

Stochastik

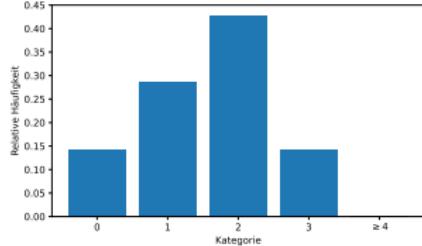
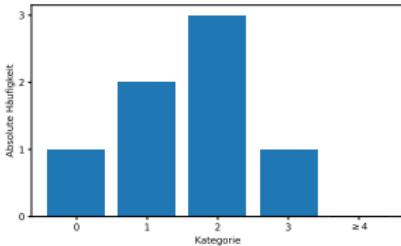
FH Aachen – Studienstandort Aachen
18. Dezember 2025

Prof. Dr. Florian Heinrichs

Wiederholung: Statistik

Deskriptive Statistik

- > Was ist die absolute Häufigkeit?
 - > Sei X ein kategoriales Merkmal mit den Ausprägungen (Kategorien) a_1, \dots, a_k
 - > Seien x_1, \dots, x_n Beobachtungen von X
 - > Die absolute Häufigkeit ist gegeben durch $h_i = \sum_{j=1}^n \mathbb{1}_{\{x_j \in a_i\}}$
 - > Entspricht "Anzahl Beobachtungen in a_i "
- > Was ist die relative Häufigkeit? $\rightarrow r_i = \frac{h_i}{n}$
- > Beispiel: Eine Person hat in den letzten Tagen 2, 3, 1, 2, 0, 1, 2 Tassen Kaffee getrunken



Lage- und Streumaße

> Sei X_1, \dots, X_n eine Stichprobe

> Was ist ...

> ... der Mittelwert? $\rightarrow \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

> ... der (empirische) Median?

$$q_{0.5} = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{falls } n \text{ ungerade} \\ \frac{1}{2}(X_{(n/2)} + X_{(n/2+1)}) & \text{falls } n \text{ gerade} \end{cases}$$

> ... das (empirische) α -Quantil?

$$q_\alpha = \begin{cases} X_{(\lfloor n \cdot \alpha + 1 \rfloor)} & \text{falls } n \cdot \alpha \notin \mathbb{N} \\ \frac{1}{2}(X_{(n \cdot \alpha)} + X_{(n \cdot \alpha + 1)}) & \text{falls } n \cdot \alpha \in \mathbb{N} \end{cases}$$

> ... die Stichprobenvarianz? $\rightarrow s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$

> ... die empirische Standardabweichung? $\rightarrow s_n$

> ... der Interquartilsabstand? $q_{0.75} - q_{0.25}$

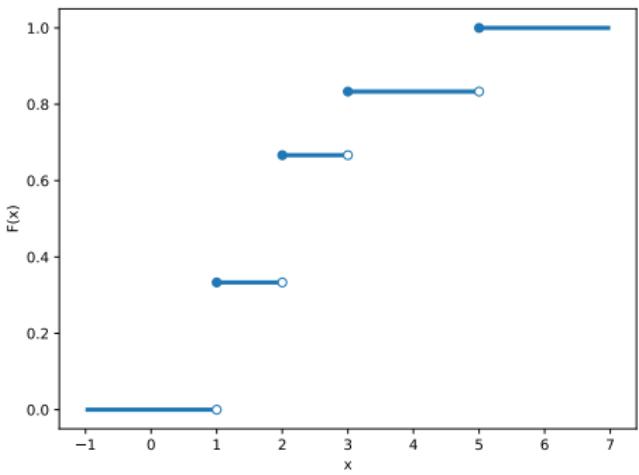
Empirische Verteilungsfunktion

Was ist die empirische Verteilungsfunktion?

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}$$

Angenommen wir kennen die Analysis I Noten von sechs Studierenden:

i	1	2	3	4	5	6
x_i	2	3	2	1	5	1



Schätzer

- > Was ist ein Schätzer?
- > **Informell:**
 - > X_1, \dots, X_n unabhängige Zufallsvariablen mit Verteilung \mathbb{P}_θ
 - > θ unbekannt, aber wir nehmen an, dass $\mathbb{P}_\theta \in \mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$
 - > Wir wollen den wahren Parameter θ basierend auf Realisierungen von X_1, \dots, X_n schätzen
- > Wann ist ein Schätzer $T(X_1, \dots, X_n)$ erwartungstreu?
 $\rightarrow \mathbb{E}_\theta[T(X_1, \dots, X_n)] = \theta$ für alle $\theta \in \Theta$
- > Was ist der Maximum-Likelihood Schätzer?
 - > Likelihood-Funktion:
$$L(\theta) = \begin{cases} \mathbb{P}_\theta(X_1 = x_1) \cdot \dots \cdot \mathbb{P}_\theta(X_n = x_n) & \text{für diskrete ZV} \\ f_\theta(x_1) \cdot \dots \cdot f_\theta(x_n) & \text{für stetige ZV} \end{cases}$$
 - > Log-Likelihood-Funktion: $\ell(\theta) = \log(L(\theta))$
 - > Das Argument $\hat{\theta}$, das $L(\theta)$ maximiert, heißt ML-Schätzer

Schätzer

- > Was ist ein $(1 - \alpha)$ -Konfidenzintervall?
 - > Eine Abbildung I , die für $\alpha \in [0, 1]$ jeder Stichprobe x_1, \dots, x_n ein Intervall zuordnet, sodass $\mathbb{P}_\theta(\theta \in I(X_1, \dots, X_n)) \geq 1 - \alpha$ für alle $\theta \in \Theta$.
- > Wichtige Konfidenzintervalle für $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ unabh.
- > Falls σ bekannt

$$\left[\bar{X}_n - \frac{q_{1-\alpha/2}\sigma}{\sqrt{n}}, \bar{X}_n + \frac{q_{1-\alpha/2}\sigma}{\sqrt{n}} \right], \quad \left(-\infty, \bar{X}_n + \frac{q_{1-\alpha}\sigma}{\sqrt{n}} \right], \quad \left[\bar{X}_n - \frac{q_{1-\alpha}\sigma}{\sqrt{n}}, \infty \right)$$

- > Falls σ unbekannt

$$\begin{aligned} &\left[\bar{X}_n - \frac{t_{n-1, 1-\alpha/2} s_n}{\sqrt{n}}, \bar{X}_n + \frac{t_{n-1, 1-\alpha/2} s_n}{\sqrt{n}} \right], \\ &\left(-\infty, \bar{X}_n + \frac{t_{n-1, 1-\alpha} s_n}{\sqrt{n}} \right], \quad \left[\bar{X}_n - \frac{t_{n-1, 1-\alpha} s_n}{\sqrt{n}}, \infty \right) \end{aligned}$$

Test

- > Was ist ein Testproblem? $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$
- > Was ist ein Fehler 1. Art? gültige Nullhypothese wird verworfen
- > Was ist ein Fehler 2. Art? ungültige Nullhypothese wird *nicht* verworfen
- > Was ist die Gütefunktion eines Tests ϕ ? $\beta_\phi(\theta) = \mathbb{P}_\theta(\phi(X) = 1)$
- > Was ist das Niveau eines Tests? $\sup_{\theta \in \Theta_0} \beta_\phi(\theta)$
 - > Anschaulich: die max. Wahrscheinlichkeit für einen Fehler 1. Art
- > Was ist die Power eines Tests? $\beta_\phi(\theta)$ für $\theta \in \Theta_1$
- > Besonders wichtig: Gauß- und t -Test
 - > Wann benutzen wir welchen?
 - > $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$
 - > $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ (bzw. \leq, \geq)
 - > σ^2 bekannt: Gauß-Test
 - > σ^2 unbekannt: t -Test

Zusammenfassung (Tests)

Vorgehen beim statistischen Testen

1. Wahl eines statistischen Modells
 - > Beispiel Münzwurf: $X \sim Ber(p)$
2. Festlegen der Nullhypothese H_0 und Alternative H_1
 - > Beispiel Münzwurf: $H_0 : p = \frac{1}{2}$ vs. $H_1 : p \neq \frac{1}{2}$
3. Wahl eines geeigneten Tests
 - > Oder Herleitung eines geeigneten Tests (vgl. Münzwurf)
4. (Oft) Berechnung einer Teststatistik T
5. (Oft) Bestimmung eines kritischen Werts c_α
6. Testentscheidung: Ablehnen (oder nicht) der Nullhypothese
 - > Oft: Falls $T \geq c_\alpha$ (oder \leq)

Gauß-Test

$$T = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma}$$

Hypothese	Verwerfe, falls	p-Wert
$H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$	$T > q_{1-\alpha}$	$1 - \Phi(t(x))$
$H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$	$T < q_\alpha$	$\Phi(t(x))$
$H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$	$ T > q_{1-\alpha/2}$	$2(1 - \Phi(t(x)))$

Gauß-Test

$$T = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma}$$

	Fehler 1. Art	Fehler 2. Art
$H_0 : \mu \leq \mu_0$	$1 - \Phi\left(q_{1-\alpha} + \sqrt{n} \frac{(\mu_0 - \mu)}{\sigma}\right)$	$\Phi\left(q_{1-\alpha} + \sqrt{n} \frac{(\mu_0 - \mu)}{\sigma}\right)$
$H_0 : \mu \geq \mu_0$	$\Phi\left(q_\alpha + \sqrt{n} \frac{(\mu_0 - \mu)}{\sigma}\right)$	$1 - \Phi\left(q_\alpha + \sqrt{n} \frac{(\mu_0 - \mu)}{\sigma}\right)$
$H_0 : \mu = \mu_0$	$2 - \Phi\left(q_{1-\frac{\alpha}{2}} + \sqrt{n} \frac{(\mu_0 - \mu)}{\sigma}\right)$ $-\Phi\left(q_{1-\frac{\alpha}{2}} - \sqrt{n} \frac{(\mu_0 - \mu)}{\sigma}\right)$	$\Phi\left(q_{1-\frac{\alpha}{2}} + \sqrt{n} \frac{(\mu_0 - \mu)}{\sigma}\right)$ $+\Phi\left(q_{1-\frac{\alpha}{2}} - \sqrt{n} \frac{(\mu_0 - \mu)}{\sigma}\right) - 1$

Einstichproben-*t*-Test

$$T = \sqrt{n} \frac{\bar{X}_n - \mu_0}{s_n}$$

Hypothese	Verwerfe, falls	p-Wert
$H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$	$T > t_{n-1,1-\alpha}$	$1 - F_{t,n-1}(t(x))$
$H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$	$T < t_{n-1,\alpha}$	$F_{t,n-1}(t(x))$
$H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$	$ T > t_{n-1,1-\alpha/2}$	$2(1 - F_{t,n-1}(t(x)))$

- > Was ist ein Zweistichproben-*t*-Test?
- > Welche Varianten des Zweistichproben-*t*-Tests gibt es?
- > Wie testen wir $H_0 : \sigma^2 = \sigma_0^2$?

Literatur I

 Bundesministerium für Gesundheit (2024).

Infektionsradar.

https:

//infektionsradar.gesund.bund.de/de/covid/inzidenz.
Abgerufen: 2024-10-15.

 Dehling, H. and Haupt, B. (2006).

Einführung in die Wahrscheinlichkeitstheorie und Statistik.

Springer-Verlag.

Literatur II

 Dinnes, J., Sharma, P., Berhane, S., van Wyk, S., Nyaaba, N., Domen, J., Taylor, M., Cunningham, J., Davenport, C., Dittrich, S., Emperador, D., Hooft, L., Leeflang, M., McInnes, M., Spijker, R., Verbakel, J., Takwoingi, Y., Taylor-Phillips, S., Van den Bruel, A., and Deeks, J. (2022).

Rapid, point-of-care antigen tests for diagnosis of sars-cov-2 infection.

Cochrane Database of Systematic Reviews, (7).

 Henze, N. et al. (1997).

Stochastik für Einsteiger, volume 4.

Springer.