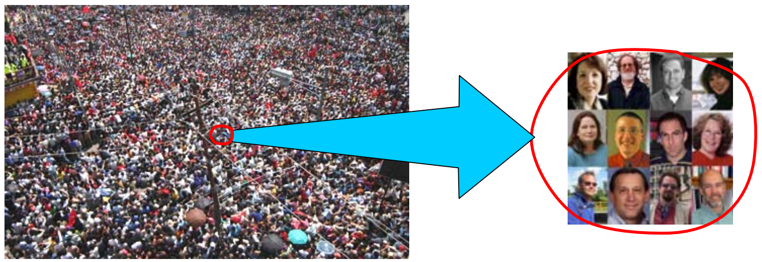
Statistische Schätzungen

(8,5±1,5) cm

Analytische Statistik



Population	Stichprobe
N = „unendlich”	n = endlich
Theoretische Verteilung	Häufigkeitsverteilung
Erwartungswert	Durchschnitt
Theoretische Streuung	Standardabweichung

?

Aufgabe der Schätztheorie

Aus einer Stichprobe Schätzwerte für

- Wahrscheinlichkeiten
- Erwartungswert
- Streuung
- oder andere Parametern

einer Verteilung zu ermitteln.

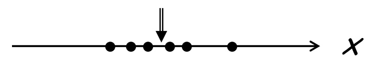
Typen der Schätzungen:

- ***Punktschätzung***
- ***Intervallschätzung***

Punktschätzungen

Wir wollen jetzt die Parameter einer Verteilung (μ,σ) aus den konkreten Werten x_1,...x_n einer Stichprobe „möglichst gut“ bestimmen, d.h. einen „Näherungswert“ errechnen.

Kriterien:



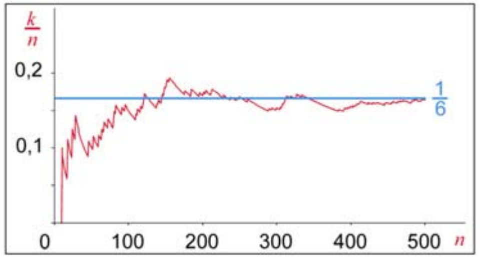
Erwartungstreue (unverzerrt)	Erwartungswert der Schätzwerte = zu schätzender Parameter
Konsistenz	n ↑ bessere Schätzung
Effizienz (wirksam)	kleine Streuung
Exhaustivität (erschöpfend)	berücksichtigt alle Informationen

Punktschätzungen

Der Parameter wird mit einem Wert geschätzt.

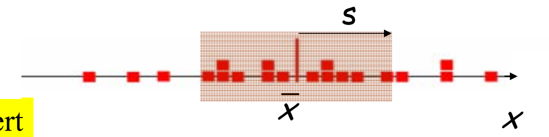
Relative Häufigkeit
ist ein Schätzwert für die Wahrscheinlichkeit

Siehe Definition der statistischen Wahrscheinlichkeit!

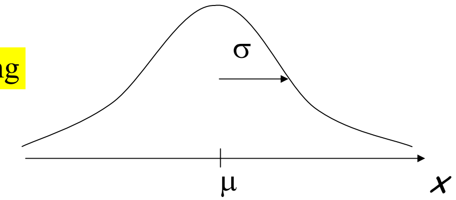


Punktschätzungen

Durchschnitt
ist ein Schätzwert
für den Erwartungswert



Standardabweichung
ist ein Schätzwert
für die theoretische Streuung



Punktschätzungen sagen
***nichts* über die Genauigkeit bzw. Sicherheit**
der Schätzung !

Intervallschätzungen

Intervallschätzung oder Konfidenzschätzung gibt zu einer vorgewählten Sicherheitswahrscheinlichkeit γ, (Konfidenzniveau) ein Intervall (c_1,c_2) an, in dem der unbekannte Parameter (zB. μ oder σ) mit einer Wahrscheinlichkeit von mindestens γ liegt.

Zb.: Erwartungswert der Pulszahl ist bei
95% Konfidenzniveau: 74±6 $^1/_{Min}$

Intervallschätzungen

Wie große γ soll gewählt werden?
Wie große sind die Schaden bei einer falschen Schätzung?

Sozialwissenschaft γ=0,9

Medizin γ=0,95

Technik γ=0,99

Einfluss der Streuung und des Stichprobenumfanges

α=1-γ Irrtumswahrscheinlichkeit

Konfidenzintervall für den Erwartungswert

Gedankenversuch:

Sei x eine Zufallsgröße (zB: Körperhöhe) mit einem beliebigen Verteilung mit einem Erwartungswert μ und Streuung σ.

Nehmen wir jetzt viele Stichproben, (zB: viele Studentengruppen) alle mit gleichem Stichprobenumfang n.

Sei $\bar{x}_i$ der Durchschnitt der i-ten Stichprobe

Wie sieht die Verteilung von $\bar{x}_i$ Werte aus?

Zentraler Grenzwertsatz: bei genug hohen n die Verteilung ist eine Normalverteilung.

Lage ($\mu_{\bar{x}}$) und Breite ($\sigma_{\bar{x}}$) der Verteilung der Durchschnittwerte?

Verteilung von Durchschnitt der Zufallsgrössen

x_1 und x_2 sind unabhängige Zufallsgrößen. Beide folgen eine Normalverteilung mit derselben Erwartungswerte μ und Streuungen σ.

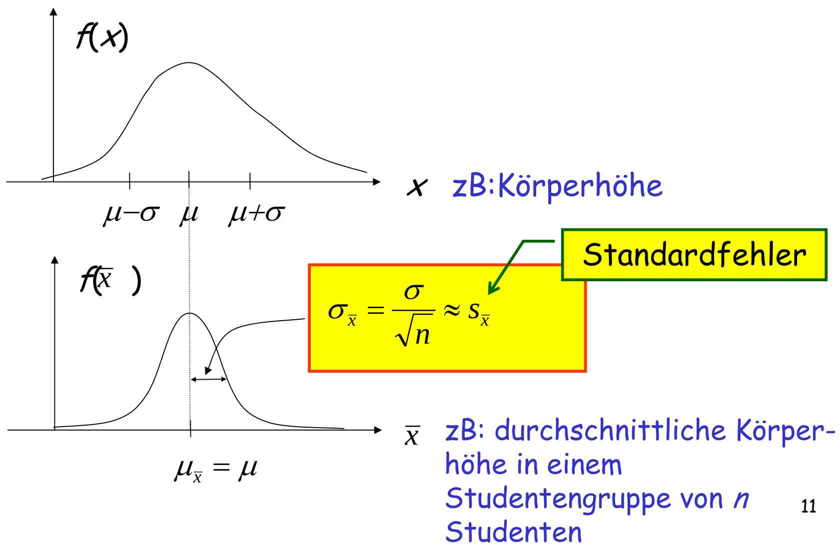
Welche Verteilung folgt der Durchschnitt $x = (x_1 + x_2)/2$?

Normalverteilung, mit den folgenden Parametern:

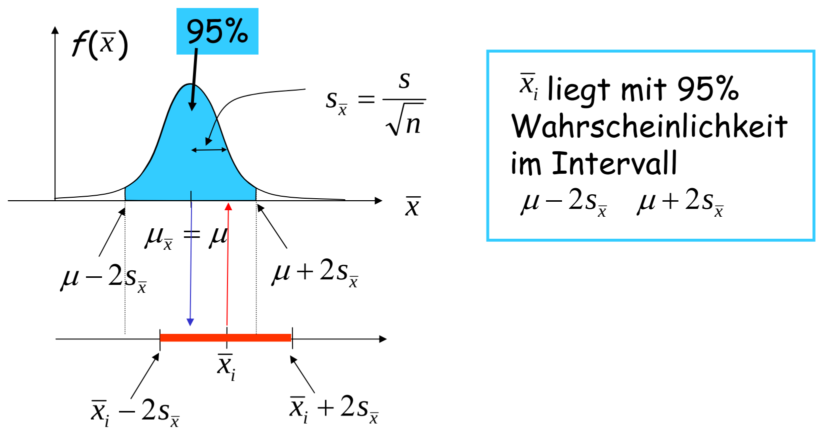
Messwerte	Summe	Durchschnitt	Allgemein für n Messwerte
x_1 , x_2	x_1+x_2	$x = (x_1 + x_2)/2$	$x = (x_1 + \ldots + x_n)/n$
μ	μ +μ=2μ	μ	μ
σ^2	$\sigma^2 + \sigma^2 = 2\sigma^2$		
σ	$\sqrt{2}\sigma$	$\sigma/\sqrt{2}$	$\sigma/\sqrt{n}$

Zur Erinnerung

Konfidenzintervall für den Erwartungswert



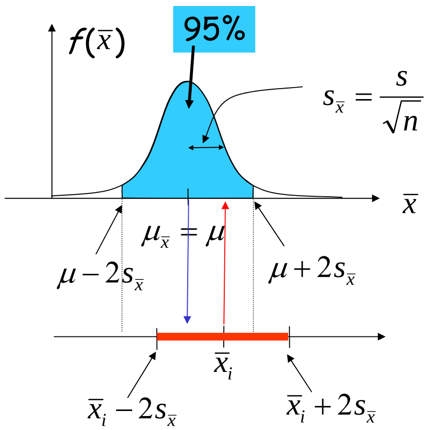
Konfidenzintervall für den Erwartungswert



wenn $\mu - 2s_{\bar{x}} \le \bar{x}_i \le \mu + 2s_{\bar{x}}$ dann $\boxed{\bar{x}_i - 2s_{\bar{x}}} \le \mu \le \boxed{\bar{x}_i + 2s_{\bar{x}}}$

95% Wahrsch. 95% Wahrsch.

Konfidenzintervall für den Erwartungswert



$\bar{x}_i$ liegt mit 5% Wahrscheinlichkeit im Intervall
$\mu - 2s_{\bar{x}}$ $\quad$ $\mu + 2s_{\bar{x}}$
nicht!

$$\bar{x}_i \le \mu - 2s_{\bar{x}} \text{ oder } \mu + 2s_{\bar{x}} \le \bar{x}_i \quad \Rightarrow \quad \mu \le \bar{x}_i - 2s_{\bar{x}} \text{ oder } \bar{x}_i + 2s_{\bar{x}} \le \mu$$

5% Wahrsch. $\qquad$ 5% Wahrsch.

Konfidenzintervall für den Erwartungswert

In dem Intervall $\bar{x} - 2s_{\bar{x}},\ \bar{x} + 2s_{\bar{x}}$ (Konfidenzintervall) liegt der Erwartungswert (μ) mit 95% Wahrscheinlichkeit

Eine ähnliche Ableitung gibt: μ liegt
-mit 68% Wahrscheinlichkeit im Intervall: $\bar{x} - s_{\bar{x}},\ \bar{x} + s_{\bar{x}}$

- mit 99,7% Wahrscheinlichkeit im Intervall: $\bar{x} - 3s_{\bar{x}},\ \bar{x} + 3s_{\bar{x}}$

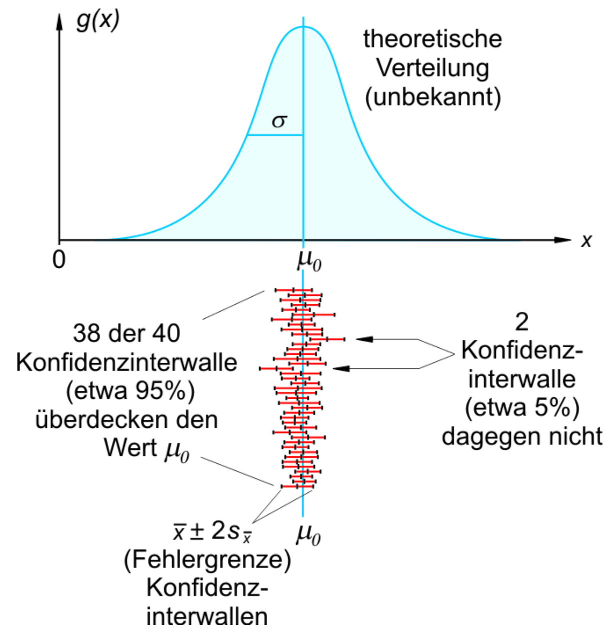
Je größer ist die Sicherheitswahrscheinlichkeit desto breiter ist das Konfidenzintervall!

Bemerkung: wenn n→∞ dann $s_{\bar{x}} \to 0$

Aufgaben zum Konfidenzintervall des Erwartungswertes

Eine Maschine herstellt Tabletten. Nominale Wirkstoffgehalt der Tabletten beträgt 20mg. Wirkstoffsgehalt der 10 Tabletten wurde gemessen. Der Durchschnitt beträgt 18,9 mg. Die Standardabweichung der Wirkstoffsgehalt beträgt 1,6 mg.
Geben Sie die mit 95% Sicherheitswahrscheinlichket gehörende Konfidenzintervall an!
Ist diese Maschine gut eingestellt?

Mit einer sehr langen Mess-Serie haben wir die Erwartungswert und die theoretische Streuung der Blutzuckerkonzentration bestimmt.
Jetzt wird die Blutzuckerkonzentration in 60 Studentengruppen bestimmt. Wir bestimmen die 95% Konfidenzintervallen für alle Gruppen. Wieviele Konfidenzintervallen enthalten den Erwartungswert?



Pr.Buch Abb. 11

Zusammenfassung der Schätzungen

Punktsätzungen:

Stich-probe	Grund-gesamtheit
$\bar{x}$ ⟶	μ
s ⟶	σ
n ⟶	∞

h ⟶	P

Intervallschätzung
für den Erwartunswert:

$\bar{x} \pm 2s_{\bar{x}}$ 95%

Bestimmung des Stichprobenumfanges

Welcher Stichprobenumfang ist notwendig zu einer bestimmten Genauigkeit?
(z.B.: Körperhöhe mit ±1cm „Genauigkeit" bei 95% Konfidenzniveau)

$$2s_{\bar{x}} = 1\,\text{cm} \Rightarrow s_{\bar{x}} = 0{,}5\ \text{cm}$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \Rightarrow s_{\bar{x}}^2 = \frac{s^2}{n} \Rightarrow n = \frac{s^2}{s_{\bar{x}}^2}$$

s = ? s kann aus einer kleineren Stichprobe geschätzt werden.

Z.B.: Körperhöhe in einer Studentengruppe (20 St.): s = 8,3 cm

$$n = \frac{s^2}{s_{\bar{x}}^2} = \frac{8.3^2\,\text{cm}^2}{0.5^2\,\text{cm}^2} \approx 276$$

Konfidenzintervall für Quotienten (Wahrscheinlichkeit)

Zwei Möglichkeiten: (E/$\bar{E}$, z.B.: Raucher/Nichtraucher)
Binomialverteilung
E kommt mit einer Wahrscheinlichkeit von p vor.
Stichprobenumfang: n
In einem Versuch E kommt n_E –mal vor (n_E aus n Personen sind Raucher)
Die relative Häufigkeit $h=n_E/n$ ist ein Schätzwert für p (Punktschätz.)
n_E folgt eine Binomialverteilung mit einem Erwartungswert von pn
Theoretische Streuung der Binomialverteilung: $\sigma_{nE}=\sqrt{np(1-p)}$ (Streuung von n_E)
p wird mit der relativen Häufigkeit geschätzt: $\sigma_{nE}\approx\sqrt{nh(1-h)}$
Weil $p\approx h = n_E/n$, Streuung von p : $\sigma = \sigma_{nE}/n = \sqrt{nh(1-h)}\,/n = \sqrt{h(1-h)/n}$
Analog zu $\bar{x}\pm 2\sigma$
p befindet sich mit 95 % Wahrscheinlichkeit in:
$h \pm 2\sqrt{h(1-h)/n}$ (95% Konfidenzniveau)
zB.: 20 Raucher aus 100 ➡ $0{,}2\pm 2\sqrt{0{,}2\cdot 0{,}8/100} = 0{,}2 \pm 0{,}08 = (20\pm 8)\%$