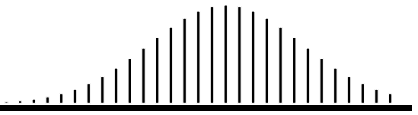
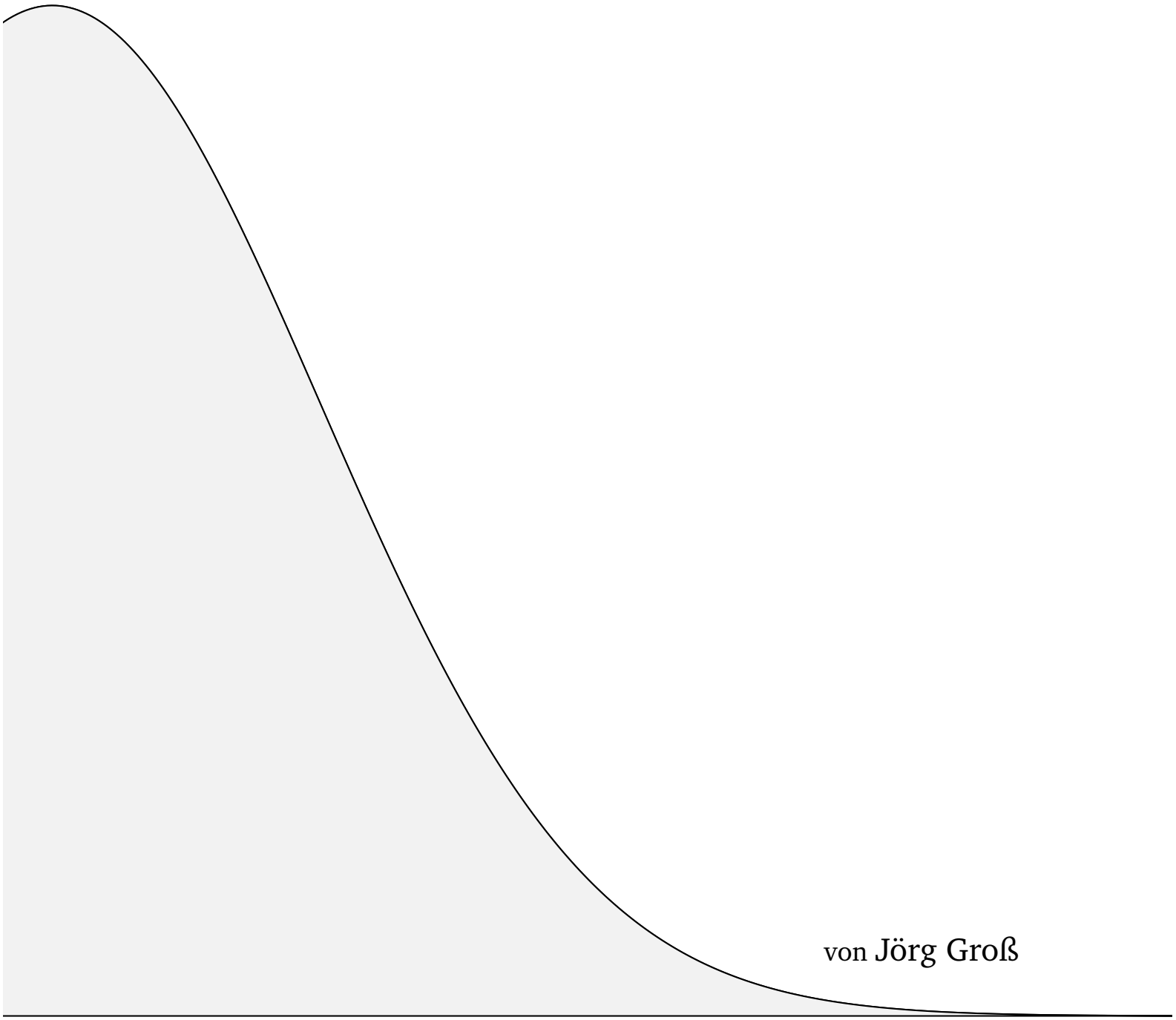
Reader

Psychologische Statistik I



Goethe-Universität, Frankfurt am Main
Fachbereich Psychologie



von Jörg Groß

1. Version: November, 2008

erstellt von: Jörg Groß
Kontakt: *joerg@licht-malerei.de*
http://www.random-stuff.de

Comic-Strips copyright by Jorge Cham
"Piled Higher and Deeper"
http://www.phdcomics.com

Dieser Reader wurde erstellt mit LaTeX
Die Abbildungen und Berechnungen wurden mit R erstellt und durchgeführt
Informationen zum Textsatzprogramm LaTeX: *http://www.latex-project.org*
Informationen zum Statistikprogramm R: *http://www.r-project.org*

Inhaltsverzeichnis

1 Grundlegendes **4**

1.1 Ein wenig über Messtheorie . . . 4
1.1.1 Homomorphe versus isomorphe Abbildung . . . 5
1.2 Skalenarten . . . 7
1.2.1 Nominalskalenniveau . . . 7
1.2.2 Ordinalskalenniveau . . . 7
1.2.3 Intervallskalenniveau . . . 7
1.2.4 Verhältnisskalenniveau . . . 8
1.2.5 Beziehungen zwischen den Skalenniveaus . . . 8
1.2.6 Erlaubte Transformationen . . . 8
1.2.7 Bedeutung des Skalenniveaus . . . 9
1.3 Wozu eigentlich Statistik? . . . 9
1.4 Statistische Analysen mit dem PC . . . 10
1.5 Mathematische Symbole . . . 12
1.5.1 Summenzeichen . . . 12
1.5.2 Fakultät . . . 13
1.A *Aufgaben* . . . 14

2 Deskriptivstatistik **16**

2.1 Maße der zentralen Tendenz . . . 16
2.1.1 Modalwert . . . 16
2.1.2 Median . . . 16
2.1.3 Arithmetisches Mittel . . . 17
2.1.4 Welches Maß der zentralen Tendenz ist das Richtige? . . . 19
2.A *Aufgaben* . . . 22
2.2 Dispersionsmaße . . . 23
2.2.1 Range . . . 23
2.2.2 Interquartilbereich . . . 23
2.2.3 Varianz und Standardabweichung . . . 24
2.2.4 Welches Dispersionsmaß ist das Richtige? . . . 26
2.B *Aufgaben* . . . 28
2.3 Deskriptive Darstellungsformen . . . 30
2.3.1 Das Histogramm . . . 30
2.3.2 Der Stem und Leaf Plot . . . 31
2.3.3 Der Box-Whisker Plot . . . 32
2.3.4 Der Mittelwertsplot . . . 33
2.C *Aufgaben* . . . 35
2.4 Kovarianz, Korrelation und Regression . . . 37
2.4.1 Kovarianz . . . 37
2.4.2 Korrelation . . . 37
2.4.3 Regression . . . 39
2.4.4 Zusammenhang von Korrelation und Regression . . . 42
2.4.5 Voraussetzungen der Korrelations- und Regressionsanalyse . . . 43
2.D *Aufgaben* . . . 45

3 Stochastik **48**

3.1 Wozu Stochastik? . . . 48
3.2 Das Zufallsexperiment . . . 48

3.2.1 Wahrscheinlichkeit für ein Ereignis . . . 51
3.2.2 Wahrscheinlichkeit für mehrere Ereignisse . . . 52
3.A *Aufgaben* . . . 56
3.3 Kombinatorik . . . 58
3.3.1 Permutation . . . 58
3.3.2 Kombination . . . 59
3.3.3 Variation . . . 60
3.3.4 Zusammenfassung . . . 62
3.3.5 Nachtrag: Das Geburtstagparadoxon . . . 62
3.B *Aufgaben* . . . 63
3.4 Wahrscheinlichkeit für eine Folge von Ereignissen . . . 65
3.4.1 Binomialverteilung . . . 66
3.C *Aufgaben* . . . 69

4 Inferenzstatistik **71**
4.1 Von der Population zur Stichprobe und zurück . . . 72
4.1.1 Schätzung des Populationsmittelwerts . . . 72
4.1.2 Schätzung der Populationsvarianz . . . 74
4.1.3 Standardfehler des Mittelwerts . . . 76
4.2 Die Normalverteilung . . . 76
4.2.1 z-Transformation . . . 77
4.2.2 Konfidenzintervalle . . . 78
4.A *Aufgaben* . . . 81
4.3 inferenzstatistische Tests . . . 83
4.3.1 Der z-Test . . . 83
4.B *Aufgaben* . . . 88
4.3.2 Der t-Test für unabhängige Stichproben . . . 90
4.C *Aufgaben* . . . 95
4.3.3 Der t-Test für gepaarte Stichproben . . . 97
4.D *Aufgaben* . . . 100
4.3.4 Der Mann-Whitney-U-Test . . . 102
4.E *Aufgaben* . . . 104
4.3.5 Der Chi-Quadrat Test . . . 105
4.F *Aufgaben* . . . 107
4.3.6 Relevanz von Mittelwertsunterschieden . . . 108
4.3.7 Zusammenfassung . . . 108
4.4 Signifikanz von Korrelationen . . . 108
4.G *Aufgaben* . . . 111

Anhang **112**
A Statistik mit dem Taschenrechner . . . 114

1 Grundlegendes

1.1 Ein wenig über Messtheorie

Die moderne Psychologie, so wie sie an dem meisten Universitäten gelehrt wird, versteht sich als Erfahrungs- oder empirische Naturwissenschaft. Das bedeutet für die Praxis psychologischer Forschung, dass es nicht ausreicht Theorien über das Erleben und Verhalten des Menschen zu entwickeln. Stattdessen müssen Theorien empirisch überprüfbar sein. Sie müssen also belegbar sein durch eine systematische Beobachtung des tatsächlichen Verhaltens.
Eine besondere Form dieser systematischen Beobachtung ist das Experiment. Das Experiment gilt als via regia, als Königsweg, des Erkenntnisgewinns in den meisten Naturwissenschaften.
Ein psychologisches Experiment läuft dabei so ab, dass Personen (auch Versuchspersonen oder Probanden genannt) in eine künstlich hergestellten Situation innerhalb einer manipulierbaren Umgebung (meist in einem Labor) gebracht werden und das Verhalten dieser Personen mittels verschiedener Methoden aufgezeichnet wird. Das Experiment gilt deswegen als Königsweg des empirischen Erkenntnisgewinns, weil sich die hergestellte Umgebung systematisch verändern lässt und man dadurch beobachten kann, wie sich Personen unter einer Bedingung A und unter einer abgewandelten Bedingung B verhalten.
Vergleicht man nun das Verhalten der Personen zwischen den verschiedenen künstlich hergestellten Bedingungen, lassen sich Rückschlüsse auf das generelle Verhalten und Erleben von Menschen außerhalb der Laborsituation ziehen. Diese Rückschlüsse kann man nun vergleichen mit den Vorhersagen, die eine Theorie über dieses Verhalten aufgestellt hat.
Ein Problem ergibt sich dabei nun dadurch, dass Theorien meist innerhalb eines sprachlichen Systems formuliert werden. Das Verhalten von Personen in einem Experiment ließe sich auch sprachlich beschreiben und mit den theoretischen Vorüberlegungen vergleichen. Diese qualitative Auswertungsmethode birgt allerdings einige Gefahren. So können Ergebnisse und Rückschlüsse sehr von der Person abhängen, die das beobachtete Verhalten der Probanden in einer experimentellen Situation interpretiert und „zur Sprache bringt“.
In der psychologischen Forschung sind daher quantitative Beobachtungs- und Auswertungsmethoden verbreiteter. Dabei wird versucht, das Verhalten der Person zu messen. Genau wie in der experimentellen Physik, oder anderen Bereichen empirischer Forschung, bedeutet Messen dabei Beobachtungen in Zahlen wiederzugeben. Man versucht, das Beobachtete mithilfe eines mathematischen Systems abzubilden.
Dieses Erkenntnissystems, bestehend aus Theorie und Experiment, verfügt also über drei Schnittstellen: die sprachliche Ebene der Theorie, die Ebene der Phänomene bzw. des Verhaltens von Personen und ein mathematisches Abbildungssystem dieser Phänomene.
Die Messtheorie befasst sich damit, wie sich diese drei Schnittstellen vereinbaren lassen und in welcher Form man beobachtbaren Phänomenen mathematische Symbole zuordnen kann. Im

Grunde geht es also darum, inwieweit man Rückschlüsse von der einen Ebene auf die andere ziehen kann, um am Ende tatsächlich menschliches Verhalten erklären, vorherzusagen oder vielleicht sogar manipulieren und modifizieren zu können.

1.1.1 Homomorphe versus isomorphe Abbildung

Bei der Umwandlung des beobachteten Verhaltens einer Person in mathematische Symbole bzw. in ein so genanntes nummerisches Relativ unterscheidet man zwei Abbildungsformen: die isomorphe und die homomorphe Abbildung.
Bei einer isomorphen Abbildung besteht eine ein-eindeutige Beziehung zwischen dem nummerischen Relativ und der empirischen Merkmalsausprägung (oder auch dem empirischen Relativ). Das bedeutet, dass jeder Merkmalsausprägung aus dem empirischen Raum ein Symbol aus dem mathematischen Raum zuordenbar ist. Von der Zahl kann man also auf die Ausprägung im Merkmal und von der Ausprägung kann man auf die Zahl schließen. Es geht bei der Zuordnung keine Information verloren.
Bei der homomorphen Abbildung ist zwar eine eindeutige Beziehung zwischen nummerischen Relativ und empirischen Relativ gegeben, dass heißt für jede Merkmalsausprägung in der Empirie gibt es genau eine Zahl oder ein mathematisches Symbol, doch kann es mehrere Ausprägungen geben, denen dieselbe Zahl zugeordnet wird. Man kann vom empirischen Relativ (Objekte, Merkmalsausprägung, Verhaltensweisen etc.) eindeutig auf das nummerische Relativ (Zahlenraum) schließen. Allerdings kann man nicht zwangsläufig umgekehrt die empirische Merkmalsausprägung von der nummerischen Zuordnung ableiten.
In Abbildung 1.1 ist dieser Unterschied zwischen einer isomorphen und einer homomorphen Relation grafisch verdeutlicht. In Abbildung 1.1.a ist eine isomorphe Abbildung gegeben: Von

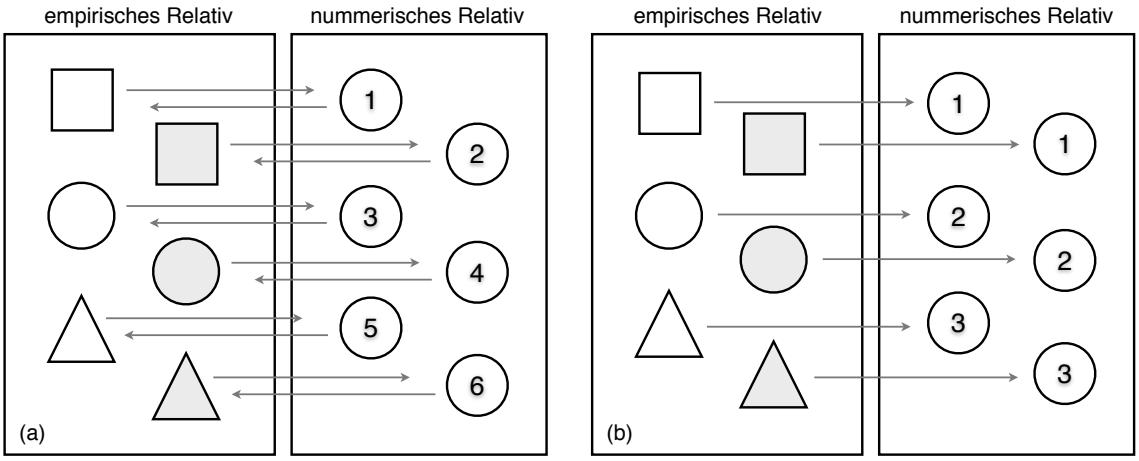
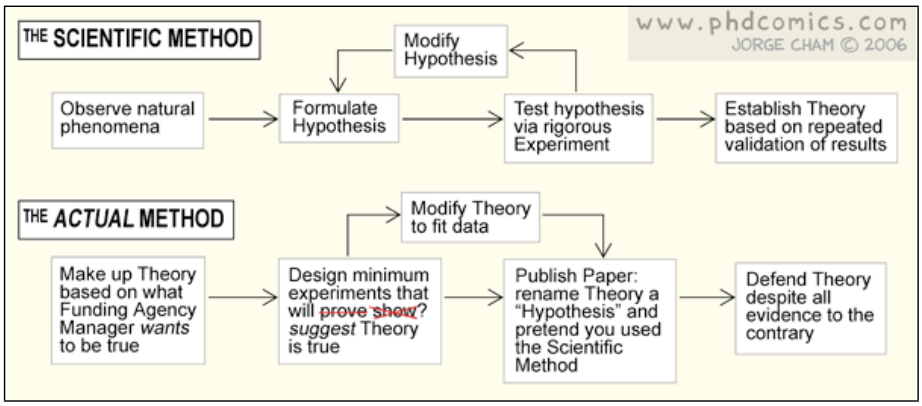


Abbildung 1.1: Umsetzung eines empirischen Relativs in ein nummerisches Relativ, via (a) einer isomorphen und (b) einer homomorphen Abbildung

der Zahl 3 lässt sich auf die Merkmalsausprägung *rund/weiß* schließen und umgekehrt.
In Abbildung 1.1.b ist dagegen eine homomorphe Abbildung gegeben: Von der Merkmalsausprägung *rund/weiß* lässt sich auf die Zahl 2 schließen. Umgekehrt geht das jedoch nicht:

Von der Zahl 2 lässt sich nicht mehr entscheiden, ob es sich im empirischen Relativ um die Merkmalsausprägung *rund/weiß* oder *rund/grau* handelte.

Um das Ganze zu verdeutlichen ein reales Beispiel: Angenommen man benutzt zur Messung eines psychologischen Merkmals einen Leistungstest mit verschiedenen Aufgaben. Wie gut man in diesem Test abschneidet, dass heißt welchen Punktewert man erreicht, hängt davon ab, wie lange man zur Bearbeitung des Tests braucht und wie viele Fehler man in den einzelnen Aufgaben macht. Der erzielte Punktewert (nummerisches Relativ) lässt keinen eindeutigen Rückschluss auf das empirische Relativ (Zeit und Fehler) zu. Da ein Punktewert von zum Beispiel 15 sich aus verschiedenen Konstellationen von benötigter Zeit und Fehleranzahl zusammensetzen kann. Die Zahl 15 kann beispielsweise entstehen, wenn man drei Minuten gebraucht hat und ein Fehler gemacht hat, genauso aber auch, wenn nur eine Minute benötigt wurde, dabei aber fünf Fehler gemacht wurden. Es handelt sich also um eine homomorphe Abbildung zwischen nummerischem und empirischem Relativ.
Angenommen man nimmt nun nur die Zeit und versucht von der Zeit auf den Punktewert zu schließen und umgekehrt, so wird dies weder in die eine noch in die andere Richtung funktionieren. Hierbei besteht also weder eine isomorphe noch eine homomorphe Beziehung. Eine Bearbeitungszeit von drei Minuten kann mit einem Punktewert von 15, genauso wie mit einem Punktewert von 23, einhergehen. Ebenso kann ein Punktewert von 23 mit einer Zeit von zwei Minuten oder sieben Minuten einhergehen.
Die Abbildung von Zeit (empirisches Relativ) auf Punktewert (nummerisches Relativ) ist in diesem Fall weder eindeutig noch ein-eindeutig.



1.2 Skalenarten

Eine (mindestens) homomorphe Abbildungsfunktion eines empirischen Relativs (Menge von Objekten) zu einem nummerischen Relativ (Menge von Zahlen) nennt man Skala (von ital. Treppe/Leiter).
Man unterscheidet im Wesentlichen vier Skalenniveaus, die sich hinsichtlich der Abbildungsart oder auch der Abbildungsgenauigkeit des empirischen Relativs unterscheiden:

- Nominalskalenniveau
- Ordinalskalenniveau
- Intervallskalenniveau
- Verhältnisskalenniveau

1.2.1 Nominalskalenniveau

Auf einer Nominalskala lassen sich qualitative Merkmale oder Objekte abbilden. Diese Merkmale lassen sich nicht quantitativ, durch größer/kleiner Beziehungen, unterscheiden. Typische empirische Merkmale, die sich in nummerische Relative auf Nominalskalenniveau umwandeln lassen, sind Studiengang oder Herkunft.
Eine Sonderform der Nominalskala ist eine dichotome Skala, die nur aus zwei Zahlen besteht und damit auch nur zwei empirische Objektausprägungen abbildet. Ein typisches Beispiel wäre hierbei das Geschlecht, das im Rahmen einer Datenerhebung durch 1 für männlich und 2 für weiblich kodiert werden könnte.

1.2.2 Ordinalskalenniveau

Auf Ordinalskalenniveau sind die einzelnen Elemente der Skala Teil einer hierarchischen Rangreihe. Jede Merkmalsausprägung des empirischen Relativs entspricht einem Rang des nummerischen Relativs. Die Abstände zwischen den Rängen müssen dabei nicht gleich sein bzw. nicht den tatsächlichen Relationen im empirischen Relativ entsprechen. Stattdessen besteht nur eine Gleichheit in der Rangfolge zwischen empirischem und nummerischem Relativ.
Es lassen sich Aussagen wie *„a ist größer b“* treffen, jedoch nicht *„a ist doppelt so groß“* oder *„a ist ein drittel so groß wie b“*.
Der Rang beim Militär oder Schulleistungen sind typische Beispiele für empirische Merkmale, die sich gut durch eine Ordinalskala (z.B. Schulnoten von 1 bis 6) abbilden lassen.

1.2.3 Intervallskalenniveau

Das nächst höhere Skalenniveau ist das Intervallskalenniveau. Über eine hierarchische Relation hinaus sind die Elemente einer Intervallskala äquidistant, dass heißt gleichabständig zueinander. Die Abstände zwischen verschiedenen Elementen lassen sich dadurch exakt bestimmen.
Bezogen auf die Merkmalsausprägung ist der Abstand zwischen a und b genau so groß, wie der Abstand zwischen b und c.
Typische Beispiele für empirische Merkmale, die man auf Intervallskalenniveau messen kann, sind Zeit, Intelligenz mittels einer IQ-Skala oder Temperatur mittels der Celsius-Skala.

1.2.4 Verhältnisskalenniveau

Das Verhältnisskalen- oder auch Rationalskalenniveau unterscheidet sich von einem Intervallskalenniveau dadurch, dass auf diesem Niveau ein absoluter Nullpunkt existiert, der durch das empirische Merkmal vorgegeben ist und nicht verschoben werden darf. Dadurch sind Aussagen wie *„a ist doppelt so groß wie b“* oder *„a ist dreifach so groß wie c“* möglich.
Beispiele sind hierbei die Messung von Körpergröße in z.B. Zentimeter, Zeitdauer in Sekunden, Preis in Euro oder Gewicht in Kilogramm. Es würde keinen Sinn machen, etwas das nichts wiegt einen anderen Wert als den Wert 0 zuzuweisen.

1.2.5 Beziehungen zwischen den Skalenniveaus

Die vier Skalenniveaus unterscheiden sich darin, welche Relationen zwischen den Elementen des nummerischen Relativs bestehen. Mit ansteigendem Skalenniveau lassen sich mehr Aussagen über die Relation zwischen den Zahlen und damit zwischen den empirischen Elementen treffen.
Hierbei erbt die nächst höhere Skala jeweils die Eigenschaften der unter ihr liegenden Skalen. Auf Nominalniveau besteht lediglich die Relation: $a \neq b$ bzw. $a_1 = a_2$. Auf Ordinalniveau gelten die Relationen der Nominalskala genauso wie die Relation: $a < b < c$.
Ordinal- und Nominalskalen sind diskrete Skalen: Zwischen den einzelnen Werten der Skala besteht kein kontinuierlicher Übergang. Stattdessen besteht eine Trennung zwischen den Werten (Diskretheit von lat. discretus = getrennt).
Auf Intervall- und Verhältnisskalenniveau lassen sich nicht nur Aussagen über die Rangfolge treffen, sondern auch über die Größe der Abstände zwischen verschiedenen Werten, aufgrund der Äquidistanzeigenschaft dieser Skalen.
Diese beiden Skalen gehören zu den kontinuierlichen oder metrischen Skalen (auch Kardinalskalen genannt). Der Übergang zwischen den Zahlen ist fließend und nicht diskret (oder kategorial).
Daten, die auf einem hohen Skalenniveau vorliegen, können auf ein niedrigeres Skalenniveau „heruntergebrochen“ werden. In Abbildung 1.2 ist dies dargestellt: Die Zeit von zehn Läufern bei einem 3000 Meterlauf wurde gestoppt. Die Daten liegen damit auf dem Verhältnisskalenniveau vor. Diese Werte kann man in eine Datenreihe auf Ordinalskalenniveau umwandeln. Man verliert, durch den Verlust der Äquidistanz, einen Teil der Information.
Es ist nicht mehr möglich zu sehen, um wie viel besser die erstplatzierte Person 1 gegenüber der zweitplatzierten Person 2 war; lediglich, dass sie besser war.
Die Rangdaten lassen sich in einem zweiten Schritt in Daten auf Nominalniveau umwandeln, indem man die fünf Besten der Gruppe 1 und die fünf schlechtesten Läufer der Gruppe 2 zuordnet.
Hierbei hat man wiederum einen Teil der Information, die noch auf Ordinalniveau vorhanden war, verloren: Aus den Daten kann man nicht mehr herauslesen, ob Person 1 aus Gruppe 1 besser oder schlechter im 3000 Meterlauf war als Person 3 der Gruppe 1.

1.2.6 Erlaubte Transformationen

Auf den verschiedenen Skalenniveaus sind unterschiedliche mathematische Transformationen zulässig. Unzulässig sind Transformationen dann, wenn die Abbildungsgenauigkeit vom empirischen auf das nummerische Relativ verloren geht.

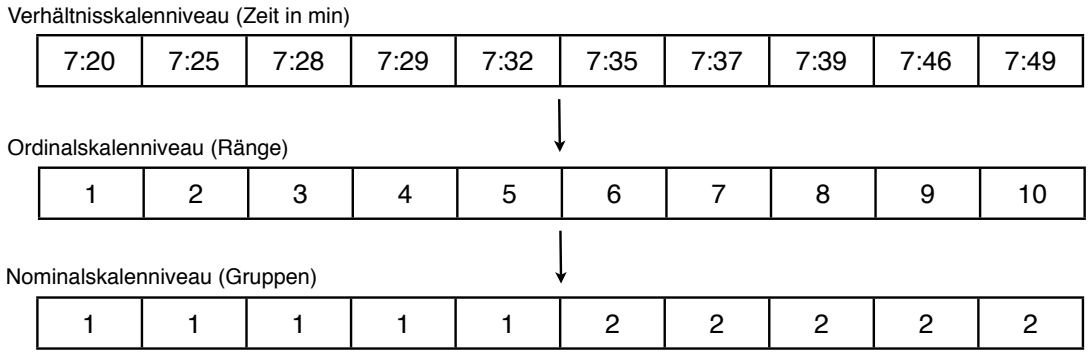


Abbildung 1.2: Beispiel für die Umwandlung von Daten aus einem höheren in ein niedrigeres Skalenniveaus

Auf Nominalniveau sind alle Transformationen erlaubt, solange die eine Zahl der einen empirischen Merkmalsausprägung entspricht und eine andere Zahl einer anderen Ausprägung.
Auf Ordinalniveau sind alle Transformationen erlaubt, welche die Rangreihe der Daten nicht verändert. So kann man beispielsweise alle Daten mit einer Zahl multiplizieren oder eine Zahl dazu addieren. Der erste Rang bleibt größer als der zweite Rang, auch wenn alle Ränge mit dem Faktor 3 multipliziert werden. Die neue Zahl für den ersten Rang wäre dann 3 und für den zweiten Rang wäre die neu zugeordnete Zahl 6. Dabei gilt immer noch: Rangwert 3 $<$ Rangwert 6.
Auf Intervallskalenniveau sind Transformationen der Form $y = bx + a$ erlaubt. Es darf also multipliziert, dividiert, addiert und subtrahiert werden. All diese Transformationen verletzen nicht die Äquidistanz der Werte. Die Gleichabständigkeit bleibt also nach Transformationen dieser Art erhalten.
Auf Verhältnisskalenniveau sind Transformationen der Form $y = bx$ zulässig. Addition und Subtraktion ist auf diesem Niveau nicht mehr zulässig, da sich dadurch der absolute Nullpunkt der Skala verschieben würde.

1.2.7 Bedeutung des Skalenniveaus

Je nachdem auf welchem Niveau die Daten sich befinden die ich auswerten will, habe ich verschiedenartige statistische Methoden zur Verfügung. Insbesondere in der Inferenzstatistik ist es von großer Bedeutung auf welchem Skalenniveau die Daten vorliegen, da viele inferenzstatistische Methoden ein bestimmtes Skalenniveau voraussetzen.

1.3 Wozu eigentlich Statistik?

Statistische Methoden sind für die psychologische Forschung in erster Linie Werkzeuge, um Beobachtungen systematisch zu untersuchen. Mithilfe bestimmter Verfahren lassen sich Unterschiede zwischen Gruppen statistisch belegen oder Zusammenhänge psychologischer Merkmale ausrechnen.
Angenommen wir wollen untersuchen, ob Intelligenz mit einem höheren Lebensstandard einhergeht. Dabei müssen wir zunächst definieren, was mit Intelligenz und höherer Lebensstan-

dard gemeint ist. Der nächste Schritt wäre diese beiden Merkmale messbar zu machen (auch operationalisieren genannt). Wir könnten uns entscheiden Lebensstandard durch monatliches Einkommen und Intelligenz durch einen standardisierten Intelligenztest zu operationalisieren. Nun könnten wir mithilfe statistischer Methoden ausrechnen, wie groß der Zusammenhang zwischen diesen beiden Merkmalen, oder statistisch ausgedrückt: Variablen, ist.
Dieser Zusammenhang wird höchstwahrscheinlich nicht perfekt sein, dass heißt ein hoher Wert im Intelligenztest muss nicht zwangsläufig mit einem hohen Einkommen einhergehen.
Zusammenhänge psychologischer Variablen sind meist nie perfekt, was man dadurch erklären kann, dass der Untersuchungsgegenstand der Psychologie, also der Mensch, sehr komplex ist und man diese Komplexität schwer als Ganzes in einem Experiment erfassen und durch Variablen abbilden bzw. messen kann.
Während beispielsweise in der Physik oft deterministische oder funktionale Beziehungen gefunden werden, handelt es sich bei psychologischen Merkmalen meist um stochastische Beziehungen.
Der IQ determiniert (zum Glück) nicht das Einkommen oder die Stellung in einer Gesellschaft. Allerdings hat man schon an vielen Stellen ein stochastische Beziehung dieser zwei Variablen finden können: Es gibt einen Zusammenhang von IQ und Einkommen. Dieser Zusammenhang ist allerdings durch eine (recht große) Störgröße überlagert, sodass man nie perfekt vorhersagen kann, welches monatliche Einkommen jemand bei IQ-Wert x bekommt oder bekommen wird.
Die Statistik ermöglicht es zu entscheiden, wie empirisch bedeutsam der Zusammenhang zweier Merkmale tatsächlich ist und wie groß die Störgröße ist, die diesen Zusammenhang beeinträchtigt. Insofern ist die Statistik ein Mittel zum Zweck, eine Methode, die dabei hilft inhaltliche Fragen wissenschaftlich fundiert zu klären.

1.4 Statistische Analysen mit dem PC

Die meisten statistischen Analysen übernimmt heutzutage der Computer. Es gibt verschiedene Programme und Programmpakete, welche die Möglichkeit bieten Daten zu analysieren und grafisch darzustellen.
Hier eine kurze Beschreibung der verbreitetsten Statistikprogramme:

Microsoft Excel Das Programm Excel eignet sich vor allem zur Datenverwaltung und Aufbereitung, sowie zur Grafikerstellung. Höhere statistische Analysen sind dagegen nur bedingt möglich. Eine kostenlose Alternative zu Excel ist das Programm Calc, das zu dem Open Source Software Paket OpenOffice.org gehört. Die Bedienung von Calc und die Möglichkeiten damit sind sehr ähnlich zu denen von Excel.

SPSS Das unter Windows, Mac OS und Linux verfügbare Programm SPSS ist für alle gängigen statistischen Analyseverfahren geeignet und im Gegensatz zu spezialisierter Statistiksoftware noch relativ benutzerfreundlich. SPSS (ursprünglich die Abkürzung für Statistical Package for Social Science) ist modular aufgebaut. Das bedeutet, dass das Programm aus einer Basisversion besteht und Zusatzmodule für spezielle Analyseverfahren extra dazu gekauft werden müssen. Es gibt eine Studentenversion für knapp 60 Euro, die in ihrem Funktionsumfang allerdings stark eingeschränkt ist.

R Eine kostenlose Alternative zu SPSS und anderen kommerziellen Statistikprogrammen ist R. Das unter der GNU-Lizenz veröffentlichte Software-Paket ist sehr flexibel, bietet sehr viele Analysemöglichkeiten und ist für die Grafikerstellung sehr gut geeignet. Dadurch das der Quellcode des Programms offen für jeden zur Verfügung steht gibt es sehr viele Zusatzmodule für speziellere Verfahren, die man kostenlos dazu installieren kann.
Der Nachteil von R liegt in der komplizierten Bedienung: R verfügt über keine grafische Oberfläche. Stattdessen müssen alle Befehle, die auf einer programmeigenen Skriptsprache basieren, eigenständig in das R-Konsolenfenster eingetippt werden.

MatLab MATrix LABoratory (kurz: MatLab) ist eine sehr mächtige Analysesoftware, die nicht nur für statistische, sondern generell mathematische Probleme und Aufgaben genutzt wird. Dateieingabe und Verarbeitung erfolgen dabei teils über eine grafische Oberfläche, teils wie bei R über Befehle, die in ein Konsolenfenster eingetippt werden müssen. Durch so genannte Toolboxen lässt sich der Funktionsumfang der Basisversion von MatLab, ähnlich wie bei SPSS, erweitern.

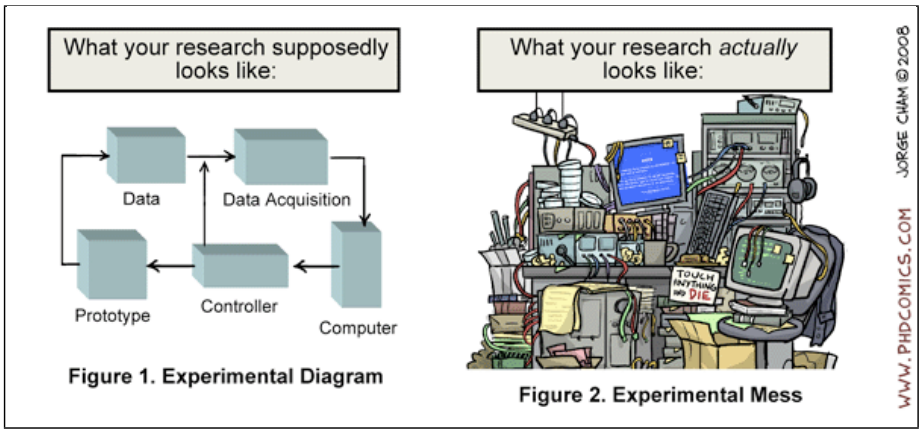


Figure 1. Experimental Diagram

Figure 2. Experimental Mess

1.5 Mathematische Symbole

1.5.1 Summenzeichen

Mithilfe des Summenzeichens lassen sich Summenberechnungen über eine lange Zahlenreihe abkürzen. Das Summenzeichen besteht aus dem großen griechischem Buchstaben Sigma. Ober- und unterhalb des Summenzeichens befinden sich so genannte Indizes.
Der Index unterhalb des Sigmas definiert die untere Summationsgrenze. Dieser Wert gibt an, ab welchem Element einer Zahlenreihe aufsummiert werden soll.
Der Index oberhalb des Sigmas definiert die obere Summationsgrenze. Dieser Wert zeigt an, bis zu welchem Wert einer Zahlenreihe aufsummiert werden soll.
Der Laufindex, der durch einen kleinen lateinischen Buchstaben symbolisiert wird, nimmt fortlaufend die Zahl zwischen Ober- und Untergrenze der Datenreihe an. In Formel 1.1 ist ein Summenzeichen dargestellt, wobei n die obere, m die untere Summationsgrenze und i der Laufindex ist.

Schreibweise und Beispiele:

$$\sum_{i=m}^{n} x_i \tag{1.1}$$

$$\sum_{i=3}^{8} x_i = x_3 + x_4 + x_5 + x_6 + x_7 + x_8$$

Angenommen folgende Daten seien gegeben:

$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$	$\mathbf{x_4}$	$\mathbf{x_5}$	$\mathbf{x_6}$	$\mathbf{x_7}$	$\mathbf{x_8}$	$\mathbf{x_9}$	$\mathbf{x_{10}}$
12	11	19	22	4	6	22	32	15	1

Das Aufsummieren dieser Daten lässt sich folgendermaßen abkürzen:

$$\sum_{i=1}^{n} x_i = 12 + 11 + 19 + 22 + 4 + 6 + 22 + 32 + 15 + 1 = 144$$

Das n oberhalb des Sigmas kodiert die Anzahl der Elemente einer Datenreihe und nimmt für obiges Beispiel den Wert 10 an, da die Datenreihe aus $n = 10$ Werten besteht.

Weitere Beispiele anhand dieser Datenreihe:

$$\sum_{i=5}^{9} x_i = 4+6+22+32+15 = 79$$

$$\sum_{i=5}^{8} \frac{x_i}{n} = \frac{4}{10}+\frac{6}{10}+\frac{22}{10}+\frac{32}{10} = 6,4$$

$$\sum_{i=1}^{3} x_i + 1 = 12+11+19+1 = 43$$

$$\sum_{i=4}^{6} (5 \cdot x_i) = 5 \cdot 22 + 5 \cdot 4 + 5 \cdot 6 = 160$$

1.5.2 Fakultät

Die Fakultät ist eine mathematische Funktion, mit deren Hilfe man die Multiplikation einer aufsteigende Reihe natürlicher Zahlen abkürzen kann.

Schreibweise und Beispiele:

$$n! \tag{1.2}$$

$$2! = 1 \cdot 2 = 2$$

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

$$4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$$

Für $n!$ gilt:

$$n! = (n-1)! \cdot n \tag{1.3}$$

Die Fakultät von n nimmt bei größer werdendem n dramatisch zu:

n	1	2	3	4	5	6	7	8	9	10
n!	1	2	6	24	120	720	5040	40320	362880	3628800

1.A Aufgaben

inhaltliche Aufgaben

1. Was bedeutet messen?
2. Was kann überhaupt gemessen werden?
3. Was ist eine Skala?
4. Welche Skalenarten gibt es?
5. Was ist ein nummerisches, was ein empirisches Relativ?
6. Welche Abbildungsarten vom empirischen auf das nummerische Relativ gibt es? Was unterscheidet sie?
7. Was bedeutet Äquidistanz?
8. Finde ein Beispiel für einen stochastischen und einen funktionalen Zusammenhang.
9. Worin besteht der Nachteil, wenn man Daten von einem höheren in ein niedrigeres Skalenniveau umwandelt?
10. Aus welchem griechischem Buchstaben besteht das Summenzeichen? Für was stehen die Indizes oberhalb und unterhalb des Summenzeichens?
11. Auf welchem Skalenniveau werden folgende Merkmale gemessen?
    a) Gewicht
    b) Berufsgruppen
    c) Länge
    d) Grad in Kelvin
    e) Promille Alkohol im Blut
    f) Platzierung bei einem Sportwettbewerb
12. Welche der folgenden Transformationen sind zulässig?
    a) Addition eines Wertes zu jedem Datenpunkt auf Intervallniveau
    b) Jeden Datenpunkt mit dem Kehrwert multiplizieren auf Ordinalniveau
    c) Multiplikation und Addition eines Wertes zu jedem Datenpunkt auf Nominalniveau
    d) Division und Subtraktion eines Wertes zu jedem Datenpunkt auf Verhältnisniveau
    e) Quadrieren der Werte auf Ordinalniveau
    f) Wurzel ziehen der Werte auf Intervallniveau

praktische Aufgaben

1. Berechne folgende Fakultäten: $7!$, $(2 \cdot 3)!$, $(5 - 1)!$

2. Gegeben sind folgende Daten:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
9	17,2	18	0	-17	3,75	8	4,4	100	3

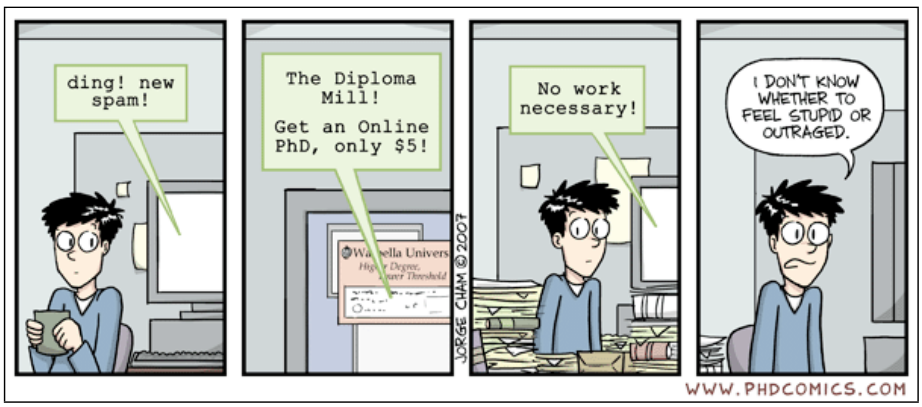
Berechne folgende Summen:

$$\sum_{i=2}^{8} x_i$$

$$\sum_{i=1}^{5} (x_1 + x_i)$$

$$\sum_{i=1}^{5} x_i \cdot \sum_{i=6}^{10} x_i$$

$$\frac{1}{n} \sum_{i=1}^{n} x_i$$



2
Deskriptivstatistik

Deskriptive Statistik, oder beschreibende Statistik, ist ein Teilgebiet der Statistik, welches sich mit der Datenaggregation und Datendarstellung befasst. Ziel ist es eine Datenmenge anhand statistischer Kennwerte zu beschreiben und mithilfe von verschiedenen Darstellungsformen grafisch aufzubereiten.

2.1 Maße der zentralen Tendenz

Eine Gesamtheit von Beobachtungen, ausgedrückt in einer Reihe von Daten, lässt sich schwer überschauen. Maße der zentralen Tendenz versuchen die Richtung oder Gemeinsamkeit der Daten in einer Kenngröße zusammenzufassen. Man unterscheidet im Wesentlichen drei Kennwerte der zentralen Tendenz, mit deren Hilfe man eine Datenreihe charakterisieren kann: Modalwert (auch Modus genannt), Median und arithmetisches Mittel.

2.1.1 Modalwert

Der Modalwert (abgekürzt: Mo) ist der Wert einer Datenreihe, der am häufigsten vorkommt. In einer stetigen Zufallsverteilung ist es der Wert, mit der höchsten Wahrscheinlichkeitsdichte. In der folgenden Datenreihe ist der Modalwert $Mo = 5$. Er kommt insgesamt sechs mal vor.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
5	3	6	8	6	5	4	5	1	6	5	5	7	4	5

2.1.2 Median

Der Median (abgekürzt: Md) ist der Wert, der eine Verteilung oder Datenreihe in zwei gleich große Teile zerlegt. Das bedeutet, dass unter- und oberhalb des Medians immer 50% der Werte liegen. Der Medianort errechnet sich über:

$$Md_{ort} = \frac{n+1}{2} \tag{2.1}$$

Der Wert, an dieser Stelle einer aufsteigend geordneten Datenreihe, ist der Median.
Um den Median der Datenreihe aus 2.1.1 zu errechnen, muss diese zunächst aufsteigend sortiert werden (siehe unten).

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
1	3	4	4	5	5	5	5	5	5	6	6	6	7	8

Der Medianort ergibt sich nun aus $Md_{ort} = \frac{15+1}{2} = 8$. Demnach nimmt der Median den Wert des Datenpunkts an achter Stelle an: $Md = x_8 = 5$.
Der Median besitzt die Eigenschaft, dass die absoluten Abweichungen aller Werte vom Median ein Minimum ergibt:

$$\sum_{i=1}^{n} |x_i - Md| = Min. \tag{2.2}$$

Man findet keinen anderen Wert, bei dem die Summe der absoluten Abweichungen kleiner ist als beim Median. Egal welchen Wert man anstelle des Medians in obige Formel einsetzt, das Ergebnis wird größer sein.

Bei einer Menge von Daten mit gerader Anzahl an Elementen liegt der Medianort zwischen zwei benachbarten Elementen. In der unten stehenden Datensequenz sind, statt 15 Elementen, wie oben, nur noch 14 vorhanden.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
1	3	4	4	5	5	6	7	7	7	8	9	9	8

Damit ergibt sich für den Medianort: $Md_{ort} = \frac{14+1}{2} = 7,5$. Der Medianwert liegt demnach zwischen dem 7. und 8. Element der Datenreihe. In diesem Fall wird der Mittelwert von diesen beiden benachbarten Werten errechnet. Im obigen Beispiel wäre der Medianwert damit $Md = \frac{6+7}{2} = 6,5$.

2.1.3 Arithmetisches Mittel

Das arithmetische Mittel (abgekürzt: M oder $\bar{x}$) ist das wohl prominenteste Maß der zentralen Tendenz. Es ergibt sich aus der Aufsummierung der Daten, relativiert an der Anzahl der Datenpunkte:

$$M = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{2.3}$$

Was man durch obige Formel erhält ist ein Wert, der dem eines Datenpunktes entsprechen würde, würde man die Summe der Datenreihe auf jedes Element gleich verteilen:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	**Summe**
0	1	2	3	4	5	6	21

$$Mittelwert = \bar{x} = \tfrac{1}{7}(0+1+2+3+4+5+6) = 3$$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	**Summe**
3	3	3	3	3	3	3	21

Das arithmetisches Mittel besitzt die Eigenschaft, dass die Summe der quadrierten Abweichungen aller Werte vom Mittelwert ein Minimum ergibt:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = Min. \tag{2.4}$$

Des Weiteren ist die Summe aller Abweichungen vom Mittelwert immer null:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \tag{2.5}$$

Gewichtetes arithmetisches Mittel

Das gewichtete arithmetische Mittel (kurz: GAM) ist eine Sonderform des Mittelwerts, die dann zum Einsatz kommt, wenn man den gemeinsamen Mittelwert über mehrere verschieden lange Datenreihen berechnen will.
Jeder Mittelwert einer Datenreihe wird mit der Anzahl der Datenpunkte dieser Datenreihe multipliziert (also gewichtet). Die Summe daraus wird an der Summe aller Datenpunkte relativiert:

$$GAM = \frac{\sum_{i=1}^{k}(n_i \cdot \bar{x}_i)}{\sum_{i=1}^{k} n_i} \tag{2.6}$$

Der Index k über dem Summenzeichen in der obigen Formel steht für die Anzahl der Datenreihen.

Beispiel:

Die unten stehenden Datenreihen bestehen aus verschieden vielen Datenpunkten:

$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$	$\mathbf{x_4}$	$\mathbf{x_5}$	$\mathbf{x_6}$	$\mathbf{x_7}$	$\bar{\boldsymbol{x}}$
1	3	1	9	6	9	4	4,71

$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$	$\mathbf{x_4}$	$\mathbf{x_5}$	$\bar{\boldsymbol{x}}$
5	7	7	6	6	6,2

$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$	$\bar{\boldsymbol{x}}$
12	17	13	14

Das gewichtete arithmetische Mittel aus den drei Datenreihen berechnet sich folgendermaßen:

$$GAM = \frac{(7 \cdot 4,71) + (5 \cdot 6,2) + (3 \cdot 14)}{7+5+3} = 7,06$$

Würde man keine Gewichtung vornehmen, sondern einfach einen Mittelwert aus den Mittelwerten berechnen, würde das Ergebnis in diesem Fall größer ausfallen:

$$\bar{x}_{\bar{x}} = \frac{4,71 + 6,2 + 14}{3} = 8,3 > GAM$$

2.1.4 Welches Maß der zentralen Tendenz ist das Richtige?

Für welches Maß der zentralen Tendenz man sich entscheidet, um eine Reihe von Daten zu charakterisieren, hängt zunächst vom Skalenniveau ab, auf dem sich die Daten befinden. Bei kategorialen Daten (Nominalniveau) macht es wenig Sinn, einen Median oder einen Mittelwert zu berechnen.
Auf Ordinalniveau ist der Mittelwert ebenfalls kein adäquates Maß zur Charakterisierung der zentralen Tendenz, da der Mittelwert die Äquidistanz der Daten voraussetzt, was bei rangskalierten Daten nicht der Fall ist. In Tabelle 2.1 sind die angemessenen zentralen Maße auf den jeweiligen Skalenniveaus aufgelistet.

Tabelle 2.1
sinnvolle Maße der zentralen Tendenz auf den verschiedenen Skalenniveaus

Skalenniveau	Maße der zentralen Tendenz
Nominalniveau	Modalwert
Ordinalniveau	Modalwert, Median
Intervallniveau	Modalwert, Median, Arithmetisches Mittel
Verhältnisniveau	Modalwert, Median, Arithmetisches Mittel

Auf dem Intervallniveau können alle drei Methoden zur Berechnung eines Maß der zentralen Tendenz genutzt werden. In der Beispieldatenreihe aus Kapitel 2.1.1 und 2.1.2 ergibt sich für Modalwert und Median der selbe Wert. Auch der Mittelwert dieser Datensequenz ist gleich dem Median und dem Modalwert. In diesem Fall würde es keinen Unterschied machen für welches Maß der zentralen Tendenz man sich, zur Beschreibung der Daten, entscheidet.
Dies liegt an der Verteilung der Daten: Sie folgen einer ganz bestimmten Art der Verteilung, die man Normalverteilung nennt. Eine grafische Veranschaulichung dieser Verteilungsform findet sich in Abbildung 2.1. Nähere Information zu den Eigenschaften dieser Verteilungsform werden in Kapitel 4 besprochen. Dass Median, Modal- und Mittelwert auf den selben Punkt fallen, liegt an der Symmetrie, die diese Verteilungsform auszeichnet. Ist eine solche Symmetrie in der Datenverteilung nicht gegeben, kommt man bei der Berechnung der drei Maße zu unterschiedlichen Ergebnissen.
Insbesondere bei schiefverteilten Daten muss man überlegen, für welches Maß man sich entscheidet, da dabei systematische Unterschiede zwischen den drei Werten entstehen.

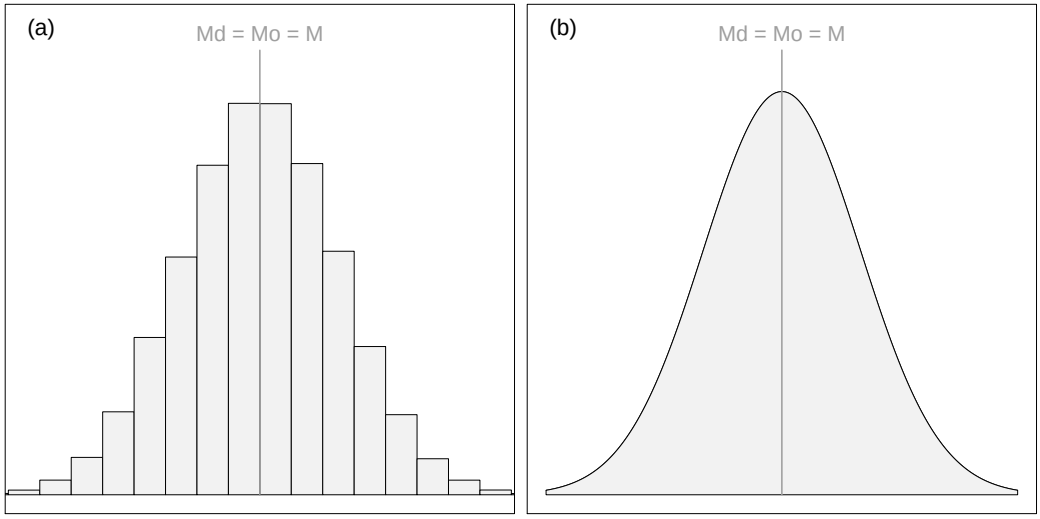


Abbildung 2.1: Lage der Maße der zentralen Tendenz bei (a) einer diskreten, normalverteilten Variable und (b) einer stetigen, normalverteilten Variable

In Abbildung 2.2 sind Daten dargestellt die, wie in Abbildung 2.2.a, entweder rechtsschief bzw. linkssteil oder, wie in Abbildung 2.2.b, linksschief bzw. rechtssteil verteilt sind.
Die Begriffe der Schiefe und Steilheit helfen eine Verteilung zu beschreiben: Eine linksschiefe Verteilung besitzt links vom Gipfel eine geringere Steigung als rechts vom Gipfel. Bei einer rechtsschiefen Verteilung ist dies umgekehrt: links vom Gipfel ist die Steigung steiler als rechts davon.
Bei einer rechtsschiefen Verteilung (wie in Abbildung 2.2.a) gilt: $M > Md > Mo$.
Bei einer linksschiefen Verteilung (Abbildung 2.2.b) ist dies umgekehrt: $Mo > Md > M$.

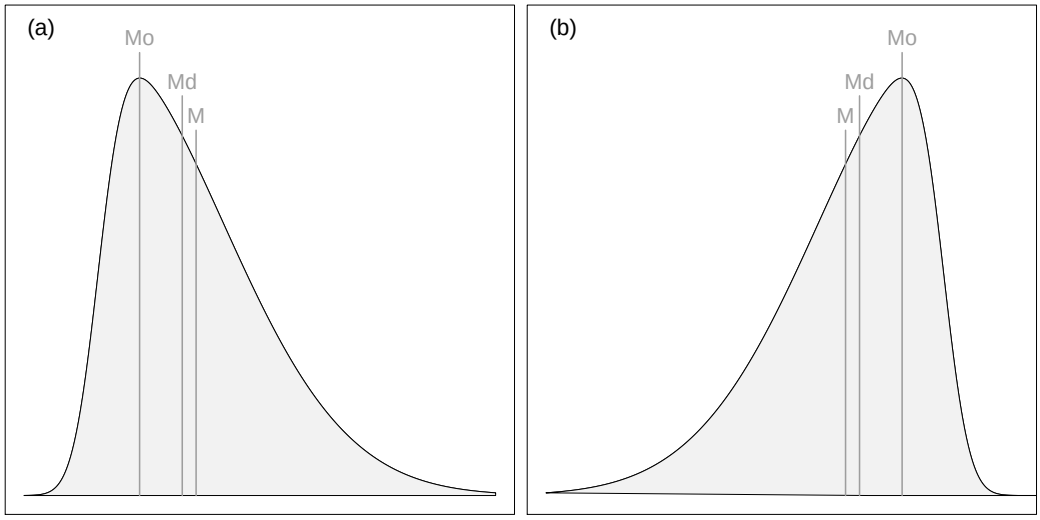


Abbildung 2.2: Lage der Maße der zentralen Tendenz bei (a) einer stetigen, rechtsschiefen Variable und (b) einer stetigen, linksschiefen Variable

Eigenschaft des Medians

Unabhängig von der Verteilungsform ist der Median, im Gegensatz zum arithmetischen Mittel, resistent gegenüber Ausreißern. Als Ausreißer bezeichnet man einen Wert, der sich extrem von den anderen Werten unterscheidet. In der unteren Datensequenz wäre dies der Wert an vierter Stelle.

x_1	x_2	x_3	x_4	x_5	x_6	x_7
2	3	4	6500	7	0	1

Der Median dieser Daten liegt bei $Md = 3$, während der Mittelwert stark von dem Ausreißer beeinflusst wird und dadurch viel höher ausfällt: $\bar{x} = 931$. Je kleiner die Datenreihe und je weiter der Ausreißer von den restlichen Daten entfernt, desto größer wird der Mittelwert von dem Ausreißer beeinflusst, während der Median konstant bleibt. Statistisch müssen Daten, wählt man das arithmetische Mittel als Maß der zentralen Tendenz, von Ausreißern bereinigt werden, was beim Median nicht notwendig ist.
Allerdings muss eine fundierte inhaltliche Begründung vorliegen, will man erhobene Daten von den weiteren statistischen Analysen und damit auch inhaltlichen Schlussfolgerungen ausschließen.

2.A Aufgaben

inhaltliche Aufgaben

1. Worin liegt der Vorteil von einem Maß der zentralen Tendenz?
2. Wann ist ein Maß der zentralen Tendenz nicht angebracht?
3. Warum muss man eine inhaltliche Begründung bei einem empirischen Datensatz haben, bevor man Ausreißer vor weiteren Analysen ausschließt?
4. Überlege dir jeweils ein Merkmal das, würde man es messen, höchstwahrscheinlich (a) rechtsschief oder (b) normalverteilt ist.
5. Welche Eigenschaften besitzt der Median? Welche das arithmetische Mittel?
6. Warum macht es keinen Sinn für ordinalskalierte Daten einen Mittelwert zu berechnen?
7. Für welches Maß der zentralen Tendenz würdest du dich bei Daten folgender Merkmale entscheiden?
   a) Gewicht (in kg)
   b) Berufsgruppen
   c) Länge (in cm)
   d) Temperatur (in Grad Kelvin)
   e) Alkohol im Blut (in Promille)
   f) Leistung bei einem Sportwettbewerb (Platzierung)

praktische Aufgaben

1. a) Berechne jeweils Median, Modalwert und arithmetisches Mittel der folgenden zwei Datenreihen.
   b) Was kannst du über die Verteilung der beiden Datenreihen sagen?
   c) Was ist das gewichtete arithmetische Mittel aus beiden Datenreihen?

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
3	3	7	1	4	4	7	5	9	6	4	4	3

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
1	8	4	0	5	0	6	3	5	2	9	6	5	4	1

2.2 Dispersionsmaße

Werden Statistiken in der Öffentlichkeit vorgestellt, werden meist nur zentrale Tendenzen, oftmals Mittelwerte, berichtet. Als Konsument solcher Statistiken ist es dadurch unmöglich zu beurteilen, wie gut dieser Mittelwert tatsächlich alle Werte repräsentiert.
Je mehr sich die einzelnen Werte unterscheiden und je größer die Abweichung der einzelnen Werte von dem Maß der zentralen Tendenz, desto größer ist die so genannte Streuung (oder Dispersion).
Mit wachsender Streuung eines Datensatzes nimmt die Aussagekraft eines Maß der zentralen Tendenz ab (und umgekehrt). Berichtet man von einem Datensatz also nur ein zentrales Tendenzmaß, so ist es für den Leser unmöglich zu entscheiden, wie gut dieser Kennwert die Werte des Datensatz abbildet.
Um dies beurteilen zu können, müssen immer Dispersionsmaße berechnet und angegeben werden, die versuchen den Streuungsgrad der Werte zu erfassen.
Die gängigsten Dispersionsmaße sind der Range, der Interquartilbereich und die Standardabweichung bzw. Varianz.

2.2.1 Range

Der Range ist, wie der Name schon sagt, die Spannweite, also der absolute Bereich zwischen dem die Werte variieren. Er errechnet sich über die Differenz aus dem größten und dem kleinsten Messwert:

$$R = x_{max} - x_{min} \tag{2.7}$$

Als Dispersionsmaß ist der Range oft ungeeignet, da im Grunde nur zwei Werte berücksichtigt werden: Der größte und der kleinste Wert. Dadurch ist der Range sehr anfällig gegenüber Ausreißern.

2.2.2 Interquartilbereich

Ein Quartil ist ein Bereich, in dem sich ein viertel der gemessenen Werte befindet. Eine Datenreihe kann man also in vier Quartile aufteilen. Der Interquartilbereich (kurz: IQR) ist der Range bzw. die Spannweite, in dem sich die mittleren zwei Quartile, also die mittleren 50% der Daten befinden.
In Abbildung 2.3 sind aufsteigend sortierte Daten auf Ordinalniveau mit den Quartilgrenzen abgebildet.
Der IQR Bereich ist der Range der mittleren 50% der Daten.

$$IQR = Q_3 - Q_1 \tag{2.8}$$

In den Daten aus Abbildung 2.3 errechnet sich ein Interquartilbereich von $IQR = 5,5 - 2,5 = 3,5$. Das bedeutet, dass innerhalb der mittleren 50% die Daten mit einer Spannweite von $3\ 1/2$ Rangplätzen streuen.
Für die Berechnung der Quartilorte gibt es zwei verschiedene Methoden. Entweder man definiert den Wertebereich ober- und unterhalb des Medians als Teilmenge, von denen man jeweils

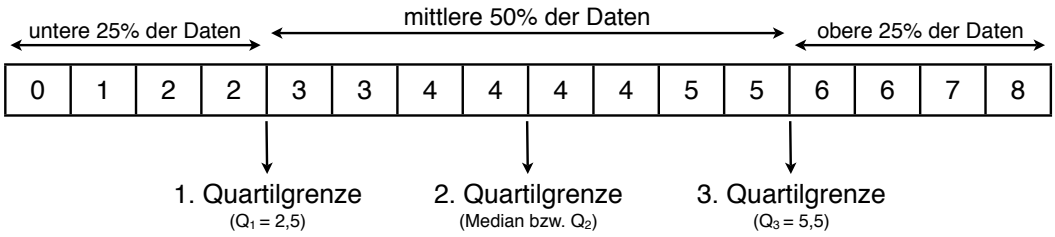


Abbildung 2.3: Quartile einer aufsteigend sortierten Rangdatenreihe

wieder eigene Mediane berechnet. Fällt der Median dabei auf genau einen Wert der Reihe, so wird dieser Wert weder dem ersten, noch dem zweiten Teilkollektiv zugeordnet (siehe Abbildung 2.4).

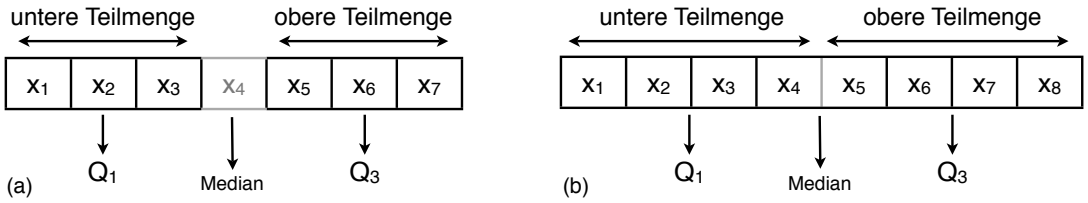


Abbildung 2.4: Quartilgrenzen bei (a) einer Datenreihe, bei der der Median genau auf einen Datenpunkt fällt und (b) bei der der Median zwischen zwei Datenpunkte fällt

In Abbildung 2.4.a errechnet sich der Q1-Ort aus: $\frac{3+1}{2} = 2$. Der Datenpunkt auf dem der Median fällt wird weder dem oberen, noch dem unteren Teilkollektiv zugerechnet. In Abbildung 2.4.b errechnet sich der Q1-Ort über: $\frac{4+1}{2} = 2,5$ (ebenso der Q3-Ort für die obere Teilmenge). Ein alternative Berechnung von Q1 und Q3-Ort ist mit folgenden Formeln möglich.

$$Q_{1(ort)} = 0,25 \cdot (n+1) \qquad Q_{3(ort)} = 0,75 \cdot (n+1) \tag{2.9}$$

Der Nachteil an diesem alternativen Weg ist, dass je nach Länge der Datenreihe Quartile eine unterschiedliche Anzahl an Elementen umfassen. Der Vorteil liegt in einer höheren Genauigkeit.
Viele Statistikprogramme berechnen Quartile nach diesen Formeln.

2.2.3 Varianz und Standardabweichung

Die Varianz (kurz: Var oder s^2) ist definiert als die mittlere quadrierte Abweichung vom Mittelwert:

$$s^2 = \frac{\sum_{i=1}^{n} (\bar{x} - x_i)^2}{n} \tag{2.10}$$

Durch die Quadrierung besitzt die Varianz eine andere Einheit (quadrierte Werte) als z.B. der Mittelwert. Dadurch ist die Varianz schwer zu interpretieren. Dem kann man abhelfen, indem man die Wurzel aus der Varianz berechnet und dadurch die Quadrierung rückgängig macht. Dieser Wert wird als Standardabweichung (kurz: s oder SD für standard deviation) bezeichnet und geht auf den Naturforscher Sir Francis Galton zurück.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(\bar{x} - x_i)^2}{n}} \tag{2.11}$$

Beispiel:

In Tabelle 2.2 sind zehn Daten, deren Mittelwert und die Berechnung der Varianz über die Mittelung der quadrierten Abweichungen dargestellt. Die Standardabweichung dieser Datenreihe wäre $s = \sqrt{6,6} = 2,57$.

Tabelle 2.2
Beispielhafte Berechnung der Varianz einer Datensequenz

	Werte	**quadrierte Abweichung**
$\mathbf{x_1}$	10	$(10-5)^2 = 25$
$\mathbf{x_2}$	7	$(7-5)^2 = 4$
$\mathbf{x_3}$	7	$(7-5)^2 = 4$
$\mathbf{x_4}$	5	$(5-5)^2 = 0$
$\mathbf{x_5}$	5	$(5-5)^2 = 0$
$\mathbf{x_6}$	5	$(5-5)^2 = 0$
$\mathbf{x_7}$	5	$(5-5)^2 = 0$
$\mathbf{x_8}$	3	$(3-5)^2 = 4$
$\mathbf{x_9}$	3	$(3-5)^2 = 4$
$\mathbf{x_{10}}$	0	$(0-5)^2 = 25$
	$\bar{x} = 5$	$s^2 = 6,6$

Warum quadriert man?

Die Logik der Varianz ist nicht gerade intuitiv nachvollziehbar. Insbesondere fragt man sich, warum die quadrierten Abweichungen gemittelt werden und nicht die absoluten Abweichungen, wie man es bei der Berechnung der mittleren Abweichung macht (kurz: MD für mean deviation - siehe unten).

$$\text{mittlere Abweichung} = MD = \frac{\sum_{i=1}^{n}|\bar{x} - x_i|}{n} \tag{2.12}$$

Immerhin könnte der Wert aus obiger Formel leichter interpretiert wären: Es wäre die durchschnittliche Abweichung jedes Datenpunkts vom Mittelwert.
Das man stattdessen die quadrierten Abweichungen errechnet und mittelt, hängt mit einer Verteilungsannahme von den Variablen oder Merkmalen, die man untersuchen will, zusammen. Man geht in den meisten Fällen davon aus, dass die Messdaten eines Merkmals normalverteilt sind. Die Form einer Normalverteilung ist in Abbildung 2.1 aufgezeichnet.
Die Normalverteilung besitzt zwei Wendepunkte, die den Wechsel von einem exponentiellen Wachstum zu einem begrenzten Wachstum charakterisieren und an dessen Stelle die Steigung maximal ist.
Zunächst nehmen die Werte also um einen bestimmten Prozentsatz stetig zu (exponentielles Wachstum), bis ein Wendepunkt erreicht wird, ab dem dieses prozentuale Wachstum wieder stetig abnimmt und ein Maximum erreicht (am Mittelwert).
An diesem Punkt beginnt ein begrenzter Abfall, gefolgt von einem exponentiellen Abfall am zweiten Wendepunkt rechts des Gipfels.
Diesem quadratischen oder exponentiellen Verlauf trägt die Quadrierung in der Varianzformel rechnung. Die Standardabweichung ist nun genau der Abstand, den der Mittelwerte von dem Wendepunkt auf der linken und dem Wendepunkt auf der rechten Seite der Normalverteilung hat.

Eigenschaft der Varianz

Aufgrund der Quadrierung der Abweichungen fallen bei der Varianz große Abweichungen stärker ins Gewicht als kleine Abweichungen (siehe Tabelle 2.2, Datum x_1 und x_{10} gegenüber x_3 oder x_8).
Dies kann dann zum Problem werden, wenn bei kleinen Datensätzen Werte mit extrem großer Abweichung vom Mittelwert auftreten. Die Varianz ist also, genau wie der Mittelwert, relativ sensibel gegenüber Ausreißern.

Standardabweichung und schiefverteilte Variablen

Die Standardabweichung ist auch dann ein sinnvolles Streuungsmaß, wenn die Daten nicht normalverteilt sind. Dies konnte durch die Tschebyscheff-Ungleichung gezeigt werden. Solange die Verteilung eine zentrale Tendenz oder einen Erwartungswert besitzt (also einen Gipfel), gilt unabhängig der Form der Verteilung: Zwischen $\bar{x} - 2 \cdot s$ und $\bar{x} + 2 \cdot s$ liegen mindestens $3/4$ der Werte.

2.2.4 Welches Dispersionsmaß ist das Richtige?

Für welchen Kennwert zur Charakterisierung der Streuung einer Datensequenz man sich entscheidet, hängt wiederum von dem Skalenniveau und von dem Maß der zentralen Tendenz, für das man sich entschieden hat, ab.
Vorraussetzung für die Berechnung von Varianz bzw. Standardabweichung ist, wie beim Mittelwert, zum Beispiel die Äquidistanz innerhalb der Metrik, die erst ab Intervallskalenniveau gegeben ist.
In Tabelle 2.3 sind sinnvolle Dispersionsmaße, in Abhängigkeit des Skalenniveaus, aufgelistet.

Tabelle 2.3
sinnvolle Dispersionsmaße auf den verschiedenen Skalenniveaus

Skalenniveau	Dispersionsmaße
Nominalniveau	Range
Ordinalniveau	Range, Interquartilbereich
Intervallniveau	Range, Interquartilbereich, Varianz
Verhältnisniveau	Range, Interquartilbereich, Varianz

Neben dem Skalenniveau ist die Wahl des zentralen Maßes entscheidend. So ergänzt die Standardabweichung oder Varianz sinnvoll das arithmetische Mittel, als Maß der zentralen Tendenz, da der Mittelwert die Eigenschaft besitzt, dass die quadrierten Abweichungen (also das, was in die Berechnung der Varianz eingeht) ein Minimum ergibt (vgl. Formel 2.4).
Bei Angabe des Medians ist der IQR-Bereich ein sinnvolles Dispersionsmaß, da hier die Spannweite der Ränge innerhalb der mittleren 50% einer Datenreihe berechnet wird, in dessen Mitte sich wiederum der Median befindet.
Bei kleinen Datensätzen mit Extremwerten sollte weder Mittelwert noch Varianz bzw. Standardabweichung genutzt werden.

2.B Aufgaben

inhaltliche Aufgaben

1. Wozu gibt man Dispersionsmaße an?
2. Was bedeutet es, wenn die Streuung der Daten besonders groß ist?
3. In welchem Verhältnis stehen Varianz und Standardabweichung?
4. Welche Dispersionsmaße sind für intervallskalierte Daten geeignet?
5. Was ist der Unterschied zwischen der mittleren Abweichung und der Standardabweichung? Warum wird meist die Standardabweichung und nicht die mittlere Abweichung als Dispersionsmaß gewählt?
6. Welche Folgen hat das Quadrieren der Abweichungen bei der Varianz?
7. Welches Dispersionsmaß ist am robustesten gegenüber Ausreißern?
8. Warum gehören Mittelwert und Standardabweichung „zusammen“?

praktische Aufgaben

1. Am Strand von Ahlbeck (Ostsee) und von Djerba (Tunesien) wird an einem Tag im Sommer stündlich die Windstärke gemessen.
   a) Berechne ein angebrachtes Maß der zentralen Tendenz und ein Maß für die Streuung.
   b) Kann man (rein deskriptiv) sagen, dass an beiden Stränden die gleichen Windverhältnisse herrschen?

Uhrzeit	8h	9h	10h	11h	12h	13h	14h	15h	16h	17h	18h	19h	20h	21h	22h
Djerba	2	2	9	3	8	4	7	9	6	10	1	6	9	9	2
Ahlbeck	5	6	4	7	6	4	5	6	3	8	5	7	6	6	5

2. Ein Psychologe führt für das Personalbüro von Lufthansa die Einstellungstests mit 14 Bewerbern durch. Eine der erhobenen Variablen ist der Intelligenzquotient. Berechne das arithmetische Mittel, die Varianz und die Standardabweichung der IQ-Werte für dieses Kollektiv.

Bewerber	1	2	3	4	5	6	7	8	9	10	11	12	13	14
IQ-Wert	93	73	77	115	111	109	100	87	123	96	100	120	97	95

3. In einer Studie über den Einfluss von Alkohol auf die Reaktionszeit wurden die Daten von Probanden vor und nach dem Trinken von zwei Gläsern Wein erhoben. Die Reaktionszeiten in msec stehen in unterer Tabelle.
   a) Auf welchem Skalenniveau befinden sich die Daten?

b) Berechne ein sinnvolles zentrales Maß und ein Dispersionsmaß.

c) Unterscheiden sich die Reaktionszeiten vor und nach dem Konsum von 2 Gläsern Wein deskriptivstatistisch?

Proband	1	2	3	4	5	6	7	8	9	10
Vor	400	350	520	300	450	600	380	570	430	610
Nach	510	690	930	740	800	620	640	550	450	710

4. In einer Untersuchung soll festgestellt werden, ob ein spezielles Lesetraining die Lesefertigkeiten legasthenischer Kinder so verbessert, dass sich die Verteilung dieses Merkmals bei legasthenischen Kindern sich der Verteilung bei gesunden Kindern annähert. Nachfolgend die Punktwerte in einem Lesetest legasthenischer und gesunder Kinder, aufgeführt nach erfolgtem Lesetraining.

   a) Stelle die Daten deskriptiv gegenüber.

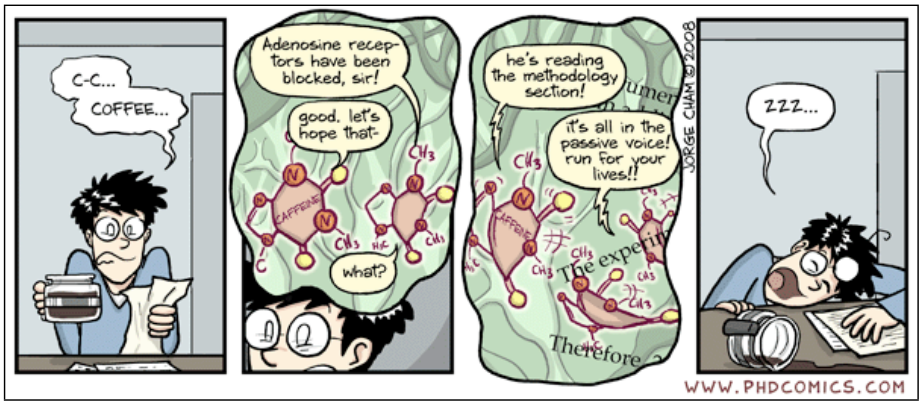
   b) Kann man rein deskriptiv sagen, dass durch das Lesetraining legasthenische Kinder genau so erfolgreich in dem Lesetest sind wie Kinder ohne Legasthenie?

Nichtlegastheniker

Proband	1	2	3	4	5	6	7	8	9	10	11	12
Testwert	17	20	15	18	18	16	12	14	19	17	15	20

Legastheniker

Proband	13	14	15	16	17	18	19	20	21	22	23	24
Testwert	14	16	12	10	16	20	13	9	14	13	15	8



2.3 Deskriptive Darstellungsformen

Die grafische Darstellung hilft die Lage und Verteilungsform von Daten schnell zu erfassen und gegenüberzustellen. Die richtige Wahl der Darstellungsform ist dabei wesentlich, da verschiedene Diagramme oder Plots unterschiedliche Vor- und Nachteile haben.

2.3.1 Das Histogramm

Bei einem Histogramm befinden sich auf der x-Achse die Werte oder Wertkategorien.
Auf der y-Achse befinden sich die beobachteten absoluten oder relativen Häufigkeiten.
Wie häufig eine bestimmte Kategorie in einem Datensatz vorkommt, wird durch Balken symbolisiert (darum nennt man das Histogramm auch Balkendiagramm). Die Balken suggerieren kategoriale Daten, doch auch metrische Daten (z.B. Daten auf Intervallniveau) können mithilfe eines Histogramms dargestellt werden. Dafür ist es zunächst notwendig, Kategorien zu bilden, die Daten den Kategorien zuzuordnen und auszuzählen.
Je größer die Kategoriegrenzen gewählt werden, desto größer ist die Gefahr, dass relevante Streuungsinformation innerhalb dieser Kategorien verloren geht. Insofern wird bei einem Histogramm zur Darstellung von intervallskalierten Daten immer ein Kompromiss zwischen praktischer Darstellbarkeit und Reduktion der Streuungsinformation geschlossen.
Eine sinnvolle Breite h für Kategorien eines Datensatzes lässt sich mithilfe der Freedman-Diaconis Regel ausrechnen, welche die Streuung der Daten mithilfe der Quartilwerte berücksichtigt:

$$h = \frac{2 \cdot (Q_3 - Q_1)}{\sqrt[3]{n}} \tag{2.13}$$

Beispiel:

Folgende Daten sollen mithilfe eines Histogramms grafisch veranschaulicht werden:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}
-2,1	0,2	3,2	3,6	4,3	4,7	4,8	5,1	6,3	6,4	6,6	6,6	7,0	7,4	7,6	7,9	8,2

Die Grenze des ersten Quartils der Datenreihe liegt zwischen dem 4. und 5. Datenpunkt. Die Grenze des zweiten Quartils zwischen dem 13. und 14. Datenpunkt. Daraus ergibt sich: $Q_1 = \frac{3,6+4,3}{2} = 3,95$ und $Q_3 = \frac{7,0+7,4}{2} = 7,2$.
Nach der Freedman-Diaconis Regel errechnet sich folgende optimale Kategorienbreite:

$$h = \frac{2 \cdot (7,2 - 3,95)}{\sqrt[3]{17}} = 2,5$$

Mithilfe der unteren Häufigkeitstabelle lässt sich das Histogramm zeichnen (siehe Abbildung 2.5).
Alternativ zu absoluten Häufigkeiten (wie in Abbildung 2.5) wäre es auch möglich und sinnvoll, relative oder prozentuale Häufigkeiten auf der Ordinate abzutragen (siehe rechte Spalte in der Häufigkeitstabelle).

Kategorie	Häufigkeit	relative Häufigkeit
-2,5 - 0	1	$1 \div 17 = 0,059$
0 - 2,5	1	$1 \div 17 = 0,059$
2,5 - 5	5	$5 \div 17 = 0,294$
5 - 7,5	7	$7 \div 17 = 0,412$
7,5 - 10	3	$3 \div 17 = 0,176$

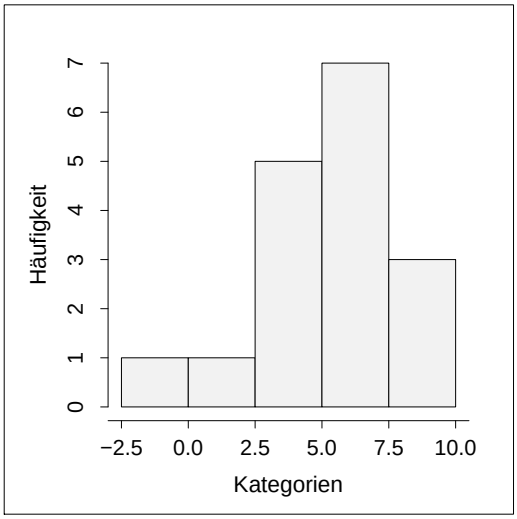


Abbildung 2.5: Histogramm mit absoluten Häufigkeiten

2.3.2 Der Stem und Leaf Plot

Der Stem und Leaf Plot eignet sich, um Häufigkeitsverteilungen von Merkmalsausprägungen zwischen zwei Gruppen oder Bedingungen zu vergleichen. Der Plot besteht aus drei Spalten. In der mittleren Spalte (dem ‚Stamm') befinden sich die Kategorien. In den Spalten links und rechts davon werden die Daten (die ‚Blätter' des Stamms) je nach Kategorienzugehörigkeit in die jeweilige Zeile eingetragen.

Beispiel:

Die unteren zwei Datenreihen sollen mit einem Stem und Leaf Plot verglichen werden.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}
59	61	59	63	57	61	46	63	58	56	72	54	45	57	57	72	66

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}
55	39	45	51	47	55	52	51	61	47	53	46	59	51	63	59	45

Der Stem und Leaf Plot für eine Kategoriengröße von 10 würde wie folgt aussehen:

Datensatz x	Kategorie	Datensatz y
22	7	
63311	6	13
99877764	5	111235599
65	4	55677
	3	9

Die zweite Dezimalzahl wurde bei den Kategorien (Stamm) weggelassen, während bei den Zahlen (Blätter) die erste Dezimalstelle weggelassen wurde. Der Plot liest sich dadurch wie folgt: In der Kategorie 4 (Werte von 40-49) befinden sich zum Beispiel das Datum 45 und 46 im Datensatz x, im Datensatz y die Daten 45, 45, 46, 47 und 47.

Der Vorteil des Stem und Leaf Plots ist, dass man zwei Stichproben in ihrer Häufigkeitsverteilung direkt vergleichen kann. Im oberen Stem und Leaf Plot sieht man, dass in Datensatz x tendenziell die Daten eher rechtsschief sind, während die Daten des Datensatz y eher rechtssteil sind (sofern man die rechte Verteilung auf die linke spiegelt).

2.3.3 Der Box-Whisker Plot

Der Box-Whisker Plot besteht aus einer Box, welche den Bereich der mittleren 50% der Werte symbolisiert und von der ersten Quartilgrenze bis zur dritten Quartilgrenze reicht. Innerhalb dieser Box wird der Median, mithilfe einer horizontalen Linie, dargestellt.
Des weiteren besteht der Box-Whisker Plot aus den Whiskern (engl.: Schnurrhaare). Die Whiskerlänge beträgt maximal $1^1/_2$ mal dem IQR-Bereich. Der maximale und minimale Whiskerwert errechnet sich folgendermaßen:

$$\text{maximal kleinster Whiskerwert} = Q_1 - (IQR \cdot 1,5) \tag{2.14}$$

$$\text{maximal größter Whiskerwert} = Q_3 + (IQR \cdot 1,5) \tag{2.15}$$

Ist der ausgerechnete Wert in der Datenreihe nicht enthalten, so nimmt man den nächst näheren. Bei dem oberen Whisker wird der nächst niedrigere und beim unteren Whisker der nächst höhere Wert genommen. Werte die außerhalb des Whiskerbereichs liegen, werden separat als Punkte eingezeichnet und damit als Ausreißer symbolisiert.
Der Box-Whisker Plot geht auf den amerikanischen Statistiker John W. Tukey zurück.

Beispiel:

Die Daten aus obigem Beispiel lassen sich auch mithilfe eines Box-Whisker Plots veranschaulichen. Dafür benötigt man vier Kennwerte je Datenreihe: Den Median, den IQR, den Q1- und

den Q3-Wert. Für Datensatz x errechnen sich die Werte folgendermaßen:

$$Md_{ort} = \frac{17+1}{17} = 9 \qquad Md = 59$$
$$Q_{1(ort)} = \frac{8+1}{2} = 4,5 \qquad Q_1 = \frac{56+57}{2} = 56,5$$
$$Q_{3(ort)} = \frac{8+1}{2} + Md = 13,5 \qquad Q_3 = \frac{63+63}{2} = 63$$

$$IQR = 63 - 56,5 = 6,5$$
$$\text{maximal kleinster Whiskerwert} = 56,5 - (6,5 \cdot 1,5) = 46,75$$
$$\text{maximal größter Whiskerwert} = 63 + (6,5 \cdot 1,5) = 72,75$$

Da der untere, errechnete Whiskerwert in der Datenreihe nicht vorkommt, nimmt man den nächst höheren, in diesem Fall 54. Auch der obere errechnete Whiskerwert kommt in der Datenreihe nicht vor. Also entscheidet man sich für den nächst niedrigeren Wert 72.
Auf die selbe Weise lassen sich die Kennwerte für Datensequenz y errechnen und die Boxplots wie in Abbildung 2.6 zeichnen.

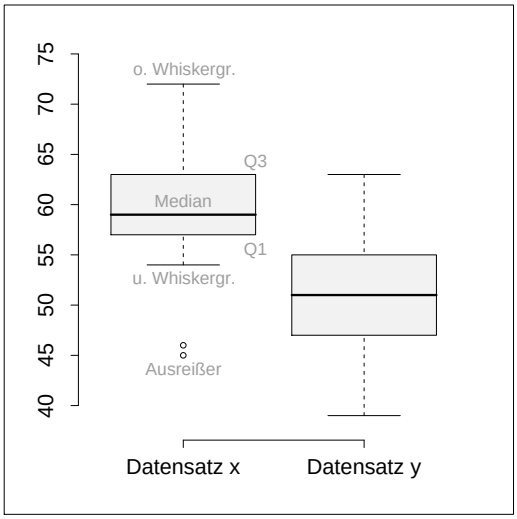


Abbildung 2.6: Boxplots für die Beispieldaten aus 2.3.2

2.3.4 Der Mittelwertsplot

Beim Mittelwertsplot wird der Mittelwert als Punkt in ein Diagramm eingezeichnet. Ober- und unterhalb des Mittelwertes werden vertikale Linien (Residuen) eingezeichnet. Diese Linien symbolisieren die Streuung der Daten und entsprechen dem Abstand von einer Standardabweichung oberhalb und einer Standardabweichung unterhalb des arithmetischen Mittels der Daten. Statt der Standardabweichung wird auch oft der sog. Standardfehler des Mittelwertes als Ober- und Untergrenze der Residuen eingesetzt (zum Standardfehler des Mittelwertes: siehe Kapitel 4).

Beispiel:

Aus den Daten von oben (siehe 2.3.2) errechnet sich für Datensatz x der Mittelwert $\bar{x} = 59,18$ und eine Standardabweichung von $s = 7,24$. Damit reicht das obere Residuum bis zum Wert $\bar{x} + s = 66,42$ und das untere Residuum bis zum Wert $\bar{x} - s = 51,94$.
In Abbildung 2.7 ist ein Mittelwertsplot für beide Datensequenzen dargestellt.

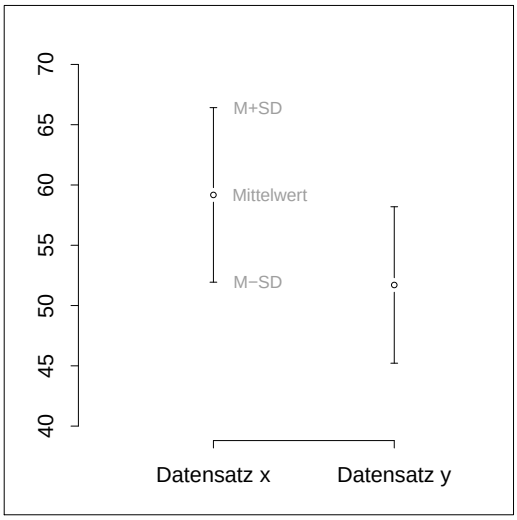


Abbildung 2.7: Mittelwertsplots für die Beispieldaten aus 2.3.2

Der Mittelwertsplot eignet sich für annähernd normalverteilte intervallskalierte Daten, während der Box-Whisker Plot auch ordinalskalierte oder schiefverteilte Daten gut abbilden kann. Bei multimodalen Verteilungen (also Verteilungen, die mehr als einen Gipfel besitzen) ist weder der Box-Whisker Plot noch der Mittelwertsplot zur Darstellung der Daten geeignet. Stattdessen können Daten, die eine multimodale Verteilung aufweisen, gut durch ein Histogramm dargestellt werden.

2.C Aufgaben

inhaltliche Aufgaben

1. In welche Darstellungsform gehen die meisten Informationen über einen Datensatz ein?
2. Vergleiche Abbildung 2.6 mit Abbildung 2.7. Was kann man über die Verteilung der Daten an den Boxplots ablesen, was bei den Mittelwertsplots verloren geht?
3. Warum ist der Boxplot für multimodale (mehrgipflige) Verteilungen nicht geeignet?
4. Ab wann gelten Daten als Ausreißer bei einem Box-Whisker Plot?
5. Alternativ zur Freedman-Diaconis Regel lässt sich über die Sturgess-Regel die Anzahl der Balken k für ein Histogramm berechnen:

$$k = 1 + 3,3 \cdot ln(n) \tag{2.16}$$

Vergleiche die Formel mit der von Freedman und Diaconis. Was ist der Nachteil der Sturgess-Regel und warum sollte man deswegen die Freedman-Diaconis Regel verwenden?

praktische Aufgaben

1. In einem Experiment soll untersucht werden, ob das subjektive Schmerzempfinden von Aufmerksamkeitsprozessen beeinflusst wird. Zwei Gruppen von Probanden wurden elektrische Schläge verabreicht, die in ihrer Intensität stufenweise erhöht wurden. Die eine Gruppe wurde durch ein Video abgelenkt, während die andere Gruppe instruiert wurde dem Schmerz Aufmerksamkeit zu schenken. Aufgezeichnet wurde die Anzahl an Elektroschocks, die ein Proband aushielt, bevor er äußerte, dass ihm weitere elektrischen Schläge zu unangenehm seien und damit das Experiment beendete.

   Die Ergebnisse sind unten tabellarisch aufgeführt.

ohne Ablenkung

Proband	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Anzahl	7	1	6	11	4	3	5	3	4	3	5	4	6	3

mit Ablenkung

Proband	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Anzahl	6	6	3	7	10	8	8	10	9	7	7	5	10	8

   a) Vergleiche die Gruppen mithilfe eines Box-Whisker Plots und eines Mittelwertsplots.
   b) Welche Darstellungsweise hältst du am angebrachtesten für die Daten? Warum?
   c) Lässt sich rein deskriptiv sagen, dass Schmerzempfindung durch Aufmerksamkeit moduliert wird?

2. Laut einer Theorie aus der Gedächtnispsychologie erinnert man gelerntes Material dann besser, wenn man sich in der selben emotionalen Verfassung befindet, wie zu der Zeit in der man das Material abgespeichert hat.

   Dies soll experimentell überprüft werden. Zwei Probandengruppen wird ein lustiger Filmausschnitt gezeigt. Anschließend sollen sie eine Reihe von sinnlosen Silben lernen. Nach 24 Stunden wird der einen Gruppe (Gruppe 1) wieder ein lustiger, der anderen (Gruppe 2) dagegen ein trauriger Filmausschnitt gezeigt. Danach werden die Probanden gebeten, alle Silben, an die sie sich erinnern können aufzusagen.

   Die Ergebnisse (Anzahl richtig erinnerter Silben) sind für die zwei Gruppen unten tabellarisch aufgeführt.

Gruppe 1

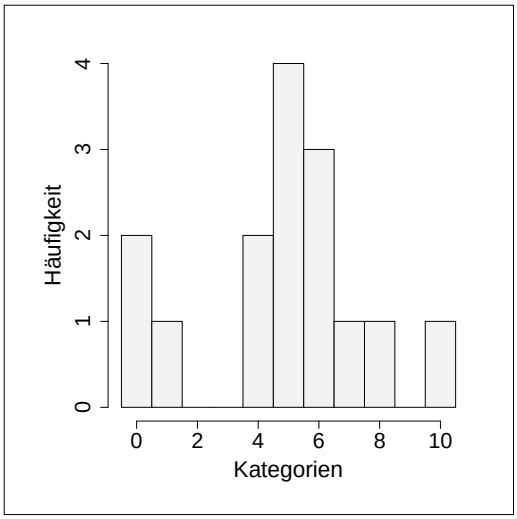
Proband	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Silben	15	9	14	8	11	7	12	12	9	8	11	11	8	9	10

Gruppe 2

Proband	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Silben	10	7	12	11	4	4	6	7	10	12	8	5	10	14	9

   a) Auf welchem Skalenniveau befinden sich die Daten?
   b) Wähle eine geeignete Darstellungsform, um die zwei Gruppen gegenüberzustellen.
   c) Belegen die Daten (rein deskriptiv) die Theorie?

3. Berechne Median und IQR für die unten in dem Histogramm dargestellten Daten.



2.4 Kovarianz, Korrelation und Regression

Angenommen man will untersuchen, ob Arbeitnehmer, die sich mit ihrer Organisation identifizieren, eine höhere Arbeitszufriedenheit aufweisen, als solche, die dies nicht tun. Um diese Frage zu beantworten, könnte man eine Gruppe von Arbeitnehmern, die sich nicht mit ihrer Organisation identifizieren, nach ihrer Arbeitszufriedenheit befragen und den Mittelwert dieser Gruppe mit dem Mittelwert einer anderen Gruppe von Arbeitnehmern vergleichen, die dagegen angeben sich mit ihrer Organisation zu identifizieren.
Was man gemessen hätte, wäre Arbeitszufriedenheit auf intervallskalenniveau (z.B. durch eine Ratingskala von 0 für ‚nicht zufrieden' bis 9 für ‚sehr zufrieden') und Identifikation auf Nominalniveau (Identifikation: ja / nein).
Ein anderer Ansatz wäre, auch den Grad an Identifikation mit einem Unternehmen exakter zu messen (z.B. wiederum mit einer Ratingskala von 0 - ‚keine Identifikation' bis 9 - ‚sehr hohe Identifikation'). Damit hätte man die Ausprägungen der zwei Merkmale auf Intervallskalenniveau abgebildet.
Nun würde man nicht mehr nach dem Unterschied, sondern nach dem Zusammenhang dieser beiden Variablen (Arbeitszufriedenheit und Identifikationsgrad) fragen.

2.4.1 Kovarianz

Um die Gemeinsamkeit von zwei Variablen zu bestimmen, kann man dessen Kovarianz (kurz: cov oder s_{xy}) errechnen. Die Kovarianz ist die gemeinsame Varianz, also ein Maß dafür, wie sehr die eine Variable mit der anderen Variable kovariiert.
Die Kovarianz errechnet sich über:

$$cov(x,y) = s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{n} \tag{2.17}$$

Die Kovarianzformel ist ähnlich aufgebaut wie die Varianzformel (siehe Formel 2.10). Statt die Abweichung eines Wertes vom Mittelwert mit sich selbst zu multiplizieren, wird bei der Kovarianzberechnung die Abweichung eines Datenpunktes x_i vom Mittelwert $\bar{x}$ mit der Abweichung eines Datenpunktes y_i vom Mittelwert $\bar{y}$ multipliziert.
Die Kovarianz kann im Gegensatz zur Varianz auch einen negativen Wert annehmen.

2.4.2 Korrelation

Mithilfe der Kovarianz kann man schwer einschätzen, ob zwei Merkmalsausprägungen oder Datenreihen stark oder schwach kovariieren. Dies liegt an der unterschiedlichen Größe der Abweichungen, die in die Formel eingehen, wenn zwei Variablen unterschiedlich stark streuen (also unterschiedlich große Varianzen besitzen).
Angenommen man will den Zusammenhang von Größe und Gewicht errechnen und man hat diese beiden Merkmale an sieben Personen gemessen. Aus den Daten in der unten stehenden Tabelle kann man ablesen, dass das Gewicht eine höhere Varianz aufweist als die Größe. Man kann allerdings nicht sagen, das Merkmal Gewicht streut zwischen den Personen mehr als die Größe. Der Unterschied entsteht stattdessen aufgrund der unterschiedlichen Skalierung

der beiden Merkmale. Während ein Unterschied von 0,1m ein enormer Zuwachs auf der Skala ‚Größe' ist, entspricht ein Zuwachs von 0,1kg dagegen einer sehr kleine Änderung in der Merkmalsausprägung auf der Gewichtsskala.
Aufgrund dieser unterschiedlichen Skalierung ist die Varianz zwischen den beiden Variablen nicht vergleichbar. Genauso ist die Kovarianz, also die gemeinsame Varianz schwer zu interpretieren. Sie muss erst normiert werden.

Proband i	Größe x_i (in m)	Gewicht y_i (in kg)	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
1	1,8	77	0,416
2	1,75	78	0,249
3	1,72	70	-0,047
4	1,71	75	0,016
5	1,69	71	0,022
6	1,65	69	0,196
7	1,6	69	0,382
	$\bar{x} = 1,70$	$\bar{y} = 72,71$	
	$s^2 = 0,0036$	$s^2 = 12,78$	$s_{xy} = 0,1765$

Dies erreicht man, indem man die Produkt-Moment-Korrelation (kurz r oder cor) ausrechnet, die auf den britischen Mathematiker Karl Pearson zurückgeht und zur Berechnung des relativen Zusammenhangs zweier, mindestens intervallskalierter, Variablen benutzt wird.

$$cor(x,y) = r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2.18}$$

In der Formel wird die Kovarianz von x und y an dem Produkt der Standardabweichung von x und y relativiert:

$$cor(x,y) = r_{xy} = \frac{cov(x,y)}{s_x \cdot s_y} \tag{2.19}$$

Die Kovarianz kann nie größer sein als das Produkt der Standardabweichungen der beiden einzelnen Variablen. Sie kann maximal genau so groß sein wie das Produkt der Standardabweichung.
r kann also nie größer als 1 werden.
Dies macht den Korrelationskoeffizienten interpretierbarer als die Kovarianz: Je näher der Korrelationskoeffizient an 1, desto größer bzw. enger ist der Zusammenhang. Für die Daten aus obiger Tabelle errechnet sich eine Korrelation von $r = 0,82$. Dies deutet auf einen Zusammenhang der Form: je größer desto schwerer hin.
Die Kovarianz kann allerdings auch einen negativen Wert annehmen. Dadurch wird auch der Korrelationskoeffizient negativ (mit einem Minimum von -1). Eine negative Korrelation deutet auf einen negativen Zusammenhang der Form „je größer Merkmalsausprägung x, desto kleiner Merkmalsausprägung y" (oder umgekehrt) hin.

Determinationskoeffizient

Quadriert man eine Korrelation, erhält man den Determinationskoeffizienten. Dieser gibt, wenn man ihn mit 100 multipliziert, Auskunft über den prozentualen Anteil gemeinsamer Varianz. Der Determinationskoeffizient des Zusammenhangs zwischen Größe und Gewicht beträgt: $r^2 = 0,82^2 = 0,668$. Das bedeutet 66,8% der Varianz der einen Variable lässt sich durch die Varianz der anderen Variable erklären. Die Information, wie groß jemand ist steckt also, zum gewissen Teil, in der Information wie viel jemand wiegt (redundante Information). 33,2% sind dagegen Rest- oder Eigenvarianz, die nicht durch die jeweilige andere Variable erklärt werden kann. Daher wird sie auch der Anteil an Fehlervarianz genannt.
In Abbildung 2.8 ist der Zusammenhang von Korrelations- und Determinationskoeffizient aufgezeichnet. Der Determinationskoeffizient ist verhältnisskaliert: Ein Determinationskoeffizient von $r^2 = 0,6$ entspricht doppelt so viel erklärter Varianz gegenüber einem Determinationskoeffizient von $r^2 = 0,3$.
Bei dem Korrelationskoeffizient ist dies nicht so, wie man an dem Verlauf der Kurve in Abbildung 2.8 ablesen kann. Die Intervalle des Korrelationskoeffizienten sind nicht äquidistant: Eine Korrelation von $r =,6$ entspricht nicht einem doppelt so großen Zusammenhang, im Vergleich zu einer Korrelation von $r =,3$.

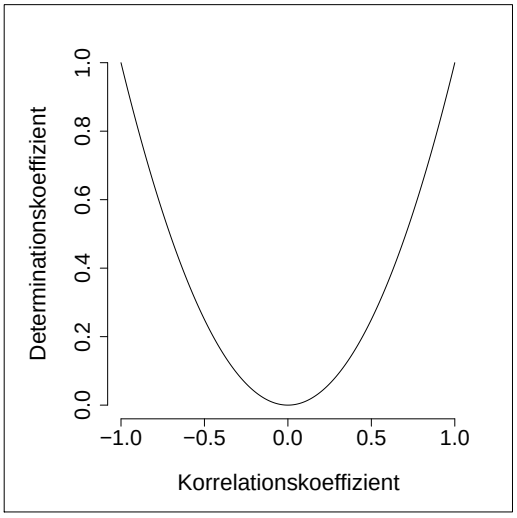


Abbildung 2.8: Zusammenhang von Korrelations- und Determinationskoeffizient

2.4.3 Regression

Der Zusammenhang von zwei stetigen Datenreihen lässt sich auch grafisch darstellen, mit einem so genannten Streudiagramm. Die Daten von oben (Größe und Gewicht) sind in Abbildung 2.9 dargestellt.
Anhand des Diagramms kann man ablesen, dass es sich bei den beiden Variablen Größe und Gewicht höchstwahrscheinlich um eine lineare Beziehung handelt; Die x-Werte steigen mit denen der y-Werte geradlinig an.
Es ist jedoch nicht möglich eine Linie zu finden, die durch alle Punkte in dem Streudiagramm führt. Stattdessen streuen die Daten unsystematisch in die eine oder andere Richtung.

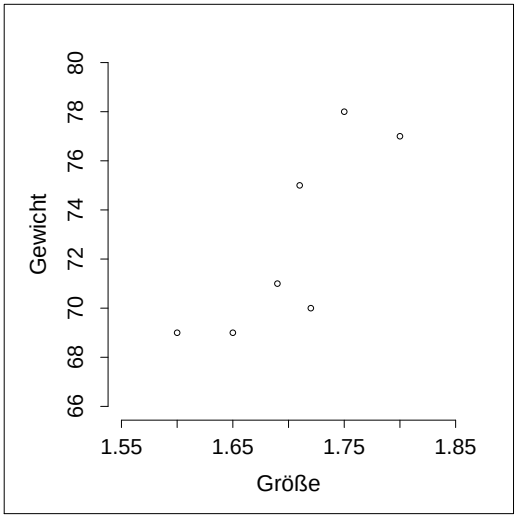


Abbildung 2.9: Streudiagramm der Daten aus 2.4.2

Die Daten lassen sich also nicht perfekt durch eine lineare Gleichung beschreiben. Man kann allerdings versuchen eine lineare Modellgleichung zu finden, welche die Datenstruktur bestmöglich beschreibt und wiedergibt. Als Modellgleichung stellen wir auf:

$$y = bx + a \tag{2.20}$$

Die Parameter b und a sind in unserem Modell zunächst unbekannt. Wir stehen also vor der Herausforderung diese beiden Parameter bestmöglich zu schätzen.
Eine Methode der bestmöglichen Anpassung (sog. ‚fitting') einer Modellgleichung und dessen Parameter an die empirischen Daten ist die von dem Mathematiker Carl Friedrich Gauß erdachte Methode der kleinsten Quadrate (engl.: Least Squares Method).
Gauß machte dabei vor allem Annahmen über die nicht erklärbare Streuung: angenommen Größe und Gewicht hängen linear miteinander zusammen (Modellannahme). Diese lineare Beziehung ist aber überlagert von einer Eigenstreuung der Variablen, die nicht durch die jeweils andere Variable erklärt werden kann. Diese Eigenstreuung (die so genannte „Fehlervarianz", da sie nicht durch das Modell erklärt wird) kann viele Gründe haben. In unserem Fall kann das Gewicht auch mit Ernährungsgewohnheiten, Genen, Alter etc. zusammenhängen, genau wie die Größe durch viele andere denkbare Variablen mit beeinflusst wird.

Wir gehen nun davon aus, dass diese Drittvariablen (bestimmte Gensequenzen, Ernährungsgewohnheiten etc.) in ihrer Ausprägung zufällig und unsystematisch auf die Personen verteilt sind. Die denkbaren Drittvariablen haben also einen Einfluss auf Gewicht und Größe und stören damit den linearen Zusammenhang. Dieser Einfluss, und das ist die zentrale Annahme der Methode der kleinsten Quadrate, ist über die Gewichts- und Größenstufen aber unsystematisch und zufällig.
Mithilfe der Wahrscheinlichkeitstheorie kann man nun nachweisen, dass diese Fehlstreuung, sofern sie tatsächlich unsystematisch ist, annähernd normalverteilt ist und einen Mittelwert bzw. Erwartungswert von 0 besitzt. Trifft dies zu, so kann man mithilfe der Methode der kleins-

ten Quadrate eine Modellgleichung finden, die die Daten bestmöglich repräsentiert. Bestmöglich bedeutet, dass die Summe der quadrierten Abweichungen der empirischen Werte von den, von der Modellgleichung vorhergesagten, Werten ein Minimum ergibt.
Dies entspricht der Logik des Mittelwerts, bezogen auf einen zweidimensionalen Raum.

Die Parameter der linearen Gleichung werden durch folgende Formeln geschätzt:

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2} \tag{2.21}$$

$$a = \bar{y} - b \cdot \bar{x} \tag{2.22}$$

Der Steigungsparameter b (engl. slope) errechnet sich also über die Kovarianz von x und y, relativiert an der Varianz von x.
Die Konstante a (engl. intercept) repräsentiert den Schneidepunkt der y-Achse der errechneten Modellgeraden und ergibt sich aus der Differenz des Mittelwerts von y und des durch den Steigungsparameter b gewichteten Mittelwerts von x.

Für die Beispieldaten von oben errechnen sich die Modellparameter folgendermaßen:

$$b = \frac{0,1765}{0,0036} = 49,03$$
$$a = 72,71 - 49,03 \cdot 1,70 = -10,64$$

Die Modellgleichung sieht also wie folgt aus:

$$\hat{y} = 49,03 \cdot x - 10,64$$

Das Dach auf dem y gibt an, dass dieser Wert mit der Modellgleichung geschätzt wird. $\hat{y}$ ist also der geschätzte oder vorhergesagte Wert und ist nicht zu verwechseln mit einem empirisch-gemessenen Wert.
In Abbildung 2.10 ist das Streudiagramm zusammen mit der Modellgeraden dargestellt. Die Regressionsgerade verläuft so durch die Punktwolke, dass die Summe der quadrierten Abweichungen der empirischen y-Werte von den, durch die Modellgerade vorhergesagten $\hat{y}$-Werten, ein Minimum ergeben.

$$\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 = Min. \tag{2.23}$$

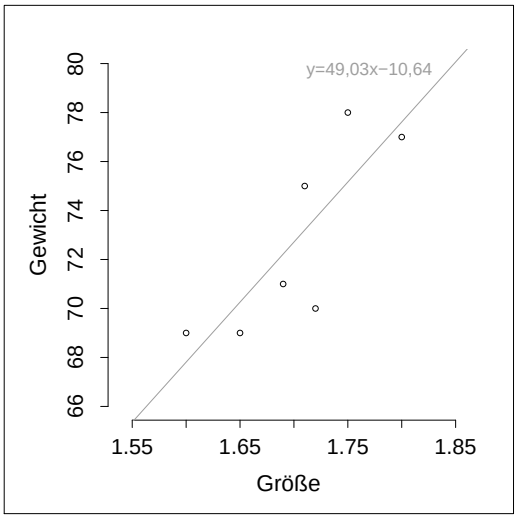


Abbildung 2.10: Streudiagramm der Daten aus 2.4.2 mit Regressionsgerade

2.4.4 Zusammenhang von Korrelation und Regression

Mithilfe einer Regressionsgeraden lassen sich auch Vorhersagen machen für Merkmalsausprägungen, die man nicht empirisch erhoben hat. So kann man errechnen, dass nach der Regressionsgleichung des Größe-Gewichtsbeispiels eine Person, die zwei Meter groß ist, schätzungsweise 87,42 Kilogramm wiegt.
Die Vorhersagegenauigkeit mit solchen Modellgleichungen hängt vor allem davon ab, wie sehr der Zusammenhang der beiden Variablen von Störvarianz überlagert ist bzw. umgekehrt: wie groß der Anteil der gemeinsamen Varianz und die Korrelation ist.

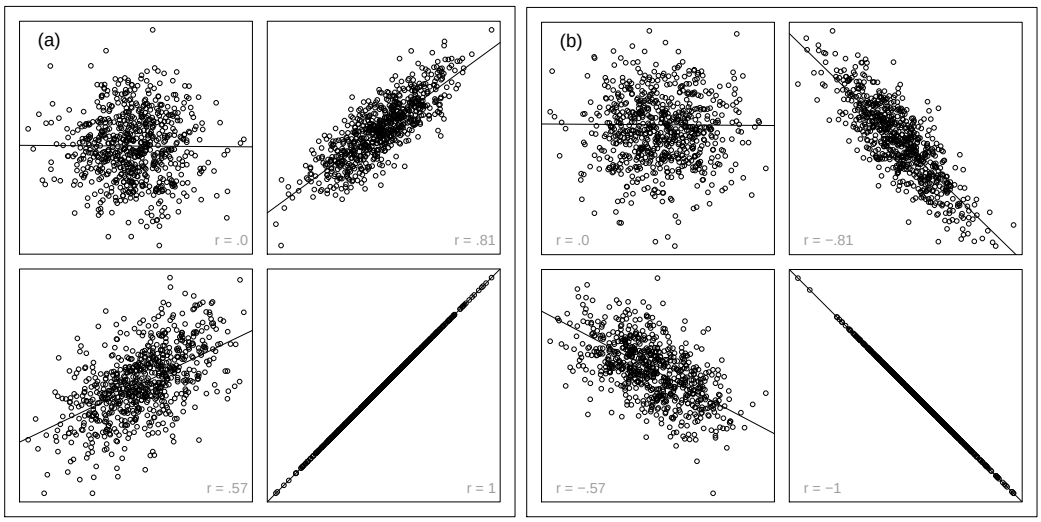


Abbildung 2.11: simulierte Daten von jeweils n=700 mit (a) aufsteigend, positiven und (b) negativen Korrelationen und geschätzten Regressionsgeraden

Anhand Abbildung 2.11 lässt sich erahnen, wie die Vorhersagesicherheit mit größer werdender Korrelation r wird. Bei einem perfekten Zusammenhang (r=1) liegen alle empirischen Messpunkte auf der theoretischen Modelllinie.

Standardschätzfehler

Errechnet man eine Regressionsgerade, so lässt sich die Varianz der abhängigen Variable y (also jener Variable, deren Werte man durch die andere Variable vorhersagen will) in einen Anteil erklärte Varianz und einen Anteil an Varianz, der nicht durch die Regressionsgerade (bzw. die Varianz von x) erklärt werden kann, zerlegen.
Es gilt:

$$\text{Gesamtvarianz von y} = \text{erklärte Varianz} + \text{Fehlervarianz} \tag{2.24}$$

$$s_y^2 = s_{\hat{y}}^2 + s_e^2 \tag{2.25}$$

$$\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n} + \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{2.26}$$

Die Fehler- oder Restvarianz s_e^2 lässt sich auch über den Korrelationskoeffizienten errechnen:

$$s_e^2 = s_y^2 \cdot (1 - r_{xy}^2) \tag{2.27}$$

Die Wurzel aus der Fehlervarianz wird Standardschätzfehler genannt. Der Standardschätzfehler ist die Standardabweichung der Residuen, also der Restvarianz. Er dient als Maß für die Genauigkeit einer Vorhersage mithilfe einer Regressionsgeraden. Je kleiner s_e, desto genauer die Vorhersage.

$$s_e = s_y \cdot \sqrt{(1 - r_{xy}^2)} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{2.28}$$

Ausreißerempfindlichkeit

Genau wie der Mittelwert ist auch die Korrelation und die Regressionsanalyse aufgrund der Methode der kleinsten Quadrate empfindlich gegenüber Ausreißern. Für die Praxis bedeutet dies, dass man sich bei kleinen Stichproben die Daten zunächst mithilfe eines Streudiagramm veranschaulichen lassen sollte, um möglichen Verzerrungen der Regressionsgerade oder der Korrelation, durch Ausreißer vorbeugen zu können.

2.4.5 Voraussetzungen der Korrelations- und Regressionsanalyse

Die oben beschriebene Produkt-Moment-Korrelation geht von der Äquidistanz der Skalen aus. Daher müssen die Daten auf Intervallskalenniveau vorliegen. Bei Verletzung dieser Vorraussetzung sollte man die Gemeinsamkeit zweier Variablen mit anderen Verfahren, wie z.B. mit der Rangkorrelation nach Spearman, errechnen, bei der die Daten lediglich auf Ordinalniveau vorliegen müssen.

Des Weiteren wird ein linearer Zusammenhang von den Variablen x und y vorausgesetzt.

Es ist jedoch vorstellbar, dass stattdessen beispielsweise ein exponentieller Zusammenhang zwischen den beiden Variablen besteht. Bei einem solchen Fall unterschätzt der Korrelationskoeffizient den wahren Zusammenhang der beiden Variablen und auch die Regressionsanalyse, wie oben beschrieben, wird die Daten nicht gut abbilden können und bei der Vorhersage von nicht gemessenen Merkmalsausprägungen zu falschen Ergebnissen kommen.
Insofern ist es wichtig, sich zunächst inhaltlich zu überlegen, in welcher Form die beiden Variablen zusammenhängen könnten. Anschließend sollte man sich empirische Datensätze immer mithilfe eines Streudiagramms veranschaulichen, um dann einschätzen zu können, ob es sinnvoll ist einen linearen Zusammenhang zu vermuten und dementsprechend die Regressionsanalyse und die Korrelation, wie oben beschrieben, zu berechnen.

Eine weitere Vorraussetzung für eine möglichst fehlerfreie Berechnung von Korrelationskoeffizienten und Regressionsparametern ist die Homoskedastizität der Residuen bzw. der Fehlervarianz.
Unter Homoskedastizität versteht man die Varianzhomogenität der Residuen. Genauer: Der durchschnittliche Abstand von dem durch die Regressionsanalyse vorhergesagten Wert $\hat{y}$ von dem gemessenen Wert y_i sollte, über die Stufen von x hinweg, konstant groß bleiben.

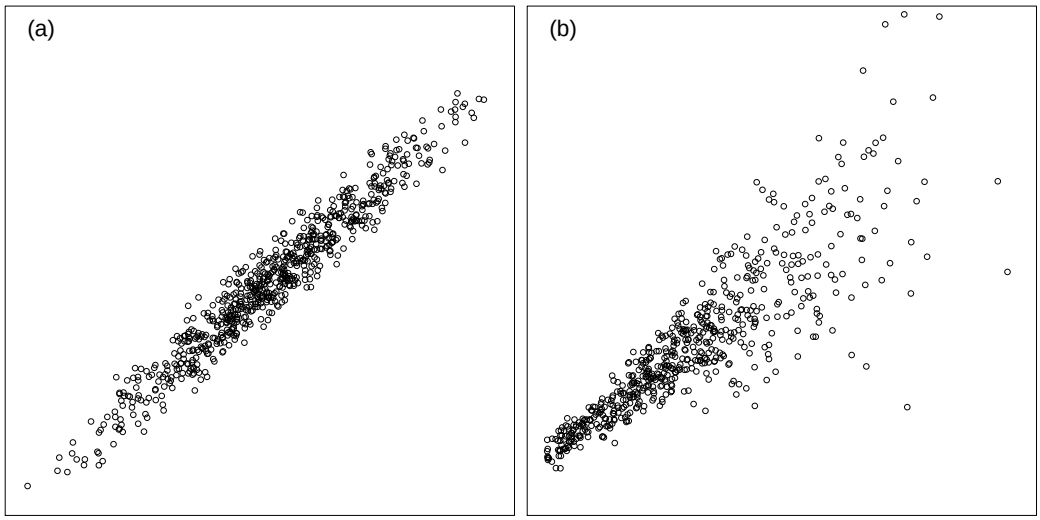


Abbildung 2.12: simulierte Daten von jeweils n=700 mit (a) Varianzhomogenität der Fehler (Homoskedastizität) und (b) Varianzheterogenität der Fehler (Heteroskedastizität)

In Abbildung 2.12 sind Datensätze dargestellt, bei denen in 2.12.a die Fehlervarianz über die Stufen von x konstant bleibt, während bei 2.12.b mit wachsender x-Ausprägung die Residualvarianz zunimmt.
Würde man nun einen Korrelationskoeffizienten der Daten aus 2.12.b berechnen, wäre dieser Koeffizient wenig repräsentativ für den tatsächlichen Zusammenhang, da dieser bei steigender Merkmalsausprägung geringer wird, also nicht konstant bleibt, wie es ein einzelner Koeffizient allerdings suggerieren würde.

2.D Aufgaben

inhaltliche Aufgaben

1. Was ist der Unterschied zwischen der Regressionsanalyse und der Korrelationsanalyse?
2. Wie hängen Korrelations- und Regressionsanalyse zusammen?
3. Was ist der Unterschied zwischen Kovarianz und Korrelation?
4. Welche Vorraussetzungen müssen erfüllt sein, damit man sicher sein kann, dass der Korrelationskoeffizient nach Pearson den tatsächlichen Zusammenhang der Daten repräsentiert? Welche Bedingungen stellt die Regressionsanalyse an die Daten?
5. Welche Methode wird bei der Regressionsanalyse verwendet? Welche Voraussetzungen und welchen Nachteil hat diese Methode?
6. Warum sollte man vor jeder Korrelations- und Regressionsanalyse sich die Daten mithilfe eines Streudiagramms veranschaulichen?
7. Was ist die Fehler-, was die erklärte Varianz?
8. Was ist der Standardschätzfehler und wozu dient er?
9. Durch was ist die Steigung einer Regressionsgerade definiert?
10. Kann man anhand der Steigung einer Regressionsgerade die Größe des Zusammenhangs von zwei Variablen ablesen?
11. Angenommen man will mithilfe einer Regressionsgleichung eine Vorhersage machen. Wie kann man die Genauigkeit der Schätzung mithilfe des Standardschätzfehlers erhöhen?
12. Wie könnte man den Zusammenhang bei bivariaten Daten mit Heteroskedastizität adäquat charakterisieren?
13. Kann man bei korrelierten Daten auf die kausale Struktur zweier Variablen mithilfe des Korrelationskoeffizienten schließen?
14. In der Regressionsanalyse wird die x-Variable auch unabhängige oder Prädiktorvariable genannt; die y-Variable dagegen abhängige bzw. Kriteriumsvariable. Bei welcher Art von Zusammenhängen sind diese Bezeichnungen sinnvoll? Bei welchen nicht?
15. Warum müssen Variablen für die Berechnung der Pearson-Produkt-Moment-Korrelation intervallskaliert sein?
16. Überlege dir zwei Merkmale, die womöglich linear miteinander zusammenhängen. Überlege dir zwei Merkmale, bei denen ein quadratischer Zusammenhang vorliegt.
17. Angenommen bei zwei Merkmalen liegt ein quadratischer Zusammenhang vor. Wie könnte man den Zusammenhang dieser zwei Variablen adäquat berechnen?

18. Welchen Wert würde man bei einer Ausprägung x_i für y_i vorhersagen, wenn eine Korrelation von r = 0 zwischen den beiden Variablen besteht?

19. Angenommen zwei Variablen kovariieren mit $s_{xy} = 0,64$. Beide Variablen haben eine Standardabweichung von 1 und einen Mittelwert von 0. Wie groß ist der b-Koeffizient der Regressionsgeraden? Wie groß ist a?

praktische Aufgaben

1. Eine Erziehungsanstalt für schwer erziehbare Kinder hat bisher immer mit Bestrafung als Motivationsmittel gearbeitet. Der Leiter der Anstalt hat letztes Wochenende zum ersten Mal seit 20 Jahren an einem Fortbildungsseminar teilgenommen. Dort hat er gehört, dass sich Bestrafung auf das Verhalten von Kindern eher negativ auswirken kann. Er mag es kaum glauben! All die Jahre ein falsches Konzept?

   Er beauftragt einen Psychologen, der herausfinden soll, ob die Kinder, die mehr bestraft werden, tatsächlich weniger sozial handeln. Die Bestrafung ist hier ausgedrückt in Stunden entzogener Freiheit, die Kriteriumsvariable ist die Anzahl an sozialen Handlungen pro Tag.

Prb Nr	**x** (Bestrafung)	**y** (Sozialverhalten)
1	10	0
2	7	2
3	1	8
4	9	1
5	2	7
6	6	4
7	3	6
8	4	6

   a) Stelle die Daten mit einem Streudiagramm dar.
   b) Wie hoch ist der Zusammenhang zwischen Bestrafung und Sozialverhalten?
   c) Bestimme die Regressionsgleichung und zeichne sie in das Streudiagramm ein.
   d) Welchen Wert für die Variable Sozialverhalten würde man einem Kind vorhersagen, dass mit fünf Stunden Freiheitsentzug bestraft wurde?
   e) Berechne den Standardschätzfehler und beurteile, wie verlässlich die Vorhersage ist.

2. In der Säuglingsforschung wollen Psychologen untersuchen, ob das Interesse für große, weiß-schwarz gescheckte Kaninchen mit dem Lebensalter wächst.

   Unten in der Tabelle sind die Daten aufgeführt, die zur Beantwortung dieser Frage erhoben wurden. Die unabhängige Variable ist das Lebensalter der Kinder (ausgedrückt in Wochen), die abhängige Variable ist das Interesse für große, schwarz-weiß gescheckte Kaninchen (ausgedrückt in Minuten pro Tag, in denen ein Kind das Kaninchen betrachtet hat oder versucht hat, Kontakt zu ihm aufzunehmen).

Prb Nr	**x** (Alter in Wochen)	**y** (Interesse in min)
1	31	15
2	100	80
3	67	35
4	46	47
5	85	75

a) Kann man einen solchen Zusammenhang (je älter, desto mehr Interesse am Kaninchen) anhand eines bivariaten Streudiagramms annehmen?

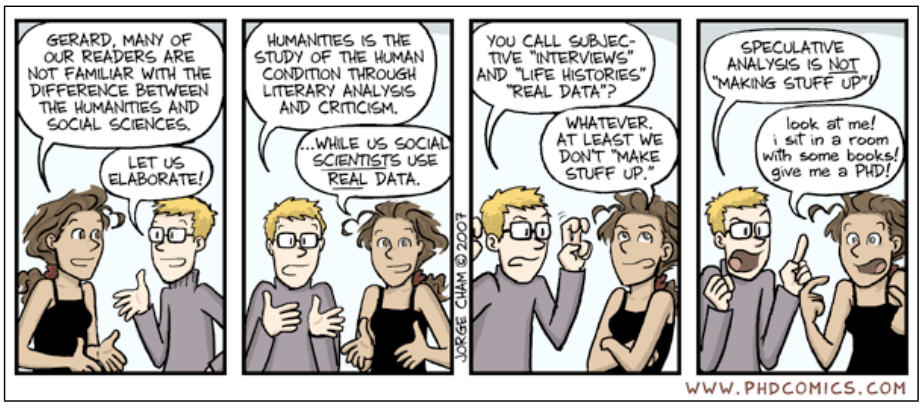
b) Bestimme die Regressionsgleichung und zeichne sie in das Streudiagramm ein.

c) Wo sieht man in diesem Streudiagramm die erzielten Werte y_i, die vorhergesagten Werte $\hat{y}_i$ und die Vorhersagefehler $y_i - \hat{y}_i$?

d) Welchen y-Wert würde man einem Kind laut der Daten aus Aufgabe 4 vorhersagen, dass 99 Wochen alt ist?

e) Welches Alter würde man einem Kind vorhersagen, dass 130 Minuten auf das Kaninchen schaut?

f) Berechne den Standardschätzfehler und beurteile, wie verlässlich die Vorhersagen sind.



3 Stochastik

3.1 Wozu Stochastik?

Stochastik oder Wahrscheinlichkeitsrechnung ist die Grundlage für die Inferenzstatistik. Mithilfe der Methoden der Inferenzstatistik (oder schließende Statistik) versucht man herauszufinden ob ein Unterschied oder ein Zusammenhang, den man in einer Stichprobe gefunden hat, rein zufällig zustande gekommen ist oder ob dieser Unterschied bzw. Zusammenhang statistisch signifikant (von lat. bedeutsam) ist.
Angenommen man will überprüfen, ob Frauen bei mentalen Rotationsaufgaben schlechter abschneiden als Männer. Bei mentalen Rotationsaufgaben geht es darum, eine dreidimensionale Figur in Gedanken um einen gewissen Winkel zu drehen und anschließend aus verschiedenen Alternativen die richtige Lösung auszuwählen.
Man könnte nun zehn solcher Aufgaben zusammenstellen und diese einer Frau und einem Mann vorlegen. Während die Frau nur sechs der Aufgaben korrekt löst, schafft es der männliche Proband acht der Aufgaben zu lösen.
Ist damit die Hypothese bestätigt, dass Männer bei Aufgaben der Raumkognition besser abschneiden als Frauen? Intuitiv würde man dem widersprechen; Das Ergebnis könnte rein zufällig entstanden sein.
Angenommen statt nur einer Frau und einem Mann testet man 15 weibliche und 15 männliche Probanden. Während die Gruppe der männlichen Probanden im Mittel sieben Aufgaben löst bei einer Standardabweichung von zwei, lösen die weiblichen Probanden im Mittel fünf Aufgaben bei einer Standardabweichung von drei.
Ist damit die Hypothese bestätigt, dass Männer bei Aufgaben der Raumkognition besser abschneiden als Frauen? Die Frage, die man sich hierbei stellen muss ist: Wie groß ist die Wahrscheinlichkeit, dass solch ein Ergebnis zustande kommt, wenn in Wirklichkeit kein Unterschied zwischen den Geschlechtern besteht?
Um diese Frage zu beantworten, sind Grundlagen der Wahrscheinlichkeitstheorie essentiell.

3.2 Das Zufallsexperiment

Unter einem Zufallsexperiment versteht man einen beliebig oft wiederholbaren Vorgang, der nach einer bestimmten gleichbleibenden Vorschrift ausgeführt wird und dessen Ergebnis vom Zufall abhängt. Zufall meint dabei, dass das Ergebnis nicht eindeutig bestimmt oder vorhergesagt werden kann.
Das Ergebnis eines Zufallsexperiments nennt man Elementarereignis.

Die Menge aller Elementarereignisse eines Zufallsexperiments bildet die Ergebnismenge Omega (Ω). Die Ergebnismenge kann sich aus verschiedenen Teilmengen (Ereignisräume - siehe Abbildung 3.1) zusammensetzen, deren Verknüpfung Implikation für die Wahrscheinlichkeit der Ziehung eines Elementarereignisses haben kann.

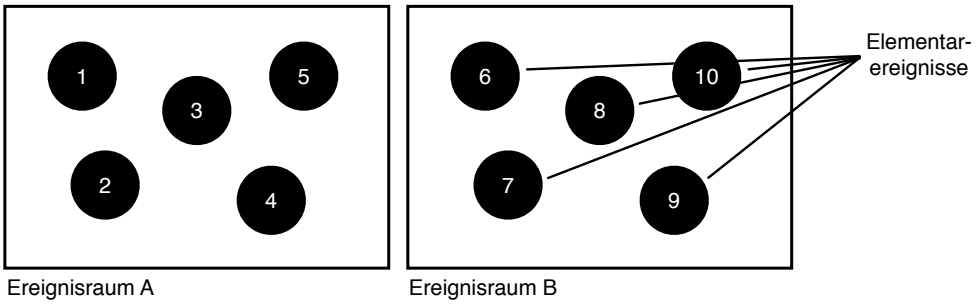


Abbildung 3.1: Darstellung von zwei Ereignisräumen und deren Elementarereignissen

In Abbildung 3.1 sind zehn Elementarereignisse (1-10) dargestellt. Jeweils fünf gehören dem Ereignisraum A und fünf dem Ereignisraum B an.

Formal schreibt man:

Ereignisraum = { zugehörige Elementarereignisse }
A = { 1 , 2 , 3 , 4 , 5 }
B = { 6 , 7 , 8 , 9 , 10 }

bzw.

Elementarereignis $\in$ (Element von) Ereignisraum
$4 \in A$
$7 \in B$

Ereignis A tritt ein, wenn 1, 2, 3, 4 oder 5 bei einem Zufallsexperiment „gezogen“ wird.

Formal geschrieben:

$A = 1 \cup 2 \cup 3 \cup 4 \cup 5$
$\cup$ steht für Vereinigung oder logische Summe.

$A \cup B = \{ 1 , 2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 , 10 \}$
Elemente aus A oder B = { Elementarereignisse von A und B }

Die Vereinigung von A und B ($A \cup B$) ergibt ein so genanntes sicheres Ereignis. Egal was gezogen wird, es ist auf jeden Fall ein Elementarereignis des Ereignisraums von A oder B.
Da $A \cup B$ die Gesamtheit aller Elementarereignisse umfasst gilt: $A \cup B = \Omega$.

Das komplementäre Ereignis zu A besteht aus den Elementarereignissen, die nicht zu dem Ereignisraum A gehören:

Nicht A ist gleich B
(bezogen auf dieses Beispiel)

Formal:

$$\bar{A} = B$$

bzw.

$$\bar{A} = \{\ 6\ ,\ 7\ ,\ 8\ ,\ 9\ ,\ 10\ \}$$

$\bar{A} \cup A =$ sicheres Ereignis $= B \cup A$

Schnittmengen

Ereignisraum A und B aus Abbildung 3.1 sind disjunkt. Das bedeutet sie sind unabhängig. Die Elementarereignisse gehören entweder zu A oder zu B.
Teilmengen der Ereignismenge Ω können auch Schnittmengen enthalten. Diese sind daher nicht disjunkt oder vereinbar.
Schnittmengen entstehen, indem Elementarereignisse sowohl zu dem einen als auch zu dem anderen Ereignisraum gehören. Veranschaulicht ist dies in Abbildung 3.2.

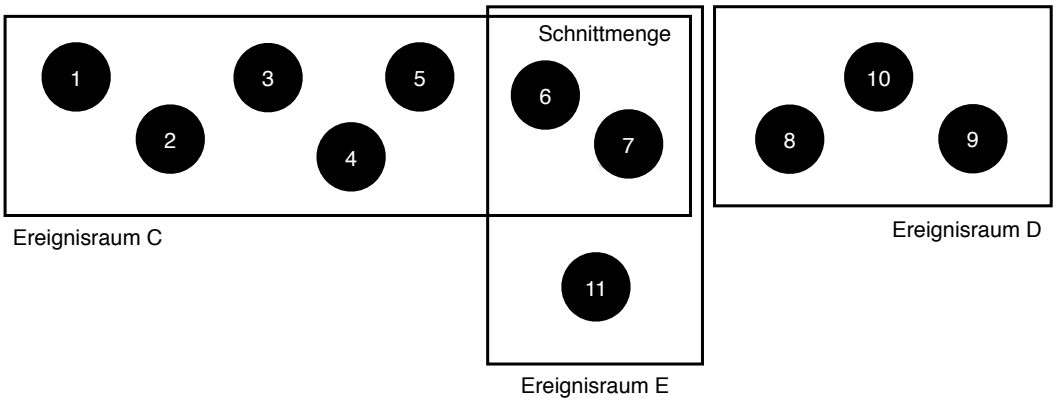


Abbildung 3.2: Darstellung von drei Ereignisräumen mit Schnittmengen

Die Elementarereignisse 6 und 7 gehören sowohl zum Ereignisraum C als auch zum Ereignisraum E. Damit sind diese beiden Ereignisräume nicht disjunkt. Demgegenüber ist der Ereignisraum D zum Ereignisraum C und E disjunkt bzw. unabhängig.

Formale Schreibweisen:

Elementarereignisse, die sowohl zu C als auch zu E gehören sind 6 und 7.

$$C \cap E = \{\ 6\ ,\ 7\ \}$$

$\cap =$ logisches Produkt;
$C \cap E$ ist die Ereignismenge, die aus der Schnittmenge von C und E besteht.
In Abbildung 3.2 besteht diese Menge aus den Elementarereignissen 6 und 7.

Dagegen gibt es keine Elementarereignisse, die sich C und D „teilen“.
C und D sind disjunkt.

$$C \cap D = \{\ \} = \varnothing \text{ (leere Menge)}$$

3.2.1 Wahrscheinlichkeit für ein Ereignis

Das Würfeln mit einem sechsäugigen Würfel ist ein klassisches Zufallsexperiment. Die Ergebnismenge Ω besteht aus sechs Elementarereignissen mit gleicher Wahrscheinlichkeit. Die Wahrscheinlichkeit das Elementarereignis {6} zu würfeln lässt sich über die so genannte Laplace-Formel berechnen:

$$p(A) = \frac{\text{Anzahl der günstigen Ereignisse (A)}}{\text{Anzahl der möglichen Ereignisse } (\Omega)} = \frac{1}{6} = 0,1\bar{6} \tag{3.1}$$

Die Wahrscheinlichkeit p (engl. für probability) für das Ereignis A ergibt sich aus der Anzahl der günstigen Ereignisse bei denen A zutrifft (in diesem Fall eins aus sechs), relativiert an der Gesamtanzahl der Elementarereignisse, die eintreffen können (in diesem Fall sechs).

Bei einem Wurf kann das Ereignis {6} entweder eintreffen oder nicht (0 oder 1). Bei zwei Würfen kann das Ereignis zweimal, einmal oder gar nicht eintreffen (2, 1, 0).
Bei n Würfen kann man die relative Häufigkeit für ein Elementarereignis ausrechnen über:

$$H(A) = \frac{n_A}{n} \tag{3.2}$$

Mit wachsendem n konvergiert H(A) auf einen konstanten Wert. Das heißt: Mit wachsenden Durchgängen wird die Fluktuationen von H(A) geringer.
In Abbildung 3.3 ist dies für das Ereignis eine Sechs zu würfeln dargestellt. Nach ungefähr 200 Durchgängen hat sich die relative Häufigkeit gegen den Wert $0,1\bar{6}$ angenähert.
H(A) konvergiert also mit wachsendem n gegen den Wert von p(A). Die Annäherung der relativen Häufigkeit an die errechnete Wahrscheinlichkeit ist bekannt als das Gesetz der großen Zahl (auch Bernoulli Theorem genannt); Je größer n desto wahrscheinlicher gilt: $H(A) = p(A)$.
In Abbildung 3.3 ist p(A) durch die konstante graue Linie gekennzeichnet.

Ein Zufallsexperiment, bei dem nur ein Ereignis betrachtet wird, das entweder zutreffen oder nicht zutreffen kann, wird auch als Bernoulli-Experiment bezeichnet, benannt nach dem Schweizer Mathematiker Jacob Bernoulli. Ein Zufallsexperiment, das aus n unabhängigen Durchgängen besteht, wird Bernoulli-Kette genannt.

Vorraussetzung der Laplace-Formel

Die Formel von Laplace setzt voraus, dass jedes Ereignis der Ereignismenge Ω mit einer gleichen Wahrscheinlichkeit auftreten kann. Es wird also eine Gleichverteilung der einzelnen Ereigniswahrscheinlichkeiten angenommen, die vor allem dann gegeben ist, wenn die einzelnen Ereignisse unabhängig voneinander sind.
Diese inhärente Bedingung der Laplace-Formel wird als Indifferenzprinzip bezeichnet und ist auch Grundlage eines „echten“ Zufallsexperiments.

Abbildung 3.3: Ergebnis eines Zufallsexperiments mit 400 Durchgängen und sechs Elementarereignissen gleicher Wahrscheinlichkeit

3.2.2 Wahrscheinlichkeit für mehrere Ereignisse

Additionstheorem *(oder)*

Die Wahrscheinlichkeit p(A) für ein Ereignis aus dem Ereignisraum A (siehe Abbildung 3.1) entspricht der Summe der Einzelwahrscheinlichkeiten für die Elementarereignisse des Ereignisraums A:

$$A = \{ 1 , 2 , 3 , 4 , 5 \}$$
$$p(A) = p(1) + p(2) + p(3) + p(4) + p(5)$$
$$p(A) = \frac{1}{10} + \frac{1}{10} + \frac{1}{10} + \frac{1}{10} + \frac{1}{10}$$
$$p(A) = \frac{5}{10} = \frac{1}{2} = 0,5$$

Die Wahrscheinlichkeit, ein Ereignis aus dem Ereignisraum A zu „ziehen“, beträgt also p = 0,5 oder 50%.
Weiterhin ergibt sich aus Abbildung 3.1:

$$p(A \cup B) = p(A) + p(B) = 1$$
(sicheres Ereignis)

$$p(A \cap B) = 0$$
(unmögliches Ereignis)

Beispiel: Die Wahrscheinlichkeit mit einem sechsäugigen Würfel eine Vier *oder* eine Sechs zu würfeln beträgt: $p = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3} = 0,\bar{3}$.

Das Ereignis eine Sechs zu würfeln schließt das Ereignis eine Vier zu würfeln aus. Die Ereignisse sind unabhängig oder disjunkt. Aus diesem Grund ist die Wahrscheinlichkeit der Summe von A und B gleich der Summe der Einzelwahrscheinlichkeiten p(A)+p(B).
Dies entspricht der Aussage des dritten Axioms der Wahrscheinlichkeit, aufgestellt von dem russischen Mathematiker A. N. Kolmogorow.

In manchen Fällen ist dies nicht der Fall. Wenn die Ereignisse nicht disjunkt sind, dass heißt Schnittmengen haben, ist die Wahrscheinlichkeit für das Eintreffen des einen Ereignis vereinbar mit dem Eintreffen von dem eines anderen Ereignis. Daher lässt sich die Wahrscheinlichkeit nicht mehr einfach durch die Addition der Einzelwahrscheinlichkeiten der Elementarereignisse errechnen.
Das Additionstheorem für nicht disjunkte Ereignisse sieht folgendermaßen aus:

$$p(A \text{ oder } B) = p(A) + p(B) - p(A \text{ und } B) \tag{3.3}$$

$$bzw.$$

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \tag{3.4}$$

Die Einzelwahrscheinlichkeiten p(A) und p(B) werden addiert und die Schnittmenge, also solche Ereignisse, bei denen sowohl A als auch B zutrifft, abgezogen.

Beispiel: Die Wahrscheinlichkeit eine Zahl zu würfeln, die durch zwei oder durch drei teilbar ist, errechnet sich folgendermaßen:

$$\text{p(durch 2 teilbar)} = \frac{3}{6}$$
$$\text{p(durch 3 teilbar)} = \frac{2}{6}$$
$$\text{p(durch 2 teilbar und durch 3 teilbar)} = \text{p(A} \cap \text{B)} = \frac{1}{6}$$
$$\text{p(durch 2 teilbar oder durch 3 teilbar)} = \text{p (A} \cup \text{B)} = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

Multiplikationstheorem *(und)*

Führt man zwei unabhängige Zufallsexperimente aus, so kann man die Wahrscheinlichkeit für das Auftreten zwei disjunkter Elementarereignisse ausrechnen:

$$p(A \cap B) = p(A) \cdot p(B) \tag{3.5}$$

Beispiel: Die Wahrscheinlichkeit bei zwei Würfen mit einem sechsäugigen Würfel zwei mal eine Sechs zu würfeln beträgt: $p = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = 0,02\bar{7}$.

bedingte Wahrscheinlichkeit

Weiterhin kann man sich fragen, wie groß die Wahrscheinlichkeit ist, ein Elementarereignis zu ziehen, wenn eine Bedingung schon erfüllt ist.
Auf Abbildung 3.2 angewandt wäre zum Beispiel eine mögliche Frage, wie groß die Wahrscheinlichkeit ist ein Objekt zu ziehen, das dem Ereignisraum C und E angehört, wenn wir schon davon ausgehen, dass ein Element aus C gezogen wurde.

$$p(E|C) = \frac{p(E) \cap p(C)}{p(C)} \tag{3.6}$$

ausgelesen: Die bedingte Wahrscheinlichkeit p(E) unter der Bedingung C

$$\mathrm{p(C)} = \frac{7}{11}$$

$$\mathrm{E} \cap \mathrm{C} = \{6, 7\}$$
$$\mathrm{p(E)} \cap \mathrm{p(C)} = \frac{2}{11}$$

$$\mathrm{p(E|C)} = \frac{\frac{2}{11}}{\frac{7}{11}} = \frac{2}{7} = 0,29$$

Bei der bedingten Wahrscheinlichkeit wird also der gesamte Ereignisraum Ω eingeschränkt.
In diesem Fall bestehen die möglichen Ereignisse nur noch aus sieben, da wir davon ausgehen, dass ein Ereignis aus Ereignisraum C schon gezogen wurde und dieser Ereignisraum aus sieben Elementarereignissen besteht.
Die möglichen bzw. gesuchten Ereignisse sind zwei, da nur zwei Elementarereignisse in die Schnittmenge von Ereignisraum C und E fallen.

Beispiel: Angenommen man will die Effektivität einer Therapiemethode untersuchen. Man befragt 112 Patienten, ob sie sich nach einem Jahr Therapie besser fühlen. Demgegenüber befragt man 96 Patienten, die unter den selben psychischen Beschwerden leiden, allerdings an keiner Therapie teilgenommen haben.
Die Ergebnisse lassen sich in folgender Tabelle anschaulich darstellen:

	Therapie (A)	**keine Therapie (B)**	**gesamt**
gesund (C)	77	38	115
nicht gesund (D)	45	58	103
gesamt	112	96	

Die Wahrscheinlichkeit, dass ein Patient angibt nach einem Jahr gesund zu sein unter der Bedingung, dass er an der Therapie teilgenommen hat, beträgt:
$p(C|A) = \frac{77}{112} = 0,6875$ oder 68,75%.

Die bedingte Wahrscheinlichkeit, dass ein Patient angibt nach einem Jahr gesund zu sein unter der Bedingung, dass er nicht an einer Therapie teilgenommen hat, beträgt dagegen:
$p(C|B) = \frac{38}{96} = 0,3958$ oder 39,58%.

3.A Aufgaben

inhaltliche Aufgaben

1. Bei welcher Art von Zusammenhängen ist die Wahrscheinlichkeitstheorie wichtig für die Statistik?
2. Was sind die Bedingungen für ein Zufallsexperiment?
3. Was ist ein Bernoulli-Experiment?
4. Was sind die Kolmogorow-Axiome der Wahrscheinlichkeitstheorie?
5. Was besagt das Gesetz der großen Zahl? Was ist die praktische Konsequenz daraus?
6. Was geschieht mit dem Ereignisraum Ω bei der bedingten Wahrscheinlichkeit?
7. Was bedeutet es, wenn zwei Ereignisräume disjunkt sind?
8. Was sind disjunkte bzw. nicht disjunkte Ereignisse?
9. Was ist das Symbol für die logische Summe? Was für das logische Produkt?
10. Was ist der Unterschied zwischen Additions- und Multiplikationstheorem?
11. Wie kann man die Wahrscheinlichkeit p(A) berechnen, wenn ein Zufallsversuch endlich viele, gleich mögliche Ausgänge hat?
12. In welchem Bereich können Wahrscheinlichkeiten liegen?
13. Was ist das sichere Ereignis?
14. Wann sind Ereignisse unabhängig voneinander?
15. Was besagt das Indifferenzprinzip?
16. Was gilt für die Auftretenswahrscheinlichkeit von Elementarereignissen, wenn das Indifferenzprinzip nicht gilt?

praktische Aufgaben

1. Wie groß ist die Wahrscheinlichkeit, mit einem Würfel entweder eine 2, 3, 4 oder 5 zu würfeln?
2. Wie groß ist die Wahrscheinlichkeit, dass ein Paar zusammen vier Kinder bekommt und alle männlich sind?
3. Wie groß ist die Wahrscheinlichkeit, dass das fünfte Kind des Paars ebenfalls ein Junge wird?
4. Wie groß ist die Wahrscheinlichkeit, dass eine zufällig aus einem Skatblatt gezogene Karte entweder schwarz oder ein König ist?

5. Wie groß ist die Wahrscheinlichkeit, dass Weihnachten (24.12.) in einem bestimmten Jahr auf einen Werktag fällt?

6. Wie groß ist die Wahrscheinlichkeit, aus einer Urne mit drei grünen, kleinen Kugeln, vier gelben, mittleren Kugeln und fünf roten, großen Kugeln eine kleine oder eine rote Kugel zu ziehen?

7. Wie groß ist die Wahrscheinlichkeit, dass man mit einem Würfel eine Sechs würfelt und aus einem Skatblatt eine rote Karte zieht?

8. Ein Bäcker backt Kreppel mit Marmeladenfüllung. Er benutzt mal Erdbeer- und mal Pfirsichmarmelade. Nun stellt er fest, dass die Verkaufszahlen oft schwanken. Er beobachtet 40 Tage lang, wie viele Kreppel er verkauft und bekommt folgendes Ergebnis:

	Erdbeermarmelade	**Pfirsichmarmelade**	**gesamt**
verkauft	610	290	900
nicht verkauft	40	60	100
gesamt	650	350	1000

   a) Wie groß ist die Wahrscheinlichkeit, einen Kreppel zu verkaufen unter der Bedingung, dass er eine Erdbeermarmeladenfüllung hat?
   b) Wie groß ist die Wahrscheinlichkeit, einen Kreppel zu verkaufen unter der Bedingung, dass er eine Pfirsichmarmeladenfüllung hat?

3.3 Kombinatorik

Neben der Frage, wie wahrscheinlich es ist, dass ein Ereignis auftritt, ist es für die Statistik oft wichtig zu wissen wie wahrscheinlich eine gewisse Abfolge von Ereignissen ist.
Die Kombinatorik, als Teilgebiet der Mathematik, beschäftigt sich mit der Bestimmung von möglichen Ereignisabfolgen und deren Wahrscheinlichkeit. Ergebnisse der Kombinatorik sind oftmals kontraintuitiv. Das bedeutet viele Menschen über- oder unterschätzen Wahrscheinlichkeiten oder Möglichkeitskombinationen aus diesem Bereich.
Dazu ein Beispiel, das als so genanntes Geburtstagsparadoxon bekannt ist:
Wie hoch würdest du die Wahrscheinlichkeit einschätzen, dass bei 23 anwesenden Personen (dies würde zwei Fußballmannschaften inklusive Schiedsrichter entsprechen) zwei von ihnen am gleichen Tag Geburtstag haben (ungeachtet des Geburtsjahres)?
Zur exakten Beantwortung dieser Frage benötigt man Methoden der Kombinatorik. Teilbereiche der Kombinatorik sind Permutation, Variation und Kombination.

3.3.1 Permutation

Unter Permutation (von lat. vertauschen) versteht man die Neuanordnung von Elementen, indem diese in ihrer Reihenfolge vertauscht werden.
Will man herausfinden, wie viele verschiedene Anordnungsmöglichkeiten es für eine Menge von n Objekten gibt, so kann man dies über die Fakultät von n ausrechnen:

$$n! \tag{3.7}$$

Beispiel: Vier verschieden farbige Karten können in 24 verschiedenen Kombinationen nacheinander auf einen Tisch gelegt werden. Errechnen kann man dies über:
$n! = 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$.
Die Wahrscheinlichkeit, eine ganz bestimmte Kombination zufällig zu legen beträgt damit $p = \frac{1}{24}$.

In manchen Fällen gibt es Gruppen von Objekten, die nicht unterscheidbar sind oder Vertauschungen innerhalb einer Gruppe gleicher Objekte nicht von Interesse ist. Um nun herauszufinden, wie viele Permutationsmöglichkeiten bei denen Vertauschungen von Elementen verschiedener Gruppen auftauchen, muss man die gesamte Anzahl an möglichen Vertauschungen relativieren an solchen Permutationen, bei denen Elemente zwischen den Gruppen neu angeordnet werden:

$$\frac{n!}{n!_{grp1} \cdot n!_{grp2} \cdot n!_{grp3}} \tag{3.8}$$

Beispiel: Statt vier verschieden farbige Karten haben wir nun drei weiße Karten, vier gelbe Karten und sieben rote Karten. Insgesamt lassen sich die 14 Karten in $14! = 87178291200$ verschiedenen Möglichkeiten auf den Tisch legen. Dabei kommt es oft vor, dass lediglich gleichfarbige Karten den Platz tauschen. Um diese Permutationen herauszurechnen und nur die Anzahl an Legemöglichkeiten zu bekommen, bei denen Karten aus verschiedenen Farbgruppen den

Platz tauschen bedienen wir uns der Formel 3.8:

$$\frac{14!}{3! \cdot 4! \cdot 7!} = \frac{87.178.291.200}{6 \cdot 24 \cdot 5040} = \frac{87.178.291.200}{725.760} = 120.120$$

Insgesamt gibt es 120.120 Anordnungsmöglichkeiten, bei denen keine gleich erscheint. 725.760 ist dagegen die Anzahl der Anordnungen, bei denen lediglich Karten der gleichen Gruppe die Plätze tauschen.

3.3.2 Kombination

Bei der Kombination geht es darum, aus einer Gesamtmenge von Elementen eine bestimmte Anzahl zu entnehmen und alle möglichen Zusammensetzungen auszurechnen. Im Gegensatz zur Permutation werden also nicht alle Elemente ausgewählt und deren Kombinationsmöglichkeiten errechnet.

Kombination mit Wiederholungen

Bei der Kombination mit Wiederholungen werden alle möglichen Auswahlen über folgende Formel errechnet. Hierbei steht n für die Gesamtanzahl der Elemente und k für die Anzahl der Ziehungen oder Durchgänge.

$$\frac{(n+k-1)!}{(n-1)! \cdot k!} \tag{3.9}$$

Beispiel: Aus einer Schale mit vier nummerierten Kugeln sollen zwei Kugeln zufällig gezogen werden, wobei nach jeder Ziehung die gezogene Kugel wieder zurückgelegt wird. Es ist also auch möglich eine Kugel zweimal hintereinander zu ziehen.

Die Gesamtmenge besteht aus vier Elementen (n):

1, 2, 3, 4

Es werden jeweils zwei Kugeln gezogen (k).
Daraus ergeben sich folgende Kombinationsmöglichkeiten:

11, 12, 13, 14, 22, 23, 24, 33, 34, 44

$$\frac{(n+k-1)!}{(n-1)! \cdot k!} = \frac{(4+2-1)!}{(4-1)! \cdot 2!} = \frac{120}{12} = 10$$

Kombination ohne Wiederholungen

Obige Formel errechnet die möglichen Kombination bei gleichbleibender Ausgangsmenge n über k hinweg. Auf das oben stehende Beispiel bezogen ist dies nur der Fall, wenn nach jeder Ziehung einer Kugel diese wieder in die Schale zurückgelegt wird.

Ist dies nicht der Fall, so sind die möglichen Kombinationen geringer und lassen sich über unten stehende Formel errechnen.

$$\frac{n!}{(n-k)! \cdot k!} \tag{3.10}$$

Der Term aus 3.10 wird auch als Binomialkoeffizient bezeichnet und lässt sich durch das Symbol in 3.11 abkürzen (gelesen: n über k).

$$\binom{n}{k} \quad \text{entspricht} \quad \frac{n!}{(n-k)! \cdot k!} \tag{3.11}$$

Beispiel: Aus einer Schale mit vier nummerierten Kugeln sollen zwei Kugeln zufällig gezogen werden, wobei diesmal die gezogene Kugel nach einer Ziehung nicht mehr zurückgelegt wird.

Die Gesamtmenge besteht aus vier Elementen (n):

1, 2, 3, 4

Es werden jeweils zwei Kugeln gezogen (k).
Daraus ergeben sich folgende Kombinationsmöglichkeiten:

12, 13, 14, 23, 24, 34

$$\frac{n!}{(n-k)! \cdot k!} = \frac{4!}{(4-2)! \cdot 2!} = \frac{24}{4} = 6$$

Die möglichen sechs Kombinationen umfassen alle Ereignisse, die schon im Beispiel zur Kombination mit Wiederholungen aufgelistet sind, außer den Kombinationen, in denen ein Element zwei mal hintereinander auftaucht.

3.3.3 Variation

Mithilfe der Kombinationsformeln lassen sich die möglichen Zusammenstellungen von Elementen errechnen. Dabei wird die Reihenfolge außer Acht gelassen.
Mithilfe der Variationsformeln errechnet man statt den möglichen Kombinationen einer Menge von Elementen zusätzlich deren Varianten. So kann die Kombination 12 in zwei Varianten auftauchen: in 12 und 21.

Variation mit Zurücklegen

Die möglichen Varianten bei gleichbleibender Gesamtmenge n über k hinweg errechnen sich folgendermaßen:

$$n^k \tag{3.12}$$

Beispiel: Aus einer Schale mit vier nummerierten Kugeln sollen zwei Kugeln zufällig gezogen werden, wobei nach jeder Ziehung die gezogene Kugel wieder zurückgelegt wird. Es ist also

auch möglich eine Kugel zweimal hintereinander zu ziehen.

Die Gesamtmenge besteht aus vier Elementen (n):

1, 2, 3, 4

Es werden jeweils zwei Kugeln gezogen (k).
Daraus ergeben sich folgende Kombinationsmöglichkeiten:

11, 12, 13, 14, 22, 23, 24, 33, 34, 44

Beachtet man nun noch die Reihenfolge, so ergeben sich folgende Ereignisvarianten:

11, 12, 21, 13, 31, 14, 41, 22, 23, 32, 24, 42, 33, 34, 43, 44

$$n^k = 4^2 = 16$$

Variation ohne Zurücklegen

Wird die Gesamtmenge n kleiner bei größer werdendem k, so errechnen sich die möglichen Variation über:

$$\frac{n!}{(n-k)!} \tag{3.13}$$

Beispiel: Aus einer Schale mit vier nummerierten Kugeln sollen zwei Kugeln zufällig gezogen werden, wobei diesmal die gezogene Kugel nach einer Ziehung nicht mehr zurückgelegt wird.

Die Gesamtmenge besteht aus vier Elementen (n):

1, 2, 3, 4

Es werden jeweils zwei Kugeln gezogen (k).
Daraus ergeben sich folgende Kombinationsmöglichkeiten:

12, 13, 14, 23, 24, 34

Beachtet man nun noch die Reihenfolge, so ergeben sich folgende Ereignisvarianten:

12, 21, 13, 31, 14, 41, 23, 32, 24, 42, 34, 43

$$\frac{n!}{(n-k)!} = \frac{4!}{(4-2)!} = \frac{24}{2} = 12$$

3.3.4 Zusammenfassung

In Tabelle 3.1 sind die Regeln der Kombinatorik zusammengefasst.

Tabelle 3.1
Formelsammlung der Kombinatorik

	Permutation	Kombination	Variation
mit Wiederholung / mit Zurücklegen	$n!$	$\frac{(n+k-1)!}{(n-1)! \cdot k!}$	n^k
ohne Wiederholung / ohne Zurücklegen	$\frac{n!}{n!_1 \cdot n!_2 \cdot n!_3}$	$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$	$\binom{n}{k} \cdot k! = \frac{n!}{(n-k)!}$

3.3.5 Nachtrag: Das Geburtstagparadoxon

Mit dem Wissen über die Regeln der Kombinatorik und der Wahrscheinlichkeitsrechnung können wir nun die eingangs gestellte Frage beantworten; Die Wahrscheinlichkeit, dass bei 23 anwesenden Personen zwei von ihnen am gleichen Tag Geburtstag haben liegt bei knapp über 50% bzw. $p = 0,5$.
Die Anzahl aller möglichen Geburtstagskombinationen errechnet sich über die erste Variationsformel:

$$365^{23}$$

Um die Anzahl der Geburtstagskombinationen zu errechnen, bei denen alle 23 Personen unterschiedliche Geburtstage haben, bedienen wir uns der zweiten Variationsregel:

$$\frac{365!}{(365-23)!}$$

Die Wahrscheinlichkeit, dass alle Personen an unterschiedlichen Tagen Geburtstag haben, lässt sich nun über die Laplace-Formel errechnen, indem die günstigen durch alle möglichen Ereignisse geteilt werden:

$$p(A) = \frac{\frac{365!}{(365-23)!}}{365^{23}}$$

Das sichere Ereignis hat nach dem zweiten Kolmogorow-Axiom den Wert 1. Das Gegenereignis (mindestens zwei Personen haben an einem selben Tag Geburtstag) errechnet sich damit über:

$$p(\bar{A}) = 1 - \frac{\frac{365!}{(365-23)!}}{365^{23}} = 1 - \frac{343 \cdot 344 \cdot \ldots \cdot 364 \cdot 365}{365^{23}} = 1 - 0,4927 = 0,5072$$

3.B Aufgaben

inhaltliche Aufgaben

1. Wozu benötigt man Kombinatorik?
2. Wie hängt Kombinatorik mit der Laplace-Formel zusammen?
3. Was ist der Unterschied zwischen Permutation und Variation?
4. Was ist der Unterschied zwischen Variation und Kombination?
5. Für was steht „n über k“?

praktische Aufgaben

1. Wie viele mögliche Ereignisabfolgen ergeben sich, wenn man sieben mal würfelt?
2. Wie groß ist die Wahrscheinlichkeit drei mal eine Sechs oder drei mal eine Eins zu würfeln?
3. 15 Schüler sollen sich hintereinander aufstellen. Wie viele verschiedene Möglichkeiten gibt es?
4. Wie viele verschiedene Reihenfolgen ergeben sich, wenn man aus fünf Kindern zwei auswählt und sie nebeneinander aufstellt?
5. Wie viele verschiedene Worte kann man aus dem Wort „Statistik“ machen, indem man die Buchstaben wild vertauscht? Wie groß ist die Wahrscheinlichkeit für jede der sich ergebenden Anordnungen?
6. Wie viele Ankreuzkombinationen sind beim Lotto (6 aus 49) möglich?
7. Ein Junge spielt mit seiner Modelleisenbahn. Er hat 15 Wagen und drei Lokomotiven. Hinter die erste will er fünf, hinter die zweite vier und hinter die dritte sechs Wagen hängen. Wie viele Möglichkeiten hat er dazu?
8. In einer Untersuchung zur Gedächtniskapazität interessiert die Forscher, ob sich die Versuchspersonen aus 60 Wörtern sechs ganz bestimmte gemerkt haben, wobei die Reihenfolge der Wörter dabei unwichtig ist. Wäre diese Leistung komplett zufallsabhängig, wie wahrscheinlich wäre es, genau diese sechs Wörter zu reproduzieren?
9. Wie groß ist die Wahrscheinlichkeit, aus vier Kartenspielen hintereinander die vier Asse zu ziehen (es wird also pro Kartenspiel einmal gezogen)?
10. Von sieben Glasperlen (vier rote, eine grüne und zwei weiße) sollen vier zusammen angeordnet werden, eine soll alleine bleiben und wiederum zwei sollen zusammen angeordnet werden.
    a) Wie viele Möglichkeiten gibt es?
    b) Wie groß ist die Wahrscheinlichkeit, dass die vier roten zusammen, die grüne alleine und die zwei weißen wieder zusammen angeordnet werden?

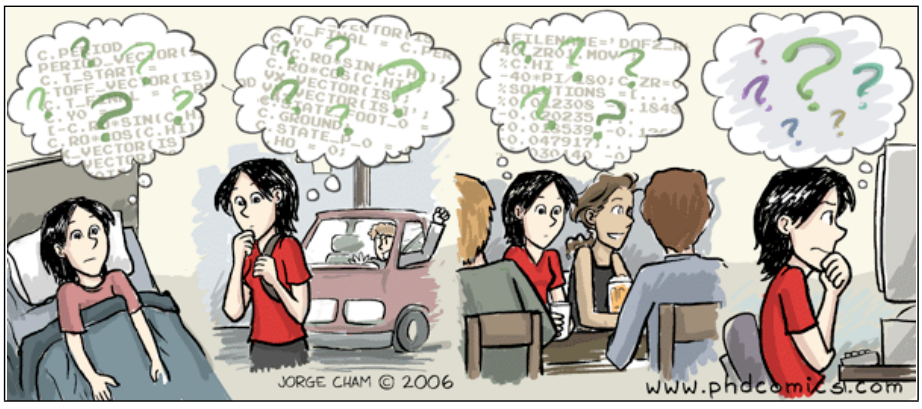
11. Frau Müller will ein Eis kaufen. Sie kann zwischen drei Waffelsorten, 25 Eissorten und sechs Arten von Schokostreuseln wählen.
    a) Wie viele verschiedene Kombinationen sind möglich?
    b) Wie groß ist die Chance, dass Frau Müller genau die Kombination wählt, die sie wählt?

12. Beim Riesenslalom im Skifahren starten 26 Skiläufer, die im Training alle gleich schnell waren.
    a) Wie groß ist die Wahrscheinlichkeit, dass sich die Fahrer (ausschließlich zufallsabhängig) genau in ihrer Startreihenfolge platzieren?
    b) Wie groß ist die Wahrscheinlichkeit, dass Fahrer 22 erster wird, Fahrer 18 zweiter und Fahrer 5 dritter?

13. Wie groß ist die Wahrscheinlichkeit, bei 10-maligem Würfeln in den ersten drei Würfen die Zahlen 1, 2 und 3 zu würfeln?

14. Zwölf Personen sollen nach der Ausprägung des Merkmals „Attraktivität“ in eine Rangreihe gebracht werden. Wie viele verschiedene Ergebnisse sind theoretisch denkbar?

15. In Urne Nr. 1 liegen sieben Kugeln, in Urne Nr. 2 liegen zwei Kugeln und in Urne Nr. 3 liegen drei Kugeln. Die Kugeln haben alle unterschiedliche Farben. Wie viele verschiedene Ereignisabfolgen sind möglich, wenn aus jeder Urne jeweils eine Kugel gezogen wird?



3.4 Wahrscheinlichkeit für eine Folge von Ereignissen

Die Wahrscheinlichkeit zwei mal eine Sechs zu würfeln, lässt sich nach dem Multiplikationstheorem errechnen über: $\frac{1}{6} \cdot \frac{1}{6} = 0,02\bar{7}$.
Alternativ kann man die möglichen Ereignisse errechnen über: $6^2 = 36$ und die Wahrscheinlichkeit über die Formel von Laplace bestimmen: $\frac{1}{36} = 0,02\bar{7}$.

Doch wie groß ist die Wahrscheinlichkeit bei zwei Würfen genau einmal eine Sechs zu würfeln? Die Fragestellung unterscheidet sich von den bisherigen dadurch, dass eine Bedingung gestellt wird. Es soll genau k-mal ein bestimmtes Ereignis bei n-Durchgängen errechnet werden.
Die möglichen Ereignisse lassen sich nach wie vor durch $6^2 = 36$ errechnen.
Die günstigen Ereignisse können in diesem einfachen Fall noch aufgezählt werden:

erster Wurf	1	2	3	4	5	6	6	6	6	6
zweiter Wurf	6	6	6	6	6	1	2	3	4	5

Die Wahrscheinlichkeit genau einmal bei zwei Würfen eine Sechs zu würfeln liegt also bei:

$$\frac{10}{36} = 0,2\bar{7}$$

Verallgemeinert lässt sich die Wahrscheinlichkeit für das k-malige Auftreten eines Ereignis mit der konstanten Wahrscheinlichkeit p bei n Durchgängen mithilfe der unten stehenden Formel bestimmen.

$$p(k) = \binom{n}{k} p^k \cdot (1-p)^{n-k} \tag{3.14}$$

Wir errechnen damit die Auftretenswahrscheinlichkeit eines Ereignis über mehrere Durchgänge hinweg, wobei die Durchgänge n einer Bernoulli-Kette bzw. mehreren Bernoulli-Experimenten entsprechen:
Die Ausgangswahrscheinlichkeit bleibt gleich, was bedeutet, dass bei jedem Durchgang n die Ereignismenge Ω konstant groß ist und die einzelnen Durchgänge voneinander unabhängig sind. Zudem interessiert lediglich, ob bei jedem Durchgang ein Ereignis eintrifft oder nicht (0 oder 1).

Beispiel

Angenommen man will einen 14-tägigen Urlaub in Schottland machen. Die Wahrscheinlichkeit zu regnen betrage pro Tag 70% bzw. $p = 0,7$. Die Wahrscheinlichkeit, dass es an zehn von den 14 Tagen regnet, lässt sich mit Formel 3.14 errechnen:

$$p(k) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

$$= \frac{14!}{(14-10)! \cdot 10!} \cdot 0,7^{10} \cdot (1-0,7)^{14-10}$$

$$= \frac{11 \cdot 12 \cdot 13 \cdot 14}{4!} \cdot 0,0282 \cdot 0,0081$$

$$= 0,2286$$

Die Wahrscheinlichkeit, dass zehn der 14 Tage ins Wasser fallen liegt also bei 22,86%.

3.4.1 Binomialverteilung

Generieren wir uns eine Variable x, die Werte von 0 bis n annimmt und setzen diese für k ein, so erhalten wir eine Reihe von Einzelwahrscheinlichkeiten p(x) für das k-malige Auftreten eines Ereignisses bei gleichbleibenden Vorraussetzungen.
Die Variable x wird auch als Zufallsvariable bezeichnet und folgt einer bestimmten Verteilung, der so genannten Binomialverteilung, sofern obige Bedingungen einer Bernoulli-Kette erfüllt sind. Die Binomialverteilung ist eine diskrete Wahrscheinlichkeitsverteilung.
Man sagt die Zufallsvariable x ist binomialverteilt oder folgt einer Binomialverteilung.

Dies lässt sich verdeutlichen, wenn man für obiges Beispiel nicht nur die Wahrscheinlichkeit für k=10 Tage, sondern für $k = 0$ bis $k = n$ (also 14) errechnet. Die Einzelwahrscheinlichkeiten und deren Verteilung ist in Abbildung 3.4 dargestellt.

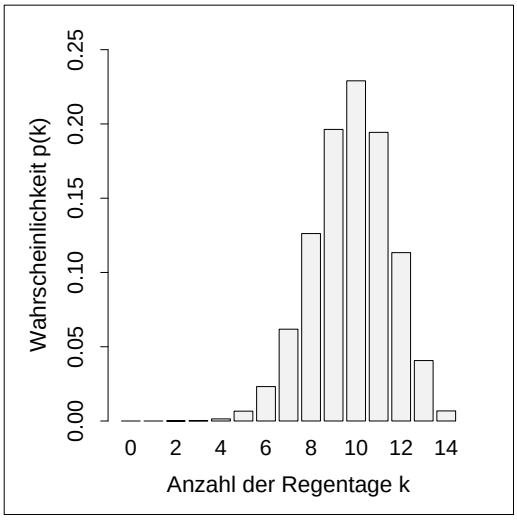


Abbildung 3.4: Binomialverteilung für $p = 0,7$

Die Binomialverteilung ist rechtsschief für $p < 0,5$ und linksschief bei $p > 0,5$. Ist $p = 0,5$ ist die Verteilung symmetrisch (siehe Abbildung 3.5.b).

Bei $n \to \infty$ konvergiert die Binomialverteilung gegen die Normalverteilung (siehe Abbildung 3.5 bzw. Abbildung 2.1). Dies gilt auch bei $p \neq 0,5$.

Anhand der Verteilung in Abbildung 3.4 lässt sich auch ablesen, wie groß die Wahrscheinlichkeit ist, dass es in dem Schottlandurlaub zehn *oder mehr* Tage regnet. Dafür müssen lediglich die Wahrscheinlichkeiten für $k = \{10, 11, 12, 13, 14\}$ aufsummiert werden:

$$p(k \geq 10) = 0,2286 + 0,1943 + 0,1134 + 0,0407 + 0,0068 = 0.5838$$

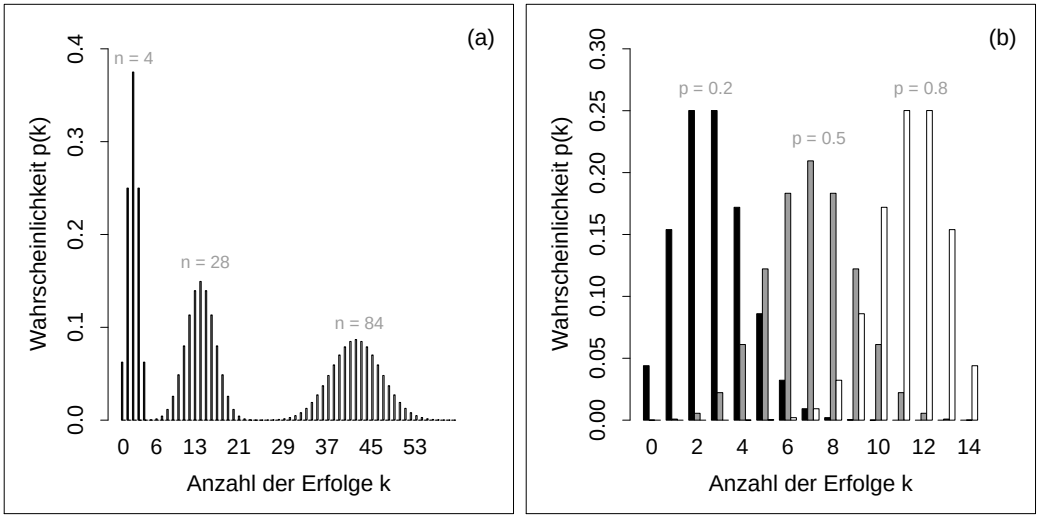


Abbildung 3.5: Binomialverteilung bei (a) konstanter Wahrscheinlichkeit $p = 0,5$ und variierendem n und (b) konstanten Durchgängen $n = 14$ und variierendem p

Gegenereignis

Des Weiteren kann man sich auch fragen, wie groß die Wahrscheinlichkeit ist, dass es weniger als zehn Tage regnet. Dies ist das so genannte Gegenereignis. Die Menge k wird dabei „umgedreht“:

$$k = \{10, 11, 12, 13, 14\} \quad \rightarrow \quad k = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

Und es gilt:

$$\text{Ereignis} + \text{Gegenereignis} = 1 \quad \text{bzw.} \quad p(k) + p(\bar{k}) = 1$$

Komplementärereignis

Bedient man sich Wahrscheinlichkeitstabellen (wie die im Bortz), so sind oft Wahrscheinlichkeiten über $p > 0,5$ nicht eingetragen, da, wie schon angesprochen, die Binomialverteilung bei hinreichend großem n symmetrisch wird und daher die Balken links von der Verteilung die selbe Höhe haben wie rechts von der Verteilung.

Will man dennoch Tabellen benutzen, statt k-mal die Formel auszurechnen, so kann man bei $p > 0,5$ statt dem gesuchten Ereignis das Komplementärereignis bzw. die Komplementärereignisse errechnen.
Dies geschieht, indem man die Frage hin zum Komplement umstellt. Für das Beispiel oben entspräche das Komplement der Wahrscheinlichkeit, dass es mehr als vier Tage *nicht* regnet.

Die einzelnen Komplementärereignisse sehen wie folgt aus:

0 Tage Regen	→	14 Tage kein Regen
1 Tage Regen	→	13 Tage kein Regen
2 Tage Regen	→	12 Tage kein Regen
3 Tage Regen	→	11 Tage kein Regen
4 Tage Regen	→	10 Tage kein Regen
5 Tage Regen	→	9 Tage kein Regen
6 Tage Regen	→	8 Tage kein Regen
7 Tage Regen	→	7 Tage kein Regen
8 Tage Regen	→	6 Tage kein Regen
9 Tage Regen	→	5 Tage kein Regen

Die Ereignismenge k besteht also aus {5, 6, 7, 8, 9, 10, 11, 12, 13, 14}. Das Ereignis ist nun nicht mehr *„Regen“*, sondern *„kein Regen“*. Daher ändert sich die Wahrscheinlichkeit von $p = 0,7$ zu $p = 1 - 0,7 = 0,3$.

Die Wahrscheinlichkeiten für k sind:

$$0,1963 + 0,1262 + 0,0618 + 0,0231 + 0,0066 + 0,0014 + 0,0002 + 0 + 0 + 0 = 0,4156$$

Das Komplementärereignis hat immer die selbe Wahrscheinlichkeit wie das gesuchte Ereignis. Damit gilt: p(mehr als vier Tage kein Regen) = p(weniger als zehn Tage Regen).

3.C Aufgaben

inhaltliche Aufgaben

1. Was ist eine Bernoulli-Kette? Was hat diese mit der Binomialverteilung zu tun?
2. Überlege dir psychologische Fragestellungen, bei denen die Binomialverteilung eine Rolle spielen könnte.
3. Wie errechnet sich die Wahrscheinlichkeit, dass von n Versuchen genau k mal ein Ereignis auftritt, welches in jedem der n Versuche mit einer Wahrscheinlichkeit von p auftritt?
4. Unter welchen Bedingungen ist die Binomialverteilung annähernd normalverteilt?
5. Die Wahrscheinlichkeit, dass man bei einem multiple choice Test eine Frage per Zufall löst möge 1/4 betragen. Insgesamt gibt es 10 Aufgaben. Die Wahrscheinlichkeit mindestens zwei Aufgaben per Zufall zu lösen beträgt damit 75,60%.
   a) Formuliere das Komplementärereignis.
   b) Formuliere das Gegenereignis.
   c) Wie groß ist die Wahrscheinlichkeit für das Gegenereignis?
   d) Wie groß ist die Wahrscheinlichkeit für das Komplementärereignis?

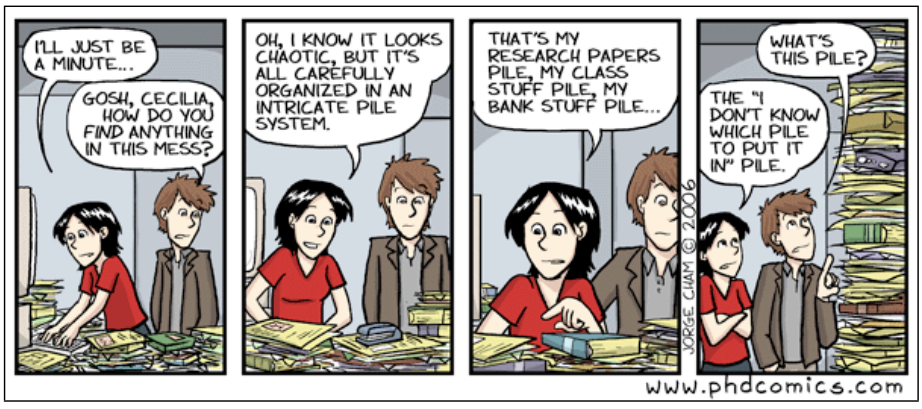
praktische Aufgaben

1. In einem entwicklungspsychologischen Experiment werden 20 Kinder gebeten, ein Puzzle zu lösen. Die Wahrscheinlichkeit, dass Puzzle zu lösen betrage $p = 0{,}4$.
   a) Wie wahrscheinlich ist es, dass 10 Kinder das Puzzle lösen?
   b) Wie wahrscheinlich ist es, dass mehr als 10 Kinder das Puzzle lösen?
2. Ein Primat soll in 18 Durchgängen aus vier vorgelegten Wörtern (Adjektiv, Substantiv, Verb, Artikel) jeweils auf das Verb deuten. Wie groß ist die Wahrscheinlichkeit von mindestens zehn richtigen Wahlen?
3. Aus Erfahrungen in der Vergangenheit ist bekannt, dass von einer unselektierten Bewerbergruppe für eine Ausbildung 30% das Ausbildungsziel nicht erreichen. Wie groß ist die Wahrscheinlichkeit, dass aus einer Gruppe von 18 Bewerbern genau 6 Bewerber das Ausbildungsziel nicht erreichen?
4. In einem wahrnehmungspsychologischen Experiment wird geprüft, welchem von zwei geometrischen Mustern A und B ein Kind mehr Aufmerksamkeit schenkt. Wie hoch ist die Wahrscheinlichkeit, dass ein Kind bei 20 Darbietungen mehr als zwölfmal Muster A auswählt?
5. In einer Meinungsumfrage werden 20 Personen gefragt, welche Koalition sie heute wählen würden. Aus früheren Umfragen ist bekannt, dass Rot-Grün derzeit von 40% der Leute gewählt wird und Schwarz-Gelb eine Chance von 60% besitzt gewählt zu werden. Wie groß ist dann die Chance, dass Schwarz-Gelb mindestens fünf Mal von den 20 Personen gewählt wird?

6. Ein Arbeits- und Betriebspsychologe wird bei einem großen Automobilhersteller angestellt. Er soll dort für mehr Arbeitssicherheit sorgen. Er findet heraus, dass die meisten Unfälle sich beim Verladen von Stahlrohren auf Karren ereignen. Die Wahrscheinlichkeit, dass ein Unfall passiert, beträgt pro Arbeitstag 0,05. Wie groß ist die Wahrscheinlichkeit, dass ...
   a) während einer Woche (5 Arbeitstage) mindestens ein Unfall auftritt?
   b) während eines Monats (20 Arbeitstage) höchstens ein Unfall auftritt?

7. Wie hoch ist die Wahrscheinlichkeit, dass unter 100 Roulette-Würfen mindestens dreimal die Zahl 13 fällt?

8. Wie groß ist die Wahrscheinlichkeit, beim 20-maligen Drehen von einem Glücksrad mindestens 15 mal eine Niete, deren Flächenanteile 70% des Glücksrads ausmachen, zu erspielen?



4
Inferenzstatistik

Die Inferenzstatistik, auch induktive Statistik genannt, versucht auf Grundlage der Stochastik und mithilfe von Wahrscheinlichkeitsverteilungen, Stichprobenergebnisse auf ihre Bedeutsamkeit und statistische Relevanz hin zu beurteilen. Konkret geht es dabei oftmals darum zu entscheiden, ob die Ausprägung eines Stichprobenkennwertes rein zufällig entstanden ist oder ob die gemessene Merkmalsausprägung aufgrund eines systematischen Einflusses entstanden ist und der reine Zufall keine gute Erklärung für den Grad der Ausprägung bietet.

Man geht davon aus, dass jeder Stichprobe eine Population zugrunde liegt. Die Population (oder Grundgesamtheit) ist die Gesamtheit aller Merkmalsausprägungen. Eine Stichprobe ist definiert als eine Menge zufällig gezogener Elemente dieser Grundgesamtheit.
Um in einer empirischen Untersuchung dieser Definition, deren wichtiger Bestandteil die zufällige Ziehung ist, gerecht zu werden, muss gewährleisten sein, dass jedes Objekt der Population die selbe Chance hat in die Stichprobe aufgenommen zu werden. Ist dies nicht der Fall, so spricht man von einer Verzerrung der Stichprobe (engl.: bias).

Mithilfe inferenzstatistischer Tests kann man beurteilen, ob eine Stichprobe einer bestimmten Population angehört oder ob zwei Stichproben aus der selben oder verschiedenen Grundgesamtheiten entstammen.
Diese Tests sind so aufgebaut, dass zunächst Annahmen oder Vermutungen über die unbekannte Population aufgestellt werden. Auf dieser Grundlage lässt sich die Wahrscheinlichkeit errechnen, mit der man annehmen kann, dass die erhobenen Stichprobenwerte dieser Population angehören.

Die Logik der inferentiellen Statistik lässt sich mithilfe des Beispiels von Kapitel 3.4.1 veranschaulichen. Angenommen in Abbildung 3.4 ist die Regenwahrscheinlichkeit für einen definierten Zeitraum von 14 Tagen abgetragen. Die Werte stammen aus vielen Messungen, die Jahr für Jahr während dieses Zeitraums gemacht wurden, so dass man davon ausgehen kann, dass diese Werte annähernd wie die Grundgesamtheit aller Werte verteilt sind.
Nun berichtet jemand, während dieses Zeitraums Urlaub gemacht zu haben. Er erzählt, dass es insgesamt sechs Tage von den 14 Tagen geregnet hat. Anhand von Abbildung 3.4 können wir mit einer relativer Sicherheit feststellen, dass er nicht an dem Ort Urlaub gemacht hat, von dem die Werte der Abbildung stammen.
Anders formuliert: Mit einer Irrtumswahrscheinlichkeit von $p = 0,0315$ kann man davon ausgehen, dass die Beobachtungen des Urlaubers (sechs Tage Regen, acht Tage kein Regen) an einem anderen Ort gemacht wurden, bei dem die Verteilung der Regentage sich bedeutsam von der uns bekannten unterscheidet.

Symbolische Abgrenzung der Population von der Stichprobe

Um mathematische Aussagen über die Population von Aussagen über die Stichprobe zu trennen, werden für die Populationsparameter, auf die man zu schließen versucht, griechische Symbole verwendet. Für den Populationsmittelwert wird das kleine griechische Mü (μ) verwendet. Für die Standardabweichung wird das kleine Sigma (σ) bzw. für die Varianz das Quadrat von Sigma (σ^2) verwendet. Sind die Populationskennwerte unbekannt, so lassen sich diese lediglich schätzen. Dies wird oft, wie in der Regressionsanalyse, mit einem Dach über dem Parameterbuchstaben symbolisiert: $\hat{\mu}$, $\hat{\sigma}$, $\hat{\sigma}^2$.

4.1 Von der Population zur Stichprobe und zurück

Will man Aussagen über eine unbekannte Population und deren deskriptive Kennwerte machen, so ist es von großer Bedeutung zu verstehen, was genau passiert, wenn man eine Stichprobe dieser Population zieht. Wichtig wäre beispielsweise zu wissen, mit welcher Wahrscheinlichkeit der Mittelwert $\bar{x}$ einer Stichprobe dem tatsächlichen Mittelwert μ entspricht.

4.1.1 Schätzung des Populationsmittelwerts

Mithilfe von Simulationen lässt sich das Verhalten von Stichprobenkennwerten untersuchen. Angenommen wir generieren uns eine Population von Werten bzw. Merkmalsausprägungen. Aus diesen Werten ziehen wir nun zufällig eine Stichprobe der Größe n und berechnen den Mittelwert. Anschließend ziehen wir wieder eine Stichprobe der Größe n, wobei die zuvor gezogenen Werte wieder zurückgelegt werden. Auch von dieser Stichprobe berechnen wir den Mittelwert. Dieses ziehen mit zurücklegen wiederholen wir tausend mal. So erhalten wir tausend Mittelwerte unabhängig gezogener Stichproben.
Die Verteilung, die sich aus diesem Simulationsexperiment ergibt, wird Stichprobenkennwerteverteilung genannt. In unserem Fall handelt es sich um die Stichprobenkennwerteverteilung des Mittelwertsmaß.

In Abbildung 4.1.a sind Stichprobenkennwerteverteilungen des Mittelwerts aufgezeichnet. Variiert wurde dabei, wieviele Werte bei einem Durchgang gezogen wurden (Stichprobengröße). Anhand der Abbildung lässt sich zeigen, dass bei konstanter Anzahl an Durchgängen (in diesem Fall tausend) mit steigender Stichprobengröße, also jene Anzahl an Elementen, die bei jedem Durchgang gezogen werden, die Stichprobenkennwerteverteilung des arithmetischen Mittels gegen eine Normalverteilung konvergiert.
Es lässt sich also feststellen: Bei hinreichend vielen Durchgängen und hinreichend großer Stichprobe je Durchgang nähert sich die Verteilung der Stichprobenkennwerte einer Normalverteilung an. Dies gilt unabhängig der Verteilung der zugrunde liegenden Population bzw. Grundgesamtheit. Auch in Abbildung 4.1.a ist die Population nicht normalverteilt. Dennoch ist die Stichprobenkennwerteverteilung bei einer Stichprobengröße von $n = 40$ bereits annähernd normalverteilt.
Diese Konvergenzaussage, dass bei theoretisch unendlichen vielen unabhängigen Ziehungen aus einer Grundgesamtheit die Verteilungsform bei wachsendem n in eine Normalverteilung übergeht, wird auch als zentraler Grenzwertsatz bezeichnet. Der zentrale Grenzwertsatz erklärt auch die Sonderstellung der Normalverteilung innerhalb der Statistik.

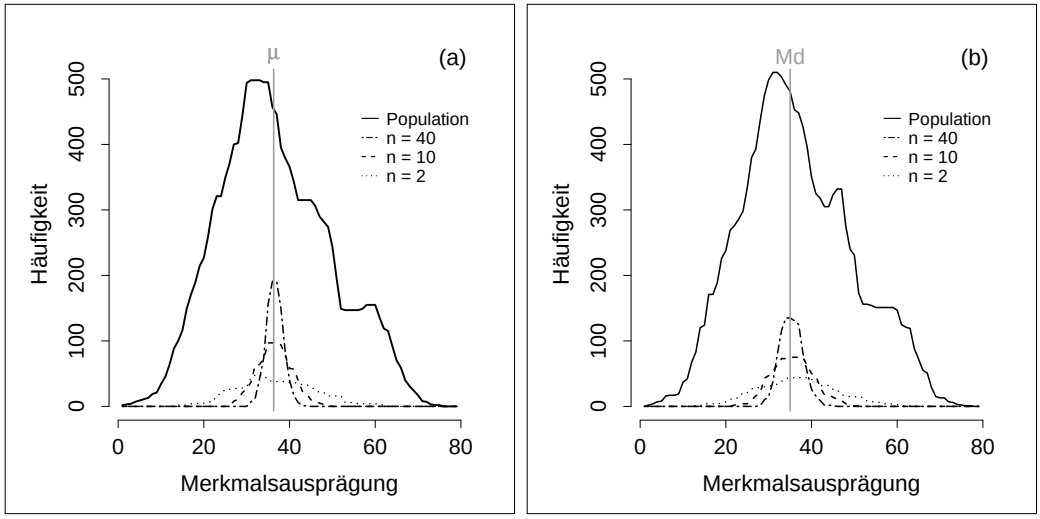


Abbildung 4.1: Stichprobenkennwerteverteilung des (a) arithmetischen Mittels und (b) Medians bei unterschiedlicher Stichprobengröße und jeweils tausend Ziehungen

Neben der Form der Verteilung ist auch deren Mittelwert für den Schluss von der Stichprobe auf die Population von Relevanz. Der Mittelwert der Stichprobenkennwerteverteilung $\bar{x}_{\bar{x}}$ entspricht bei hinreichend großer Stichprobe dem Populationsmittelwert μ (siehe Abbildung 4.1.a). Man sagt, der Mittelwert ist erwartungstreu. Weder wird der Populationsmittelwert von der Gesamtheit der Stichproben überschätzt noch unterschätzt.
Daraus leitet man ab, dass der Mittelwert einer Stichprobe $\bar{x}$ ein guter, da erwartungstreuer Schätzer des Populationsmittelwerts μ ist. Die Wahrscheinlichkeit, dass bei der Ziehung einer Stichprobe der Stichprobenmittelwert dem Populationsmittelwert sehr nahe liegt, die Differenz aus beiden also sehr gering ist, ist viel höher als die Wahrscheinlichkeit eine Stichprobe zu ziehen, bei dem die Differenz zwischen Stichprobenmittelwert und Populationsmittelwert sehr hoch ist.

Der beste Schätzer für das arithmetische Mittel der Population ist daher das arithmetische Mittel der gezogenen Stichprobe.

$$E(\bar{x}) = \mu \tag{4.1}$$
$$(\text{E} = \text{Erwartungswert})$$

Die Wahrscheinlichkeit, dass der Mittelwert der Stichprobe den Populationsmittelwert gut schätzt, wächst mit steigender Stichprobengröße. Die Stichprobenkennwerteverteilung konvergiert nicht nur gegen eine Normalverteilung bei größer werdenden Stichproben, sie wird auch steiler. Die Streubreite wird geringer (siehe Abbildung 4.1.a). Damit nimmt die Fehlerwahrscheinlichkeit bzw. die Wahrscheinlichkeit sich im großen Maße zu verschätzen ab. Bei großen Stichproben ist die Schätzgüte des Mittelwerts größer als bei kleinen Stichproben.

Auch der Median einer Stichprobe ist ein erwartungstreuer Schätzer für den Median der Population (siehe Abbildung 4.1.b). Vergleicht man jedoch die Stichprobenkennwerteverteilungen von Mittelwert und Median, so fällt auf, dass bei gleich vielen Ziehungen und gleich großen

Stichproben die Kennwerteverteilung des Medians flacher und breiter ist, während die des Mittelwerts steiler und schmaler ist. Die Streuung der Stichprobenkennwerte im Falle des Medians ist also größer als beim arithmetischen Mittel.
Daraus kann man ableiten, dass die Schätzung des Medians mit einer größeren Fehlerwahrscheinlichkeit bzw. Verschätzungswahrscheinlichkeit behaftet ist. Bei gleich großer Stichprobe ist die Wahrscheinlichkeit größer, dass die Differenz aus Stichprobenmittelwert und Populationsmittelwert geringer ist als die Differenz aus Stichprobenmedian und Populationsmedian.
Man sagt, der Median einer Stichprobe ist zwar ein erwartungstreuer Schätzer des Populationsmedians, er ist allerdings nicht so effizient bei der Schätzung. Wegen diesem Effizienzmangel spielt der Mittelwert in der Inferenzstatistik eine weitaus größere Rolle als der Median.

4.1.2 Schätzung der Populationsvarianz

Im Gegensatz zum arithmetischen Mittel und dem Median ist die Varianz nicht erwartungstreu. Die Berechnung der Varianz, wie in 2.2.3 beschrieben oder in Formel 4.2 angegeben, führt zu einer systematischen Unterschätzung der Populationsvarianz um den Faktor $n-1$.

$$s^2 = \frac{\sum_{i=1}^{n}(\bar{x} - x_i)^2}{n} \quad \text{bzw.} \quad s^2 = \frac{1}{n} \cdot \sum_{i=1}^{n}(\bar{x} - x_i)^2 \tag{4.2}$$

(Stichprobenvarianz)

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(\hat{\mu} - x_i)^2}{n-1} \quad \text{bzw.} \quad \hat{\sigma}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n}(\hat{\mu} - x_i)^2 \tag{4.3}$$

(geschätzte Populationsvarianz - wobei $\hat{\mu} = \bar{x}$)

Bei einem Simulationsexperiment mit tausend Durchgängen (wie oben erläutert) würde der Gipfel der Kennwerteverteilung der Varianz, berechnet über Formel 4.2, um $n-1$ neben der tatsächlichen Populationsvarianz liegen.
Deswegen führt man den Korrekturfaktor $n-1$ in die Formel ein und erhält eine erwartungstreue Schätzung mithilfe von Formel 4.3.

Diese Korrektur verliert bei wachsendem n an Bedeutung. Die Differenz des Varianzwertes, berechnet nach Formel 4.2 und Formel 4.3, wird immer kleiner bei größeren Stichproben. Daher sagt man, dass die Varianz (berechnet nach Formel 4.2) asymptotisch erwartungstreu ist. Bei großen Stichproben nähern sich die Werte an, sodass auch Formel 4.2 eine erwartungstreue Schätzung der Populationsvarianz abgibt.
Diese asymptotische Annäherung bei größer werdendem n ist in Abbildung 4.2 veranschaulicht.

Warum n-1? Das Konzept der Freiheitsgrade

Eine heuristische Erklärung für diese Verschätzung gibt das Konzept der *Freiheitsgrade*. Als Freiheitsgrade (engl.: degrees of freedom, oft df abgekürzt) wird die Anzahl an Elementen

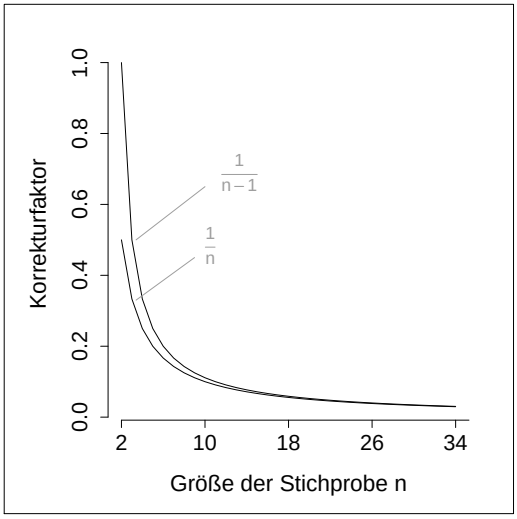


Abbildung 4.2: Darstellung der asymptotischen Annäherung des Faktors $1/n$ an den Korrekturfaktor $1/n\text{-}1$ bei ansteigender Stichprobengröße n

aus einer Menge von Elementen bezeichnet, die frei variieren können, bzw. deren Wert frei wählbar und nicht vordeterminiert ist.
Angenommen die folgenden Werte seien gegeben:

$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$	$\mathbf{x_4}$	$\mathbf{x_5}$	$\mathbf{x_6}$	
1	5	4	2	?	?	$\bar{x} = 3$

Die Werte von x_5 und x_6 sind unbekannt. Für eines der beiden Elemente der Datenreihe kann nun ein beliebiger Wert eingesetzt werden. Man hat also einen Freiheitsgrad, oder auch eine freie Wahlmöglichkeit bei der Festlegung der Werte für die zwei Unbekannten in obiger Reihe zur Verfügung.
Angenommen man entscheidet sich, für x_5 den Wert 5 einzusetzen. Durch diese Festlegung sinkt die Anzahl der Freiheitsgrade auf $df = 0$. Da der Mittelwert der Datenreihe schon bekannt ist, darf x_6 ausschließlich den Wert 1 annehmen. Nur in diesem Fall gilt für obige Datenreihe: $\bar{x} = 3$.

Bei der Schätzung der Populationsvarianz besitzt man statt n Freiheitsgraden lediglich n-1 Freiheitsgrade, also immer einen weniger als die Gesamtanzahl an Elementen. Dies liegt daran, dass in die Schätzung der Varianz bereits ein geschätzter Wert, nämlich der Populationsmittelwert $\hat{\mu}$, eingeht und die Summe aller Abweichungen der n Werte von μ immer null sein muss (siehe Formel 4.4).

$$\sum_{i=1}^{n}(x_i - \mu) = 0 \tag{4.4}$$

Daher können n-1 Werte frei variieren. Der n-te Wert ist dagegen festgelegt, sofern die Aussage der Formel 4.4 gelten soll.

4.1.3 Standardfehler des Mittelwerts

Anhand der Abbildung 4.1.a lässt sich mithilfe der Streubreite der Stichprobenkennwerteverteilungen abschätzen, wie oft man bei der Schätzung des Populationsmittelwertes μ mithilfe des Stichprobenmittelwertes $\bar{x}$ einer einzelnen Stichprobe ein falsches Ergebnis erhält.
Bei großen Stichproben (z.B. n = 40) ist die Streuung der Stichprobenkennwerte wesentlich geringer als bei kleinen Stichproben. Dementsprechend ist die Verschätzungsgefahr von μ auch geringer.

Der Standardfehler des Mittelwerts ist ein Streuungsmaß der Kennwerteverteilung des Mittelwerts. Er dient zur Einschätzung, wie gut ein einzelner Mittelwert der Stichprobe mit Größe n den Populationsparameter schätzen kann, bzw. mit welcher Wahrscheinlichkeit und in welchem Ausmaß das arithmetische Mittel $\bar{x}$ der Stichprobe den Populationsmittelwert μ über- oder unterschätzt.

Die Breite der Stichprobenkennwerteverteilung des Mittelwerts ist dabei nicht nur von der Größe der gezogenen Stichproben abhängig, sondern auch von der Streubreite der Population, gemessen mithilfe der Populationsvarianz σ^2 bzw. deren Standardabweichung σ.
Der Standardfehler des Mittelwerts lässt sich berechnen, indem man die Streuung an der Stichprobengröße relativiert:

$$\sqrt{\sigma_{\bar{x}}^2} = \sqrt{\frac{\sigma^2}{n}} \quad \text{bzw.} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{4.5}$$

Der Standardfehler des Mittelwerts besitzt, bei hinreichend großem n, die selbe Aussagekraft wie die Standardabweichung bei einer normalverteilten Variable: In dem Bereich von einer Standardfehler-Abweichung links vom Mittelwert bis einer Standardfehler-Abweichung rechts vom Mittelwert ($\mu \pm 1 \cdot \sigma_{\bar{x}}$) liegen 68,27% der Werte.
In dem Bereich von zwei Standardfehler-Abweichungen links vom Mittelwert bis zwei Standardfehler-Abweichungen rechts vom Mittelwert ($\mu \pm 2 \cdot \sigma_{\bar{x}}$) liegen 95,45% der Werte.
Im Englischen wird der Standardfehler oft mit SE (standard error) abgekürzt.

4.2 Die Normalverteilung

Da die Normalverteilung, auch Glockenverteilung oder Gauß-Verteilung genannt (nach dem Mathematiker Carl Gauß), einen wichtigen Stellenwert in der Inferenzstatistik einnimmt, soll sie an dieser Stelle detaillierter erläutert werden.
Die Normalverteilung ist symmetrisch, wobei die Ausläufer links und rechts der Verteilung sich der Abszisse (der x-Achse) asymptotisch annähern. Wie bereits angesprochen befinden sich eine Standardabweichung links und rechts des Mittelwerts 68,27% der Verteilung. Zwischen dem Abstand $\mu - 2 \cdot \sigma_{\bar{x}}$ und $\mu + 2 \cdot \sigma_{\bar{x}}$ liegen 95,45% der Verteilung. Im Bereich von drei Standardabweichungen links und rechts des Mittelwerts liegen 99,73% der Verteilung.
Eine Normalverteilung mit Angabe der Abweichungsbereiche ist in Abbildung 4.3 dargestellt.

Die Dichtefunktion der Normalverteilung ist in Formel 4.6 aufgestellt.

$$f(x) = \frac{1}{\sigma_x \cdot \sqrt{2\pi}} \cdot e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{4.6}$$

$$\text{e} = \text{eulersche Zahl } (\approx 2,718)$$

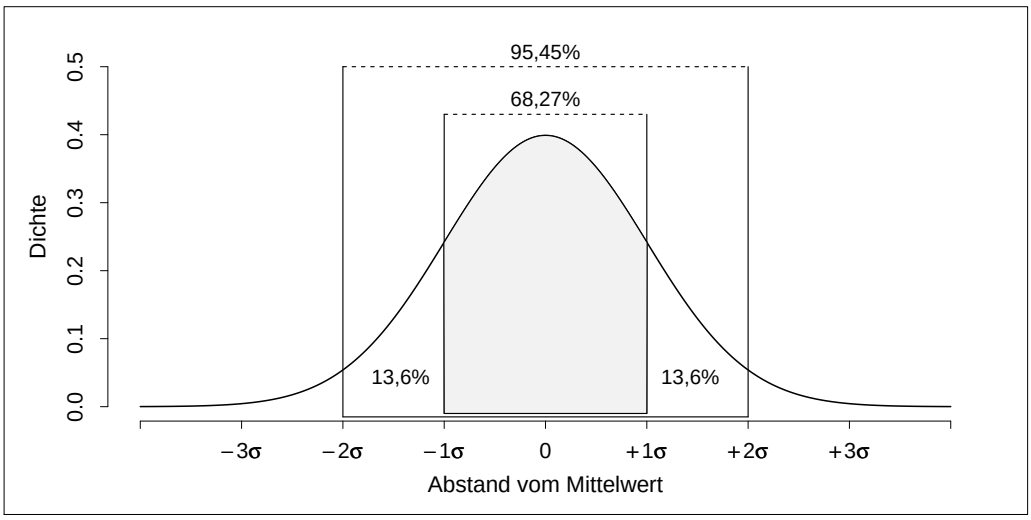


Abbildung 4.3: Die Normalverteilung mit dem Abweichungsmaß σ auf der Abszisse

4.2.1 z-Transformation

Aufgrund von Skalierungsunterschieden können Normalverteilungen im Vergleich zueinander unterschiedliche Breiten aufweisen. Bei Intelligenz geht man beispielsweise davon aus, dass diese in der Population (z.B. der deutschen Bevölkerung) normalverteilt ist. IQ-Tests, die versuchen Intelligenz messbar zu machen, benutzen teils unterschiedliche Skalierungen. Der Mittelwert von IQ-Tests liegt meist bei 100, wobei manche Tests eine Standardabweichung von 10 Intelligenzpunkten, andere von 15 festlegen.
Würde man die Verteilungen von Intelligenztests unterschiedlicher Skalierung in einer Abbildung vergleichen, wäre die Verteilung eines Intelligenztests mit großer Standardabweichung (z.B. 15 IQ-Punkte), breiter und flacher im Gegensatz zu Verteilungen von Tests mit kleiner Standardabweichung (z.B. 10 IQ-Punkte).
Um nun zwei Personen, bei denen unterschiedliche IQ-Tests mit unterschiedlicher Skalierung angewandt wurden, in ihrer Intelligenzausprägung vergleichen zu können, kann man sich eines Tricks bedienen: man standardisiert die Werte mithilfe der z-Transformation.

Führt man die z-Transformation bei allen empirischen Werten durch und folgen die empirischen Werte einer Normalverteilung, so geht diese Verteilung in eine Standardnormalverteilung über. Die Standardnormalverteilung ist definiert als eine Normalverteilung mit dem Mittelwert 0 und einer Standardabweichung von 1.

Bei der z-Transformation wird schlicht die Differenz eines gemessenen Wertes vom Mittelwert an der Standardabweichung relativiert (siehe Formel).

$$z_i = \frac{x_i - \bar{x}}{s} \quad \text{bzw.} \quad z_i = \frac{x_i - \mu}{\sigma} \tag{4.7}$$

Angenommen Person A (x_1) erreichte einen IQ von 110 bei einem IQ-Test mit einem Mittelwert von 100 und einer Standardabweichung von 10. Person B (x_2) erreichte einen IQ von 112 bei einem IQ-Test mit einem Mittelwert von 100 und einer Standardabweichung von 15. Die beiden Punktwerte sind aufgrund der unterschiedlichen Skalierung der Tests nicht direkt vergleichbar.
Berechnen wir die z-Werte beider Personen, so sehen wir, dass Person B mit einem z-Wert von $z_2 = 0,8$ eine schlechtere Leistung erbrachte als Person A mit einem z-Wert von $z_1 = 1$.
Beide liegen mit ihren Intelligenzqotienten über dem Durchschnitt. Person B erreicht jedoch einen IQ der eine Standardabweichung über dem Durchschnitt liegt, während Person A nur ein $4/5$ einer Standardabweichung über dem Mittelwert liegt.

4.2.2 Konfidenzintervalle

Anhand der z-Transformation und der Standardnormalverteilung lassen sich auch auf einfache Weise so genannte Konfidenzintervalle errechnen. Ein Konfidenzintervall, auch Vertrauensintervall genannt, ist ein Wertebereich, in dem mit einer gewissen Sicherheit ein Einzelwert liegt. Zieht man ein Stichprobe, so lassen sich mithilfe der Stichprobenkennwerte die Populationsparameter $\hat{\mu}$ und $\hat{\sigma}^2$ schätzen. Wir wissen jedoch nicht, ob überhaupt und inwieweit der Stichprobenmittelwert den Populationsmittelwert repräsentiert.
Zwar ist der Mittelwert im Durchschnitt erwartungstreu, doch setzt man einen Stichprobenmittelwert mit dem Populationsmittelwert gleich, so kann man bei dieser so genannten Punktschätzung auch weit daneben liegen.
Anhand der Schätzung des Standardfehlers der Stichprobenkennwerteverteilung lässt sich jedoch ein Vertrauensintervall berechnen, welches angibt, in welchem Bereich und mit welcher Wahrscheinlichkeit der tatsächliche Mittelwert μ liegt.

Beispiel

Angenommen das arithmetische Mittel einer hinreichen großen Stichprobe (n=30) betrage $\bar{x}$=12,5. Die Varianz der Stichprobe liegt bei s^2=4. Die hinreichend große Stichprobe gewährleistet, dass wir davon ausgehen können, dass der gezogene Mittelwert aus einer annähernd normalverteilten Stichprobenkennwerteverteilung stammt.
Wir schätzen nun den Populationsmittelwert $\hat{\mu}$, aufgrund der Erwartungstreue des Mittelwertsmaß, auf 12,5. Dies Punktschätzung ist uns allerdings zu ungenau und fehleranfällig.

Daher schätzen wir die Populationsvarianz und den Standardfehler des Mittelwerts folgendermaßen:

$$\hat{\sigma}^2 = s^2 \cdot \frac{n}{n-1} = 4 \cdot \frac{30}{29} = 4,14$$
$$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{\hat{\sigma}^2}{n}} = \sqrt{\frac{4,14}{30}} = 0,38$$

Mithilfe dieser Werte lässt sich ein Konfidenzintervall definierter Größe errechnen. Angenommen wir wollen wissen, in welchem Bereich der wahre Mittelwert μ mit einer Wahrscheinlichkeit von 95% liegt, so lassen sich Ober- und Untergrenze dieses Bereichs über folgende Formeln errechnen:

$$\text{Untere Grenze} = UG = \bar{x} - z_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\bar{x}} \tag{4.8}$$
$$\text{Obere Grenze} = OG = \bar{x} + z_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\bar{x}} \tag{4.9}$$

Der z-Wert aus obiger Formel ist jener Wert der z-Verteilung, der an den beiden Grenzen liegt, die wir für unser Intervall festgelegt haben. Der kleine griechische Buchstabe alpha symbolisiert den Anteil der z-Verteilung, der nicht in unserer Vertrauensintervall mit aufgenommen wird. In unserem Fall sind dies die äußeren 5%.
Diese 5% teilen sich auf in die äußeren 2,5% der Verteilung auf der linken Seite und die äußeren 2,5% rechts von der z-Verteilung (daher $\alpha/2$).
Schaut man in einer z-Tabelle nach, in der die Dichteanteile der Verteilung angegeben sind (z.B. im Bortz), so findet man bei dem Dichtewert 0,025 (also die Schwelle bei der sich die äußeren 2,5% der Verteilung aufkumuliert) einen z-Wert von -1,96. Aufgrund der Symmetrie der z-Verteilung liegen die extremsten 2,5% der Werte hinter dem z-Wert 1,96.
Wir errechnen nun das 1,96-fache einer Standardfehler-Abweichung und ziehen dies vom Stichprobenmittelwert ab, um die untere Grenze unseres Vertrauensintervalls zu errechnen:

$$UG = 12,5 - 1,96 \cdot 0,38 = 12,5 - 0,745 = 11,78$$

So errechnen wir auch die obere Grenze des Intervalls:

$$OG = 12,5 + 1,96 \cdot 0,38 = 12,5 + 0,745 = 13,25$$

Unser Konfidenzintervall liegt also zwischen 11,78 und 13,25. Wir können nun von der Punktschätzung zu einer Intervallschätzung übergehen und mit einer Wahrscheinlichkeit von 95% annehmen, dass der wahre Populationsmittelwert μ irgendwo im Bereich von 11,78 und 13,25 liegt.

Die Bedeutung von Konfidenzintervallen lässt sich anhand von Abbildung 4.4 verdeutlichen. Angenommen in Abbildung 4.4 ist die Kennwerteverteilung für einen Mittelwert μ dargestellt. Man zieht aus dieser Population nun zufällig eine Stichprobe und erhält den Mittelwert $\bar{x}$.
Würde man sich auf eine Punktschätzung verlassen, so läge man in diesem Fall recht weit da-

neben. Der Stichprobenmittelwert liegt über eine Standardfehler-Abweichung über dem tatsächlichen Mittelwert.
Berechnet man jedoch ein Vertrauensintervall der Größe $z \pm 1,96$, so sieht man, dass der wahre Mittelwert tatsächlich in diesem Bereich liegt.

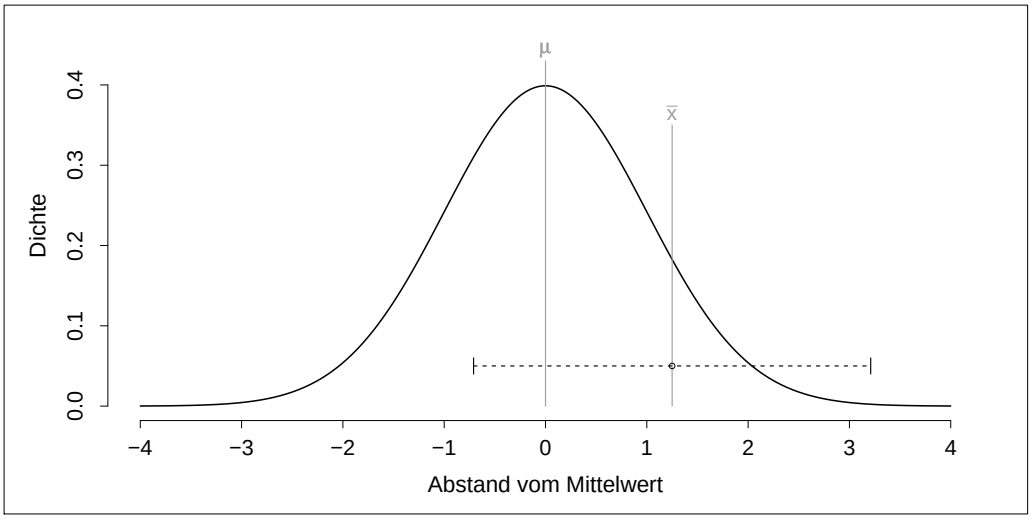


Abbildung 4.4: Darstellung eines 95%-Konfidenzintervalls um den Mittelwert $\bar{x}$

4.A Aufgaben

inhaltliche Aufgaben

1. Was bedeutet der Ausdruck „Inferenzstatistik“ und welche Fragen lassen sich mithilfe der Inferenzstatistik beantworten?
2. Was ist der Unterschied zwischen der Inferenz- und Deskriptivstatistik?
3. Was ist eine Stichprobe? Was eine Population?
4. Welche Faktoren beeinflussen, wie gut eine Stichproben die Population repräsentiert?
5. Was versteht man unter dem Begriff „Stichprobenkennwerteverteilung“?
6. Was besagt das zentrale Grenzwerttheorem?
7. Wie sind Mittelwerte um μ verteilt?
8. Was versteht man unter einer Normalverteilung?
9. Was ist die Standardnormalverteilung? Welche Eigenschaften weist die Standardnormalverteilung auf?
10. Was ist die z-Verteilung?
11. Wie groß ist die Fläche der z-Verteilung?
12. Welcher z-Wert teilt die Fläche unter der Verteilung in zwei gleich große Hälften?
13. Welche Wahrscheinlichkeitsaussage lässt sich mit diesem Wert verbinden?
14. Wie groß ist die Wahrscheinlichkeit, einen z-Wert zwischen -1 und +1 zu erhalten?
15. Wie hängen μ und $\bar{x}$ zusammen?
16. Was versteht man unter Erwartungstreue eines Kennwertes, was unter Effizenz?
17. Was ist der Standardfehler des Mittelwerts? Welche Aussagen lassen sich mithilfe dieses Kennwertes treffen?
18. Wie lässt sich eine Normalverteilung in eine Standardnormalverteilung umwandeln?
19. In einer Klinik findet ein Forschungsprojekt statt, bei dem man herausfinden will, ob ein neues Antidepressivum wirkt. Die Wirksamkeit wird an der Dauer in Tagen bis zum Erreichen eines bestimmten Wertes im BDI (Becks Depressions Inventar) gemessen. Verglichen werden die Ergebnisse der mit dem Antidepressivum behandelten Patienten mit der durchschnittlichen Dauer bis zum Erreichen des festgelegten Wertes bei Depression ohne Behandlung. Definiere die Stichprobe und die Population. Was entspricht μ und $\bar{x}$?
20. Was ist ein Konfidenzintervall?
21. Erstelle eine schrittweise Anleitung zur Berechnung eines Konfidenzintervalls.
22. Wie errechnen sich die Grenzen eines 50% Konfidenzintervall für einen Stichprobenmittelwert?

praktische Aufgaben

1. Wie groß ist $\hat{\mu}$ bei einem Stichprobenmittelwert von $\bar{x} = 12$ und einer Stichprobengröße von n = 100?

2. Wie groß ist $\hat{\sigma}$ bei einer Stichprobengröße n = 10 und einer Stichprobenvarianz von $s^2 = 9,6$?

3. In einer Zufallsstichprobe mit n = 20 Messwerten hat man folgende Daten erhoben:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
100	150	130	130	78	90	99	145	123	119

x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
83	134	97	70	171	89	114	104	94	81

   Wie groß ist die Stichprobenvarianz, das arithmetische Mittel, die geschätzte Populationsvarianz und der Standardfehler des Mittelwertes?

4. In einer Stichprobe (n = 100) wurde eine Varianz von s^2=368,37 ermittelt. Wie groß ist der Standardfehler des Mittelwerts?

5. Errechne bitte das 95%ige Konfidenzintervall für folgende Mittelwerte:
   a) $\bar{x} = 85$, $s^2 = 11$, n = 50
   b) $\bar{x} = 60$, $s^2 = 6$, n = 50

   Könnten die beiden Stichproben aus der selben Population gezogen worden sein?

6. Bei einem Fragebogen zur Einstellung gegenüber Hundehaltung auf Bauernhöfen mit einer Skala von 1 = „Absolut dagegen“ bis 20 = „Absolut dafür“ hat man 30 Personen befragt und folgende Werte erhalten:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
14	13	16	19	17	17	8	4	16	18

x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
19	15	15	13	17	19	5	7	5	16

x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	x_{30}
7	3	19	5	10	11	10	13	8	3

   In welchem Bereich liegt mit 99%iger Wahrscheinlichkeit der Populationsmittelwert?

4.3 inferenzstatistische Tests

4.3.1 Der z-Test

Der z-Test, so wie er hier besprochen wird, setzt voraus, dass die Population sowie deren Kennwerte bekannt sind. Da dies selten der Fall ist, ist der z-Test vorrangig von theoretischem Interesse, da sich an ihm die Logik inferenzstatistischer Tests gut veranschaulichen lässt.

Mithilfe der z-Transformation lässt sich einschätzen, wie weit einzelne Messwerte vom Mittelwert entfernt sind, indem die Skalierung auf den Mittelwert 0 bei einer Standardabweichung von 1 normiert wird.
Beim z-Test wird im Prinzip auch eine z-Transformation durchgeführt. Doch statt einen Messwert an den Verteilungsparametern zu relativieren, wird ein gezogener Stichprobenmittelwert ins Verhältnis zu dem tatsächlichen Populationsmittelwert und derem Standardfehler gesetzt. Der daraus resultierende z-Wert gibt an, wie weit ein gezogener Stichprobenmittelwert vom Populationsmittelwert entfernt ist, gemessen in Standardfehler-Abweichungen.
Die Formel für den z-Wert ähnelt daher sehr der obigen Formel der z-Transformation:

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \tag{4.10}$$

Da wir wissen, dass bei hinreichen großem n die Stichprobenkennwerteverteilung von gezogenen Mittelwerten normalverteilt ist, können wir einschätzen wie wahrscheinlich es ist, dass der Mittelwert der Population entstammt.
Liegt ein gezogener Mittelwert beispielsweise drei Standardfehler-Abweichungen vom tatsächlichen Populationsmittelwert entfernt, so können wir mit relativer Sicherheit sagen, dass es ziemlich unwahrscheinlich ist, dass die Stichprobenwerte aus der uns bekannten Population gezogen wurden.

statistische Signifikanz

Um zu entscheiden, ab wann man sich dafür entscheidet, die Hypothese abzulehnen, dass der Stichprobenmittelwert der Population entstammt, haben sich wissenschaftliche Konventionen entwickelt. Ein Mittelwert gilt dann als bedeutsam verschieden vom Populationsmittelwert, wenn die Wahrscheinlichkeit, dass er aus der Population stammt, kleiner als 5% ist. Die Wahrscheinlichkeit, dass der Mittelwert aus einer anderen Population stammt, die sich in ihrer Verteilung von der uns bekannten wesentlich unterscheidet, ist in diesem Fall viel größer.
Man sagt dann, der Mittelwert unterscheidet sich signifikant vom Populationsmittelwert (signifikant von lat.: bedeutsam oder bezeichnend). Ist die Wahrscheinlichkeit, dass der Mittelwert einer bekannten Population entspringt, kleiner als 1%, so spricht man von einem hoch signifikanten Ergebnis.

Die Grenze, ab der man davon ausgeht, dass der Mittelwert nicht aus der bekannten Population stammt, wird auch Signifikanzschranke oder akzeptierte Irrtumswahrscheinlichkeit bezeichnet. Wenn man hundert Stichproben ziehen würde, und alle Mittelwerte dieser Stichproben würden in dem äußeren 5% Bereich liegen, so würde man sich wahrscheinlichkeitstheoretisch

in ungefähr fünf Fällen irren; Bei fünf Stichproben würde man irrtümlicherweise davon ausgehen, dass die Stichprobenwerte nicht aus der Population gezogen wurden, was jedoch eine Fehlentscheidung war. Dagegen wäre in 95 Fällen die Entscheidung richtig.

einseitige versus zweiseitige Testung

Bei den meisten inferenzstatistischen Tests legt man vorher fest, ob man einseitig oder zweiseitig testen will. Dies hängt damit zusammen, ob man eine gerichtete oder eine ungerichtete Hypothese formuliert. Bei einer ungerichteten Hypothese geht man davon aus, dass der Stichprobenmittelwert sich vom Populationsmittelwert unterscheidet. Man weiß jedoch nicht, ob dieser signifikant höher oder niedriger als der Populationsmittelwert ist.
Bei gerichteter Hypothese hat man aufgrund von Vorannahmen oder bereits vorliegenden Forschungsergebnissen die Vermutung, dass der Stichprobenmittelwert sich nicht nur bedeutsam vom Populationsmittelwert unterscheidet, sondern auch in welche Richtung er sich unterscheidet, also entweder signifikant größer oder kleiner ist.

Da die Irrtumswahrscheinlichkeit in beiden Fällen nur 5% betragen soll, wird bei der einseitigen Testung der äußerst rechte oder linke Bereich, der 5% der Verteilung umfasst, festgelegt und der z-Wert an dieser Stelle als Schwellenwert oder kritischer z-Wert definiert, der von dem durch Formel 4.10 errechneten, so genanntem empirischen z-Wert überschritten werden muss, um von einem signifikanten Ergebnis sprechen zu können.
In Abbildung 4.5.a ist der Bereich dargestellt, in dem der empirische z-Wert eines Stichprobenmittelwertes liegen muss, unter Annahme der Hypothese, dass dieser signifikant größer ist als der Populationsmittelwert.
In Abbildung 4.5.b sind links und rechts von der Verteilung die jeweils äußeren 2,5% dargestellt. Damit wird auch bei zweiseitiger Testung eine Fehlerwahrscheinlichkeit von insgesamt 5% akzeptiert. Der empirische z-Wert eines Stichprobenmittelwerts muss demnach entweder über 1,96 oder unter -1,96 liegen. Diese beiden Werte sind demnach die kritischen z-Werte bei zweiseitiger Testung und einer akzeptierten Irrtumswahrscheinlichkeit von 5%.

Hypothesen

Üblicherweise wird vor der Durchführung eines inferenzstatistischen Tests das Signifikanzniveau festgelgt, symbolisiert durch den kleinen griechischen Buchstaben alpha (α). Des Weiteren werden die formalen Hypothesen aufgestellt. Das Signifikanzniveau liegt, wie bereits angesprochen, üblicherweise bei α=5% oder α=1%.

Hypothesen werde entweder gerichtet oder ungerichtet aufgestellt. Gerichtete Hypothesen werden dann formuliert, wenn man nicht nur einen Mittelwertsunterschied, sondern auch dessen Richtung postuliert. Ungerichtete Hypothesen werden dann aufgestellt, wenn man lediglich einen Mittelwertsunterschied vermutet, ohne zu wissen in welche Richtung.
Bei gerichteter Hypothese wird einseitig, bei ungerichteter Hypothese dagegen zweiseitig getestet.

Bei der formalen Aufstellung von Hypothesen unterscheidet man meist zwei Einzelhypothesen: Die Nullhypothese (H_0) und die Alternativhypothese (H_1).
Die Nullhypothese trifft bei ungerichteten Hypothesen dann zu, wenn der Stichprobenmittelwert sich, im Fall des z-Tests, nicht bedeutsam vom Populationsmittelwert unterscheidet. Die

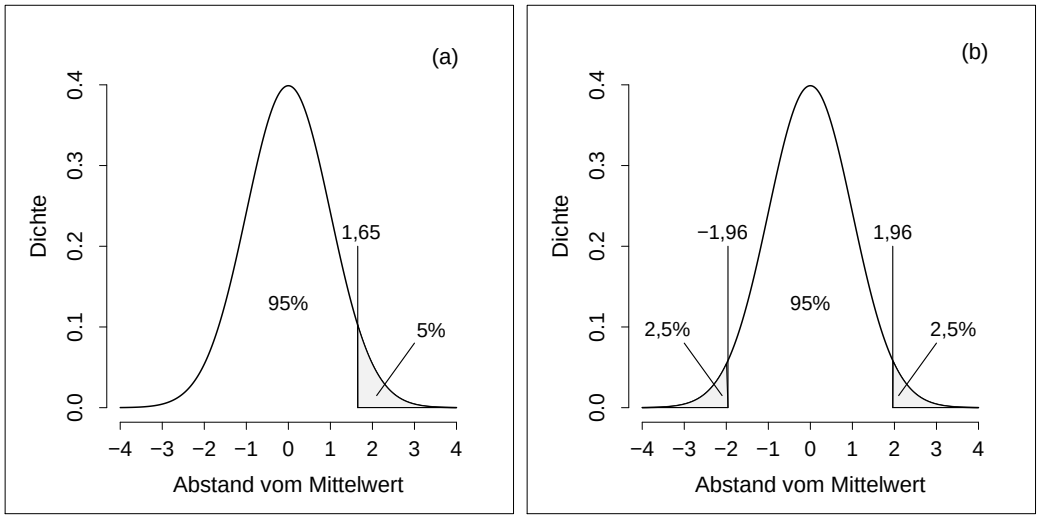


Abbildung 4.5: z-Verteilung und Bereich der äußeren 5% der Verteilung bei (a) einseitiger Testung und (b) zweiseitiger Testung

Alternativhypothese gilt dann, wenn der z-Wert des Stichprobenmittelwerts sich in einem der äußeren Bereiche befinden, die jeweils die äußeren 2,5% der Verteilung ausmachen (siehe 4.5.b). Es gilt dann: $\mu_0 \neq \mu_1$; wobei μ_0 der Mittelwert der Population und μ_1 ein unbekannter Parameter einer unbekannten Population ist, dessen Verteilung sich von der bekannten Population unterscheidet.
Die formale Schreibweise der Hypothesen bei einem z-Test sieht folgendermaßen aus:

$$H_0 : \mu_0 = \mu_1$$
$$H_1 : \mu_0 \neq \mu_1$$
(ungerichtete Hypothesen)

Bei gerichteten Hypothesen wird dagegen die Richtung des Unterschieds angegeben:

$$H_0 : \mu_0 \geq \mu_1$$
$$H_1 : \mu_0 < \mu_1$$
(gerichtete Hypothesen)

In diesem Fall wird definiert: Der Mittelwert der gezogenen Stichprobe stammt aus einer Population dessen Mittelwert μ_1 signifikant größer ist als der Mittelwert der Ausgangspopulation μ_0 (Alternativhypothese).
Die Nullhypothese dagegen sagt aus, dass der gezogene Mittelwert sich nicht von dem Populationsmittelwert unterscheidet oder signifikant kleiner ist.

Alpha- und Beta-Fehler

Angenommen die rechte Verteilung in Abbildung 4.6 (durchgezogene Linie) ist die z-transformierte Stichprobenkennwerteverteilung des Mittelwerts einer uns bekannten Population. Die linke Verteilung (gestrichelte Linie) ist eine Alternativverteilung (die unter realen Bedingung unbekannt ist).

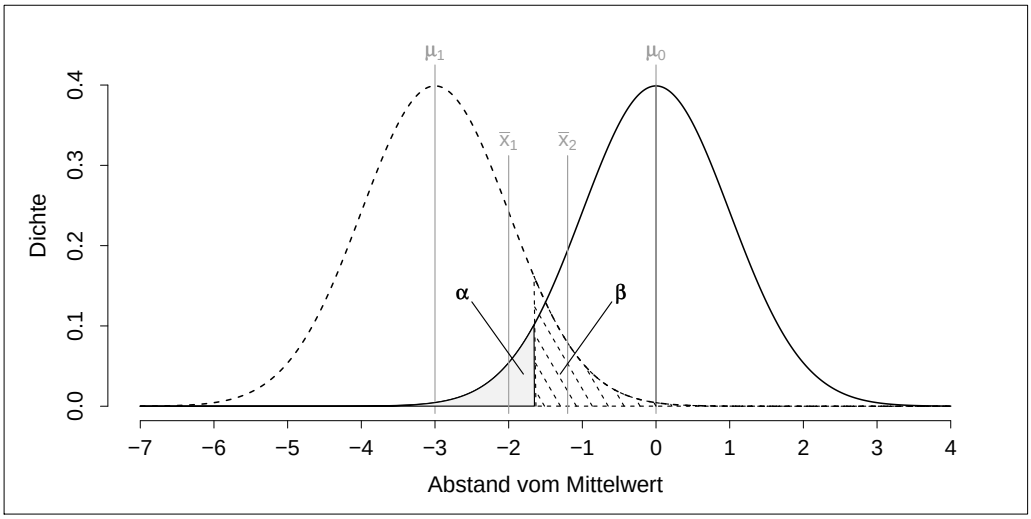


Abbildung 4.6: Darstellung der Alpha- und Beta-Fehlerproblematik

Wir ziehen nun eine Stichprobe und errechnen den Mittelwert $\bar{x}_1$. Wir wissen nicht, ob diese Werte der Stichprobe aus der Verteilung der H0-Population oder der H1-Population stammt.
Da der z-Wert dieses Mittelwerts allerdings weit von dem Populationsmittelwert μ_0 entfernt liegt und wir eine α-Fehlerwahrscheinlichkeit von 5% akzeptieren (grauer Bereich), nehmen wir an, die Werte dieser Stichprobe stammen aus der Alternativverteilung. Die Wahrscheinlichkeit, dass die Stichprobenwerte aus der H1-Verteilung stammen ist weitaus größer als die Wahrscheinlichkeit, dass sie aus der H0-Verteilung stammen. Dennoch können wir uns bei dieser Entscheidung irren. Dies ist das so genannte Alpha-Fehlerrisiko; also jenes Risiko, dass wir uns für die H1-Hypothese entscheiden (die Werte stammen aus einer anderen als der H0-Population), obwohl die Werte tatsächlich aus der H0-Population gezogen wurden und die weite Entfernung des Stichprobenmittelwerts von μ_0 reiner Zufall war.

Angenommen wir ziehen eine Stichprobe mit dem Mittelwert $\bar{x}_2$. Aufgrund unseres α-Fehlerniveaus würden wir uns dafür entscheiden, dass dieser Wert nicht aus einer Alternativverteilung stammt. Die Wahrscheinlichkeit dafür ist geringer als die Wahrscheinlichkeit, dass die Werte aus der H0-Verteilungen gezogen wurden.
Dennoch liegt der z-Wert von $\bar{x}_2$ noch innerhalb des schraffierten Bereichs. Es ist durchaus möglich, dass die Stichprobenwerte der H1-Verteilung (gestrichelte Linie) angehören.
Dies ist das so genannte Beta-Fehlerrisiko; also jene Gefahr, sich für die H0-Hypothese zu entscheiden, obwohl die Werte tatsächlich aus einer alternativen H1-Verteilung gezogen wurden.

Im Falle von $\bar{x}_2$ könnte man das Beta-Fehlerrisiko verringern, indem man die Alpha-Schranke oder die akzeptierte Alpha-Fehlerwahrscheinlichkeit erhöht, also den grauen Bereich in Abbildung 4.6 erweitert.

Die Wahrscheinlichkeit, dass man $\bar{x}_2$ irrtümlicherweise der H0-Population zuordnet, wird damit also geringer. Im Gegenzug steigt jedoch die Wahrscheinlichkeit, dass man die Werte von $\bar{x}_2$ fälschlicherweise der H1-Population zurechnet, obwohl sie eigentlich der H0-Population angehören.

Diese Dilemma ist als Alpha- und Beta-Fehlerproblematik bekannt. Die beiden Fehlerarten bedingen sich gegenseitig; erhöht man das akzeptierte Alpha-Risiko von beispielsweise 5% auf 10%, so steigt das Risiko einen so genannten Fehler erster Art zu begehen (Alpha-Fehler). Man entscheidet sich für die H1, obwohl die H0 gilt.
Umgekehrt steigt die Wahrscheinlichkeit zweiter Art (Beta-Fehler), verringert man die akzeptierte Irrtumswahrscheinlichkeit von beispielsweise 5% auf 1%. Die Gefahr hierbei ist, dass man sich für die H0 entscheidet, obwohl die Stichprobenwerte aus einer anderen Verteilung, als der H0-Verteilungen entstammen.

Aufgrund des Beta-Fehlerrisikos ist es auch nicht möglich aufgrund eines insignifikanten Ergebnisses auf dem 5%-Niveau darauf zu schließen, dass die Stichprobenwerte der H0-Population angehören. Um zu testen, ob ein Mittelwert aus einer bekannten Population stammt, muss das Alpha-Risiko erhöht werden, um das Beta-Fehlerrisiko möglichst zu minimieren. Es gibt allerdings für diesen Fall keine strengen verbindlichen Konventionen, welches Alpha-Niveau man für diesen Fall festlegen sollte, da die Alternativverteilung meist unbekannt ist und daher auch die Größe des Überschneidungsbereichs der Verteilungen nicht bekannt ist.

4.B Aufgaben

inhaltliche Aufgaben

1. Wozu wird die Verteilungsfunktion der Normalverteilung herangezogen?
2. Bestimmte mithilfe einer z-Werte Tabelle ...
   a) wie groß die Fläche ist, die bei der Standardnormalverteilung zwischen z $= +1$ und z = -1 liegt.
   b) wie groß die Fläche ist, die bei der Standardnormalverteilung zwischen z $= +3$ und z $= \infty$ liegt.
   c) wie groß die Fläche ist, die bei einer Standardnormalverteilung zwischen z = -1,5 und z = 2 liegt.
3. Welche z-Werte trennen ...
   a) jeweils die äußersten 5% rechts und die äußersten 5% links ab?
   b) die 5% extremsten Werte ab? (5% auf beide Seiten verteilt)
   c) jeweils die äußersten 1% rechts und die äußersten 1% links ab?
   d) die 1% extremsten Werte ab?
4. Wann wird ein z-Test gerechnet?
5. Welche Schritte werden bei der Berechnung eines z-Tests durchgeführt?
6. Was ist eine Null- und was eine Alternativhypothese?
7. Ein Krebsforscher arbeitet an einem neuen Medikament und vergleicht es mit dem im Moment marktführenden Medikament. Wie sehen wahrscheinlich seine Hypothesen aus? Formuliert diese bitte in je einem Satz und statistisch (besser = höhere Werte).
8. Was ist der α- und der β-Fehler? Wie hängen sie zusammen?
9. Wieso ist die β-Fehlerwahrscheinlichkeit meistens unbekannt?
10. Wann stellt man gerichtete, wann ungerichtete Hypothesen auf?
11. Was ist eine einseitige, was eine zweiseitige Testung? Was passiert dabei mit der α-Fehlerwahrscheinlichkeit?

praktische Aufgaben

1. Ein musikpädagogischer Forscher möchte die Fragestellung untersuchen, ob Kinder, die ein Musikinstrument beherrschen, emotional weniger labil sind als der Durchschnitt. In der Population von Kindern sei ein durchschnittlicher Neurotizismuswert von $\mu = 50$ bei einer Populationsstreuung von $\sigma = 20$ bekannt. Der Musikpädagoge findet in einer Stichprobe von 40 klavier- und flötespielenden Kindern einen durchschnittlichen Neurotizismuswert von $\bar{x} = 40$ ($s = 15$). Leider weiß er nicht, ob man aufgrund dieser Stichprobendaten tatsächlich sagen kann, dass musizierende Kinder generell emotional stabiler sind.

Führe einen z-Test durch, um diese Frage inferenzstatistisch zu überprüfen, sowohl bei einem Fehlerniveau von $\alpha = 1\%$ als auch einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$. Formuliere dabei die Hypothesen sowohl inhaltlich als auch statistisch.

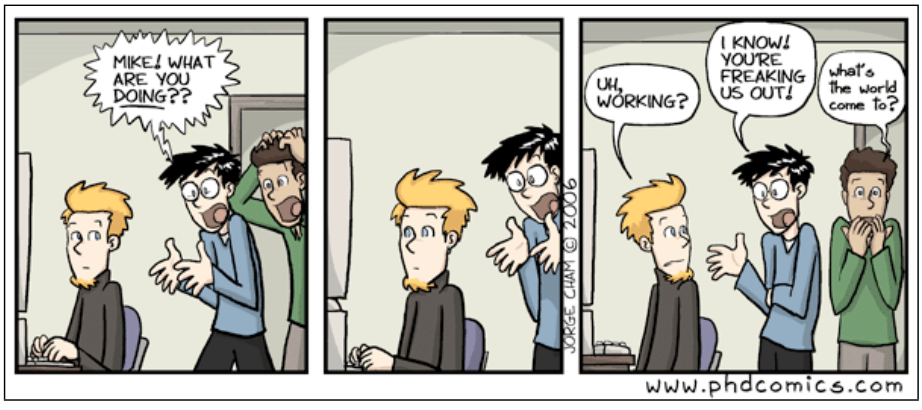
2. Es ist bekannt, dass die durchschnittliche Körpergröße der Deutschen 1,72m ist, die Standardabweichung beträgt 10cm. Es soll nun geklärt werden, ob die Menschen, die in der Nähe einer Chemiefabrik leben, hochsignifikant von der Durchschnittsgröße abweichen, bei einer Irrtumswahrscheinlichkeit von $\alpha = 1\%$.

   Hier die Daten der Versuchspersonen:

Proband	1	2	3	4	5	6	7	8	9	10
Größe	1,75	1,63	1,85	1,52	1,71	1,76	1,87	1,68	1,68	1,84

Proband	11	12	13	14	15	16	17	18	19	20
Größe	1,67	1,65	1,74	1,60	1,72	1,72	1,77	1,57	1,60	1,80

Proband	21	22	23	24	25	26	27	28	29	30
Größe	1,68	1,65	1,78	1,89	1,48	1,73	1,69	1,57	1,67	1,80



4.3.2 Der t-Test für unabhängige Stichproben

Bei den meisten empirischen Untersuchungen ist weder die H0-Population noch eine Alternativpopulation bekannt. Der z-Test ist daher oft nicht anwendbar. Stattdessen liegen in der experimentellen Psychologie typischerweise Daten von zwei Probandengruppen vor. Die Gruppen unterscheiden sich beispielsweise darin, unter welcher experimentellen Bedingung sie während eines psychologischen Experiments beobachet oder getestet wurden.
Nun gilt es herauszufinden, ob die Gruppen sich signifikant, also statistisch bedeutsam, in der erhobenen Merkmalsausprägung unterscheiden. Ist ein signifikanter Mittelwertsunterschied vorhanden schließt man daraus, dass die Probanden der einen Gruppe einer Merkmalsverteilung angehören, die sich von der Verteilung des Merkmals der zweiten Gruppe bzw. deren Population unterscheidet.
Da keine Populationsparameter bekannt sind, wird zunächst die Hypothese aufgestellt, dass die beiden Stichproben aus derselben, unbekannten Population stammen.
Unter dieser Annahme gilt:

$$H_0 : \mu_1 = \mu_2$$
(Nullhypothesen)

Eine Alternative zu der Nullhypothese wäre, dass die Probanden zwischen den Gruppen sich in ihrer Merkmalsausprägung bedeutsam unterscheiden, also verschiedenen Populationen angehören, deren Verteilungsparameter μ sich unterscheidet:

$$H_1 : \mu_1 \neq \mu_2$$
(ungerichtete Alternativhypothese)

Gilt die H0 Hypothese, so sollte die Differenz von $\bar{x}_1$ und $\bar{x}_2$ relativ gering sein, da ja gilt: $\mu_1 = \mu_2$. Eine große Differenz der beiden Parameter (relativiert an der Streuung) wäre dagegen höchst unwahrscheinlich. In diesem Fall wäre die Alternativhypothese plausibler.

Nach dieser Logik geht der t-Test vor. In Abschnitt 4.1.1 wurde die Entstehung der Stichprobenkennwerteverteilung für den Mittelwert beschrieben. Eine ähnliches Simulationsexperiment ließe sich mit den Differenzwerten von Stichproben errechnen; dabei zieht man jeweils zwei Stichproben der Größe n aus einer Population und errechnet die Differenz der Mittelwerte dieser beiden Stichproben. Diesen Vorgang wiederholt man beispielsweise tausend mal und erhält eine Verteilung, die als t-Verteilung bekannt ist.
Allerdings gibt es „die“ t-Verteilung streng genommen nicht. Wie man in Abbildung 4.1 sehen kann, unterscheiden sich Kennwerteverteilungen je nach Stichprobengröße in ihrer Streubreite.
Bei der t-Verteilung wird dem Rechnung getragen. Je nach Größe der Stichproben bzw. der zur Verfügung stehenden Freiheitsgrade ist die t-Verteilung unterschiedlich breit (siehe Abbildung 4.7). Wie man anhand von Abbildung 4.7 erkennen kann, nähert sich die t-Verteilung bei größer werdendem n (bzw. df) der z-Verteilung an.
Dies hat den Vorteil, dass man bei großen Stichproben (ungefähr $n_1+n_2 \geq 30$) für den kritischen Grenzwert statt einem t-Wert sich einem z-Wert (wie in 4.3.1 besprochen) bedienen kann.

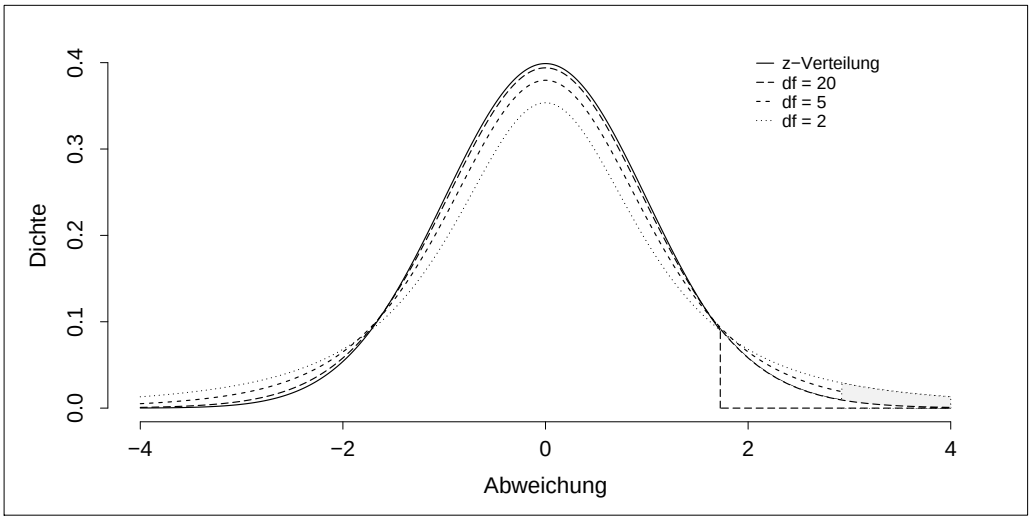


Abbildung 4.7: Darstellung einer z-Verteilung im Vergleich zu verschiedenen t-Verteilungen unterschiedlicher Freiheitsgrade

Mithilfe der t-Formel lässt sich nun errechnen, an welcher Stelle der t-Verteilung jener empirische t-Wert liegt der sich bei der Differenz der Stichprobenmittelwerte relativiert an der Standardfehler-Abweichung der Differenzwerte-Verteilung ergibt:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}} \tag{4.11}$$

Die Wahrscheinlichkeit, mit der ein solcher t-Wert bei geltender H0-Hypothese vorkommt, lässt sich nun an der t-Verteilung ablesen. Legt man eine Irrtumswahrscheinlichkeit von α=5% fest und möchte die H0-Hypothese wiederlegen, so muss der t-Wert, der sich anhand der empirischen Daten errechnete, über dem kritischen t-Wert liegen, der sich allerdings je nach Stichprobengröße bzw. Freiheitsgraden verändert, aufgrund der sich ändernden Breite der t-Verteilung.

In Abbildung 4.7 ist dies dargestellt. Bei df=2 muss ein kritischer t-Wert von 2,92 überschritten werden (siehe grauen Bereich), um mit einer Irrtumswahrscheinlichkeit von 5% (einseitig) behaupten zu können, die Mittelwerte der beiden Stichproben unterscheiden sich statistisch bedeutsam. Dagegen reicht bei df=20 ein Wert größer als t=1,725 bereits aus, um von einem signifikanten Ergebnis zu sprechen.
Das bedeutet, dass bei kleinen Stichproben größere Mittelwertsunterschiede (relativ zur Streuung) gefunden werden müssen, um einen signifikanten Unterscheid festzumachen, da bei kleinen Stichproben ein großer Mittelwertsunterschied mit höherer Wahrscheinlichkeit durch reinen Zufall entstehen kann, während bei großen Stichproben ein großer Mittelwertsunterschied aufgrund von Zufall weniger wahrscheinlich ist.

Die für einen t-Test zur Verfügung stehenden Freiheitsgrade errechnen sich über:

$$\mathrm{df} = (n_1 + n_2) - 2 \tag{4.12}$$

Es gehen zwei Freiheitsgrade verloren, da in die Schätzung von $\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)}$ zwei geschätzte Werte eingehen; die jeweils geschätzten Populationsvarianzen σ_1^2 und σ_2^2.

Der Standardfehler der Differenzverteilung von $\bar{x}_1$ und $\bar{x}_2$ lässt sich bei gleich großen Stichproben (n_1=n_2) folgendermaßen schätzen:

$$\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{\hat{\sigma}_1^2+\hat{\sigma}_2^2}{2}} \cdot \sqrt{\frac{2}{n}} \tag{4.13}$$

Bei ungleich großen Stichproben ($n_1 \neq n_2$) werden die geschätzten Populationsvarianzen durch die jeweils zur Verfügung stehenden Freiheitsgrade gewichtet:

$$\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{(n_1-1)\cdot\hat{\sigma}_1^2+(n_2-1)\cdot\hat{\sigma}_2^2}{(n_1-1)+(n_2-1)}} \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \tag{4.14}$$

Vorraussetzungen des t-Tests

Die t-Verteilung ist auch bei wenigen Freiheitsgraden normalverteilt (siehe Abbildung 4.7). Wie man aber in Abbildung 4.1 sehen kann, sind Kennwerteverteilungen bei kleinem n keineswegs normalverteilt. Sie folgen stattdessen ungefähr der Verteilung der Population.

Dass t-Verteilungen bei kleinem n dennoch normalverteilt sind, hängt mit einer Annahme zusammen, die mit dem t-Test verbunden ist. Der t-Test geht davon aus, dass die Stichproben aus normalverteilten Grundgesamtheiten stammen.

Diese Vorraussetzung kann oft nicht überprüft werden, da die Grundgesamtheit bzw. Grundgesamheiten nicht bekannt sind. In der Praxis wird diese Bedingung daher oft vernachlässigt oder schlicht ignoriert, was bei großen Stichproben auch weniger problematisch ist, da hier die Kennwerteverteilung unabhängig der Populationsverteilung sich annähernd normalverteilt.
Bei kleinen Stichproben sollte dagegen lieber auf ein so genanntes nonparametrischens Verfahren, wie den Mann-Whitney U-Test, ausgewichen werden. Nonparametrisch bedeutet verteilungsfrei. Der U-Test stellt keine Bedingung an die Verteilung der Daten bzw. der Population. Der t-Test ist dagegen ein parametrisches Verfahren.

Des Weiteren müssen die Werte bzw. die dahinter liegenden gemessenen Merkmalsausprägungen, wie auch im z-Test, äquidistant zueinander sein. Dies ist nur gewährleistet auf Intervall- oder Verhältnisskalenniveau.

Neben der Verteilungsannahme und dem Skalenniveau setzt der t-Test voraus, dass, unterscheiden sich die zugrunde liegenden Grundgesamtheiten in ihrem Maß der zentralen Tendenz signifikant, diese sich in ihrer Varianz zumindest annähernd gleichen.
Ist die Streuung der Werte der einen Population weitaus größer oder kleiner als die Streuung der zweiten Population, ist diese Bedingung verletzt und das Ergebnis des t-Tests verliert an statistischer Validität (Gültigkeit), was zu Fehlurteilen führen kann.

Die Varianzhomogenitätsbedingung kann mit einem weiteren statistischem Testverfahren überprüft werden: dem F-Test. Beim F-Test, konzipiert von dem Statistiker Ronald A. Fisher, wird

die geschätzte Populationsvarianz der einen Stichprobe an der geschätzten Populationsvarianz der zweiten Stichprobe relativiert:

$$F_{emp} = \frac{\sigma_1^2}{\sigma_2^2} \tag{4.15}$$

Hierbei wird im Zähler die größere, im Nenner stets die niedrigere Populationsvarianz eingesetzt.

Der F-Test besitzt jeweils $df_1 = n_1$-1 und $df_2 = n_2$-1 Freiheitsgrade, da pro Stichprobe jeweils die Schätzung der Populationsvarianz in den F-Bruch eingeht.
Sind die geschätzten Populationsvarianzen der beiden Stichproben gleich, so ergibt sich ein F-Wert von 1. Je weiter der F-Wert von 1 entfernt ist, desto unwahrscheinlicher ist es, dass die tatsächlichen Populationsvarianzen gleich sind.
Wir stellen also die Hypothesen folgendermaßen auf:

$$H_0 : \sigma_1^2 \leq \sigma_2^2$$
$$H_1 : \sigma_1^2 > \sigma_2^2$$

Diese sind, im Falle der Überprüfung einer Voraussetzungsverletzung des t-Tests, stets gerichtet, da beim F-Bruch in obiger Formel nur Werte größer 1 errechnet werden können, setzt man im Nenner die größere geschätzte Populationsvarianz ein.

Auch die F-Verteilung trägt der wahrscheinlichkeitstheoretischen Überlegung Rechnung, dass große Varianzunterschiede bei kleinen Stichproben öfter aufgrund des Zufalls auftreten, während dies bei großen Stichproben unwahrscheinlicher ist. Daher ändert sich auch die F-Verteilung je nach Anzahl der Freiheitsgrade. In Abbildung 4.8 ist dies verdeutlicht.

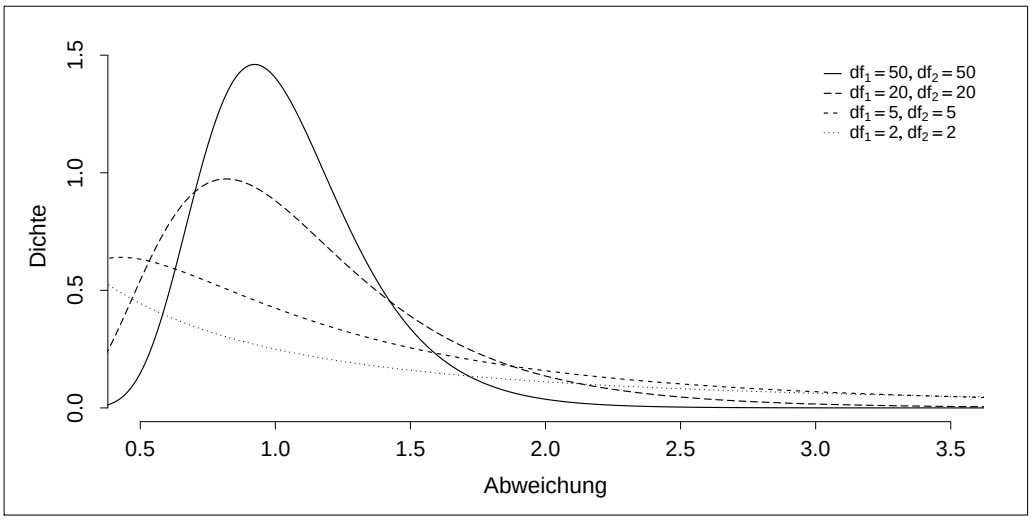


Abbildung 4.8: Darstellung der F-Verteilung in Abhängigkeit der Stichprobengrößen bzw. der zur Verfügung stehenden Freiheitsgraden

Während bei Stichprobengrößen von $n_1 = 51$ und $n_2 = 51$ ($df_1 = 50$ und $df_2 = 50$) die F-Verteilung annähernd normalverteilt ist und ein bedeutsamer Unterschied sich schon bei einem F-Wert größer als 1,6 einstellt, beginnen die äußeren 5% der F-Verteilung mit jeweils 2 Freiheitsgraden (also jeweils drei Messwerten pro Stichprobe) erst bei einen F-Wert von 5,05.

Will man einen t-Test rechnen, so ist Varianzhomogenität eine Bedingung für ein valides Ergebnis. Von Varianzhomogenität kann dann mit höherer Wahrscheinlichkeit ausgegangen werden, wenn das Ergebnis des F-Tests *nicht* signifikant ist, sich also kein großer F-Wert aus obiger Formel (relativ zur F-Verteilung) errechnet.
Ist der empirische F-Wert größer als der kritische F-Wert, der die akzeptierte Irrtumswahrscheinlichkeit angibt (meist α=5%), so muss man stattdessen von Varianzheterogenität ausgehen. In diesem Fall sollte man wiederum auf ein nonparametrisches Verfahren ausweichen, um die Bedeutsamkeit eines Mittelwertsunterschied statistisch valide zu prüfen.

4.C Aufgaben

inhaltliche Aufgaben

1. Was ist der Grundgedanke des t-Tests?
2. Wie verändert sich die t-Verteilung in Abhängigkeit der Stichprobengrößen? Warum?
3. Warum ist die Normalverteilung der Populationswerte eine Vorraussetzung des t-Tests? Wann ist diese Vorraussetzung zu vernachlässigen?
4. Welches Skalenniveau müssen die Daten mindestens haben, damit sie sich mithilfe eines t-Tests vergleichen lassen?
5. Welche Vorraussetzung hat der t-Test noch? Wie lässt sich diese testen?
6. Warum macht es keinen Sinn beim F-Test zweiseitig zu testen?
7. Wie errechnen sich die Freiheitsgrade beim t-Test? Wie beim F-Test?
8. Wie ändert sich die F-Verteilung bei wachsenden Stichprobengrößen? Welche Konsequenz hat dies für den kritischen F-Wert und die Varianzdifferenz zweier Stichproben?

praktische Aufgaben

1. Zur Überprüfung der Fragestellung, ob die Einnahme eines bestimmten Medikamentes die durchschnittlichen Schlafzeiten verlängert, wurden 32 Versuchspersonen per Zufall einer Versuchs- und einer Kontrollgruppe zugeordnet. Für die Versuchsgruppe (n = 16) ergaben sich ein Mittelwert von 9 Zeiteinheiten und einer geschätzten Standardabweichung von $\hat{\sigma}_1$= 4, für die Kontrollgruppe (n = 16) ein Mittelwert von 7 Zeiteinheiten und eine Standardabweichung von $\hat{\sigma}_2$= 3. Beurteile bitte die eingangs gestellte Frage bei einem Risiko von 5%.
2. Ein Psychologiestudent hat in einer Zeitung über eine Studie gelesen, in der gefunden wurde, dass selbstständige Unternehmer eine bessere Allgemeinbildung haben als Studenten. Um die Ehre der Studierenden zu retten, will er untersuchen, ob dieses Ergebnis eventuell auf das höhere Lebensalter (und damit auf mehr Lebenserfahrung) der Unternehmer im Vergleich zu den Studierenden zurückgeführt werden kann. Unten sind die Altersangaben der beiden Stichproben aufgeführt, die der Psychologiestudent freundlicherweise vom Autor der Studie erhalten hat:

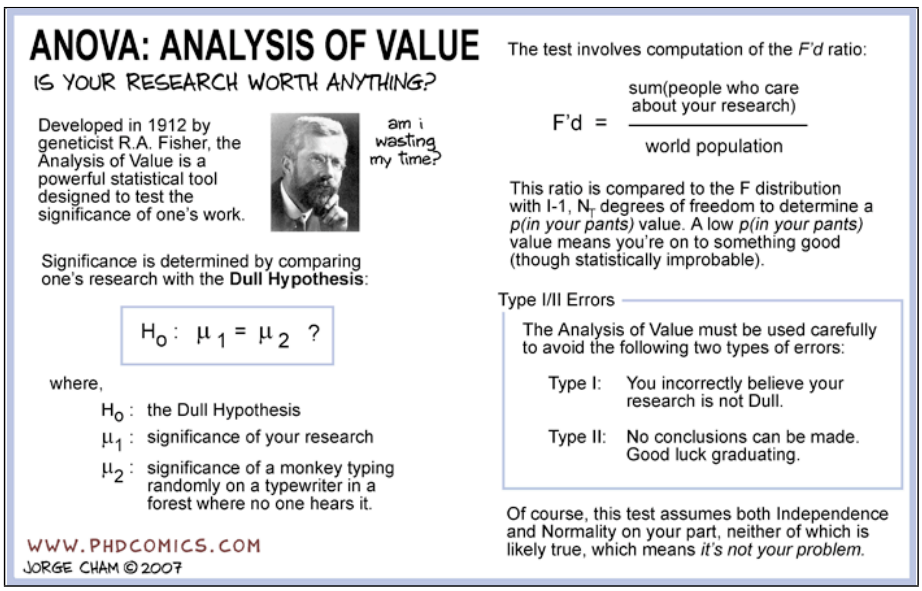
Studierende	22	23	31	35	25	27	25	28	23	21	42	24
Unternehmer	26	31	38	29	39	29	45	57	31	46	43	38

   Von Normalverteilung der Daten in der Population kann ausgegangen werden. Unterscheiden sich die Studierenden und die Unternehmer hinsichtlich des Alters (α=0,05)?
3. Es wird vermutet, dass das Bildungsniveau den Termin der 1. Schwangerschaft bei Frauen beeinflusst. Bei zwei Stichproben von jeweils 121 Frauen wurde das Alter erhoben, in

dem sie erstmals ein Kind gebaren. Die erste Stichprobe (n_1) umfasste Frauen, die maximal einen Realschulabschluss nachweisen konnten, Frauen der zweiten Stichprobe (n_2) besaßen dagegen alle die Hochschulreife. Folgende deskriptivstatistischen Kennwerte wurden ermittelt:

$$\bar{x}_1 = 18,5 \quad \hat{\sigma}_1 = 5,8 \qquad \bar{x}_2 = 20,7 \quad \hat{\sigma}_2 = 6,6$$

Prüfe inferenzstatistisch, ob zwischen den beiden Stichproben bedeutsame Unterschiede bezüglich des Alters bestehen, mit dem erstmals ein Kind geboren wurde. Von Normalverteilung des Merkmals kann ausgegangen werden.



4.3.3 Der t-Test für gepaarte Stichproben

Beim t-Test für gepaarte Stichproben, auch t-Test für abhängige Stichproben genannt, sind die Messwerte der einen Stichprobe nicht unabhängig von denen der zweiten Stichprobe. Stattdessen liegen Wertpaare vor. Dies ist zum Beispiel der Fall bei Untersuchungen mit Messwiederholung. Dabei wird ein Merkmal x mehrmals bei einem Proband i gemessen. x_{1i} und x_{2i} korrelieren dabei typischerweise miteinander.
Wertpaare, die abhängig voneinander sind, entstehen jedoch nicht nur bei Messwiederholungsdesigns. Sobald Stichproben nach einem bestimmten Kriterium „gematcht“ werden, entsteht die für einen gepaarten t-Test charakteristische Abhängigkeit der Werte.

Beispiel:

Angenommen man stellt die Theorie auf, dass erstgeborene Geschwister autonomer sind als zweitgeborene, weil Erstgeborene öfter und früher Verantwortung in der Familie übernehmen müssen.
Will man diese Theorie empirisch untersuchen, so könnte man Geschwisterteile beispielsweise mit einem standardisierten Fragebogen untersuchen, welcher versucht den Grad an Autonomie zu erfassen.
In Tabelle 4.1 ist ein Datensatz wie er aus einer solchen Untersuchung sich hätte ergeben können dargestellt. In den Spalten sind die Daten von jeweils zwei Geschwisterpaaren enthalten, wobei die Spalte x_1 den erreichten Punktwert in dem Fragebogen des erstgeborenen Geschwisters, die Spalte x_2 den erreichten Punktwert des zweitgeborenen Geschwisters angibt. Die Skala des Fragebogens besitzt einen Minimalwert von null und einen Maximalwert von 20, wobei ein niedriger Wert eine geringe Autonomie-Ausprägung kodiert und umgekehrt.

Beim t-Test für gepaarte Stichproben werden nun Differenzwerte aus den Wertpaaren errechnet, wie geschehen in der dritten Zeile der unteren Tabelle. Der Index d steht für Differenz. Wie man sehen kann liegen die meisten Werte im positivem Bereich. Auch der Mittelwert der Differenzwerte $\bar{x}_d$ liegt im positiven Bereich. Dies deutet rein deskriptiv darauf hin, dass sich das erstgeborene Geschwisterteil tendenziell als autonomer beschrieb relativ zum zweitgeborenem Geschwisterteil.
Dieser Trend lässt sich mithilfe von Abbildung 4.9 darstellen. Die Datenpunkte in der Abbildung entsprechen den Wertpaaren aus Tabelle 4.1. Die durchgezogene Linie ist nicht zu verwechseln mit einer geschätzten Regressionsgeraden. Die Linie symbolisiert vielmehr die Verteilung der Werte, die man unter Annahme der H0 Hypothese annehmen würde. In diesem Fall sollten die Geschwisterpaare jeweils den selben Punktwert in dem Fragebogen erreichen. Stattdessen liegen viele Datenpunkte über der Geraden, woraus man eine positive Differenz ableiten kann.

Die eingangs beschriebene Theorie über Autonomie bei Geschwisterpaaren wurde gerichtet formuliert. Dementsprechend werden in unserem Beispiel auch die H0-Hypothese und die Alternativhypothese in formalisierter Schreibweise gerichtet aufgestellt:

$$H_0 : \mu_d \leq 0 \tag{4.16}$$

$$H_1 : \mu_d > 0 \tag{4.17}$$

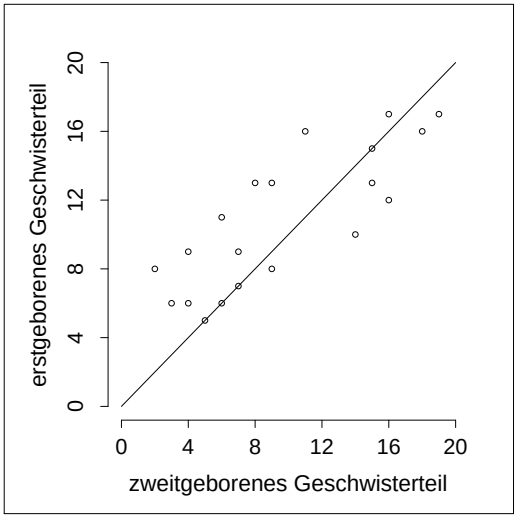


Abbildung 4.9: Darstellung der Wertpaare mit „Erwartungsline“ unter Annahme der H0

Tabelle 4.1
Beispieldatensatz mit 20 Wertpaaren

$\mathbf{x_1}$	6	13	17	6	16	7	16	6	15	12	13
$\mathbf{x_2}$	4	9	16	6	18	7	11	3	15	16	8
$\mathbf{x_d}\ (\mathbf{x_1} - \mathbf{x_2})$	2	4	1	0	-2	0	5	3	0	-4	5

$\mathbf{x_1}$	5	8	11	9	17	10	13	9	8	
$\mathbf{x_2}$	5	9	6	4	19	14	15	7	2	
$\mathbf{x_d}\ (\mathbf{x_1} - \mathbf{x_2})$	0	-1	5	5	-2	-4	-2	2	6	$\bar{x}_d = 1,15$

Wie beim t-Test für unabhängige Stichproben lässt sich auch beim t-Test für abhängige Stichproben ein t-Wert errechnen, den man mithilfe der t-Verteilung auf Signifikanz prüft;

$$t = \frac{\bar{x}_d - \mu_d}{\hat{\sigma}_{\bar{x}_d}} \tag{4.18}$$

Da wir μ_d unter der H0-Hypothese einen Wert von null (oder kleiner) erwarten, verkürzt sich die Formel folgendermaßen:

$$t = \frac{\bar{x}_d}{\hat{\sigma}_{\bar{x}_d}} \tag{4.19}$$

Würden wir eine nummerisch klar definierte Differenz erwarten, so könnte man diese für μ_d auch einsetzen. Dies würde beispielsweise Sinn machen, wenn man die Auswirkung eines Konzentrationstrainings auf die Leistung in einem IQ-Test testen will. Man könnte zunächst einen IQ-Test bei den Probanden durchführen, anschließend das Konzentrationstraining, und da-

nach wiederum einen IQ-Test. Die verbesserte Leistung im zweiten IQ-Test könnte durch einen reinen Übungseffekt entstanden sein, der nichts mit dem Konzentrationstraining zu tun hat. Angenommen man hat aus früheren Untersuchungen herausgefunden, dass dieser Übungseffekt sich in 5 IQ-Punkten niederschlägt. So könnte man unter der H0 eine generelle Differenz von 5 IQ Punkten annehmen und für μ_d in obige Formel 5 einsetzen. Erst Differenzen darüber hinaus würden als bedeutsam erachtet werden und mit der Auswirkung des Konzentrationstraining in Verbindung gebracht werden.

Der Standardfehler der Mittelwerte der Differenzen $\hat{\sigma}_{\bar{x}_d}$ lässt sich folgendermaßen schätzen:

$$\hat{\sigma}_{\bar{x}_d} = \frac{\hat{\sigma}_d}{\sqrt{n}} \tag{4.20}$$

wobei:

$$\hat{\sigma}_d = \sqrt{\frac{\sum_{i=1}^{n} d_i^2 - \frac{(\sum_{i=1}^{n} d_i)^2}{n}}{n-1}} \quad \text{bzw.} \quad \hat{\sigma}_d = \sqrt{\frac{\sum_{i=1}^{n} (\bar{x}_d - x_{di})^2}{n-1}} \tag{4.21}$$

Zu beachten hierbei ist, dass n der Anzahl der Messwertpaare entspricht, also der Wertemenge von x_d und nicht der Wertemenge von x_1 und x_2.

Der kritische t-Wert, also jener Wert der t-Verteilung, welcher die extremsten 5% bzw. 1% von dem Rest der Verteilung trennt, lässt sich wie beim t-Test für unabhängige Stichproben mithilfe einer t-Werte Tabelle ablesen. Hierbei ist wieder zu beachten, ob ein- oder zweiseitig getestet werden soll.
Die Freiheitsgrade ergeben sich aus der Anzahl der Messwertpaare n minus 1.

Beim t-Test für abhängige Stichproben gelten die selben Vorraussetzung wie beim t-Test für unabhängige Stichproben.

4.D Aufgaben

inhaltliche Aufgaben

1. Wann rechnet man einen t-Test für abhängige Stichproben?

praktische Aufgaben

1. Errechne mithilfe der angegeben Formeln und den Beispieldaten, ob erstgeborene Geschwister angeben eine signifikant höhere Ausprägung in dem Merkmal Autonomie zu haben als ihre zweitgeborenen Geschwister (α=0,05). Von Intervallskalenniveau der Daten und Normalverteilung in der Population kann ausgegangen ausgehen.

2. In einer Untersuchung zu figuralen Nachwirkungen sollten 25 Versuchsteilnehmer zunächst den Durchmesser eines schwarzen Kreises auf weißem Hintergrund schätzen. Anschließend mussten sie einen einen größeren Kreis eine Minute lang fixieren. Nach dem Verschwinden der Figur wurde wieder der ursprüngliche, kleinerer Kreis an gleicher Stelle präsentiert. Dieser Kreis sollte nun wiederum hinsichtlich seines Durchmessers geschätzt werden. Erwartet wurde, dass die Durchmesserschätzungen nun deutlich geringer ausfallen würden.

   Folgende Differenzen (Vorher – Nachher in Millimetern) wurden ermittelt:

-2	10	0	9	-2	7	0	4	-2	0	17	12	-3
5	5	7	0	3	1	2	4	0	8	1	-2	

   a) Beschreibe die Stichprobenverteilung auf der Basis eines Box-Whiskers-Plots.
   b) Prüfe die Hypothese inferenzstatistisch (α=0,05). Von intervallskalierten Werten und Normalverteilung in der Population kann ausgegangen werden.

3. Der leitende Psychologe einer Reha-Klinik möchte nachweisen, dass die von ihm angewandte Biofeedback-Therapie dauerhaft zu Blutdrucksenkungen bei Patienten mit essentieller Hypertonie führt. Zu Beginn der klinischen Therapie und 6 Monate nach deren Abschluss werden die folgenden systolischen Blutdruckwerte erhoben (von intervallskalierten Daten kann ausgegangen werden). Prüfe inferenzstatistisch, ob die Erwartungen bestätigt werden können ($\alpha = 5\%$).

VpNr	1	2	3	4	5	6	7	8	9	10
Vorher	170	180	172	176	176	176	172	178	182	164
Nachher	168	172	162	170	158	166	168	152	164	172

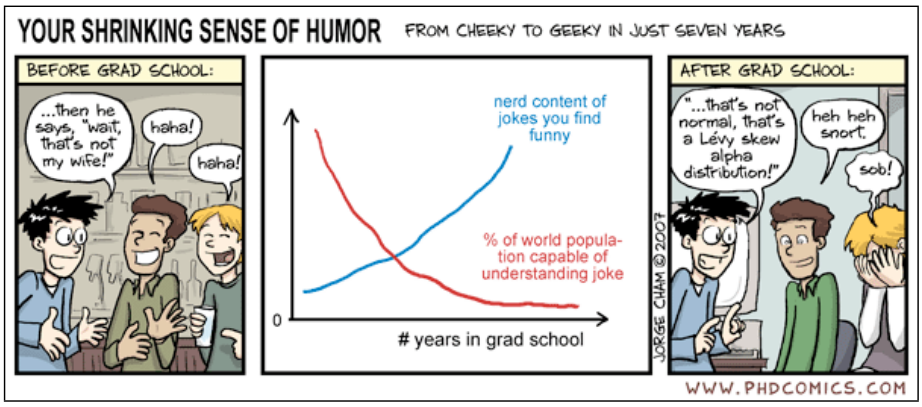
VpNr	11	12	13	14	15
Vorher	174	192	174	168	184
Nachher	174	160	174	172	174

4. Ein pädagogischer Psychologe nimmt an, dass freies Arbeiten in der Schule die Freude am Unterricht erhöht. Er lässt daher in verschiedenen Schulen fünfte Klassen die Lehrkraft auswählen, bei der der Unterricht am meisten bzw. am wenigsten Spaß macht und misst anschließend in den entsprechenden Schulstunden, wie viele Minuten pro Stunde freie Aufgaben gemacht werden. Er erhält folgende Werte (+ und – stehen jeweils für die beste/schlechteste Lehre):

Unterricht	5a+	5a-	5b+	5b-	5c+	5c-	5d+	5d-
freie Arbeit (Min./Std)	20	5	10	11	5	45	15	0

Unterricht	5e+	5e-	5f+	5f-	5g+	5g-	5h+	5h-
freie Arbeit (Min./Std)	7	1	12	3	22	6	18	12

Was kann er über seine Hypothese auf 5%-Niveau sagen?



4.3.4 Der Mann-Whitney-U-Test

Ist eine der Vorraussetzungen des t-Tests verletzt, so kann man auf alternative Testverfahren ausweichen. Im Falle von unabhängigen Stichproben wäre der so genannte Mann-Whitney-U-Test eine Alternative, benannt nach den Statistikern Henry Mann und Donald Whitney sowie Frank Wilcoxon (daher wird der U-Test auch manchmal Mann-Whitney-Wilcoxon-Test genannt - kurz: MWW-Test).
Der U-Test stellt keine Verteilungsannahme über die Daten auf und gehört daher zu der Kategorie der so genannten nonparametrischen Verfahren. Darüber hinaus verliert der U-Test auch bei Varianzheterogenität nicht an statistischer Validität. Die Daten müssen allerdings auf Ordinalniveau vorliegen.
Vergleicht man zwei unabhängige Stichproben auf Intervallskalenniveau, bei denen keine Varianzhomogenität gegeben ist (signifikantes Ergebnis beim F-Test) so wäre eine Möglichkeit, die Daten auf Ordinalniveau „runter zu brechen“ (vgl. Abschnitt 1.2.5) und einen U-Test durchzuführen.

Beim U-Test nach Mann und Whitney wird getestet, ob sich die Ränge in ihre Größe überwahrscheinlich ungleich auf die Gruppen verteilen (ungerichtet formuliert). Die Hypothesen im ungerichteten Fall werden folgendermaßen aufgestellt:

$$H_0 : E(\bar{R}_1) = E(\bar{R}_2) \tag{4.22}$$

$$H_1 : E(\bar{R}_1) \neq E(\bar{R}_2) \tag{4.23}$$

wobei E($\bar{R}$) für den Erwartungswert der Rangmittelwerte steht

Zur Berechnung eines empirischen U-Wertes werden zunächst die Rangsummen T_1 und T_2 für die beiden Stichproben seperat errechnet. In die U-Formel wird nun immer der kleinere T-Wert eingesetzt, also jener T-Wert der Stichprobe, bei der in der Summe weniger große Ränge vorhanden sind. In der unteren Formel wäre das der T-Wert der ersten Stichprobe.

$$U = T_1 - \frac{n_1 \cdot (n_1 + 1)}{2} \tag{4.24}$$

Berechnet man auch einen U-Wert für die zweite Stichprobe, deren Rangsumme größer ist, so erhält meinen einen U'-Wert. Vergleicht man U und U', so lässt sich schon deskriptiv feststellen, ob überhaupt ein Unterschied zwischen den Gruppen in ihren Rangsummen besteht. Verteilen sich die Ränge in ihrer Größe gleichmäßig auf die beiden Gruppen, so gilt: U = U'.

$$U' = T_2 - \frac{n_2 \cdot (n_2 + 1)}{2} \tag{4.25}$$

Rechnet man sowohl U als auch U' aus, so lassen sich auch Rechenfehler leicht überprüfen, da auf jeden Fall gelten muss:

$$U' + U = n_1 \cdot n_2 \tag{4.26}$$

In manchen U-Test Tabellen kann die Höhe der Fehlerwahrscheinlichkeit bei gegebenem n_1, n_2 und U direkt abgelesen werden. Liegt diese Fehlerwahrscheinlichkeit unter 0,05, so würde man die Alternativhypothese annehmen, vorausgesetzt man akzeptiert eine Irrtumswahrscheinlichkeit von $\alpha = 5\%$.
Liest man die Fehlerwahrscheinlichkeit in einer Tabelle direkt ab, so muss man darauf achten, ob die angegeben Werte sich auf eine zweiseitige oder einseitige Testung beziehen. Angenommen die Wahrscheinlichkeitswerte beziehen sich auf eine einseitige Testung, so muss man die Fehlerwahrscheinlichkeit verdoppeln, wenn man zweiseitig testen will, da in diesem Fall die akzeptierte Irrtumswahrscheinlichkeit sich auf beide Seiten der Verteilung verteilt (siehe Abbildung 4.5).

Zieht man eine U-Tabelle heran, in der statt der Fehlerwahrscheinlichkeit ein kritischer U-Wert bei festgelegtem α-Niveau angegeben ist, so wird die H1-Hypothese dann akzeptiert, wenn gilt:

$$U_{emp} < U_{krit} \quad \rightarrow \quad H1 \tag{4.27}$$

Dies ist genau konträr zu den bisher besprochenen Tests. Ist der empirische U-Wert nach Formel 4.24 kleiner als der kritische U-Wert aus der U-Tabelle, so verteilen sich die Ränge ungleichmäßig auf die beiden Gruppen.

Dies wird klarer, setzt man den U-Wert aus Formel 4.24 in eine z-Formel ein:

$$z = \frac{U - \hat{\mu}_U}{\hat{\sigma}_U} \tag{4.28}$$

Einen z-Wert zu errechnen ist erst bei größeren Stichproben legitim (n_1 oder $n_2 > 10$), da erst dann die theoretisch möglichen z-transformierten Rangwerte U annähernd um den Wert μ_U normalverteilt sind.

Für $\hat{\mu}_U$ errechnen wir die größtmögliche Zahl an Rangvermischungen:

$$\hat{\mu}_U = \frac{n_1 \cdot n_2}{2} \tag{4.29}$$

U entspricht also $\hat{\mu}_U$ wenn die Ränge der beiden Stichproben (relativ zur Stichprobengröße) sich gleichmäßig verteilen (wenn $n_1 = n_2$ gilt U = $\hat{\mu}_U$ wenn $T_1 = T_2$).
Der Standardfehler des U-Wertes $\hat{\sigma}_U$ errechnet sich folgendermaßen:

$$\hat{\sigma}_U = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}} \tag{4.30}$$

Der kritische Wert, den es zu überschreiten gilt, soll die Alternativhypothese bestätigt werden, ist nun je nach Stichprobengröße der t- oder z-Verteilung zu entnehmen. Bei $n_1 + n_2 > 50$ sind die theoretisch möglichen U-Werte bei geltender H0-Hypothese annähernd standardnormalverteilt (also z-verteilt), so dass auf die z-Tabelle zurückgegriffen werden kann.
Bei $n_1 + n_2 \leq 50$ sollte eine t-Tabelle zur Ermittlung der kritischen Grenze bei gegebenem α-Niveau herangezogen werden.

4.E Aufgaben

inhaltliche Aufgaben

1. Wann wird ein U-Test gerechnet?
2. Welche Verteilungstabellen kann ich zur Bestimmung des kritischen U-Wertes heranziehen? Unter welchen Bedingungen entscheide ich mich für welche Verteilungstabelle?

praktische Aufgaben

1. Es werden zwei Trainingsformen für stotternde Kinder verglichen. Nach dem Training wird in jeder Gruppe die durchschnittliche Anzahl von Versprechern pro Stunde erhoben. Es stellt sich heraus, dass die erhobenen Daten stark von der Normalverteilung abweichen.

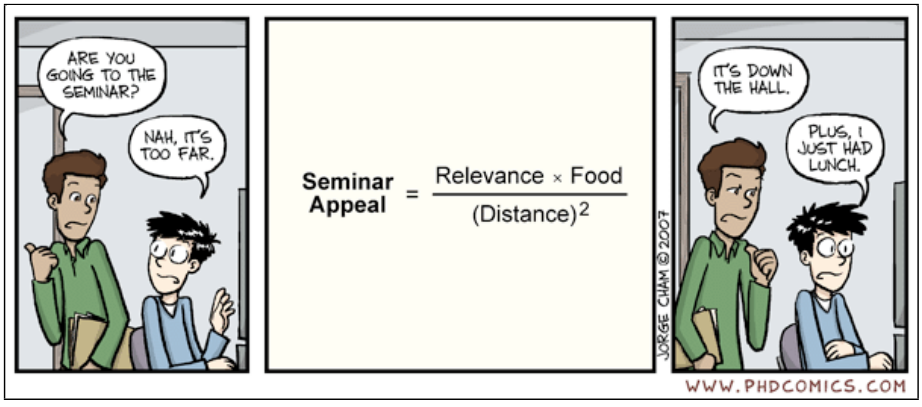
   Transformiere die Daten, sodass sie auf Ordinalniveau vorliegen.

   Unterscheiden sich die Trainings signifikant voneinander?

Gruppe 1	13,9	32,2	17,5	17,4	28,7	38,7	17,9	20,8	12,5	32,4
Gruppe 2	24,5	22,7	26,5	23,9	40,6	41,1	46,1	31,2	35,4	41,4

2. Es wird vermutet, dass Zimmerpflanzen das allgemeine Wohlbefinden steigern. Ein Psychologe vergleicht dazu das allgemeine Stressniveau von Menschen, die keine Zimmerpflanzen besitzen, mit Personen, die mehr als drei Pflanzen in ihrem Zimmer haben. Als Indikator für Stress wird der Gehalt von Stresshormonen im Blut gemessen (in mg/100l). Dabei stellt sich heraus, dass der Stresshormonspiegel nicht der Normalverteilung entspricht. Prüfe die Signifikanz der Ergebnisse.

Mehr als drei Pflanzen	10	7,5	8	12	15	6,5	7	6
Keine Pflanzen	12	11	18	9	13	5,5	13,5	17



4.3.5 Der Chi-Quadrat Test

Mithilfe des Chi-Quadrat Test (abgekürzt: χ^2-Test) lässt sich überprüfen, ob sich die Häufigkeit des Auftretens einer Merkmalsausprägung überzufällig zwischen Gruppen unterscheidet. Der χ^2-Test kommt also dann zur Anwendung, wenn Daten auf Nominalniveau vorliegen.
Der χ^2-Test ist, ähnlich wie der U-Test, auch anwendbar bei Daten, die auf einem höheren Skalenniveau vorliegen. Diese Daten müssen jedoch dann in das Nominalniveau transformiert werden, indem man beispielsweise Wertkategorien bildet (vgl. Abschnitt 2.3.1).

eindimensionaler Chi-Quadrat Test

Man verwendet eindimensionale χ^2-Methoden, wenn man Häufigkeitsunterschiede auf den verschiedenen Stufen *eines* Merkmals analysieren will.

Beispiel: Bei der Messung der Auftretenshäufigkeit eines Merkmals oder einer Merkmalsausprägung könnte man als weitere Variable das Geschlecht mit den Stufen männlich und weiblich erfassen. Daraus würde sich folgende Häufigkeitstabelle ergeben:

	männlich	**weiblich**
Häufigkeit	a	b

Die Hypothesen werden im ungerichteten Fall beim χ^2-Test folgendermaßen aufgestellt:

$$H_0 : p_{ij} = p_i p_j \tag{4.31}$$

$$H_1 : p_{ij} \neq p_i p_j \tag{4.32}$$

Wobei unter Annahme der H0 gilt: Die Wahrscheinlichkeit p für die Auftretenswahrscheinlichkeit der Merkmalsausprägung i entspricht der Auftretenswahrscheinlichkeit p für die Merkmalsausprägung j, während in der Alternativhypothese postuliert wird, dass die Einzelwahrscheinlichkeiten p_i und p_j für das Auftreten einer Merkmalsausprägung i bzw. j sich unterscheiden.

Die erwartete Häufigkeit für obiges Beispiel, unter Annahme der H0, wäre, dass sowohl bei Männern als auch bei Frauen gleich oft das Merkmal bzw. die Merkmalsausprägung x beobachtet wird;

$$f_e = \frac{f_{b(1)} + f_{b(2)}}{2} \tag{4.33}$$

b = beobachtete Häufigkeiten

b(1) = beobachtete Häufigkeit in dem Merkmal männlich

b(2) =beobachtete Häufigkeit in dem Merkmal weiblich

Die erwarte Häufigkeit wird in der χ^2-Formel mit der tatsächlichen Häufigkeit in Beziehung gebracht:

$$\chi^2 = \sum_{j=1}^{k} = \frac{(f_{b(j)} - f_{e(j)})^2}{f_{e(j)}} \tag{4.34}$$

k = Anzahl der Stufen des Merkmals (oben: Geschlecht; k = 2)

Der empirische χ^2-Wert wird nun mit dem kritischen χ^2 verglichen, wobei sich der kritische χ^2 je nach akzeptierter Irrtumswahrscheinlichkeit verändert. Der eindimensionaler Chi-Quadrat Test orientiert sich an der χ^2-Verteilung mit einem Freiheitsgrad, da nur ein Häufigkeitswert frei variieren kann.

mehrdimensionaler Chi-Quadrat Test

Mehrdimensionale χ^2-Methoden werden dann eingesetzt, wenn man Häufigkeitsunterschiede auf den verschiedenen Stufen *mehrerer* Merkmale analysieren will.

Beispiel: Bei der Messung der Auftretenshäufigkeit eines Merkmals oder einer Merkmalsausprägung werden als weitere Variablen die Konfession mit den Stufen evangelisch und katholisch und das Alter mit den Kategorien alt und jung erfasst. Es ergibt sich folgende Häufigkeitstabelle mit den Häufigkeiten a, b, c und d im Merkmal x, die miteinander verglichen werden können:

	katholisch	**evangelisch**
jung	a	b
alt	c	d

Bei zwei Merkmalen mit jeweils zwei Ausprägungen ergibt sich ein Vierfelder-Schema und die Wahrscheinlichkeit einer Gleichverteilung der Häufigkeiten über die Merkmale hinweg lässt sich mithilfe des Vierfelder-χ^2-Test errechnen. Die Hypothesen werden dabei wie oben aufgestellt.
Der empirische χ^2-Wert errechnet sich über Formel 4.35.

$$\chi^2 = \frac{n \cdot (ad - bc)^2}{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)} \tag{4.35}$$

Die H0 Hypthese wird dann verworfen, wenn bei festgelegtem α-Niveau gilt;

$$\chi^2_{emp} > \chi^2_{krit} \tag{4.36}$$

Auch der Vierfelder-χ^2-Test besitzt nur einen Freiheitsgrad. Nur eine Merkmalshäufigkeit kann frei variieren. Ist beispielsweise die Häufigkeit der Merkmalskombination jung und katholisch auf einen Wert festgelegt, so sind alle weiteren Häufigkeiten von diesem Wert abhängig.

4.F Aufgaben

inhaltliche Aufgaben

1. Wann rechnen man einen χ^2-Test?
2. Was sind eindimensionale, was mehrdimensionale χ^2-Methoden?
3. Wieviel Dimensionen kann man mithilfe eines χ^2-Tests vergleichen?

praktische Aufgaben

1. In einer Kundenbefragung bei Douglas an n=1000 Kunden wurde festgestellt, dass 740 der befragten Personen Frauen waren und der Rest Männer. Handelt es sich hier um eine signifikante Differenz?
2. Was würde sich an der Berechnung in der oberen Aufgabe ändern, wenn die Fragestellung lautete, ob sich das Verhältnis, mit dem Frauen und Männer bei Douglas einkaufen gehen von dem unterscheidet, mit dem Frauen sonst im Verhältnis zu Männern einkaufen (dieses Verhältnis betrage 3:1)?
3. Eine Untersuchung in einem Betrieb wurde die Arbeitszufriedenheit und die Zugehörigkeit zu einer Altersgruppe erhoben. Die Tabelle zeigt die Ergebnisse:

	Alter < 40	**Alter > 40**
eher zufrieden	200	65
eher unzufrieden	100	65

   Unterscheiden sich jüngere und ältere Beschäftigte hinsichtlich der Arbeitszufriedenheit statistisch bedeutsam?
4. Briefträger werden häufig von Hunden gebissen. Im letzten Jahr trugen von 700 Briefträgern, die gebissen wurden, 290 lange Hosen und 410 kurze Hosen. Ist dies ein statistisch bedeutsamer Unterschied?

4.3.6 Relevanz von Mittelwertsunterschieden

Nach der Logik der hier besprochenen Signifikanztests sind auch kleine Unterschiede zwischen Stichproben statistisch signifikant, ist die Stichprobe hinreichend groß.

Das Unterschiede statistisch bedeutsam sind, bedeutet daher nicht, dass sie auch praktisch relevant sind. Man fand beispielsweise heraus, dass Säuglinge, die mit der Flasche aufgezogen wurden, später bei IQ-Tests durchschnittlich um ein bis zwei IQ-Punkte schlechter abschneiden als Säuglinge, die von der Mutter gestillt werden. Dieser Unterschied war statistisch bedeutsam, da eine relativ große Stichprobe zur Überprüfung dieses Unterschieds herangezogen wurde.

Dieser sehr kleine Unterschied ist von keiner großen praktischen Relevanz. Die Ernährungsweise im Säuglingsalter ist keine sehr ertragreiche Variable zur Erklärung von unterschiedlichen Leistungen in IQ-Tests im späteren Alter.

Der Statistiker Jacob Cohen hat daher einen Koeffizienten vorgeschlagen, mit dem man die praktische Relevanz von Mittelwertsunterschieden einschätzen kann. Der Koeffizient ergibt sich aus der Differenz der Mittelwerte zweier Stichproben, relativiert an der Streuung beider Stichproben;

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(s_1^2 + s_2^2)}{2}}} \tag{4.37}$$

Cohens d gibt die relative Stärke eines Effektes an und wird daher auch als Effektstärkemaß bezeichnet. Nach Cohen gilt ein $d \leq 0{,}3$ als kleiner Effekt, ein d zwischen 0,31 und $< 0{,}8$ gilt als mittlerer Effekt und ein $d > 0{,}8$ ist nach Cohen's Definition ein großer Effekt.

4.3.7 Zusammenfassung

Die hier besprochenen Signifikanztest zur Überprüfung von Unterschiedshypothesen unterscheiden sich darin, auf welchem Skalenniveau die Daten vorliegen müssen. Darüber hinaus stellen manche Tests Verteilungsannahmen über die Daten auf, sowie Annahmen über die Dispersion der Daten in der Population.

In Abbildung 4.10 ist ein Entscheidungsbaum dargestellt, der dabei helfen kann, welcher Test bei welchen Annahmen und Skalenniveaus statistisch valide Ergebnisse liefert bei der Beurteilung von der Bedeutsamkeit von Unterschieden zwischen Stichproben, wobei der Wilcoxon Rangsummentest und der McNemar-Test hier nicht besprochen wurde, der Vollständigkeit halber aber mit aufgenommen wurde.

4.4 Signifikanz von Korrelationen

Neben der statistischen Bedeutsamkeit von Unterschieden zwischen zwei Stichproben lässt sich auch die statistische Bedeutsamkeit von Zusammenhängen errechnen. Angenommen die Population zweier Variablen sei bekannt. Des Weiteren ist bekannt, dass diese zwei Variablen nicht miteinander korrelieren. Würde man nun tausend mal Wertpaare der Größe n aus beiden Populationen ziehen und jeweils Korrelationskoeffizienten errechnen, so würden sich diese

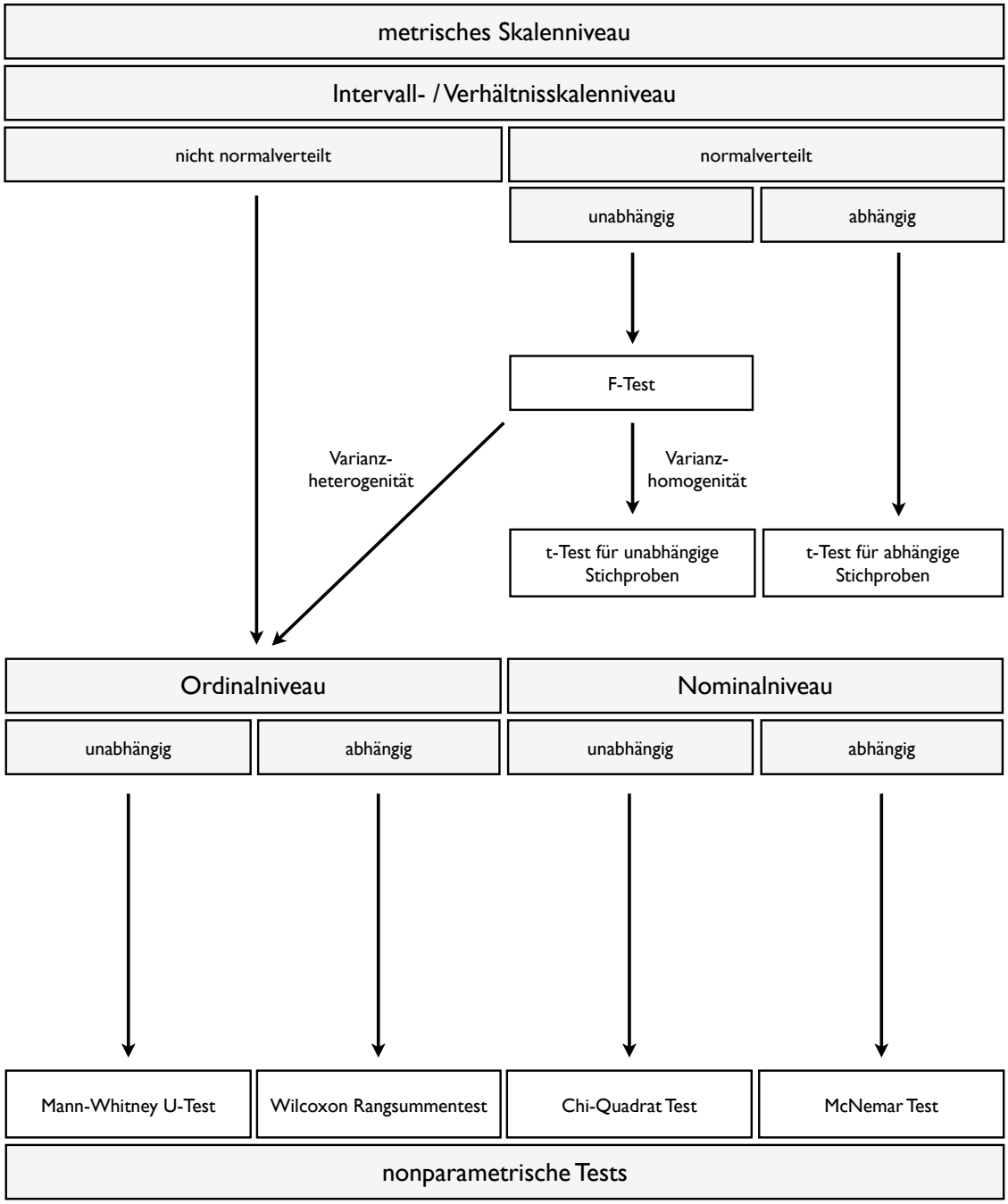


Abbildung 4.10: Entscheidungsbaum für inferenzstatistische Testverfahren

Korrelationskoeffizienten um den Nullpunkt t-verteilen. Bei $n > 30$ würde diese t-Verteilung ungefähr einer z-Verteilung entsprechen.

Mit dem Wissen darüber, wie sich Korrelationskoeffizienten bei einem solchen Zufallsexperiment verteilen, lässt sich entscheiden, ob ein einzelner Korrelationskoeffizient einer Stichprobe der Größe n sich statistisch bedeutsam von 0 verschieden ist.

Die Hypothesen werden dabei wie in Formel 4.37 und 4.38 aufgestellt, wobei ρ (griechisches rho) für den wahren Korrelationskoeffizient in der Population steht.

$$H_0 : \rho = 0 \tag{4.38}$$

$$H_1 : \rho \neq 0 \tag{4.39}$$

Der empirische t-Wert lässt sich folgendermaßen errechnen:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{4.40}$$

Um einzuschätzen, wie wahrscheinlich es ist, dass ein solcher t-Wert unter Annahme der H0 auftritt, wird wie beim t-Test die t-Verteilung herangezogen. Wobei die t-Verteilung df = n-2 Freiheitsgrade besitzt. Bei großem n (ungefähr $n > 30$) kann man auch die z-Verteilung heranziehen.

Legt man einen kritischen t-Wert fest, also einen Schwellenwert, über und unter dem der prozentuale Anteil der Verteilung liegt, den man als Irrtumswahrscheinlichkeit akzeptiert (bezogen auf eine zweiseitige Testung), so wird, wie beim t-Test, die H0-Hypothese dann abgelehnt, wenn der empirische t-Wert, über (bzw. unter) dem kritischen Wert liegt.

Der Korrelationskoeffizient der Stichprobe gilt dann als statistisch bedeutsam. Die Wahrscheinlichkeit, dass dieser durch reinen Zufall entstanden ist und die wahre Korrelation ρ der beiden Variablen 0 ist, ist so gering, dass man eher die H1 annehmen kann.

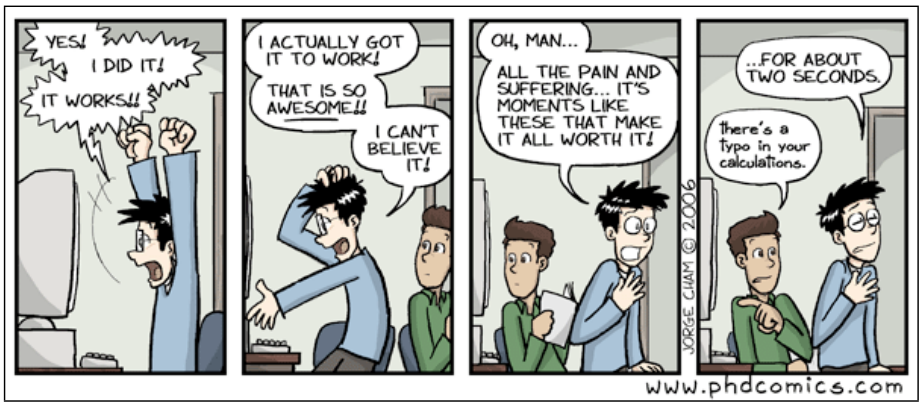
4.G Aufgaben

inhaltliche Aufgaben

1. Warum werden Korrelationen auf Signifikanz geprüft?
2. Wie geht man bei der Signifikanzprüfung von Korrelation vor? Erstelle eine schrittweise Anleitung.

praktische Aufgaben

1. In einer Stichprobe von 23 männlichen Personen zeigt sich ein Zusammenhang zwischen den Variablen „Anzahl der gerauchten Zigaretten“ und „Anzahl der getrunkenen Flaschen Bier“ von r = 0,49. Überprüfe die Hypothese, dass es auch in der Gesamtpopulation der Männer einen Zusammenhang zwischen den Variablen „Anzahl der gerauchten Zigaretten“ und „Anzahl der getrunkenen Flaschen Bier“ gibt (Signifikanzniveau $\alpha = 5\%$ und $\alpha = 1\%$).
2. In einer Stichprobe (n = 17) wurde ein Zusammenhang von r = .51 gefunden. Ist das ein signifikanter Zusammenhang bei einem α von 10%?



ANHANG

Seminar BINGO!

To play, simply print out this bingo sheet and attend a departmental seminar.

Mark over each square that occurs throughout the course of the lecture.

The first one to form a straight line (or all four corners) must yell out BINGO!! to win!

SEMINAR BINGO

Speaker bashes previous work	Repeated use of "um…"	Speaker sucks up to host professor	Host Professor falls asleep	Speaker wastes 5 minutes explaining outline
Laptop malfunction	Work ties in to Cancer/HIV or War on Terror	"…et al."	You're the only one in your lab that bothered to show up	Blatant typo
Entire slide filled with equations	"The data *clearly* shows…"	FREE Speaker runs out of time	Use of Powerpoint template with blue background	References Advisor (past or present)
There's a Grad Student wearing same clothes as yesterday	Bitter Post-doc asks question	"That's an interesting question"	"Beyond the scope of this work"	Master's student bobs head fighting sleep
Speaker forgets to thank collaborators	Cell phone goes off	You've no idea what's going on	"Future work will…"	Results conveniently show improvement

JORGE CHAM © 2007

WWW.PHDCOMICS.COM

A Statistik mit dem Taschenrechner

Die folgenden Erläuterungen zur Berechnung von statistischen Kennwerten mit dem Taschenrechner beziehen sich auf das Modell FX-350MS der Marke Casio. Sie sollten auch für ähnliche Modelle von Casio funktionieren. Bei Taschenrechner anderer Hersteller können die Tastenbezeichnungen und Eingaberoutinen von den hier beschriebenen abweichen.

Mittelwert und Standardabweichung

Über die Taste [Mode] gelangt man in ein Menü, in dem man verschiedene Modi einstellen kann. Bei zweimaligem Drucken der [Mode]-Taste erscheint an zweiter Stelle der Modus SD, den man mit Drücken der [2] aktivieren kann.

Nun lassen sich Daten in den Speicher eingeben, indem man die Zahlen der Reihe nach eingibt und zwischendrin die Taste [DT] (für Data) drückt. Das DT ist ein wenig versteckt; Es befindet sich blau gedruckt unter der Taste [M+]. Zwischen den Eingaben erscheint im Display ein n, gefolgt von der Anzahl der bisher eingegebenen Daten.

Sind alle Daten eingegeben, so kann man über [Shift] und der Taste [2] in ein Verarbeitungsmenü gelangen. An erster Stelle steht die Berechnung des Mittelwerts „$\bar{x}$“, die man mit der [1] und [=] bestätigt.

Die Standardabweichung lässt sich auch wiederum über das Menü errechnen, indem man [Shift] und anschließend die Taste [2] drückt und mit [=] bestätigt. Das Symbol für die Standardabweichung ist „$x\sigma n$“.
Soll die Varianz berechnet werden, so quadriert man einfach den Wert mit der Taste [x^2].
Die korrigierte Standardabweichung bzw. geschätzte Standardabweichung der Population (siehe 4.1.2) wird über den Menüpunkt „$x\sigma n$-1“ bzw. über das Drücken der [3] errechnet.

bivariate Regression

Den Modus Regression erreicht man über drücken der Taste [Shift]. An dritter Stelle sollte nun der Eintrag „Reg“ für Regression im Display stehen, den man über [3] auswählt. Will man Parameter für eine lineare Regression berechnen, so wählt man nun die [1] („Lin“).

Diesmal sollen statt einer Variablen zwei Variablen (x und y) in den Speicher gelesen werden. Die Wertpaare von x und y gibt man ein, indem man zuerst den Wert von x_1 eingibt, anschließend ein [,] setzt und dann den Wert von y_1 einträgt. Das Wertpaar speichert man mit [DT] und gibt dann sukzessive auf diese Weise die weiteren Wertpaare ein.

Den Steigungsparameter b kann man sich nun errechnen lassen, indem man [Shift] und die Taste [2] betätigt und solange mit den Pfeiltasten nach rechts drückt, bis der Eintrag „B“ im

Menü erscheint. Diesen wählt man mit der 2 aus und bestätigt mit =.
Die Konstante a der Regressionsgleichung lässt sich über das selbe Menü auswählen. Statt „A“ wählt man den Eintrag „B“ mit der Taste 2 und bestätigt ebenfalls mit =.

bivariate Korrelation

Die Produkt-Moment Korrelation nach Pearson lässt sich errechnen, indem man zunächst die Wertpaare von x und y wie oben beschrieben in den Taschenrechner eingibt. Anschließend wählt man den Eintrag „r“ aus dem Menü (über Shift) aus.

löschen des Speichers

Will man eine neue Datenreihe eintippen, so kann es passieren, dass noch alte Daten im Speicher vorhanden sind. Um sicher zu sein, dass der Speicher leer ist muss über Shift + Mode der Modus neu eingestellt werden. Im Display erscheint „Stat clear“ als Zeichen dafür, dass die alten Werte nicht übernommen werden.