

## Parameterschätzung

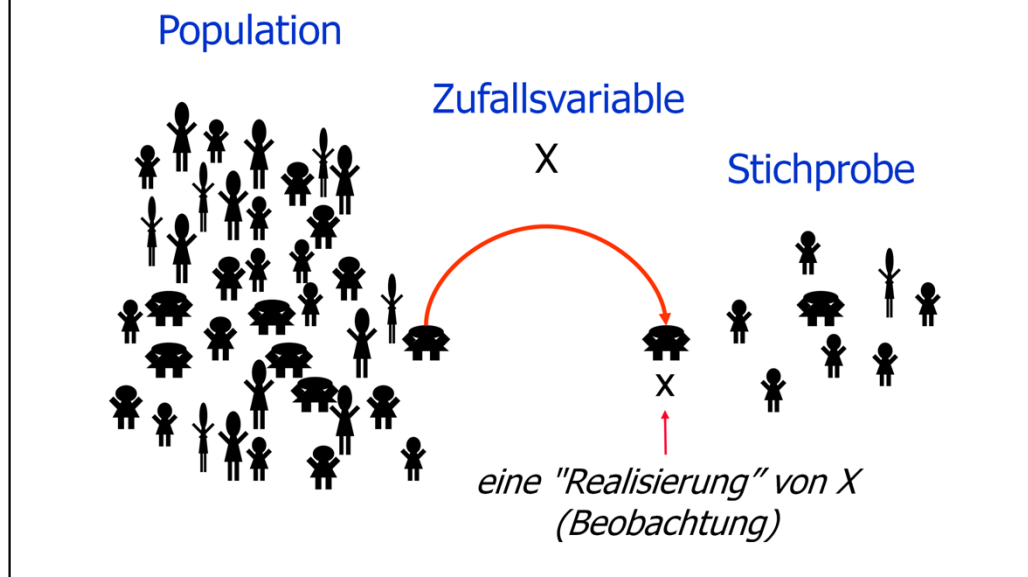


Numero, pondere et mensura Deus omnia condidit

Populationsparameter müssen immer dann geschätzt werden, wenn die jeweilige Population bezüglich des interessierenden Charakteristikums nicht vollständig untersucht werden kann. Ein vollständiger Zensus liefert demgegenüber z.B. das exakte durchschnittliche Jahreseinkommen von Familien oder das exakte Geschlechtsverhältnis oder den exakten Anteil von Ruheständlern an der Bevölkerung eines Landes. Schätzen ist auch immer dann erforderlich, wenn die Zielpopulation (zumindest im Prinzip) unendlich groß ist. So ist beispielsweise die Anzahl der wiederholten Würfe eines einzelnen Würfels potenziell unbegrenzt, was bedeutet, dass deren erwartete Augenzahl nur geschätzt, aber niemals exakt bestimmt werden kann.

Die Frage, ob eine einmalige Schätzung "gut" oder "schlecht" ist, lässt sich solange nicht beantworten, wie der wahre Wert des zu schätzenden Parameters unbekannt ist. In der wissenschaftlichen Forschung können deshalb nur Regeln (so genannte "Schätzer") dafür angegeben werden, wie ein Parameter aus einer Stichprobe geschätzt werden sollte. Im Zusammenhang mit Parameterschätzungen ist "Qualität" also eine Eigenschaft von Regeln, nicht von Ergebnissen, die durch die Anwendung dieser Regeln erzielt werden.

## Population, Zufallsvariable, Stichprobe



Im Kontext wissenschaftlicher Forschung versteht man unter einem "Populationsparameter" eine bestimmte Eigenschaft der Verteilung einer Zufallsvariablen. Parameter von praktischem Interesse sind im Wesentlichen Wahrscheinlichkeiten (für qualitative Zufallsvariable) sowie Erwartungswerte, Quantile und Varianzen (für quantitative Zufallsvariable). Die Schätzung eines Populationsparameters erfolgt aus einer Stichprobe, worunter man in der Statistik eine Menge von Realisierungen der zu untersuchenden Zufallsvariablen versteht.



alle männlichen Rekruten der US Armee  
(Population)

die ersten 10 männlichen Wehrpflichtigen, die am 11. Mai 2004  
das Rekrutierungsbüro der US Armee in Concord NH betreten  
(Stichprobe)

BMI eines zu einem zufälligen Zeitpunkt in einem zufällig  
gewählten Rekrutierungsbüro anzutreffenden Rekruten  
(Zufallsvariable)

## Einige Konventionen

Zufallsvariable werden mit **Großbuchstaben** bezeichnet.

Mögliche Werte oder tatsächliche Beobachtungen ("Realisierungen") werden mit **kleinen Buchstaben** bezeichnet.

$$P(X \leq x)$$

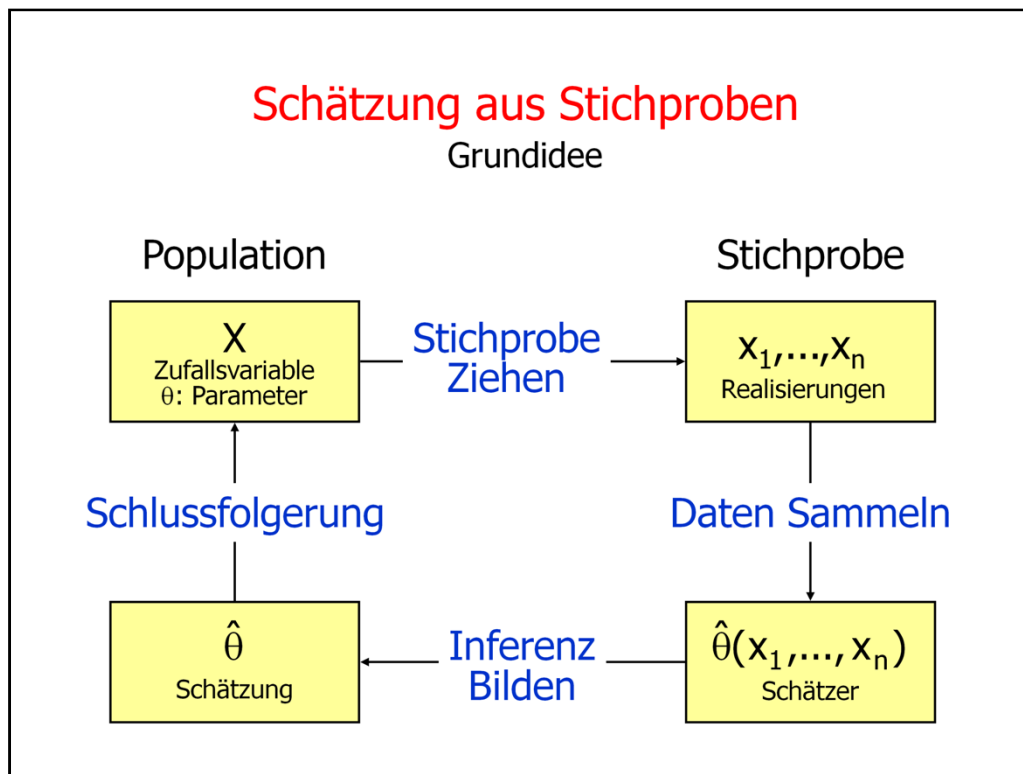
Wahrscheinlichkeit, dass der BMI eines zufällig ausgewählten Rekruten höchstens den Wert  $x$  annimmt

$$P(X > 18)$$

Wahrscheinlichkeit, dass ein zufällig ausgewählter Rekrut nicht untergewichtig ist

$$P(24 < X < 30)$$

Wahrscheinlichkeit, dass ein zufällig ausgewählter Rekrut übergewichtig, aber nicht fettleibig ist



Dieses Diagramm spiegelt ein wiederkehrendes Thema wissenschaftlicher Forschung wider, nämlich den Versuch, Inferenz hinsichtlich eines Populationscharakteristikums auf der Grundlage einer Stichprobe aus eben dieser Population zu bilden. Im Zusammenhang mit Schätzungen von Parametern heißt das, aus der Stichprobe einen "guten Anhalt" für den fraglichen Parameter zu bekommen.

Es gibt viele Wege, um von einer Stichprobe zur Schätzung eines Parameters zu gelangen. Als Schätzung des Erwartungswertes könnte z.B. der größte Wert der Stichprobe dienen oder der kleinste oder der als letztes gezogene. Statt sich aber auf den gesunden Menschenverstand, Eingebung oder Glück zu verlassen, verlangen Wissenschaftler klare Regeln, wie sie sinnvolle Schätzungen aus ihren Daten gewinnen sollen. Diese Regeln heißen "Schätzer". Schätzer sind im Grunde Kochrezepte, die angeben, wie man aus guten Zutaten (Daten oder Realisierungen) ein gutes Essen (d.h. Schätzungen) zubereitet.

## Schätzung aus Stichproben

### Grundidee

Parameter $\theta$	Beobachtungen $x_1, \dots, x_n$	Schätzer $\hat{\theta}(x_1, \dots, x_n)$
$\pi$ Wahrscheinlichkeit	0,0,1,1,0,1,...	$\hat{\pi} = k / n$ Anteil
$\mu$ Erwartungswert	1.23,4.81,7.55,...	$\hat{\mu} = \bar{x}$ Stichprobenmittel
$\sigma^2$ Varianz	12.4,19.6,20.4,...	$\hat{\sigma}^2 = s^2$ Stichprobenvarianz

In der Praxis konzentriert sich das wissenschaftliche Interesse im Wesentlichen auf drei Arten von Parametern: Wahrscheinlichkeiten, Erwartungswerte und Varianzen. Die besten Schätzer hierfür sind Anteile (d.h. relative Häufigkeiten), Stichprobenmittel und Stichprobenvarianzen. Punkt!

Der Rest der Vorlesung dient lediglich der Erläuterung, warum es sich bei den genannten Schätzern um gute Schätzer handelt, und insbesondere welche Kriterien in der Statistik zur Bewertung der Qualität eines Schätzers herangezogen werden.

## Werfen einer Münze



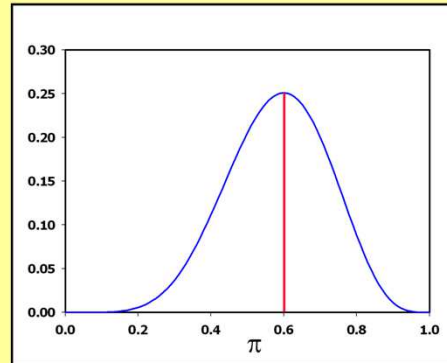
$\pi$ : Wahrscheinlichkeit für Kopf

$$\hat{\pi} = \frac{k}{n} = \frac{6}{10} = 0.6$$

X: Anzahl Kopf in 10 Würf

X hat eine  $\text{Bin}(\pi, 10)$  Verteilung

$$P_{\pi}(X = 6) = 210 \cdot \pi^6 \cdot (1 - \pi)^4$$



Selbst ein statistischer Laie würde die gleiche Antwort geben: Wenn eine Münze sechs mal unter 10 Würf Kopf gezeigt hat, dann beträgt die Wahrscheinlichkeit für Kopf im nächsten Wurf eben 0.6. Wie wir gleich sehen werden, folgt diese "Schätzung" unbewusst einem grundlegenden Prinzip der Inferenzbildung.

Der Münzwurf ist ein typisches Binomialexperiment, d.h. es gibt 10 unabhängige Versuche, die jeweils mit Erfolgswahrscheinlichkeit  $\pi$  durchgeführt werden. Die Wahrscheinlichkeit für sechs Erfolge bei 10 Versuchen kann man also aus der Wahrscheinlichkeitsfunktion der Binomialverteilung berechnen:

$$P_{\pi}(X = 6) = 210 \cdot \pi^6 \cdot (1 - \pi)^4.$$

Dieser Zusammenhang zwischen  $\pi$  und  $P_{\pi}(X=6)$  lässt sich auch als mathematische Funktion auffassen und in Form einer Kurve graphisch darstellen. Beachten Sie, dass der Gipfel der Kurve genau bei  $\pi=0.6$  liegt. Aber warum?

## Likelihood und Wahrscheinlichkeit

Die **Likelihood** eines Parameters,  
gegeben die Beobachtungen,  
ist die **Wahrscheinlichkeit** der Beobachtungen,  
gegeben der Parameter.

$$L(\theta | x) = P_{\theta}(X = x)$$

In seinem 1921 verfassten Artikel "On the 'probable error' of a coefficient of correlation deduced from a small sample" machte Ronald A. Fisher die folgende Bemerkung: "Was wir einer Stichprobe entnehmen können, ist die Mutmaßlichkeit (engl. likelihood) eines jeden Wertes [eines Parameters], wenn wir die Mutmaßlichkeit (von nun an "Likelihood") proportional zur Wahrscheinlichkeit definieren, mit der man aus einer Population mit diesem Wert [des Parameters] die beobachteten Daten erhält. So definiert sind Wahrscheinlichkeit und Likelihood zwei Zahlen mit vollkommen unterschiedlicher Bedeutung."

Der Hauptunterschied zwischen Wahrscheinlichkeit und Likelihood besteht also darin, dass bei der Likelihood der Parameter (oder später ein ganzes statistisches Modell) die veränderliche Größe ist, während die eigentlichen Beobachtungen festliegen. Gegenstand einer Wahrscheinlichkeitsaussage ist demgegenüber eine Vielzahl möglicher Ausgänge, und der oder die Parameter liegen fest. So wird z.B. der Hersteller eines fabrikneuen Würfels verständlicher Weise behaupten, dass mit diesem Würfel alle Augenzahlen die gleiche Wahrscheinlichkeit von 1/6 haben. Ein Kasinoinspektor wird demgegenüber die Likelihood unterschiedlicher Kombinationen von Wahrscheinlichkeiten für die einzelnen Augenzahlen anhand einer tatsächlichen Beobachtungsreihe berechnen und bewerten.



## Das Maximum-Likelihood-Prinzip

Auf der Basis von Beobachtungen wird ein Parameter durch dessen mutmaßlichsten Wert geschätzt, d.h. durch den Wert, der die Wahrscheinlichkeit der Beobachtungen maximiert.

D.h.  $\hat{\theta}$  wird so gewählt, dass

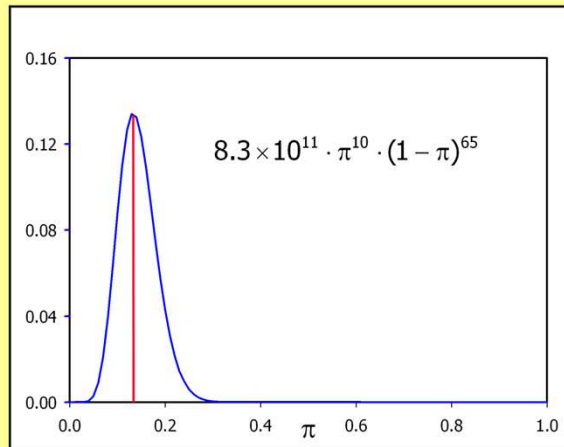
$$L(\hat{\theta} | x) = \max_{\theta} L(\theta | x)$$

Um das einfache Prinzip der Maximum-Likelihood-Schätzung noch einmal zu wiederholen: Schätze einen Parameter durch den Wert, für den die gemachten Beobachtungen am wahrscheinlichsten sind, und der somit angesichts der gemachten Beobachtungen der mutmaßlichste ist.

## AB0-Blutgruppen

In einer Stichprobe von 75 Individuen aus einer bestimmten Population hatten 10 Personen die Blutgruppe B. Wie groß ist die Häufigkeit  $\pi$  der Blutgruppe B in dieser Population?

$$\hat{\pi} = \frac{10}{75} = 0.133$$



## Das Maximum-Likelihood-Prinzip

### Binomialverteilung

$$L(\pi | k) = \binom{n}{k} \cdot \pi^k \cdot (1 - \pi)^{n-k} \quad \longrightarrow$$

$$\log\{L(\pi | k)\} = \text{const} + k \cdot \log(\pi) + (n - k) \cdot \log(1 - \pi)$$

$$\frac{\delta \log\{L(\pi | k)\}}{\delta \pi} = \frac{k}{\pi} - \frac{n - k}{1 - \pi} = 0 \quad \longleftrightarrow \quad \pi = \frac{k}{n}$$

$\hat{\pi} = k / n$  ist der Maximum-Likelihood-Schätzer von  $\pi$

Für qualitative Zufallsvariable liefert der relative Anteil eines bestimmten Ausgangs an einer Stichprobe den Maximum-Likelihood-Schätzer für dessen Wahrscheinlichkeit. Diese Übereinstimmung lässt sich mit mathematischen Standardmethoden relativ einfach beweisen.

Die Maximum-Likelihood-Schätzung von  $\pi$  bezeichnet den Wert von  $\pi$ , der die Likelihood-Funktion  $L(\pi|k)$  maximiert. Hierbei ist  $k$  die beobachtete Anzahl der Erfolge unter  $n$  Versuchen. Wie können wir diesen Wert von  $\pi$  berechnen? Nun, wir brauchen dafür nur  $L(\pi|k)$  einmal nach  $\pi$  abzuleiten und diese neue Funktion (die erste Ableitung also) gleich null zu setzen (Hinweis: Das ist so, weil eine Funktion dort ihr Maximum annimmt, wo sie weder steigt noch fällt, d.h. wo die Steigung gleich null ist). Das Lösen der genannten Gleichung ist jedoch einfacher, wenn man statt  $L(\pi|k)$  den Logarithmus von  $L(\pi|k)$  maximiert. Dadurch wird das Endergebnis nicht verändert, da die Likelihood und der log-Likelihood ihr Maximum beim gleichen Wert von  $\pi$  haben. Die erste Ableitung von  $\log L(\pi|k)$  sieht jedoch sehr einfach aus:

$$\frac{k}{\pi} - \frac{n - k}{1 - \pi},$$

und man kann leicht zeigen, dass diese tatsächlich für  $\pi = k/n$  den Wert null annimmt.

### Weiblicher Body-Mass-Index (BMI)

Welchen Erwartungswert  $\mu$   
hat der BMI einer US  
Schönheitskönigin?

Liegt  $\mu$  ungefähr bei  
 $\bar{x} = 18.6$ ?

Jahr	Name	BMI
1984	Suzette Charles	17.7
1985	Sharlene Wells	18.2
1986	Susan Akin	16.8
1987	Kellye Cash	17.6
1988	Kaye Lani Rae Rafko	18.8
1989	Gretchen Carlson	19.1
1990	Debbye Turner	17.9
1991	Marjorie Vincent	17.8
1998	Kate Shindle	20.2
1999	Nicole Johnson	19.6
2001	Angela Perez Baraquio	20.3
2002	Katie Harman	19.5

Der Durchschnitt der 12 BMI-Werte in der vorliegenden Tabelle beträgt 18.6. Handelt es sich dabei um eine gute Schätzung des Erwartungswerts für den BMI irgendeiner zufällig ausgewählten US-amerikanischen Schönheitskönigin? Die Antwort lautet "ja, aber".

Wenn es überhaupt so etwas wie "die amerikanische Schönheitskönigin" gibt, d.h. wenn die Miss Americas hinsichtlich ihrer äußeren Erscheinung eine homogene Population bilden, dann gibt der Wert 18.6 in der Tat einen guten Eindruck davon, wo der BMI in 2003 oder 2005 ungefähr gelegen haben sollte. Allerdings scheinen die BMI in der Tabelle einen Trend in Richtung höherer Werte gegen Ende der 1990-er Jahre aufzuweisen. Wenn es sich dabei um ein reales Phänomen handelt, das z.B. auf einer Veränderung der generellen Wahrnehmung von Schönheit und Gesundheit beruhen könnte, dann wäre unsere ursprüngliche Homogenitätsannahme nicht gerechtfertigt. In diesem Fall müsste man kompliziertere statistische Methoden anwenden (die allerdings jenseits des Rahmens dieser Vorlesung liegen), um zu vernünftigen Schätzungen der für 2003 oder 2005 zu erwartenden BMI-Werte zu gelangen.

## Das Maximum-Likelihood-Prinzip

Normalverteilung

$$L(\mu \mid x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_i - \mu)^2}{\sigma^2}} \quad \longrightarrow$$

$$\log\{L(\mu \mid x_1, \dots, x_n)\} = \text{const} - \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

$$\frac{\delta \log\{L(\mu \mid x_1, \dots, x_n)\}}{\delta \mu} = -\frac{2}{\sigma^2} \left( n \cdot \mu - \sum_{i=1}^n x_i \right) = 0$$

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

$\hat{\mu} = \bar{x}$  ist der Maximum-Likelihood-Schätzer von  $\mu$

Durch Ableiten des Logarithmus der Dichtefunktion und durch Gleichsetzen der Ableitung mit null kann man für eine normalverteilte Zufallsvariable  $X$  leicht zeigen, dass das Stichprobenmittel der Maximum-Likelihood-Schätzer des Erwartungswertes von  $X$  ist.

## Schätzer als Zufallsvariable

Wegen der zufälligen Natur von Stichproben ist jeder Schätzer  $\hat{\theta}$  selbst eine Zufallsvariable mit Erwartungswert  $E(\hat{\theta})$  und Varianz  $\text{Var}(\hat{\theta})$ .

$$\sqrt{\text{Var}(\hat{\theta})}$$

wird als "Standardfehler" von  $\hat{\theta}$  bezeichnet.

Da das Ergebnis eines Schätzvorgangs sowohl zufällig als auch variabel ist, bildet jeder Schätzer selbst wieder eine Zufallsvariable mit zugehörigem Erwartungswert ("Wo liegt die Schätzung im Mittel?") und zugehöriger Varianz ("Um wieviel streut die Schätzung bei mehrfacher Wiederholung des Stichprobenziehens und des anschließenden Schätzens?").

Die Standardabweichung eines Schätzers wird als dessen "Standardfehler" bezeichnet.

### Verteilung des Stichprobenmittels

Seien  $X_1, \dots, X_n$  unabhängig und identisch verteilt mit  $E(X_i) = \mu$  und  $\text{Var}(X_i) = \sigma^2$ .

$$E(\bar{X}) = E\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \frac{1}{n} \cdot \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n \cdot \mu = \mu$$

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \cdot \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

$$\text{Standardfehler: } \frac{\sigma}{\sqrt{n}}$$

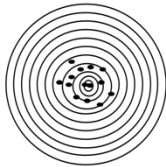
Es lässt sich leicht zeigen, dass der Erwartungswert des Stichprobenmittels gleich dem Erwartungswert der ursprünglichen Zufallsvariablen ist, dem Parameter also, den man schätzen möchte. Mit anderen Worten liefert das Stichprobenmittel "im Durchschnitt" den richtigen Parameterwert. Einen Schätzer mit dieser höchst wünschenswerten Eigenschaft nennt man "unverzerrt" (engl. unbiased).

Die Varianz des Stichprobenmittels ist gleich der ursprünglichen Varianz, geteilt durch den Stichprobenumfang. Das bedeutet, dass das Stichprobenmittel zwischen mehrfachen Wiederholungen ein und des selben Experiments um so weniger streut, je größer die zugehörige Stichprobe ist. Theoretisch würden die Stichprobenmittel sogar konstant ausfallen (d.h. sie hätten eine Varianz von null), wenn die Stichprobe jeweils unendlich groß wäre. Ein Schätzer mit dieser nützlichen Eigenschaft heißt "konsistent".

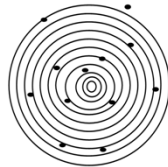
## Genauigkeit und Präzision

**Genauigkeit** bezieht sich auf die Differenz zwischen dem Erwartungswert eines Schätzers und dem wahren Parameter.

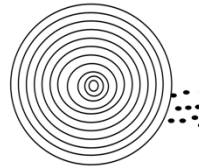
**Präzision** bezieht sich auf die Varianz eines Schätzers.



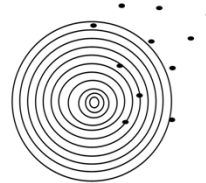
genau  
präzise



genau  
nicht präzise



nicht genau  
präzise



nicht genau  
nicht präzise



## "Gute" Schätzer

Ein guter Schätzer ist

**unverzerrt:**  $E(\hat{\theta}) = \theta$

"100% genau, d.h. er liefert im Durchschnitt den wahren Parameter"

**konsistent:**  $P(|\hat{\theta}_n - \theta| > \varepsilon) \xrightarrow{n} 0$

"liefert mit wachsendem Stichprobenumfang immer genauere und präzisere Schätzungen, die dem wahren Parameter zustreben"

**effizient:**  $\text{Var}(\hat{\theta})$  minimal

"kein anderer unverzerrter (d.h. 100% genauer) Schätzer liefert präzisere Schätzungen"

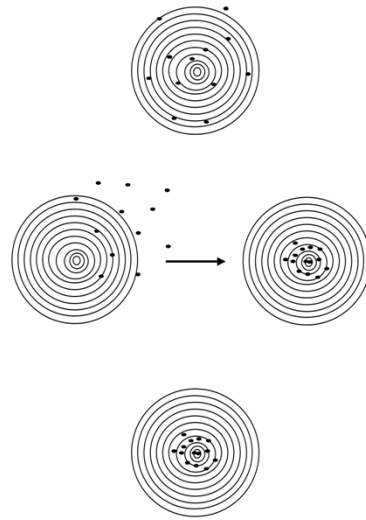
Die dritte Eigenschaft, die neben Unverzerrtheit und Konsistenz über die Nützlichkeit eines Schätzers entscheidet, ist dessen Effizienz. Für einen Statistiker ist es in der Regel einfach, die Unverzerrtheit und Konsistenz eines neu entwickelten Schätzers nachzuweisen. Ungleich schwieriger ist meistens der Beweis, dass der Schätzer auch "effizient" ist, d.h. dass es keinen anderen unverzerrten Schätzer mit kleinerer Varianz gibt.

## "Gute" Schätzer

unverzerrt (100% genau)

konsistent (Genauigkeit und Präzision streben mit wachsendem Umfang der Stichprobe 100% zu)

effizient (der präziseste Schätzer unter allen 100% genauen Schätzern)



## "Gute" Schätzer

Maximum-Likelihood

Maximum-Likelihood-Schätzer sind generell

konsistent

asymptotisch effizient

aber NICHT immer

unverzerrt

Wenn man die Kriterien "unverzerrt", "konsistent" und "effizient" anlegt, so erweisen sich die meisten Maximum-Likelihood-Schätzer als "gute" Schätzer. Es gibt aber keine generelle Garantie dafür, dass Maximum-Likelihood-Schätzer alle drei Eigenschaften gleichzeitig erfüllen. Während sie immer konsistent sind, gilt ihre Effizienz im Allgemeinen nur asymptotisch, d.h. sie sind nur dann am präzisesten unter den 100% genauen Schätzern, wenn die zugehörige Stichprobe sehr groß ist. Unglücklicherweise sind viele Maximum-Likelihood-Schätzer nicht unverzerrt. Dieser Nachteil lässt sich jedoch meistens durch nachträgliche Korrektur (engl. bias correction) der Schätzung wettmachen.

## Unverzerrte Schätzer

Wahrscheinlichkeit

X habe eine  $\text{Bin}(\pi, n)$  Verteilung

$$\hat{\pi} = \frac{k}{n} \quad \longrightarrow$$

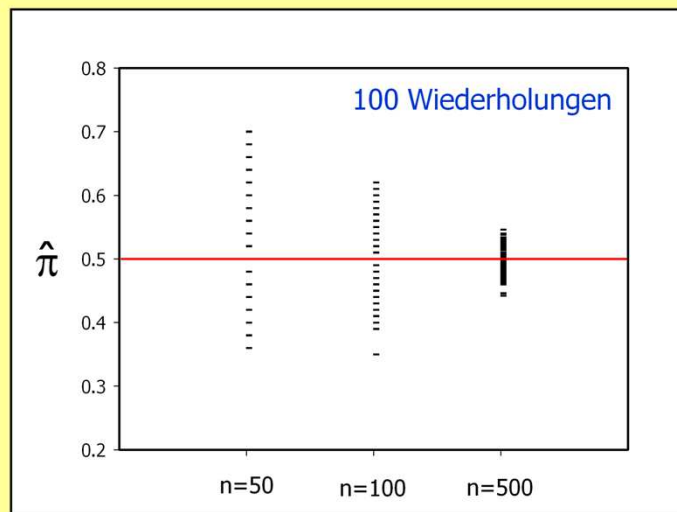
$$\begin{aligned} E(\hat{\pi}) &= E\left(\frac{X}{n}\right) = \frac{1}{n} \cdot E(X) = \frac{1}{n} \cdot \sum_{k=0}^n k \cdot \binom{n}{k} \cdot \pi^k \cdot (1-\pi)^{n-k} = \dots \\ &= \frac{1}{n} \cdot n \cdot \pi = \pi \end{aligned}$$

$\hat{\pi} = k / n$  ist ein unverzerrter Schätzer von  $\pi$

Es überrascht nicht, dass Anteile bzw. relative Häufigkeiten unverzerrte Schätzer von Wahrscheinlichkeiten sind.

## Werfen einer Münze

$\hat{\pi}$  : Anteil von Kopf unter n Würf



## Unverzerrte Schätzer

Erwartungswert

$X_1, \dots, X_n$  seien identisch verteilt mit  $E(X_i) = \mu$

$$\hat{\mu} = \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \longrightarrow$$

$$E(\hat{\mu}) = E\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \frac{1}{n} \cdot \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n \cdot \mu = \mu$$

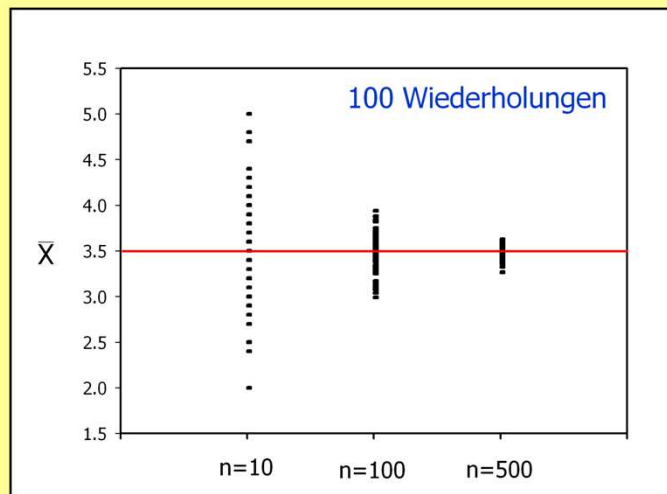
$\hat{\mu} = \bar{x}$  ist ein unverzerrter Schätzer von  $\mu$

Wie wir bereits gesehen haben, ist das Stichprobenmittel ein unverzerrter Schätzer des Erwartungswertes einer Zufallsvariablen  $X$ . Das ist immer so und hängt nicht von der Verteilung von  $X$  ab, da der notwendige Beweis nur auf allgemein gültige Rechenregeln für Erwartungswerte zurückgreift.

### Würfelspiel

$X_i$ : Augenzahl eines einzelnen Wurfs ( $i=1,\dots,n$ )

$\bar{X}$ : durchschnittliche Augenzahl aus  $n$  Würfeln



Diese Abbildung verdeutlicht die Unverzerrtheit des Stichprobenmittels, da die durchschnittliche Augenzahl aus einer wechselnden Anzahl von Würfeln immer bei 3.5 lokalisiert ist, also beim wahren Erwartungswert.

## Unverzerrte Schätzer

### Varianz

$X_1, \dots, X_n$  seien unabhängig und identisch verteilt  
mit  $\text{Var}(X_i) = \sigma^2$

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

$$\begin{aligned} E(\hat{\sigma}^2) &= E\left(\frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X})^2\right) = \dots \\ &= \frac{1}{n-1} \cdot (n-1) \cdot [E(X_1^2) - \mu^2] = \sigma^2 \end{aligned}$$

$\hat{\sigma}^2 = s^2$  ist ein unverzerrter Schätzer von  $\sigma^2$

Die Stichprobenvarianz ist ein unverzerrter Schätzer der Varianz einer Zufallsvariablen. Beachten Sie jedoch, dass im Nenner von  $s^2$  nicht der Stichprobenumfang  $n$  steht, sondern  $n-1$ . Diese Modifikation ist ein typisches Beispiel für eine Bias-Korrektur zur Sicherstellung der Unverzerrtheit. Der Maximum-Likelihood-Schätzer der Varianz  $\sigma^2$  einer Normalverteilung ist nämlich gleich

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

aber dieser Schätzer ist nicht unverzerrt. Wie wir gleich sehen werden, ändert die Bias-Korrektur nichts an den anderen "schönen" Eigenschaften von  $s^2$ , also an dessen Effizienz und Konsistenz.

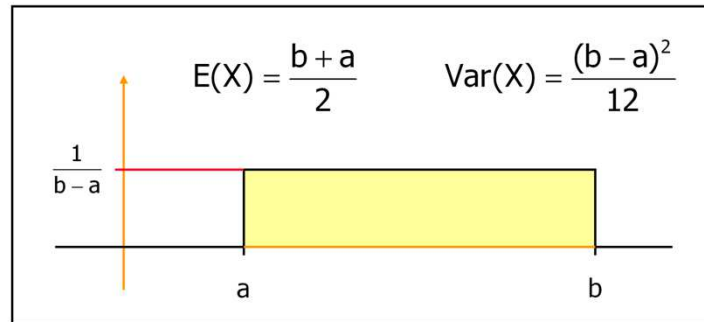
Aus praktischen Gründen (d.h. der bequemen Berechnung wegen) sollte man sich merken, dass die Summe der quadratischen Abweichung vom Mittelwert in der Formel für  $s^2$  durch einen Term ersetzt werden kann, der stattdessen nur den Mittelwert und die Quadratsumme enthält:

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



## Die (stetige) Gleichverteilung

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

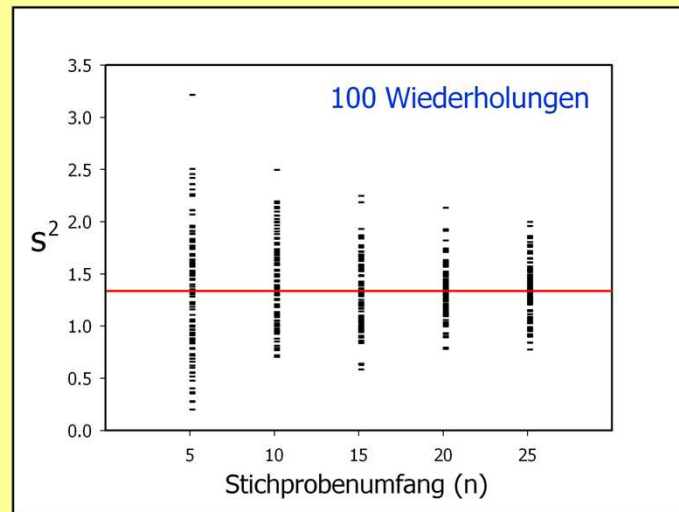


Die Unverzerrtheit von  $s^2$  wollen wir am Beispiel einer anderen, praktisch relevanten Verteilung quantitativer Zufallsvariabler verdeutlichen, nämlich der stetigen "Gleichverteilung". Eine Zufallsvariable  $X$  hat definitionsgemäß eine Gleichverteilung, wenn alle möglichen Werte in einem gegebenen Intervall  $[a, b]$  gleich wahrscheinlich sind, und wenn alle Werte außerhalb des Intervalls für  $X$  unmöglich sind. Die Annahme der gleichen Wahrscheinlichkeiten hat zur Folge, dass die Dichte der Gleichverteilung innerhalb des Intervalls konstant ist. Da die gesamte Fläche unter der Kurve gleich eins sein muss, beträgt diese Konstante zwangsläufig  $1/(b-a)$ .

Der Erwartungswert einer stetigen Gleichverteilung in dem Intervall  $[a, b]$  ist der Mittelpunkt des Intervalls, also  $(a+b)/2$ , da dort die Realisierungen von  $X$  "im Durchschnitt" lokalisiert sind.

## Gleichverteilung

$$a=4, b=8 \longrightarrow \mu=6, \sigma^2=1.33$$



## Konsistente Schätzer

Ein Schätzer heißt "konsistent", wenn seine Genauigkeit und Präzision mit zunehmendem Stichprobenumfang wachsen und jeweils gegen 100% streben.

$$P(|\hat{\theta}_n - \theta| > \varepsilon) \xrightarrow{n} 0$$

$\hat{\pi} = k / n$  ist ein konsistenter Schätzer von  $\pi$

$\hat{\mu} = \bar{x}$  ist ein konsistenter Schätzer von  $\mu$

$\hat{\sigma}^2 = s^2$  ist ein konsistenter Schätzer von  $\sigma^2$

Alle drei bislang betrachteten Schätzer sind konsistent, was schon aus den Abbildungen zur Verdeutlichung ihrer Unverzerrtheit ersichtlich wurde. Die darin gezeigten simulationsbasierten Schätzungen waren nicht nur beim Originalparameter lokalisiert, sondern streuten auch mit zunehmendem Stichprobenumfang immer weniger um diesen.

## Effiziente Schätzer

Ein unverzerrter Schätzer heißt "effizient", wenn jeder andere unverzerrte Schätzer mehr streut.

$$\text{Var}(\hat{\theta}) \text{ minimal}$$

$\hat{\pi} = k / n$  ist ein effizienter Schätzer von  $\pi$   
 $\hat{\mu} = \bar{X}$  ist meistens ein effizienter Schätzer von  $\mu$   
 $\hat{\sigma}^2 = s^2$  ist meistens ein effizienter Schätzer von  $\sigma^2$

Stichprobenmittel und Stichprobenvarianz sind leider nicht immer effizient. Man kann sich in der Tat obskure Verteilungen denken, für die andere Schätzer eine geringere Varianz haben. Für die Normalverteilung und andere "anständige" Verteilungen von praktischem Interesse sind die beiden Schätzer jedoch effizient.

## Konfidenzintervall

Für die meisten stetigen Zufallsvariablen gilt

$$P(\hat{\theta} = \theta) = 0$$

d.h. es ist unmöglich, dass ein Schätzer  
den wahren Parameter "auf den Kopf trifft".



Meistens ist es sinnvoller,  $\theta$  durch ein Intervall zu  
schätzen, das  $\theta$  mit einer gewissen "Sicherheit"  
enthält.

## Konfidenzintervall

### Definition

Ein Konfidenzintervall ist eine Vorschrift, die einer Stichprobe  $x$  ein Intervall  $I(x)$  so zuordnet, dass für jeden möglichen Wert  $\theta$  des zu schätzenden Parameters

$$P_{\theta}(x : \theta \in I(x)) \geq 1 - \alpha$$

gilt. Wurde die Stichprobe  $x$  erhoben und das Konfidenzintervall  $I(x)$  berechnet, so besteht ein **Vertrauen**  $1-\alpha$ , dass  $I(x)$  das wahre  $\theta$  auch enthält.

$1-\alpha$  heißt "Konfidenzniveau" (üblicherweise 0.95)

Konfidenzintervalle kommen durch die Anwendung einer festen Regel auf Stichprobendaten zustande. Bevor wir diese Regel auf die Stichprobe anwenden, wissen wir bereits, dass wir am Ende des Experiments mit einer Wahrscheinlichkeit von mindesten  $1-\alpha$  (meistens 0.95) ein Intervall erhalten, das den wahren Parameterwert einschließt. Das ist die Aussage der etwas kompliziert anmutenden Formel im Zentrum dieser Folie.

Ist die Stichprobe erst einmal gezogen und das Konfidenzintervall berechnet worden, dann transformiert sich die vor dem Experiment bestehende Wahrscheinlichkeit von z.B. 0.95 in ein 95%-iges Vertrauen, dass der wahre Parameterwert in dem errechneten Intervall enthalten ist.

## Konfidenzintervall

Erwartungswert

Aus dem Zentralen Grenzwertsatz folgt, dass

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

für großes  $n$  eine  $N(0,1)$ -Verteilung hat.

$$\longrightarrow P(-1.96 \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq 1.96) = 0.95$$

$$\longrightarrow P(\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}) = 0.95$$

Über Parameter können in der Regel keine Wahrscheinlichkeitsaussagen der folgenden Art getroffen werden: "Die Wahrscheinlichkeit, dass  $\mu$  größer ist als 10, beträgt 0.05". Solche Aussagen würden Informationen über die *a priori* Wahrscheinlichkeit von Parametern erfordern, die normalerweise nicht verfügbar sind.

Über Zufallsvariable lassen sich jedoch Wahrscheinlichkeitsaussagen treffen, und wie wir bereits gesehen haben, ist auch das Stichprobenmittel eine Zufallsvariable. Zudem besagt der Zentrale Grenzwertsatz, dass das standardisierte Stichprobenmittel bei hinreichend großem Stichprobenumfang einer Standard-Normalverteilung folgt. Das standardisierte Stichprobenmittel nimmt also mit Wahrscheinlichkeit 0.95 Werte zwischen -1.96 und +1.96 an. Dies ist die wesentliche Botschaft der ersten Formel im unteren Teil der Folie.

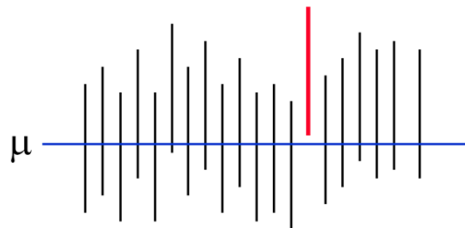
Die zweite Formel ist nichts weiter als eine mathematische Umformung der ersten; sie besagt das Gleiche, nur auf andere Weise. Allerdings lässt sich die zweite Formel auch als Aussage über eine Prozedur interpretieren, die so genannte "Konfidenzintervalle" dadurch erzeugt, dass zum Stichprobenmittel 1.96 mal den Standardfehler hinzuaddiert bzw. davon abgezogen wird. Die zweite Formel besagt, dass man bei Befolgen dieser Regel mit Wahrscheinlichkeit 0.95 ein Intervall erhält, das den wahren Parameter (in diesem Fall den Erwartungswert) einschließt.

## Konfidenzintervall

Erwartungswert

$$\bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

markiert ein Intervall, das den wahren Erwartungswert mit Wahrscheinlichkeit 0.95 (d.h. in 95% aller unabhängigen Wiederholungen des Experiments) enthalten **wird**.



Das 95% Konfidenzintervall (95% KI) für den Erwartungswert einer quantitativen Zufallsvariablen erhält man also dadurch, dass man zum Stichprobenmittel den folgenden Wert hinzuaddiert bzw. davon abzieht:

$$1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

(diese Faustregel ist auch als "plus-minus zwei Standardfehler" bekannt). Das resultierende Konfidenzintervall ist wegen der Art der Berechnung natürlich symmetrisch um das Stichprobenmittel.



## Weiblicher Body-Mass-Index (BMI)

95% Konfidenzintervall  
für den  
Erwartungswert  
(Annahme:  $\sigma=1.2$ )

$$18.6 \pm 1.96 \cdot \frac{1.2}{\sqrt{12}}$$

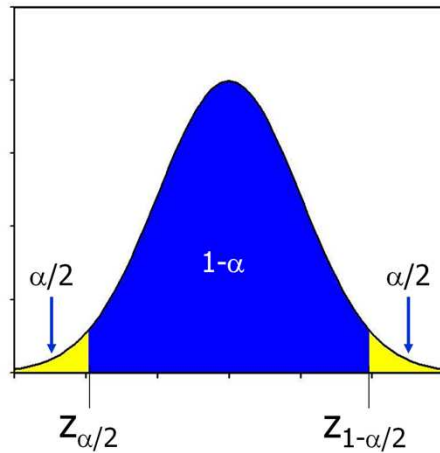
*oder*

(17.9,19.3)

Jahr	Name	BMI
1984	Suzette Charles	17.7
1985	Sharlene Wells	18.2
1986	Susan Akin	16.8
1987	Kellye Cash	17.6
1988	Kaye Lani Rae Rafko	18.8
1989	Gretchen Carlson	19.1
1990	Debbye Turner	17.9
1991	Marjorie Vincent	17.8
1998	Kate Shindle	20.2
1999	Nicole Johnson	19.6
2001	Angela Perez Baraquio	20.3
2002	Katie Harman	19.5

## Konfidenzintervall

Erwartungswert



$\sigma$  bekannt:

$$\bar{x} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$\sigma$  unbekannt:

$$\bar{x} \pm t_{1-\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}$$

wobei  $t_{1-\alpha/2, n-1}$  das Quantil einer t-Verteilung mit  $n-1$  Freiheitsgraden ist.

Bislang sind wir bei der Berechnung von Konfidenzintervallen für Erwartungswerte immer davon ausgegangen, dass (i) das vom Forscher gewünschte Vertrauen 95% beträgt, und dass (ii) die Standardabweichung der zu analysierenden Zufallsvariablen bekannt ist. Diese Annahmen müssen in der Praxis jedoch nicht notwendigerweise erfüllt sein.

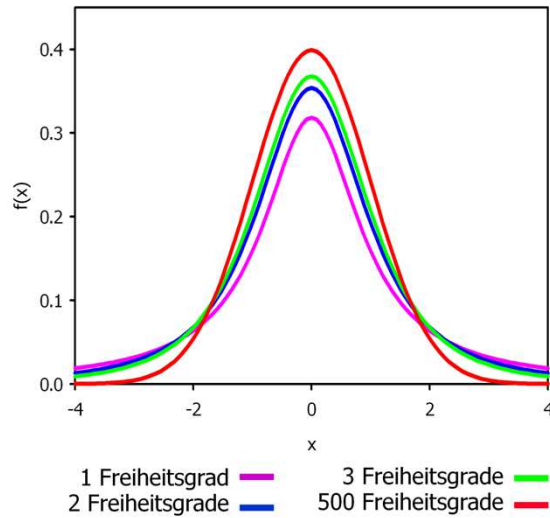
Die in den vorangegangenen Berechnungen verwendete Zahl 1.96 entspricht dem 0.975-Quantil der Standard-Normalverteilung, d.h. 97.5% der Fläche unter der Gauß'schen Glockenkurve liegen links von 1.96, und 2.5% der Fläche liegen rechts davon. Formal:  $z_{0.975} = 1.96$ . Wegen der Symmetrie der Normalverteilung liegen deshalb 95% der Fläche innerhalb des Intervalls  $[z_{0.025}, z_{0.975}]$ , und 5% liegen außerhalb. Für andere Fehlerwahrscheinlichkeiten  $\alpha$  lässt sich dieser Zusammenhang dadurch verallgemeinern, dass statt -1.96 und 1.96 die entsprechenden Quantile  $z_{\alpha/2}$  und  $z_{1-\alpha/2}$  verwendet werden. Tabellen mit den Quantilen der Standard-Normalverteilung finden sich in jedem Statistiklehrbuch.

Wenn die Standardabweichung  $\sigma$  nicht bekannt ist, dann muss sie durch  $s$  aus der gleichen Stichprobe geschätzt werden, aus der auch das Konfidenzintervall berechnet werden soll. In diesem Fall sind allerdings statt  $z_{\alpha/2}$  und  $z_{1-\alpha/2}$  die Quantile  $t_{\alpha/2, n-1}$  bzw.  $t_{1-\alpha/2, n-1}$  einer anderen Verteilung, nämlich der t-Verteilung, zu verwenden.

## Student t-Verteilung



William S. Gosset  
(1876-1937)



William Gosset wurde am 13. Juni 1876 in Canterbury, England, geboren. Er besuchte das Winchester College und das New College in Oxford, wo er 1899 in Statistik und Chemie graduierte. Nach seinem Abschluss begann er für eine Firma namens "Dublin Brewery of Arthur Guinness and Son" zu arbeiten. Mit seinem statistischen Wissen half Gosset der Brauerei u.a. bei der Auswahl der ertragreichsten Gerstensorten.

Da Guinness seine Angestellten daran hinderte, eigene wissenschaftliche Arbeiten zu verfassen, musste Gosset unter falschem Namen publizieren. Sein Pseudonym lautete "Student", was zur heute gängigen Verwendung der Bezeichnung "Student'sche t-Verteilung" für eine seiner wichtigsten wissenschaftlichen Leistungen führte.

In einer 1908 erschienenen Arbeit zeigte Gosset, dass eine kleine Stichprobe von im Wesentlichen normalverteilten Daten eine Verteilung hat, die allein durch die Stichprobengröße charakterisiert wird. Diese Verteilung, die t-Verteilung, entsteht, wenn (wie häufig in der Statistik) die Standardabweichung eines Populationsparameters nicht bekannt ist und stattdessen aus den gleichen Daten geschätzt werden muss wie der Parameter selbst. Die t-Verteilung lässt sich durch einen einzigen Parameter  $\nu$  charakterisieren, der "Anzahl der Freiheitsgrade" genannt wird. Die Quantile der t-Verteilung findet man ebenfalls in den meisten Statistiklehrbüchern tabelliert.

## Wie man bessere Schätzungen bekommt

Erwartungswert

$$\sqrt{\text{Var}(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

d.h. der Standardfehler des Stichprobenmittels sinkt mit wachsendem Stichprobenumfang.

$$\text{KI} : \bar{X} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\text{KI} : \bar{X} \pm t_{1-\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}$$

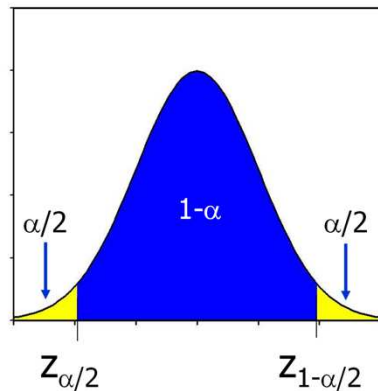
d.h. die Breite eines Konfidenzintervalls verringert sich mit wachsendem Stichprobenumfang.

Der Standardfehler des Stichprobenmittels (als Schätzer des Erwartungswertes  $\mu$ ) ergibt sich aus der Standardabweichung der ursprünglichen Zufallsvariablen, geteilt durch die Wurzel aus der Stichprobengröße. Daher verbessert sich die Präzision der Schätzung des Erwartungswertes mit zunehmendem Stichprobenumfang. Je größer die Stichprobe, umso geringer der Standardfehler, d.h. umso weniger wird das Stichprobenmittel um den (wahren, aber unbekannten) Erwartungswert streuen.

In gleicher Weise kann auch die Qualität des Konfidenzintervalls für einen Erwartungswert dadurch erhöht werden, dass die Stichprobe vergrößert wird. Ein schmales Konfidenzintervall ist besser als ein breites, weil Ersteres weniger Unsicherheit über die Lage des wahren Parameters hinterlässt als Letzteres. Die Intervallbreite ist jedoch proportional zum Standardfehler, so dass sich die Breite verringert wenn  $n$  wächst (und zwar unabhängig davon, ob  $\sigma$  bekannt ist oder erst aus der Stichprobe geschätzt werden muss).

## Wie man bessere Schätzungen bekommt

Erwartungswert



Breite W des KI:

$$W = 2 \cdot z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$1-\alpha$  steigt

→  $z_{1-\alpha/2}$  steigt

→ W steigt

Die Breite eines Konfidenzintervalls steigt mit steigender Sicherheit.

Die Breite eines Konfidenzintervalls hängt auch von der jeweiligen Fehlerwahrscheinlichkeit  $\alpha$  ab. Je kleiner  $\alpha$  ist, d.h. je mehr Vertrauen in das errechnete Intervall gesetzt werden soll, umso breiter muss es sein. In dieser Hinsicht lässt sich ein Konfidenzintervall mit einem Fischernetz vergleichen. Je sicherer man sein möchte, dass der Fisch ins Netz gegangen ist, umso größer muss das Netz sein.

## Konfidenzintervall

### Stichprobenumfang

$$W = 2 \cdot z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \longrightarrow \quad n = \left( \frac{2 \cdot z_{1-\alpha/2} \cdot \sigma}{W} \right)^2$$

Wie viele Beobachtungen sind nötig, um den Erwartungswert des männlichen BMI mit einem 95% Konfidenzintervall von höchstens 2 kg/m<sup>2</sup> Breite zu schätzen (Annahme: BMI ist normalverteilt mit  $\sigma=2.5$ )?

$$\text{Antwort: } n = \left( \frac{2 \cdot 1.96 \cdot 2.5}{2} \right)^2 = 24.01$$

### Die mathematische Beziehung

$$W = 2 \cdot z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

zwischen der Breite eines Konfidenzintervalls und dem Stichprobenumfang kann auch zur Planung wissenschaftlicher Experimente herangezogen werden. Sobald der Stichprobenumfang und das gewünschte Ausmaß an Vertrauen festliegen, ergibt sich die Intervallbreite automatisch. Unterliegt der Stichprobenumfang jedoch keinerlei Beschränkung, dann kann die Intervallbreite als Maß der gewünschten Präzision bei vorgegebener Fehlerwahrscheinlichkeit frei gewählt werden. Der dafür notwendige Stichprobenumfang ergibt sich dann aus der transformierten Formel

$$n = \left( \frac{2 \cdot z_{1-\alpha/2} \cdot \sigma}{W} \right)^2.$$

## Zusammenfassung

- Schätzen bezeichnet den wissenschaftlichen Vorgang des **Erschließens von Populationsparametern** aus Stichproben.
- Ein **Schätzer** ist eine mathematische Vorschrift für die Berechnung von Parameterschätzungen aus Daten.
- Schätzer wie z.B. das Stichprobenmittel sind selbst wieder **Zufallsvariable**, mit Erwartungswert und Varianz.
- Gute Schätzer sollten **unverzerrt** (genau), **effizient** (am präzisesten) und **konsistent** (zunehmend präzise) sein.
- Statt "Punktschätzungen" liefern **Konfidenzintervalle** Bereiche, die einen gesuchten Parameter mit bestimmter Sicherheit enthalten.