

Dieses Skript richtet sich nach den Anforderungen, die die Abteilung Methodenlehre & Diagnostik für das Vordiplom stellt. Das Skript erhebt keinen Anspruch auf Vollständigkeit und Richtigkeit, ist aber sicher eine gute Hilfe zum Verständnis der Thematiken sowie zur Prüfungsvorbereitung.

Bortz: Statistik

SKRIPTLISTE:				
Grundstudium:		erstellt	Hauptstudium:	
		erstellt		
B – Methodenlehre (+ Bortz)	2002			
C – Allgemeine Psychologie 1	2001			
D – Wahrnehmung & Denken	2002			
KEINE weiteren Skripte geplant			weitere Skripte in Vorbereitung	

Viel Spaß!

Manuel Ulrich

Zur Überprüfung: 151 Seiten

Erstellt wurde das Skript Januar bis April 2002.

Inhaltsverzeichnis:

Diverses: S. 5

(Konzeption des Skriptes; Literaturliste; griechisches Alphabet; Warnhinweis)

Statistik (B2):

Deskriptive Statistik: S. 8

(Skalenniveaus; Maße der zentralen Tendenz: Modus, Median & Arithmetisches Mittel; Dispersionsmaße: Variationsbreite und Perzentile, Quadratsummen, Varianz und Standardabweichung; z-Werte)

Merkmalszusammenhänge & -vorhersagen: S. 12

(Kovarianz und Korrelation; Korrelation und Kausalität; Lineare Regression; Beobachtete, vorhergesagte und Residualwerte; Determinationskoeffizient)

Wahrscheinlichkeitstheorie und -verteilungen: S. 16

(Zufallsexperimente; Grundbegriffe der Wahrscheinlichkeitsrechnung; Axiome der Wahrscheinlichkeitstheorie; Das Theorem von Bayes; Variationen, Permutationen, Kombinationen; Verteilungsfunktionen; Diskrete Verteilungen; Stetige Verteilungen 1: Normalverteilung & Standardnormalverteilung; Stetige Verteilungen 2: Chi²-Verteilung, t-Verteilung sowie F-Verteilung; Zusammenhänge von NV, STNV, Chi²-, t- und F-Verteilung)

Stichprobe und Grundgesamtheit: S. 27

(Begriffe der Inferenzstatistik; Stichprobenarten; Verteilung von Stichprobenkennwerten; Kriterien der Parameterschätzung: Erwartungstreue, Konsistenz, Effizienz, Suffizienz; Methoden der Parameterschätzung: Methode der kleinsten Quadrate, Maximum-likelihood-Methode; Intervallschätzung: Wertebereich, Konfidenzintervalle; Bedeutung des Stichprobenumfangs)

Formulierung und Überprüfung von Hypothesen: S. 35

(Alternativhypothesen; Varianten von Alternativhypothesen; Die Nullhypothese; Fehlerarten bei statistischen Entscheidungen: Alpha- & Beta-Fehler; Signifikanzaussagen; Bestimmung der Irrtumswahrscheinlichkeit; Einseitige und zweiseitige Tests; Statistische Signifikanz und praktische Bedeutsamkeit; Effektgröße; Der Beta-Fehler; Indifferenzbereiche; Teststärke („power“); Bedeutung der Stichprobengröße; Monte-Carlo-Studien, Bootstrap-Technik)

Verfahren zur Überprüfung von Unterschiedshypothesen: S.44

(Verfahren für Intervalldaten; z-Test, t-Test; t-Test für unabhängige Stichproben; t-Test für abhängige Stichproben; chi²-Test für eine Varianz; F-Test; Verfahren für Ordinaldaten; Randomisationstest nach Fisher; U-Test; Wilcoxon-Test; Verfahren für Nominaldaten; Vergleich der Häufigkeit eines zweifach gestuften Merkmals; Vergleich der Häufigkeit eines k-fach gestuften Merkmals; 4-Felder-chi²-Test; k x l-chi²-Test; Konfigurationsanalyse; Allgemeine Bemerkungen zu den chi²-Techniken)

Verfahren zur Überprüfung von Zusammenhangshypothesen: S.65

(Die Statistische Absicherung von linearen Regressionen; Determinanten der Vorhersagegenauigkeit; Nonlineare Regression; Interpretationshilfen für r; k-fach gestufte Merkmale; Korrelation für nonlineare Zusammenhänge; Fishers Z-Transformation; Spezielle Korrelationstechniken)

Varianzanalyse (B3):

Einfaktorielle Versuchspläne: S. 75

(Begriffe der Varianzanalyse; Grundprinzip der einfaktoriellen Varianzanalyse; Quadratsummenzerlegung; Varianzschätzungen; Entscheidung; Effektgrößen; Varianzaufklärung; Voraussetzungen; Einzelvergleiche; Alpha-Fehler-Korrektur; Scheffé-Test; Trendtests; Zusammenhänge; Rangvarianzanalyse)

Mehrfaktorielle Versuchspläne: S. 88

(Zweifaktorielle Varianzanalyse; Quadratsummen; Freiheitsgrade und Varianzschätzungen; Hypothesenprüfung; Varianzaufklärung; Interaktionsdiagramme; Feste und zufällige Faktoren; Effektgrößen und Einzelvergleiche; Drei- und mehrfaktorielle Varianzanalysen; Ungleiche Stichprobengrößen; Voraussetzungen)

Versuchspläne mit Messwiederholungen: S. 97

(Einfaktorielle Varianzanalyse mit Messwiederholung; Datenschema; Hypothesen; Quadratsummen; Varianzschätzungen; Prüfung der Nullhypothese; Trendtests und Einzelvergleiche; Mehrfaktorielle Varianzanalyse mit Messwiederholung; Mathematik total; Voraussetzungen)

Kovarianzanalyse: S. 103

(Einfaktorielle Kovarianzanalyse; Datenschema; Hypothesen; Quadratsummen; Freiheitsgrade und Varianzschätzungen; Hypothesenprüfung; Unterschiedliche Stichprobengrößen, A-priori Einzelvergleiche; Effektgrößen; Zusammenfassung; Voraussetzungen; Mehrfaktorielle Kovarianzanalyse; Kovarianzanalyse mit Messwiederholung)

Unvollständige, mehrfaktorielle Versuchspläne: S. 109

(Hierarchische und teilhierarchische Versuchspläne; Lateinische Quadrate; Griechisch-lateinische Quadrate; Quadratische Anordnungen mit Messwiederholungen)

Theoretische Grundlagen der Varianzanalyse: S. 111

(Einfaktorielle Varianzanalyse; Mehrfaktorielle Varianzanalyse; Varianzanalyse mit Messwiederholung; Einfaktorielle Kovarianzanalyse; Weitere Verfahren; Allgemeine Regeln)

Multivariate Verfahren (B4):

Partialkorrelation und Multiple Korrelation: S. 114

(Partialkorrelation; Berechnung der Partialkorrelation; Semipartialkorrelation; Berechnung der Semipartialkorrelation; Partialkorrelationen höherer Ordnung; Signifikanztests; Multiple Korrelation und Regression; Voraussetzungen und Signifikanztests; Schrumpfungskorrektur; Beta-Koeffizienten; Strukturkoeffizienten; Effektgröße; Lineare Strukturgleichungsmodelle; Zeit & Korrelation)

Das Allgemeine Lineare Modell (ALM): S. 127

(Grundprinzip des Allgemeinen Linearen Modells; Dummycodierung; Effektcodierung; Kontrastcodierung; Zusammenfassung)

Faktorenanalyse: S. 130

(Erste Schritte zur Faktorenanalyse; Grundprinzip der Faktorenanalyse; Zusammenfassung; Die „richtige“ Faktorenanalyse; Kennwerte der Faktorenanalyse: Eigenwert, Faktorladung & Kommunalität; Die Kennwerte Eigenwert, Faktorladung & Kommunalität am empirischen Beispiel; Kriterien für relevante Faktoren; Interpretation der Faktoren; Rotationsverfahren: Orthogonal vs. Oblique; Weiteres)

Clusteranalyse: S. 141

(Grundprinzip der Clusteranalyse; Grundprinzip hierarchischer clusteranalytischer Methoden; Fusionskriterien; Nichthierarchische clusteranalytische Verfahren; Die k-means-Methode; Probleme bei der Clusteranalyse; Zusammenfassung)

Multivariate Mittelwertvergleiche: S. 146

(Mehrfach univariate Analysen vs. eine multivariate Analyse)

Diskriminanzanalyse: S. 148

(Grundprinzip der Diskriminanzanalyse)

Kanonische Korrelationsanalyse: S. 153

(Grundprinzip der kanonischen Korrelationsanalyse; Schlussbemerkung)

© Manuel Ulrich

Gegen eine Vervielfältigung dieses Skriptes für private Zwecke bestehen keine Einwände. Die kommerzielle Nutzung ist untersagt.

Zusammenfassung der gebrauchten Abkürzungen:

NV	=	Normalverteilung
STNV	=	Standardnormalverteilung
SKV	=	Stichprobenkennwerteverteilung
H_0	=	Nullhypothese
H_1	=	Alternativhypothese
KI	=	Konfidenzintervall
df	=	Freiheitsgrade
r	=	Korrelation
Abszisse	=	x-Achse
AV	=	abhängige Variable
UV	=	unabhängige Variable
KV	=	Kontrollvariable
QS	=	Quadratsumme
ANOVA	=	Varianzanalyse
ANCOVA	=	Kovarianzanalyse
MANOVA	=	Multivariate Varianzanalyse
FA	=	Faktoranalyse
Vp(n)	=	Versuchsperson(en)
VT	=	Vortest
NT	=	Nachtest
QUEX	=	Quasi-Experiment
Reliabilität	=	Zuverlässigkeit der Daten
Validität	=	Gültigkeit inhaltlicher Schlussfolgerungen
Axiom	=	nicht beweisbare Grundannahme
DK	=	Determinationskoeffizient (der %-Anteil der Varianz, der vorhergesagt wird)
EG	=	Experimentalgruppe
KG	=	Kontrollgruppe
SP	=	Stichprobe
IA	=	Interaktion

Konzeption des Skriptes

Das Skript folgt beim Mathematikteil (B2, B3 & B4) den Kapiteln des Bortz (sofern nicht anders angegeben), ergänzend dazu habe ich die Folien von B2, B3 & B4 nachgearbeitet.

Ein Eingang auf SPSS – Ausdrücke (wäre wohl eher B5) erfolgt im B4 Teil.

Da ich im Mathematikteil auf den Bortz als Grundlage, sowie zur Absicherung auf die offiziellen Folien und ein Skript aus der Fachschaft zurückgreifen konnte, meine ich, dass ich alle relevanten Informationen (eher zuviel) für B2, B3 & B4 in diesen Skript stehen habe.

Den Methodenlehreteil (B1) habe ich in einem gesonderten Skript stehen, da die Anforderungen & Themen pro Uni unterschiedlich sind. In der Psychologiefachschaft der Uni Bonn ist aber der B1-Teil zu finden (oder schreibt mir eine eMail mit der Bitte um Sendung des B1-Teils).

Das Erstellen eines eigenen Lernskriptes – was den eigentlichen Lernerfolg bringt, da ihr die Sachen nicht nur verstehen, sondern auch erklären müsst – ersetzt das Lesen dieses Skriptes mit Sicherheit nicht.

Benutzte Literatur für das Skript:

Statistik (B2), Varianzanalyse (B3) & Multivariate Verfahren (B4):

- Bortz (1999): Statistik für Sozialwissenschaftler (5. Auflage) [Unsere Bibel]
- Die Folien der Lehrveranstaltungen B2 (WS 01/02), B3 (SS 01) & B4 (WS 01/02)
- Das Statistik-Skript von Benjamin Zeller (B2 & B3)

Griechisches Alphabet:

A	α	Alpha	N	ν	Ny
B	β	Beta	Ξ	ξ	Xi
Γ	γ	Gamma	O	o	Omikron
Δ	δ	Delta	Π	π	Pi
E	ϵ	Epsilon	P	ρ	Rho
Z	ζ	Zeta	Σ	σ	Sigma
H	η	Eta	T	τ	Tau
Θ	ϑ	Theta	Y	υ	Ypsilon
I	ι	Jota	Φ	ϕ	Phi
K	κ	Kappa	X	χ	Chi
Λ	λ	Lambda	Ψ	ψ	Psi
M	μ	My	Ω	ω	Omega

ACHTUNG!

Bevor es losgeht:

In diesem Skript stehen viele mit dem Formeleditor von Windows erstellte Formeln sowie einige Formeln aus dem B2 & B3-Skript. Es kann sein, dass die Formeln sich verändern, da der Formeleditor von Windows nicht bei allen PCs die Formeln genau gleich darstellt (dies kann beim „Anklicken“ der Formeln geschehen).

Deswegen ist das Skript schreibgeschützt. Bitte also nur Änderungen speichern, wenn Ihr – zu eurem eigenen Interesse – eine anders benannte Sicherheitskopie erstellt habt.

Wenn gar nichts klappt, kopiert Euch mein Skript ausgedruckt aus der Fachschaft.

Zur Kontrolle: es folgen eine fehlerfreie Kovarianzformel und ein z-Test (für eine Population).

$$\text{COV}_{(x,y)} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$z_{\text{emp}} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

Statistik

Deskriptive Statistik:

Deskriptive Statistik = beschreibende Statistik, d.h. sie dient zur Zusammenfassung und Darstellung von Daten (z.B. durch schöne Grafiken, Kennwerte wie Mittelwert, Streuung, etc.).

Skalenniveaus:

Terminologie: um Menschen empirisch untersuchen zu können, müssen wir ihnen Zahlen zuordnen können (nach bestimmten, vorgegebenen Regeln), so dass wir auf dieser Basis Berechnungen anstellen können:

Die z.B. Größe von Herrn Schmitz (ein **empirisches Relativ**) lässt sich durch Zuordnung von Zahlen in einem **numerischen Relativ** (z.B. 1,85 m) ausdrücken.

Wenn ein empirisches Relativ nur durch bestimmte Zahlen ausgedrückt werden kann und nicht auch durch andere, dann ist es eindeutig bzw., anders ausgedrückt: **homomorph**.

D.h. auf das Beispiel bezogen: die Größe von Herrn Schmitz beträgt 1,85 m; oder mathematisch formuliert:

$$\text{Herr Schmitz} = \text{Größe in Metern (1,85)}$$

Die oben zu sehende Gleichung ist eine homomorphe Abbildungsfunktion zusammen mit einem empirischen und einem numerischen Relativ. Dies bezeichnet man auch als **SKALA**.

Die Funktionswerte (nur der Teil der Gleichung rechts vom „=") werden auch als **Skalenniveaus** oder **Messwerte** bezeichnet.

Messen ist also eine strukturerhaltende Zuordnung von Zahlen zu Objekten und Ereignissen.

Diese Zuordnung kann Aussagen unterschiedlicher „Güte“ erlauben, wobei die jeweiligen Gütestufen als Skalenniveaus bezeichnet werden:

Grundsätzlich wird unterschieden zwischen

- Nominalskalen
- Ordinalskalen
- Intervallskalen
- Ratio- oder Verhältnisskalen

Eselsbrücke: Wenn Ihr bei der Reihenfolge schwarz (französisch: „**noir**“) seht.

Die Skalenniveaus unterscheiden sich in ihrem Informationsgehalt. Dieser determiniert die erlaubten (mathematischen) Operationen und damit die statistischen Verfahren, die auf die jeweiligen Daten angewendet werden dürfen.

Nominalskala:	Objekte mit gleicher Merkmalsausprägung erhalten gleiche Zahlen, Objekte mit verschiedenen Merkmalsausprägung erhalten verschiedene Zahlen. Die einzelnen Ausprägungen des Merkmals und damit die Stufen der Skala sind ungeordnet (gleichberechtigt). Beispiele: Augenfarbe, Geschlecht (solange es ohne Wertung bleibt)
Ordinalskala:	Eine Ordinalskala ordnet Objekten Zahlen zu, die so geartet sind, dass von jeweils zwei Objekten das Objekt mit der größeren Merkmalsausprägung die größere Zahl erhält. Über das Ausmaß des Unterschiedes zwischen den einzelnen Ausprägungen ist nichts bekannt. Beispiele: Windstärken, Militärische Ränge, Ranglisten beim Sport
Intervallskala:	Wie bei der Ordinalskala, so sind auch bei der Intervallskala die Stufen geordnet. Zusätzlich sind hier die Abstände zwischen den einzelnen Stufen alle gleich groß. Hat z.B. ein Objekt den Skalenwert 1, so ist es von dem Objekt mit dem Skalenwert 2 genauso weit entfernt, wie dieses von einem Objekt mit dem Skalenwert 3. Der Nullpunkt einer Intervallskala ist relativ. Beispiele: Temperatur in Celsius oder Fahrenheit
Verhältnisskala:	Eine Verhältnisskala ist eine Intervallskala mit absolutem Nullpunkt, der das „Nichtvorhandensein“ des Merkmals ausdrückt. Beispiele: Temperatur in Kelvin, Größe in m; Gewicht in kg, Zeit in sec. Null Kelvin (ca. -273 °C): das Fehlen jedweder atomarer Bewegung

	Mögliche Aussagen:	Mögliche Transformationen:
Nominalskala:	gleich / ungleich	Alle, die die Unterscheidung der Werte nicht verhindern.
Ordinalskala:	gleich / ungleich; größer/ kleiner	Alle, die die Ordnung der Werte nicht verändern.
Intervallskala:	gleich / ungleich; größer/ kleiner; Abstandsvergleiche	Multiplikation/ Division, Addition/ Subtraktion
Verhältnisskala:	gleich / ungleich; größer/ kleiner; Abstands-; Verhältnisvergleiche	nur Multiplikation/ Division (Erhaltung des Nullpunktes)

	Mögliche zentrale Tendenzen:	Mögliche Dispersionsmaße:
Nominalskala:	Modus	---
Ordinalskala:	Modus, Median	Variationsbreite
Intervallskala:	Modus, Median, Arithmetisches Mittel	Variationsbreite, AD-Streuung, Varianz
Verhältnisskala:	Modus, Median, Arithmetisches Mittel	Variationsbreite, AD-Streuung, Varianz

Um noch vernünftige Berechnungen anstellen zu können, geht man u.a. in der psychologischen Forschung meist von Intervallskalenniveau aus, auch wenn das streng mathematisch gesehen nicht vorhanden ist.

[Die Kultusministerkonferenz definiert Schulnoten als intervallskaliert, um so einen Durchschnitt u.a. beim Abitur berechnen zu können, auch wenn z.B. zum Erreichen einer „Vier gut“ 50 % oder zwei Drittel der Punkte zu erreichen sind.]

Maße der zentralen Tendenz: Modus, Median & Arithmetisches Mittel:

Durch welchen Wert wird eine Vielzahl von Werten am besten repräsentiert?

Hierzu geben Maße der zentralen Tendenz Auskunft über die *Lage* bzw. *Lokation* einer Verteilung in Termini der zugrundeliegenden Skala.

Modalwert (Modus):

Der Modus ist der Wert, der in einer Verteilung am häufigsten vorkommt. Es kann auch mehr als einen Modus geben (durch mehrere Maxima; Ausdruck bei Kurvendiskussion: Hochpunkte).

Statistische Abkürzung: **Mo**

Median:

Der Median ist der Wert, der eine der Größe nach geordnete Merkmalsverteilung (also mind. Ordinalskalenniveau) in zwei gleichgroße Hälften teilt. Bei gerader Anzahl von Werten wird das arithmetische Mittel aus dem größten Wert der unteren Hälfte und dem kleinsten Wert der oberen Hälfte bestimmt. Bei ungerader Anzahl wird der Wert in der Mitte genommen.

Statistische Abkürzung: **Md**

Statistische Eigenschaften:

- Die Summe der absoluten Abweichungen vom Median ist minimal.

Arithmetisches Mittel

Das arithmetische Mittel bzw. der Mittelwert ist die Summe aller Messwerte dividiert durch die Anzahl der Messwerte n :

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

Statistische Eigenschaften:

- Die Summe der Abweichungen aller Messwerte vom Mittelwert ist Null.
- Die Summe der quadrierten Abweichungen der Messwerte vom Mittelwert ist ein Minimum.

Modus, Median und Arithmetisches Mittel sind im wissenschaftlichen, populärwissenschaftlichen und alltäglichen Gebrauch beliebte Werkzeuge zum Beweisen eigener Thesen:
 „Die meisten Abiturienten wählen Jura.“ (auch wenn es nur 8,6 % sind) [Zahl fiktiv!]
 „Die besten 50% haben einen Schnitt von 3,4 oder besser.“ (Aha! Und wie verteilt sich das?)
 „Im Gegensatz zu den 50er können die Hauptschüler heute im Durchschnitt nicht sauber schreiben.“ (In den 50ern gingen auch fast 90% aller Schüler auf die Hauptschule; heute 30%.)

Weitere Maße der zentralen Tendenz:

Geometrisches Mittel, Harmonisches Mittel, Gewichtetes Mittel (siehe Bortz S. 40)

Dispersionsmaße: u.a. Perzentile, Quadratsummen & Varianzen:

Zwei Verteilungen mit ähnlichen oder sogar gleichen Maßen der zentralen Tendenz können trotzdem sehr unterschiedlich sein, und zwar aufgrund ungleicher Streuungen (Dispersionen) ihrer einzelnen Werte.

Dispersionsmaße geben also Auskunft über die Streuung bzw. Variabilität der Werte.

„Für die empirische Forschung sind Dispersionswerte den Maßen der zentralen Tendenz zumindest ebenbürtig.“ (Bortz)

Perzentile und Variationsbreite:

Das Perzentil $P_{X\%}$ ist der Punkt auf einer Skala, unterhalb dessen $X\%$ der aufsteigend angeordneten Messwerte einer Stichprobe liegen. Der Median ist das 50. Perzentil einer Verteilung.

Die **Variationsbreite** (engl. Range) ist die Differenz zwischen dem größten (Maximum) und dem kleinsten Wert (Minimum).

Wenn dieser Wert stark von Extremwerten abhängt, betrachtet man nur einen „eingeschränkten“ Streubereich, z.B. nur die mittleren 90% aller Werte (also den Bereich zwischen dem 5. und 95. Perzentil.). [Bei einer Standardnormalverteilung liegen die Werte zwischen $-\infty$ und $+\infty$, betrachtet man aber „nur“ 99, 74% aller Werte, liegen die Werte zwischen $-3,00$ und $+3,00$!]

Die **Quadratsumme (QS)** ist die Summe der quadrierten Abweichungen aller Messwerte vom Mittelwert.

Die **Varianz** ist die Summe der quadrierten Abweichungen aller Messwerte vom Mittelwert, dividiert durch die Anzahl der Messwerte n :

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

Die **Standardabweichung** oder **Streuung** ist die Wurzel aus der Varianz:

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

z-Werte:

In der psychologischen Diagnostik ergibt sich oftmals die Aufgabe, Testwerte zweier Personen aus verschiedenen Kollektiven (oder Gruppen, etc.) zu vergleichen.

Deswegen werden die individuellen Leistungen an der Gesamtleistung des (eigenen) Kollektivs relativiert, d.h. standardisiert.

Dies geschieht über eine z-Transformation:

Die z-Transformation ermöglicht den *relativen* Vergleich von Variablenausprägungen, da sie Mittelwerts- und Streuungsunterschiede „wegrelativiert“:

$$z_i = \frac{x_i - \bar{x}}{s_x}$$

Die z-Transformation überführt jede Verteilung in eine Verteilung mit Mittelwert 0 und Streuung bzw. Varianz 1.

Merkmalszusammenhänge & -vorhersagen:

Ich ziehe hier die Bereiche „Kovarianz und Korrelation“, „Korrelation und Kausalität“, „Lineare Regression“, „Beobachtete, vorhergesagte und Residualwerte“ und „Determinationskoeffizient“ aus Kapitel 6 im Bortz vor, da mir eine Thematisierung in Kapitel 6 als zu spät erscheint.

Kovarianz und Korrelation:

Mit der **Kovarianz** lässt sich bestimmen, wie sich die relativen Positionen von gepaarten Messwerten aus zwei Variablen zueinander verhalten:

$$\text{COV}_{(x,y)} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Die Kovarianz kann nur hinsichtlich ihres Vorzeichens, nicht aber hinsichtlich ihrer Größe interpretiert werden. [Da man auch mit Inches statt cm oder m statt cm arbeiten kann, kann die Größe der Kovarianz stark schwanken.] [Die Kovarianz ist bei Monte-Carlo-Studien sehr robust.]

- Ein positives Vorzeichen der Kovarianz zeigt an, dass sowohl überdurchschnittliche Werte von X stärker mit überdurchschnittlichen Werten von Y gepaart sind als auch, dass unterdurchschnittliche Werte von X stärker mit unterdurchschnittlichen Werten von Y zusammen auftreten.
- Ein negatives Vorzeichen zeigt entsprechend an, dass überdurchschnittliche Werte von X stärker mit unterdurchschnittlichen Werten von Y gepaart sind und umgekehrt.

- Keine Kovarianz entsteht dann, wenn sich die Werte von X und Y ausgleichen. Dies macht eine Interpretation unmöglich.
- Nur zur Erinnerung: selbst der kleinste positive lineare Zusammenhang ergibt eine positive Kovarianz! Damit man das auch erkennen kann, nutzt man Korrelationen.

Man standardisiert eine Kovarianz und erhält eine Korrelation:

Die Kovarianz kann maximal den Wert $s_x \cdot s_y$ annehmen. Dies ist damit ein sehr schöner Term zur Relativierung bzw. Standardisierung von Kovarianzen. (Man macht so was wie eine z-Transformation, nur für beide Variablen (X, Y) gleichzeitig.)

Diese **Standardisierung** nennt man *Produkt-Moment-Korrelation* $r_{(x,y)}$.

$$r_{(x,y)} = \frac{COV_{(x,y)}}{s_x \cdot s_y}$$

Die Korrelation $r_{x,y}$ kann Werte zwischen -1 und $+1$ annehmen, wobei

- -1 einen perfekt negativen linearen Zusammenhang anzeigt,
- $+1$ einen perfekt positiven linearen Zusammenhang anzeigt und
- 0 keinen linearen Zusammenhang anzeigt.

Perfekter linearer Zusammenhang: Die Werte der einen Variablen lassen sich durch eine Lineartransformation in die der anderen überführen:

$$y_i = a + b \cdot x_i$$

(Und schon schließt sich der Kreis zu Linearer Regression.)

[Die Abkürzung „r“ bei der Korrelation ist auf das Wort „Regression“ zurückzuführen, da Korrelations- wie Regressionsrechnung eng miteinander verknüpft sind.]

Korrelation und Kausalität:

Die Kovariation zweier Variablen kann auf eine kausale Beziehung der beiden Variablen oder auf die Beeinflussung durch eine oder mehrere *Drittvariablen* zurückgehen. Auf Basis einer Korrelation kann **nicht** auf die Art der Verursachung geschlossen werden. (Z.B. korrelieren bei Kindern Intelligenz und Schuhgröße.) **Kausalität lässt sich nur widerlegen, nicht beweisen.**

Lineare Regression:

Hängen zwei Variablen (wie z.B. Körperlänge und Gewicht) zusammen, so ist es möglich, eine Variable jeweils auf Basis der anderen vorherzusagen.

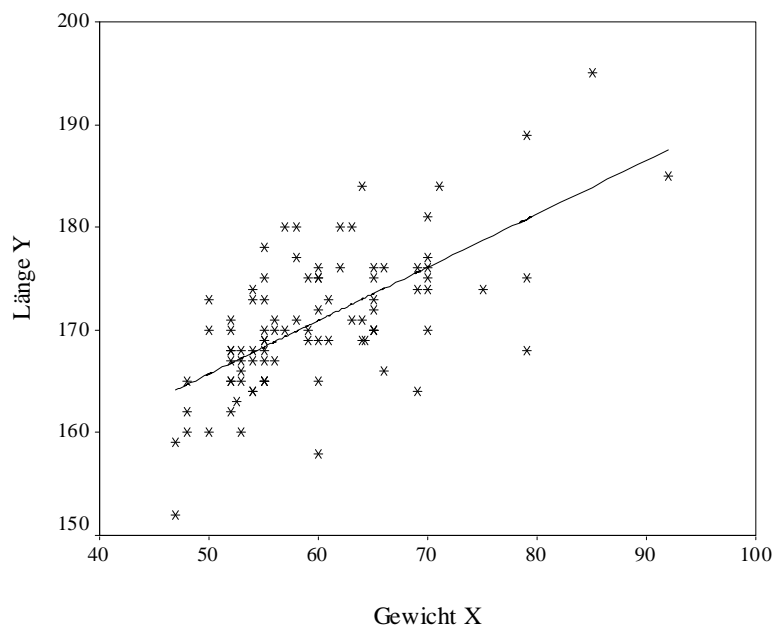
Die Vorhersagevariable wird als *unabhängige Variable* oder *Prädiktorvariable* (X) bezeichnet, die Variable, die vorhergesagt werden soll, als *abhängige Variable* oder *Kriteriumsvariable* (Y):

(Prädiktorvariablen sind im allgemeinen einfacher und billiger zu messen; z.B. IQ-Tests)
Im Rahmen der bivariaten Regression wird die Vorhersage über folgende lineare Beziehung vorgenommen:

$$\hat{y}_i = a + b \cdot x_i, \text{ wobei}$$

\hat{y}_i = vorhergesagter Wert, a = Schnittpunkt der Geraden mit der y -Achse (Ordinatenabschnitt, Interzept) und b = Steigung der Geraden.

Ziel ist die Bestimmung einer Geraden, die den Gesamttrend aller Punkte am besten wiedergibt.



Gleichungen:

$$b_{yx} = \frac{\text{COV}_{(x,y)}}{s_x^2}; \quad a_{yx} = \bar{y} - b_{yx} \cdot \bar{x}$$

(Grafiken im Bortz S. 181)

Bei der linearen Regression wird die Gleichung gesucht, für die die Summe der quadrierten Abweichungen zwischen vorhergesagten und tatsächlichen Werten minimal ist (= Kriterium

der kleinsten Quadrate). $\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \min.$, wobei \hat{y}_i = Wert auf der Geraden

Beobachtete, vorhergesagte und Residualwerte:

Die Abweichungen der beobachteten Werte von den vorhergesagten Werten heißen *Regressionsresiduen*:

$$y_i^* = y_i - \hat{y}_i.$$

Die Residuen enthalten die Anteile der Kriteriumsvariablen (Y), die durch die Prädiktorvariable nicht erfasst werden:

- Die Mittelwerte der Kriteriumsvariablen und der vorhergesagten Werte sind gleich.
- Der Mittelwert der Residuen ist 0.

Diese Anteile der Kriteriumsvariablen sind auf Messfehler, vor allem aber auch durch Bestandteile des Kriteriums, die durch andere, mit der Prädiktorvariablen nicht zusammenhängende Merkmale, erklärt werden können (z.B. Störvariablen).

Zerlegung der Kriteriumsvarianz:

Die Varianz der y -Werte setzt sich additiv aus der Varianz der vorhergesagten \hat{y} -Werte und der Varianz der Residuen y^* zusammen:

$$s_y^2 = s_{\hat{y}}^2 + s_{y^*}^2.$$

Determinationskoeffizient:

= Varianzanteil der abhängigen Variablen, der mit der unabhängigen vorhergesagt bzw. erklärt werden kann:

$$r_{(y,\hat{y})}^2 = \frac{s_{\hat{y}}^2}{s_y^2}.$$

Dabei kann man die „Power“ einer Regressionsgeraden erfassen, d.h. wie viel Prozent sagt sie vorher und wie viel bleibt im Dunkeln.

Korrelationen:

- Die Korrelation zwischen Kriterium und Vorhersage ist gleich der Korrelation zwischen Kriterium und Prädiktor:

$$r_{(y,\hat{y})} = r_{(y,x)}.$$

- Die Korrelation zwischen Prädiktor (X) und Residuen (von Y) ist 0: (muss ja, die Residuen sind ja der Teil, den der Prädiktor X nicht aufklärt)

$$r_{(x,y^*)} = 0.$$

- Für die Korrelation zwischen Kriterium und Residuen gilt:

$$r_{(y,y^*)}^2 = 1 - r_{(x,y)}^2.$$

Wahrscheinlichkeitstheorie und -verteilungen:

Eine der wichtigsten Eigenschaften des Menschen ist es, Redundanzen (unnötige Ausführlichkeit einer Information) zu erkennen und zu erlernen. Dies verhindert, dass der Mensch in dem Chaos von Zufälligkeiten und in der Fülle von Informationen zugrunde geht.

Der Mensch schafft sich ein Ordnungssystem (z.B. Zahlen, Buchstaben, Gesetze), an dem er sein Verhalten orientiert – in (teils blindem) Vertrauen darauf, dass das System auch funktioniert.

Jedoch gibt es in der Welt keine absoluten Wahrheiten und Wahrscheinlichkeiten. Wir regulieren unser Verhalten nach einem System unterschiedlich wahrscheinlicher Hypothesen, die teils völlig blöde sind (Geld in Glücksspiele investieren), teils sehr vernünftig sind (wir steigen in ein Flugzeug, weil es sehr unwahrscheinlich ist, dass es abstürzt.).

Der Mensch schafft sich Grundsätze (in der Mathematik: Axiome), die er beweislos voraussetzt. Aus mehreren Axiomen (einem Axiomssystem) können Lehrsätze nach den Regeln der Logik hergeleitet werden. Wichtig dabei:

- 1.) **Widerspruchsfreiheit**
- 2.) **Unabhängigkeit:** kein Axiom lässt sich aus einem anderen herleiten.
- 3.) **Vollständigkeit:** alle Axiome reichen aus, um alle inhaltlich richtigen Sätze zu beweisen.

Zufallsexperimente:

Für die Definition objektiver Wahrscheinlichkeiten (das will ja der Methodiker) ist der Begriff des „Zufallsexperimentes“ zentral:

Ein Zufallsexperiment ist ein beliebig oft wiederholbarer wohldefinierter Vorgang, dessen Ausgang nicht vorhersagbar ist.

Ein Ausgang wird als *Ergebnis* oder *Elementarereignis* ω , die Menge aller *Elementarereignisse* wird als *Ergebnisraum* $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ bezeichnet.

Bei der *theoretischen (kombinatorischen) Bestimmung* einer *Wahrscheinlichkeit* geht man oft von der Gleichwahrscheinlichkeit aller Ergebnisse aus:

$$p(A) = \frac{\text{Anzahl günstiger Ergebnisse}}{\text{Anzahl möglicher Ergebnisse}}$$

Günstige Ergebnisse sind solche, bei denen sich A ereignet, mögliche Ergebnisse sind alle Ergebnisse des endlichen Ergebnisraumes (A kann durchaus mehr als ein einziges ω_i umfassen).

Ein Beispiel:

Ich werfe einen perfekten sechsseitigen Würfel: Die Chance, das ich eine „sechs“ erhalte, beträgt also ein Sechstel, oder, mathematisch formuliert:

$$p(\text{Wurf}) = \frac{1}{6}$$

Grundbegriffe der Wahrscheinlichkeitsrechnung:

disjunkt = unvereinbar; die Ereignisse haben keine gemeinsame Menge, keinen **Durchschnitt**

p = Possibility (Wahrscheinlichkeit)

$p(A)$ = Wahrscheinlichkeit von A

$p(\bar{A})$ = Wahrscheinlichkeit des Fehlens von A

$p(A \cap B)$ = die Wahrscheinlichkeit von A und B (gleichzeitig)

$p(A \cup B)$ = die Wahrscheinlichkeit von A und/oder B (also entweder A oder B oder A&B gleichzeitig!)

- Das Ereignis, dass sich A und B ereignen, wird als **Durchschnitt** bezeichnet.
Symbol: \cap (= logisches Produkt)
- Das Ereignis, dass sich A und/oder B ereignen, wird als **Vereinigung** bezeichnet.

Bei unvorhersehbaren Wahrscheinlichkeiten muss man über viele Versuche die Wahrscheinlichkeiten schätzen (z.B. Reiszwecken werfen: zu wie viel Prozent landet es mit der Spitze nach oben, zu wie viel mit der Spitze schräg nach unten?) (je größer n , desto genauer):

Ein *statistischer Schätzer* der Wahrscheinlichkeit eines Ereignisses A ist die relative Häufigkeit eben dieses Ereignisses:

$$h(A) = \frac{f(A)}{n}.$$

Axiome der Wahrscheinlichkeitstheorie:

- 1.) Für die Wahrscheinlichkeit eines zufälligen Ereignisses A gilt $p(A) \geq 0$ (Nicht- Negativität).
- 2.) Die Wahrscheinlichkeit eines sicheren Ereignisses ist gleich 1 (Normierung).
- 3.) Die Wahrscheinlichkeit, dass eines der disjunkten Ereignisse A_1 oder A_2 oder ... A_k auftritt, ist gleich der Summe der Einzelwahrscheinlichkeiten $p(A_1), p(A_2), \dots p(A_k)$.

Additionstheorem für vereinbare bzw. nicht disjunkte Ereignisse:

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

Bei disjunkten bzw. nicht vereinbaren Ereignissen:

- $p(A \cup B) = p(A) + p(B)$.

Die bedingte Wahrscheinlichkeit $p(A|B)$ ist die Wahrscheinlichkeit von A, sofern B bereits eingetreten ist:

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

Multiplikationstheorem für **abhängige** Ereignisse:

$$p(A \cap B) = p(A) \cdot p(B|A) \quad \text{oder}$$

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

Multiplikationstheorem für **unabhängige** Ereignisse:

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

Zwei Ereignisse A und B sind *stochastisch unabhängig*, wenn A unabhängig von B eintreten kann, d.h. das Eintreten von B ändert nichts an der Eintretenswahrscheinlichkeit von A, und umgekehrt.

Das Theorem von Bayes:

Das Theorem von Bayes verknüpft die bedingten Wahrscheinlichkeiten $p(A|B)$ und $p(B|A)$ unter Verwendung des Satzes der totalen Wahrscheinlichkeit (Bortz S. 57). Eine der beiden bedingten Wahrscheinlichkeiten kann damit von der anderen aus erschlossen werden.

In der statistischen Entscheidungstheorie hat somit das Theorem von Bayes eine besondere Bedeutung; statistischen Entscheidungen werden immer aufgrund bedingter Wahrscheinlichkeiten getroffen.

[Die Wahrscheinlichkeit für das Auftreten bestimmter Daten (D) unter der Bedingung, dass eine bestimmte Hypothese (H) richtig ist: $p(D|H)$]

Variationen, Permutationen, Kombinationen:

Insbesondere durch Glücksspiele wurde eine Reihe von Rechenregeln erarbeitet, mit denen die Wahrscheinlichkeit bestimmter Ereigniskombinationen von gleichwahrscheinlichen Elementarereignissen ermittelt wird.

1. Variationsregel: Wenn jedes von k sich gegenseitig ausschließenden Ereignissen bei jedem Versuch auftreten kann, ergeben sich bei n Versuchen k^n verschiedene Ereignisabfolgen.

Beispiel: die Chance, fünf mal hintereinander beim Münzwurf „Zahl“ zu erhalten (2 mögliche Ereignisse), beträgt $k^n = 2^5 = 32$ verschiedene Ereignisabfolgen. Da es sich um ein günstiges Ereignis handelt, beträgt die Wahrscheinlichkeit $p = 1/32 = 0,031$

2. Variationsregel: Werden n voneinander unabhängige Zufallsexperimente durchgeführt und besteht die Ereignismenge des 1. Zufallsexperiments aus k_1 , die Ereignismenge des 2. Zufallsex-

periments aus k_2, \dots und die Ereignismenge des n -ten Zufallsexperiments aus k_n verschiedenen Elementarereignissen, sind $k_1 \times k_2 \times \dots \times k_n$ Ereignisabfolgen möglich (x bedeutet „mal“).

Beispiel: Ein Münzwurf und ein Würfelwurf, man möchte Zahl und eine 6 erhalten. Es ergeben sich $2 \times 6 = 12$ Ereignismöglichkeiten. Die Wahrscheinlichkeit beträgt also $p = 1/12 = 0.08$.

Permutationsregel: n verschiedene Objekte können in $n! = 1 \times 2 \times 3 \times \dots \times (n-1) \times n$ verschiedenen Abfolgen angeordnet werden. ($n!$ bedeutet n Fakultät)

Beispiel: In einer Urne befinden sich 6 unterschiedlich schwere Kugeln. Wie groß ist die Chance, die Kugeln in der Reihenfolge ihres Gewichtes herauszunehmen? (Es ist also ein unterschiedliches Ereignis ohne Zurücklegen; ergo ein abhängiges Ereignis)
 $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$ mögliche Abfolgen. Nur eine ist richtig, also $p = 1/720 = 0,0014$

1. Kombinationsregel: Wählt man aus n verschiedenen Objekten r zufällig aus, ergeben sich $n! / (n-r)!$ verschiedene Reihenfolgen der r Objekte.

Beispiel: Ich will aus einem Kartenspiel (32 Karten) in vier Zügen die Asse herausziehen. Die Reihenfolge der Asse ist **festgelegt**:

$$\frac{32!}{(32-4)!} = 32 \times 31 \times 30 \times 29 = 863040$$

Die Wahrscheinlichkeit beträgt also $1/863040 = 1,16 \times 10^{-6}$ Möglichkeiten

2. Kombinationsregel: Wählt man aus n verschiedenen Objekten r zufällig aus und lässt hierbei die Reihenfolge außer acht, ergeben sich für die r Objekte $\binom{n}{r}$ verschiedene Kombinationen.

Beispiel: Wie groß ist die Wahrscheinlichkeit, 6 Richtige aus 49 zu haben? (Reihenfolge ist im Gegensatz zur 1. Kombinationsregel ja egal)

Da wir noch alle wissen, dass

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \text{ ist also eingesetzt: } \binom{49}{6} = \frac{49!}{6!(49-6)!} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} =$$

$$= 13.983.816 \quad (\text{Noch genauere Rechenschritte: Bortz S. 61})$$

Die Wahrscheinlichkeit, sechs Richtige im Lotto zu haben beträgt also $1/13.983.816 = 7,15 \times 10^{-8}$ Möglichkeiten.

(Und die Lottobranche boomt. Sollten wir Psychologen da nicht mal eingreifen?)

3. Kombinationsregel: Sollen n Objekte in k Gruppen der Größen n_1, n_2, \dots, n_k eingeteilt werden (wobei $n_1 + n_2 + \dots + n_k = n$), ergeben sich $n! / (n_1! \dots n_k!)$ Möglichkeiten.

Beispiel: In einer Urne sind 4 rote, 3 blaue und 3 grüne Kugeln. Wie groß ist die Wahrscheinlichkeit, erst die 4 roten Kugeln zusammen, denn die 3 blauen und zuletzt die 3 grünen Kugeln zu entnehmen?

$$\frac{10!}{4!3!3!} = 4200$$

Die Wahrscheinlichkeit, erst die 4 roten Kugeln zusammen, denn die 3 blauen und zuletzt die 3 grünen Kugeln zu entnehmen, beträgt also $1/4200 = 2,38 \times 10^{-4}$ Möglichkeiten.

Verteilungsfunktionen:

Wenn man Experimente durchführt, erhält man jede Menge Daten (z.B. Studiengang von allen Studenten der Uni Bonn). Diesen Daten kann man bestimmte Kennungen (z.B. 1, 2, 3, etc.) hinzufügen, und so erhält man eine **Tabelle**. Das ist eine **diskrete** Wahrscheinlichkeitsfunktion.*

Diskrete Wahrscheinlichkeitsfunktionen bestehen aus einer endlichen Anzahl von Ereignissen. Auftretenswahrscheinlichkeiten einzelner Ereignisse lassen sich damit bestimmen.

Beispiele: Würfelergebnisse, Noten (manche haben gleiche Ergebnisse)

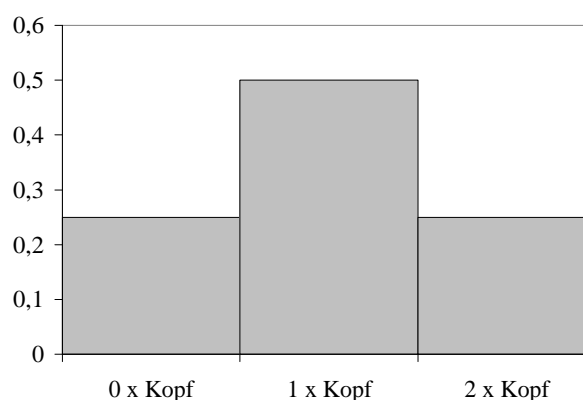
Man kann aber Daten auf bestimmte Kennungen zufügen, d.h. z.B. ich ordne alle Studenten der Uni Bonn der Größe nach. So erhält man eine **Kurve**. Das ist eine **stetige** Wahrscheinlichkeitsfunktion.*

Stetige Wahrscheinlichkeitsfunktionen bestehen aus einer unendlichen Anzahl von Ereignissen. Auftretenswahrscheinlichkeiten einzelner Ereignisse lassen sich damit nicht bestimmen.

Beispiele: Körpergröße, Geschwindigkeit (Niemand ist gleich groß, wenn man es genau nimmt.)

* Diese Beispiele sind von mir sind konstruiert. Ich kann z.B. den Zahlen 1 bis X die Studiengänge der Studenten der Uni Bonn zuordnen, bin deswegen aber nicht bei einer stetigen Wahrscheinlichkeitsfunktion.

Der eigentliche Unterschied zwischen STETIG und DISKRET ist in dem **fett** Geschrieben festgehalten. (Bei Stetigkeit Kurven und bei Diskretheit Tabellen zu benutzen stimmt aber.)

Diskrete Verteilungen:

\bar{x} und s^2 einer empirischen Verteilung / Stichprobe werden durch μ (Erwartungswert) und σ^2 bei einer theoretischen Verteilung einer Zufallsvariablen / Population ersetzt, bei diskreter Verteilung

$$\text{ergibt sich } \mu = \sum_{i=1}^N x_i \cdot p_i \text{ und } \sigma^2 = \sum_{i=1}^N p_i \cdot (x_i - \mu)^2 .$$

Binomialverteilung:

Ein Experiment, welches in der n -fachen Wiederholung von unabhängigen Einzelexperimenten mit nur zwei möglichen Ausgängen besteht, heißt **Bernoulli-Experiment**. Wenn ein Ereignis X in einem Einzelexperiment mit einer Wahrscheinlichkeit p auftritt, dann kann die Wahrscheinlichkeit P , dass X in n Wiederholungen k -mal auftritt, allgemein bestimmt werden:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} . \quad [1-p = q]$$

Diese **Wahrscheinlichkeitsfunktion** heißt *Binomialverteilung* und gibt an, wie wahrscheinlich das Eintreten eines Ereignisses $X = k$ ist.

Beispiel: Ich möchte wissen, wie groß die Chance ist in 10 Münzwürfen genau 7x „Zahl“ zu erhalten. Die Wahrscheinlichkeit einer Zahl beträgt 50%, mathematisch formuliert:

$$p = 0,5 = 1/2$$

Wenn ich jetzt einsetze: [$n = 10$, da 10 Würfe; $k = 7$, da 7x „Zahl“ gewünscht]

$$p(X = 7) = \binom{10}{7} \cdot \left(\frac{1}{2}\right)^7 \cdot \left(\frac{1}{2}\right)^3 = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot \left(\frac{1}{2}\right)^{10}}{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 0,117$$

Die Wahrscheinlichkeit, mit 10 Münzen genau 7 mal Zahl zu treffen, beträgt 11,7%.

Die **Verteilungsfunktion** hingegen gibt an, mit welcher Wahrscheinlichkeit ein Ereignis höchstens k -mal auftritt:

$$P(X \leq k) = \sum_{j=0}^k \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j} .$$

Die Verteilungsfunktion ist die kumulierte Wahrscheinlichkeitsfunktion.

Andere diskrete Verteilungen:

- Hypergeometrische Verteilung: wird benutzt bei **schwankenden Wahrscheinlichkeiten** (z.B. durch Ziehen ohne Zurücklegen)
- Poisson-Verteilung: **Verteilung seltener Ereignisse**; wird benutzt bei großem n & kleinem k .
- Multinominale Verteilung: auch Polynominalverteilung genannt; ermöglicht **mehrere unterschiedlich große k_x** .
- Negative Binomialverteilung: ermittelt die Wahrscheinlichkeit, dass ein Ereignis nach x Versuchen erst dann eintritt, wird häufig benutzt zur **Analyse von Wartezeiten**.

Stetige Verteilungen 1: Normalverteilung & Standardnormalverteilung:**Für alle stetigen Verteilungen geltend:**

- Die Gesamtfläche unter der Kurve ist auf 1 normiert, was bedeutet, dass mit einer Wahrscheinlichkeit 1 irgendein Ergebnis eintritt. Die Fläche über einem Intervall von zwei Werten gibt an, mit welcher Wahrscheinlichkeit Werte innerhalb dieses Intervalls eintreten.

Normalverteilung(NV):

Die Normalverteilung ist die für die Statistik wichtigste Verteilung; aus ihr werden weitere stetige Verteilungen wie u.a. die χ^2 -Verteilung, die t-Verteilung sowie die F-Verteilung abgeleitet; diskrete Verteilungen wie die Binomial- und Poisson-Verteilung konvergieren gegen sie. Die Zusammenhänge zwischen χ^2 -Verteilung, t-Verteilung sowie F-Verteilung folgen.

Alle Normalverteilungen haben typische Eigenschaften:

- Die Verteilung hat einen glockenförmigen Verlauf.
- Die Verteilung ist symmetrisch.
- Modus, Median und Mittelwert fallen zusammen.
- Es gibt nur einen Modus (wie auch bei Chi²-Verteilung, t-Verteilung sowie F-Verteilung.)
- Die Verteilung nähert sich asymptotisch der x-Achse (geht also von $-\infty$ bis $+\infty$).
- Zwischen den zu den Wendepunkten gehörenden x-Werten befinden sich ungefähr 2/3 der Gesamtfläche.
- Zwischen den Werten $\bar{x} + s$ (= Streuung) und $\bar{x} - s$ liegen 68,26% aller Werte, im Bereich $\bar{x} \pm 2s$ liegen 95,44%.
- Binomfinalverteilung ist für $n \Rightarrow \infty$ gleich der Normalverteilung (S. 78 Bortz)

Fesche Bilder von Normalverteilungen im Bortz auf S. 74.

Normalverteilungen unterscheiden sich durch

- a.) unterschiedliche Erwartungswerte (μ) [also Mittelwerte] und
- b.) unterschiedliche Streuungen (σ).

Zwei Normalverteilungen mit gleichem σ und μ sind identisch. Die Normalverteilung wird somit durch die beiden Parameter σ und μ festgelegt. Ihre Dichtefunktion lautet:

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

Damit klar ist, was diese Funktion da oben soll:

Mithilfe von Funktionen kann man durch Einsetzen des x-Wertes den y-Wert bestimmen und erhält somit eine wunderschöne Kurve, wie z.B. bei der Parabel: $f(x) = x^2$

Diese Funktion da oben gibt dann den y-Wert an, der aber gleichzeitig die Wahrscheinlichkeit des Eintretens dieses (eingesetzten) x-Wertes (in einem Intervall ΔX) angibt. (in der Tabelle als „Ordinate“ bezeichnet.) Über ein Integral dieser Dichtefunktion kann die Fläche berechnet werden (das ist ja der Sinn von Integralen). Alles klar?

Grundprinzip mathematisch ausgedrückt:

Die Wahrscheinlichkeitsfunktion einer stetigen Verteilung wird als Wahrscheinlichkeitsdichte oder **Dichtefunktion** bezeichnet. Erst über Integrale dieser Dichtefunktion können Wertebereichen Wahrscheinlichkeiten zugeordnet werden.

Standardnormalverteilung (STNV):

Die Standardnormalverteilung (STNV) ist eine ganz normale Normalverteilung; wird aber als Grundlage aller Vergleiche zwischen Normalverteilungen genutzt, da sie einen Mittelwert $\mu = 0$ und eine Streuung $\sigma = 1$ [Abk.: $N(0, 1)$] hat.

Durch eine z-Transformation kann jede Normalverteilung in die Standardnormalverteilung überführt werden; somit sind Vergleiche zwischen allen Normalverteilungen möglich!!

Ihre Dichtefunktion lautet:

$$f(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2}.$$

[Man hat hier x durch z ersetzt, um zum Ausdruck zu bringen, dass diese Funktion nur zur Standardnormalverteilung gehört.]

Diese Dichte (siehe „Ordinate“) und deren Verteilungsfunktion (siehe „Fläche“) ist im Bortz in Tabelle B (S. 768-772) wiedergegeben. [Einen z -Wert in der Tabelle nachzusehen setzte ich als bekannt voraus.]

Das hier ist übrigens die Integralformel, die benötigt wird, um die zu den z -Werten zugehörigen Flächenanteile [von $-\infty$ bis a] bei der Standardnormalverteilung (STNV) auszurechnen:

$$p(z < a) = \int_{-\infty}^a \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2} dz$$

Schön, nicht? Die anderen Formeln sind noch komplizierter (das ist aber, ebenso wie diese Formel, nicht relevant für die Prüfung – nur zur Anschauung gedacht :-).

Die **Bedeutsamkeit der Normalverteilung (NV)** lässt sich auf 4 Aspekte zurückführen:

- die NV als empirische Verteilung: eine Reihe von Messungen verteilt sich meistens **annähernd** normal, z.B. Körpergröße, -gewicht, Testleistungen. (Im 19. Jahrhundert entdeckt, heute relativiert.)
- die NV als Verteilungsmodell für statistische Kennwerte: **wenn man genug Stichproben zieht verteilen sich die Mittelwerte der Stichproben normal (Inferenzstatistik!)**.
- die NV als mathematische Basisverteilung: Verknüpfungen u.a. zu χ^2 -, t - sowie F -Verteilung.

- die NV in der statistischen Fehlertheorie: kaum eine Messungen ist absolut fehlerfrei. Misst man andauernd, erhält man irgendwann normalverteilte Werte. Der Mittelwert ist das wahrscheinlichste richtige Ergebnis (Strategie der alten Römer u.a. beim Aquäduktbau).

Stetige Verteilungen 2: Chi²-Verteilung, t-Verteilung sowie F-Verteilung:

χ²-Verteilung:

Gegeben sei eine normalverteilte Zufallsvariable z mit $\mu = 0$ und $\sigma = 1$, also eine STNV. Das Quadrat dieser Zufallsvariablen bezeichnen wir als χ_1^2 -verteilte Zufallsvariable.

$$\chi_1^2 = z^2$$

(Die untere Zahl am Chi² kennzeichnet die Anzahl der normalverteilten Zufallsvariablen z .)

Wenn wir jetzt (theoretisch unendlich) viele χ_1^2 -Werte aus zufällig gezogenen z -Werten nach der oben stehenden Formel ermitteln, erhalten wir eine stetige χ_1^2 -Verteilung. (Dichtefunktion steht nicht im Bortz) [Ein Bild verschiedener χ^2 -Verteilungen im Bortz auf S. 80.]

$$\chi_n^2 = z_1^2 + z_2^2 + \dots + z_n^2.$$

χ²-Verteilungen haben:

- in Abhängigkeit von der Anzahl der z^2 - Variablen unterschiedliche Freiheitsgrade bzw. df (degrees of freedom)= n
- eine Variationsbreite von 0 bis $+\infty$
- bei $df = n$ einen Erwartungswert von $\mu = n$, eine Streuung von $\sqrt{2n}$ und eine Schiefe (ob es linkssteil oder rechtssteil ist.) von $\sqrt{8/n}$.

Mit größer werdendem n nähert sich die χ^2 -Verteilung einer NV mit $\mu = n$ und $\sigma = \sqrt{2n}$ an.

t-Verteilung:

Aus einer STNV wird ein z -Wert und aus einer hiervon unabhängigen χ^2 -Verteilung ein χ_n^2 -Wert gezogen. Der folgende Quotient definiert einen t_n -Wert:

$$t_n = \frac{z}{\sqrt{\chi_n^2 / n}}$$

Ein Bild einer t-Verteilung im Bortz auf S. 81.

t-Verteilungen:

- unterscheiden sich durch Anzahl der Freiheitsgrade (= Freiheitsgrade der χ^2 -Verteilung; die steckt ja in den Werten drinnen – siehe Formel).
- nähert sich mit wachsenden Freiheitsgraden (n) der Normalverteilung an.

- haben einen Erwartungswert $\mu = 0$
- sind symmetrisch
- gehen bei $df \rightarrow \infty$ in eine NV über (Grenzpunkt Definitionssache)

F-Verteilung:

Gegeben sei eine χ^2 -Verteilung mit $df_1 = n_1$ und eine weitere, unabhängige χ^2 -Verteilung mit $df_2 = n_2$. Der Quotient von 2 zufällig aus diesen beiden Verteilungen entnommenen χ^2 -(Einzel-) Werten, multipliziert mit dem Kehrwert des Quotienten ihrer Freiheitsgrade (d.h. die df sind Zähler/Nenner-mäßig vertauscht), wird als F-Wert bezeichnet:

$$F_{(n_1, n_2)} = \frac{\chi_{n_1}^2}{\chi_{n_2}^2} * \frac{n_2}{n_1}$$

F-Verteilungen:

- sind asymmetrische, stetige Verteilungen mit einer Variationsbreite von 0 bis $+\infty$.
- unterscheiden sich durch Zähler- (n_1) und Nenner- (n_2) Freiheitsgrade

Ein Bild von F-Verteilungen ist im Bortz auf S.82.

Noch ein paar Zusammenhänge:

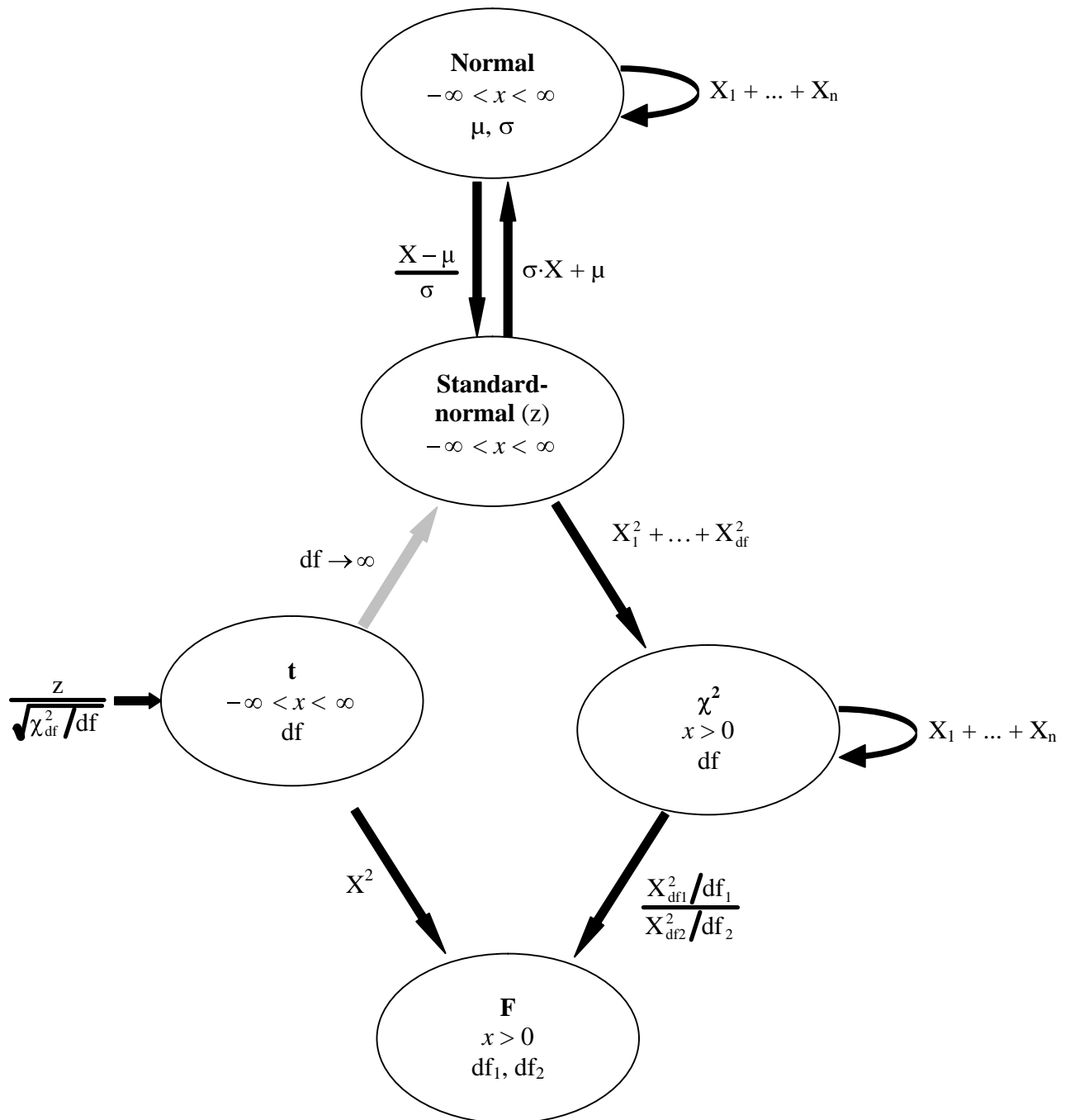
- Ein F-Wert in der Tabelle bei $df_1 = 1$ und $df_2 = \infty$ entspricht einem z-Wert der STNV bei halbiertem Alphaniveau (da die STNV zweiseitig ist muss man beide Grenzwerte aufaddieren.).
- Eine quadrierte t-Verteilung mit $df = n$ ist mit der F-Verteilung für einen Zähler- und n Nennerfreiheitsgrade identisch. Als Formel:

$$t_n^2 = \frac{\chi_1^2}{\chi_2^2} \cdot \frac{n}{1} = F_{(1, n)}$$

- Zwischen einer χ^2 -Verteilung und einer F-Verteilung besteht der folgende Zusammenhang:

$$F_{(n, \infty)} = \chi_n^2 / n$$

Zusammenhänge von NV, STNV, Chi²-, t- und F-Verteilung:



Stichprobe und Grundgesamtheit:

Die meisten Untersuchungen in u.a. der Psychologie sind dahin ausgerichtet, aufgrund einer kleinen (in einem Experiment) untersuchten Stichprobe Schlüsse auf die Allgemeinheit zu ziehen. Der sich hiermit befassende Teilbereich der Statistik wird als **Inferenz-** oder **schließende Statistik** bezeichnet.

Inferenzstatistik hat das Ziel, auf Basis von Stichprobenergebnissen Aussagen über die zugrundeliegende Population zu treffen.

Zur Inferenzstatistik zählen

- die Schätzung von Populationsparametern und
- das Testen (Überprüfen) von Hypothesen.

Begriffe der Inferenzstatistik:

- **Grundgesamtheit** oder **Population** bezeichnet alle potenziell untersuchbaren Elemente (Personen, Objekte), die ein gemeinsames Merkmal oder eine gemeinsame Merkmalskombination aufweisen (z.B. Männer, Studenten, Deutsche, VW Golf, Gartenzwerge, etc.).
- **Stichprobe** bezeichnet eine nach einer bestimmten Auswahlmethode gewonnene Teilmenge von Elementen aus der Population.
Inferenzstatistische Verfahren erfordern, dass die Stichprobe mittels einer Zufallsprozedur gewonnen wird.
- **Statistische Kennwerte**, wie die Maße der zentralen Tendenz oder die Maße der Dispersion können für Stichproben wie für Populationen (Grundgesamtheiten) ermittelt werden. Die Kennwerte einer Population bezeichnen wir als **Parameter**, bzw. **Populationsparameter**.
Ein Stichprobenkennwert wird als **Punktschätzer** bezeichnet.
- **Unbekannte Parameter** (sind in der Regel unbekannt) werden mit den Stichprobenkennwerten geschätzt.
- **Zur Unterscheidung** von Stichprobenkennwerten und Populationsparametern werden Kennwerte mit lateinischen und Parameter mit griechischen Buchstaben notiert. (manchmal werden für Parameter einfach die großen Buchstaben der Kennwerte genutzt; also N statt n . Die Bezeichnung der Kennwerte variiert nicht.)

	Stichprobenkennwert:	Populationsparameter:
Mittelwert:	\bar{x}	μ
Varianz:	s^2	σ^2
Streuung:	s	σ
Korrelation:	r	ρ

Mit endlichen Populationen – z.B. die Psychologiestudenten im 3. Semester der Uni Bonn – zu arbeiten ist ausgesprochen schwierig (letztlich ist jede Population endlich, aber ab gewissen Größen – z.B. alle Studenten Deutschlands – gilt sie statistisch als unendlich), dementsprechend versucht man mit unendlichen Populationen zu arbeiten.

- Damit eine Population als unendlich gilt, sollte sie mindestens 100x so groß sein wie die Stichprobe.

Stichprobenarten:

Wie genau kann gewährleistet werden, dass eine Stichprobe eine Grundgesamtheit möglich genau repräsentiert?

Eine Stichprobe kann in bezug auf die Grundgesamtheit entweder in bezug auf

- a.) alle Merkmale (globale Repräsentativität) oder
- b.) bestimmte Merkmale (spezifische Repräsentativität)

repräsentativ sein.

- **Zufallsstichprobe:** Jedes Element der Grundgesamtheit kann mit gleicher Wahrscheinlichkeit ausgewählt werden (Ziehen mit Zurücklegen).
Eine absolut fehlerfreie Zufallsstichprobe ist fast unmöglich; meist aufgrund zu kleiner Stichprobengrößen, mangelnder Repräsentativität und mangelnder Bereitschaft der Menschen, an einem Experiment teilzunehmen.
- **Klumpenstichprobe:** Elemente werden nicht einzeln, sondern in natürlichen Gruppen (Klumpen) ausgewählt [z.B. alle Alkoholiker in Kliniken].
Hierbei ist zu beachten, dass die Klumpen zufällig ausgewählt werden (vgl. mit dem Bsp.: die Auswahl der Kliniken müsste zufällig sein.), ein einzelner Klumpen reicht nicht aus.
- **Geschichtete Stichprobe:** Elemente werden nach relevanten Merkmalen in Schichten (Strata) vorgruppiert. Man überlegt, welche Merkmale relevant für die Untersuchung sind (z.B. Geschlecht, etc.), und wählt innerhalb dieser Merkmale zufällig die Leute aus (mit dem Klumpen-Verfahren geht es auch, ist aber statistisch unsauberer.).
Proportional geschichtete Stichprobe: entsprechend der Merkmalsverteilung in Population zusammengestellt, z.B. 40% Singles, 60% Ehepaare.
Generell führt eine nach relevanten Merkmalen geschichtete Stichprobe zu besseren Schätzwerten der Populationsparameter als eine einfache Zufallsstichprobe.

Verteilung von Stichprobenkennwerten:

Wir berechnen bei einer Stichprobe der Größe n den Mittelwert \bar{x} . Wie genau können wir nun daraus den Mittelwert der Population erkennen?

- bei zwei Stichproben aus einer Population: je weiter die Mittelwerte der Stichprobe voneinander entfernt sind, desto geringer die vermutlich richtige Schätzung des Populationsmittelwertes.
- wenn wir unendlich viele Stichproben haben, dann entsteht eine **Stichprobenkennwerteverteilung**, je geringer die Streuung (= Standardfehler $\sigma_{\bar{x}}$) ist, desto genauer ist die Schätzung.

Die Stichprobenkennwerteverteilung (SKV):

Gedankenexperiment: Eine Population mit μ und σ sei gegeben. Aus dieser Population werden wiederholt (theoretisch unendlich oft) Stichproben derselben Größe n mit Zurücklegen gezogen und innerhalb jeder Stichprobe wird z.B. der Stichprobenkennwert \bar{x} berechnet. Die Verteilung der so gewonnenen Kennwerte ist die Stichproben(kennwerte)verteilung von \bar{x} .

Wenn man genug Stichprobenkennwerte hat, ist deren Mittelwert mit dem der Population identisch.

Der Mittelwert einer Stichprobenverteilung (SKV) eines Kennwertes T heißt Erwartungswert E(T).

Entspricht der Erwartungswert eines Kennwertes dem Wert des zugehörigen Populationsparameters, $E(T) = \theta$, so wird dieser Kennwert als erwartungstreuer Schätzer bezeichnet.

Mittelwert:

Der Mittelwert ist ein erwartungstreuer Schätzer für μ , da gilt: $E(\bar{x}) = \mu$.

Varianz:

Die Stichprobenvarianz s^2 ist kein erwartungstreuer Schätzer: $E(s^2) \neq \sigma^2$.

s^2 unterschätzt σ^2 systematisch um den Faktor: $\frac{n-1}{n}$

Daher muss s^2 wie folgt korrigiert werden (doch, der Bruch ist richtig herum*):

$$\hat{\sigma}^2 = s^2 \cdot \frac{n}{n-1}$$

Das Dach über dem Sigma zeigt an, dass es sich um den Schätzer eines Parameters handelt.

$$\hat{\sigma}^2 \text{ ist erwartungstreu: } E(\hat{\sigma}^2) = \sigma^2.$$

Die Populationsvarianz schätzt man also durch die Summe der quadrierten Abweichungen aller Messwerte vom Mittelwert (Quadratsumme) geteilt durch $n-1$ (Freiheitsgrade der Varianz):

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

Jetzt ist eine Berechnung des geschätzten Standardfehlers möglich (siehe nächste Seite)

* **Einfach:** Weil s^2 unser Sigma systematisch unterschätzt, müssen wir s^2 vergrößern.

Mathematisch korrekt: s^2 unterschätzt systematisch das Sigma:

$$\hat{\sigma}^2 \cdot \frac{n-1}{n} = s^2$$

Und nun stellt die Formel mal um.

Der Standardfehler:

Ist die Streuung einer \bar{x} -Verteilung ziemlich klein, ist die Chance, mithilfe eines Stichprobenkennwertes \bar{x} den Mittelwert der Population μ zu treffen, relativ groß. Ist jedoch die Streuung groß, ist die Chance auf einen Treffer gering. Deswegen benötigt man den Wert der Streuung der Stichprobenkennwerteverteilung, und das ist der Standardfehler.

Die Streuung einer Stichprobenkennwerteverteilung wird als **Standardfehler des Mittelwertes** (also des Kennwertes) bezeichnet.

Der Standardfehler des Mittelwertes hängt von der Populationsvarianz σ^2 und der Stichprobengröße n ab:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$$

- Standardfehler verändert sich proportional zur Streuung in Population.
- Standardfehler verringert sich mit zunehmendem Stichprobenumfang.

Schätzung des Standardfehlers des Mittelwertes:

In der Regel ist die Varianz einer Population σ^2 nicht bekannt. Daher wird der Standardfehler aus den Stichprobendaten geschätzt: (eine Schätzung über den Durchschnitt der Stichprobenvarianzen wäre ja nicht erwartungstreu.)

$$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{\hat{\sigma}^2}{n}} = \sqrt{\frac{s^2}{n-1}}$$

Zentrales Grenzwerttheorem:

Der zentrale Grenzwertsatz sagt nun etwas über die Form aus:

Verteilungen von Mittelwerten aus Stichproben des Umfangs n , die sämtlich derselben Grundgesamtheit entnommen werden, **gehen** mit wachsendem Stichprobenumfang **in eine Normalverteilung über**.

Vereinbarungsgemäß hält man ein $n \geq 30$ für hinreichend, um die Stichprobenverteilung des Mittelwertes als Normalverteilung zu behandeln, ungeachtet der Verteilungsform der zugrundeliegenden Population.

Ein Beispiel: man werfe zwei sechsseitige Würfel, und lese deren Ergebnis ab. Die 7 kommt z.B. häufiger vor als die 2, da es bei der 7 mehrere Kombinationsmöglichkeiten gibt (Bortz S. 93). Es ergibt sich eine in Richtung NV bewegende Verteilung:

Ergebnis:	2	3	4	5	6	7	8	9	10	11	12
p(x):	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Entsprechendes gilt für **JEDE** beliebige Verteilungsform (**bei Mittelwerten**).

Stichprobenkennwerte müssen nicht nur Mittelwerte sein, sondern es sind auch Mediane und Varianzen möglich. **Bei Varianzen s^2 ist die Stichprobenverteilung χ^2 -verteilt.**

Mittelwert (\bar{x}) der Stichprobenkennwerteverteilung der Mittelwerte:

- entspricht μ der Population bei hinreichend großen Stichprobenumfängen.
- Hat ebenfalls bei $\mu \pm \sigma_{\bar{x}}$ 68,26% und bei $\mu \pm 2\sigma_{\bar{x}}$ 95,44% aller Messwerte liegen.

[wobei $\sigma_{\bar{x}}$ = Streuung der Mittelwerte der Stichproben (= Standardfehler)]

Das bedeutet man kann: „Bei $\mu \pm 2\sigma_{\bar{x}}$ liegen 95,44% aller Messwerte“, auch mathematisch formulieren, und zwar:

$$\mu - 2 \cdot \sigma_{\bar{x}} \leq \bar{x} \leq \mu + 2 \cdot \sigma_{\bar{x}}$$

Und schon sind wir bei einem **Wertebereich** (mehr dazu später).

Kriterien der Parameterschätzung: Erwartungstreue, Konsistenz, Effizienz, Suffizienz:

Nach welchen Kriterien können wir entscheiden, ob ein statistischer Kennwert einen brauchbaren Schätzer für einen Parameter darstellt?

Was sind die Eigenschaften, die eine gute Schätzung auszeichnen?

Erwartungstreue:

Ein statistischer Kennwert schätzt einen Populationsparameter erwartungstreu, wenn der Mittelwert der Kennwerteverteilung (**Kennwerte sind nicht nur Mittelwerte!**) bzw. deren Erwartungswert dem Populationsparameter entspricht (siehe auch „Stichprobenkennwerteverteilung“).

Konsistenz:

Von einem konsistenten Schätzwert sprechen wir, wenn sich ein statistischer Kennwert mit wachsendem Stichprobenumfang dem Parameter, den er schätzen soll, nähert.

[Demzufolge ist z.B. die Standardabweichung konsistent, aber nicht erwartungstreu; der Mittelwert hingegen konsistent und erwartungstreu.]

Effizienz:

Je größer die Varianz der SKV, desto geringer ist die Effizienz des entsprechenden Schätzwertes. [konkret: es geht hier also um die **Präzision** einer Schätzung.]

[z.B. schätzt eine Mittelwertverteilung effizienter μ als eine Medianverteilung. (= eine Verteilung aus den Medianen der Stichproben).]

Suffizienz:

Ein Schätzwert ist suffizient oder erschöpfend, wenn er alle in den Daten einer Stichprobe enthaltenden Informationen berücksichtigt, so dass durch Berechnung eines weiteren statistischen Kennwertes keine zusätzlichen Informationen über den zu schätzenden Parameter gewonnen werden kann.

[Da der Median nur ordinale Dateninformationen berücksichtigt, der Mittelwert aber intervallskalierte Dateninformationen berücksichtigt, ist der Mittelwert der erschöpfendere Schätzer.]

Betrachtet man alle vier Kriterien, sieht man, dass $\hat{\sigma}^2$ und \bar{x} die besten Schätzer für σ^2 und μ sind. Sie sind erwartungstreu, konsistent und erschöpfend. Bei NVen sind sie auch am Effizientesten.

Methoden der Parameterschätzung: Meth. d. kleinsten Quadrate, Maximum-likelihood-Methode:

Wie finde ich den besten Schätzer für eine Population?

Methode der kleinsten Quadrate:

Die Methode der kleinsten Quadrate wird noch im Punkt „Verfahren zur Überprüfung von Zusammenhangshypothesen“ thematisiert, trotzdem eine kurze Beschreibung:

Nehmen wir an, wir suchen den besten Schätzer für μ . (Wir wissen natürlich, dass das \bar{x} ist.) Den gesuchten Wert nennen wir a .

$$f(a) = \sum_{i=1}^n (x_i - a)^2 = \min.$$

Wir lösen das Ganze auf (Rechnung im Bortz S.97 & 98), und erhalten:

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

Demnach ist \bar{x} der beste Schätzer für μ .

Maximum-likelihood-Methode:

Mit der Maximum-likelihood-Methode finden wir für die Schätzung unbekannter Parameter Stichprobenkennwerte, die so geartet sind, dass sie die Wahrscheinlichkeit (engl.: likelihood) des Auftretens in der Population maximieren.

Messen wir 5x, und zwar

$x_1 = 11$, $x_2 = 8$, $x_3 = 12$, $x_4 = 9$ und $x_5 = 10$

Wenn wir jetzt eine NV haben, ist es unwahrscheinlich, dass die Population einen Mittelwert von $\mu = 20$ hat.

Nach der Maximum-likelihood-Methode würde sich herausstellen, dass $\bar{x} = 10$ der beste Schätzer für μ ist.

[Eingehendere Beschreibung im Bortz auf den S. 98 & 99]

Intervallschätzung: Wertebereich, Konfidenzintervalle:

Bei einem \bar{x} -Wertebereich versucht man herauszufinden, ob eine Stichprobe (d.h. eigentlich der Mittelwert \bar{x} einer Stichprobe) zu einer Population gehört.

D.h. bei einer NV, also $n \geq 30$, liegen Werte zu bestimmten Bedingungen um den Mittelwert μ herum.

Bei $\mu \pm \sigma_{\bar{x}}$ liegen 68,26% und

bei $\mu \pm 2\sigma_{\bar{x}}$ 95,44% aller zufällig gezogenen Stichprobenmittelwerte in dem Bereich ± 2 Streuungen um den Mittelwert μ .

D.h. bei $\mu = 100$ und $\sigma_{\bar{x}} = 5$, läge bei ± 2 Streuungen zu 95,44 % die gezogene Stichprobe zwischen 90 und 110. [eigentlich ist das falsch, ein Wert kann nur drin oder draußen liegen, also ist die Wahrscheinlichkeit entweder 1 oder 0 – ist aber nicht so wichtig (= Mathematiker-Pedanterie)].

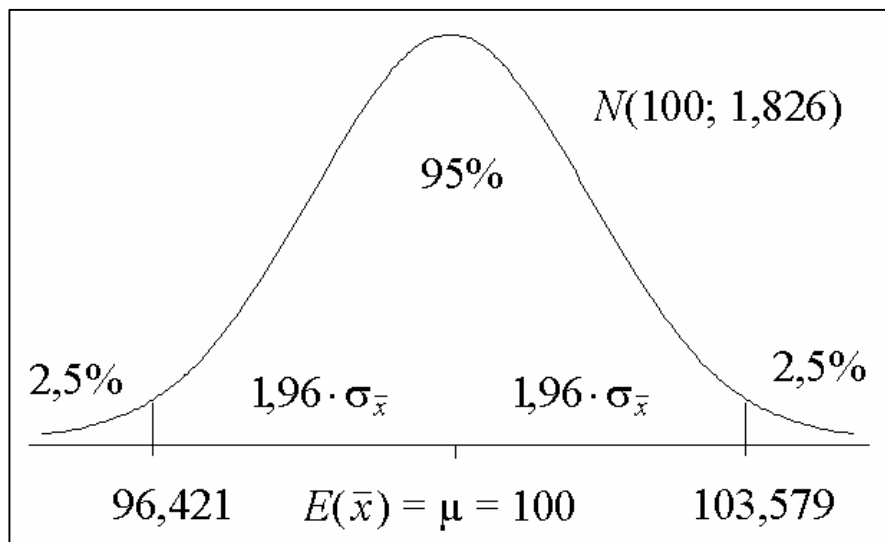
Diesen Bereich – hier 90 bis 110 – nennt man \bar{x} -Wertebereich.

μ und $\sigma_{\bar{x}}$ sind hierbei festgelegte Werte. Das Einzige, was variieren kann, ist \bar{x} . (Weswegen man es ja auch \bar{x} -Wertebereich nennt.)

Da man aber lieber mit standardisierten NVen rechnet, arbeitet man meist mit z-Werten. Die Gleichung sieht folgendermaßen aus:

$$\mu + z_{(\alpha/2)} \cdot \sigma_{\bar{x}} \leq \bar{x} \leq \mu + z_{(1-\alpha/2)} \cdot \sigma_{\bar{x}}$$

- $(\alpha/2)$ haben wir hier, da ein zweiseitiges Intervall gesucht wird, also z.B. auf jeder Seite 2,5% abgeschnitten wird.



In der Standardnormalverteilung liegen 95% der Fläche zwischen $z = \pm 1,96$.

Konfidenzintervalle:

Konfidenzintervalle verhalten sich genau umgekehrt zum \bar{x} -Wertebereich:

Bei einem Konfidenzintervall versucht man herauszufinden, ob eine Population zu einer Stichprobe gehört (die möglichen μ -Werte werden gesucht); und nicht wie beim Wertebereich, ob eine Stichprobe zu einer Population gehört (die möglichen \bar{x} -Werte werden gesucht).

- Konfidenzintervalle sind demnach Bereiche, in denen sich Populationsparameter befinden, die als Erzeuger einer Stichprobe (mit xx% Wahrscheinlichkeit) in Frage kommen.
- Man nimmt hier 95%ige oder 99%ige Wahrscheinlichkeiten (= **Konfidenzkoeffizienten**).

Das Konfidenzintervall kennzeichnet denjenigen Bereich eines Merkmals, in dem sich 95% oder 99% (legt man vorher fest) aller möglichen Populationsparameter befinden, die den empirisch ermittelten Stichprobenkennwert \bar{x} erzeugt haben können.

$$\bar{x} + z_{(\alpha/2)} \cdot \sigma_{\bar{x}} \leq \mu \leq \bar{x} + z_{(1-\alpha/2)} \cdot \sigma_{\bar{x}}$$

Gesucht sind nun hierbei die Grenzen des Konfidenzintervalls.

- Ein Konfidenzintervall Δ_{crit} wird bestimmt nach $\Delta_{\text{crit}} = \mu + z_{(\alpha/2)} \cdot \sigma_{\bar{x}}$

Ein Delta steht immer für eine Differenz, d.h. die (Unter-) Grenze des Konfidenzintervalls hat den Abstand Delta entweder zum Mittelwert oder zur anderen (Ober-)Grenze.

- bei kleinen Stichproben (unter 30) sind die „z-Werte“ nicht standardnormalverteilt, sondern t-verteilt (wenn die Population normalverteilt ist), also muss man t-Werte aus der Tabelle einsetzen; Freiheitsgrade = n-1 (siehe „t-Test“)

Bedeutung des Stichprobenumfangs:

Eine (weitere) Bestimmungsgröße für die Breite eines Konfidenzintervalls ist der Stichprobenumfang: **Je größer die untersuchte Stichprobe, desto kleiner das Konfidenzintervall.**

- Die Halbierung eines Konfidenzintervalls macht einen vierfachen Stichprobenumfang erforderlich.
- Die benötigten Stichprobenumfänge können erheblich gesenkt werden, wenn statt einer reinen Zufallsstichprobe eine sinnvoll geschichtete Stichprobe gezogen wird (=> Standardfehler wird kleiner => kleinere Stichprobenumfänge möglich.)

Formulierung und Überprüfung von Hypothesen:

Bei der Formulierung und Überprüfung von Hypothesen entsteht ein zentrales Problem:

Inwieweit können postulierte Eigenschaften (= Hypothesen) der Population (Theorie) durch Stichproben (Empirie) bestätigt werden?

Welche Kriterien gibt es dort? Bis wann ist ein Stichprobenwert (z.B. \bar{x} , s^2 , etc.) „gerade noch mit der Theorie übereinstimmend“?

Um diese u.ä. Fragen geht es in diesem Kapitel.

Alternativhypothesen:

Die Brauchbarkeit von Theorien ist davon abhängig, inwieweit sie sich in Untersuchungen von Teilaussagen, die aus ihr abgeleitet wurden, bewährt.

Ist eine Theorie (noch) nicht empirisch bewiesen worden, stellt man **Hypothesen** auf, um präzisere Vorhersagen machen zu können. Eine Hypothese ist also eine Deduktion der Theorie (Deduktion = Ableitung des Einzelfalls aus dem Allgemeinen). [Deswegen soll man Hypothesen ja auch vorher aufstellen, sonst leitet man nämlich aus dem Einzelfall bzw. den Einzelfällen das Allgemeine ab.]

Hypothesen beinhalten Aussagen, die mit anderen Theorien in Widerspruch stehen können oder den Wissensstand erweitern.

Hypothesen, die in diesem Sinne „innovative“ Aussagen beinhalten, werden als Gegen- oder Alternativhypothesen bezeichnet.

Diese Hypothesen müssen dann (nur noch) überprüft werden.

Varianten von Alternativhypothesen:

Je nach Art der Hypothesenformulierung unterscheidet man zwischen

- Unterschiedshypothesen und
- Zusammenhangshypothesen.

Unterschiedshypothesen prüft man bei *Häufigkeitsvergleichen* und *Mittelwertvergleichen*

(siehe „Verfahren zur Überprüfung von Unterschiedshypothesen“) und

Zusammenhangshypothesen bei *Korrelationsrechnung*

(siehe „Verfahren zur Überprüfung von Zusammenhangshypothesen“)

Gerichtete und ungerichtete Hypothesen:

Gerichtete Hypothesen haben irgendeine Richtungsvorgabe, d.h. es sind Formulierungen z.B. „größer, besser, schneller, schlechter, wandelbarer, etc. als“ vorhanden.

Ungerichtete Hypothesen haben keine Richtungsvorgabe, sie postulieren nur irgendeinen Unterschied. Typische Formulierungen wären: „ist ungleich, verschieden, anders, etc. zu“

Spezifische und unspezifische Hypothesen:

Wenn wir eine bestimmte Größe bei einer Hypothese angeben, haben wir eine **spezifische Hypothese**. Z.B. „Es ist (mindestens) um den Betrag x besser als vorher.“ (= Unterschiedshypothese) oder „Der Zusammenhang (=Korrelation) beträgt (mindestens) x.“ (=Zusammenhangshypothese).

Unspezifische Hypothesen haben demnach keine expliziten Angaben.

- Spezifische Hypothesen kommen in der psychologische Forschung meistens nur in Verbindung mit gerichteten Hypothesen vor.

Die Alternativhypothese sollte – soweit sich das inhaltlich rechtfertigen lässt – so präzise wie möglich formuliert sein. Die wenigsten Vorkenntnisse verlangt eine

unspezifische ungerichtete Hypothese	gefolgt von einer
unspezifischen gerichteten Hypothese	gefolgt von einer
spezifischen gerichteten Hypothese	

Wie das bei einer spezifischen ungerichteten Hypothese (Es unterscheidet sich mindestens um x.) aussieht, war nicht erwähnt (ist wohl auch eher selten.).

Statistische Hypothesen:

- Zur Überprüfung müssen wissenschaftliche Hypothesen **vor** einer Untersuchung aufgestellt und in eine statistische Hypothese überführt werden.
- Hypothesen werden für Populationen, also mit Parametern formuliert.
- Diese beiden Punkte gelten auch für die Nullhypothese.

Alternativhypothesen heißen üblicherweise H_1 .

Bei einer Unterschiedshypothese (hier Mittelwert) heiße sie z.B.

„Psychologiestudenten sind mit diesem neuen Skript (μ_1) besser in der Methodenlehreprüfung als vorher (μ_0).“ Mathematisch formuliert:

$$H_1 : \mu_0 < \mu_1$$

Bei einer Zusammenhangshypothese heiße sie z.B.

„Psychologiestudenten werden eher wahnsinnig (ρ_1) (korrelieren mehr mit Irrenhausbesuchen) als andere Studenten (ρ_0).“ Mathematisch formuliert:

$$H_1 : \rho_0 < \rho_1$$

Die Nullhypothese:

In Abhängigkeit zur Alternativhypothese wird eine konkurrierende Hypothese gebildet, die sogenannte **Nullhypothese (H_0)**. Die **Nullhypothese** ist eine Negativhypothese, die **behauptet, dass die zur H_1 komplementäre Aussage richtig sei. Die Nullhypothese ist die Basis der klassischen Prüfstatistik, da von dort aus entschieden wird, ob die H_1 stimmt oder nicht.**

(Mehr dazu im Bortz S. 109 & 110)

Fehlerarten bei statistischen Entscheidungen: Alpha- & Beta-Fehler:

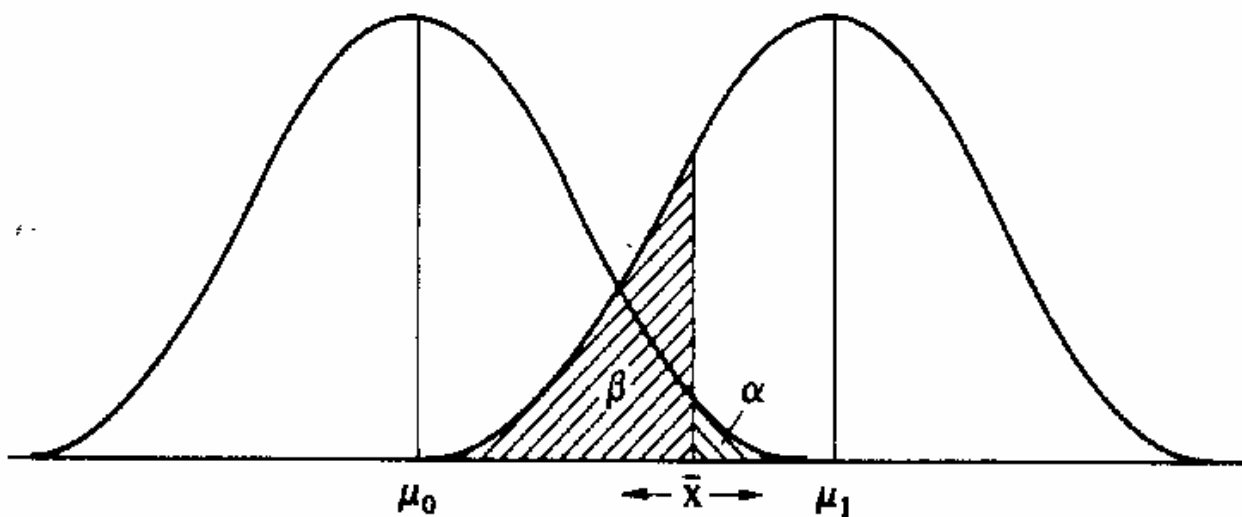
Stehen Null- und Alternativhypothese fest, kann die Untersuchung beginnen. Aber welche der Hypothesen gilt nun? **Die Entscheidung hierüber wird noch erschwert, da sich das Ergebnis der Untersuchung auf eine Stichprobe bezieht, während wir eigentlich Aussagen über eine Population treffen wollen.**

Damit ist nicht auszuschließen, dass wir aufgrund der Stichprobenauswahl Fehler machen: Insgesamt gibt es 4 Möglichkeiten:

		In der Population („in Wahrheit“) gilt ...	
		H_0	H_1
Entscheidung für ...	H_0	Richtige Entscheidung	β -Fehler
	H_1	α -Fehler	Richtige Entscheidung

α -Fehler oder Fehler erster Art: Die H_1 wird angenommen, obwohl eigentlich die H_0 gilt.

β -Fehler oder Fehler zweiter Art: Die H_0 wird beibehalten, obwohl eigentlich die H_1 gilt.



α - und β -Fehlerwahrscheinlichkeit bei einer (rechts-) gerichteten Hypothese.

Signifikanzaussagen:

Signifikanzaussagen sind Aussagen, ob eine Hypothese – meist die H_1 – auf der Basis festgelegter Grenzen (meist bei 5% oder 1% Irrtumswahrscheinlichkeit) gilt oder nicht.

Da Stichprobenergebnisse (zufällig) stark von der Population abweichen können, setzt man Signifikanzen fest, um zu einer bestimmten Wahrscheinlichkeit eine Hypothese (Aussage) beweisen zu können.

Man bezeichnet ein **Signifikanzniveau von $\alpha = 5\%$ als signifikant**, ein **Signifikanzniveau von $\alpha = 1\%$ als sehr signifikant**. (Das ist so festgelegt worden.)

Inwieweit man α festlegt, ist eine methodische Sache (Brisanz der Untersuchung => kleines α , junge Wissenschaft mit innovativen Ideen => großes α)

Da die H_0 die Basis der klassischen Prüfstatistik ist und wir die H_1 beweisen wollen, ist für uns (erst einmal) der α -Fehler von Bedeutung. Deswegen wird erst einmal die Wahrscheinlichkeit des α -Fehlers – die sogenannte **Irrtumswahrscheinlichkeit** – bestimmt.

Bestimmung der Irrtumswahrscheinlichkeit:

Irrtumswahrscheinlichkeiten sind bedingte Wahrscheinlichkeiten, d.h. Wahrscheinlichkeiten für das Auftreten eines Ereignisses (hier das Ergebnis unserer Stichprobenuntersuchung) unter der Bedingung, dass die Nullhypothese zutrifft.

Aus der Population ziehen wir (theoretisch) unendlich viele Stichproben – es ergibt sich eine SKV mit dem Mittelwert μ_0 , der dem der Population entspricht.

Bei dieser SKV trennen wir (bei rechtsseitiger H_1 & $\alpha = 0,05$) rechts 5% ab.

Diese Teilfläche stellt den Bereich dar, bei der wir uns zu α % (hier 5%) irrtümlicherweise zugunsten der H_1 entscheiden würden. Die Größe der Irrtumswahrscheinlichkeit wird mit **P gekennzeichnet.**

Mittels einer z-Transformation lässt sich dann der genaue (Grenz-) Wert errechnen.

Nun ändern wir bei einer Stichprobe die Bedingungen, damit unsere H_1 wirkt (z.B. H_1 : neue Lehrmethode => die wird nun angewandt). Landet der Mittelwert der Stichprobe im Bereich P, nehmen wir die H_1 an.

Das Ganze war die Theorie.

Da aber die Kosten ins Astronomische steigen würden (unendlich viele Stichproben), gehen wir von einer NV der Mittelwerte aus (die SKV soll also eine NV sein).

Dann berechnen wir eine z-Transformation mit dem Wert unserer – unter H_1 -Bedingungen getesteten – Stichprobe. Diesen z-Wert schlagen wir in der Tabelle nach und sehen, wie signifikant er ist. Ist er geringer als unsere Signifikanzniveau α , können wir die H_1 annehmen.

Beispiel: Nach der z-Transformation unser Experimentalergebnisse

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \quad [\sigma_{\bar{x}} \text{ kann auch geschätzt sein.}]$$

erhalten wir den z-Wert 2,29.

Dieser Wert schneidet links von sich 98,90% aller z-Werte ab. D.h. bei unserem rechtsseitigem Test und einem Signifikanzniveau von $\alpha = 5\%$ könnten wir die H_1 annehmen, bei einem Signifikanzniveau von $\alpha = 1\%$ müsste die H_0 beibehalten werden (obwohl sie nur zu 1,1% stimmen würde).

- **Ein nicht-signifikantes Ergebnis heißt nicht, dass die H_0 richtig ist;** z.B. bei einer Verfehlung des 5%-Niveaus um 1% ist die H_0 nur mit einer Wahrscheinlichkeit von 6% richtig !! Deswegen sagt man auch „Beibehaltung“ statt „Annahme“ der H_0 .

Die Irrtumswahrscheinlichkeit α ist immer vorhanden, sie verringert sich bei

- 1.) größer werdender Diskrepanz von $\bar{x} - \mu_0$ (je ausgefallener die Stichprobe)
- 2.) Vergrößerung des Stichprobenumfangs: nur wenn n so groß wie die Population ist, ist die Entscheidung für H_1 oder H_0 sicher (aber da alles bekannt ist fällt sie so und so weg).
- 3.) kleiner werdender Populationsstreuung σ bzw. $\hat{\sigma}$ (je weniger sich die Leute unterscheiden.)

Die statistische Hypothesentestung führt zu keinen Wahrheiten, nur zu Wahrscheinlichkeiten.

Unspezifische Nullhypothesen:

Meist haben wir eine spezifische H_0 , also $\mu_1 = \mu_0$. Es kann aber auch unspezifische Nullhypothesen geben, also bei einem rechtsseitigem Test $H_0: \mu_1 > \mu_0$. Hat man jedoch festgestellt, dass $\mu_1 = \mu_0$ verworfen werden kann, so gilt das auch für $\mu_1 > \mu_0$.

Es genügt also, wenn eine unspezifische H_1 an einer spezifischen H_0 getestet wird.

Einseitige und zweiseitige Tests:

Zweiseitige Tests sind genauer als Einseitige, da das α -Niveau konstant bei z.B. 1% liegt (es wird also 1% der Verteilung abgeschnitten), und eine zweiseitige Testung durchgeführt wird (es werden also auf beiden Seiten $\alpha/2$ % abgeschnitten).

Zweiseitige Tests haben den Nachteil, dass sie keine Richtungen, sondern nur Ungleichheiten formulieren.

Einseitige Test werden eher signifikant als Zweiseitige – was in der psychologischen Forschung ja auch gerne gesehen wird.

Kann nicht klar entschieden werden, ob ein Sachverhalt besser durch eine gerichtete (einseitig) oder ungerichtete (zweiseitig) Hypothese erfasst wird, muss in jeden Fall zweiseitig getestet werden.

Statistische Signifikanz und praktische Bedeutsamkeit:

Ist n hinreichend groß und also $\hat{\sigma}_{\bar{x}}$ entsprechend klein, wird jeder noch so kleine hypothesenkonforme Unterschied signifikant.

Ein Beispiel: eine neue Lehrmethode führte bei 100 Schülern zu durchschnittlich 42 Punkten im Abschlusstest. Der Mittelwert aller Tests vor der neuen Lehrmethode betrug 40 Punkte, die Streuung 0,8 Punkte. Die H_1 (die neue Lehrmethode ist besser) wäre hierbei auf einem 1% tigen Alphaniveau signifikant.

Wenn wir jetzt statt 100 Schülern 10.000 untersuchen würden, wäre selbst ein Unterschied von 0,19 Punkten bei einem 1% tigen Alphaniveau signifikant

Aber hat das noch irgendeine Bedeutung?

Umgekehrt gilt auch, dass große (und bedeutsame) Unterschiede bei zu kleinen Stichproben nicht signifikant werden, also „unentdeckt“ bleiben.

Deswegen benötigt man Verfahren, die praktische Bedeutsamkeit von Untersuchungen erkennen zu können. Das geht über die Feststellung der:

- Effektgrößen und
- Teststärke

Die korrekte Anwendung eines Signifikanztests und die Interpretation der Ergebnisse unter dem Blickwinkel der praktischen Bedeutsamkeit sind essentielle und gleichwertige Bestandteile der empirischen Hypothesenprüfung.

Effektgröße:

Der in der H_1 behauptete Unterschied wird als Effekt bezeichnet. Um Effekte verschiedener Untersuchungen vergleichbar zu machen, werden Effekte zu Effektgrößen standardisiert:

$$\varepsilon = \frac{\mu_1 - \mu_0}{\hat{\sigma}}$$

Das Epsilon (ε) steht für Effektgröße.

Die Effektgröße gibt also an, inwieweit – in einem speziellen Test – μ_1 von μ_0 entfernt sein sollte, um von einem praktisch bedeutsamen Effekt sprechen zu können.

Wenn man vor einer Untersuchung eine Effektgröße festlegt, hat dies den Vorteil (neben dem größeren Aufwand), dass der Stichprobenumfang kalkulierbar ist:

$$n = \frac{(z_{(1-\alpha)} - z_{\beta})^2}{\varepsilon^2}.$$

Um eine geeignete Teststärke zu erzielen, sollte vor einer Untersuchung eine Effektgröße festgelegt und n darauf abgestimmt werden.

Der Beta-Fehler:

Die Festlegung des β -Fehlers erfolgt wie die Festlegung des α -Fehlers (zu finden unter dem Punkt „Bestimmung der Irrtumswahrscheinlichkeit“), nur dass das Ganze mit vertauschten Hypothesen stattfindet.

D.h. eigentlich sollte man unendlich viele H_1 -Stichproben (= unter der Bedingung H_1) durchführen, man geht aber von einer NV aus und macht einen z-Test.

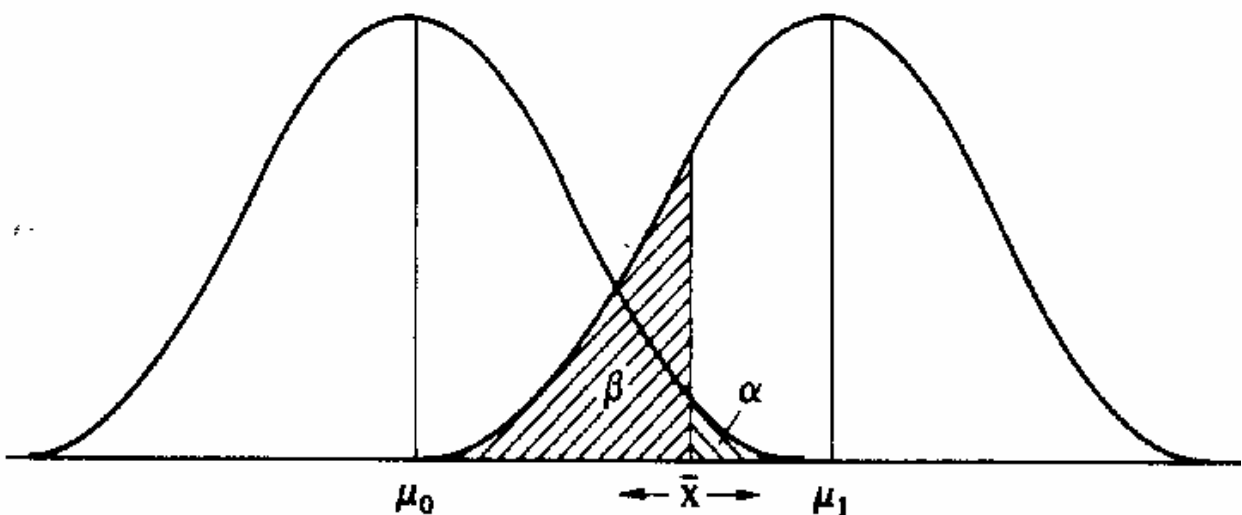
(z.B. bei $H_1: \mu_1 \geq 110$ wäre $\mu_1 = 110$)

Der Beta-Fehler stellt den Bereich dar, bei der wir uns zu β % irrtümlicherweise zugunsten der H_0 entscheiden würden.

Unspezifische Alternativhypothesen:

Die Festlegung des β -Fehlers, die mit einer Entscheidung zugunsten der H_0 verbunden ist, kann bei unspezifischen H_1 nicht bestimmt werden.

Hier das Bild einer spezifischen H_1 samt Alpha- & Beta-Fehler-Bereich:



Teils ist der Beta-Fehler durch die Wahl des Alpha-Fehlers festgelegt, d.h. er fängt an der kritischen Grenze (hier links davon) an, die durch die Wahl des Alpha-Niveaus entstanden ist. Der β -Fehler kann aber auch festgelegt sein. So können Indifferenzbereiche entstehen.

Indifferenzbereiche:

Wenn Alpha- und Beta-Niveau festgelegt sind, kann es bei sehr kleinen bzw. sehr großen Effektgrößen (entweder große bzw. kleine Streuung und/oder nahe beieinander bzw. weit entfernt liegenden Mittelwerte von H_1 und H_0) zu Indifferenzbereichen kommen.

Das Stichprobenergebnis befindet sich dann in einem Bereich, für den

- weder die H_0 noch die H_1 abgelehnt werden können (Überschneidung der Kurven) oder
- sowohl die H_0 als auch die H_1 abgelehnt werden müssen (kaum Kurvenüberschneidung, d.h. nur noch die ganz ganz flachen Bereiche der Kurven von H_1 und H_0 überschneiden sich).

Teststärke („power“):

Die Teststärke ist die der korrekten Entscheidung für die H_1 zugeordnete Wahrscheinlichkeit, d.h. **die Teststärke gibt an, mit welcher Wahrscheinlichkeit ein Signifikanztest zugunsten einer spezifischen H_1 entscheidet.**

Die Teststärke hat den Wert $1-\beta$

Da sich α und β gegenseitig beeinflussen, ist die Teststärke für $\alpha = 0,05$ größer als für $\alpha = 0,01$.

Eigenschaften der Teststärke:

Sie vergrößert sich:

- mit wachsender Differenz von μ_0 und μ_1 (also auch mit wachsender Effektgröße)
- bei kleiner werdender Merkmals (=Populations-) streuung.
- mit größer werdendem α (sofern β nicht festgelegt ist).
- mit wachsendem Stichprobenumfang (da der Standardfehler kleiner ist) $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$
- Bei gerichteter Hypothese hat ein einseitiger Test größere Power als zweiseitiger Test.

Bedeutung der Stichprobengröße:

Im Allgemeinen kann man sagen, dass große Stichproben immer günstiger sind als kleine, da sie zu genaueren Ergebnissen führen. Da große Stichproben auch mehr kosten, reicht ein „optimaler“ Stichprobenumfang.

Stichprobenumfänge sind dann optimal, wenn sie bei gegebenem α , β und ε eine eindeutige Entscheidung über die Gültigkeit von H_0 oder H_1 sicherstellen.

Für Stichprobenumfänge, die kleiner sind als der „optimale“, existiert ein \bar{x} -Wertebereich, der sowohl mit H_0 als auch mit H_1 vereinbar ist. Für größere Stichproben gibt es Werte, die weder mit H_0 noch mit H_1 vereinbar sind.

Monte-Carlo-Studien, Bootstrap-Technik:

Es gibt statistische Kennwerte, deren mathematischer Aufbau so kompliziert ist, dass es bis heute nicht gelungen ist, deren Standardfehler auf analytischem Wege zu entwickeln.

In diesen Fällen können sogenannte Monte-Carlo-Studien oder die Bootstrap-Technik eingesetzt werden, mit denen die unbekannte H_0 -Verteilung des jeweiligen Kennwertes auf einem PC simuliert wird.

Die Monte-Carlo-Methode hat zwei (für uns) vorrangige Anwendungsvarianten:

- die Erzeugung der H_0 -Verteilung eines statistischen Kennwertes
(man simuliert per PC eine Population, und der PC zieht immer weiter Stichproben, er simuliert eine SKV. Mit dieser künstlichen SKV kann man überprüfen, ob der reale untersuchte Kennwert signifikant ist oder nicht.)
- die Überprüfung der Folgen, die mit der Verletzung von Voraussetzungen eines statistischen Tests verbunden sind.
(da in der empirischen Forschung oftmals Testvoraussetzungen verletzt werden, kann man einen Test auf seine „Robustheit“ testen, d.h. beeinflusst die Verletzung der Testvoraussetzungen das Ergebnis des Tests oder nicht.)

Monte-Carlo-Studien sind für die empirische Forschung äußerst wichtig, da sie –teils – die Entscheidung darüber erleichtern, wann ein bestimmter Test eingesetzt werden kann oder nicht.

Die Bootstrap-Technik:

Die Bootstrap-Technik ist den Monte-Carlo-Studien sehr ähnlich; sie unterscheidet sich von der Monte-Carlo-Studie in einem wesentlichen Punkt:

Die Monte-Carlo-Studien kommen zu generalisierbaren Ergebnissen, die Ergebnisse der Bootstrap-Techniken beziehen sich immer nur auf eine bestimmte Untersuchung.

Die Bootstrap-Technik simuliert keine Population, sie verwendet ausschließlich Informationen einer empirisch untersuchten Stichprobe mit dem Ziel, eine Vorstellung über die Variabilität des zu prüfenden Stichprobenkennwertes zu gewinnen.

Die Bootstrap-Technik zieht aus den einzelnen Werten der empirischen Stichprobe immer neue Stichproben, bildet davon Mittelwert und Varianz und berechnet schließlich ein Konfidenzintervall für diese Stichproben. Daran wird dann die H_0 überprüft.

Verfahren zur Überprüfung von Unterschiedshypothesen:

Bei den Verfahren zur Überprüfung von Unterschiedshypothesen werden normalerweise zwei Stichprobenergebnisse miteinander verglichen; Vergleiche eines Stichprobenmittelwertes mit einem Populationsmittelwert gibt es aber auch (meist ob ein Stichprobenkennwert zu einer Population gehört oder nicht).

Dabei gibt es eine Menge von Testverfahren. Bevor man eines auswählt, muss noch das Skalenniveau festgelegt werden; kann man sich nicht zwischen zweien entscheiden, sollten Tests (also unterschiedliche Tests, die dasselbe messen) für beiden Skalenniveaus durchgeführt werden. Sind die Ergebnisse gleich, kann das höhere Skalenniveau nicht angezweifelt werden. Unterscheiden sich die Ergebnisse, muss das anforderungslosere Skalenniveau gewählt werden.

Verfahren für Intervalldaten:

Die folgenden Verfahren wie z-Test, t-Test, F-Test und χ^2 -Test für eine Varianz, vorgestellt in den Punkten:

- „Vergleich eines Stichprobenmittelwertes mit einem Populationsparameter“
- „Vergleich zweier Stichprobenmittelwerte aus unabhängigen Stichproben“
- „Vergleich zweier Stichprobenmittelwerte aus abhängigen Stichproben“
- „Vergleich einer Stichprobenvarianz mit einer Populationsvarianz“
- „Vergleich zweier Stichprobenvarianzen“

sind Verfahren für Intervalldaten.

Vergleich eines Stichprobenmittelwertes mit einem Populationsparameter (z-Test, t-Test):

Eine Zufallsstichprobe des Umfangs n mit einem Mittelwert \bar{x} wird berechnet.

Nun soll die Hypothese geprüft werden, ob die Stichprobe zu einer Grundgesamtheit mit bekanntem Populationsparameter μ_0 gehört.

Aufstellen der Hypothesen:

$$H_0: \mu_0 = \mu_1$$

$$H_1: \mu_0 \neq \mu_1$$

Bei $n \geq 30$ und bekannter Populationsvarianz σ^2 (und somit bekanntem Standardfehler) sowie bei $n \geq 50$ und geschätzter Populationsvarianz $\hat{\sigma}^2$ (und somit geschätztem Standardfehler) kann man einen **z-Test** durchführen:

Also entweder

$$z_{emp} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \quad \text{oder} \quad z_{emp} = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}}$$

Man berechnet z_{emp} und z_{krit} . Wenn $z_{\text{emp}} > z_{\text{krit}}$, nehmen wir (bei rechts- und ungerichteter H_1 , also auch hier) die H_1 an.

Zu beachten ist, dass sich das α bei z_{krit} bei zweiseitigen Tests halbiert.

Kleine Stichproben:

Erfüllen unsere Stichprobe und die Population die oben genannten Kriterien für den z-Test nicht, kann man auf einen **t-Test** ausweichen.

Die Hypothesenaufstellung und das Verfahren erfolgen wie beim z-Test, nur halt jetzt mit t-Werten statt z-Werten.

Nicht zu vergessen:

Werden Stichproben des Umfangs n aus einer normalverteilten Population (Grundgesamtheit) gezogen, verteilen sie sich entsprechend einer t-Verteilung mit $n-1$ Freiheitsgraden (df = degrees of freedom).

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}}, \text{ mit } df = n-1.$$

Jetzt werden t_{emp} und t_{krit} berechnet, nur abreitet man mit der t-Tabelle statt der z-Tabelle, der Rest läuft wie beim z-Test.

Zur Erinnerung: je nach t-Test (links- / rechtsseitig / ungerichtet) ist:

$$t_{\text{krit}} = t_{(df; \alpha)} \quad \text{oder} \quad t_{\text{krit}} = t_{(df; 1-\alpha)} \quad \text{oder} \quad t_{\text{krit}} = t_{(df; \alpha/2)}$$

Anzahl der Freiheitsgrade:

Freiheitsgrad = Die Anzahl der bei der Berechnung des Kennwertes frei variierbaren Werte.

Die Anzahl der Freiheitsgrade hängt davon ab, wie viel Werte noch frei variieren können. D.h., wenn 4 von 5 Werten bei einer Additionsaufgabe und das Ergebnis feststehen, dann ist der Wert des 5. Wertes festgelegt, ergo unfrei.

Beispiel: die 4 Werte ergeben aufaddiert 9. Das Endergebnis ist 10. Welche Zahl hat wohl der 5. Wert?

Vergleich zweier Stichprobenmittelwerte aus unabhängigen Stichproben (t-Test):

Sind bei Gruppenvergleichen die Daten beider Populationen unbekannt, vergleicht man anhand von Stichproben.

Der t-Test für unabhängige Stichproben prüft zwei voneinander unabhängige Stichproben des Umfangs n_1 und n_2 , die aus zwei verschiedenen Grundgesamtheiten gezogen wurden.

Die Hypothesen lauten:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0 \text{ (ungerichtet)}$$

[Nullhypothese ausformuliert: Wenn sich beide Populationen nicht unterscheiden, müsste bei (unendlich) vielen Stichproben aus beiden Populationen auch kein Unterschied zwischen den Mittelwerten der beiden SKVen bestehen.]

Die eigentliche Formel lautet (ist eine Art z-Transformation für zwei Stichproben):

$$t_{emp} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}} \quad df = n_1 + n_2 - 2$$

Denn im Nenner steht erneut nur der Standardfehler, nur diesmal für die Differenz aus beiden Stichproben.

Und dieser Standardfehler berechnet sich aus dieser einfachen formschönen Formel:

$$\begin{aligned} \hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} &= \sqrt{\frac{(n_1 - 1) \cdot \hat{\sigma}_1^2 + (n_2 - 1) \cdot \hat{\sigma}_2^2}{(n_1 - 1) + (n_2 - 1)}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\ &= \sqrt{\frac{\sum (x_{i1} - \bar{x}_1)^2 + \sum (x_{i2} - \bar{x}_2)^2}{(n_1 - 1) + (n_2 - 1)}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \end{aligned}$$

(Beide Versionen – also die obere und untere – sind mögliche Abwandlungen. Die obere Version kann man anwenden, wenn beide Populationsvarianzen schon geschätzt wurden.)

Voraussetzungen (für einen t-Test für unabhängige Stichproben):

- Die Verteilungen der Stichprobenmittelwerte sind NV, bei kleineren Stichproben müssen die Populationen (goodness-of-fit-Test) normalverteilt sein.
- homogene Varianzen
- Unabhängigkeit der Stichproben
- Gleiche Stichprobengrößen (im Allgemeinen)

Noch etwas: ab $n_1 + n_2 \geq 50$ kann man auch in der z-Tabelle statt in der t-Tabelle nachsehen.

Vergleich zweier Stichprobenmittelwerte aus abhängigen Stichproben (t-Test):

Abhängig sind Stichproben, deren Elemente paarweise einander zugeordnet sind. Dies geschieht durch Messwiederholung oder Zuordnung (Parallelisierung) nach sinnvollen Kriterien:

Beispiele für „paarweise“ wären:

- Intelligenz wird bei denselben Probanden zweimal erhoben und es soll untersucht werden, ob sich der Intelligenztestwert von Messung zu Messung verändert hat (**Messwiederholung**).
- Intelligenz wird bei Zwillingen gemessen.
- Stichprobengruppen werden **parallelisiert** (matched samples)

Beim t-Test für abhängige Stichproben wird berücksichtigt, dass die Varianz des ersten Messzeitpunktes die des zweiten beeinflusst (oder umgekehrt oder beides).

Dadurch können Mehrfachmessungen (z.B. durch Vorwissen bei einem Wissenstest) vermieden werden.

Voraussetzungen (für einen t-Test für abhängige Stichproben)

- Die Stichprobenverteilung der mittleren Differenzen ist eine NV (uns interessiert also nicht die Verteilung von Mittelwerten im möglichst vielen Stichproben, sondern die Verteilung der Differenzen der Mittelwerte.).
- Bei kleineren Stichproben müssen die Populationen normalverteilt sein.

Die Hypothesen lauten:

ungerichtet	rechtsseitig	linksseitig
$H_0: \mu_d = 0$	$H_0: \mu_d \leq 0$	$H_0: \mu_d \geq 0$
$H_1: \mu_d \neq 0$	$H_1: \mu_d > 0$	$H_1: \mu_d < 0$

Vorgehen:

Zunächst werden die Differenzen zwischen den Messwerten der ersten und der zweiten Erhebung berechnet:

$$d_i = x_{i1} - x_{i2}.$$

Dann wird der Mittelwert der Differenzen berechnet:

$$\bar{x}_d = \frac{\sum_{i=1}^n d_i}{n}, \text{ mit } n: \text{ Anzahl der Messwertpaare.}$$

Dann der geschätzte Standardfehler ermittelt:

$$\hat{\sigma}_{\bar{x}_d} = \sqrt{\frac{\hat{\sigma}_d^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{x}_d)^2}{n-1} \cdot \frac{1}{n}} = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{x}_d)^2}{n \cdot (n-1)}}$$

Die eigentliche Formel lautet (ist eine Art z-Transformation ohne μ_d , da $\mu_d = 0$):

$$t_{emp} = \frac{\bar{x}_d}{\hat{\sigma}_{\bar{x}_d}}, \text{ mit } df = n-1 \text{ und dem Standardfehler}$$

Statistische Entscheidung (für alle t-Tests):

	ungerichtet	rechtsseitig	linksseitig
Annahme der H_1 , falls ...	$ t_{emp} > t_{(df; \alpha/2)} $	$t_{emp} > t_{(df; 1-\alpha)}$	$t_{emp} < t_{(df; \alpha)}$

Bei einem gerichteten t-Test für abhängige Stichproben ist darauf zu achten, dass die inhaltliche und die statistische Alternativhypothese dieselbe Richtung behaupten; das hängt davon ab, wie man die Differenzen bildet (der Fehler passiert bei Aufgabenblatt 10 in B2).

Deswegen bei $d_i = x_{i1} - x_{i2}$ immer die H_1 -Werte bei x_1 einsetzen (dann behaupten die inhaltliche und die statistische Alternativhypothese immer dieselbe Richtung.).

Aus Monte-Carlo-Studien geht hervor, dass der t-Test für abhängige Stichproben wie auch der t-Test für unabhängige Stichproben äußerst robust auf Verletzungen seiner Voraussetzungen reagiert.

Bei einem t-Test für unabhängige Stichproben geht das bei

- ähnlichen Grundgesamtheiten, möglichst eingipflig und symmetrisch
- großen Stichprobenunterschieden, solange die Varianzen gleich sind.

Bei einem t-Test für abhängige Stichproben sollte

- die Korrelation positiv sein.

Vergleich einer Stichprobenvarianz mit einer Populationsvarianz (chi²-Test für eine Varianz):

Manchmal kann die Varianz einer Stichprobe stark von der der Population abweichen, der Mittelwert z.B. aber identisch sein. U.a. aufgrund dessen kann es interessant sein, Varianzvergleiche durchzuführen.

Der folgende Test – ein χ^2 -Test für eine Varianz – überprüft, ob eine Stichprobe aus einer Population mit der Varianz $\sigma^2 = a$ stammt.

Voraussetzung:

- Sofern die Messwerte (der Population & der Stichprobe) normalverteilt sind, verteilt sich die Prüfgröße (=die eigentliche Formel) gemäß einer χ^2 -Verteilung mit $n-1$ Freiheitsgraden.

Die Nullhypothese (ist immer so):

$$H_0: \sigma^2 = a$$

Die Alternativhypothesen lauten:

	ungerichtet	rechtsseitig	linksseitig
falls ...	$\chi_{\text{emp}}^2 < \chi_{df, \alpha/2}^2$ oder $\chi_{\text{emp}}^2 > \chi_{df, 1-\alpha/2}^2$	$\chi_{\text{emp}}^2 > \chi_{df, 1-\alpha}^2$	$\chi_{\text{emp}}^2 < \chi_{df, \alpha}^2$

Vorgehen:

Standardfehler ausrechnen, Werte einsetzen nach folgender Formel.

Die eigentliche Formel lautet:

$$\chi_{\text{emp}}^2 = \frac{(n-1) \cdot \hat{\sigma}^2}{a}, \text{ mit } df = n-1.$$

Statistische Entscheidung:

	ungerichtet	rechtsseitig	linksseitig
Annahme der H_1 , falls ...	$H_1: \sigma^2 \neq a$	$H_1: \sigma^2 > a$	$H_1: \sigma^2 < a$

Vergleich zweier Stichprobenvarianzen (F-Test):

Der Vergleich von Stichprobenvarianzen mit der Populationsvarianz kommt in der Praxis nicht so häufig vor. Es werden eher zwei Stichprobenvarianzen miteinander verglichen. Dies geschieht über einen F-Test.

Voraussetzungen

- Normalverteilung der Grundgesamtheiten,
- Unabhängigkeit der Stichproben.

Die Hypothesen lauten:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \quad (\text{ungerichtet})$$

$$H_1 : \sigma_1^2 > \sigma_2^2 \quad (\text{gerichtet})$$

[Die H_0 ausformuliert: es gibt keinen Unterschied in den Populationsvarianzen.]

[Diese Hypothese wird nun anhand von Stichprobenvarianzen überprüft.]

Bei einem einseitigen Test (es kommen fast nur einseitige F-Tests vor) kommt die als größer erwartete Varianz in den Zähler, bei einem zweiseitigen Test (selten) kommt die empirisch größere Varianz in den Zähler.

Die eigentliche Formel lautet:

$$F_{\text{emp}} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \quad \text{mit } df_{\text{Zähler}} = n_1 - 1 \text{ und } df_{\text{Nenner}} = n_2 - 1$$

(Da die Populationsvarianzen nicht bekannt sind, schätzen wir diese.)

Statistische Entscheidung:

	ungerichtet	gerichtet*
falls ...	$F_{\text{emp}} > F_{n_1-1, n_2-1, 1-\alpha/2}$	$F_{\text{emp}} > F_{n_1-1, n_2-1, 1-\alpha}$
Annahme der ...	$H_1: \sigma_1^2 \neq \sigma_2^2$	$H_1: \sigma_1^2 > \sigma_2^2$

* es gibt nur rechtsseitige gerichtete Hypothesen.

Verfahren für Ordinaldaten:

Werden in bei den Verfahren für Intervalldaten einige Voraussetzungen nicht erfüllt, oder hat man in der Untersuchung nur Rangreihen erhoben, muss man auf Verfahren für Ordinaldaten zurückgreifen.

Verfahren für Ordinaldaten sind „verteilungsfrei“, sie vergleichen nur Relationen (= Ordinaldaten) = Rangdaten.

Die folgenden Verfahren U-Test und Wilcoxon-Test, vorgestellt in den Punkten:

„Vergleich von zwei unabhängigen Stichproben hinsichtlich ihrer zentralen Tendenz“ und
 „Vergleich von zwei abhängigen Stichproben hinsichtlich ihrer zentralen Tendenz“

sind Verfahren für Ordinaldaten.

Randomisationstest nach Fisher:

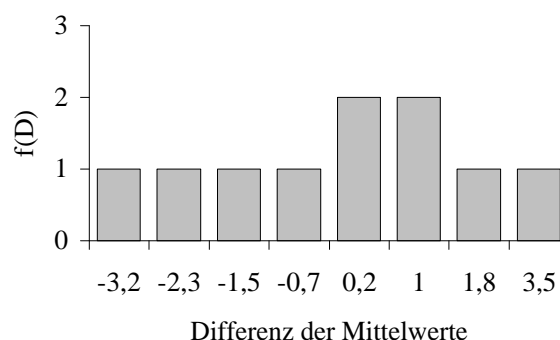
Da die hier vorgestellten Verfahren (U-Test & Wilcoxon-Test) auf der Randomisationsidee von Fisher basieren, wird diese kurz erklärt (am Beispiel unabhängiger Gruppen).

Das ganze Verfahren steht nicht im Bortz (in B3-Folien) und ist auch nur als Einleitung gedacht.

In einem Experiment wurden die Vbn zufällig auf zwei Gruppen (EG und KG) aufgeteilt. Die H_0 wäre: Die Daten stammen aus derselben Population.

Fishers erstellt eine diskrete Verteilung einer interessierenden Prüfgröße (z.B. $D = \bar{x}_1 - \bar{x}_2$), die auf allen möglichen Aufteilungen aller Messwerte auf die Gruppen basieren. Im Gegensatz zur normalen Stichprobenverteilung, konstruiert Fisher eine Verteilung von Prüfgrößen, die auf Ziehungen **ohne** Zurücklegen basiert.

Konkret geht Fisher alle möglichen Kombinationen durch und bildet pro Gruppe jeweils den Mittelwert. Dann subtrahiert er den einen Mittelwert vom anderen, und erhält eine Verteilung.



Diese Verfahren für Ordinaldaten erstellt also eine Häufigkeitstabelle, an der man mit simpler Wahrscheinlichkeitsrechnung feststellen kann, wie groß die Irrtumswahrscheinlichkeit eines gemessenen Wertes ist.

Vergleich von zwei unabhängigen Stichproben hinsichtlich ihrer zentralen Tendenz (U-Test):

Wenn wir die Beeinträchtigung der Reaktionszeit unter Alkohol mit der Beeinträchtigung der Reaktionszeit unter Alkohol unter Einnahme des Präparates A (soll den negativen Effekt des Alkohols neutralisieren) vergleichen wollen, können wir nicht davon ausgehen, dass die Reaktionszeiten normalverteilt sind.

Dementsprechend können wir nur ein Verfahren für Ordinaldaten anwenden, den sogenannten **U-Test** (der funktioniert auch für ungleich große Stichproben).

Der U-Test ist so etwas ähnliches wie der t-Test für unabhängige Stichproben

Als Beispiel:

Wir messen bei insgesamt 27 Personen die Reaktionszeiten, bei 12 unter Alkohol und bei 15 unter Alkohol unter Einnahme des Präparates A.

Mit Alkohol		Mit Alkohol und Präparat A	
Reaktionszeit (ms)	Rangplatz	Reaktionszeit (ms)	Rangplatz
85	4	96	10
106	17	105	16
118	22	104	15
81	2	108	19
138	27	86	5
90	8	84	3
112	21	99	12
119	23	101	13
107	18	78	1
95	9	124	25
88	7	121	24
103	14	97	11
		129	26
		87	6
		109	20
	$T_1 = 172$		$T_2 = 206$

T_x = Summe der Rangplätze der Bedingung x

\bar{R} = Durchschnitt (Mittelwert) der Rangplätze

$\bar{R}_1 = 14,33$

$\bar{R}_2 = 13,73$

Bestimmung des U-Wertes:

Am Beispiel:

Die erste Person in Gruppe 1 hat den Rangplatz 4. In Gruppe zwei gibt es 13 Personen mit einem höheren Platz. Die zweite Person in Gruppe 1 hat den Rangplatz 17, der von 5 aus Gruppe 2 übertroffen wird.

Die Summe der Rangplatzüberschreitungen ist der U-Wert (also hier: 13+5+..... etc.).

Die Summe der Rangplatzunterschreitungen wird mit U' bezeichnet: $U = n_1 \cdot n_2 - U'$

Da das zählen sehr langwierig ist, geht das auch mit dieser Formel:

$$U = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - T_1$$

Eingesetzt erhalten wir auf das Beispiel bezogen $U = 86$ und $U' = 94$. [U' erhält man, wenn man in der Formel im Zähler n_2 statt n_1 und am Ende T_2 statt T_1 benutzt.] [Zur Kontrolle kann man $U = n_1 \cdot n_2 - U'$ anwenden.]

Wenn n_1 oder n_2 größer als 10 sind, sind die Werte annähernd normal:

Dann erfolgt die Ermittlung des Populationsmittelwertes und des Standardfehlers:

$$\mu_U = \frac{n_1 \cdot n_2}{2} \quad \sigma_U = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}$$

Die eigentliche Formel lautet dann:

$$z = \frac{U - \mu_U}{\sigma_U}$$

Bei unserem Beispiel erhalten wir $z = -0,20$

Dann entnehmen wir der z-Tabelle, dass auf einem Alphaniveau von 5% dieser (negative) Wert kleiner ist als unser empirischer z-Wert => Beibehaltung der H_0 .

Kleinere Stichproben:

Bei kleineren Stichproben bestimmt man nur U und schaut dann in der U-Tabelle nach.

Wichtig:

- Bei verbundenen Rangplätzen (mehrere Personen auf einem Rangplatz) muss die Streuung korrigiert werden (siehe Bortz S. 148)
- **Beim U-Test muss bei ungerichteter Hypothese das Alphaniveau verdoppelt werden.**

Vergleich von zwei abhängigen Stichproben hinsichtl. ihrer zentralen Tendenz (Wilcoxon-Test):

Der Wilcoxon-Test ist so etwas ähnliches wie der t-Test für abhängige Stichproben.

Zuerst wird die Differenz zwischen den zwei Messungen d_i (z.B. „vorher“ & „nachher“) ermittelt. Danach werden die Rangplätze vom Betrag von d_i aufgestellt.

Dann werden die Rangplätze speziell gekennzeichnet, deren Vorzeichen (das Vorzeichen bei d_i) weniger vorkommt.

Dann werden die Summen der speziell gekennzeichneten und der restlichen Rangplätze aufgestellt.

T = Summe der speziell gekennzeichneten Rangplätze

T' = Summe der restlichen Rangplätze

(Paare mit $d_i = 0$ bleiben unberücksichtigt!)

Verbundenen Rängen (z.B. drei Personen auf Platz 2) werden gemittelte Rangplätze vergeben.

Wenn es **keinen** Unterschied gibt, erwarten wir als T-Wert (nicht t-Wert, sondern „T“):

$$\mu_T = \frac{n \cdot (n + 1)}{4}$$

n = Anzahl der Messwertpaare, wobei $d_i = 0$ unberücksichtigt bleibt.

Je deutlicher der empirische T-Wert (das ist „T“) von μ_T abweicht, desto geringer ist die Chance, dass der gefundene Unterschied zufällig zustande kam.

Wenn wir die Summe der speziell gekennzeichneten Rangplätze (T) ermittelt haben, schauen wir in der T-Tabelle (NEIN, NICHT t-Tabelle) beim kritischen Alphaniveau nach. Ist der T-Wert größer, müssen wir die H_0 beibehalten.

Größere Stichproben:

Bei Stichprobengrößen $n > 25$ geht die T-Verteilung (auch jetzt NICHT t-Verteilung) in eine NV über, so dass man mit der z-Tabelle arbeiten kann.

Dazu muss nur noch die Streuung errechnet werden:

$$\sigma_T = \sqrt{\frac{n \cdot (n + 1) \cdot (2 \cdot n + 1) - \sum_{i=1}^k \frac{t_i^3 - t_i}{2}}{24}}$$

k = Anzahl der Rangbindungen und t_i = Länge der Rangbindung i

Und schon kann man die Formel

$$z = \frac{T - \mu_T}{\sigma_T} \quad \text{angewendet werden.}$$

(Diesen Formelwald muss man aber nicht können.)

Der Wilcoxon-Test setzt im Gegensatz zum U-Test Intervallskalenniveau voraus. Trotzdem ist er ein Verfahren für Ordinaldaten. Wieso?

Der Wilcoxon-Test vergleicht Ränge, er kommt im Prinzip mit Ordinaldaten aus. Das Problem ist nur, dass er Rangplatzdifferenzen bildet, und so Ordinaldatenniveau dafür nicht mehr ausreicht. Trotzdem ist er jedoch ein nichtparametrisches Verfahren und wird somit unter Verfahren für Ordinaldaten gehandelt.

Verfahren für Nominaldaten:

Nominaldaten sind dann nötig, wenn Häufigkeitsunterschiede im Auftreten bestimmter Merkmale bzw. Merkmalskombinationen analysiert werden sollen

Da in fast allen im Bortz in Kapitel 5.3 behandelten Verfahren Prüfstatistiken vermittelt werden, die (annähernd) χ^2 -verteilt sind, werden die Verfahren zur Analyse von Häufigkeiten gelegentlich vereinfachend als χ^2 -Methoden bezeichnet.

 χ^2 -Verfahren dienen also der Analyse von Häufigkeiten.

χ^2 -Verfahren können natürlich auch für ordinal- und intervallskalierte Variablen eingesetzt werden, wenn die Häufigkeit bestimmter Merkmale untersucht werden soll. Die Merkmale werden dann wie nominalskalierte Merkmale behandelt, und somit ist auch das Ordinal- oder Intervallskalenniveau im Eimer.

- Die erwarteten Häufigkeiten repräsentieren immer die jeweilige Nullhypothese.

Übersicht über die χ^2 -Verfahren

	1 Merkmal	2 Merkmale	M Merkmale
2fach gestuft	a) einmalige Untersuchung: eindimensionaler χ^2 -Test b) zweimalige Untersuchung: McNemar χ^2 -Test c) mehrmalige Untersuchung: Cochran-Q-Test	4-Felder χ^2 - Test	Konfigurationsanalyse für alternative Merkmale
mehrfach gestuft	eindimensionaler χ^2 -Test: Vgl. einer empirischen mit einer theoretischen Verteilung	$k \times l$ - χ^2 -Test	Konfigurationsanalyse für mehrfach gestufte Merkmale

Vergleich der Häufigkeit eines zweifach gestuften Merkmals (eindimensionaler χ^2 -Test):**Einmalige Untersuchung:**

An einer Uni sind im Fach Psychologie 869 männliche (Wer's glaubt, wird selig.) und 576 weibliche Studenten immatrikuliert.

Kann man davon ausgehen, dass dieser Unterschied zufällig zustande kam?

Die Antwort auf diese Frage (und somit der ganze Rechenweg) ist davon abhängig, wie die Nullhypothese formuliert ist. Fragen wir danach, ob das Verhältnis männlich/weiblich an der ganzen Uni so ist, oder, ob das Geschlechterverhältnis im allgemeinen im Bereich Psychologie 50 : 50 ist.

Um es (mir) einfacher zu machen, nehme ich die zweite Hypothese (50 : 50) an.

Aufstellen der Hypothesen:

H_0 : Die empirische Verteilung entspricht der erwarteten Verteilung.

H_1 : Die empirische Verteilung weicht von der erwarteten Verteilung ab.

Zeichenerklärung:

f = Häufigkeit

f_{bj} = beobachtete Häufigkeit in Kategorie j und

f_{ej} = erwartete Häufigkeit in Kategorie j ,

k = Anzahl der Kategorien des Merkmals.

p_{ej} = erwartete Wahrscheinlichkeit des Merkmals j .

Dann Abstimmung der erwarteten Häufigkeiten:

$$f_{ej} = p_{ej} \cdot n.$$

D.h. bei unserem Beispiel wäre die Wahrscheinlichkeit für (z.B.) Männer 50%, es gibt insgesamt 1445 Psychologiestudenten, also

$$f_{ej} = 0,5 \cdot 1445 = 722,5$$

Eigentliche Formel:

$$\chi^2_{\text{emp}} = \sum_{j=1}^k \frac{(f_{bj} - f_{ej})^2}{f_{ej}} \quad \text{mit } df = k - 1$$

An dieser Formel erkennt man die Grundstruktur aller χ^2 -Verfahren: Alle χ^2 -Verfahren laufen auf einen Vergleich von beobachteten und erwarteten Häufigkeiten hinaus, wobei die erwarteten Häufigkeiten jeweils die geprüfte Nullhypothese repräsentieren.

Unsere Werte eingesetzt:

$$\chi^2 = \frac{(869 - 722,5)^2}{722,5} + \frac{(576 - 722,5)^2}{722,5} = 59,41$$

Feststellen der Freiheitsgrade:

df = k - 1 es gibt 2 Kategorien (Männer/Frauen), also df = 1

Die Freiheitsgrade eines χ^2 -Wertes entsprechen der Anzahl der Summanden (wie viele Werte gibt es für f_b ? Hier zwei, nämlich Männer und Frauen) **abzüglich der Bestimmungsstücke der erwarteten Häufigkeiten, die aus den beobachteten Häufigkeiten abgeleitet wurden** (es gibt hier nur ein gemeinsames Bestimmungsstück: der Stichprobenumfang n : Die Summe der beobachteten und erwarteten Häufigkeiten ergibt jeweils n).

Feststellen von χ^2_{krit} :

$$\chi^2_{krit} = \chi^2_{(df; \alpha)}$$

Bei einer gerichteten Hypothese in einem χ^2 -Verfahren wird das Alphaniveau beim Nachschlagen in der Tabelle verdoppelt; soll heißen: bei $\alpha = 0,05$ schlage ich in der Tabelle bei 10% nach. Das liegt daran, dass alle χ^2 -Verfahren einseitig getestet werden, denn man betrachtet nur die rechte Seite der Verteilungen.

Die Überprüfung einer gerichteten Hypothese ist bei χ^2 -Verfahren nur möglich, wenn der resultierende χ^2 -Wert einen Freiheitsgrad hat.

Deswegen kann man bei einer gerichteten Hypothese auch mit der STNV testen, sofern man aus dem empirischen und kritischen χ^2 -Wert die Wurzel zieht.

Das Ergebnis eines χ^2 -Tests erhalten wir auch, wenn die Häufigkeit der Alternative 1 (z.B. hier weiblich) über die Binomialverteilung geprüft wird.

Ab $n < 10$ ist das nötig (bei zwei Merkmalen).

Entscheidung:

Falls $\chi^2_{emp} < \chi^2_{k-1, 1-\alpha}$, Beibehaltung der H_0 : Es ist davon auszugehen, dass die Verteilung des untersuchten Merkmals nicht von der unter der H_0 erwarteten Verteilung abweicht.

Entscheidung bei unserem Beispiel:

χ^2_{krit} ist bei uns 3,84; χ^2_{emp} ist bei uns 59,41 => Annahme der H_1

Voraussetzungen:

- Jede untersuchte Einheit kann eindeutig einer Kategorie zugeordnet werden.
- Die erwarteten Häufigkeiten sind in jeder Kategorie größer als 5. Im Falle von $k = 2$ sollten in jeder Kategorie 10 Elemente erwartet werden.

Weitere tolle Modifikationen und Verfahren:**Kontinuitätskorrektur:**

Die Kontinuitätskorrektur, auch Yates-Korrektur genannt, schätzt χ^2 besser, da Häufigkeiten diskret, χ^2 -Werte aber stetig verteilt sind. D.h. jedes $|f_b - f_e|$ wird um $-0,5$ vermindert. Die Kontinuitätskorrektur wirkt der Tendenz nach konservativ (schützt H_0).

Der McNemar-Test:

Dieser Test wird bei zweimaliger Untersuchung und Prozentunterschieden benutzt, **er berücksichtigt nur diejenigen Fälle, bei denen eine Veränderung eingetreten ist.**

Der Cochran-Q-Test:

ist eine Erweiterung des McNemar-Tests, gilt auch für mehrmalige Untersuchungen.
Nur anzuwenden bei $n \cdot m > 30$

(Wer sich das alles – Kontinuitätskorrektur, McNemar-Test & Cochran-Q-Test – antun möchte: steht alles im Bortz auf S. 155-158)

Vergleich der Häufigkeit eines k-fach gestuften Merkmals (eindimensionaler χ^2 -Test):

Diese Variation des eindimensionalen χ^2 -Tests ist mit dem eindimensionalen χ^2 -Test für zweifachgestufte Merkmale ziemlich identisch.

Es gibt zwei Möglichkeiten:

- 1.) **Test auf Gleichverteilung (z.B. je 50% Männer/Frauen) und**
- 2.) **Test auf andere Verteilungsformen.**

Für beide Möglichkeiten gilt:

Die Formel ist mit dem des eindimensionalen χ^2 -Test für zweifachgestufte Merkmale identisch, es gibt halt nur mehr Bedingungen j , ergo wird die Berechnung langwieriger.

$$\chi_{\text{emp}}^2 = \sum_{j=1}^k \frac{(f_{bj} - f_{ej})^2}{f_{ej}}, \text{ mit } df = k-1$$

Bei der Frage, ob sich Variable 1 von den anderen unterscheidet, wird von den Variablen 2 bis k der Durchschnitt gebildet (fungiert als f_e) & mit Variable 1 (als f_b) über den χ^2 -Test verglichen.

Die Voraussetzungen sind ebenso identisch, nur dass man bei weniger als fünf erwarteten Häufigkeiten pro Kategorie (wie beim eindimensionalen χ^2 -Test für zweifachgestufte Merkmale) anstelle der Binomialverteilung nun eine **Multinomialverteilung** rechnen sollte.

All das oben genannt galt noch für beide Möglichkeiten. Die erste Möglichkeit ist hiermit abgeschlossen.

Das Folgende bezieht sich auf:

Tests auf andere Verteilungsformen.

Sofern die H_0 nicht behauptet, dass eine Gleichverteilung vorherrscht (z.B. je 50% Männer/Frauen), haben wir es mit Test für andere Verteilungsformen zu tun, denn **es ergibt sich bei steigender Variablenanzahl eine theoretische Verteilung, die mit einer empirischen (beobachteten) Verteilung verglichen wird.**

Der Vergleich einer empirischen mit einer theoretischen Verteilung heißt Goodness-of-fit-test!

Beim Goodness-of-fit-Test werden die erwarteten Häufigkeiten aus einer theoretischen Verteilung abgeleitet:

Man kann so z.B. prüfen, ob sich eine empirische Verteilung normalverteilt oder z.B. nach der Poisson-Verteilung verteilt.

- Der Goodness-of-fit-Test dient also z.B. der Überprüfung der Voraussetzung, dass Grundgesamtheit normalverteilt (H_0 : normalverteilt, H_1 : nicht normalverteilt)

Problem: die H_0 als „Wunschhypothese“

- wenn der χ^2 -Wert auf 5%-Niveau signifikant ist, heißt das, dass die Wahrscheinlichkeit für eine Normalverteilung der Werte kleiner als 5% ist, wenn es nicht signifikant ist, bedeutet es lediglich, dass es mit mehr als 5% Wahrscheinlichkeit normalverteilt ist - mehr nicht.
Wir wollen aber die H_0 „beweisen“ (sonst will man normalerweise ja die H_1 „beweisen“), deshalb versucht man den Beta-Fehler möglichst klein zu halten, nicht Alpha (ist wurscht).
Da aber die Hypothese in diesem Fall nicht spezifisch ist, kann der Beta-Fehler nicht bestimmt werden, deswegen wird einfach der Alpha-Fehler enorm vergrößert, z.B. auf 25%
Wenn wir auf 25% tigem Alphaniveau testen und die H_1 dann immer noch nicht signifikant ist, haben wir H_0 gesichert.
- Beachtet: die H_0 als Wunschhypothese beizubehalten, wird mit wachsendem Stichprobenumfang unwahrscheinlicher!

Der Goodness-of-fit-Test:

Zuerst werden die Daten in Kategorien eingeteilt, d.h. wenn wir einen Wert haben, können wir ihn in eine Kategorie packen. Ein gutes Beispiel wären die gemessenen km/h der Wagen, eingeteilt in 10er-Kategorien, beginnend von 60km/h bis (sagen wir mal) 220 km/h.
(Das Ganze läuft unter der H_0 , das unsere Kategorien normalverteilt sind.)

Also eine Tabelle:

Intervall	Beobachtete Häufigkeiten	Erwartete Häufigkeiten	$\frac{(f_b - f_e)^2}{f_e}$
60,0-69,9 km/h	4	3	0,33
70,0-79,9 km/h	10	6	2,67
etc.	etc.	etc.	etc.
			Die Summe davon ergibt den empirischen χ^2 -Wert

Der kritische χ^2 -Wert lautet ausgeschrieben $\chi^2_{(k-3, 1-\alpha)}$,

somit sind die Freiheitsgrade $df = k - 3$, wobei k die Anzahl der Kategorien angibt.

Poisson-Verteilung:

Bei einer Frage nach einer Poisson-Verteilung sind die Freiheitsgrade $df = k - 2$.
Vom Prinzip läuft es wie oben, nur werden die erwarteten Häufigkeiten nach der Poisson-Verteilung ermittelt (Formel im Bortz S. 71; der Rest S. 161 & 162)

Vergleich der Häufigkeiten von zwei alternativen Merkmalen (4-Felder- χ^2 -Test):

Werden n voneinander unabhängigen Beobachtungen nicht nur einer, sondern zwei Merkmalsalternativen zugeordnet, erhalten wir eine **4-Felder-Kontingenztafel** bzw. eine **bivariate Häufigkeitsverteilung**.

Und schon sind wir bei einem 4-Felder- χ^2 -Test.

(den kann man auch schon $k \times l$ - χ^2 -Test nennen)

Ein Beispiel: 100 Personen sollen auf die Merkmalsalternativen männlich/weiblich und Brillenträger/Nicht-Brillenträger untersucht werden.

	männlich	weiblich	
mit Brille	25	10	35
ohne Brille	25	40	35
	50	50	100

Jetzt gibt es zwei Möglichkeiten:

- 1.) Vorgegebene erwartete Häufigkeiten und
- 2.) Geschätzte erwartete Häufigkeiten

Vorgegebene erwartete Häufigkeiten:

Dann könnte unsere vorgegebene H_0 lauten:

Brillenträger = 30% in der Bevölkerung

Männer/Frauen ja 50% in der Bevölkerung

Dementsprechend wären unsere Wahrscheinlichkeiten

$p(\text{Brille, Mann}) = 0,15$

$p(\text{Brille, Frau}) = 0,15$

$p(\text{ohne Brille, Mann}) = 0,35$

$p(\text{ohne Brille, Frau}) = 0,35$

Dann setzen wir das ein in folgende Formel:

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(f_{b(i,j)} - f_{e(i,j)})^2}{f_{e(i,j)}}$$

und bekommen ein $\chi^2 = 11,90$

Die Freiheitsgrade $df = 2 \times 2 - 1 = 3$

Dementsprechend ist unser kritisches $\chi^2 = 11,34 \Rightarrow$ Annahme der H_1 , Frauen tragen weniger Brillen, Männer mehr Brillen als erwartet.

Geschätzte erwartete Häufigkeiten:

Beim 4-Felder- χ^2 -Test entsprechen die erwarteten Zellenhäufigkeiten den bei stochastischer Unabhängigkeit zu erwartenden:

$$f_{eij} = p_i \cdot p_j \cdot n = \frac{(\text{Summe Zeile } i) \cdot (\text{Summe Spalte } j)}{n} .$$

Der Rest läuft wie bei vorgegebenen Nullhypothesen, nur sind die Freiheitsgrade $df = 1$ df .

Dementsprechend ist auch das kritische χ^2 anders.

Auch hier wird übrigens die H_1 angenommen (Bortz S. 164)

Kontinuitätskorrektur:

- Auch hier ist eine Kontinuitätskorrektur nach Yates möglich

Prozentunterschiede:

- laufen genauso wie beim eindimensionalen χ^2 -Test

Vergleich der Häufigkeiten von zwei mehrfach gestuften Merkmalen (k x l-chi²-Test):

Im vorherigen Punkt waren beide Merkmale jeweils zweifach gestuft. Beim $k \times l$ - χ^2 -Test ist die Anzahl der Stufen unwichtig

Das Merkmal A ist k-fach gestuft und das Merkmal B l-fach.

Der $k \times l$ - χ^2 -Test überprüft die Alternativhypothese:

- H_1 , dass zwischen zwei Merkmalen ein *irgendwie gearteter* Zusammenhang besteht,

gegen die Nullhypothese

- H_0 , dass es zwischen den beiden Merkmalen keinen Zusammenhang gibt bzw. die Merkmale stochastisch unabhängig sind.

Das Verfahren läuft wie beim 4-Felder- χ^2 -Test:

$$\chi_{\text{emp}}^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(f_{bij} - f_{eij})^2}{f_{eij}}$$

Beim $k \times l$ - χ^2 -Test entsprechen die erwarteten Zellenhäufigkeiten den bei stochastischer Unabhängigkeit zu erwartenden:

$$f_{eij} = p_i \cdot p_j \cdot n = \frac{(\text{Summe Zeile } i) \cdot (\text{Summe Spalte } j)}{n} .$$

Freiheitsgrade:

Bei geschätzten erwarteten Häufigkeiten: $df = (k - 1) \cdot (l - 1)$

Bei vorgegebenen erwarteten Häufigkeiten: $df = (k - 1) \cdot (l - 1) - 1$

Entscheidung

Falls $\chi_{\text{emp}}^2 > \chi_{(k-1)(l-1), 1-\alpha}^2$, Entscheidung für die H_1 : Es ist davon auszugehen, dass die Variablen voneinander stochastisch (irgendwie) abhängig sind.

Voraussetzungen

- Jede untersuchte Einheit kann eindeutig den Kategorien (Zellen) zugeordnet werden.
- Die erwarteten Häufigkeiten sind alle größer als 5.

Der $k \times l$ - χ^2 -Test und der 4-Felder- χ^2 -Test lassen sich auch mit Prozentwerten durchführen.

Vgl. d. Häufigkeiten von m alternativ oder mehrf. gestuften Merkmalen (Konfigurationsanalyse):

Verallgemeinern wir das 4-Felder χ^2 auf m alternative Merkmale, erhalten wir eine **mehrdimensionale Kontingenztafel**, die mit der **Konfigurationsanalyse** analysiert werden kann.

Ein Beispiel:

Wir überprüfen, ob Frauen, die in der Stadt leben, überzufällig berufstätig sind ($\alpha = 0.01$).

Also haben wir drei alternative Merkmale ($m = 3$):

A= Stadt (+) vs. Land (-)

B = männlich (+) vs. weiblich (-)

C = berufstätig (+) vs. nicht berufstätig (-)

Merkmal			Häufigkeiten		
A	B	C	f_b	f_e	$\frac{(f_b - f_e)^2}{f_e}$
+	+	+	120	86,79	12,71
+	+	-	15	63,33	36,88
+	-	+	70	95,32	6,73
+	-	-	110	69,56	23,51
-	+	+	160	89,54	55,45
-	+	-	10	65,34	46,87
-	-	+	20	98,35	62,42
-	-	-	135	71,77	55,71
			$n_b = 640$	$n_e = 640$	$\chi^2 = 300,28$

Die erwarteten Häufigkeiten werden geschätzt mit:

$$f_{e(i,j,k)} = \frac{(\text{Summe } A_i) \cdot (\text{Summe } B_j) \cdot (\text{Summe } C_k)}{n^2}$$

In unserem Beispiel lauten die Summen

$$\begin{array}{lll} A(+) = 315 & B(+) = 305 & C(+) = 370 \\ A(-) = 325 & B(-) = 335 & C(-) = 270 \end{array}$$

$$\text{(z.B. } f_{e(+++)} = 315 \times 305 \times 370 / 640^2 = 86,79)$$

Freiheitsgrade: $df = 1$

Der kritische Wert (5,43) ist kleiner als der Empirische (6,73) für unsere These, dass in der Stadt mehr Frauen berufstätig sind.

Da das aber unsere H_0 war, müssen wir nun die H_0 annehmen bzw. beibehalten (nicht davon verwirren lassen, ist ein Fehler bei der Hypothesenaufstellung).

Allgemeine Bemerkungen zu den χ^2 -Techniken:

χ^2 -Techniken gehören – von der Durchführung her – zu den einfachsten Verfahren der Elementarstatistik, wenngleich der mathematische Hintergrund dieser Verfahren komplex ist.

Mit Hilfe der χ^2 -Verfahren werden die Wahrscheinlichkeiten multinomialverteilter Ereignisse geschätzt (wobei die Schätzung erst bei $n \rightarrow \infty$ exakt übereinstimmen).

Deshalb sollte beachtet werden:

- Die einzelnen Beobachtungen müssen voneinander unabhängig sein (Ausnahme: McNemar- & Cochran-Test)
- Jede untersuchte Einheit muss eindeutig einer Kategorie zugeordnet werden können.
- Nicht mehr als 20% der erwarteten Häufigkeiten in den Kategorien sollten kleiner als 5 sein.

Verfahren zur Überprüfung von Zusammenhangshypothesen:

Die Verfahren zur Überprüfung von Zusammenhangshypothesen haben der u.a. psychologischen Forschung so viel Impulse wie kein anderes statistisches Verfahren verliehen. Zusammenhänge sind in der Mathematik und in den Naturwissenschaften bekannt, z.B. verändert sich der Umfang eines Kreises proportional zu seinem Umfang. Solche exakten, funktionalen Zusammenhänge haben wir in der Psychologie nicht. Deswegen sprechen wir von **stochastischen (zufallsabhängigen) Zusammenhängen**, die nach Höhe des Zusammenhangs unterschiedlich präzise Vorhersagen zulassen. (Bei einer Korrelation von +1 oder -1 geht unser stochastischer Zusammenhang über in einen funktionalen, deterministischen Zusammenhang.)

Im Kapitel „Merkmalszusammenhänge & –vorhersagen“ habe ich einige Punkte aus dem 6. Kapitel im Bortz vorgezogen – also bitte nicht wundern, wenn jetzt was „fehlt“. Die behandelten Punkte waren, bezogen auf die Bortz-Unterkapitel:

- 6.1 Lineare Regression, dabei: Kovarianz und Regression
- 6.2.1 Kovarianz und Korrelation, dabei: Produkt-Moment-Korrelation; Korrelation und Regression; Beobachtete, vorhergesagte und Residualwerte; Determinationskoeffizient
- 6.4 Korrelation und Kausalität

Die Statistische Absicherung von linearen Regressionen:

Regressionsgleichungen werden auf der Grundlage einer repräsentativen Stichprobe bestimmt, um sie auch auf andere Untersuchungseinheiten, die zur selben Population gehören, anwenden zu können.

Das bedeutet aber, dass eine (auf Basis einer Stichprobe gebildete) Regressionsgleichung generalisierbar sein muss. Aus dieser Population können wir eine durch (unendlich) viele Stichproben wieder eine SKV bilden.

Je größer aber die Streuung (der Standardfehler) dieser SKV ist, desto weniger werden die Werte einer einzelnen Stichprobe (hier: einer einzelnen Regressionsgerade) die Regressionsgerade der Population vorhersagen können.

Die nach der Methode der kleinsten Quadrate ermittelte Regressionsgleichung stellt nur eine Schätzung der folgenden, in der Population gültigen, Regressionsgerade dar:

$$\hat{y}_j^* = \beta_{yx} \cdot x_j + \alpha_{yx}$$

\hat{y}_j^* kennzeichnet hierbei einen \hat{y} -Wert, der aufgrund der Populations-Regressionsgleichung vorhergesagt wurde.

- Wichtig: nicht die Punktabstände zur Geraden minimieren, sondern die Abstände in y-Richtung, damit für die Vorhersage x auf y optimiert, deswegen gibt es für eine Regression von y auf x auch andere Koeffizienten a & b.

Voraussetzungen:

- Unsere SKV verteilt sich (gemäß Abb. 6.7 im Bortz S. 183) als bivariate Häufigkeitsverteilung:
 - a.) Die Verteilung der x-Werte muss für sich genommen normal sein.
 - b.) Die Verteilung der y-Werte muss für sich genommen normal sein.
 - c.) Die zu einem x-Wert gehörenden y-Werte (Arrayverteilungen)* müssen normalverteilt sein.
 - d.) Die Mittelwerte der Arrayverteilungen müssen auf einer Gerade liegen.

* Ein Arrayverteilung von y-Werten (zu natürlich jeweils zu einem x-Wert gehören) bedeutet, das bei nicht perfektem Zusammenhang gleiche Untersuchungen eines bestimmten x-Merkmals verschiedene y-Merkmale aufweisen.

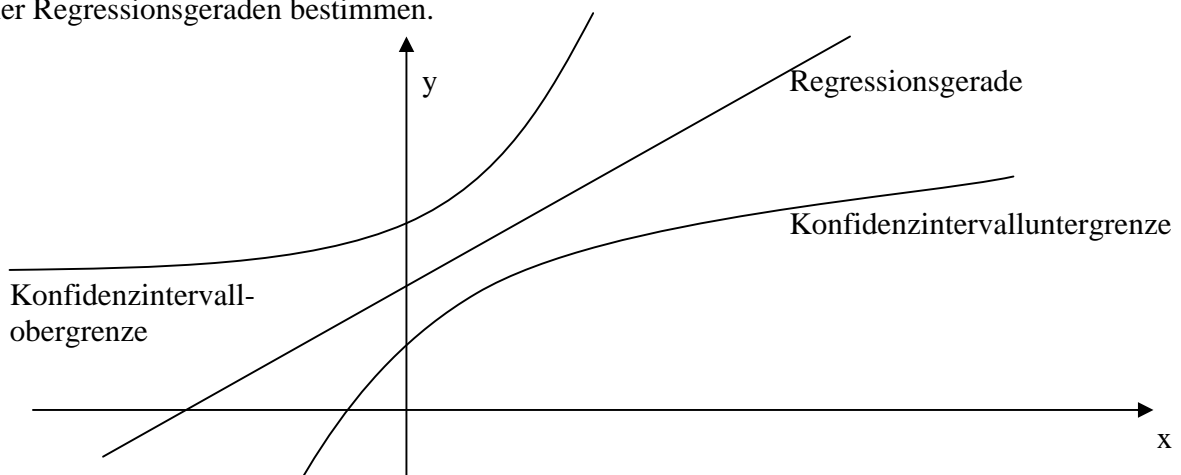
Die Genauigkeit der Regressionsvorhersagen:

Die Genauigkeit der Regressionsvorhersagen wird durch den Standardschätzfehler (STSF) bestimmt. (Formel siehe Bortz S. 185)

Der Standardschätzfehler kennzeichnet die Streuung der y-Werte um die Regressionsgerade und ist damit ein Gütemaßstab für die Genauigkeit der Regressionsvorhersagen. Die Genauigkeit einer Regressionsvorhersage wächst mit kleiner werdendem Standardschätzfehler.

Determinanten der Vorhersagegenauigkeit (durch Konfidenzintervalle):

Durch Konfidenzintervalle kann man die Vorhersagegenauigkeit einzelner Punkt um und auf einer Regressionsgeraden bestimmen.

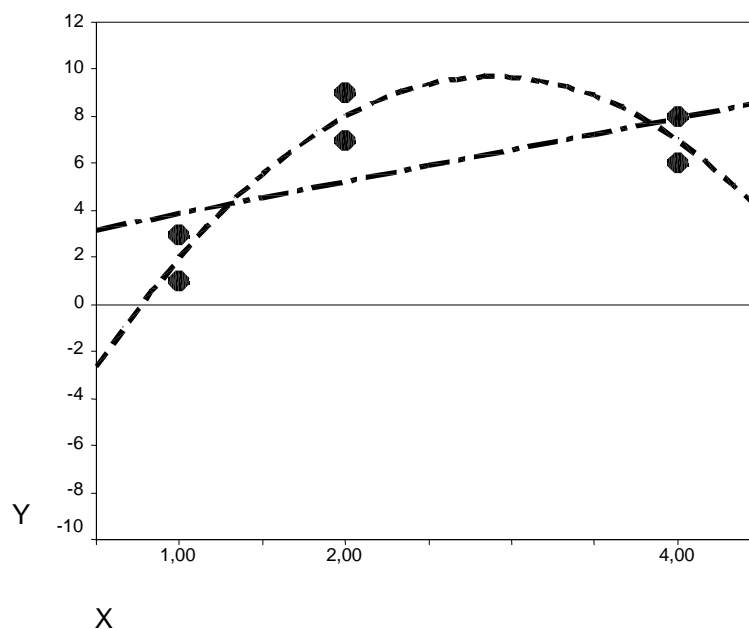
**Bestimmung der Größe des Konfidenzintervalls (KI):**

- Je kleiner der Konfidenzkoeffizient ($1 - \alpha$), desto kleiner KI
- Je kleiner der Standardschätzfehler (STSF), desto kleiner KI.
- Je kleiner n , desto größer KI.
- Je kleiner die Varianz x (y ist im STSF), desto kleiner KI.

Nonlineare Regression:

Gelegentlich erwartet man, dass eine nicht-lineare Beziehung eine bessere Vorhersage macht als eine lineare Beziehung. Als Beispiel mit einer nicht- und einer linearen Regression:

X	1	1	2	2	4	4
Y	1	3	7	9	6	8



Hier dürfte die nicht-lineare Beziehung die Punkte besser vorhersagen (Grafik aus den B3-Folien, leicht abgewandelt)

Varianten (einige ausgewählte):**Umgekehrt U-förmige Beziehung** (Bortz S. 189):

Eine umgekehrt U-förmige bzw. parabolische Beziehung wird durch eine quadratische Regressionsgleichung oder ein Polynom 2. Ordnung modelliert.

$$\hat{y} = a + b_1 \cdot x + b_2 \cdot x^2$$

Polynome höherer Ordnung:

Ein (umgekehrt) S-förmiger Zusammenhang lässt sich durch eine kubische Regressionsgleichung bzw. ein Polynom 3. Ordnung anpassen:

$$\hat{y} = a + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3$$

Nonlineare Zusammenhänge, die über ein Polynom 3. Ordnung hinausgehen, können in der u.a. psychologischen Forschung nur sehr selten begründet werden. Die Aufstellung eines solchen Polynoms kann bestenfalls ex post (= ohne theoretische Vorannahmen) sinnvoll sein.

[Die Regressionsgleichung wird einfach um weitere b und x erweitert, wobei der Exponent pro weiterem b um 1 erhöht wird.]

Linearisierung:

Über Logarithmierung kann man die Funktion von Gleichungen ermitteln.

Geht man von einem exponentiellen Zusammenhang von zwei Variablen aus, erhält man die Gleichung:

$$\hat{y} = a \cdot x^b$$

Durch Logarithmierung (siehe Bortz S. 192) gelangt man schlussendlich zu folgender Gleichung:

$$\hat{y}' = a' + b' \cdot x$$

Was verteuft wie eine lineare Regression aussieht.

[Ganz exakt sind die so ermittelten Regressionskoeffizienten leider nicht, sie unterscheiden sich von denen, die wir mit der Methode der kleinsten Quadrate ermitteln würden.]

Interpretationshilfen für r:

Angenommen ein Schulpsychologo ermittelt eine Korrelation von $r = 0,60$ zwischen Abiturschnitt (y) und Intelligenzquotienten (x).

Was bedeutet diese Zahl?

Um das zu veranschaulichen, dichotomisieren wir beide Variablen am Median und erhalten eine 4-Felder-Tafel.

(dichotom = ein Merkmal mit nur zwei Ausprägungen, z.B. Geschlecht. Man kann aber auch intervallskalierte Daten dichotomisieren, d.h. zweiteilen. D.h. hier teilen wir am Median in zwei Hälften.)

Wenn wir dem Schulpsychologen jetzt auftragen, mal zu schätzen, wie sich die Abiturschnitte und IQs bei 200 Schülern verteilen, wenn er am Median teilen soll, so werden bei keinem Zusammenhang ($r = 0$) je 50 Schüler pro Kategorie verteilt sein.

Aber es verteilt sich so bei $r = 0,60$:

		Abischnitt (y)		
		< Md _y	> Md _y	
IQs (x)	< Md _x	80	20	100
	> Md _x	20	80	100
		100	100	200

Wieso?

Bei zwei symmetrischen, mediandichotomisierten Merkmalen gibt r an, um wie viel der Fehler minimiert wird.

Im Beispiel betrug der Fehler bei Zufallsverteilung (je 50 n pro Kategorie) 50% Diese 50% reduziert um $r = 0,60$ bzw. 60% von $50 = 30$. Diese 30 werden einer Korrelation von $r = 0,60$ nicht mehr falsch eingeordnet, sondern richtig, d.h.

Die Fehler (50%), die wir bei der Vorhersage von x durch y machen (großes y = großes x), werden bei einer Korrelation von 0,60 um 30 % vermindert.

Mehr Infos Bortz S. 201/202.

k-fach gestufte Merkmale:

Diese Variante ist einfach eine Erweiterung der 4-Felder-Matrix um weitere Kategorien. Eine Besonderheit dabei ist, dass größere Abweichungen um mehrere Kategorien stärker „bestraft“ werden als andere. Vom Prinzip wie die 4-Felder-Matrix, jedoch aufwendiger zu rechnen. Wer mehr wissen will: Steht alles im Bortz auf S. 203/204.

Korrelation für nonlineare Zusammenhänge:

Der Zusammenhang zwischen nonlinearen Variablen lässt sich ermitteln, indem man die Varianz der vorhergesagten Werte durch die Varianz der tatsächlichen Werte teilt:

$$r^2 = \frac{S_{\hat{y}}^2}{S_y^2}$$

Überprüfung von Korrelationshypotesen:

Wird aus einer bivariaten, intervallskalierten Grundgesamtheit eine Stichprobe gezogen, kann ein Korrelationskoeffizient r berechnet werden. Er zeigt den linearen Zusammenhang zwischen den zwei Merkmalen an, bzw., quadriert (r^2), den Anteil gemeinsamer Varianz.

Voraussetzungen:

- Grundgesamtheit bivariat normalverteilt
- Kovarianz & Korrelation der Grundgesamtheit bekannt
- Die einzelnen Merkmalsstufenkombinationen müssen voneinander unabhängig sein.

Die ganzen Voraussetzungen sind aber ein bisschen schwierig zu erfüllen (siehe Bortz S.205), dementsprechend beschränkt man sich darauf, dass beide einzelnen Merkmale NV sind, die Arrayverteilungen Normalverteilt sind und die Varianzen der Arrayverteilungen homogen ist.

Wenn man diese Voraussetzungen missachtet, muss man mit größerem Alpha- und Beta-Fehler rechnen. Trotzdem ist der Test relativ robust.

Signifikanztest:

Ob eine empirisch ermittelte Korrelation r mit der Hypothese

$$H_0: \rho = 0$$

zu vereinbaren ist, lässt sich mit folgender Prüfgröße testen:

$$t_{emp} = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}, \text{ mit } df = n-2.$$

Ab $n > 3$ kann man zeigen, dass der Ausdruck bei $df = n - 2$ t-verteilt ist.

Entscheidung:

	ungerichtet	rechtsseitig	linksseitig
falls ...	$ t_{emp} > t_{(df; \alpha/2)} $	$t_{emp} > t_{(df; 1-\alpha)}$	$t_{emp} < t_{(df; \alpha)}$
Annahme ...	$H_1: \rho \neq 0$	$H_1: \rho > 0$	$H_1: \rho < 0$

Für den zweiseitigen Test einer Korrelation sind in der Tabelle D (Bortz 1999, S. 775) kritische Wert für $\alpha = 0,05$ und $\alpha = 0,01$ angegeben. Diese Werte erhält man, wenn man obige Prüfgröße nach r auflöst:

$$r_{krit} = \frac{t_{krit}}{\sqrt{n-2 + t_{krit}^2}}.$$

Fischers Z-Transformation:

Besteht in der Grundgesamtheit ein Zusammenhang $\rho \neq 0$, erhalten wir für (theoretisch unendlich) viele Stichproben eine rechts- ($\rho > 0$) oder linkssteile ($\rho < 0$) SKV der Korrelationen. Eine Nullhypothese mit einem $\rho \neq 0$ kann somit nichts mehr messen (da asymmetrisch).

Diese rechts- oder linkssteile Verteilung lässt sich durch Fishers Z-Transformation wieder in Normalverteilung transformieren:

$$Z = \frac{1}{2} \cdot \ln\left(\frac{1+r}{1-r}\right) \quad \text{aufgelöst nach } r \quad r = \frac{e^{2Z} - 1}{e^{2Z} + 1}$$

wobei \ln = Logarithmus zur Basis e ($\approx 2,718$).

Je weiter ρ von +/-1 entfernt und je größer n ist, desto normalverteilter ist die Kurve nach Fishers Z-Transformation

Weiterhin sind die transformierten Werte Teil einer Verhältnisskala (fester Nullpunkt), man kann nämlich bei einer Korrelation nicht davon ausgehen, dass der Zusammenhang bei $r = 0,4$ halb so groß ist wie bei $r = 0,8$!

Wenn es transformiert wird zeigt sich: die Zuwachsraten im oberen Korrelations-Bereich sind bedeutsamer.

Auch Mittelwerte & Varianzen von einer Korrelation sind nicht interpretierbar, sofern sie nicht transformiert wurden.

**Bei Mittelwertbildung von (z.B. 3) Korrelationen also erst transformieren, dann Mittel der Fisher-Z-Werte bilden, dann dieses Mittel rücktransformieren:
Dabei werden höhere Korrelationen stärker berücksichtigt als niedrige.**

Es gibt mehrere Variationsmöglichkeiten von Fischers Z-Transformation, sie stehen im Bortz auf den Seiten 210 – 214. Wenn es interessiert, der möge es nachschlagen.

Ich glaube aber nicht, dass das für die Prüfung in irgendeiner Form relevant sein könnte.

Spezielle Korrelationstechniken:

Unter dem Punkt „Überprüfung von Korrelationshypotesen“ ging es um die Produkt-Moment-Korrelation zweier intervallskalierter Merkmale.

Das Ganze geht aber natürlich auch mit ordinal- und nominalskalierten sowie dichotomen Merkmalen.

Übersicht der bivariaten Korrelationsarten:

Merkmal y	Merkmal x		
	Intervallskala	dichotomes Merkmal	Ordinalskala
Intervallskala	Produkt-Moment-Korrelation	Punkt-biseriale Korrelation	Rangkorrelation
dichotomes Merkmal	-	Φ -Koeffizient	Biseriale Rangkorrelation
Ordinalskala	-	-	Rangkorrelation

Dazu gibt es noch den Kontingenzkoeffizienten, der den Zusammenhang zweier nominalskalierter Merkmale bestimmt. Da dieser Koeffizient aber kein Korrelationsmaß im engeren Sinn darstellt, ist er nicht in der Tabelle.

I.) Korrelation zweier Intervallskalen:

Berechnung der Produkt-Moment-Korrelation, siehe „Überprüfung von Korrelationshypotesen“

II.) Korrelation einer Intervallskala mit einem dichotomen Merkmal:

A.) Punktbiseriale Korrelation:

Ein Zusammenhang zwischen einem dichotomen Merkmale (z.B. männlich/weiblich) und einem intervallskalierten Merkmal (z.B. Körpergewicht) wird durch eine punktbiseriale Korrelation (auch Produkt-Moment-biseriale Korrelation genannt) erfasst.

Eine punktbiseriale Gleichung erhält man, wenn die Werte 0 und 1 für die beiden dichotomen Variablen in die Produkt-Moment-Korrelations-Gleichung eingesetzt werden.

Dadurch vereinfacht sich die Korrelationsformel:

$$r_{pb} = \frac{\bar{y}_1 - \bar{y}_0}{s_y} \cdot \sqrt{\frac{n_0 \cdot n_1}{n^2}}$$

wobei

n_0, n_1 = Anzahl der Untersuchungseinheiten in den Merkmalskategorien x_0 und x_1 .

\bar{y}_0, \bar{y}_1 = durchschnittliche Ausprägung des Merkmals y (Körpergewicht) bei den Untersuchungseinheiten x_0 (Männer) und x_1 (Frauen).

$n = n_0 + n_1$ = Gesamtstichprobe

s_y = Streuung der kontinuierlichen y -Variablen

Die Signifikanzüberprüfung ($H_0: \rho = 0$) erfolgt wie bei der Produkt-Moment-Korrelation durch:

$$t = \frac{r_{pb}}{\sqrt{(1 - r_{pb}^2)/(n - 2)}}$$

Der so ermittelte t -Wert ist mit $n - 2$ Freiheitsgraden versehen und in der t -Tabelle mit t_{krit} verglichen.

Die punktbiseriale Korrelation entspricht (als Verfahren zur Überprüfung von Zusammenhangshypothesen) **dem t-Test für unabhängige Stichproben** (als Verfahren zur Überprüfung von Unterschiedshypothesen).

B.) Biseriale Korrelation:

Bei einem künstlich dichotomisiertem Merkmal und einem intervallskalierten Merkmal wird eine biseriale Korrelation (r_{bis}) durchgeführt.

- Biseriale Korrelation (r_{bis}) und Punktbiseriale Korrelation (r_{pb}) hängen statistisch zusammen.
- Es gibt auch noch triseriale und polyseriale Korrelation
- Das steht alles im Bortz auf S. 216-218

III.) Korrelation einer Intervallskala mit einer Ordinalskala:

- Richtige Verfahren hierfür gibt es (noch) nicht. Für die Praxis: Herabstufung der Intervallskala auf Ordinalskalenniveau, Anwendung der Korrelation zweier Ordinalskalen.

IV.) Korrelation für zwei dichotome Variablen:

Wenn die Merkmale x und y jeweils dichotom sind, kann ihr Zusammenhang durch den Phi-Koeffizienten (Φ) ermittelt werden

Da es nur 4 Häufigkeiten gibt...

		Merkmal (y)	
		0	1
Merkmal (x)	0	a	b
	1	c	d

....ist die Berechnung von Φ einfach:

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

Es gibt einen engen Zusammenhang mit dem 4-Felder- χ^2 , deshalb Signifikanzprüfung mittels

$$\chi^2 = n \cdot \Phi^2 \text{ mit } df = 1$$

Der Zusammenhang zwischen dem 4-Felder- χ^2 und Phi-Koeffizienten:

$$\Phi = \sqrt{\frac{\chi^2}{n}}$$

Achtung:

Der Φ -Koeffizient liegt nur im üblichen Wertebereich von -1 bis +1, wenn die Aufteilung der Stichprobe in Alternativen von x der in die Alternativen von y entspricht, also bei identischen Randverteilungen.

Ansonsten gibt es einen geringeren Φ_{maximal} -Wert (ist mit Formel berechenbar)

Bei der Bestimmung von Φ_{max} ist darauf zu achten, dass

- das Vorzeichen von Φ_{max} mit dem Φ_{emp} von übereinstimmt.
- das Vorzeichen von Φ_{max} beliebig ist.

Gelegentlich versuchen Einige die Aufwertung eines empirischen Φ -Wertes am Φ_{max} , um es vergleichbar mit der Produkt-Moment-Korrelation zu machen, das ist aber fragwürdig, da auch die Produkt-Moment-Korrelation nur bei identischen Randverteilungen einen Wertebereich von -1 bis +1 aufweist.

Sind beide Variablen künstliche Dichotomien, arbeitet man mit einer **Tetrachorischen Korrelation**. (siehe Bortz S. 220/221)

V.) Korrelation eines dichotomen Merkmals mit einer Ordinalskala:

Die Korrelation eines dichotomen Merkmals mit einer Ordinalskala – auch biserial Rangkorrelation (r_{bisR}) genannt – wird berechnet, wenn ein Merkmal (x) in künstlicher oder natürlicher Dichotomie vorliegt und das andere Merkmal (y) rangskaliert. (ähnlich dem U-Test)

$$r_{bisR} = \frac{2}{n} \cdot (\bar{y}_1 - \bar{y}_2)$$

wobei

\bar{y}_1 = durchschnittlicher Rangplatz der zu x_1 gehörenden Untersuchungseinheiten

\bar{y}_2 = durchschnittlicher Rangplatz der zu x_2 gehörenden Untersuchungseinheiten

n = Umfang der Stichprobe

Der kritische Wert wird – bei hinreichend großen Stichproben – mit dem U-Test berechnet:

$$z = \frac{U - \mu_U}{\sigma_U}$$

VI.) Korrelation zweier Ordinalskalen (Spearman's Rho):

Der Zusammenhang zweier ordinalskalierte Merkmale wird durch die **Rangkorrelation** nach Spearman (r_s oder ρ) erfasst.

Eigentliche Formel:

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

Errechnung des kritischen Wertes ab $n \geq 30$

$$t = \frac{r_s}{\sqrt{(1 - r_s^2)/(n - 2)}} \quad df = n - 2$$

(Dieselbe Formel wie bei der punktbiserialen Korrelation)

- bei verbundenen Rängen gibt es Modifikationen (Bortz S. 224)

VII.) „Korrelation“ zweier Nominalskalen (Kontingenzkoeffizient):

Das bekannteste Maß zur Berechnung des Zusammenhangs zweier Nominalskalen ist der Kontingenzkoeffizient C.

Dieser Test hängt eng mit dem $k \times l$ - χ^2 -Test zusammen, mit dem wir auch die H_0 überprüfen, ob zwei nominalskalierte Merkmale stochastisch unabhängig sind.

Eigentliche Formel:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

wobei $\chi^2 = -\chi^2$ Wert des $k \times l$ - χ^2 -Tests.