Allgemeines Lineares Modell

BSc Psychologie SoSe 2023

Prof. Dr. Dirk Ostwald

(9) T-Tests

Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

Überblick

Modul B2 Inferenzstatistik | Allgemeines Lineares Modell

Datum	Einheit	Thema
13.04.2023	Grundlagen	(1) Regression
20.04.2023	Grundlagen	(2) Korrelation
27.04.2023	Grundlagen	(3) Matrizen
04.05.2023	Grundlagen	(4) Normalverteilungen
11.05.2023	Theorie	(5) Modellformulierung
18.05.2023	Theorie	(6) Modellschätzung
25.05.2023	Theorie	(7) T-Statistiken
01.06.2023	Theorie	(8) F-Statistiken
08.06.2021	Anwendung	(9) T-Tests
15.06.2021	Anwendung	(10) Einfaktorielle Varianzanalyse
22.06.2021	Anwendung	(11) Zweifaktorielle Varianzanalyse
29.06.2023	Anwendung	(12) Partielle Korrelation
06.07.2023	Anwendung	(13) Multiple Regression
13.07.2023	Anwendung	(14) Kovarianzanalyse
20.07.2023	Klausurtermin	
März 2024	Klausurwiederholungstermin	

Kontinuum von ALM Designs

Extremszenario (1) Die Erwartungswerte aller Datenvariablen sind identisch.

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

$\Rightarrow$ Jegliche Datenvariabilität wird dem Fehlerterm zugeschrieben.

Extremszenario (2) Die Erwartungswerte aller Datenvariablen sind paarweise verschieden

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

$\Rightarrow$ Jegliche Datenvariabilität wird dem Erwartungswertparameter zugeschrieben.

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

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.

Überblick

Faktorielle und Parametrische ALM Designs

Faktorielle ALM Designs

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

Überblick

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.

Überblick

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 (Behrends-Fischer Problem).
- Zweistichproben T-Tests bei abhängigen Stichproben.

Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

Einstichproben-T-Tests

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendungsszenario

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

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

Gruppenanalysen mit Wechsler Adult Intelligence Scale

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

Gruppenanalysen in der funktionellen Kernspintomographie

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

Anwendungsbeispiel

Face-to-face PreBDI

PostBDI

Wir nehmen an, dass die Datenpunkte u.i.v. Realisierungen von ZVen $v_i \sim N(\mu, \sigma^2)$ 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.

Anwendungsszenario

Dateneinlesen

```
fname        = file.path(getwd(), "9_Daten", "data_9_t_tests.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

Anwendungsszenario

```r
# 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_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(0,0.15)                     # 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,                            # 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       = "dBDI",                          # 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)
```

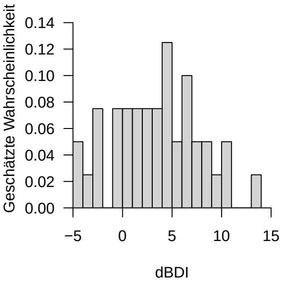
Anwendungsszenario

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

# Initialisierung eines Dataframes
tp            = c("F2F")                         # Therapiebedingungen
ntp           = length(tp)                       # Anzahl Therapiebedingungen
S             = data.frame(                      # Dataframeerzeugung
                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 3.92 19.4 4.41
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Modellformulierung

Definition (Einstichproben-T-Test-Modell)

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

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

die Datenverteilungsform

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

und für den Datenvektor $\upsilon = (\upsilon_1, ..., \upsilon_n)^T$ die Designmatrixform

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

Bemerkungen

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

Datensimulation (vgl. Einheit (5) Modellformulierung)

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

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} \upsilon_i =: \bar{\upsilon}, \tag{6}$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\upsilon_i - \bar{\upsilon})^2 =: s_\upsilon^2 \tag{7}$$

Bemerkungen

- Die Formen von $\hat{\beta}$ und $\hat{\sigma}^2$ wurden in Einheit (6) Parameterschätzung hergeleitet.
- $\bar{\upsilon}$ und s_υ^2 bezeichnen das Stichprobenmittel und die Stichprobenvarianz der $\upsilon_1, ..., \upsilon_n$.

Modellschätzung

```r
# Dateneinlesen
fname       = file.path(getwd(), "9_Daten", "data_9_t_tests.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}   :  3.92
> bar{y}      :  3.92
> hat{sigsqr} :  19.4
> s_y^2       :  19.4
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

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

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 Einheit (12))

(1) Statistisches Modell ✓

(2) Testhypothesen ✓

(3) Teststatistik

(4) Test

(5) Analyse der Testgütefunktion

(6) Testumfangkontrolle

(7) p-Wert

(8) Analyse der Powerfunktion

Bezeichungskonventionen für t-Zufallsvariablen

- Für eine (nichtzentrale) t-Zufallsvariable ξ schreiben wir $\xi \sim t(n)$ ($\xi \sim t(\delta, n)$).
- Die WDF einer (nichtzentralen) t-Zufallsvariable bezeichnen wir mit $t(\cdot; n)$ ($t(\cdot; \delta, n)$).
- Die KVF einer (nichtzentralen) t-Zufallsvariable bezeichnen wir mit $\Psi(\cdot; n)$ ($\Psi(\cdot; \delta, n)$).
- Die Teststatistik eines T-Test-Designs Hypothesentests 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

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

dass

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

und es gilt

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

Bemerkungen

- Das Theorem basiert auf dem T-Statistik Theorem in Einheit (7) T-Statistiken.
- Wir erinnern an das verwandte populäre stichprobengrößenunabhängige Effekstärke Maß *Cohen's* d bei Einstichproben-T-Test-Designs,

$$d := \frac{\bar{\upsilon}}{s_\upsilon}. \tag{11}$$

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

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

Modellevaluation (3) Teststatistik

Beweis

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

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

Weiterhin gilt mit demselben Theorem

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

Modellevaluation (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

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \tag{15}$$

definiert. Weiterhin sei die T-Teststatistik definiert als

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

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

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

Bemerkungen

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

Theorem (Testgütefunktion)

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

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

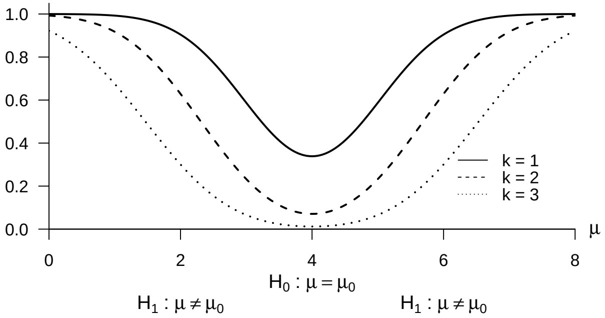
wobei $\psi(\cdot; \delta, n-1)$ die KVF der nichtzentralen t-Verteilung mit Nichtzentralitätsparameter

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

und Freiheitsgradparameter $n-1$ bezeichnet.

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

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



Modellevaluation (5) Analyse der Testgütefunktion

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 wird die also zunächst die Verteilung der Teststatistik. Wir haben oben bereits gesehen, dass die T-Teststatistik

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

unter der Annahme $\upsilon_i \sim N(\mu, \sigma^2)$ 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

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

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.

□

Modellevaluation (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

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

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

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

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

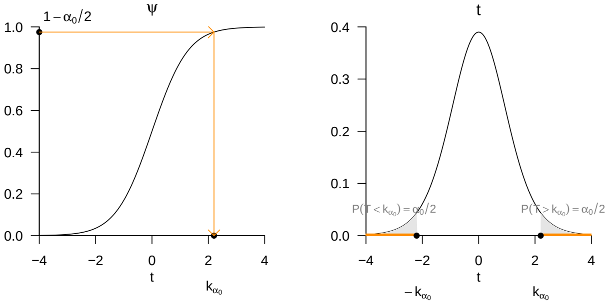
wobei $\psi(\cdot; \delta, n-1)$ und $\psi(\cdot; n-1)$ die KVF der nichtzentralen t-Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparameter $n-1$ sowie der t-Verteilung mit Freiheitsgradparameter $n-1$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

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

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.

Modellevaluation (6) Testumfangkontrolle

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

Praktisches Vorgehen

- Man nimmt an, dass ein Datensatz $\upsilon_1, ..., \upsilon_n$ eine Realisation von $\upsilon_i \sim N(\mu, \sigma^2)$ 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 $n = 12$, also Freiheitsgradparameter 11, dass $k_{0.05} = \psi^{-1}(1 - 0.05/2; 11) \approx 2.20$ ist.
- Anhand von $n, \mu_0, \bar{\upsilon}$ und s_υ berechnet man die Realisierung der T-Teststatistik

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

- 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

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

- Dies aber folgt aus

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

Modellevaluation

```r
# Dateneinlesen
fname       = file.path(getwd(), "9_Daten", "data_9_t_tests.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

# 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 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|}
pval       = 2*(1 - pt(abs(t), n-1))                               # p-Wert
d          = t/sqrt(n)                                             # Cohen's d
```

```
> fg        =  39
> kappa_1   =  2.52 5.33
> t         =  5.64
> alpha_0   =  0.05
> k_alpha_0 =  2.02
> phi       =  1
> p-Wert    =  1.66e-06
> Cohen's d =  0.891
```

Modellevaluation

Anwendungszenario

```
# Automatischer Einstichproben-T-Test
varphi    = t.test(                               # ?t.test für Details
            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 (sic!)

# Ausgabe
print(varphi)
```

```
>
>   One Sample t-test
>
> data:  y
> t = 6, df = 39, p-value = 2e-06
> alternative hypothesis: true mean is not equal to 0
> 95 percent confidence interval:
>  2.52 5.33
> sample estimates:
> mean of x
>      3.92
```

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

```
> [1] "c(t = 5.63531986397201)"
```

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

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

Wir betrachten die Testgütefunktion

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

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} auch von n ab.

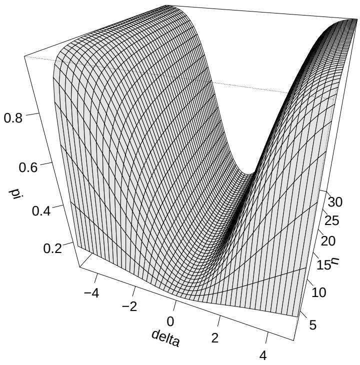
Es ergibt sich die bivariate reellwertige Funktion

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

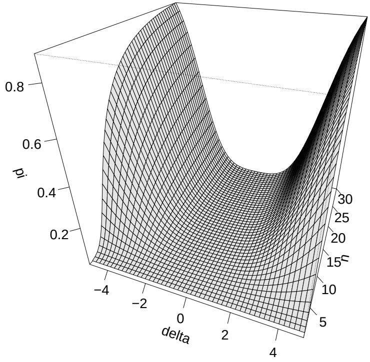
Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekannten Wert δ und von der Stichprobengröße n ab. Wir visualisieren diese Abhängigkeiten untenstehend.

Modellevaluation (8) Analyse der Powerfunktion

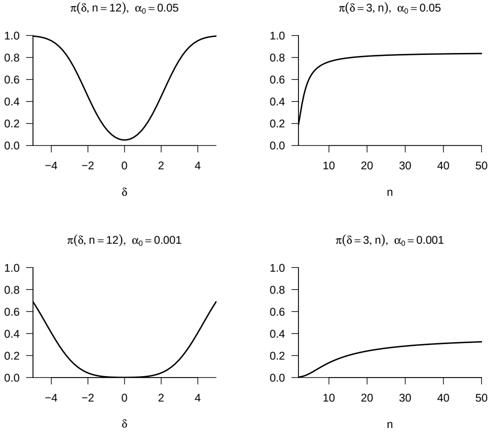
Powerfunktion für $\alpha_0 = 0.05$



Powerfunktion für $\alpha_0 = 0.001$



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.

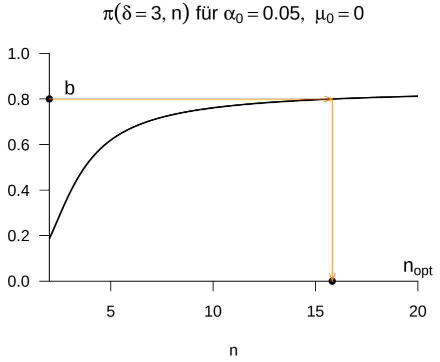
Die Powerfunktion hängt vom wahren, aber unbekannten, Wert $\delta = \sqrt{n}\frac{\mu - \mu_0}{\sigma}$ ab.

$\Rightarrow$ 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.

Praktisches Vorgehen



Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

Zweistichproben-T-Tests

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendungsszenario

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

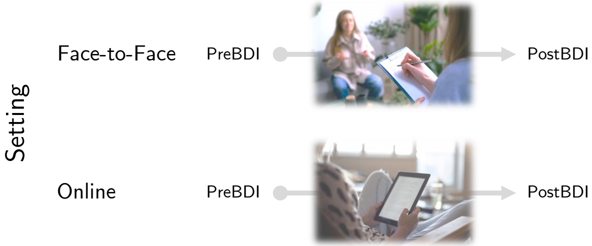
BDI Differenzwert Datenanalyse bei zwei Gruppen von Patient:innen

- 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?

Anwendungsbeispiel



Wir nehmen an, dass die Datenpunkte der Face-to-Face Therapiegruppe u.i.v. Realisierungen von ZVen $\upsilon_{1j} \sim N(\mu_1, \sigma^2)$ für $j = 1, ..., 40$ und dass die Datenpunkte der Online Therapiegruppe u.i.v. Realisierungen von ZVen $\upsilon_{2j} \sim N(\mu_2, \sigma^2)$ für $j = 1, ..., 40$ 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.

Dateneinlesen | $j = 1, \ldots 15$ für jede Gruppe

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

	X	ID	Setting	PreBDI	PostBDI	dBDI
1	1	1	F2F	29	25	4
2	2	2	F2F	32	27	5
3	3	3	F2F	28	31	-3
4	4	4	F2F	36	22	14
5	5	5	F2F	32	29	3
6	6	6	F2F	28	28	0
7	7	7	F2F	33	30	3
8	8	8	F2F	33	26	7
9	9	9	F2F	33	28	5
10	10	10	F2F	30	28	2
11	11	11	F2F	36	25	11
12	12	12	F2F	32	31	1
13	13	13	F2F	29	31	-2
14	14	14	F2F	24	29	-5
15	15	15	F2F	35	32	3
41	41	41	ONL	31	24	7
42	42	42	ONL	31	30	1
43	43	43	ONL	34	25	9
44	44	44	ONL	34	25	9
45	45	45	ONL	30	26	4
46	46	46	ONL	30	28	2
47	47	47	ONL	33	26	7
48	48	48	ONL	34	26	8
49	49	49	ONL	32	24	8
50	50	50	ONL	35	25	10
51	51	51	ONL	33	26	7
52	52	52	ONL	30	24	6
53	53	53	ONL	33	28	5
54	54	54	ONL	28	21	7
55	55	55	ONL	37	27	10

Anwendungsszenario

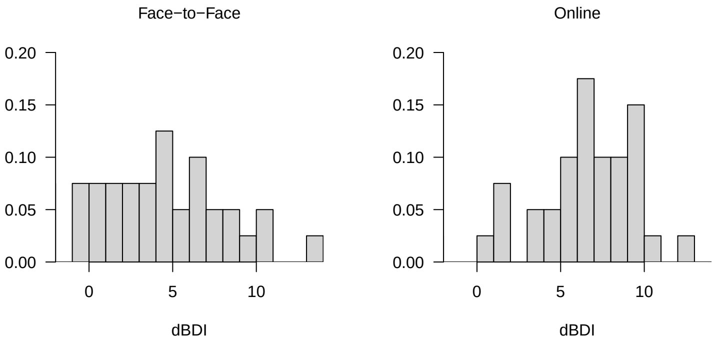
```r
# Histogrammparameter
h           = 1                                           # gewünschte Klassenbreite
b_0         = min(D$dBDI)                                 # b_0
b_k         = max(D$dBDI)                                 # b_0
k           = ceiling((b_k - b_0)/h)                      # Anzahl der Klassen
b           = seq(b_0, b_k, by = h)                       # Klassen [b_{j-1}, b_j[
ylimits     = c(0,.2)                                     # 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)                                          # Titeltextvergrößerungsfaktor
# Iteration über Therapiebedingungen
for(i in 1:2){
  hist(
  D$dBDI[D$Setting == therapie[i]],                       # Werte von Therapiebedingung i
  breaks    = b,                                          # Histogrammklassen
  freq      = F,                                          # normierte relative Häufigkeit
  xlim      = xlimits,                                    # x-Achsenlimits
  ylim      = ylimits,                                    # y-Achsenlimits
  xlab      = "dBDI",                                     # 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)
```

Anwendungsszenario

```r
# Initialisierung eines Dataframes
tp            = c("F2F", "ONL")                          # Therapiebedingungen
ntp           = length(tp)                               # Anzahl Therapiebedingungen
S             = data.frame(                              # Dataframeerzeugung
                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 PostBDI-PreBDI Differenzen bei Face-to-Face und Online Therapie




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

```
>       n Max Min Median Mean  Var  Std
> F2F 40  14  -5      4 3.92 19.4 4.41
> ONL 40  20  -7      7 7.40 23.8 4.88
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Modellformulierung

Definition (Zweistichproben-T-Test-Modell)

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

$$\upsilon_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, 2, j = 1, ..., n_i \text{ mit } \mu_i \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (35)$$

die Datenverteilungsform

$$\upsilon_{ij} \sim N(\mu_i, \sigma^2) \text{ u.i.v. für } i = 1, 2, j = 1, ..., n_i \text{ mit } \mu_i \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (36)$$

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

$$\upsilon = X\beta + \varepsilon \text{ mit } X := \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2, \varepsilon \sim N(0_n, \sigma^2 I_n), \sigma^2 > 0. \quad (37)$$

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.
- Die Anzahl der Betaparameter ist $p = 2$.
- Die Äquivalenz der drei Modellformen ergibt sich mit den Ergebnissen Einheit (5) Modellformulierung.

Datensimulation (vgl. Einheit (5) Modellformulierung)

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

# Modellformulierung
n_1    = 40                                     # Anzahl von Datenpunkten Gruppe 1
n_2    = 40                                     # Anzahl von Datenpunkten Gruppe 2
n      = n_1 + n_2                              # Gesamtanzahl Datenpunkte
p      = 2                                      # Anzahl von Betaparameter
X      = matrix(c(rep(1,n_1), rep(0,n_1),      # Designmatrix
                  rep(0,n_2), rep(1,n_2)),
                  nrow  = n)
I_n    = diag(n)                                # n x n Einheitsmatrix
beta   = matrix(c(1,2), nrow = p)               # 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 Zweistichproben-T-Test-Modell)

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

$$\hat{\beta} = \begin{pmatrix} \frac{1}{n_1}\sum_{j=1}^{n_1} \upsilon_{1j} \\ \frac{1}{n_2}\sum_{j=1}^{n_2} \upsilon_{2j} \end{pmatrix} =: \begin{pmatrix} \bar{\upsilon}_1 \\ \bar{\upsilon}_2 \end{pmatrix} \tag{38}$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{n_1} (\upsilon_{1j} - \bar{\upsilon}_1)^2 + \sum_{j=1}^{n_2} (\upsilon_{2j} - \bar{\upsilon}_2)^2}{n_1 + n_2 - 2} =: s_{12}^2 \tag{39}$$

Bemerkungen

- $\bar{\upsilon}_1$ und $\bar{\upsilon}_2$ bezeichnen die gruppenspezifischen Stichprobenmittel.
- s_{12}^2 wird als *gepoolte Stichprobenvarianz* bezeichnet.
- Für einen Datensatz $\upsilon = (\upsilon_1, \upsilon_2)$ gilt im Allgemeinen, dass $s_\upsilon^2 \neq s_{12}^2$; 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 $\upsilon_i := (\upsilon_{i1}, ..., \upsilon_{in_i})^T$. Dann ergibt sich für den Betaparameterschätzer

$$
\begin{aligned}
\hat{\beta} &= (X^T X)^{-1} X^T y \\
&= \left(\begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix} \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \right)^{-1} \begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \upsilon_1 \\ \upsilon_2 \end{pmatrix} \\
&= \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{n_1} \upsilon_{1j} \\ \sum_{j=1}^{n_2} \upsilon_{2j} \end{pmatrix} \\
&= \begin{pmatrix} n_1^{-1} & 0 \\ 0 & n_2^{-1} \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n_1} \upsilon_{1j} \\ \sum_{j=1}^{n_2} \upsilon_{2j} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} \upsilon_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} \upsilon_{2j} \end{pmatrix} \\
&=: \begin{pmatrix} \bar{\upsilon}_1 \\ \bar{\upsilon}_2 \end{pmatrix}.
\end{aligned}
\tag{40}
$$

Beweis (fortgeführt)

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

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \\
&= \frac{1}{n_1 + n_2 - 2} \left(\begin{pmatrix} \upsilon_1 \\ \upsilon_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{\upsilon}_1 \\ \bar{\upsilon}_2 \end{pmatrix} \right)^T \left(\begin{pmatrix} \upsilon_1 \\ \upsilon_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{\upsilon}_1 \\ \bar{\upsilon}_2 \end{pmatrix} \right) \\
&= \frac{1}{n_1 + n_2 - 2} \begin{pmatrix} \upsilon_{11} - \bar{\upsilon}_1 \\ \vdots \\ \upsilon_{1n_1} - \bar{\upsilon}_1 \\ \upsilon_{21} - \bar{\upsilon}_2 \\ \vdots \\ \upsilon_{2n_2} - \bar{\upsilon}_2 \end{pmatrix}^T \begin{pmatrix} \upsilon_{11} - \bar{\upsilon}_1 \\ \vdots \\ \upsilon_{1n_1} - \bar{\upsilon}_1 \\ \upsilon_{21} - \bar{\upsilon}_2 \\ \vdots \\ \upsilon_{2n_2} - \bar{\upsilon}_2 \end{pmatrix} \\
&= \frac{\sum_{j=1}^{n_1} (\upsilon_{1j} - \bar{\upsilon}_1)^2 + \sum_{j=1}^{n_2} (\upsilon_{2j} - \bar{\upsilon}_2)^2}{n_1 + n_2 - 2} \\
&=: s_{12}^2 .
\end{aligned}
\tag{41}
$$

Modellschätzung

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

# Modellformulierung
n_1        = length(y_1)                                          # Anzahl Datenpunkte Gruppe 1 (F2F)
n_2        = length(y_1)                                          # 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 konkatenierten Datensatzes
```

```
> hat{beta}          :  3.925 7.4
> bar{y}_1, bar{y_2} :  3.925 7.4
> hat{sigsqr}        :  21.5945512820513
> s_12^2             :  21.5945512820513
> s_y^2              :  24.3783227848101
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

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.
- 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

- $H_0 : \mu_1 - \mu_2 = 0 \Leftrightarrow H_0 : \mu_1 = \mu_2$
- $H_1 : \mu_1 - \mu_2 \neq 0 \Leftrightarrow H_1 : \mu_1 \neq \mu_2$

Gliederung (vgl. WTFI Einheiten (12))

(1) Statistisches Modell ✓

(2) Testhypothesen ✓

(3) Teststatistik

(4) Test

(5) Analyse der Testgütefunktion

(6) Testumfangkontrolle

(7) p-Wert

(8) Analyse der Powerfunktion

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

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

$$c := (1, -1)^T \text{ und } c^T\beta_0 =: \mu_0, \tag{42}$$

dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{\upsilon}_1 - \bar{\upsilon}_2 - \mu_0}{s_{12}} \right) \tag{43}$$

und es gilt

$$T \sim t(\delta, n_1 + n_2 - 2) \text{ mit } \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \tag{44}$$

Bemerkungen

- Das Theorem basiert auf dem T-Statistik Theorem in Einheit (7) T-Statistiken.
- Wir erinnern an das verwandte populäre stichprobengrößenunabhängige Effekstärke Maß *Cohen's* d bei Zweistichproben-T-Test-Designs,

$$d := \frac{\bar{\upsilon_1} - \bar{\upsilon_2}}{s_{12}}. \tag{45}$$

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

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} d \Leftrightarrow d = T / \sqrt{\frac{n_1 n_2}{n_1 + n_2}}. \tag{46}$$

Modellevaluation (3) Teststatistik

<u>Beweis</u>

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

$$c^T\hat{\beta} - c^T\beta_0 = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \bar{\upsilon}_1 \\ \bar{\upsilon}_2 \end{pmatrix} - \mu_0 = \bar{\upsilon}_1 - \bar{\upsilon}_2 - \mu_0 \tag{47}$$

und

$$c^T\beta - c^T\beta_0 = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} - \mu_0 = \mu_1 - \mu_2 - \mu_0, \tag{48}$$

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

$$c^T(X^TX)^{-1}c = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} n_1^{-1} & 0 \\ 0 & n_2^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} n_1^{-1} & -n_2^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{n_1} + \frac{1}{n_2} \tag{49}$$

Außerdem gilt

$$\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} = \left(\frac{n_2}{n_1 n_2} + \frac{n_1}{n_1 n_2}\right)^{-\frac{1}{2}} = \left(\frac{n_1 + n_2}{n_1 n_2}\right)^{-\frac{1}{2}} = \left(\frac{n_1 n_2}{n_1 + n_2}\right)^{\frac{1}{2}} \tag{50}$$

Zusammengenommen folgt direkt, dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{\upsilon}_1 - \bar{\upsilon}_2 - \mu_0}{s_{12}}\right) \text{ und } \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\mu_1 - \mu_2 - \mu_0}{\sigma}\right). \tag{51}$$

Definition (Zweiseitiger Zweistichproben-T-Test)

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

$$H_0 : \mu_1 - \mu_2 = \mu_0 \Leftrightarrow \Theta_0 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 | \mu_1 - \mu_2 = \mu_0\} \tag{52}$$

und

$$H_1 : \mu_1 - \mu_2 \neq \mu_0 \Leftrightarrow \Theta_1 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 | \mu_1 - \mu_2 \neq \mu_0\}, \tag{53}$$

respektive. Weiterhin sei die T-Teststatistik definiert durch

$$T := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{\upsilon}_1 - \bar{\upsilon}_2 - \mu_0}{s_{12}}\right) \tag{54}$$

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

$$\phi(\upsilon) := 1_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases}. \tag{55}$$

Bemerkungen

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

Theorem (Testgütefunktion)

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

$$
\begin{aligned}
q_\phi : \mathbb{R}^2 \to [0,1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\
:= 1 - \psi(k; \delta, n_1 + n_2 - 2) + \psi(-k; \delta, n_1 + n_2 - 2)
\end{aligned}
\tag{56}
$$

wobei $\psi(\cdot; \delta, n_1 + n_2 - 2)$ die KVF der nichtzentralen t-Verteilung mit Nichtzentralitätsparameter

$$
\delta := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\mu_1 - \mu_2 - \mu_0}{\sigma}
\tag{57}
$$

und Freiheitsgradparameter $n_1 + n_2 - 2$ bezeichnet.

Bemerkungen

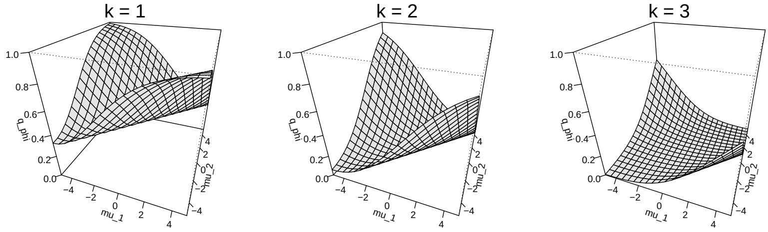
- q_ϕ ist eine bivariate reellwertige Funktion.
- q_ϕ kann alternativ als univariate reellwertige Funktion von $\Delta := \mu_1 - \mu_2$ konzipiert werden.
- Im Vergleich zum Einstichprobenszenario gelten

$$
n - 1 \hookrightarrow n_1 + n_2 - 2, \qquad \sqrt{n} \hookrightarrow \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \qquad \mu - \mu_0 \hookrightarrow \mu_1 - \mu_2 - \mu_0
\tag{58}
$$

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

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

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



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

$$k_{\alpha_0} := \psi^{-1}\left(1 - \frac{\alpha_0}{2}; n_1 + n_2 - 2\right), \tag{59}$$

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

$$n - 1 \hookrightarrow n_1 + n_2 - 2. \tag{60}$$

Modellevaluation (6) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass die Daten zweier Gruppen $v_{11}, \ldots, v_{1n_1}$ und $v_{21}, \ldots, v_{2n_2}$ Realisationen von $v_{1j} \sim N(\mu_1, \sigma^2)$ u.i.v. für $j = 1, \ldots, n_1$ und $v_{2j} \sim N(\mu_2, \sigma^2)$ u.i.v. für $j = 1, \ldots, n_2$ 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 $12 + 12 - 2 = 22$, so 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

$$t := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{v}_1 - \bar{v}_2 - \mu_0}{s_{12}} \right) \tag{61}$$

- 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.

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

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \tag{62}$$

- 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

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n_1 + n_2 - 2)). \tag{63}$$

- Im Vergleich zum Einstichprobenfall gilt lediglich $n - 1 \hookrightarrow n_1 + n_2 - 2$.

```r
# Modellevaluation
fname       = file.path(getwd(), "9_Daten", "data_9_t_tests.csv")  # Dateiname
D           = read.table(fname, sep = ",", header = TRUE)           # Dataframe
y_1         = D$dBDI[D$Setting == "F2F"]                            # BDI Differenzwerte in der F2F Gruppe
y_2         = D$dBDI[D$Setting == "ONL"]                            # BDI Differenzwerte in der ONL Gruppe
n_1         = length(y_1)                                           # Anzahl Datenpunkte Gruppe 1 (F2F)
n_2         = length(y_1)                                           # 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_2),                      # Designmatrix
                       rep(0,n_1), rep(1,n_2)), nrow  = n)
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
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
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){phi = 1} else {phi = 0}                     # Test 1_{|T(X) >= k_alpha_0|}
pval        = 2*(1-pt(abs(t), n_1+n_2-2))                           # p-Wert
d           = t/sqrt((n_1*n_2)/(n_1 + n_2))                         # Cohen's d
print(beta_hat)
```

```
>      [,1]
> [1,] 3.92
> [2,] 7.40
> fg        =  78
> kappa_1   =  2.46 5.39
> kappa_2   =  5.94 8.86
> t         =  -3.34
> alpha_0   =  0.05
> k_alpha_0 =  1.99
> phi       =  1
> p-Wert    =  0.00127
> Cohen's d =  -0.748
```

Modellevaluation

```r
# Automatischer Zweistichproben-T-Test
varphi    = t.test(                              # ?t.test für Details
            y_1,                                 # Datensatz y_1
            y_2,                                 # Datensatz y_2
            var.equal   = TRUE,                  # \sigma_1^2 = \sigma_2^2
            alternative = c("two.sided"),        # H_1: \mu_1 \neq \mu_2
            conf.level  = 1-alpha_0)             # \delta = 1 - \alpha_0 (sic!)

# Ausgabe
print(varphi)
```

```
>
>   Two Sample t-test
>
> data:  y_1 and y_2
> t = -3, df = 78, p-value = 0.001
> alternative hypothesis: true difference in means is not equal to 0
> 95 percent confidence interval:
>  -5.54 -1.41
> sample estimates:
> mean of x mean of y
>      3.92      7.40
```

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

```
> [1] "c(t = -3.34424213733072)"
```

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

```
> [1] "0.00127017790178135"
```

- R nutzt hier eine alternative Parameterisierung des Zweistichproben-T-Test Szenarios, die sogenannten *Effektdarstellung*.
- Wir werden die Effektdarstellung im Kontext der Einfaktoriellen Varianzanalyse ausführlich diskutieren.

Wir betrachten die Testgütefunktion

$$
\begin{aligned}
q_\phi : \mathbb{R}^2 \to [0,1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\
:= 1 - \psi(k; \delta, n_1 + n_2 - 2) + \psi(-k; \delta, n_1 + n_2 - 2) \quad (64)
\end{aligned}
$$

als Funktion des Nichtzentralitätsparameters $\delta := \sqrt{\frac{n_1 n_2}{n_1+n_2}} \frac{\mu_1 - \mu_2 - \mu_0}{\sigma}$ 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

$$
\pi : \mathbb{R} \times \mathbb{N} \to [0,1], (\delta, n) \mapsto \pi(\delta, n) := 1 - \psi(k_{\alpha_0}; \delta, n-2) + \psi(-k_{\alpha_0}; \delta, n-2) \quad (65)
$$

Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekannten Wert δ 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.

Modellevaluation (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 ausglichen, dass Datenpunkte einer Stichproben auch zur Varianzschätzung in der anderen Stichprobe beitragen, da eine identische Varianz vorausgesetzt wurde.

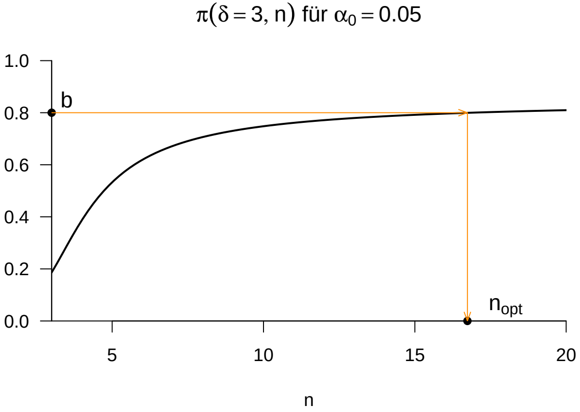
Die Powerfunktion hängt vom wahren, aber unbekannten, Wert $\delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\mu_1 - \mu_2 - \mu_0}{\sigma}$ ab.

$\Rightarrow$ 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.

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. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests.
18. Geben Sie die Definition des Zweistichproben-T-Test-Modells wieder.
19. Geben Sie das Theorem zur Parameterschätzung im Zweistichproben-T-Test-Modell wieder.
20. Geben Sie das Theorem zur T-Teststatistik des Zweistichproben-T-Tests wieder.
21. Erläutern Sie mögliche Hypothesenszenarien eines Zweistichproben-T-Tests.
22. Geben Sie die Definition des zweiseitigen Zweistichproben-T-Tests (mit ungerichteter Hypothese) wieder.
23. Geben Sie das Theorem zur Testumfangkontrolle im zweiseitigen Level-α_0-Zweistichproben-T-Test wieder.
24. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level-α_0-Zweistichproben-T-Tests.
25. Geben Sie die Definition des p-Wertes für einen zweiseitigen Zweistichproben-T-Test wieder.

References

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