



Grundlagen der Inferenzstatistik I: Punkt und Intervallschätzungen


"Die Statistik ist das Märchen der Vernunft." Martin Kessel (1901 - 1990)

- 1 Einführung
- 2 Stichprobenkennwerteverteilungen
und Standardfehler
- 3 Punktschätzungen
- 4 Intervallschätzungen

Einführende Literatur

-  Bortz, J. & Schuster, Ch. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Auflage). Berlin: Springer. [Kap. 6]
-  Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden*. Weinheim: Beltz. [Kap. 8, v.a. 8.5]

Weiterführende Literatur

-  Levy, P. S. & Lemeshow, S. (2008). *Sampling of populations. Methods and applications* (4th ed.). Hoboken, NJ: Wiley.

Inferenzstatistik: Einführung

- Die **Inferenzstatistik (schließende Statistik)** beschäftigt sich mit der Frage, wie man aufgrund von Daten, die an einer Stichprobe erhoben wurden, Schlüsse auf die **Population** (d.h. die **Grundgesamtheit** aller Personen) ziehen kann.
- Wir wollen also verallgemeinerte Aussagen über die untersuchten Personen hinaus treffen. Dabei werden sich die Ergebnisse von Stichprobe zu Stichprobe unterscheiden und von den Ergebnissen in der Population abweichen (**Stichprobenfehler, sampling error**).
- Zwei Teilbereiche der Inferenzstatistik:
 - **Schluss auf Parameter in der Population**, z.B. wollen wir wissen, wie hoch die Lesekompetenzen (der Population) aller 15-jährigen Schüler in Deutschland im Durchschnitt sind. Der Punktwert, ermittelt in der PISA-Studie 2009 anhand einer Stichprobe von knapp 5000 Schülerinnen und Schülern, betrug 497.
 - **Hypothesentesten**, z.B. die Hypothese, dass sich Schülerinnen und Schüler in diesen Kompetenzen unterscheiden (**Unterschiedshypothese**; in der Stichprobe waren die Mädchen um 40 Punkte besser) oder dass ein Zusammenhang zwischen der Kompetenz und der sozialen Herkunft der Schülerinnen und Schüler besteht (**Zusammenhangshypothese**; $r = 0.35$ in der Stichprobe).

- Eine **Population** Ω ist die Menge aller N Person (bzw. allgemeiner: Untersuchungseinheiten, auch: eineiige Zwillinge, Ratten, Glühbirnen usw.), über die eine Aussage gemacht werden soll. Populationen können auch hypothetischer Natur sein.
- **Beispiele** für Populationen: alle Menschen; alle Patienten mit Lungenkrebs; alle Studierenden der Universität Osnabrück; alle Menschen, bei denen Therapie A angewendet wird bzw. alle, bei denen das nicht der Fall ist.
- Eine **Stichprobe** ist die Teilmenge der Population, die in die Untersuchung tatsächlich eingeht, d.h. diejenigen n Untersuchungseinheiten, an denen Messwerte erhoben wurden.
- Stichproben können auf verschiedene Weisen aus der Population gezogen werden. Wir gehen im Folgenden von der Ziehung einer **einfachen Zufallsstichprobe** aus (zu anderen Stichproben vgl. Vorlesung *Forschungsmethoden*): Aus einer Grundgesamtheit werden n Personen zufällig gezogen, d.h. die Wahrscheinlichkeit in die Stichprobe zu gelangen, ist für jede Person aus der Population gleich.
- Wir interessieren uns dabei für eine (oder auch mehrere) Variablen X . Man kann X als **Zufallsvariable** ansehen: Je nachdem, welches $\omega \in \Omega$ zufällig gezogen wird, nimmt X einen anderen Wert an. Dasselbe gilt für Kennwerte, die über Personen bestimmt werden, z.B. den Mittelwert über die n unabhängigen Zufallsvariablen X_1, X_2, \dots, X_n der Stichprobe.

- Nehmen wir an, wir interessieren uns für den Mittelwert (als Beispiel für einen Kennwert) eines Merkmals X in der Grundgesamtheit, haben dazu aus der interessierenden Population eine einfache Zufallsstichprobe der Größe n gezogen und den Mittelwert \bar{x} bestimmt.
- Wir interessieren uns also für den Schluss von dem Stichprobenkennwert \bar{x} auf den Populationsparameter μ . Diesen Schluss bezeichnet man auch als **Schätzen (estimation)**.
- Man schreibt auch $\hat{\mu} = \bar{x}$ („mue geschätzt“), und bei anderen Kennwerten entsprechend z.B. $\hat{\sigma}^2 = s^2, \hat{\rho} = r$.
- **Beispiel:** Wir interessieren uns für die durchschnittliche Intelligenz aller Studierenden in Deutschland. Dazu haben wir eine einfache Zufallsstichprobe von $n = 200$ Studierenden gezogen und bei ihnen die Intelligenz gemessen. Empirisch liegt nun der Mittelwert der Intelligenz dieser Stichprobe vor, z.B. $\bar{x} = 108$.

Stichprobenkennwerteverteilung

- Frage: Wie brauchbar ist der an der Stichprobe erhobene Mittelwert \bar{x} für den Schluss auf den Mittelwert μ in der Population? Wie gut schätzt \bar{x} den Parameter μ ?
- Wir stellen dazu folgende Überlegung an. Nehmen wir an, wir ziehen zufällig eine Stichprobe des Umfangs n und berechnen \bar{x} . Dann ziehen wir erneut zufällig eine Stichprobe derselben Größe n und berechnen wieder \bar{x} , usw. Die dabei resultierenden \bar{x} können mehr oder weniger stark untereinander und von μ abweichen.
- Ziehen wir theoretisch unendlich viele Stichproben einer konstanten Größe n , so bezeichnet man die resultierende theoretische Verteilung des an diesen Stichproben berechneten Kennwertes als **Stichproben(kennwerte)verteilung** (sampling distribution) des Kennwertes (hier: des Mittelwertes).

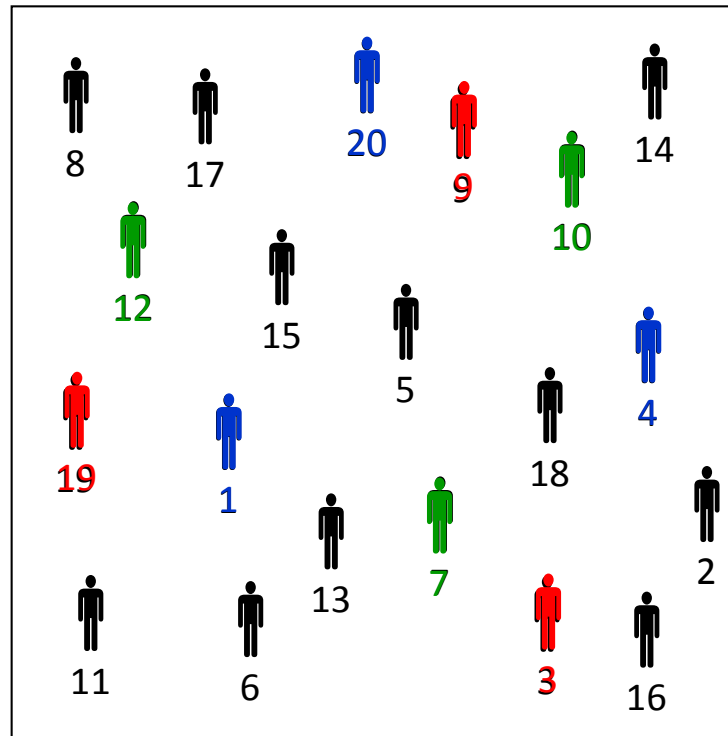
Stichprobenkennwerteverteilung

Variable X in Population mit $N = 20$ und Parametern $\mu = 100.0$ und $\sigma = 14.5$

Stichprobenziehung mit $n = 3$

Vp	X
1	96.4
2	101.6
3	71.0
4	88.5
5	99.2
6	101.3
7	98.9
8	108.0
9	129.3
10	98.9

Vp	X
11	116.7
12	94.9
13	117.4
14	106.0
15	82.8
16	109.2
17	112.3
18	109.9
19	88.0
20	70.5



Ziehung $k = 1$: $S = \{3, 9, 19\}$

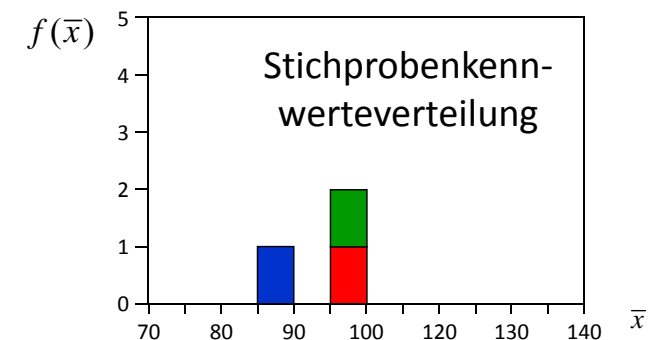
$$\bar{x}_1 = \frac{71.0 + 129.3 + 88.0}{3} = 96.10$$

Ziehung $k = 2$: $S = \{1, 4, 20\}$

$$\bar{x}_2 = \frac{96.4 + 88.5 + 70.5}{3} = 85.13$$

Ziehung $k = 3$: $S = \{7, 10, 12\}$

$$\bar{x}_3 = \frac{98.9 + 98.9 + 94.9}{3} = 97.57$$

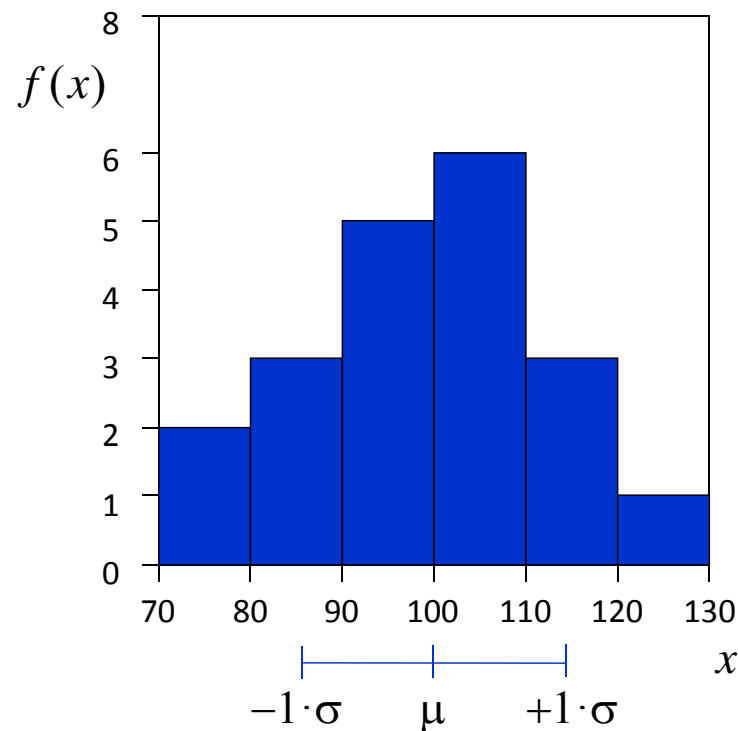


Stichprobenkennwerteverteilung

Variable X in Population mit $N = 20$ und Parametern $\mu = 100.0$ und $\sigma = 14.5$

Vp	X
1	96.4
2	101.6
3	71.0
4	88.5
5	99.2
6	101.3
7	98.9
8	108.0
9	129.3
10	98.9

Vp	X
11	116.7
12	94.9
13	117.4
14	106.0
15	82.8
16	109.2
17	112.3
18	109.9
19	88.0
20	70.5



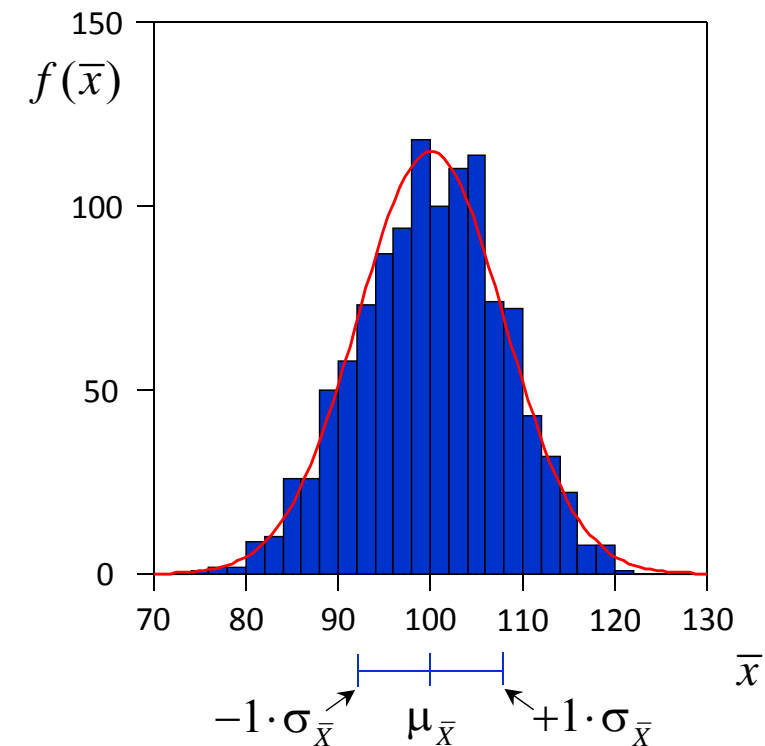
Stichprobenziehung mit $n = 3$

$$\binom{20}{3} = 1140$$

(...)

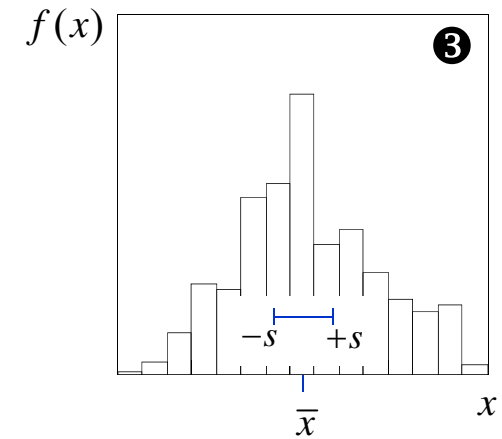
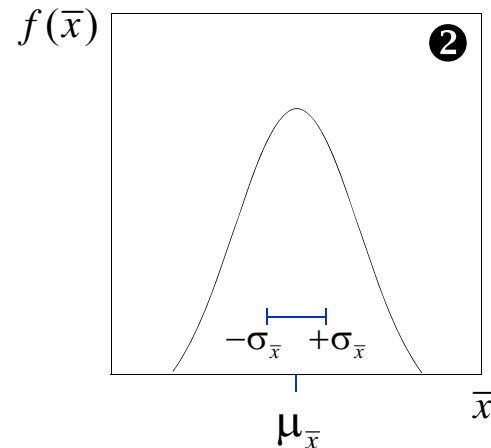
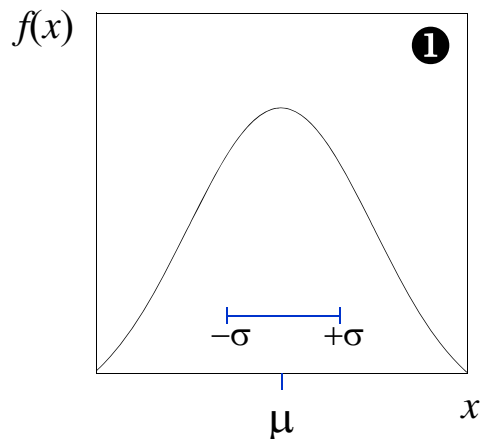
Ziehung $k = 1140$: $S = \{2, 6, 9\}$

$$\bar{x}_{1140} = \frac{101.6 + 101.3 + 129.3}{3} = 110.73$$



Stichprobenkennwerteverteilung

- Es ist wichtig, die folgenden drei Verteilungen nicht zu verwechseln:
- ➊ Die **Verteilung des Merkmals in der Population** mit ihrem Mittelwert μ und der Standardabweichung σ .
 - ➋ Die **Stichprobenkennwerteverteilung** eines bestimmten Kennwertes (z.B. des Mittelwertes, der Varianz, der Produkt-Moment Korrelation ...), die entstehen würde, wenn wir immer wieder Stichproben einer konstanten Größe n aus der Population ziehen und den Kennwert an der Stichprobe berechnen würden. Der Mittelwert wird hier als $\mu_{\bar{x}}$ und die Standardabweichung als $\sigma_{\bar{x}}$ bezeichnet.
 - ➌ Die **Verteilung des Merkmals in einer Stichprobe** mit Mittelwert \bar{x} und Standardabweichung s .



Stichprobenkennwerteverteilung

- Der **Erwartungswert** der Stichprobenkennwerteverteilung des Mittelwertes entspricht (unabhängig von n) dem Erwartungswert der Verteilung des Merkmals X in der Population, μ :

$$E(\bar{X}) = \mu_{\bar{x}} = \mu$$

- Die **Varianz** der Stichprobenkennwerteverteilung des Mittelwertes ist um den Faktor n kleiner als die Varianz der Verteilung des Merkmals X in der Population σ^2 :

$$\text{Var}(\bar{X}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

- Statt der Varianz der Stichprobenkennwerteverteilung des Mittelwertes verwendet man häufiger deren **Standardabweichung**, die auch als **Standardfehler** des Mittelwertes bezeichnet wird:

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

- Der Standardfehler ist ein Maß für die Präzision der Schätzung. Je kleiner $\sigma_{\bar{x}}$, desto genauer schätzt \bar{x} den Parameter μ . Die Präzision können wir durch die Erhöhung von n vergrößern (entsprechend den Stichprobenfehler verringern); auf σ haben wir keinen Einfluss.

Stichprobenkennwerteverteilung

- Wenn man also den Erwartungswert und die Varianz der Verteilung des Merkmals in der Population kennt, kann man also auch den Erwartungswert der Stichprobenkennwerteverteilung des Mittels sowie den Standardfehler bestimmen (ohne die Verteilungsform des Merkmals zu kennen).
- Was weiß man noch über die Verteilungsform der Stichprobenkennwerteverteilung?
- Ist die Verteilung des Merkmals in der Population eine Normalverteilung, so kann man zeigen, dass auch die Stichprobenkennwerteverteilung eine Normalverteilung aufweist (unabhängig von n):

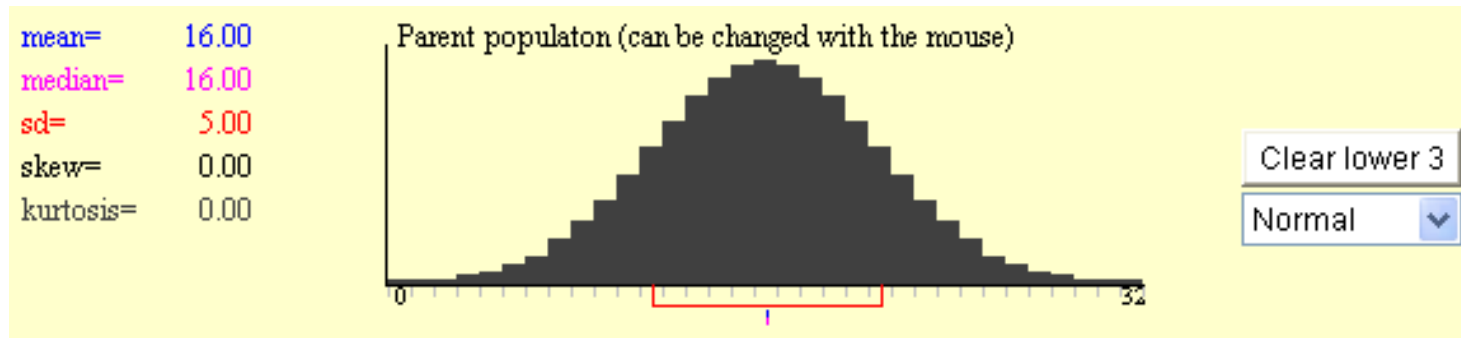
$$X \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N(\mu, \sigma_{\bar{x}})$$

- Ist die Verteilung des Merkmals in der Population keine Normalverteilung, so geht die Stichprobenkennwerteverteilung des Mittelwertes mit wachsendem n in eine Normalverteilung über (**zentrales Grenzwerttheorem** oder **zentraler Grenzwertsatz**, central limit theorem).

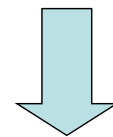
Ab $n > 30$ ist die Annäherung für die meisten Zwecke „ausreichend gut“.

Stichprobenkennwerteverteilung

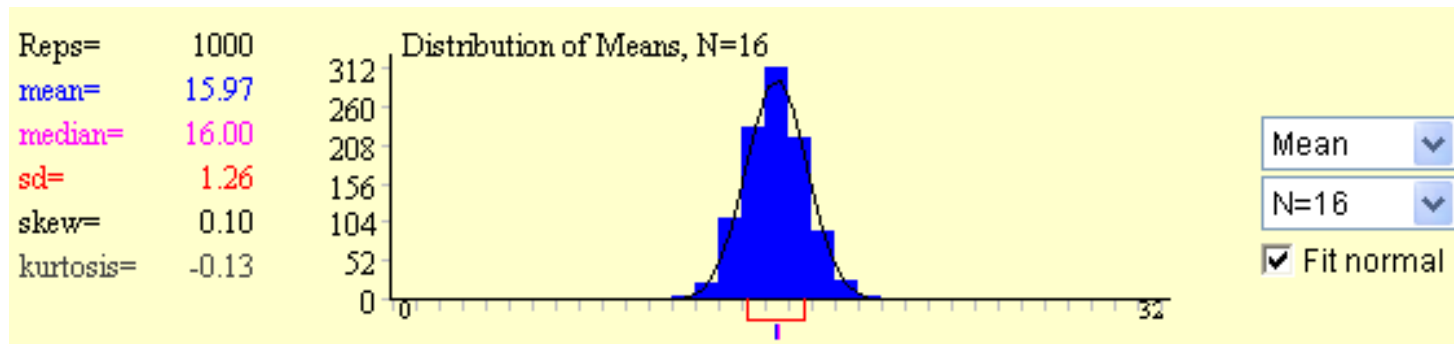
- **Beispiel 1:** Ziehen von 1000 Zufallsstichproben der Größe $n = 16$ aus einer normalverteilten Population mit $\mu = 16$ und $\sigma = 5$.



Verteilung des Merkmals in der Population $N(16,5)$



1000 Ziehungen des Stichprobenumfangs $n = 16$

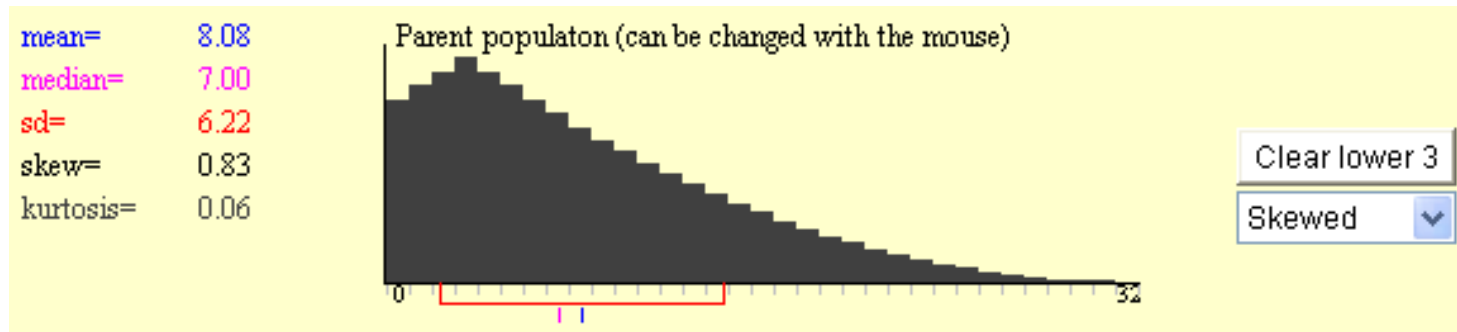


Stichprobenkennwerteverteilung des Mittels

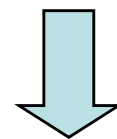
Simulationsprogramm unter: http://onlinestatbook.com/stat_sim/sampling_dist/index.html

Stichprobenkennwerteverteilung

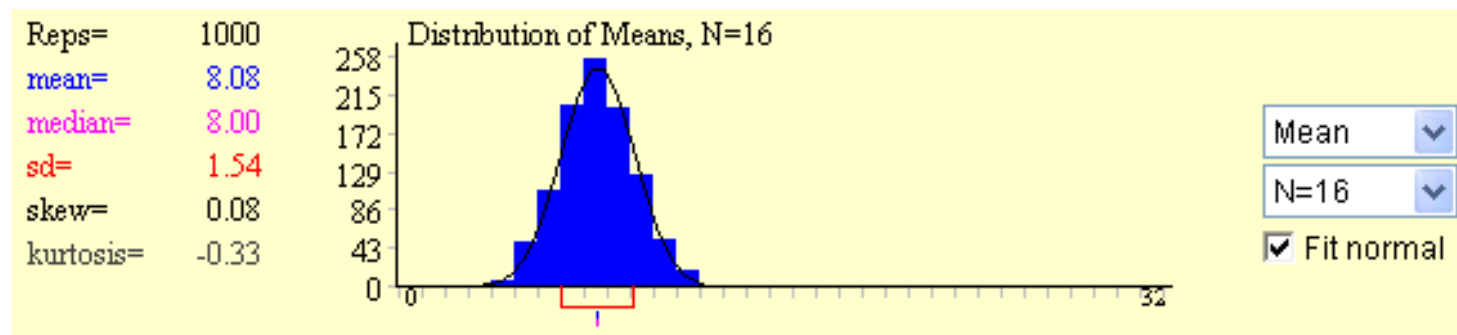
- **Beispiel 2:** Ziehen von 1000 Zufallsstichproben der Größe $n = 16$ aus einer Population mit einer linkssteilen (nicht normalverteilten) Verteilung mit $\mu = 8.1$ und $\sigma = 6.2$.



Verteilung des Merkmals in der Population



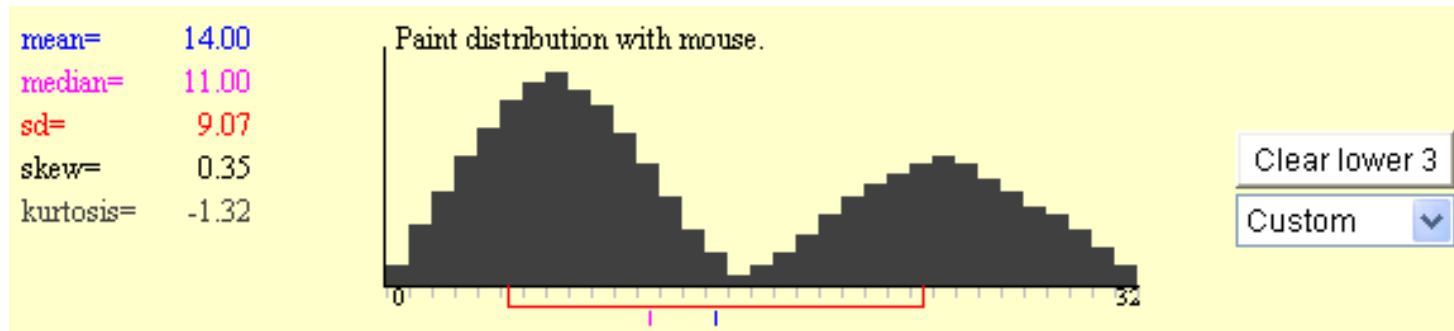
1000 Ziehungen des Stichprobenumfangs $n = 16$



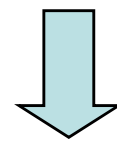
Stichprobenkennwerteverteilung des Mittels

Stichprobenkennwerteverteilung

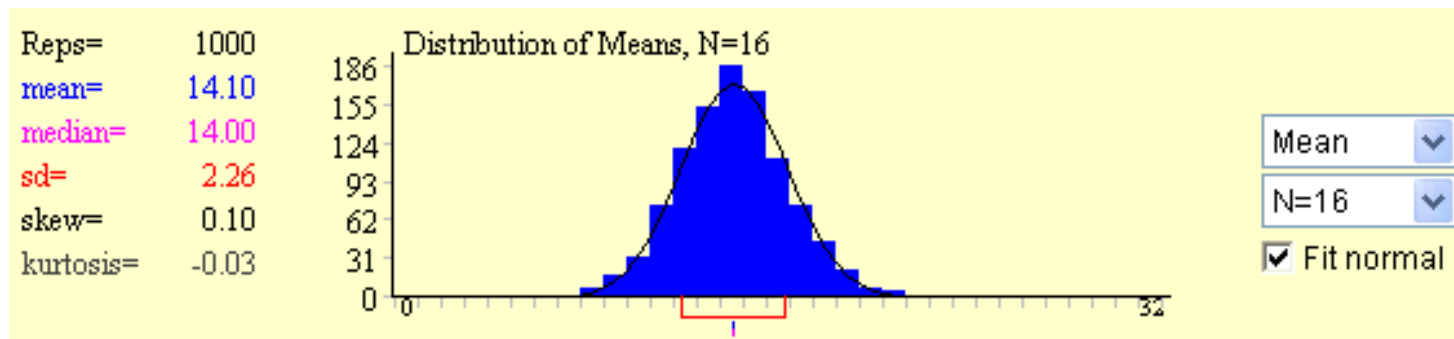
- **Beispiel 3:** Ziehen von 1000 Zufallsstichproben der Größe $n = 16$ aus einer Population mit einer bimodalen Verteilung mit $\mu = 14$ und $\sigma = 9.1$.



Verteilung des Merkmals in der Population

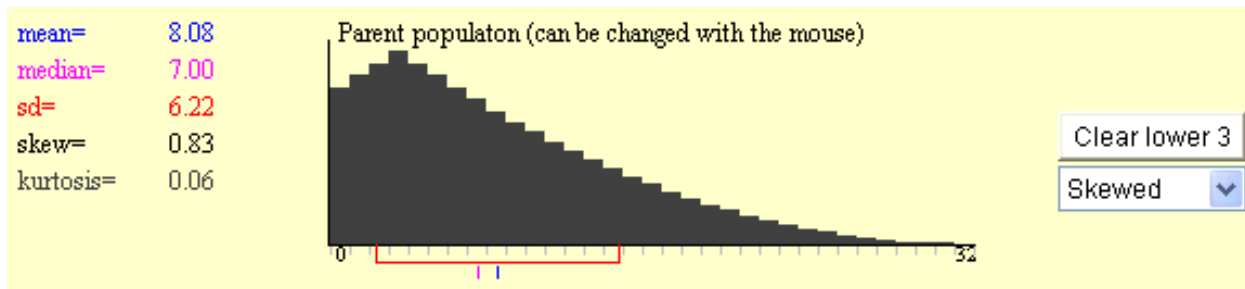


1000 Ziehungen des Stichprobenumfangs $n = 16$

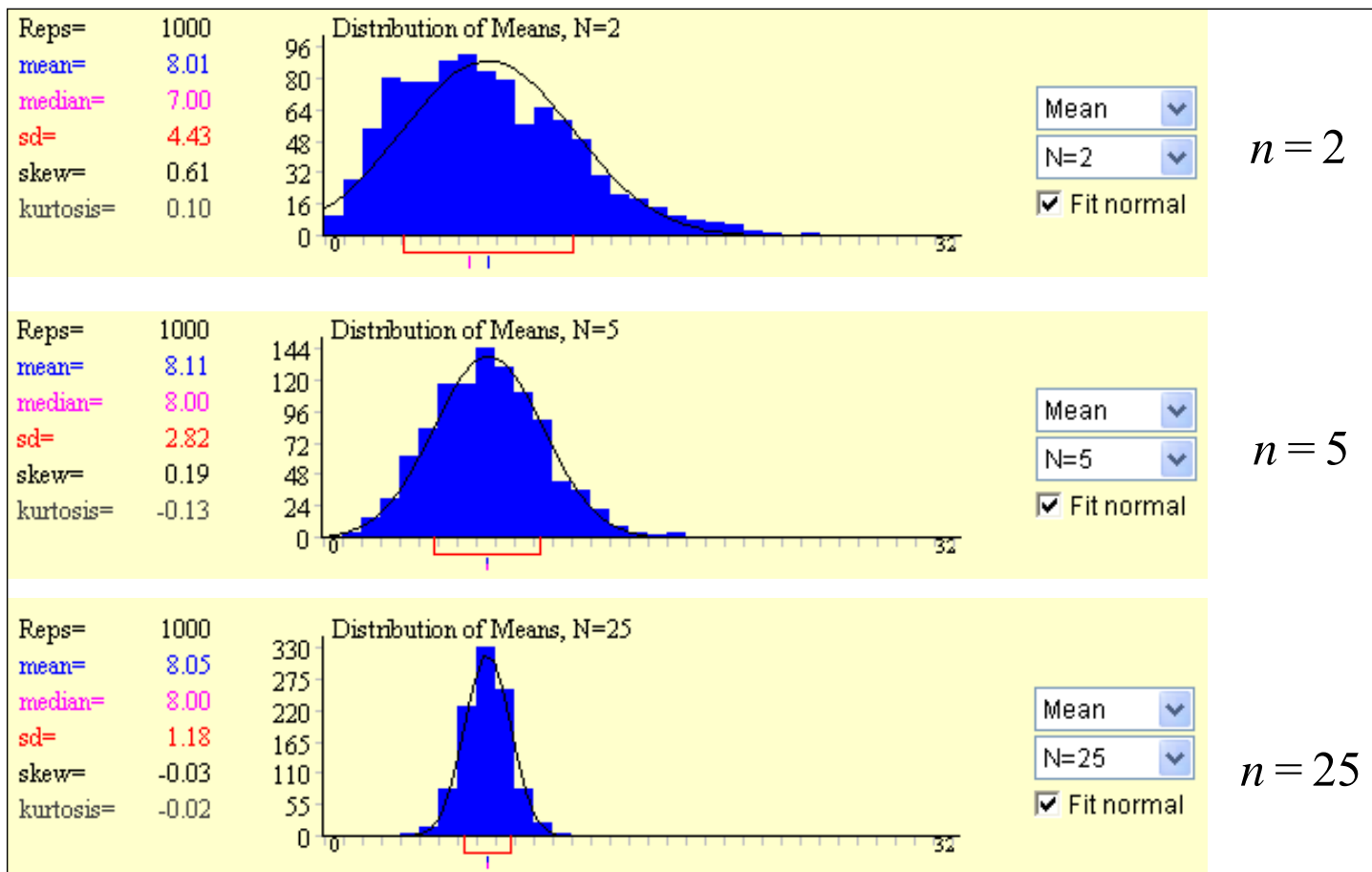


Stichprobenkennwerteverteilung des Mittels

Stichprobenkennwerteverteilung



Beispiel 2: Linkssteile
Verteilung des
Merkmals in der
Population



Annäherung der
Verteilungsform der
Stichprobenkenn-
werteverteilung
des Mittels an die
Normalverteilung
mit wachsendem n .

Standardfehler

- Kennen wir (wie fast immer) den Erwartungswert und die Standardabweichung in der Population nicht, so lassen sich beide durch die Stichprobenkennwerte \bar{x} und s schätzen. Entsprechend resultiert dann als Schätzung für den Standardfehler des Mittelwertes $s_{\bar{x}}$:

$$\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

- Für andere Kennwerte können ebenfalls Standardfehler abgeleitet werden. So wird z.B. als Standardfehler für eine relative Häufigkeit $\hat{\pi} = p = h_k / n$ des Auftretens eines Ereignisses k (den Anteil p) bei ausreichend großem n häufig herangezogen (vgl. Walter, 2010):

$$\hat{\sigma}_p = \sqrt{\frac{\hat{\pi} \cdot (1 - \hat{\pi})}{n}} = s_p = \sqrt{\frac{p \cdot (1 - p)}{n}}$$

- **Beispiel:** Bei einer Meinungsumfrage bei $n = 800$ Personen haben 20% der Befragten eine bestimmte Frage positiv beantwortet. Der Standardfehler für diese Variable beträgt:

$$s_p = \sqrt{\frac{0.20 \cdot (1 - 0.20)}{800}} = 0.014$$

- Für andere Arten der Ziehung von Stichproben (z.B. die geschichtete Zufallsauswahl) ergeben sich ebenfalls andere Standardfehler (vgl. Levy & Lemeshow, 2008).

Standardfehler

- Bei der Ableitung des Standardfehlers wird eine **Zufallsziehung mit Zurücklegen** zugrunde gelegt (bzw. eine **unendlich große Population**). In der Regel ziehen wir aber Stichproben ohne Zurücklegen aus endlichen Populationen, da wir nicht Personen mehrfach in der Stichprobe haben (und untersuchen) wollen.
- Für den Fall, dass die Population „sehr viel“ größer als die Stichprobe ist, also das Verhältnis n / N hinreichend klein, so spielt das praktisch keine große Rolle. (Das ist meistens der Fall.)
- Ist hingegen $n / N > 0.05$ (sind also mehr als 5% der Population in der Stichprobe enthalten), so sollte der Standardfehler des Mittelwertes mit dem folgenden Korrekturfaktor multipliziert werden (**Endlichkeitskorrektur**; finite population correction, fpc):

$$f = \sqrt{\frac{N-n}{N-1}} < 1 \text{ bei } n > 1$$

- **Beispiel:** In einer Mitarbeiterbefragung in einem Unternehmen wird die Hälfte aller 200 Beschäftigten danach befragt, wie zufrieden sie mit ihrer Bezahlung sind. Die Standardabweichung der Antworten auf die Frage beträgt $s = 1.2$.

Durch die Endlichkeitskorrektur verringert sich der Standardfehler von $s_{\bar{x}} = s / \sqrt{n} = 1.2 / \sqrt{100} = 0.12$ auf $0.12 \cdot 0.71 = 0.085$.

$$f = \sqrt{\frac{200-100}{200-1}} = 0.71$$

Qualitätsmerkmale von Punktschätzern

- Die an einer Stichprobe gewonnenen Kennwerte wie f_k , \bar{x} , s^2 , r etc. werden also nicht nur zur Beschreibung einer Stichprobe verwendet, sondern auch zur Schätzung von Parametern π_k , μ , σ^2 , ρ etc. der Population.
- Bei der Entscheidung, welcher Kennwert am besten zur Schätzung eines Parameters geeignet ist, können folgende Qualitätskriterien herangezogen werden: **Konsistenz**, **Erwartungstreue**, **Effizienz** und **Suffizienz**.
- Ein Kennwert ist **konsistent**, wenn er sich bei wachsendem Stichprobenumfang n immer mehr dem zu schätzenden Parameter annähert. Dies ist, anders ausgedrückt, dann der Fall, wenn der Standardfehler des Schätzers für $n \rightarrow \infty$ gegen 0 geht.
- **Beispiel:** \bar{x} , s und s^2 sind konsistente Schätzer für μ , σ und σ^2 . Ebenso $f_k = h_k / n$ für π_k .

Qualitätsmerkmale von Punktschätzern

- Ein Schätzer ist **erwartungstreu (unverzerrt, unbiased)**, wenn der Erwartungswert der Stichprobenkennwerteverteilung des Schätzers dem Parameter entspricht. Eine bestehende Abweichung von beiden bezeichnet man als **Verzerrung** oder **Bias**. Tritt Bias auf, so wird der Parameter also systematisch über- oder unterschätzt.
- **Beispiel:** \bar{x} und s^2 sind erwartungstreue Schätzer für μ und σ^2 . Bei einer symmetrischen Populationsverteilung ist auch der *Md* ein erwartungstreuer Schätzer für μ .
- **Gegenbeispiel:** Die **mittlere quadratische Abweichung** zur Erfassung der Variabilität

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

ist keine erwartungstreue Schätzung für σ^2 . Der Verschätzungsfaktor beträgt $n / (n - 1)$. (*MQA* ist aber ein konsistenter Schätzer, da für $n \rightarrow \infty$ der Bias gegen 0 geht.)

- Daher haben wir in der deskriptiven Statistik auch die Varianz einer Stichprobe definiert als

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

Qualitätsmerkmale von Punktschätzern

- Die **Effizienz** eines Schätzers gibt an, wie genau bzw. präzise ein Kennwert den Parameter schätzt. Bei erwartungstreuen Schätzern ist der Standardfehler (oder die Varianz der Stichprobenkennwerteverteilung) ein Maß für die Effizienz.
- **Beispiel:** Bei vorliegender Normalverteilung in der Population sind der Mittelwert \bar{x} und der Median Md erwartungstreue und konsistente Schätzer des Populationsmittelwertes μ . Die Varianz der Stichprobenkennwerteverteilung des Medians ist aber um den Faktor 1.56 größer als die des Mittelwertes. Der Mittelwert ist also der effizientere Schätzer.
- Zum Vergleich wird häufiger die relative Effizienz zweier erwartungstreuer Schätzer a und b durch den Quotienten der beiden quadrierten Standardfehler $\sigma_a^2 / \sigma_b^2 \cdot 100\%$ ausgedrückt. Im obigen Beispiel ist dann die relative Effizienz des Medians zum Mittelwert $1 / 1.56 \cdot 100\% = 64\%$. Dies kann so interpretiert werden, dass der Median einer Stichprobe von $n = 100$ den Parameter μ genauso präzise schätzt wie der Mittelwert einer Stichprobe von $n = 64$.
- **Beispiel** für $\sigma = 15, n = 50$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{15^2}{50} = 4.50$$

$$\sigma_{Md}^2 = \frac{\sigma^2}{n} \cdot 1.56 = 7.08$$

$$\frac{\sigma_{\bar{x}}^2}{\sigma_{Md}^2} = \frac{4.50}{7.08} = 0.64$$

Qualitätsmerkmale von Punktschätzern, Intervallschätzung

- Ein Schätzer wird dann als **suffizient** (erschöpfend) bezeichnet, wenn er alle in der Stichprobe vorliegenden Informationen bezüglich des Parameters verwendet.
- **Beispiel:** \bar{x} und s^2 sind suffiziente Schätzer für μ und σ^2 , da alle n Werte in ihren konkreten Ausprägungen in die Berechnungen eingehen. Dies trifft beispielweise nicht für den Median zu.
- Die Schätzung von Populationsparametern durch einen einzelnen Kennwert/Schätzer aus den Stichprobendaten bezeichnet man auch als **Punktschätzung**.
- Wenngleich diese Schätzungen hinsichtlich der genannten Kriterien die bestmöglichen sind, so werden sie doch von Stichprobe zu Stichprobe Schwankungen unterliegen.
- Daher ist es informativer, einen Bereich anzugeben, in dem der unbekannte Parameter mit einer gewissen statistischen Sicherheit liegt. Einen solchen Bereich bezeichnen wir als **Konfidenzintervall** (**Vertrauensintervall**, Mutungsintervall) und sprechen dann auch von **Intervallschätzung**.

- 1 Einführung
- 2 Stichprobenkennwerteverteilungen
und Standardfehler
- 3 Punktschätzungen
- 4 Intervallschätzungen

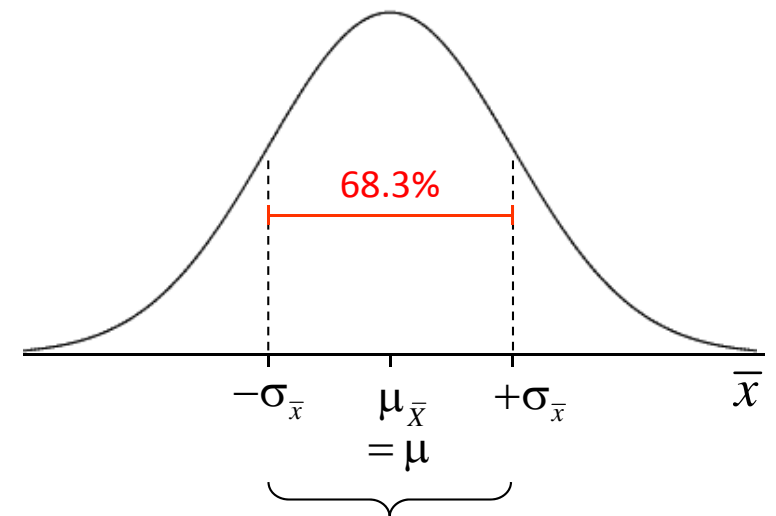
Intervallschätzung

- Bei der Intervallschätzung will man beispielsweise wissen, innerhalb welcher Grenzen der Parameter μ mit einer gewissen Wahrscheinlichkeit liegt.
- Gehen wir zunächst davon aus, das betrachtete Merkmal X sei normalverteilt und die Standardabweichung σ bekannt. Dann ist die Stichprobenkennwerteverteilung ebenfalls normalverteilt mit Erwartungswert μ und Standardfehler $\sigma_{\bar{x}} = \sigma / \sqrt{n}$.
- Dann wissen wir ferner aus der Kenntnis der Eigenschaft der Normalverteilung, dass zwischen plus und minus $\sigma_{\bar{x}}$ um μ 68.3% der Fläche liegen. In 68.3% der Fälle fällt also ein Stichprobenmittelwert in diesen Bereich, was man formal schreiben kann als

$$P(\mu - \sigma_{\bar{x}} \leq \bar{x} \leq \mu + \sigma_{\bar{x}}) = 0.683$$

- Stellen wir innerhalb der Klammer um, so gilt auch:

$$P(\bar{x} - \sigma_{\bar{x}} \leq \mu \leq \bar{x} + \sigma_{\bar{x}}) = 0.683$$



In 68.3% der Fälle fällt \bar{x} in das Intervall $\mu \pm \sigma_{\bar{x}}$

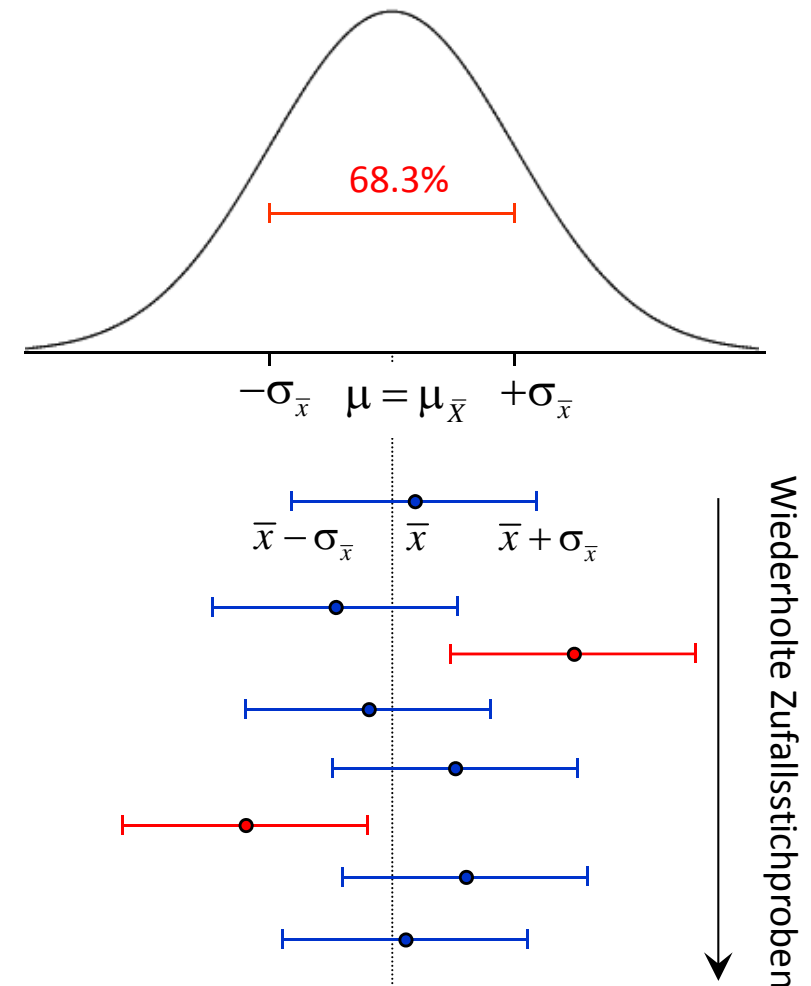
Intervallschätzung

- Anschaulich bedeutet

$$P(\bar{x} - \sigma_{\bar{x}} \leq \mu \leq \bar{x} + \sigma_{\bar{x}}) = 0.683$$

dass wenn \bar{x} in dieses Intervall fällt (in 68.3% der Fälle), \bar{x} niemals weiter als $\sigma_{\bar{x}}$ von μ entfernt ist, also in diesen 68.3% der Fälle μ im Intervall $\bar{x} \pm \sigma_{\bar{x}}$ liegt.

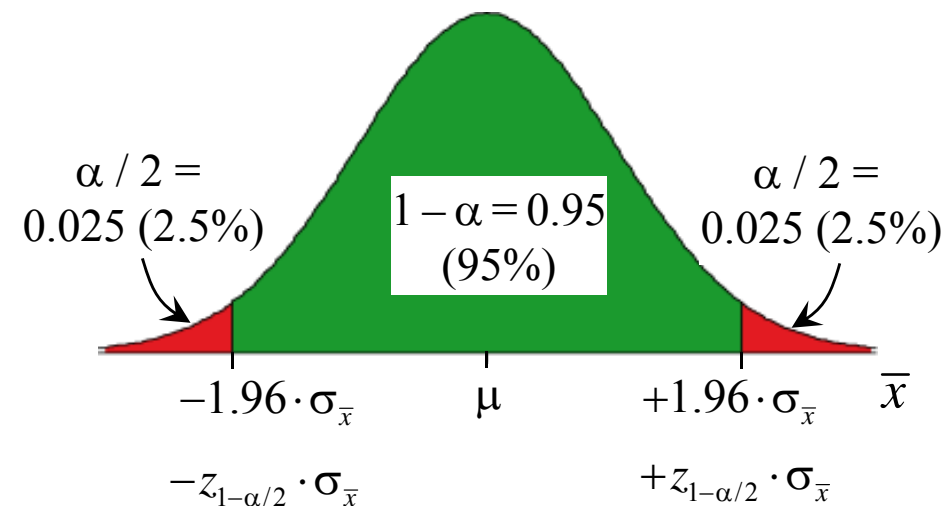
- Zieht man also immer wieder eine Zufallsstichprobe und berechnet \bar{x} und das Intervall $[\bar{x} - \sigma_{\bar{x}}, \bar{x} + \sigma_{\bar{x}}]$, so liegt μ in 68.3% der Fälle in diesem Intervall.



In 68.3% der Fälle fällt \bar{x} in das Intervall $\mu \pm \sigma_{\bar{x}}$ und μ in das Intervall $\bar{x} \pm \sigma_{\bar{x}}$

Intervallschätzung

- In der Regel wird man wollen, dass μ mit einer größeren Wahrscheinlichkeit als 68.3% im Vertrauensintervall liegt. Üblicherweise wählt man als **Konfidenzkoeffizienten** einen höheren Wert, z.B. 95% und entsprechend als Wahrscheinlichkeit, mit der der Parameter außerhalb des Intervalls liegt, auch als **Irrtumswahrscheinlichkeit** bezeichnet, $\alpha = 0.05$ (5%).
- Zwischen welchen Werten liegen 95% der Fläche einer Normalverteilung? Der Standardnormalverteilungstabelle entnimmt man, dass außerhalb der z -Werte von ± 1.96 jeweils 2.5% der Fläche liegen ($z_{0.975} = 1.96 = -z_{0.025}$).



- Entsprechend gilt für das 95%-Konfidenzintervall:

$$P(\bar{x} - 1.96 \cdot \sigma_{\bar{x}} \leq \mu \leq \bar{x} + 1.96 \cdot \sigma_{\bar{x}}) = 0.95$$

bzw. allgemein

$$P(\bar{x} - z_{1-\alpha/2} \cdot \sigma_{\bar{x}} \leq \mu \leq \bar{x} + z_{1-\alpha/2} \cdot \sigma_{\bar{x}}) = 1 - \alpha$$

- Das Konfidenzintervall wird immer breiter, also die Vorhersage unpräziser, je kleiner die Irrtumswahrscheinlichkeit α , je kleiner n und je größer die Streuung des Merkmals σ ist.

- **Beispiel:** Bei einem Intelligenztest seien die Werte in der Population normalverteilt mit $\sigma = 15$. In einer Stichprobe von 400 Studierenden wird ein mittlerer IQ-Wert von 108 bestimmt. Wie groß ist das 95%-Konfidenzintervall?

Für den Konfidenzkoeffizienten 95% mit $\alpha = 0.05$ ist $z_{1-\alpha/2} = 1.96$ (z-Wert für p-Wert 0.975)

Ferner ist: $\sigma_{\bar{x}} = \sigma / \sqrt{n} = 15 / \sqrt{400} = 0.75$

$$z_{1-\alpha/2} \cdot \sigma_{\bar{x}} = 1.96 \cdot 0.75 = 1.47$$

$$\bar{x} \pm z_{1-\alpha/2} \cdot \sigma_{\bar{x}} = 108 \pm 1.47 \Rightarrow [106.53, 109.47]$$

Man kann also schließen, dass bei einer großen Zahl von Wiederholungen dieser Studie mit $n = 400$ etwa 95% der berechneten Konfidenzintervalle mit 1.47 Intelligenzpunkten um \bar{x} den Mittelwert μ der Population enthalten. Man sagt auch, der Mittelwert der Population liegt mit einer 95%igen **Sicherheit** in dem Intervall $[106.53, 109.47]$.

- **Beispiel:** Wie sieht das Konfidenzintervall bei 99%iger Sicherheit aus?

Für den Konfidenzkoeffizienten 99% mit $\alpha = 0.01$ ist $z_{1-\alpha/2} = 2.58$ (z-Wert für p-Wert 0.995)

$$\text{Entsprechend ist } \bar{x} \pm z_{1-\alpha/2} \cdot \sigma_{\bar{x}} = 108 \pm 2.58 \cdot 0.75 \Rightarrow [106.07, 109.94]$$

Intervallschätzung

- Ist die Variable in der Population normalverteilt, die **Standardabweichung** σ aber **unbekannt**, so lässt sich das Konfidenzintervall wie folgt bestimmen:

$$\bar{x} \pm t_{n-1;1-\alpha/2} \cdot \hat{\sigma}_{\bar{x}} = \bar{x} \pm t_{n-1;1-\alpha/2} \cdot s_{\bar{x}} = \bar{x} \pm t_{n-1;1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$t_{n-1;1-\alpha/2}$ bezeichnet den t -Wert einer t -Verteilung, der bei $df = n - 1$ Freiheitsgraden $1 - \alpha/2$ der Verteilung abschneidet.

- **Beispiel:** An einer Stichprobe von $n = 25$ Bewerbern auf einen Job als Auszubildende in der Metallverarbeitung wurde das mechanisch-technische Verständnis erfasst. Es ergab sich ein Mittelwert von 24.4 und eine Standardabweichung von 5.0. Das Merkmal sei in der Population normalverteilt. Wie groß ist das 99%-Konfidenzintervall?

Für $\alpha = 0.01$ ist $t_{25-1;1-\alpha/2} = 2.80$ (t -Wert für p -Wert 0.995 und $df = 24$ aus Tabelle für t -Verteilung).

Entsprechend ist:

$$24.4 \pm 2.80 \cdot \frac{5.0}{\sqrt{25}} \Rightarrow [21.6, 27.2]$$

Intervallschätzung

- Aufgrund des zentralen Grenzwertsatzes geht bei großem n ($n > 30$) die Stichprobenkennwerteverteilung des Mittelwertes in eine Normalverteilung über, so dass in diesem Fall ohne Annahmen über die Verteilungsform des Merkmals das Konfidenzintervall wie folgt bestimmt werden kann (bei unbekanntem σ):

$$\bar{x} \pm z_{1-\alpha/2} \cdot \hat{\sigma}_{\bar{x}} = \bar{x} \pm z_{1-\alpha/2} \cdot s_{\bar{x}} = \bar{x} \pm z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

- Ist die Stichprobe klein, die Verteilung des Merkmals nicht normal und σ unbekannt, müssen hier nicht näher dargestellte alternative Methoden der Konfidenzintervallbestimmung herangezogen werden (vgl. z.B. Walter, 2010).

- Konfidenzintervalle lassen sich auch für andere Statistiken, z.B. Varianzen, Korrelationen etc. bestimmen. Das **Konfidenzintervall um den Anteil** $\hat{\pi} = p = h/n$ lautet bei ausreichend großem n :

$$\hat{\pi} \pm z_{1-\alpha/2} \cdot \hat{\sigma}_p = p \pm z_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

Anders als beim Mittelwert hängt die Breite des Konfidenzintervalls hier vom Kennwert ab und ist am größten, wenn $p \cdot (1-p)$ maximal ist. Dies ist bei $p = 0.5$ der Fall.

Intervallschätzung

- Man kann die obigen Zusammenhänge auch nutzen, um für eine gewünschte Präzision und Irrtumswahrscheinlichkeit den erforderlichen Stichprobenumfang zu bestimmen. Die Breite des Konfidenzintervalls beträgt beim Mittelwert: $2 \cdot z_{1-\alpha/2} \cdot \sigma / \sqrt{n}$
- Soll diese Breite maximal Δ betragen, so erhält man durch Auflösen nach n :

$$n \geq \left(\frac{2 \cdot z_{1-\alpha/2} \cdot \sigma}{\Delta} \right)^2 \quad \text{bzw. bei Schätzung durch } s: \quad n \geq \left(\frac{2 \cdot z_{1-\alpha/2} \cdot s}{\Delta} \right)^2 = \frac{4 \cdot z_{1-\alpha/2}^2 \cdot s^2}{\Delta^2}$$

- bzw. analog für den Anteil p :


$$n \geq \frac{4 \cdot z_{1-\alpha/2}^2 \cdot p \cdot (1-p)}{\Delta^2}$$

- **Beispiel:** Wie groß muss die Stichprobe in einer Meinungsumfrage sein, damit die Prozentwerte (bei $\alpha = 0.05$) maximal eine Fehlermarge von ± 3 Prozentpunkten aufweisen?

Wir gehen von dem ungünstigsten Fall aus und setzen $p = 0.50$ ein:

$$n \geq \frac{4 \cdot 1.96^2 \cdot 0.50 \cdot (1-0.50)}{0.06^2} = 1067$$

Zitierte Quellen

-  Walter, O. B. (2010). Konfidenzintervalle. In H. Holling & B. Schmitz (Hrsg.), *Handbuch Statistik, Methoden und Evaluation* (S. 335-349). Göttingen: Hogrefe.

Änderungen am 13.12.2016

- Folie 18: Notation für Standardfehler des Anteils geändert.
- Folien 30 & 31: Notation entsprechend angepasst.