

Statistik: Prinzip des Konfidenzintervalls

Aus Wikibooks

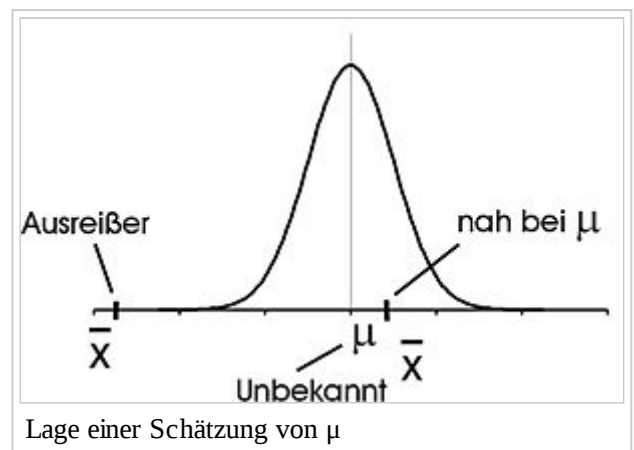
Inhaltsverzeichnis

- 1 Beispiel mit Absatz von Kaffeepaketen
 - 1.1 Plausible Überlegungen zur Schätzung
 - 1.2 Verteilung des Merkmals und der Schätzfunktion
 - 1.3 Herleitung des Intervalls
 - 1.4 Konkretes 95%-Konfidenzintervall
 - 1.5 Einfluss der Varianz auf das Konfidenzintervall
 - 1.6 Mindest erforderlicher Stichprobenumfang
 - 1.7 90%-Konfidenzintervall

Beispiel mit Absatz von Kaffeepaketen

Beispiel:

Eine Kaffeerösterei möchte eine neue Röstanlage anschaffen. Um beurteilen zu können, ob die Firma den aufzunehmenden Kredit tilgen kann, braucht sie Informationen über den durchschnittlichen monatlichen Absatz an Kaffeepaketen. Pro Monat muss die Firma 20.000 € Annuität zahlen. Zusammen mit den Produktionskosten sollte sie im Durchschnitt auf einen Absatz von 100.000 Kaffeepaketen im Monat kommen. Die Frage ist nun, wird sie es schaffen?



Plausible Überlegungen zur Schätzung

Der durchschnittliche monatliche Absatz von Kaffeepaketen ist unbekannt. Wie könnte man den Durchschnitt ermitteln? Man könnte eine Stichprobe mit z.B. $n = 50$ Beobachtungen ziehen und versuchen, aus dem arithmetischen Mittel \bar{x} auf den durchschnittlichen monatlichen Absatz der Grundgesamtheit zu schließen. Ist die Stichprobe groß genug, kann man vermuten, dass der Durchschnitt $E\bar{X}$ in der Grundgesamtheit, hier μ , in der Nähe von \bar{x} liegen müsste. Meistens wird \bar{x} in der Nähe von μ liegen, da aber \bar{x} die Realisation einer Zufallsvariablen ist, kann in sehr wenigen Fällen \bar{x} auch extrem weit von μ weg liegen, so daß man dann μ verkehrt einschätzt.

Wir betrachten nun den monatlichen Absatz von Kaffeepaketen (in 1000). Wir bezeichnen ihn als Zufallsvariable X . Es soll der monatliche durchschnittliche Absatz der Kaffeepäckchen geschätzt werden. Bekannt ist lediglich, dass die Zahl der verkauften Kaffeepakete normalverteilt ist mit einer Varianz 200 $[1000^2 \text{ Stück}^2]$.

Wie sollen wir nun μ eingrenzen? Wir könnten etwa ein Intervall bestimmen, in dem z.B. 95% aller möglichen \bar{x} -Werte liegen, also

$$P(\bar{x}_u \leq \bar{X} \leq \bar{x}_o) = 0,95.$$

Damit man dieses Intervall berechnen kann, müssen Informationen über die Verteilung von \bar{X} verfügbar sein. Es soll eine Stichprobe von $n = 50$ gezogen werden, d.h. es werden die verkauften Kaffeepakete der letzten 50 Monate erfasst:

$$\bar{X} = \frac{1}{50} \sum_{i=1}^{50} X_i .$$

Verteilung des Merkmals und der Schätzfunktion

Die Zufallsvariable X in der Grundgesamtheit soll normalverteilt sein mit dem Durchschnitt $EX = \mu$ und der Varianz $\text{Var}X = \sigma^2$. Die Varianz soll bekannt sein.

Es wird eine Stichprobe vom Umfang n gezogen. Der Stichprobendurchschnitt \bar{X} ist selbst eine Zufallsvariable und ist als lineare Transformation von X wiederum normalverteilt, und zwar mit den Parametern

$$E\bar{X} = \mu \text{ und } \text{var}\bar{X} = \frac{\sigma^2}{n} .$$

Hier ist

$$\frac{\sigma^2}{n} = \frac{200}{50} = 4 .$$

Herleitung des Intervalls

Ausgegangen wird von

$$P(\bar{x}_u \leq \bar{X} \leq \bar{x}_o) = 0,95 .$$

Untergrenze u und Obergrenze o sollen nun bestimmt werden. Wir standardisieren zunächst

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{X} - \mu}{\sqrt{4}} ,$$

so dass sich analog zu oben

$$P(z_u \leq Z \leq z_o) = 0,95$$

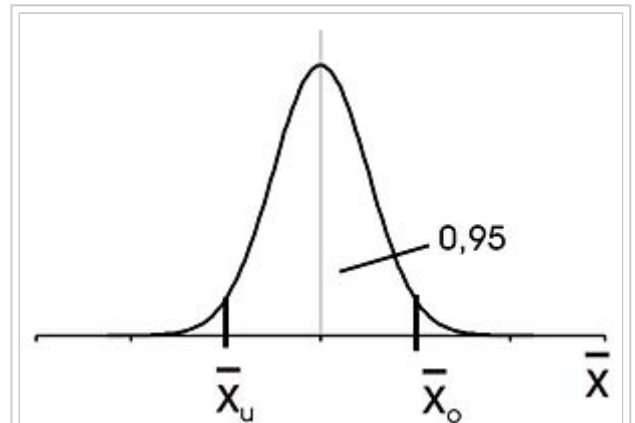
ergibt. z_o ist hier das 0,975-Quantil der Standardnormalverteilung. Ein Blick in die Normalverteilungstabelle verrät uns, dass der z -Wert, der zur Wahrscheinlichkeit 0,975 gehört, 1,96 ist.

Wir können jetzt das entsprechende Intervall für Z

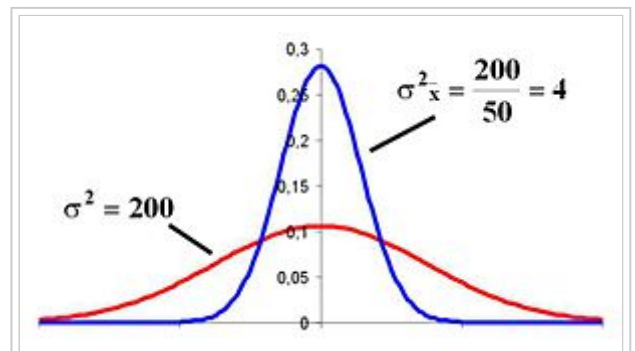
$$P(-1,96 \leq Z \leq 1,96) = 0,95$$

angeben. Die Ungleichung wird bezüglich μ aufgelöst:

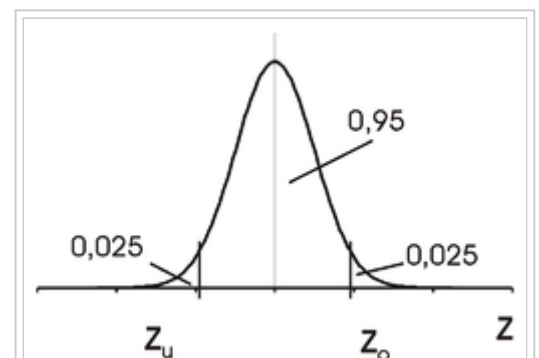
$$P\left(-1,96 \leq \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq 1,96\right) = 0,95 .$$



95%-Intervall des durchschnittlichen monatlichen Absatzes



Vergleich: Normalverteilung der Zufallsvariablen Absatz X und der Zufallsvariablen Durchschnittlicher Absatz \bar{X}



Ober- und Untergrenze der standardnormalverteilten Zufallsvariablen Z

$$P\left(-1,96 \leq \frac{\bar{X} - \mu}{\sqrt{4}} \leq 1,96\right) = 0,95.$$

$$P(-1,96 \cdot 2 \leq \bar{X} - \mu \leq 1,96 \cdot 2) = 0,95.$$

$$P(-\bar{X} - 1,96 \cdot 2 \leq -\mu \leq -\bar{X} + 1,96 \cdot 2) = 0,95.$$

$$P(\bar{X} + 1,96 \cdot 2 \geq \mu \geq \bar{X} - 1,96 \cdot 2) = 0,95.$$

$$P(\bar{X} - 1,96 \cdot 2 \leq \mu \leq \bar{X} + 1,96 \cdot 2) = 0,95.$$

Dieses Intervall wird Zufallsintervall genannt, weil es von einer Zufallsvariablen (\bar{X}) gebildet wird. Wir schreiben jetzt dieses Intervall mit Symbolen:

$$P\left(\bar{X} - z\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Wir bezeichnen $1 - \alpha = 0,95$ als Konfidenzkoeffizient. $\alpha = 0,05$ dagegen ist die Irrtumswahrscheinlichkeit oder das Signifikanzniveau.

Die Breite des Intervalls ist hier

$$2 \cdot (2 \cdot 1,96) = 2 \cdot \frac{\sigma}{\sqrt{n}} \cdot z(0,975) = 7,84.$$

Also schwankt eine \bar{X} -Schätzung für μ mit einer 95%igen Wahrscheinlichkeit in einem Intervall der Breite von 7840 Kaffeepaketen, d.h. μ befindet sich mit einer 95%igen Wahrscheinlichkeit in diesem Intervall.

Es kann aber passieren, dass die Schätzung \bar{x} extrem daneben liegt. In der Grafik wurde mit \bar{x} daneben gegriffen; dieser Fall durch diese restlichen 5% abgedeckt.

Konkretes 95%-Konfidenzintervall

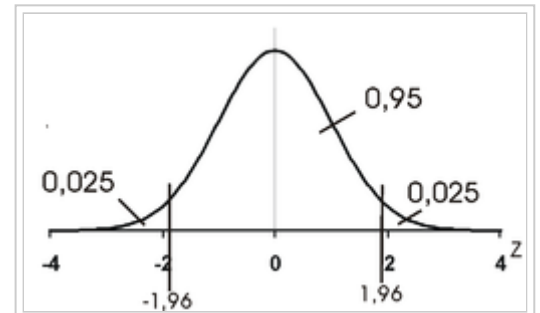
Es liegt nun ein konkreter Schätzwert von $\bar{x} = 98$ vor. Wir erhalten das Konfidenzintervall

$$\begin{aligned} & [\bar{x} - 2 \cdot 1,96; \bar{x} + 2 \cdot 1,96] \\ &= [98 - 2 \cdot 1,96; 98 + 2 \cdot 1,96] \\ &= [98 - 3,92; 98 + 3,92] \\ &= [94,08; 101,92]. \end{aligned}$$

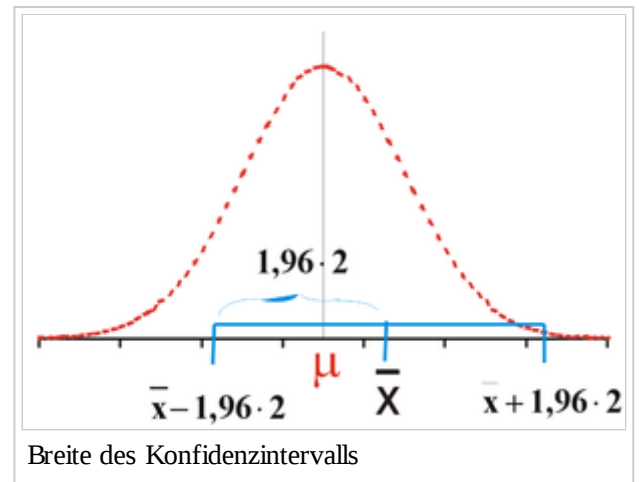
Entscheidung: μ kann bei einer Wahrscheinlichkeit von 95% unter 100 liegen, also kann der mittlere Umsatz unter 100.000 liegen. Deshalb sollte die Firma von dieser Investition absehen.

Was wäre, wenn man $[101; 108,84]$ erhalten hätte? Dann wäre eine dauerhafte Liquidität zu vermuten.

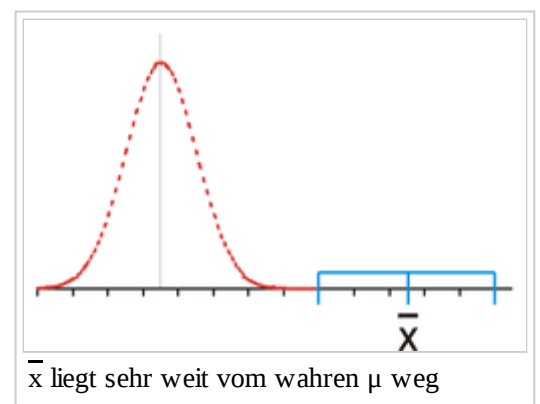
Einfluss der Varianz auf das Konfidenzintervall



(1- α /2)-Quantil der Standardnormalverteilung



Breite des Konfidenzintervalls



\bar{x} liegt sehr weit vom wahren μ weg

Was wäre, wenn σ^2 statt 200 den Wert 5000 hätte? Dann wäre

$$\bar{X} \rightarrow N\left(\mu; \frac{5000}{50} = 100\right).$$

Wir erhielten das Konfidenzintervall

$$\begin{aligned} & [\bar{x} - 1,96 \cdot \sqrt{100}; \bar{x} + 1,96 \cdot \sqrt{100}] \\ &= [98 - 19,6; 98 + 19,6] \\ &= [78,4; 117,6]. \end{aligned}$$

Das hieße, der wahre durchschnittliche Absatz läge mit einer Wahrscheinlichkeit von 95% zwischen 78 400 und 117 600 Päckchen. Dieses Intervall wäre eine sehr grobe Abschätzung. Mit so etwas kann man nicht mehr vernünftig planen.

Also wird das Konfidenzintervall mit steigender Varianz breiter, die Schätzungen werden schlechter, ungenauer. Hier könnte man als Abhilfe den Stichprobenumfang erhöhen.

Mindest erforderlicher Stichprobenumfang

Wie groß muss die Stichprobe mindestens sein, damit die Breite des Konfidenzintervalls höchstens 10 ist?

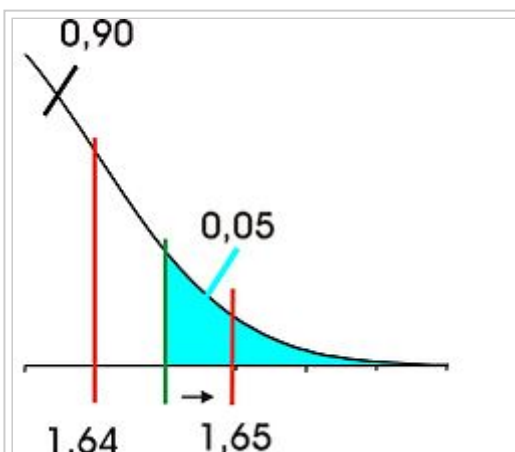
Die Breite des Konfidenzintervalls ist

$$\begin{aligned} 2 \cdot 1,96 \cdot \sqrt{\frac{5000}{n}} &\leq 10 \rightarrow \\ \sqrt{n} &\geq \frac{2 \cdot 1,96 \cdot \sqrt{5000}}{10} = 27,71. \end{aligned}$$

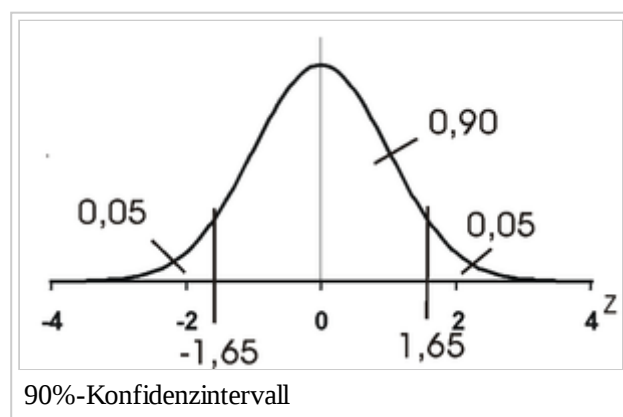
Man müsste also mindestens $n = 769$ Monate erheben, über 64 Jahre!

90%-Konfidenzintervall

Es soll nun ein 90%-Konfidenzintervall für μ bestimmt werden.



Wenn die vorgegebene Wahrscheinlichkeit zwischen zwei Quantile fällt, rückt man auf das äußere Quantil



90%-Konfidenzintervall

$$\begin{aligned} & \left[\bar{x} - z(0,95) \cdot \frac{\sigma}{\sqrt{n}}; \bar{x} + z(0,95) \cdot \frac{\sigma}{\sqrt{n}} \right] \\ = & [98 - 2 \cdot 1,65; 98 + 2 \cdot 1,65] \\ = & [98 - 3,3; 98 + 3,3] \\ = & [94,7; 101,3] . \end{aligned}$$

Dieses Intervall ist schmaler als das 95%-Intervall.

↓ Ausgewählte Konfidenzintervalle

↑ Konfidenzintervalle

↑↑ Inhaltsverzeichnis Statistik

Abgerufen von „https://de.wikibooks.org/w/index.php?title=Statistik:_Prinzip_des_Konfidenzintervalls&oldid=743472“

-
- Diese Seite wurde zuletzt am 31. März 2015 um 16:53 Uhr geändert.
 - Der Text ist unter der Lizenz "Creative Commons" „Namensnennung – Weitergabe unter gleichen Bedingungen“ verfügbar. Zusätzliche Bedingungen können gelten. Einzelheiten sind in den Nutzungsbedingungen beschrieben.