

Litter Mining

oder

Why most published research findings are false.

Jörg Pflüger

Litteratur

; -)

- [1] Ioannidis, J.P.: Why Most Published Research Findings Are False, PLoS Medicine, August 2005
- [2] Wacholder, Sh. et. al.: Assessing the Probability That a Positive Report is False: An Approach for Molecular Epidemiology Studies, Journal of the National Cancer Institute, March 2004
- [3] Sterne, J.A.C., Smith, G.D.: Sifting the evidence – what’s wrong with significance tests?, BMJ January 2001
- [4] Winkler, R.L., Hays, W.L.: Statistics: Probability, Inference, and Decision, 1971/75
- [5] Cattell, R.B.: Handbook of Multivariate Experimental Psychology, 1966
- [6]
- [7]

Mining

»Mining is a dangerous profession - and data mining is just the same.«
Ben Goldacre, badscience-Blog

Data Mining
Web Mining
Reality Mining
Visual Datamining

...

Unterscheidung zwischen
Bastelkästen (Mining-Tools)
und
automatischem Profiling
(Registrierkassen, RFIDs,
Websites, Terroristenhatz)

The screenshot shows a Microsoft Internet Explorer browser window displaying the Amazon.com product page for 'Practical Internet Groupware'. The browser's address bar shows the URL: <http://www.amazon.com/exec/obidos/ASIN/1565925378/102-4289198-1281743>. The page features the Amazon.com logo, navigation links (VIEW CART, WISH LIST, YOUR ACCOUNT, HELP), and a search bar. The main content area displays the book 'Practical Internet Groupware' by Jon Udell and Tim O'Reilly, priced at \$7.19. The page also includes a 'RATE THIS ITEM' section with a 5-star rating scale and a 'Customers who shopped for this item also wear:' section with a list of related products.

Mining

»Mining is a dangerous profession - and data mining is just the same.«
Ben Goldacre, badscience-Blog

Data Mining
Web Mining
Reality Mining
Visual Datamining

...

Unterscheidung zwischen
Bastelkästen (Mining-Tools)
und
automatischem Profiling
(Registrierkassen, RFIDs,
Websites, Terroristenhatz)

The screenshot shows a Microsoft Internet Explorer browser window displaying the Amazon.com product page for 'Practical Internet Groupware' by Jon Udell and Tim O'Reilly. The browser's address bar shows the URL: <http://www.amazon.com/exec/obidos/ASIN/1565925378/102-4289198-1281743>. The page features the Amazon.com logo, navigation links (VIEW CART, WISH LIST, YOUR ACCOUNT, HELP), and a search bar. The product title is 'Practical Internet Groupware' by Jon Udell, Tim O'Reilly. The price is listed as 'Used & new from \$7.19'. The book is described as 'Usually ships in 1-2 business days' and 'Edition: Paperback'. There are buttons for 'Add to Wish List', 'Add to Wedding Registry', and 'See more product details'. A section titled 'Customers who shopped for this item also wear:' lists related items like 'Clean Underwear', 'Cashmere Sweaters', and 'Tommy Bahama Shirts'. The browser window title is 'Amazon.com: Books: Practical Internet Groupware - Microsoft Internet Explorer'.

Data Mining – Google-Definitionen

The ability to query very large databases in order to **satisfy a hypothesis** ("top down" data mining); or to interrogate a database in order to **generate new hypotheses** based on rigorous **statistical correlations** ("bottom-up" data mining).

Data mining is the sports-betting equivalent of sitting a huge number of monkeys down at keyboards, and then reporting on the monkeys who happened to type actual words.

Searching large volumes of data looking for patterns that accurately **predict behavior in customers and prospects**.

As the term suggests, data mining is the analysis of data to **establish relationships and identify patterns**.

The analysis of data for relationships that have **not previously been discovered**.

Nontrivial **extraction of implicit, previously unknown and potentially useful information** from data, or the search for **relationships and global patterns** that exist in databases.

The process of using **statistical techniques** to discover subtle relationships between data items, and the **construction of predictive models** based on them.

Extraction of useful information from data sets. Data mining serves to find **information that is hidden** within the available data.

The unguided (or minimally guided) application of a collection of mathematical procedures to a company's data warehouse in an effort to **find "nuggets"** in the form of **statistical relationships**.

Data mining is just what its name implies - searching through mountains of data to **find important values** or key relationships. Data mining can be performed in many ways from brute force calculations through **smart algorithms** to **visualisation**.

Using **advanced statistical tools** to identify **commercially useful patterns** in databases.

Data mining, also known as **knowledge-discovery** in databases (KDD), is the practice of automatically searching large stores of data for patterns. To do this, data mining uses computational techniques from **statistics and pattern recognition**.

A process of analysing **business data** (often stored in a data warehouse) to **uncover hidden trends and patterns** and establish relationships. Data mining is normally **performed by expert analysts** who use specialist software tools.

Data Mining – Google-Definitionen

The ability to query very large databases in order to **satisfy a hypothesis** ("top down" data mining); or to interrogate a database in order to **generate new hypotheses** based on rigorous **statistical correlations** ("bottom-up" data mining).

Data mining is the sports-betting equivalent of sitting a huge number of monkeys down at keyboards, and then reporting on the monkeys who happened to type actual words.

Searching large volumes of data looking for patterns that accurately **predict behavior in customers and prospects**.

As the term suggests, data mining is the analysis of data to **establish relationships and identify patterns**.

The analysis of data for relationships that have **not previously been discovered**.

Nontrivial **extraction of implicit, previously unknown and potentially useful information** from data, or the search for **relationships and global patterns** that exist in databases.

The process of using **statistical techniques** to discover subtle relationships between data items, and the **construction of predictive models** based on them.

Extraction of useful information from data sets. Data mining serves to find **information that is hidden** within the available data.

The unguided (or minimally guided) application of a collection of mathematical procedures to a company's data warehouse in an effort to **find "nuggets"** in the form of **statistical relationships**.

Data mining is just what its name implies - searching through mountains of data to **find important values** or key relationships. Data mining can be performed in many ways from brute force calculations through **smart algorithms** to **visualisation**.

Using **advanced statistical tools** to identify **commercially useful patterns** in databases.

Data mining, also known as **knowledge-discovery** in databases (KDD), is the practice of automatically searching large stores of data for patterns. To do this, data mining uses computational techniques from **statistics and pattern recognition**.

A process of analysing **business data** (often stored in a data warehouse) to **uncover hidden trends and patterns** and establish relationships. Data mining is normally **performed by expert analysts** who use specialist software tools.

Data Mining – Google-Definitionen

The ability to query very large databases in order to **satisfy a hypothesis** ("top down" data mining); or to interrogate a database in order to **generate new hypotheses** based on rigorous **statistical correlations** ("bottom-up" data mining).

Data mining is the sports-betting equivalent of sitting a huge number of monkeys down at keyboards, and then reporting on the monkeys who happened to type actual words.

Searching large volumes of data looking for patterns that accurately **predict behavior in customers and prospects**.

As the term suggests, data mining is the analysis of data to **establish relationships and identify patterns**.

The analysis of data for relationships that have **not previously been discovered**.

Nontrivial **extraction of implicit, previously unknown and potentially useful information** from data, or the search for **relationships and global patterns** that exist in databases.

The process of using **statistical techniques** to discover subtle relationships between data items, and the **construction of predictive models** based on them.

Extraction of useful information from data sets. Data mining serves to find **information that is hidden** within the available data.

The unguided (or minimally guided) application of a collection of mathematical procedures to a company's data warehouse in an effort to **find "nuggets"** in the form of **statistical relationships**.

Data mining is just what its name implies - searching through mountains of data to **find important values** or key relationships. Data mining can be performed in many ways from brute force calculations through **smart algorithms** to **visualisation**.

Using **advanced statistical tools** to identify **commercially useful patterns** in databases.

Data mining, also known as **knowledge-discovery** in databases (KDD), is the practice of automatically searching large stores of data for patterns. To do this, data mining uses computational techniques from **statistics and pattern recognition**.

A process of analysing **business data** (often stored in a data warehouse) to **uncover hidden trends and patterns** and establish relationships. Data mining is normally **performed by expert analysts** who use specialist software tools.

Schatz- und Sinnsuche, nützlich, Muster, Entbergen, Statistik

Statistische Aussagen – allenthalben

Bei großen Datenmengen sind die meisten Aussagen statistischer Art.
In fast allen Wissensgebieten und in der Lebenswelt spielen statistische Modelle eine immer fundamentalere Rolle:

Physik (statistische Mechanik, Quantenmechanik)

Sozialwissenschaften (Psychologie, Soziologie, Ökonomie, Politik)

Medizin, Biologie, Gentechnologie

Paläontologie, Ökologie

Informatik (Routing, ??)

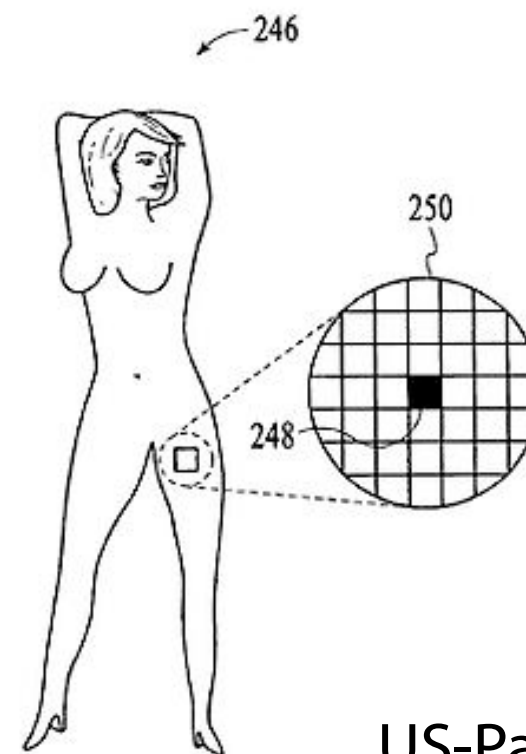
Information Retrieval (Clustering)

Business (Profiling, Bewertungen)

Google Ranking (random walk)

Sicherheitstechnologie

➔ Data Mining ~ Statistik



$$V = \frac{\sum_i^P \sqrt{\text{colour}^2}}{P}$$

V is variance

P is number of pixels

$$\text{colour}^2 = (L - \text{mean } L)^2 + (a - \text{mean } a)^2 + (b - \text{mean } b)^2$$

$$\text{mean } L = \frac{\sum_1^P L}{P} \quad \text{mean } a = \frac{\sum_1^P a}{P}$$

$$\text{mean } b = \frac{\sum_1^P b}{P}$$

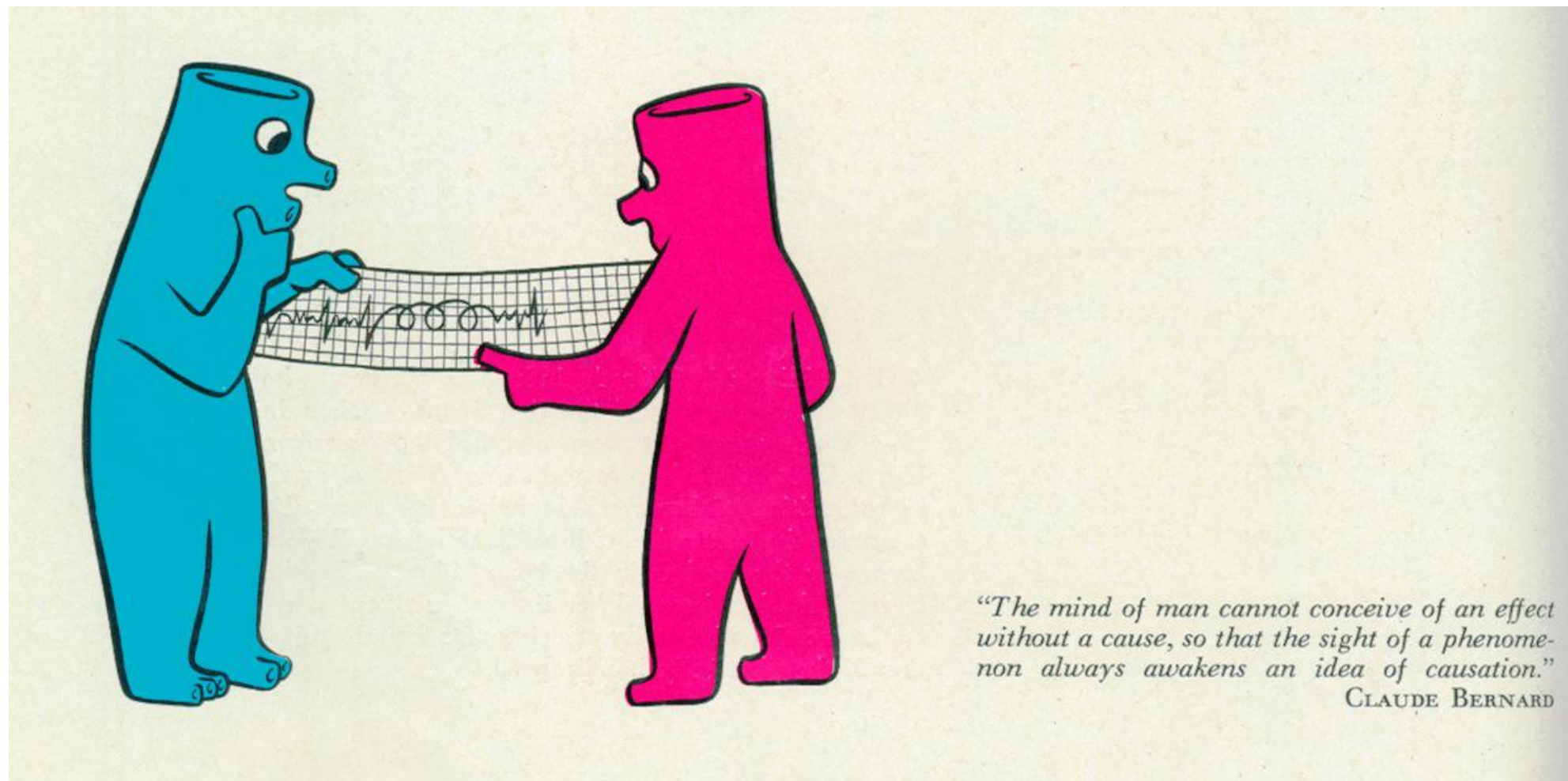
$V \geq 10 \Rightarrow$ not skin

$V < 10 \Rightarrow$ skin

US-Patent 6751348 zur automatischen Erkennung von Pornos

Wer schätzt die Schätzer?

Im Verhältnis zur Bedeutung von statistischen Aussagen gibt es erstaunlich wenig Gefühl und Bewußtheit von deren Bedeutung. Das schlichte Stereotyp des »Lügens mit Statistik« korrespondiert mit mißverstehendem Glauben an statistische Behauptungen; die Wenigsten können Schätzungen richtig einschätzen.



Schwarting, Heart Bulletin 1956

**Raucher
sterben
früher**

Wer das Rauchen
aufgibt, verringert
das Risiko tödlicher
Herz- und Lungen-
erkrankungen

Veränderung des Denkens

Ich glaube, daß Data Mining eine der fundamentalen epistemischen Veränderungen repräsentiert:

top-down → bottom-up,

vom (systematischen) Ausdenken zum Überprüfen und Bewerten.

(Analoges in der Kunstproduktion und beim Web:

»the pre-Internet world was one of "select, then publish." ... The Internet world is one of "publish, then select."« (Boing Boing, gestern)

Charakteristisch für die ›neue‹ (wissenschaftliche) Sicht auf die Welt:
von (technisch aufbereiteten) Daten zu Information/Wissen,
weder deduktiv (keine Ausgangsthese) noch induktiv (kein Exemplum).

Wissenschaftlichkeit und Statistik

Klassische Wissenschaftsvorstellung (hard science): nur deduktiv;
induktiv ist problematisch, zweifelhaft
(wissenschaftstheoretische Diskussionen ohne Ende).

Wiederholt sich in der Statistik (Psychologie):

Klassische Inferenzstatistik:

Hypothese → Datenerhebung → Auswertung (hypothetico-deduktive)
über

Bayessche Statistik: Einbeziehung von ›subjektiver‹, externer Information;
durchbricht die streng deduktive Objektivitätsvorstellung

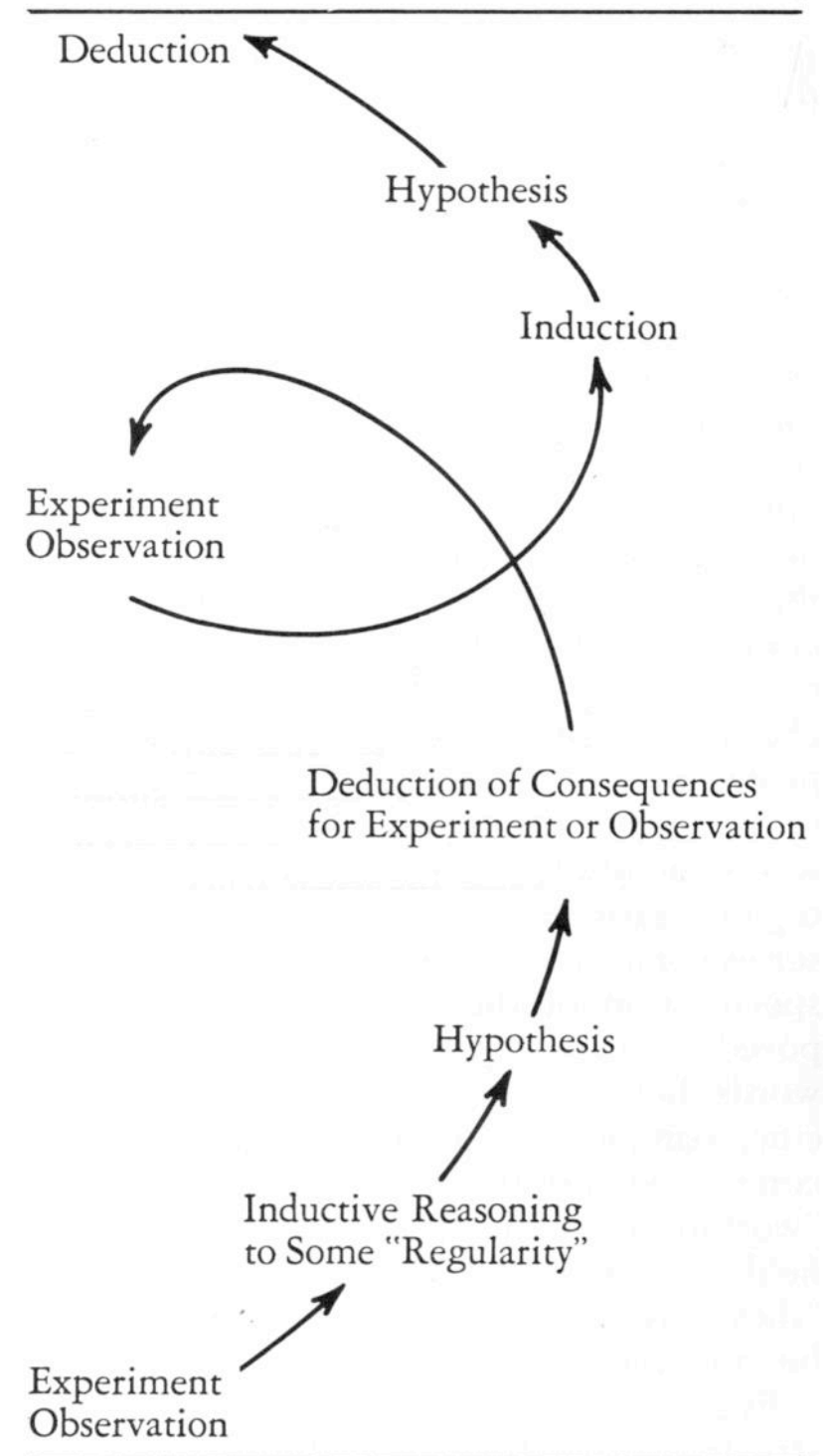


Hypothesen-generierende Verfahren (multivariate Verfahren wie
Faktorenanalyse, Cluster, nur möglich durch den Computer) und
explorative Datenanalyse → Data Mining

Inductive-hypothetico-deductive spiral

»The standard cliché of textbook scientific method thus commonly misrepresents to this extent the living psychological process of hypothetico-deductive research and does so in two major respects. First, it chops out a single link – the testing of a specific hypothesis – from what is truly a chain, or more exactly, a spiral. Second, it speaks of *the* hypothesis. ... The initial surmises grow, through a long series of inspired guesses – checking against facts, deducing further properties, and checking again – into a *series* of links which eventually come to consist of well-structured hypothesis in a sort of organic growth. ... It is for this reason that the traditional term “hypothetico-deductive method” is so misleading. For this describes only one part of the cycle – the legalistic and disputative rather than the explorative and more scientifically creative part. If there is any part of the spiral which can be called the scientific beginning, it is the induction rather than the deduction. But what we can be certain about is that the complete cycle is an inductive-hypothetico-deductive-experimental-inductive one, no matter where we decide to cut it. [5, 1966, S.15]

The Inductive-Hypothetico-Deductive Spiral



Der Computer

»The rapid growth of computer size and complex program construction in this generation is going to overcome many of the obstacles of sweat and cost which for years have kept many fairly energetic psychologists rationalizing against multivariate designs. Inevitably it may also cause much pointless tossing about of masses of data by uninspired researchers; but this is less important than the encouragement it will give to correct and potent uses. Notably we can expect a richer flow of well-born hypotheses, with a healthy and realistic aptness to the data, from the beginning.« [5, 1966, S. 63]

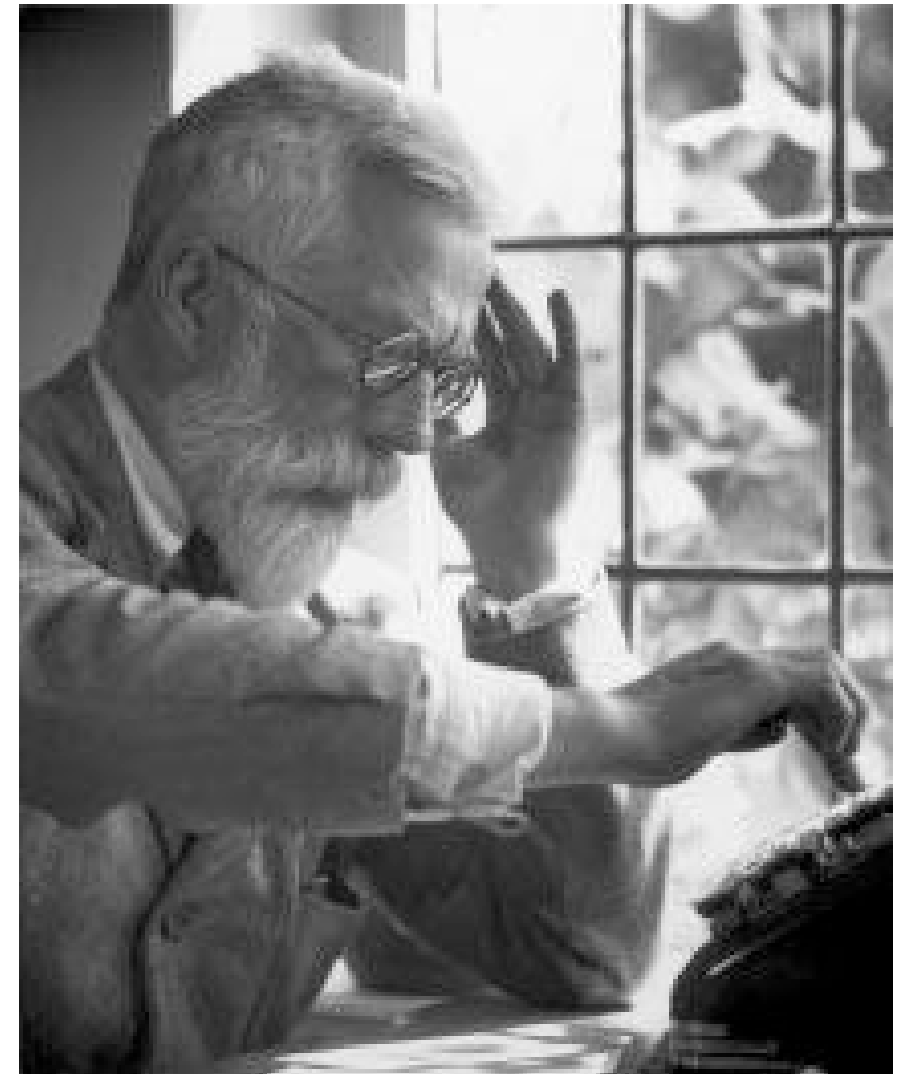
Der Computer

»The evolution of standard statistical computer programs has greatly eased the computational burden on the statistician and has made it possible to analyze larger and larger problems, problems that would be impossible to handle other than by computer. This development has not been entirely without drawbacks, however. With some programs, the printed output of results is almost as massive as the data used as input. The statistician is now faced with the task of sorting out meaningful results from relatively unimportant results. ... If we know exactly what we are looking for, then we should have little difficulty, ... On the other hand, in performing hand calculations or working on a calculating machine, the statistician is able to follow the analysis step by step. In so doing, a “feel” for the problem (and for the statistical techniques involved) is acquired, a “feel” which is often hard to get from reading printed output from computer. ... In some respects, the standard statistical programs are almost too *easy* to use. By this we mean that it is possible for a person with very little knowledge of statistics to use one of these complex programs even though that person has little idea as to what the program does or what much of the output means. This is a weakness not of the computer programs themselves, but of the users of the program. ...

In the computer age, we can spend more time concentrating on the concepts and less time on the calculations.« [4, 1971/75, S. 701]

Der Computer

»In the days when perhaps only a few hundred statistical hypotheses were being tested each year, and when calculations had to be done laboriously with mechanical hand calculators (as in Fisher's photograph), it seemed reasonable that a 5% false positive rate would screen out most of the random errors. With many thousands of journals publishing a myriad hypothesis tests each year and the ease of use of statistical software it is likely that the proportion of tested hypotheses that are meaningful (in the sense that the effect is large enough to be of interest) has decreased, leading to a finding of $P < 0.05$ having low predictive value for the appropriate rejection of the null hypothesis.« [3, 2001, S. 228]



»R.A. Fisher, the founder of statistical inference, working on a mechanical calculator«

Der Computer

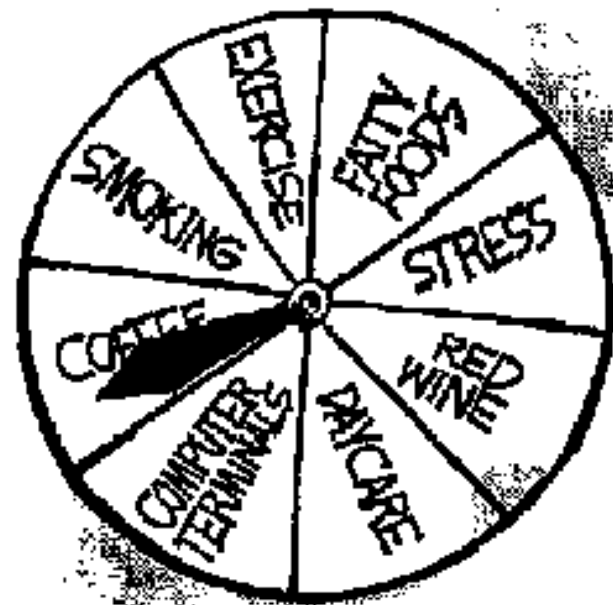
»Commercially available “data mining” packages actually are proud of their ability to yield statistically significant results through data dredging.« [1, 2005, S.699]

Statistik Crash-Kurs

Today's Random Medical News

from the New England Journal of Panic-Inducing Gobbledygook

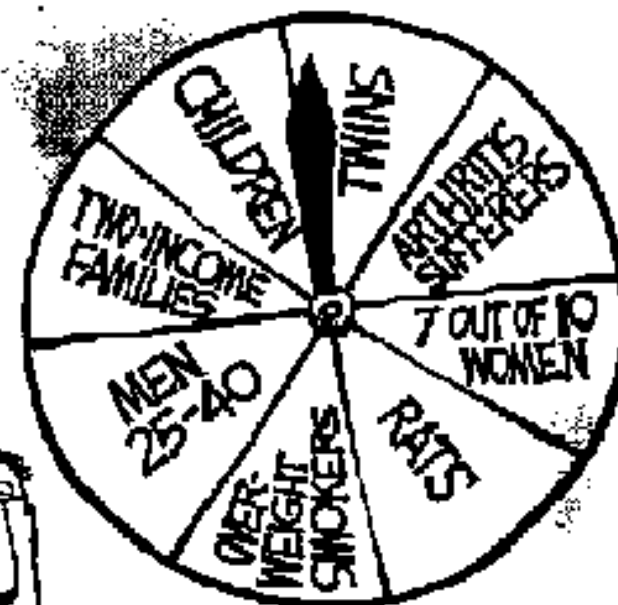
WILSON



CAN CAUSE



IN



Beispielhaftes

Experiment: *Wurf mit zwei unterscheidbaren Würfeln*

Elementare Ereignisse: $\mathcal{E} = \{ \text{red 1, green 1}, \text{red 1, green 2}, \dots, \text{red 6, green 6} \}$

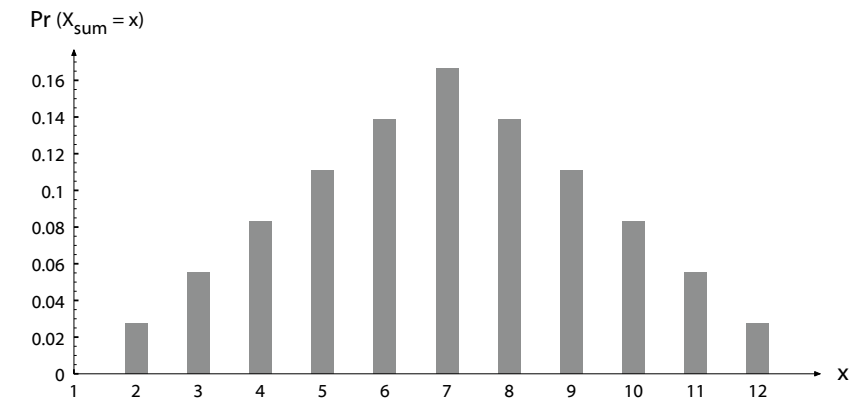
Ereignis: z.B. $A_{Pasch} = \{ \text{red 1, green 1}, \text{red 1, green 2}, \text{red 1, green 3}, \dots, \text{red 6, green 6} \}$

Wahrscheinlichkeit: $\Pr(\text{red 1, green 1}) = \Pr(\text{red 1, green 2}) = \dots = \Pr(\text{red 6, green 6}) = 1/36$
 $\Pr(A_{Pasch}) = 1/6$

Zufallsvariable: $X_{sum} = \text{Augensumme}$

Wahrscheinlichkeitsverteilung
der Zufallsvariablen X_{sum} :

$\Pr(2) = \Pr(12) = 1/36$
$\Pr(3) = \Pr(11) = 2/36$
$\Pr(4) = \Pr(10) = 3/36$
$\Pr(5) = \Pr(9) = 4/36$
$\Pr(6) = \Pr(8) = 5/36$
$\Pr(7) = 6/36$



Wahrscheinlichkeitsverteilung der Würfelsumme

Mittelwert von X_{sum} :

$$\mu_{sum} = E(X_{sum}) = (2 \cdot 1 + 3 \cdot 2 + 4 \cdot 3 + 5 \cdot 4 + 6 \cdot 5 + 7 \cdot 6 + 8 \cdot 5 + 9 \cdot 4 + 10 \cdot 3 + 11 \cdot 2 + 12 \cdot 1) \cdot 1/36 = 7$$

Standardabweichung von X_{sum} :

$$\sigma_{sum} = \sqrt{E([X_{sum} - 7]^2)} = \sqrt{((-5)^2 \cdot 1 + (-4)^2 \cdot 2 + (-3)^2 \cdot 3 + (-2)^2 \cdot 4 + (-1)^2 \cdot 5 + 0^2 \cdot 6 + 1^2 \cdot 5 + 2^2 \cdot 4 + 3^2 \cdot 3 + 4^2 \cdot 2 + 5^2 \cdot 1) \cdot 1/36} \approx 2.415$$

Begriffe

Experiment: irgendeine Aktivität mit eindeutigen Resultaten

Menge der elementaren Ereignisse: \mathcal{E} (= alle möglichen Resultate)

Menge der Ereignisse: $\mathcal{A} = \mathcal{P}(\mathcal{E}) = \{A \subseteq \mathcal{E}\}$ (= alle Teilmengen von \mathcal{E})

Wahrscheinlichkeit: $\text{Pr} : \mathcal{A} \longrightarrow [0, 1]$ mit

1. $\text{Pr}(\mathcal{E}) = 1$

2. $\text{Pr}\left(\bigsqcup_{i \in I} A_i\right) = \sum_{i \in I} \text{Pr}(A_i)$ ($A_i \cap A_j = \emptyset$, wenn $i \neq j$.)

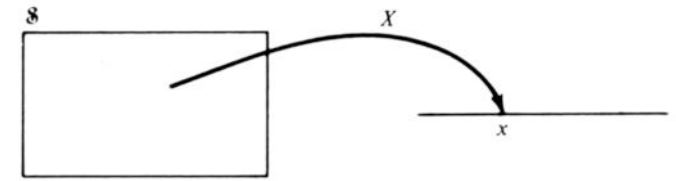
(Speziell: $\text{Pr}(A) = \frac{|A|}{|\mathcal{E}|}$, falls alle Elementarereignisse gleichwahrscheinlich.)

Wahrscheinlichkeitsverteilung: $\left\{ \text{Pr}(A_j) \mid A_j \in \mathcal{A}, j \in J \text{ und } \bigsqcup_{j \in J} A_j = \mathcal{E} \right\}$

(Spezialfall der Elementarereignisse-Partition: $\{ \text{Pr}(E_j) \mid E_j \in \mathcal{E} \}$.)

Begriffe

Zufallsvariable: $X : \mathcal{E} \longrightarrow \mathfrak{R}$ (bzw. in \mathcal{E}')



A Random Variable as a Function from the Sample Space to the Real Line

Zufallsvariable induziert Wahrscheinlichkeit auf ihrem Bildbereich:

$\Pr : X(\mathcal{E}) \longrightarrow [0, 1]$, mit $\Pr(x) = \Pr(X = x) = \sum_{X(E_i)=x} \Pr(E_i) = \Pr(A_x = \{E_i \mid X(E_i) = x\})$

(Über Zufallsvariablen können Ereignisse definiert werden,

z.B. $\Pr(a < X < b) = \Pr(A_{(a,b)} = \{E_i \mid a < X(E_i) < b\})$.)

Verteilung einer Zufallsvariablen: $\{ \langle x_i, \Pr(X = x_i) \rangle \mid x_i \in X(\mathcal{E}) \}$

Erwartungswert: $E(X) = \sum_x x \cdot \Pr(X = x) = \sum_x x \cdot \Pr(X)$

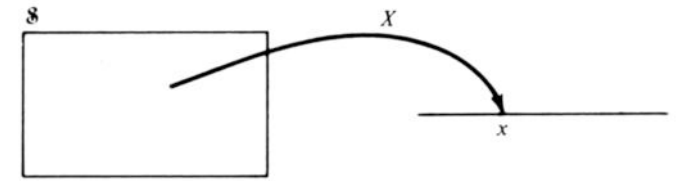
Mittelwert: $\mu = E(X)$

Varianz: $\sigma^2 = V(X) = E([X - \mu]^2) = E(X^2) - \mu^2$ (mtl. quadr. Abweichung)

Standardabweichung: $\sigma = \sqrt{\sigma^2}$

Begriffe

Zufallsvariable: $X : \mathcal{E} \longrightarrow \mathfrak{R}$ (bzw. in \mathcal{E}')



A Random Variable as a Function from the Sample Space to the Real Line

Zufallsvariable induziert Wahrscheinlichkeit auf ihrem Bildbereich:

$$\Pr : X(\mathcal{E}) \longrightarrow [0, 1], \text{ mit } \Pr(x) = \Pr(X = x) = \sum_{X(E_i)=x} \Pr(E_i) = \Pr(A_x = \{E_i \mid X(E_i) = x\})$$

(Über Zufallsvariablen können Ereignisse definiert werden,

$$\text{z.B. } \Pr(a < X < b) = \Pr(A_{(a,b)} = \{E_i \mid a < X(E_i) < b\}).)$$

Verteilung einer Zufallsvariablen: $\{ \langle x_i, \Pr(X = x_i) \rangle \mid x_i \in X(\mathcal{E}) \}$

$$\text{Erwartungswert: } E(X) = \sum_x x \cdot \Pr(X = x) = \sum_x x \cdot \Pr(X)$$

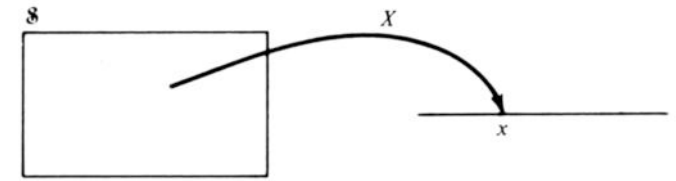
$$\text{Mittelwert: } \mu = E(X)$$

$$\text{Varianz: } \sigma^2 = V(X) = E([X - \mu]^2) = E(X^2) - \mu^2 \quad (\text{mtl. quadr. Abweichung})$$

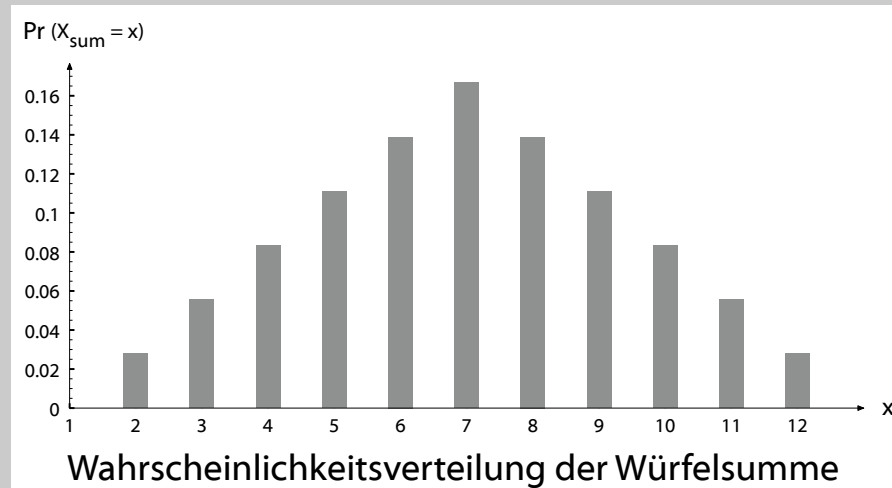
$$\text{Standardabweichung: } \sigma = \sqrt{\sigma^2}$$

Begriffe

Zufallsvariable (bzw. in \mathcal{E}')



A Random Variable as a Function from the Sample Space to the Real Line



Wahrscheinlichkeit auf ihrem Bildbereich:

$$= \sum_{X(E_i)=x} \Pr(E_i) = \Pr(A_x = \{E_i \mid X(E_i) = x\})$$

(Übersetzungsergebnisse definiert werden,

z.B. $\Pr(a < X(E_i) < b)$.)

Verteilung einer Zufallsvariablen: $\{ \langle x_i, \Pr(X = x_i) \rangle \mid x_i \in X(\mathcal{E}) \}$

Erwartungswert: $E(X) = \sum_x x \cdot \Pr(X = x) = \sum_x x \cdot \Pr(X)$

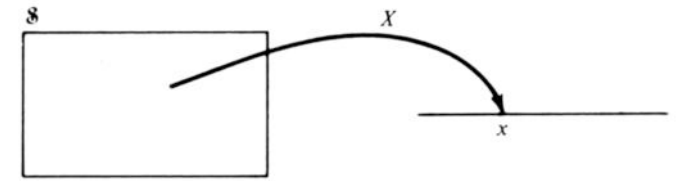
Mittelwert: $\mu = E(X)$

Varianz: $\sigma^2 = V(X) = E([X - \mu]^2) = E(X^2) - \mu^2$ (mtl. quadr. Abweichung)

Standardabweichung: $\sigma = \sqrt{\sigma^2}$

Begriffe

Zufallsvariable: $X : \mathcal{E} \longrightarrow \mathfrak{R}$ (bzw. in \mathcal{E}')



A Random Variable as a Function from the Sample Space to the Real Line

Zufallsvariable induziert Wahrscheinlichkeit auf ihrem Bildbereich:

$\Pr : X(\mathcal{E}) \longrightarrow [0, 1]$, mit $\Pr(x) = \Pr(X = x) = \sum_{X(E_i)=x} \Pr(E_i) = \Pr(A_x = \{E_i \mid X(E_i) = x\})$

(Über Zufallsvariablen können Ereignisse definiert werden,

z.B. $\Pr(a < X < b) = \Pr(A_{(a,b)} = \{E_i \mid a < X(E_i) < b\})$.)

Verteilung einer Zufallsvariablen: $\{ \langle x_i, \Pr(X = x_i) \rangle \mid x_i \in X(\mathcal{E}) \}$

Erwartungswert: $E(X) = \sum_x x \cdot \Pr(X = x) = \sum_x x \cdot \Pr(X)$

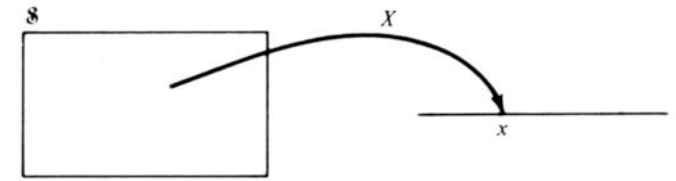
Mittelwert: $\mu = E(X)$

Varianz: $\sigma^2 = V(X) = E([X - \mu]^2) = E(X^2) - \mu^2$ (mtl. quadr. Abweichung)

Standardabweichung: $\sigma = \sqrt{\sigma^2}$

Begriffe

Zufallsvariable: $X : \mathcal{E} \longrightarrow \mathfrak{R}$ (bzw. in \mathcal{E}')



A Random Variable as a Function from the Sample Space to the Real Line

Zufallsvariable induziert Wahrscheinlichkeit auf ihrem Bildbereich:

$\text{Pr} : X(\mathcal{E}) \longrightarrow [0, 1]$, mit $\text{Pr}(x) = \text{Pr}(\{E_i \mid X(E_i) = x\})$

(Über Zufallsvariablen)

z.B. $\text{Pr}(a < X < b) = \text{Pr}(\{E_i \mid a < X(E_i) < b\})$

Verteilung einer Zufallsvariable

Eine Zufallsvariable X beschreibt eine Messung **vor** der Messung; wir wissen nicht, welcher Wert x sich im Einzelfall ergeben wird, aber wir können auf die Dauer einen Durchschnittswert von $E(X)$ **erwarten**.
($E(X)$ muß nicht der häufigste Wert sein!)

Erwartungswert: $E(X) = \sum_x x \cdot \text{Pr}(X = x) = \sum_x x \cdot \text{Pr}(X)$

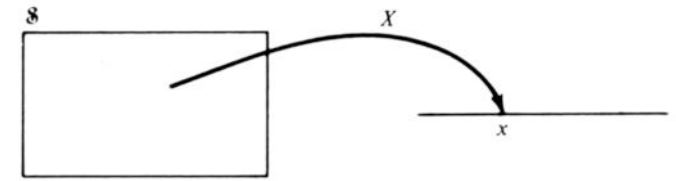
Mittelwert: $\mu = E(X)$

Varianz: $\sigma^2 = V(X) = E([X - \mu]^2) = E(X^2) - \mu^2$ (mtl. quadr. Abweichung)

Standardabweichung: $\sigma = \sqrt{\sigma^2}$

Begriffe

Zufallsvariable: $X : \mathcal{E} \longrightarrow \mathfrak{R}$ (bzw. in \mathcal{E}')



A Random Variable as a Function from the Sample Space to the Real Line

Zufallsvariable induziert Wahrscheinlichkeit auf ihrem Bildbereich:

$\Pr : X(\mathcal{E}) \longrightarrow [0, 1]$, mit $\Pr(x) = \Pr(X = x) = \sum_{X(E_i)=x} \Pr(E_i) = \Pr(A_x = \{E_i \mid X(E_i) = x\})$

(Über Zufallsvariablen können Ereignisse definiert werden,

z.B. $\Pr(a < X < b) = \Pr(A_{(a,b)} = \{E_i \mid a < X(E_i) < b\})$.)

Verteilung einer Zufallsvariablen: $\{ \langle x_i, \Pr(X = x_i) \rangle \mid x_i \in X(\mathcal{E}) \}$

Erwartungswert: $E(X) = \sum_x x \cdot \Pr(X = x) = \sum_x x \cdot \Pr(X)$

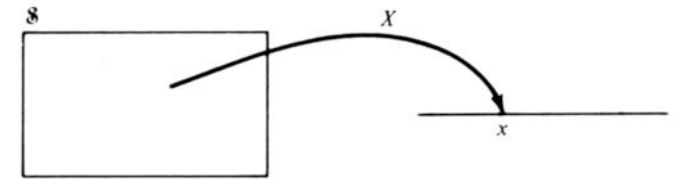
Mittelwert: $\mu = E(X)$

Varianz: $\sigma^2 = V(X) = E([X - \mu]^2) = E(X^2) - \mu^2$ (mtl. quadr. Abweichung)

Standardabweichung: $\sigma = \sqrt{\sigma^2}$

Begriffe

Zufallsvariable: $X : \mathcal{E} \longrightarrow \mathfrak{R}$ (bzw. in \mathcal{E}')



A Random Variable as a Function from the Sample Space to the Real Line

Zufallsvariable induziert Wahrscheinlichkeit auf ihrem Bildbereich:

$\Pr : X(\mathcal{E}) \longrightarrow [0, 1]$, mit $\Pr(x) = \Pr(X = x) = \sum_{X(E_i)=x} \Pr(E_i) = \Pr(A_x = \{E_i \mid X(E_i) = x\})$

(Über Zufallsvariablen können Ereignisse definiert werden,

z.B. $\Pr(a < X < b) = \Pr(A_{(a,b)} = \{E_i \mid a < X(E_i) < b\})$

Verteilung einer Zufallsvariable

Erwartungswert: $E(X) = \sum_x$

Die Varianz mißt die ›Streuung‹ der Werte; durch die Quadrierung gehen positive und negative Abweichung vom Mittelwert gleich ein [$E(X - \mu) = 0$], außerdem werden große Abweichungen stärker berücksichtigt.

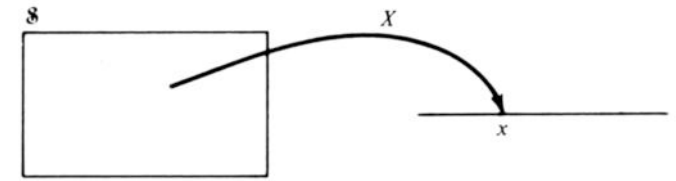
Mittelwert: $\mu = E(X)$

Varianz: $\sigma^2 = V(X) = E([X - \mu]^2) = E(X^2) - \mu^2$ (mtl. quadr. Abweichung)

Standardabweichung: $\sigma = \sqrt{\sigma^2}$

Begriffe

Zufallsvariable: $X : \mathcal{E} \longrightarrow \mathfrak{R}$ (bzw. in \mathcal{E}')



A Random Variable as a Function from the Sample Space to the Real Line

Zufallsvariable induziert Wahrscheinlichkeit auf ihrem Bildbereich:

$\Pr : X(\mathcal{E}) \longrightarrow [0, 1]$, mit $\Pr(x) = \Pr(X = x) = \sum_{X(E_i)=x} \Pr(E_i) = \Pr(A_x = \{E_i \mid X(E_i) = x\})$

(Über Zufallsvariablen können Ereignisse definiert werden,

z.B. $\Pr(a < X < b) = \Pr(A_{(a,b)} = \{E_i \mid a < X(E_i) < b\})$.)

Verteilung einer Zufallsvariablen: $\{ \langle x_i, \Pr(X = x_i) \rangle \mid x_i \in X(\mathcal{E}) \}$

Erwartungswert: $E(X) = \sum_x x \cdot \Pr(X = x) = \sum_x x \cdot \Pr(X)$

Mittelwert: $\mu = E(X)$

Varianz: $\sigma^2 = V(X) = E([X - \mu]^2) = E(X^2) - \mu^2$ (mtl. quadr. Abweichung)

Standardabweichung: $\sigma = \sqrt{\sigma^2}$

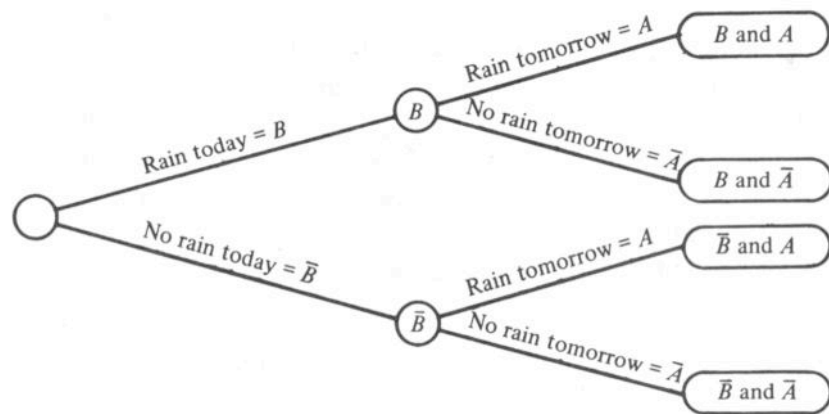
Begriffe

Produktwahrscheinlichkeit: $\Pr(A, B)$ bzw. $\Pr(X = a, Y = b)$

Bedingte Wahrscheinlichkeit: $\Pr(A|B)$ bzw. $\Pr(X = a | Y = b)$

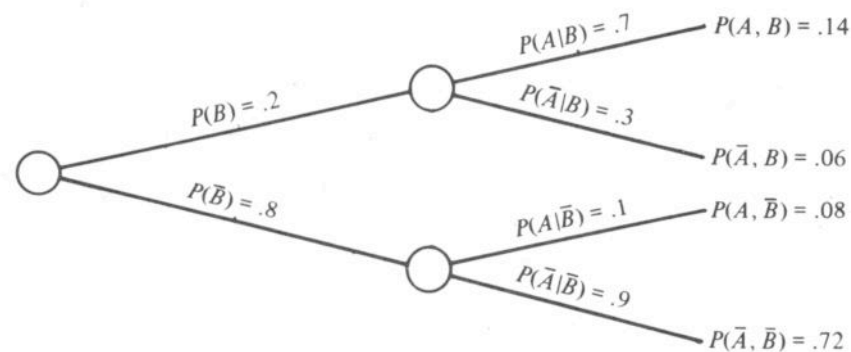
Zusammenhang von bedingter und Produktwahrscheinlichkeit:

$$\Pr(A, B) = \Pr(A|B) \cdot \Pr(B) \text{ bzw. } \Pr(X = a, Y = b) = \Pr(X = a | Y = b) \cdot \Pr(Y = b)$$



A Tree Diagram for the Occurrence or Nonoccurrence of Rain

Baum- und Tafeldarstellung



A Tree Diagram with Probabilities for the Rain Example

Joint Probabilities in Tabular Form

	A	\bar{A}	
B	$P(A, B)$	$P(\bar{A}, B)$	$P(B)$
\bar{B}	$P(A, \bar{B})$	$P(\bar{A}, \bar{B})$	$P(\bar{B})$
	$P(A)$	$P(\bar{A})$	

Joint Probabilities for Rain Example

	Rain tomorrow	No rain tomorrow	
Rain today	.14	.06	.20
No rain today	.08	.72	.80
	.22	.78	

Verschiedene Interpretationen des formalen Modells

1. traditionell (Glücksspielproblemstellung)

$\Pr(E)$ = günstige Ergebnisse / alle Ergebnisse,
setzt Gleichverteilung der Elementarereignisse voraus.

2. long-run frequency, frequentist (verbreitetstes Modell)

$\Pr(E) = 0.9$ bedeutet: E tritt in 90% aller Fälle auf.
»objektiv«, weil nur beobachtete Ereignisse berücksichtigt werden,
Grenzwert von relativen Häufigkeiten (Gesetz der großen Zahl),
setzt beliebig wiederholbare Ereignisse voraus.

3. »subjektiv« (Bayessche Statistik)

$\Pr(E)$ = »degree of belief«: »Ich wette 5:1, daß es morgen regnet.«,
geht auch für einmalige Ereignisse. In Verbindung mit Bayesscher
Statistik werden zusätzliche apriori Informationen einbezogen.

A priori und a posteriori

Aus $\Pr(A|B) \cdot \Pr(B) = \Pr(A, B) = \Pr(B, A) = \Pr(B|A) \cdot \Pr(A)$ folgt:

Bayessches Theorem:

$$\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B)} = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B|A) \cdot \Pr(A) + \Pr(B|\bar{A}) \cdot \Pr(\bar{A})}$$

Wenn θ ein Parameter und y ein beobachtetes Ereignis ist, gilt somit:

$$\Pr(\theta_i|y) = \frac{\Pr(y|\theta_i) \cdot \Pr(\theta_i)}{\Pr(y)} = \frac{\Pr(y|\theta_i) \cdot \Pr(\theta_i)}{\sum_{j=1}^J \Pr(y|\theta_j) \Pr(\theta_j)}$$

In der Bayesschen Statistik wird das als Wahrscheinlichkeit vor (a priori) und nach einem Experiment (a posteriori) interpretiert:

A priori und a posteriori

Aus $\Pr(A|B) \cdot \Pr(B) = \Pr(A, B) = \Pr(B, A) = \Pr(B|A) \cdot \Pr(A)$ folgt:

Bayessches Theorem:

$$\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B)} = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B|A) \cdot \Pr(A) + \Pr(B|\bar{A}) \cdot \Pr(\bar{A})}$$

Wenn θ ein Parameter und y ein beobachtetes Ereignis ist, gilt somit:

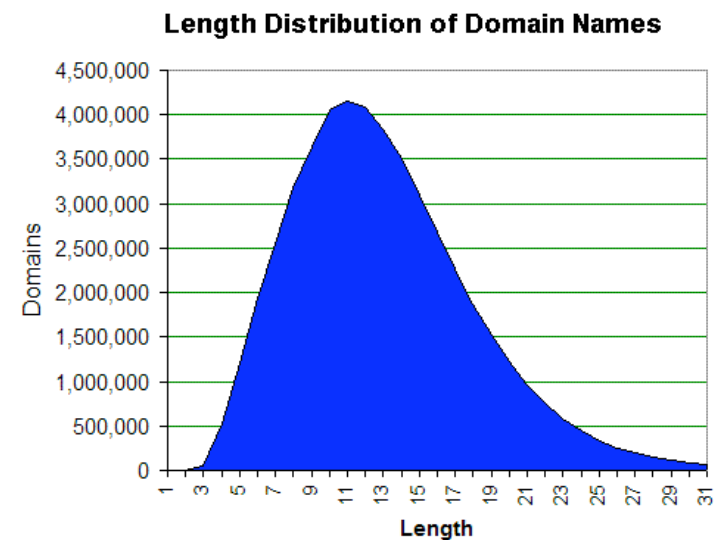
$$\Pr(\theta_i|y) = \frac{\Pr(y|\theta_i) \cdot \Pr(\theta_i)}{\Pr(y)} = \frac{\Pr(y|\theta_i) \cdot \Pr(\theta_i)}{\sum_{j=1}^J \Pr(y|\theta_j) \Pr(\theta_j)}$$

In der Bayesschen Statistik wird das als Wahrscheinlichkeit vor (a priori) und nach einem Experiment (a posteriori) interpretiert:

$$\text{posterior probability} = \frac{\text{likelihood} \cdot \text{prior probability}}{\sum \text{likelihood} \cdot \text{prior probability}}$$

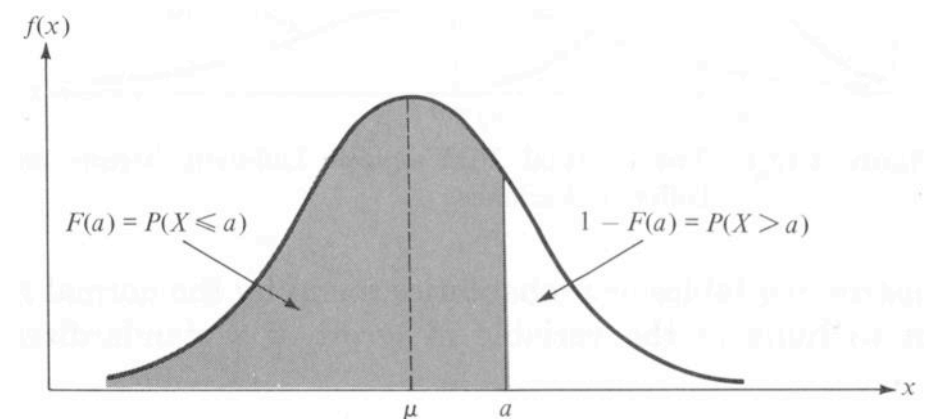
Kontinuierliche Idealisierung

Bei großen Wertemengen ist es oft einfacher mit einer kontinuierlichen Verteilung zu rechnen, die idealisiert als Grenzverteilung einer diskreten Verteilung angenommen wird.



Bei einer kontinuierlichen Zufallsvariablen X hat jedes von Null verschiedene Intervall $[a, b]$ eine Wahrscheinlichkeit $\Pr(a < X < b) > 0$.

Da für einzelne Werte $\Pr(X = a) = 0$ ist, wird die Wahrscheinlichkeit eines Wertintervalls als Fläche (Integral) unter einer Dichtefunktion bestimmt.

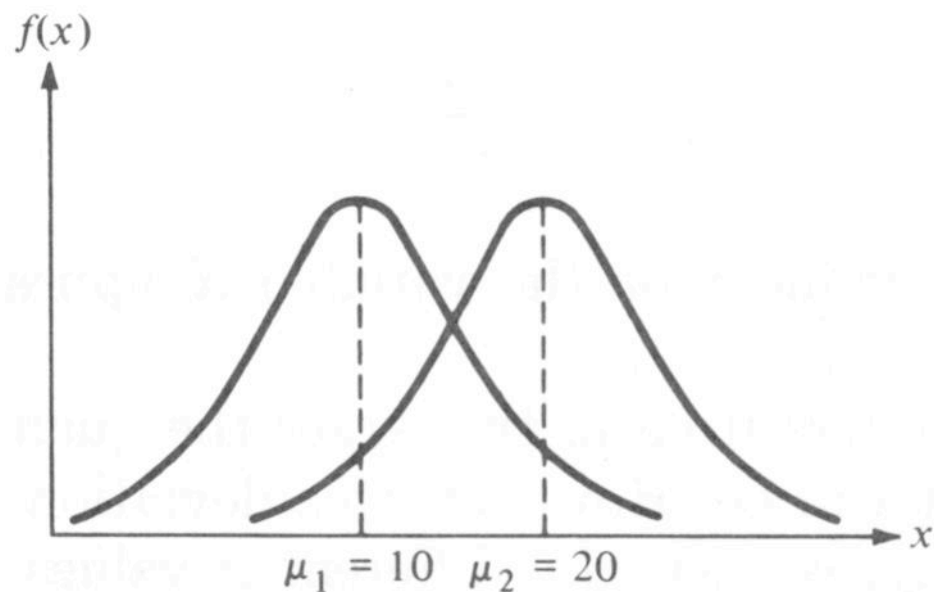


Probabilities as Areas Under a Normal Density Function

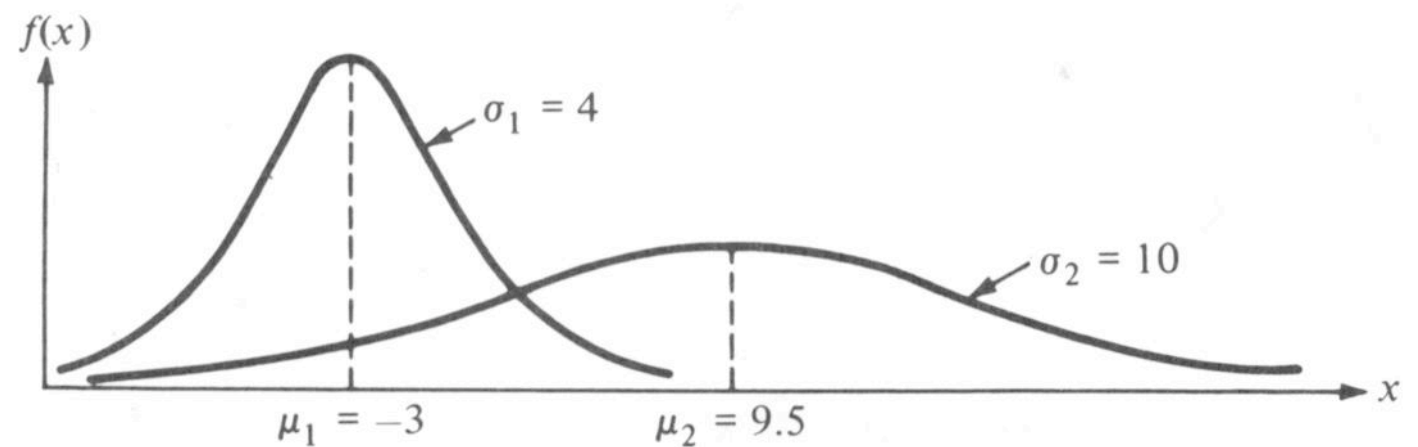
$$\Pr(a \leq X \leq b) = \int_a^b f(x) dx = \Pr(X \leq b) - \Pr(X < a)$$

Die Verteilung mit den nettesten Eigenschaften ist die Normalverteilung.

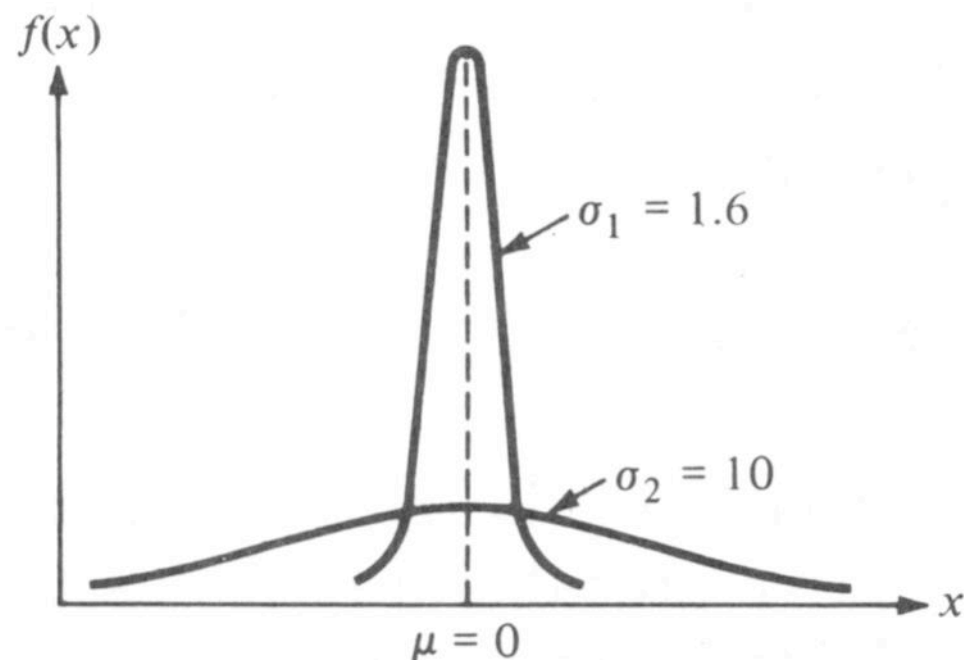
Unterschiedlich parametrisierte Normalverteilungen



Two Normal PDF's with Different Means



Two Normal PDF's with Different Means *and* Different Variances



Two Normal PDF's with Different Variances

Durch die Parameter μ und σ wird eine Normalverteilung eindeutig festgelegt.

Da die Fläche unter der Glockenkurve immer 1 ist, ergibt eine kleinere Varianz eine höhere Dichtefunktion.

Vergleichbarkeit durch Standardisierung

Die Transformation

$$X \mapsto Z, \text{ mit } Z = \frac{X - \mu}{\sigma}$$

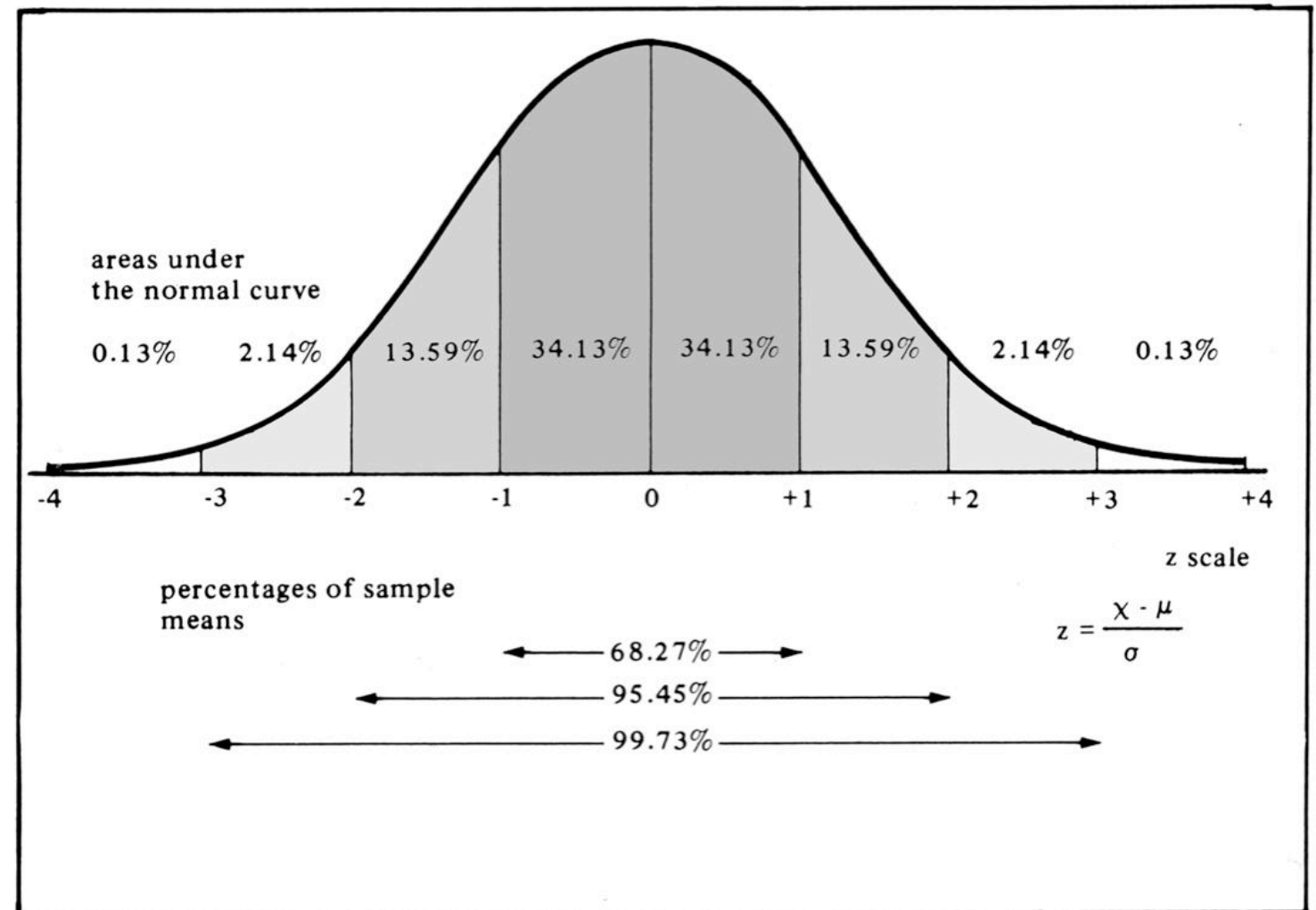
führt zur

Standardnormalverteilung
mit $\mu_Z = 0$ und $\sigma_Z = 1$.

Rückwärts: $X = \mu + \sigma \cdot Z$.

(Die Sampling Distribution

wird zu $Z = \frac{M - \mu}{\sigma/\sqrt{n}}$.)



Vergleichbarkeit durch Standardisierung

Die Transformation

$$X \mapsto Z, \text{ mit } Z = \frac{X - \mu}{\sigma}$$

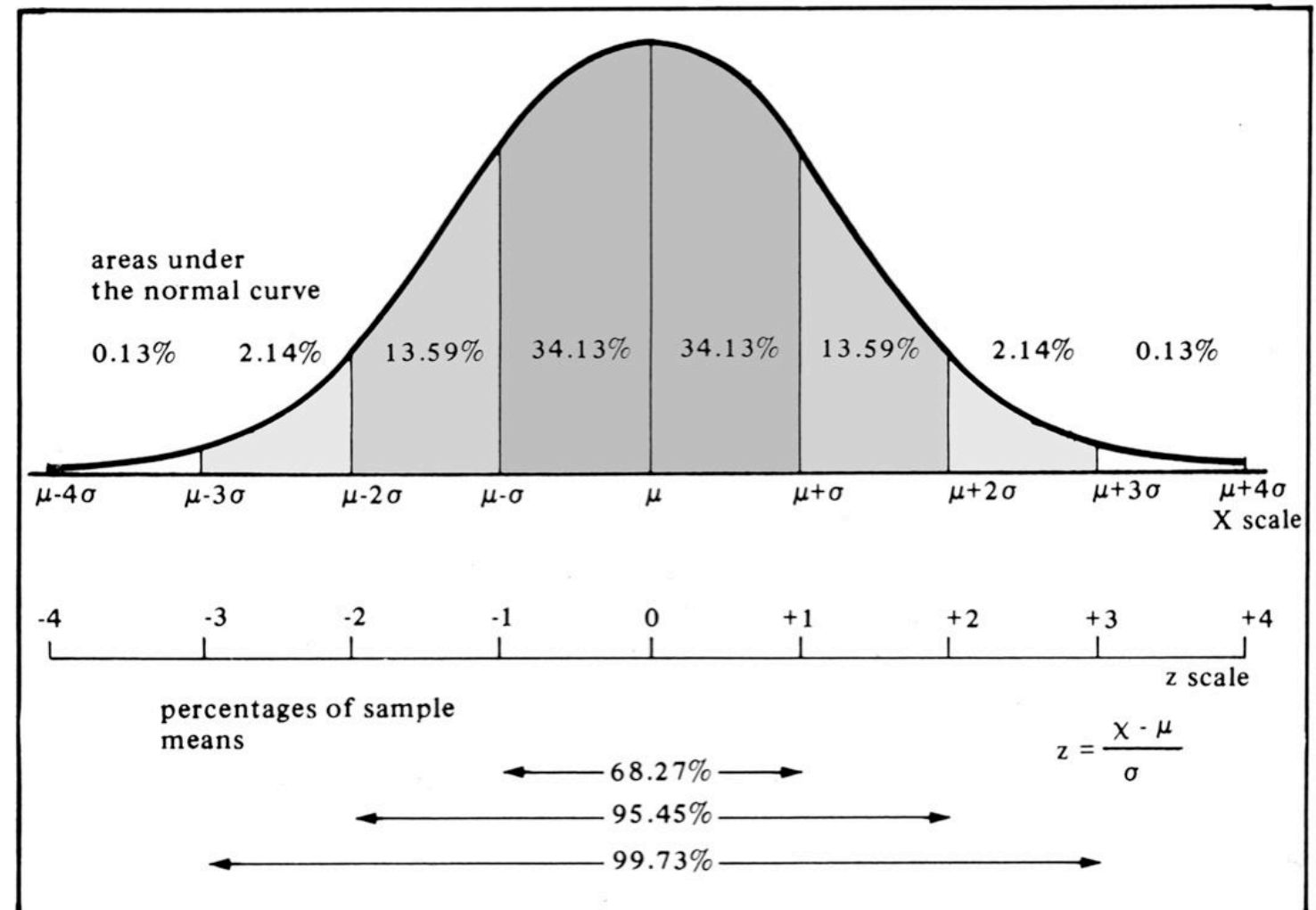
führt zur

Standardnormalverteilung
mit $\mu_Z = 0$ und $\sigma_Z = 1$.

Rückwärts: $X = \mu + \sigma \cdot Z$.

(Die Sampling Distribution

wird zu $Z = \frac{M - \mu}{\sigma/\sqrt{n}}$.)



Vergleichbarkeit durch Standardisierung

Die Transformation

$$X \mapsto Z, \text{ mit } Z = \frac{X - \mu}{\sigma}$$

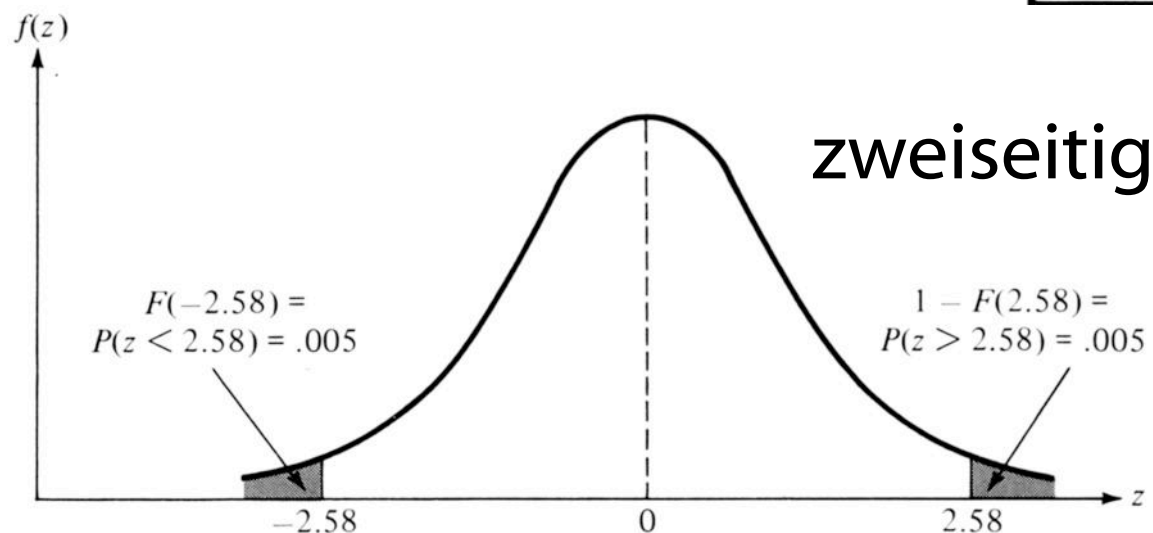
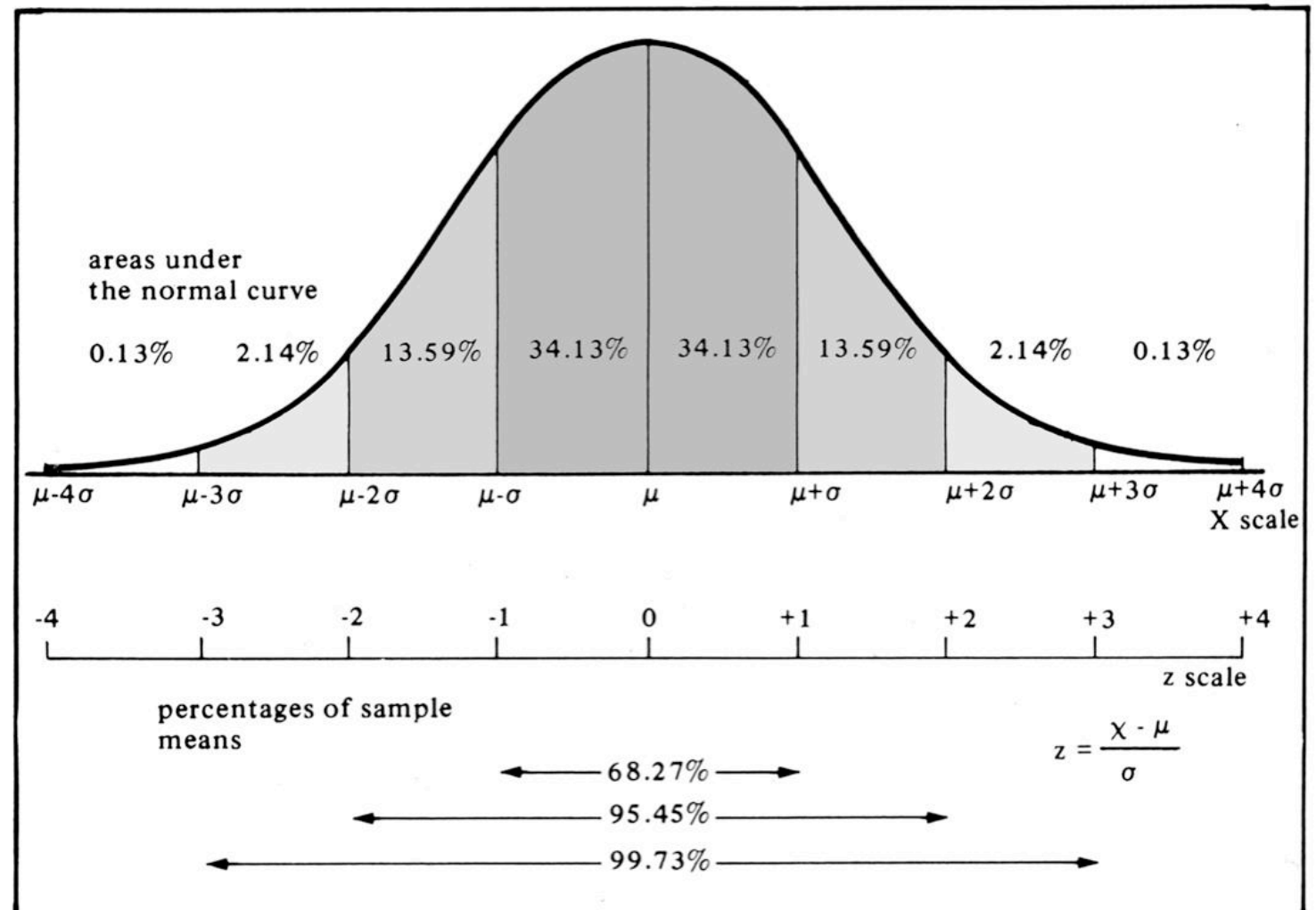
führt zur

Standardnormalverteilung
mit $\mu_Z = 0$ und $\sigma_Z = 1$.

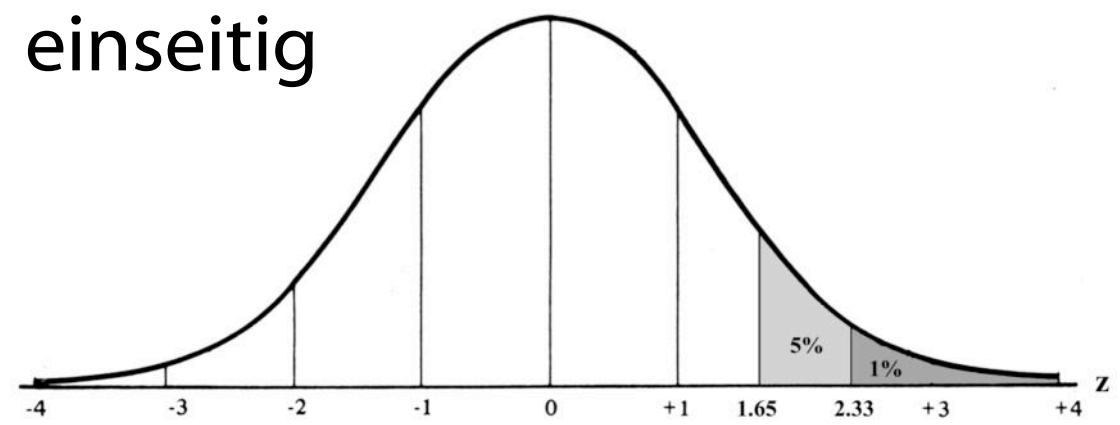
Rückwärts: $X = \mu + \sigma \cdot Z$.

(Die Sampling Distribution

wird zu $Z = \frac{M - \mu}{\sigma/\sqrt{n}}$.)



$$\Pr(|Z| > 2.58) = 0.01$$



$$\Pr(Z \geq c) = 0.01? \Rightarrow c = 2.33$$

Vergleichbarkeit durch Standardisierung

Die Transformation

$X \mapsto Z$, mit $Z = \frac{X - \mu}{\sigma/\sqrt{n}}$

führt zur Standardnormalverteilung mit $\mu_Z = 0$

Rückwärts: $X = \mu + Z \cdot \frac{\sigma}{\sqrt{n}}$

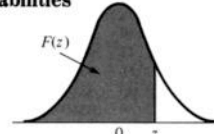
(Die Samplingverteilung)

wird zu $Z = \frac{M - \mu}{\sigma/\sqrt{n}}$

Table I. Cumulative Standard Normal Probabilities

This table gives values of the cumulative distribution function of the standard normal distribution,

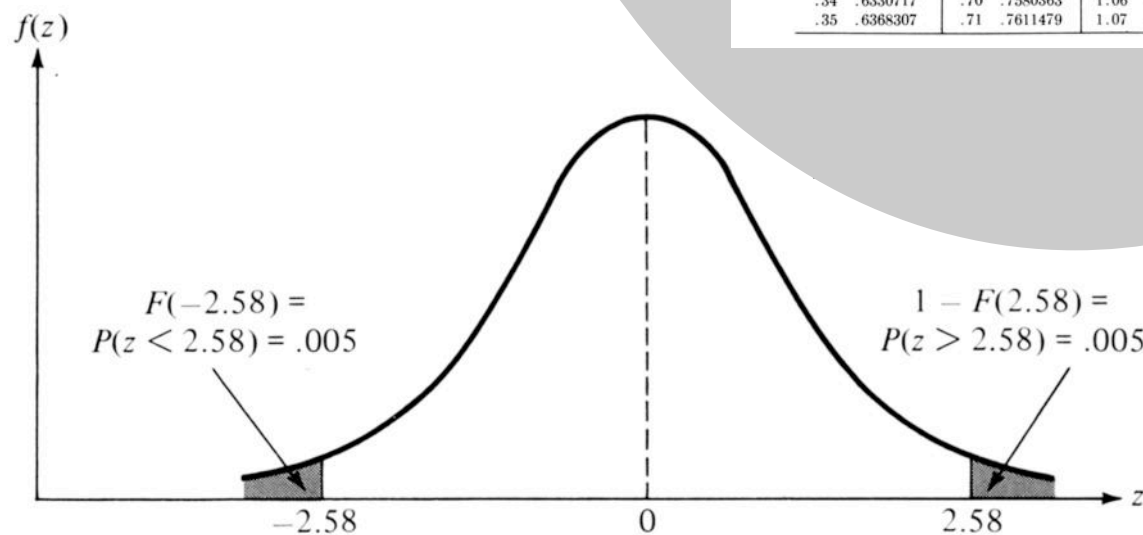
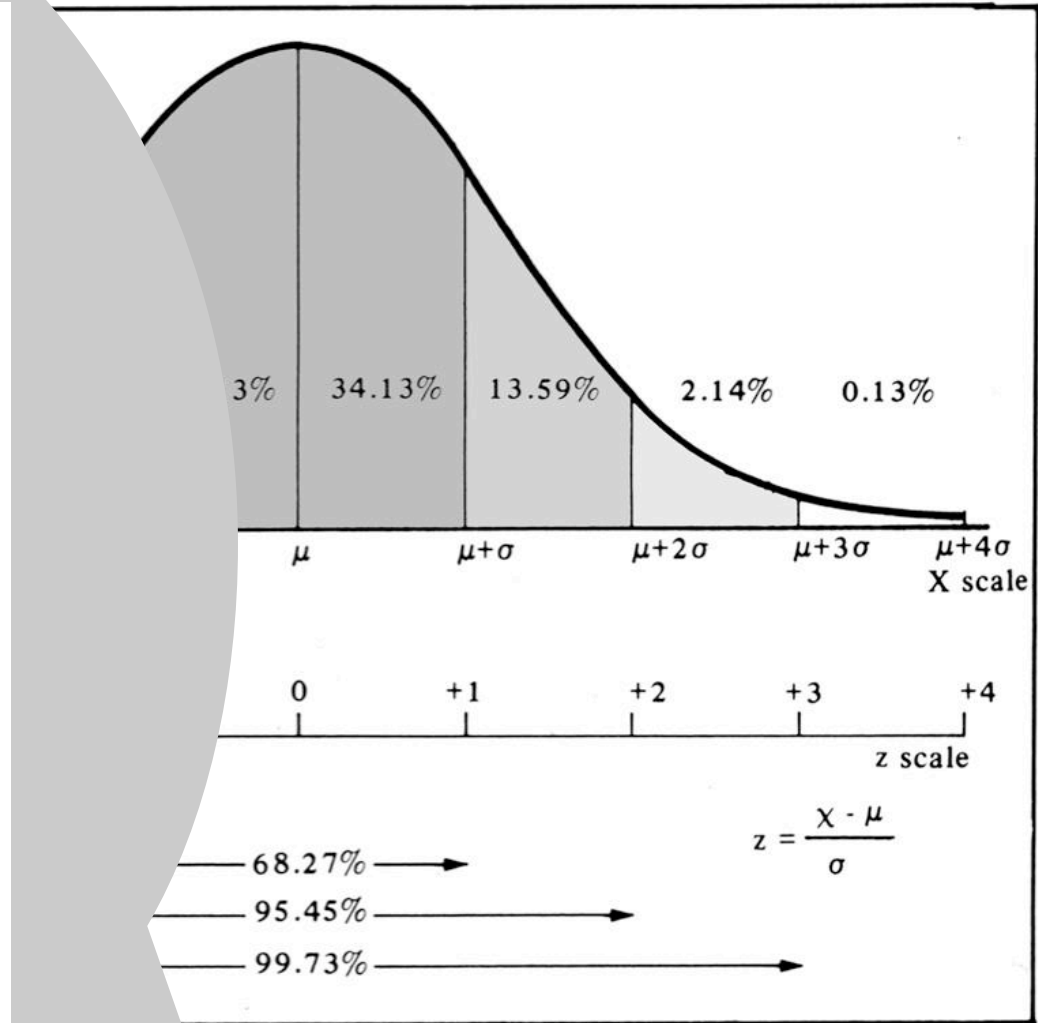
$$F(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz,$$



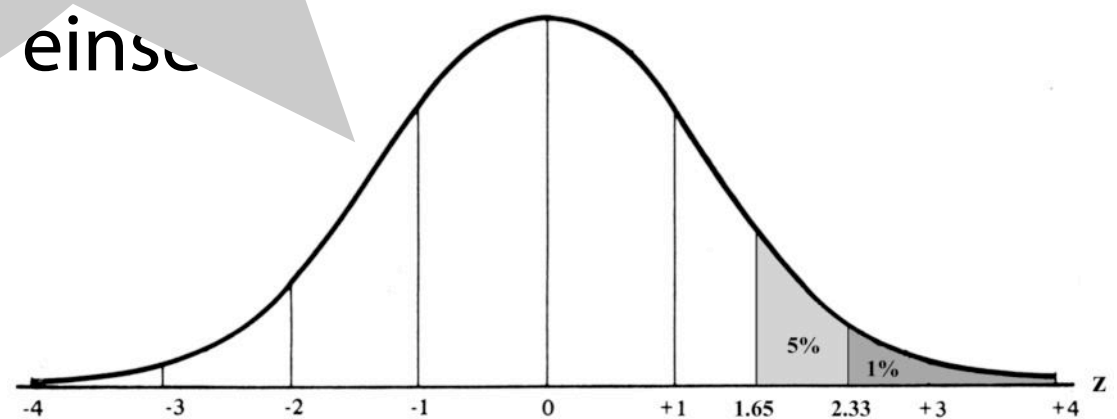
for $z = 0(.01)2.60(.10)3.00(.20)4.00(.50)5.50$. For $z < 0$, take $F(z) = 1 - F(-z)$.

Examples: $F(1.12) = .8686431$,
and $F(-.67) = 1 - F(.67) = 1 - .7485711 = .2514289$.

z	F(z)	z	F(z)	z	F(z)	z	F(z)
.00	.5000000	.36	.6405764	.72	.7642375	1.08	.8599289
.01	.5039894	.37	.6443088	.73	.7673049	1.09	.8621434
.02	.5079783	.38	.6480273	.74	.7703500	1.10	.8643339
.03	.5119665	.39	.6517317	.75	.7733726	1.11	.8665005
.04	.5159534	.40	.6554217	.76	.7763727	1.12	.8686431
.05	.5199388	.41	.6590970	.77	.7793501	1.13	.8707619
.06	.5239222	.42	.6627573	.78	.7823046	1.14	.8728568
.07	.5279032	.43	.6664022	.79	.7852361	1.15	.8749281
.08	.5318814	.44	.6700314	.80	.7881446	1.16	.8769756
.09	.5358564	.45	.6736448	.81	.7910299	1.17	.8789995
.10	.5398278	.46	.6772419	.82	.7938919	1.18	.8809999
.11	.5437953	.47	.6808225	.83	.7967306	1.19	.8829768
.12	.5477584	.48	.6843863	.84	.7995458	1.20	.8849303
.13	.5517168	.49	.6879331	.85	.8023375	1.21	.8868606
.14	.5556700	.50	.6914625	.86	.8051055	1.22	.8887676
.15	.5596177	.51	.6949743	.87	.8078498	1.23	.8906514
.16	.5635595	.52	.6984682	.88	.8105703	1.24	.8925123
.17	.5674949	.53	.7019440	.89	.8132671	1.25	.8943502
.18	.5714237	.54	.7054015	.90	.8159399	1.26	.8961653
.19	.5753454	.55	.7088403	.91	.8185887	1.27	.8979577
.20	.5792597	.56	.7122603	.92	.8212136	1.28	.8997274
.21	.5831662	.57	.7156612	.93	.8238145	1.29	.9014747
.22	.5870604	.58	.7190427	.94	.8263912	1.30	.9031995
.23	.5909541	.59	.7224047	.95	.8289439	1.31	.9049021
.24	.5948349	.60	.7257469	.96	.8314724	1.32	.9065825
.25	.5987063	.61	.7290691	.97	.8339768	1.33	.9082409
.26	.6025681	.62	.7323711	.98	.8364569	1.34	.9098773
.27	.6064199	.63	.7356527	.99	.8389129	1.35	.9114920
.28	.6102612	.64	.7389137	1.00	.8413447	1.36	.9130850
.29	.6140919	.65	.7421539	1.01	.8437524	1.37	.9146565
.30	.6179114	.66	.7453731	1.02	.8461358	1.38	.9162067
.31	.6217195	.67	.7485711	1.03	.8484950	1.39	.9177356
.32	.6255158	.68	.7517478	1.04	.8508300	1.40	.9192433
.33	.6293000	.69	.7549029	1.05	.8531409	1.41	.9207302
.34	.6330717	.70	.7580363	1.06	.8554277	1.42	.9221962
.35	.6368307	.71	.7611479	1.07	.8576903	1.43	.9236415



$$\Pr(|Z| > 2.58) = 0.01$$



$$\Pr(Z \geq c) = 0.01? \Rightarrow c = 2.33$$

Vergleichbarkeit durch Standardisierung

Die Transformation

$$X \mapsto Z, \text{ mit } Z = \frac{X - \mu}{\sigma}$$

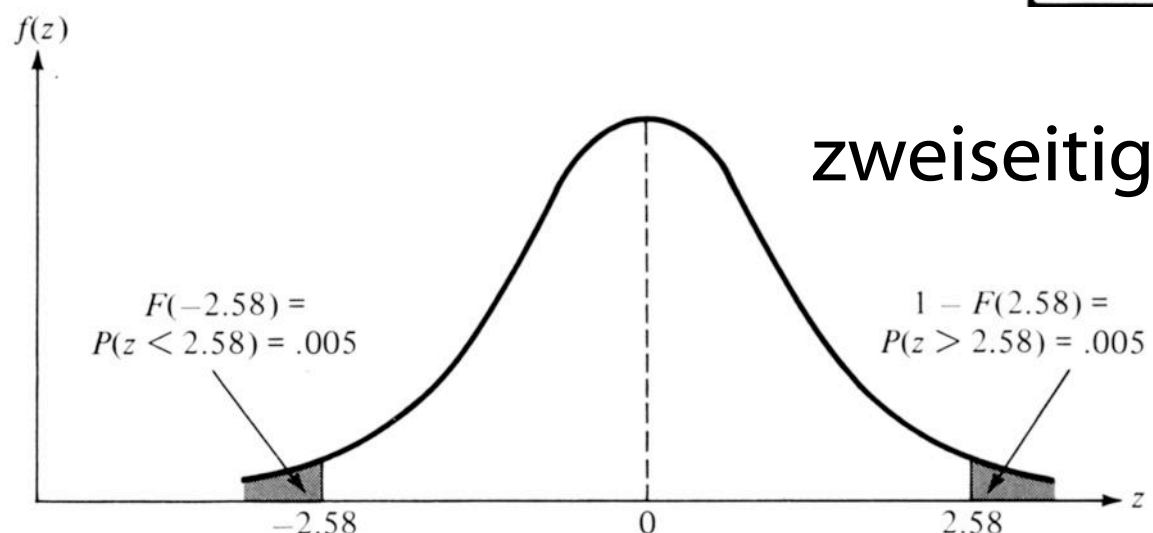
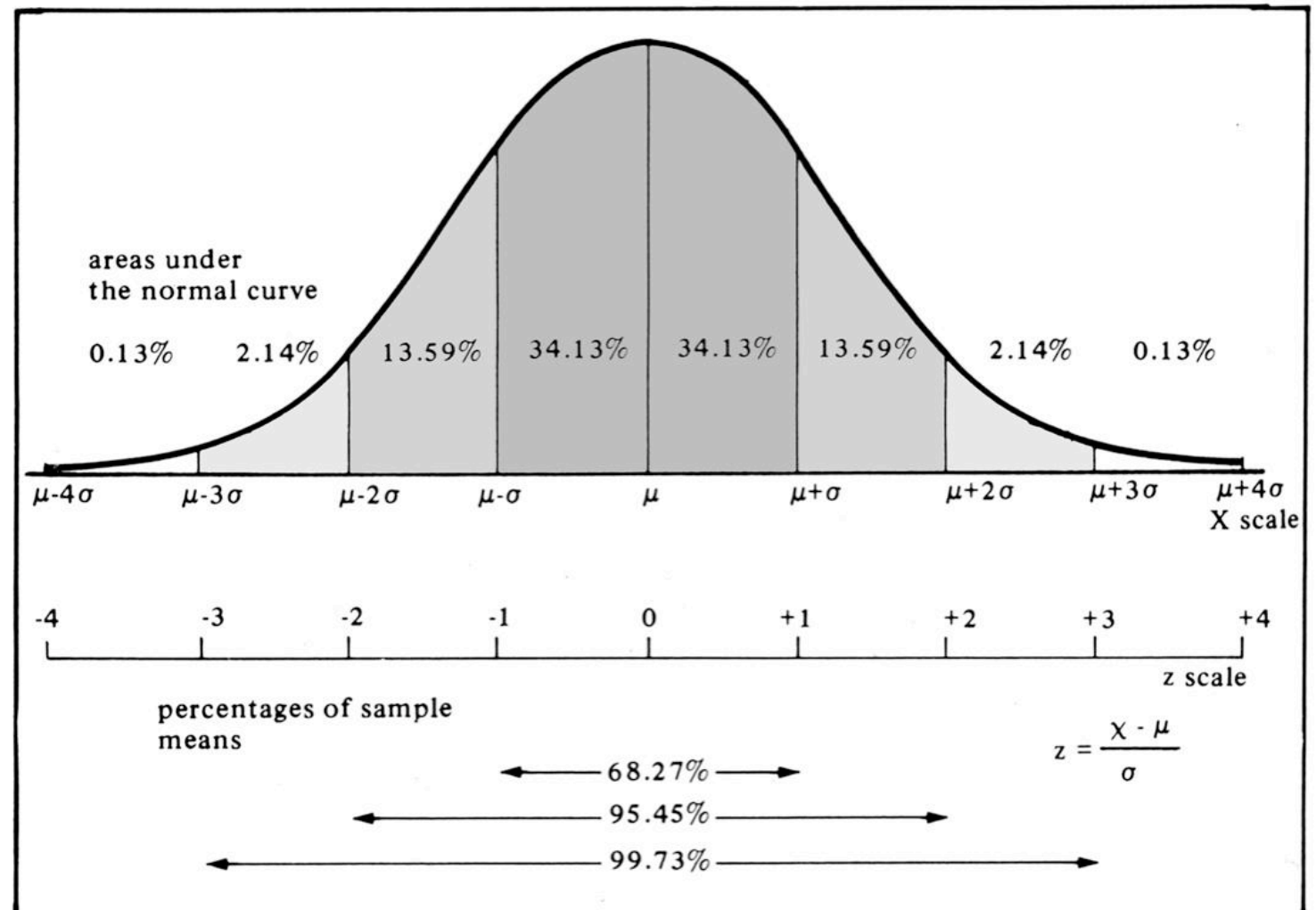
führt zur

Standardnormalverteilung
mit $\mu_Z = 0$ und $\sigma_Z = 1$.

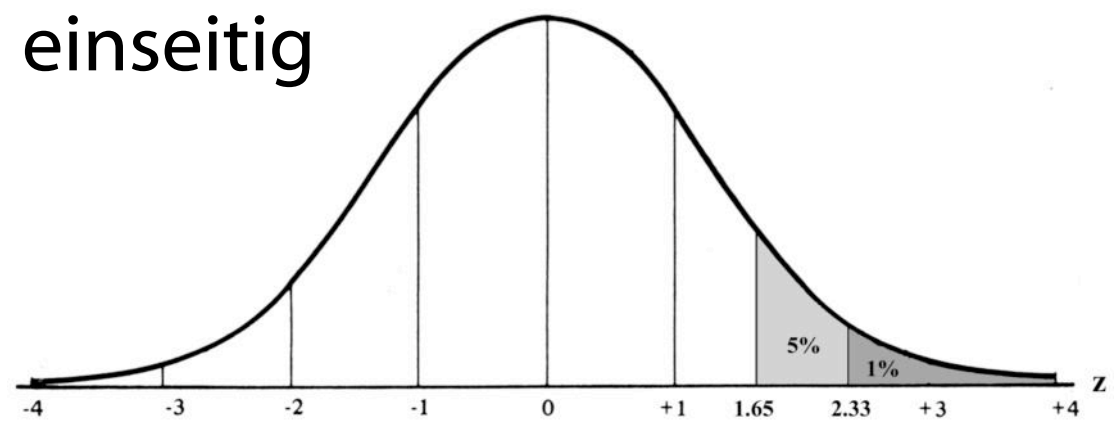
Rückwärts: $X = \mu + \sigma \cdot Z$.

(Die Sampling Distribution

wird zu $Z = \frac{M - \mu}{\sigma/\sqrt{n}}$.)



$$\Pr(|Z| > 2.58) = 0.01$$



$$\Pr(Z \geq c) = 0.01? \Rightarrow c = 2.33$$

Drei zu unterscheidende Verteilungen

Wahrscheinlichkeitsverteilung in der Population:

theoretische Verteilung mit den Parametern μ und σ , die i. A. nicht bekannt sind und mittels einer Stichprobe geschätzt werden müssen. Praktisch besteht umgekehrt die Population aus allen Sub/Objekten, die die gleiche Chance haben, in der Stichprobe aufzutreten.

Häufigkeitsverteilung in einer Stichprobe:

beschreibende Statistik mit Kennwerten $m = \frac{1}{n} \cdot \sum_{i=1}^n x_i$ und $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - m)^2$.

Stichprobenkennwerteverteilung (Sampling Distribution):

theoretische Verteilung, die die Verbindung zwischen Stichprobe und Population angibt und für Inferenzen die wichtigste Verteilung ist. Sie beschreibt mittels der Zufallsvariablen M , mit welcher Wahrscheinlichkeit (Dichte) der Mittelwert einer Stichprobe unter allen möglichen Stichproben der Größe n aus der gleichen Population vorkommt.

$$\mu_M = E(M) = E\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \mu \quad \sigma_M^2 = V(M) = V\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \frac{\sigma^2}{n}$$

Etwas pingeliger, aber immer noch nicht ganz sauber

Basis-Experiment: Kaufe zufällig eine Packung UGLI Tiefkühlerbsen.

Elementarereignisse: $\mathcal{U} = \text{Population}$ aller UGLI Erbsenpackungen

Messung: $X : \mathcal{U} \rightarrow \mathbb{R}$, mit $X(U_j) = \text{Anzahl der Erbsen in } U_j$

Populationsverteilung: Verteilung von X mit Parametern μ und σ

Sekundär-Experiment: Ziehe eine Zufallsstichprobe der Größe n

Elementarereignisse: $\mathcal{S}^{(n)} \subseteq \mathcal{U}^n = \underbrace{\mathcal{U} \times \mathcal{U} \times \dots \times \mathcal{U}}_{n\text{-mal}}$, $\mathcal{S}^{(n)}$ ist gleichverteilt. (random sampling)

Messung: $X^{(n)} = X_1 \times \dots \times X_n : \mathcal{S}^{(n)} \rightarrow \mathbb{R}^n$, mit $X^{(n)}(\langle U_1, \dots, U_n \rangle) = \langle x_1, \dots, x_n \rangle$, alle X_i sind unabhängig.

Mittelwertzufallsvariable: $M^{(n)} : \mathcal{S}^{(n)} \rightarrow \mathbb{R}$, mit $M^{(n)}(S_j) = m_j = \frac{1}{n} \cdot \sum_{i=1}^n x_i$.

Verteilung von $M^{(n)}$: $\{\langle m, \Pr(M^{(n)} = m) \rangle\}$ oder $\{\langle m, f(m) \rangle\}$, mit

$$\mu_{M^{(n)}} = E(M^{(n)}) = E\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \mu \quad \text{und} \quad \sigma_{M^{(n)}} = \frac{\sigma}{\sqrt{n}}.$$

Grenzfälle:

$n = 1 \Rightarrow$ Verteilung von $M^{(n)} = \text{Populationsverteilung}$ und $\mu_{M^{(1)}} = \mu, \sigma_{M^{(1)}} = \sigma$

$n = |\mathcal{U}| \Rightarrow$ Verteilung von $M^{(n)}$ besteht nur aus 1 Element und $\mu_{M^{(1)}} = \mu, \sigma_{M^{(1)}} = 0$

die i. A. nicht
werden müssen.
Sub/Objekten,
treten.

$$d \quad s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - m)^2.$$

n):

Stichprobe und
erteilung ist.

her Wahrschein-
allen möglichen
vorkommt.

$$\mu_M = E(M) = E\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \mu \quad \sigma_M^2 = V(M) = V\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \frac{\sigma^2}{n}$$

Drei zu unterscheidende Verteilungen

Wahrscheinlichkeitsverteilung in der Population:

theoretische Verteilung mit den Parametern μ und σ , die i. A. nicht bekannt sind und mittels einer Stichprobe geschätzt werden müssen. Praktisch besteht umgekehrt die Population aus allen Sub/Objekten, die die gleiche Chance haben, in der Stichprobe aufzutreten.

Häufigkeitsverteilung in einer Stichprobe:

beschreibende Statistik mit Kennwerten $m = \frac{1}{n} \cdot \sum_{i=1}^n x_i$ und $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - m)^2$.

Stichprobenkennwerteverteilung (Sampling Distribution):

theoretische Verteilung, die die Verbindung zwischen Stichprobe und Population angibt und für Inferenzen die wichtigste Verteilung ist. Sie beschreibt mittels der Zufallsvariablen M , mit welcher Wahrscheinlichkeit (Dichte) der Mittelwert einer Stichprobe unter allen möglichen Stichproben der Größe n aus der gleichen Population vorkommt.

$$\mu_M = E(M) = E\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \mu \quad \sigma_M^2 = V(M) = V\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \frac{\sigma^2}{n}$$

Zentraler Grenzwertsatz

Für eine beliebige Populationsverteilung mit Mittelwert μ und Varianz σ entspricht die Verteilung der Mittelwerte aller Stichproben der Größe $n \gg 1$ ungefähr einer Normalverteilung mit

$$\mu_M = \mu \quad \text{und} \quad \sigma_M = \frac{\sigma}{\sqrt{n}}.$$

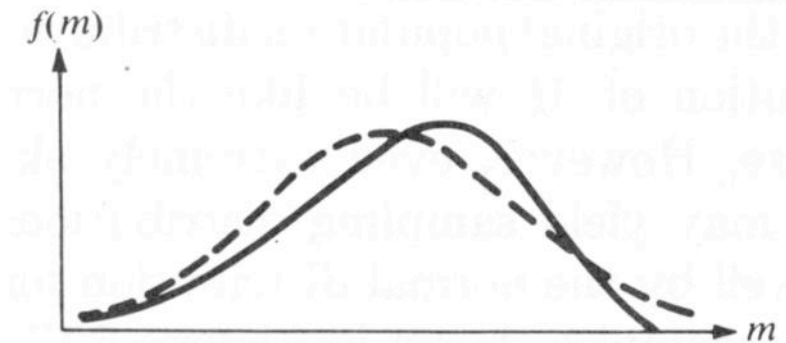
(Die Sampling Distribution wird also

zu $Z = \frac{M - \mu}{\sigma/\sqrt{n}}$ standardisiert.)

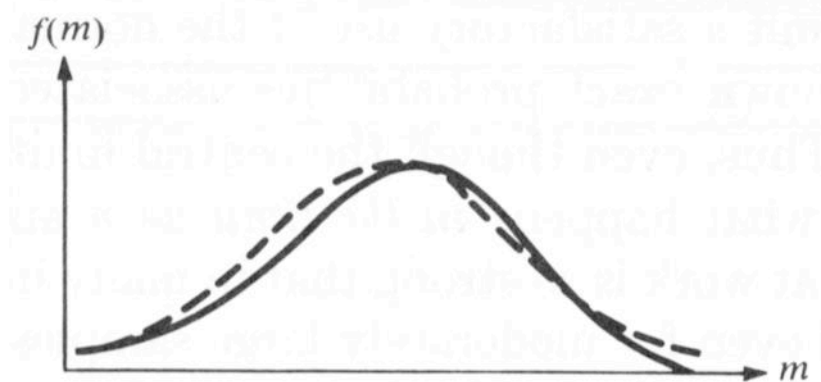
$n \geq 30$ gilt normalerweise als ausreichend.

Wenn σ nicht bekannt ist, kann es mittels s , das T-verteilt ist, geschätzt werden.

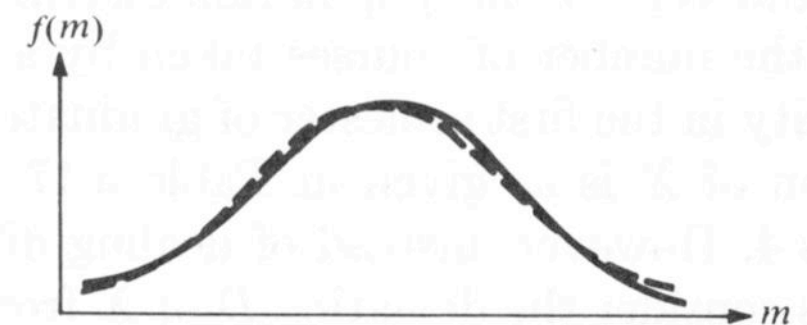
σ_M heißt auch Standard-Fehler.



A Sampling Distribution of M for $n = 2$ with a Comparable Normal Distribution



A Sampling Distribution of M for $n = 4$ with a Comparable Normal Distribution



A Sampling Distribution of M for $n = 10$ with a Comparable Normal Distribution

Zentraler Grenzwertsatz

Für eine beliebige Populationsverteilung mit Mittelwert μ und Varianz σ entspricht die Verteilung der Mittelwerte aller Stichproben der Größe $n \gg 1$ ungefähr einer Normalverteilung mit

$$\mu_M = \mu \text{ und } \sigma_M = \frac{\sigma}{\sqrt{n}}$$

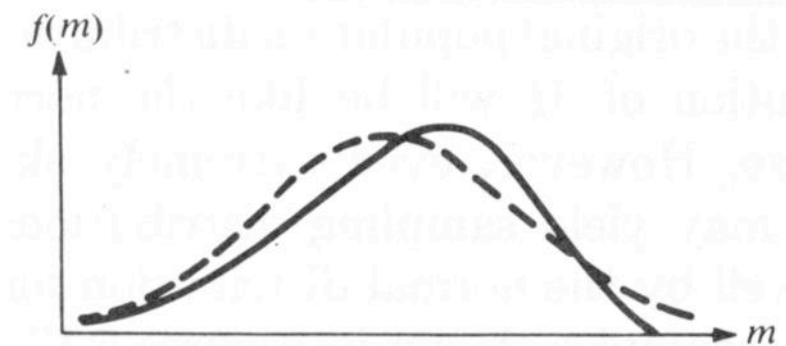
(Die Sampling

zu $Z = \frac{M - \mu}{\sigma/\sqrt{n}}$

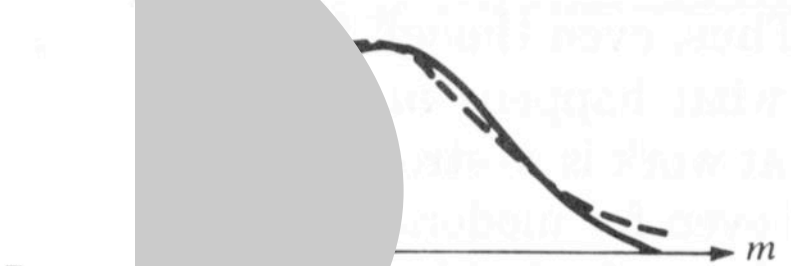
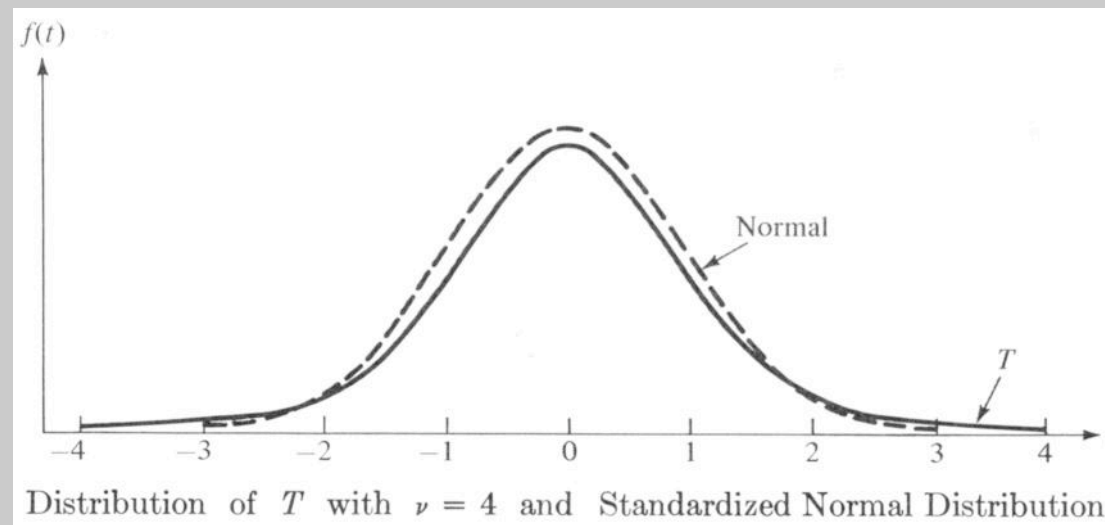
$n \geq 30$ gilt nicht

Wenn σ nicht bekannt ist, wird s , das T-verteilt ist, geschätzt werden.

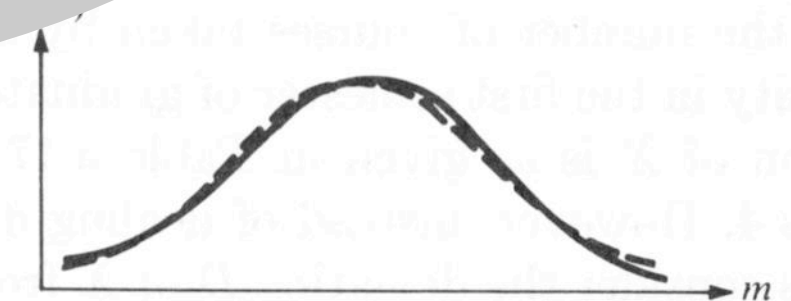
σ_M heißt auch Standard-Fehler.



A Sampling Distribution of M for $n = 2$ with a Comparable Normal Distribution



A Sampling Distribution of M for $n = 4$ with a Comparable Normal Distribution



A Sampling Distribution of M for $n = 10$ with a Comparable Normal Distribution

Zentraler Grenzwertsatz

Für eine beliebige Populationsverteilung mit Mittelwert μ und Varianz σ entspricht die Verteilung der Mittelwerte aller Stichproben der Größe $n \gg 1$ ungefähr einer Normalverteilung mit

$$\mu_M = \mu \quad \text{und} \quad \sigma_M = \frac{\sigma}{\sqrt{n}}.$$

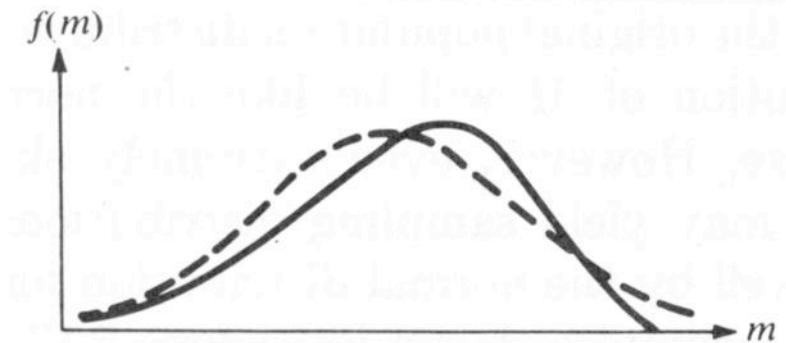
(Die Sampling Distribution wird also

zu $Z = \frac{M - \mu}{\sigma/\sqrt{n}}$ standardisiert.)

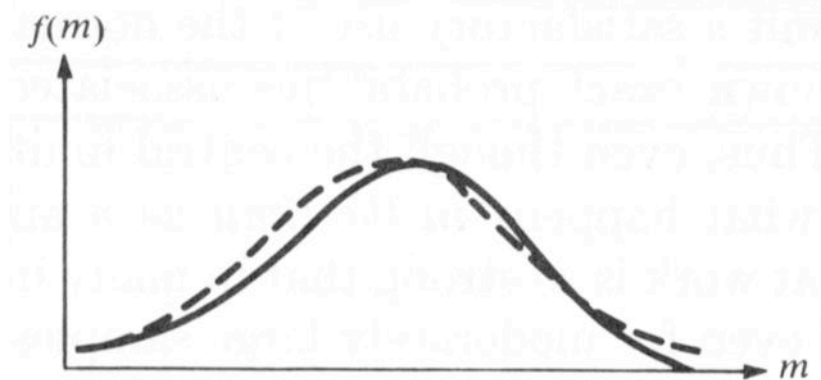
$n \geq 30$ gilt normalerweise als ausreichend.

Wenn σ nicht bekannt ist, kann es mittels s , das T-verteilt ist, geschätzt werden.

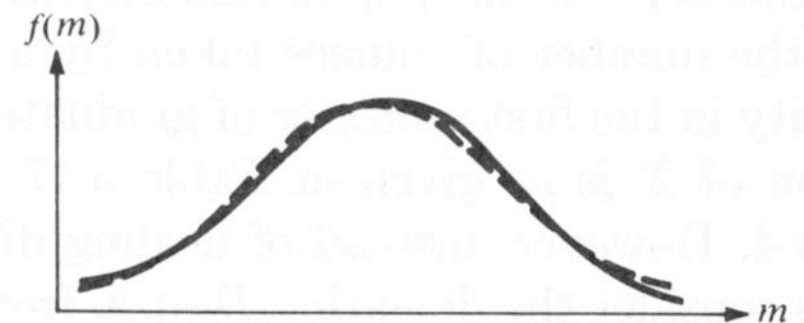
σ_M heißt auch Standard-Fehler.



A Sampling Distribution of M for $n = 2$ with a Comparable Normal Distribution



A Sampling Distribution of M for $n = 4$ with a Comparable Normal Distribution



A Sampling Distribution of M for $n = 10$ with a Comparable Normal Distribution

Konfidenzintervalle für Schätzungen

Aus $\Pr\left(-1.96 \leq Z = \frac{M - \mu}{\sigma_M} \leq +1.96\right) = 0.95$ folgt durch Umrechnen

$\Pr(M - 1.96 \cdot \sigma_M \leq \mu \leq M + 1.96 \cdot \sigma_M) = 0.95$, ein 95%-Konfidenzintervall.

Allgemeiner ergibt sich für $\Pr(-a \leq Z \leq +a) = (1 - \alpha)$ das

$(1 - \alpha) \cdot 100\%$ Konfidenzintervall durch $CI = \left(M - a \cdot \frac{\sigma}{\sqrt{n}}, M + a \cdot \frac{\sigma}{\sqrt{n}}\right)$.

Frequentist-Interpretation: 95% aller Zufallsstichproben enthalten μ in ihrem 95%-Konfidenzintervall, μ ist fest.

›Subjektive‹ Interpretation: μ wird als Zufallsvariable aufgefaßt, die mit 95%-iger Wahrscheinlichkeit im Konfidenzintervall der gemessenen Stichprobe liegt.

Durch Vergrößern der Stichprobe kann man μ beliebig präzise schätzen und für eine gesuchte Genauigkeit die Stichprobengröße ausrechnen, z.B.

$\Pr(|M - \mu| \leq 0.1 \cdot \sigma) = 0.99 \Rightarrow M \pm 2.58 \cdot \sigma / \sqrt{n}$, also $0.1 \cdot \sigma = 2.58 \cdot \sigma / \sqrt{n} \Rightarrow n = 666$.

Konfidenzintervalle für Schätzungen

Aus $\Pr\left(-1.96 \leq Z = \frac{M - \mu}{\sigma_M} \leq +1.96\right) = 0.95$ folgt durch Umrechnen

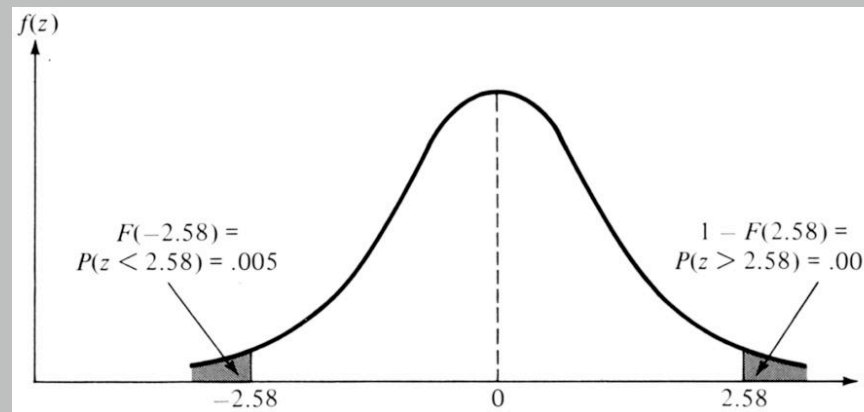
$\Pr(M - 1.96 \cdot \sigma_M \leq \mu \leq M + 1.96 \cdot \sigma_M) = 0.95$, ein 95%-Konfidenzintervall.

Allgemeiner ergibt sich für $\Pr(-a \leq Z \leq +a) = (1 - \alpha)$ das

$(1 - \alpha) \cdot 100\%$ Konfidenzintervall durch $CI = \left(M - a \cdot \frac{\sigma}{\sqrt{n}}, M + a \cdot \frac{\sigma}{\sqrt{n}}\right)$.

Frequentist-Interpretation: 95% aller Konfidenzintervalle enthalten μ , μ ist fest.

›Subjektive‹ Interpretation: 95% aller Stichproben umfassen μ , μ ist unbekannt, die mit Konfidenzintervall ermittelte Stichprobe liegt.



Durch Vergrößern der Stichprobe kann man μ präziser schätzen und für eine gesuchte Genauigkeit die Stichprobengröße ausrechnen, z.B.

$\Pr(|M - \mu| \leq 0.1 \cdot \sigma) = 0.99 \Rightarrow M \pm 2.58 \cdot \sigma / \sqrt{n}$, also $0.1 \cdot \sigma = 2.58 \cdot \sigma / \sqrt{n} \Rightarrow n = 666$.

Konfidenzintervalle für Schätzungen

Aus $\Pr\left(-1.96 \leq Z = \frac{M - \mu}{\sigma_M} \leq +1.96\right) = 0.95$ folgt durch Umrechnen

$\Pr(M - 1.96 \cdot \sigma_M \leq \mu \leq M + 1.96 \cdot \sigma_M) = 0.95$, ein 95%-Konfidenzintervall.

Allgemeiner ergibt sich für $\Pr(-a \leq Z \leq +a) = (1 - \alpha)$ das

$(1 - \alpha) \cdot 100\%$ Konfidenzintervall durch $CI = \left(M - a \cdot \frac{\sigma}{\sqrt{n}}, M + a \cdot \frac{\sigma}{\sqrt{n}}\right)$.

Frequentist-Interpretation: 95% aller Zufallsstichproben enthalten μ in ihrem 95%-Konfidenzintervall, μ ist fest.

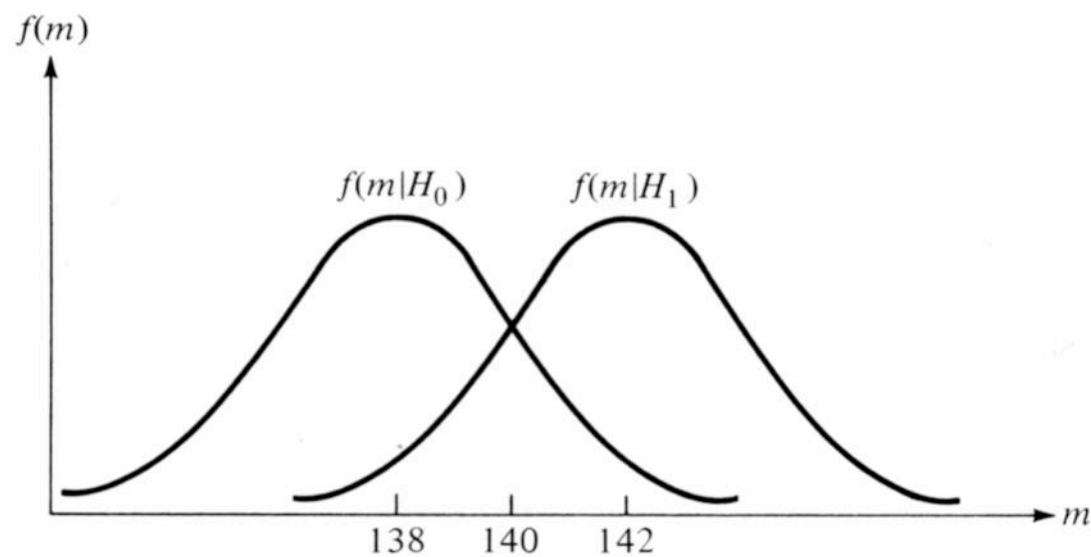
›Subjektive‹ Interpretation: μ wird als Zufallsvariable aufgefaßt, die mit 95%-iger Wahrscheinlichkeit im Konfidenzintervall der gemessenen Stichprobe liegt.

Durch Vergrößern der Stichprobe kann man μ beliebig präzise schätzen und für eine gesuchte Genauigkeit die Stichprobengröße ausrechnen, z.B.

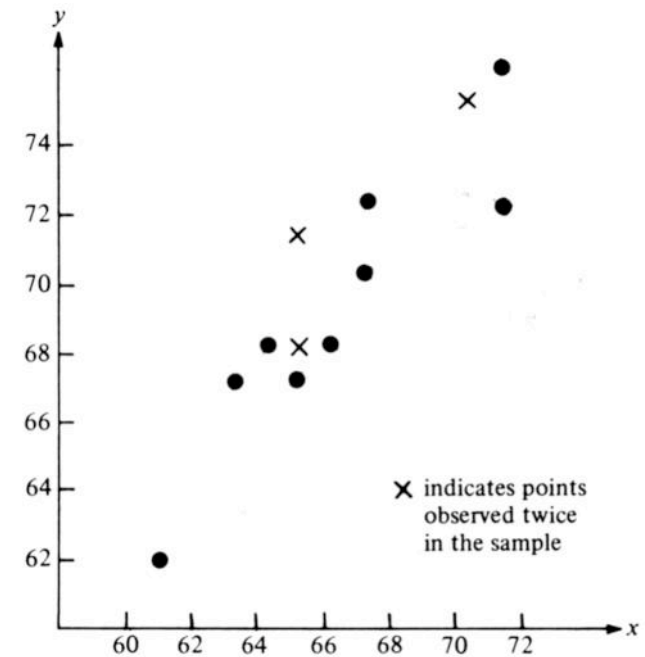
$\Pr(|M - \mu| \leq 0.1 \cdot \sigma) = 0.99 \Rightarrow M \pm 2.58 \cdot \sigma / \sqrt{n}$, also $0.1 \cdot \sigma = 2.58 \cdot \sigma / \sqrt{n} \Rightarrow n = 666$.

Tests

Typische Fragestellungen:



Sampling Distributions of M Under H_0 and H_1



A Scatter Diagram of Heights of Wives and Husbands

Mittelwertvergleich/einordnung

Korrelationen

Wir rechnen mit Sampling Distributions, weil wir unsere beobachtete Stichprobe in eine Stichprobenverteilung einordnen wollen.

Es muß zwischen zwei alternativen Hypothesen entschieden werden:

Nullhypothese H_0 : z.B. $\mu = 138, \rho = 0, RR = 1$

Alternativhypothese H_1 : z.B. $\mu = 142, \rho \neq 0, RR = 1.5$

Tests

Unterschiedliche Hypothesenarten

spezifische Hypothesen: $H_0 : \mu = \mu_0$
 $H_1 : \mu = \mu_1$

unspezifische Hypothesen

einseitig: $H_0 : \mu = \mu_0$ oder $H_0 : \mu \leq \mu_0$
 $H_1 : \mu > \mu_0$ $H_1 : \mu > \mu_0$

zweiseitig: $H_0 : \mu = \mu_0$
 $H_1 : \mu \neq \mu_0$

Fehler (bei der Bestimmung eines Mittelwertes)

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu = \mu_1 \quad (\mu_1 > \mu_0)$$

Alles sei normalverteilt (M sowieso).
Stichprobe liefert Mittelwert m .

Gehört die Stichprobe eher zur Population mit μ_0 oder einer mit μ_1 ?

α -Fehler: H_0 wird fälschlich verworfen, $\alpha = \Pr(\text{verwerfe } H_0 \mid H_0 \text{ ist wahr})$

β -Fehler: H_0 wird fälschlich akzeptiert, $\beta = \Pr(\text{akzeptiere } H_0 \mid H_0 \text{ ist falsch})$.

Errors in Hypothesis-Testing Situation

		TRUE SITUATION	
		H_0	H_1
DECISION	H_0	correct	error
	H_1	error	correct

Probabilities of Erroneous and Correct Decisions

		TRUE SITUATION	
		H_0	H_1
DECISION	Accept H_0	$1 - \alpha$	β
	Reject H_0	α	$1 - \beta$

Signifikanzniveau α wird (vor dem Test) festgelegt, abhängig von Kosten.

Power (Stärke, Trennschärfe):

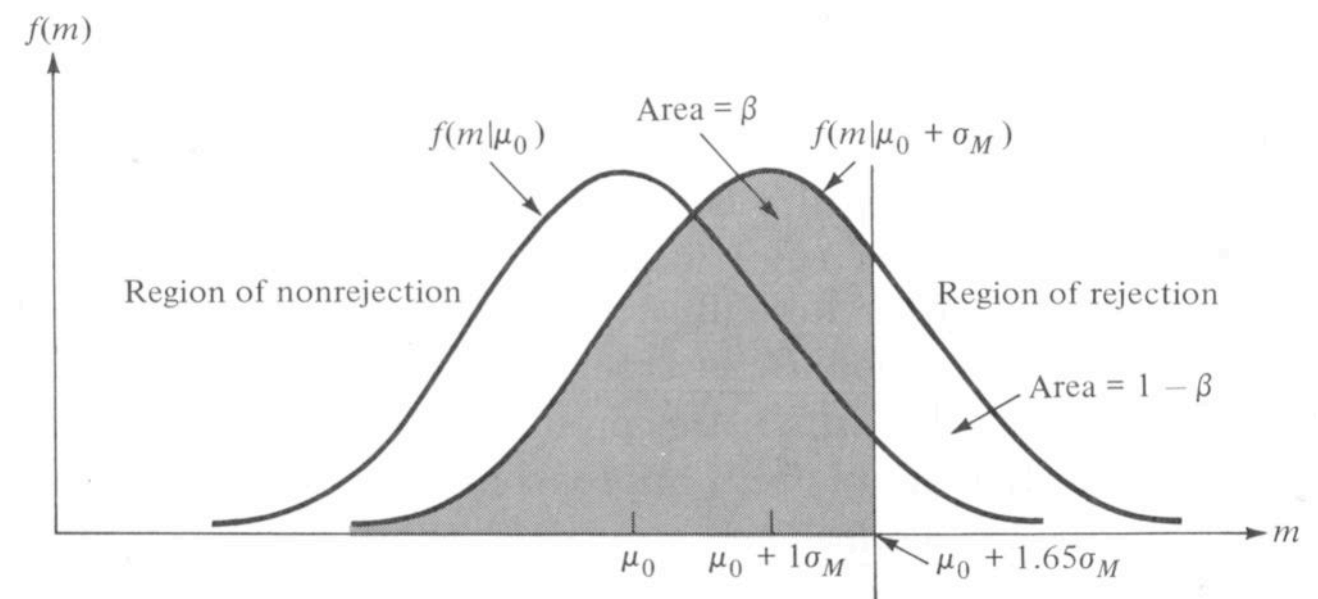
$$\Pr(\text{verwerfe } H_0 \mid H_1 \text{ ist wahr}) = 1 - \Pr(\text{akzeptiere } H_0 \mid H_1 \text{ ist wahr}) = 1 - \beta$$

Power und α -Fehler, »region of rejection«

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu = \mu_1 \stackrel{z.B.}{=} \mu_0 + \sigma_M$$

Durch α wird eine »region of rejection« festgelegt; für jeden Wert $m \geq c_\alpha$ der Stichprobe wird die H_0 verworfen.



Power as Area Under a Sampling Distribution

Der kritische Wert c_α mit $\Pr(M \geq c_\alpha | H_0) = \alpha = \Pr\left(Z \geq \frac{c_\alpha - \mu_0}{\sigma_M}\right)$

läßt sich aus der Tabelle der Standardnormalverteilung gewinnen; für $\alpha = 0.05$ ergibt sich $c_{0.05} = \mu_0 + 1.65 \cdot \sigma_M$.

Bei einer spezifischen Alternativhypothese läßt sich β und $1-\beta$ ausrechnen:

$$\beta = \Pr(M < c_\alpha | H_1) = \Pr\left(Z < \frac{c_\alpha - \mu_1}{\sigma_M}\right).$$

Power und α -Fehler, »region of rejection«

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu = \mu_1 \stackrel{z.B.}{=} \mu_0 + \sigma_M$$

Durch α wird eine »region of rejection« festgelegt.

Werte

von

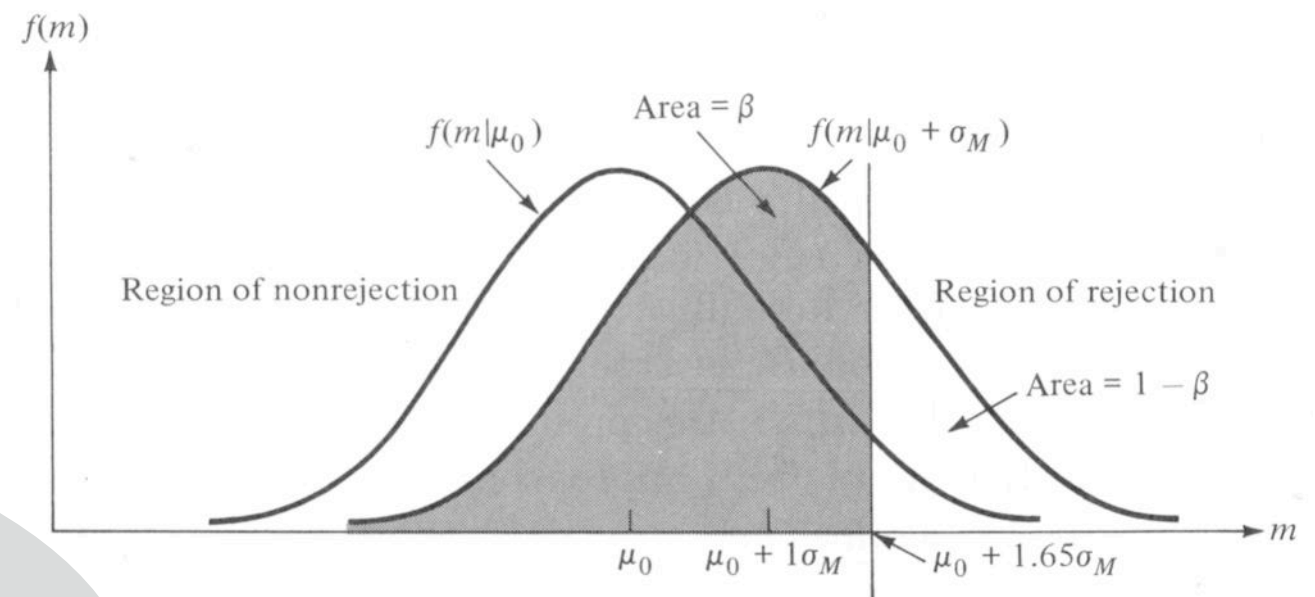
der

läßt sich aus einer standardisierten Normalverteilung gewinnen;

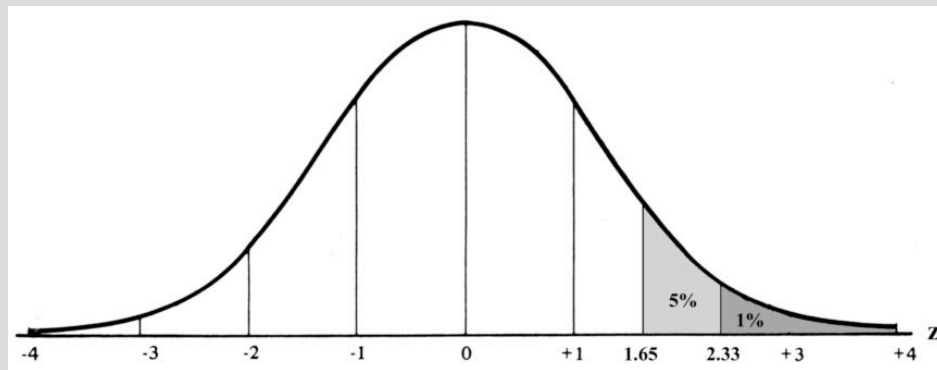
für $\alpha = 0.05$ ergibt sich $c_{0.05} = \mu_0 + 1.65 \cdot \sigma_M$.

Bei einer spezifischen Alternativhypothese läßt sich β und $1-\beta$ ausrechnen:

$$\beta = \Pr(M < c_\alpha | H_1) = \Pr\left(Z < \frac{c_\alpha - \mu_1}{\sigma_M}\right).$$



as Area Under a Sampling Distribution

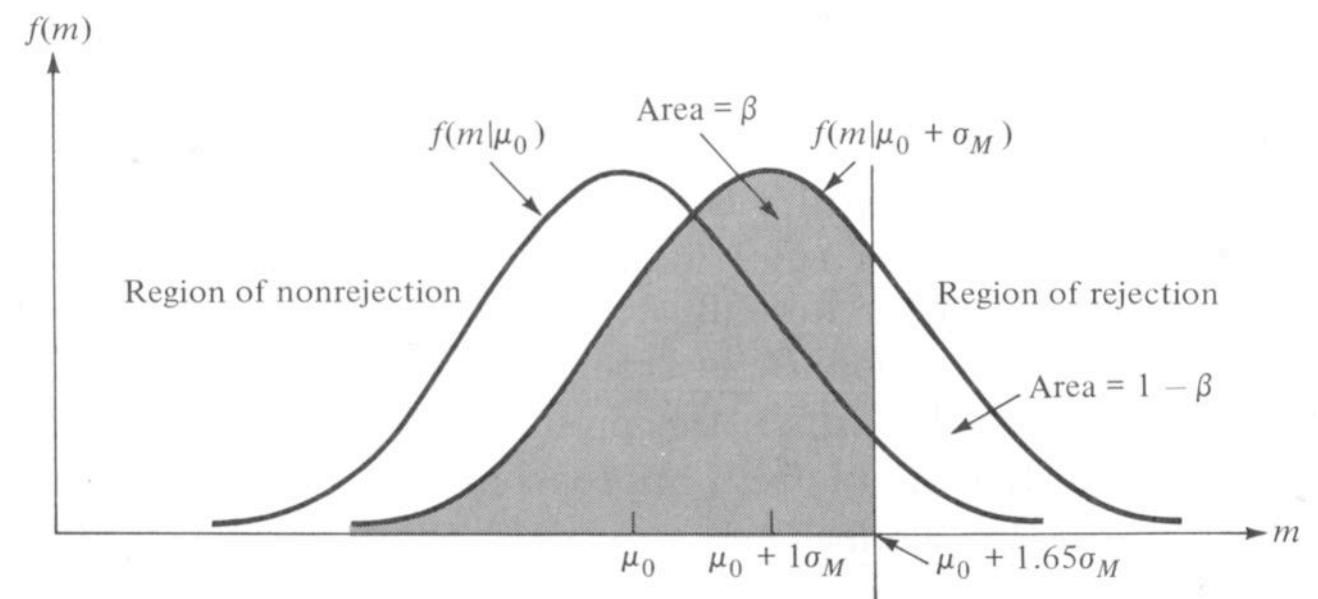


Power und α -Fehler, »region of rejection«

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu = \mu_1 \stackrel{z.B.}{=} \mu_0 + \sigma_M$$

Durch α wird eine »region of rejection« festgelegt; für jeden Wert $m \geq c_\alpha$ der Stichprobe wird die H_0 verworfen.



Power as Area Under a Sampling Distribution

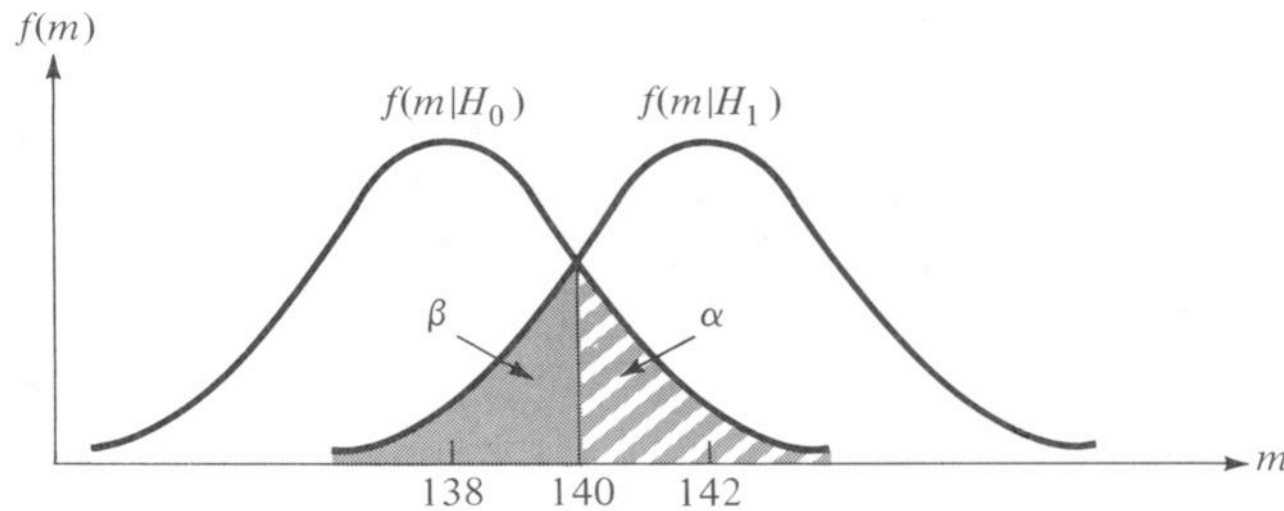
Der kritische Wert c_α mit $\Pr(M \geq c_\alpha | H_0) = \alpha = \Pr\left(Z \geq \frac{c_\alpha - \mu_0}{\sigma_M}\right)$

läßt sich aus der Tabelle der Standardnormalverteilung gewinnen; für $\alpha = 0.05$ ergibt sich $c_{0.05} = \mu_0 + 1.65 \cdot \sigma_M$.

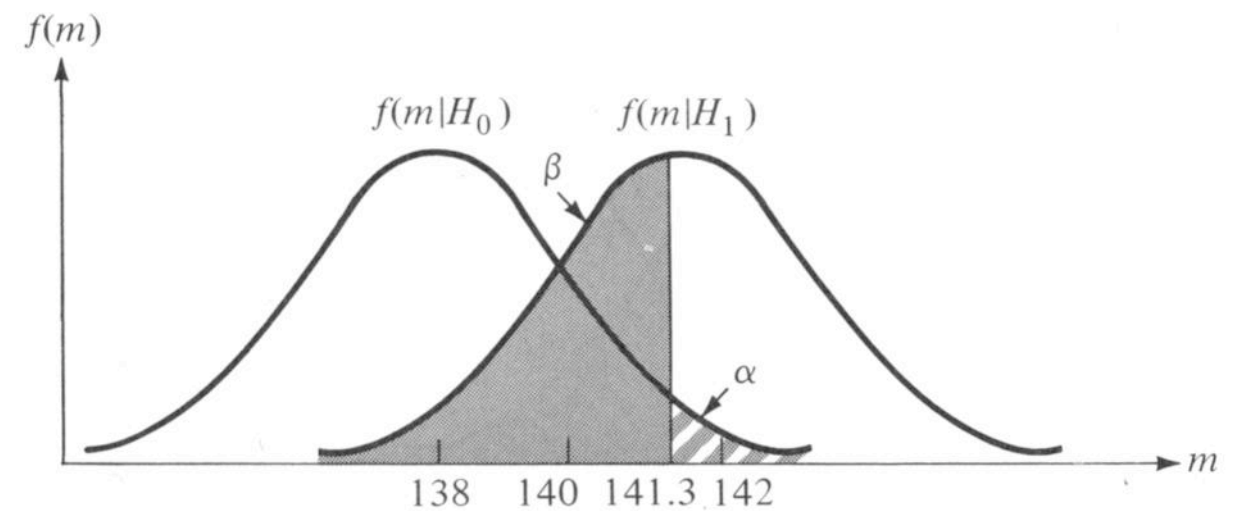
Bei einer spezifischen Alternativhypothese läßt sich β und $1 - \beta$ ausrechnen:

$$\beta = \Pr(M < c_\alpha | H_1) = \Pr\left(Z < \frac{c_\alpha - \mu_1}{\sigma_M}\right).$$

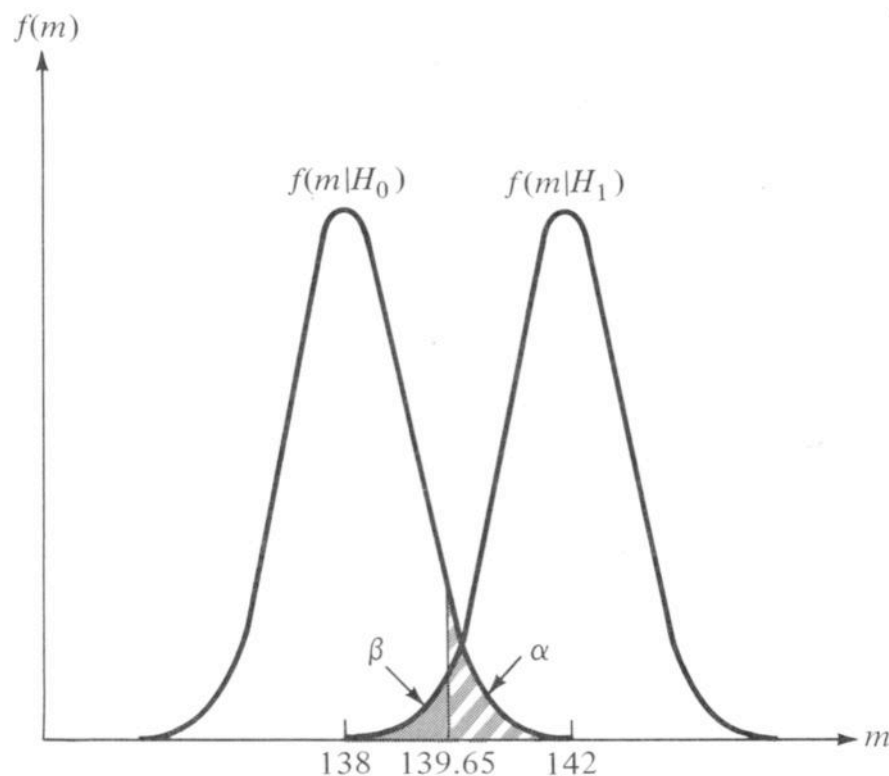
Fehlerverhalten bei spezifischen Hypothesen



Probabilities of Error as Areas Under Sampling Distributions for $c = 140$, $n = 100$



Probabilities of Error as Areas Under Sampling Distributions for $c = 141.3$, $n = 100$



Probabilities of Error as Areas Under Sampling Distributions for $c = 139.65$, $n = 400$

α und β verhalten sich gegenläufig; zu einem kleineren α gehört ein größerer kritischer Wert c_α und somit ein größeres β und umgekehrt. Bei gleichem α kann $1-\beta$ durch Vergrößern der Stichprobe verbessert werden, ebenso wird die Trennschärfe durch einen größeren Abstand besser.

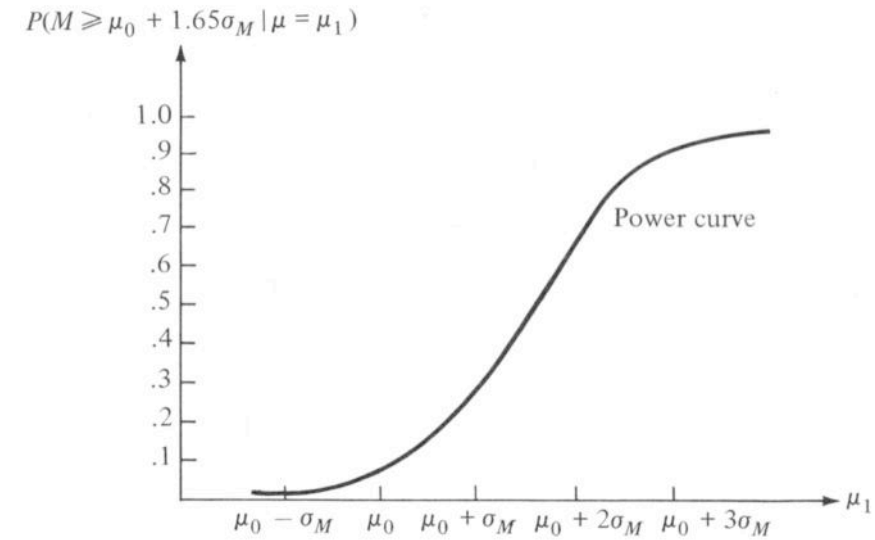
Unspezifische Hypothesen

$$H_0 : \mu_0 = 100$$

$$H_1 : \mu_1 > 100$$

Die »region of rejection« ist allein durch α bestimmt.

Bei einer unspezifischen H_1 kann β nicht berechnet werden, weil $\Pr(\text{akzeptiere } H_0 \mid \mu > 100)$ keine bedingte Wahrscheinlichkeit ist. Man betrachtet dann Power-Kurven.

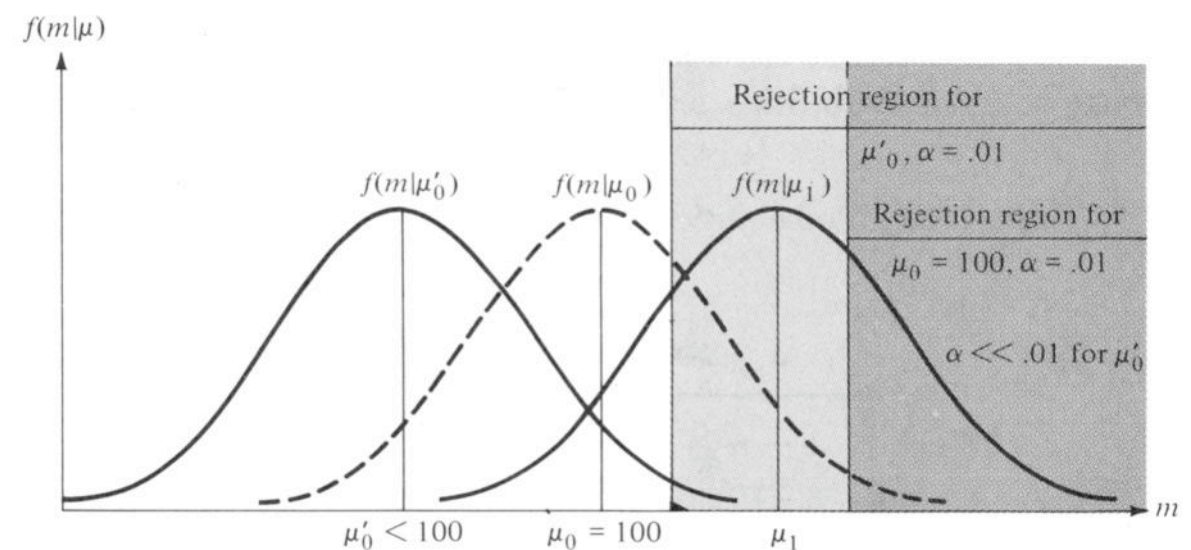


A Power Curve

$$H_0 : \mu_0 \leq 100$$

$$H_1 : \mu_1 > 100$$

Wenn $\mu_0 = 100$ mit α verworfen werden kann, dann auch jedes $\mu_0 < 100$, und die Power ist für 100 kleiner als für jedes kleinere μ_0 .



Rejection Regions for Different Means Under H_0

Fürbitte als relatives Risiko

Studie mit ca. 1800 Patienten, die eine Bypass-Operation bekamen (Coronary Artery Bypass Graft); Gebete von 3 christl. Vereinen.

Eingeteilt in 3 Gruppen:

1. möglicherweise, und gebetet
2. möglicherweise, aber nicht
3. versichert, daß gebetet.



In der Medizin übliches Maß: Relative Risk $RR = \frac{P_{\text{Probanden}}}{P_{\text{Kontrollgruppe}}}$.

Die Sampling Distribution von $\log(RR)$ ist annähernd normalverteilt.

Konfidenzintervall: $\log-CI_{95\%} = (\log(RR) - 1.96 \cdot \sigma_M, \log(RR) + 1.96 \cdot \sigma_M)$,
woraus sich das eigentliche (asymmetrische) Konfidenzintervall ergibt:

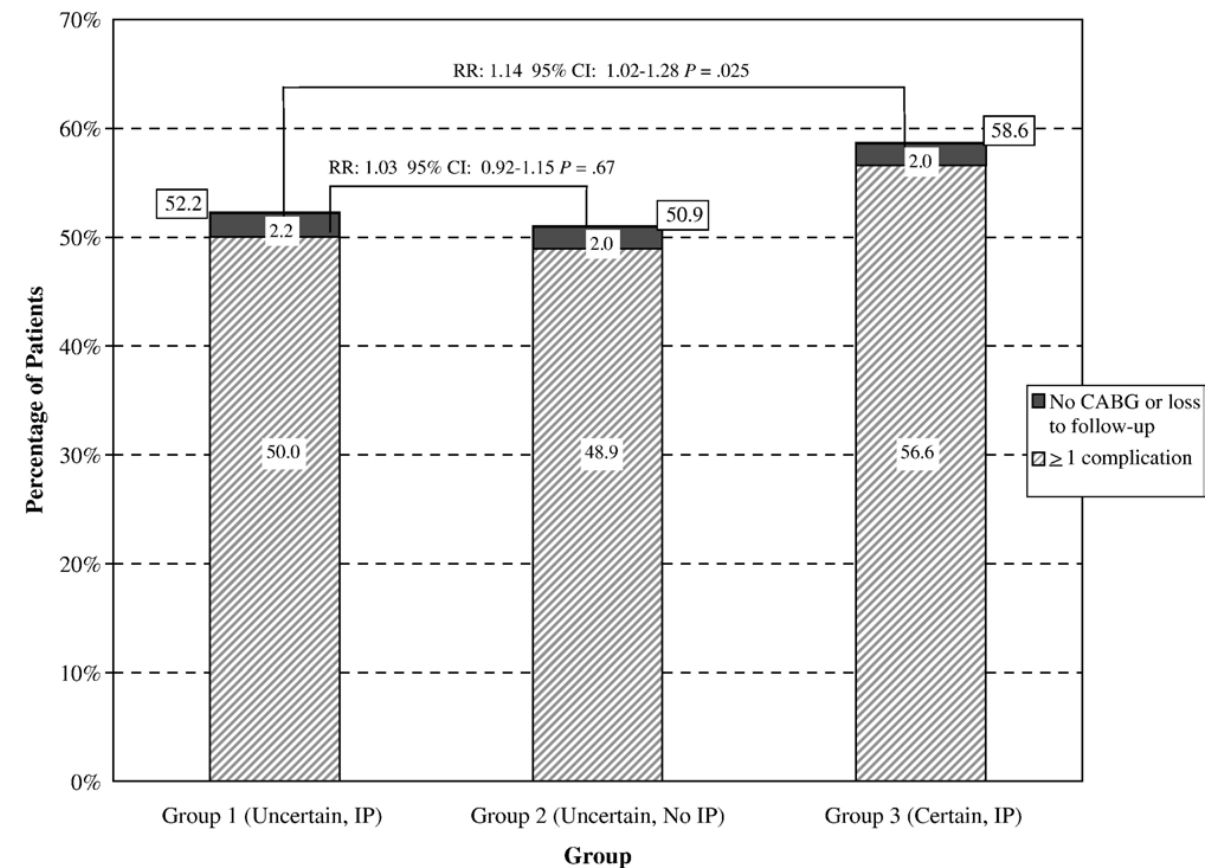
$$CI_{95\%} = (10^{\log(RR) - 1.96 \cdot \sigma_M}, 10^{\log(RR) + 1.96 \cdot \sigma_M})$$

Fürbitte als relatives Risiko

Studie mit ca. 1800 Patienten, die eine Bypass-Operation bekamen (Coronary Artery Bypass Graft); Gebete von 3 christl. Vereinen.

Eingeteilt in 3 Gruppen:

1. möglicherweise, und gebetet
2. möglicherweise, aber nicht
3. versichert, daß gebetet.



In der Medizin übliches Maß: Relative Risk $RR = \frac{P_{\text{Probanden}}}{P_{\text{Kontrollgruppe}}}$.

Die Sampling Distribution von $\log(RR)$ ist annähernd normalverteilt.

Konfidenzintervall: $\log-CI_{95\%} = (\log(RR) - 1.96 \cdot \sigma_M, \log(RR) + 1.96 \cdot \sigma_M)$,
woraus sich das eigentliche (asymmetrische) Konfidenzintervall ergibt:

$$CI_{95\%} = (10^{\log(RR) - 1.96 \cdot \sigma_M}, 10^{\log(RR) + 1.96 \cdot \sigma_M})$$

Fürbitte als relatives Risiko

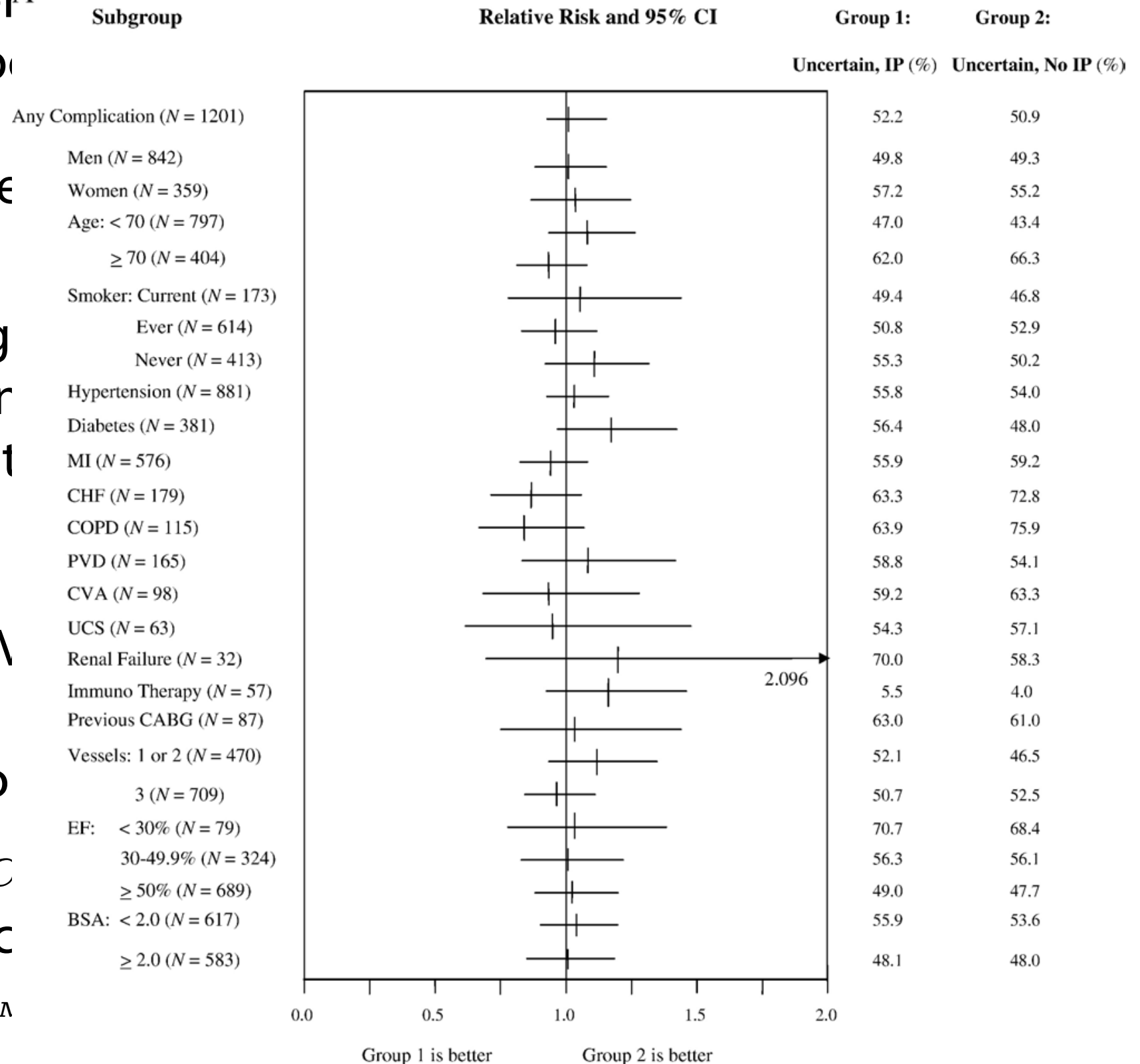
Studie mit ca. 1800 Patienten
 eine Bypass-Operation bei
 (Coronary Artery Bypass
 Gebete von 3 christl. Vere
 Eingeteilt in 3 Gruppen:
 1. möglicherweise, und g
 2. möglicherweise, aber r
 3. versichert, daß gebetet

In der Medizin übliches N

Die Sampling Distributio

Konfidenzintervall: $\log-C$
 woraus sich das eigentlic

$$CI_{95\%} = \left(10^{\log(RR) - 1.96 \cdot \sigma_{\log}}\right)$$

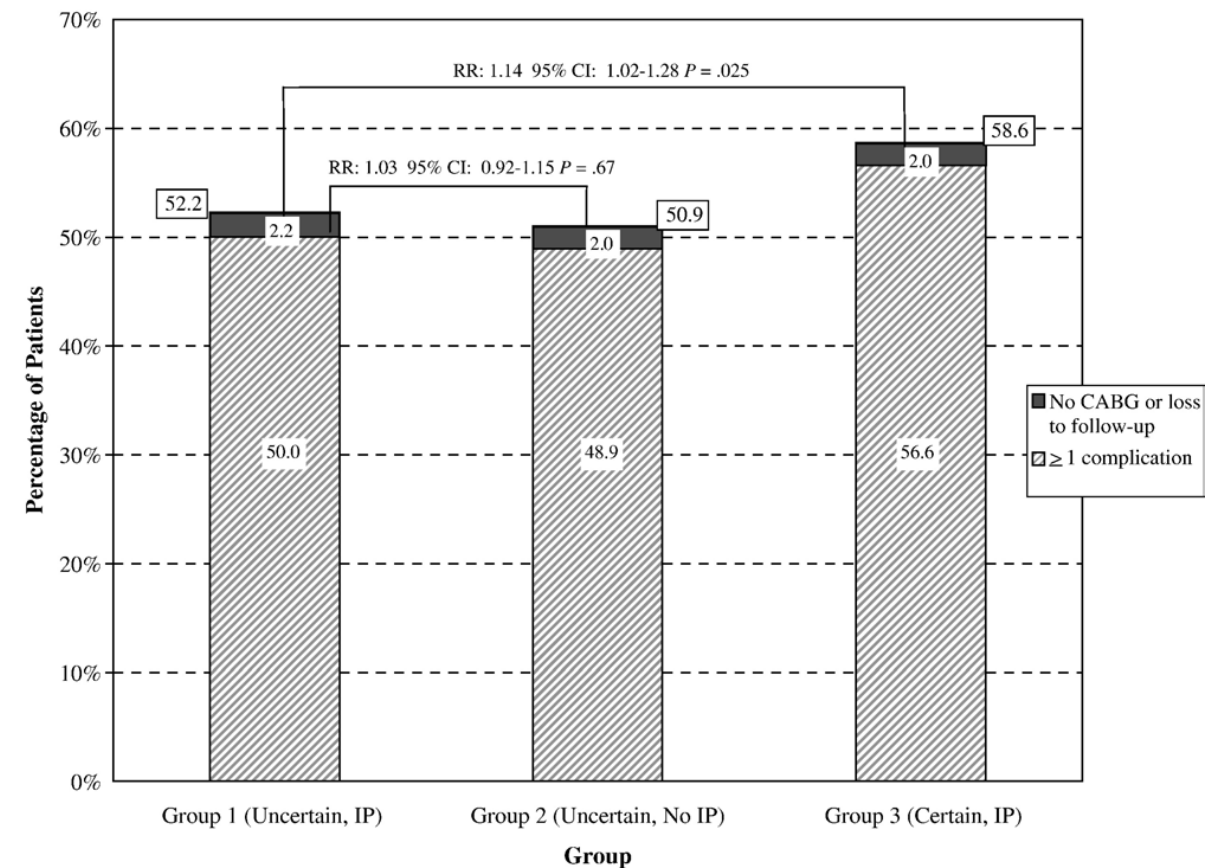


Fürbitte als relatives Risiko

Studie mit ca. 1800 Patienten, die eine Bypass-Operation bekamen (Coronary Artery Bypass Graft); Gebete von 3 christl. Vereinen.

Eingeteilt in 3 Gruppen:

1. möglicherweise, und gebetet
2. möglicherweise, aber nicht
3. versichert, daß gebetet.



In der Medizin übliches Maß: Relative Risk $RR = \frac{P_{\text{Probanden}}}{P_{\text{Kontrollgruppe}}}$.

Die Sampling Distribution von $\log(RR)$ ist annähernd normalverteilt.

Konfidenzintervall: $\log-CI_{95\%} = (\log(RR) - 1.96 \cdot \sigma_M, \log(RR) + 1.96 \cdot \sigma_M)$,
woraus sich das eigentliche (asymmetrische) Konfidenzintervall ergibt:

$$CI_{95\%} = (10^{\log(RR) - 1.96 \cdot \sigma_M}, 10^{\log(RR) + 1.96 \cdot \sigma_M})$$

Likelihood Ratios und Bayessche Inferenz

posterior odds ratio = likelihood ratio · prior odds ratio ($\Omega'' = \Omega \cdot \Omega'$)

Berechnet sich zu:
$$\frac{\Pr(H_0 | RR)}{\Pr(H_1 | RR)} = \frac{\Pr(RR | H_0)}{\Pr(RR | H_1)} \cdot \frac{\Pr(H_0)}{\Pr(H_1)} .$$

Test, ob sich durch Gebete mindestens 10% Komplikationen vermeiden lassen, unter unterschiedlichen subjektiven apriori-Annahmen:

$$H_0 : RR = 1$$

$$H_1 : RR \leq 0.9$$

Thomas:
$$\begin{array}{l} \Pr(H_0) = 0.9 \\ \Pr(H_1) = 0.1 \end{array} \Rightarrow \Omega'' = \frac{0.7}{0.008} \cdot \frac{0.9}{0.1} = 787.5 = \frac{0.9987}{0.0013}$$

Peter:
$$\begin{array}{l} \Pr(H_0) = 0.1 \\ \Pr(H_1) = 0.9 \end{array} \Rightarrow \Omega'' = \frac{0.7}{0.008} \cdot \frac{0.1}{0.9} = 9.72 = \frac{0.906}{0.094}$$

Oh weh

Table III. Independent predictors of complications after CABG (Society of Thoracic Surgeons Adult Cardiac Surgery Database Definitions)

Variable associated with having a complication	Relative risk	95% CI
Certain of receiving intercessory prayer	1.27	1.03-1.57
Older age	1.04	1.03-1.05
Prior myocardial infarction	1.45	1.18-1.77
History of chronic obstructive pulmonary disease	1.61	1.14-2.27
History of congestive heart failure	1.67	1.23-2.26
History of hypertension	1.39	1.11-1.74

Model includes 1684 patients (93% of total) with a complete data and no missing values. *IP*, Intercessory prayer.

»We have no clear explanation for the observed excess of complications in patients who were certain that intercessors would pray for them ... the excess may be a chance finding.«
Benson et al., 2006

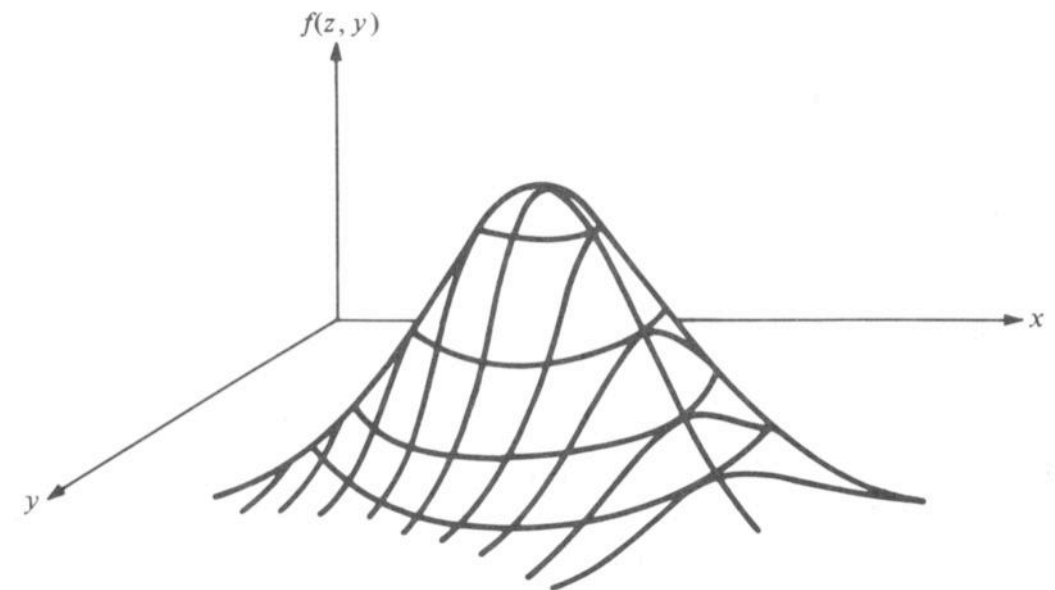
»While presenting these results clearly and noting them in discussion, the investigators take an almost casual approach toward any explanation, stating only that it "may have been a chance finding." It is rather unusual to attribute a statistically significant result in the primary end point of a prospective, multicenter randomized trial to "chance." In fact, such attribution is antithetical to the very definition of what α error and statistical certainty imply: that the worse outcomes are almost certainly related to the therapy and not the play of chance. If the results had shown benefit rather than harm, would we have read the investigators' conclusion that this effect "may have been a chance finding," with absolutely no other comments, insight, or even speculation?«
Editorial American Heart Journal

Korrelationen (ganz ähnlich)

Vor.: Zufallsvariablen X und Y
sind bivariat normalverteilt.

Ein theoretisches Maß der Korrelation
– des Kovariierens – von X und Y ist:

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{E[(X - \mu_X) \cdot (Y - \mu_Y)]}{\sigma_X \cdot \sigma_Y}$$



A Bivariate Normal Density Function

Der aus der Stichprobe berechnete Korrelationskoeffizient

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - m_X) \cdot (y_i - m_Y)}{n \cdot s_X \cdot s_Y}$$

ist ein Schätzwert für ρ_{XY} .

Hypothesen: $H_0 : \rho_{XY} = 0$
 $H_1 : \rho_{XY} \neq 0$

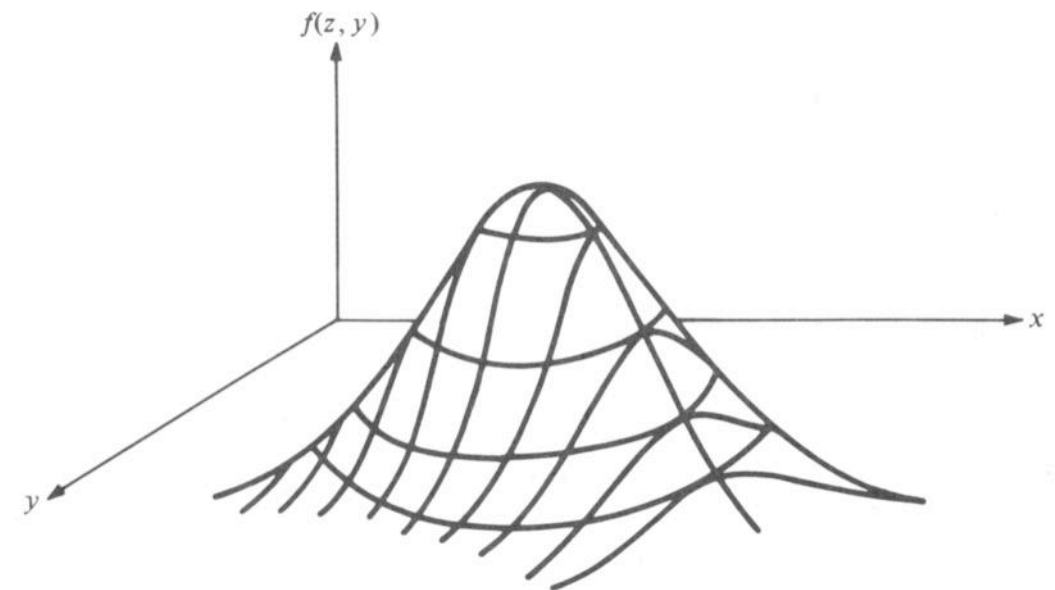
Der T-verteilte Wert $t = \frac{\rho_{XY} \cdot \sqrt{n-2}}{\sqrt{1-\rho_{XY}^2}}$ wird auf Signifikanz geprüft.

Korrelationen (ganz ähnlich)

Vor.: Zufallsvariablen X und Y
sind bivariat normalverteilt.

Ein theoretisches Maß der Korrelation
– des Kovariierens – von X und Y ist:

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{E[(X - \mu_X) \cdot (Y - \mu_Y)]}{\sigma_X \cdot \sigma_Y}$$



A Bivariate Normal Density Function

Der aus der Stichprobe berechnete Korrelationskoeffizient

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - m_X) \cdot (y_i - m_Y)}{n \cdot s_X \cdot s_Y}$$

ist ein Schätzwert für ρ_{XY} .

Hypothesen: $H_0 : \rho_{XY} = 0$
 $H_1 : \rho_{XY} \neq 0$

Bei n Zufallsvariablen gibt es $\frac{n \cdot (n - 1)}{2}$
zu testende Korrelationen.

Der T-verteilte Wert $t = \frac{\rho_{XY} \cdot \sqrt{n - 2}}{\sqrt{1 - \rho_{XY}^2}}$ wird auf Signifikanz geprüft.

Published Research Findings

1. Signifikanzniveaus:

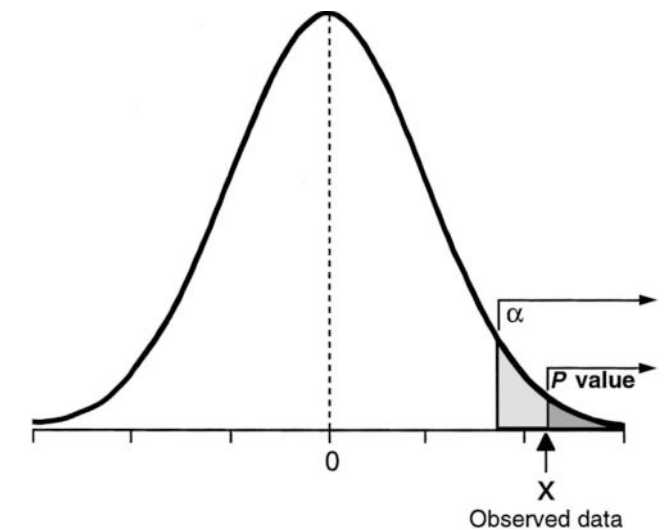
signifikant: $\alpha = 0.05$ hochsignifikant: $\alpha = 0.01$

2. Konfidenzintervall (Vertrauensintervall)

3. p-Wert (p-Value, P) p_α :

p_α ist die kleinste Irrtumswahrscheinlichkeit α , für die mit dem Kennwert x der Stichprobe die H_0 verworfen werden kann: $p_\alpha = \min \{ \alpha \mid x \geq c_\alpha \}$.

Einfacher ausgedrückt: $p_\alpha = \Pr (X \geq x \mid H_0)$.



Bei vorgegebenem Signifikanzniveau α lautet die Entscheidung jetzt:

$p_\alpha \leq \alpha \Rightarrow$ verwerfe die H_0

$p_\alpha > \alpha \Rightarrow$ akzeptiere die H_0 .

p_α liefert mehr Information als die Signifikanzschwellen, er gibt an, wie »ungewöhnlich« die Stichprobe unter Voraussetzung der H_0 ist.

Published Research Findings

1. Signifikanzniveaus:

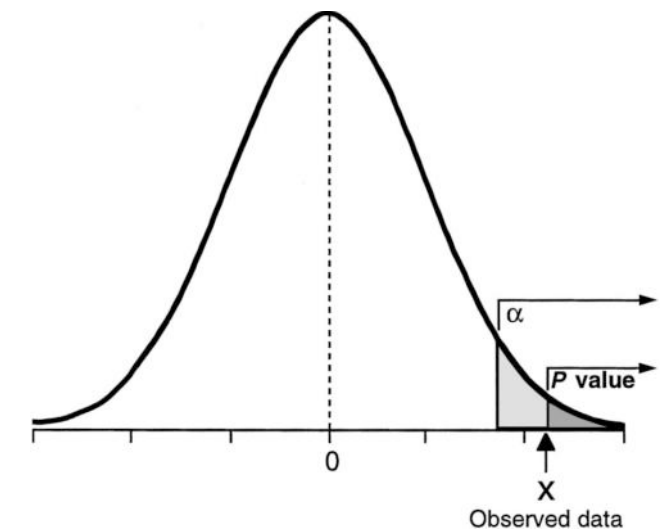
signifikant: $\alpha = 0.05$ hochsignifikant: $\alpha = 0.01$

2. Konfidenzintervall (Vertrauensintervall)

3. p-Wert (p-Value, P) p_α :

p_α ist die kleinste Irrtumswahrscheinlichkeit α , für die mit dem Kennwert x der Stichprobe die H_0 verworfen werden kann: $p_\alpha = \min \{ \alpha \mid x \geq c_\alpha \}$.

Einfacher ausgedrückt: $p_\alpha = \Pr (X \geq x \mid H_0)$.



Bei vorgegebenem Signifikanzniveau α lautet die Entscheidung jetzt:

$p_\alpha \leq \alpha \Rightarrow$ verwerfe die H_0

$p_\alpha > \alpha \Rightarrow$ akzeptiere die H_0 .

Aber:

p_α liefert mehr Information als die Signifikanzschwellen, er gibt an, wie »ungewöhnlich« die Stichprobe unter Voraussetzung der H_0 ist.

1. Signifikante Bedeutungslosigkeit

Bei der Angabe des p-Wertes erscheint nicht die Stichprobengröße, aber die Relevanz eines p_α -Ergebnisses kann sehr davon abhängen.

Beispiel:

$$H_0 : \mu = 100$$

$$H_1 : \mu > 100$$

Population sei normalverteilt mit $\sigma = 10$.

1. Stichprobe: $n = 9, m = 107.77 \Rightarrow z_1 = \frac{7.77}{10/\sqrt{9}} = 2.33 \Rightarrow p_\alpha = 0.01$

Vergleich mit $\mu = 101$: $z'_1 = \frac{6.77}{10/\sqrt{9}} = 2.03 \Rightarrow \beta_1 = 0.979$ bzw. $1 - \beta_1 = 0.021$

2. Stichprobe: $n = 900, m = 100.777 \Rightarrow z_2 = \frac{0.777}{10/\sqrt{900}} = 2.33 \Rightarrow p_\alpha = 0.01$

Vergleich mit $\mu = 101$: $z'_2 = \frac{-0.223}{10/\sqrt{9}} = -0.669 \Rightarrow \beta_2 = 0.245$ bzw. $1 - \beta_2 = 0.755$



(Fast) jeder Unterschied kann signifikant gemacht werden,

weil $\lim_{n \rightarrow \infty} z_M = \frac{m - \mu}{\sigma_M} = \frac{m - \mu}{\sigma/\sqrt{n}} \rightarrow \infty$, z. B. Differenzen ♀ vs. ♂.

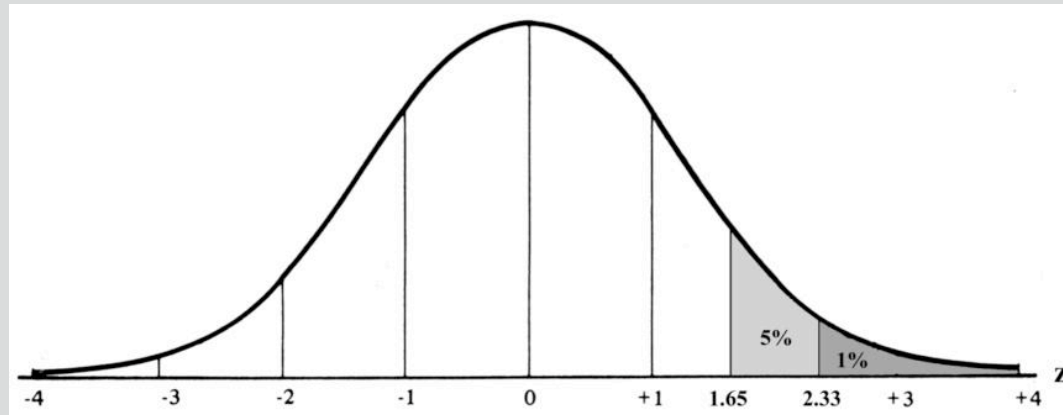
1. Signifikante Bedeutungslosigkeit

Bei der Anzahl n der Stichproben und der Stichprobengröße, aber die σ und μ sind für die Stichprobengröße, n und σ sind für n und σ davon abhängen.

Beispiele

$$H_0 : \mu = 100$$

$$H_1 : \mu > 100$$



$$\sigma = 10.$$

1. Stichprobe: $n = 9, m = 107.77 \Rightarrow z_1 = \frac{7.77}{10/\sqrt{9}} = 2.33 \Rightarrow p_\alpha = 0.01$

Vergleich mit $\mu = 101$: $z'_1 = \frac{6.77}{10/\sqrt{9}} = 2.03 \Rightarrow \beta_1 = 0.979$ bzw. $1 - \beta_1 = 0.021$

2. Stichprobe: $n = 900, m = 100.777 \Rightarrow z_2 = \frac{0.777}{10/\sqrt{900}} = 2.33 \Rightarrow p_\alpha = 0.01$

Vergleich mit $\mu = 101$: $z'_2 = \frac{-0.223}{10/\sqrt{9}} = -0.669 \Rightarrow \beta_2 = 0.245$ bzw. $1 - \beta_2 = 0.755$



(Fast) jeder Unterschied kann signifikant gemacht werden,

weil $\lim_{n \rightarrow \infty} z_M = \frac{m - \mu}{\sigma_M} = \frac{m - \mu}{\sigma/\sqrt{n}} \rightarrow \infty$, z. B. Differenzen ♀ vs. ♂.

1. Signifikante Bedeutungslosigkeit

Bei der Angabe des p-Wertes erscheint nicht die Stichprobengröße, aber die Relevanz eines p_α -Ergebnisses kann sehr davon abhängen.

Beispiel:

$$H_0 : \mu = 100$$

$$H_1 : \mu > 100$$

Population sei normalverteilt mit $\sigma = 10$.

$$1. \text{ Stichprobe: } n = 9, m = 107.77 \Rightarrow z_1 = \frac{7.77}{10/\sqrt{9}} = 2.33 \Rightarrow p_\alpha = 0.01$$

$$\text{Vergleich mit } \mu = 101: z'_1 = \frac{6.77}{10/\sqrt{9}} = 2.03 \Rightarrow \beta_1 = 0.979 \text{ bzw. } 1 - \beta_1 = 0.021$$

$$2. \text{ Stichprobe: } n = 900, m = 100.777 \Rightarrow z_2 = \frac{0.777}{10/\sqrt{900}} = 2.33 \Rightarrow p_\alpha = 0.01$$

$$\text{Vergleich mit } \mu = 101: z'_2 = \frac{-0.223}{10/\sqrt{9}} = -0.669 \Rightarrow \beta_2 = 0.245 \text{ bzw. } 1 - \beta_2 = 0.755$$



(Fast) jeder Unterschied kann signifikant gemacht werden,

weil $\lim_{n \rightarrow \infty} z_M = \frac{m - \mu}{\sigma_M} = \frac{m - \mu}{\sigma/\sqrt{n}} \rightarrow \infty$, z. B. Differenzen ♀ vs. ♂.

1. Signifikante Bedeutungslosigkeit

Bei der Angabe des p-Wertes erscheint nicht die Stichprobengröße, aber die Relevanz eines p_α -Ergebnisses kann sehr davon abhängen.

Beispiel:

$$H_0 : \mu = 100$$

$$H_1 : \mu > 100$$

Population sei normalverteilt

1. Stichprobe: $n = 9, m = 107.77 \Rightarrow z = \frac{6.77}{10/\sqrt{9}} = 2.031$

Vergleich mit $\mu = 101$: $z'_1 = \frac{6.77}{10/\sqrt{9}} = 2.031$

2. Stichprobe: $n = 900, m = 100.777 \Rightarrow z = \frac{-0.223}{10/\sqrt{900}} = -0.223$

Vergleich mit $\mu = 101$: $z'_2 = \frac{-0.223}{10/\sqrt{900}} = -0.223$

»When the sample size is very large, there is a real danger of detecting trivial associations as significant results. ... , all that a "significant" result or a small p-value implies is that we have observed something relatively unlikely given the hypothetical situation. Statistical significance is a statement about conditional probability, nothing else. It does not guarantee that something important, or even meaningful, has been found.«

[4, S. 445f]



(Fast) jeder Unterschied kann signifikant gemacht werden,

weil $\lim_{n \rightarrow \infty} z_M = \frac{m - \mu}{\sigma_M} = \frac{m - \mu}{\sigma/\sqrt{n}} \rightarrow \infty$, z. B. Differenzen ♀ vs. ♂.

1. Signifikante Bedeutungslosigkeit

Bei der Angabe des p-Wertes erscheint nicht die Stichprobengröße, aber die Relevanz eines p_α -Ergebnisses kann sehr davon abhängen.

Beispiel:

$$H_0 : \mu = 100$$

$$H_1 : \mu > 100$$

Population sei normalverteilt mit $\sigma = 10$.

$$1. \text{ Stichprobe: } n = 9, m = 107.77 \Rightarrow z_1 = \frac{7.77}{10/\sqrt{9}} = 2.33 \Rightarrow p_\alpha = 0.01$$

$$\text{Vergleich mit } \mu = 101: z'_1 = \frac{6.77}{10/\sqrt{9}} = 2.03 \Rightarrow \beta_1 = 0.979 \text{ bzw. } 1 - \beta_1 = 0.021$$

$$2. \text{ Stichprobe: } n = 900, m = 100.777 \Rightarrow z_2 = \frac{0.777}{10/\sqrt{900}} = 2.33 \Rightarrow p_\alpha = 0.01$$

$$\text{Vergleich mit } \mu = 101: z'_2 = \frac{-0.223}{10/\sqrt{9}} = -0.669 \Rightarrow \beta_2 = 0.245 \text{ bzw. } 1 - \beta_2 = 0.755$$



(Fast) jeder Unterschied kann signifikant gemacht werden,

weil $\lim_{n \rightarrow \infty} z_M = \frac{m - \mu}{\sigma_M} = \frac{m - \mu}{\sigma/\sqrt{n}} \rightarrow \infty$, z. B. Differenzen ♀ vs. ♂.

Koksfluten

The Times

March 25, 2006

Cocaine floods the playground

BY RICHARD FORD AND SEAN O'NEILL

- Use of the addictive drug by children doubles in a year

COCAINE use among children has doubled in a year as the fashionable drug of the middle classes extends its reach from the dinner party to the playground.

Hundreds of thousands of 11 to 15-year-olds are being offered the Class A drug, which is flooding into the country, according to government figures published yesterday.

$n \approx 9000$ in 305 Schulen,
 $1.4\% \rightarrow 1.9\%$ ($\Delta n = 45$),
 $p\text{-Wert} < 0.05$

Problematisches:

sehr großes n, kleine Prozentzahlen
 kein random sampling, sondern ›Cluster‹
 multiple comparisons

Table 5

Whether pupils had taken individual drugs in the last year: 1998-2005^a

Type of drug	Year								
	1998	1999	2000	2001*	2002	2003	2004	2005	
	%	%	%	%	%	%	%	%	
Cannabis	10	11	12	13	13	13	11	12	
Any stimulants	3	3	4	6	6	6	5	6	
Cocaine	1	1	1	1	1	1	1	2	
Crack	0	1	1	1	1	1	1	1	
Ecstasy	1	1	1	2	1	1	1	1	
Amphetamines ^b	2	1	1	1	1	1	1	1	
Poppers	1	2	2	3	4	4	3	4	
Any psychedelics^c	1	1	2	2	2	2	2	2	
LSD	1	1	1	1	1	1	1	1	
Magic mushrooms	1	1	1	2	1	2	2	2	
Ketamine ^d	d	d	d	d	d	d	d	0	
Any opiates	0	0	1	1	1	1	1	1	
Heroin	0	0	1	1	1	1	1	1	
Methadone	0	0	0	0	0	0	0	0	
Glue, gas, aerosols or solvents ^e	1	3	3	7	6	8	6	7	
Tranquillisers	0	0	0	0	0	0	0	0	
Anabolic steroids	0	0	0	0	0	0	0	0	
Other drugs	1	0	1	1	0	1	0	1	
Any Class A drug^f	3	2	4	4	4	4	4	4	
Any drug	11	12	14	20	20	21	18	19	
Any drug (excluding volatile substances)	11	11	13	16	16	16	14	15	
Bases	4647	9053	6862	9357	9830	10371	9666	9174	

^a Estimates from 2001 onwards are not comparable with estimates from previous years because of the change in the way that drug taking was measured. See note 6.

^b The 2004 survey onward asked about 'speed and other amphetamines'. See note 7.

^c The 2005 figure for psychedelics includes ketamine.

^d Ketamine was measured for the first time in 2005.

^e The 1998 survey asked about Glue or Solvents only, Gas was added as a separate category in 1999, and from 2000 questions were asked about Glue, Gas, Aerosols and other solvents as one category.

^f See note 5 for a definition of Class A drugs.

1. Signifikante Bedeutungslosigkeit

Bei der Angabe des p-Wertes erscheint nicht die Stichprobengröße, aber die Relevanz eines p_α -Ergebnisses kann sehr davon abhängen.

Beispiel:

$$H_0 : \mu = 100$$

$$H_1 : \mu > 100$$

Population sei normalverteilt mit $\sigma = 10$.

1. Stichprobe: $n = 9, m = 107.77 \Rightarrow z_1 = \frac{7.77}{10/\sqrt{9}} = 2.33 \Rightarrow p_\alpha = 0.01$

Vergleich mit $\mu = 101$: $z'_1 = \frac{6.77}{10/\sqrt{9}} = 2.03 \Rightarrow \beta_1 = 0.979$ bzw. $1 - \beta_1 = 0.021$

2. Stichprobe: $n = 900, m = 100.777 \Rightarrow z_2 = \frac{0.777}{10/\sqrt{900}} = 2.33 \Rightarrow p_\alpha = 0.01$

Vergleich mit $\mu = 101$: $z'_2 = \frac{-0.223}{10/\sqrt{9}} = -0.669 \Rightarrow \beta_2 = 0.245$ bzw. $1 - \beta_2 = 0.755$



(Fast) jeder Unterschied kann signifikant gemacht werden,

weil $\lim_{n \rightarrow \infty} z_M = \frac{m - \mu}{\sigma_M} = \frac{m - \mu}{\sigma/\sqrt{n}} \rightarrow \infty$, z. B. Differenzen ♀ vs. ♂.

2. (Miß-)Verständnis

Ein p-Wert von 0.05 bedeutet, daß die Nullhypothese nach dem Test nur eine Wahrscheinlichkeit von 5% hat, also unwahrscheinlich ist und somit wahrscheinlich ein Effekt existiert.

- stimmt
- stimmt nicht

2. (Miß-)Verständnis

Ein p-Wert von 0.05 bedeutet, daß die Nullhypothese nach dem Test nur eine Wahrscheinlichkeit von 5% hat, also unwahrscheinlich ist und somit wahrscheinlich ein Effekt existiert.

- stimmt
 stimmt nicht

2. (Miß-)Verständnis

Ein p-Wert von 0.05 bedeutet, daß die Nullhypothese nach dem Test nur eine Wahrscheinlichkeit von 5% hat, also unwahrscheinlich ist und somit wahrscheinlich ein Effekt existiert.

- stimmt
 stimmt nicht

2. (Miß-)Verständnis

Ein p-Wert von 0.05 bedeutet, daß die Nullhypothese nach dem Test nur eine Wahrscheinlichkeit von 5% hat, also unwahrscheinlich ist und somit wahrscheinlich ein Effekt existiert.

- stimmt
 stimmt nicht

Ein signifikantes Ergebnis bzw. ein signifikanter p-Wert wird unter Annahme der H_0 berechnet, gibt also die bedingte Wahrscheinlichkeit $\Pr(p_\alpha \leq 0.05 | H_0)$ an, obiges Mißverständnis behauptet aber $\Pr(H_0 | p_\alpha \leq 0.05) \leq 0.05$.

$$\Pr(H_0|y) \cdot \Pr(y) = \Pr(y|H_0) \cdot \Pr(H_0),$$

aber im Allgemeinen ist

$$\begin{array}{ccc} \Pr(y) \neq \Pr(H_0) & \longrightarrow & \Pr(H_0|\text{Ergebnis}) \neq \Pr(\text{Ergebnis}|H_0) \\ \sim & & \\ \Pr(H_0|y) \neq \Pr(y|H_0) & & \end{array}$$

2. (Miß-)Verständnis

Ein p-Wert von 0.05 bedeutet, daß die Nullhypothese nach dem Test nur eine Wahrscheinlichkeit von 5% hat, also unwahrscheinlich ist und somit wahrscheinlich ein Effekt existiert.

stimmt
 stimmt nicht

Ein signifikantes Ergebnis bzw. ein signifikanter p-Wert wird unter Annahme der H_0 berechnet, gibt also die bedingte Wahrscheinlichkeit $\Pr(p_\alpha \leq 0.05 | H_0)$ an. Das obige Mißverständnis behauptet aber

$$\Pr(H_0|y) \cdot \Pr(y) = \Pr(y|H_0) \cdot \Pr(H_0),$$

aber im Allgemeinen ist

$$\Pr(y) \neq \Pr(H_0) \quad \rightarrow \quad \Pr(H_0|y) \neq \Pr(y|H_0)$$

\rightsquigarrow

$$\Pr(H_0|y) \neq \Pr(y|H_0)$$

Probabilities of Erroneous and Correct Decisions

		TRUE SITUATION	
		H_0	H_1
DECISION	Accept H_0	$1 - \alpha$	β
	Reject H_0	α	$1 - \beta$

täuscht!

Betrachten wir die Situation also genauer.

Kontingenztafel der Produktwahrscheinlichkeiten

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$\Pr(p_\alpha \leq \alpha H_1) \cdot \Pr(H_1)$	$\Pr(p_\alpha > \alpha H_1) \cdot \Pr(H_1)$	$\Pr(H_1)$
falsch	(H_0)	$\Pr(p_\alpha \leq \alpha H_0) \cdot \Pr(H_0)$	$\Pr(p_\alpha > \alpha H_0) \cdot \Pr(H_0)$	$\Pr(H_0)$
	Σ	$\Pr(p_\alpha \leq \alpha)$	$\Pr(p_\alpha > \alpha)$	1

Betrachte die Maße:

$$FPRP = \frac{\Pr(p_\alpha \leq \alpha | H_0) \cdot \Pr(H_0)}{\Pr(p_\alpha \leq \alpha)} = \Pr(H_0 | p_\alpha \leq \alpha) \quad (\mathbf{False\ Positive\ Report\ Probability, [2]})$$

$$FPRP + PPV = 1$$

$$PPV = \frac{\Pr(p_\alpha \leq \alpha | H_1) \cdot \Pr(H_1)}{\Pr(p_\alpha \leq \alpha)} = \Pr(H_1 | p_\alpha \leq \alpha) \quad (\mathbf{Positive\ Predictive\ Value, [1]})$$

Kontingenztafel der Produktwahrscheinlichkeiten

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$\Pr(p_\alpha \leq \alpha H_1) \cdot \Pr(H_1)$	$\Pr(p_\alpha > \alpha H_1) \cdot \Pr(H_1)$	$\Pr(H_1)$
falsch	(H_0)	$\Pr(p_\alpha \leq \alpha H_0) \cdot \Pr(H_0)$	$\Pr(p_\alpha > \alpha H_0) \cdot \Pr(H_0)$	$\Pr(H_0)$
	Σ	$\Pr(p_\alpha \leq \alpha)$	$\Pr(p_\alpha > \alpha)$	1

Betrachte die Maße:

$$FPRP = \frac{\Pr(p_\alpha \leq \alpha | H_0) \cdot \Pr(H_0)}{\Pr(p_\alpha \leq \alpha)} = \Pr(H_0 | p_\alpha \leq \alpha) \quad (\text{False Positive Report Probability, [2]})$$

$$FPRP + PPV = 1$$

$$PPV = \frac{\Pr(p_\alpha \leq \alpha | H_1) \cdot \Pr(H_1)}{\Pr(p_\alpha \leq \alpha)} = \Pr(H_1 | p_\alpha \leq \alpha) \quad (\text{Positive Predictive Value, [1]})$$

Kontingenztafel der Produktwahrscheinlichkeiten

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$\Pr(p_\alpha \leq \alpha H_1) \cdot \Pr(H_1)$	$\Pr(p_\alpha > \alpha H_1) \cdot \Pr(H_1)$	$\Pr(H_1)$
falsch	(H_0)	$\Pr(p_\alpha \leq \alpha H_0) \cdot \Pr(H_0)$	$\Pr(p_\alpha > \alpha H_0) \cdot \Pr(H_0)$	$\Pr(H_0)$
	Σ	$\Pr(p_\alpha \leq \alpha)$	$\Pr(p_\alpha > \alpha)$	1

Betrachte die Maße:

$$FPRP = \frac{\Pr(p_\alpha \leq \alpha | H_0) \cdot \Pr(H_0)}{\Pr(p_\alpha \leq \alpha)} = \Pr(H_0 | p_\alpha \leq \alpha) \quad (\text{False Positive Report Probability, [2]})$$

$$FPRP + PPV = 1$$

$$PPV = \frac{\Pr(p_\alpha \leq \alpha | H_1) \cdot \Pr(H_1)}{\Pr(p_\alpha \leq \alpha)} = \Pr(H_1 | p_\alpha \leq \alpha) \quad (\text{Positive Predictive Value, [1]})$$

Kontingenztafel der Produktwahrscheinlichkeiten

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$\Pr(p_\alpha \leq \alpha H_1) \cdot \Pr(H_1)$	$\Pr(p_\alpha > \alpha H_1) \cdot \Pr(H_1)$	$\Pr(H_1)$
falsch	(H_0)	$\Pr(p_\alpha \leq \alpha H_0) \cdot \Pr(H_0)$	$\Pr(p_\alpha > \alpha H_0) \cdot \Pr(H_0)$	$\Pr(H_0)$
	Σ	$\Pr(p_\alpha \leq \alpha)$	$\Pr(p_\alpha > \alpha)$	1

Betrachte die Maße:

$$\Pr(\theta_i|y) = \frac{\Pr(y|\theta_i) \cdot \Pr(\theta_i)}{\Pr(y)} = \frac{\Pr(y|\theta_i) \cdot \Pr(\theta_i)}{\sum_{j=1}^J \Pr(y|\theta_j) \Pr(\theta_j)}$$

$$FPRP = \frac{\Pr(p_\alpha \leq \alpha | H_0) \cdot \Pr(H_0)}{\Pr(p_\alpha \leq \alpha)} = \Pr(H_0 | p_\alpha \leq \alpha) \quad (\text{False Positive Report Probability, [2]})$$

$$FPRP + PPV = 1$$

$$PPV = \frac{\Pr(p_\alpha \leq \alpha | H_1) \cdot \Pr(H_1)}{\Pr(p_\alpha \leq \alpha)} = \Pr(H_1 | p_\alpha \leq \alpha) \quad (\text{Positive Predictive Value, [1]})$$

Die Maße können als a posteriori Wahrscheinlichkeiten aufgefaßt werden.

Kontingenztafel der Produktwahrscheinlichkeiten

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$\Pr(p_\alpha \leq \alpha H_1) \cdot \Pr(H_1)$	$\Pr(p_\alpha > \alpha H_1) \cdot \Pr(H_1)$	$\Pr(H_1)$
falsch	(H_0)	$\Pr(p_\alpha \leq \alpha H_0) \cdot \Pr(H_0)$	$\Pr(p_\alpha > \alpha H_0) \cdot \Pr(H_0)$	$\Pr(H_0)$
	Σ	$\Pr(p_\alpha \leq \alpha)$	$\Pr(p_\alpha > \alpha)$	1

Betrachte die Maße:

$$FPRP = \frac{\Pr(p_\alpha \leq \alpha | H_0) \cdot \Pr(H_0)}{\Pr(p_\alpha \leq \alpha)} = \Pr(H_0 | p_\alpha \leq \alpha) \quad (\text{False Positive Report Probability, [2]})$$

$$FPRP + PPV = 1$$

$$PPV = \frac{\Pr(p_\alpha \leq \alpha | H_1) \cdot \Pr(H_1)}{\Pr(p_\alpha \leq \alpha)} = \Pr(H_1 | p_\alpha \leq \alpha) \quad (\text{Positive Predictive Value, [1]})$$

Die Maße können als a posteriori Wahrscheinlichkeiten aufgefaßt werden.

Ein erstes (fiktives) Beispiel aus der Medizin

Table 2 Number of times we accept and reject null hypothesis, under plausible assumptions regarding conduct of medical research (adapted from Oakes²⁵)

Result of experiment	Null hypothesis true (treatment doesn't work)	Null hypothesis false (treatment works)	Total
Accept null hypothesis	855	50	905
Reject null hypothesis	45	50	95
Total	900	100	1000

Von den 95 signifikanten Ergebnissen sind 45 (= 47%) falsch, also $FPRP = 0.47$ und $PPV = 0.53$.

Dabei wurde angenommen, daß 10% echte Effekte existieren und die Tests eine Trennschärfe von 50% haben.

Produktwahrscheinlichkeiten

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$(1 - \beta) \cdot \pi$	$\beta \cdot \pi$	π
falsch	(H_0)	$\alpha \cdot (1 - \pi)$	$(1 - \alpha) \cdot (1 - \pi)$	$1 - \pi$
	Σ	$(1 - \beta) \cdot \pi + \alpha \cdot (1 - \pi)$	$\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)$	1

Sei $\Pr(H_1) = \pi$, $\Pr(H_0) = (1 - \pi)$; π ist normalerweise nicht bekannt, aber oft dimensionsmäßig abzuschätzen. Dann ergibt sich:

$$FPRP = \frac{\alpha \cdot (1 - \pi)}{\alpha \cdot (1 - \pi) + (1 - \beta) \cdot \pi} = \frac{(1 - \pi)}{(1 - \pi) + \frac{(1 - \beta)}{\alpha} \cdot \pi} = \frac{1}{1 + \frac{(1 - \beta)}{\alpha} \cdot \frac{\pi}{1 - \pi}}$$

$$PPV = \frac{(1 - \beta) \cdot \pi}{\alpha \cdot (1 - \pi) + (1 - \beta) \cdot \pi} = \frac{\pi}{\pi + \frac{\alpha}{(1 - \beta)} \cdot (1 - \pi)} = \frac{1}{1 + \frac{\alpha}{(1 - \beta)} \cdot \frac{(1 - \pi)}{\pi}}$$

Es gilt: $PPV > 0.5 \Leftrightarrow \frac{\pi}{1 - \pi} > \frac{\alpha}{1 - \beta} \Leftrightarrow \pi > \frac{\alpha}{(1 - \beta) + \alpha}$,

was nur schwer zu erreichen ist, z. B.: $\alpha = 0.05$ und $(1 - \beta) = 0.95 \Rightarrow \pi > 0.05$.

Produktwahrscheinlichkeiten

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$(1 - \beta) \cdot \pi$	$\beta \cdot \pi$	π
falsch	(H_0)	$\alpha \cdot (1 - \pi)$	$(1 - \alpha) \cdot (1 - \pi)$	$1 - \pi$
	Σ	$(1 - \beta) \cdot \pi + \alpha \cdot (1 - \pi)$	$\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)$	1

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$\Pr(p_\alpha \leq \alpha H_1) \cdot \Pr(H_1)$	$\Pr(p_\alpha > \alpha H_1) \cdot \Pr(H_1)$	$\Pr(H_1)$
falsch	(H_0)	$\Pr(p_\alpha \leq \alpha H_0) \cdot \Pr(H_0)$	$\Pr(p_\alpha > \alpha H_0) \cdot \Pr(H_0)$	$\Pr(H_0)$
	Σ	$\Pr(p_\alpha \leq \alpha)$	$\Pr(p_\alpha > \alpha)$	1

kannt,

was nur schwer zu erreichen ist, z. B.: $\alpha = 0.05$ und $(1 - \beta) = 0.95 \Rightarrow \pi > 0.05$.

Produktwahrscheinlichkeiten

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$(1 - \beta) \cdot \pi$	$\beta \cdot \pi$	π
falsch	(H_0)	$\alpha \cdot (1 - \pi)$	$(1 - \alpha) \cdot (1 - \pi)$	$1 - \pi$
	Σ	$(1 - \beta) \cdot \pi + \alpha \cdot (1 - \pi)$	$\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)$	1

Sei $\Pr(H_1) = \pi$, $\Pr(H_0) = (1 - \pi)$; π ist normalerweise nicht bekannt, aber oft dimensionsmäßig abzuschätzen. Dann ergibt sich:

$$FPRP = \frac{\alpha \cdot (1 - \pi)}{\alpha \cdot (1 - \pi) + (1 - \beta) \cdot \pi} = \frac{(1 - \pi)}{(1 - \pi) + \frac{(1 - \beta)}{\alpha} \cdot \pi} = \frac{1}{1 + \frac{(1 - \beta)}{\alpha} \cdot \frac{\pi}{1 - \pi}}$$

$$PPV = \frac{(1 - \beta) \cdot \pi}{\alpha \cdot (1 - \pi) + (1 - \beta) \cdot \pi} = \frac{\pi}{\pi + \frac{\alpha}{(1 - \beta)} \cdot (1 - \pi)} = \frac{1}{1 + \frac{\alpha}{(1 - \beta)} \cdot \frac{(1 - \pi)}{\pi}}$$

Es gilt: $PPV > 0.5 \Leftrightarrow \frac{\pi}{1 - \pi} > \frac{\alpha}{1 - \beta} \Leftrightarrow \pi > \frac{\alpha}{(1 - \beta) + \alpha}$,

was nur schwer zu erreichen ist, z. B.: $\alpha = 0.05$ und $(1 - \beta) = 0.95 \Rightarrow \pi > 0.05$.

Produktwahrscheinlichkeiten

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$(1 - \beta) \cdot \pi$	$\beta \cdot \pi$	π
falsch	(H_0)	$\alpha \cdot (1 - \pi)$	$(1 - \alpha) \cdot (1 - \pi)$	$1 - \pi$
	Σ	$(1 - \beta) \cdot \pi + \alpha \cdot (1 - \pi)$	$\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)$	1

Sei $\Pr(H_1) = \pi$, $\Pr(H_0) = (1 - \pi)$; π ist normalerweise nicht bekannt, aber oft dimensionsmäßig abzuschätzen. Dann ergibt sich:

$$FPRP = \frac{\alpha \cdot (1 - \pi)}{\alpha \cdot (1 - \pi) + (1 - \beta) \cdot \pi} = \frac{(1 - \pi)}{(1 - \pi) + \frac{(1 - \beta)}{\alpha} \cdot \pi} = \frac{1}{1 + \frac{(1 - \beta)}{\alpha} \cdot \frac{\pi}{1 - \pi}}$$

$$PPV = \frac{(1 - \beta) \cdot \pi}{\alpha \cdot (1 - \pi) + (1 - \beta) \cdot \pi} = \frac{\pi}{\pi + \frac{\alpha}{(1 - \beta)} \cdot (1 - \pi)} = \frac{1}{1 + \frac{\alpha}{(1 - \beta)} \cdot \frac{(1 - \pi)}{\pi}}$$

Es gilt: $PPV > 0.5 \Leftrightarrow \frac{\pi}{1 - \pi} > \frac{\alpha}{1 - \beta} \Leftrightarrow \pi > \frac{\alpha}{(1 - \beta) + \alpha}$,

was nur schwer zu erreichen ist, z. B.: $\alpha = 0.05$ und $(1 - \beta) = 0.45 \Rightarrow \pi > 0.10$.

Koksfluten revisited

The Times

March 25, 2006

Cocaine floods the playground

BY RICHARD FORD AND SEAN O'NEILL

- Use of the addictive drug by children doubles in a year

COCAINE use among children has doubled in a year as the fashionable drug of the middle classes extends its reach from the dinner party to the playground.

Hundreds of thousands of 11 to 15-year-olds are being offered the Class A drug, which is flooding into the country, according to government figures published yesterday.

$n \approx 9000$ in 305 Schulen,
 $1.4\% \rightarrow 1.9\%$ ($\Delta n = 45$),
 $p\text{-Wert} < 0.05$

Problematisches:

sehr großes n , kleine Prozentzahlen
 kein random sampling, sondern ›Cluster‹
 multiple comparisons

$PPV = 0.17$, für $\alpha = 0.05$, $1 - \beta = 1$, $\pi = 0.01$

Table 5

Whether pupils had taken individual drugs in the last year: 1998-2005^a

Type of drug	Year								
	1998	1999	2000	2001*	2002	2003	2004	2005	
	%	%	%	%	%	%	%	%	
Cannabis	10	11	12	13	13	13	11	12	
Any stimulants	3	3	4	6	6	6	5	6	
Cocaine	1	1	1	1	1	1	1	2	
Crack	0	1	1	1	1	1	1	1	
Ecstasy	1	1	1	2	1	1	1	1	
Amphetamines ^b	2	1	1	1	1	1	1	1	
Poppers	1	2	2	3	4	4	3	4	
Any psychedelics^c	1	1	2	2	2	2	2	2	
LSD	1	1	1	1	1	1	1	1	
Magic mushrooms	1	1	1	2	1	2	2	2	
Ketamine ^d	d	d	d	d	d	d	d	0	
Any opiates	0	0	1	1	1	1	1	1	
Heroin	0	0	1	1	1	1	1	1	
Methadone	0	0	0	0	0	0	0	0	
Glue, gas, aerosols or solvents ^e	1	3	3	7	6	8	6	7	
Tranquillisers	0	0	0	0	0	0	0	0	
Anabolic steroids	0	0	0	0	0	0	0	0	
Other drugs	1	0	1	1	0	1	0	1	
Any Class A drug^f	3	2	4	4	4	4	4	4	
Any drug	11	12	14	20	20	21	18	19	
Any drug (excluding volatile substances)	11	11	13	16	16	16	14	15	
Bases	4647	9053	6862	9357	9830	10371	9666	9174	

^a Estimates from 2001 onwards are not comparable with estimates from previous years because of the change in the way that drug taking was measured. See note 6.

^b The 2004 survey onward asked about 'speed and other amphetamines'. See note 7.

^c The 2005 figure for psychedelics includes ketamine.

^d Ketamine was measured for the first time in 2005.

^e The 1998 survey asked about Glue or Solvents only, Gas was added as a separate category in 1999, and from 2000 questions were asked about Glue, Gas, Aerosols and other solvents as one category.

^f See note 5 for a definition of Class A drugs.

Idealtypische Beispielrechnungen

1. Gentechnologie

100000 Gene, ≈ 10 Gen-Polymorphismen, $RR = 1.3$, $\alpha = 0.05$

$\Rightarrow \pi = 10^{-4}$, $1 - \beta = 0.60$ und $PPV = 12 \cdot 10^{-4}$

+ $u = 10\% \Rightarrow PPV = 4.4 \cdot 10^{-4}$

+ $k = 10 \Rightarrow PPV = 1.5 \cdot 10^{-4}$

2. Terroristen-Fahndung mit Face-Recognition an Flughäfen

Terroristen : Bürger = 1 : 1000000 $\Rightarrow \pi = 10^{-6}$

$\alpha = 0.10$, $1 - \beta = 0.50$ (technisch und sozial bedingt, beides utopisch)

$\Rightarrow PPV = 0.000005$

$n = 10^9$ Fluggäste/Jahr $\Rightarrow 10^8$ Fehlalarme/Jahr

Idealtypische Beispielrechnungen

1. Gentechnologie

Neues Gen für Übergewicht entdeckt

Bonn/Washington (dpa) - Ein internationales Forscherteam hat eine neue Genvariante für Übergewicht entdeckt. Menschen, die dieses Gen in sich tragen, seien zu 30 Prozent häufiger übergewichtig als andere, teilte das Nationale Genomforschungsnetz am Donnerstag in Bonn mit.

Etwa zehn Prozent der Bevölkerung besäßen diese genetische Variante, die das Risiko für Übergewicht und Fettleibigkeit im Kindes- und Erwachsenenalter erhöhe.

Das Team unter Mitarbeit von Prof. Thomas Meitinger und Prof. Erich Wichmann vom Forschungszentrum für Umwelt und Gesundheit (GSF Neuherberg) sowie Forschern aus Essen und den USA untersuchte rund 100 000 Genvarianten im menschlichen Erbgut. Dabei entdeckten die Forscher eine Genveränderung in der Nähe eines Gens, das den Fettstoffwechsel steuert. Die gefundene Genvariante komme sowohl bei Menschen mit westeuropäischer als auch mit afroamerikanischer Abstammung vor.

Es sind bereits mehrere Gene bekannt, die das Gewicht beeinflussen. Die Forscher weisen jedoch auch darauf hin, dass Gene allein nicht dick machen. Vielmehr führe die Kombination aus genetischen Faktoren und Lebensstil dazu, dass Menschen zunehmen. Die Ergebnisse der Untersuchung stellen die Forscher in der US-Fachzeitschrift "Science" (Bd. 312, S. 279) vom Freitag vor.



Study reveals new genes for excessive alcohol drinking

Researchers supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA), part of the National Institutes of Health (NIH), have identified new genes that may contribute to excessive alcohol consumption. The new study, conducted with strains of animals that have either a high or low innate preference for alcohol, provides clues about the molecular mechanisms that underlie the tendency to drink heavily. A report of the findings appears in the April 18, 2006 issue of Proceedings of the National Academy of Sciences.

"These findings provide a wealth of new insights into the molecular determinants of excessive drinking, which could lead to a better understanding of alcoholism," notes NIAAA Director Ting-Kai Li, M.D. "They also underscore the value that animal models bring to the investigation of complex human disorders such as alcohol dependence."

Idealtypische Beispielrechnungen

1. Gentechnologie

100000 Gene, ≈ 10 Gen-Polymorphismen, $RR = 1.3$, $\alpha = 0.05$

$\Rightarrow \pi = 10^{-4}$, $1 - \beta = 0.60$ und $PPV = 12 \cdot 10^{-4}$

+ $u = 10\% \Rightarrow PPV = 4.4 \cdot 10^{-4}$

+ $k = 10 \Rightarrow PPV = 1.5 \cdot 10^{-4}$

2. Terroristen-Fahndung mit Face-Recognition an Flughäfen

Terroristen : Bürger = 1 : 1000000 $\Rightarrow \pi = 10^{-6}$

$\alpha = 0.10$, $1 - \beta = 0.50$ (technisch und sozial bedingt, beides utopisch)

$\Rightarrow PPV = 0.000005$

$n = 10^9$ Fluggäste/Jahr $\Rightarrow 10^8$ Fehlalarme/Jahr

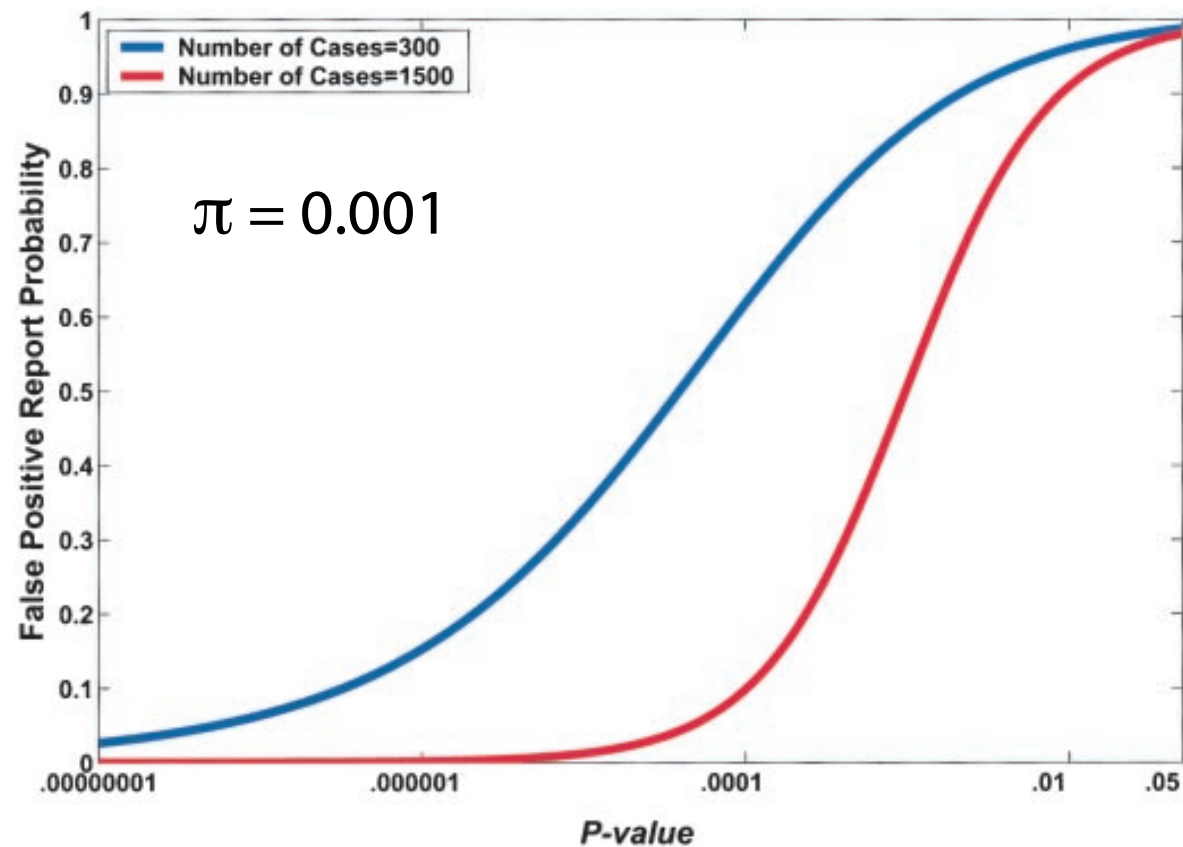
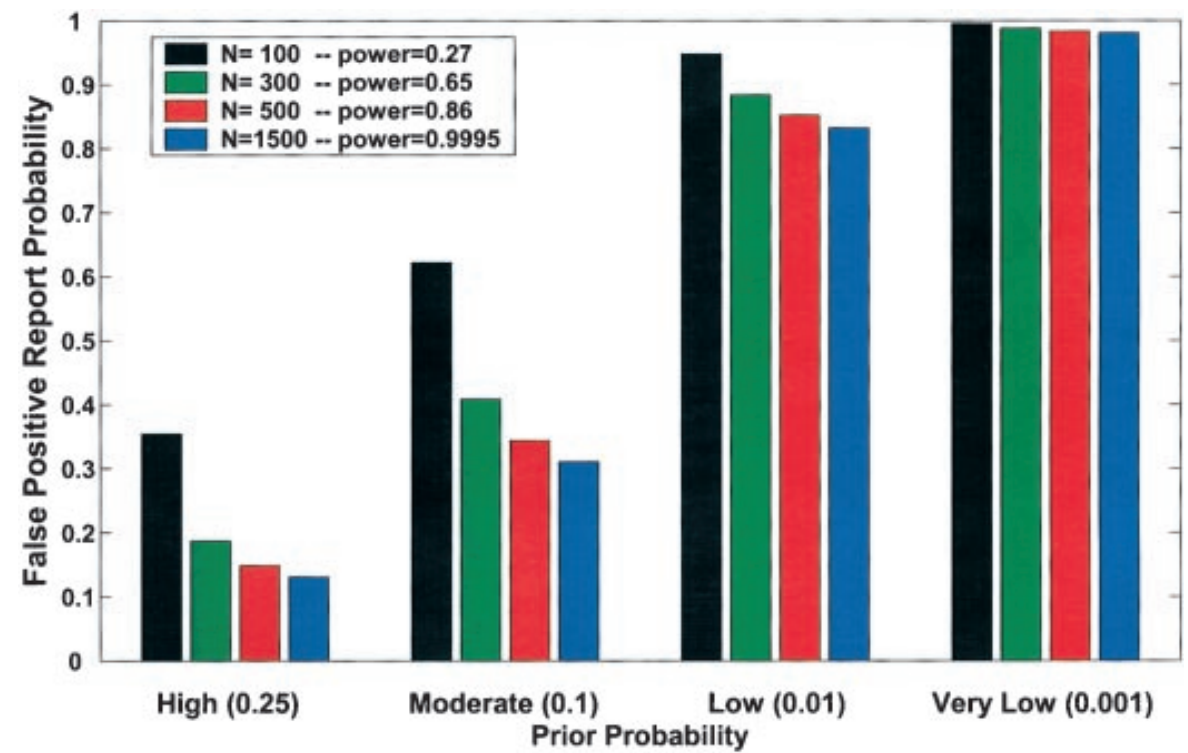
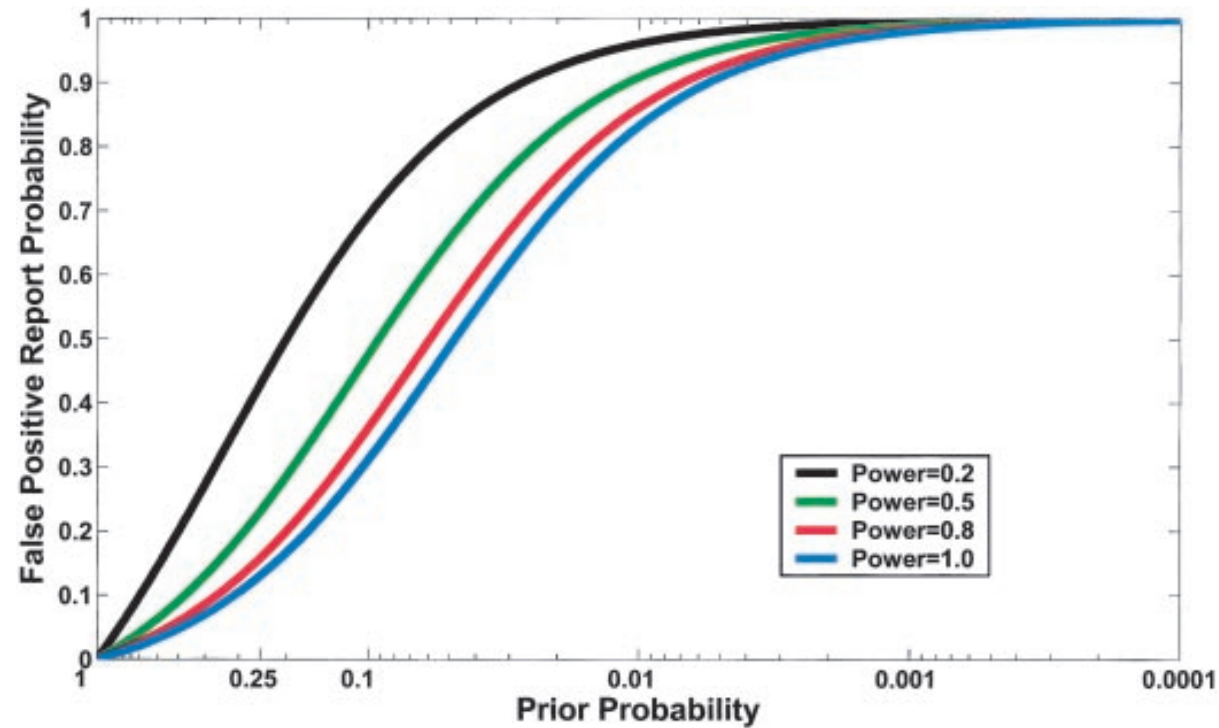
FPRP für Kombinationen von π , α und $1-\beta$

Table 3 Proportion of false positive significant results with three different criteria for significance

Power of study (proportion (%) of time we reject null hypothesis if it is false)	Percentage of “significant” results that are false positives		
	P=0.05	P=0.01	P=0.001
80% of ideas correct (null hypothesis false)			
20	5.9	1.2	0.10
50	2.4	0.5	0.05
80	1.5	0.3	0.03
50% of ideas correct (null hypothesis false)			
20	20.0	4.8	0.50
50	9.1	2.0	0.20
80	5.9	1.2	0.10
10% of ideas correct (null hypothesis false)			
20	69.2	31.0	4.30
50	47.4*	15.3	1.80
80	36.0	10.1	1.10
1% of ideas correct (null hypothesis false)			
20	96.1	83.2	33.10
50	90.8	66.4	16.50
80	86.1	55.3	11.00

*Corresponds to assumptions in table 2.

FPRP für Kombinationen von π , n und $1-\beta$ ($\alpha = 0.05$)



Durch Vergrößern der Stichprobe kann $1-\beta$ verbessert werden, was aber nichts bringt, wenn π klein ist.

Produktwahrscheinlichkeiten

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$(1 - \beta) \cdot \pi$	$\beta \cdot \pi$	π
falsch	(H_0)	$\alpha \cdot (1 - \pi)$	$(1 - \alpha) \cdot (1 - \pi)$	$1 - \pi$
	Σ	$(1 - \beta) \cdot \pi + \alpha \cdot (1 - \pi)$	$\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)$	1

Sei $\Pr(H_1) = \pi$, $\Pr(H_0) = (1 - \pi)$; π ist normalerweise nicht bekannt, aber oft dimensionsmäßig abzuschätzen. Dann ergibt sich:

$$FPRP = \frac{\alpha \cdot (1 - \pi)}{\alpha \cdot (1 - \pi) + (1 - \beta) \cdot \pi} = \frac{(1 - \pi)}{(1 - \pi) + \frac{(1 - \beta)}{\alpha} \cdot \pi} = \frac{1}{1 + \frac{(1 - \beta)}{\alpha} \cdot \frac{\pi}{1 - \pi}}$$

$$PPV = \frac{(1 - \beta) \cdot \pi}{\alpha \cdot (1 - \pi) + (1 - \beta) \cdot \pi} = \frac{\pi}{\pi + \frac{\alpha}{(1 - \beta)} \cdot (1 - \pi)} = \frac{1}{1 + \frac{\alpha}{(1 - \beta)} \cdot \frac{(1 - \pi)}{\pi}}$$

Es gilt: $PPV > 0.5 \Leftrightarrow \frac{\pi}{1 - \pi} > \frac{\alpha}{1 - \beta} \Leftrightarrow \pi > \frac{\alpha}{(1 - \beta) + \alpha}$,

was nur schwer zu erreichen ist, z. B.: $\alpha = 0.05$ und $(1 - \beta) = 0.45 \Rightarrow \pi > 0.10$.

Übergang zu (absoluten) Häufigkeiten

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$h_{11} = n \cdot (1 - \beta) \cdot \pi$	$h_{12} = n \cdot \beta \cdot \pi$	$n \cdot \pi$
falsch	(H_0)	$h_{21} = n \cdot \alpha \cdot (1 - \pi)$	$h_{22} = n \cdot (1 - \alpha) \cdot (1 - \pi)$	$n \cdot (1 - \pi)$
	Σ	$n \cdot [(1 - \beta) \cdot \pi + \alpha \cdot (1 - \pi)]$	$n \cdot [\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)]$	n

Es seien n Versuche mit gleichem α und β ausgeführt worden.

Bias

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$h'_{11} = n \cdot [(1 - \beta) + u \cdot \beta] \cdot \pi$	$h'_{12} = (1 - u) \cdot n \cdot \beta \cdot \pi$	$n \cdot \pi$
falsch	(H_0)	$h'_{21} = n \cdot [\alpha + u \cdot (1 - \alpha)] \cdot (1 - \pi)$	$h'_{22} = (1 - u) \cdot n \cdot (1 - \alpha) \cdot (1 - \pi)$	$n \cdot (1 - \pi)$
	Σ	$n \cdot [(1 - \beta) \cdot \pi + \alpha \cdot (1 - \pi) + u \cdot (\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi))]$	$(1 - u) \cdot n \cdot [\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)]$	n

Es wird ein Anteil von $u \cdot 100\%$ nichtsignifikanten Ergebnissen zu den signifikanten ›verschoben‹.

Bias

Fakt	Test	nichtsignifikant	signifikant	Σ
wahr (π)		$(1 - \beta) \cdot n \cdot \pi$	$\beta \cdot n \cdot \pi$	$n \cdot \pi$
falsch ($1 - \pi$)		$(1 - \alpha) \cdot (1 - \pi) \cdot n$	$\alpha \cdot (1 - \pi) \cdot n$	$n \cdot (1 - \pi)$
	Σ	$(1 - \alpha) \cdot (1 - \pi) \cdot n + (1 - \beta) \cdot n \cdot \pi$	$\alpha \cdot (1 - \pi) \cdot n + \beta \cdot n \cdot \pi$	n

... let us define bias as the combination of various design, data, analysis, and presentation factors that tend to produce research findings when they should not be produced. [1, S.697]

Commercially available "data mining" packages actually are proud of their ability to yield statistically significant results through data dredging. [1, S.699]

Es wird ein Anteil von $\alpha \cdot 100\%$ nichtsignifikanten Ergebnissen zu den signifikanten ›verschoben‹.

Bias

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$h'_{11} = n \cdot [(1 - \beta) + u \cdot \beta] \cdot \pi$	$h'_{12} = (1 - u) \cdot n \cdot \beta \cdot \pi$	$n \cdot \pi$
falsch	(H_0)	$h'_{21} = n \cdot [\alpha + u \cdot (1 - \alpha)] \cdot (1 - \pi)$	$h'_{22} = (1 - u) \cdot n \cdot (1 - \alpha) \cdot (1 - \pi)$	$n \cdot (1 - \pi)$
	Σ	$n \cdot [(1 - \beta) \cdot \pi + \alpha \cdot (1 - \pi) + u \cdot (\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi))]$	$(1 - u) \cdot n \cdot [\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)]$	n

Es wird ein Anteil von $u \cdot 100\%$ nichtsignifikanten Ergebnissen zu den signifikanten ›verschoben‹.

Bias

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$f'_{11} = \frac{n}{n} \cdot [(1 - \beta) + u \cdot \beta] \cdot \pi$	$f'_{12} = \frac{(1 - u) \cdot n}{n} \cdot \beta \cdot \pi$	π
falsch	(H_0)	$f'_{21} = \frac{n}{n} \cdot [\alpha + u \cdot (1 - \alpha)] \cdot (1 - \pi)$	$f'_{22} = \frac{(1 - u) \cdot n}{n} \cdot (1 - \alpha) \cdot (1 - \pi)$	$1 - \pi$
	Σ	$(1 - \beta) \cdot \pi + \alpha \cdot (1 - \pi) + u \cdot [\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)]$	$(1 - u) \cdot [\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)]$	1

Es werden neue relative Häufigkeiten für die Bias-Verteilung gebildet.

Bias

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$f'_{11} = (1 - \beta') \cdot \pi$	$f'_{12} = \beta' \cdot \pi$	π
falsch	(H_0)	$f'_{21} = \alpha' \cdot (1 - \pi)$	$f'_{22} = (1 - \alpha') \cdot (1 - \pi)$	$1 - \pi$
	Σ	$(1 - \beta') \cdot \pi + \alpha' \cdot (1 - \pi)$	$\beta' \cdot \pi + (1 - \alpha') \cdot (1 - \pi)$	1

Die Bias-Verteilung entspricht Tests mit kleinerem β' und größerem α' :

$$\beta' = (1 - u) \cdot \beta$$

$$\alpha' = \alpha + u \cdot (1 - \alpha).$$

Bias

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$[(1 - \beta) + u \cdot \beta] \cdot \pi$	$(1 - u) \cdot \beta \cdot \pi$	π
falsch	(H_0)	$[\alpha + u \cdot (1 - \alpha)] \cdot (1 - \pi)$	$(1 - u) \cdot (1 - \alpha) \cdot (1 - \pi)$	$1 - \pi$
	Σ	$(1 - \beta) \cdot \pi + \alpha \cdot (1 - \pi) + u \cdot [\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)]$	$(1 - u) \cdot [\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)]$	1

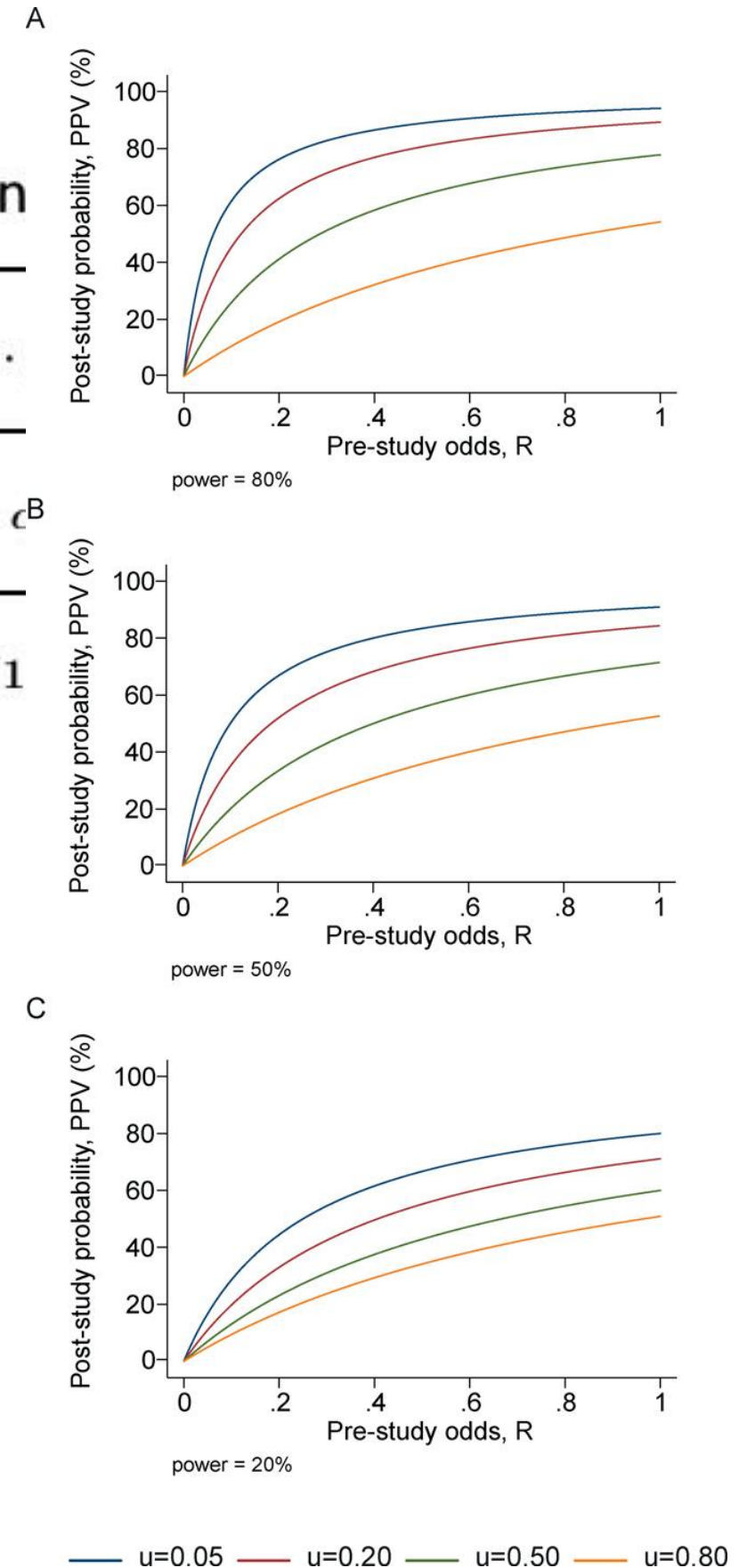
Für die Tests mit Bias ergeben sich also neue Wahrscheinlichkeiten und

$$PPV_B = \frac{[(1 - \beta) + u \cdot \beta] \cdot \pi}{(1 - \beta) \cdot \pi + \alpha \cdot (1 - \pi) + u \cdot [\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)]} = \frac{\pi}{\pi + \frac{[\alpha + u \cdot (1 - \alpha)] \cdot (1 - \pi)}{(1 - \beta) + u \cdot \beta}}$$

Wenn sonst alles gleich bleibt, gilt: $1 - \beta > \alpha \Rightarrow PPV_B \xrightarrow{u \rightarrow 1} \pi$.

Bias

Fakt	Test	signifikant	nicht sign
wahr (H_1)		$[(1 - \beta) + u \cdot \beta] \cdot \pi$	$(1 - u) \cdot \pi$
falsch (H_0)		$[\alpha + u \cdot (1 - \alpha)] \cdot (1 - \pi)$	$(1 - u) \cdot (1 - \pi)$
	Σ	$(1 - \beta) \cdot \pi + \alpha \cdot (1 - \pi) + u \cdot [\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)]$	$(1 - u) \cdot [\beta \cdot \pi + (1 - \alpha) \cdot (1 - \pi)]$



R ist ein zu π alternatives apriori-Maß:

$$R = \frac{\text{Treffer}}{\text{Niete}(n)} \Rightarrow \pi = \frac{R}{R + 1} \text{ und } R = \frac{\pi}{1 - \pi} .$$

Verschiedene Kombinationen von R , $1-\beta$ und u

Table 4. PPV of Research Findings for Various Combinations of Power ($1 - \beta$), Ratio of True to Not-True Relationships (R), and Bias (u)

$1 - \beta$	R	u	Practical Example	PPV
0.80	1:1	0.10	Adequately powered RCT with little bias and 1:1 pre-study odds	0.85
0.95	2:1	0.30	Confirmatory meta-analysis of good-quality RCTs	0.85
0.80	1:3	0.40	Meta-analysis of small inconclusive studies	0.41
0.20	1:5	0.20	Underpowered, but well-performed phase I/II RCT	0.23
0.20	1:5	0.80	Underpowered, poorly performed phase I/II RCT	0.17
0.80	1:10	0.30	Adequately powered exploratory epidemiological study	0.20
0.20	1:10	0.30	Underpowered exploratory epidemiological study	0.12
0.20	1:1,000	0.80	Discovery-oriented exploratory research with massive testing	0.0010
0.20	1:1,000	0.20	As in previous example, but with more limited bias (more standardized)	0.0015

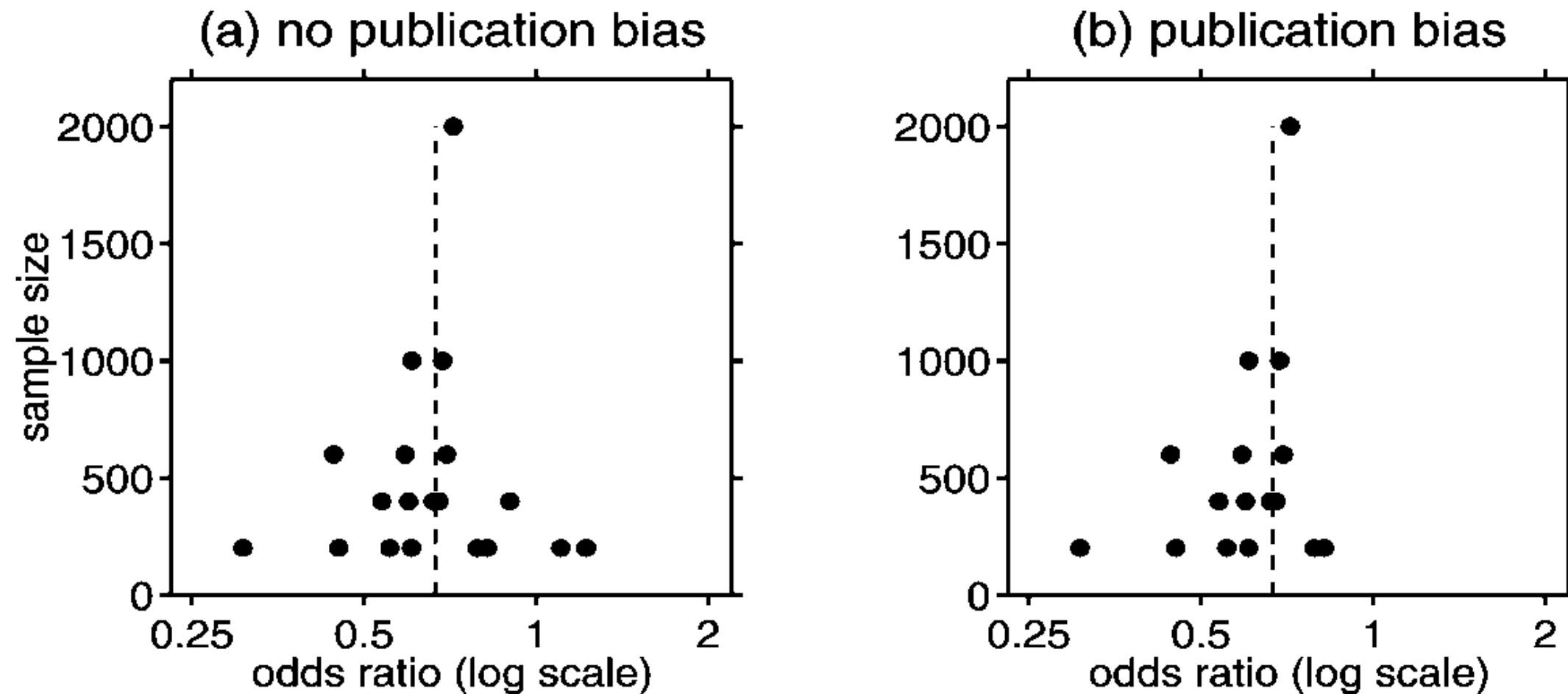
The estimated PPVs (positive predictive values) are derived assuming $\alpha = 0.05$ for a single study.

RCT, randomized controlled trial.

DOI: 10.1371/journal.pmed.0020124.t004

Publication Bias

Problem bei Meta-Studien, daß viele Studien mit nicht-signifikanten Ergebnissen nicht publiziert werden, dadurch der Gesamteindruck verfälscht. Wie erkennen?



Funnel-Diagramme

Konkurrenzkampf

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$(1 - \beta^k) \cdot \pi$	$\beta^k \cdot \pi$	π
falsch	(H_0)	$[1 - (1 - \alpha)^k] \cdot (1 - \pi)$	$(1 - \alpha)^k \cdot (1 - \pi)$	$1 - \pi$
	Σ	$(1 - \beta^k) \cdot \pi + [1 - (1 - \alpha)^k] \cdot (1 - \pi)$	$\beta^k \cdot \pi + (1 - \alpha)^k \cdot (1 - \pi)$	1

Wahrscheinlichkeiten für ein signifikantes Resultat, wenn k Teams an einer Fragestellung arbeiten. Es ergibt sich:

$$PPV_K = \frac{(1 - \beta^k) \cdot \pi}{(1 - \beta^k) \cdot \pi + [1 - (1 - \alpha)^k] \cdot (1 - \pi)} = \frac{\pi}{\pi + \frac{1 - (1 - \alpha)^k}{1 - \beta^k} \cdot (1 - \pi)}$$

Wenn sonst alles gleich bleibt, gilt: $1 - \beta > \alpha \Rightarrow PPV_K \xrightarrow{k \rightarrow \infty} \pi$.

Konkurrenzkampf

Fakt	Test		Σ
wahr (H_1)			π
falsch (H_0)		$[1 - \dots]$	$1 - \pi$
	Σ	$(1 - \beta^k) \cdot \pi + [1 - (1 - \alpha)^k] \cdot (1 - \pi)$	1

Unfortunately, in some areas, the prevailing mentality until now has been to focus on isolated discoveries by single teams and interpret research experiments in isolation. An increasing number of questions have at least one study claiming a research finding, and this receives unilateral attention. [1, S.696]

Wahrscheinlichkeiten für ein signifikantes Resultat, wenn k Teams an einer Fragestellung arbeiten. Es ergibt sich:

$$PPV_K = \frac{(1 - \beta^k) \cdot \pi}{(1 - \beta^k) \cdot \pi + [1 - (1 - \alpha)^k] \cdot (1 - \pi)} = \frac{\pi}{\pi + \frac{1 - (1 - \alpha)^k}{1 - \beta^k} \cdot (1 - \pi)}$$

Wenn sonst alles gleich bleibt, gilt: $1 - \beta > \alpha \Rightarrow \underset{k \rightarrow \infty}{PPV_K} \rightarrow \pi$.

Konkurrenzkampf

Fakt	Test	signifikant	nicht signifikant	Σ
wahr	(H_1)	$(1 - \beta^k) \cdot \pi$	$\beta^k \cdot \pi$	π
falsch	(H_0)	$[1 - (1 - \alpha)^k] \cdot (1 - \pi)$	$(1 - \alpha)^k \cdot (1 - \pi)$	$1 - \pi$
	Σ	$(1 - \beta^k) \cdot \pi + [1 - (1 - \alpha)^k] \cdot (1 - \pi)$	$\beta^k \cdot \pi + (1 - \alpha)^k \cdot (1 - \pi)$	1

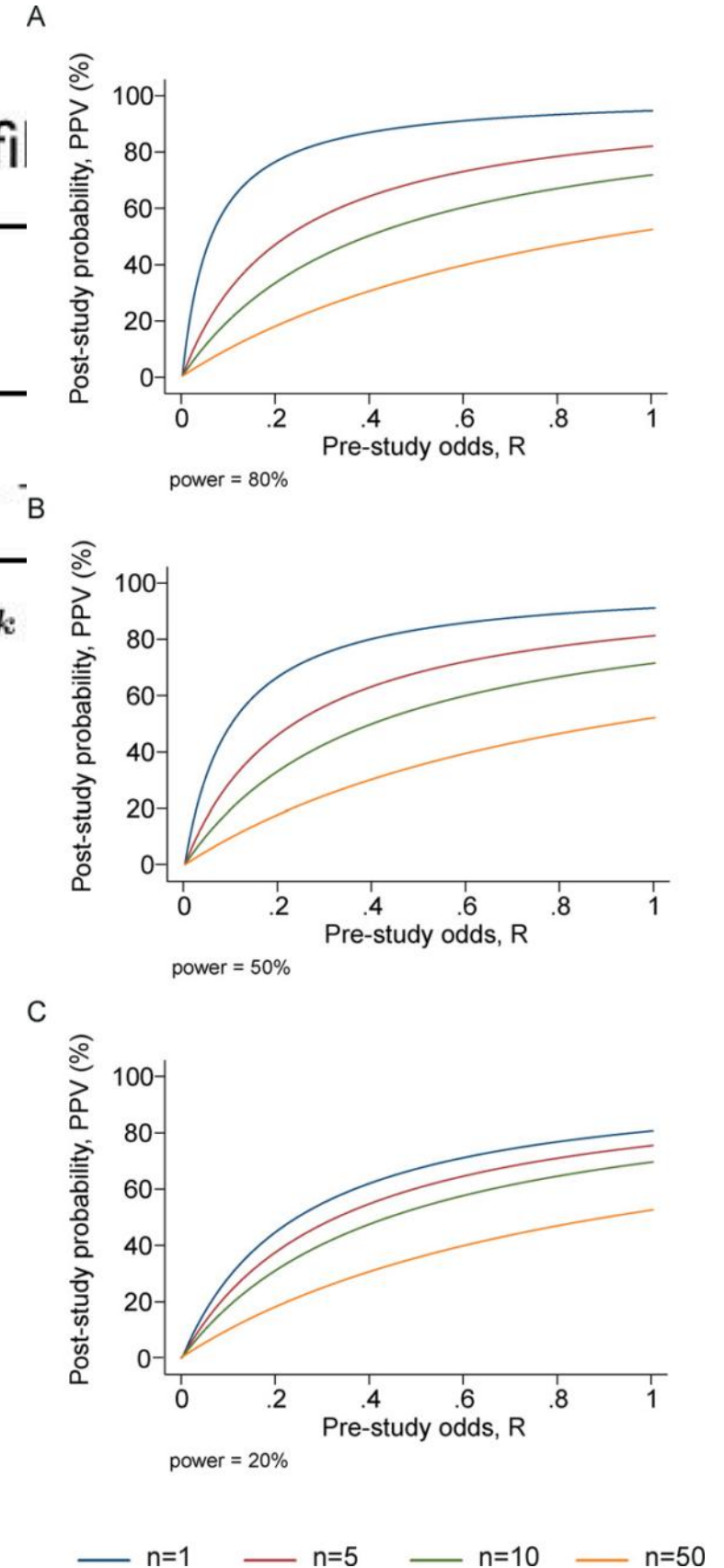
Wahrscheinlichkeiten für ein signifikantes Resultat, wenn k Teams an einer Fragestellung arbeiten. Es ergibt sich:

$$PPV_K = \frac{(1 - \beta^k) \cdot \pi}{(1 - \beta^k) \cdot \pi + [1 - (1 - \alpha)^k] \cdot (1 - \pi)} = \frac{\pi}{\pi + \frac{1 - (1 - \alpha)^k}{1 - \beta^k} \cdot (1 - \pi)}$$

Wenn sonst alles gleich bleibt, gilt: $1 - \beta > \alpha \Rightarrow PPV_K \xrightarrow{k \rightarrow \infty} \pi$.

Konkurrenzkampf

Fakt	Test	signifikant	nicht signifi
wahr (H_1)		$(1 - \beta^k) \cdot \pi$	$\beta^k \cdot \pi$
falsch (H_0)		$[1 - (1 - \alpha)^k] \cdot (1 - \pi)$	$(1 - \alpha)^k \cdot (1 - \pi)$
	Σ	$(1 - \beta^k) \cdot \pi + [1 - (1 - \alpha)^k] \cdot (1 - \pi)$	$\beta^k \cdot \pi + (1 - \alpha)^k \cdot (1 - \pi)$



Idealtypische Beispielrechnungen

1. Gentechnologie

100000 Gene, ≈ 10 Gen-Polymorphismen, $RR = 1.3$, $\alpha = 0.05$

$\Rightarrow \pi = 10^{-4}$, $1 - \beta = 0.60$ und $PPV = 12 \cdot 10^{-4}$

+ $u = 10\% \Rightarrow PPV = 4.4 \cdot 10^{-4}$

+ $k = 10 \Rightarrow PPV = 1.5 \cdot 10^{-4}$

2. Terroristen-Fahndung mit Face-Recognition an Flughäfen

Terroristen : Bürger = 1 : 10000000 $\Rightarrow \pi = 10^{-6}$

$\alpha = 0.10$, $1 - \beta = 0.50$ (technisch und sozial bedingt, beides utopisch)

$\Rightarrow PPV = 0.000005$

$n = 10^9$ Fluggäste/Jahr $\Rightarrow 10^8$ Fehllarme/Jahr

Korollare

Corollary 1: The smaller the studies conducted in a scientific field, the less likely the research findings are to be true.

Corollary 2: The smaller the effect sizes in a scientific field, the less likely the research findings are to be true.

Corollary 3: The greater the number and the lesser the selection of tested relationships in a scientific field, the less likely the research findings are to be true.

Corollary 4: The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true.

Corollary 5: The greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true.

Corollary 6: The hotter a scientific field (with more scientific teams involved), the less likely the research findings are to be true. [1.S.697ff]

Vom Kopf auf die Füße gestellt?

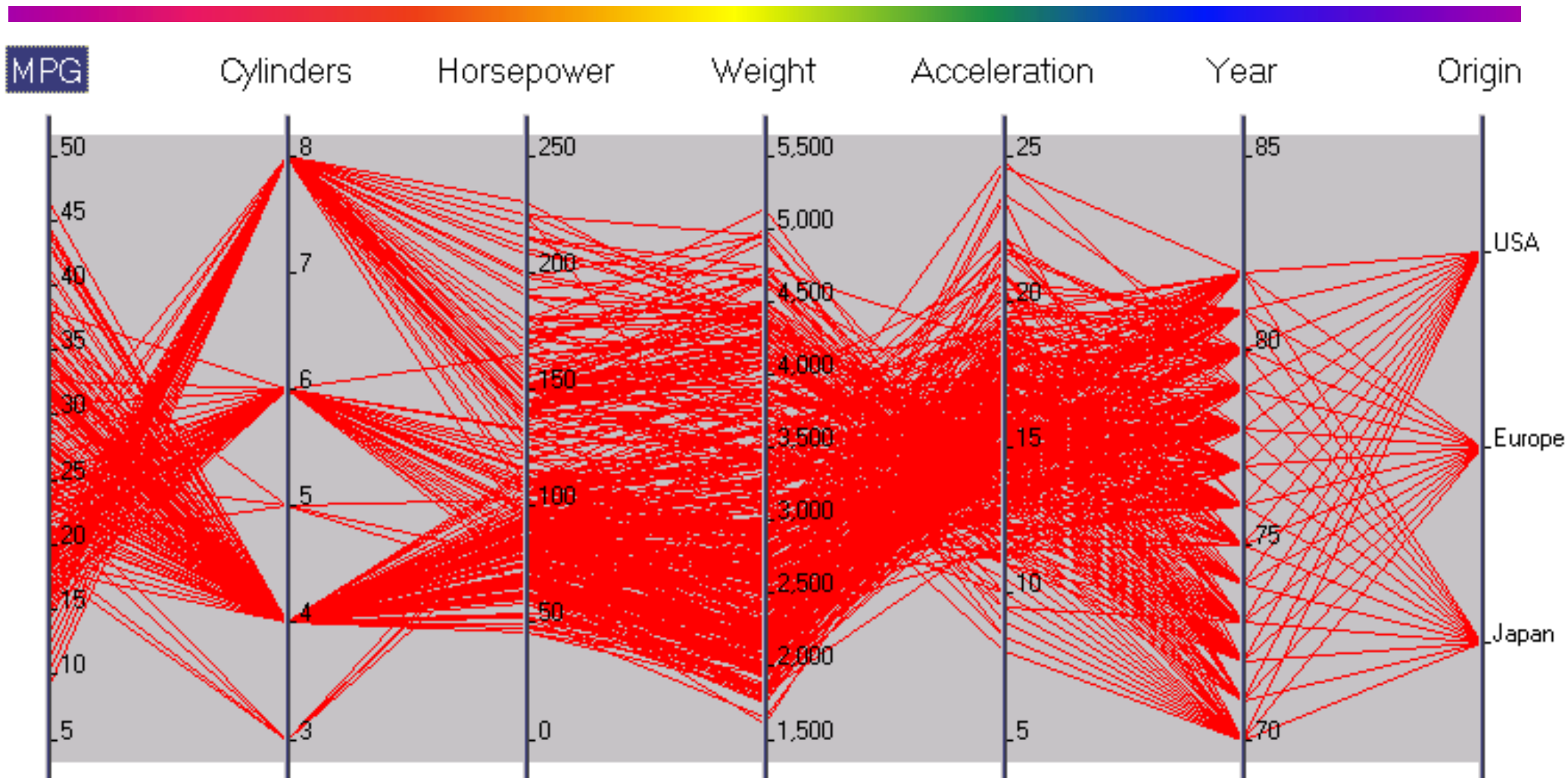
»History of science teaches us that scientific endeavor has often in the past wasted effort in fields with absolutely no yield of true scientific information, at least based on our current understanding. In such a “null field,” one would ideally expect all observed effect sizes to vary by chance around the null in the absence of bias. The extent that observed findings deviate from what is expected by chance alone would be simply a pure measure of the prevailing bias. ... For fields with very low PPV, the few true relationships would not distort this overall picture much. Even if a few relationships are true, the shape of the distribution of the observed effects would still yield a clear measure of the biases involved in the field.« [1, S.700]

→ neues Forschungsprogramm?



Visuelles Datamining – Parallelkoordinaten

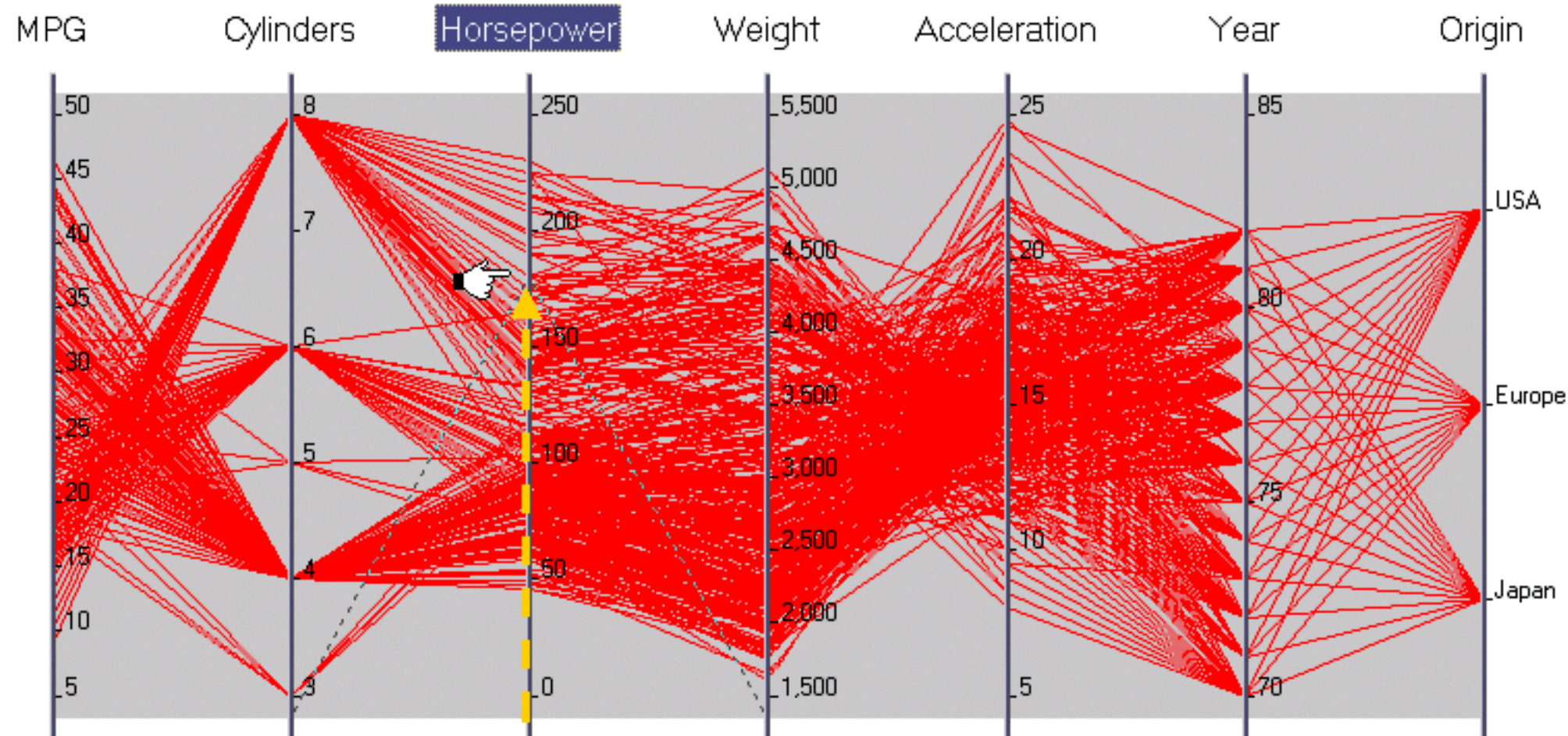
Visualization result



Each red line above indicates that a particular value category on an axis corresponds to those on its adjacent axes on a one-to-one basis.

Visuelles Datamining – Parallelkoordinaten

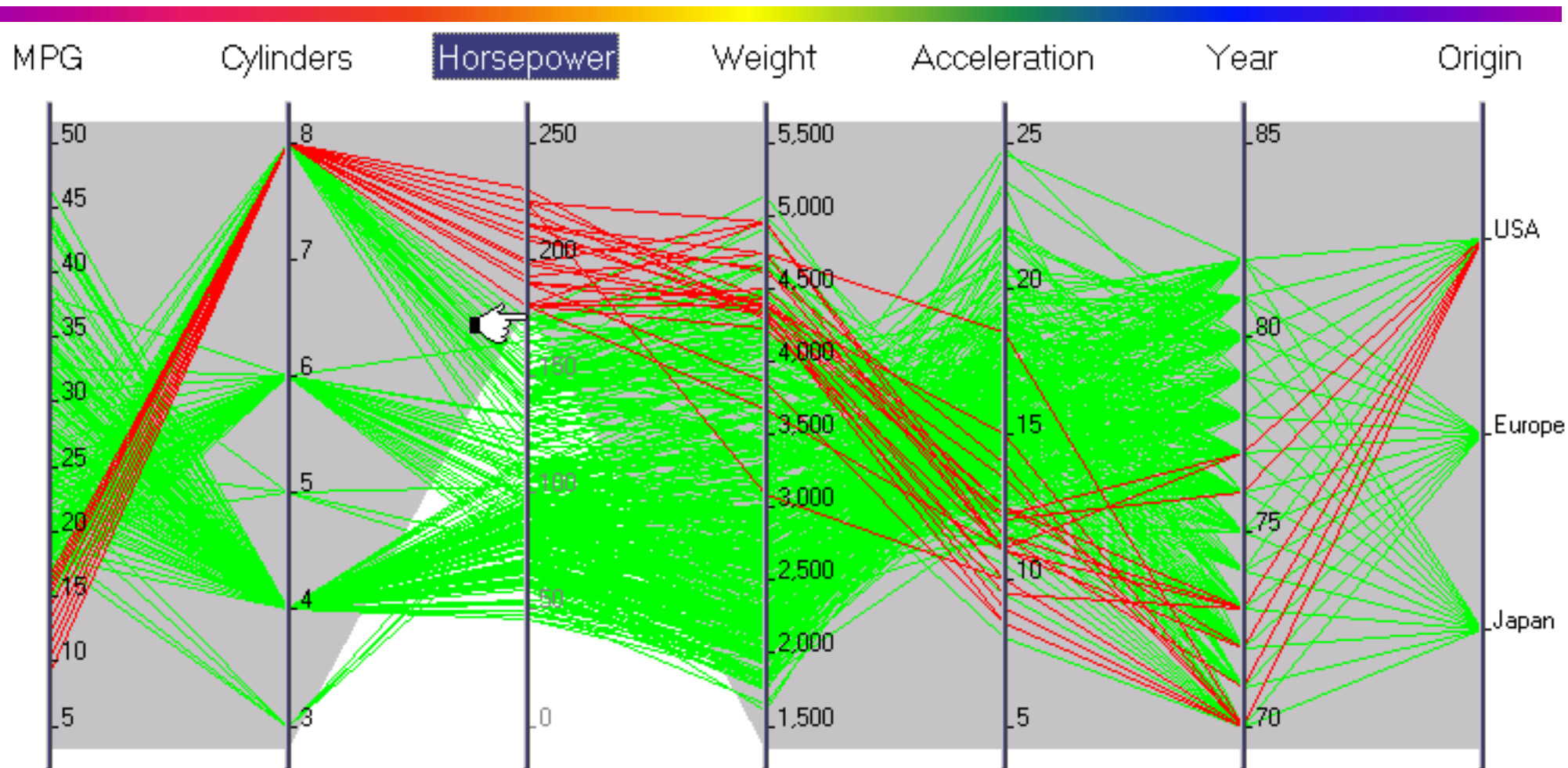
Selecting a range of records



To select high horsepower readings above 160, for example, place the cursor on the Horsepower axis below the gray rectangle (selection area), and drag the cursor, while holding it, along the axis to the point where the horsepower reading is more than 160. Then release the mouse button.

Visuelles Datamining – Parallelkoordinaten

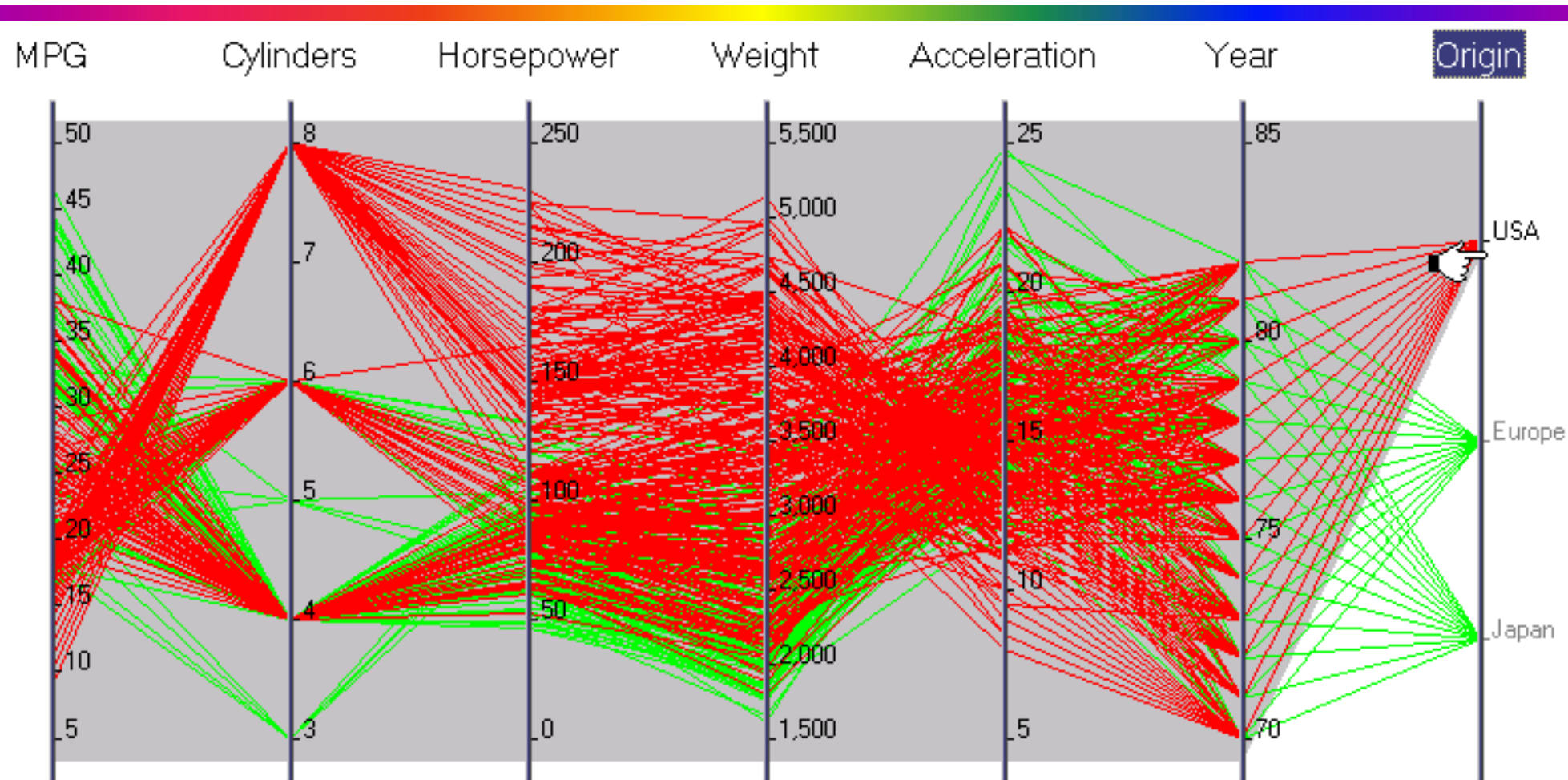
Selecting records



Let's interpret the result. The cars with high horsepower readings have 8 cylinders and they also have higher fuel consumption (low MPG). These cars are all 1970s models made in the USA. Based on this observation result, you might jump to the conclusion that all American-made cars have high-horsepower readings and low mileage per gallon. Now let's narrow down the countries of automobile origin to the USA alone.

Visuelles Datamining – Parallelkoordinaten

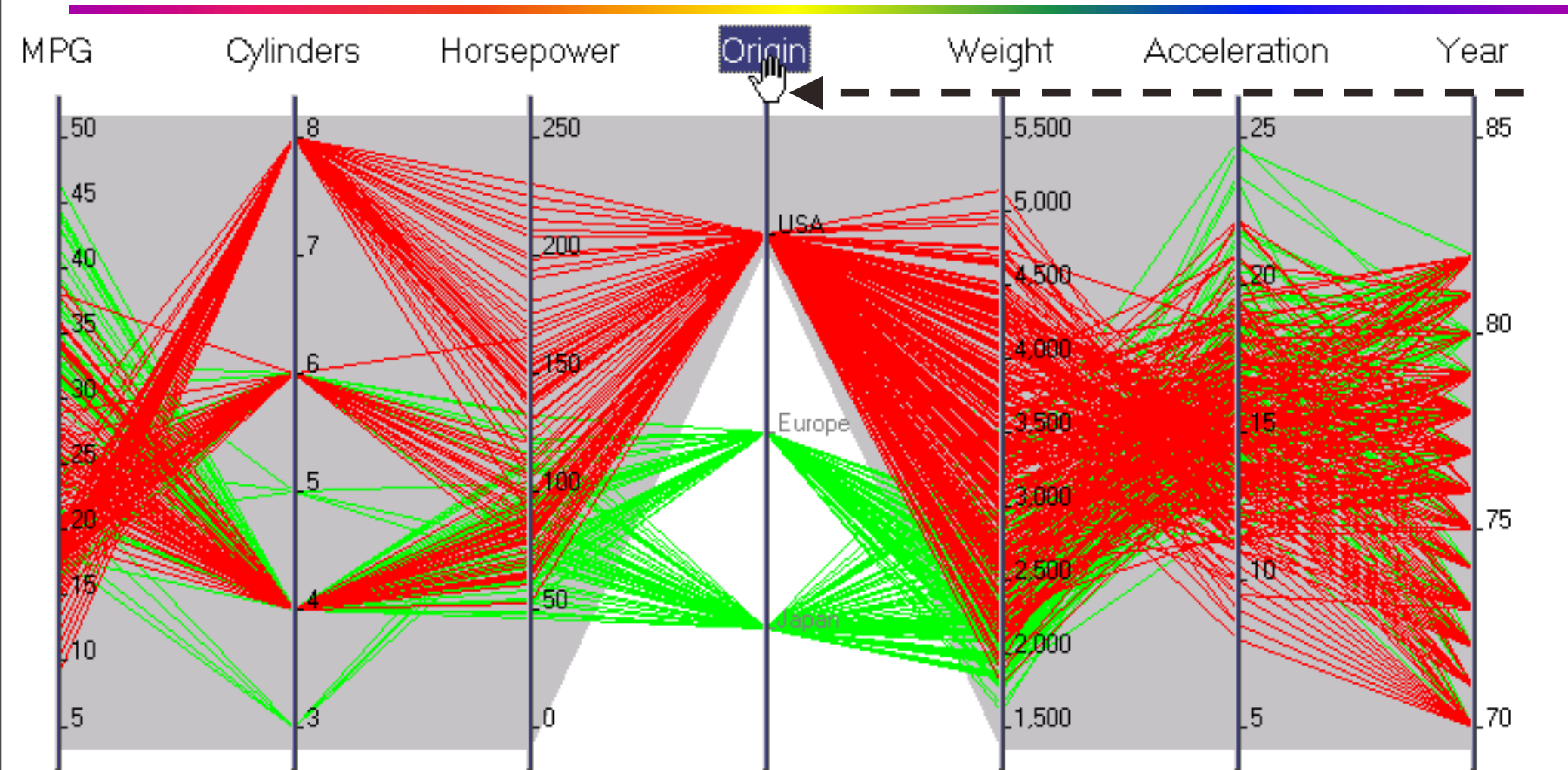
Verifying a hypothesis



The hypothesis that all American-made cars have high horsepower readings and high fuel consumption rates cannot be verified from the given dataset. To highlight the relationships between the country of origin and the horsepower, you can move the Origin axis to the right or left of the Horsepower axis with simple drag-and-drop.

Visuelles Datamining – Parallelkoordinaten

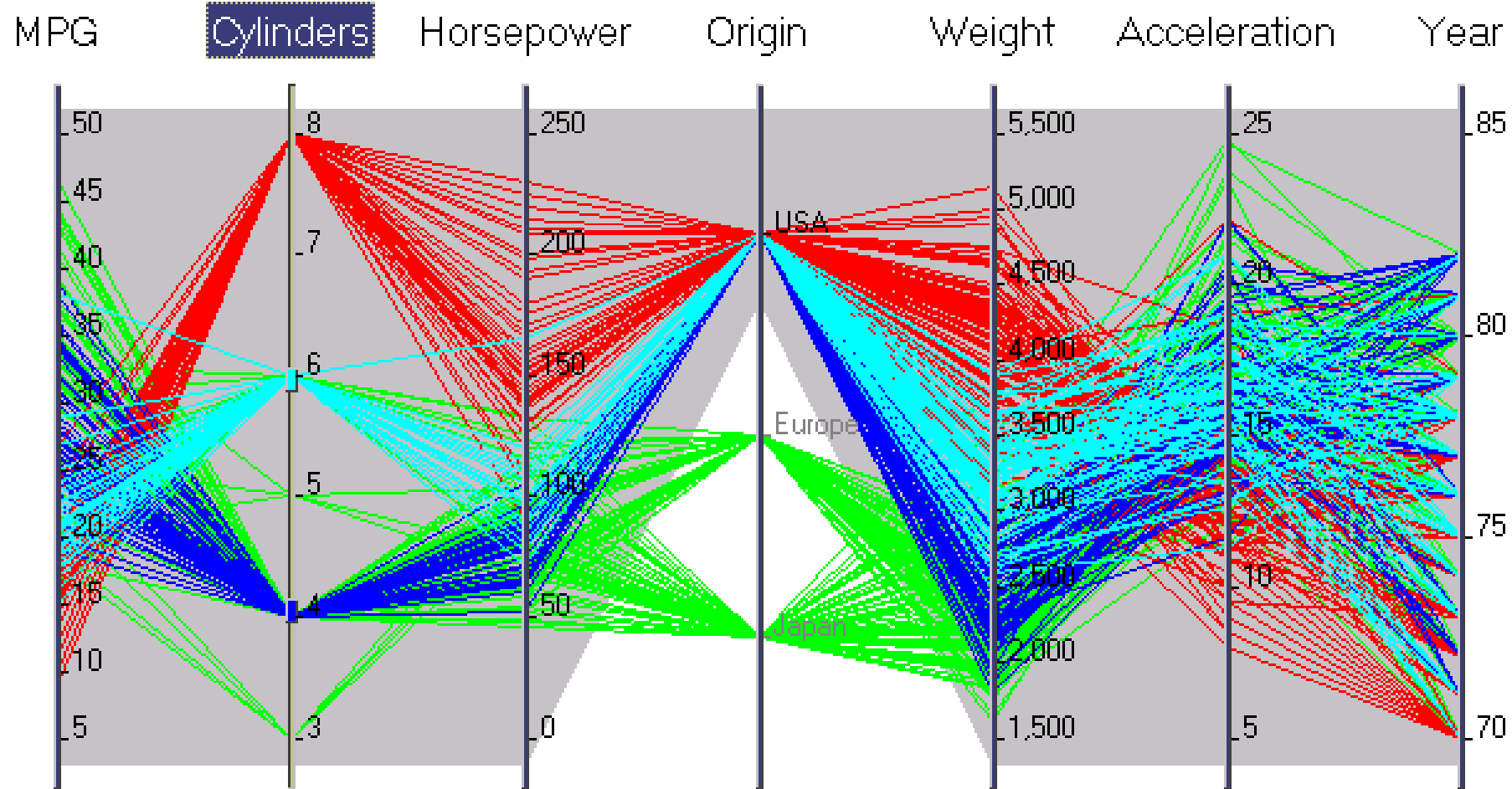
Highlighting relationships



We can now compare the relationships between the two parameters more closely. This type of verification model used to take a lot of technical efforts in database query and calculation. VisualMiner makes it easily accessible by business users with the parallel coordinates display.

Visuelles Datamining – Parallelkoordinaten

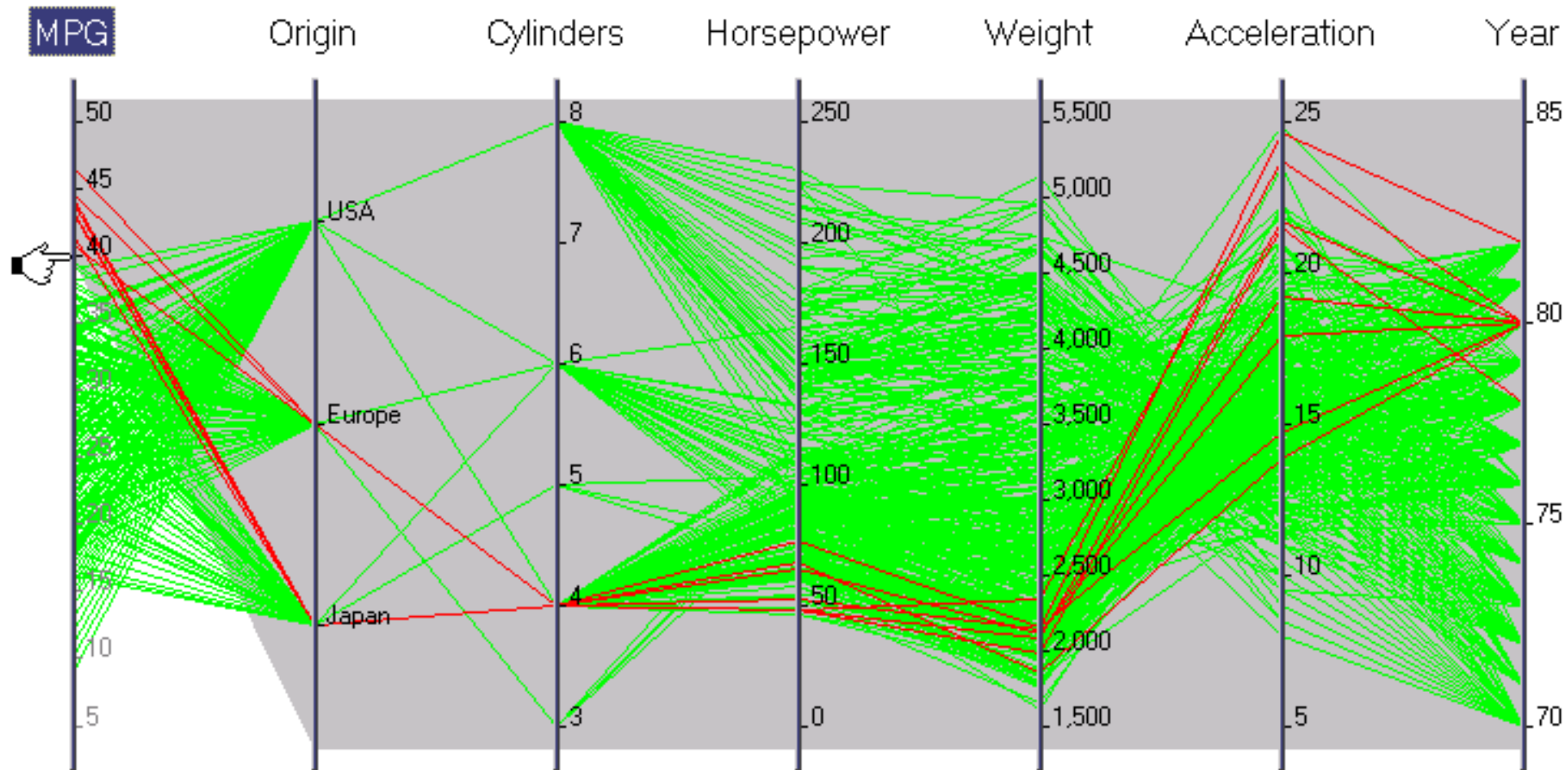
Separating different record groups



Another nice feature is to separate different record groups by assigning different colors. In this example, the red lines are for the selected records with 8 cylinders, the cyan lines are for those with 6 cylinders, and the dark blue lines are for those with 4 cylinders. The green lines represent the records that are not selected.

Visuelles Datamining – Parallelkoordinaten

Another observation



The Origin axis has moved to the right of the MPG axis and the selected MPG values are 40 or above. It is interesting to observe that cars with high MPG are made in either Japan or Europe during late 1970s and early 1980s, and they are all 4 cylinder cars with low weight and high acceleration.