

Statistische Schätzungen,

László Smeller

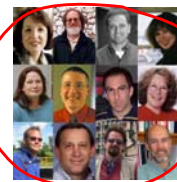
Statistische Schätzungen



Analytische Statistik (induktive o. schließende Statistik)



Population
N = „unendlich“



Stichprobe
n = endlich

Theoretische Verteilung
Erwartungswert
Theoretische Streuung



Häufigkeitsverteilung
Durchschnitt
Standardabweichung

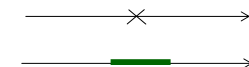
Aufgabe der Schätztheorie

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, \dots, x_n einer Stichprobe „möglichst gut“ bestimmen, d.h. einen „Näherungswert“ errechnen.

Kriterien:



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

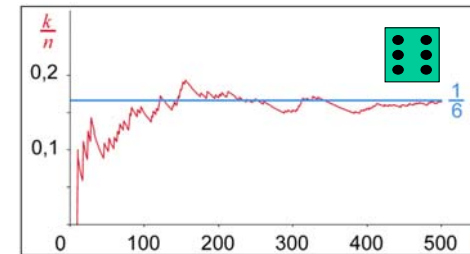
Punktschätzungen

Der Parameter wird mit **einem Wert** geschätzt.

Relative Häufigkeit

ist ein Schätzwert für die **Wahrscheinlichkeit**

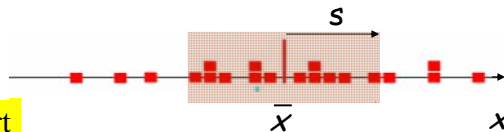
Siehe Definition der statistischen Wahrscheinlichkeit!



Punktschätzungen

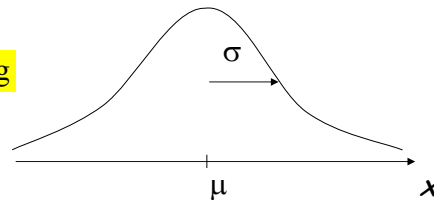
Durchschnitt

ist ein Schätzwert
für den **Erwartungswert**



Standardabweichung

ist ein Schätzwert
für die **theoretische Streuung**



Punktschätzungen sagen

nicht 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 $1/_{Min}$

$\alpha = 1 - \gamma$ Irrtumswahrscheinlichkeit

Intervallschätzungen

Wie große γ Sicherheitswahrscheinlichkeit (Konfidenzniveau) soll gewählt werden?

Wichtige Faktoren:

- Streuung der Daten
- Stichprobenumfang
- Größe der Schaden bei einer falschen Schätzung

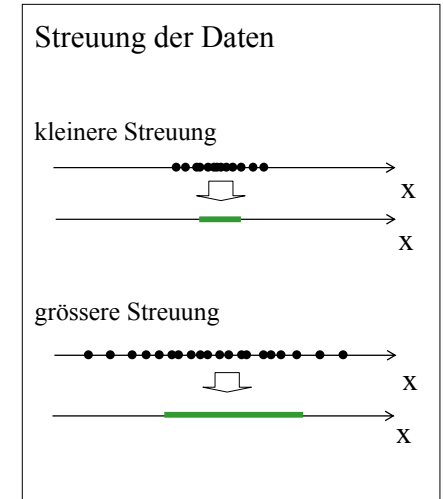
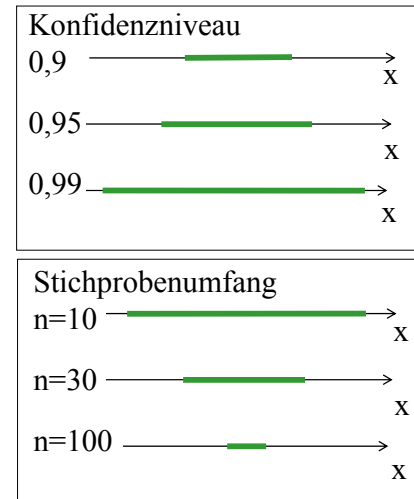
Sozialwissenschaft $\gamma=0,9$

Medizin $\gamma=0,95$

Technik $\gamma=0,99$

9

Einfluss des Konfidenzniveaus, der Streuung und des Stichprobenumfanges auf die Breite des Konfidenzintervalles



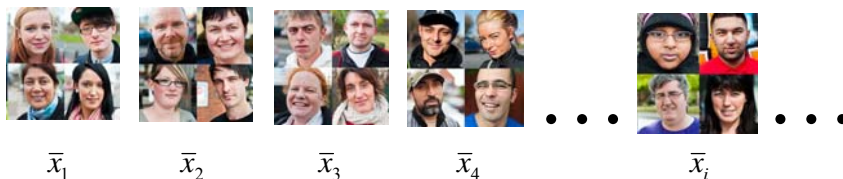
Konfidenzintervall für den Erwartungswert

Wir wollen eine Intervallschätzung für den Erwartungswert (μ) einer Zufallsgröße (zB: Körperhöhe) geben.

Gedankenexperiment:

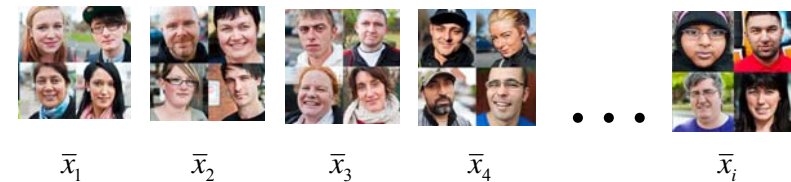
Nehmen wir jetzt viele Stichproben, (zB: viele Studentengruppen) alle mit gleichem Stichprobenumfang n .

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



11

Konfidenzintervall für den Erwartungswert



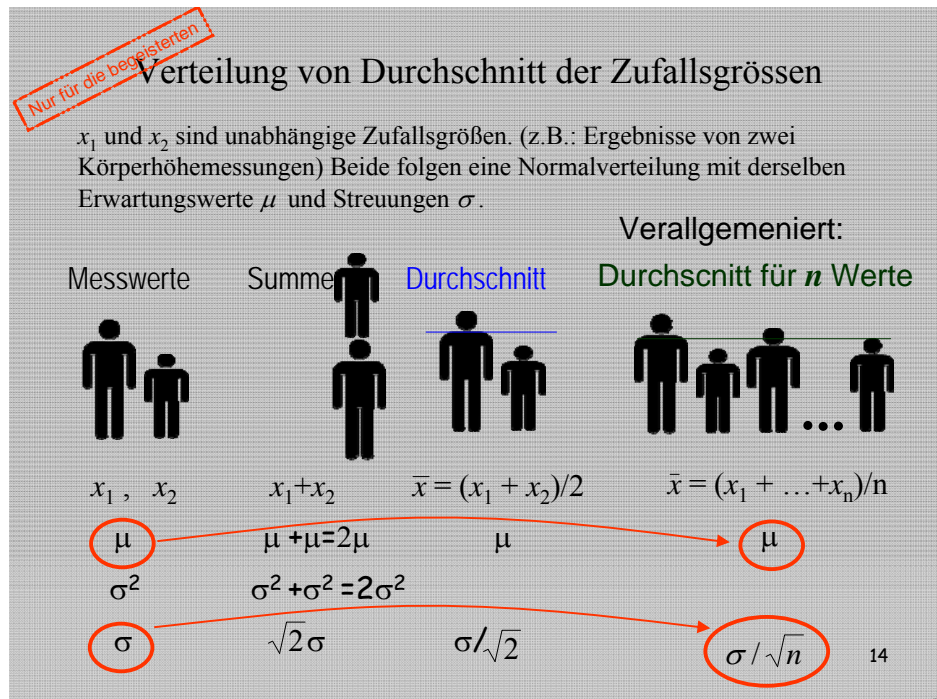
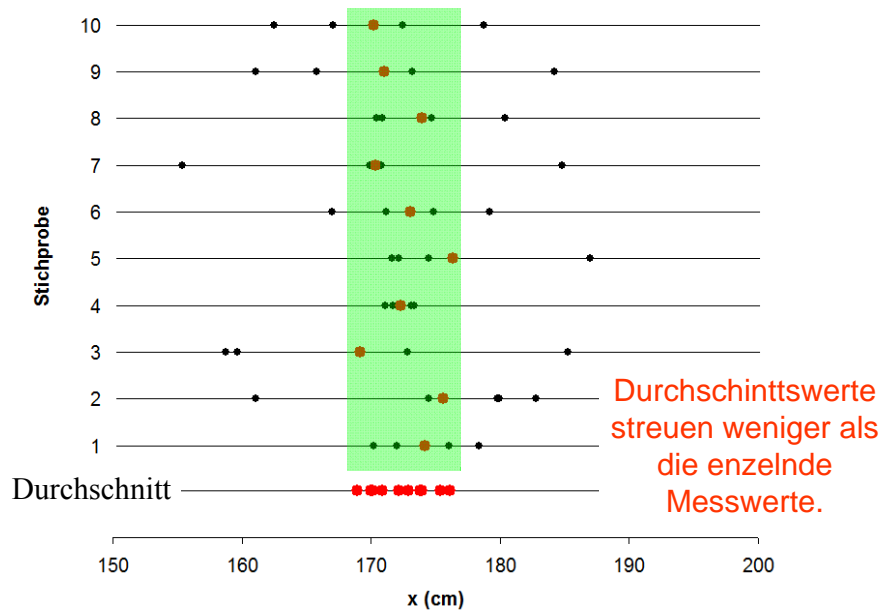
Wie sieht die Verteilung von \bar{x}_i Werte aus?

Zentraler Grenzwertsatz: bei genug hohen n die Verteilung der Durchschnittswerte (\bar{x}_i) ist eine Normalverteilung.

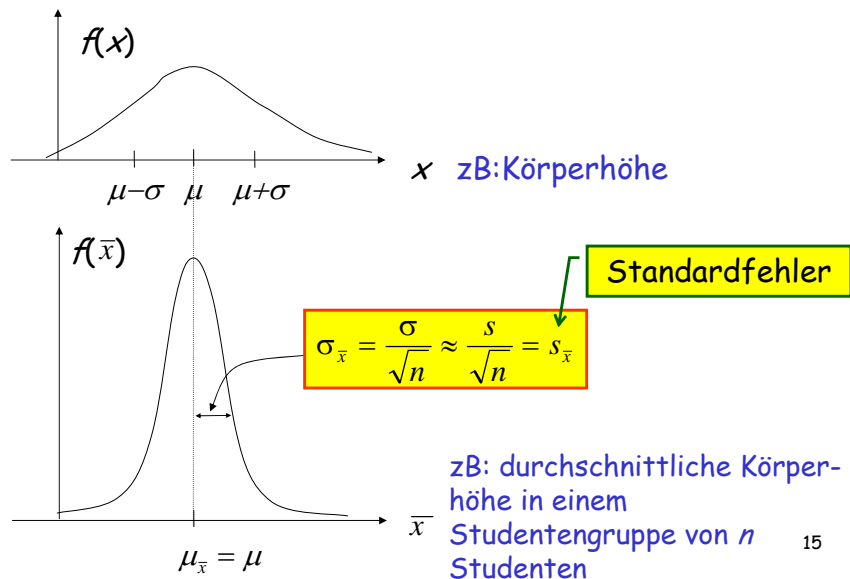
Lage ($\mu_{\bar{x}}$) und Breite ($\sigma_{\bar{x}}$) der Verteilung der Durchschnittswerte (\bar{x}_i)?

12

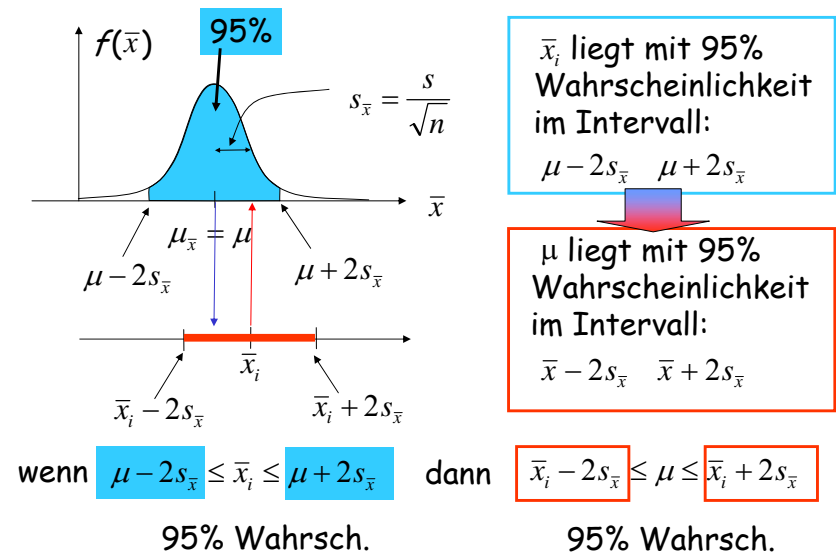
Daten und ihre Durchschnittswerte



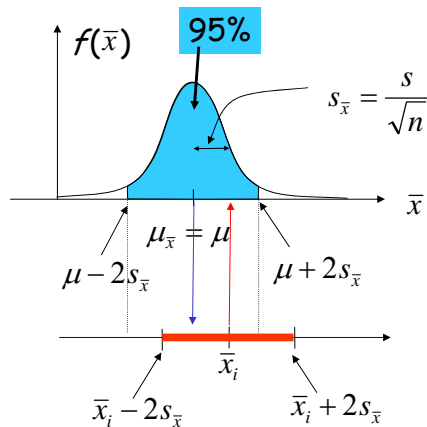
Konfidenzintervall für den Erwartungswert



Konfidenzintervall für den Erwartungswert



Konfidenzintervall für den Erwartungswert



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

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

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

5% Wahrsch. 5% Wahrsch.

17

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: $\bar{x} - s_{\bar{x}}$, $\bar{x} + s_{\bar{x}}$

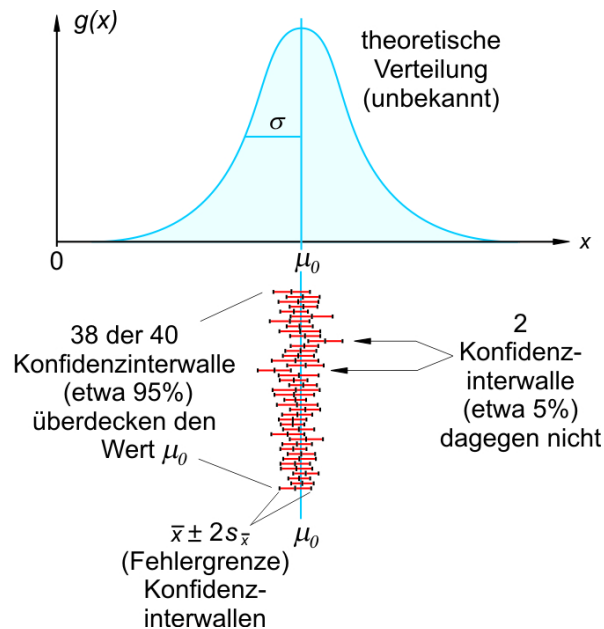
- mit 99,7% Wahrscheinlichkeit im Intervall:

$$\bar{x} - 3s_{\bar{x}}, \bar{x} + 3s_{\bar{x}}$$

Je größer ist die Sicherheitswahrscheinlichkeit desto breiter ist das Konfidenzintervall!

Bemerkung: wenn $n \rightarrow \infty$ dann $s_{\bar{x}} \rightarrow 0$

18



19

Bestimmung des Stichprobenumfangs

Welcher Stichprobenumfang ist notwendig zu einer bestimmten Genauigkeit? (z.B.: Körperhöhe mit ± 1 cm „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.): $s = 8,3 \text{ cm}$

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

Konfidenzintervall für Quotienten (Wahrscheinlichkeit)

Zwei Möglichkeiten: (E/Ē, z.B.: Raucher/Nichtraucher)

Binomialverteilung

E kommt mit einer Wahrscheinlichkeit von p vor.

Stichprobenumfang: n

In einem Versuch E kommt k-mal vor (k aus n Personen sind Raucher)

Die relative Häufigkeit $h=k/n$ ist ein Schätzwert für p (Punktschätzung.)

k folgt eine Binomialverteilung mit einem Erwartungswert von pn

Theoretische Streuung der Binomialverteilung: $\sigma_k = \sqrt{np(1-p)}$ (Streuung von k)

p wird mit der relativen Häufigkeit geschätzt: $\sigma_k \approx \sqrt{nh(1-h)}$

Weil $p \approx h = k/n$, Streuung von p: $\sigma = \sigma_k/n = \sqrt{nh(1-h)}/n = \sqrt{h(1-h)/n}$

Analog zu $\bar{x} \pm 2\sigma$

p befindet sich mit 95 % Wahrscheinlichkeit in:

$h \pm 2\sqrt{h(1-h)/n}$ (95% Konfidenzniveau)

z.B.: 20 Raucher aus 100 $\rightarrow 0,2 \pm 2\sqrt{0,2 \cdot 0,8/100} = 0,2 \pm 0,08 = (20 \pm 8)\%$

Zusammenfassung der Schätzungen

Punktschätzungen:

| Stich-probe | Grund-gesamtheit |
|-------------|------------------|
| \bar{x} | μ |
| s | σ |
| n | ∞ |
| h | P |

Intervallschätzung mit 95% Konfidenzniveau

für den Erwartungswert (μ):

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

für die Wahrscheinlichkeit (P):

$$h \pm 2\sqrt{h(1-h)/n}$$

