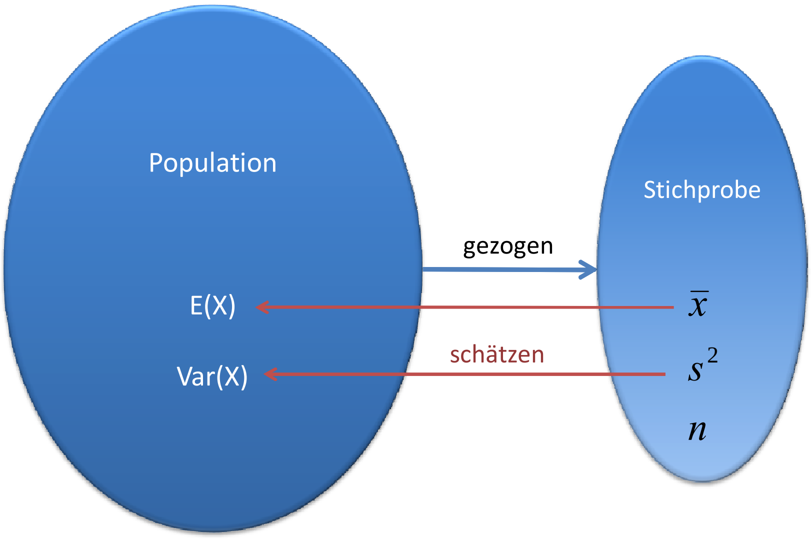
Statistik
Teil 5 - Statistisches Schließen

Ulrich Schrader
http://info.ulrich-schrader.de

Statistisches Schätzen

- Situation
  - Aus einer Population wird eine Stichprobe gezogen. Ein Merkmal X wird beobachtet.
- Frage
  - Wie kann der Erwartungswert E(X) des Merkmals aus der Stichprobe geschätzt werden?

Aufgabenstellung



Konfidenzintervall

- Beim statistischen Schätzen wird ein (zufälliges) Intervall bestimmt, in dem der Erwartungswert E(X) mit einer vorgegebenen Sicherheit 1-α enthalten ist.
  - α : Irrtumswahrscheinlichkeit
    Wahrscheinlichkeit, dass der Erwartungswert E(X) nicht in dem Intervall ist. Wird festgelegt.
- Synonyme
  - Vertrauensbereich
  - Vertrauensintervall

Konfidenzintervalle im täglichen Leben

- *"Im Mittel brauche ich für den Weg von der Wohnung zur Fachhochschule zwischen 75 und 105 Minuten. Ich bin mir dabei 95% sicher, dass das stimmt."*
- *"Der Kuchen braucht immer zwischen 60 und 75 Minuten bis er fertig ist."*
- Umgangssprachlich nennt man meist nicht die Sicherheit der Aussage.

Konfidenzintervalle

- Für die Konfidenzintervalle des Erwartungswertes (es gibt auch andere!) gilt ...
  - Schätzer ist das arithmetische Mittel der Stichprobe
- Allgemein gilt:
  - Konfidenzintervalle werden kleiner bei wachsendem Stichprobenumfang (n größer).
  - Konfidenzintervalle werden kleiner bei geringerer Streuung der beobachteten Werte.
  - Konfidenzintervalle werden um so größer je mehr Sicherheit man haben möchte.

Berechnung des Konfidenzintervalls

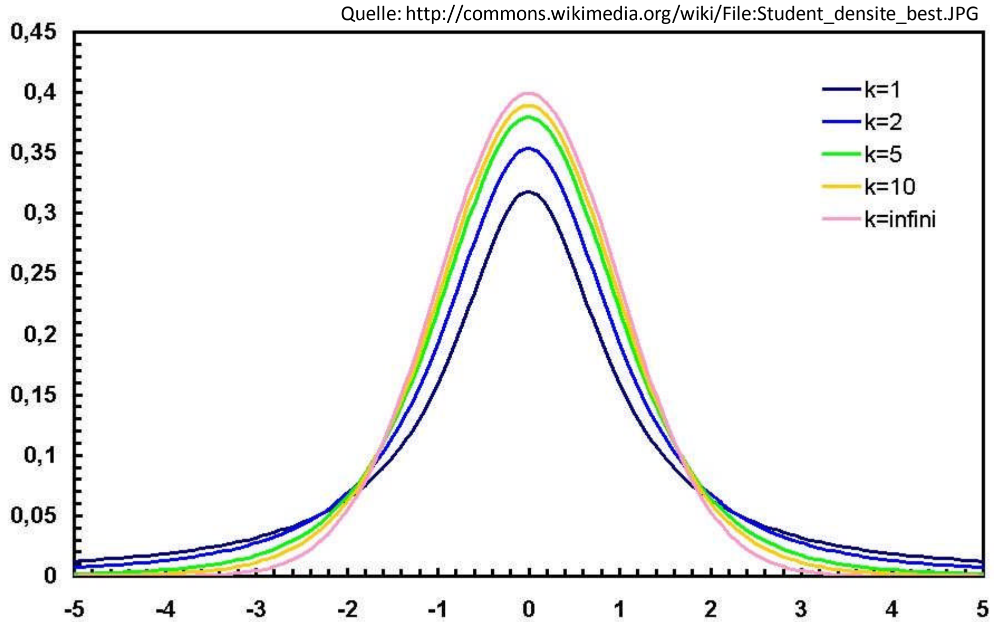
$$(1\text{-}\alpha) - \mathrm{KI} = \left[\bar{x} - t_{n-1,1-\alpha/2} \cdot \frac{s}{\sqrt{n}},\quad \bar{x} + t_{n-1,1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \right]$$

linke Grenze — rechte Grenze

- α: Irrtumswahrscheinlichkeit. Festgelegt aus dem Kontext (Was sind die Folgen eines Fehlers?)
- $\bar{x}, s, n$: Kennzahlen der Stichprobe
- $t_{n-1,1-\alpha/2}$: berücksichtigt die Stichprobengröße, die Irrtumswahrscheinlichkeit α und die Schätzung der unbekannten Streuung in der Population durch die Standardabweichung in der Stichprobe.

t-Verteilung mit k Freiheitsgraden

Quelle: http://commons.wikimedia.org/wiki/File:Student_densite_best.JPG



Konfidenzintervall (Beispiel)

Eine Person benötigt zur Durchführung einer bestimmten Aufgabe (Knöpfen) an verschiedenen Tagen jeweils:
92, 90, 89, 85, 51, 74, 45 Sekunden

$$\bar{x} = 75.14; \quad s = 19.52; \quad n = 7; \quad \alpha = 0.05$$

$$t_{7-1,1-0,05/2} = t_{6,\,0.975} = 2.45 \quad \text{(aus Tabelle)}$$

$$\begin{aligned}(1\text{-}\alpha) - \mathrm{KI} &= \left[\bar{x} - t_{n-1,1-\alpha/2} \cdot \frac{s}{\sqrt{n}}, \quad \bar{x} + t_{n-1,1-\alpha/2} \cdot \frac{s}{\sqrt{n}}\right] \\ &= \left[75.14 - 2.45 \cdot \frac{19.52}{\sqrt{7}}, \quad 75.14 + 2.45 \cdot \frac{19.52}{\sqrt{7}}\right] \\ &= \left[75.14 - \frac{47.82}{2.65}, \quad 75.14 + \frac{47.82}{2.65}\right] \\ &= [75.14 - 18.05, \quad 75.14 + 18.05] \\ &= [57.09, \quad 93.19]\end{aligned}$$

Der Erwartungswert für die Dauer zur Durchführung der Aufgabe (Knöpfen) bewegt sich mit 95% Sicherheit zwischen 57,86 und 90,42 Sekunden.

Anwendungsbeispiel

- Ein neues Verfahren zur Senkung der Körpertemperatur ist nur klinisch relevant, wenn die Senkung mindestens 1.5°C beträgt.
  - **1. Fall: 95%-KI = [1.7°C; 2.4°C]**
    Mit 95% Sicherheit kann davon ausgegangen werden, dass der Erwartungswert für die Temperatursenkung klinisch relevant ist. Im schlimmsten Fall beträgt sie nur 1.7°C, im besten Fall 2.4°C.
  - **2. Fall: 95%-KI = [1.2°C, 2.9°C]**
    Es kann nicht mit 95% Sicherheit davon ausgegangen werden, dass der Erwartungswert für die Temperatursenkung klinisch relevant ist. Im schlimmsten Fall beträgt sie nur 1.2°C.