FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

Stochastik

FH Aachen – Studienstandort Aachen
18. Dezember 2025

Prof. Dr. Florian Heinrichs

Statistische Tests

Münzwurf

Beispiel 148

- Wir werfen eine Münze 100 Mal und erhalten 33 Mal Kopf
- Falls die Münze fair ist ($p = \frac{1}{2}$), ist die Wahrscheinlichkeit dafür ca. 0.023%
- Das ist unwahrscheinlich, aber möglich!
- Können wir überprüfen, ob $p = \frac{1}{2}$?
- Sei $X \sim Bin(100, p)$
- Unter der Hypothese $p = \frac{1}{2}$ gilt

$$\mathbb{P}(X = 50) = \binom{100}{50}\left(\frac{1}{2}\right)^{50}\left(1 - \frac{1}{2}\right)^{100-50} \approx 0.079 = 7.9\%$$

- Die Wahrscheinlichkeit für **genau** 50 Mal Kopf ist also auch klein

Münzwurf

Beispiel 148

- Ab welcher Abweichung von 50 Mal Kopf/Zahl glauben wir nicht mehr an Zufall?
- Wir können die Wahrscheinlichkeit für Abweichungen der Größe $k \leq 50$ explizit berechnen

$$p_k = \mathbb{P}(50 - k \leq X \leq 50 + k)$$
$$= \sum_{i=50-k}^{50+k} \binom{100}{i} \left(\tfrac{1}{2}\right)^i \left(1 - \tfrac{1}{2}\right)^{100-i} = \tfrac{1}{2^{100}} \sum_{i=50-k}^{50+k} \binom{100}{i}$$

- Es ergeben sich folgende Wahrscheinlichkeiten

k	0	1	2	3	5	9	10	11	12	13	14
p_k	0.08	0.236	0.383	0.516	0.729	0.943	0.965	0.979	0.988	0.993	0.996

Münzwurf

Beispiel 148

- Es ergeben sich folgende Wahrscheinlichkeiten

k	0	1	2	3	5	9	10	11	12	13	14
p_k	0.08	0.236	0.383	0.516	0.729	0.943	0.965	0.979	0.988	0.993	0.996

- Mit einer Wahrscheinlichkeit von über 99% liegt X zwischen $50 - 13$ und $50 + 13$
- Die beobachtete Anzahl 33 liegt außerhalb des Bereichs $[37, 63]$
- Die ursprüngliche Hypothese ($p = \frac{1}{2}$) ist also unwahrscheinlich und wir “verwerfen” sie

Testproblem

Definition 66

Sei $X : \Omega \to \mathcal{X}$ eine Zufallsvariable mit statistischem Modell $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$. Seien $\Theta_0, \Theta_1 \subset \Theta$ zwei disjunkte Teilmengen. Die Aussagen $H_0 : \theta \in \Theta_0$ und $H_1 : \theta \in \Theta_1$ heißen *(Null-)Hypothese* bzw. *Alternative*. Das Entscheidungsproblem

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1$$

heißt *Testproblem*.

- $(x_1, \ldots, x_n) = X(\omega)$ ist die Stichprobe, $\mathcal{X}$ der Stichprobenraum.
- Manchmal werden die (Null-)Hypothese und Alternative auch mit H und A bezeichnet.

Münzwurf

Beispiel 148

Bei der Überprüfung, ob eine Münze fair ist, ergeben sich:

- $\Theta = [0, 1]$
- $\Theta_0 = \{0.5\}$
- $\Theta_1 = [0, 0.5) \cup (0.5, 1]$
- Testproblem

$$H_0 : p \in \Theta_0 \quad \text{vs.} \quad H_1 : p \in \Theta_1$$

oder äquivalent dazu

$$H_0 : p = \tfrac{1}{2} \quad \text{vs.} \quad H_1 : p \neq \tfrac{1}{2}.$$

Test

Definition 67

Sei $X : \Omega \to \mathcal{X}$ eine Zufallsvariable mit statistischem Modell $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$. Ein *(statistischer) Test* (auch *Hypothesentest*) ist eine Abbildung $\phi : \mathcal{X} \to \{0, 1\}$. $\phi(x) = 1$ bedeutet, dass wir die Nullhypothese *verwerfen* und $\phi(x) = 0$, dass wir sie *nicht verwerfen*.

- Achtung:
  - Wir verwerfen H_0 oder verwerfen H_0 nicht
  - **Nicht:** Wir "verwerfen H_1" oder "akzeptieren H_0"
- H_0 ist die Aussage, die wir überprüfen/widerlegen wollen
- Wenn ausreichend Daten gegen H_0 sprechen, verwerfen wir H_0
  - Die Daten "bestätigen" nicht H_1, sondern widersprechen H_0
- **Informell:** Es gilt die Unschuldsvermutung (H_0) bis genug Informationen dagegen sprechen.

Test

- Im Beispiel des Münzwurfs haben wir ein Intervall angegeben, in dem der beobachtete Anteil an Würfen mit “Kopf” mit hoher Wahrscheinlichkeit liegt
- Genauer: Unter H_0 ($p = 0.5$) gilt $\mathbb{P}(X \in [37, 63]) \geq 0.99$
- Falls bei weniger/mehr Würfen “Kopf” geworfen wurde, sprechen die Daten **gegen** H_0
- Wichtig: Die Aussage beruht auf Wahrscheinlichkeiten
- Grundsätzlich ist $X = 33$ auch für $p = 0.5$ möglich (nur eben unwahrscheinlich)
- Wir können zwei Fehler machen
  1. Wir verwerfen H_0, falls H_0 wahr ist
  2. Wir verwerfen H_0 nicht, falls H_1 wahr ist

Fehler

Definition 68

Sei $X : \Omega \to \mathcal{X}$ eine Zufallsvariable mit statistischem Modell $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ und sei $\phi : \mathcal{X} \to \{0, 1\}$ ein Test für das Testproblem $H_0 : \theta \in \Theta_0$ *vs.* $H_1 : \theta \in \Theta_1$.

Fehler 1. Art: Eine gültige Nullhypothese wird verworfen, d.h. es gilt $\theta \in \Theta_0$, aber $\phi(x) = 1$.

Fehler 2. Art: Eine ungültige Nullhypothese wird *nicht* verworfen, d.h. es gilt $\theta \notin \Theta_0$, aber $\phi(x) = 0$.

	H_0 ist wahr	H_1 ist wahr
Test verwirft H_0 nicht	Richtige Entscheidung (richtig negativ)	Fehler 2. Art (falsch negativ)
Test verwirft H_0	Fehler 1. Art (falsch positiv)	Richtige Entscheidung (richtig positiv)

FH AACHEN UNIVERSITY OF APPLIED SCIENCES

Fehler

- Natürlich wollen wir (die Wahrscheinlichkeit für) beide Fehler minimieren
- Im Idealfall wäre $\phi(x) = 0$ für $\theta \in \Theta_0$ und $\phi(x) = 1$ für $\theta \in \Theta_1$
- Das ist in vielen Fällen nicht möglich und wir müssen eine Balance finden
  - Ein Test, der den Fehler 1. Art minimiert, verwirft seltener H_0 und macht tendenziell mehr Fehler 2. Art
  - Ein Test, der den Fehler 2. Art minimiert, verwirft öfter H_0 und macht tendenziell mehr Fehler 1. Art
- Wahrscheinlichkeit für einen Fehler:
  - $\theta \in \Theta_0$:
$$\mathbb{P}(\text{“Fehler 1. Art”}) = \mathbb{P}_\theta(\phi(X) = 1).$$
  - $\theta \in \Theta_1$:
$$\mathbb{P}(\text{“Fehler 2. Art”}) = \mathbb{P}_\theta(\phi(X) = 0) = 1 - \mathbb{P}_\theta(\phi(X) = 1).$$

Güte & Niveau

Definition 69

Sei $X : \Omega \to \mathcal{X}$ eine Zufallsvariable mit statistischem Modell $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ und sei $\phi : \mathcal{X} \to \{0,1\}$ ein Test für das Testproblem $H_0 : \theta \in \Theta_0$ *vs.* $H_1 : \theta \in \Theta_1$.

1. Die Funktion $\beta_\phi : \Theta \to [0,1]$, definiert durch
$$\beta_\phi(\theta) = \mathbb{P}_\theta(\phi(X) = 1)$$
ist die *Gütefunktion* des Tests ϕ.
2. Für $\alpha \in [0,1]$, heißt der Test ϕ *Test zum Niveau* α, falls
$$\sup_{\theta \in \Theta_0} \beta_\phi(\theta) \leq \alpha.$$
$\sup_{\theta \in \Theta_0} \beta_\phi(\theta)$ heißt *Niveau* des Tests.
3. Für $\theta \in \Theta_1$ heißt $\beta_\phi(\theta)$ die *Macht* des Tests (auch: *Power*).

Güte & Niveau

Bemerkung 31

> Die Wahrscheinlichkeit für einen Fehler 1. Art ist gegeben durch $\beta_\phi(\theta) = \mathbb{P}_\theta(\phi(X) = 1)$ für $\theta \in \Theta_0$

> Die Wahrscheinlichkeit für einen Fehler 2. Art ist gegeben durch $1 - \beta_\phi(\theta) = \mathbb{P}_\theta(\phi(X) = 0)$ für $\theta \in \Theta_1$

> In der Praxis können wir oft nicht beide Fehlerwahrscheinlichkeiten minimieren

> Wir geben ein Niveau α vor (max. Wahrscheinlichkeit für Fehler 1. Art) und nutzen einen Test mit möglichst viel Power (kleine Wahrscheinlichkeit für Fehler 2. Art)

> Es ist oft schwierig die Power eines Tests zu berechnen

Güte & Niveau

Bemerkung 31

- Wir erhalten durch einen statistischen Test also eine Aussage über H_0 (verwerfen oder nicht)
- Die Wahl von H_0 und H_1 ist **nicht** symmetrisch
  - Wir wählen H_0 und H_1 so, dass Fehler 1. Art gravierender sind
- Ein "guter" Test sollte für $n \to \infty$ immer besser werden
  - Da das Niveau α vorgegeben ist, bedeutet "besser", dass die Power gegen 1 konvergiert: $\lim_{n\to\infty} \beta_\phi(\theta) = 1$ für $\theta \in \Theta_1$
  - Diese Eigenschaft heißt *Konsistenz*

	H_0 ist wahr ($\theta \in \Theta_0$)	H_1 ist wahr ($\theta \in \Theta_1$)
Test verwirft H_0 nicht	$1 - \beta_\phi(\theta) \geq 1 - \alpha$	$1 - \beta_\phi(\theta)$
Test verwirft H_0	$\beta_\phi(\theta) \leq \alpha$	$\beta_\phi(\theta)$

Güte & Niveau

Beispiel 148: (Münzwurf - Fortsetzung)

Bei der Überprüfung, ob eine Münze fair ist, ergeben sich:

- Anzahl "Kopf" im n-fachen Münzwurf: $X \sim Bin(n,p)$
- Wir verwerfen die Nullhypothese $p = \frac{1}{2}$, falls $|X - \frac{n}{2}| > k$.
  - k ist ein *kritischer Wert*
  - $\frac{n}{2}$ ist der Erwartungswert von X unter H_0.
- Für die Gütefunktion ergibt sich

$$\beta_\phi(p) = \mathbb{P}_p(\phi(X) = 1) = \mathbb{P}_p(|X - \tfrac{n}{2}| > k) = 1 - \mathbb{P}_p(|X - \tfrac{n}{2}| \leq k)$$

$$= 1 - \mathbb{P}_p(\tfrac{n}{2} - k \leq X \leq \tfrac{n}{2} + k) = 1 - \sum_{i=\lceil \frac{n}{2} \rceil - k}^{\lfloor \frac{n}{2} \rfloor + k} \binom{n}{i} p^i (1-p)^{n-i}$$

Güte & Niveau

Beispiel 148: (Münzwurf - Fortsetzung)

Bei der Überprüfung, ob eine Münze fair ist, ergeben sich:

- Anzahl "Kopf" im n-fachen Münzwurf: $X \sim Bin(n, p)$
- Wir verwerfen $H_0 : p = \frac{1}{2}$, falls $|X - \frac{n}{2}| > k$
- Gütefunktion: $\beta_\phi(p) = 1 - \sum_{i=\lceil \frac{n}{2} \rceil - k}^{\lfloor \frac{n}{2} \rfloor + k} \binom{n}{i} p^i (1-p)^{n-i}$
- Für $n = 100$ und $k = 13$ ist das Niveau des Tests

$$\begin{aligned} \beta_\phi(\tfrac{1}{2}) &= 1 - \sum_{i=37}^{63} \binom{100}{i} (\tfrac{1}{2})^i (1 - \tfrac{1}{2})^{100-i} \\ &= 1 - \tfrac{1}{2^{100}} \sum_{i=37}^{63} \binom{100}{i} = 0.0066 = 0.66\% \end{aligned}$$

Güte & Niveau

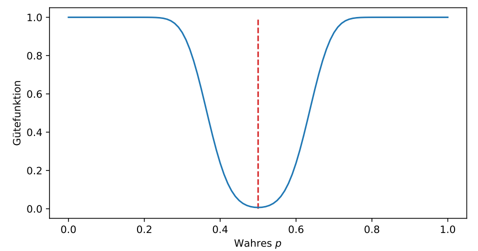
Beispiel 148: (Münzwurf - Fortsetzung)

Bei der Überprüfung, ob eine Münze fair ist, ergeben sich:

> Anzahl "Kopf" im 100-fachen Münzwurf: $X \sim Bin(100, p)$

> Wir verwerfen $H_0 : p = \frac{1}{2}$, falls $|X - 50| > 13$

> Gütefunktion: $\beta_\phi(p) = 1 - \sum_{i=37}^{63} \binom{100}{i} p^i (1-p)^{100-i}$



Tests & Konfidenzbereiche

Wie hängen Tests und Konfidenzbereiche zusammen?

> Sei θ ein unbekannter Parameter

> $I(X_1, \ldots, X_n)$ ist ein $(1-\alpha)$-Konfidenzbereich für θ, falls

$$\mathbb{P}_\theta(\theta \in I(X_1, \ldots, X_n)) \geq 1 - \alpha.$$

> Wir erhalten einen Test für $H_0 : \theta = \theta_0$, wenn wir H_0 verwerfen, falls $\theta_0 \notin I(X_1, \ldots, X_n)$ und andernfalls H_0 nicht verwerfen:

$$\mathbb{P}_{\theta_0}(\text{Verwerfe } H_0) = \mathbb{P}_{\theta_0}(\theta_0 \notin I(X_1, \ldots, X_n)) \leq \alpha$$

> Andersrum definiert ein Test mit Niveau α einen $(1-\alpha)$-Konfidenzbereich

$$I(X_1, \ldots, X_n) = \{\theta_0 \in \Theta : \text{Verwerfe nicht } H_0 : \theta = \theta_0\}$$

denn $\mathbb{P}_{\theta_0}(\theta_0 \in I(X_1, \ldots, X_n)) \geq 1 - \alpha$.

p-Wert

- Sei $X \hat{=}$ "Anzahl Kopf" beim 100-fachen Münzwurf
- Für $\alpha = 1\%$, verwerfen wir $H_0 : p = \frac{1}{2}$, falls $X \notin [37, 63]$
- Aber: $X = 10$ spricht stärker gegen H_0 als $X = 33$, wie können wir das messen?
- Oft formulieren wir Tests mit Hilfe einer *Teststatistik* $t : \mathcal{X} \to \mathbb{R}$ und lehnen H_0 ab, falls $t(x) > c_\alpha$, wobei x die beobachtete Stichprobe bezeichnet und c_α einen kritischen Wert, der vom Niveau α abhängt
- Der p-*Wert* gibt die Wahrscheinlichkeit an, dass die Teststatistik $t(X)$ Werte annimmt, die so extrem oder extremer als die beobachtete Teststatistik $t(x)$ sind
- Wenn der p-Wert $\mathbf{p}$ klein ist, ist die Wahrscheinlichkeit für die beobachtete Teststatistik unter H_0 klein
- Wir verwerfen $H_0 \iff \mathbf{p} < \alpha$

p-Wert

Definition 70

Sei $X : \Omega \to \mathcal{X}$ eine Zufallsvariable mit statistischem Modell $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$. Weiter sei $t : \mathcal{X} \to \mathbb{R}$ eine *Teststatistik* für das Testproblem $H_0 : \theta \in \Theta_0$ *vs.* $H_1 : \theta \in \Theta_1$ und c_α ein *kritischer Wert*, sodass die Entscheidungsregel

"Verwerfe H_0, falls $t(x) > c_\alpha$"

einen Test zum Niveau α definiert. Der p-*Wert* ist definiert als

$$\mathbf{p} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(t(X) \geq t(x)).$$

Interpretation: Wahrscheinlichkeit, dass $t(X)$ Werte annimmt, die so extrem oder extremer sind als die beobachtete Teststatistik $t(x)$.

p-Wert

Beispiel 149

Betrachte beim 100-fachen Münzwurf $H_0 : p = \frac{1}{2}$ vs. $H_1 : p \neq \frac{1}{2}$

> Niveau: $\alpha = 0.01$

> Stichprobe: $x = (x_1, \ldots, x_{100}) \in \{0, 1\}^{100}$

> Teststatistik: $t(x) = \left| \sum_{i=1}^{100} x_i - 50 \right|$

> Kritischer Wert: $c_\alpha = 13$

> Falls 33 Mal "Kopf" geworfen wurde, verwerfen wir H_0, da

$$t(x) = |33 - 50| = 17 > 13 = c_\alpha.$$

> Der p-Wert ergibt sich als

$$\mathbf{p} = \mathbb{P}_{0.5}(t(X) \geq 17) = \tfrac{1}{2^{100}} \sum_{i=0}^{33} \binom{100}{i} + \tfrac{1}{2^{100}} \sum_{i=67}^{100} \binom{100}{i} = 0.00087 = 0.087\%$$

Da $p \ll 1\%$, sprechen die Daten stark gegen H_0.

Wichtige Tests

- Seien $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ unabhängige Zufallsvariablen
- Oft wollen wir Hypothesen der Form

$$H_0 : \mu \leq \mu_0 \text{ vs. } H_1 : \mu > \mu_0$$
$$H_0 : \mu \geq \mu_0 \text{ vs. } H_1 : \mu < \mu_0$$
$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0$$

  testen
- Falls σ^2 bekannt ist: **Gauß-Test**
- Falls σ^2 unbekannt ist: t-**Test**
- Die Tests gehen im Wesentlichen auf die zugehörigen Konfidenzintervalle zurück
- Die Herleitungen sind analog zu denen der Konfidenzintervalle

Gauß-Tests

- Seien $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ unabh. und σ^2 bekannt
  - Formal: $X = (X_1, \dots, X_n) \in \mathcal{X}$
- Bezeichne mit q_α das α-Quantil von $\mathcal{N}(0,1)$
- Die Teststatistik ist

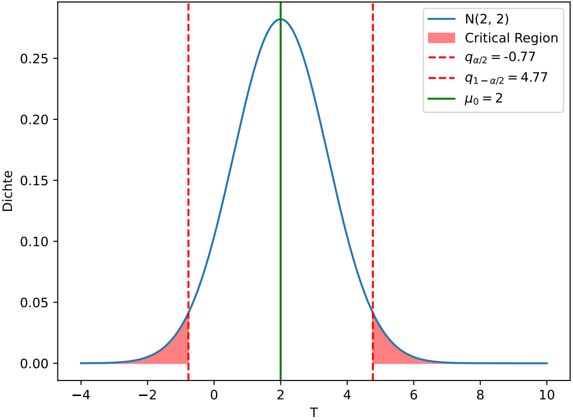
$$T = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma}$$

- Falls $\mu = \mu_0 : T \sim \mathcal{N}(0,1)$

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

Gauß-Tests

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



t-Tests

- Seien $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ unabh. und σ^2 **unbekannt**
- Bezeichne mit $t_{n-1,\alpha}$ das α-Quantil der t-Verteilung mit $(n-1)$ Freiheitsgraden und s_n^2 die Stichprobenvarianz
- Die Teststatistik ist

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

- Falls $\mu = \mu_0 : T \sim t_{n-1}$

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

Übungen

Übung 89

Die Körpergröße des Menschen folgt näherungsweise einer Normalverteilung $\mathcal{N}(\mu, \sigma^2)$. Es sei bekannt, dass die (wahre) Varianz $\sigma^2 = 100\,\text{cm}^2$ beträgt. Wir vermuten, dass die mittlere Körpergröße von Basketballspielern bei $194\,$cm liegt. In einer Studie wurde die Körpergröße von 81 Basketballspielern gemessen, wobei sich ein Mittelwert in Höhe von $189\,$cm und eine empirische Varianz in Höhe von $81\,\text{cm}^2$ ergeben hat.

1. Testen Sie unsere Vermutung mit einem geeigneten Test zum Niveau $\alpha = 0.05$:
$$H_0 : \mu = 194\,\text{cm} \quad \text{vs.} \quad H_1 : \mu \neq 194\,\text{cm}$$
Berechnen Sie anschließend den p-Wert $\mathbf{p}$.
2. Im Folgenden sei die wahre Varianz σ^2 unbekannt. Testen Sie unsere Vermutung mit einem geeigneten Test bei unbekannter Varianz.

Gauß-Test

Wie kommen wir zur Entscheidungsregel des Gauß-Tests?

> Seien $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ unabh. und σ^2 bekannt

> Testproblem $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

> Bezeichne mit q_α das α-Quantil von $\mathcal{N}(0,1)$

> Ein zweiseitiges $(1-\alpha)$-Konfidenzintervall für μ ist gegeben durch $[\bar{X}_n - \frac{q_{1-\alpha/2}\sigma}{\sqrt{n}}, \bar{X}_n + \frac{q_{1-\alpha/2}\sigma}{\sqrt{n}}]$, d.h.

$$\mathbb{P}\Big(\mu \in \Big[\bar{X}_n - \frac{q_{1-\alpha/2}\sigma}{\sqrt{n}}, \bar{X}_n + \frac{q_{1-\alpha/2}\sigma}{\sqrt{n}}\Big]\Big)$$

$$= \mathbb{P}\Big(-q_{1-\alpha/2} \le \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \le q_{1-\alpha/2}\Big) = 1 - \alpha$$

> Sei $T = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma}$ die Teststatistik

> Unter H_0 gilt $\mu = \mu_0$ und damit $\mathbb{P}(|T| > q_{1-\alpha/2}) = \alpha$

Gauß-Test

Wie hoch sind jeweils die Wahrscheinlichkeiten für Fehler 1./2. Art?

> Berechne zunächst die Gütefunktion $\beta_\phi(\mu)$ des Gauß-Tests

> Betrachte $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$

> Wir lehnen H_0 ab, falls $T > q_{1-\alpha}$, also

$$\begin{aligned}\beta_\phi(\mu) &= \mathbb{P}_\mu(T > q_{1-\alpha}) = \mathbb{P}_\mu\Big(\sqrt{n}\tfrac{\bar{X}_n-\mu_0}{\sigma} > q_{1-\alpha}\Big)\\ &= \mathbb{P}_\mu\Big(\sqrt{n}\tfrac{\bar{X}_n-\mu}{\sigma} > q_{1-\alpha} + \tfrac{\sqrt{n}}{\sigma}(\mu_0-\mu)\Big)\\ &= 1 - \Phi\Big(q_{1-\alpha} + \tfrac{\sqrt{n}}{\sigma}(\mu_0-\mu)\Big)\end{aligned}$$

> Analog gilt für den Gauß-Test zur Nullhypothesen $H_0 : \mu \geq \mu_0$

$$\beta_\phi(\mu) = \mathbb{P}_\mu(T < q_\alpha) = 1 - \Phi\Big(q_{1-\alpha} - \tfrac{\sqrt{n}}{\sigma}(\mu_0-\mu)\Big)$$

> Und für den Gauß-Test zur zweiseitigen Hypothese $H_0 : \mu = \mu_0$

$$\beta_\phi(\mu) = 2 - \Phi\Big(q_{1-\alpha/2} + \tfrac{\sqrt{n}}{\sigma}(\mu_0-\mu)\Big) - \Phi\Big(q_{1-\alpha/2} - \tfrac{\sqrt{n}}{\sigma}(\mu_0-\mu)\Big)$$

Gauß-Test

Wie hoch sind jeweils die Wahrscheinlichkeiten für Fehler 1./2. Art?

- Die Wahrscheinlichkeit für Fehler 1. Art ist $\beta_\phi(\mu)$, für $\mu \in \Theta_0$
  - Außerdem: $\beta_\phi(\mu) \leq \alpha$ für $\mu \in \Theta_0$
- Die Wahrscheinlichkeit für Fehler 2. Art ist $1 - \beta_\phi(\mu)$, für $\mu \in \Theta_1$
- Für die drei Gauß-Tests ergeben sich folgende Fehlerwahrscheinlichkeiten

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

Gauß-Test

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

Beispiel 150

Betrachte das Testproblem $H_0 : \mu \leq 0$ vs. $H_1 : \mu > 0$.

> Sei $\sigma^2 = 1, n = 100$ und das wahre $\mu = 1$

> Das Niveau sei $\alpha = 0.05$

> Es gilt $\mu_0 = 0$

> Wie groß ist die Wahrscheinlichkeit für einen Fehler 2. Art?

$$\begin{aligned}\Phi\big(q_{1-\alpha} + \sqrt{n}\tfrac{(\mu_0-\mu)}{\sigma}\big) &= \Phi\big(1.64 + \tfrac{\sqrt{100}}{1}(0-1)\big)\\ &= \Phi(1.64 - 10) = 3.14 \cdot 10^{-17} \approx 0\end{aligned}$$

Gauß-Tests

Optional 10

Woher kommen die p-Werte?

> Betrachte $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$

> Falls $\mu = \mu_0$ gilt $T \sim \mathcal{N}(0, 1)$

> Für eine beobachtete Stichprobe x gilt

$$\sup_{\mu \leq \mu_0} \mathbb{P}_\mu(T \geq t(x)) = \mathbb{P}_{\mu_0}(T \geq t(x)) = 1 - \Phi(t(x)),$$

wobei $t(x) = \sqrt{n}\frac{\bar{x}_n - \mu_0}{\sigma}$

Übungen

Übung 90

In einer Abfüllanlage werden 1 l Flaschen mit Wasser befüllt. Die Füllmenge ist normalverteilt mit Varianz $400\,\text{ml}^2$, d.h. $\mathcal{N}(\mu, 400)$. Der Erwartungswert sollte mindestens $1000\,\text{ml}$ betragen. Bei einer Kontrolle von 20 Flaschen wurde jedoch eine mittlere Füllmenge von $990\,\text{ml}$ gemessen. Wird der geforderte Erwartungswert trotzdem eingehalten?

1. Überprüfen Sie die Vermutung mit einem geeigneten Test zum Niveau $\alpha = 0.01$:
$$H_0 : \mu \geq 1000\,\text{ml} \quad \text{vs.} \quad H_1 : \mu < 1000\,\text{ml}$$
2. Bestimmen Sie den p-Wert $\mathbf{p}$.
3. Wie hoch ist die W'keit für einen Fehler 1. Art, falls $\mu = 1000$?
4. Wie hoch ist die W'keit für einen Fehler 1. Art, falls $\mu = 1010$?
5. Wie hoch ist die W'keit für einen Fehler 2. Art, falls $\mu = 990$?

Zweistichproben-t-Test

- Bisher hatten wir eine Stichprobe und wollten den Erwartungswert testen.
- Manchmal haben wir aber auch zwei Stichproben und wollen die Erwartungswerte der Stichproben vergleichen.
  - Beispiel: Testen der Wirkung eines Medikaments.
- Die beiden Stichproben können abhängig oder unabhängig sein.
- Abhängig:
  - Gepaarte Daten $(x_1, y_1), \ldots, (x_n, y_n)$
  - Im Beispiel: Wert eines Hormons vor (x_i) bzw. nach (y_i) Einnahme des Medikaments.
- Unabhängig:
  - Zwei Stichproben $x_1, \ldots, x_n$ und $y_1, \ldots, y_m$
  - Im Beispiel: Wert eines Hormons mit (x_i) bzw. ohne (y_i) Einnahme des Medikaments.

Zweistichproben-t-Test

Wie können wir für **gepaarte Daten** entscheiden, ob sich die Erwartungswerte der Gruppen unterscheiden?

- Seien $X_1, \ldots, X_n \sim \mathcal{N}(\mu_X, \sigma_X^2)$ und $Y_1, \ldots, Y_n \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$
- Vergleiche Differenzen: $D_i := X_i - Y_i \sim \mathcal{N}(\mu_X - \mu_Y, \sigma_{X+Y}^2)$
- Nullhypothese: $H_0 : \mu_X - \mu_Y = 0$.
- Annahme: $D_1, \ldots, D_n$ unabhängig.
- Nutze (Einstichproben) t-Test.

Zweistichproben-t-Test

Wie können wir für **unabhängige Stichproben** entscheiden, ob sich die Erwartungswerte der Gruppen unterscheiden?

> Seien $X_1, \ldots, X_n \sim \mathcal{N}(\mu_X, \sigma_X^2)$ und $Y_1, \ldots, Y_n \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ paarweise unabhängige Zufallsvariablen.

> Zweiseitiges Testproblem: $H_0 : \mu_X = \mu_Y$ vs. $H_1 : \mu_X \neq \mu_Y$.

> Aufgrund der Unabhängigkeit gilt
$$\bar{X}_n - \bar{Y}_n \sim \mathcal{N}(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}).$$

> Teststatistik
$$T := \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim \mathcal{N}\Big(\frac{\mu_X - \mu_Y}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}, 1\Big).$$

> Unter H_0 ist $T \sim \mathcal{N}(0,1)$.

> Verwerfe H_0, falls $|T| > q_{1-\alpha/2}$, wobei $q_{1-\alpha/2}$ das $(1-\frac{\alpha}{2})$-Quantil der Standardnormalverteilung bezeichnet.

Zweistichproben-t-Test

- Seien $X_1, \dots, X_n \sim \mathcal{N}(\mu_X, \sigma^2)$ und $Y_1, \dots, Y_n \sim \mathcal{N}(\mu_Y, \sigma^2)$ paarweise unabhängige Zufallsvariablen.
- Zweiseitiges Testproblem: $H_0 : \mu_X = \mu_Y$ vs. $H_1 : \mu_X \neq \mu_Y$.
- Falls die Varianzen **unbekannt** sind, müssen wir sie schätzen!
- Es gilt: $\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} = \frac{\sigma^2}{n} + \frac{\sigma^2}{m} = \frac{m+n}{mn}\sigma^2$.
- Seien s_X^2 und s_Y^2 die empirischen Varianzen von X bzw. Y.
- Schätzer: $s^2 := \frac{1}{n+m-2}((n-1)s_X^2 + (m-1)s_Y^2)$
- Dann gilt $\frac{n+m-2}{\sigma^2}s^2 \sim \chi^2_{n+m-2}$
- Teststatistik
$$T := \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{m+n}{mn}s^2}}.$$
- Unter H_0 ist $T \sim t_{n+m-2}$.
- Verwerfe H_0, falls $|T| > t_{n+m-2,1-\alpha/2}$.

Zweistichproben-t-Test

Übung 91

In einer Studie wird die Wirkung eines Konzentrationstrainings untersucht. Von 10 Teilnehmern wird jeweils vor und nach dem Training ein Konzentrationstest durchgeführt. Die gemessenen Punktzahlen lauten:

Person	1	2	3	4	5	6	7	8	9	10
Vorher (X_i)	12	15	14	13	16	11	17	14	15	13
Nachher (Y_i)	14	16	15	15	17	13	18	15	16	14

Testen Sie zum Niveau $\alpha = 0.05$, ob das Training einen Einfluss auf die Konzentrationsleistung hat, d.h. ob die Erwartungswerte verschieden sind.

Chi-Quadrat-Test

- Für den Gauß-Test haben wir angenommen, dass σ^2 bekannt ist.
- Wie können wir überprüfen, ob σ^2 wirklich mit einem bestimmten Wert σ_0^2 übereinstimmt? $\rightarrow$ Testen!
- Seien $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ unabhängige Zufallsvariablen.
- Testproblem: $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 \neq \sigma_0^2$.
- Falls μ bekannt:

$$T_0 := \hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 \sim \frac{\sigma^2}{n} \chi_n^2.$$

- Falls μ unbekannt:

$$T_1 := s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \sim \frac{\sigma^2}{n-1} \chi_{n-1}^2.$$

- Verwerfe H_0, falls $\frac{n-i}{\sigma_0^2} T_i \notin [\chi^2_{n-i,\alpha/2}, \chi^2_{n-i,1-\alpha/2}]$, für $i = 0, 1$.

Zweistichproben-t-Test

Übung 92

Eine Maschine füllt Joghurtbecher ab. Die Füllmenge ist normalverteilt mit unbekanntem Erwartungswert und soll eine Standardabweichung von höchstens $2\,g$ aufweisen. Bei einer Stichprobe von $n = 8$ Bechern wurden folgende Füllmengen (in g) gemessen:

$$108, \quad 106, \quad 109, \quad 107, \quad 110, \quad 105, \quad 111, \quad 108$$

Berechnen Sie die empirische Varianz und testen Sie mit einem Chi-Quadrat-Test zum Niveau $\alpha = 0.05$, ob die Standardabweichung ungleich $2\,g$ ist, d.h. $H_0 : \sigma^2 = 4$ vs. $H_1 : \sigma^2 \neq 4$.

Weitere wichtige Tests

Es gibt noch einige weitere wichtige Tests

- Seien $X_1, \ldots, X_n \sim \mathcal{N}(\mu_X, \sigma_X^2)$ und $Y_1, \ldots, Y_m \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ unabhängige Zufallsvariablen
- Der "Zweistichproben t-Test" (falls $\sigma_X^2 = \sigma_Y^2$)
  - Teste $H_0 : \mu_X = \mu_Y$ vs. $H_1 : \mu_X \neq \mu_Y$
  - Alternativ: $\leq$ / $\geq$
  - Falls $\sigma_X^2 \neq \sigma_Y^2$: Welch-Test
- F-Test: $H_0 : \sigma_X^2 = \sigma_Y^2$ vs. $H_1 : \sigma_X^2 \neq \sigma_Y^2$
- Chi-Quadrat-Test: $H_0 : \sigma_X^2 = \sigma_0^2$ vs. $H_1 : \sigma_X^2 \neq \sigma_0^2$
  - Alternativ: $\leq$ / $\geq$
- Chi-Quadrat-Test: $H_0 : X, Y$ unabhängig vs. $H_1 : X, Y$ nicht unabhängig
- Chi-Quadrat-Test: $H_0 : X \sim F$ vs. $H_1 :$ nicht $X \sim F$

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 $\leq$)

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.

FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

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.