	100	150	200	500	∞
1	0.95	249.00	250.00	251.00	252.00	253.00	253.00	253.00	254.00	254.00	254.00
	0.99	6240.00	6261.00	6287.00	6303.00	6324.00	6334.00	6345.00	6350.00	6360.00	6366.00
2	0.95	19.50	19.50	19.50	19.50	19.50	19.50	19.50	19.50	19.50	19.50
	0.99	99.50	99.50	99.50	99.50	99.50	99.50	99.50	99.50	99.50	99.50
3	0.95	8.63	8.62	8.59	8.58	8.56	8.55	8.54	8.54	8.53	8.53
	0.99	26.60	26.50	26.40	26.40	26.30	26.20	26.20	26.20	26.10	26.10
4	0.95	5.77	5.75	5.72	5.70	5.68	5.66	5.65	5.65	5.63	5.63
	0.99	13.90	13.80	13.70	13.70	13.60	13.60	13.50	13.50	13.50	13.50
5	0.95	4.52	4.50	4.46	4.44	4.42	4.41	4.39	4.39	4.36	4.36
	0.99	9.45	9.38	9.29	9.24	9.17	9.13	9.09	9.08	9.02	9.02
6	0.95	3.83	3.81	3.77	3.75	3.73	3.71	3.70	3.69	3.68	3.67
	0.99	7.30	7.23	7.14	7.09	7.02	6.99	6.95	6.93	6.90	6.88
7	0.95	3.40	3.38	3.34	3.32	3.29	3.27	3.26	3.25	3.24	3.23
	0.99	6.06	5.99	5.91	5.86	5.79	5.75	5.72	5.70	5.67	5.65
8	0.95	3.11	3.08	3.04	3.02	2.99	2.97	2.96	2.95	2.94	2.93
	0.99	5.26	5.20	5.12	5.07	5.00	4.96	4.93	4.91	4.88	4.86
9	0.95	2.89	2.86	2.83	2.80	2.77	2.76	2.74	2.73	2.72	2.71
	0.99	4.71	4.65	4.57	4.52	4.45	4.42	4.38	4.36	4.33	4.31
10	0.95	2.73	2.70	2.66	2.64	2.60	2.59	2.57	2.56	2.55	2.54
	0.99	4.31	4.25	4.17	4.12	4.05	4.01	3.98	3.96	3.93	3.91
11	0.95	2.60	2.57	2.53	2.51	2.47	2.46	2.44	2.43	2.42	2.40
	0.99	4.01	3.94	3.86	3.81	3.74	3.71	3.67	3.66	3.62	3.60
12	0.95	2.50	2.47	2.43	2.40	2.37	2.35	2.33	2.32	2.31	2.30
	0.99	3.76	3.70	3.62	3.57	3.50	3.47	3.43	3.41	3.38	3.36
13	0.95	2.41	2.38	2.34	2.31	2.28	2.26	2.24	2.23	2.22	2.21
	0.99	3.57	3.51	3.43	3.38	3.31	3.27	3.24	3.22	3.19	3.17
14	0.95	2.34	2.31	2.27	2.24	2.21	2.19	2.17	2.16	2.14	2.13
	0.99	3.41	3.35	3.27	3.22	3.15	3.11	3.08	3.06	3.03	3.00
15	0.95	2.28	2.25	2.20	2.18	2.14	2.12	2.10	2.10	2.08	2.07
	0.99	3.28	3.21	3.13	3.08	3.01	2.98	2.94	2.92	2.89	2.87
16	0.95	2.23	2.19	2.15	2.12	2.09	2.07	2.05	2.04	2.02	2.01
	0.99	3.16	3.10	3.02	2.97	2.90	2.86	2.83	2.81	2.78	2.75
17	0.95	2.18	2.15	2.10	2.08	2.04	2.02	2.00	1.99	1.97	1.96
	0.99	3.07	3.00	2.92	2.87	2.80	2.76	2.73	2.71	2.68	2.65
18	0.95	2.14	2.11	2.06	2.04	2.00	1.98	1.96	1.95	1.93	1.92
	0.99	2.98	2.92	2.84	2.78	2.71	2.68	2.64	2.62	2.59	2.57
19	0.95	2.11	2.07	2.03	2.00	1.96	1.94	1.92	1.91	1.89	1.88
	0.99	2.91	2.84	2.76	2.71	2.64	2.60	2.57	2.55	2.51	2.49
20	0.95	2.07	2.04	1.99	1.97	1.93	1.91	1.89	1.88	1.86	1.84
	0.99	2.84	2.78	2.69	2.64	2.57	2.54	2.50	2.48	2.44	2.42
21	0.95	2.05	2.01	1.96	1.94	1.90	1.88	1.86	1.84	1.83	1.81
	0.99	2.79	2.72	2.64	2.58	2.51	2.48	2.44	2.42	2.38	2.36
22	0.95	2.02	1.98	1.94	1.91	1.87	1.85	1.83	1.82	1.80	1.78
	0.99	2.73	2.67	2.58	2.53	2.46	2.42	2.38	2.36	2.33	2.31
23	0.95	2.00	1.96	1.91	1.88	1.84	1.82	1.80	1.79	1.77	1.76
	0.99	2.69	2.62	2.54	2.48	2.41	2.37	2.34	2.32	2.28	2.26
24	0.95	1.97	1.94	1.89	1.86	1.82	1.80	1.78	1.77	1.75	1.73
	0.99	2.64	2.58	2.49	2.44	2.37	2.33	2.29	2.27	2.24	2.21
25	0.95	1.96	1.92	1.87	1.84	1.80	1.78	1.76	1.75	1.73	1.71
	0.99	2.60	2.54	2.45	2.40	2.33	2.29	2.25	2.23	2.19	2.17

Tabelle B: F-Werte, einseitig, Zähler-df 25-∞, Nenner-df 26-∞ (Fortsetzung)

Nenner-df,	Fläche	Zähler-df, 25	30	40	50	75	100	150	200	500	∞
26	0.95	1.94	1.90	1.85	1.82	1.78	1.76	1.74	1.73	1.71	1.69
	0.99	2.57	2.50	2.42	2.36	2.29	2.25	2.21	2.19	2.16	2.13
27	0.95	1.92	1.88	1.84	1.81	1.75	1.73	1.70	1.69	1.67	1.65
	0.99	2.54	2.47	2.38	2.33	2.23	2.19	2.15	2.13	2.09	2.06
28	0.95	1.91	1.87	1.82	1.79	1.73	1.71	1.69	1.67	1.65	1.64
	0.99	2.51	2.44	2.35	2.30	2.20	2.16	2.12	2.10	2.06	2.03
29	0.95	1.89	1.85	1.81	1.77	1.72	1.70	1.67	1.66	1.64	1.62
	0.99	2.48	2.41	2.33	2.27	2.17	2.13	2.09	2.07	2.03	2.01
30	0.95	1.88	1.84	1.79	1.76	1.69	1.67	1.64	1.63	1.61	1.59
	0.99	2.45	2.39	2.30	2.25	2.12	2.08	2.04	2.02	1.98	1.96
32	0.95	1.85	1.82	1.77	1.74	1.67	1.65	1.62	1.61	1.59	1.57
	0.99	2.41	2.34	2.25	2.20	2.08	2.04	2.00	1.98	1.94	1.91
34	0.95	1.83	1.80	1.75	1.71	1.65	1.62	1.60	1.59	1.56	1.55
	0.99	2.37	2.30	2.21	2.16	2.04	2.00	1.96	1.94	1.90	1.87
36	0.95	1.81	1.78	1.73	1.69	1.63	1.61	1.58	1.57	1.54	1.53
	0.99	2.33	2.26	2.18	2.12	2.01	1.97	1.93	1.90	1.86	1.84
38	0.95	1.80	1.76	1.71	1.68	1.61	1.59	1.56	1.55	1.53	1.51
	0.99	2.30	2.23	2.14	2.09	1.98	1.94	1.90	1.87	1.83	1.80
40	0.95	1.78	1.74	1.69	1.66	1.60	1.57	1.55	1.53	1.51	1.49
	0.99	2.27	2.20	2.11	2.06	1.95	1.91	1.87	1.85	1.80	1.78
42	0.95	1.77	1.73	1.68	1.65	1.59	1.56	1.56	1.52	1.49	1.48
	0.99	2.25	2.18	2.09	2.03	1.93	1.89	1.84	1.82	1.78	1.75
44	0.95	1.76	1.72	1.67	1.63	1.57	1.55	1.52	1.51	1.48	1.46
	0.99	2.22	2.15	2.07	2.01	1.91	1.86	1.82	1.80	1.76	1.73
46	0.95	1.75	1.71	1.65	1.62	1.56	1.54	1.51	1.49	1.47	1.45
	0.99	2.20	2.13	2.04	1.99	1.89	1.84	1.80	1.78	1.73	1.70
48	0.95	1.74	1.70	1.64	1.61	1.55	1.52	1.50	1.48	1.46	1.44
	0.99	2.18	2.12	2.02	1.97	1.87	1.82	1.78	1.76	1.71	1.68
50	0.95	1.73	1.69	1.63	1.60	1.53	1.50	1.47	1.46	1.43	1.41
	0.99	2.17	2.10	2.01	1.95	1.83	1.78	1.74	1.71	1.67	1.64
55	0.95	1.71	1.67	1.61	1.58	1.51	1.48	1.45	1.44	1.41	1.39
	0.99	2.13	2.06	1.97	1.91	1.79	1.75	1.70	1.68	1.63	1.60
60	0.95	1.69	1.65	1.59	1.56	1.49	1.46	1.44	1.42	1.39	1.37
	0.99	2.10	2.03	1.94	1.88	1.77	1.72	1.67	1.65	1.60	1.57
65	0.95	1.68	1.63	1.58	1.54	1.48	1.45	1.42	1.40	1.37	1.35
	0.99	2.07	2.00	1.91	1.85	1.74	1.70	1.65	1.62	1.57	1.54
70	0.95	1.66	1.62	1.57	1.53	1.45	1.43	1.39	1.38	1.35	1.32
	0.99	2.05	1.98	1.89	1.83	1.70	1.65	1.61	1.58	1.53	1.49
80	0.95	1.64	1.60	1.54	1.51	1.42	1.39	1.36	1.34	1.31	1.28
	0.99	2.01	1.94	1.85	1.79	1.65	1.60	1.55	1.52	1.47	1.43
100	0.95	1.62	1.57	1.52	1.48	1.40	1.36	1.33	1.31	1.27	1.25
	0.99	1.97	1.89	1.80	1.74	1.60	1.55	1.50	1.47	1.41	1.37
125	0.95	1.59	1.55	1.49	1.45	1.38	1.34	1.31	1.29	1.25	1.22
	0.99	1.93	1.85	1.46	1.69	1.57	1.52	1.46	1.43	1.38	1.33
150	0.95	1.58	1.54	1.48	1.44	1.36	1.33	1.29	1.27	1.23	1.20
	0.99	1.90	1.83	1.73	1.66	1.55	1.50	1.44	1.41	1.35	1.30
175	0.95	1.57	1.52	1.46	1.42	1.35	1.32	1.28	1.26	1.22	1.19
	0.99	1.88	1.81	1.71	1.64	1.53	1.48	1.42	1.39	1.33	1.28
200	0.95	1.56	1.52	1.46	1.41	1.28	1.24	1.20	1.17	1.11	1.00
	0.99	1.87	1.79	1.69	1.63	1.42	1.36	1.29	1.25	1.15	1.00
∞	0.95	1.51	1.46	1.39	1.35						
	0.99	1.77	1.70	1.59	1.52						

Tabelle C: Kritische t-Werte (einseitig, 1-α)

			Fläche		
df	0.90	0.95	0.975	0.99	0.995
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617
z	1,282	1,645	1,960	2,326	2,576

Tabelle D: Kritische χ^2-Werte (zweiseitig, 1-α)

			Fläche		
df	0.90	0.95	0.975	0.99	0.995
1	2,706	3,841	5,024	6,635	7,879
2	4,605	5,991	7,378	9,210	10,597
3	6,251	7,815	9,348	11,345	12,838
4	7,779	9,488	11,143	13,277	14,860
5	9,236	11,071	12,833	15,086	16,750
6	10,645	12,592	14,449	16,812	18,548
7	12,017	14,067	16,013	18,475	20,278
8	13,362	15,507	17,535	20,090	21,955
9	14,684	16,919	19,023	21,666	23,589
10	15,987	18,307	20,483	23,209	25,188
11	17,275	19,675	21,920	24,725	26,757
12	18,549	21,026	23,337	26,217	28,300
13	19,812	22,362	24,736	27,688	29,819
14	21,064	23,685	26,119	29,141	31,319
15	22,307	24,996	27,488	30,578	32,801
16	23,542	26,296	28,845	32,000	31,267
17	24,769	27,587	30,191	33,409	35,719
18	25,989	28,869	31,526	34,805	37,156
19	27,204	30,144	32,852	36,191	38,582
20	28,412	31,410	34,170	37,566	39,997
21	29,615	32,671	35,479	38,932	41,401
22	30,813	33,924	36,781	40,289	42,796
23	32,007	35,173	38,076	41,638	44,181
24	33,196	36,415	39,364	42,980	45,559
25	34,382	37,653	40,647	44,314	46,928
26	35,563	38,885	41,923	45,642	48,290
27	36,741	40,113	43,194	46,963	49,645
28	37,916	41,337	44,461	48,278	50,993
29	39,088	42,557	45,722	49,588	52,336
30	40,256	43,773	46,979	50,892	53,672
40	51,805	55,759	59,342	63,691	66,766
50	63,167	67,505	71,420	76,154	79,490
60	74,397	79,082	83,298	88,379	91,952
70	85,527	90,531	95,023	100,425	104,215
80	96,578	101,879	106,629	112,329	116,321
90	107,565	113,145	118,136	124,116	128,299
100	118,498	124,342	129,561	135,807	140,169
z	1,282	1,645	1,960	2,326	2,576

Tabelle E: Kritische U-Werte, einseitige Wahrscheinlichkeiten

	n2 = 3			n2 = 4			
	n1			n1			
U	1	2	3	1	2	3	4
0	0,250	0,100	0,050	0,200	0,067	0,028	0,014
1	0,500	0,200	0,100	0,400	0,133	0,057	0,029
2	0,750	0,400	0,200	0,600	0,267	0,114	0,057
3		0,600	0,350		0,400	0,200	0,100
4			0,500		0,600	0,314	0,171
5			0,650			0,429	0,243
6						0,571	0,343
7							0,443
8							0,557

	n2=5					n2=6					
	n1					n1					
U	1	2	3	4	5	1	2	3	4	5	6
0	0,167	0,047	0,018	0,008	0,004	0,143	0,036	0,012	0,005	0,002	0,001
1	0,333	0,095	0,036	0,016	0,008	0,286	0,071	0,024	0,010	0,004	0,002
2	0,500	0,190	0,071	0,032	0,016	0,428	0,143	0,048	0,019	0,009	0,004
3	0,667	0,286	0,125	0,056	0,028	0,571	0,214	0,083	0,033	0,015	0,008
4		0,429	0,196	0,095	0,048		0,21	0,131	0,057	0,026	0,013
5		0,571	0,286	0,143	0,075		0,429	0,190	0,086	0,041	0,021
6			0,393	0,206	0,111		0,571	0,274	0,129	0,063	0,032
7			0,500	0,278	0,155			0,357	0,176	0,089	0,047
8			0,607	0,365	0,210			0,452	0,238	0,123	0,066
9				0,452	0,274			0,548	0,305	0,165	0,090
10				0,548	0,345				0,381	0,214	0,120
11					0,421				0,457	0,268	0,155
12					0,500				0,545	0,331	0,197
13					0,579					0,396	0,242
14										0,465	0,294
15										0,535	0,350
16											0,409
17											0,468
18											0,531

	n2 = 7						
	n1						
U	1	2	3	4	5	6	7
0	0,125	0,028	0,008	0,003	0,001	0,001	0,000
1	0,250	0,056	0,017	0,006	0,003	0,001	0,001
2	0,375	0,111	0,033	0,012	0,005	0,002	0,001
3	0,500	0,167	0,058	0,021	0,009	0,004	0,002
4	0,625	0,250	0,092	0,036	0,015	0,007	0,003
5		0,333	0,133	0,055	0,024	0,011	0,006
6		0,444	0,192	0,082	0,037	0,017	0,009
7		0,556	0,258	0,115	0,053	0,026	0,013
8			0,333	0,158	0,074	0,037	0,019
9			0,417	0,206	0,101	0,051	0,027
10			0,500	0,264	0,134	0,069	0,036
11			0,583	0,324	0,172	0,090	0,049
12				0,394	0,216	0,117	0,064
13				0,464	0,265	0,147	0,082
14				0,538	0,319	0,183	0,104
15					0,378	0,223	0,130
16					0,438	0,267	0,159
17					0,500	0,314	0,191
18					0,562	0,365	0,228
19						0,418	0,267
20						0,473	0,310
21						0,527	0,355
22							0,402
23							0,451
24							0,500
25							0,549

Tabelle E: Kritische U-Werte, einseitige Wahrscheinlichkeiten (Fortsetzung)

	n2 = 8									
	n1									
U	1	2	3	4	5	6	7	8	t	Normal
0	0,111	0,022	0,006	0,002	0,001	0,000	0,000	0,000	3,308	0,001
1	0,222	0,044	0,012	0,004	0,002	0,001	0,000	0,000	3,203	0,001
2	0,333	0,089	0,024	0,008	0,003	0,001	0,001	0,000	3,098	0,001
3	0,444	0,133	0,042	0,014	0,005	0,002	0,001	0,001	2,993	0,001
4	0,556	0,200	0,067	0,024	0,009	0,004	0,002	0,001	2,888	0,002
5		0,267	0,097	0,036	0,015	0,006	0,003	0,001	2,783	0,003
6		0,356	0,139	0,055	0,023	0,010	0,005	0,002	2,678	0,004
7		0,444	0,188	0,077	0,033	0,015	0,007	0,003	2,573	0,005
8		0,556	0,248	0,107	0,047	0,021	0,010	0,005	2,468	0,007
9			0,315	0,141	0,064	0,030	0,014	0,007	2,363	0,009
10			0,387	0,184	0,085	0,041	0,020	0,010	2,258	0,012
11			0,461	0,230	0,111	0,054	0,027	0,014	2,153	0,016
12			0,539	0,285	0,142	0,071	0,036	0,019	2,048	0,020
13				0,341	0,177	0,091	0,047	0,025	1,943	0,026
14				0,404	0,217	0,114	0,060	0,032	1,838	0,033
15				0,467	0,262	0,141	0,076	0,041	1,733	0,041
16				0,533	0,311	0,172	0,095	0,052	1,628	0,052
17					0,362	0,207	0,116	0,065	1,523	0,064
18					0,416	0,245	0,140	0,080	1,418	0,078
19					0,472	0,286	0,168	0,097	1,313	0,094
20					0,528	0,331	0,198	0,117	1,208	0,113
21						0,377	0,232	0,139	1,102	0,135
22						0,426	0,268	0,164	0,998	0,159
23						0,475	0,306	0,191	0,893	0,185
24						0,525	0,347	0,221	0,788	0,215
25							0,389	0,253	0,683	0,247
26							0,433	0,287	0,578	0,282
27							0,478	0,323	0,473	0,318
28							0,522	0,360	0,368	0,356
29								0,399	0,263	0,396
30								0,439	0,158	0,437
31								0,480	0,052	0,481
32								0,520		

Lösungen

Kapitel 2

001 a) $\Omega = \{(K,K,K),(K,K,Z),(K,Z,K),(K,Z,Z),(Z,K,K),(Z,K,Z),(Z,Z,K),(Z,Z,Z)\}$

b) Ω enthält 8 Elemente.

c) $\mathrm{A} = \{(K,K,K),(K,K,Z),(K,Z,K),(K,Z,Z)\}$,
$\mathrm{B} = \{(\mathrm{K},\mathrm{K},\mathrm{Z}),(\mathrm{K},\mathrm{Z},\mathrm{Z}),(\mathrm{Z},\mathrm{K},\mathrm{Z}),(\mathrm{Z},\mathrm{Z},\mathrm{Z})\}$

d) $\mathrm{A} \cap \mathrm{B} =$ „Beim 1. Wurf erscheint Kopf und beim 3. Wurf erscheint Zahl“;
$\mathrm{A} \cap \mathrm{B} = \{(\mathrm{K},\mathrm{K},\mathrm{Z}),(\mathrm{K},\mathrm{Z},\mathrm{Z})\}$

$\mathrm{A} \cup \mathrm{B} =$ „Beim 1. Wurf erscheint Kopf oder beim 3. Wurf erscheint Zahl“;
$\mathrm{A} \cup \mathrm{B} = \{(K,K,K),(K,K,Z),(K,Z,K),(K,Z,Z),(Z,K,Z),(Z,Z,Z)\}$

$\overline{\mathrm{A}} =$ „Beim 1. Wurf erscheint kein Kopf“; $\overline{\mathrm{A}} = \{(Z,K,K),(Z,K,Z),(Z,Z,K),(Z,Z,Z)\}$

$\mathrm{A} \cap \overline{\mathrm{B}} =$ „Beim 1. Wurf erscheint Kopf und beim 3. Wurf erscheint nicht Zahl“;
$\mathrm{A} \cap \overline{\mathrm{B}} = \{(\mathrm{K},\mathrm{K},\mathrm{K}),(\mathrm{K},\mathrm{Z},\mathrm{K})\}$

$\overline{\mathrm{A}} \cap \overline{\mathrm{B}} =$ „Beim 1. Wurf erscheint nicht Kopf und beim 3. Wurf erscheint nicht Zahl“;
$\overline{\mathrm{A}} \cap \overline{\mathrm{B}} = \{(\mathrm{Z},\mathrm{K},\mathrm{K}),(\mathrm{Z},\mathrm{Z},\mathrm{K})\}$

e) $\mathrm{A} \cup \mathrm{B}$ und $\overline{\mathrm{A}} \cap \overline{\mathrm{B}}$ sind Gegenereignisse voneinander (vgl. De Morgan’sches Gesetz aus der Mengenlehre)

002 a) $\Omega = \{(1,1),(1,2),(1,3),\ldots,(6,5),(6,6)\}$

b) $|\Omega| = 36$

c) $\mathrm{A} = \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}$

$\mathrm{B} = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$

$\mathrm{C} = \{(1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(2,4),(3,1),(3,2),(3,3),(3,4),$
$(4,1),(4,2),(4,3),(4,4)\}$

d) Es gilt:

$$\mathrm{P}(\mathrm{A}) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{|\mathrm{A}|}{|\Omega|} = \frac{9}{36};$$

$$\mathrm{P}(\mathrm{B}) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{|\mathrm{B}|}{|\Omega|} = \frac{5}{36};$$

$$\mathrm{P}(\mathrm{C}) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{|\mathrm{C}|}{|\Omega|} = \frac{16}{36};$$

$$\mathrm{P}(\overline{\mathrm{A}}) = 1 - \mathrm{P}(\mathrm{A}) = 1 - \frac{9}{36} = \frac{27}{36};$$

$$P(A \cap B) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{|A \cap B|}{|\Omega|} = \frac{3}{36},$$

da $A \cap B = \{(2,6),(4,4),(6,2)\}$;

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{9}{36} + \frac{5}{36} - \frac{3}{36} = \frac{11}{36};$$

$$P(A \cap B \cap C) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{|A \cap B \cap C|}{|\Omega|} = \frac{1}{36}$$

da $A \cap B \cap C = \{(4,4)\}$;

$$P(A \backslash C) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{|A \backslash C|}{|\Omega|} = \frac{5}{36}$$

da $A \backslash C = \{(2,6),(4,6),(6,2),(6,4),(6,6)\}$.

003 a) Da es an der Uni insgesamt gleichviel Studentinnen wie Studenten gibt, ist die Wahrscheinlichkeit P(weiblich)=0.5.

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

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(Person ist weiblich, studiert Psychologie und besitzt gute stat. Kenntnisse) = 0.066×0.4 = 0.0264.

004 P(verstehen)=0.8 P(nicht verstehen)=q=0.2 N=25

a) alle → k=25

Binomialformel:

$$P(k=25) = \binom{n}{k} \times p^k \times q^{n-k} = \frac{n!}{k! \times (n-k)!} \times p^k \times q^{n-k}$$
$$P(k=25) = \frac{25!}{25! \times (25-25)!} \times 0.8^{25} \times 0.2^{25-25}$$
$$P(k=25) = \frac{25 \times 24 \times 23 \times \ldots \times 1}{25 \times 24 \times 23 \times \ldots \times 1 \times 0!} \times 0.0038 \times 1$$
$$P(k=25) = 0.0038$$

Die Wahrscheinlichkeit, dass alle 25 Personen den Unterricht verstehen, beträgt 0.38%.

b) keine → k=0

Binomialformel:

$$P(k=0)=\binom{n}{k}\times p^k\times q^{n-k}=\frac{n!}{k!\times(n-k)!}\times p^k\times q^{n-k}$$

$$P(k=0)=\frac{25!}{0!\times(25-0)!}\times 0.8^0\times 0.2^{25-0}$$

$$P(k=0)=\frac{25\times 24\times 23\times\ldots\times 1}{1\times 25\times 24\times 23\times\ldots\times 1}\times 1\times 0.00000000000000000033554432$$

$$P(k=0)=0.00000000000000000033554432$$

Die Wahrscheinlichkeit dafür, dass keine der 25 Personen den Unterricht versteht, ist verschwindend gering (nahe Null).

005 P(richtiger Ort)=0.3 → P(falscher Ort)=q=0.7
P(richtige Zeit)=0.5 → P(falsche Zeit)=q=0.5

a) P(richtige Zeit UND falscher Ort)=0.5×0.7=0.35

b) P(falsche Zeit UND richtiger Ort)=0.5×0.3=0.15

c) Binomialformel
n=5, k=mindestens 3 d.h. 3, 4, 5
P(richtige Zeit UND richtiger Ort)=0.5×0.3=0.15 → q=0.85

zuerst für k=3

$$P(k=3)=\binom{5}{3}\times 0.15^3\times 0.85^{5-3}$$

$$P(k=3)=\frac{5!}{3!\times 2!}\times 0.003375\times 0.7225=10\times 0.0024384375=0.0244$$

jetzt für k=4

$$P(k=4)=\binom{5}{4}\times 0.15^4\times 0.85^{5-4}$$

$$P(k=4)=\frac{5!}{4!\times 1!}\times 0.00050625\times 0.85=5\times 0.0004303125=0.00215$$

jetzt für k=5

$$P(k=5)=\binom{5}{5}\times 0.15^5\times 0.85^{5-5}$$

$$P(k=5)=\frac{5!}{5!\times 0!}\times 0.000076\times 1=0.000076$$

Und jetzt alles zusammen:
P(k=mindestens 3)=0.0244+0.00215+0.000076=0.026626

006 a) P=1/5=0.2

b) P=4/5=0.8

c) P=0.2*0.2*0.8*0.8=0.0256

d) P=0.2*0.8*0.2*0.8=0.0256

e) Binomialverteilung
n=4, k=2, p=0.2, q=0.8

$$P(k=2)=\binom{4}{2}\times 0.2^2\times 0.80^{4-2}$$

$$P(k=2)=\frac{4!}{2!\times 2!}\times 0.0256=6\times 0.0256=0.1536$$

007 a) Binomialverteilung
n=5, k=2, p=1/20=0.05, q=0.95

$$P(k=2)=\binom{5}{2}\times 0.05^2\times 0.95^{5-2}$$

$$P(k=2)=\frac{5!}{2!\times 3!}\times 0.0025\times 0.8574=10\times 0.00214=0.0214$$

b) Binomialverteilung
n=5, k=höchstens 2 d.h. k=0, 1, 2, p=0.05, q=0.95

zuerst k=0

$$P(k=0)=\binom{5}{0}\times 0.05^0\times 0.95^{5-0}$$

$$P(k=0)=\frac{5!}{0!\times 5!}\times 1\times 0.77378=0.77378$$

jetzt für k=1

$$P(k=1)=\binom{5}{1}\times 0.05^1\times 0.95^{5-1}$$

$$P(k=1)=\frac{5!}{1!\times 4!}\times 0.05\times 0.81451=5\times 0.04073=0.20365$$

- Und jetzt alles zusammen:
  P(k höchstens 2)=0.77378+0.20365+0.0214=0.99883

008 Die Kennzeichen der Standardnormalverteilung sind
- ein Mittelwert von Null und
- eine Standardabweichung von Eins.

Kapitel 3

009 Entscheidet man sich auf Grund der Stcihprobenergebnisse für die Nullhypothese H0, obwohl in der Grundgesamtheit die Alternativhypothese H1 gilt, so begeht man einen Fehler, den sogenannten Fehler 2. Art (β-Fehler).

010 Entscheidet man sich auf Grund der Stcihprobenergebnisse für die Alternativhypothese H1, obwohl in der Grundgesamtheit die Nullhypothese H0 gilt, so begeht man einen Fehler, den sogenannten Fehler 1. Art (α-Fehler).

011 Unter einer Stichprobe versteht man eine Auswahl von Objekten (wie z.B. Personen, Haushalte, Unternehmen etc.) aus einer Grundgesamtheit. Die Auswahl der Objekte kann dabei durch den Zufall gesteuert sein oder nach anderen Auswahlkriterien erfolgen.

012 Es wird verglichen, wie Personen vor und nach einer Maßnahme (z.B. Weiterbildung) in einem Test abschneiden

013

Nominalskalenniveau	Geschlecht
Ordinalskalenniveau	Position in einer Firma
Intervallskalenniveau	Intelligenzquotient
Verhältnisskalenniveau	Körpergröße

014 Es liegt eine Ordinalskala vor.

015 Nominalskalenniveau

Kapitel 4

016 arithmetisches Mittel (Mittelwert), Median, Modalwert
Mittelwert = Durchschnitt
Median teilt eine geordnete Liste in zwei gleich große Hälften
Modalwert = häufigster Wert

017 Z.B. Minimum, Maximum, Range, Varianz, Standardabweichung

018 Die Variable "Körpergröße" ist verhältnisskaliert. Das arithmetische Mittel, der Median und der Modus machen somit Sinn. Die Variable "Zufriedenheit" ist offensichtlich ordinalskaliert. Daher ist es nicht sinnvoll das arithmetische Mittel zu berechnen (obwohl es rechnerisch möglich wäre)! Bei ordinalskalierten Variablen greift man bei den Lageparametern auf Median und Modus zurück. Die Variable "Geschlecht" ist nominalskaliert. Daher macht nur die Angabe des Modus Sinn. Der Median kann bei nominalskalierten Merkmalen übrigens nicht angegeben werden, da sich die Beobachtungswerte nicht in eine Rangfolge bringen lassen. Könnte man dies, dann wäre das Merkmal nicht nominalskaliert.

An dieser Stelle sei noch bemerkt, dass die Angabe der Streuungsparameter Varianz und Standardabweichung entsprechend auch nur bei mindestens intervallskalierten Variablen Sinn macht.

019 Das arithmetische Mittel der Beobachtungswerte erhält man, indem man alle Zahlen zusammenzählt und das Ergebnis durch 10 teilt. Das arithmetische Mittel beträgt somit 175 [cm]. Sortiert man die Werte der Größe nach, erhält man die folgende geordnete Rangreihe: 160, 164, 164, 170, 175, 176, 178, 182, 189, 192. Da man eine gerade Anzahl an Beobachtungswerten hat, bestimmt sich der Median als arithmetisches Mittel des 5. und des 6. ten Wertes. Der Median ist somit 175,5 [cm]. Der Modus ist diejenige Ausprägung, die am häufigsten auftritt. In diesem Falle liegt der Modus bei 164 [cm].

020 Als Rechenfehler; ein Quadrat kann nicht negativ sein!

021 Beide Variablen sind nominalskaliert und liegen in jeweils zwei Ausprägungen vor. Der geeignete Korrelationskoeffizient wäre dann die Phi-Korrelation. Dazu sollte erst einmal eine Vierfeldertafel erstellt werden.

	Geschlecht		
Abteilung	m	w	
Personal	1	3	4
EDV	3	1	4
	4	4	8

$$Phi_{emp} = \frac{b \times c - a \times d}{\sqrt{(a+c) \times (b+d) \times (a+b) \times (c+d)}}$$

$$Phi_{emp} = \frac{9-1}{\sqrt{4 \times 4 \times 4 \times 4}} = \frac{8}{16} = 0.5$$

Vierfeldertafel mit maximiertem Zusammenhang

	Geschlecht		
Abteilung	m	w	
Personal	0	4	4
EDV	4	0	4
	4	4	8

$$Phi_{max} = \frac{16-0}{\sqrt{4 \times 4 \times 4 \times 4}} = \frac{16}{16} = 1$$

Phi_{korr}=0.5 -> 25% gem. Variation

Es besteht mittlerer Zusammenhang zwischen Geschlecht und Abteilungszugehörigkeit.

Kapitel 5

022 a) Mittelwert, Standardabweichung, Varianz für Frauen und Männer

Frauen:
MW=17/4 = 4.25

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n} = \frac{91 - \frac{17^2}{4}}{4} = \frac{91 - 72.25}{4} = 4.69 \Rightarrow S = 2.17$$

Männer:
MW=18/4 = 4.5

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n} = \frac{110 - \frac{18^2}{4}}{4} = \frac{110 - 81}{4} = 7.25 \Rightarrow S = 2.69$$

b) Prüfung auf Varianzgleichheit: F-Test
Hypothesen:

$H1: \sigma^2_{Frauen} \neq \sigma^2_{Männer}$

$H0: \sigma^2_{Frauen} = \sigma^2_{Männer}$

$$\hat{\sigma}^2 = S^2 \times \frac{n}{n-1}$$

$$\hat{\sigma}^2_{Frauen} = 4.69 \times \frac{4}{4-1} = 6.25$$

$$\hat{\sigma}^2_{Männer} = 7.25 \times \frac{4}{4-1} = 9.67$$

$$F_{emp} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{9.67}{6.25} = 1.55$$

$F_{krit(df\text{-}Zähler=3;\ df\text{-}Nenner=3;\ alpha=5\%,\ zweiseitig)} = 9.28$

Es kann nicht behauptet werden, dass sich die Varianzen von Frauen und Männern unterscheiden, die Varianzen sind homogen.

c) Prüfung auf Mittelwertsunterschiede, homogene Varianzen (siehe b) → t-Test
Hypothesen:

H0: $\mu_w = \mu_m$ Frauen und Männer unterscheiden sich (im Durchschnitt / Mittel) nicht voneinander bei den Fähigkeitswerten.

H1: $\mu_A \neq \mu_B$ Frauen und Männer unterscheiden sich (im Durchschnitt / Mittel) voneinander bei den Fähigkeitswerten.

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

$$\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{3\times 6.25 + 3\times 9.67}{4+4-2} \times \left(\frac{1}{4}+\frac{1}{4}\right)} = \sqrt{\frac{18.75+29}{6}\times 0.5}$$

$$\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{47.75}{6}\times 0.5} = \sqrt{3.98} = 1.99$$

$$t_{emp} = \frac{\bar{x}_1-\bar{x}_2}{\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)}} = \frac{4.5-4.25}{1.99} = \frac{0.25}{1.99} = 0.126$$

$t_{krit(df=6;\ 5\%\ zweiseitig)} = 2.447$

Es kann nicht behauptet werden, dass sich Frauen und Männer hinsichtlich der Fähigkeit im Umgang mit dem PC unterscheiden.

d) „Fähigkeit im Umgang mit dem PC (x)" intervallskaliert,
„Anzahl besuchter Fortbildungen (y)" mindestens intervallskaliert,
da natürlicher Nullpunkt: Produkt-Moment-Korrelation

Zur Berechnung müssen noch einige Summen gebildet werden:

$$\sum_{i=1}^{n} x_i = 35 \qquad \sum_{i=1}^{n} x_i^2 = 201 \qquad \sum_{i=1}^{n} y_i = 19 \qquad \sum_{i=1}^{n} y_i^2 = 71 \qquad \sum_{i=1}^{n} x_i \times y_i = 112$$

$$r_{PM} = \frac{8\times 112 - 35\times 19}{\sqrt{\left[8\times 201 - 35^2\right]\times\left[8\times 71 - 19^2\right]}}$$

$$r_{PM} = \frac{896-665}{\sqrt{[1608-1225]\times[568-361]}}$$

$$r_{PM} = \frac{231}{\sqrt{383\times 207}} = \frac{231}{\sqrt{79281}} = \frac{231}{281.57} = 0.820 \quad \Rightarrow \quad r^2 = 0.6724$$

Prüfung auf Signifikanz: t-Test für Korrelationen:

Hypothesen:

H0: $\rho = 0$ Es besteht kein Zusammenhang zwischen Fähigkeit und Anzahl besuchter Fortbildungen.

H1: $\rho \neq 0$ Es besteht ein Zusammenhang zwischen Fähigkeit und Anzahl besuchter Fortbildungen..

$$t_{emp} = \frac{r\times\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.82\times\sqrt{6}}{\sqrt{1-0.6724}} = \frac{0.82\times 2.45}{0.572} = 3.51$$

$t_{krit(df=6;\ 5\%\ zweiseitig)} = 2.447$

Mit einer Irrtumswahrscheinlichkeit von 5% kann behauptet werden, dass es einen Zusammenhang zwischen den Fähigkeiten im Umgang mit dem PC und der Anzahl besuchter Fortbildungen gibt.

023 Prüfende Statistik, schließende Statistik. Oder: „Wie sind Stichprobenergebnisse auf die betreffende Grundgesamtheit übertragbar?"

024 Pädagogischer Ansatz des Kindergartens und Schuleignung
Es kann nicht von einer Normalverteilung der Daten ausgegangen werden: Also kein Intervallskalenniveau, aber sicherlich Ordinalskalenniveau. Die Stichproben sind unabhängig voneinander, also Mann-Whitney-U-Test.

Vergabe von Rangplätzen
Montessori-Ansatz: n=7, Rangplätze: 1., 5., 4., 9., 11., 12., 6. → Rangplatzsumme T=48
Situations-Ansatz: n=5, Rangplätze: 2., 3., 7., 10., 8. -> Rangplatzsumme T=30

Hypothesen:
H0: Bezogen auf die Schuleignung unterscheiden sich Montessori-Ansatz und Situations-Ansatz nicht voneinander.
H1: Bezogen auf die Schuleignung unterscheiden sich Montessori-Ansatz und Situations-Ansatz voneinander.

Berechnung von U:

Montessori-Ansatz
$$U = n_1 \times n_2 + \frac{n_1 \times (n_1 + 1)}{2} - T1 = 7 \times 5 + \frac{7 \times 8}{2} - 48 = 35 + 28 - 48 = 15$$
$$U' = n_1 \times n_2 - U = 7 \times 5 - 15 = 20$$

Situations-Ansatz
$$U = n_1 \times n_2 + \frac{n_1 \times (n_1 + 1)}{2} - T1 = 5 \times 7 + \frac{5 \times 6}{2} - 30 = 35 + 15 - 30 = 20$$
$$U' = n_1 \times n_2 - U = 5 \times 7 - 20 = 15$$

Mit dem kleineren U-Wert in Tabelle E gehen und Wahrscheinlichkeit ablesen.
n_1=7, n_2=5, U=15

p = 0,378, d.h. die Wahrscheinlichkeit dafür, dass diese Rangplatzverteilung auftritt, beträgt 37,8%, die Rangplatzverteilung ist also nicht signifikant unterschiedlich! (größer als 5%!!)

Es kann nicht behauptet werden, dass sich die beiden Kindergarten-Konzepte hinsichtlich der Schuleignung voneinander unterscheiden!

Kapitel 6

025 Bierausschank

a) Formeln für Einfachregression

$$b = \frac{n \times \sum_{i=1}^{n} x_i \times y_i - \sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n \times \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{r_{PM} \times S_y}{S_x} \qquad a = \bar{y} - b \times \bar{x}$$

	Woche	(Woche)²	Verkauftes Bier (in Hundert Litern)	(Verk. Biere)²	(Woche*Verk. Biere)
	1	1	3	9	3
	2	4	4	16	8
	3	9	4	16	12
	4	16	5	25	20
	5	25	4	16	20
	6	36	6	36	36
	7	49	5	25	35
	8	64	7	49	56
	9	81	6	36	54
	10	100	8	64	80
Summe	55	385	52	292	324

$$b = \frac{n \times \sum_{i=1}^{n} x_i \times y_i - \sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n \times \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{10 \times 324 - 55 \times 52}{10 \times 385 - 55^2}$$

$$b = \frac{3240 - 2860}{3850 - 3025} = \frac{380}{825} = 0{,}4606$$

$$a = \bar{y} - b \times \bar{x} = 5{,}2 - 0{,}4606 \times 5{,}5 = 5{,}2 - 2{,}5333$$

$$a = 2{,}6667$$

Regressionsgleichung: $\hat{y} = 0{,}4606 \times x + 2{,}6667$

b) 20. Woche → x=20

$$\hat{y} = 0{,}4606 \times x + 2{,}6667$$

$$\hat{y} = 0{,}4606 \times 20 + 2{,}6667 = 9{,}212 + 2{,}6667$$

$$\hat{y} = 11{,}8787$$

In den Formeln für das Konfidenzintervall tauchen dummerweise auch die Varianzen der beiden Variablen auf … Die müssen wir zuerst noch berechnen. Aber wir haben ja schon die Summen gebildet, daher hält sich auch das in Grenzen.

Woche:
$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n} = \frac{385 - \frac{55^2}{10}}{10} = \frac{385 - 302{,}5}{10} = \frac{82{,}5}{10} = 8{,}25$$

Verkaufte Biere:
$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n} = \frac{292 - \frac{52^2}{10}}{10} = \frac{292 - 270{,}4}{10} = \frac{21{,}6}{10} = 2{,}16$$

So, dann zum Standardschätzfehler:

$$\hat{\sigma}_{yx} = \sqrt{\frac{n \times S_y^2 - n \times b_{yx}^2 \times S_x^2}{n-2}} = \sqrt{\frac{10 \times 2{,}16 - 10 \times 0{,}4606^2 \times 8{,}25}{10-2}}$$

$$\hat{\sigma}_{yx} = \sqrt{\frac{21{,}6 - 17{,}5026}{8}} = \sqrt{\frac{4{,}0974}{8}} = \sqrt{0{,}5122} = 0{,}7157$$

Obergrenze: $\hat{y}_i + t_{(\alpha/2)} \times \hat{\sigma}_{yx} \times \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{n \times S_x^2}}$

t=2,306

$$t_{(\alpha/2)} \times \hat{\sigma}_{yx} \times \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{n \times S_x^2}}$$

$$2{,}306 \times 0{,}7157 \times \sqrt{\frac{1}{10} + \frac{(20 - 5{,}5)^2}{10 \times 8{,}25}} = 1{,}6504 \times \sqrt{0{,}1 + \frac{210{,}25}{82{,}5}}$$

$$1{,}6504 \times \sqrt{2{,}6485} = 1{,}6504 + 1{,}6274 = 2{,}6859$$

Damit ermitteln wir folgende Grenzen für das Konfidenzintervall:
Untergrenze: 11,8787-2,6859=9,1928
Obergrenze: 11,8787+2,6859=14,5646

Mit einer Wahrscheinlichkeit von 95% befindet sich der geschätzte wahre Wert für „Verkaufte Biere" in der 20ten Woche in dem Bereich von 9,1928 bis 14,5646.

c) $$b = \frac{n \times \sum_{i=1}^{n} x_i \times y_i - \sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n \times \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \frac{10 \times 324 - 52 \times 55}{10 \times 292 - 52^2} = \frac{3240 - 2860}{2920 - 2704} = \frac{380}{216} = 1{,}7593$$

$$a = \bar{y} - b \times \bar{x} = 5{,}5 - 1{,}7593 \times 5{,}2 = 5{,}5 - 9{,}1484$$

$$a = -3{,}6484$$

Regressionsgleichung: $\hat{y} = 1{,}7593 \times x - 3{,}6484$

d) 4.000 Liter -> Einheit in 100 Litern: x=40

$$\hat{y} = 1{,}7593 \times x - 3{,}6484 = 1{,}7593 \times 40 - 3{,}6484$$

$$\hat{y} = 70{,}372 - 3{,}6484 = 66{,}7236$$

Unser Wirt Alex kann darauf hoffen – wenn alles so weitergeht – ab der 67ten Woche die Sonderkonditionen zu erhalten. Drücken wir ihm die Daumen!

Kapitel 7

026 Das Fundamentaltheorem der Faktorenanalyse lautet: R = AA´. Es beschreibt, dass im Idealfall bei der Datenreduktion keine Information verloren geht.

027 Der Eigenwert eines Faktors beschreibt den Varianzanteil, der durch den Faktor erklärt wird; allgemein ausgedrückt beschreibt der Eigenwert die Anzahl an Variablen, die durch den Faktor ersetzt werden.

028 Die Kommunalität eines Items beschreibt, wie viel Prozent der ursprünglich in dem Item steckenden Information nach der Datenreduktion noch vorhanden sind.

029 Kaiser- und Scree-Kriterium dienen zur Bestimmung der Anzahl zu bildender (zu extrahierender) Faktoren. Das Kaiser-Kriterium lautet: Der Eigenwert eines Faktors muss mindestens 1 betragen. Das Scree-Kriterium bezieht sich auf eine Grafik, in welcher der Eigenwerteverlauf der Faktoren dargestellt ist. Ein solcher Eigenwerteverlauf geht an einer bestimmten Stelle in eine Gerade über. Zu berücksichtigen sind nur solche Faktoren vor diesem Knick.

Kapitel 8

030 Eine multiple Regressionsanalyse dient zur Vorhersage / zur Schätzung einer Kriteriumsvariablen (y) mittels mehr als einer Prädiktorvariablen (x1, x2, x3...).

031 Allgemein gilt: $\hat{y} = b_0 + b_1 \times x_1 + b_2 \times x_2$

Für den beschriebenen Fall gilt: $\hat{y} = 2 + 1.6 \times x_1 + 1 \times x_2 \rightarrow x_1 = 5,\ x_2 = 3$

$$\hat{y} = 2 + 1.6 \times 5 + 1 \times 3 = 2 + 8 + 3 = 13$$

Für den neuen Bewerber würde man eine Vorgesetztenbeurteilung (y) von 13 Punkten schätzen.

Kapitel 9

032 Eine Varianzanalyse vergleicht Mittelwerte.

033 Die totale Quadratsumme QS_{tot} kann zerlegt werden in determinierte Quadratsumme QS_{det} und Fehlerquadratsumme QS_{error}

034 Bei einer zweifaktoriellen Varianzanalyse kann die determinierte Quadratsumme QS_{det} zerlegt werden in $QS_{Faktor\ A}$, $QS_{Faktor\ B}$ und Wechselwirkung $QS_{Faktor\ A*Faktor\ B}$.

035 Kundenzufriedenheitsstudie

Quelle der Variation	Quadratsumme	Freiheits-grade	F	Sig.
Faktor A	48	?=2	?=2,72	n.s.
Faktor B	18	?=1	?=2,04	?=n.s.
Wechselwirkung A*B	?=144	2	8,15	?=sig.
determinierte / Modell	210	?=5	?=4,76	?=sig.
Fehler	?=96	12		
total	306	?=17		

$QS_{tot}=QS_{det}+QS_{err}$ -> $QS_{error}=316 - 210 =106$

$QS_{det}=QS_{Faktor\ A} + QS_{Faktor\ B} + QS_{A*B}$ -> $QS_{A*B}=210-48-18=144$

Faktor A: 3 Stufen -> $df_A=2$; Faktor B: 2 Stufen -> $df_B=1$

$df_{det}=df_A+df_B+df_{A*B}=2+1+2+=5$; $df_{total}=df_{det}+df_{error}=5+12=17$

Signifikanzprüfungen

Faktor A: $F_{emp} = \frac{QS_h / df_h}{QS_e / df_e} = \frac{48/2}{96/12} = \frac{24}{8} = 3; F_{krit} = 3.89 \rightarrow n.s.$

Faktor B: $F_{emp} = \frac{QS_h / df_h}{QS_e / df_e} = \frac{18/1}{96/12} = \frac{18}{8} = 2.25; F_{krit} = 4.75 \rightarrow n.s.$

Wechselwirkung A*B: $F_{emp} = \frac{QS_h / df_h}{QS_e / df_e} = \frac{144/2}{96/12} = \frac{72}{8} = 9; F_{krit} = 3.89 \rightarrow sig.$

Determinierte: $F_{emp} = \frac{QS_h / df_h}{QS_e / df_e} = \frac{210/5}{96/12} = \frac{42}{8} = 5.25; F_{krit} = 3.11 \rightarrow sig.$

Kapitel 10

036 Mindestens ordinalskalierte Variablen.

037 Bewerberauswahl
H0: Die Interviewer unterscheiden sich nicht
H1: Die Interviewer unterscheiden sich.

$$H_{emp} = \frac{12}{N \times (N+1)} \times \sum_{j=1}^{k} \frac{T_j^2}{n_j} - 3 \times (N+1)$$

$$H_{emp} = \frac{12}{18 \times 19} \times \left(\frac{61^2}{6} + \frac{70^2}{6} + \frac{40^2}{6} \right) - 3 \times 19$$

$$H_{emp} = \frac{12}{342} \times (620.17 + 816.67 + 266.67) - 57$$

$$H_{emp} = 0.0351 \times 1703.50 - 57 = 59.79 - 57 = 2.79$$

$H_{krit(df=2)} = 5.99$

Nein, es gibt keinen signifikanten Interviewereffekt.

Glossar

Abhängige Stichproben
Werden zwei oder mehr Gruppen miteinander verglichen und lassen sich jeweils Wertepaare bzw. Wertereihen über die Gruppen hinweg bilden, so wird von abhängigen Stichproben gesprochen.

Additionstheorem
Wahrscheinlichkeitsrechnung; sind zwei Bedingungen durch 'ODER' verknüpft, werden die Einzelwahrscheinlichkeiten addiert.

Allgemeines lineares Modell (ALM)
Formalsatz in Matrixalgebra, mittels welchem z.B. Varianz- und Regressionsanalysen gerechnet werden können.

alpha-Fehler
Man begeht einen "α-Fehler " (auch Fehler 1. Art genannt), wenn man bei einem statistischen Hypothesentest die Nullhypothese H_0 ablehnt, obwohl sie wahr ist. Diese Fehlentscheidung wird als Fehler 1. Art bezeichnet. Über das Signifikanzniveau α legt man im Vorfeld der Testdurchführung fest, mit welcher maximalen Wahrscheinlichkeit dieser Fehler 1. Art auftreten darf.

beta-Fehler
Man begeht einen "β-Fehler" (auch Fehler 2. Art genannt), wenn man bei einem statistischen Hypothesentest die Nullhypothese annimmt, obwohl sie falsch ist (also H_1 wahr wäre).

Deskriptivstatistik
Beschreibende Statistik; Oberbegriff für alle Kennwerte, welche zur Beschreibung der Stichproben bzw. Variablen herangezogen werden.

Faktorenanalyse
Rechentechnik zur Gruppierung von Variablen an Hand deren Korrelationen.

Inferenzstatistik
Schließende Statistik; Oberbegriff für alle Testverfahren, mittels derer der Schluss von Stichprobe auf Grundgesamtheit durchgeführt werden kann.

Intervallskalenniveau
Datenniveau; Daten unterscheiden sich nach gleich, ungleich, größer, kleiner sowie plus und minus; kein natürlicher Nullpunkt vorhanden.

Konfidenzintervall
Bereich, innerhalb dessen ein geschätzter Wert mit einer gewissen Wahrscheinlichkeit zu Liegen kommt.

Korrelation / Korrelationskoeffizient
Eine Korrelation beschreibt eine Beziehung zwischen zwei oder mehreren Merkmalen. Der Korrelationskoeffizient (auch Produkt-Moment-Korrelation, Pearson-Korrelation, Korrelationswert oder auch kurz Korrelation genannt) ist eine Messgröße für den Grad des linearen Zusammenhangs zwischen zwei mindestens intervallskalierten Variablen

Kriterium
Jene Variable, deren Ausprägung mittels eines oder mehrerer Prädiktoren geschätzt wird.

Median
Der Median ist derjenige Wert, der in einer Rangreihe (geordnete Stichprobe) mit einer ungeraden Anzahl von Beobachtungswerten in der Mitte liegt. Rechts und links vom Median liegen somit gleich viele Merkmalswerte. Bei einer geraden Anzahl von Beobachtungswerten wird der Median als arithmetisches Mittel aus den beiden in der Mitte liegenden Werten der geordneten Rangreihe berechnet. Zur Bestimmung des Medians müssen die zugrundeliegenden Daten mindestens ordinalskaliert sein.

Mittelwert
Durchschnittswert; Summe der Einzelwerte dividiert durch die Anzahl der Werte.

Modalwert / Modus
Diejenige Merkmalsausprägung oder diejenigen Merkmalsausprägungen, die am häufigsten vorkommt bzw. vorkommen. Der Modus kann bei allen Skalenniveaus angewandt werden, also auch bei nominalskalierten Merkmalen.

Multiplikationstheorem
Wahrscheinlichkeitsrechnung; sind zwei Bedingungen durch 'UND' verknüpft, werden die Einzelwahrscheinlichkeiten multipliziert.

Nominalskalenniveau
Datenniveau; Daten unterscheiden sich lediglich nach gleich vs. ungleich.

Ordinalskalenniveau
Datenniveau; Daten unterscheiden sich nach gleich, ungleich, größer und kleiner.

Phi-Korrelation
Korrelationstechnik, wenn jene beiden Merkmale, die in Zusammenhang gebracht werden sollen, nominalskaliert sind und jeweils in zwei Stufen auftreten.

Prädiktor
Jene Variable, mittels deren Ausprägung ein Kriterium geschätzt wird.

Produkt-Moment-Korrelation
Korrelationstechnik, wenn jene beiden Merkmale, die in Zusammenhang gebracht werden sollen, intervallskaliert sind.

Regressionsanalyse
Ziel der Regressionsanalyse (kurz auch Regression genannt) ist es, Beziehungen zwischen einer abhängigen und einer oder mehreren unabhängigen Variablen festzustellen.

Spearman-Rang-Korrelation
Korrelationstechnik, wenn jene beiden Merkmale, die in Zusammenhang gebracht werden sollen, ordinalskaliert sind.

Standardabweichung
Positive Quadratwurzel aus der Varianz

Unabhängige Stichproben
Werden zwei oder mehr Gruppen miteinander verglichen und werden keine Wertepaare bzw. Wertereihen über die Gruppen hinweg gebildet, so wird von unabhängigen Stichproben gesprochen.

Varianz
Die Varianz ist ein Maß für die Streuung von Beobachtungswerten. Sie wird berechnet als durchschnittliche quadratische Abweichung der einzelnen Beobachtungswerte vom arithmetischen Mittel. Die positive Wurzel aus der Varianz wird als Standardabweichung bezeichnet.

Varianzanalyse
Rechentechnik zum Vergleich mehrerer Mittelwerte.

Verhältnisskalenniveau
Datenniveau; Daten besitzen einen natürlichen Nullpunkt; alle mathematischen Operationen sind erlaubt.

Literaturverzeichnis

Atteslander, P.: Methoden der empirischen Sozialforschung. 12., durchgesehene Auflage. Erich Schmidt. Berlin 2008

Backhaus, K./Erichson, B./Plinke, W./Weiber, R.: Multivariate Analysemethoden - Eine anwendungsorientierte Einführung. 12., vollständig überarbeitete Auflage. Springer. Berlin 2008

Bortz, J.: Statistik für Sozialwissenschaftler. 5., vollständig überarbeitete und aktualisierte Auflage. Springer. Berlin 1999

Claus, G./Ebner, H.: Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen. VEB. Berlin 1974

Papula, L.: Mathematische Formelsammlung: für Ingenieure und Naturwissenschaftler. 10., überabeitete und erweiterte Auflage. Vieweg + Teubner. Wiesbaden 2009

Pospeschill, M.: Statistische Methoden. Strukturen, Grundlagen, Anwendungen in Psychologie und Sozialwissenschaften. Elsevier. München 2006

Siegel, S.: Nichtparametrische statistische Methoden. 5. Aufl. Verl. Dietmar Klotz. Eschborn 2001

Abbildungsverzeichnis

Tabellenverzeichnis