



Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(9) T-Tests

Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

Kontinuum von ALM Designs

Extremszenario (1) Die Erwartungswerte aller Datenvariablen sind identisch.

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

\Rightarrow Jegliche Datenvariabilität wird dem Fehlerterm zugeschrieben.

Extremszenario (2) Die Erwartungswerte aller Datenvariablen sind paarweise verschieden

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

\Rightarrow Jegliche Datenvariabilität wird dem Erwartungswertparameter zugeschrieben.

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

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

Faktorielle und Parametrische ALM Designs

Faktorielle ALM Designs

- Designmatrizen mit 1en und 0en, manchmal -1 en.
- Betaparameter repräsentieren Gruppenerwartungswerte.
- Betaparameterschätzer repräsentieren Gruppenstichprobenmittel.
- \Rightarrow T-Tests, Einfaktorielle Varianzanalyse, Mehrfaktorielle Varianzanalyse

Parametrische ALM Designs

- Designmatrizen besitzen Spalten mit kontinuierlichen reellen Werten.
- Die Designmatrixspalten werden *Regressoren*, *Prädiktoren*, oder *Kovariaten* genannt.
- Betaparameter repräsentieren Steigungsparameter.
- Betaparameterschätzer ergeben sich als normalisierte Regressor-Daten Kovarianzen.
- 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

ALM Designs als Hypothesentestverfahren*

Testen von Unterschiedshypothesen

- T-Tests
- Einfaktorielle Varianzanalyse
- Mehrfaktorielle Varianzanalyse
- Kovarianzanalyse

Testen von Zusammenhangshypothesen

- Einfache lineare Regression/Korrelation
- Multiple lineare Regression/Multiple Korrelation

*Diese Sichtweise durch den Lehrenden nicht favorisiert.

T-Tests

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

Wir fokussieren hier auf die Erkenntnis von T-Tests als Spezialfälle 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.

Wir behandeln nicht

- T-Tests mit gerichteten Hypothesen oder einfachen Null- und Alternativhypothesen.
- Zweistichproben-T-Tests bei Annahme nicht identischer Varianzen (Behrend's-Fischer Problem).
- Zweistichproben T-Tests bei abhängigen Stichproben.

Für dieses Themengebiet wird auf WTFI Einheiten (13) und (14) verwiesen.

Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

Einstichproben-T-Tests

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Eine Gruppe (Stichprobe) randomisierter experimenteller Einheiten.

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

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

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

Gruppenanalysen mit Wechsler Adult Intelligence Scale

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

Gruppenanalysen in der funktionellen Kernspintomographie

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

Wir betrachten das Anwendungsbeispiel aus Einheit (8) Studiendesign und fokussieren auf die Gruppe (= Stichprobe, Experimentalbedingung) der Face-to-Face Therapie. Wir betrachten also den Datensatz der negativen PostBDI-PreBDI Differenzwerte mit Variablennamen "BDI."

Wir nehmen an, dass diese Datenpunkte 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.

Im Folgenden evaluieren diesen Datensatz zunächst im Sinne deskriptiver Statistiken, siehe dazu Einheit (11) Anwendungsbeispiel in Programmierung und Deskriptive Statistik.

Anwendungsszenario

Dateneinlesen

```
fname = file.path(getwd(), "9_Daten", "data_9_t_tests.csv")
D = read.table(fname, sep = ",", header = TRUE)
```

X	ID	Condition	PreBDI	PostBDI	BDI	Age	Duration
1	1	F2F	29	25	4	66	22
2	2	F2F	32	31	1	29	23
3	3	F2F	28	26	2	71	14
4	4	F2F	36	26	10	77	21
5	5	F2F	32	27	5	55	24
6	6	F2F	28	29	-1	50	24
7	7	F2F	33	27	6	31	16
8	8	F2F	33	27	6	20	17
9	9	F2F	33	25	8	73	23
10	10	F2F	30	26	4	28	13
11	11	F2F	36	27	9	21	23
12	12	F2F	32	25	7	76	16
13	13	F2F	29	29	0	38	18
14	14	F2F	24	22	2	30	18
15	15	F2F	35	28	7	44	21
16	16	F2F	31	22	9	48	17
17	17	F2F	31	26	5	46	14
18	18	F2F	34	25	9	51	16
19	19	F2F	34	25	9	71	20
20	20	F2F	33	27	6	23	13
21	21	F2F	34	21	13	53	22
22	22	F2F	33	31	2	61	15
23	23	F2F	31	22	9	59	18
24	24	F2F	25	26	-1	60	14
25	25	F2F	33	23	10	48	16
26	26	F2F	31	25	6	78	19
27	27	F2F	31	34	-3	44	15
28	28	F2F	26	27	-1	71	20
29	29	F2F	29	23	6	65	18
30	30	F2F	32	22	10	52	23
31	31	F2F	35	28	7	72	22
32	32	F2F	31	27	4	48	16
33	33	F2F	32	26	6	20	16
34	34	F2F	31	24	7	64	22
35	35	F2F	27	22	5	63	17
36	36	F2F	30	24	6	31	16
37	37	F2F	30	30	0	59	13
38	38	F2F	31	25	6	53	13
39	39	F2F	34	23	11	40	20
40	40	F2F	33	33	0	58	20

Anwendungsszenario

Histogramm

```
# Datensatz von Interesse
BDI_F2F      = D$BDI[D$Condition == "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_0
k           = ceiling((b_k - b_0)/h)      # Anzahl der Klassen
b           = seq(b_0, b_k, by = h)        # Klassen [b_{j-1}, b_j[
ylimits     = c(0,0.25)                   # 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,                         # Non-Bold Titel
  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     = "BDI",                           # x-Achsenbeschriftung
  ylab     = "Geschätzte Wahrscheinlichkeit", # y-Achsenbeschriftung
  main     = "")                             # Titelbeschriftung

# PDF Speicherung
dev.copy2pdf(
  file      = file.path(getwd(), "9_Abbildungen", "alm_9_F2F_histogramm.pdf"),
  width     = 4,
  height    = 4)
```


Anwendungsszenario

Deskriptive Statistiken

```
# Initialisierung eines Dataframes
tp      = c("F2F")
ntp     = length(tp)
S       = data.frame(
  n      = rep(NaN,ntp),
  Max    = rep(NaN,ntp),
  Min    = rep(NaN,ntp),
  Median = rep(NaN,ntp),
  Mean   = rep(NaN,ntp),
  Var    = rep(NaN,ntp),
  Std    = rep(NaN,ntp),
  row.names = tp)

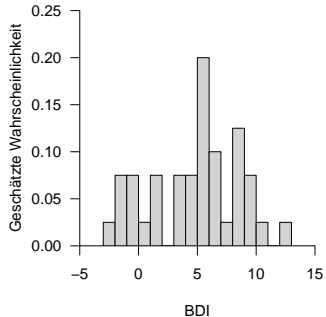
# Therapiebedingungen
# Anzahl Therapiebedingungen
# Dataframeerzeugung
# Stichprobengrößen
# Maxima
# Minima
# Mediane
# Mittelwerte
# Varianzen
# Standardabweichungen
# Therapiebedingungen

# Iterationen über Therapiebedingungen
for(i in 1:ntp){
  data = D$BDI[D$Condition == tp[i]]
  print(data)
  S$n[i] = length(data)
  S$Max[i] = max(data)
  S$Min[i] = min(data)
  S$Median[i] = median(data)
  S$Mean[i] = mean(data)
  S$Var[i] = var(data)
  S$Std[i] = sd(data)
}

# Daten
# Stichprobengröße
# Maxima
# Minima
# Mediane
# Mittelwerte
# Varianzen
# Standardabweichungen

> [1] 4 1 2 10 5 -1 6 6 8 4 9 7 0 2 7 9 5 9 9 6 13
> [22] 2 9 -1 10 6 -3 -1 6 10 7 4 6 7 5 6 0 6 11 0
```

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 13  -3      6 5.28 14.8 3.85
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Definition (Einstichproben-T-Test Modell)

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

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

die Datenverteilungsform

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

und für den Datenvektor $y = (y_1, \dots, y_n)^T$ die Designmatrixform

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

Bemerkungen

- Das Modell ist identisch mit dem Modell unabhängiger und identisch normalverteilter Zufallsvariablen.
- Es ist $p = 1$.
- Die Äquivalenz der drei Modellformen wurde in Einheit (5) Modellformulierung ausführlich diskutiert.

Datensimulation (vgl. Einheit (5) Modellformulierung)

```
# Libraries
library(MASS)                                # Multivariate Normalverteilung

# Modellformulierung
n      = 40                                  # Anzahl von Datenpunkten
p      = 1                                  # Anzahl von Betaparameter
X      = matrix(rep(1,n), nrow = n)         # 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 eines n-dimensionalen ZVs
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Theorem (Parameterschätzung im Einstichproben-T-Test Modell)

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

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n y_i =: \bar{y}, \quad (6)$$

und für den Varianzparameterschätzer

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

Bemerkungen

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

Modellschätzung

```
# Dateneinlesen
fname      = file.path(getwd(), "9_Daten", "data_9_t_tests.csv") # Dateiname
D          = read.table(fname, sep = ",", header = TRUE)         # Dataframe
y          = D$BDI[D$Condition == "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}   :  5.28
> bar{y}      :  5.28
> hat{sigsqr} : 14.8
> s_y^2       : 14.8
```


Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Überblick

- Wir gruppieren frequentistische Konfidenzintervalle und Hypothesentests unter Modellevaluation.
- Zu Konfidenzintervallen im Szenario u.i.n.v. ZVen siehe WTFI Einheit (11) Konfidenzintervalle.
- In der Praxis zielt die Evaluation von Einstichproben-T-Tests ALM Designs meist auf einen Hypothesentest.
- Die Theorie des Einstichproben-T-Tests ist umfangreich, siehe dazu WTFI (13) Einstichproben T-Tests.
- Wir fokussieren hier auf zweiseitige Einstichproben-T-Tests mit ungerichteter Hypothese.
- Ein gutes Verständnis von WTFI Einheit (12) Hypothesentests wird im Folgenden vorausgesetzt.

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

Hypothesenszenarien

Im Folgenden näher betrachtetes Hypothesenszenario

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. WTFI Einheiten (12) - (14))

- (1) Statistisches Modell ✓
- (2) Testhypothesen ✓
- (3) Teststatistik
- (4) Test
- (5) Analyse der Testgütefunktion
- (6) Testumfangkontrolle
- (7) p-Werte
- (8) Analyse der Powerfunktion

(3) Teststatistik

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

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

$$c := 1 \text{ und } c^T \beta_0 =: \mu_0, \quad (8)$$

dass

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

und es gilt

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

Bemerkungen

- Das Theorem basiert auf dem T-Teststatistik Theorem in Einheit (7) Modellevaluation.
- Die eng verwandte T-Statistik wurde bereits in Einheit (7) Modellevaluation hergeleitet.

(3) Teststatistik

Beweis

Mit dem T-Teststatistik Theorem in Einheit (7) Modellevaluation gilt

$$T = \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} = \frac{1^T \bar{y} - 1^T \mu_0}{\sqrt{s_y^2 1^T (1_n^T 1_n)^{-1} 1}} = \sqrt{n} \left(\frac{\bar{y} - \mu_0}{s_y} \right). \quad (11)$$

Weiterhin gilt mit demselben Theorem

$$\delta = \frac{c^T \beta - c^T \beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}} = \frac{1^T \mu - 1^T \mu_0}{\sqrt{\sigma^2 1^T (1_n^T 1_n)^{-1} 1}} = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (12)$$

(4) Test

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

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \quad (13)$$

definiert. Weiterhin sei die T-Teststatistik definiert als

$$T := \sqrt{n} \left(\frac{\bar{y} - \mu_0}{s_y} \right) \quad (14)$$

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

$$\phi(y) := 1_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases}. \quad (15)$$

Bemerkungen

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

(5) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

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

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k; d_\mu, n - 1) + \psi(-k; d_\mu, n - 1) \quad (16)$$

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

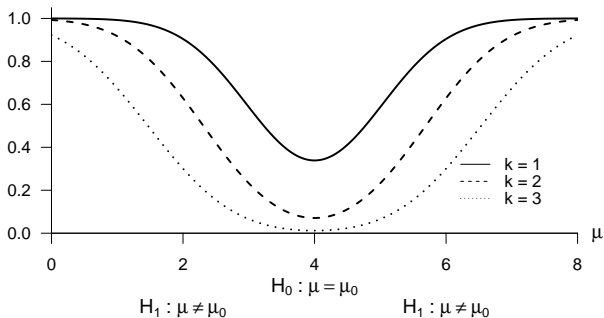
$$d_\mu := \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (17)$$

und Freiheitsgradparameter $n - 1$ bezeichnet.

(5) Analyse der Testgütefunktion

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

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



(5) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testszenario ist definiert als

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \quad (18)$$

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

$$T := \sqrt{n} \left(\frac{\bar{y} - \mu_0}{s_y} \right) \quad (19)$$

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

$$d = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (20)$$

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

$$A =] - \infty, -k] \cup]k, \infty[. \quad (21)$$

(5) Analyse der Testgütefunktion

Beweis (fortgeführt)

Mit diesem Ablehungsbereich ergibt sich dann

$$\begin{aligned}q_{\phi}(\mu) &= \mathbb{P}_{\mu}(\phi = 1) \\&= \mathbb{P}_{\mu}(T \in]-\infty, -k] \cup]k, \infty[) \\&= \mathbb{P}_{\mu}(T \in]-\infty, -k]) + \mathbb{P}_{\mu}(T \in [k, \infty[) \\&= \mathbb{P}_{\mu}(T \leq -k) + \mathbb{P}_{\mu}(T \geq k) \\&= \mathbb{P}_{\mu}(T \leq -k) + (1 - \mathbb{P}_{\mu}(T \leq k)) \\&= 1 - \mathbb{P}_{\mu}(T \leq k) + \mathbb{P}_{\mu}(T \leq -k) \\&= 1 - \psi(k; d_{\mu}, n - 1) + \psi(-k; d_{\mu}, n - 1),\end{aligned}\tag{22}$$

wobei $\psi(\cdot; d_{\mu}, n - 1)$ die KVF der nichtzentralen T-Verteilung mit Nichtzentralitätsparameter d_{μ} und Freiheitsgradparameter $n - 1$ bezeichnet.

□

(6) Testumfangkontrolle

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

$$k_{\alpha_0} := \psi^{-1} \left(1 - \frac{\alpha_0}{2}; n - 1 \right), \quad (23)$$

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

(6) Testumfangkontrolle

Beweis

Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich $q_\phi(\mu) \leq \alpha_0$ für alle $\mu \in \{\mu_0\}$, also hier $q_\phi(\mu_0) \leq \alpha_0$, 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

$$\begin{aligned} q_\phi(\mu_0) &= 1 - \psi(k; d_{\mu_0}, n-1) + \psi(-k; d_{\mu_0}, n-1) \\ &= 1 - \psi(k; 0, n-1) + \psi(-k; 0, n-1) \\ &= 1 - \psi(k; n-1) + \psi(-k; n-1), \end{aligned} \tag{24}$$

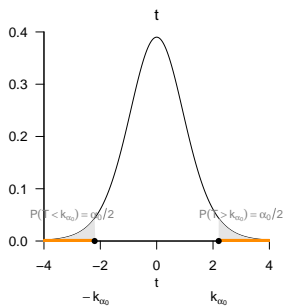
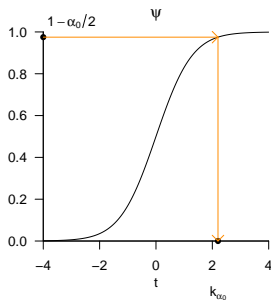
wobei $\psi(\cdot; d, n-1)$ und $\psi(\cdot; n-1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter d und Freiheitsgradparameter $n-1$ sowie der t -Verteilung mit Freiheitsgradparameter $n-1$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

$$\begin{aligned} q_\phi(\mu_0) &= 1 - \psi(k_{\alpha_0}, n-1) + \psi(-k_{\alpha_0}, n-1) \\ &= 1 - \psi(k_{\alpha_0}, n-1) + (1 - \psi(k_{\alpha_0}, n-1)) \\ &= 2(1 - \psi(k_{\alpha_0}, n-1)) \\ &= 2 \left(1 - \psi \left(\psi^{-1}(1 - \alpha_0/2, n-1), n-1 \right) \right) \\ &= 2(1 - 1 + \alpha_0/2) \\ &= \alpha_0, \end{aligned} \tag{25}$$

wobei die zweite Gleichung mit der Symmetrie der t -Verteilung folgt. Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \leq \alpha_0$ 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.

(6) Testumfangkontrolle

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



(6) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass ein Datensatz v_1, \dots, v_n eine Realisation von $y_i \sim N(\mu, \sigma^2)$ u.i.v. für $i = 1, \dots, n$ 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 $n = 12$, also Freiheitsgradparameter 11, dass $k_{0.05} = \psi^{-1}(1 - 0.05/2; 11) \approx 2.20$ ist.
- Anhand von n, μ_0, \bar{v} und s_v berechnet man die Realisierung der T-Teststatistik

$$t := \sqrt{n} \left(\frac{\bar{v} - \mu_0}{s_v} \right) \quad (26)$$

- 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.

(7) p-Werte

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,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (27)$$

- 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

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n - 1)). \quad (28)$$

- 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.

(7) p-Werte

Bestimmung des p-Wertes

- Es bleibt zu zeigen, dass gilt

$$|t| \geq \psi^{-1}(1 - \alpha_0/2; n - 1) \Leftrightarrow \alpha_0 \geq 2\mathbb{P}(T \geq |t|) \quad (29)$$

- Dies aber folgt aus

$$\begin{aligned} |t| &\geq \psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right) \\ \Leftrightarrow \psi(|t|; n - 1) &\geq \psi\left(\psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right); n - 1\right) \\ \Leftrightarrow \psi(|t|; n - 1) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \mathbb{P}(T \leq |t|) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq 1 - \mathbb{P}(T \leq |t|) \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq \mathbb{P}(T \geq |t|) \\ \Leftrightarrow \alpha_0 &\geq 2\mathbb{P}(T \geq |t|). \end{aligned} \quad (30)$$

Modellevaluation

Anwendungsszenario

```
# Dateneinlesen
fname      = file.path(getwd(), "9_Daten", "data_9_t_tests.csv") # Dateiname
D          = read.table(fname, sep = ",", header = TRUE)         # Dataframe
y          = D$BDI[D$Condition == "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

# Modellevaluation
c          = matrix(c(1), nrow = p)                            # Kontrastgewichtsvektor
mu_0       = 0                                                  # Nullhypothese H_0
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}
# Ablehnen von H_0
# Nicht Ablehnen von H_0
pval       = 2*(1 - pt(abs(t), n-1))                           # p-Wert

> fg      = 39
> t       = 8.67
> alpha_0 = 0.05
> k_alpha_0 = 2.02
> phi     = 1
> p-Wert  = 1.25e-10
```

Modellevaluation

Anwendungsszenario

```
# Automatischer Einstichproben-T-Test
varphi = t.test(
  y,
  alternative = c("two.sided"),
  mu = 0,
  conf.level = 1-alpha_0

# ?t.test für Details
# Datensatz
# H_1: \mu \neq \mu_0
# \mu_0 (sic!)
# \delta = 1 - \alpha_0 (sic!)

# Ausgabe
print(varphi)
```

```
>
> One Sample t-test
>
> data: y
> t = 9, df = 39, p-value = 1e-10
> alternative hypothesis: true mean is not equal to 0
> 95 percent confidence interval:
>  4.04 6.51
> sample estimates:
> mean of x
>      5.28
```

```
# Genauere Ausgabe t
paste(varphi[1])
```

```
> [1] "c(t = 8.66623050649246)"
# Genauere Ausgabe p
paste(varphi[3])
```

```
> [1] "1.25177126014618e-10"
```

(8) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k_{\alpha_0}; d_\mu, n - 1) + \psi(-k_{\alpha_0}; d_\mu, n - 1) \quad (31)$$

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} auch von n ab.

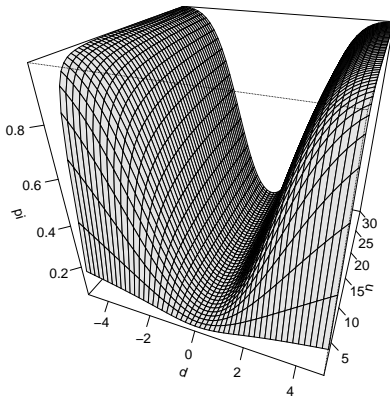
Es ergibt sich die bivariate reellwertige Funktion

$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (d, n) \mapsto \pi(d, n) := 1 - \psi(k_{\alpha_0}; d, n - 1) + \psi(-k_{\alpha_0}; d, n - 1) \quad (32)$$

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

(8) Analyse der Powerfunktion

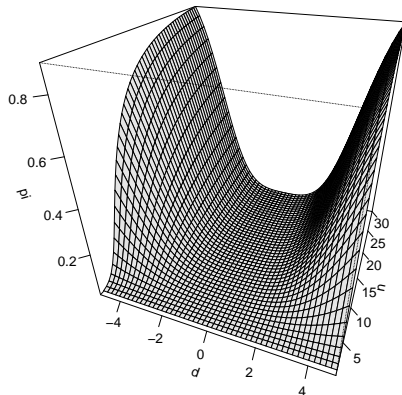
Powerfunktion für $\alpha_0 = 0.05$



Zweiseitige Einstichproben-T-Tests

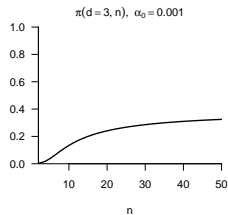
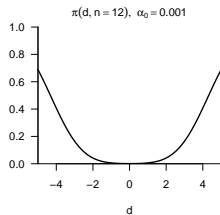
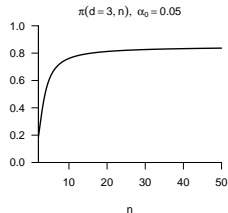
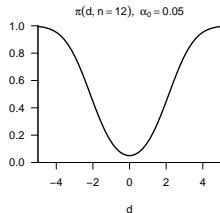
(8) Analyse der Powerfunktion

Powerfunktion für $\alpha_0 = 0.001$



(8) Analyse der Powerfunktion

Powerfunktionen für $\mu_0 = 0$



(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.

⇒ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

Die Powerfunktion hängt vom wahren, aber unbekannten, Parameterwert $d = \sqrt{n}(\mu - \mu_0)/\sigma$ ab.

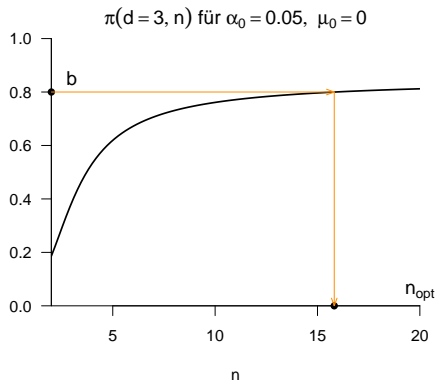
⇒ Wenn man d schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert

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

(8) Analyse der Powerfunktion

Praktisches Vorgehen



Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

Zweistichproben-T-Tests

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Zwei Gruppen (Stichproben) randomisierter experimenteller Einheiten.

Annahme unabhängiger identischer Normalverteilungen $N(\mu_1, \sigma^2)$ und $N(\mu_2, \sigma^2)$.

μ_1, μ_2 und σ^2 unbekannt.

Annahme eines identischen Varianzparameters für beide Gruppen.

Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ_1 mit μ_2 beabsichtigt.

Anwendungsbeispiele

BDI Differenzwert Datenanalyse bei zwei Gruppen von Patient:innen

- Gruppe 1 Face-to-Face Therapie, Gruppe 2 Online Therapie
- $\mu_1 \neq \mu_2 \Leftrightarrow$ Unterscheiden sich die Therapiewirksamkeiten?

Forcierte Schwimmtestdatenanalyse bei zwei Gruppen genmanipulierter Mäuse

- Gruppe 1 Wildtyp, Gruppe 2 Serotoninrezeptormutation
- $\mu_1 \neq \mu_2 \Leftrightarrow$ Trägt Serotoninrezeptor zum Schwimmtestverhalten bei?

Wir betrachten das Anwendungsbeispiel aus Einheit (8) Studiendesign und fokussieren auf die Gruppen (= Stichproben, Experimentalbedingungen) der Face-to-Face und der Online Therapie. Wir betrachten dabei wieder den Datensatz der negativen PostBDI-PreBDI Differenzwerte mit Variablennamen "BDI."

Wir nehmen an, dass die Datenpunkte der Face-to-Face Therapiegruppe u.i.v. Realisierungen von ZVen $y_{1j} \sim N(\mu_1, \sigma^2)$ für $j = 1, \dots, 40$ und dass die Datenpunkte der Online Therapiegruppe u.i.v. Realisierungen von ZVen $y_{2j} \sim N(\mu_2, \sigma^2)$ für $j = 1, \dots, 40$ sind. Wir nehmen weiter an, dass wir an der Quantifizierung der Unsicherheit beim inferentiellen Vergleich der wahren, aber unbekannt, Erwartungswertparameter μ_1 und μ_2 im Sinne eines Hypothesentests interessiert sind.

Im Folgenden evaluieren den hier betrachteten Datensatz zunächst im Sinne deskriptiver Statistiken, siehe dazu auch Einheit (11) Anwendungsbeispiel in Programmierung und Deskriptive Statistik.

Anwendungsszenario

Dateneinlesen | $j = 1, \dots, 20$ für jede Gruppe

```
fname = file.path(getwd(), "9_Daten", "data_9_t_tests.csv")
D = read.table(fname, sep = ",", header = TRUE)
```

	X	ID	Condition	PreBDI	PostBDI	BDI	Age	Duration
1	1	1	F2F	29	25	4	66	22
2	2	2	F2F	32	31	1	29	23
3	3	3	F2F	28	26	2	71	14
4	4	4	F2F	36	26	10	77	21
5	5	5	F2F	32	27	5	55	24
6	6	6	F2F	28	29	-1	50	24
7	7	7	F2F	33	27	6	31	16
8	8	8	F2F	33	27	6	20	17
9	9	9	F2F	33	25	8	73	23
10	10	10	F2F	30	26	4	28	13
11	11	11	F2F	36	27	9	21	23
12	12	12	F2F	32	25	7	76	16
13	13	13	F2F	29	29	0	38	18
14	14	14	F2F	24	22	2	30	18
15	15	15	F2F	35	28	7	44	21
16	16	16	F2F	31	22	9	48	17
17	17	17	F2F	31	26	5	46	14
18	18	18	F2F	34	25	9	51	16
19	19	19	F2F	34	25	9	71	20
20	20	20	F2F	33	27	6	23	13
41	41	41	ONL	31	27	4	70	23
42	42	42	ONL	31	25	6	63	16
43	43	43	ONL	34	29	5	41	23
44	44	44	ONL	34	29	5	28	14
45	45	45	ONL	30	24	6	43	22
46	46	46	ONL	30	33	-3	76	15
47	47	47	ONL	33	25	8	68	12
48	48	48	ONL	34	21	13	66	22
49	49	49	ONL	32	26	6	77	20
50	50	50	ONL	35	27	8	80	23
51	51	51	ONL	33	33	0	56	20
52	52	52	ONL	30	26	4	22	22
53	53	53	ONL	33	27	6	40	16
54	54	54	ONL	28	26	2	37	17
55	55	55	ONL	37	25	12	27	19
56	56	56	ONL	38	26	12	23	13
57	57	57	ONL	31	28	3	42	14
58	58	58	ONL	29	33	-4	40	19
59	59	59	ONL	34	29	5	30	21
60	60	60	ONL	32	30	2	57	22

Anwendungsszenario

Histogramme

```
# Histogrammparameter
h          = 1                                # gewünschte Klassenbreite
b_0        = min(D$BDI)                       # b_0
b_k        = max(D$BDI)                       # b_0
k          = ceiling((b_k - b_0)/h)           # Anzahl der Klassen
b          = seq(b_0, b_k, by = h)            # Klassen [b_{j-1}, b_j[
ylimits    = c(0,.25)                         # y-Achsenlimits
xlimits    = c(-2,14)                         # x-Achsenlimits
therapie   = c("F2F", "ONL")                 # Therapiebedingungen
labs       = c("Face-to-Face",               # Abbildungslabel
              "Online")

# Abbildungsparameter
par(                                               # für Details siehe ?par
mfcol      = c(1,2),                             # 1 x 2 Panelstruktur
family     = "sans",                             # Serif-freier Fonttyp
pty       = "m",                                 # Maximale 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,                                  # Non-Bold Titel
cex       = 1,                                  # Textvergrößerungsfaktor
cex.main   = 1)                                 # Titelttextvergrößerungsfaktor

# Iteration über Therapiebedingungen
for(i in 1:2){
  hist(
    D$BDI[D$Condition == therapie[i]],          # Werte von Therapiebedingung i
    breaks = b,                                 # Histogrammklassen
    freq   = F,                                 # normierte relative Häufigkeit
    xlim   = xlimits,                           # x-Achsenlimits
    ylim   = ylimits,                           # y-Achsenlimits
    xlab   = TeX("BDI"),                       # x-Achsenbeschriftung
    ylab   = "",                               # y-Achsenbeschriftung
    main   = labs[i])                          # Titelbeschriftung
}

# PDF Speicherung
dev.copy2pdf(
  file     = file.path(getwd(), "9_Abbildungen", "alm_9_F2F_ONL_histogramme.pdf"),
  width    = 8,
  height   = 4)
```

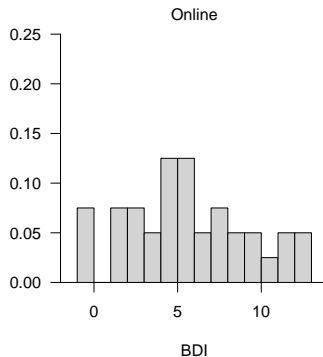
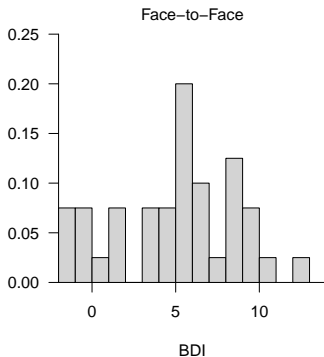
Deskriptive Statistiken

```
# Initialisierung eines Dataframes
tp      = c("F2F", "ONL")
ntp     = length(tp)
S       = data.frame(
  n      = rep(NaN,ntp),
  Max    = rep(NaN,ntp),
  Min    = rep(NaN,ntp),
  Median = rep(NaN,ntp),
  Mean   = rep(NaN,ntp),
  Var    = rep(NaN,ntp),
  Std    = rep(NaN,ntp),
  row.names = tp)

# Deskriptive Statistiken
# Therapiebedingungen
# Anzahl Therapiebedingungen
# Dataframeerzeugung
# Stichprobengrößen
# Maxima
# Minima
# Mediane
# Mittelwerte
# Varianzen
# Standardabweichungen
# Therapiebedingungen

# Iterationen über Therapiebedingungen
for(i in 1:ntp){
  data      = D$BDI[D$Condition == tp[i]]
  S$n[i]    = length(data)
  S$Max[i]  = max(data)
  S$Min[i]  = min(data)
  S$Median[i] = median(data)
  S$Mean[i] = mean(data)
  S$Var[i]  = var(data)
  S$Std[i]  = sd(data)
}
```

Deskriptive Statistiken der PostBDI-PreBDI Differenzen bei Face-to-Face und Online Therapie



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

```
>      n Max Min Median Mean Var Std  
> F2F 40 13 -3      6 5.28 14.8 3.85  
> ONL 40 18 -5      6 5.92 25.7 5.07
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Definition (Zweistichproben-T-Test Modell)

y_{ij} mit $i = 1, 2$ und $j = 1, \dots, n_i$ seien Zufallsvariablen, die die Datenpunkte eines Zweistichproben-T-Test Anwendungsszenarios modellieren. Dann hat das *Zweistichproben-T-Test Modell* die strukturelle Form

$$y_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, 2, j = 1, \dots, n_i \text{ mit } \mu_i \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (33)$$

die Datenverteilungsform

$$y_{ij} \sim N(\mu_i, \sigma^2) \text{ u.i.v. für } i = 1, 2, j = 1, \dots, n_i \text{ mit } \mu_i \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (34)$$

und für den Datenvektor $y = (y_{11}, \dots, y_{1n_1}, y_{21}, \dots, y_{2n_2})^T$ und $n := n_1 + n_2$ die Designmatrixform

$$y = X\beta + \varepsilon \text{ mit } X := \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2, \varepsilon \sim N(0_n, \sigma^2 I_n), \sigma^2 > 0. \quad (35)$$

Bemerkungen

- i indiziert die Gruppen, j indiziert die Daten in jeder Gruppe.
- n_1 und n_2 repräsentieren die Gruppengrößen, n repräsentiert die Gesamtanzahl an Datenpunkten.
- Es ist $p = 2$.
- Die Äquivalenz der drei Modellformen ergibt sich mit den Ergebnissen Einheit (5) Modellformulierung.

Datensimulation (vgl. Einheit (5) Modellformulierung)

```
# Libraries
library(MASS)

# Modellformulierung
n_1 = 40
n_2 = 40
n = n_1 + n_2
p = 2
X = matrix(c(rep(1,n_1), rep(0,n_1),
              rep(0,n_2), rep(1,n_2)),
           nrow = n)

I_n = diag(n)
beta = matrix(c(1,2), nrow = p)
sigsqr = 14

# Datenrealisierung
y = mvrnorm(1, X %*% beta, sigsqr*I_n)
```

Multivariate Normalverteilung

Anzahl von Datenpunkten Gruppe 1
Anzahl von Datenpunkten Gruppe 2
Gesamtanzahl Datenpunkte
Anzahl von Betaparameter
Designmatrix

n x n Einheitsmatrix
wahrer, aber unbekannter, Betaparameter
wahrer, aber unbekannter, Varianzparameter

eine Realisierung eines n-dimensionalen ZVs

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Theorem (Parameterschätzung im Zweistichproben-T-Test Modell)

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

$$\hat{\beta} = \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} =: \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \quad (36)$$

und für den Varianzparameterschätzer

$$\hat{\sigma}_{12}^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} =: s_{12}^2 \quad (37)$$

Bemerkungen

- \bar{y}_1 und \bar{y}_2 bezeichnen die gruppenspezifischen Stichprobenmittel.
- s_{12}^2 wird als *gepoolte Stichprobenvarianz* bezeichnet.
- Für einen Datensatz $y = (y_1, y_2)^T \in \mathbb{R}^{n_1+n_2}$ gilt allgemein, dass $s_y^2 \neq s_{12}^2$; die gepoolte Stichprobenvarianz und die Stichprobenvarianz eines konkatenierten Datensatzes sind im Allgemeinen also nicht identisch. Wir wollen das Konzept der gepoolten Stichprobenvarianz hier aber nicht weiter vertiefen.

Beweis

Für $i = 1, 2$ sei $y_i := (y_{i1}, \dots, y_{in_i})^T$. Dann ergibt sich für den Betaparameterschätzer

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T y \\&= \left(\begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix} \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \right)^{-1} \begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\&= \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\&= \begin{pmatrix} n_1^{-1} & 0 \\ 0 & n_2^{-1} \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \tag{38} \\&= \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\&=: \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}.\end{aligned}$$

Beweis (fortgeführt)

Gleichsam ergibt sich für Varianzparameterschätzer mit $n = n_1 + n_2$ und $p = 2$

$$\begin{aligned}\hat{\sigma}^2 &= \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \\&= \frac{1}{n_1 + n_2 - 2} \left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \right)^T \left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \right) \\&= \frac{1}{n_1 + n_2 - 2} \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \end{pmatrix}^T \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \end{pmatrix} \\&= \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} \\&=: s_{12}^2.\end{aligned}\tag{39}$$

Modellschätzung

```
# Dateneinlesen
fname      = file.path(getwd(), "9_Daten", "data_9_t_tests.csv") # Dateiname
D          = read.table(fname, sep = ",", header = TRUE)         # Dataframe
y_1        = D$BDI[D$Condition == "F2F"]                         # BDI Differenzwerte in der F2F Gruppe
y_2        = D$BDI[D$Condition == "ONL"]                         # BDI Differenzwerte in der ONL Gruppe

# Modellformulierung
n_1        = length(y_1)                                         # Anzahl Datenpunkte Gruppe 1 (F2F)
n_2        = length(y_2)                                         # Anzahl Datenpunkte Gruppe 2 (ONL)
n          = n_1 + n_2                                           # Gesamtanzahl Datenpunkte
y          = matrix(c(y_1, y_2), nrow = n)                       # Datenvektor
p          = 2                                                    # Anzahl Betaparameter
X          = matrix(c(rep(1,n_1), rep(0,n_1),                    # Designmatrix
                      rep(0,n_2), rep(1,n_2)),
                    nrow = n)

# 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
s_sqr_12    = ((n_1-1)*var(y_1) + (n_2-1)*var(y_2)) / (n_1+n_2-2) # gepoolte Stichprobenvarianz

# Ausgabe
cat("hat{beta}          : ", paste(beta_hat),                  # Betaparameterschätzer
    "\nbar{y}_1, bar{y}_2 : ", paste( c(mean(y_1), mean(y_2))), # Stichprobenmittel
    "\nhat{sigsqr}       : ", paste(sigsqr_hat),               # Varianzparameterschätzer
    "\ns_12^2           : ", paste(s_sqr_12),                  # gepoolte Stichprobenvarianz
    "\ns_y^2            : ", paste(var(y)))                    # Stichprobenvarianz des konkatinierten Datensatzes

> hat{beta}          :  5.275  5.925
> bar{y}_1, bar{y}_2 :  5.275  5.925
> hat{sigsqr}       :  20.2403846153846
> s_12^2           :  20.2403846153846
> s_y^2            :  20.0911392405063
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Überblick

- Wir gruppieren frequentistische Konfidenzintervalle und Hypothesentests unter Modellevaluation.
- Wir verzichten an dieser Stelle auf eine Diskussion von Konfidenzintervallen
- In der Praxis zielt die Evaluation von Zweistichproben-T-Tests ALM Designs meist auf einen Hypothesentest.
- Die Theorie der Zweistichproben-T-Tests ist umfangreich, siehe dazu WTFI (14) Zweistichproben-T-Tests.
- Ein gutes Verständnis von WTFI Einheit (12) Hypothesentests wird im Folgenden vorausgesetzt.

Im Zweistichproben-T-Test ALM Design ergeben sich folgende Hypothesenszenarien

- $H_0 : \mu_1 - \mu_2 = \mu_0$ und $H_1 : \mu_1 - \mu_2 \neq \mu_0$
- $H_0 : \mu_1 - \mu_2 \leq \mu_0$ und $H_1 : \mu_1 - \mu_2 > \mu_0$
- $H_0 : \mu_1 - \mu_2 \geq \mu_0$ und $H_1 : \mu_1 - \mu_2 < \mu_0$

Wir betrachten hier nur exemplarisch $H_0 : \mu_1 - \mu_2 = \mu_0$ und $H_1 : \mu_1 - \mu_2 \neq \mu_0$

Für $\mu_0 := 0$ gelten dabei insbesondere

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

Gliederung (vgl. WTFI Einheiten (12) - (14))

- (1) Statistisches Modell ✓
- (2) Testhypothesen ✓
- (3) Teststatistik
- (4) Test
- (5) Analyse der Testgütefunktion
- (6) Testumfangkontrolle
- (7) p-Werte
- (8) Analyse der Powerfunktion

(3) Teststatistik

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

Gegeben sei die Designmatrixform des Zweistichproben-T-Tests. Dann ergibt sich für die T-Teststatistik mit

$$c := (1, -1)^T \text{ und } c^T \beta_0 =: \mu_0, \quad (40)$$

dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad (41)$$

und es gilt

$$T \sim t(\delta, n_1 + n_2 - 2) \text{ mit } \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (42)$$

Bemerkungen

- Das Theorem basiert auf dem T-Teststatistik Theorem in Einheit (7) Modellevaluation.

(3) Teststatistik

Beweis

Mit dem T-Teststatistik Theorem in Einheit (7) Modellevaluation gilt zunächst für die Zähler von T und δ , dass

$$c^T \hat{\beta} - c^T \beta_0 = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} - \mu_0 = \bar{y}_1 - \bar{y}_2 - \mu_0 \quad (43)$$

und

$$c^T \beta - c^T \beta_0 = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} - \mu_0 = \mu_1 - \mu_2 - \mu_0, \quad (44)$$

respektive. Weiterhin gilt für die Nenner von T und δ , dass

$$c^T (X^T X)^{-1} c = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} n_1^{-1} & 0 \\ 0 & n_2^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} n_1^{-1} & -n_2^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{n_1} + \frac{1}{n_2} \quad (45)$$

Außerdem gilt

$$\left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-\frac{1}{2}} = \left(\frac{n_2}{n_1 n_2} + \frac{n_1}{n_1 n_2} \right)^{-\frac{1}{2}} = \left(\frac{n_1 + n_2}{n_1 n_2} \right)^{-\frac{1}{2}} = \left(\frac{n_1 n_2}{n_1 + n_2} \right)^{\frac{1}{2}} \quad (46)$$

Zusammengenommen folgt direkt, dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \text{ und } \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (47)$$

(4) Test

Definition (Zweiseitiger Zweistichproben-T-Test)

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

$$H_0 : \mu_1 - \mu_2 = \mu_0 \Leftrightarrow \Theta_0 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 - \mu_2 = \mu_0\} \quad (48)$$

und

$$H_1 : \mu_1 - \mu_2 \neq \mu_0 \Leftrightarrow \Theta_1 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 - \mu_2 \neq \mu_0\}, \quad (49)$$

respektive. Weiterhin sei die T-Teststatistik definiert durch

$$T := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{y}_1 - \bar{y}_2}{s_{12}} \right) \quad (50)$$

Dann ist der zweiseitige *Zweistichproben-T-Teststatistik* definiert als der kritischen Wert-basierte Test

$$\phi(y) := 1_{\{|T| \geq k\}}. \quad (51)$$

Bemerkungen

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

(5) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

Es sei ϕ der im obigen Modell formulierte Zweistichproben-T-Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$\begin{aligned} q_\phi : \mathbb{R}^2 &\rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\ &:= 1 - \psi(k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) + \psi(-k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) \end{aligned} \quad (52)$$

wobei $\psi(\cdot; d_{\mu_1, \mu_2}, n_1 + n_2 - 2)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter

$$d_{\mu_1, \mu_2} := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\mu_1 - \mu_2}{\sigma} \quad (53)$$

und Freiheitsgradparameter $n_1 + n_2 - 2$ bezeichnet.

Bemerkungen

- q_ϕ ist eine bivariate reellwertige Funktion.
- q_ϕ kann alternativ als univariate reellwertige Funktion von $\Delta := \mu_1 - \mu_2$ konzipiert werden.
- Im Vergleich zum Einstichprobenszenario gelten

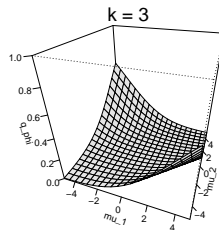
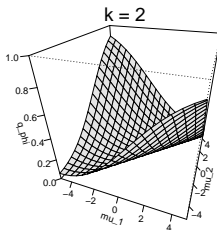
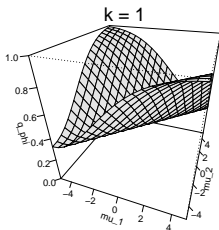
$$n \hookrightarrow n_1 + n_2 - 2, \quad \sqrt{n} \hookrightarrow \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \quad \mu - \mu_0 \hookrightarrow \mu_1 - \mu_2 \quad (54)$$

- Für einen Beweisansatz, siehe DeGroot and Schervish (2012) Seite 591.

(5) Analyse der Testgütefunktion

Testgütefunktion q_ϕ für $\sigma^2 = 9, n_1 = 12, n_2 = 12$.

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



(6) Testumfangkontrolle

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

$$k_{\alpha_0} := \psi^{-1} \left(1 - \frac{\alpha_0}{2}; n_1 + n_2 - 2 \right), \quad (55)$$

wobei $\psi^{-1}(\cdot; n_1 + n_2 - 2)$ die inverse KVF der t -Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden ist.

Bemerkungen

- Das Resultat folgt in Analogie zum Einstichproben-T-Test.
- Im Vergleich zum Einstichproben-T-Testfall gilt lediglich

$$n - 1 \hookrightarrow n_1 + n_2 - 2. \quad (56)$$

(6) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass die Daten zweier Gruppen v_{11}, \dots, v_{1n_1} und v_{21}, \dots, v_{2n_2} Realisationen von $y_{1j} \sim N(\mu_1, \sigma^2)$ u.i.v. für $j = 1, \dots, n_1$ und $y_{2j} \sim N(\mu_2, \sigma^2)$ u.i.v. für $j = 1, \dots, n_2$ mit unbekannten Parametern μ_1, μ_2, σ^2 sind.
- Man möchte entscheiden, ob eher $H_0 : \mu_1 - \mu_2 = \mu_0$ oder $H_1 : \mu_1 - \mu_2 \neq \mu_0$ 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 $n_1 = 12, n_2 = 12$, also Freiheitsgradparameter $12+12-2 = 22$, dass $k_{0.05} = \psi^{-1}(1 - 0.05/2; 22) \approx 2.07$ ist.
- Anhand von $n_1, n_2, \bar{v}_1, \bar{v}_2$ und der gepoolten Stichprobenstandardabweichung s_{12} berechnet man die Realisierung der Zweistichproben-T-Teststatistik

$$t := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{v}_1 - \bar{v}_2}{s_{12}} \right) \quad (57)$$

- 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 des Zweistichproben-T-Tests garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

(7) p-Werte

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 $|t| \geq \psi^{-1}(1 - \alpha_0/2; n_1 + n_2 - 2)$ abgelehnt werden. Für diese α_0 gilt, wie bereits mehrfach gezeigt,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (58)$$

- 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

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n_1 + n_2 - 2)). \quad (59)$$

- Im Vergleich zum Einstichprobenfall gilt lediglich $n \hookrightarrow n_1 + n_2 - 2$.

Modellevaluation

Anwendungsszenario

```
# Dateneinlesen
fname = file.path(getwd(), "9_Daten", "data_9_t_tests.csv") # Dateiname
D = read.table(fname, sep = ",", header = TRUE) # Dataframe
y_1 = D$BDI[D$Condition == "F2F"] # BDI Differenzwerte in der F2F Gruppe
y_2 = D$BDI[D$Condition == "ONL"] # BDI Differenzwerte in der ONL Gruppe

# Modellformulierung
n_1 = length(y_1) # Anzahl Datenpunkte Gruppe 1 (F2F)
n_2 = length(y_2) # Anzahl Datenpunkte Gruppe 2 (ONL)
n = n_1 + n_2 # Gesamtanzahl Datenpunkte
y = matrix(c(y_1, y_2), nrow = n) # Datenvektor
p = 2 # Anzahl Betaparameter
X = matrix(c(rep(1, n_1), rep(0, n_1), # Designmatrix
             rep(0, n_2), rep(1, n_2)),
           nrow = n)

# 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

# Modellevaluation
c = matrix(c(1, -1), nrow = 2) # Kontrastgewichtsvektor
mu_0 = 0 # Nullhypothese H_0
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){ # Test  $1_{|T(X)| \geq k_{\alpha_0/2}}$ 
  phi = 1 # Ablehnen von  $H_0$ 
} else {
  phi = 0 # Nicht Ablehnen von  $H_0$ 
}
pval = 2*(1-pt(abs(t), n_1+n_2-2)) # p-Wert

> fg = 78
> t = -0.646
> alpha_0 = 0.05
> k_alpha_0 = 1.99
> phi = 0
> p-Wert = 0.52
```


Modellevaluation

Anwendungsszenario

```
# Automatischer Zweistichproben-T-Test
varphi = t.test(
  y_1,
  y_2,
  var.equal = TRUE,
  alternative = c("two.sided"),
  conf.level = 1-alpha_0)

# ?t.test für Details
# Datensatz y_1
# Datensatz y_2
# \sigma_1^2 = \sigma_2^2
# H_1: \mu_1 \neq \mu_2
# \delta = 1 - \alpha_0 (sic!)

# Ausgabe
print(varphi)
```

```
>
> Two Sample t-test
>
> data: y_1 and y_2
> t = -0.6, df = 78, p-value = 0.5
> alternative hypothesis: true difference in means is not equal to 0
> 95 percent confidence interval:
> -2.65 1.35
> sample estimates:
> mean of x mean of y
> 5.28 5.92
```

```
# Genauere Ausgabe t
paste(varphi[1])
```

```
> [1] "c(t = -0.646128613569434)"
```

```
# Genauere Ausgabe p
paste(varphi[3])
```

```
> [1] "0.520092604281206"
```

(8) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$\begin{aligned} q_\phi : \mathbb{R}^2 &\rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\ &:= 1 - \psi(k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) + \psi(-k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) \end{aligned} \quad (60)$$

als Funktion des Nichtzentralitätsparameters und der Summe der Stichprobenumfänge $n := n_1 + n_2$ bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 2)$ mit festem α_0 .

Es ergibt sich die multivariate reellwertige Funktion

$$\begin{aligned} \pi : \mathbb{R} \times \mathbb{N}^2 &\rightarrow [0, 1], (d, n) \mapsto \\ \pi(d, n) &:= 1 - \psi(k_{\alpha_0}; d, n - 2) + \psi(-k_{\alpha_0}; d, n - 2) \end{aligned} \quad (61)$$

Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekannten Wert d und von der Summe der Stichprobengrößen n ab. De-facto handelt es sich also um die gleiche Powerfunktion wie beim zweiseitigen Einstichproben-T-Test mit dem einzigen Unterschied, dass für den Freiheitsgradparameter $n - 2$ anstelle von $n - 1$ gilt. Wir verzichten auf eine erneute Visualisierung.

(8) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem $n = n_1 + n_2$ 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.
- Ungleichgewichte zwischen n_1 und n_2 werden durch die Tatsache ausglich, dass Datenpunkte einer Stichproben auch zur Varianzschätzung in der anderen Stichprobe beitragen, da eine identische Varianz vorausgesetzt wurde.

Die Powerfunktion hängt vom wahren, aber unbekannten, Parameterwert $d = \sqrt{n}(\mu_1 - \mu_0)/\sigma$ ab.

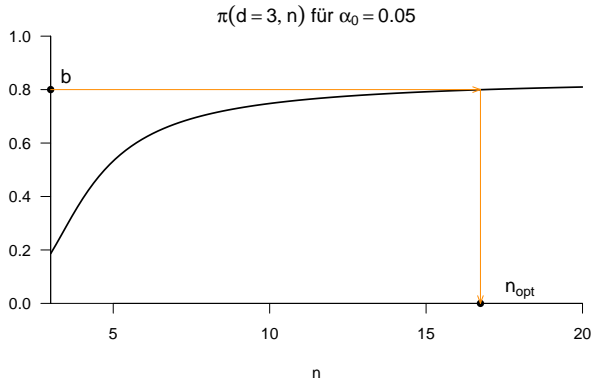
⇒ Wenn man d schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert

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

(8) Analyse der Powerfunktion

Praktisches Vorgehen



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 zweiseitigen Einstichproben-T-Tests bei fester Stichprobengröße.
16. Skizzieren Sie die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei festem Nichtzentralitätsparameter.
17. Betrachten Sie die PostBDI-PreBDI Differenzwertdaten der Waitlist Control Bedingung im Beispieldatensatz. Erstellen Sie ein Histogramm dieser Daten und evaluieren Sie Ihnen bekannte deskriptive Statistiken zu diesen Daten. Führen Sie einen zweiseitigen Einstichproben-T-Test mit Nullhypothesenparameter $\mu_0 = 0$ durch. Dokumentieren Sie Ihre Ergebnisse. Was folgern Sie aus den sich ergebenden Resultaten?

18. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests.
19. Geben Sie die Definition des Zweistichproben-T-Test Modells wieder.
20. Geben Sie das Theorem zur Parameterschätzung im Zweistichproben-T-Test Modell wieder.
21. Geben Sie das Theorem zur T-Teststatistik des Zweistichproben-T-Tests wieder.
22. Erläutern Sie mögliche Hypothesenszenarien eines Zweistichproben-T-Tests.
23. Geben Sie die Definition des zweiseitigen Zweistichproben-T-Tests (mit ungerichteter Hypothese) wieder.
24. Skizzieren Sie die Testgütefunktion des zweiseitigen Zweistichproben-T-Tests.
25. Geben Sie das Theorem zur Testumfangkontrolle im zweiseitigen Level- α_0 -Zweistichproben-T-Test wieder.
26. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- α_0 -Zweistichproben-T-Tests.
27. Geben Sie die Definition des p-Wertes Werts für einen zweiseitigen Zweistichproben-T-Test wieder.
28. Von welchen Werten hängt die Powerfunktion eines zweiseitigen Zweistichproben-T-Tests ab?
29. Skizzieren Sie die Powerfunktion des zweiseitigen Zweistichproben-T-Tests bei fester Stichprobengröße.
30. Skizzieren Sie die Powerfunktion des zweiseitigen Zweistichproben-T-Tests bei festem Nichtzentralitätsparameter.
31. Betrachten Sie die Daten zum Alter der Patient:innen in der Face-to-Face und Online Therapie Bedingung im Beispieldatensatz. Erstellen gruppenspezifische Histogramme dieser Daten und evaluieren Sie gruppenspezifische deskriptive Statistiken zu diesen Daten. Führen Sie einen zweiseitigen Zweistichproben-T-Test mit Nullhypothesenparameter $\mu_0 = 0$ durch. Dokumentieren Sie Ihre Ergebnisse. Was folgern Sie aus den sich ergebenden Resultaten?

References

DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.