Formelsammlung Statistik

Lukas Warode

Maße der zentralen Tendenz

Modus

- Nominales Skalenniveau
- Häufigster Wert

$$x_{mod}$$

Median

- Ordinales Skalenniveau
- Mittlere Ausprägung bei Anordnung der Variable

Ungerade Anzahl an Fällen (n):

$$\tilde{x} = x_{(\frac{n+1}{2})}$$

```r
set.seed(42)

random_sample <- sample(1:42, 11)

sort(random_sample)
```

```
##  [1]  1  7 10 18 20 24 25 26 36 37 40
```

```r
median(random_sample)
```

```
## [1] 24
```

Gerade Anzahl an Fällen (n):

$$\tilde{x} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$$

```r
random_sample <- random_sample[random_sample != max(random_sample)]

sort(random_sample)
```

```
##  [1]  1  7 10 18 20 24 25 26 36 37
```

```r
median(random_sample)
```

```
## [1] 22
```

Arithmetisches Mittel

- Metrisches Skalenniveau
- Summe aller Fälle durch Anzahl der Fälle teilen

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n} \sum_{i=i}^{n} x_i$$

```r
random_sample
```

```
##  [1] 37  1 25 10 36 18 24  7 20 26
```

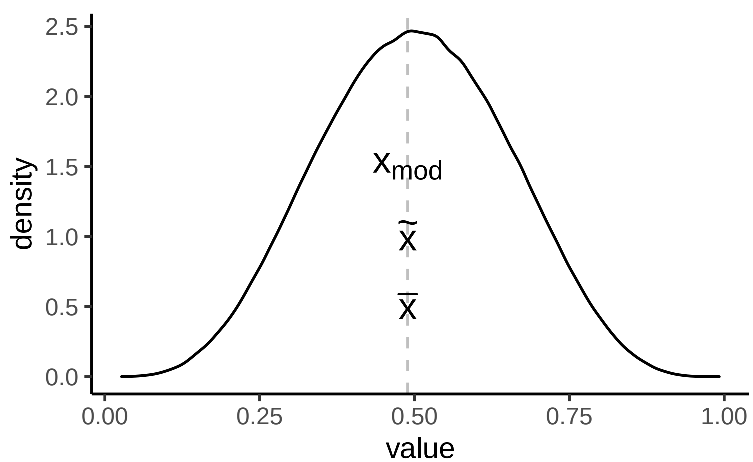
```r
mean(random_sample)
```

```
## [1] 20.4
```

Verteilungsformen

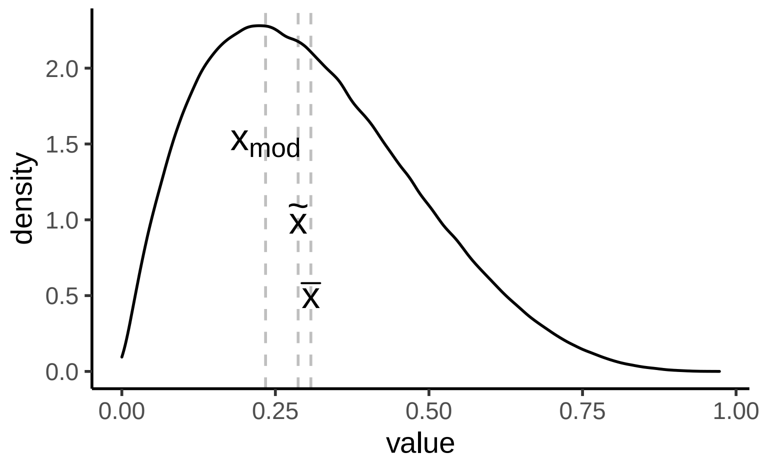
Symmetrisch (Normalverteilung)

$$x_{mod} = \tilde{x} = \bar{x}$$



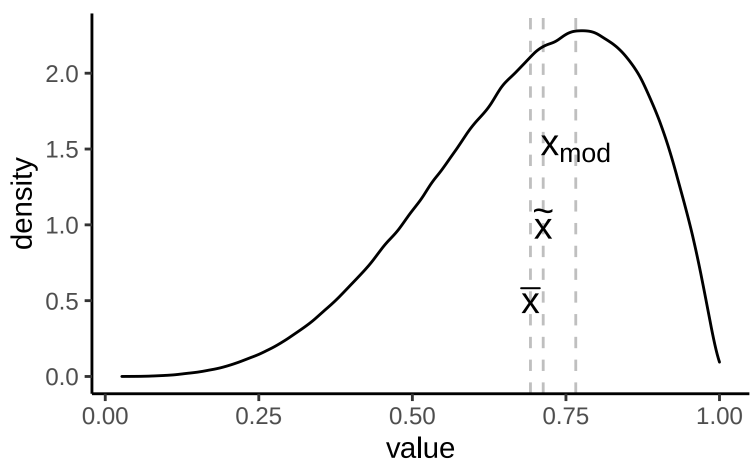
Linkssteil / Rechtsschief

$$x_{mod} < \tilde{x} < \bar{x}$$



Rechtssteil / Linksschief

$$x_{mod} > \tilde{x} > \bar{x}$$



Streuungsmaße

Spannweite

- Ordinales Skalenniveau
- Differenz zwischen größter und kleinster Ausprägung

$$R = x_{max} - x_{min}$$

Interquartilsabstand (IQR)

- Ordinales Skalenniveau
- Intervall der mittleren 50% der Stichprobe

$$IQR = Q_{0.75} - Q_{0.25}$$

Variation (Summe der Abweichungsquadrate)

- Metrisches Skalenniveau
- Englisch: *Sum of squares / sum of squared deviations*

$$SS_x = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Varianz

- Metrisches Skalenniveau
- Standardisierte Variation

$$s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Standardabweichung

- Metrisches Skalenniveau
- Quadratwurzel der Varianz
- Durchschnittliche Abweichung von Werten zum Arithmetischen Mittel

$$\sigma_{x_{Population}} = \sqrt{\sigma_x^2} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$
$$s_{x_{Stichprobe}} = \sqrt{s_x^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Variationskoeffizient (Abweichungskoeffizient)

- Metrisches Skalenniveau
- Relatives Streuungsmaß, d.h. nicht abhängig von der Maßeinheit der Variable
- Standardbweichung in Relation zum Arithmetischen Mittel

$$V_x = \frac{s_x}{\bar{x}}$$

Standardfehler

- (Durchschnittliche) Abweichung von Stichprobenkennwerten zu Populationskennwerten, z.B. vom Mittelwert

Standardfehler des Arithmetischen Mittelwertes (***Standard error of the mean***)

- Symbol des Populationsmittelwertes: μ
- Symbol des Stichprobenmittelwertes $\bar{x}$

$$\sigma_{\bar{x}_{Population}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\sigma^2}{n}}$$

$$s_{\bar{x}_{Stichprobe}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$$

$$\hat{\sigma}_{\bar{x}_{Population,\ geschätzt}} = \frac{s}{\sqrt{n-1}} = \sqrt{\frac{s^2}{n-1}}$$

Standardfehler des Anteilswertes

- Symbol des Populationsanteilswertes: π
- Symbol des Stichprobenanteilswertes: p_x

$$\sigma(p_x)_{Population} = \sqrt{\frac{\pi_x \cdot (1-\pi_x)}{n}}$$

Schätzung der Populationsvarianz: $\pi_x \cdot (1-\pi_x)$ aus der Stichprobenvarianz: $p_x \cdot (1-p_x)$

$$\hat{\sigma}(p_x)_{Population} = \sqrt{\frac{p_x \cdot (1-p_x)}{n}}$$

Konfidenzintervall

- Generalisierbarkeit von Parametern aus der Stichprobe (auf die Population)
- Geschätzter Intervallbereich, in dem Parameter der Grundgesamtheit mit einer bestimmten Wahrscheinlichkeit liegen
- Z-Standardisierung: $z = \frac{x-\mu}{\sigma}$

Konfidenzintervall des Populationsmittelwertes (μ_x)

- Kleine Stichproben: t-Verteilung
- Große Stichproben: Standardnormalverteilung

Bestimmung der Intervallgrenzen:

$$\bar{x} - \frac{s_x}{\sqrt{n-1}} \cdot z_{(1-\frac{\alpha}{2})} < \mu_x < \bar{x} + \frac{s_x}{\sqrt{n-1}} \cdot z_{(1-\frac{\alpha}{2})}$$

Konfidenzintervall des Populationsanteilswertes (π_x)

- Standardnormalverteilung (wenn Stichprobe ausreichend groß)

Bestimmung der Intervallgrenzen:

$$p_x - \sqrt{\frac{p_x \cdot (1-p_x)}{n}} \cdot z_{(1-\frac{\alpha}{2})} < \pi_x < p_x + \sqrt{\frac{p_x \cdot (1-p_x)}{n}} \cdot z_{(1-\frac{\alpha}{2})}$$

t-Test

t-Test: Mittelwert

t-Test für einen Mittelwert

- H_0: $\mu_1 = \mu$
- H_A: $\mu_1 \neq \mu$

Teststatistik:

$$Z = \frac{\bar{x} - \mu}{\frac{s_x}{\sqrt{n-1}}}$$

- Kritische Testwerte bei z_α und $z_{(1-\alpha)}$

t-Test für 2 Mittelwerte

- H_0: $\mu_1 - \mu_2 = 0$
- H_A: $\mu_1 - \mu_2 \neq 0$

Teststatistik:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_{x_1}^2}{n_1-1} - \frac{s_{x_2}^2}{n_2-1}}}$$

- Kritische Testwerte bei $z_{\frac{\alpha}{2}}$ und $z_{(1-\frac{\alpha}{2})}$

t-Test: Populationsanteil

t-Test für einen Populationsanteil

- H_0: $\pi_1 = \pi$
- H_A: $\pi_1 \neq \pi$

Teststatistik:

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi \cdot (1-\pi)}{n}}}$$

- Kritische Testwerte bei z_α und $z_{(1-\alpha)}$

t-Test für 2 Populationsanteile

- H_0: $\pi_1 - \pi_2 = 0$
- H_A: $\pi_1 - \pi_2 \neq 0$

Teststatistik:

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 \cdot (1-p_1)}{n_1} + \frac{p_2 \cdot (1-p_2)}{n_2}}}$$

- Kritische Testwerte bei $z_{\frac{\alpha}{2}}$ und $z_{(1-\frac{\alpha}{2})}$