Konfidenzintervall für μ_X einer NV bei unbekannter Varianz

Zur Erstellung eines Konfidenzintervalls für den Erwartungswert einer Normalverteilung benötigt man die Populationsvarianz. Im besten Fall kennt man diese bereits, das KI lässt sich dann einfach berechnen (siehe Abschnitt *2.4. Konfidenzintervall für μ_x einer Normalverteilung bei bekannter Varianz*). Ist die Varianz jedoch nicht bekannt, so muss diese ebenfalls anhand der vorliegenden Stichprobendaten geschätzt werden:

$$\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$$

$$\hat{\sigma}_X = \sqrt{\frac{1}{n-1}\sum_i (x_i - \bar{x})^2}$$

Statt

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}}$$

verwendet man nun also:

$$T = \frac{\bar{X} - \mu_{\bar{X}}}{\hat{\sigma}_{\bar{X}}} = \frac{\bar{X} - \mu_X}{\frac{\hat{\sigma}_X}{\sqrt{n}}}$$

Z war standardnormalverteilt, für die Varianz war die Zufallsschwankung von $\bar{X}$ verantwortlich. Da aber die Varianzschätzung ebenfalls einem Stichprobenfehler unterworfen ist, muss die Varianz von T größer als jene von Z sein. Folglich kann T nicht standardnormalverteilt sein. Allerdings wird der Unterschied für sehr große Stichproben ($n \to \infty$) immer kleiner, da die Schätzung immer genauer wird.

WELCHE VERTEILUNG HAT T?

Aus Abschnitt *2.5. Die Verteilung von Stichprobenvarianzen einer normalverteilten Variablen* wissen wir:

$$S_X^2 = \frac{\sigma_X^2}{n-1} \cdot \chi^2[n-1]$$

Verwendet man S_X^2 als Schätzung für σ_X^2, also

$$S_X^2 = \hat{\sigma}_X^2$$
$$S_X = \hat{\sigma}_X$$

so ergibt sich für T:

$$T = \frac{\bar{X} - \mu_X}{\frac{\sigma_X \sqrt{\frac{\chi^2[n-1]}{n-1}}}{\sqrt{n}}}$$

Nach Umformung zu

$$T = \frac{\bar{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \cdot \frac{1}{\sqrt{\frac{\chi^2[n-1]}{n-1}}}$$

erkennt man:

$$T = \frac{Z}{\sqrt{\frac{\chi^2[n-1]}{n-1}}}$$

Da $\chi^2[n-1]$ den Erwartungswert $n-1$ hat, gilt:

$$\frac{\chi^2[n-1]}{n-1} \approx 1$$

$$\Rightarrow T \approx Z$$

Dennoch weist T eine andere Verteilung als Z auf. Die Variable

$$t = \frac{\bar{x} - \mu_X}{\frac{s_X}{\sqrt{n}}} \quad \left(mit \ \ s_X = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} \right)$$

hat „t-Verteilung“ (oft auch als „Student-Verteilung“ bezeichnet) mit $df = n - 1$ Freiheitsgraden. Kritische Werte dieser Verteilung sind für verschiedene df und für verschiedene Irrtumswahrscheinlichkeiten α tabelliert.

Anmerkungen

t ist symmetrisch.

$E(t) = 0$

Für $df \to \infty$ strebt die t-Verteilung gegen $N(0,1)$.

BERECHNEN DES KI

Mit Wahrscheinlichkeit $1 - \alpha$ gilt:

$$|t| \leq t_{1-\frac{\alpha}{2}}$$

Dies gilt auch für unser T:

$$|T| = \left| \frac{\bar{X} - \mu_X}{\frac{\hat{\sigma}_X}{\sqrt{n}}} \right| = \frac{|\bar{X} - \mu_X|}{\frac{\hat{\sigma}_X}{\sqrt{n}}} \leq t_{1-\frac{\alpha}{2}}, \quad df = n - 1$$

$$|\bar{X} - \mu_X| \leq t_{1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}_X}{\sqrt{n}}, \quad df = n - 1$$

Die Grenzen des KI

basierend auf einem Stichprobenmittelwert $\bar{x}$ und der Varianzschätzung $\hat{\sigma}_X$ sind:

$$\mu_{1,2} = \bar{x} \pm t_{1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}_X}{\sqrt{n}}, \quad df = n - 1$$