Allgemeines Lineares Modell

BSc Psychologie, SoSe 2026

Joram Soch

(9) Einstichproben-T-Tests

Überblick

Modul B2: Inferenzstatistik | Allgemeines Lineares Modell

Datum	Einheit	Do, 11-15 (ca. 11:15-14:00)	
09.04.2026	Grundlagen	(0) Formalia	(1) Regression
16.04.2026	Grundlagen	(2) Korrelation	
23.04.2026	Grundlagen	(3) Matrizen	
30.04.2026	Grundlagen / Theorie	(4) Normalverteilungen	(5) Modellformulierung
07.05.2026	Theorie	(6) Parameterschätzung	
14.05.2026	– Feiertag –	– keine Vorlesung –	
21.05.2026	Theorie	(7) T-Statistiken	
28.05.2026	Theorie	(8) F-Statistiken	
04.06.2026	Anwendung	(9) Einstichproben-T-Tests	(10) Zweistichproben-T-Tests
11.06.2026	Anwendung	(11) Einfaktorielle Varianzanalyse	
18.06.2026	Anwendung	(12) Zweifaktorielle Varianzanalyse	– Vorlesung online –
25.06.2026	Anwendung	(13) Partielle Korrelation	
02.07.2026	Anwendung	(14) Multiple Regression	
09.07.2026	Anwendung	(15) Kovarianzanalyse	
17.07.2026	Klausurtermin		
Februar 2027	Klausurwiederholungstermin		

Überblick

Kontinuum von ALM-Designs

Extremszenario (1) Die Erwartungswerte aller Datenvariablen sind identisch.

$$
\begin{aligned}
&y_i \sim N(\mu, \sigma^2) \quad \text{u.i.v. für} \quad i = 1, ..., n \quad \Leftrightarrow \\
&y = X\beta + \varepsilon, \quad X := 1_n \in \mathbb{R}^{n \times 1}, \quad \beta := \mu \in \mathbb{R}, \quad \varepsilon \sim N(0_n, \sigma^2 I_n)
\end{aligned}
\tag{1}
$$

$\Rightarrow$ Jegliche Datenvariabilität wird dem Fehlerterm zugeschrieben.

$\Rightarrow$ Es gilt $\quad \hat{\beta} = (1_n^{\mathrm{T}} 1_n)^{-1} 1_n^{\mathrm{T}} y = \bar{y} \quad$ und $\quad \hat{\sigma}^2 = \frac{1}{n-p}(y - 1_n \bar{y})^{\mathrm{T}}(y - 1_n \bar{y}) = s_y^2.$

Extremszenario (2) Die Erwartungswerte aller Datenvariablen sind paarweise verschieden.

$$
\begin{aligned}
&y_i \sim N(\mu_i, \sigma^2) \quad \text{u.v. für} \quad i = 1, ..., n \quad \Leftrightarrow \\
&y = X\beta + \varepsilon, \quad X := I_n \in \mathbb{R}^{n \times n}, \quad \beta := (\mu_1, ..., \mu_n)^{\mathrm{T}} \in \mathbb{R}^n, \quad \varepsilon \sim N(0_n, \sigma^2 I_n)
\end{aligned}
\tag{2}
$$

$\Rightarrow$ Jegliche Datenvariabilität wird dem Erwartungswertparameter zugeschrieben.

$\Rightarrow$ Es gilt $\quad \hat{\beta} = (I_n^{\mathrm{T}} I_n)^{-1} I_n^{\mathrm{T}} y = y \quad$ und $\quad \hat{\sigma}^2 = \frac{1}{n-p}(y - I_n y)^{\mathrm{T}}(y - I_n y) = 0.$

Beide Extremszenarien sind wissenschaftlich nicht ergiebig, da sie keine theoriegeleitete systematische Abhängigkeit zwischen unabhängigen und abhängigen Variablen (UV, AV) annehmen. Die im weiteren Verlauf betrachteten ALM-Designs liegen zwischen den beiden Extremszenarien und repräsentieren verschiedene Formen der systematischen Abhängigkeit zwischen UV und AV.

Überblick

Faktorielle und parametrische ALM-Designs

Faktorielle ALM-Designs:

- Designmatrizen mit 1en und 0en, manchmal -1en.
- Betaparameter repräsentieren Erwartungswerte für Gruppen.
- Betaparameterschätzer repräsentieren Stichprobenmittel für Gruppen.
- $\Rightarrow$ T-Tests, einfaktorielle Varianzanalyse, zweifaktorielle Varianzanalyse

Parametrische ALM-Designs:

- Designmatrizen besitzen Spalten mit kontinuierlichen reellen Werten.
- Die Designmatrixsspalten werden *Regressoren*, *Prädiktoren*, oder *Kovariaten* genannt.
- Betaparameter repräsentieren Steigungsparameter.
- Betaparameterschätzer ergeben sich als normalisierte Kovarianzen von Regressoren und Daten.
- Es besteht ein enger Bezug zur Theorie der Korrelation.
- $\Rightarrow$ einfache lineare Regression, multiple lineare Regression

Faktoriell-parametrische ALM-Designs:

- Designmatrizen mit mehreren faktoriellen und parametrischen Werten.
- Die parametrischen Regressoren werden oft als kontrollierte Kovariaten betrachtet.
- $\Rightarrow$ Kovarianzanalyse, (partielle Korrelation)

Überblick

ALM-Designs als Hypothesentestverfahren*

Testen von Unterschiedshypothesen:

- Einstichproben-T-Tests
- Zweistichproben-T-Tests
- einfaktorielle Varianzanalyse
- zweifaktorielle Varianzanalyse
- Kovarianzanalyse

Testen von Zusammenhangshypothesen:

- einfache lineare Regression/Korrelation
- multiple lineare Regression/partielle Korrelation

* Diese Sichtweise wird durch Prof. Ostwald nicht favorisiert.

Überblick

T-Tests

Es gibt viele T-Test-Varianten, jeweils mit eigenen Testgütefunktionen.

Wir fokussieren hier auf die Darstellung von T-Tests als Spezialfällen des ALMs.

Wir behandeln im Detail:

- Einstichproben-T-Tests mit einfacher Nullhypothese und ungerichteter Alternativhypothese.
- Zweistichproben-T-Tests bei unabhängigen Stichproben unter Annahme identischer Varianzen mit einfacher Nullhypothese und ungerichteter Alternativhypothese (siehe Einheit (10) in *Allgemeines Lineares Modell*)

Wir behandeln nicht:

- T-Tests mit gerichteten Hypothesen oder einfachen Null- und Alternativhypothesen.
- Zweistichproben-T-Tests bei Annahme verschiedener Varianzen (Behrends-Fischer-Problem).
- Zweistichproben T-Tests bei abhängigen Stichproben.

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

Selbstkontrollfragen

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

Selbstkontrollfragen

Anwendungsszenario

Eine Gruppe (Stichprobe) randomisierter experimenteller Einheiten.

Annahme der unabhängigen und identischen Normalverteilung $N(\mu, \sigma^2)$.

μ und σ^2 unbekannt.

Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ mit μ_0 beabsichtigt.

Anwendungsbeispiele

Pre-Post-Psychotherapie: BDI-Differenzanalyse einer Gruppe von Patient:innen

- $\mu \neq \mu_0 := 0 \quad \Rightarrow$ Evidenz für Depressionsymptomatikveränderung

Gruppenanalysen mit *Wechsler Adult Intelligence Scale* (WAIS)

- $\mu \neq \mu_0 := 100 \Rightarrow$ Evidenz für über- oder unterdurchschnittliche WAIS-Performanz

Gruppenanalysen in der funktionellen Kernspintomographie (fMRT)

- $\mu > \mu_0 := 0 \quad \Rightarrow$ Evidenz für regionale Gehirnaktivierung

Anwendungsszenario

Anwendungsbeispiel

Face-to-face

PreBDI  PostBDI

Wir nehmen an, dass die Datenpunkte unabhängige und identisch verteilte (u.i.v.) Realisierungen von ZVen $y_i \sim N(\mu, \sigma^2)$ sind und nehmen weiter an, dass wir sind an der Quantifizierung der Unsicherheit beim inferentiellen Vergleich des wahren, aber unbekannten Erwartungswertparameters μ im Sinne eines Hypothesentests interessiert sind.

Definition (Einstichproben-T-Test-Modell)

$y_i, i = 1, ..., n$ seien Zufallsvariablen, die die n Datenpunkte eines Anwendungsszenarios für den Einstichproben-T-Test modellieren. Dann hat das *Einstichproben-T-Test-Modell* die strukturelle Form

$$y_i = \mu + \varepsilon_i \quad \text{mit} \quad \varepsilon_i \sim N(0, \sigma^2) \quad \text{u.i.v. für} \quad i = 1, ..., n \quad \text{mit} \quad \mu \in \mathbb{R} \quad \text{und} \quad \sigma^2 > 0, \tag{3}$$

die Datenverteilungsform

$$y_i \sim N(\mu, \sigma^2) \quad \text{u.i.v. für} \quad i = 1, ..., n \quad \text{mit} \quad \mu \in \mathbb{R} \quad \text{und} \quad \sigma^2 > 0, \tag{4}$$

und für den Datenvektor $y = (y_1, ..., y_n)^{\mathrm{T}}$ die Designmatrixform

$$y = X\beta + \varepsilon \quad \text{mit} \quad X := 1_n \in \mathbb{R}^{n \times 1}, \quad \beta := \mu \in \mathbb{R}, \quad \varepsilon \sim N(0_n, \sigma^2 I_n), \quad \text{und} \quad \sigma^2 > 0. \tag{5}$$

Bemerkungen

- Das Modell ist identisch mit dem Modell unabhängiger und identisch normalverteilter Zufallsvariablen.
- Die Äquivalenz der drei Modellformen wurde bereits diskutiert (vgl. Einheit (5) in *Allgemeines Lineares Modell*).
- Die Anzahl der Betaparameter ist $p = 1$.

Modellformulierung

Designmatrix des Einstichproben-T-Test-Modells ($n = 20$, $p = 1$)

Datensimulation (vgl. Einheit (5) in *Allgemeines Lineares Modell*)

```
# Modellformulierung
library(MASS)                                  # multivariate Normalverteilung
n      = 20                                    # Anzahl von Datenpunkten
p      = 1                                     # Anzahl von Betaparametern
X      = matrix(rep(1,n), nrow = n)            # n x p Designmatrix
I_n    = diag(n)                               # n x n Einheitsmatrix
beta   = 5                                     # wahrer, aber unbekannter Betaparameter
sigsqr = 14                                    # wahrer, aber unbekannter Varianzparameter

# Datenrealisierung
y      = mvrnorm(1, X %*% beta, sigsqr*I_n)  # eine Realisierung des n-dimensionalen ZVs y
print(y)
```

```
 [1]  6.3323967  6.8906198  3.8857604  9.2885623 -2.8381190  2.8728300
 [7] -0.8463879  4.3471780 13.2448606  9.3029791 -2.7044583  3.0756980
[13]  2.4637458  1.0785485  4.4239466  4.8236078  2.5477300  7.1895089
[19]  2.3182476  9.2401002
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

Selbstkontrollfragen

Theorem (Parameterschätzung im Einstichproben-T-Test-Modell)

Gegeben sei die Designmatrixform des Einstichproben-T-Test-Modells. Dann ergeben sich für den Betaparameterschätzer

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} y_i =: \bar{y} \tag{6}$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 =: s_y^2 \,. \tag{7}$$

Bemerkungen

- Die Formen von $\hat{\beta}$ und $\hat{\sigma}^2$ wurden bereits hergeleitet (siehe Einheit (6) in *Allgemeines Lineares Modell*).
- $\bar{y}$ und s_y^2 bezeichnen das Stichprobenmittel und die Stichprobenvarianz der $y_1, ..., y_n$.

Modellschätzung

```r
# Daten einlesen
fname       = "Daten/T-Tests_Daten.csv"                               # Dateiname
D           = read.table(fname, sep = ",", header = TRUE)             # Dataframe
y           = D$dBDI[D$Setting == "F2F"]                              # BDI-Differenzwerte in der F2F-Gruppe

# Modellformulierung
n          = length(y)                                                # Anzahl Datenpunkte
p          = 1                                                        # Anzahl Betaparameter
X          = matrix(rep(1,n), nrow = n)                               # Designmatrix

# Modellschätzung
beta_hat    = solve(t(X) %*% X) %*% t(X) %*% y                        # Betaparameterschätzer
eps_hat     = y - X %*% beta_hat                                      # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) /(n-p)                          # Varianzparameterschätzer

# Ausgabe
cat(  "hat{beta}   : ", beta_hat,                                     # Betaparameterschätzer
    "\nbar{y}      : ", mean(y),                                      # Stichprobenmittel
    "\nhat{sigsqr} : ", sigsqr_hat,                                   # Varianzparameterschätzer
    "\ns_y^2       : ", var(y))                                       # Stichprobenvarianz
```

```
hat{beta}   :  3.925
bar{y}      :  3.925
hat{sigsqr} :  19.40449
s_y^2       :  19.40449
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

Selbstkontrollfragen

Modellevaluation

Hypothesenszenarien

Einfache Nullhypothese, einfache Alternativhypothese $H_0 : \mu = \mu_0$, $H_1 : \mu = \mu_1$

- theoretisch wichtiges Szenario (Neymann-Pearson-Lemma)
- praktische Relevanz eher gering

Einfache Nullhypothese, zusammengesetzte Alternativhypothese $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$

- zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- ungerichtete Fragestellung nach einem Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \leq \mu_0$, $H_1 : \mu > \mu_0$

- einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- gerichtete Fragestellung nach einem positiven Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \geq \mu_0$, $H_1 : \mu < \mu_0$

- gerichtete Fragestellung nach einem negativen Unterschied
- qualitativ äquivalente Theorie zum umgekehrten Fall

Modellevaluation

Hypothesenszenarien

Einfache Nullhypothese, einfache Alternativhypothese $H_0 : \mu = \mu_0$, $H_1 : \mu = \mu_1$

- theoretisch wichtiges Szenario (Neymann-Pearson Lemma)
- praktische Relevanz eher gering

Einfache Nullhypothese, zusammengesetzte Alternativhypothese $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$

- zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- ungerichtete Fragestellung nach einem Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \leq \mu_0$, $H_1 : \mu > \mu_0$

- einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- gerichtete Fragestellung nach einem positiven Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \geq \mu_0$,$H_1 : \mu < \mu_0$

- gerichtete Fragestellung nach einem negativen Unterschied
- qualitativ äquivalente Theorie zum umgekehrten Fall

Gliederung (vgl. Einheit (12), um 01:58:00 in *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*)

(1) Statistisches Modell ✓

(2) Testhypothesen ✓

(3) Teststatistik

(4) Test

(5) Analyse der Testgütefunktion

(6) Testumfangkontrolle

(7) p-Wert

(8) Analyse der Powerfunktion

Modellevaluation (3) Teststatistik

Bezeichungskonventionen für t-Zufallsvariablen

- Für eine (nichtzentrale) t-Zufallsvariable ξ schreiben wir $\xi \sim t(n)$ (oder $\xi \sim t(\delta, n)$).
- Die WDF einer (nichtzentralen) t-Zufallsvariable ist $t(\cdot; n)$ (oder $t(\cdot; \delta, n)$).
- Die KVF einer (nichtzentralen) t-Zufallsvariable ist $\psi(\cdot; n)$ (oder $\psi(\cdot; \delta, n)$).
- Die Teststatistik eines T-Test-Designs/Hypothesentests bezeichnen wir mit T.
- Die Realisierung der T-Teststatistik für einen Datensatz bezeichnen wir mit t.

Theorem (T-Teststatistik des Einstichproben-T-Tests)

Gegeben sei die Designmatrixform des Einstichproben-T-Test-Modells. Dann ergibt sich für die T-Teststatistik mit

$$c := 1 \quad \text{und} \quad \beta_0 =: \mu_0, \tag{8}$$

dass

$$T = \sqrt{n}\left(\frac{\bar{y} - \mu_0}{s_y}\right) \tag{9}$$

und es gilt

$$T \sim t(\delta, n-1) \quad \text{mit} \quad \delta = \sqrt{n}\left(\frac{\mu - \mu_0}{\sigma}\right). \tag{10}$$

Bemerkungen

- Das Theorem basiert auf der T-Statistik im Rahmen des ALM.
- Wir erinnern an das verwandte populäre und von der Stichprobengröße unabhängige Effekstärke-Maß *Cohen's* d bei Einstichproben-T-Test-Designs,

$$d := \frac{\bar{y}}{s_y}. \tag{11}$$

- Offenbar gilt für *Cohen's* d, dass mit $\mu_0 := 0$

$$T = \sqrt{n}d \quad \Leftrightarrow \quad d = T/\sqrt{n}. \tag{12}$$

Modellevaluation (3) Teststatistik

Beweis

Mit dem Theorem zur Verteilung der T-Statistik (siehe Einheit (7) in *Allgemeines Lineares Modell*) gilt

$$
T = \frac{c^{\mathrm{T}}\hat{\beta} - c^{\mathrm{T}}\beta_0}{\sqrt{\hat{\sigma}^2 c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}c}} = \frac{1^{\mathrm{T}}\bar{y} - 1^{\mathrm{T}}\mu_0}{\sqrt{s_y^2 1^{\mathrm{T}}(1_n^{\mathrm{T}}1_n)^{-1}1}} = \sqrt{n}\left(\frac{\bar{y}-\mu_0}{s_y}\right). \tag{13}
$$

Weiterhin gilt mit demselben Theorem

$$
\delta = \frac{c^{\mathrm{T}}\beta - c^{\mathrm{T}}\beta_0}{\sqrt{\sigma^2 c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}c}} = \frac{1^{\mathrm{T}}\mu - 1^{\mathrm{T}}\mu_0}{\sqrt{\sigma^2 1^{\mathrm{T}}(1_n^{\mathrm{T}}1_n)^{-1}1}} = \sqrt{n}\left(\frac{\mu-\mu_0}{\sigma}\right). \tag{14}
$$

□

Definition (Zweiseitiger Einstichproben-T-Tests)

Gegeben sei das Einstichproben-T-Test-Modell. Für ein $\mu_0 \in \mathbb{R}$ seien die einfache Nullhypothese und die zusammengesetzte Alternativhypothese als

$$H_0 : \mu = \mu_0 \quad \text{und} \quad H_1 : \mu \neq \mu_0 \tag{15}$$

definiert. Weiterhin sei die T-Teststatistik definiert als

$$T := \sqrt{n}\left(\frac{\bar{y} - \mu_0}{s_y}\right) \tag{16}$$

Dann ist der *zweiseitige Einstichproben-T-Tests* definiert als der kritische-Wert-basierte Test

$$\phi(y) := 1_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases}. \tag{17}$$

Bemerkungen

- Ausführlicher handelt es sich um den *zweiseitigen Einstichproben-T-Test mit ungerichteter Hypothese*.

Theorem (Testgütefunktion)

ϕ sei der im obigen Testszenario definierte Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R} \to [0,1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k; \delta, n-1) + \psi(-k; \delta, n-1) \tag{18}$$

wobei $\psi(\cdot; \delta, n-1)$ die KVF der nichtzentralen t-Verteilung mit Nichtzentralitätsparameter

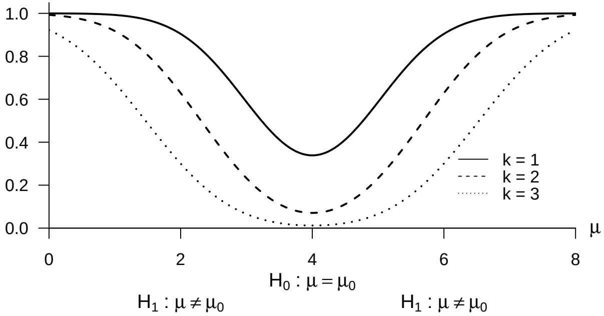
$$\delta := \sqrt{n}\left(\frac{\mu - \mu_0}{\sigma}\right) \tag{19}$$

und Freiheitsgradparameter $n-1$ bezeichnet.

Modellevaluation (5) Analyse der Testgütefunktion

Testgütefunktion q_ϕ für $\sigma^2 = 9, \mu_0 = 4, n = 12$ und $k = 1, 2, 3$.

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1)$$



Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testszenario ist definiert als

$$q_\phi : \mathbb{R} \to [0,1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \tag{20}$$

Da die Wahrscheinlichkeiten für $\phi = 1$ und dafür, dass die zugehörige Teststatistik im Ablehnungsbereich des Tests liegt, gleich sind, benötigen wir also zunächst die Verteilung der Teststatistik. Wir haben oben bereits gesehen, dass die T-Teststatistik

$$T := \sqrt{n}\left(\frac{\bar{y} - \mu_0}{s_y}\right) \tag{21}$$

unter der Annahme $y_i \sim N(\mu, \sigma^2)$ u.i.v. für $i = 1, ..., n$ nach einer nichtzentralen t-Verteilung $t(\delta, n-1)$ mit Nichtzentralitätsparameter

$$\delta = \sqrt{n}\left(\frac{\mu - \mu_0}{\sigma}\right) \tag{22}$$

verteilt ist. Der Ablehnungsbereich des zweiseitigen T-Tests ergibt sich, wie in ähnlicher Form bei der Betrachtung des zweiseitigen Z-Tests gesehen, zu

$$A =]-\infty, -k] \cup]k, \infty[. \tag{23}$$

Modellevaluation (5) Analyse der Testgütefunktion

Beweis (fortgeführt)

Mit diesem Ablehungsbereich ergibt sich dann

$$
\begin{aligned}
q_\phi(\mu) &= \mathbb{P}_\mu(\phi = 1) \\
&= \mathbb{P}_\mu\left(T \in]-\infty, -k] \cup]k, \infty[\right) \\
&= \mathbb{P}_\mu\left(T \in]-\infty, -k]\right) + \mathbb{P}_\mu\left(T \in [k, \infty[\right) \\
&= \mathbb{P}_\mu(T \le -k) + \mathbb{P}_\mu(T \ge k) \\
&= \mathbb{P}_\mu(T \le -k) + (1 - \mathbb{P}_\mu(T \le k)) \\
&= 1 - \mathbb{P}_\mu(T \le k) + \mathbb{P}_\mu(T \le -k) \\
&= 1 - \psi(k; \delta, n-1) + \psi(-k; \delta, n-1),
\end{aligned}
\tag{24}
$$

wobei $\psi(\cdot; \delta, n-1)$ die KVF der nichtzentralen T-Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparameter $n-1$ bezeichnet. □

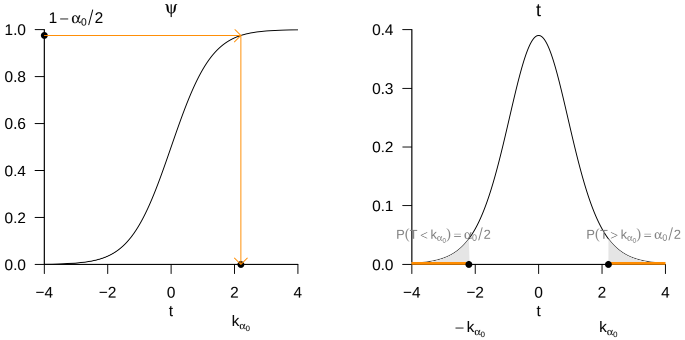
Theorem (Testumfangkontrolle)

ϕ sei der im obigen Testszenario definierte Test. Dann ist ϕ ein Level-α_0-Test mit Testumfang α_0, wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \psi^{-1}\left(1 - \frac{\alpha_0}{2}; n-1\right), \tag{25}$$

wobei $\psi^{-1}(\cdot; n-1)$ die inverse KVF der t-Verteilung mit $n-1$ Freiheitsgraden ist.

Wahl von $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n-1)$ mit $n = 12$, $\alpha_0 := 0.05$ und Ablehnungsbereich



Modellevaluation (6) Testumfangkontrolle

Beweis

Damit der betrachtete Test ein Level-α_0-Test ist, muss bekanntlich $q_\phi(\mu) \le \alpha_0$ für alle $\mu \in \{\mu_0\}$, also hier $q_\phi(\mu_0) \le \alpha_0$, gelten. Weiterhin ist der Testumfang des betrachteten Tests durch $\alpha = \max_{\mu \in \{\mu_0\}} q_\phi(\mu)$, also hier durch $\alpha = q_\phi(\mu_0)$ gegeben. Wir müssen also zeigen, dass die Wahl von k_{α_0} garantiert, dass ϕ ein Level-α_0-Test mit Testumfang α_0 ist. Dazu merken wir zunächst an, dass für $\mu = \mu_0$ gilt, dass

$$
\begin{aligned}
q_\phi(\mu_0) &= 1 - \psi(k; \delta, n-1) + \psi(-k; \delta, n-1) \\
&= 1 - \psi(k; 0, n-1) + \psi(-k; 0, n-1) \\
&= 1 - \psi(k; n-1) + \psi(-k; n-1),
\end{aligned}
\tag{26}
$$

wobei $\psi(\cdot; \delta, n-1)$ und $\psi(\cdot; n-1)$ die KVF der nichtzentralen t-Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparameter $n-1$ bzw. der t-Verteilung mit Freiheitsgradparameter $n-1$ bezeichnen. Mit $k := k_{\alpha_0}$ gilt dann

$$
\begin{aligned}
q_\phi(\mu_0) &= 1 - \psi(k_{\alpha_0}; n-1) + \psi(-k_{\alpha_0}; n-1) \\
&= 1 - \psi(k_{\alpha_0}; n-1) + (1 - \psi(k_{\alpha_0}; n-1)) \\
&= 2(1 - \psi(k_{\alpha_0}; n-1)) \\
&= 2\left(1 - \psi\left(\psi^{-1}\left(1 - \alpha_0/2; n-1\right); n-1\right)\right) \\
&= 2\left(1 - (1 - \alpha_0/2)\right) \\
&= \alpha_0,
\end{aligned}
\tag{27}
$$

wobei die zweite Gleichung mit der Symmetrie der t-Verteilung folgt. Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \le \alpha_0$ ist und der betrachtete Test somit ein Level-α_0-Test ist. Weiterhin folgt direkt, dass der Testumfang des betrachteten Tests bei der Wahl von $k = k_{\alpha_0}$ gleich α_0 ist. $\square$

Modellevaluation (6) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass ein Datensatz $y_1, ..., y_n$ eine Realisation von $y_i \sim N(\mu, \sigma^2)$ u.i.v. für $i = 1, ..., n$ mit unbekannten Parametern μ und $\sigma^2 > 0$ ist.
- Man möchte entscheiden ob für ein $\mu_0 \in \mathbb{R}$ eher $H_0 : \mu = \mu_0$ oder $H_1 : \mu \neq \mu_0$ zutrifft.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0}. Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$ und $n = 12$, dass $k_{0.05} = \psi^{-1}(1 - 0.05/2; 12 - 1) \approx 2.20$ ist.
- Anhand von $n, \mu_0, \bar{y}$ und s_y berechnet man die Realisierung der T-Teststatistik

$$t := \sqrt{n}\left(\frac{\bar{y} - \mu_0}{s_y}\right) \tag{28}$$

- Wenn t größer-gleich k_{α_0} ist oder wenn t kleiner-gleich $-k_{\alpha_0}$ ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab.
- Die oben entwickelte Theorie garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

Modellevaluation (7) p-Wert

Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0, bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.
- Bei $T = t$ würde H_0 für jedes α_0 mit $|t| \geq \psi^{-1}(1 - \alpha_0/2; n-1)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt,
$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \tag{29}$$
- Das kleinste $\alpha_0 \in [0,1]$ mit $\alpha_0 \geq 2\mathbb{P}(T \geq |t|)$ ist dann $\alpha_0 = 2\mathbb{P}(T \geq |t|)$, also folgt
$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n-1)). \tag{30}$$
- Im Gegensatz zum Z-Test hängt bei T-Tests der p-Wert auch von der Stichprobengröße ab.
- Zum Beispiel ist für $t = 2.00$ und $n = 10$ der p-Wert 0.076, für $t = 2.00$ und $n = 100$ ist der p-Wert dagegen 0.048.

Modellevaluation (7) p-Wert

Bestimmung des p-Wertes

- Es bleibt zu zeigen, dass gilt

$$|t| \ge \psi^{-1}(1-\alpha_0/2; n-1) \quad \Leftrightarrow \quad \alpha_0 \ge 2\mathbb{P}(T \ge |t|) \, . \tag{31}$$

- Dies aber folgt aus

$$\begin{aligned}
& |t| \ge \psi^{-1}\left(1-\frac{\alpha_0}{2}; n-1\right) \\
\Leftrightarrow\ & \psi(|t|; n-1) \ge \psi\left(\psi^{-1}\left(1-\frac{\alpha_0}{2}; n-1\right); n-1\right) \\
\Leftrightarrow\ & \psi(|t|; n-1) \ge 1-\frac{\alpha_0}{2} \\
\Leftrightarrow\ & \mathbb{P}(T \le |t|) \ge 1-\frac{\alpha_0}{2} \\
\Leftrightarrow\ & \frac{\alpha_0}{2} \ge 1-\mathbb{P}(T \le |t|) \\
\Leftrightarrow\ & \frac{\alpha_0}{2} \ge \mathbb{P}(T \ge |t|) \\
\Leftrightarrow\ & \alpha_0 \ge 2\mathbb{P}(T \ge |t|) \, .
\end{aligned} \tag{32}$$

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R} \to [0,1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k_{\alpha_0}; \delta, n-1) + \psi(-k_{\alpha_0}; \delta, n-1) \tag{33}$$

bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \psi^{-1}(1-\alpha_0/2; n-1)$ mit festem α_0 als Funktion des Nichtzentralitätsparameters und des Stichprobenumfangs. Namentlich hängt hier k_{α_0} sowohl von α_0 als auch von n ab.

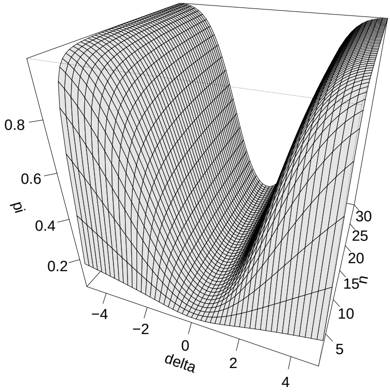
Es ergibt sich die bivariate reellwertige Funktion

$$\pi : \mathbb{R} \times \mathbb{N} \to [0,1], (\delta, n) \mapsto \pi(\delta, n) := 1 - \psi(k_{\alpha_0}; \delta, n-1) + \psi(-k_{\alpha_0}; \delta, n-1) \tag{34}$$

Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekannten Wert δ und von der Stichprobengröße n ab. Wir visualisieren diese Abhängigkeiten untenstehend.

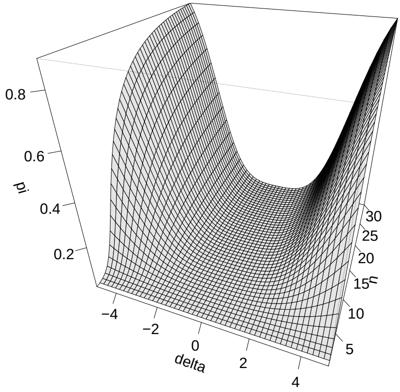
Modellevaluation (8) Analyse der Powerfunktion

Powerfunktion für $\alpha_0 = 0.05$



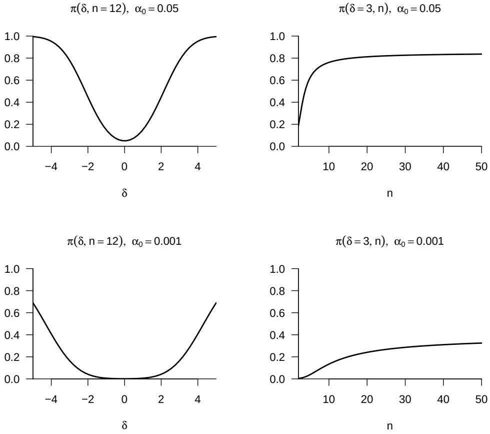
Modellevaluation (8) Analyse der Powerfunktion

Powerfunktion für $\alpha_0 = 0.001$



Modellevaluation (8) Analyse der Powerfunktion

Powerfunktionen für $\mu_0 = 0$



Modellevaluation (8) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem n steigt die Powerfunktion des Tests an:

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.

$\Rightarrow$ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

Die Powerfunktion hängt vom wahren, aber unbekannten Wert $\delta = \sqrt{n}\left(\frac{\mu-\mu_0}{\sigma}\right)$ ab.

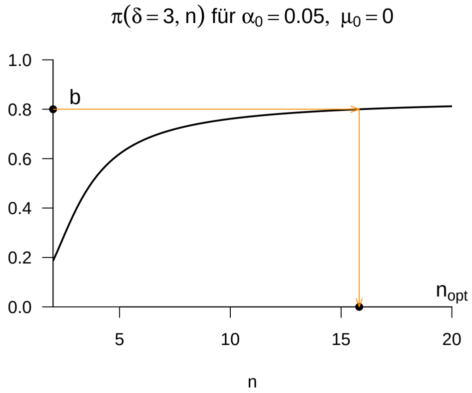
$\Rightarrow$ Wenn man δ schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert:

- Man legt das Signifikanzniveau α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert δ^*, den man mit $\pi(\delta, n) = b$ detektieren möchte.
- Ein konventioneller Wert ist $b = 0.8$.
- Man liest die für $\pi(\delta = \delta^*, n) = b$ nötige Stichprobengröße n ab.

Modellevaluation (8) Analyse der Powerfunktion

Praktisches Vorgehen



Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

Selbstkontrollfragen

Anwendungsbeispiel

Face-to-face PreBDI  PostBDI

Wir nehmen an, dass die Datenpunkte unabhängige und identisch verteilte (u.i.v.) Realisierungen von ZVen $y_i \sim N(\mu, \sigma^2)$ sind und nehmen weiter an, dass wir sind an der Quantifizierung der Unsicherheit beim inferentiellen Vergleich des wahren, aber unbekannten Erwartungswertparameters μ im Sinne eines Hypothesentests interessiert sind.

Anwendung/Praxis

Daten einlesen

```
fname        = "Daten/T-Tests_Daten.csv"
D            = read.table(fname, sep = ",", header = TRUE)
```

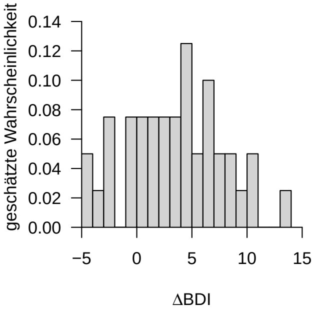
X	ID	Setting	PreBDI	PostBDI	dBDI
1	1	F2F	29	25	4
2	2	F2F	32	27	5
3	3	F2F	28	31	-3
4	4	F2F	36	22	14
5	5	F2F	32	29	3
6	6	F2F	28	28	0
7	7	F2F	33	30	3
8	8	F2F	33	26	7
9	9	F2F	33	28	5
10	10	F2F	30	28	2
11	11	F2F	36	25	11
12	12	F2F	32	31	1
13	13	F2F	29	31	-2
14	14	F2F	24	29	-5
15	15	F2F	35	32	3
16	16	F2F	31	29	2
17	17	F2F	31	23	8
18	18	F2F	34	25	9
19	19	F2F	34	23	11
20	20	F2F	33	26	7
21	21	F2F	34	25	9
22	22	F2F	33	27	6
23	23	F2F	31	24	7
24	24	F2F	25	27	-2
25	25	F2F	33	25	8
26	26	F2F	31	33	-2
27	27	F2F	31	29	2
28	28	F2F	26	30	-4
29	29	F2F	29	28	1
30	30	F2F	32	32	0
31	31	F2F	35	25	10
32	32	F2F	31	26	5
33	33	F2F	32	32	0
34	34	F2F	31	25	6
35	35	F2F	27	26	1
36	36	F2F	30	26	4
37	37	F2F	30	26	4
38	38	F2F	31	26	5
39	39	F2F	34	29	5
40	40	F2F	33	26	7

Anwendung/Praxis

```
# Datensatz von Interesse
BDI_F2F     = D$dBDI[D$Setting == "F2F"]        # BDI-Differenzwerte in der F2F-Gruppe

# Histogrammparameter
h           = 1                                 # gewünschte Klassenbreite
b_0         = min(BDI_F2F)                      # b_0
b_k         = max(BDI_F2F)                      # b_k
k           = ceiling((b_k - b_0)/h)            # Anzahl der Klassen
b           = seq(b_0, b_k, by = h)             # Klassen [b_0, ..., b_k]
ylimits     = c(0,0.15)                         # y-Achsenlimits
xlimits     = c(-5,15)                          # x-Achsenlimits

# Abbildungsparameter
par(                                            # für Details siehe ?par
    mfcol       = c(1,1),                       # 1 x 1 Panelstruktur
    family      = "sans",                       # Serif-freier Fonttyp
    pty         = "s",                          # quadratische Abbildungsregion
    bty         = "l",                          # L-förmige Box
    las         = 1,                            # horizontale Achsenbeschriftung
    xaxs        = "i",                          # x-Achse bei y = 0
    yaxs        = "i",                          # y-Achse bei x = 0
    font.main   = 1,                            # Titel nicht fett
    cex         = 1,                            # Textvergrößerungsfaktor
    cex.main    = 1)                            # Titeltextvergrößerungsfaktor

# Histogramm
hist(BDI_F2F,                                   # Delta-BDI_Werte von Therapiebedingung i
    breaks    = b,                              # Histogrammklassen
    freq      = F,                              # normierte relative Häufigkeit
    xlim      = xlimits,                        # x-Achsenlimits
    ylim      = ylimits,                        # y-Achsenlimits
    xlab      = TeX("$\\Delta BDI$"),           # x-Achsenbeschriftung
    ylab      = "geschätzte Wahrscheinlichkeit",# y-Achsenbeschriftung
    main      = "")                             # Titelbeschriftung

# Speichern
dev.copy2pdf(
    file      = "Abbildungen/F2F_histogramm.pdf",
    width     = 4,
    height    = 4)
```

```
# Einlesen der Daten
fname         = "Daten/T-Tests_Daten.csv"
D             = read.table(fname, sep = ",", header = TRUE)

# Initialisierung eines Dataframes
tp            = c("F2F")                              # Therapiebedingungen
ntp           = length(tp)                            # Anzahl Therapiebedingungen
S             = data.frame(                           # Dataframe-Erzeugung
                    n         = rep(NaN,ntp),         # Stichprobengrößen
                    Max       = rep(NaN,ntp),         # Maxima
                    Min       = rep(NaN,ntp),         # Minima
                    Median    = rep(NaN,ntp),         # Mediane
                    Mean      = rep(NaN,ntp),         # Mittelwerte
                    Var       = rep(NaN,ntp),         # Varianzen
                    Std       = rep(NaN,ntp),         # Standardabweichungen
                    row.names = tp)                   # Therapiebedingungen

# Iterationen über Therapiebedingungen
for(i in 1:ntp){
    data        = D$dBDI[D$Setting == tp[i]]          # Daten
    S$n[i]      = length(data)                        # Stichprobengröße
    S$Max[i]    = max(data)                           # Maxima
    S$Min[i]    = min(data)                           # Minima
    S$Median[i] = median(data)                        # Mediane
    S$Mean[i]   = mean(data)                          # Mittelwerte
    S$Var[i]    = var(data)                           # Varianzen
    S$Std[i]    = sd(data)                            # Standardabweichungen
}
```

Deskriptive Statistiken der negativen PostBDI-PreBDI-Differenzen bei Face-to-Face-Therapie




```
# Ausgabe
print.AsIs(S)
```

```
     n Max Min Median  Mean      Var      Std
F2F 40  14  -5      4 3.925 19.40449 4.405052
```

Anwendung/Praxis

```
# Daten einlesen
fname       = "Daten/T-Tests_Daten.csv"                              # Dateiname
D           = read.table(fname, sep = ",", header = TRUE)            # Dataframe
y           = D$dBDI[D$Setting == "F2F"]                             # BDI-Differenzwerte in der F2F-Gruppe

# Modellformulierung
n           = length(y)                                              # Anzahl Datenpunkte
p           = 1                                                      # Anzahl Betaparameter
X           = matrix(rep(1,n), nrow = n)                             # n x p Designmatrix

# Parameterschätzung
beta_hat    = solve(t(X) %*% X) %*% t(X) %*% y                       # Betaparameterschätzer
eps_hat     = y - X %*% beta_hat                                     # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) /(n-p)                         # Varianzparameterschätzer

# Konfidenzintervall
delta       = 0.95                                                   # Konfidenzbedingung
t_delta     = qt((1+delta)/2,n-1)                                    # \Psi^{-1}((1+\delta)/2; n-1)
lambda      = diag(solve(t(X) %*% X))                                # \lambda_j Werte
kappa       = matrix(rep(NaN,p*2), nrow = p)                         # \beta_j Konfidenintervall-Array
for(j in 1:p){                                                       # Iteration über \beta_j
    kappa[j,1] = beta_hat[j]-sqrt(sigsqr_hat*lambda[j])*t_delta      # untere KI-Grenze
    kappa[j,2] = beta_hat[j]+sqrt(sigsqr_hat*lambda[j])*t_delta      # obere  KI-Grenze
}

# Hypothesentest
c           = matrix(c(1),nrow = p)                                  # Kontrastgewichtsvektor
mu_0        = 0                                                      # Nullhypothese
alpha_0     = 0.05                                                   # Signifikanzniveau
k_alpha_0   = qt(1 - (alpha_0/2), n-1)                               # kritischer Wert
t_num       = t(c) %*% beta_hat - mu_0                               # T-Teststatistik-Zähler
t_den       = sqrt(sigsqr_hat %*% t(c)*solve(t(X) %*% X)%*%c)        # T-Teststatistik-Nenner
t           = t_num/t_den                                            # T-Teststatistik
if(abs(t) >= k_alpha_0){phi = 1} else {phi = 0}                      # Test 1_{|T(X) >= k_alpha_0|}
pval        = 2*(1 - pt(abs(t), n-1))                                # p-Wert
d           = t/sqrt(n)                                              # Cohen's d
```

```
fg         =  39
kappa_1    =  2.516196 5.333804
t          =  5.63532
alpha_0    =  0.05
k_alpha_0 =  2.022691
phi        =  1
p-Wert     =  1.662163e-06
Cohen's d =  0.8910223
```

Anwendung/Praxis

Anwendungszenario

```
# automatischer Einstichproben-T-Test
varphi = t.test(y,                                # Datensatz
             alternative = c("two.sided"),        # H_1: \mu \neq \mu_0
             mu          = 0,                     # \mu_0 (sic!)
             conf.level  = 1-alpha_0)             # \delta = 1 - \alpha_0

# Ausgabe
print(varphi)
```

```

    One Sample t-test

data:  y
t = 5.6353, df = 39, p-value = 1.662e-06
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 2.516196 5.333804
sample estimates:
mean of x
    3.925
```

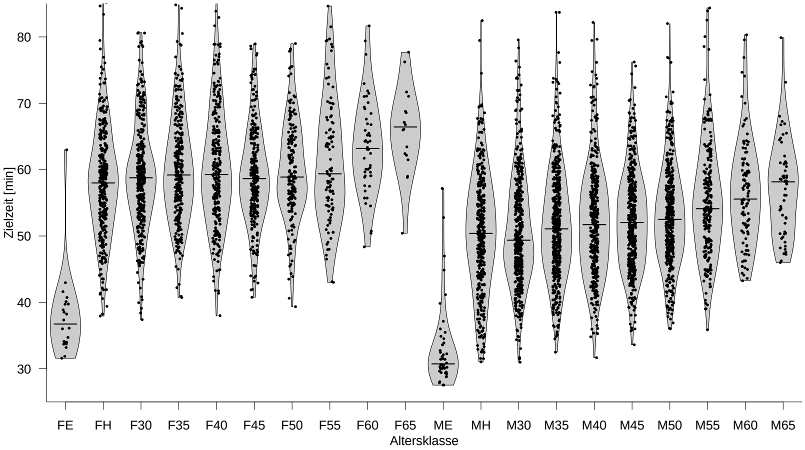
```
# genauere Ausgabe von t
paste(varphi[1]$statistic)
```

```
[1] "5.63531986397201"
```

```
# genauere Ausgabe von p
paste(varphi[3]$p.value)
```

```
[1] "1.66216308541e-06"
```

Datensatz: Zielzeiten beim Great 10k (08.10.2017), getrennt nach Altersklasse ($n = 4883$)

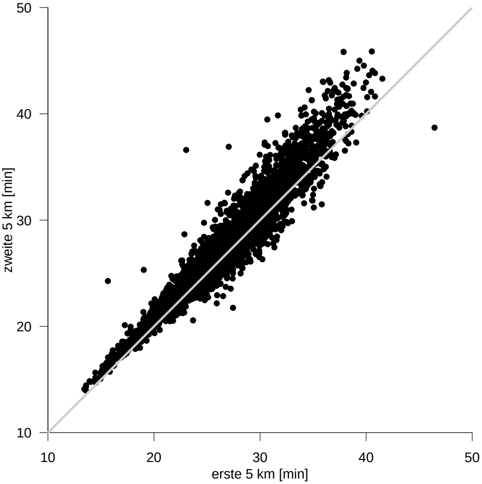


Einstichproben-T-Test: 10-km-Zeit, alle Teilnehmer ($n = 4883$)

```
# Daten extrahieren
D$T10k = t10_min                                                  # 10-km-Zeit [min]

# Einstichproben-T-Test (falsch)
n      = nrow(D)                                                  # Anzahl Datenpunkte
p      = 1                                                        # Anzahl Regressoren
y      = matrix(D$T10k, nrow = n)                                 # Datenvektor
X      = matrix(rep(1,n), ncol = p)                               # Designmatrix
```

```
Betaparameterschätzer          :  54.901
Varianzparameterschätzer       :  86.6
Einstichproben-T-Teststatistik :  412.257
Cohen's  d                     :  5.9
p-Wert                         :  0
```

Datensatz: erste vs. zweite 5 km beim Great 10k (08.10.2017), alle Teilnehmer ($n = 4883$)



Anwendung/Praxis

Einstichproben-T-Test: Differenz zweite minus erste 5 km, alle Teilnehmer ($n = 4883$)

```
# Daten extrahieren
D$T5k1 = t5_min                                                    # erste  5 km [min]
D$T5k2 = t10_min-t5_min                                            # zweite 5 km [min]
D$Tdiff= D$T5k2 - D$T5k1                                           # Differenz [min]

# Einstichproben-T-Test (richtig)
n       = nrow(D)                                                  # Anzahl Datenpunkte
p       = 1                                                        # Anzahl Regressoren
y       = matrix(D$Tdiff, nrow = n)                                # Datenvektor
X       = matrix(rep(1,n), ncol = p)                               # Designmatrix
```

```
Betaparameterschätzer            :  0.708
Varianzparameterschätzer         :  2.208
Einstichproben-T-Teststatistik :  33.28
Cohen's d                        :  0.476
p-Wert                           :  0
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie die Extremszenarien im Kontinuum von ALM-Designs.
2. Erläutern Sie die Begriffe der faktoriellen und parametrischen ALM-Designs.
3. Nennen Sie Beispiele für faktorielle, parametrische und faktoriell-parametrische ALM-Designs.
4. Erläutern Sie das Anwendungsszenario eines Einstichproben-T-Tests.
5. Erläutern Sie mögliche Hypothesenszenarien eines Einstichproben-T-Tests.
6. Geben Sie die Definition des Einstichproben-T-Test-Modells wieder.
7. Geben Sie das Theorem zur Parameterschätzung im Einstichproben-T-Test-Modell wieder.
8. Geben Sie das Theorem zur T-Teststatistik des Einstichproben-T-Tests wieder.
9. Geben Sie die Definition des zweiseitigen Einstichproben-T-Tests (mit ungerichteter Hypothese) wieder.
10. Skizzieren Sie die Testgütefunktion des zweiseitigen Einstichproben-T-Tests.
11. Geben Sie das Theorem zur Testumfangkontrolle im zweiseitigen Level-α_0-Einstichproben-T-Test wieder.
12. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level-α_0-Einstichproben-T-Tests.
13. Geben Sie die Definition des p-Wertes Werts für einen zweiseitigen Einstichproben-T-Test wieder.
14. Von welchen Werten hängt die Powerfunktion eines zweiseitigen Einstichproben-T-Tests ab?
15. Skizzieren Sie die Powerfunktion des Einstichproben-T-Tests bei fester Stichprobengröße.
16. Skizzieren Sie die Powerfunktion des Einstichproben-T-Tests bei festem Nichtzentralitätsparameter.