

Peter Hautz

Statistik

**Skript zur Einführung
Version 3.1**

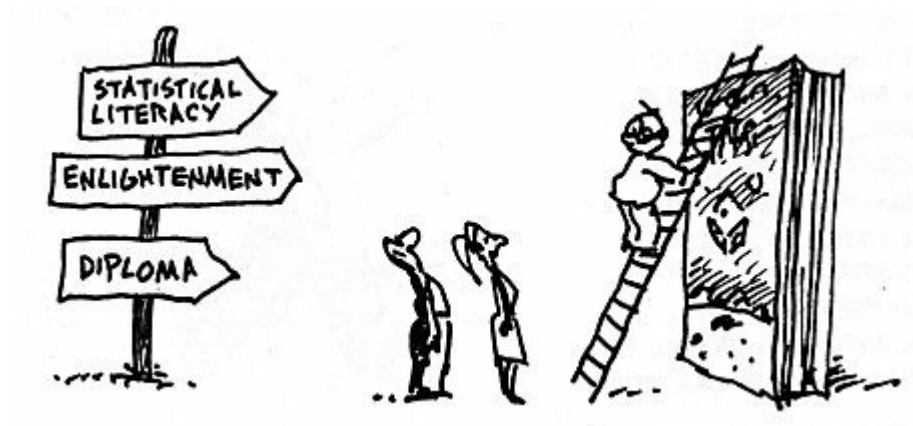


Abb. aus: The Cartoon Guide to Statistics by Larry Gonick & Woollcott Smith (1993)

Statistik
Skript zur Einführung
Version 3.1

Universität der Künste Berlin
Studiengang Gesellschafts- und Wirtschaftskommunikation
Peter Hautz
E-Mail: hautz@udk-berlin.de
Stand: 20.10.2004

Inhalt

Grundlegende Konzepte	1
Merkmale und Variablen	1
Abhängige und unabhängige Variablen	2
Operationalisierung und Messen.....	3
Skalenniveaus	4
Diskrete und stetige Variablen.....	5
Konstrukte und Indikatoren	6
Aufgaben	6
 Deskriptive Statistik	 8
Häufigkeitsverteilung.....	8
Kategorisierung (Klassifizierung)	9
Grafische Darstellungen von Häufigkeiten.....	10
Balken- und Säulendiagramm (bar chart)	11
Liniendiagramm (line chart, line plot)	13
Histogramm (histogram).....	14
Polygonzug (frequency polygon).....	15
Kreis- oder Tortendiagramm (pie chart)	15
Grafische Darstellungen anderer Kennwerte.....	16
Statistische Kennwerte	16
Maße der zentralen Tendenz (Lagemaße)	16
Dispersionsmaße (Streuungsmaße).....	17
Aufgaben	18
Exkurs: Rechnen mit dem Summenzeichen	18
Aufgaben	19
z-Transformation.....	20
Aufgabe	21
 Wahrscheinlichkeitstheorie	 22
Zufallsexperiment	22
Wahrscheinlichkeitskonzepte	23
1. Subjektive Wahrscheinlichkeit	23
2. Empirische Wahrscheinlichkeit	23
3. Theoretische Wahrscheinlichkeit.....	23
Aufgaben	24
Rechenregeln.....	24
Komplement, Komplementärereignis.....	24
Additionssatz	25
Bedingte Wahrscheinlichkeit	25
Multiplikationssatz	25
Aufgaben	26
Zufallsvariable und ihre Funktionen	26

Zufallsvariable	26
Wahrscheinlichkeitsfunktion	27
Wahrscheinlichkeitsdichte (Dichtefunktion)	27
Verteilungsfunktionen	28
Aufgaben	28
Verteilungsmodelle	29
Gleichverteilung.....	29
Normalverteilung.....	29
Wozu brauchen wir die Normalverteilung?.....	30
Standardnormalverteilung	31
Aufgabe	32
Stichprobentheorie	33
Stichprobenarten (Auswahlverfahren)	33
Zufallsverfahren (Wahrscheinlichkeitsauswahlverfahren)	33
Nicht-Zufallsverfahren (bewusste Auswahlverfahren)	34
Stichprobenkennwerte und Parameter	35
Stichprobenkennwerte-Verteilung.....	35
Die Streuung der Stichprobenkennwerte-Verteilung	36
Aufgaben	37
Die Form der Stichprobenkennwerte-Verteilung	38
Aufgabe	38
Der Mittelwert der Stichprobenkennwerte-Verteilung	38
Zusammenfassung.....	39
Intervallschätzung	39
Aufgaben	40
Hypothesen und ihre Überprüfung	41
Hypothesenarten	41
Forschungshypothesen und statistische Hypothesen	41
Gerichtete und ungerichtete Hypothesen.....	42
Spezifische und unspezifische Hypothesen	42
Überprüfen von Hypothesen.....	43
Prinzip des Signifikanztests.....	44
Entscheidung und mögliche Fehlentscheidungen	44
Berechnung der Irrtumswahrscheinlichkeit	45
Aufgaben	48
Signifikanz und Relevanz	49
Mittelwertunterschiede.....	50
Deskriptive und inferenzstatistische Ebene	50
t-Tests	50
t-Test für unabhängige Stichproben.....	51
t-Test für abhängige Stichproben.....	53
Voraussetzungen für t-Tests	55

Lineare Zusammenhänge.....	56
Lineare Regression.....	56
Aufgaben	56
Scatterplot	57
Kovarianz.....	57
Aufgabe.....	57
Regressionsgerade.....	57
Aufgaben	58
Korrelation (deskriptive Ebene)	58
Aufgaben	59
Korrelationstest (inferenzstatistische Ebene)	60
Voraussetzungen für den Korrelationstest	60
Aufgaben	60
 Analyse von Häufigkeiten	 62
Eindimensionale Chi-Quadrat-Tests	62
Allgemeiner eindimensionaler χ^2 -Test	62
Goodness-of-fit- χ^2 -Test (χ^2 -Anpassungstest)	64
Ein dichotomes Merkmal bei zweimaliger Untersuchung (McNemar- χ^2 -Test).....	65
Zweidimensionale Häufigkeitstabellen: Kreuztabellen (deskriptive Ebene).....	66
Zweidimensionaler Chi-Quadrat-Test.....	70
Voraussetzungen für χ^2-Tests.....	71
Aufgaben	71
 Literatur und Websites	 73
SPSS (Statistik-Software)	75
Der Rahmen: Forschungsmethoden (hier ist nur Literatur zu quantitativen Methoden aufgelistet).....	75

Grundlegende Konzepte

Merkmale und Variablen

Da wir in der quantitativen empirischen Forschung die Realität nicht als Ganzes erfassen können, wählen wir als Realitätsausschnitte bestimmte *Merkmale* aus. Das sind Aspekte oder Eigenschaften, hinsichtlich derer eine Person oder ein Objekt in einer empirischen Untersuchung beschrieben wird, z.B. „Umweltbewusstsein“ oder „Toleranz“.

Die konkrete Erscheinungsform eines Merkmals, die man bei einer Person oder einem Objekt feststellt, nennen wir *Merkmalsausprägung*. Für die statistische Analyse werden die Merkmalsausprägungen für jede Person bzw. jedes Objekt durch Zahlenwerte codiert. In der Statistik bezeichnet man Merkmale als *Variablen* und die codierten (Merkmals-) Ausprägungen als *Werte* (*values*).

Beispiele:

- Das Merkmal Geschlecht mit den möglichen Merkmalsausprägungen {weiblich, männlich} kann durch eine Variable X mit den möglichen (Zahlen-)Werten {1, 2} codiert werden: 1 für weiblich, 2 für männlich.
- Das Alter in Jahren ist ein Merkmal mit den möglichen Ausprägungen {1 Jahr, 2 Jahre, 3 Jahre, ..., 25 Jahre, 26 Jahre, ..., 70 Jahre, ...} und wird als Variable Y behandelt, die die Werte {1, 2, 3, ..., 25, 26, ..., 70, ...} annimmt.

Die in einer Stichprobe erfassten Ausprägungen bzw. Werte der einzelnen Personen oder Objekte werden mit einem tiefgestellten Index (Identifikationsnummer) durchnummeriert.

Beispiel (Forts.):

- Vier Personen werden nach ihrem Alter gefragt, und wir notieren die Alterswerte $y_1 = 24$, $y_2 = 21$, $y_3 = 30$, $y_4 = 21$. Zur Notation: Y bezeichnet hier das Alter, weil im obigen Beispiel so gewählt; y_1 ist der Wert der ersten Person auf der Variablen „Alter“, y_2 der Wert der zweiten Person etc.. Die Anzahl der Personen oder Objekte in einer Stichprobe (der so genannte *Stichprobenumfang*) wird immer mit n abgekürzt. In diesem Beispiel ist $n = 4$.

Abhängige und unabhängige Variablen

Eine *unabhängige Variable* ist ein Merkmal, dessen Auswirkungen auf andere Merkmale (nämlich auf die abhängigen Variablen) untersucht werden sollen. Eine unabhängige Variable gehört zum „Wenn“-Teil bzw. zum „Je“-Teil einer Hypothese.

In *abhängigen Variablen* zeigen sich die Wirkungen der unabhängigen Variablen. Eine abhängige Variable gehört zum „Dann“-Teil bzw. zum „Desto“-Teil einer Hypothese.

Welche Variablen wir in einer Untersuchung als unabhängig und welche wir als abhängig betrachten, hängt von unserer Fragestellung bzw. von unserem Erkenntnisinteresse ab.

Zudem gibt es *Störvariablen*, die zwar in der Hypothese und im Forschungsplan nicht als unabhängige Variablen vorkommen, aber möglicherweise die abhängigen Variablen beeinflussen. Es gibt die Strategien, Störvariablen konstant zu halten oder zu kontrollieren und später statistisch „herauszurechnen“.

Diese Entscheidungen – und dazu gehört auch, wie wir Variablen behandeln, die weder zu den abhängigen noch zu den unabhängigen Variablen zählen – münden in das *Forschungsdesign* oder *Untersuchungsdesign* (*research design*), die Art, wie eine empirische Untersuchung angelegt ist. Das Forschungsdesign kann experimentell oder quasi-experimentell sein, und es kann sich um eine Labor- oder eine Felduntersuchung handeln. Zudem kann man dieselben Untersuchungs„objekte“ einmalig untersuchen oder mehrfach in zeitlichem Abstand, z.B. vor und nach einer Intervention (etwa einer Kampagne).

In einem *Experiment* werden die Untersuchungsbedingungen – und damit die Untersuchungsgruppen, die unter den verschiedenen Bedingungen untersucht werden – durch die VersuchsleiterIn gezielt hergestellt. Den Untersuchungsbedingungen entsprechen verschiedene Ausprägungen der unabhängigen Variablen. Die untersuchten Personen werden den Untersuchungsgruppen *per Zufall* zugeordnet (*Randomisierung*).

Quasi-experimentelle Untersuchungen dagegen arbeiten mit *natürlichen Gruppen*. Personen werden anhand vorher existierender Merkmale Gruppen zugeordnet, die den Ausprägungen der unabhängigen Variablen entsprechen.

Die abhängigen Variablen werden in beiden Fällen – ob Experiment oder Quasi-Experiment – nur erfasst.

Eine *Felduntersuchung* findet im natürlichen Umfeld statt; eine *Laboruntersuchung* unter genau kontrollierten Bedingungen.

Operationalisierung und Messen

Im Alltag werden Merkmalsausprägungen nach persönlichem Eindruck eingeschätzt und verbal beschrieben, z.B. „sehr umweltbewusst“, „wenig tolerant“. In der empirischen Forschung gibt es dafür bestimmte *Methoden der Datenerhebung*: Befragung, Beobachtung und Inhaltsanalyse (die drei klassischen Arten der Datenerhebung; es gibt hier jeweils Varianten für quantitative und für qualitative Forschung).

Dabei werden *Datenerhebungsinstrumente* verwendet, mit deren Hilfe Merkmale erfasst und – im Falle quantitativer Forschung – durch Zahlenwerte codiert werden. Beispiele für Erhebungsinstrumente sind: Fragebogen, Interview, Beobachtungsschema oder physiologische Messung.

Wie die Merkmalsausprägungen genau festgestellt werden, wird durch die *Operationalisierung* festgelegt, eine Anweisung für „Forschungsoperationen“. Die Operationalisierung umfasst die Wahl des Datenerhebungsinstruments und die Festlegung der Zahlencodierung (mehr zur Zahlencodierung s.u.: Skalenniveaus). Die Arbeitszufriedenheit könnte etwa trivial operationalisiert werden durch die Frage „Wie zufrieden sind Sie mit Ihrer Arbeit?“ plus Rating-Skala (Likert-Skalierung): nicht/wenig/mittelmäßig/ziemlich/sehr. Die Schicht könnte durch Fragen nach Bildung, Einkommen und beruflicher Position operationalisiert werden.

Wenn den Merkmalsausprägungen nach bestimmten Regeln Zahlen zugeordnet werden (*Quantifizierung*, z.B. 1/2/3/4/5 bei einer Rating-Skala), handelt es sich um eine *Messung*. *Messen* ist das Zuordnen von Zahlen zu Personen oder Objekten in Abhängigkeit von deren Merkmalsausprägungen. Dabei muss die Zuordnung strukturerhaltend sein („homomorphe Abbildung“), d.h. die Relationen zwischen den Messwerten müssen die Relationen der Objekte untereinander widerspiegeln.

Beispiel: In einer Statistik-Klausur werden den TeilnehmerInnen Punktwerte so zugeordnet, dass die Relation „besser in Statistik“ zwischen zwei beliebigen TeilnehmerInnen repräsentiert wird durch die Relation „größer“ der zugeordneten Punktwerte.

Quantitative Daten sind die Ergebnisse von Messungen, also die (numerischen) Messwerte oder Zahlen, die den Individuen oder Objekten zugeordnet werden. Beispiel „Geschlecht“ von S. 1: Jede Frau einer Stichprobe bekommt eine 1 zugeordnet, jeder Mann eine 2.

Beim Messen geht es nicht darum, Menschen auf Zahlen zu reduzieren, sondern darum, Merkmalsausprägungen systematischer zu erfassen und zu beschreiben als im Alltag. Ob Messen sinnvoll ist, hängt von der Fragestellung ab und von dem Kontext, in dem ein Messinstrument eingesetzt wird.

Skalenniveaus

Das Skalenniveau einer Variablen gibt an, wie viel und welche Information über Relationen zwischen den Personen/Objekten wir den Messwerten entnehmen können bzw. dürfen. Je mehr Informationen es sind (je informativer also die Messwerte sind), desto höher ist das *Skalenniveau* der Variablen (synonym: *Skalendignität*, *Messniveau*, engl. *measurement scale*). Man unterscheidet folgende Skalenniveaus:

- Die *Nominalskala* drückt eine Verschiedenartigkeit aus. Bei der Messung erhalten Objekte mit *gleicher* Merkmalsausprägung *gleiche* Zahlen und Objekte mit *verschiedener* Merkmalsausprägung *verschiedene* Zahlen.
Sonderfall der Nominalskala: Ein nominalskaliertes Merkmal, das nur zwei mögliche Merkmalsausprägungen hat, heißt *dichotomes* oder *binäres* Merkmal.
- Die *Ordinalskala* (= Rangskala) drückt eine Rangordnung oder Reihenfolge aus. Messung: Von jeweils zwei Objekten erhält das Objekt mit der *größeren* Merkmalsausprägung die *größere* Zahl.
- Die *Intervallskala* drückt quantifizierbare Unterschiede oder Abstände aus. Die durch die Messung zugeordneten Zahlen sollen über die Rangordnung hinaus auch die *Größendifferenzen* der Ausprägungen der gemessenen Eigenschaft abbilden. Skala mit gleich großen Abschnitten, ohne natürlichen Nullpunkt.
- *Verhältnisskala*: Skala mit gleich großen Abschnitten, mit natürlichem Nullpunkt. Bei der Messung werden die Zahlen so zugeordnet, dass das Verhältnis zwischen zwei Zahlen dem Verhältnis der Merkmalsausprägungen entspricht, d.h. auch *Größenverhältnisse* der Ausprägungen der gemessenen Eigenschaft werden durch die Zahlen abgebildet.

Intervall- und Verhältnisskala werden zusammenfassend als *metrische Skala* oder *Kardinalskala* bezeichnet. Wir sprechen dann von *metrischen Daten*.

In der Literatur werden manchmal noch *qualitative* und *quantitative* Merkmale unterschieden. Unter *qualitativen* oder *kategorialen* Merkmalen versteht man dann nominalskalierte (oft auch ordinalskalierte) Merkmale, während metrische Merkmale dann als *quantitative* Merkmale bezeichnet werden. Vom Begriffspaar „qualitative vs. quantitative Merkmale“ rate ich eher ab, weil es zu Verwirrung führen kann.¹

Die Nominalskala hat das niedrigste Skalenniveau, die Verhältnisskala das höchste. Je höher das Skalenniveau, desto höher ist der Informationsgehalt der jeweiligen Daten und

¹ Eine Gleichsetzung von *Nominaldaten* mit *qualitativen Daten* ist nur bedingt richtig:

Qualitative Daten sind nicht-numerische Daten, v.a. verbales Material (z.B. Interviews, Briefe, Artikel, Beobachtungsprotokolle), aber visuelles Material wie auch Bilder, Filme etc.. Verbale (also qualitative) Daten können durch eine *quantitative* Inhaltsanalyse in Nominaldaten überführt werden (vgl. Bortz & Döring, 1995). Im Allgemeinen werden verbale Daten aber interpretativ ausgewertet (*qualitative* Forschung).

Quantitative Daten sind numerische Daten (Zahlen), die statistisch analysiert werden können. Nominaldaten sind Häufigkeitsdaten, denn bei der Messung werden die Objekte verschiedenen Merkmalskategorien zugeordnet, deren Häufigkeiten dann statistisch weiterverarbeitet werden können; insofern kann man Nominaldaten als quantitative Daten auffassen.

desto mehr Relationen, Rechenoperationen und statistische Kennwerte (Kennwerte: s. S. 16f.) kann man auf die Daten anwenden:

Skalenniveau	sinnvolle Relationen und Operationen			
	Gleichheit/ Verschiedenheit; Auszählen	größer/kleiner; Ordnen, Rang- reihe bilden	Differenzen/ Abstände berechnen	Quotienten (Verhältnis- zahlen) bilden
nominal	ja	nein	nein	nein
ordinal	ja	ja	nein	nein
intervall	ja	ja	ja	nein
verhältnis	ja	ja	ja	ja

Man kann erhobene Daten jederzeit auf ein niedrigeres Skalenniveau transformieren; eine nachträgliche Transformation auf ein höheres Niveau ist aber nicht möglich. *Vom Skalenniveau des Datenmaterials hängt es ab, welche statistischen Verfahren wir anwenden können.*

Diskrete und stetige Variablen

Diskret (discrete) nennt man eine Variable, die als mögliche Werte nur bestimmte Abstufungen hat, z.B. nur ganze Zahlen. Zur Benennung ihrer Ausprägungen genügt die Menge der natürlichen Zahlen. (Exakte Definition: Eine diskrete Variable ist eine Variable, die nur endlich viele oder abzählbar unendlich viele Werte annehmen kann.)

Stetig (continuous) nennt man eine Variable, bei der zwischen zwei Werten auch jeder Zwischenwert möglich ist, sei das Intervall auch noch so klein. Zur Benennung ihrer Ausprägungen ist die Menge der rationalen Zahlen notwendig.

In der Praxis resultiert jedoch jede empirische Messung in diskreten Messwerten, auch wenn im Einzelfall sehr viele unterschiedliche Messwerte möglich sind. Beispiel: Das Alter – obwohl als Zeit eigentlich eine stetige Variable – wird i.A. in Jahren gemessen. Jahre sind eine diskrete Zeiteinheit. Daran würde auch eine feinere Messung, z.B. in Sekunden, nichts ändern.

Dennoch ist die Unterscheidung „diskret vs. stetig“ auch für die angewandte Statistik relevant. Die mathematische Statistik hat nämlich für diskrete und stetige Variablen („Zufallsvariablen“, s. Kap. zur Wahrscheinlichkeitstheorie) theoretische Verteilungsmodelle entwickelt, die wir für die Inferenzstatistik benötigen. Für das Alter in Jahren z.B. benutzt man theoretische Verteilungsmodelle für stetige Variablen, weil ein kontinuierliches Merkmal zugrunde liegt.

Konstrukte und Indikatoren

Merkmale, die unmittelbar wahrnehmbar (z.B. biologisches Geschlecht), durch einfache Fragen direkt erfassbar (z.B. Einkommen) oder physiologisch messbar sind (z.B. Blutdruck), bezeichnet man als *manifeste Merkmale*. Merkmale dagegen, die nicht direkt wahrnehmbar/erfassbar/messbar sind, bezeichnet man als *latente Merkmale* oder theoretische *Konstrukte*: gedankliche Konstruktionen, die erst vor dem Hintergrund theoretischer Überlegungen eine Bedeutung erhalten, z.B. „Toleranz“, „Intelligenz“, bestimmte „Einstellungen“, „Image“ oder „Sozialschicht“.

In der empirischen Forschung haben wir es meist mit latenten Merkmalen zu tun. Ihre Operationalisierung ist aufwändiger als bei manifesten Merkmalen. Latente Merkmale oder theoretische Konstrukte versucht man durch *Indikatoren* greifbar zu machen. Das sind direkt wahrnehmbare Phänomene, die man als manifeste Variablen misst und mit deren Hilfe man auf das Vorhandensein eines Konstruktes schließt. Indikatoren für das Konstrukt „Feindschaft“ könnten etwa sein: (1) Anna gibt Berta Ohrfeigen, (2) Anna sagt: „Ich hasse Berta“, (3) Anna beleidigt Berta bei einem Wortwechsel, (4) Anna lehnt die von Berta erbetene Hilfeleistung ab (Beispiel von Prim & Tilmann, 1989).

Achtung: „A lehnt die von B erbetene Hilfeleistung ab“ muss kein Zeichen von Feindschaft sein, sondern kann gerade eines von Freundschaft sein, die der Partnerin zu größerer Selbstständigkeit verhelfen will. Oder die Ohrfeige kann an einer nervlichen Überlastung liegen. Hier geht es um die *Validität*, das ist das Ausmaß, in dem Indikatoren tatsächlich das messen, was sie messen wollen. Die Beobachtung „A lehnt die von B erbetene Hilfeleistung ab“ ist für sich genommen kein valider Indikator für das Konstrukt „Feindschaft“.

Latente Merkmale sind meistens komplex, so dass wir mehrere Indikatoren brauchen. „Offenheit gegenüber Ausländern“ wird z.B. nicht nur aus der Anzahl ausländischer Freunde erschlossen, sondern auch aus anderen Indikatoren, die dann durch entsprechende Fragen und/oder Verhaltensbeobachtungen operationalisiert werden. Verschiedene Einzelindikatoren müssen dann wieder zu einem Gesamtwert (Index) zusammengefasst werden.

Aufgaben

1. Was ist ein Stichprobenumfang?
2. Was sind abhängige Variablen, unabhängige Variablen, Störvariablen?
Konstruiere ein Beispiel.
3. Worin besteht der Unterschied zwischen einer experimentellen und einer quasi-experimentellen Untersuchung?
4. Was ist eine Messung?
5. Was hat Messen mit den Skalenniveaus zu tun?
6. Überlege dir zu jedem Skalenniveau ein Beispiel.

7. Welche Operationen (+, −, •, /) und Relationen (=, ≠, <, >) sind mit den Daten der verschiedenen Skalenniveaus möglich?
8. Was sind diskrete, was sind stetige Variablen? Bitte je ein Beispiel.
9. Was sind Indikatoren, und wozu braucht man sie?

Deskriptive Statistik

Die *deskriptive Statistik* (*descriptive statistics*) dient der Beschreibung quantitativ-empirischer Daten. Es geht darum, wie wir Daten und ihnen zugrunde liegende Muster sinnvoll darstellen und zusammenfassen können. Darstellungsmittel der deskriptiven Statistik sind:

- Tabellen (z.B. Häufigkeitstabellen, oft kategorisiert)
- Grafiken (z.B. Balkendiagramm, Kreisdiagramm)
- statistische Kennwerte

Die durch deskriptive Statistik gewonnenen Aussagen gelten nur für die untersuchte Stichprobe. „Verallgemeinernde Interpretationen der deskriptiven statistischen Analyse, die über das erhobene Material hinausgehen, sind spekulativ“ (Bortz, 1999, S. 1).

Die *Inferenzstatistik* (*schließende Statistik, analytische Statistik; inferential statistics*) ist der Gegenbegriff zur deskriptiven Statistik. Bei der Inferenzstatistik geht es darum, wie und unter welchen Voraussetzungen man von den Ergebnissen einer Stichprobe auf die „Grundgesamtheit“ schließen oder hochrechnen kann:

- Auf der Basis von Stichprobenergebnissen kann man auf die Verteilung der Merkmale von Personen/Objekten in der Grundgesamtheit schließen.
- Auf der Basis von Stichprobenergebnissen kann man Hypothesen überprüfen, die sich auf die Grundgesamtheit beziehen.

Der Schluss von der Stichprobe auf die Grundgesamtheit ist immer mit einem Fehlerrisiko behaftet. Dessen Höhe lässt sich mit Hilfe der Wahrscheinlichkeitsrechnung bestimmen.

Häufigkeitsverteilung

Merkmalsausprägungen bzw. Messwerte treten in einer Stichprobe meistens mehrfach auf (Häufung). Um sich eine Übersicht über das Datenmaterial zu verschaffen, zählt man, wie oft jeder Messwert auftritt. Man unterscheidet dabei *absolute Häufigkeiten* (Beispiel von S. 1: Das Alter 21 tritt zweimal auf) und *relative Häufigkeiten* (= Anteile oder %-Werte; Beispiel: 50% unserer Ministichprobe sind 21 Jahre alt).

Das Ergebnis einer solchen Auszählung ist eine *Häufigkeitsverteilung (frequency distribution)*, d.h. die Zuordnung von Häufigkeiten zu den verschiedenen Messwerten der Stichprobe bzw. zu den verschiedenen Merkmalsausprägungen. Das kann mit Hilfe einer Strichliste geschehen, die als Basis für eine Häufigkeitstabelle dient. Absolute und relative Häufigkeiten können grafisch dargestellt werden, z.B. in Balkendiagrammen, Liniendiagrammen, Histogrammen, Polygonzügen oder Kreisdiagrammen (s.u.).

Kategorisierung (Klassifizierung)

Wenn eine Variable sehr viele mögliche Merkmalsausprägungen hat bzw. wenn sie kontinuierlich ist, wäre eine solche Häufigkeitsverteilung unübersichtlich und wenig aussagekräftig. Deshalb fasst man in diesen Fällen ähnliche bzw. benachbarte Merkmalsausprägungen zu Kategorien (Intervallen, Klassen) zusammen (*classified variable*). Erst nach der Kategorisierung ordnet man die Häufigkeiten den (wenigen) Kategorien zu.

Da eine Kategorisierung immer einen Informationsverlust bedeutet, ist es meist besser, die Daten nicht bereits kategorisiert zu erheben (eine Ausnahme sind sensible Daten), sondern die feiner aufgelösten, realen Daten selbst, und sie erst später je nach Bedarf zu kategorisieren. Der Grund dafür ist, dass wir bei nicht-kategorisierten Werten mehr Möglichkeiten der Datenauswertung haben.

Beispiel für eine Kategorisierung: Nehmen wir an, 130 Personen geben in einer Befragung ihr Haushalts-Nettoeinkommen an ($n = 130$). Person Nr. 1: 1.450 €, Person Nr. 2: 3.300 €, ..., Person Nr. 130: 2.100 €. Bevor es EDV gab, hat man die Ergebnisse einfach nacheinander in eine so genannte *Urliste* geschrieben: 1.450, 3.300, ..., 2.100 (hier bitte 130 Zahlen denken). Man kann sich dann sinnvolle Kategoriengrenzen überlegen, die Urliste durchgehen, dabei jeden Wert mithilfe einer Strichliste einer Kategorie zuordnen und erhält so die absoluten und relativen Häufigkeiten:

Häufigkeitstabelle: Haushalts-Nettoeinkommen in Euro (fiktives Beispiel)

Kategorie k	Strichliste	abs. Häufigkeit (absolute Werte)	rel. Häufigkeit (%-Werte)
bis unter 500		1	0,8
500 bis unter 1.000		12	9,2
1.000 bis unter 1.500		20	15,4
1.500 bis unter 2.000		24	18,5
2.000 bis unter 2.500		22	16,9
2.500 bis unter 3.000		18	13,8
3.000 bis unter 3.500		11	8,5
3.500 bis unter 4.000		9	6,9
4.000 und mehr		13	10,0
		$n = 130$	100

Heute gibt man die Daten in eine Datei ein, definiert die Kategorien und lässt ein Statistik-Programm (z.B. SPSS = Statistical Package for the Social Sciences) die Häufigkeitstabelle erstellen.

Wie viele Kategorien bildet man, bzw. welche Kategoriengrenzen definiert man? Wenn man viele Messwerte/Daten hat, kann man mehr Kategorien bilden als bei einem kleinen Datenset. Faustregeln für die Anzahl der zu bildenden Kategorien:

- Bezeichnen wir mit m die Anzahl der Kategorien und mit n die Kollektivgröße, dann ist eine Faustregel für eine optimale Anzahl von Kategorien: $2^m < n$.
- Eine genauere Formel lautet: $m \approx 1 + 3,32 \cdot \lg n$ (Bortz, 1999, S. 30), wobei \lg der Logarithmus zur Basis 10 ist.
Unser Beispiel: $n = 130$ Personen. Dann ist $m \approx 1 + 3,32 \cdot \lg 130 = 8,018 \approx 8$.
Man sollte nach dieser Formel also 8 Kategorien bilden.
- Oft ändert man die so erhaltene Anzahl der Kategorien etwas, um eingängige und sinnvolle Kategoriengrenzen zu haben. Im Beispiel mit dem Nettoeinkommen folgen die Kategoriengrenzen in 500er-Schritten aufeinander, wir haben dann 9 statt 8 Kategorien.

Anforderungen an Kategorien:

- Die Kategorien müssen eindeutig definiert sein und sich gegenseitig ausschließen (disjunkt sein), d.h. jeder Messwert kann nur in eine Kategorie gesteckt werden.
- Die Kategorien müssen erschöpfend sein, d.h. jeder Messwert muss einer Kategorie zuordenbar sein.
- Man sollte gleich breite Intervalle bilden, mit runden Zahlen als Intervallmitten, oder mit runden Zahlen als Intervallgrenzen (z.B. Einkommen 1.500 bis 2.000 €)
- Man sollte zu einer Seite offene Klassen vermeiden. Wenn aber Minimum bzw. Maximum unbekannt sind, muss die unterste Klasse nach unten bzw. die oberste Klasse nach oben offen bleiben (z.B. oberste Kategorie der Variablen Nettoeinkommen: „4.000 und mehr“).

Grafische Darstellungen von Häufigkeiten

Im Folgenden geht es um Darstellungen von absoluten oder relativen Häufigkeiten für *einzelne* Variablen (genauer: um die Häufigkeiten der *Ausprägungen* einzelner Variablen).

Die grafische Darstellung von Zusammenhängen zwischen *zwei oder mehr* Variablen behandeln wir bei der Analyse von Zusammenhängen (Kontingenz, Korrelation; dafür gibt es z.B. gestapelte Balkendiagramme für nominale und Scatterplots für metrische Daten).

Balken- und Säulendiagramm (bar chart)

Balken- und Säulendiagramme sind gut zur Darstellung der Häufigkeiten von Ausprägungen **nominaler** und **ordinaler** Merkmale geeignet. Bei metrischen Merkmalen können wir es verwenden, wenn es nur sehr wenige Ausprägungen gibt und keine Lücken in den Werten auftreten.²

In Abb. 1 repräsentieren die Balken die Häufigkeiten der Ausprägungen oder Kategorien des Merkmals „Werthaltungen“. Bei waagerechten Balken, wie hier, spricht man von einem *Balkendiagramm*, bei senkrechten von einem *Säulendiagramm* (z.B. bei den Antworten auf die Sonntagsfrage).

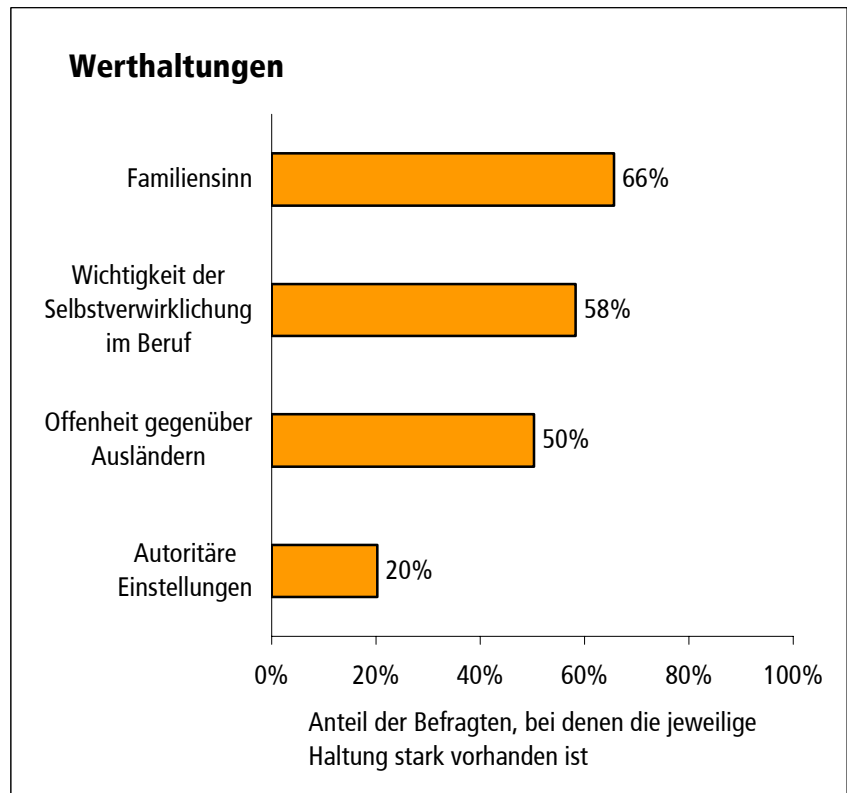


Abbildung 1: Häufigkeiten verschiedener Werthaltungen. Quelle: Befragung von 163 Personen im Seminar Fragebogenkonstruktion (P. Hautz, UdK Berlin, Januar 2001). Die Erfassung jeder Werthaltung geschah mit einer Batterie von Statements.

Die Längen der Balken sind proportional zu den Häufigkeiten. Die Balken sollten vom Nullpunkt ausgehen, da sonst ein verzerrter Eindruck entsteht. Es werden nur die tatsächlich vorkommenden Ausprägungen oder Kategorien dargestellt, keine „unbesetzten“ Ausprägungen (im Ggs. zum Histogramm). Zwischen den Balken gibt es Zwischenräume: Die Balken berühren einander nicht (im Ggs. zum Histogramm). Bei nominalen Merkmalen ist die Anordnung der Balken willkürlich, die Reihenfolge der Balken darf vertauscht werden (im Ggs. zum Histogramm), z.B. ordnet man die Balken manchmal nach der Häufigkeit.

Wie das obige Beispiel (Abb. 1) zeigt, kann man mit einem Balkendiagramm auch *Mehrfachantworten* visualisieren. Dabei addieren sich die Häufigkeiten der einzelnen Antwortmöglichkeiten zu *mehr* als 100% (das geht ja beim Kreisdiagramm nicht).

² Wenn eine Variable dagegen sehr viele Ausprägungen hat, ist eine Klassifizierung der Variablenwerte empfehlenswert, damit Häufungen in bestimmten Wertebereichen überhaupt erkennbar sind. Dann ist ein Histogramm (s.u.) die passende grafische Darstellung.

Mit Säulendiagrammen lassen sich auch Entwicklungen im Zeitverlauf darstellen. Die Zeitabschnitte werden dann auf der x-Achse abgetragen, die Säulen stehen senkrecht (s. Abb. 2).

Säulendiagramme sind zur Visualisierung von Zeitverläufen nur dann geeignet, wenn es um *wenige* Zeitabschnitte geht. Für *längere* Zeitreihen benutzt man Liniendiagramme (nächste Seite).

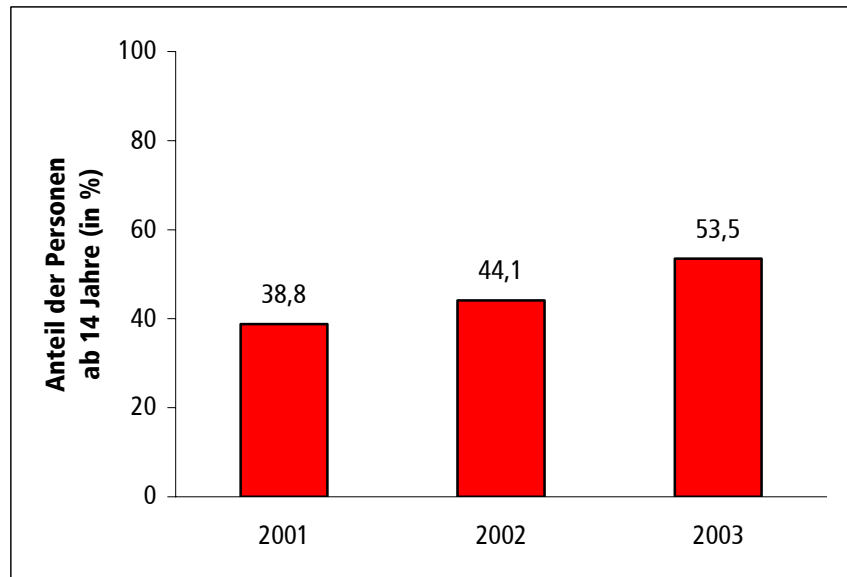


Abbildung 2: Entwicklung der Internet-Nutzung in Deutschland in den letzten drei Jahren (ARD/ZDF-Online-Studien 1998-2003)

Abb. 3 zeigt ein *gruppiertes Balkendiagramm*, in dem das Merkmal Geschlecht hinzugefügt ist.

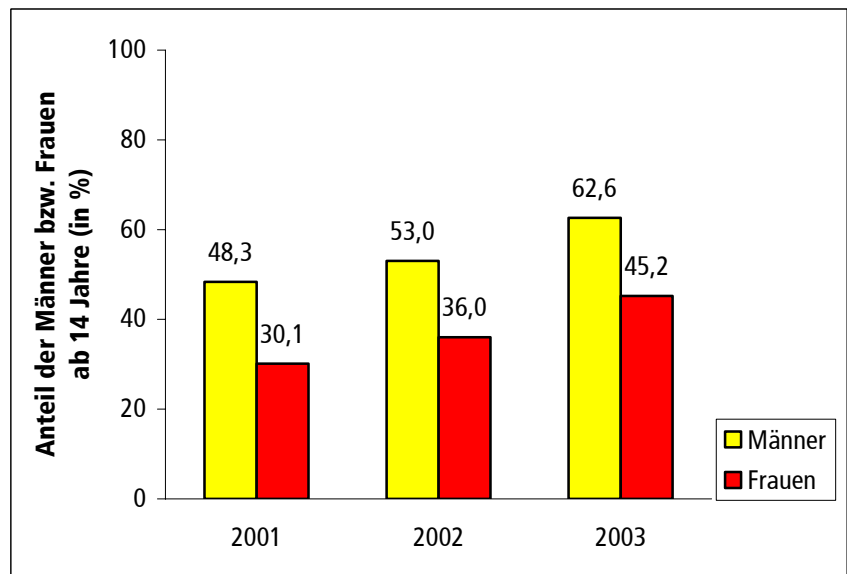


Abbildung 3: Entwicklung der Internet-Nutzung von Männern und Frauen in Deutschland in den letzten drei Jahren (ARD/ZDF-Online-Studien 1998-2003)

Später wird als weitere Variante noch das *gestapelte Balkendiagramm* behandelt. Es wird häufig verwendet, wenn man die Merkmalsausprägungen, deren Häufigkeiten man dargestellt, durch ein *zweites* Merkmal in Gruppen einteilt (s. Kap. Analyse von Häufigkeiten, Abschnitt Kreuztabellen).

Liniendiagramm (line chart, line plot)

Das Liniendiagramm kann Häufigkeiten von Ausprägungen **metrischer** Merkmale darstellen. Es wird sehr häufig zur Visualisierung von Veränderungen im Zeitverlauf benutzt (der Zeitverlauf ist ja metrisch, wenn die Zeitpunkte auf der x-Achse gleichabständig sind). Für die Wahrnehmung von Verläufen ist das Liniendiagramm besser geeignet als das Säulendiagramm.

Die Häufigkeiten werden durch miteinander verbundene Punkte repräsentiert. Die y-Achse sollte am Nullpunkt beginnen, da sonst ein verzerrter Eindruck entsteht.

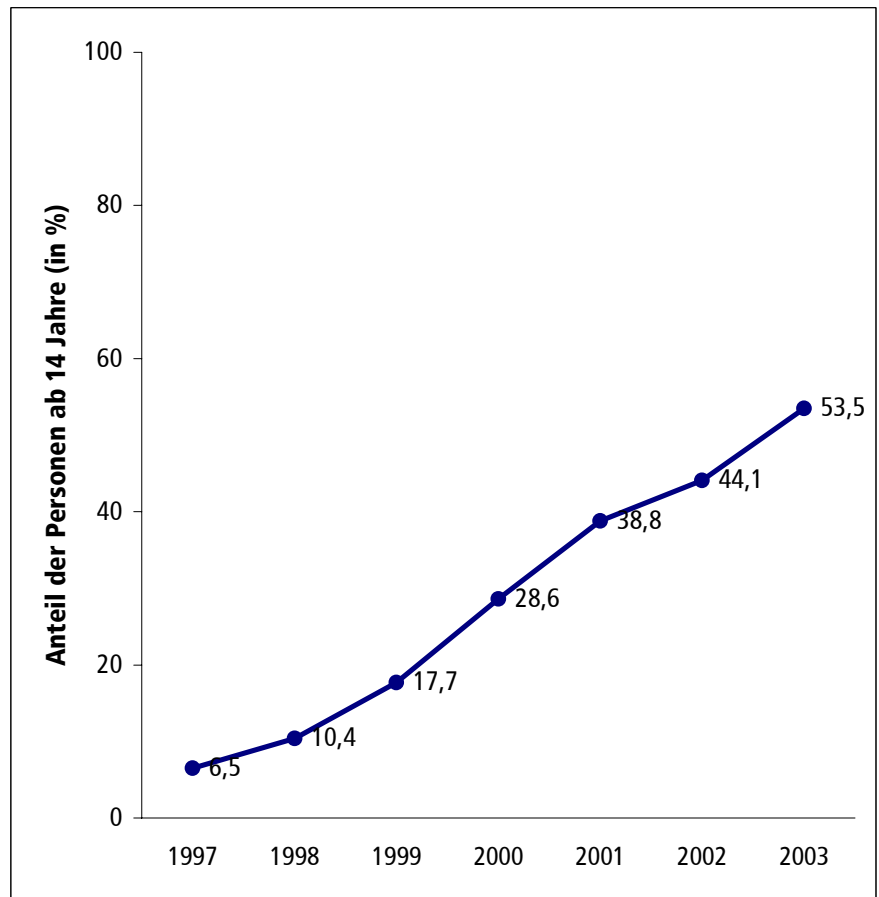


Abbildung 4: Entwicklung der Internet-Nutzung in Deutschland 1997 bis 2003 (ARD-Online-Studie 1997, ARD/ZDF-Online-Studien 1998-2003)

Anmerkungen zum Liniendiagramm:

- Natürlich kann man mit einem Liniendiagramm nicht nur Häufigkeiten darstellen, sondern auch andere Größen, wie z.B. Aktienkurse.
- Eine gute Möglichkeit, mehrere gleichartige Datensätze visuell miteinander zu vergleichen, ist ein Liniendiagramm mit mehreren Linienzügen.
- Bei mehreren Datensätzen bzw. mehreren Linien „darf“ man ein solches Diagramm auch dann verwenden, wenn das Merkmal auf der x-Achse nominal oder ordinal ist, denn die in diesem Fall sonst verwendeten Säulen wären dann unübersichtlich. Achtung: Wenn das Merkmal auf der x-Achse nominalskaliert ist, kann durch Vertauschen der Ausprägungen auf der x-Achse ein ganz anderer visueller Eindruck entstehen, was Irreführungen ermöglicht.

Histogramm (histogram)

Die Häufigkeiten von Kategorien **metrischer** Merkmale werden in Histogrammen oder in Polygonzügen (s.u.) dargestellt. Nach Bortz (1999) sollte das Histogramm nur für diskrete Variablen verwendet werden; es wird aber oft auch für stetige Variablen verwendet. Histogramme sehen zwar ähnlich aus wie Balkendiagramme, sind aber keine. Die x-Achse eines Histogramms ist im Ggs. zum Balkendiagramm eine durchgängige, metrische Skala.

Der Bereich zwischen dem kleinsten und größten Wert des Merkmals (x-Achse) wird in gleich große Intervalle aufgeteilt; meist bildet man 5 bis 15 Intervalle/Klassen. Zur Berechnung einer günstigen Intervall- oder Klassenanzahl m gibt es die oben genannten Faustformeln (s. S. 10).

Die Häufigkeiten des Auftretens von Werten in den Teilintervallen werden durch säulenartige Rechtecke dargestellt. Die Fläche eines Rechtecks ist proportional zur Häufigkeit des Auftretens der betreffenden Klasse. Die Rechtecke sollten vom Nullpunkt ausgehen, da sonst ein verzerrter Eindruck entsteht. Keine Leerräume zwischen den Rechtecken lassen (im Ggs. zum Balkendiagramm). Auf der x-Achse treten – im Ggs. zum Balkendiagramm – *alle* Klassen auf. Für die „unbesetzten“ Klassen (d.h. die mit einer Häufigkeit von Null) schrumpft die Höhe des Rechtecks auf die Nulllinie. Die Reihenfolge der Rechtecke bzw. Teilflächen wird durch die Ordnung der Skala bestimmt und darf *nicht* vertauscht werden (im Ggs. zum Balkendiagramm, wenn man nominale Variablen hat).

Vor jeder Analyse metrischer Variablen sollte man sich die Verteilung der Werte mit Hilfe eines Histogramms veranschaulichen.

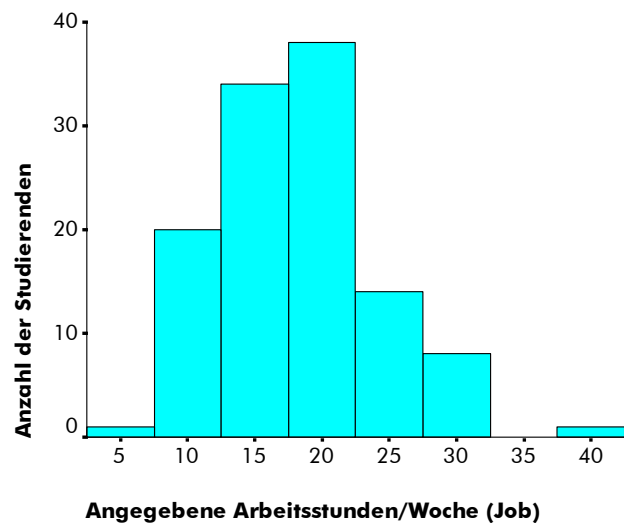


Abbildung 5: Erwerbstätigkeit neben dem Studium. Antworten von 116 GWK-Studierenden auf die Frage „Wenn Sie regelmäßig arbeiten, wie viele Stunden die Woche?“ (GWK-Evaluationsbericht 2002, an der Befragung nahmen 247 Studierende teil). Auf der x-Achse sind die Intervallmitten angegeben.

Polygonzug (frequency polygon)

Für Häufigkeiten von Kategorien **metrischer** Variablen, die stetig oder künstlich diskret sind. Ziel ist die Darstellung der Häufigkeitsverteilung einer Variablen (näherungsweise) als kontinuierliche Funktion.

Wir gehen von einem Histogramm aus (es erscheint in der Abb. rechts zusätzlich zum Polygonzug) und nähern die kontinuierliche Funktion durch einen Linienzug an, der die Mittelpunkte der oberen Säulenkanten verbindet. Außen endet der Linienzug jeweils eine halbe Intervallbreite weiter auf der x-Achse.

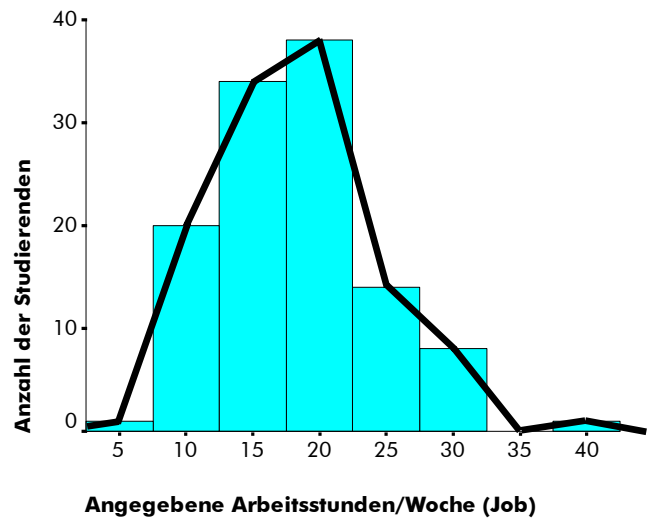


Abbildung 6: Histogramm und Polygonzug

Der Polygonzug ist auch ein Liniendiagramm, nur dass die Merkmalsausprägungen hier keine Zeitpunkte, sondern gleichabständige Kategorienmitten sind.

Kreis- oder Tortendiagramm (pie chart)

Das Kreisdiagramm eignet sich für %-Anteile von Ausprägungen **nominaler** Merkmale (es ist zwar auch für höher skalierte kategoriale Daten möglich, aber nicht zu empfehlen, weil im Kreisdiagramm die ordinale Reihenfolge nicht wiedergegeben wird; im Gegensatz dazu kann das Balkendiagramm von links nach rechts ordinal ansteigende Werte repräsentieren).

Der Flächenanteil eines Kreissektors entspricht der relativen Häufigkeit (%-Anteil). Die einzelnen Sektoren dürfen nicht zu klein sein, weil es sonst unübersichtlich wird.

Die Anteile müssen sich zu 100% addieren (im Ggs. zum Balkendiagramm).

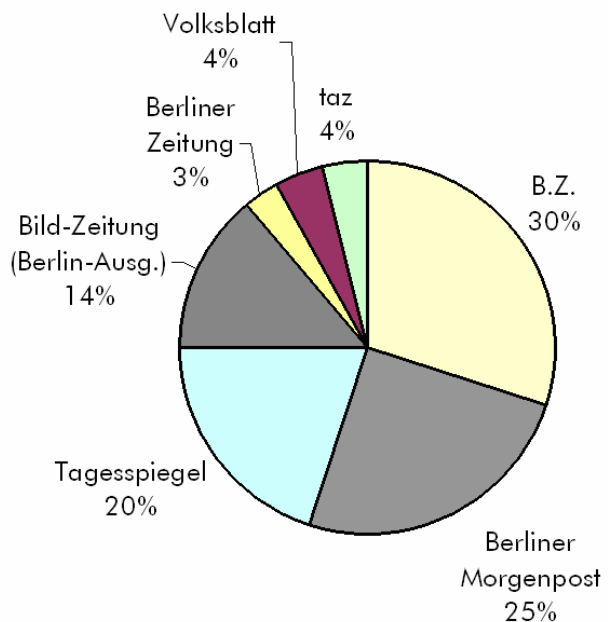


Abbildung 7: Marktanteile West-Berliner Zeitungen im Jahr 1993 (nach Held & Simeon, 1994)

Grafische Darstellungen anderer Kennwerte

Nicht nur Häufigkeiten, sondern auch andere Kennwerte werden in Grafiken dargestellt, z.B. Indizes im Zeitverlauf oder Mittelwerte in verschiedenen Gruppen.

Hierzu eignen sich besonders

- Balken- und Säulendiagramm
- bei Zeitverläufen: Liniendiagramm
- bei zwei metrischen Variablen: Scatterplot (s. Kap. Lineare Zusammenhänge)
- bei drei metrischen Variablen: Blasendiagramm (*bubble chart*), bei dem die Ausprägungen der dritten Variable durch die Größe der Blasen dargestellt wird.

Statistische Kennwerte

Statistische Kennwerte sind Maße, die wichtige Eigenschaften einer Merkmalsverteilung komprimiert ausdrücken. Man unterscheidet Maße der zentralen Tendenz und Dispersionsmaße.

Maße der zentralen Tendenz (Lagemaße)

Die zentrale Tendenz ist die Position besonders häufiger oder typischer Werte. Ludwig-Mayerhofer (1998–2004) spricht vom „Schwerpunkt“ eines Datenbündels. Maße der zentralen Tendenz (*measures of location, measures of central tendency*) repräsentieren alle Messwerte zusammenfassend. Die wichtigsten Maße der zentralen Tendenz sind:

- Modalwert = Modus (*mode*), Notation: Mo
Der Modalwert ist der in einer Stichprobe am häufigsten vorkommende Wert einer Variablen.
- Medianwert = Median = Zentralwert (*median*), Notation: Md
Der Median ist der Wert einer Variablen, unterhalb und oberhalb dessen jeweils die Hälfte der Werte aller anderen Fälle liegt. (Exakte Formulierung: Mindestens 50% der Werte sind kleiner oder gleich dem Md, und mindestens 50% der Werte sind größer oder gleich dem Md.)
- Mittelwert = arithmetisches Mittel (*mean, arithmetic mean*), Notation: \bar{x} , M, AM
Der Mittelwert einer Variablen ist die Summe der Werte bei allen Fällen, dividiert durch die Anzahl der Fälle („Durchschnitt“).

Dispersionsmaße (Streuungsmaße)

Dispersionsmaße (*measures of dispersion, measures of spread*) geben Auskunft über die Unterschiedlichkeit der Messwerte. Bei einer kleinen Dispersion (Streuung) verteilen sich die meisten Werte eng um einen Wert der zentralen Tendenz, bei einer großen Dispersion weit. Die wichtigsten Dispersionsmaße sind:

- Range = Variationsbreite = Spannweite
Der Range ist der Abstand zwischen dem niedrigsten und dem höchsten gemessenen Wert einer Variablen, also zwischen Minimum und Maximum.
- Varianz (*variance*), Notation: s^2
Die empirische Varianz ist eine Kennzahl für die Dispersion von gemessenen Werten um ihren Mittelpunkt herum. Je stärker die Messwerte der einzelnen Fälle vom Mittelwert abweichen, desto größer ist die Varianz einer Variablen. In die Berechnung der empirischen Varianz gehen die quadrierten Abweichungen der einzelnen Werte von ihrem Mittelwert ein:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad \text{empirische Varianz der Variablen } x$$

Zum Σ -Zeichen, dem so genannten Summenzeichen, siehe den Exkurs auf der nächsten Seite.

Achtung: In der angloamerikanischen Literatur (und danach richtet sich auch Statistik-

Software wie SPSS) wird die Varianz oft durch den Ausdruck $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$ definiert.

Die Formel mit $n - 1$ im Nenner wird benutzt, wenn man aus den Daten einer Stichprobe die Varianz in der Grundgesamtheit schätzen will (s. Kap. Stichprobentheorie). Für

die empirische Varianz einer Stichprobe verwenden wir jedoch $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$.

Bei großen n ergibt sich praktisch kein Unterschied zwischen den beiden Formeln.³

³ Ludwig-Mayerhofer schreibt in seinem *Internet-Lexikon der Methoden der empirischen Sozialforschung* (1998–2004) beim Stichwort Varianz: „Hinsichtlich der sprachlichen Bezeichnung der beiden angegebenen Berechnungsmöglichkeiten für die Varianz hat sich leider eine offensichtlich unaufhebbare Konfusion eingebürgert. Wie man durch eine schnelle Internetrecherche ebenso wie durch einen vergleichenden Blick in eine Reihe von Lehrbüchern feststellen kann, werden die Begriffe "empirische Varianz", "Stichprobenvarianz" und manchmal auch nur "Varianz" alternativ für beide Formeln verwendet. Es kann also vorkommen, dass ein Buch oder eine sonstige Quelle den Begriff "empirische Varianz" für die erste und den Begriff "Stichprobenvarianz" für die zweite Formel verwendet und das nächste Buch genau umgekehrt verfährt! Wieder andere geben nur die eine oder die andere Formel an, ohne auf deren exakte Bedeutung hinzuweisen. Beim Konsultieren statistischer Texte ist also höchste Vorsicht geboten und es ist dringend erforderlich, sich jeweils ein genaues Bild darüber zu verschaffen, was mit "Varianz", "empirischer Varianz" oder "Stichprobenvarianz" gemeint ist.“

- Standardabweichung (*standard deviation, std. dev.*), Notation: s oder SD
Die Standardabweichung ist die Quadratwurzel aus der Varianz. Oft wird nur kurz von *Streuung* gesprochen. Der Begriff Streuung dient also einerseits als Synonym zum Begriff Standardabweichung, andererseits als Oberbegriff für alle Dispersionsmaße (Streuungsmaße). Meist meint man mit Streuung die Standardabweichung.
- AD-Streuung (*average deviation, mean absolute deviation*), Notation: AD, MAD
Die AD-Streuung ist die durchschnittliche Abweichung vom Mittelwert.

Aufgaben

1. Definiere die statistischen Kennwerte Mo, Md, AM, Range, AD, s^2 , s
 - a) in Worten,
 - b) durch die entsprechende Formel, falls es eine gibt.
2. Warum reichen die Maße der zentralen Tendenz zur Charakterisierung einer Häufigkeitsverteilung nicht aus?
3. Welchen Vorteil hat der Median gegenüber dem arithmetischen Mittel?
4. Berechne Mittelwert, Standardabweichung und AD-Streuung folgender (metrischer) Werte: 7 3 5 6 8
5. Welche statistischen Kennwerte kann man für
 - a) nominalskalierte Daten
 - b) ordinalskalierte Daten
 - c) metrische Daten
 sinnvoll berechnen?

Exkurs: Rechnen mit dem Summenzeichen

Ein in der Statistik häufig benutztes Operationszeichen ist das Summenzeichen, ein großes griechisches Sigma (Σ). Es dient dazu, Summen abkürzend zu notieren. Wir schreiben z.B.:

$$x_1 + x_2 + x_3 + x_4 = \sum_{i=1}^4 x_i \quad (\text{lies: „Summe aller } x_i\text{-Werte für } i \text{ von 1 bis 4“})$$

Das i ist der so genannte *Laufindex*. Man kann dafür auch beliebige andere Buchstaben benutzen. Unterhalb des Summenzeichens wird die untere Grenze des Laufindex angegeben (erster Wert, im Beispiel: 1), oberhalb des Summenzeichens die obere Grenze (letzter Wert, im Beispiel: 4).

Leider müssen nicht immer nur einfache Werte einer Variablen aufsummiert werden. Häufig handelt es sich um komplexere mathematische Ausdrücke. Beispiel:

$$\sum_{i=1}^n (x_i^2 + 3a - x_i)$$

Hier kann man das Summenzeichen behandeln, als ob es eine vor der Klammer stehende Konstante wäre, die zur Auflösung der Klammer mit jedem in der Klammer stehenden Wert multipliziert werden muss.

Im Beispiel ergibt sich:

$$\sum_{i=1}^n (x_i^2 + 3a - x_i) = \sum_{i=1}^n x_i^2 + \sum_{i=1}^n 3a - \sum_{i=1}^n x_i$$

Nun sind die einzelnen Summationen leichter durchzuführen.

Achtung: Auch Konstanten (z.B. $3a$) müssen summiert werden. Da sich hier nichts verändert, kann statt $3a + 3a + \dots + 3a$ auch $n \cdot 3a$ geschrieben werden:

$$\sum_{i=1}^n (x_i^2 + 3a - x_i) = \sum_{i=1}^n x_i^2 + n \cdot 3a - \sum_{i=1}^n x_i$$

Aufgaben

Bitte forme folgende Ausdrücke um: Wo kein Summenzeichen ist (wie bei Aufgabe 1), kürze den Ausdruck mit Hilfe eines Summenzeichens ab. Wo bereits ein Summenzeichen steht (wie bei Aufgabe 2), löse den Ausdruck auf.

1. $B_1 + B_2 + B_3 + B_4 =$

2. $c \cdot \sum_{i=1}^n z_i =$

3. $\sum_{i=1}^3 (x_i - b) =$

4. $\sum_{i=1}^4 x_i y_i =$

5. $(x_1 + x_2 + x_3)^2 =$

6. $\sum_{i=1}^n dx_i + a =$

Achtung bei Aufgabe 6: Die Schreibweise $\sum_{i=1}^n dx_i + a$ ist zu interpretieren als $\left(\sum_{i=1}^n dx_i \right) + a$

z-Transformation

Eine Transformation ist eine Umwandlung von Zahlenwerten. Die *z-Transformation* eines Zahlenwertes besteht darin, dass seine Abweichung vom Mittelwert berechnet und dann durch die Standardabweichung dividiert wird.

Formel für die z-Transformation (bitte eintragen):

$$z_i =$$

Mit der z-Transformation werden metrische Zahlen- oder Messwerte *standardisiert* oder *normiert*. Aus den ursprünglichen Werten werden standardisierte Werte, so genannte *z-Werte* (*z-scores*).

Die Verteilung der so entstandenen z-Werte hat immer einen Mittelwert von 0 und eine Standardabweichung von 1. Mit anderen Worten: Durch eine z-Transformation wird jede beliebige Verteilung in eine Verteilung mit $\bar{z} = 0$ und $s_z = 1$ überführt. Der Verteilungstyp ändert sich dabei nicht: Die Form der Verteilung bleibt im Prinzip erhalten, nur wird die Verteilung erstens zum Nullpunkt verschoben und zweitens gestaucht oder gedehnt.

Anwendung:

- Wenn man zwei Messwerte aus zwei verschiedenen Stichproben vergleichen will, geht das nicht unmittelbar, denn wenn man ein Merkmal in zwei verschiedenen Stichproben erhebt, resultieren in den beiden Stichproben i.a. verschiedene Mittelwerte und Standardabweichungen. Um nun einen Messwert aus der einen Stichprobe mit einem Messwert aus der anderen Stichprobe vergleichen zu können, berechnen wir die beiden entsprechenden z-Werte und vergleichen diese.
- Nach demselben Prinzip kann man auch die Messwerte zweier verschiedener Variablen, die in *einer* Stichprobe erhoben wurden, aber unterschiedliche Mittelwerte und Standardabweichungen aufweisen, vergleichbar machen.
- Allgemein: Ein z-Wert drückt aus, wo der ursprüngliche Messwert innerhalb seiner Verteilung liegt.

Mit einer z-Transformation werden wir auch so genannte Normalverteilungen auf einen Standardmaßstab bringen und damit Rechenarbeit sparen (später, s. die Kap. „Wahrscheinlichkeitstheorie“, „Stichprobentheorie“ und „Hypothesen und ihre Überprüfung“). Die z-Transformation wird *auch, aber nicht nur* auf Normalverteilungen angewendet.

(Anmerkung: Von der z-Transformation muss eine andere Transformation, die so genannte Fisher's Z-Transformation, unterschieden werden. Hierauf gehen wir hier nicht ein.)

Aufgabe

Nehmen wir an, wir können mit physiologischen Messungen den Grad der Aktivierung beim Betrachten eines Werbespots registrieren (von 0 in vollkommener Ruhe bis 100 als höchstmögliche Aktivierung).

Eine Testperson sieht nun einen Werbespot für eine Sportbekleidungsmarke und einen Werbespot für einen Mobilfunkanbieter. Die Person gehört bei beiden Produktbereichen zur jeweiligen Zielgruppe. Beim Betrachten des Sportbekleidungssports zeigt unsere Testperson eine Aktivierung von 50, beim Mobilfunkspot eine von 43. Aufgrund von Untersuchungen, die mit vielen Betrachtern der Zielgruppen durchgeführt wurden, kennen wir die durchschnittlichen Aktivierungsgrade

- bei der Sportbekleidungsmarke: $\bar{x}_1 = 44$ (mit einer Streuung von $s_1 = 12$) und
- bei dem Mobilfunkanbieter: $\bar{x}_2 = 40$ (mit einer Streuung von $s_2 = 4$).

Von welchem Spot wurde die Testperson – gemessen am Aktivierungspotential des Produktbereichs – stärker aktiviert?

Wahrscheinlichkeitstheorie

Die Wahrscheinlichkeit P (P steht für *probabilité*) ist ein quantitatives Maß, das die Sicherheit oder Unsicherheit über das Eintreten eines Ereignisses ausdrückt. Ereignisse werden im Rahmen von Zufallsexperimenten modelliert:

Zufallsexperiment

Ein *Zufallsexperiment* ist ein reales oder gedachtes Experiment, das beliebig oft wiederholt werden kann, nach einer bestimmten Vorschrift ausgeführt wird und dessen Ergebnis nicht mit Sicherheit vorhersagbar ist (also „vom Zufall abhängt“).

Das Ergebnis oder der Ausgang eines Zufallsexperiments heißt *Ereignis*. Ein *Elementarereignis* ist ein Ereignis, das nicht weiter in andere Ereignisse zerlegbar ist. Der *Ereignisraum* (*sample space*) ist die Menge aller möglichen Elementarereignisse. Als Symbol für den Ereignisraum dient Ω oder S .

Beispiel:

„Werfen eines Würfels“ ist das Zufallsexperiment,
 $\{6\}$, $\{\text{gerade Augenzahl}\}$, $\{5\}$, $\{\text{Augenzahl kleiner als } 3\}$ sind Ereignisse,
 $\{6\}$, $\{5\}$ sind Elementarereignisse (es gibt noch vier weitere),
 $\Omega = \{1, 2, 3, 4, 5, 6\}$ ist der Ereignisraum.

Jedes Ereignis ist also eine Teilmenge des Ereignisraums und setzt sich aus einem oder mehreren Elementarereignissen zusammen.

Wahrscheinlichkeitskonzepte

Die Wahrscheinlichkeit für das Eintreten eines Ereignisses A , symbolisiert durch $P(A)$, ist eine reelle Zahl zwischen 0 und 1 (oder zwischen 0% und 100%), die dem Ereignis zugeordnet wird und die bestimmten Axiomen genügt (Näheres zu den Axiomen s. z.B. Bortz, 1999, S. 53). Man unterscheidet drei Konzepte von Wahrscheinlichkeit:

1. Subjektive Wahrscheinlichkeit

Die *subjektive* Wahrscheinlichkeit charakterisiert die Stärke der inneren Überzeugung darüber, ob ein Ereignis eintritt. Es gibt sie bei wiederholbaren und bei nicht-wiederholbaren (d.h. nur einmalig auftretenden) Ereignissen.

2. Empirische Wahrscheinlichkeit

Die *empirische (statistische)* Wahrscheinlichkeit gilt für wiederholbare Ereignisse, die nicht unbedingt gleich wahrscheinlich sein müssen. Die empirische Wahrscheinlichkeit eines Ereignisses A wird aufgrund empirischer Daten geschätzt, wobei als Schätzwert die *relative Häufigkeit* bei einem n -mal wiederholten Zufallsexperiment dient:

n	Anzahl der Wiederholungen (Versuche) des Zufallsexperiments
n_A	absolute Häufigkeit des Auftretens des Ereignisses A bei n Wiederholungen
$\frac{n_A}{n}$	relative Häufigkeit des Ereignisses A = Schätzwert für $P(A)$ (Exakt: $P(A)$ = Grenzwert der relativen Häufigkeit des Ereignisses wenn n gegen unendlich geht).

3. Theoretische Wahrscheinlichkeit

Die *theoretische (klassische)* Wahrscheinlichkeit gilt für wiederholbare, gleich wahrscheinliche und einander ausschließende Ereignisse. Sie lässt sich rein theoretisch berechnen:

k	Anzahl der möglichen Elementarereignisse
k_A	Anzahl der Elementarereignisse, aus denen das Ereignis A besteht
$P(E) = \frac{1}{k}$	Wahrscheinlichkeit für das Auftreten des Elementarereignisses E
$P(A) = \frac{k_A}{k}$	Wahrscheinlichkeit für das Auftreten des Ereignisses A

Beispiel:

Im Zufallsexperiment „Roulette“ ist $k = 37$.

Für das Ereignis „Rot“ gilt $k_{\text{Rot}} = 18$.

Die Wahrscheinlichkeit für eine bestimmte Zahl beträgt jeweils $P(E) = \frac{1}{37}$.

Die Wahrscheinlichkeit für „Rot“ beträgt $P(\text{Rot}) = \frac{k_{\text{Rot}}}{k} = \frac{18}{37} = 48,65\%$.

Schreibweise: Für die theoretische Wahrscheinlichkeit wird statt P manchmal auch das Symbol π (griech. pi) verwendet, was aber nichts mit der Kreiszahl π zu tun hat.

Zusammenfassung: Der Berechnung sowohl empirischer als auch theoretischer Wahrscheinlichkeiten liegt folgender Gedanke zugrunde:

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

Aufgaben

1. Wir werfen zwei Münzen. Wie sieht der Ereignisraum dieses Zufallsexperiments aus?
2. Überlege dir zu jedem der drei Wahrscheinlichkeitskonzepte ein Beispiel.

Rechenregeln

Aus dem Abschnitt „Rechenregeln“ reicht für die Prüfung/Klausur aus: Die Berechnung der Wahrscheinlichkeit eines Komplementärereignisses.

Komplement, Komplementärereignis

Schreibweise: A steht für ein beliebiges Ereignis. \bar{A} symbolisiert das Komplement oder Komplementärereignis des Ereignisses A (also „nicht A “, „das Gegenteil von A “). Die Wahrscheinlichkeit des Komplementärereignisses wird so berechnet:

$$P(\bar{A}) = 1 - P(A)$$

Andersherum geht's auch: $P(A) = 1 - P(\bar{A})$

Additionssatz

Für zwei einander *ausschließende* Ereignisse A und B eines Zufallsexperiments gilt:
Die Wahrscheinlichkeit, dass A oder B (logisches ODER) eintreten, ist

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

Allgemein: Für zwei beliebige Ereignisse A und B eines Zufallsexperiments gilt:
Die Wahrscheinlichkeit, dass A oder B oder beide (logisches ODER) eintreten, ist

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

Die zweite Formel schließt die erste mit ein.

Bedingte Wahrscheinlichkeit

Die bedingte Wahrscheinlichkeit $P(B | A)$ (lies: „P von B unter der Bedingung A“) ist die Wahrscheinlichkeit für das Ereignis B unter der Bedingung, dass das Ereignis A bereits eingetreten ist:

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

Für die bedingte Wahrscheinlichkeit eines Ereignisses B unter der Bedingung A ist der zugehörige Ereignisraum nicht mehr der ursprüngliche (also nicht mehr Ω), sondern eingeschränkt auf das Ereignis A.

Multiplikationssatz

Für zwei voneinander *unabhängige* Ereignisse A und B eines Zufallsexperiments gilt:
Die Wahrscheinlichkeit, dass sowohl A als auch B eintreten (logisches UND), ist

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

Allgemein: Für zwei beliebige Ereignisse A und B eines Zufallsexperiments gilt:
Die Wahrscheinlichkeit, dass sowohl A als auch B eintreten (logisches UND), ist

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

Die zweite Formel schließt die erste mit ein.

Aufgaben

1. Es wird mit einem idealen Spielwürfel gespielt. Es sei dreimal hintereinander eine „6“ erschienen. Wie groß ist die Wahrscheinlichkeit dafür, dass auch der vierte Wurf eine „6“ ist?
2. Wie groß ist die Wahrscheinlichkeit für ein unmögliches Ereignis?
3. Wie nennt man ein Ereignis A, dessen Auftretenswahrscheinlichkeit $P(A) = 1$ ist?
4. Was ist ein Komplementärereignis? Beispiel?
5. Wie groß ist die Wahrscheinlichkeit für das Komplementärereignis von A, wenn $P(A) = 0,3$ ist?
6. Wie groß ist die Wahrscheinlichkeit, aus einem gut gemischten Skatspiel (32 Karten) mit einmaligem Ziehen einen König zu ziehen?
7. Wie groß ist die Wahrscheinlichkeit, aus einem gut gemischten Skatspiel mit einmaligem Ziehen einen König oder eine Dame zu ziehen?
8. Bei einer Prüfung wird für jeden Prüfling eine von zehn vorgegeben Prüfungsfragen per Los gezogen. Wie groß ist die Wahrscheinlichkeit dafür, dass von drei Studierenden
 - a) alle die erste Frage gestellt bekommen?
 - b) alle die gleiche Frage gestellt bekommen?
9. Aus einem gut gemischten Skatspiel werden nacheinander zwei Karten gezogen. Die zuerst gezogene Karte wird nicht zurückgesteckt. Berechne die Wahrscheinlichkeiten für folgende Ereignisse:
 - a) Beim ersten Ziehen kommt Kreuz und beim zweiten Ziehen Herz.
 - b) Beim ersten Ziehen kommt Kreuz oder Herz.

Zufallsvariable und ihre Funktionen

Zufallsvariable

Oft betrachtet man nicht den Ereignisraum selbst, sondern stattdessen eine Zahlenmenge: Man ordnet jedem möglichen Elementarereignis des Ereignisraums eine reelle Zahl zu. Durch diese Abbildung entsteht eine *Zufallsvariable (random variable)*, deren Zahlenwerte für einzelne Elementarereignisse stehen. Einen konkreten Zahlenwert nennt man *Realisation der Zufallsvariablen*.

Schreibweise: Zufallsvariablen werden mit Großbuchstaben gekennzeichnet (X oder Y etc.); die dazugehörigen möglichen Werte oder Realisationsmöglichkeiten der Zufallsvariablen mit Kleinbuchstaben (x_1, x_2, x_3, \dots oder y_1, y_2, y_3, \dots etc.).

Jedes Elementarereignis bzw. jede Realisation x_i der Zufallsvariablen X tritt mit einer bestimmten Wahrscheinlichkeit auf, die dem x_i -Wert zugeordnet wird.

Wahrscheinlichkeitsfunktion

Eine *Wahrscheinlichkeitsfunktion* (*probability function*) gibt für jeden möglichen x_i -Wert der *diskreten* Zufallsvariablen X an, mit welcher Wahrscheinlichkeit er auftritt. D.h. die Wahrscheinlichkeitsfunktion f ordnet jedem möglichen x_i -Wert (und damit jedem Element des Ereignisraumes) eine Wahrscheinlichkeit P zu:

$$f(x_i) = P(X = x_i) \quad \text{oder kurz: } f(x_i) = p_i$$

Beispiel: Beim Zufallsexperiment „Werfen eines Würfels“ lautet die Wahrscheinlichkeit, eine „5“ zu würfeln, also die Wahrscheinlichkeit des Ereignisses $\{5\}$, wie folgt:

$$f(5) = P(X = 5) = \frac{1}{6} = 0,1667 = 16,67\%$$

Grafisch lässt sich die Wahrscheinlichkeitsfunktion durch ein Stabdiagramm darstellen. In unserem Beispiel resultiert ein Diagramm mit sechs gleich hohen Stäben der Höhe 0,1667.

Die Summe aller Einzelwahrscheinlichkeiten ist immer gleich 1: $\sum_{\text{alle } i} f(x_i) = \sum_{\text{alle } i} p_i = 1$

Wahrscheinlichkeitsdichte (Dichtefunktion)

Bei stetigen Zufallsvariablen kann *nicht* die Wahrscheinlichkeit für das Auftreten einzelner Elementarereignisse berechnet werden, sondern die Wahrscheinlichkeit für das Auftreten von Ereignissen, die sich in einem bestimmten Intervall befinden.

Eine *Wahrscheinlichkeitsdichte* oder *Dichtefunktion* (*probability density function*) gibt an, mit welcher Wahrscheinlichkeit die *stetige* Zufallsvariable X einen Wert zwischen a und b annimmt. Diese Wahrscheinlichkeit berechnet sich als Integral der Dichtefunktion f zwischen den Intervallgrenzen a und b :

$$P(a \leq X \leq b) = \int_a^b f(x) \, dx$$

Grafisch wird die Dichtefunktion als stetige Kurve wiedergegeben.

Die Gesamtfläche zwischen x -Achse und der Kurve der Dichtefunktion ist auf 1 normiert:

$$\int_{-\infty}^{+\infty} f(x) \, dx = 1$$

Verteilungsfunktionen

Die *Verteilungsfunktion* (*distribution function*) $F(x)$ einer Zufallsvariablen X ist die Funktion, die die Wahrscheinlichkeit dafür angibt, dass die Zufallsvariable X höchstens den Wert x annimmt:

$$F(x) = P(X \leq x)$$

Verteilungsfunktionen gibt es für diskrete und für stetige Zufallsvariablen:

Die Verteilungsfunktion einer *diskreten* Zufallsvariablen ist die kumulierte Wahrscheinlichkeitsfunktion. Kumuliert wird durch Summieren der Einzelwahrscheinlichkeiten:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i) \quad \text{für eine diskrete Zufallsvariable } X$$

Die Verteilungsfunktion einer *stetigen* Zufallsvariablen ist das Integral der Dichtefunktion:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt \quad \text{für eine stetige Zufallsvariable } X$$

D.h. bei der Verteilungsfunktion für eine stetige Zufallsvariable ist $F(x)$ gleich dem Flächeninhalt unter dem Graphen der Funktion f von $-\infty$ bis zum Wert x .

Aufgaben

1. Was ist eine Zufallsvariable? Was sind diskrete, was sind stetige Zufallsvariablen?
2. Stelle eine Wahrscheinlichkeitsfunktion grafisch dar und veranschauliche die Wahrscheinlichkeit für einen ganz bestimmten Wert der Variablen X .
3. Zeichne den Graphen einer Dichtefunktion und veranschauliche die Wahrscheinlichkeit für das Auftreten eines Wertes der Variablen X , der einen bestimmten Wert x nicht überschreitet.
4. Wie Aufgabe 3, aber veranschauliche die Wahrscheinlichkeit für das Auftreten eines Wertes der Variablen X in einem bestimmten Intervall.
5. Mache dir an der Grafik aus Aufgabe 4 klar, dass gilt:
$$P(a \leq X \leq b) = F(b) - F(a)$$

Verteilungsmodelle

Die meisten Zufallsvariablen können durch einige Verteilungsmodelle beschrieben werden. Das trifft für viele empirisch beobachtbare Verteilungen zu, v.a. aber auch für Verteilungen von statistischen Kennwerten (→ Stichprobentheorie). Hier werden nur zwei Verteilungsmodelle behandelt.

Gleichverteilung

Gleichverteilung einer *diskreten* Zufallsvariablen heißt, dass alle k möglichen Ereignisse bzw. x_i -Werte gleich wahrscheinlich sind:

$$f(x_i) = \frac{1}{k} \text{ für alle } i = 1, \dots, k$$

Beispiel: Beim Würfel haben alle sechs Elementarereignisse jeweils die Wahrscheinlichkeit $f(x_i) = 1/6 = 0,1667$.

Grafisch lässt sich eine solche Wahrscheinlichkeitsfunktion als Stabdiagramm mit lauter gleich hohen Stäben oder Säulen darstellen.

Bei einer *stetigen* Zufallsvariablen hat die Funktionsvorschrift für eine Gleichverteilung einen konstanten Wert (welchen, interessiert uns hier nicht). Der Graph der Dichtefunktion ist deshalb eine Gerade parallel zur x-Achse.

Es gibt weitere Verteilungsformen für diskrete Zufallsvariablen und wiederum andere Verteilungsformen für stetige. Im Folgenden betrachten wir das wichtigste Verteilungsmodell der Inferenzstatistik.

Normalverteilung

Die Normalverteilung ist ein Verteilungsmodell für *stetige* Zufallsvariablen und wurde von Carl Friedrich Gauß entwickelt („Gaußsche Glockenkurve“). Der Begriff bezeichnet eine ganze Klasse von symmetrischen, glockenförmigen Verteilungen, die sich in ihrem Mittelwert (Erwartungswert) μ und in ihrer Streuung σ unterscheiden (μ : griech. my; σ : griech. sigma).

μ und σ sind die so genannten *Parameter* in der Funktionsgleichung der Normalverteilung (s. Abb. 8 auf der nächsten Seite). Der Wert von σ bestimmt, ob die Verteilung breit und flach oder eng und hoch ist. Normalverteilungen mit gleichem μ und σ sind identisch.

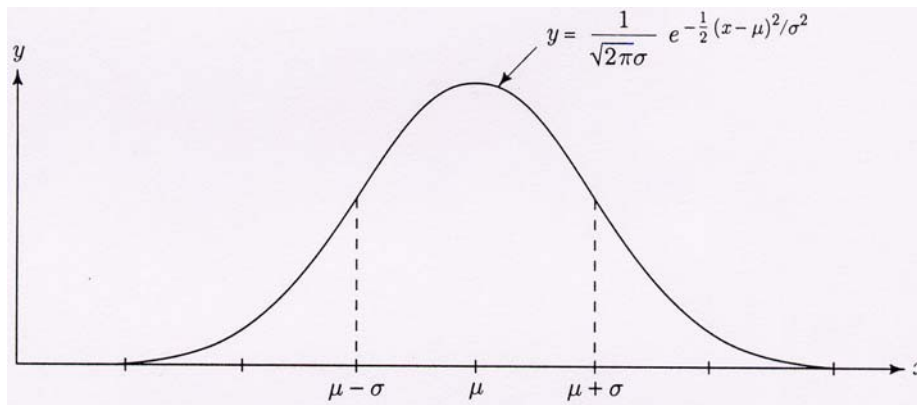


Abbildung 8: Normalverteilung (aus Pitman, 1993, S. 93)

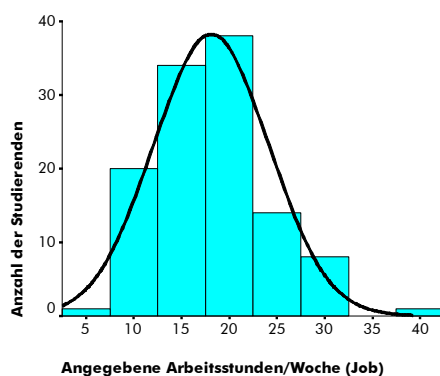
Alle Normalverteilungen haben folgende Eigenschaften:

- Sie sind glockenförmig und symmetrisch.
- Bei Normalverteilungen ist $\mu_o = \mu_d = \mu$. Das Maximum ist also an der Stelle μ .
- Die Kurve der Dichtefunktion nähert sich asymptotisch der x-Achse.
- Zwischen den zu den Wendepunkten gehörenden x-Werten, das ist der Bereich von $\mu - \sigma$ bis $\mu + \sigma$ (kurz: der Bereich $\mu \pm \sigma$), befinden sich ca. 68% der Fläche. In Abbildung 8 ist dieser Bereich markiert. Die Wahrscheinlichkeit, dass ein Wert einer normalverteilten Variablen in diesen Bereich fällt, beträgt ca. 68%.
- Im Bereich $\mu \pm 2\sigma$ befinden sich sogar ca. 95,5% der Fläche.

Wozu brauchen wir die Normalverteilung?

1. Die Normalverteilung ist eine empirisch häufig beobachtbare Verteilung

- Wenn wir wissen, dass sich die Messwerte eines Merkmals normalverteilen, können wir ausrechnen, wie viele der Messwerte in bestimmte Bereiche fallen. Zum Beispiel: Wie groß ist eine Personengruppe, etwa eine Zielgruppe, die einen Index über einem bestimmten Wert aufweist (s. dazu die Aufgabe auf S. 32).
- Wir können prüfen, ob ein empirisch gemessenes Merkmal sich annähernd normal verteilt. Mit den Daten, die z.B. dem Histogramm von S. 14 zugrunde liegen, können wir prüfen, ob sich die wöchentliche Arbeitszeit neben dem Studium bei GWK-Studierenden normalverteilt. Hierfür gibt es zwei Möglichkeiten: Man kann es zum einen optisch prüfen (grobe Prüfung, s. Abbildung links). Zum anderen kann man auch ausrechnen, ob die Merkmalsverteilung einer Normal-



verteilt. Mit den Daten, die z.B. dem Histogramm von S. 14 zugrunde liegen, können wir prüfen, ob sich die wöchentliche Arbeitszeit neben dem Studium bei GWK-Studierenden normalverteilt. Hierfür gibt es zwei Möglichkeiten: Man kann es zum einen optisch prüfen (grobe Prüfung, s. Abbildung links). Zum anderen kann man auch ausrechnen, ob die Merkmalsverteilung einer Normal-

verteilung entspricht. Das ist relevant, weil viele Verfahren des statistischen Hypothesentestens nur anwendbar sind, wenn die Merkmale annähernd normalverteilt sind.

2. Die Normalverteilung ist ein Verteilungsmodell für Stichprobenkennwerte

- Mithilfe eines Verteilungsmodells für Stichprobenkennwerte (was das genau ist: → Kap. Stichprobentheorie) können wir die Fehlertoleranzen errechnen, die sich beim Hochrechnen von Stichproben auf die Grundgesamtheit ergeben (→ Abschnitt Intervallschätzung im Kap. Stichprobentheorie).
- Mithilfe eines Verteilungsmodells für Stichprobenkennwerte können wir das Fehlerisiko berechnen, das wir eingehen, wenn wir uns für oder gegen eine Hypothese entscheiden (→ Kap. Hypothesen und ihre Überprüfung).

Standardnormalverteilung

Wichtig für uns ist, dass wir normalverteilte Werte – wie andere Werte auch – mit einer z-Transformation standardisieren können. Weil Mittelwert und Streuung bei Normalverteilungen mit μ und σ bezeichnet werden, sieht die Formel für die z-Transformation dann anders aus als bisher, nämlich

$$z_i = \frac{x_i - \mu}{\sigma} \quad \text{z-Transformation für normalverteilte } x_i\text{-Werte}$$

Durch eine z-Transformation wird jede Normalverteilung in die *Standardnormalverteilung* überführt, das ist die Normalverteilung mit $\mu = 0$ und $\sigma = 1$ (Abb. 9). Beliebige Normalverteilungen können dadurch standardisiert bzw. vergleichbar gemacht werden.

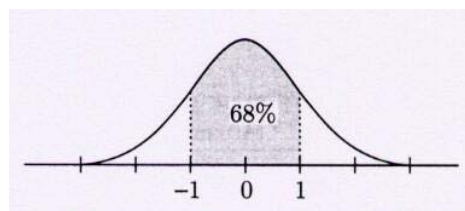


Abbildung 9: Standardnormalverteilung (aus Pitman, 1993, S. 96)

Vorteil: Wir können alle Normalverteilungen, mit denen wir es zu tun haben, in die Standardnormalverteilung umwandeln. Dabei verschieben sich alle Werte maßstabsgetreu, und wir brauchen nur mit der Standardnormalverteilung zu rechnen. Die Werte der Verteilungsfunktion der Standardnormalverteilung, also die zu den x-Werten gehörenden Flächenanteile, gibt es fertig in Tabellen (z.B. Tabelle B in Bortz, 1993, 4. Aufl., S. 694–698 oder Bortz, 1999, 5. Aufl., S. 768–772).

Aufgabe

Für diese Aufgabe brauchst du eine Tabelle mit der Verteilungsfunktion (Flächenanteile) der Standardnormalverteilung (z.B. Tabelle B in Bortz, 1999, S. 768ff.):

In einer Untersuchung zur Markenaffinität werden verschiedene Konstrukte durch Items in einem Fragebogen erfasst. Items für das Konstrukt Extraversion sind z.B. „Ich habe gern viele Leute um mich herum“ oder „Ich ziehe es gewöhnlich vor, Dinge allein zu tun“. Die Befragten kreuzen an, inwieweit die jeweiligen Aussagen auf sie zutreffen. Aus vielen zum Konstrukt Extraversion gehörenden Items wird für jede Person ein Extraversions-Rohwert berechnet. Diese Werte verteilen sich in der Bevölkerung normal. Sie werden dann der Übersichtlichkeit halber normiert, so dass ein mit $\mu = 100$ und $\sigma = 10$ normalverteilter Extraversions-Index (E-Index) entsteht.

- a) Wie viel Prozent der Bevölkerung haben dann einen E-Index von über 100?
- b) Wie viele haben einen E-Index zwischen 90 und 100?
- c) Wie groß ist der Prozentsatz derer, die einen E-Index zwischen 103 und 105,7 haben?
- d) Wie groß ist die Wahrscheinlichkeit dafür, dass jemand einen „von den meisten Menschen abweichenden“ E-Index unter 80 oder über 120 hat?
- e) Wie viel Prozent haben einen „überdurchschnittlichen“ E-Index, soll hier bedeuten: über 110?

Stichprobentheorie

Die Gesamtheit aller Personen oder Objekte, über die etwas ausgesagt werden soll, heißt *Grundgesamtheit* oder *Population*. Populationen sind meistens so groß, dass man keine Vollerhebung durchführt, sondern nur eine Stichprobe untersucht.

Eine *Stichprobe* ist eine Teilmenge der Population, die für eine statistische Untersuchung ausgewählt wird. Sie soll die Population möglichst genau abbilden, d.h. sie soll repräsentativ für die Population sein. Eine Stichprobe kann für eine Population in Bezug auf alle Merkmale repräsentativ sein (*globale Repräsentativität*) oder nur in Bezug auf bestimmte Merkmale (*spezifische Repräsentativität*). Die Anzahl der Personen oder Objekte in einer Stichprobe heißt *Stichprobenumfang* und wird mit n symbolisiert.

Die Repräsentativität einer Stichprobe für die Population, aus der sie stammt und über die Aussagen gemacht werden sollen, hängt weniger vom Stichprobenumfang ab als vom Auswahlverfahren. Eine hohe Repräsentativität erreicht man v.a. durch Zufallsverfahren (z.B. einfache Zufallsstichprobe, geschichtete Stichprobe), aber auch mit dem Quotenverfahren. Im Folgenden werden verschiedene Verfahren beschrieben.

Stichprobenarten (Auswahlverfahren)

Zufallsverfahren (Wahrscheinlichkeitsauswahlverfahren)

- *Einfache Zufallsstichprobe (simple random sample)* heißt: zufällige Auswahl von Elementen aus einer Population, jedes Element kann mit gleicher Wahrscheinlichkeit ausgewählt werden. Man geht dabei vom Urnenmodell aus: Alle Elemente der Population sind in einer Liste/Datei verzeichnet, aus der dann bestimmte Elemente per Zufall ausgewählt werden, z.B. per Los oder mithilfe von Zufallszahlen.

Ökonomischer bzw. durch die Einbeziehung von Vorkenntnissen über die relevanten Merkmale in der Population spezifischer und präziser sind folgende Verfahren:

- *Geschichtete Stichprobe (stratified sample)*: Wenn man Merkmale (Determinanten) kennt, die die untersuchungsrelevanten Merkmale beeinflussen, kann man eine Stichprobe zusammenzustellen, die in Bezug auf diese Determinanten repräsentativ für die Population ist (spezifische Repräsentativität). Welche Determinanten man dazu auswählt, hängt ab vom Thema der Untersuchung, von der Fragestellung und von bereits

vorhandenen Theorien und Erkenntnissen. Determinanten können z.B. Alter oder Einkommen sein. Die Population wird anhand der Determinanten in Schichten (Strata) aufgeteilt; jeder Merkmalsausprägung bzw. -kombination entspricht eine Teilpopulation/Schicht. Aus jeder Schicht wird jeweils eine Zufallsstichprobe (oder, wenn das nicht geht, eine Klumpenstichprobe, s.u.) gezogen. Im Idealfall der *proportional geschichteten Stichprobe* wird die Stichprobe so ausgewählt, dass die prozentuale Verteilung der Schichtungsmerkmale in der Stichprobe der prozentualen Verteilung in der Population entspricht.

- Eine *Klumpenstichprobe (cluster sample)* besteht aus mehreren zufällig ausgewählten Teilmengen (Klumpen, cluster), die schon vorgruppiert sind. Z.B. werden *mehrere* Schulklassen anhand eines vollständigen Verzeichnisses von Schulklassen *zufällig* ausgewählt und dann *vollständig* befragt. Anderes Beispiel: Wahlkreise.
- *Mehrstufige Stichprobe (multi-stage sample)*: Mehrere Zufallsauswahlen werden hintereinander geschaltet. So kann man zunächst eine Klumpenstichprobe mit großen Klumpen ziehen, dann aus den Klumpen Zufallsstichproben ziehen. Auch mehr als zwei Stufen sind möglich.

Nicht-Zufallsverfahren (bewusste Auswahlverfahren)

- *Ad-hoc-Stichprobe*: Auswahl aufs Geratewohl oder willkürlich, d.h. Auswahl weder per Zufall noch mit einem festgelegten Auswahlplan. Beispiele: willkürliche Straßenbefragungen; „zufällig“ anwesende TeilnehmerInnen eines Seminars. Vor solchen Erhebungen wird gewarnt. Wenn ausnahmsweise Ad-hoc-Stichproben erhoben werden, müssen die Besonderheiten der jeweiligen Stichprobe explizit diskutiert und Verallgemeinerungen sehr vorsichtig formuliert werden.
- *Auswahl typischer Fälle (theoretical sampling)*: Auswahl von Fällen, die als besonders charakteristisch angesehen werden. So geht man z.B. im qualitativen Forschungsansatz *Grounded Theory* vor.
- *Quotenstichprobe*: Auch hier kennt man die Merkmale, bezüglich derer die Stichprobe repräsentativ sein sollte. Die Stichprobe wird so zusammengestellt und angepasst, dass die Verteilung der relevanten Merkmale in der Stichprobe der Verteilung dieser Merkmale in der Population genau entspricht. Bei der Quotenstichprobe handelt es sich *nicht* um eine geschichtete Stichprobe, sondern um eine bewusste Auswahl von Personen in gewünschter Häufigkeit, die die entsprechenden Merkmalsausprägungen haben. Dazu erhält der/die ForscherIn oder InterviewerIn eine Quotenanweisung mit vorgegebenen Prozentanteilen (Quoten) für bestimmte Merkmalsausprägungen. Das Quotenverfahren ist sehr ökonomisch, seine Repräsentativität aber umstritten. In empirischen Vergleichen mit Zufallsverfahren schneidet das Quotenverfahren jedoch gar nicht schlecht ab, so dass ich es – unter der Voraussetzung, dass *mehrere* InterviewerInnen zum Einsatz kommen – als pragmatische Lösung empfehlen kann.

Stichprobenkennwerte und Parameter

Um die Merkmalsverteilung einer Stichprobe zu charakterisieren, benutzen wir statistische *Stichprobenkennwerte*, die wir mit lateinischen Buchstaben bezeichnen, z.B.

- \bar{x} Mittelwert in der Stichprobe
- s Streuung in der Stichprobe
- p relative Häufigkeit (Prozentanteil) in der Stichprobe
- r Korrelation in der Stichprobe (Korrelation: → Kap. Lineare Zusammenhänge)

Bei Populationen sprechen wir stattdessen von *Parametern* oder *Populationsparametern* und bezeichnen sie mit griechischen Buchstaben, z.B.

- μ Mittelwert in der Population
- σ Streuung in der Population
- π relative Häufigkeit (Prozentanteil) in der Population
- ρ Korrelation in der Population (ρ : griech. rho)

(Von Parametern sprechen wir übrigens nicht nur bei Populationen, sondern auch bei theoretischen Verteilungen, → Normalverteilung).

Anhand der Ergebnisse einer Stichprobe (Stichprobenkennwerte) wollen wir einen Rückschluss auf die unbekannten Verhältnisse in der Population (Parameter) ziehen. D.h. wir wollen z.B. den Mittelwert μ der Population schätzen, indem wir aus der Stichprobe den Mittelwert \bar{x} berechnen. Bei einer solchen Schätzung von μ gibt es ein Fehlerisiko, dessen Höhe mit Hilfe der Wahrscheinlichkeitsrechnung quantifiziert werden kann.

Die folgenden Überlegungen dienen der Beantwortung der Frage: Wie brauchbar ist der Stichprobenkennwert \bar{x} als Schätzwert für den Populationsparameter μ ? Mit anderen Worten: Wie genau ist die Schätzung und wie groß die Fehlertoleranz?

Stichprobenkennwerte-Verteilung

Gedankenexperiment: Wir ziehen aus einer Population sehr viele (theoretisch unendlich viele) gleich große Zufallsstichproben und errechnen aus jeder einzelnen Stichprobe einen Stichprobenkennwert. Wir wählen als Stichprobenkennwert den Mittelwert, weil er am häufigsten verwendet wird; es geht natürlich auch mit anderen Kennwerten. Die Stichprobenkennwerte unserer vielen Stichproben, also die vielen Mittelwerte, werden nicht alle gleich sein, sondern wir erhalten von Stichprobe zu Stichprobe etwas unterschiedliche Mit-

telwerte. Damit entsteht eine *Verteilung* von Mittelwerten. Eine solche Verteilung heißt *Stichprobenkennwerte-Verteilung* (*sampling distribution*).

Damit haben wir wieder eine Zufallsvariable, aber sie heißt hier \bar{X} statt X . Die Zufallsvariable \bar{X} besteht aus lauter Mittelwerten $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N$ aus N Stichproben mit jeweils demselben Stichprobenumfang n . (N symbolisiert hier die sehr große Anzahl der Stichproben; dagegen haben wir das Symbol n für den Stichprobenumfang reserviert, d.h. für die Anzahl der Objekte in einer Stichprobe. Achtung: Das Symbol N wird manchmal auch benutzt für die Anzahl der Objekte in einer Population).

Die Streuung der Stichprobenkennwerte-Verteilung

Die Streuung oder Standardabweichung der Stichprobenkennwerte-Verteilung nennt man *Standardfehler* (*standard error*, Notation: SE). Wenn es sich um eine Verteilung von Stichprobenmittelwerten handelt, sprechen wir vom *Standardfehler des Mittelwertes* (= Standardabweichung der Stichprobenmittelwerte, abgekürzt $\sigma_{\bar{x}}$).

Wenn wir andere Stichprobenkennwerte nehmen (z.B. Median oder Prozentwert), haben wir eine Verteilung von Medianwerten bzw. von Prozentwerten und dementsprechend einen Standardfehler des Medians σ_{Md} bzw. einen Standardfehler des Prozentwertes $\sigma_{\%}$. Wir beschränken uns hier aber auf die Verteilung von Stichprobenmittelwerten.

Der Standardfehler des Mittelwertes gibt an, wie gut ein einzelner Stichprobenmittelwert den unbekannten Populationsparameter μ schätzt: Je geringer die Streuung oder Standardabweichung der Stichprobenkennwerte-Verteilung ist, desto genauer schätzt ein einzelner Stichprobenkennwert den gesuchten Populationsparameter.

Je stärker das gemessene Merkmal in der Population streut, desto unterschiedlicher fallen auch die Stichprobenergebnisse aus, d.h. desto größer ist die Streuung der Stichprobenmittelwerte. Es handelt sich dabei um einen proportionalen Zusammenhang. Kurz: $\sigma_{\bar{x}}$ ist proportional zu σ .

Ferner hängt $\sigma_{\bar{x}}$ vom Stichprobenumfang n ab: Je größer der Stichprobenumfang ist, desto genauer ist die Schätzung des Populationsparameters μ durch einen Stichprobenmittelwert, d.h. desto kleiner ist $\sigma_{\bar{x}}$. n und $\sigma_{\bar{x}}$ hängen jedoch nicht proportional zusammen.

Die Zusammenhänge werden in der Berechnungsformel für den Standardfehler des Mittelwertes sichtbar:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \quad \text{Standardfehler des Mittelwertes}$$

Problem: Im Allgemeinen kennen wir die Populationsstreuung σ bzw. -varianz σ^2 gar nicht und müssen sie deshalb aufgrund der Stichprobenvarianz s^2 schätzen. Die Varianzen von Stichproben sind im Durchschnitt kleiner als die Populationsvarianz; sie unterschätzen die Populationsvarianz σ^2 , so dass für deren Schätzung ein Korrekturfaktor $n/(n-1)$ nötig ist:

$$\hat{\sigma}^2 = s^2 \cdot \frac{n}{n-1} \quad \text{geschätzte Populationsvarianz}$$

„ $\hat{\sigma}^2$ “ spricht man „Sigma Dach Quadrat“. Ein Dach symbolisiert immer, dass es sich um eine Schätzung handelt.

Wenn wir für s^2 die entsprechende Formel (vgl. S. 17) einsetzen, ergibt sich für die geschätzte Populationsvarianz:

$$\hat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n} \cdot \frac{n}{n-1} = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Wenn wir statt mit der Varianz mit der Streuung rechnen wollen, ziehen wir die Wurzel:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} \quad \text{geschätzte Populationsstreuung}$$

Damit können wir dann den Standardfehler schätzen:

$$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{\hat{\sigma}^2}{n}} = \frac{\hat{\sigma}}{\sqrt{n}} \quad \text{geschätzter Standardfehler des Mittelwertes}$$

(Für %-Anteile, z.B. bei Wahlumfragen, ist der geschätzte Standardfehler des Prozentwertes $\hat{\sigma}_{\%}$ relevant. Die Formel findet sich z.B. bei Bortz, 1999, S. 92).

Aufgaben

1. Welche Stichprobenarten kennst du? Was sind ihre Besonderheiten?
2. Was ist eine Stichprobenkennwerte-Verteilung?
3. Was ist ein Standardfehler?
4. Wenn wir anhand eines Stichprobenmittelwertes den unbekannten Populationsmittelwert schätzen wollen: Ist es dann günstiger, wenn der Standardfehler groß oder wenn er klein ist?
5. Die Größe des Standardfehlers hängt ab ...
erstens von:
zweitens von:
6. Was ist größer: Die Stichprobenvarianz oder die geschätzte Populationsvarianz?

7. Wie kann man die (unbekannte) Populationsvarianz σ^2 schätzen?
8. Was bedeuten die folgenden Symbole?
- a) \bar{x}
 - b) μ
 - c) s und s^2
 - d) σ und σ^2
 - e) $\sigma_{\bar{x}}$
 - f) $\hat{\sigma}$ und $\hat{\sigma}^2$
 - g) $\hat{\sigma}_{\bar{x}}$

Die Form der Stichprobenkennwerte-Verteilung

Die Verteilung von Mittelwerten aus *theoretisch* unendlich vielen Stichproben des Umfangs n , die sämtlich der derselben Population entnommen wurden, geht mit wachsendem Stichprobenumfang n ($n \geq 30$) in eine *Normalverteilung* über, und zwar *unabhängig* von der Verteilungsform des Merkmals in der Population (Zentraler Grenzwertsatz). Bereits ab $n \geq 30$ ist \bar{X} in guter Näherung normalverteilt.

Achtung: n = Stichprobenumfang = Stichprobengröße = Anzahl der Objekte in einer Stichprobe (nicht die Anzahl der Stichproben!).

Daraus folgt: Wenn wir den unbekannten Populationsparameter μ durch einen Stichprobenkennwert \bar{x} schätzen wollen, können wir uns die Eigenschaften der Normalverteilung zunutze machen. Das tun wir im Abschnitt „Intervallschätzung“.

Aufgabe

Versuche, die Aussage des Zentralen Grenzwertsatzes mit eigenen Worten auszudrücken.

Der Mittelwert der Stichprobenkennwerte-Verteilung

Zwar gibt *ein* Stichprobenmittelwert im Allgemeinen nicht den wahren Populationsmittelwert μ wieder. Aber der Mittelwert *sehr vieler* Stichprobenmittelwerte ist gleich dem Populationsmittelwert. „Der Mittelwert von Mittelwerten rutscht immer in die Mitte“, oder mathematisch formuliert: Der Stichprobenmittelwert \bar{x} ist ein *erwartungstreuer* Schätzer des Populationsmittelwertes μ .

Wir halten fest: Der Mittelwert der Stichprobenkennwerte-Verteilung ist gleich dem Mittelwert μ der Verteilung des Merkmals in der Population.

Zusammenfassung

- Mittelwerte aus Zufallsstichproben ($n \geq 30$) verteilen sich
 - normal
 - um den Populationsmittelwert μ
 - mit der Streuung $\sigma_{\bar{x}}$(das ergibt sich aus den drei genannten Punkten: Streuung, Form und Mittelwert).
- Bei allen Normalverteilungen befinden sich im Bereich $\mu \pm 2\sigma$ ca. 95,5% aller Werte (das ergibt sich aus der Verteilungsfunktion).

Daraus ergibt sich der maximale Abstand, den Mittelwerte aus Zufallsstichproben „meistens“ (nämlich zu ca. 95,5%) vom Populationsmittelwert μ haben. Dieser maximale Abstand beträgt $2\sigma_{\bar{x}}$ (bzw. $2\hat{\sigma}_{\bar{x}}$). Anders formuliert: Zu ca. 95,5% liegen Mittelwerte aus Zufallsstichproben zwischen dem Wert $\mu - 2\hat{\sigma}_{\bar{x}}$ und dem Wert $\mu + 2\hat{\sigma}_{\bar{x}}$. Die Ungleichung $\mu - 2\hat{\sigma}_{\bar{x}} \leq \bar{x} \leq \mu + 2\hat{\sigma}_{\bar{x}}$ gilt mit einer Wahrscheinlichkeit von ca. 95,5%.

Intervallschätzung

Die Schätzung eines unbekannten Populationsparameters auf der Basis von Stichprobenkennwerten heißt *Parameterschätzung*. Man unterscheidet dabei Punktschätzung und Intervallschätzung. *Punktschätzung* ist Schätzung von Populationsparametern durch einen einzigen Wert. Nachteil: Punktschätzungen schwanken stark von Zufallsstichprobe zu Zufallsstichprobe. Bei der *Intervallschätzung* wird ein geschätzter Bereich (so genanntes Konfidenzintervall) für den gesuchten Populationsparameter angegeben. Vorteil: Es sind Wahrscheinlichkeitsangaben für die Genauigkeit der Schätzung möglich.

Wie brauchbar/genau ist nun die Schätzung des Populationsparameters μ durch einen \bar{x} ?

Von Mittelwerten aus Zufallsstichproben (mit $n \geq 30$) wissen wir, dass sie sich mit einer Wahrscheinlichkeit von ca. 95,5% im Bereich $\mu \pm 2\hat{\sigma}_{\bar{x}}$ befinden (s.o.). Im Allgemeinen interessiert uns aber nicht, wo Mittelwerte sich „meistens“ aufhalten, sondern wir kennen den Mittelwert \bar{x} einer *einzigsten* Stichprobe und wollen daraus den unbekannten Parameter μ schätzen. Die obige Ungleichung, $\mu - 2\hat{\sigma}_{\bar{x}} \leq \bar{x} \leq \mu + 2\hat{\sigma}_{\bar{x}}$, ergibt umgeformt:

$$\bar{x} - 2\hat{\sigma}_{\bar{x}} \leq \mu \leq \bar{x} + 2\hat{\sigma}_{\bar{x}} \quad \text{ist bei 95,5\% aller Stichproben richtig}$$

Es ist also sehr wahrscheinlich (95,5%), dass der Bereich $\bar{x} \pm 2\hat{\sigma}_{\bar{x}}$ den gesuchten Parameter μ umschließt.

Solche Bereiche heißen *Konfidenzintervalle* und werden durch Δ_{crit} symbolisiert (Δ_{crit} ist die Abkürzung für „kritisches Intervall“, Δ ist ein großes griechisches Delta und symboli-

siert immer Differenzen oder Intervalle). $\Delta_{\text{crit}} = \bar{x} \pm 2 \cdot \hat{\sigma}_{\bar{x}}$ ist das 95,5%ige Konfidenzintervall für den Populationsmittelwert.

Man kann ein Konfidenzintervall auch für andere Wahrscheinlichkeiten berechnen. Die Wahrscheinlichkeit, für die man das Konfidenzintervall berechnet, heißt *Konfidenzkoeffizient*. Für das 95%ige Konfidenzintervall z.B. nehmen wir nicht den Faktor 2, sondern den Faktor 1,96:

$$\Delta_{\text{crit}} = \bar{x} \pm 1,96 \cdot \hat{\sigma}_{\bar{x}} \quad \text{95\%iges Konfidenzintervall}$$

Warum 1,96? Wir benötigen die Grenzen des Intervalls, in dem bei einer Normalverteilung 95% der Gesamtfläche liegen. Die Tabelle der Verteilungsfunktion der Standardnormalverteilung zeigt, dass die Werte $z = -1,96$ bzw. $z = +1,96$ von der Standardnormalverteilungsfläche an beiden Seiten jeweils 2,5% abschneiden, also zusammen 5%. Damit bleiben dazwischen 95% der Fläche übrig.

Aufgaben

1. Wovon hängt die Breite eines Konfidenzintervalls ab?
2. Die bayerische Landesregierung möchte wissen, wie viel Bier im Freistaat getrunken wird. Die Besucher mehrerer zufällig ausgewählter Biergärten wurden zu einer Klumpenstichprobe zusammengefasst und nach ihrem wöchentlichen Bierkonsum befragt ($n = 30$). Wöchentlicher Bierkonsum in Litern:

0,0	2,0	3,5	4,5	6,5
0,0	2,5	4,0	4,5	7,0
0,0	2,5	4,0	5,0	7,5
0,5	3,0	4,0	5,5	8,0
1,0	3,5	4,0	5,5	8,0
1,5	3,5	4,5	6,0	8,0

Bestimme das 95%ige Konfidenzintervall für den Mittelwert!

3. Es wurde das Durchschnittsalter von 36 zufällig ausgewählten StudentInnen im ersten Semester erhoben. Die Stichprobenkennwerte sind $\bar{x} = 21$ und $s^2 = 5$. Wie lautet das 99%ige Konfidenzintervall für den Mittelwert und was bedeutet das?
4. (aus: Moosbrugger & Müller, 1990, *Psychologische Statistik*): 400 zufällig ausgewählte 10-Jährige haben an einem Rechentest teilgenommen. Im Durchschnitt hatten sie 28 richtige Lösungen. Die Standardabweichung der Anzahl richtiger Lösungen betrug 6,8.
 - a) Welches sind die besten Schätzwerte für die Parameter μ und σ^2 der Population?
 - b) Wie groß ist der Standardfehler des Mittelwertes?
 - c) Bestimme das Konfidenzintervall für den Mittelwertparameter der Population bei einem Konfidenzkoeffizienten von 99%.
 - d) Welche Möglichkeiten hast du bei einer praktischen Untersuchung, das Konfidenzintervall enger zu machen?

Hypothesen und ihre Überprüfung

Wenn wir eine Untersuchung durchführen wollen, verfolgen wir meist eine Idee, auf die wir z.B. durch eine Theorie oder durch Beobachtungen gekommen sind und die wir nun empirisch überprüfen wollen. Die Aussagen, die wir überprüfen wollen, heißen Hypothesen. Eine *Hypothese* ist eine Vermutung über die Art der Beziehung von Merkmalen in bestimmten Populationen, d.h. über Zusammenhänge, Unterschiede oder Veränderungen von Merkmalen in den Populationen. Dementsprechend unterscheidet man Zusammenhangs-, Unterschieds- und Veränderungshypothesen.

Während wir beim Schätzen (s. S. 39f.) zunächst die Daten einer Stichprobe betrachten und dann anhand dieser Daten auf die Population schließen, gehen wir beim Testen von Hypothesen umgekehrt vor: Wir formulieren Hypothesen über Eigenschaften der Population (Theorie) und prüfen dann, ob diese Hypothesen durch die Daten einer Stichprobe (Empirie) bestätigt werden können.

Hypothesenarten

Forschungshypothesen und statistische Hypothesen

Die allgemeine inhaltliche Formulierung der Hypothese nennt man *Forschungshypothese*. Beispiel: „Computervermittelte Kommunikation verdrängt andere Kommunikationsmedien.“ Das ist eine sehr allgemeine Aussage und z.B. für den Bereich „E-Mail und Brief“ empirisch nicht haltbar. Wir grenzen ein und formulieren die Hypothese: „Kommunikation über PC konkurriert mit SMS.“ oder genauer: „PC-Nutzer simsen weniger als Nicht-PC-Nutzer“. Das ist ein Beispiel für eine Unterschiedshypothese.

Eine Konkretisierung der Forschungshypothese ist die *operationale Hypothese* (empirische Vorhersage), bei deren Formulierung das Untersuchungsdesign (s. S. 2) und die Operationalisierung (s. S. 3) der Merkmale berücksichtigt werden. Beispiel: „Wenn man 50 zufällig ausgewählte PC-Nutzer und 50 zufällig ausgewählte Nicht-PC-Nutzer eine Woche lang bittet, ihre versendeten SMS zu zählen, werden die PC-Nutzer am Ende der Woche im Durchschnitt weniger SMS versendet haben als die Nicht-PC-Nutzer.“

Statistische Hypothesen beziehen sich auf die quantitativen Maße (Populationsparameter, s. S. 35) für die behaupteten Merkmalsbeziehungen. In diesem Beispiel handelt es sich um zwei Populationsparameter, nämlich erstens um die durchschnittliche Anzahl von SMS im

Laufe einer Woche in der Population der PC-Nutzer (diesen Durchschnittswert in der Population kürzen wir ab mit μ_1) und zweitens um die durchschnittliche Anzahl in der Population der Nicht-PC-Nutzer (abgekürzt mit μ_2).

Man prüft immer zwei sich gegenseitig ausschließende statistische Hypothesen: Die *Alternativhypothese* (H_1) und die *Nullhypothese* (H_0). Die Alternativhypothese behauptet immer Zusammenhänge, Unterschiede oder Veränderungen von Merkmalen in Populationen und entspricht im Allgemeinen der Forschungshypothese. Die Nullhypothese widerspricht ihr und behauptet immer, dass es *keinen* Zusammenhang, keinen Unterschied, keine Veränderung in der Population gibt.

In unserem Beispiel lautet die Alternativhypothese $H_1: \mu_1 < \mu_2$. Das bedeutet: „Die durchschnittliche Anzahl von SMS im Laufe einer Woche in der Population der PC-Nutzer ist kleiner als in der Population der Nicht-PC-Nutzer.“ Die H_0 behauptet, dass es *nicht* so ist, dass also die SMS-Nutzung in beiden Gruppen gleich ist oder sogar bei den PC-Nutzern noch höher als bei den Nicht-PC-Nutzern. Die Nullhypothese lautet damit in unserem Beispiel $H_0: \mu_1 \geq \mu_2$. Alternative Schreibweise: $H_1: \mu_1 - \mu_2 < 0$. $H_0: \mu_1 - \mu_2 \geq 0$.

Gerichtete und ungerichtete Hypothesen

Die oben formulierten Hypothesen H_1 und H_0 geben die Richtung des behaupteten Unterschieds an: Sie sagen, in welcher Gruppe die SMS-Nutzung höher ist. Man nennt sie *gerichtete Hypothesen*.

Dagegen sagen *ungerichtete Hypothesen* nichts über die Richtung des behaupteten Zusammenhanges, Unterschiedes oder der Veränderung aus. Eine ungerichtete Hypothese wäre es, wenn wir behaupteten, dass die beiden Gruppen sich in ihrer SMS-Nutzung unterscheiden, ohne zu sagen, welche Gruppe stärker SMS nutzt.

Die ungerichtete H_1 wäre in unserem Beispiel $H_1: \mu_1 \neq \mu_2$. Die ungerichtete H_0 lautet im Beispiel $H_0: \mu_1 = \mu_2$. Alternative Schreibweise: $H_1: \mu_1 - \mu_2 \neq 0$. $H_0: \mu_1 - \mu_2 = 0$.

Spezifische und unspezifische Hypothesen

Die bisher betrachteten Hypothesen sind *unspezifisch*, d.h. die Höhe oder *Größe* eines für die Population behaupteten Zusammenhanges, eines Unterschiedes oder einer Veränderung (*Effektstärke*, *Effektgröße*, *effect size*) wird *nicht* angegeben.

Wenn eine Hypothese auch die Effektgröße angibt, handelt es sich um eine *spezifische Hypothese*.

Beispiel: „Im Durchschnitt versenden PC-Nutzer mindestens 20 SMS pro Woche weniger als Nicht-PC-Nutzer“, oder kurz, $H_1: \mu_1 \leq \mu_2 - 20$. Die entsprechende H_0 sagt, dass es nicht so ist, oder kurz, $H_0: \mu_1 > \mu_2 - 20$.

Überprüfen von Hypothesen

Eine Hypothese zu testen heißt sie an der Realität zu überprüfen: Ein Realitätsausschnitt in Form einer Stichprobe, die empirische Daten liefert, wird der Hypothese gegenübergestellt.

Einschub für WissenschaftstheoretikerInnen: Wir setzen hier im Sinne des Kritischen Rationalismus voraus, dass es eine Realität gibt und dass wir Poppers Ausweg aus dem *Basisatzproblem* akzeptieren. Basissätze sind aus Beobachtungen abgeleitete Aussagen, „die behaupten, dass sich in einem individuellen Raum-Zeit-Gebiet ein beobachtbarer Vorgang abspielt“ (Popper, 1982, S. 69). Das Basissatzproblem bezieht sich darauf, inwieweit empirische Beobachtungen und Beschreibungen tatsächlich mit der „Realität“ übereinstimmen. Das Problem liegt darin, dass auch intersubjektiv übereinstimmende empirische Beobachtungen „falsch“ sein könnten, wenn sie denselben Wahrnehmungs- oder Messverzerrungen unterliegen. Nach Popper können Basissätze aber als vorläufig gültig betrachtet werden, wenn deren Zustandekommen bestimmten methodologischen Kriterien genügt, die wiederum von einem Konsens innerhalb der Scientific Community abhängen.

Wenn nun die Stichprobendaten tatsächlich so ausfallen, wie wir es aufgrund unserer Hypothese bzw. unserer Theorie erwarten (wenn es z.B. um eine Unterschiedshypothese geht und dann in den Stichproben tatsächlich zwei unterschiedliche Gruppenmittelwerte auftreten), können wir noch nicht sagen, dass die Hypothese damit bestätigt ist. Denn das deskriptive Stichprobenergebnis sagt uns nicht, ob das Ergebnis auch für die Population gilt oder ob es nur *zufällig* aufgrund der Besonderheiten dieser Stichprobe zustande gekommen ist.

Prinzipiell kann man Hypothesen nicht endgültig beweisen (verifizieren), denn eine Stichprobe erfasst die Realität der Population nie vollständig, sondern immer nur ausschnitthaft. Zudem sind sozialwissenschaftliche Hypothesen – im Gegensatz zu den meisten naturwissenschaftlichen Hypothesen – Wahrscheinlichkeitsaussagen, die sich durch konträre Einzelfälle prinzipiell nicht widerlegen (falsifizieren) lassen. Empirische Daten können uns also nicht sagen, ob eine Hypothese „wahr“ ist oder nicht, sondern sie können uns nur eine Entscheidungsgrundlage für die Annahme oder Ablehnung einer Hypothese geben.

Ein Prüfkriterium für Hypothesen ist die statistische *Signifikanz*. Signifikanz bedeutet, dass der Realitätsausschnitt in Form des Stichprobenergebnisses „kaum“ mit der Nullhypothese vereinbar ist – was dann gegen die Nullhypothese spricht. Ob Signifikanz vorliegt, lässt sich mit einem statistischen *Signifikanztest* prüfen. Wenn dieser Test ergibt, dass das Stichprobenergebnis signifikant ist, verwerfen wir die Nullhypothese und entscheiden uns für die Alternativhypothese. Diese ist jetzt untermauert (nicht bewiesen!). Wenn dagegen das Stichprobenergebnis nicht signifikant ist, also „ausreichend gut“ zur Nullhypothese passt, behalten wir diese bei und lehnen die Alternativhypothese ab.

Im Folgenden sind die Begriffe „kaum“ und „ausreichend gut“ zu präzisieren. Damit wird auch der Begriff der Signifikanz präzisiert.

Prinzip des Signifikanztests

Um festzustellen, wie gut ein Stichprobenergebnis mit der H_0 vereinbar ist, machen wir einen Signifikanztest. Ein Signifikanztest ermittelt, mit welcher Wahrscheinlichkeit das gefundene (oder ein extremeres) Stichprobenergebnis auftreten kann, wenn die H_0 zutrifft. Wenn diese Wahrscheinlichkeit p „klein genug“ ist, d.h. per Konvention „kleiner oder gleich 5%“, spricht man von einem *signifikanten Ergebnis*. Kurz: Ein Ergebnis wird als signifikant bezeichnet, wenn gilt: $p \leq 5\%$ bzw. $p \leq 0,05$.

Die Wahrscheinlichkeits-Höchstgrenze von 5% heißt *Signifikanzniveau* (*significance level*). Das Signifikanzniveau wird durch α symbolisiert (α : griech. alpha) und ist per Konvention auf $\alpha = 5\%$ festgelegt, um Forschungsergebnisse vergleichbar zu machen.

Es gibt noch zwei weitere Konventionen: Signifikanzniveau $\alpha = 1\%$ und $\alpha = 0,1\%$. Wenn die Wahrscheinlichkeit p für das gefundene Stichprobenergebnis (unter der Bedingung, dass die H_0 zutrifft) höchstens 1% bzw. 0,01 beträgt, spricht man von einem *sehr signifikanten Ergebnis*; bei 0,1% von einem *höchst signifikanten Ergebnis*.

Signifikante Ergebnisse werden oft mit Sternchen gekennzeichnet:

Wahrscheinlichkeit, mit der das gefundene (oder ein extremeres) Stichprobenergebnis auftreten kann, wenn die H_0 zutrifft	Bezeichnung des Stichprobenergebnisses als ...	Kennzeichnung des Stichprobenergebnisses
$p > 0,05$	nicht signifikant	n.s.
$p \leq 0,05$	signifikant	*
$p \leq 0,01$	sehr signifikant	**
$p \leq 0,001$	höchst signifikant	***

Entscheidung und mögliche Fehlentscheidungen

Wenn ein Ergebnis signifikant ist, kann die H_0 nicht mehr als plausibel gelten. Man entscheidet sich deshalb dafür, die H_0 zu verwerfen und die Alternativhypothese zu akzeptieren. Bei einem nicht-signifikanten Ergebnis wird man sich für die Nullhypothese und gegen die Alternativhypothese entscheiden.

Bei diesen Entscheidungen gibt es zwei Möglichkeiten, einen Fehler zu begehen:

- *Alpha-Fehler* (α -Fehler, Fehler erster Art, Fehler Typ I): Man trifft eine Entscheidung für die Alternativhypothese, obwohl in der Population die H_0 gilt.
- *Beta-Fehler* (β -Fehler, Fehler zweiter Art, Fehler Typ II): Man trifft eine Entscheidung für die Nullhypothese, obwohl in der Population die H_1 gilt.

Welcher der beiden Fehler bei einer Entscheidung zugunsten einer Hypothese bedeutsamer wäre bzw. welcher Fehler möglichst klein gehalten werden soll, ist von den praktischen Konsequenzen der Entscheidung abhängig (vgl. als Beispiel Aufgabe 2 auf S. 49).

Da eine falsche Entscheidung zugunsten der H_1 ein α -Fehler ist, heißt die Wahrscheinlichkeit p , mit der das gefundene (oder ein extremeres) Stichprobenergebnis auftreten kann, wenn die H_0 zutrifft, und mit der wir uns dann *fälschlicherweise* für die H_1 entscheiden würden, α -Fehler-Wahrscheinlichkeit oder *Irrtumswahrscheinlichkeit: Wahrscheinlichkeit eines Irrtums* beim Verwerfen der H_0 bzw. beim Annehmen der H_1 .

Ein Signifikanztest ist also ein Verfahren zur Berechnung der Irrtumswahrscheinlichkeit p . Man spricht auch vom p -Wert (p -value).

Achtung, zur Schreibweise: In einigen, wenigen Statistik-Büchern wird die Irrtums- oder α -Fehler-Wahrscheinlichkeit leider nicht durch p , sondern durch α symbolisiert → nicht mit dem Signifikanzniveau α verwechseln!

Berechnung der Irrtumswahrscheinlichkeit

Zur Berechnung der Irrtumswahrscheinlichkeit vergleicht man das gefundene Stichprobenergebnis, also den Stichprobenkennwert, mit der theoretischen Stichprobenkennwert-Verteilung, die sich unter der Annahme ergibt, dass die H_0 gilt (H_0 -Verteilung oder H_0 -Modell).

Abbildung 10 zeigt die H_0 -Verteilung von theoretisch unendlich vielen Stichprobenmitteln. Der Mittelwert \bar{x} einer Stichprobe ist eingezeichnet und, schattiert, die Wahrscheinlichkeit, mit der dieses (oder ein extremeres) Stichprobenergebnis \bar{x} auftreten kann, wenn die H_0 zutrifft. Das ist die Irrtumswahrscheinlichkeit p .

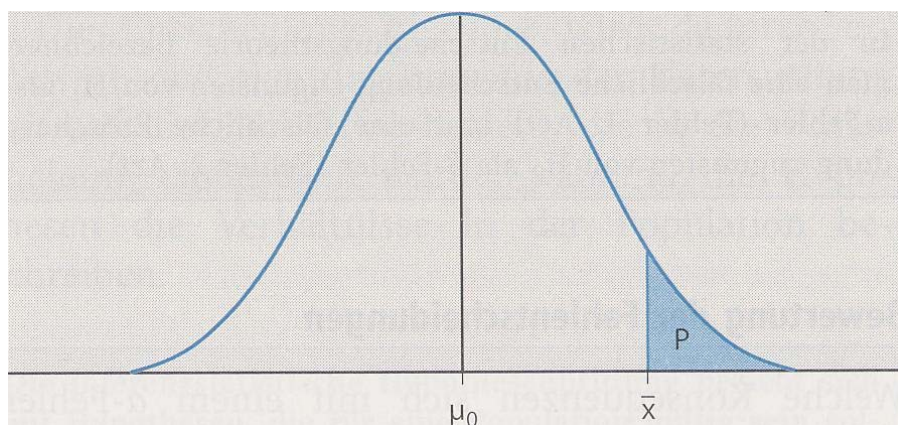


Abb. 10: Irrtumswahrscheinlichkeit beim Verwerfen der H_0 (aus Bortz, 1999, S. 112)

Da es sich um die Wahrscheinlichkeit handelt, dass das gefundene Stichprobenergebnis \bar{x} auftritt oder überschritten wird, nennt man den p-Wert auch *Überschreitungswahrscheinlichkeit*. Folgende Begriffe sind also synonym: α -Fehler-Wahrscheinlichkeit, Wahrscheinlichkeit für einen Fehler erster Art, Irrtumswahrscheinlichkeit, Überschreitungswahrscheinlichkeit, p-Wert.

Ein Beispiel zur Berechnung der Irrtumswahrscheinlichkeit: Der Hersteller einer neuen Sportlernahrung wirbt für sein Produkt mit der Behauptung, dass sie die Leistungsfähigkeit gegenüber der herkömmlichen, sportmedizinisch empfohlenen Ernährung erhöhe. Er möchte seine Werbebotschaft empirisch untermauern. Die Forschungshypothese lautet, dass die KonsumentInnen der neuen Sportlernahrung im Durchschnitt leistungsfähiger sind als diejenigen, die herkömmliche Nahrung essen. Der Leistungsdurchschnitt in der Grundgesamtheit der herkömmlich ernährten Sportler ist aus vielen Untersuchungen bekannt und möge $\mu_0 = 10$ Leistungspunkte betragen (mit einer Streuung von $\sigma = 8$).

Wenn wir mit μ_1 den Leistungsdurchschnitt bei den potentiellen KonsumentInnen des neuen Produkts bezeichnen, ist folgende Alternativhypothese zu prüfen: $H_1 : \mu_1 > \mu_0$. Es ist also die Nullhypothese $H_0 : \mu_1 \leq \mu_0$ zu testen, die besagt, dass die neue Sportlernahrung gegenüber der herkömmlichen Ernährung nichts bringt (oder sogar weniger leistungsfähig macht).

100 zufällig ausgewählte SportlerInnen haben nun das neue Produkt gegessen und sich danach einem Leistungstest unterzogen. Die durchschnittliche Leistung dieser Stichprobe beträgt $\bar{x} = 11,36$ Punkte (hohe Punktzahl entspricht hoher Leistung). Auf Stichprobenebene haben wir damit eine Abweichung gegenüber μ_0 in die erwünschte Richtung, in Richtung der Forschungshypothese. Ist diese Abweichung nur ein zufälliges Ergebnis dieser Stichprobe, oder ist der höhere Leistungswert der Esser des neuen Produkts signifikant?

Das beantwortet der Signifikanztest. Wenn das Stichprobenergebnis $\bar{x} = 11,36$ „kaum“ mit der H_0 vereinbar ist, d.h. wenn $p \leq 5\%$ ist, verwerfen wir die H_0 und akzeptieren die H_1 .

Folgendes ist bekannt:

- Die H_0 -Verteilung hat die Streuung $\sigma_{\bar{x}}$ (vgl. S. 36f.),
- die H_0 -Verteilung entspricht für $n \geq 30$ einer Normalverteilung (vgl. S. 38),
- der Mittelwert der H_0 -Verteilung ist μ (vgl. S. 38), hier: $\mu_0 = 10$, und entspricht der Leistung bei herkömmlicher Ernährung.

Wir fragen, wie wahrscheinlich es ist, dass man die Leistung von $\bar{x} = 11,36$ (oder eine noch bessere) auch mit herkömmlicher Ernährung hätte erreichen können. Diese Wahrscheinlichkeit p entspricht dem schattierten Flächenanteil, den der \bar{x} -Wert in Abbildung 10 (S. 45) rechts von der H_0 -Verteilung „abschneidet“.

Wir erhalten p , indem wir ermitteln, welchen Flächenanteil der zu \bar{x} gehörige z -Wert rechts von der Standardnormalverteilung abschneidet. Dazu transformieren wir in das Stichprobenergebnis, also den empirischen \bar{x} -Wert, in einen empirischen z -Wert der Standardnormalverteilung. Achtung: Es handelt sich jetzt um eine z -Transformation von \bar{x} -Werten einer Verteilung mit dem Mittelwert μ_0 und der Streuung $\sigma_{\bar{x}}$, deshalb benutzen wir folgende Abwandlung der Formel von S. 31:

$$z_{\text{emp}} = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \quad \text{empirischer } z\text{-Wert}$$

Um diese Formel anwenden können, müssen wir zuerst $\sigma_{\bar{x}}$ berechnen:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{8}{10} = 0,8$$

Durch Einsetzen ergibt sich:

$$z_{\text{emp}} = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{11,36 - 10}{0,8} = 1,7$$

In einer Tabelle der Verteilungsfunktion (Flächenanteile) der Standardnormalverteilung (z.B. Tabelle B bei Bortz, 1993, 4. Aufl., S. 694–698 oder Bortz, 1999, 5. Aufl., S. 768–772) finden wir dann den gesuchten Flächenanteil, den der z_{emp} -Wert von der Standardnormalverteilung abschneidet, also die gesuchte Wahrscheinlichkeit p . Der Flächenanteil beträgt in unserem Beispiel laut der Verteilungsfunktion der Standardnormalverteilung: $1 - 0,9554 = 0,0446 = 4,46\%$. Also $p \leq 5\%$, das Ergebnis ist signifikant, so dass wir die H_1 akzeptieren können.

So macht es übrigens das Programm SPSS: Es gibt die exakte Irrtumswahrscheinlichkeit von 0,0446 aus.

Es gibt eine alternative Vorgehensweise, die man meistens wählt, wenn man den Signifikanztest ohne EDV macht: Man vergleicht den empirischen z -Wert mit dem *kritischen* z -Wert, der gerade 5% abschneidet. Falls der empirische z -Wert in der Verteilung weiter außen liegt, heißt das, dass die ihm zugeordnete Wahrscheinlichkeit $p \leq 5\%$ sein muss. Der kritische z -Wert lautet in unserem Beispiel $z_{\text{crit}} = 1,65$ (vgl. Tabelle der Standardnormalverteilung). Der Wert $z_{\text{emp}} = 1,7$ liegt weiter außen, schneidet also weniger als 5% der Verteilung ab, d.h. das Ergebnis ist signifikant.

Wir haben jetzt eine gerichtete Hypothese getestet. Wenn wir eine *ungerichtete* Hypothese testen ($H_1 : \mu_1 \neq \mu_0$; $H_0 : \mu_1 = \mu_0$), machen wir einen *zweiseitigen Test*, d.h. wir teilen die „erlaubte“ Irrtumswahrscheinlichkeit auf beide Seiten der Verteilung auf (*two-tailed probability*). Bei einem zweiseitigen Test auf dem 5%-Niveau sind dann pro Seite maximal

2,5% Irrtumswahrscheinlichkeit erlaubt, bzw. wir definieren als kritische z-Werte diejenigen Werte, die an beiden Seiten der Verteilung jeweils 2,5% abschneiden.

Der hier dargestellte Signifikanztest dient zum Test der Nullhypothese, dass eine Zufallsstichprobe zu einer Population mit einem bekannten Populationsparameter μ_0 gehört. Mit anderen Worten, dass die Population, aus der die Zufallsstichprobe stammt, ebenfalls den Populationsparameter μ_0 hat (und nicht einen davon abweichenden μ_1 , wie es die H_1 behauptet). Dieser Test heißt *z-Test* (auch Gauß-Test genannt), denn er verwendet als *Prüfgröße* den standardnormalverteilten z-Wert. Unter Prüfgröße (engl. *test statistic*) verstehen wir eine Größe, die wir aus dem jeweiligen Stichprobenkennwert (hier: \bar{x}) nach einer bestimmten Formel (hier: z-Transformation in der Version von S. 47) berechnen, und von der wir wissen, welcher theoretischen Verteilung sie folgt (hier: Standardnormalverteilung).

Es gibt es viele verschiedene Signifikanztests, z.B. t-Test, χ^2 -Test, Korrelationstest. Nicht durchführen werden wir z.B. U-Test, Wilcoxon-Test, F-Test oder H-Test. Die verschiedenen Signifikanztests basieren auf verschiedenen Prüfgrößen mit verschiedenen Verteilungen. Welchen Test man auswählt, hängt ab von folgenden Faktoren:

- Art der Hypothese,
- Skalenniveau der Variablen,
- Verteilungseigenschaften der Variablen,
- Stichprobenumfang.

Aufgaben

1. Betrachte das Sportlernahrungs-Beispiel von S. 46f.

- Welcher Fehler wäre hier deiner Meinung nach schwerwiegender: ein α - oder ein β -Fehler?
- Veranschauliche den Signifikanztest von S. 47 grafisch: Zeichne in ein H_0 -Modell und in ein z-standardisiertes H_0 -Modell (letzteres entspricht der Standardnormalverteilung) Folgendes ein:
 - den empirischen \bar{x} -Wert (nur in das H_0 -Modell einzeichnen),
 - den empirischen z-Wert = z-Wert des Stichprobenergebnisses (nur in das z-standardisierte H_0 -Modell einzeichnen),
 - die Irrtumswahrscheinlichkeit p (in beide Modelle einzeichnen),
 - den kritischen \bar{x} -Wert, der 5% abschneidet (nur in das H_0 -Modell einzeichnen),
 - den kritischen z-Wert, der 5% abschneidet (nur in das z-standardisierte H_0 -Modell einzeichnen),
 - das Signifikanzniveau α (in beide Modelle einzeichnen).
- Wie kann man das Ergebnis des Signifikanztests interpretieren?

2. Überlege dir zu den folgenden Fragestellungen die Hypothesen: inhaltliche Alternativhypothese (Forschungshypothese), inhaltliche Nullhypothese, statistische H_1 und statistische H_0 . Beurteile außerdem inhaltlich, ob du jeweils einen α -Fehler oder einen β -Fehler für schwerwiegender halten würdest:
- a) Der Direktor einer kleinen Schule beschließt, den Englischlehrer zu entlassen, da er vermutet, dass die Klassendurchschnitte durch einen neuen Lehrer um mindestens eine Note angehoben werden.
 - b) Unterscheiden sich die Diplomnoten von StudentInnen, die in der Zeit vor den Prüfungen regelmäßig jobben, von den Noten derjenigen StudentInnen, die nicht jobben?
3. Die kritischen Grenzen bei einem z-Test auf dem 5%-Niveau sind
- bei einseitigem z-Test: $+1,65$ (oder, falls der empirische z-Wert negativ ist: $-1,65$),
 - bei zweiseitigem z-Test: $-1,96$ links und $+1,96$ rechts.

Wie lauten die kritischen Grenzen für $\alpha = 1\%$?

- bei einseitigem z-Test:
- bei zweiseitigem z-Test:

Signifikanz und Relevanz

Wenn ein Ergebnis statistisch bedeutsam ist (Signifikanz), bedeutet das noch nicht, dass es auch praktisch bedeutsam ist (Relevanz). Denn: Wenn man sehr große Stichproben bildet, bekommt man auch sehr kleine Effekte (Unterschiede, Zusammenhänge) signifikant. Das liegt am Standardfehler, der sich verkleinert, was wiederum den p-Wert verkleinert. Die als signifikant geprüften Effekte sind dann zwar gegen den Zufall abgesichert, können aber so klein sein, dass sie für die Praxis nicht relevant sind. Deshalb sollte man bei großen Stichproben (grobe Faustregel: bei Stichprobenumfängen um die Tausend) nicht nur einen Signifikanztest machen, sondern auch die Effektgröße berechnen.

Wir müssen hier zwei Verwendungskontexte der Effektgröße unterscheiden:

1. Wie auf S. 42 definiert: Effektgröße als gerade akzeptable Mindestgröße eines behaupteten Zusammenhanges oder Unterschiedes in der Population. D.h. man postuliert sinnvollerweise einen Effekt in einer Größe, die man für praktisch bedeutsam hält.
2. (Darum geht es hier:) Nach Abschluss der Untersuchung kann man die tatsächliche Größe des empirisch auftretenden Effekts berechnen (Berechnung *ex post*). Das sollte man bei großen Stichproben tun. Hierfür gibt es je nach Signifikanztest verschiedene Formeln. Für den t-Test für unabhängige Stichproben (\rightarrow Mittelwertunterschiede) beispielsweise ist „Cohens d“ die Effektgröße. Cohens d = Differenz zwischen den beiden Gruppenmittelwerten dividiert durch die gepoolte Standardabweichung beider Gruppen. Über die einzelnen Formeln, und welche Effektgrößen als klein, mittel und groß einzuschätzen sind, informieren Bortz und Döring (2002, Kap. 9).

Mittelwertunterschiede

Deskriptive und inferenzstatistische Ebene

Wenn wir zwei oder mehr Gruppen miteinander vergleichen wollen (z.B. vier Gruppen, die vier unterschiedlichen Werbespots für ein neues Produkt ausgesetzt werden), betrachten wir diese Gruppen als die Ausprägungen einer unabhängigen Variablen (hier: Werbespot). Das Merkmal, hinsichtlich dessen wir die Gruppen vergleichen, betrachten wir als abhängige Variable (z.B. Einstellung zu dem neuen Produkt auf einer Skala).

Wenn die abhängige Variable metrisch ist, berechnen wir für jede der vier Gruppen einen Mittelwert der abhängigen Variablen. Diese Gruppenmittelwerte können wir dann vergleichen und z.B. in einem Balkendiagramm mit vier Balken visualisieren.

Die so gewonnenen Aussagen sind jedoch zunächst rein *deskriptiv*: Sie gelten nur für die Personenstichprobe, denen man die Werbespots gezeigt hat, und sind nicht auf die KonsumentInnen verallgemeinerbar.

Auf der *inferenzstatistischen* Ebene geht es darum, das Stichprobenergebnis statistisch abzusichern, d.h. – mit einer gewissen Irrtumswahrscheinlichkeit – auszuschließen, dass es sich nur um ein stichprobenbedingtes Zufallsergebnis handelt: Wir testen die Hypothese, dass die Gruppen sich (auch) in der Population unterscheiden. Genauer: Wir testen die Alternativhypothese, dass die (im Beispiel) vier Stichproben aus Populationen mit verschiedenen Populationsparametern μ_1 , μ_2 , μ_3 und μ_4 stammen, bzw. die Nullhypothese, dass sie aus Populationen mit identischen Populationsparametern stammen.

Für solche Unterschiedshypothesen bei metrischen Daten gibt es verschiedene Varianten – je nachdem, um wie viele Gruppen es sich handelt, und ob die Stichproben in einer bestimmten Beziehung zueinander stehen. Im Folgenden werden Verfahren für *zwei* Gruppen behandelt. Bei mehr als zwei Gruppen sind so genannte Varianzanalysen indiziert, die aber nicht zum Stoff dieser Veranstaltung gehören.

t-Tests

Der z-Test (Gauß-Test) diene zum Test der Nullhypothese, dass eine Zufallsstichprobe zu einer Population mit einem bekannten Populationsparameter μ_0 gehört. In der Praxis ist

aber oft kein μ_0 bekannt. Deshalb gibt es Signifikanztests, die nicht eine Zufallsstichprobe mit einem bekannten Populationsparameter, sondern zwei Zufallsstichproben miteinander vergleichen: Die *t-Tests*. Sie heißen so, weil sie als Prüfgröße einen Wert verwenden, der der so genannten t-Verteilung folgt.

Der Begriff *t-Verteilung* steht für eine Familie von symmetrischen und glockenförmigen Verteilungen mit $\mu = 0$. Die Form von t-Verteilungen ähnelt der Form der Standardnormalverteilung. t-Verteilungen sind aber schmalgipfliger, in der Mitte weniger hoch und an den Rändern höher als die Standardnormalverteilung. Die t-Verteilungen unterscheiden sich in einem Parameter: in den Freiheitsgraden (*degrees of freedom*, kurz df). Je mehr Freiheitsgrade eine t-Verteilung hat, desto näher kommt sie der Standardnormalverteilung.

Die t-Werte für bestimmte Flächenanteile bei t-Verteilungen verschiedener Freiheitsgrade sind tabelliert (z.B. Tabelle D bei Bortz, 1993, S. 701, oder Bortz, 1999, S. 775). Wir benutzen diese t-Werte bei den Signifikanztests als kritische t-Werte, analog den kritischen z-Werten beim z-Test.

t-Test für unabhängige Stichproben

Der t-Test für unabhängige Stichproben (*t-test for independent samples*) prüft die H_1 , dass zwei unterschiedliche Mittelwerte aus zwei unabhängigen Stichproben sich auch in den entsprechenden Populationen unterscheiden, dass es sich also nicht nur um einen zufälligen Stichprobenunterschied handelt. Oder andersherum: Er prüft die H_0 , dass die beiden unabhängigen Stichproben aus Populationen mit identischen Parametern μ_1 und μ_2 stammen.

Beispiel: Unterscheidet sich der Fernsehkonsum in Deutschland und Frankreich? Je fünf zufällig ausgewählte Deutsche und Franzosen werden gefragt, wie viele Stunden sie wöchentlich vor dem Fernsehgerät zubringen. Die Daten: Deutschland: 5, 15, 16, 5 und 4 Stunden pro Woche. Frankreich: 6, 7, 12, 9 und 6 Stunden.⁴

Bitte formuliere und teste die Hypothese, indem du im Folgenden alle Leerstellen ausfüllst (beginne mit der Formulierung des Hypothesenpaars H_1/H_0 , und führe dann die Berechnungen durch). Lege dabei ein Signifikanzniveau von $\alpha = 5\%$ zugrunde.

H_1 :

H_0 :

Als Stichprobenkennwert verwenden wir $\bar{x}_1 - \bar{x}_2$, also die *Differenz der Mittelwerte*, und berechnen daraus die Prüfgröße t, die an der t-Verteilung auf Signifikanz getestet wird:

⁴ Die kleinen Stichproben in diesem Beispiel dienen dazu, den Rechenaufwand gering zu halten und entsprechen nicht der Praxis. Wenn man aber nur sehr kleine Stichproben bilden kann, muss man sie sehr gut auswählen und andere, selten durchgeführte Signifikanztests aus der Gruppe der so genannten nonparametrischen oder verteilungsfreien Verfahren anwenden. Für Zweigruppenvergleiche ist dann der U-Test geeignet. Solche Tests gehören aber nicht zum Stoff dieser Veranstaltung.

$$t_{\text{emp}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} \quad \text{vgl. } z_{\text{emp}} = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{x}}} \quad (\text{Analogie!})$$

Wegen der $H_0: \mu_1 - \mu_2 = 0$ vereinfacht sich die Formel für die Prüfgröße t zu:

$$\boxed{t_{\text{emp}} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}}} \quad \text{empirischer t-Wert (bei unabhängigen Stichproben)}$$

$$\text{Dabei ist } \hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2}{(n_1 - 1) + (n_2 - 1)} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Standardfehler der Differenz von Mittelwerten (also die Streuung der Verteilung der Differenzen von Mittelwerten); darin steckt schon die Schätzung der gemeinsamen Populationsvarianz aus den Daten beider Stichproben.

Die Berechnung dieses Standardfehlers wird durch folgendes Rechenschema erleichtert:

x_{i1}	$x_{i1} - \bar{x}_1$	$(x_{i1} - \bar{x}_1)^2$	x_{i2}	$x_{i2} - \bar{x}_2$	$(x_{i2} - \bar{x}_2)^2$
		$\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 =$			$\sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2 =$

Ergebnis: Empirischer t-Wert (Prüfgröße; Ergebnis bitte selbst eintragen): $t_{\text{emp}} =$

Der *kritische t-Wert*, der von der Fläche der t-Verteilung 5% abschneidet, ist in der t-Verteilungs-Tabelle ablesbar. Dabei hängt die Anzahl der Freiheitsgrade von der Anzahl der Messwerte ab. Mit den Freiheitsgraden beschäftigen wir uns hier nicht näher; sie interessieren uns nur, um in der Tabelle die Zeile mit der gesuchten t-Verteilung zu finden.

Freiheitsgrade beim t-Test für unabhängige Stichproben: $df = n_1 + n_2 - 2$

Kritischer t-Wert: $t_{\text{crit}} =$

Entscheidung:

Interpretation:

t-Test für abhängige Stichproben

Wenn eine Stichprobe zweimal untersucht wird, kann man mit dem t-Test für abhängige Stichproben (*t-test for dependent samples, paired samples t-test*) prüfen, ob der Mittelwert sich signifikant verändert hat. Der Test prüft also die Veränderungshypothese, dass zwei Mittelwerte einer Variablen, die bei denselben Personen zu zwei verschiedenen Zeitpunkten gemessen wurde (*Messwiederholung*), sich in der Population unterscheiden bzw. dass der Mittelwert μ_d der einzelnen zeitlichen Differenzen ungleich Null ist (ungerichtete H_1). Der t-Test für abhängige Stichproben wird außerdem benutzt, um *matched samples* zu vergleichen, d.h. die Untersuchungsteilnehmer der beiden Stichproben sind einander paarweise zugeordnet, etwa bei Geschwisterpaaren.

Der t-Test für abhängige Stichproben berücksichtigt in den auf S. 54 angegebenen Formeln, dass die Varianz der einen Messwertreihe von der Varianz der anderen Messwertreihe beeinflusst wird, und unterscheidet sich dadurch vom t-Test für unabhängige Stichproben.

Als Beispiel testen wir die gerichtete Veränderungshypothese, dass eine Kampagne gegen das Rauchen das Rauchverhalten senkt. Eine Zufallsstichprobe von sieben RaucherInnen wird vor und nach der Kampagne gefragt, wie viele Zigaretten sie täglich rauchen:

RaucherIn	vorher x_{i1}	nachher x_{i2}
1.	20	17
2.	18	15
3.	19	20
4.	16	18
5.	18	17
6.	19	18
7.	20	19

Lässt sich für die Population der RaucherInnen behaupten, dass die Kampagne wirkt? Teste die Hypothese auf dem 5%-Niveau. (Die Differenzen der Messwerte sind normalverteilt.)

H_1 :

H_0 :

Als Stichprobenkennwert, aus dem der empirische Wert der Prüfgröße t berechnet wird, verwenden wir \bar{x}_d , den *Mittelwert der Differenzen* der Messwertpaare. Dieser Mittelwert der Differenzen wird wie folgt in eine Prüfgröße t verwandelt:

$$t_{\text{emp}} = \frac{\bar{x}_d - \mu_d}{\hat{\sigma}_{\bar{x}_d}} \quad \text{vgl.} \quad z_{\text{emp}} = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{x}}} \quad (\text{Analogie!})$$

Wegen der H_0 : $\mu_d = 0$ vereinfacht sich die Formel für die Prüfgröße t zu:

$$t_{\text{emp}} = \frac{\bar{x}_d}{\hat{\sigma}_{\bar{x}_d}}$$

empirischer t-Wert (bei abhängigen Stichproben)

Dabei ist $\hat{\sigma}_{\bar{x}_d} = \frac{\hat{\sigma}_d}{\sqrt{n}}$ der geschätzte *Standardfehler des Mittelwertes der Differenzen*, also die Streuung der Verteilung der Mittelwerte von Differenzen (beachte die Analogie zum geschätzten Standardfehler des Mittelwertes: $\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}}$).

Im Zähler steht $\hat{\sigma}_d$, also die geschätzte Streuung der Differenzen in der Population. Sie wird berechnet nach der Formel:

$$\hat{\sigma}_d = \sqrt{\frac{\sum_{i=1}^n d_i^2 - \frac{\left(\sum_{i=1}^n d_i\right)^2}{n}}{n-1}}$$

Die Berechnung der Prüfgröße t wird durch folgendes Rechenschema erleichtert:

RaucherIn	vorher x_{i1}	nachher x_{i2}	Differenz $d_i = x_{i1} - x_{i2}$	quadrierte Differenz d_i^2
1.	20	17		
2.	18	15		
3.	19	20		
4.	16	18		
5.	18	17		
6.	19	18		
7.	20	19		
Summe			$\sum_{i=1}^n d_i =$	$\sum_{i=1}^n d_i^2 =$
Mittelwert			$\bar{x}_d =$	

Ergebnis: Empirischer t-Wert (Prüfgröße; bitte selbst eintragen): $t_{\text{emp}} =$

Beim Heraussuchen des kritischen t-Werts aus t-Verteilungs-Tabelle sind wieder die Freiheitsgrade zu beachten. Sie hängen hier von der Anzahl n der Messwertpaare ab.

Freiheitsgrade beim t-Test für abhängige Stichproben: $df = n_{\text{Paare}} - 1$

Kritischer t-Wert: $t_{\text{crit}} =$

Entscheidung:

Interpretation:

Voraussetzungen für t-Tests

Voraussetzungen, die für die Anwendung des t-Tests für *unabhängige Stichproben* erfüllt sein müssen, sind über das Kardinalskalenniveau hinaus:

- Falls $n_1 + n_2 < 50$ ist, müssen sich die Messwerte in den beiden Populationen normalverteilen.
- Annähernd gleiche Varianzen in den beiden Populationen (Nachprüfung per F-Test; der F-Test wird hier nicht weiter behandelt)
- Unabhängigkeit der Stichproben, d.h. sie haben nichts miteinander zu tun.

Beim t-Test für *abhängige Stichproben*:

- Falls $n < 30$ ist, müssen sich die Differenzen der Messwerte in der Population normalverteilen (n ist dabei die Anzahl der Messwertpaare).

Lineare Zusammenhänge

Lineare Regression

Die Regressionsrechnung dient zur Vorhersage eines Merkmals (Kriterium) aufgrund eines anderen Merkmals (Prädiktor). Sie ist für metrische Merkmale geeignet.

Beispiel: Angenommen, wir fragen vier zufällig ausgewählte Personen nach ihrer Selbstsicherheit (= Merkmal x) und nach der Anzahl ihrer sehr engen Freunde (= Merkmal y). Das Merkmal x wird erhoben durch eine Selbsteinschätzung von 0: „Ich fühle mich sehr unsicher“ bis 10: „Ich fühle mich völlig selbstsicher“. Das Merkmal y wird erhoben durch die Frage: „Wie viele sehr enge Freunde hast du?“ – Die Daten (in der Praxis hat man mehr Messwertpaare, aber zur Veranschaulichung des Rechenweges reichen diese vier):

Person	Selbstsi. x_i	Freunde y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
1.	2	1			
2.	4	5			
3.	1	0			
4.	9	6			
Summe					
Mittel					

Aufgaben

1. Was fällt dir auf? Inwiefern scheinen die beiden Merkmale zusammenzuhängen?
2. Zeichne die vier Datenpaare als Punkte in das Koordinatensystem (nimm die x-Achse für die Selbstsicherheit, die y-Achse für die Freunde).



3. Zeichne eine Gerade ein, die deinem Augenmaß nach den Trend gut wiedergibt.

Scatterplot

Die grafische Darstellung des Zusammenhangs zweier Variablen in einem Koordinatensystem nennt man *Scatterplot* (Streudiagramm). Die x-Achse und die y-Achse des zweidimensionalen Koordinatensystems repräsentieren die Variablen x und y. Die Koordinaten x_i und y_i eines Datenpunktes entsprechen den beiden Messwerten des Objektes i.

Kovarianz

Der Zusammenhang zwischen x und y lässt sich durch die *Kovarianz* quantifizieren. Die Kovarianz ist der Mittelwert aller Produkte von korrespondierenden Abweichungen:

$$\text{cov}_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n} \quad \text{Kovarianz der Variablen x und y}$$

Ein auf der Kovarianz aufbauendes Maß, das den Zusammenhang zwischen x und y quantifiziert, ist die Korrelation (s.u.).

Aufgabe

Berechne mit Hilfe des Datenschemas auf S. 56 die Kovarianz zwischen Selbstsicherheit und Anzahl der sehr engen Freunde.

Regressionsgerade

Die Gerade, die die Punktwolke am besten repräsentiert, veranschaulicht den linearen Zusammenhang zwischen x und y. Diesen Zusammenhang machen wir uns zunutze, um y vorherzusagen, wenn wir nur x kennen („Vorhersage von y aufgrund von x“).

Gesucht wird diejenige Gerade, bei der die quadrierten Abweichungen der vorhergesagten y-Werte (kurz: \hat{y}_i) von den tatsächlichen y-Werten (kurz: y_i) am kleinsten sind (Kriterium der kleinsten Quadrate). Das ist die *Regressionsgerade*. Sie repräsentiert den linearen Zusammenhang zwischen den beiden Merkmalen x und y.

$$\hat{y}_i = b \cdot x_i + a \quad \text{Gleichung der Regressionsgeraden, Regressionsgleichung} \\ \text{(genauer: } \hat{y}_i = b_{yx} \cdot x_i + a_{yx} \text{)}$$

$$b = \frac{\text{cov}_{xy}}{s_x^2} \quad \text{Berechnung von b = Steigung der Regressionsgeraden}$$

$$\text{oder: } b = r \cdot \frac{s_y}{s_x}$$

Berechnung von b, wenn die Korrelation r (s.u.) bekannt ist

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

Berechnung von a = Schnittpunkt mit der y-Achse

Aufgaben

1. Bestimme anhand der Stichprobendaten von S. 56 die Regressionsgleichung zur Vorhersage von y aufgrund von x.
2. Zeichne die Regressionsgerade in das Koordinatensystem auf S. 56 ein.
3. Zeichne auch die Abweichungen der \hat{y}_i -Werte von den y_i -Werten ein!
4. Wie sähe die Punktwolke im Koordinatensystem aus, wenn man bei jedem Messwertpaar den y-Wert durch den vorhergesagten y-Wert (also durch den \hat{y}_i -Wert) ersetzen würde?
5. Wenn kein linearer Zusammenhang zwischen den Variablen besteht, welcher Wert wird dann durch die Regressionsgleichung vorhergesagt? Wie sieht dann die Regressionsgerade aus?
6. Was bedeutet es inhaltlich, wenn zwischen zwei Variablen eine negative Kovarianz besteht? Beispiel!
7. Was bedeutet eine negative Kovarianz für die Steigung der Regressionsgeraden?
8. Wovon hängt die Güte der Vorhersage ab?

Korrelation (deskriptive Ebene)

Die Korrelation ist ein Maß für den Zusammenhang zwischen Merkmalen. Bei Zusammenhängen müssen wir unterscheiden:

- kontingenter Zusammenhang bei Nominaldaten (*Kontingenz*)
- monotoner Zusammenhang bei Ordinaldaten (*Rangkorrelation*)
- linearer Zusammenhang bei Kardinaldaten (*Bravais-Pearson-Korrelation*)

Der häufigste Fall ist ein linearer Zusammenhang zwischen kardinalskalierten Merkmalen: die *Bravais-Pearson-Korrelation*, auch *Produkt-Moment-Korrelation* oder *lineare Korrelation* genannt, kurz *Korrelation*.

Der Zahlenwert für die Bravais-Pearson-Korrelation heißt *Korrelationskoeffizient* und wird durch r symbolisiert. Der Korrelationskoeffizient gibt die Enge und die Richtung des linearen Zusammenhangs zwischen zwei Merkmalen an.

Wertebereich von r: Der Korrelationskoeffizient kann sich zwischen -1 und +1 bewegen.

$r = -1$ perfekter negativer linearer Zusammenhang,

$r = 0$ kein linearer Zusammenhang,

$r = 1$ perfekter positiver linearer Zusammenhang.

Allgemein geben Werte von $r > 0$ einen positiven linearen Zusammenhang an, Werte von $r < 0$ einen negativen linearen Zusammenhang. Je größer der absolute Betrag von r, desto enger ist der lineare Zusammenhang.

Berechnung des Korrelationskoeffizienten:

$$r = \frac{\text{cov}_{xy}}{s_x \cdot s_y} \quad \text{Korrelationskoeffizient r (Korrelation in der Stichprobe)}$$

Durch Einsetzen der Formel für die Kovarianz ergeben sich die äquivalenten, aber weniger leicht zu merkenden Formeln:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot s_x \cdot s_y} \quad \text{oder} \quad r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$r^2 \quad \text{Determinationskoeffizient (Varianzerklärung oder Anteil der gemeinsamen Varianz der beiden Variablen; wird oft in \% angegeben)}$$

Aufgaben

1. Wir lassen die vier Personen (von S. 56) einen psychometrischen Test ausfüllen, der Selbstsicherheit erfasst (= Merkmal w, von 0 Punkte: 'sehr unsicher' bis 50 Punkte: 'völlig selbstsicher') und erhalten die Messwerte 10, 20, 5, 45; Anzahlen der sehr engen Freunde bleiben wie oben. Wie hoch ist *jetzt* die Kovarianz zwischen Selbstsicherheit und Anzahl der sehr engen Freunde?

$$\text{cov}_{wy} =$$

2. Wie hoch ist die Korrelation zwischen Selbstsicherheit und Anzahl der sehr engen Freunde? Berechne r zweimal:
 - a) Nimm als Maß für die Selbstsicherheit das Merkmal x (= Selbsteinschätzung),
 - b) Nimm als Maß für die Selbstsicherheit das Merkmal w (= psychometrischer Test).

$$\text{a) } r_{xy} =$$

$$\text{b) } r_{wy} =$$

3. Mach dir anhand der zweiten auf dieser Seite genannten Korrelationsformel (= die Korrelationsformel, in die die Formel für die Kovarianz eingesetzt wurde) klar, worin der Unterschied zwischen Kovarianz und Korrelation besteht. (Tipp: z-Transformation!)

Korrelationstest (inferenzstatistische Ebene)

Der *Korrelationstest* (Signifikanztest für Korrelationen) überprüft Zusammenhangshypothesen, die sich auf die Korrelation in der Population beziehen. Das Symbol für die Korrelation in der Population ist ρ (griech. rho).

Die ungerichteten statistischen Hypothesen lauten $H_0: \rho = 0$ und $H_1: \rho \neq 0$.

Eine gerichtete H_1 postuliert $\rho < 0$ bzw. > 0 , je nach inhaltlicher Forschungshypothese.

Für den Signifikanztest wird der Stichprobenkennwert r in die folgende t-verteilte Prüfgröße umgewandelt:

$$t_{\text{emp}} = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}} \quad (\text{df} = n-2)$$

Dieser empirische t-Wert wird wie gewohnt mit dem kritischen t-Wert verglichen, der der t-Verteilungs-Tabelle zu entnehmen ist.

Nur wenn zwischen x und y ein signifikanter linearer Zusammenhang besteht, ist es sinnvoll, y -Werte durch x -Werte mit Hilfe einer Regressionsgleichung vorherzusagen!

Voraussetzungen für den Korrelationstest

Die Population muss bivariat normalverteilt sein. Das gilt als erfüllt, wenn (a) das Merkmal x für sich genommen normalverteilt ist und (b) das Merkmal y für sich genommen normalverteilt ist und (c) die Verteilung der zu einem x -Wert gehörenden y -Werte jeweils normal ist und umgekehrt.

Der Korrelationstest ist aber sehr robust gegenüber Verletzungen der Voraussetzungen. Zumindest sollte man optisch (mit einem Scatterplot) überprüfen, ob die Punktwolke ungefähr die Form einer Ellipse hat.

Aufgaben

1. Ist die Korrelation im Beispiel S. 56ff. signifikant? Formuliere eine gerichtete Zusammenhangshypothese, sowohl in Worten als auch statistisch, und teste sie auf einem Signifikanzniveau von $\alpha = 5\%$ (bivariate Normalverteilung kann angenommen werden).
2. Welcher Anteil der Varianz von y ist in unserem Beispiel aufgrund von x vorhersagbar?
3. Was sagt dieses Ergebnis über Kausalität aus?

4. Ein Medienforscher will den evtl. Zusammenhang zwischen Gewaltvideos und Aggression untersuchen. Bei einer Stichprobe von 10 Personen wurde der Konsum von gewalthaltigen Videos (Anzahl pro Woche, Variable X) und das Ausmaß aggressiven Verhaltens (Rating durch Beobachter von 1=niedrig bis 10=hoch, Variable Y) gemessen.

Person	Videos pro Woche x_i	aggressives Verhalten y_i
1.	7	6
2.	8	4
3.	10	9
4.	6	2
5.	7	1
6.	8	5
7.	10	6
8.	9	4
9.	9	8
10.	8	10
	$\Sigma = 82$	$\Sigma = 55$

- Zeichne in ein Koordinatensystem die zehn Punkte und die Gerade, die nach Deinem Augenmaß den Trend am besten wiedergibt!
- Wie hoch ist in dieser Stichprobe der lineare Zusammenhang zwischen den beiden Merkmalen?
- Ist der Zusammenhang signifikant? Formuliere und teste eine gerichtete Hypothese ($\alpha = 5\%$, X und Y sind normalverteilt).
- Berechne die Gleichung für die Regressionsgerade und vergleiche die errechnete mit der graphischen Lösung.
- Welcher Wert für die Aggressionsbereitschaft ist bei einem Konsum von 7 gewalthaltigen Videos pro Woche zu erwarten?

Analyse von Häufigkeiten

In diesem Kapitel geht es um das Überprüfen von Hypothesen, die sich auf nominalskalierte Merkmale beziehen. Es werden *Häufigkeitsunterschiede* im Auftreten bestimmter Merkmalsausprägungen oder deren Kombinationen analysiert. Die dazu verwendeten Tests heißen Chi-Quadrat-Tests, weil die dabei benutzte Prüfgröße der so genannten χ^2 -Verteilung folgt (χ : griech. chi).

Es gibt ein- und zweidimensionale Chi-Quadrat-Tests. Die eindimensionalen Tests dienen der Analyse nur *eines* Merkmals; der zweidimensionale Chi-Quadrat-Test wird verwendet, wenn wir *zwei* nominalskalierte Merkmale *gleichzeitig* untersuchen.

Prinzip aller Chi-Quadrat-Tests: Die empirisch beobachteten Häufigkeiten werden mit den gemäß der H_0 erwarteten Häufigkeiten verglichen. Aus den Abweichungen wird die Prüfgröße χ^2 berechnet, deren empirischer Wert an der χ^2 -Verteilung (Tabelle z.B. bei Bortz, 1993, S. 699f., oder Bortz, 1999, S. 773f.) auf Signifikanz getestet wird.

Eindimensionale Chi-Quadrat-Tests

Allgemeiner eindimensionaler χ^2 -Test

Der allgemeine eindimensionale χ^2 -Test vergleicht die Häufigkeiten der k Kategorien eines nominalskalierten Merkmals und testet Häufigkeitsunterschiede bzw. Abweichungen der beobachteten von den erwarteten Häufigkeiten auf Signifikanz.

Beispiel (nach Bortz, 1999): In einem Warenhaus soll untersucht werden, ob vier Produkte unterschiedlich häufig verkauft werden bzw. ob die Verkaufszahlen zueinander sich anders verhalten, als man es erwarten könnte (vier Produkte, also Anzahl der Kategorien $k = 4$). Es wurden $n = 400$ Verkäufe registriert:

Produkt 1	Produkt 2	Produkt 3	Produkt 4	
70	120	110	100	(Verkaufszahlen)

Man kann mit χ^2 -Tests anhand derselben Daten verschiedene Fragestellungen untersuchen, d.h. verschiedene Varianten der H_0 testen.

Erste Fragestellung: Wir prüfen, ob die Verkaufszahlen der vier Produkte in unserem Warenhaus sich signifikant voneinander unterscheiden ($\alpha = 5\%$), oder ob die Unterschiede nur zufällig zustande gekommen sind. Die entsprechende H_0 lautet: Gleichverteilung, d.h. gleich häufiger Verkauf der vier Produkte in der „Population“ aller Verkäufe des Warenhauses.

Der empirische Wert der Prüfgröße χ^2 wird wie folgt berechnet:

$$\chi^2_{\text{emp}} = \sum_{j=1}^k \frac{(f_{b(j)} - f_{e(j)})^2}{f_{e(j)}} \quad \text{empirischer } \chi^2\text{-Wert}$$

Dabei ist $f_{b(j)}$ die *beobachtete* Häufigkeit in der Kategorie j
und $f_{e(j)}$ die gemäß der H_0 *erwartete* Häufigkeit in der Kategorie j .

Die erwarteten Häufigkeiten werden wie folgt berechnet:

$$f_{e(j)} = p_j \cdot n$$

n ist die Gesamtanzahl der beobachteten Fälle;
 p_j ist die Wahrscheinlichkeit für die Kategorie j bei Gültigkeit der H_0 .

In unserem Beispiel (Gleichverteilung auf vier Kategorien) ist für jede Kategorie $p_j = 1/4$.

Die Berechnung χ^2 -Wertes ist mit einem Datenschema einfach. Vervollständige das Datenschema und berechne den empirischen χ^2 -Wert als Summe der letzten Spalte:

Kategorie/ Feld Nr. j	$f_{b(j)}$	$f_{e(j)}$	$f_{b(j)} - f_{e(j)}$	$(f_{b(j)} - f_{e(j)})^2$	$\frac{(f_{b(j)} - f_{e(j)})^2}{f_{e(j)}}$
1	70	100	-30	900	9
2					
3					
4					

Freiheitsgrade (bitte im Folgenden selbst ausfüllen): $df = k - 1 =$

Kritischer χ^2 -Wert: $\chi^2_{\text{crit}} =$

Entscheidung und Interpretation:

Zweite Fragestellung (immer noch Beispiel nach Bortz, 1999): Sind die Verkaufszahlen der Produkte in unserem Warenhaus anders als in der Stadt allgemein? Die entsprechende H_0 lautet: Die Verkaufszahlen verteilen sich auf die Produkte wie in der Stadt allgemein, nämlich folgendermaßen:

Produkt 1	Produkt 2	Produkt 3	Produkt 4	
560	680	640	700	(Verkaufszahlen der Stadt)

Wir prüfen bei dieser zweiten Variante nicht, ob die Produkte unterschiedlich häufig verkauft werden, sondern ob die Verkaufszahlen unseres Warenhauses sich signifikant von den allgemeinen Verkaufszahlen in der Stadt unterscheiden ($\alpha = 5\%$). Die erwarteten Häufigkeiten für die Produkte (immer noch unseres Warenhauses!) richten sich jetzt nach den anteilmäßigen Verkaufszahlen der Stadt, z.B. für Produkt 1:

$$p_1 = 560/2580 = 0,22 \text{ und deshalb } f_{e(1)} = p_1 \cdot n = 0,22 \cdot 400 = 88.$$

Vervollständige wieder das Datenschema und berechne den empirischen χ^2 -Wert:

Kategorie/ Feld Nr. j	$f_{b(j)}$	$f_{e(j)}$	$f_{b(j)} - f_{e(j)}$	$(f_{b(j)} - f_{e(j)})^2$	$\frac{(f_{b(j)} - f_{e(j)})^2}{f_{e(j)}}$
1	70	88	-18	324	3,68
2					
3					
4					

Freiheitsgrade (im Folgenden bitte wieder selbst ausfüllen): $df = k - 1 =$

Kritischer χ^2 -Wert: $\chi^2_{crit} =$

Entscheidung und Interpretation:

Goodness-of-fit- χ^2 -Test (χ^2 -Anpassungstest)

Im Prinzip wie beim vorigen Abschnitt, aber: Die empirische Verteilung (beobachtete Häufigkeiten) wird daraufhin überprüft, ob sie sich einer theoretischen Verteilung anpasst, z.B. einer Normalverteilung. Dazu muss das gemessene Merkmal metrisch (!) sein.

Die Bedeutung dieses Tests liegt darin, dass viele Signifikanztests normalverteilte Merkmale voraussetzen. Diese Voraussetzung kann mit dem Goodness-of-fit- χ^2 überprüft werden.

Bei der Überprüfung der H_0 , dass die Messwerte normalverteilt sind, geht man folgendermaßen vor: Zunächst werden die Messwerte in Kategorien aufgeteilt; man kann dann die empirischen (beobachteten) Häufigkeiten der einzelnen Kategorien auszählen. Die Wahrscheinlichkeit p_j , dass ein Messwert in eine bestimmte Kategorie j fällt, wird anhand der z-transformierten Kategoriengrenzen und der entsprechenden Standardnormalverteilungs-Flächenanteile ermittelt. Diese Wahrscheinlichkeiten dienen als Basis für die erwarteten Häufigkeiten, die wieder mit der Formel $f_{e(j)} = p_j \cdot n$ berechnet werden. – Besonderheiten beim Goodness-of-fit- χ^2 -Test: Die Freiheitsgrade betragen $df = k - 3$. Außerdem ist beim Ermitteln des kritischen χ^2 -Wertes zu beachten, dass man die H_0 (Normalverteilung) beibehalten möchte. Deshalb muss man sich gegen einen Beta-Fehler absichern, was man dadurch erreicht, dass man das Signifikanzniveau auf $\alpha = 25\%$ setzt. Bei $p > 25\%$ kann man dann von einer Normalverteilung der Messwerte ausgehen.

Hinweis für die Datenanalyse mit SPSS: Eine andere Möglichkeit zur Überprüfung auf Normalverteilung ist der im Programmpaket enthaltene Kolmogoroff-Smirnov-Test.

Ein dichotomes Merkmal bei zweimaliger Untersuchung (McNemar- χ^2 -Test)

Mit dem McNemar-Test werden die Häufigkeiten bei einem dichotomen Merkmal zu zwei Zeitpunkten (vorher und nachher) untersucht. Es wird geprüft, ob die Häufigkeitsverteilung sich signifikant verändert (*significance of change*).

Beispiel: Zur Prüfung der Effektivität einer Kampagne für Kondomverwendung wurden 60 Personen vor und nach der Kampagne gefragt, ob sie Kondome benutzen. Vor der Werbekampagne gaben 20 Personen an, Kondome zu benutzen; nach der Kampagne 28 Personen. 15 Personen sagten sowohl vorher als auch nachher, dass sie Kondome verwenden.

Trage die Häufigkeiten bitte in das Datenschema ein. Formuliere dann in Worten (a) die ungerichtete Forschungshypothese und (b) die gerichtete Forschungshypothese, und teste beide auf einem Niveau von $\alpha = 5\%$.

		nachher	
		+	-
vorher	+	a	b
	-	c	d

+ = Kondombenutzung
- = keine Kondombenutzung

H_0 (beim McNemar-Test immer ungerichtet): Die Veränderungen sind rein zufällig, d.h. es gibt genauso viele Veränderungen von Plus nach Minus wie umgekehrt.

Deshalb erwarten wir, dass in den Feldern b und c gleich viele Personen sind; wir erwarten also die Häufigkeiten $f_{e(b)} = f_{e(c)} = \frac{b+c}{2}$.

Der empirische Testwert ergibt sich durch Einsetzen der erwarteten Häufigkeiten in die normale χ^2 -Formel von S. 63. Durch Ausmultiplizieren und Zusammenfassen resultiert (bitte Zahlenwerte des Beispiels einsetzen):

$$\chi^2_{\text{emp}} = \frac{(b-c)^2}{b+c} =$$

Freiheitsgrade: $df = 1$

Bei der Benutzung der χ^2 -Tabelle ist zu beachten, dass sie anders konstruiert ist als die übrigen Tabellen, nämlich extra für *ungerichtete* Hypothesen. Man nimmt dann den kritischen χ^2 -Wert, der von der χ^2 -Verteilung eine Fläche von α abschneidet (nicht wie sonst $\alpha/2$).

Bei *gerichteten* Hypothesen, was bei χ^2 -Tests selten vorkommt, nimmt man den Wert, der eine Fläche von $2 \cdot \alpha$ abschneidet. Für alle χ^2 -Tests gilt: Grundsätzlich lassen sich gerichtete Hypothesen nur bei $df = 1$ testen.

Kritische χ^2 -Werte (bitte selbst ausfüllen):

a) falls man eine ungerichtete Hypothese testet: $\chi^2_{\text{crit}} =$

b) falls man eine gerichtete Hypothese testet: $\chi^2_{\text{crit}} =$

Entscheidung und Interpretation:

Zweidimensionale Häufigkeitstabellen: Kreuztabellen (deskriptive Ebene)

Kreuztabellen (*crosstabulations*) stellen die Häufigkeitsverteilung zweier nominalskalierter Merkmale deskriptiv dar. (Die inferenzstatistische Überprüfung geschieht mit dem zweidimensionalen χ^2 -Test, s.u.).

Beispiel: 240 WählerInnen werden nach ihrer Parteipräferenz und nach ihrer Einstellung zur Ökosteuer gefragt. Die resultierenden Häufigkeiten sind in Tabelle 1a zu sehen (fiktive Daten).

Tabelle 1a Kreuztabelle Ökosteuer x Partei: Häufigkeiten (count)

	SPD	CDU	Grüne	FDP	total
dagegen	31	72	1	13	117
dafür, wie jetzt	9	2	3	1	15
dafür, zweckgebunden	64	24	16	4	108
total	104	98	20	18	240

Den Randsummen (total) ist zu entnehmen, dass insgesamt 117 der befragten Personen gegen die Ökosteuer sind, 15 wollen die Ökosteuer und ihre jetzige Verwendung beibehalten, und 108 Personen wollen eine zweckgebundene Ökosteuer (Förderung erneuerbarer Energiequellen und öffentlicher Verkehrsmittel). Insgesamt wurden 104 SPD-SympathisantInnen befragt, 98 CDU-SympathisantInnen etc.

Die zwölf Felder in der Mitte geben die Häufigkeiten der Merkmalskombinationen („Kreuzungen“) wieder. Die Schattierung in Tabelle 1b zeigt z.B., dass 24 der Befragten die CDU und zugleich eine zweckgebundene Ökosteuer präferieren.

Tabelle 1b Kreuztabelle Ökosteuer x Partei: Häufigkeiten (count)

	SPD	CDU	Grüne	FDP	total
dagegen	31	72	1	13	117
dafür, wie jetzt	9	2	3	1	15
dafür, zweckgebunden	64	24	16	4	108
total	104	98	20	18	240

Tabelle 2 enthält zusätzlich zu den absoluten auch die relativen Häufigkeiten (Prozentwerte), bezogen auf die Gesamtanzahl der Befragten, $n = 240$:

Tabelle 2 Kreuztabelle Ökosteuer x Partei:
Häufigkeiten und Gesamtprozentwerte (count and total percentage)

	SPD	CDU	Grüne	FDP	total
dagegen	31 12,9%	72 30,0%	1 0,4%	13 5,4%	117 48,8%
dafür, wie jetzt	9 3,8%	2 0,8%	3 1,3%	1 0,4%	15 6,3%
dafür, zweckgebunden	64 26,7%	24 10,0%	16 6,7%	4 1,7%	108 45,0%
total	104 43,3%	98 40,8%	20 8,3%	18 7,5%	240 100,0%

Oft ist es aber inhaltlich sinnvoll, die Prozentwerte nicht auf den gesamten Stichprobenumfang (hier $n = 240$) zu beziehen, sondern auf die jeweiligen Merkmalsausprägungen, sprich auf die Zeilen bzw. Spalten. Das wird in Tabelle 3 bzw. Tabelle 4 gemacht. Tabelle 3 enthält die Zeilenprozentwerte. Man sieht dort z.B., dass von den 117 Befragten, die gegen die Ökosteuer sind, 61,5% die CDU wählen, aber nur 0,9% die Grünen.

Tabelle 3 Kreuztabelle Ökosteuer x Partei:
Häufigkeiten und Zeilenprozentwerte (count and row percentage)

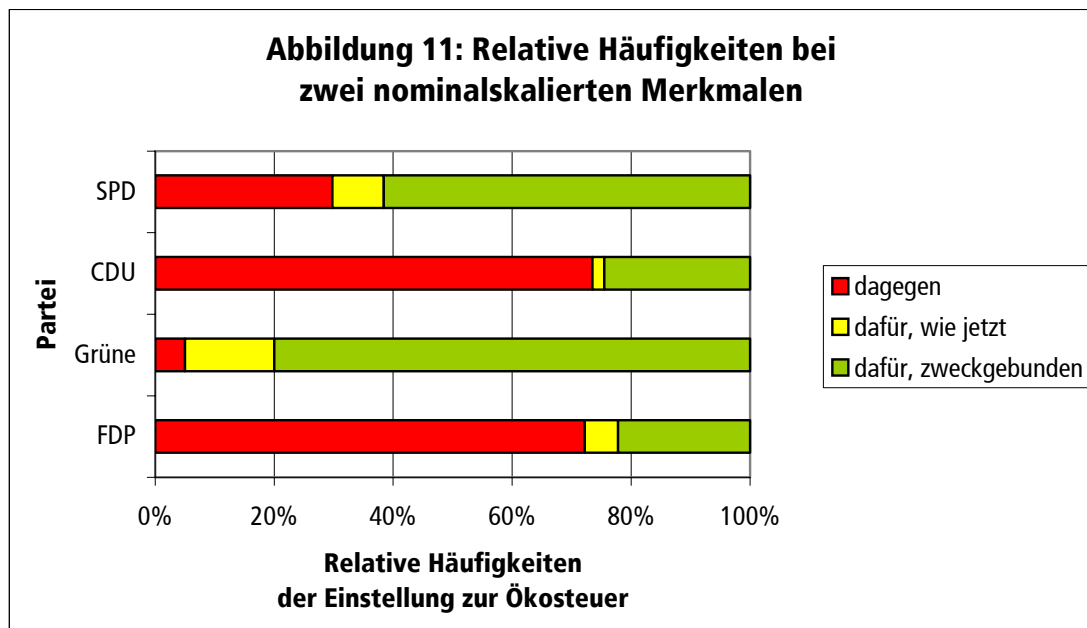
	SPD	CDU	Grüne	FDP	total
dagegen	31 26,5%	72 61,5%	1 0,9%	13 11,1%	117 100,0%
dafür, wie jetzt	9 60,0%	2 13,3%	3 20,0%	1 6,7%	15 100,0%
dafür, zweckgebunden	64 59,3%	24 22,2%	16 14,8%	4 3,7%	108 100,0%
total	104 43,3%	98 40,8%	20 8,3%	18 7,5%	240 100,0%

In Tabelle 4 mit den Spaltenprozentwerten sieht man z.B., dass von den befragten Grünen-Wählern 80% für eine zweckgebundene Ökosteuer sind, aber nur 15% für eine Ökosteuer in der jetzigen Form.

Tabelle 4 Kreuztabelle Ökosteuer x Partei:
Häufigkeiten und Spaltenprozentwerte (count and column percentage)

	SPD	CDU	Grüne	FDP	total
dagegen	31 29,8%	72 73,5%	1 5,0%	13 72,2%	117 48,8%
dafür, wie jetzt	9 8,7%	2 2,0%	3 15,0%	1 5,6%	15 6,3%
dafür, zweckgebunden	64 61,5%	24 24,5%	16 80,0%	4 22,2%	108 45,0%
total	104 100,0%	98 100,0%	20 100,0%	18 100,0%	240 100,0%

Eine Möglichkeit, Tabelle 4 zu visualisieren, ist ein gestapeltes Balkendiagramm (Abbildung 11):



Um zu untersuchen, ob die Einstellung zur Ökosteuer in den verschiedenen Parteien unterschiedlich ist, bzw. ob sie mit der Partei zusammenhängt (was gleichbedeutend ist), ver-

gleicht man die empirisch vorgefundenen, d.h. beobachteten Häufigkeiten mit denjenigen Häufigkeiten, die unter der Annahme zu *erwarten* sind, dass die beiden Merkmale „Einstellung“ und „Partei“ *unabhängig* voneinander seien. Tabelle 5 enthält in Klammern diese erwarteten Häufigkeiten (zu deren Berechnung s. S. 71).

Tabelle 5 Kreuztabelle Ökosteuer x Partei:
beobachtete (und erwartete) Häufigkeiten [observed (and expected) count]

	SPD	CDU	Grüne	FDP	total
dagegen	31 (50,7)	72 (47,8)	1 (9,8)	13 (8,8)	117
dafür, wie jetzt	9 (6,5)	2 (6,1)	3 (1,3)	1 (1,1)	15
dafür, zweckgebunden	64 (46,8)	24 (44,1)	16 (9,0)	4 (8,1)	108
total	104	98	20	18	240

Die Differenz zwischen den vorgefundenen und den erwarteten Häufigkeiten ist – wie im eindimensionalen Fall – die Basis für den χ^2 -Wert, den wir für den Signifikanztest (hier: zweidimensionaler χ^2 -Test) brauchen.

Zweidimensionaler Chi-Quadrat-Test

Der zweidimensionale χ^2 -Test prüft Häufigkeitsunterschiede bei zwei Merkmalen bzw. die Abweichungen der beobachteten von den erwarteten Häufigkeiten auf Signifikanz.

Er heißt auch $k \cdot \ell$ - χ^2 -Test, weil wir ein Merkmal mit k Kategorien und eines mit ℓ Kategorien haben.

Die zu testende *Unterschiedshypothese* lautet, dass k Stichproben sich unterschiedlich (bzw. nicht so, wie man es erwartet) auf die ℓ Ausprägungen eines nominalskalierten Merkmals verteilen. Die Hypothese ist auch als *Zusammenhangshypothese* zweier nominalskalierten Merkmale formulierbar: H_1 : Die beiden Merkmale sind abhängig voneinander bzw. hängen zusammen (*Kontingenz*, vgl. die Einteilung auf S. 58). H_0 : Die beiden Merkmale sind unabhängig voneinander bzw. hängen nicht miteinander zusammen.

Der empirische χ^2 -Wert wird nach folgender Formel berechnet:

$$\chi_{\text{emp}}^2 = \sum_{i=1}^k \sum_{j=1}^{\ell} \frac{(f_{b(i,j)} - f_{e(i,j)})^2}{f_{e(i,j)}}$$

Durch die beiden Summenzeichen läuft die Summierung über die Zeilen und die Spalten. Das Vorgehen ist aber dasselbe wie beim eindimensionalen χ^2 -Test, nur dass wir im Parteien-Beispiel zwölf Felder haben, während es im Warenhaus-Beispiel vier waren.

Die Berechnung der erwarteten Häufigkeiten hängt davon ab, ob die Wahrscheinlichkeiten aus den Daten geschätzt oder durch die H_0 vorgegeben werden; meistens ist Ersteres der Fall.

$$f_{e(i,j)} = \frac{\text{Zeilensumme } i \cdot \text{Spaltensumme } j}{n}, \text{ wenn Schätzung aus den Daten,}$$

$$f_{e(i,j)} = p_{i,j} \cdot n, \text{ wenn die } p_{i,j} \text{ durch die } H_0 \text{ bereits vorgegeben sind.}$$

Freiheitsgrade:

$$df = (k - 1) \cdot (\ell - 1), \text{ wenn Schätzung der Wahrscheinlichkeiten aus den Daten,}$$

$$df = k \cdot \ell - 1, \text{ wenn die Wahrscheinlichkeiten durch die } H_0 \text{ vorgegeben sind.}$$

Voraussetzungen für χ^2 -Tests

- Nominaldaten (Ausnahme: Metrische Daten beim Goodness-of-fit- χ^2 -Test).
- Jede untersuchte Einheit muss eindeutig einer Merkmalskategorie (bzw. bei zweidimensionalen Tests: einer Kombination von Merkmalskategorien) zuzuordnen sein.
- Minimum für erwartete Häufigkeiten: $f_e \geq 5$

(wenn $df = 1$ ist, benötigen wir sogar $f_e \geq 10$)

Diese Regel muss nicht streng eingehalten werden, da χ^2 -Tests relativ robust sind. Das angegebene Minimum für erwartete Häufigkeiten darf aber in höchstens 20% der Felder unterschritten werden.

Aufgaben

1. Sind beim Parteien-Beispiel die Voraussetzungen für einen χ^2 -Test erfüllt?

2. Berechne anhand der beobachteten und der erwarteten Häufigkeiten (Tabelle 5, S. 70) den empirischen χ^2 -Wert.
3. Nimm an, die 240 WählerInnen wurden zufällig ausgewählt, um ein Meinungsbild der Deutschen zu erhalten. Teste die Hypothese, dass die Einstellung der Deutschen zur Ökosteuer mit ihrer Parteipräferenz zusammenhängt ($\alpha = 1\%$).
4. Ein privater Fernsehsender lässt untersuchen, ob das Anschauen bzw. Nichtanschauen seines Programms vom Geschlecht der jeweiligen Fernsehzuschauer abhängt. (Dies ist ein Beispiel für ein *4-Felder- χ^2* , so nennt man ein $k \cdot \ell$ - χ^2 mit zwei dichotomen Merkmalen.)

An einer Stichprobe von 100 Personen werden folgende Häufigkeiten festgestellt:

	männlich	weiblich
sieht das Programm	10	40
sieht das Programm nicht	20	30

Formuliere die Alternativhypothese und teste sie auf dem 1%-Niveau.

Literatur und Websites

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2003). Multivariate Analysemethoden. Eine anwendungsorientierte Einführung (10. Aufl.). Berlin: Springer.

Praxisorientiertes, einführendes Standardwerk in Methoden wie Faktorenanalyse, Clusteranalyse und Multidimensionale Skalierung, die u.a. in der Marktforschung angewandt werden. Mit SPSS-Beispielen. Die mathematischen Hintergründe werden nur oberflächlich dargestellt. Didaktisch sehr gut aufbereitet.

Bortz, J. (1999). Statistik für Sozialwissenschaftler (5. Aufl.). Berlin: Springer.
(oder 4. Aufl. von 1993)

Einführung in elementarstatistische, in varianzanalytische sowie in multivariate Methoden. Mit SPSS-Beispielen. Durchgängig wird Wert auf die Anwendungsweisen der Verfahren gelegt. Die mathematischen Hintergründe werden dargestellt, trotzdem auch für Nicht-Mathematiker verständlich, didaktisch sehr gut aufbereitet. Die Kapitel 1 bis 6 enthalten die Themen der Veranstaltung „Statistik“ im Studiengang GWK. Der Stoff wird jedoch in unserer Veranstaltung weniger tiefgehend behandelt als im Bortz-Buch, d.h. aus den Kapiteln 1 bis 6 des Bortz-Buchs müsst ihr nicht alles wissen, sondern nur das, was im Seminar behandelt wird. Tipp: Zunächst das Skript durcharbeiten. Dann ggf. vertiefend im Bortz-Buch nachlesen.

Fahrmeir, L., Künstler, R., Pigeot, I. & Tutz, G. (2003). Statistik. Der Weg zur Datenanalyse (4. Aufl.). Berlin: Springer.

Standardwerk zur Einführung in die Statistik. Multivariate Verfahren werden aber so gut wie gar behandelt.

Kennedy, G. (1993). Einladung zur Statistik (2. Aufl.). Frankfurt/M.: Campus.

Führt in die Denkweise der Statistik ein, sympathisches Buch, macht Appetit auf Statistik und ist deshalb gut zur Einführung geeignet. Reicht allerdings für die Veranstaltung „Statistik“ bei GWK nicht aus.

Kühnel, S. M. & Krebs, D. (2001). Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen. Reinbek: Rowohlt.

Nicht umfassend, aber zur Einführung geeignet.

Lane, D. M. (1993–2003). HyperStat Online Textbook.
URL: <http://davidmlane.com/hyperstat/>

Sehr gutes Online-Statistikbuch in 14 Kapiteln, mit interaktiven Simulationen/Demonstrationen, Glossar und Übungen.

Ludwig-Mayerhofer, W. (1998–2004). ILMES – Internet-Lexikon der Methoden der empirischen Sozialforschung. URL: <http://www.lrz-muenchen.de/~wlm/ilmes.htm>

sehr zu empfehlen

Pitman, J. (1993). Probability. New York: Springer.

Wie der Titel sagt: Wahrscheinlichkeitsrechnung (also nicht als Statistik-Buch geeignet). Aus diesem Buch werden hier v.a. Abbildungen für Folien verwendet.

StatSoft, Inc. (1984–2004). Electronic Statistics Textbook.
URL: <http://www.statsoft.com/textbook/stathome.html>

Für die verschiedenen statistischen Verfahren werden jeweils Prinzip und Anwendungsmöglichkeiten dargestellt. Mit Glossar.

SurfStat.australia. An Online Text in Introductory Statistics (1994–2004).
URL: <http://www.anu.edu.au/nceph/surfstat/surfstat-home/surfstat.html>

Auf dieser Seite gibt es ein sehr empfehlenswertes Glossar.

Stelzl, I. (1982). Fehler und Fallen der Statistik. Bern: Huber.

Zeigt an Beispielen, wie statistische Methoden wissentlich oder unwissentlich falsch angewendet werden.

SPSS (Statistik-Software)

Bühl, A. & Zöfel, P. (2002). SPSS Version 11. Einführung in die moderne Datenanalyse unter Windows. München: Pearson.

Der Rahmen: Forschungsmethoden (hier ist nur Literatur zu quantitativen Methoden aufgelistet)

Bortz, J. & Döring, N. (2002). Forschungsmethoden und Evaluation (3. Aufl.). Berlin: Springer. (oder die 2. Aufl. von 1995)

Umfassend, gut zur Einführung, als Überblick, zum Nachschlagen. Auch qualitative Methoden werden behandelt (wenn auch nicht ausführlich). Spezielle Literaturhinweise zu qualitativen Forschungsmethoden gibt es bei mir auf Anfrage.

Diekmann, A. (2000). Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen (6. Aufl.). Reinbek: Rowohlt.

Gut lesbar, qualitative Forschung kommt allerdings zu kurz.

Ludwig-Mayerhofer, W. (1998–2004). ILMES – Internet-Lexikon der Methoden der empirischen Sozialforschung. URL: <http://www.lrz-muenchen.de/~wlm/ilmes.htm>

sehr zu empfehlen

Popper, K. R. (1982). Logik der Forschung (7. Aufl.). Tübingen: Mohr. (Die 1. Aufl. ist 1934 erschienen.)

Grundlegung des Kritischen Rationalismus

Prim, R. & Tilmann, H. (1989). Grundlagen einer kritisch-rationalen Sozialwissenschaft (6. Aufl.). Heidelberg: Quelle & Meyer.

Einführung in die empirische Forschung nach dem Modell des Kritischen Rationalismus (Popper).

Schnell, R., Hill, P. B. & Esser, E. (1999). Methoden der empirischen Sozialforschung. München: Oldenbourg.

Gut strukturiertes Standardwerk. Qualitative Forschung wird allerdings unterschlagen.