§ 10 Statistische Tests

Setting: Gegeben sind Daten $x_1, \dots, x_N \in \mathbb{R}$.
Wir schreiben $x := (x_1, \dots, x_N) \in \mathbb{R}^N$.

Grundidee:

1. Wahl eines Modells: Wir annehmen an, dass $x_1 = X_1(\omega), \dots, x_N = X_N(\omega)$ Realisationen von unabhängigen Zufallsvariablen sind, so dass $X_i : \Omega \to \mathbb{R}$ und $\omega \in \Omega$.
   (Beispiel: $X_1, \dots, X_N \overset{iid}{\sim} N(\mu, \sigma^2)$ ⟵ das ist ein Modell).
2. Wahl einer Nullhypothese H_0. Die Nullhypothese stellt eine Behauptung über das Modell dar, die wir anhand der Daten überprüfen möchten.
   (Beispiel: Falls $X_1, \dots, X_N \overset{iid}{\sim} N(\mu, \sigma^2)$, dann $H_0 := \{\mu \leq \mu_0\}$ für ein festes $\mu_0 \in \mathbb{R}$).
3. Definition einer Alternativhypothese: H_1 = Komplement / Gegenteil von H_0.
   (Beispiel: Falls $X_1, \dots, X_N \overset{iid}{\sim} N(\mu, \sigma^2)$, und $H_0 = \{\mu \leq \mu_0\}$ dann $H_1 = \{\mu > \mu_0\}$.
4. Berechnung einer Teststatistik $T(x) \in \mathbb{R}$ (i.A. ist auch $T(x) \in \mathbb{R}^m$ möglich).
   ↑ abhängig von den Daten $x \in \mathbb{R}^N$.
5. Bestimmung eines Ablehnungsbereich $A \in \mathbb{R}$.
6. Falls $T(x) \in A$, dann lehnen wir die Nullhypothese H_0 ab.

Zusammengefasst: Wir haben Daten. Wir haben ein Modell. Wir wollen mit Hilfe der Daten eine Aussage über das Modell treffen.

Fehler 1. Art und Fehler 2. Art

Seien H_0 die gewählte Nullhypothese und H_1 die entsprechende Alternativhypothese.

	Entscheidung: H_0 wird nicht abgelehnt	H_0 wird abgelehnt
H_0 gilt	✓	Fehler 1. Art
H_1 gilt	Fehler 2. Art.	✓

→ Fehler 1. Art = Fehler H_0 abzulehnen, obwohl H_0 gilt.
Fehler 2. Art = Fehler H_0 nicht abzulehnen, obwohl H_1 gilt.

Im statistischen Test wird der Fehler 1. Art kontrolliert, indem wir einen entsprechenden Ablehnungsbereich A formulieren. Idee:
$P(T(X) \in A \mid H_0) = P(\text{Fehler 1. Art})$ ist klein.
X = Vektor der Zufallsvariablen $X_1, \dots, X_N$
→ $T(X)$ = Zufallsvariable.

D.h. wir wählen A, so dass $P(\text{Fehler 1. Art})$ klein ist.
Das können wir erreichen, indem wir die Verteilung der Teststatistik $T(X)$ verstehen.

<u>Definition</u> (Signifikanzniveau)
Sei $\alpha \in (0,1)$. Sei der Ablehnungsbereich A so gewählt, dass $P(\text{Fehler 1. Art}) = \alpha$. Dann nennen wir α das Signifikanzniveau.

Beispiel $X_1, \dots, X_N \sim N(\mu, \sigma^2)$ und σ^2 ist bekannt, aber μ ist unbekannt.
Das ist das Modell.

$H_0 := \{\mu \le \mu_0\}$, $\quad H_1 := \{\mu > \mu_0\}$.

Sei $\alpha = 0{,}05$ das Signifikanzniveau.

Dann ist $\quad \alpha = P(\text{Fehler 1. Art}) = P(T(X) \in A \mid H_0)$.

Wir wählen als Teststatistik $T(X)$ für μ: $\quad T(X) = \frac{1}{N}(X_1 + \dots + X_N)$

$\Rightarrow \quad \mathbb{E}(T(X)) = \mu \underset{\text{unter } H_0}{\le} \mu_0$

$\rightsquigarrow$ Wir wählen $A = \{z \in \mathbb{R} \mid z \ge c\}$ für einen kritischen Wert $c \in \mathbb{R}$,

so dass $\quad P(T(X) \ge c \mid H_0) = \alpha$.

$\rightsquigarrow$ Dann lehnen wir H_0 ab, genau dann wenn $\underbrace{T(x)}_{\text{Teststatistik für die Daten}} \ge c$

(Dieses Beispiel ist der "Gauss-Test").

Tests, wobei das Modell eine Zufallsverteilung für die $X_1, \dots, X_N$ annimmt, die durch Parameter bestimmt ist, und wobei H_0 eine Behauptung über die Parameter trifft, nennt man parametrische Tests.

Bei parametrischen Tests möchten wir etwas über die Parameter wissen.

Alle Tests in dieser Vorlesung sind parametrisch.

Definition (Quantile)

Sei X eine Zufallsvariable mit Verteilungsfunktion $F(x) = P(X \le x)$.

Sei $\alpha \in (0,1)$. Das α-Quantil von X ist der kleinste Wert $q_\alpha \in \mathbb{R}$, für den gilt: $\quad F(q_\alpha) \ge \alpha$.

Bemerkung: (1) Falls F stetig ist und streng monoton wachsend ist (z.B. $X \sim N(\mu,\sigma^2)$, $X \sim t_n$ oder $X \sim \chi_n^2$), so ist das α-Quantil definiert durch $F(q_\alpha) = \alpha$.

(2) Die α-Quantile der Standardnormalverteilung $N(0,1)$, der t-Verteilung t_n und der Chi-Quadrat-Verteilung χ_n^2 werden mit z_α, $t_{n,\alpha}$ und $\chi^2_{n,\alpha}$ bezeichnet.

(3) Es gilt: $z_\alpha = -z_{1-\alpha}$ und $t_{n,\alpha} = -t_{n,1-\alpha}$.
(das liegt daran, dass die Dichten symmetrisch um 0 sind).

<u>Interpretation von Testergebnissen:</u>

- Führt ein Test mit Signifikanzniveau α zur Ablehnung von H_0, so können mit einer Wahrscheinlichkeit von $1-\alpha$ sicher sein, dass die Entscheidung richtig war und dass H_1 gilt. Die Daten sprechen signifikant für H_1.
- Führt der Test nicht zur Ablehnung von H_0, so bedeutet dies nur, dass die Daten nicht signifikant für H_0 sprechen. Das <u>heisst nicht</u>, dass wir schliessen können, <u>dass H_0 gilt</u>, weil der Fehler 2. Art immer noch groß sein könnte.

<u>Folgerung:</u> Wenn wir nachweisen wollen, dass eine Aussage über das Modell gilt, müssen wir sie als Alternativhypothese H_1 formulieren.

Definition (Rechts-, Links- oder zweiseitiger Test).

Falls der Ablehnungsbereich A

- $A = \{T(x) < c\}$, nennen wir den Test linksseitig.
- $A = \{T(x) > c\}$, nennen wir den Test rechtsseitig.
- $A = \{T(x) < c_1$ oder $T(x) > c_2\}$, nennen wir den Test zweiseitig.

Definition (p-Wert).

Sei $T(X)$ eine Teststatistik und $\underbrace{x = (x_1, \dots, x_N)}_{= \text{Daten}}$ und $\underbrace{X = (X_1, \dots, X_N)}_{\text{Zufallsvariablen aus dem Modell.}}$

Dann ist der p-Wert p, wie folgt definiert:

- falls der Test linksseitig ist $p_L := P(T(X) \leq T(x) \mid H_0)$

  $T(X)$: Zufallsvariablen, $T(x)$: Daten.

- falls der Test rechtsseitig ist $p_R := P(T(X) \geq T(x) \mid H_0)$
- falls der Test zweiseitig ist: $p := 2 \min\{p_L, p_R\}$.

Motivation Falls α das Signifikanzniveau des Tests ist. Dann gilt:

$p < \alpha \iff T(x) \in A$, d.h. die Teststatistik ist im Ablehnungs-bereich

$\iff H_0$ wird abgelehnt.

- In R wird bei Tests meistens der p-Wert ausgegeben.
- In vielen Studien wird der p-Wert angegeben.

Beliebte Missinterpretation: $p =$ Wahrscheinlichkeit, dass H_1 nicht gilt.
← FALSCH!

Spezifische (parametrische) Tests:

1. Der Gauss-Test:

(a) Einstichproben Gauss-Test:

Modell: $X_1, \dots, X_N \sim N(\mu, \sigma^2)$ mit μ unbekannt, σ^2 bekannt.

Hypothesen: (1) linksseitig $H_0 = \{\mu \geq \mu_0\}$

(2) rechtsseitig $H_0 = \{\mu \leq \mu_0\}$

(3) zweiseitig $H_0 = \{\mu = \mu_0\}$

Teststatistik: $T(X) = \sqrt{n} \cdot \dfrac{\bar{x} - \mu_0}{\sigma}$, wobei $\bar{x} = \frac{1}{n}(X_1 + \dots + X_N)$.

(für den zweiseitigen Test gilt: Unter H_0 gilt $T(x) \sim N(0,1)$ nach den Rechenregeln für normalverteilte Zufallsvariablen).

$\leadsto$ Bei Signifikanzniveau α lehnen wir H_0 ab, falls:

(1) $T(x) < z_\alpha$

(2) $T(x) > z_{1-\alpha}$

(3) $|T(x)| > z_{1-\alpha/2}$

(b) Zweistichproben Gauss-Test.

Modell: $X_1, \dots, X_N \sim N(\mu_1, \sigma^2)$ mit μ_1 unbekannt, σ^2 bekannt.

$Y_1, \dots, Y_M \sim N(\mu_2, \sigma^2)$ mit μ_2 unbekannt.

Hypothesen: (1) linksseitig $H_0 = \{\mu_1 \geq \mu_2\}$

(2) rechtsseitig $H_0 = \{\mu_1 \leq \mu_2\}$

(3) zweiseitig $H_0 = \{\mu_1 = \mu_2\}$

In R: Gauss.test.

2. t-Test

(a) Einstichproben t - Test:

Modell: $X_1, \dots, X_N \sim N(\mu, \sigma^2)$ mit μ unbekannt, σ^2 unbekannt

(ist beim Gauss-Test bekannt).

Hypothesen: (1) linksseitig $H_0 = \{\mu \geq \mu_0\}$

(2) rechtsseitig $H_0 = \{\mu \leq \mu_0\}$

(3) zweiseitig $H_0 = \{\mu = \mu_0\}$

Teststatistik: $T(x) = \sqrt{n} \dfrac{\bar{x} - \mu_0}{S}$, wobei $\bar{x} = \frac{1}{n}(X_1 + \dots + X_N)$ und $S = \left(\frac{1}{n-1} \sum_{i=1}^{N} (x_i - \bar{x})^2\right)^{1/2}$

(Unter H_0 für zweiseitigen Test: $T(x) \sim t_{n-1}$).

(approximativ, da $\bar{x}$ und S nicht unabh.)

⇝ Bei Signifikanzniveau α lehnen wir H_0 ab, falls:

(1) $T(x) < t_{n-1, \alpha}$

(2) $T(x) > t_{n-1, 1-\alpha}$

(3) $|T(x)| > t_{n-1, 1-\alpha/2}$

(b) Zweistichproben t-Test

Modell: $X_1, \dots, X_N \sim N(\mu_1, \sigma^2)$ mit μ_1 unbekannt, σ^2 unbekannt

$Y_1, \dots, Y_M \sim N(\mu_2, \sigma^2)$ mit μ_2 unbekannt.

Hypothesen: (1) linksseitig $H_0 = \{\mu_1 \geq \mu_2\}$

(2) rechtsseitig $H_0 = \{\mu_1 \leq \mu_2\}$

(3) zweiseitig $H_0 = \{\mu_1 = \mu_2\}$

In R: t.test.

<u>Bemerkung</u> Falls die Varianzen von $X_1, \dots, X_N$ und $Y_1, \dots, Y_M$ nicht als gleich angenommen werden können, macht man den Welch-Test.
(R macht das automatisch).
(D.h. $X_1, \dots, X_N \sim N(\mu_1, \sigma_1^2)$ und $Y_1, \dots, Y_M \sim N(\mu_2, \sigma_2^2)$ und wir wissen nicht ob $\sigma_1 = \sigma_2$).