Statistik – Einführung

Stichprobenverteilung
Kapitel 6

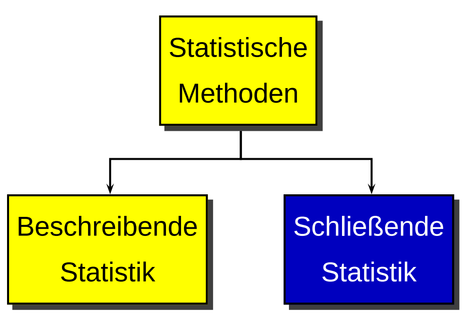
Statistik – WU Wien

Gerhard Derflinger · Michael Hauser · Jörg Lenneis · Josef Leydold ·
Günter Tirler · Rosmarie Wakolbinger

Lernziele

1. Beschreiben Eigenschaften von Schätzern.
2. Beschreiben Stichprobenverteilungen.
3. Beschreiben den Zusammenhang zwischen Grundgesamtheit und Stichprobenverteilung.
4. Formulieren den zentralen Grenzwertsatz.

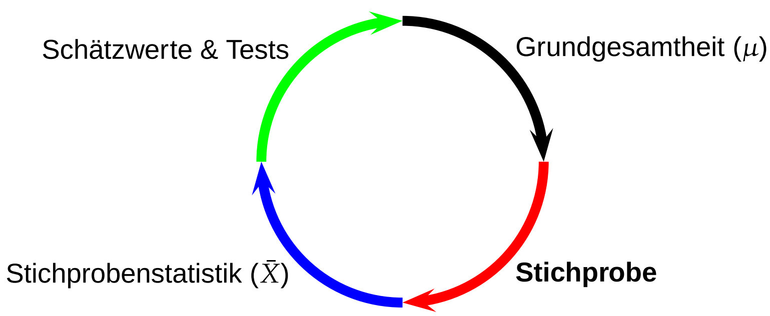
Statistische Methoden



- Umfasst
  - **Schätzen**
  - **Hypothesen testen**
- Zweck
  - mittels einer Stichprobe auf eine Grundgesamtheit schließen.

    Beispiel:
    100 Telefoninterviews nach einer TV-Diskussion:
    Wer hat besser abgeschnitten, Kandidat A oder Kandidat B?
    Angenommen Kandidat A wird von 60% der Befragten bevorzugt. Gilt dies auch für die 2 Mio. Seher?

Statistisches Schließen



Schätzer

- Zufallsvariable werden benutzt um Parameter der Grundgesamtheit zu schätzen.
  - Stichprobenmittelwert, Stichprobenanteil, Stichprobenvarianz sind Zufallsvariable.
- Beispiel: Stichprobenmittelwert $\bar{X}$ ist Schätzer für den Erwartungswert μ der Grundgesamtheit.
  - Ist $\bar{X} = 3$, so ist 3 ein (Punkt-)Schätzer für μ.
    Ist $\bar{X} = 4$, so ist 4 ein Schätzer für μ.
- Die theoretische Basis für die Eigenschaften unserer Schätzer bildet die zugehörige Stichprobenverteilung.

- Die Stichprobenverteilung ist eine **theoretische** Verteilung.
- Die Stichprobenstatistik ist eine Zufallsvariable. Ihr Wert hängt von der jeweils gezogenen Stichprobe ab.
Z.B.: Stichprobenmittelwert, Stichprobenanteilswert, ...
- Die Stichprobenverteilung erhält man durch Ziehen **aller** möglichen Stichproben **fixer** Größe.
Für jede mögliche Stichprobe berechnet man z.B. das Stichprobenmittel.
- Die Stichprobenverteilung ist die Liste $(\bar{x}, f_{\bar{x}}(\bar{x}))$ für z.B. das Stichprobenmittel.

Bestimmung der Stichprobenverteilung

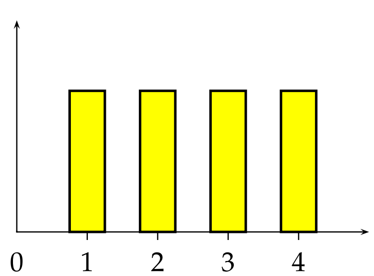
Beispiel:

- Urne mit 4 Kugeln mit Gewinnzahlen 1 bis 4. Ziehen mit Zurücklegen.
- Zufallsvariable X, 1-maliges Ziehen, ist diskret gleich verteilt.
- Werte für X: $x = 1, 2, 3, 4$
- $P(X = 1) = \ldots = P(X = 4) = \dfrac{1}{4}$

Bestimmung der Stichprobenverteilung

Parameter der Grundgesamtheit:

$$\mu = \sum_{i=1}^{N} x_i \frac{1}{N} = \frac{1+2+3+4}{4} = 2.5$$



$$\sigma = \sqrt{\sum_{i=1}^{N} (x_i - \mu)^2 \frac{1}{N}} = 1.12$$

16 Stichprobenmuster

1. Beobachtung		2. Beobachtung			
		1	2	3	4
	1	1/1	1/2	1/3	1/4
	2	2/1	2/2	2/3	2/4
	3	3/1	3/2	3/3	3/4
	4	4/1	4/2	4/3	4/4

16 Stichprobenmittelwerte

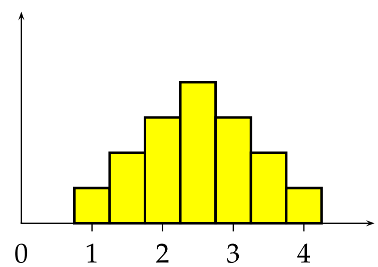
1. Beobachtung		2. Beobachtung			
		1	2	3	4
	1	1.0	1.5	2.0	2.5
	2	1.5	2.0	2.5	3.0
	3	2.0	2.5	3.0	3.5
	4	2.5	3.0	3.5	4.0

Ziehen mit Zurücklegen

Verteilung aller Stichprobenmittelwerte

16 Stichprobenmittelwerte

1. Beobachtung		2. Beobachtung			
		1	2	3	4
	1	1.0	1.5	2.0	2.5
	2	1.5	2.0	2.5	3.0
	3	2.0	2.5	3.0	3.5
	4	2.5	3.0	3.5	4.0

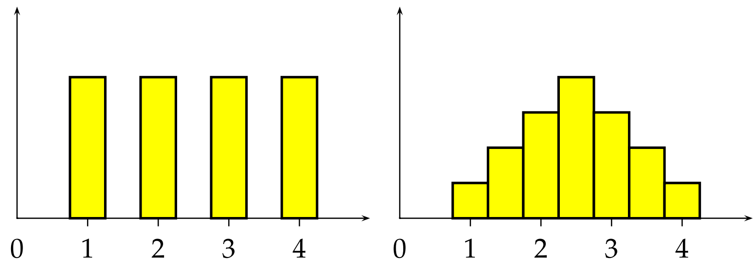


Parameter Stichprobenverteilung

$$\mu_{\bar{x}} = \frac{\sum_{i=1}^{N} \bar{x}_i}{N} = \frac{1.0 + 1.5 + 1.5 + 2.0 + \ldots + 4.0}{16}$$
$$= 2.5$$

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{N} (\bar{x}_i - \mu_{\bar{x}})^2}{N}$$
$$= \frac{(1.0 - 2.5)^2 + (1.5 - 2.5)^2 + (1.5 - 2.5)^2 + \ldots + (4.0 - 2.5)^2}{16}$$
$$= 0.79^2$$

Grundgesamtheit | Stichprobe



$\mu = 2.5$ | $\mu_{\bar{x}} = 2.5$

$\sigma = 1.12$ | $\sigma_{\bar{x}} = 0.79$

Standardfehler des Mittelwertes

- Ist die Standardabweichung aller möglichen Stichprobenmittelwerte $\bar{X}$.
  - Misst die Streuung unter allen möglichen (verschiedenen) Stichprobenmittelwerten.
- Kleiner als Standardabweichung der Grundgesamtheit (außer $n = 1$).
- Formel (für Ziehen mit Zurücklegen):

$$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

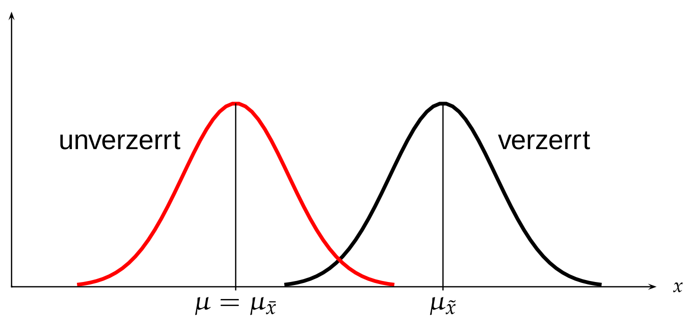
Eigenschaften der Stichprobenverteilung

- **Unverzerrtheit**
  Der Mittelwert der Stichprobenverteilung ist gleich dem Mittelwert der Grundgesamtheit (Erwartungswert).

- **Effizienz**
  Der Mittelwert der Stichprobenverteilung ist näher am Mittelwert der Grundgesamtheit als irgendein anderer unverzerrter Schätzer.

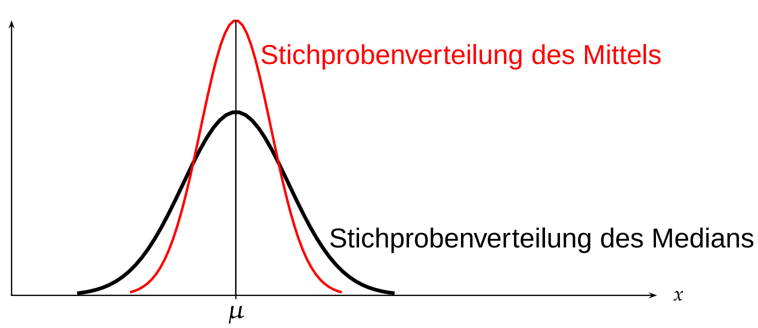
- **Konsistenz**
  Bei steigender Stichprobengröße wird die Abweichung des Stichprobenmittelwertes vom Mittelwert der Grundgesamtheit kleiner und geht mit n gegen unendlich gegen Null.

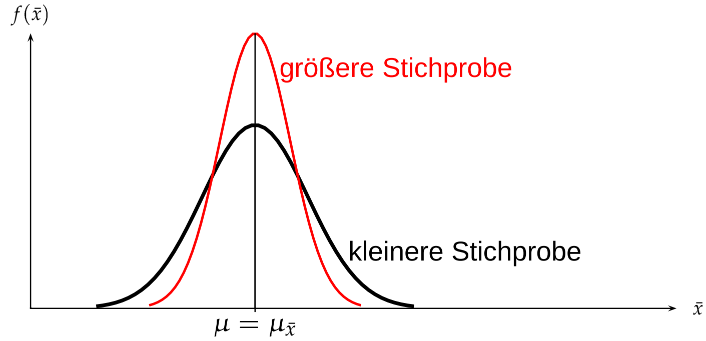
Unverzerrtheit



Vergleich zweier Schätzer für μ: $\bar{x}$ und $\tilde{x}$.

Effizienz (Vergleich bei festem n)





Normalverteilte Grundgesamtheit

Mittelwert:

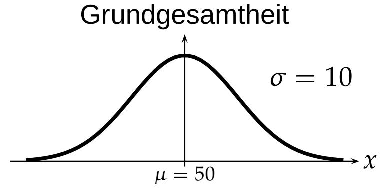
$$\boxed{\mu_{\bar{x}} = \mu}$$

Standardabweichung:

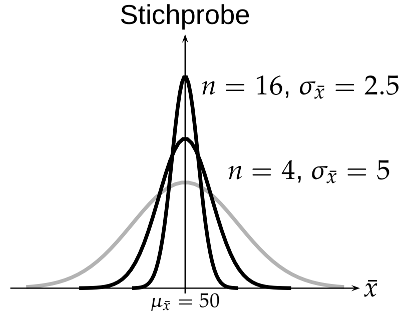
$$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

(Ziehen mit Zurücklegen, bzw. sehr große Grundgesamtheit)

Grundgesamtheit

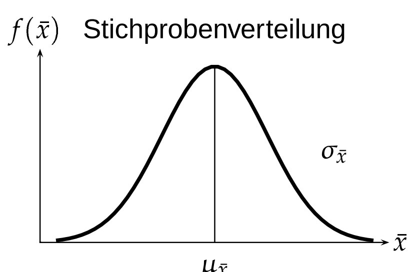


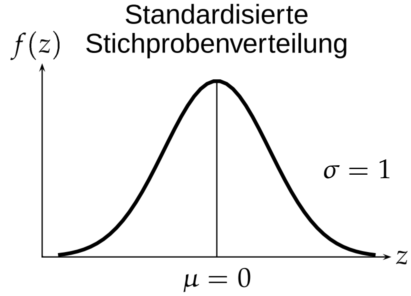
Stichprobe



Standardisieren der Stichprobenverteilung

$$\boxed{Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}$$





Ferngespräche, die von einer Telefongesellschaft abgewickelt werden, seien normalverteilt mit $\mu = 8$ min und $\sigma = 2$ min.

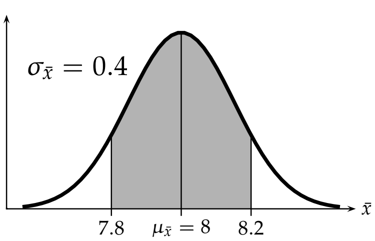
Wie groß ist die Wahrscheinlichkeit, dass die durchschnittliche Gesprächsdauer von 25 zufällig ausgewählten Telefonaten zwischen 7.8 und 8.2 Minuten liegt?

Stichprobenverteilung // Lösung

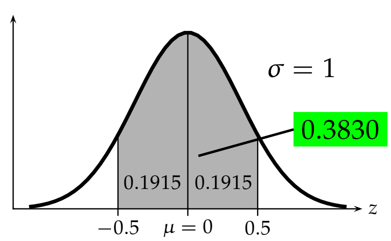
$$z_1 = \frac{\bar{x}_1 - \mu}{\sigma/\sqrt{n}} = \frac{7.8 - 8}{2/\sqrt{25}} = -0.5$$

$$z_2 = \frac{\bar{x}_2 - \mu}{\sigma/\sqrt{n}} = \frac{8.2 - 8}{2/\sqrt{25}} = 0.5$$

Stichprobenverteilung



Standardisierte Stichprobenverteilung



Nicht-Normalverteilte Grundgesamtheit

Mittelwert:

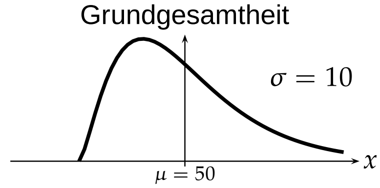
$$\boxed{\mu_{\bar{x}} = \mu}$$

Standardabweichung:

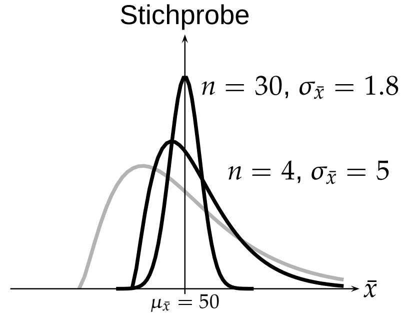
$$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

(Ziehen mit Zurücklegen, bzw. sehr große Grundgesamtheit)

Grundgesamtheit



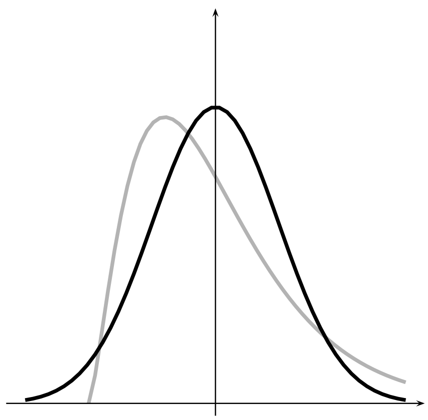
Stichprobe



Zentraler Grenzwertsatz

Zentraler Grenzwertsatz

Bei hinreichend großer Stichprobengröße ($n \gtrsim 30$) wird Stichprobenverteilung annäherend normalverteilt mit $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.



Zusammenfassung

1. Beschrieben Eigenschaften von Schätzern.
2. Beschrieben Stichprobenverteilungen.
3. Beschrieben Zusammenhang zwischen Grundgesamtheit und Stichprobenverteilung.
4. Formulierten zentralen Grenzwertsatz.