Vorlesung 8b

Mittelwerte

Eine große Population von Werten

$$w_1, \ldots, w_g$$

ist auf der Zahlengeraden verteilt.

Man interessiert sich für den Populationsmittelwert:

$$\mu := \frac{1}{g} \sum_{j=1}^{g} w_j.$$

Zur Verfügung stehen die Werte einer aus der Population gezogenen Stichprobe

$$x_1, \ldots, x_n.$$

Als *Schätzwert* für μ bietet sich an:

$$m := \frac{1}{n}(x_1 + \ldots + x_n)$$

Wie zuverlässig ist diese Schätzung?

Goldene Idee der Statistik:

Man fasst $x_1, \dots, x_n$ auf als Ergebnis eines rein zufälligen Ziehens aus der Population:

$$X_1 := w_{J_1},\ X_2 := w_{J_2}, \dots$$

mit $J_1, J_2, \dots$ rein zufällige Wahl aus $\{1, \dots, g\}$ ("Ziehen mit Zurücklegen").

Wir setzen hier g als sehr groß gegenüber n voraus,

damit entstehen *auch*

beim n-maligen Ziehen *ohne* Zurücklegen

Kollisionen nur mit verschwindend kleiner Wahrscheinlichkeit.

Das führt auf die Vorstellung:

$x_1, \ldots, x_n$ sind entstanden

durch n-maliges unabhängiges Ziehen $X_1, \ldots, X_n$

aus einer Verteilung ρ auf $\mathbb{R}$

$$m = \frac{1}{n}(x_1 + \ldots + x_n)$$

fasst man also auf

als *eine* Realisierung (*einen* Ausgang)

der **Zufallsvariable**

$$M := \bar{X} := \frac{1}{n}(X_1 + \ldots + X_n)$$

(Mittelwert der zufälligen Stichprobe $(X_1, \ldots, X_n)$)

Wie ist $\bar{X}$ verteilt?

$$\mathbf{E}[X_1] = \mu$$

$$\mathbf{Var}[X_1] = \frac{1}{g}\sum_{j=1}^{g}(w_j - \mu)^2 =: \sigma^2$$

Dieses σ^2 heißt auch die **Populationsvarianz**.

$$\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$$

$$\mathbf{E}[\bar{X}] = \mu$$

Ist die Populationsgröße g nicht “sehr groß” gegenüber der Stichprobengröße n, dann hat es Sinn, die *Korrektur für endliche Populationen* zu berücksichtigen (vgl. Aufgaben 17 und 32):

$$\mathbf{Var}[\bar{X}] = \frac{\sigma^2}{n} \cdot \frac{g-n}{g-1}$$

Diese Korrektur werden wir für den Rest dieser Vorlesung vernachlässigen (wir denken an $g = \infty$, bzw. – wie schon gesagt – an ein wiederholtes unabhängiges Ziehen aus einer Verteilung.

Dann gilt:

$$\mathbf{Var}[\bar{X}] = \frac{\sigma^2}{n},$$

die Standardabweichung des Stichprobenmittels ist $\frac{\sigma}{\sqrt{n}}$.

Wie ist $\bar{X}$ verteilt?

Der Zentrale Grenzwertsatz gibt eine Antwort:

$\bar{X}$ ist approximativ $N(\mu, \frac{\sigma^2}{n})$-verteilt.

Ein Problem in der Praxis: I. A. kenn man σ^2 nicht.

Auch σ^2 muss man dann schätzen.

Zwei Vorschläge für die

(aus der Stichprobe) **geschätzte** (Populations-)**Varianz**:

(i) die *Stichprobenvarianz*

$$\hat{\sigma}^2 := \frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2.$$

(ii) die *modifizierte Stichprobenvarianz*

$$s^2 := \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)^2$$

Es gibt theoretische Begründung für beide Vorschläge

(vgl. Buch S. 124, S. 138).

Wir kommen darauf später zurück

und halten uns erst einmal an den Vorschlag (ii):

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)^2$$

Die Standardabweichung des Stichprobenmittels $\bar{X}$ ist $\frac{\sigma}{\sqrt{n}}$.

Die geschätzte Standardabweichung

des Stichhprobenmittels $\bar{X}$ ist

$$\boxed{s/\sqrt{n} =: f}$$

Diese Größe nennen wir auch den *Standardfehler*.

$\bar{X}$ ist approximativ $N(\mu, \frac{\sigma^2}{n})$-verteilt.

Und:

$\bar{X}$ ist approximativ $N(\mu, f^2)$-verteilt.