

Vorlesung 8a

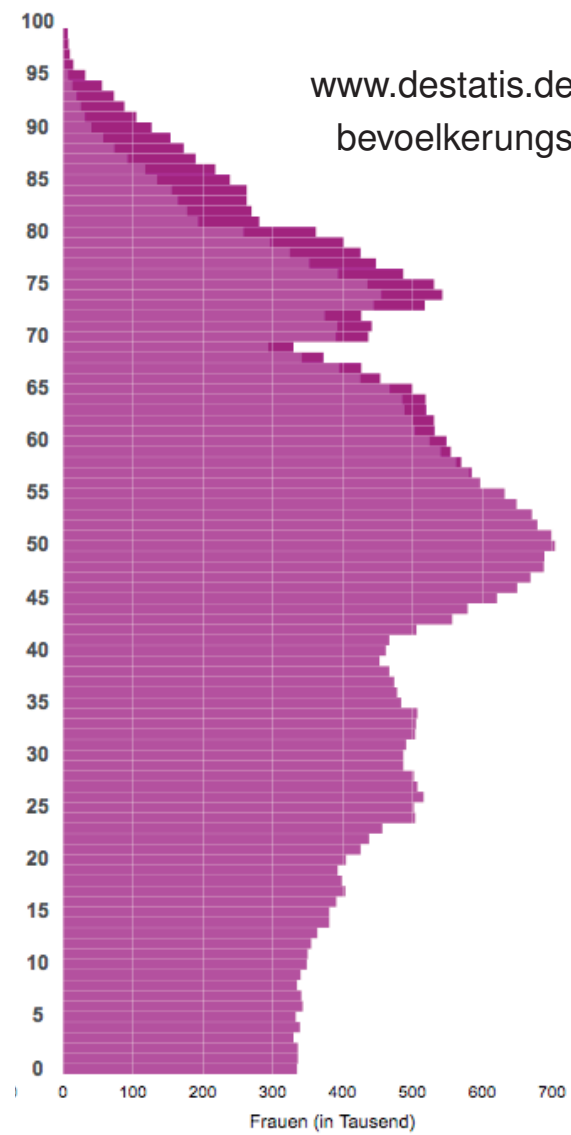
Mittelwerte

1. Populationsmittelwert und Stichprobenmittelwert

Denken wir an eine Liste (eine “Population”)
von reellen Daten

$$w_1, \dots, w_g$$

z. B. die Lebensalter aller Frauen
in der deutschen Bevölkerung 2014



Angenommen man möchte den **Populationsmittelwert**

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

schätzen,

und zwar aus den Werten einer
aus der Population gezogenen **Stichprobe**

$x_1, \dots, x_n$.

Als *Schätzwert* für μ bietet sich an:

$$\bar{x} := \frac{1}{n}(x_1 + \dots + x_n)$$

Wie zuverlässig ist diese Schätzung?

2. Die goldene Idee der Statistik

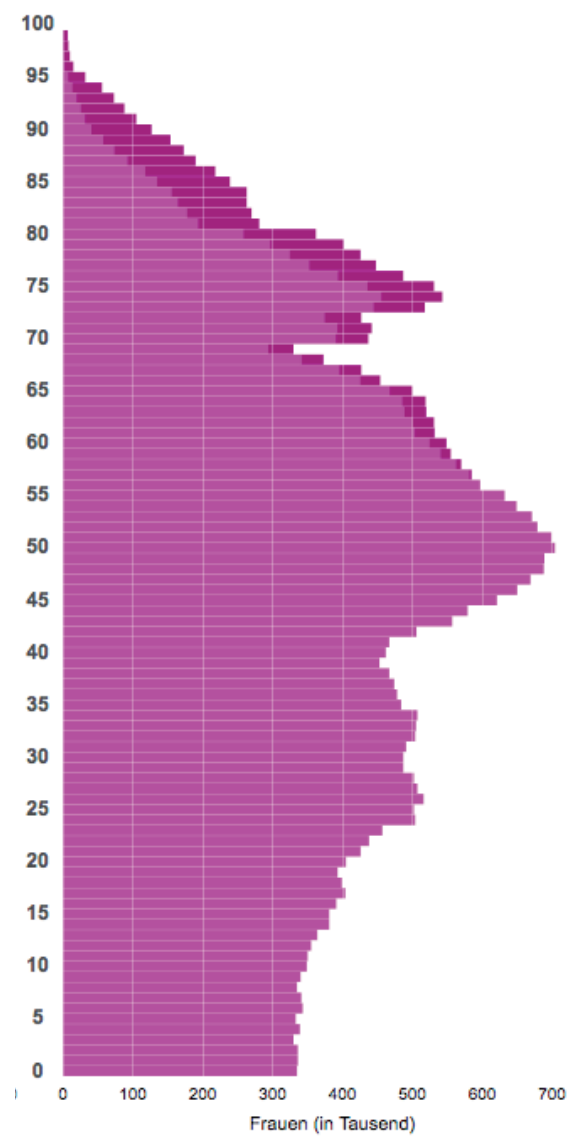
Man fasst $x_1, \dots, x_n$ auf als Ergebnis eines rein zufälligen Ziehens aus der Population:

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

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:
 $x_1, \dots, x_n$ sind entstanden
durch n -maliges unabhängiges Ziehen $X_1, \dots, X_n$
aus der Verteilung ρ auf $\mathbb{R}$
mit $\rho([c, d]) := \frac{1}{g} \#\{i : w_i \in [c, d]\}$



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

fasst man also auf
als *eine* Realisierung (*einen* Ausgang)
der **Zufallsvariable**

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

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

Wie ist $\bar{X}$ verteilt?

3. Erwartungswert, Varianz und approximative Verteilung des Stichprobenmittelwerts

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

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

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

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

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

Ist die Populationsgröße g nicht sehr groß gegenüber der Stichprobengröße n , dann hat es Sinn, die *Korrektur für endliche Populationen* zu berücksichtigen (vgl. Aufgabe 19):

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

Diese Korrektur werden wir für den Rest dieser Vorlesung vernachlässigen (wir denken an großes g , 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 Stichprobenmittelwertes $\bar{X}$
ist also $\frac{\sigma}{\sqrt{n}}$.

Wie ist $\bar{X}$ verteilt?

Der Zentrale Grenzwertsatz gibt eine Antwort:

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

4. Geschätzte Populationsvarianz und Standardfehler

Ein Problem in der Praxis: I. A. kann 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 Stichprobenmittels $\bar{X}$ ist

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

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.

Eine anschauliche Präsentation der Thematik
“Wie genau ist eine Schätzung des Mittelwertes?”
von einem elementaren
und anwendungsorientierten Standpunkt aus
finden Sie auf den klassischen Ferebee’schen Folien

[http://www.math.uni-frankfurt.de/~ismi/
wakolbinger/teaching/Stofl1617/
statbioStandardfehler.pdf](http://www.math.uni-frankfurt.de/~ismi/wakolbinger/teaching/Stofl1617/statbioStandardfehler.pdf)