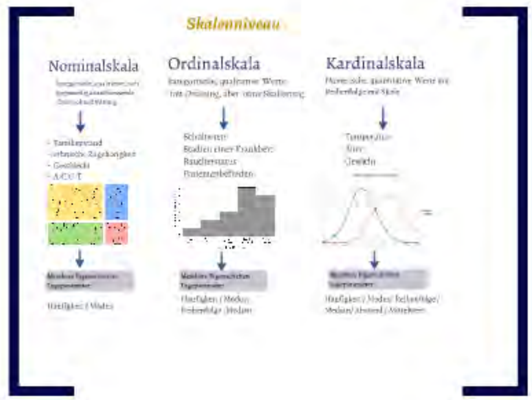
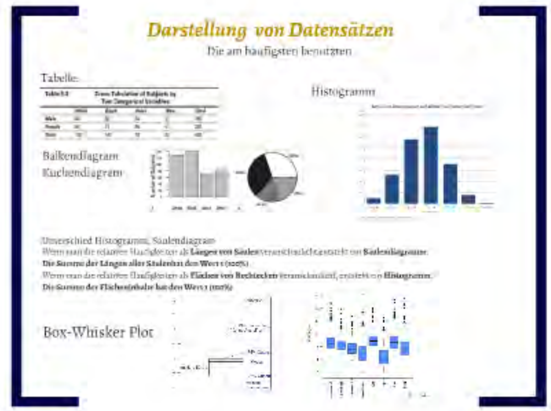
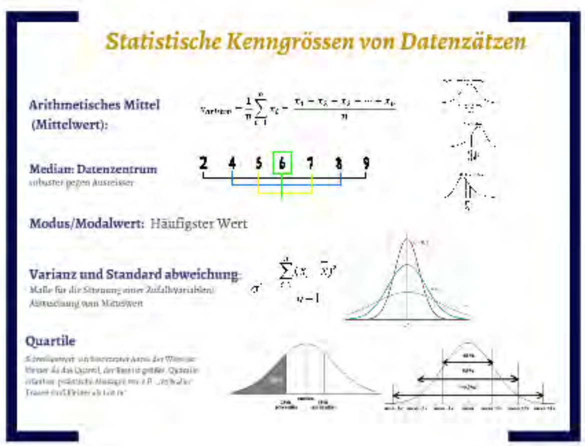
Deskriptive Statistik

Ziel: Empirische Daten durch Tabellen, Kennzahlen und Grafiken übersichtlich darzustellen und zu ordnen.







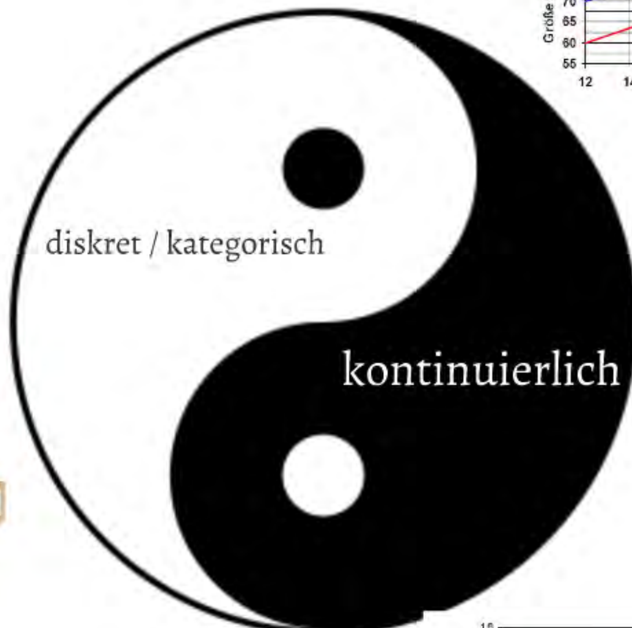
Was habe ich überhaupt für Daten ?

Nominal

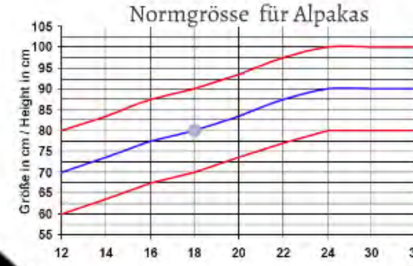
Nominal

durchschnittliches masernkranke Europaer ??

Nominal

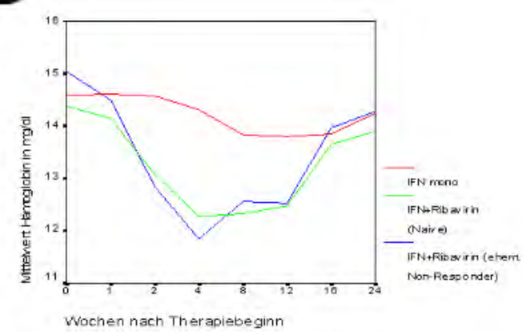


Normgrösse für Alpakas

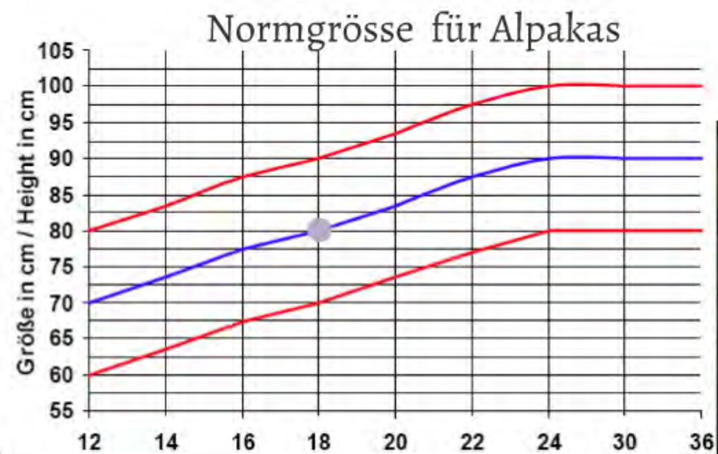


durchschnittliches Alpaka ?

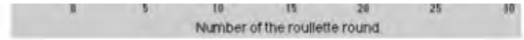
Ordinal



be ich überhaupt für Daten ?



durchschnittliches Alpaka ?



Nominal

Masernfälle in Europa
von März 2012 bis März 2013

durchschnittlicher masernkranke Europäer ??

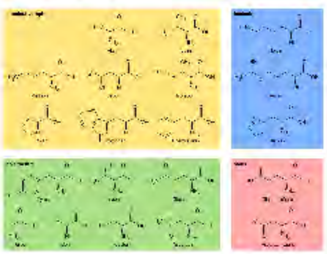
Nominal

Skalenniveau

Nominalskala

kategorische, qualitative, sich gegenseitig ausschliessende Daten ohne Ordnung

- Familienstand
- ethnische Zugehörigkeit
- Geschlecht
- A/C/G/T



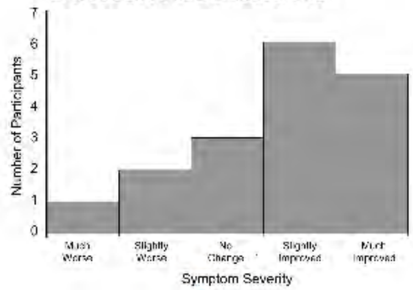
Messbare Eigenschaften
Lageparameter

Häufigkeit / Modus

Ordinalskala

kategorische, qualitative Werte mit Ordnung, aber ohne Skalierung

Schulnoten
Stadien einer Krankheit
Raucherstatus
Patientenbefinden



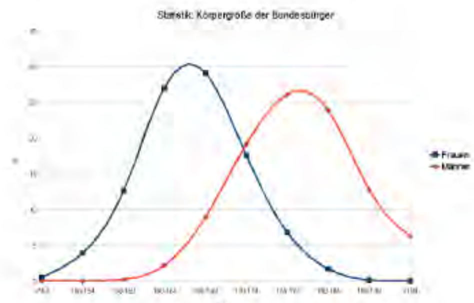
Messbare Eigenschaften
Lageparameter

Häufigkeit / Modus/
Reihenfolge /Median

Kardinalskala

Numerische, quantitative Werte mit Reihenfolge mit Skala

Temperatur
Alter
Gewicht



Messbare Eigenschaften
Lageparameter

Häufigkeit / Modus/ Reihenfolge /
Median/ Abstand / Mittelwert

Darstellung von Datensätzen

Die am häufigsten benutzten

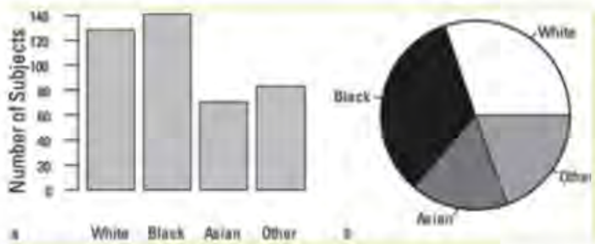
Tabelle:

Table 8-2	Cross-Tabulation of Subjects by Two Categorical Variables				
	White	*Black*	*Asian*	*Other*	*Total*
Male	60	60	34	42	196
Female	68	81	36	41	226
Total	128	141	70	83	422

Histogramm

Balkendiagram

Kuchendiagram



Unterschied Histogramm, Säulendiagram

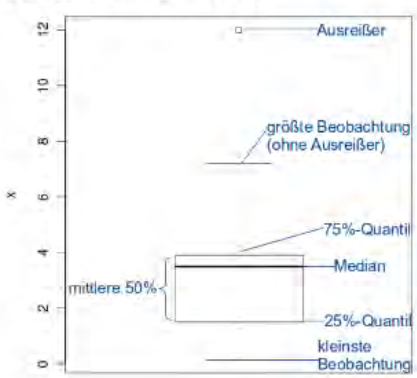
Wenn man die relativen Häufigkeiten als **Längen von Säulen** veranschaulicht,entsteht ein **Säulendiagramm**.

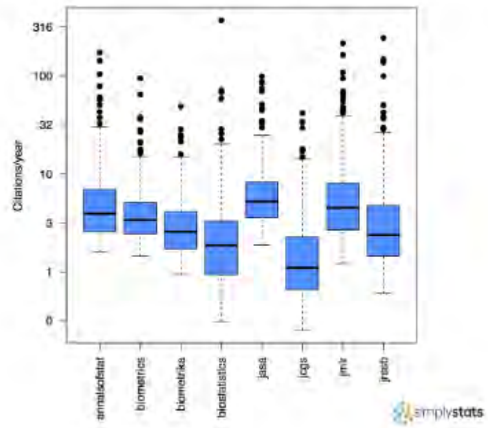
Die Summe der Längen aller Säulenhat den Wert 1 (100%)

Wenn man die relativen Häufigkeiten als **Flächen von Rechtecken** veranschaulicht, entsteht ein **Histogramm**.

Die Summe der Flächeninhalte hat den Wert 1 (100%)

Box-Whisker Plot

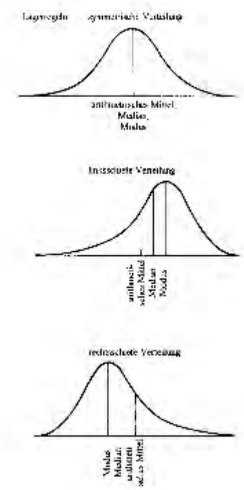




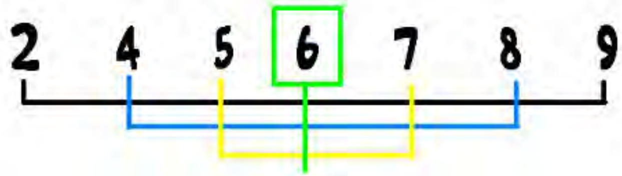
Statistische Kenngrössen von Datenzätzen

Arithmetisches Mittel (Mittelwert):

$$\overline{x}_{arithm} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$



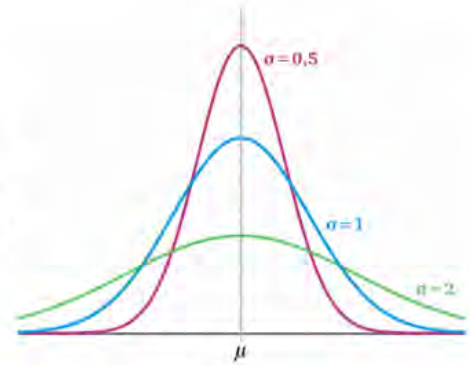
Median: Datenzentrum
robuster gegen Ausreisser



Modus/Modalwert: Häufigster Wert

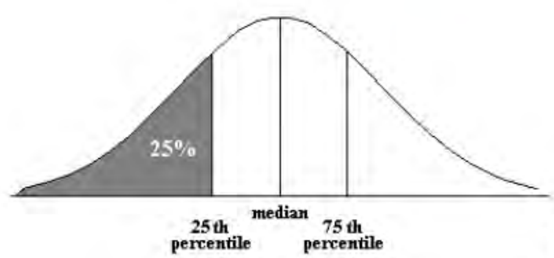
Varianz und Standard abweichung:
Maße für die Streuung einer Zufallsvariablen/ Abweichung vom Mittelwert

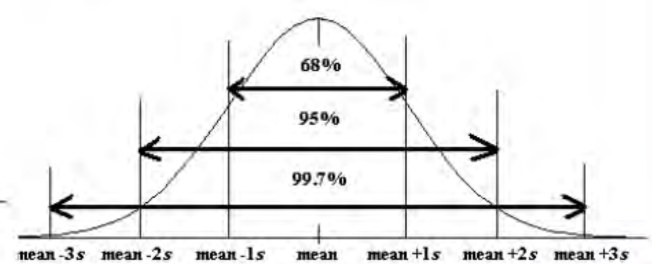
$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$



Quartile

Schwellenwert: ein bestimmter Anteil der Werte ist kleiner als das Quantil, der Rest ist größer. Quantile erlauben praktische Aussagen wie z.B. „25 % aller Frauen sind kleiner als 1,62 m"

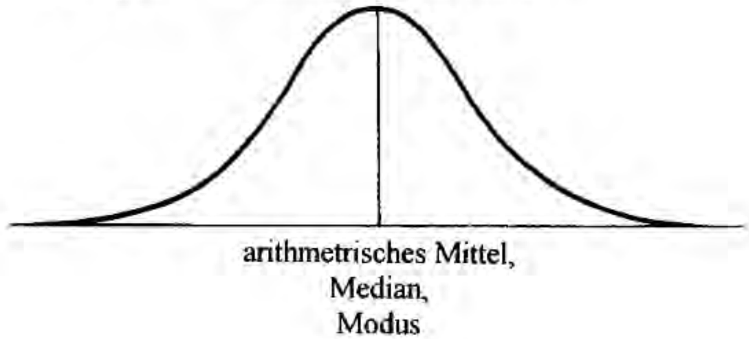




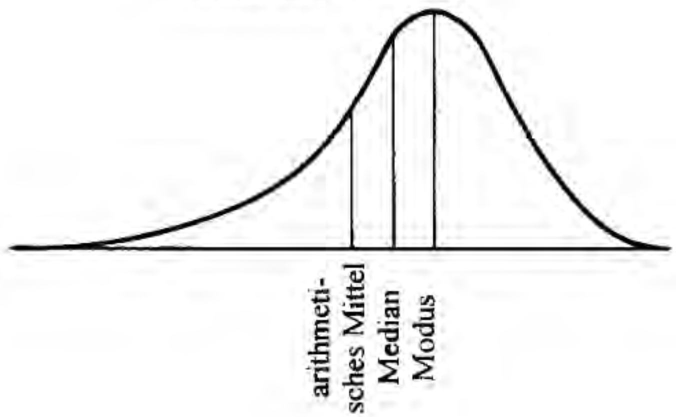
$$+ x_n$$

Lagerregeln

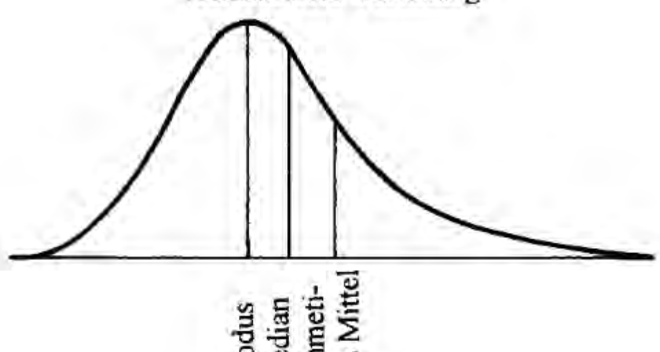
symmetrische Verteilung

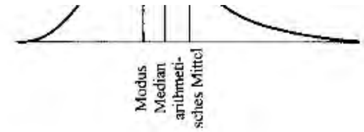


linksschiefe Verteilung

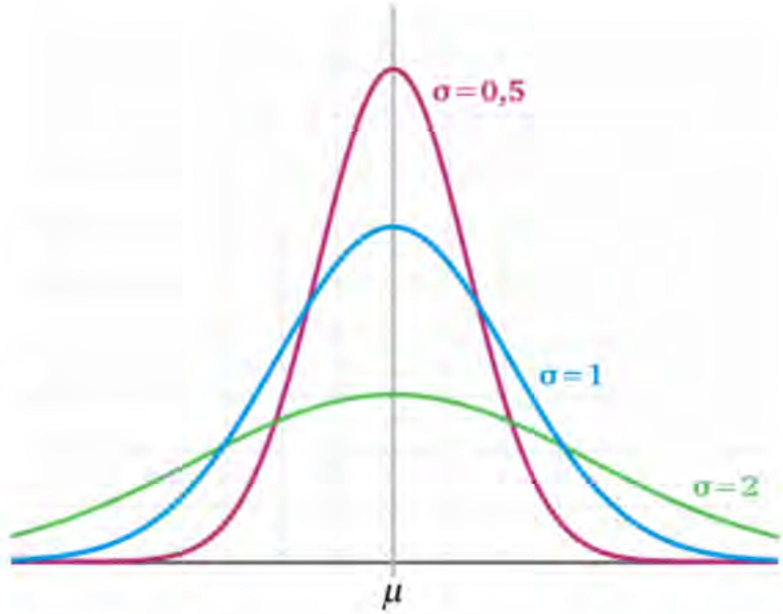


rechtsschiefe Verteilung





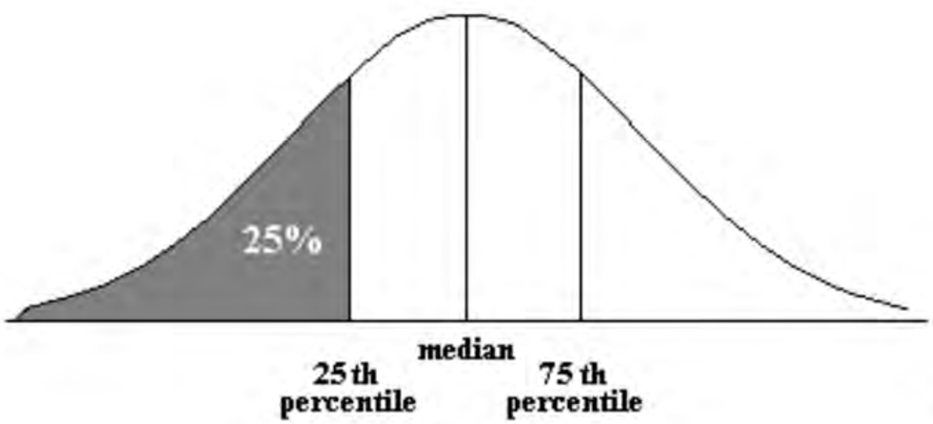
ter Wert

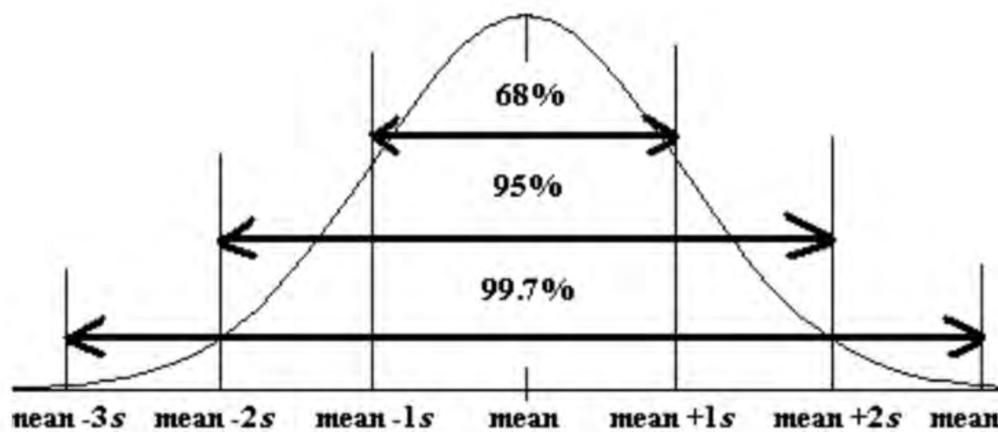


chung:

len/

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





Statistische Tests

Grundbegriffe:

Population — Alle Individuen haben (meist) bestimmte Merkmale gemeinsam (z.B. Übergewicht, Blutgruppe, sozialer Hintergrund, Geschlecht, Krankheit,..).

Stichprobe — Eine Teilmenge der obigen Population

Beziehung des zu untersuchenden Merkmales zu einer/mehreren Variablen:

Der Hauptzweck fast aller Statistiken: Herauszufinden, welche (externen/internen) Faktoren (Medikament/SNP) das betrachtete Merkmal beeinflussen

Hypothese — eine Annahme, die mit Methoden der Statistik auf Basis empirischer Daten geprüft wird.

Nullhypothese: *H0* eine Annahme über die Wahrscheinlichkeitsverteilung einer oder mehrerer Zufallsvariablen (Eine bestimmte Annahme (H1) trifft nicht zu) -> Das Medikament ist wirkungslos, und der Effekt/ Zusammenhang, den ich vermeine in den Daten zu sehen entsteht durch Rauschen/Zufall/ Ungenauigkeit bei der Erhebung der Daten.

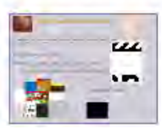
Gegenhypothese/AlternativhypotheseArbeitshypothese: H1 - die Vermutung, die man eigentlich überprüfen möchte.
-> Das Medikament zeigt Wirkung und der Effekt, den ich sehe ist NICHT durch verrauschte Daten zu erklären.
->Disjunkt von H0, aber geschlossene Aussage

Signifikanzniveau: Festgelegte Schwelle, ab der das beobachtete Verhalten nicht mehr als Zufall gilt (z.B. p - Wert)

Confounding/ Störfaktor: (Unbekannter)Faktor, der sowohl abhängige als auch unabhängige Variablen beeinflussen kann und das Test-Ergebniss dadurch verzerrt.

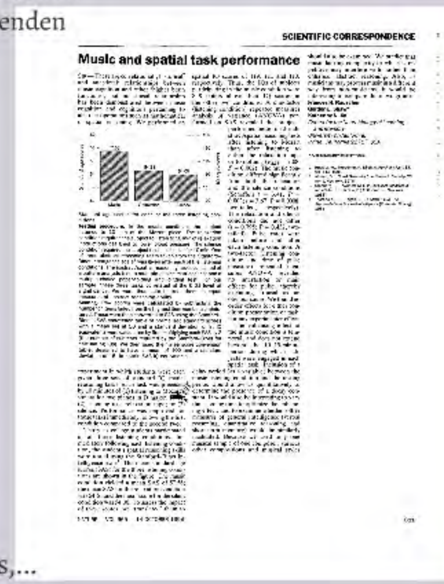
repräsentiv: Eigenschaft von Erhebungen, dass diese Aussagen über eine Grundgesamtheit zulassen
streng genommen meist nicht realisierbar -> Quotenstichproben/Randomisierte Stichprobe

Fehler erster und zweiter Ordnung: Risiko für falsch-positiven, bzw. falsch negativem Test



Macht Mozart hören Babys schlauer ?

Eine Studie (Rauscher, et.al, Nature, 1993) beschreibt den Zusammenhang zwischen dem hören einer Mozart Sonate (KV 448) und einer erstaunlich gesteigerten Leistung im räumlichen Vorstellungsvermögen während eines direkt im Anschluss stattfindenden IQ-Tests. (8-9 IQP)


SCIENTIFIC CORRESPONDENCE

Music and spatial task performance

Mediales Echo -> Mozart macht intelligent !

Folgestudien - insbesondere an Kindern und Babys schienen das zu belegen

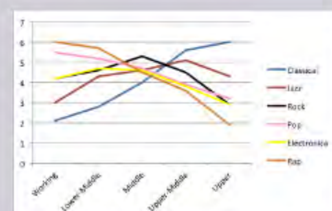
- privater Musikunterricht bei Kindergartenkindern (6-8 IQP)
- Kinder, die zu Hause regelmässig Klassik hören sind klüger als andere
- …

Mütter bekamen CDs zur Geburt,
Florida: Gesetz 1 Stunde Klassik pro Tag in den Kitas,…

Wo liegt das Problem bei der Studie ?

Tipp:

- Freiwillige Teilnahme mit aktiver Bewerbung
- Teilnahme unvergütet
- Studiendesign und Studienziel waren bekannt



from Rentfrow et al.'s 2009 data, plotting social class against genre preference luss stattfindenden

n

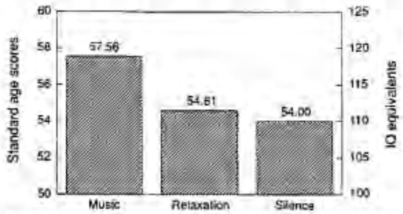
dere

in den Kitas,...

Music and spatial task performance

SIR — There are correlational[1], historical[2] and anecdotal[3] relationships between music cognition and other 'higher brain functions', but no causal relationship has been demonstrated between music cognition and cognitions pertaining to abstract operations such as mathematical or spatial reasoning. We performed an experiment in which students were each given three sets of standard IQ spatial reasoning tasks; each task was preceded by 10 minutes of (1) listening to Mozart's sonata for two pianos in D major, [redacted] K448 (2) listening to a relaxation tape; or (3) silence. Performance was improved for those tasks immediately following the first condition compared to the second two.

Thirty-six college students participated in all three listening conditions. Immediately following each listening condition, the student's spatial reasoning skills were tested using the Stanford–Binet intelligence scale[4]. The mean standard age scores (SAS) for the three listening conditions are shown in the figure. The music condition yielded a mean SAS of 57.56; the mean SAS for the relaxation condition was 54.61 and the mean score for the silent condition was 54.00. To assess the impact of these scores, we 'translated' them to spatial IQ scores of 119, 111 and 110, respectively. Thus, the IQs of subjects participating in the music condition were 8–9 points above their IQ scores in the other two conditions. A one-factor (listening condition) repeated measures analysis of variance (ANOVA) performed on SAS revealed that subjects performed better on the abstract/spatial reasoning tests after listening to Mozart than after listening to either the relaxation tape or to nothing ($F_{2,35} = 7.08$; $P = 0.002$). The music condition differed significantly from both the relaxation and the silence conditions (Scheffe's $t = 3.41$, $P = 0.002$; $t = 3.67$, $P = 0.0008$, two-tailed, respectively). The relaxation and silence conditions did not differ ($t = 0.795$; $P = 0.432$, two-tailed). Pulse rates were taken before and after each listening condition. A two-factor (listening condition and time of pulse measure) repeated measures ANOVA revealed no interaction or main effects for pulse, thereby excluding arousal as an obvious cause. We found no order effects for either condition presentation or task, nor any experimenter effect.



Standard age scores for each of the three listening conditions.

Testing procedure. In the music condition, the subject listened to 10 min of the Mozart piece. The relaxation condition required the subject to listen to 10 min of relaxation instructions designed to lower blood pressure. The silence condition required the subject to sit in silence for 10 min. One of three abstract reasoning tests taken from the Stanford–Binet intelligence scale[4] was given after each of the listening conditions. The abstract/spatial reasoning tasks consisted of a pattern analysis test, a multiple-choice matrices test and a multiple-choice paper-folding and cutting test. For our sample, these three tasks correlated at the 0.01 level of significance. We were thus able to treat them as equal measures of abstract reasoning ability.

Scoring. Raw scores were calculated by subtracting the number of items failed from the highest item number administered. These were then converted to SAS using the Stanford–Binet's SAS conversion table of normalized standard scores with a mean set at 50 and a standard deviation of 8. IQ equivalents were calculated by first multiplying each SAS by 3 (the number of subtests required by the Stanford–Binet for calculating IQs). We then used their area score conversion table, designed to have a mean of 100 and a standard deviation of 16, to obtain SAS IQ equivalents.

The enhancing effect of the music condition is temporal, and does not extend beyond the 10–15-minute period during which subjects were engaged in each spatial task. Inclusion of a delay period (as a variable) between the music listening condition and the testing period would allow us quantitatively to determine the presence of a decay constant. It would also be interesting to vary the listening time to optimize the enhancing effect, and to examine whether other measures of general intelligence (verbal reasoning, quantitative reasoning and short-term memory) would be similarly facilitated. Because we used only one musical sample of one composer, various other compositions and musical styles should also be examined. We predict that music lacking complexity or which is repetitive may interfere with, rather than enhance, abstract reasoning. Also, as musicians may process music in a different way from non-musicians, it would be interesting to compare these two groups.

Frances H. Rauscher
Gordon L. Shaw*
Katherine N. Ky
Center for the Neurobiology of Learning and Memory, University of California, Irvine, California 92717, USA

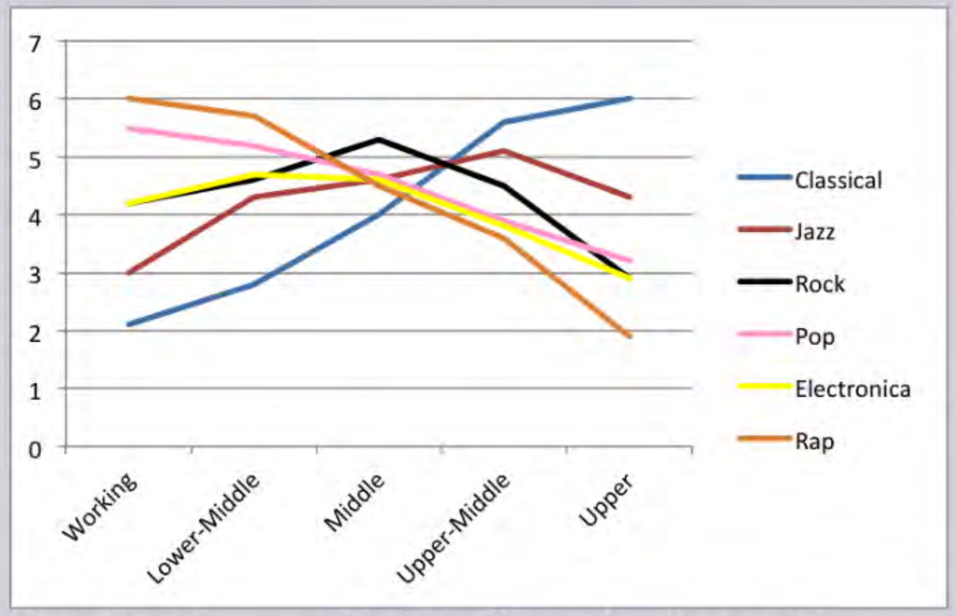
* Also at Department of Physics

1. Hassler, M., Birbaumer, N. & Feil, A. *Psychol. Music* **13**, 99–113 (1985).
2. Allman, G. J. *Greek Geometry from Thales to Euclid* p. 23 (Arno, New York, 1976).
3. Cranberg, L. D. & Albert, M. L. in *The Exceptional Brain* (eds Obler, L. K. & Fein, D.) 156 (Guilford, New York, 1988).
4. Thorndike, R. L., Hagen, E. P. & Sattler, J. M. *The Stanford–Binet Scale of Intelligence* (Riverside, Chicago, 1986).

Wo liegt das Problem bei der Studie ?

Tipp:

- Freiwillige Teilnahme mit aktiver Bewerbung
- Teilnahme unvergütet
- Studiendesign und Studienziel waren bekannt



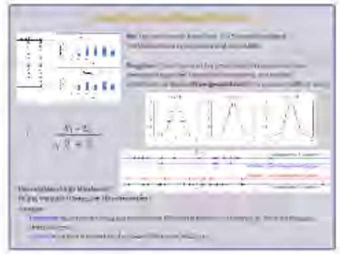
from Rentfrow et al.'s 2009 data, plotting social class against genre preference

Statistische Tests II

Vergleiche von Gruppen/Messreihen:

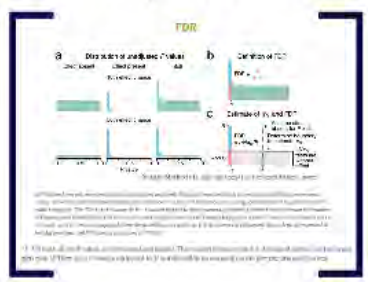
Wahl des geeigneten Tests ist abhängig von der spezifischen Fragestellung und dem Datensatz.

Student t-test: Ist der beobachtete Effekt grösser als von einer zufälligen Abweichung erwarten würde ?



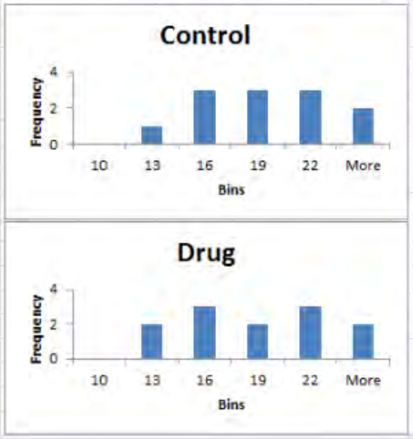
Exakter Test von Fischer: Besonders geeignet bei Kreuzversuchen (z.B. Stichproben von Patienten mit/ohne Krankheit und mit/ohne Medikament). Oder Vergleich von Gen-Sets mit oder ohne Mutationen.

False discovery rate (FDR): Methode um p-values zu adjustieren.



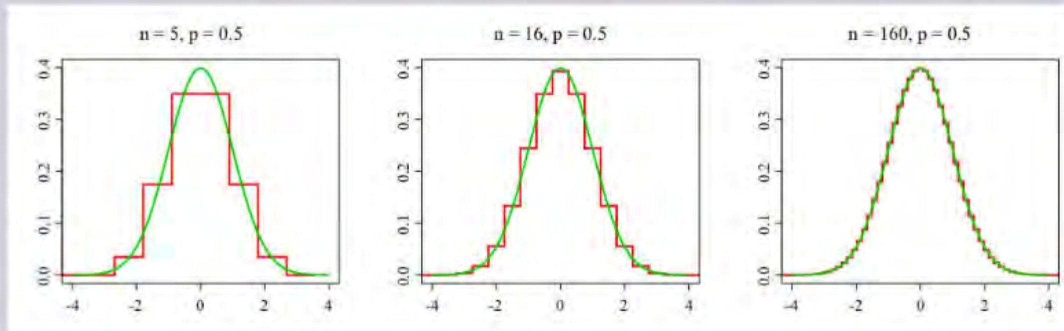
Beispiel: t-test für unabhängige Stichproben

Control	Drug
23	16
15	21
16	16
25	11
20	24
17	21
18	18
14	15
12	19
19	22
21	13
22	24

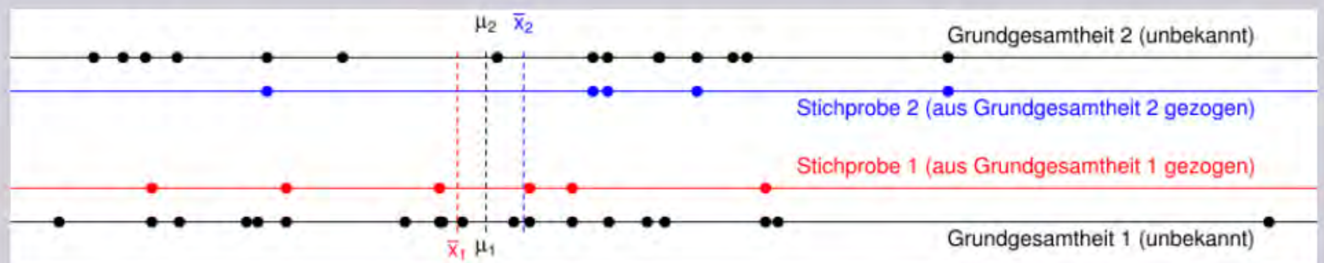


Ho: Das Medikament beeinflusst das Fahrverhalten **nicht**.
Die Unterschiede in den Scores sind rein zufällig

Vorgehen: Es wird (anhand der gewichteten Varianzen und unter Berücksichtigung der Grösse der Stichproben), wie weit die Mittelwerte der beiden **Grundgesamtheiten** von einander entfernt sind.



$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$



Unterscheiden sich die Mittelwerte ?

Falls ja, wie stark ? Genug, um Ho zu verwerfen ?

Aussagen:

T-Statistik: Mass für die Grösse des betrachteten Effektes (Mittelwert_1-Mittlwert_2 / Innerere Standard Abweichungen)

p-Wert: Mass für die Sicherheit, das dieser Effekt nicht zufällig ist.

Der exakte Fisher Test

Signifikanztest auf Unabhängigkeit in einer Kontingenztafel

Beispiel: Studie mit dem Ziel ein neues Medikament für eine Krankheit mit hoher Sterberate auf Wirksamkeit zu testen

Observed Counts		Outcome		
		Lived	Died	Total
Treatment	Drug:	33	27	60
	Placebo:	10	30	40
	Total:	43	57	100

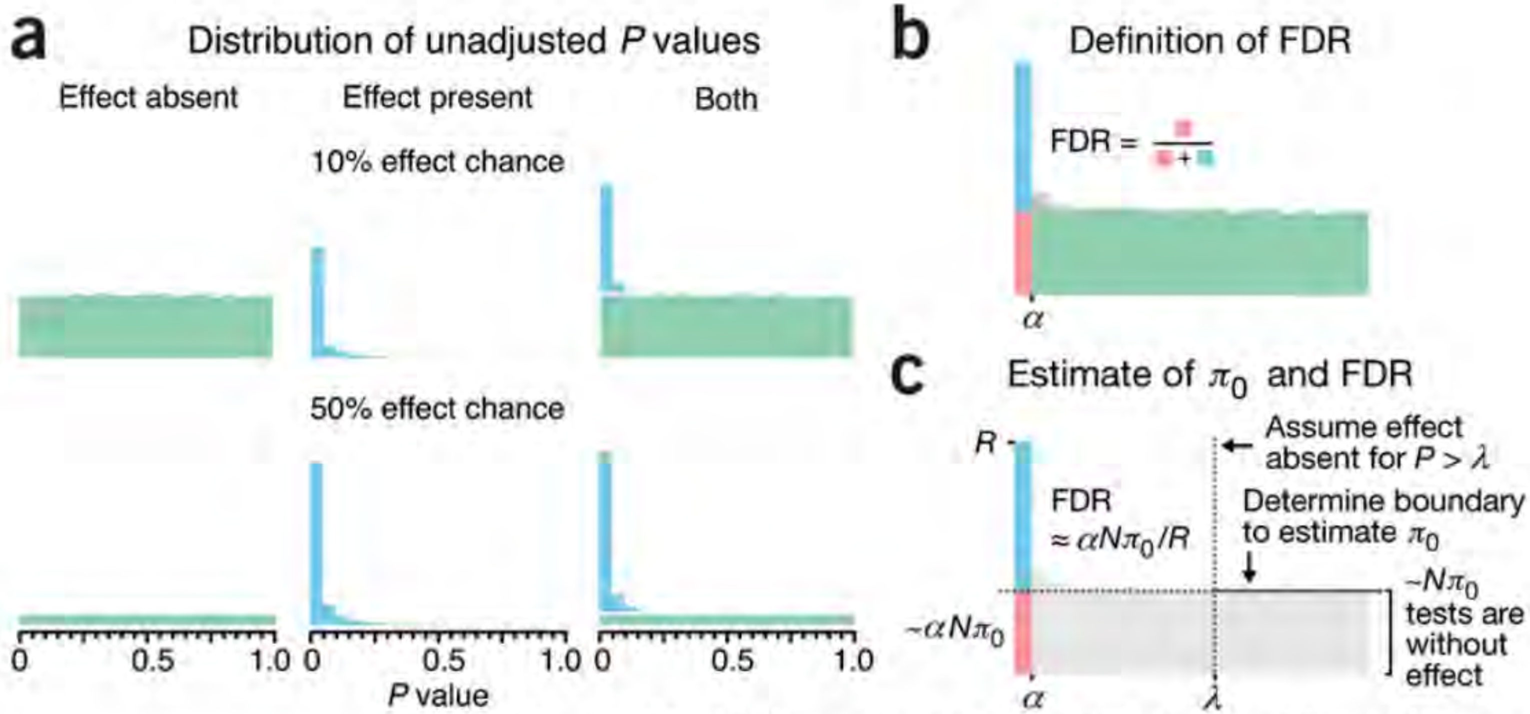
H0: Es gibt **keinen** Effekt auf die Überlebenschance und das Medikament ist daher wirkungslos

-> H0 wird verworfen, wenn der p-Wert klein genug ist

Erwartete Häufigkeiten bei Gültigkeit der Nullhypothese.					Beobachtete Häufigkeiten in der Stichprobe.			
	A	nicht A	$\sum$			A	nicht A	$\sum$
B	h_a	h_c	h_B		B	a	c	a+c
nicht B	h_b	h_d	$h_{\bar{B}}$		nicht B	b	d	b+d
$\sum$	h_A	$h_{\bar{A}}$	n		$\sum$	a+b	c+d	n=a+b+c+d

Idee: wie gross ist die Wahrscheinlichkeit, das das obige Resultat durch Zufall

FDR



Nature Methods 11, 355–356 (2014) doi:10.1038/nmeth.2900

(a) P values from null are expected to be distributed uniformly, whereas those for which the null is false will have more small values. Shown are distributions from the simulation for N = 1,000. (b) Inference types using color scheme of Figure 1b on the P value histogram. The FDR is the fraction of P < that correspond to false positives. (c) Storey's method first estimates the fraction of comparisons for which the null is true, π0, by counting the number of P values larger than a cutoff (such as 0.5) relative to (1 –)N (such as N/2), the count expected when the distribution is uniform. If R discoveries are observed, about Nπ0 are expected to be false positives, and FDR can be estimated by Nπ0/R.

In this method, the P-values are first sorted and ranked. The smallest value gets rank 1, the second rank 2, and the largest gets rank N. Then, each P-value is multiplied by N and divided by its assigned rank to give the adjusted P-values.