

# SPSS UE07

## Konfidenzintervalle

Konfidenzintervalle (Vertrauensintervalle) bieten die Möglichkeit unbekannte Parameter der Population (z.B. Mittelwert, Standardabweichung, Anteil) auf der Basis von bekannten Daten einer gegebenen Stichprobe vorherzusagen. Das Vertrauensintervall schließt einen Bereich um den geschätzten Wert des Parameters ein, der – vereinfacht gesprochen – mit einer zuvor festgelegten Wahrscheinlichkeit/Sicherheit (dem Konfidenzniveau) die wahre Lage des Parameters trifft. Übliche Werte für das Konfidenzniveau sind 0.99 bzw. 0.95, d.h., der Bereich enthält in 99% bzw. 95% der Fälle die wahre Lage des Parameters.

### Mittelwerte

Als erstes Beispiele werden die Konfidenzintervalle (untere Grenze UG, obere Grenze OG) für den Mittelwert  $\mu$  der Population für das Konfidenzniveau 0.99 bzw. 0.95 angeführt. Gegeben sind der Mittelwert  $\bar{x}$  und Standardabweichung  $s$  einer Stichprobe mindestens vom Umfang 30. Die  $t$ -Verteilung wird hier durch die Standard-Normalverteilung approximiert.

$$95\% - KI: \quad UG = \bar{x} - 1,96 \frac{\hat{\sigma}}{\sqrt{n}}, \quad OG = \bar{x} + 1,96 \frac{\hat{\sigma}}{\sqrt{n}} \quad (1)$$

$$99\% - KI: \quad UG = \bar{x} - 2,58 \frac{\hat{\sigma}}{\sqrt{n}}, \quad OG = \bar{x} + 2,58 \frac{\hat{\sigma}}{\sqrt{n}} \quad (2)$$

mit

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{n}{n-1} s^2. \quad (3)$$

### Anteile

Als zweites Beispiele werden die Konfidenzintervalle (untere Grenze UG, obere Grenze OG) für den Anteil  $\pi$  in einer Population für das Konfidenzniveau 0.99 bzw. 0.95 angeführt. Gegeben ist der Anteil  $p$  in einer Stichprobe.

$$95\% - KI: \quad UG = p - 1,96 \sqrt{\frac{p(1-p)}{n}}, \quad OG = p + 1,96 \sqrt{\frac{p(1-p)}{n}} \quad (4)$$

$$99\% - KI: \quad UG = p - 2,58 \sqrt{\frac{p(1-p)}{n}}, \quad OG = p + 2,58 \sqrt{\frac{p(1-p)}{n}} \quad (5)$$

## Konfidenzintervalle in SPSS

- Konfidenzintervalle für den Mittelwert werden automatisch über das Menü: **Analysieren - Explorative Datenanalyse** berechnet; die Wahl des Konfidenzniveaus erfolgt über den Auswahlpunkt: **Statistik**
- Konfidenzintervalle für Anteile: Daten werden geeignet kodiert (1 interessierende Eigenschaft, 0 Rest), der Anteil wird dann als Mittelwert behandelt.

## Übungen

1. Ein guter Weitspringer und ein guter Sprinter streiten, wer der bessere Sportler ist. Für beide kennt man die Leistungen ihrer Altersgenossen. Welche Leistung ist auf Basis nachfolgender Daten als besser einzustufen?
  - (a)
    - Weitsprung: Mittelwert der Altersgruppe  $\mu_1 = 5,37$  m, Standardabweichung  $\sigma_1 = 26$  cm
    - Bestleistung des guten Weitspringers:  $x_1 = 5,63$  m
  - (b)
    - Sprint: Mittelwert der Altersgruppe  $\mu_2 = 11,50$ s, Standardabweichung  $\sigma_2 = 0,25$  s
    - Bestleistung des guten Sprinters:  $x_2 = 11,125$  s.

Unter welchen Voraussetzungen ist ihre Einstufung gültig?

2. **A:** Beim Ausschank von Bier ist die Füllmenge je Glas normalverteilt. 33 Kontrollmessungen ergaben folgende Werte für die Füllmenge (in ml):  
500 503 498 497 502 505 495 499 505 495 499  
510 509 499 506 495 490 505 507 506 495 490  
495 499 505 507 495 499 505 495 499 503 496  
Berechnen sie ein 95%-KI und ein 99%-KI für die mittlere Füllmenge pro Glas
  - mittels SPSS und Excel
3. **A:** Die Umfrageergebnisse einer Zufallsstichprobe on  $n = 2000$  Personen aus der Grundgesamtheit der Wahlberechtigten liegt vor. Es gaben 910 Befragte an, die Partei X wählen zu wollen. Berechnen Sie das
  - 99% Konfidenzintervall
  - 95% Konfidenzintervall
  - 90% Konfidenzintervall

für den Anteil der Wähler der Partei X in der Gesamtbevölkerung

- händisch
- mittels Excel. Variieren sie in Excel die Größe der Stichprobe ( $n=4000, 8000, 16000$ ) und analysieren sie die Intervallgrenzen der Konfidenzintervalle.
- (+) mittels SPSS.