Multivariate Datenanalyse

MSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(10) T^2-Tests

Vorbemerkungen

Einstichproben-T^2-Tests

Zweistichproben-T^2-Tests

Univariates vs. multivariates Testen

Selbstkontrollfragen

Vorbemerkungen

Ausgewählte multivariate Methoden der Frequentistischen Inferenz

Multivariate Generalisierungen bekannter Frequentistischer Verfahren

$\mathbf{T}^2$-Tests als Generalisierung von T-Tests

- Inferentieller Vergleich von ein bis zwei Gruppen multivariater Daten

Multivariate **Einfaktorielle Varianzanalyse** als Generalisierung der einfaktoriellen Varianzanalyse

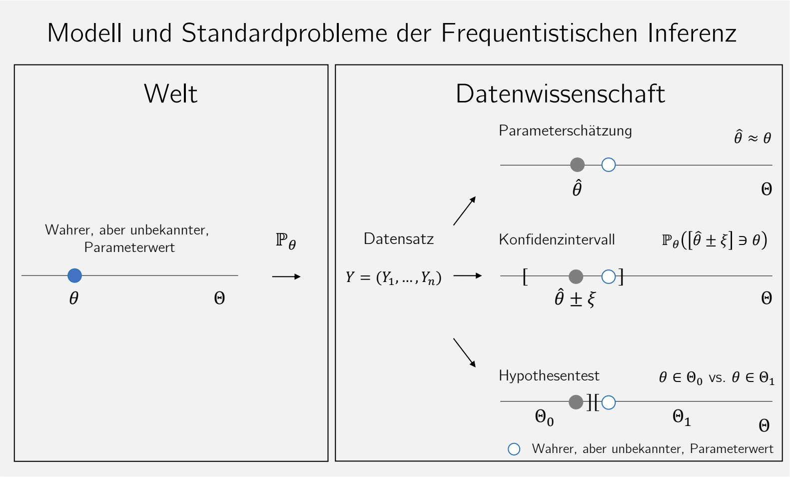
- Inferentieller Vergleich von drei oder mehr Gruppen multivariater Daten

Kanonische Korrelationsanalyse als Generalisierung (multipler) Korrelation

- Korrelative Zusammenhänge von Zufallsvektoren

Zur Revision univariater Frequentistischer Verfahren

- Inferenzstatistik SoSe 21
- Wahrscheinlichkeitstheorie und Frequentistische Inferenz WiSe 21/22


Modell und Standardprobleme der Frequentistischen Inferenz
Welt
Wahrer, aber unbekannter, Parameterwert
θ
Θ
$\mathbb{P}_\theta$
Datenwissenschaft
Datensatz
$Y = (Y_1, \dots, Y_n)$
Parameterschätzung
$\hat{\theta} \approx \theta$
$\hat{\theta}$
Θ
Konfidenzintervall
$\mathbb{P}_\theta([\hat{\theta} \pm \xi] \ni \theta)$
$\hat{\theta} \pm \xi$
Θ
Hypothesentest
$\theta \in \Theta_0$ vs. $\theta \in \Theta_1$
Θ_0
Θ_1
Θ
Wahrer, aber unbekannter, Parameterwert

Vorbemerkungen

Standardprobleme Frequentistischer Inferenz

(1) **Parameterschätzung**

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für den wahren, aber unbekannten, Parameterwert (oder eine Funktion dessen) abzugeben, typischerweise basierend auf der Beobachtung einer Realisierung von $Y_1, ..., Y_n \sim \mathbb{P}_\theta$.

(2) **Konfidenzintervalle**

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) **Hypothesentests**

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Beobachtungen $Y_1, ..., Y_n$ in einer möglichst sinnvollen Form zu entscheiden, ob der wahre, aber unbekannte Parameterwert, in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Standardannahmen Frequentistischer Inferenz

$\mathcal{M}$ sei ein statistisches Modell mit unabhängig und identisch verteilten Zufallsvektoren $Y_1, ..., Y_n \sim \mathbb{P}_\theta$. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen von $Y_1, ..., Y_n \sim \mathbb{P}_\theta$ ist. Aus frequentistischer Sicht kann man die Erhebung von Datensätzen unendlich oft wiederholen und zu jedem Datensatz Statistiken auswerten.

Datensatz (1) : $y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, ..., y_n^{(1)}\right)$, Statistik (1): $S : \mathbb{R}^{m \times n} \to \Sigma, y^{(1)} \mapsto S\left(y^{(1)}\right)$

Datensatz (2) : $y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, ..., y_n^{(2)}\right)$, Statistik (2): $S : \mathbb{R}^{m \times n} \to \Sigma, y^{(2)} \mapsto S\left(y^{(2)}\right)$

Datensatz (3) : $y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, ..., y_n^{(3)}\right)$, Statistik (3): $S : \mathbb{R}^{m \times n} \to \Sigma, y^{(3)} \mapsto S\left(y^{(3)}\right)$

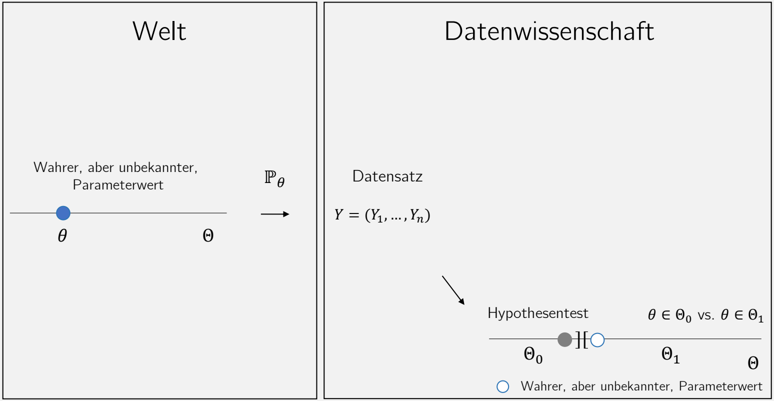
Datensatz (4) : $y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, ..., y_n^{(4)}\right)$, Statistik (4): $S : \mathbb{R}^{m \times n} \to \Sigma, y^{(4)} \mapsto S\left(y^{(4)}\right)$

...

Um die Qualität statistischer Methoden zu beurteilen betrachtet die frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Statistiken und Schätzern unter der Annahme von $Y_1, ..., Y_n \sim \mathbb{P}_\theta$.

Wenn eine statistische Methode im Sinne der frequentitischen Standardannahmen “gut” ist, dann heißt das also, dass sie bei häufiger Anwendung “im Mittel gut” ist. Im Einzelfall, also im realen Normalfall nur eines vorliegenden Datensatzes, kann sie auch “schlecht” sein.

Modell und Standardprobleme der Frequentistischen Inferenz



Definition (Mahalanobis Distanz)

ξ_1 sei ein Zufallsvektor, eine Realisiation eines Zufallsvektors, ein multivariater Erwartungswert oder ein multivariates Stichprobenmittel, ξ_2 sei ein Zufallsvektor, eine Realisiation eines Zufallsvektors, ein multivariater Erwartungswert oder ein multivariates Stichprobenmittel und Ξ sei eine Kovarianzmatrix oder eine Stichprobenkovarianzmatrix. Dann heißt

$$D = (\xi_1 - \xi_2)^T \, \Xi^{-1} \, (\xi_1 - \xi_2) \tag{1}$$

Mahalanobis Distanz von ξ_1 und ξ_2 hinsichtlich Ξ.

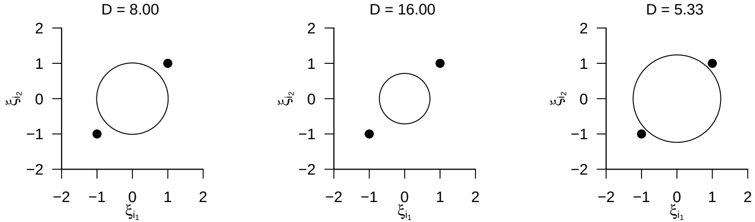
Bemerkungen

- Eine Mahalanobis Distanz ist eine Kovarianzmatrix-normalisierte quadrierte Euklidische Distanz.
- Ähnliche Maße in der univariaten Statistik sind die z-Transformation $z = \frac{y-\mu}{\sigma}$ und Cohen's $d = \frac{\bar{y}_1 - \bar{y}_2}{s_{12}}$.
- Ähnlich wie bei z-Werten wird bei der Mahalanobis Distanz in "Einheiten von Kovarianzen" gemessen.
- Stark variante Komponenten von ξ_1 und ξ_2 tragen weniger zur Distanz bei.
- Stark kovariante Komponenten von ξ_1 und ξ_2 tragen weniger zur Distanz bei.

Vorbemerkungen

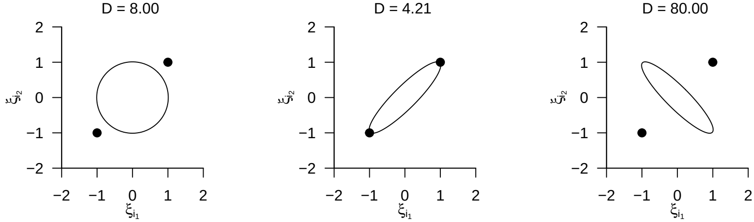
Mahalanobis Distanzen als Funktion von Komponentenvarianzen

$$\Sigma := \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix} \qquad \Sigma := \begin{pmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{pmatrix} \qquad \Sigma := \begin{pmatrix} 1.5 & 0.0 \\ 0.0 & 1.5 \end{pmatrix}$$



Mahalanobis Distanzen als Funktion von Komponentenkovarianzen

$$\Sigma := \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix} \qquad \Sigma := \begin{pmatrix} 1.0 & 0.9 \\ 0.9 & 1.0 \end{pmatrix} \qquad \Sigma := \begin{pmatrix} 1.0 & -0.9 \\ -0.9 & 1.0 \end{pmatrix}$$



Definition (f-Zufallsvariable)

X sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und Wahrscheinlichkeitsdichtefunktion

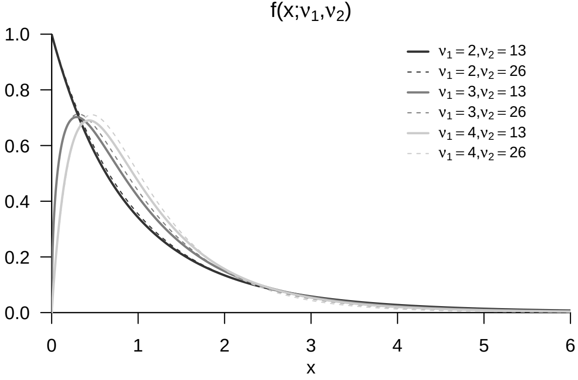
$$p_X : \mathbb{R} \to \mathbb{R}_{>0}, x \mapsto p_X(x) := \nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}} \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \frac{x^{\frac{\nu_1}{2}-1}}{(\nu_1 x + \nu_2)^{\frac{\nu_1+\nu_2}{2}}}, \tag{2}$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass X einer f-Verteilung mit Freiheitsgradparametern ν_1 und ν_2 unterliegt und nennen X eine f-Zufallsvariable mit Freiheitsgradparametern ν_1 und ν_2. Wir kürzen dies mit $X \sim f(\nu_1, \nu_2)$ ab. Die Wahrscheinlichkeitsdichtefunktion (WDF) einer f-Zufallsvariable bezeichnen wir mit $f(x; \nu_1, \nu_2)$, die kumulative Verteilungsfunktion (KVF) einer f-Zufallsvariable bezeichnen wir mit $F(x; \nu_1, \nu_2)$, und die inverse kumulative Verteilungsfunktion einer f-Zufallsvariable bezeichnen wir mit $F^{-1}(x; \nu_1, \nu_2)$.

Bemerkungen

- Im univariaten Fall ist die F-Statistik der Varianzanalyse bei Zutreffen der Nullhypothese f-verteilt
- Im multivariaten Fall ist z.B. die T^2-Statistik bei Zutreffen der Nullhypothese f-verteilt.

Wahrscheinlichkeitsdichtefunktionen von f-Verteilungen



Definition (Nichtzentrale f-Zufallsvariable)

X sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und Wahrscheinlichkeitsdichtefunktion

$$p_X : \mathbb{R} \to \mathbb{R}_{>0}, x \mapsto$$

$$p_X(x) := \sum_{k=0}^{\infty} \frac{e^{-\delta/2}(\delta/2)^k}{\frac{\Gamma(\nu_2/2)\Gamma(\nu_1/2+k)}{\Gamma(\nu_2/2+\nu_1/2+k)}k!} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2+k} \left(\frac{\nu_2}{\nu_2+\nu_1 x}\right)^{(\nu_1+\nu_2)/2+k} x^{\nu_1/2-1+k} \quad (3)$$

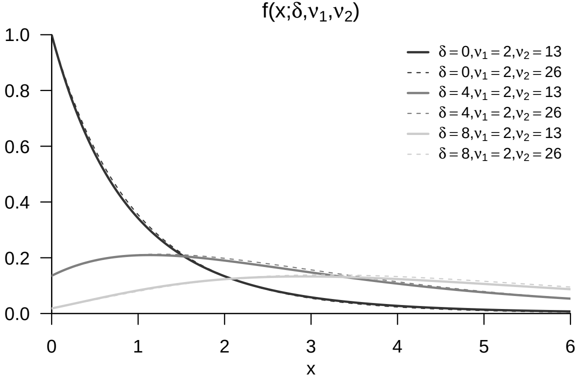
wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass X einer nichtzentralen f-Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparametern ν_1 und ν_2 unterliegt und nennen X eine nichtzentrale f-Zufallsvariable mit Nichtzentralitätsparameter δ und Freiheitsgradparametern ν_1 und ν_2. Wir kürzen dies mit $X \sim f(\delta, \nu_1, \nu_2)$ ab. Die Wahrscheinlichkeitsdichtefunktion (WDF) einer f-Zufallsvariable bezeichnen wir mit $f(x; \delta, \nu_1, \nu_2)$, die kumulative Verteilungsfunktion (KVF) einer nichtzentralen f-Zufallsvariable bezeichnen wir mit $F(x; \delta, \nu_1, \nu_2)$, und die inverse kumulative Verteilungsfunktion einer nichtzentralen f-Zufallsvariable bezeichnen wir mit $F^{-1}(x; \delta, \nu_1, \nu_2)$.

Bemerkungen

- Es gilt $f(0, \nu_1, \nu_2) = f(\nu_1, \nu_2)$.
- Im univariaten Fall ist die F-Statistik bei Nichtzutreffen der Nullhypothese nichtzentral f-verteilt
- Im multivariaten Fall ist z.B. die T^2-Statistik bei Nichtzutreffen der Nullhypothese nichtzentral f-verteilt.

Vorbemerkungen

WDFen von nichtzentralen f-Verteilungen



Vorbemerkungen

Einstichproben-$\mathbf{T}^2$-Tests

Zweistichproben-T^2-Tests

Univariates vs. multivariates Testen

Selbstkontrollfragen

Einstichproben-T^2-Tests

Anwendungszenario

- **Eine Stichprobe experimenteller Einheiten**.
- Annahme unabhängiger und identisch nach $N(\mu, \Sigma)$ multivariat normalverteilter Daten.
- μ und Σ unbekannt.
- Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ mit μ_0 beabsichtigt.

Anwendungsbeispiele

- Gruppenanalyse von BDI und Glukokortikoid Daten
  - $\mu \neq \mu_0$ als Evidenz für eine multivariate Abweichung von einem Normwert μ_0.
- Gruppenanalyse von Kognitionstestdaten
  - $\mu \neq \mu_0$ als Evidenz für eine multivariate Abweichung von einem Normwert μ_0.

Anwendungsbeispiel

VP	IQ Test Score	Math Test Score
1	54	44
2	60	20
3	67	36
4	41	39
5	66	57
6	51	28
7	51	46
8	37	46
9	57	54
10	47	12
11	50	67
12	42	63
13	60	64
14	36	64
15	60	71

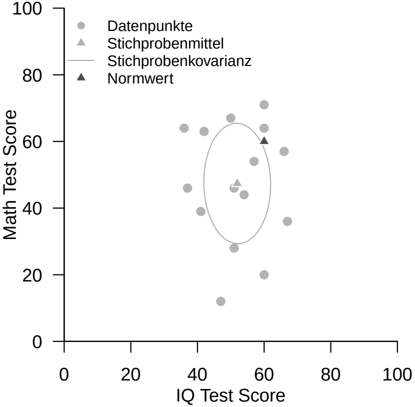
Abweichung des (IQ Test Score, Math Test Score) Erwartungswertparameters vom Normwert $\mu_0 = \begin{pmatrix} 60 \\ 60 \end{pmatrix}$?

Anwendungsbeispiel

```r
# Deskriptivstatistik
library(foreign)                                                         # R Paket
library(ellipse)                                                         # R Paket
library(matlib)                                                          # R Paket
mu_0   = matrix(c(60,60), nrow = 2)                                      # Normwert
fname  = file.path(getwd(), "10_Daten", "studienerfolg.sav")             # Dateinamen
D      = read.spss(fname, to.data.frame = T)                             # Dateneinlesen
Y      = rbind(D$X1[D$Gruppe == "ungenügend"],                           # Y_{i_1} IQ Test Score
               D$X2[D$Gruppe == "ungenügend"])                           # Y_{i_2} Math Test Score
n      = ncol(Y)                                                         # Anzahl Datenpunkte
j_n    = matrix(rep(1,n), nrow = n)                                      # 1_n
I_n    = diag(n)                                                         # I_n
J_n    = matrix(rep(1,n^2), nrow = n)                                    # 1_{nn}
Y_bar  = (1/n)*(Y %*% j_n)                                               # Stichprobenmittel (cf.(3) Wahrscheinlichkeitstheorie)
C      = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))                      # Stichprobenkovarianzmatrix (cf. ibid.)
D      = t(Y_bar - mu_0) %*% inv(C) %*% (Y_bar - mu_0)                   # Mahalanobis Distanz

# Ausgabe
cat("Y_bar =", Y_bar,
    "\nD     =", D)
```

```
Y_bar = 51.9 47.4
D     = 1.19
```

Anwendungsbeispiel



Einstichproben-T^2-Tests

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

Einstichproben-T^2-Tests

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

Einstichproben-T^2-Tests

Gliederung (vgl. (12) Hypothesentests)

(1) Statistisches Modell in klassischer Form

(2) Statistisches Modell in generativer Form

(3) Testhypothesen, Teststatistik, Test

(4) Analyse der Teststatistik

(5) Analyse der Testgütefunktion

(6) Testumfangkontrolle

(7) p-Werte

(8) Analyse der Powerfunktion

Zur Wiederholung des univariaten Falls, siehe (13) Einstichproben-T-Tests.

(1) Statistisches Modell in klassischer Form

$Y_1, ..., Y_n \sim N(\mu, \Sigma)$ sei die Stichproben eines m-dimensionalen Normalverteilungsmodells mit unbekanntem Erwartungswertparameter $\mu \in \mathbb{R}^m$ und unbekanntem Kovarianzmatrixparameter $\Sigma \in \mathbb{R}^{m \times m}$ p.d. Als Parameter von Interesse betrachten wir $\theta = \mu$, so dass sich der Parameterraum von Interesse zu $\Theta = \mathbb{R}^m$ ergibt.

(2) Statistisches Modell in generativer Form

Es sei

$$Y_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0_m, \Sigma) \text{ für } i = 1, ..., n \tag{4}$$

wobei

- $Y_i, i = 1, ..., m$ beobachtbare Zufallsvektoren,
- $\mu \in \mathbb{R}^m$ den festen und identischen Erwartungswertparameter über Zufallsvektoren und
- $\varepsilon_i, i = 1, ..., n$ unabhängige normalverteilte nicht beobachtbare Zufallsvektoren

bezeichnen.

(2) Statistisches Modell in generativer Form (fortgeführt)

Die generative Form betont, dass im vorliegenden Modell beobachtete Daten durch einen systematischen deterministischen Prozess (hier $\mu \in \mathbb{R}^m$) unter dem additiven Einfluss einer Vielzahl unabhängiger und deshalb in ihrer Summe normalverteilter Störprozesse (hier in der Summe $\varepsilon_i, i = 1, ..., n$) erzeugt konzipiert werden.

Beweis

Wir halten zunächst fest, dass wenn $\xi \sim N(\alpha, S)$ ein m_1-dimensionaler normalverteilter Zufallsvektor mit Erwartungswertparameter $\alpha \in \mathbb{R}^{m_1}$ und Kovarianzmatrixparameter $S \in \mathbb{R}^{m_1 \times m_1}$ p.d. ist und für $A \in \mathbb{R}^{m_2 \times m_1}$ und $b \in \mathbb{R}^{m_2}$ ein m_2-dimensionaler Zufallsvektor definiert ist als

$$\zeta := A\xi + b, \tag{5}$$

dann gilt, dass

$$\zeta \sim N\left(A\alpha + b, ASA^T\right) \tag{6}$$

(vgl. Anderson (2003), Section 2.4). Aus $\varepsilon \sim N(0_m, \Sigma)$ folgt hier mit

$$Y_i = I_m \varepsilon + \mu \tag{7}$$

dann aber sofort, dass

$$Y \sim N\left(I_m 0_m + \mu, I_m \Sigma I_m^T\right) = N(\mu, \Sigma). \tag{8}$$

□

Einstichproben-T^2-Tests

(3) Testhypothesen, Teststatistik, Test

Für ein $\mu_0 \in \mathbb{R}^m$ betrachten wir die einfache Nullhypothese und die zusammengesetzte Alternativhypothese

$$H_0 := \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_0 := \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R}^m \setminus \{\mu_0\} \tag{9}$$

Weiterhin betrachten wir die Einstichproben-T^2-Teststatistik

$$T^2 := n(\bar{Y} - \mu_0)^T C^{-1} (\bar{Y} - \mu_0) \tag{10}$$

wobei $\bar{Y}$ und C das Stichprobenmittel und die Stichprobenkovarianzmatrix der $Y_1, ..., Y_n$, respektive, bezeichnen (vgl. (3) Wahrscheinlichkeitstheorie).

- T^2 ist die mit dem Stichprobenumfang skalierte Mahalanobis Distanz von $\bar{Y}$ und μ_0 hinsichtlich C.
- $T^2 \uparrow$ für $||\bar{Y} - \mu_0|| \uparrow$, $C \downarrow$ und $n \uparrow$.

Schließlich definieren wir den kritischen Wert-basierten Test

$$\phi(Y) := 1_{\{T^2 > k\}} := \begin{cases} 1 & T^2 > k \\ 0 & T^2 \leq k \end{cases}. \tag{11}$$

wobei wie üblich 1 den Vorgang des Ablehnens von H_0 und 0 den Vorgang des Nichtablehnens von H_0 repräsentieren.

(4) Analyse der Teststatistik

Theorem (Verteilung der Einstichproben-T^2-Teststatistik)

Es seien $Y_1, ..., Y_n \sim N(\mu, \Sigma)$ mit $\mu \in \mathbb{R}^m$ und $\Sigma \in \mathbb{R}^{m \times m}$ p.d.,

$$\nu := \frac{n - m}{(n - 1)m} \tag{12}$$

und für $\mu \in \mathbb{R}^m$ sei die Einstichproben-T^2-Teststatistik definiert als

$$T^2 := n(\bar{Y} - \mu_0)^T C^{-1} (\bar{Y} - \mu_0). \tag{13}$$

Dann gilt

$$\nu T^2 \sim f(\delta, m, n - m) \tag{14}$$

wobei $f(\delta, m, n - m)$ die nichtzentrale f-Verteilung mit Nichtzentralitätsparameter

$$\delta := n(\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0) \tag{15}$$

sowie mit Freiheitsgradparametern m und $n - m$ bezeichnet.

Bemerkungen

- Für einen Beweis von (14) verweisen wir auf Anderson (2003) und Hotelling (1931).
- Für $\mu = \mu_0$ und damit $\delta = 0$ entspricht $f(m, n - m, \delta)$ der f-Verteilung $f(m, n - m)$

Theorem (WDF und KVF der Einstichproben-T^2-Teststatistik)

Im Einstichproben-T^2-Testszenario sei

$$\nu := \frac{n-m}{(n-1)m} \tag{16}$$

Dann ist eine WDF der Einstichproben-T^2-Teststatistik gegeben durch

$$p_{T^2} : \mathbb{R}_{\geq 0} \to \mathbb{R}, t^2 \mapsto p_{T^2}(t^2) := \nu f(\nu t^2; \delta, m, n-m) \tag{17}$$

und eine WDF der Einstichproben-T^2-Teststatistik ist gegeben durch

$$P_{T^2} : \mathbb{R}_{\geq 0} \to [0,1], t^2 \mapsto P_{T^2}(t^2) := F(\nu t^2; \delta, m, n-m) \tag{18}$$

Bemerkungen

- νT^2 hat die WDF $f(\delta, m, n-m)$, T^2 dagegen hat die WDF $\nu f(\nu t^2; \delta, m, n-m)$.
- Für $m := 1$ ist $\nu = (n-1)/(n-1) \cdot 1 = 1$ und mit der Stichprobenvarianz S^2 gilt

$$T^2 = n\frac{(\bar{Y}-\mu_0)^2}{S^2} = \left(\sqrt{n}\frac{\bar{Y}-\mu_0}{S}\right)^2 \tag{19}$$

- Das Quadrat der univariate Einstichproben-T-Teststatistik $T := \sqrt{n}\frac{\bar{Y}-\mu_0}{S}$ ist also $f(\delta, 1, n-1)$ verteilt.

Einstichproben-T^2-Tests

Beweis

Wir halten zunächst fest, dass das Theorem zur univariate WDF Transformation bei linear-affinen Abbildungen (vgl. (7) Transformationen der Normalverteilung) besagt, dass für eine Zufallsvariable X mit WDF p_X und der Definition $Y = f(X)$ mit $f(X) := aX + b$ für $a \neq 0$ eine WDF von Y definiert ist $p_Y(y) := (1/|a|)p_X((y-b)/a)$. Im vorliegenden Fall ist $X = \nu T^2$ mit WDF $f(\delta, m, n-m)$ und $Y := T^2 = \frac{1}{\nu}\nu T^2$, also $a = 1/\nu$ und $b = 0$. Mit $\nu > 0$ ergibt sich (17) also aus

$$p_{T^2}(t^2) = \frac{1}{a} p_X\left(\frac{t^2}{a}\right) = \nu f(\nu t^2; m, n-m) \tag{20}$$

(18) folgt dann mit der Tatsache, dass WDFen bei kontinuierlichen Zufallsvariablen die Ableitungen der entsprechenden KVF sind, sowie der Kettenregel der Differentiation

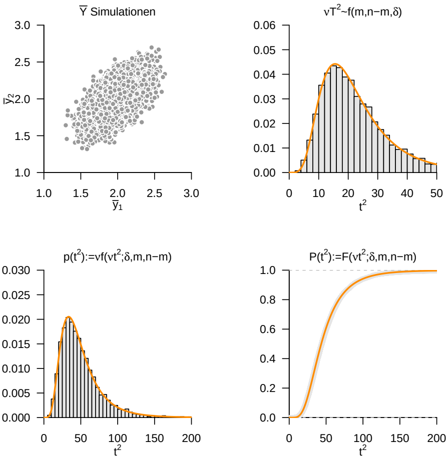
$$\begin{aligned}
\frac{d}{dt^2} P_{T^2}\left(t^2\right) &= \frac{d}{dt^2}\left(F(\nu t^2; m, n-m, \delta)\right) \\
&= \frac{d}{dt^2} F(\nu t^2; m, n-m, \delta) \frac{d}{dt^2}\left(\nu t^2\right) \\
&= \nu f(\nu t^2; m, n-m, \delta) \\
&= p_{T^2}(t^2)
\end{aligned} \tag{21}$$

□

(4) Analyse der Teststatistik

```
# Modellparameter
m     = 2                                                   # Dimensionalität der Zufallsvektoren/Daten
n     = 15                                                  # Anzahl der Datenpunkte
mu_0  = matrix(c(1,1) , nrow = 2)                           # H0 Hypothesenparameter
mu    = matrix(c(2,2) , nrow = 2)                           # wahrer, aber unbekannter, Erwartungswertparameter
Sigma = matrix(c(0.5,0.3, 0.3,0.5), nrow = 2, byrow = TRUE) # wahrer, aber unbekannter, Kovarianzmatrixparameter

# Simulation
library(MASS)                                               # R Paket für multivariate Normalverteilungen
library(matlib)                                             # R Paket für Matrizenrechnung
nsim  = 1e4                                                 # Anzahl Simulationen/Datensatzrealisierungen
Yb    = matrix(rep(NaN,m*nsim), nrow = 2)                   # Stichprobenmittelarray
T2    = rep(NaN,nsim)                                       # T2 Statistik Array
j_n   = matrix(rep(1,n), nrow = n)                          # 1_n
I_n   = diag(n)                                             # I_n
J_n   = matrix(rep(1,n^2), nrow = n)                        # 1_{nn}
for(s in 1:nsim){                                           # Simulationsiterationen
    Y       = t(mvrnorm(n,mu,Sigma))                        # Y_i \sim N(\mu,\Sigma), i = 1,...,n
    Y_bar   = (1/n)*(Y %*% j_n)                             # Stichprobenmittel
    C       = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))    # Stichprobenkovarianzmatrix
    T2[s]   = n*t(Y_bar - mu_0) %*% inv(C) %*% (Y_bar - mu_0) # T2 Statistik
    Yb[,s]  = Y_bar                                         # Stichprobenmittel für Visualisierung
}
```

(4) Analyse der Teststatistik



(5) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

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

$$q_\phi : \mathbb{R}^m \to [0,1], \mu \mapsto q_\phi(\mu) := 1 - F(\nu k; \delta_\mu, m, n-m) \tag{22}$$

wobei $F(\cdot; \delta_\mu, m, n-m)$ die KVF der nichtzentralen f-Verteilung mit Freiheitsgradparametern$m$ und $n-m$ sowie mit Nichtzentralitätsparameter

$$\delta_\mu := n(\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0) \tag{23}$$

bezeichnet.

Bemerkungen

- q_ϕ kann zur Bestimmung kritischer Werte für einen erwünschten Testumfang genutzt werden.
- q_ϕ kann zur Bestimmung der Testpower genutzt werden.

(5) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Tests im vorliegenden Testszenario ist definiert als

$$q_\phi : \mathbb{R}^m \to [0,1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1) \tag{24}$$

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

$$\frac{n-m}{m(n-1)}T^2 \sim f(m, n-m, \delta_\mu) \text{ mit } \delta_\mu := n(\mu - \mu_0)^T\Sigma^{-1}(\mu - \mu_0) \tag{25}$$

gilt. Der Ablehnungsbereich des betrachteten Tests ist $A :=]k, \infty[$. Also ergibt sich

$$\begin{aligned} q_\phi(\mu) &= \mathbb{P}_\mu(\phi = 1) \\ &= \mathbb{P}_\mu\left(T^2 \in \;]k,\infty[\right) \\ &= \mathbb{P}_\mu\left(T^2 > k\right) \\ &= 1 - \mathbb{P}_\mu\left(T^2 \le k\right) \\ &= 1 - F(\nu k; \delta_\mu, m, m-n) \end{aligned} \tag{26}$$

(5) Analyse der Testgütefunktion

Beispiele

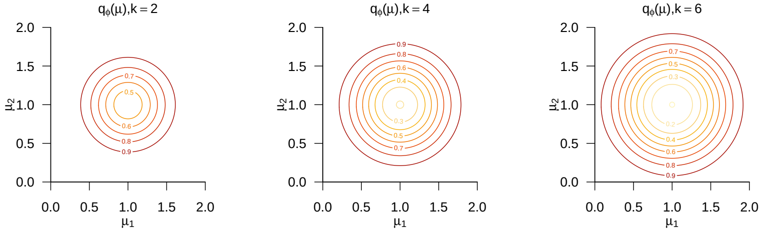
$$m := 2, n := 15, \Sigma := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mu_0 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

```
# Modellparameter
library(matlib)                                                          # R Paket
m            = 2                                                         # m
n            = 15                                                        # n
nu           = (n-m)/((n-1)*m)                                           # \nu
Sigma        = diag(m)                                                   # \Sigma = I_2
iSigma       = inv(Sigma)                                                # \Sigma^{-1}

# Testparameter
mu_0         = matrix(c(1,1), nrow = 2)                                  # \mu_0
k_all        = c(2,4,6)                                                  # k <-> \phi
n_k          = length(k_all)                                             # Anzahl k Werte/Tests

# q_\phi(\mu) Evaluation
mu_min       = 0                                                         # \mu_i Minimum
mu_max       = 2                                                         # \mu_i Maximum
mu_res       = 1e3                                                       # \mu_i Auflösung
mu_i         = seq(mu_min,mu_max,len = mu_res)                           # mu_i
q_phi        = array(dim = c(mu_res, mu_res, length(k_all)))             # q_\phi Array
for(k in 1:n_k){
    for(i in 1:mu_res){
        for(j in 1:mu_res){
            mu           = matrix(c(mu_i[i],mu_i[j]), nrow = 2)          # \mu
            delta_mu     = n*t(mu - mu_0) %*% iSigma %*% (mu -mu_0)      # \delta_\mu
            q_phi[i,j,k] = 1 - pf(nu*k_all[k], m, n-m, delta_mu)}}}      # q_\phi(\mu)
```

(5) Analyse der Testgütefunktion

Beispiele

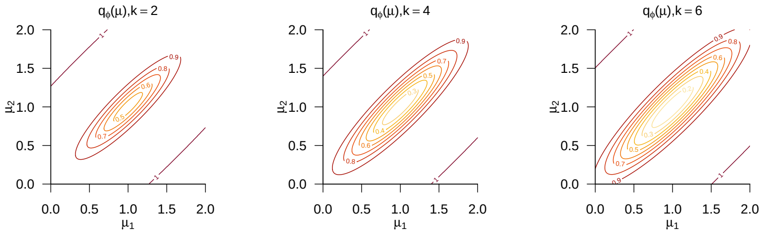
$$m := 2, n := 15, \Sigma := \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix}, \mu_0 := \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix}$$



(5) Analyse der Testgütefunktion

Beispiele

$$
m := 2, n := 15, \Sigma := \begin{pmatrix} 1.0 & 0.9 \\ 0.9 & 1.0 \end{pmatrix}, \mu_0 := \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix}
$$



(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

$$k_{\alpha_0} := \nu^{-1} F^{-1}\left(1 - \alpha_0; m, n - m\right) \tag{27}$$

wobei $\nu := (n-m)/((n-1)m)$ und $F^{-1}(\cdot; m, n-m)$ die inverse KVF der f-Verteilung mit Freiheitsgradparametern m und $n - m$ ist.

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 wird zunächst an, dass für $\mu = \mu_0$ gilt, dass

$$q_\phi(\mu_0) = 1 - F(\nu k; m, n-m, \delta_{\mu_0}) = 1 - F(\nu k; m, n-m, 0) = 1 - F(\nu k; m, n-m) \tag{28}$$

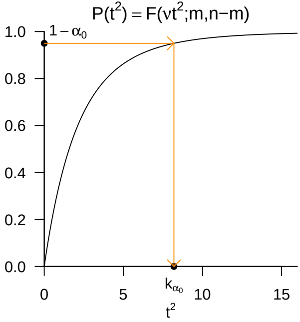
wobei$F(\nu k; \delta, m, n-m)$ und $F(\nu k; m, n-m)$ die KVF der nichtzentralen f-Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparametern m und $n-m$ sowie der f-Verteilung mit Freiheitsgradparametern m und $n-m$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

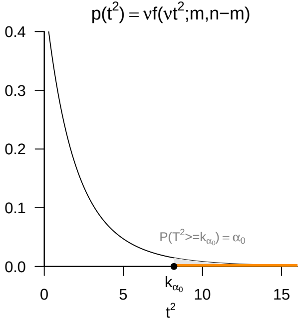
$$\begin{aligned}
q_\phi(\mu_0) &= 1 - F(\nu k_{\alpha_0}; m, n-m) \\
&= 1 - F\left(\nu\nu^{-1}F^{-1}\left(1-\alpha_0; m, n-m\right); m, n-m\right) \\
&= 1 - F\left(F^{-1}\left(1-\alpha_0; m, n-m\right); m, n-m\right) \\
&= 1 - (1-\alpha_0) = \alpha_0
\end{aligned} \tag{29}$$

Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \le \alpha_0$ ist 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 $k_{\alpha_0} := \nu^{-1}F^{-1}\left(1-\alpha_0; m, n-m\right)$ mit $m = 2, n = 15$ und $\alpha_0 := 0.05$





(6) Testumfangkontrolle

```r
# Modellparameter
m          = 2                                                            # Dimensionalität der Zufallsvektoren/Daten
n          = 15                                                           # Anzahl der Datenpunkte
nu         = (n-m)/(m*(n-1))                                              # Parameter
mu_0       = matrix(c(1,1) , nrow = 2)                                    # H0 Hypothesenparameter
mu         = mu_0                                                         # w.a.u. Erwartungswertparameter bei Zutreffen von H0
Sigma      = matrix(c(0.5,0.3, 0.3,0.5), nrow = 2, byrow = TRUE)          # wahrer, aber unbekannter, Kovarianzmatrixparameter

# Testparameter
alpha_0    = 0.05                                                         # Signifikanzlevel
k_alpha_0 = (1/nu)*qf(1-alpha_0, m,n-m)                                   # kritischer Wert

# Simulation der Testumfangkontrolle
library(MASS)                                                             # R Paket für multivariate Normalverteilungen
library(matlib)                                                           # R Paket für Matrizenrechnung
nsim   = 1e5                                                              # Testentscheidungsarray
phi    = rep(NaN,nsim)                                                    # Testentscheidungsarray
j_n    = matrix(rep(1,n), nrow = n)                                       # 1_n
I_n    = diag(n)                                                          # I_n
J_n    = matrix(rep(1,n^2), nrow = n)                                     # 1_{nn}
for(s in 1:nsim){                                                         # Simulationsiterationen
    Y      = t(mvrnorm(n,mu,Sigma))                                       # Y_i \sim N(\mu,\Sigma), i = 1,...,n
    Y_bar  = (1/n)*(Y %*% j_n)                                            # Stichprobenmittel
    C      = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))                   # Stichprobenkovarianzmatrix
    T2     = n*t(Y_bar - mu_0) %*% inv(C) %*% (Y_bar - mu_0)              # T^2 Statistik
    if(T2 > k_alpha_0){                                                  # Test 1_{T^2 >= k_alpha_0}
        phi[s] = 1                                                        # Ablehnen von H_0
    } else {
        phi[s] = 0                                                        # Nicht Ablehnen von H_0
    }
}
cat("\nKritischer Wert                = ", k_alpha_0,
    "\nGeschätzter Testumfang alpha = ", mean(phi))                       # Ausgabe
```

```
Kritischer Wert               =  8.2
Geschätzter Testumfang alpha =  0.0494
```

(6) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz $y_1, ..., y_n$ eine Realisation von $Y_1, ..., Y_n \sim N(\mu, \Sigma)$ mit unbekannten Parametern $\mu \in \mathbb{R}^m$ und $\Sigma \in \mathbb{R}^{m \times m}$p.d. ist.
- Man möchte entscheiden ob für ein $\mu_0 \in \mathbb{R}^m$ eher $H_0 : \mu = \mu_0$ oder $H_1 : \mu \neq \mu_0$ zutrifft.
- Man wählt ein Signifikanzlevel α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0}. Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05, m = 2$ und $n = 15$, also Freiheitsgradparametern 2 und 13, dass $k_{0.05} = \nu^{-1} F^{-1}(1-0.05; 2, 13) \approx 8.2$.
- Anhand von $m, n, \mu_0, \bar{Y}$ und C berechnet man die Realisierung der Einstichproben-T^2-Teststatistik
$$T^2 := n(\bar{Y} - \mu_0)^T C^{-1} (\bar{Y} - \mu_0) \tag{30}$$
- Wenn T^2 größer als k_{α_0} ist, lehnt man die Nullhypothese ab, andernfalls nicht.
- 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 vorliegenden Wert der Teststatistik ablehnen würde.
- Bei $T^2 = t^2$ würde H_0 für jedes α_0 mit $t^2 \geq \nu^{-1} F^{-1}(1-\alpha_0; m, n-m)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt

$$\alpha_0 \geq \mathbb{P}\left(T^2 \geq t^2\right) \tag{31}$$

- Das kleinste $\alpha_0 \in [0,1]$ mit $\alpha_0 \geq \mathbb{P}\left(T^2 \geq t^2\right)$ ist dann $\alpha_0 = \mathbb{P}(T^2 \geq t^2)$, also folgt

$$\text{p-Wert } = \mathbb{P}\left(T^2 \geq t^2\right) = 1 - F(\nu t^2; m, n-m) \tag{32}$$

- Zum Beispiel ergibt sich bei

- $m = 2$ und $n = 15$ der p-Wert für $t^2 = 7.00$ zu 0.071
- $m = 2$ und $n = 15$ der p-Wert für $t^2 = 9.00$ zu 0.040
- $m = 2$ und $n = 99$ der p-Wert für $t^2 = 7.00$ zu 0.035
- $m = 4$ und $n = 15$ der p-Wert für $t^2 = 7.00$ zu 0.304

Einstichproben-T^2-Tests

(7) p-Werte

Bestimmung des p-Wertes

- Es bleibt zu zeigen, dass gilt

$$
\begin{aligned}
t^2 &\geq \nu^{-1} F^{-1}(1-\alpha_0; m, n-m) \\
\Leftrightarrow \nu t^2 &\geq F^{-1}(1-\alpha_0; m, n-m) \\
\Leftrightarrow \alpha_0 &\geq \mathbb{P}\left(T^2 \geq t^2\right)
\end{aligned}
\tag{33}
$$

- Dies aber folgt aus

$$
\begin{aligned}
t^2 &\geq \nu^{-1} F^{-1}(1-\alpha_0; m, n-m) \\
\nu t^2 &\geq F^{-1}(1-\alpha_0; m, n-m) \\
F(\nu t^2; m, n-m) &\geq F\left(F^{-1}(1-\alpha_0; m, n-m); m, n-m\right) \\
F(\nu t^2; m, n-m) &\geq 1-\alpha_0 \\
\mathbb{P}\left(T^2 \leq t^2\right) &\geq 1-\alpha_0 \\
\alpha_0 &\geq 1-\mathbb{P}\left(T^2 \leq t^2\right)
\end{aligned}
\tag{34}
$$

Anwendungsbeispiel

```
# R Pakete
library(foreign)                                                  # Dateneinlesen
library(matlib)                                                   # Matrixalgebra

# Datenpräprozessierung
fname     = file.path(getwd(), "10_Daten", "studienerfolg.sav") # Dateinamen
D         = read.spss(fname, to.data.frame = T)                  # Dateneinlesen
Y         = rbind(D$X1[D$Gruppe == "ungenügend"],               # Y_{i_1} IQ Test Score
                  D$X2[D$Gruppe == "ungenügend"])                # Y_{i_2} Math Test Score

# Testparameter
m         = nrow(Y)                                               # Dimensionalität der Zufallsvektoren/Daten
n         = ncol(Y)                                               # Anzahl der Datenpunkte
nu        = (n-m)/(m*(n-1))                                       # Parameter
mu_0      = matrix(c(60,60) , nrow = 2)                           # H0 Hypothesenparameter ("Normwert")
alpha_0   = 0.05                                                  # Signifikanzlevel
k_alpha_0 = (1/nu)*qf(1-alpha_0,m,n-m)                            # kritischer Wert

# Testevaluation
j_n       = matrix(rep(1,n), nrow = n)                            # 1_n
I_n       = diag(n)                                               # I_n
J_n       = matrix(rep(1,n^2), nrow = n)                          # 1_{nn}
Y_bar     = (1/n)*(Y %*% j_n)                                     # Stichprobenmittel
C         = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))            # Stichprobenkovarianzmatrix
T2        = n*t(Y_bar - mu_0) %*% inv(C) %*% (Y_bar - mu_0)       # T^2 Statistik
if(T2 > k_alpha_0){                                              # Test 1_{T^2 >= k_alpha_0}
    phi = 1                                                       # Ablehnen von H_0
} else {
    phi = 0                                                       # Nicht Ablehnen von H_0
}
p         = 1 - pf(nu*T2,m,n-m)                                   # p-Wert
```

Einstichproben-T^2-Tests

Anwendungsbeispiel

```
# Ausgabe
cat("Y_bar   = ", Y_bar,
    "\nC       = ", C,
    "\nT^2     = ", T2,
    "\nalpha_0 = ", alpha_0,
    "\nk       = ", k_alpha_0,
    "\nphi     = ", phi,
    "\np       = ", p)
```

```
Y_bar   =  51.9 47.4
C       =  98.2 -4.11 -4.11 319
T^2     =  17.8
alpha_0 =  0.05
k       =  8.2
phi     =  1
p       =  0.00482
```

Anwendungsbeispiel mit MVTests::OneSampleHT2()

```
library(MVTests)                                   # R Pakete
Y     = D[1:15,2:3]                                # Dataframe von Interesse
phi   = OneSampleHT2(Y, mu_0, alpha_0)             # Einstichproben-T^2-Test

# Ausgabe
cat("Y_bar   = "  , phi$Descriptive[2,],
    "\nT^2     = ", phi$HT2,
    "\nalpha_0 = ", phi$alpha,
    "\nk       = " ,phi$F,
    "\np       = ", phi$p.value)
```

```
Y_bar   =  51.9 47.4
T^2     =  17.8
alpha_0 =  0.05
k       =  8.27
p       =  0.00482
```

(8) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R}^m \to [0,1], \mu \mapsto q_\phi(\mu) := 1 - F(\nu k; \delta_\mu, m, n-m) \tag{35}$$

bei kontrolliertem Testumfang, also für

$$k_{\alpha_0} := \nu^{-1} F^{-1}(1-\alpha_0; m, n-m) \tag{36}$$

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

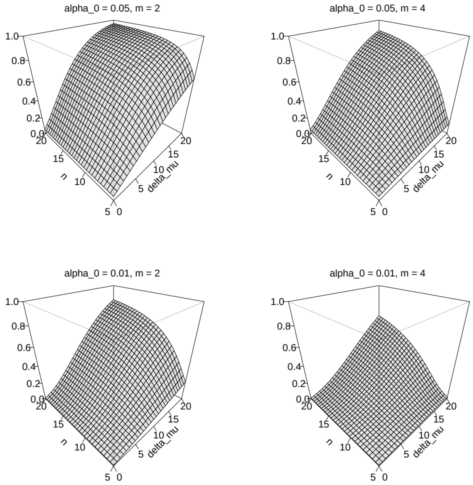
$$\pi : \mathbb{R} \times \mathbb{N} \to [0,1], (\delta_\mu, n) \mapsto \pi(\delta_\mu, n) := 1 - F(\nu k_{\alpha_0}; \delta_\mu, m, n-m) \tag{37}$$

Bei festgelegtem α_0 hängt die Powerfunktion des Einstichproben-T^2-Tests also vom unbekannten Wert δ_μ, von der Datendimensionalität m und von der Stichprobengröße n ab. Wir evaluieren und visualisieren diese Abhängigkeiten untenstehend.

(8) Analyse der Powerfunktion

```
# Szenariospezifikationen
a_0_all   = c(0.05,0.01)                                             # \alpha_0 raum
d_mu_min  = 0                                                        # \delta_\mu Minimum
d_mu_max  = 20                                                       # \delta_\mu Maximum
d_mu_res  = 30                                                       # \delta_\mu Auflösung
d_mu_all  = seq(d_mu_min, d_mu_max, len = d_mu_res)                  # \delta_\mu d Raum
n_min     = 5                                                        # n Minimum
n_max     = 20                                                       # n Maximum
n_res     = 30                                                       # n Auflösung
n_all     = seq(n_min,n_max, len = n_res)                            # n Raum
m_all     = c(2,4)                                                   # m Raum

# Evaluation der Powerfunktion
pi        = array(dim = c(d_mu_res, n_res, 2,2))                     # Powerfunktionsarray
for (a in 1:length(a_0_all)){
    for (l in 1:length(m_all)){                                      # m Iterationen
        for(i in 1:length(d_mu_all)){                                # \delta_\mu Iterationen
            for(j in 1:length(n_all)){                               # n Iterationen
                m           = m_all[l]                               # Datendimensionalität
                n           = n_all[j]                               # Stichprobenumfang
                d_mu        = d_mu_all[i]                            # wahrer, aber unbekannter, Parameter
                nu          = (n-m)/(m*(n-1))                        # Parameter
                alpha_0     = a_0_all[a]                             # Signifikanzlevel
                k_alpha_0   = (1/nu)*qf(1-alpha_0,m,n-m)             # kritischer Wert
                pi[i,j,l,a] = 1 - pf(nu*k_alpha_0, m, n-m, d_mu)}}}} # Powerfunktionswert
```

(8) Analyse der Powerfunktion



(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, Parameterwert $\delta_\mu = n(\mu - \mu_0)^T \Sigma^{-1}(\mu - \mu_0)$ ab.

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

Generell wird folgendes Vorgehen favorisiert

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

(7) Analyse der Powerfunktion

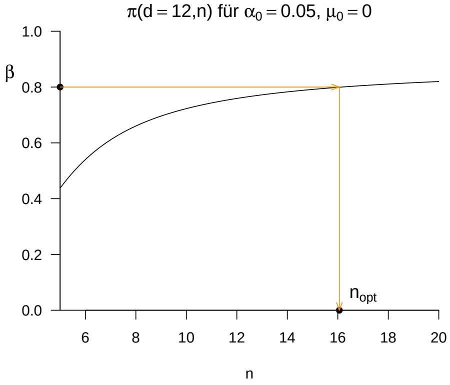
Praktisches Vorgehen

```
# Szenariospezifikation
n_min      = 5                                              # n Minimum
n_max      = 20                                             # n Maximum
n_res      = 1e2                                            # n Auflösung
n          = seq(n_min,n_max, len = n_res)                  # n Raum
alpha_0    = 0.05                                           # Signifikanzlevel

# Poweranalyse
m          = 2                                              # Datendimensionalität
d_mu_fix   = 12                                             # fester Nichtzentralitätsparameter
nu         = (n-m)/(m*(n-1))                                # Parameter
k_alpha_0  = (1/nu)*qf(1-alpha_0,m,n-m)                     # kritischer Wert
pi_n       = 1 - pf(nu*k_alpha_0, m, n-m, d_mu_fix)         # Powerfunktionswert
beta       = 0.8                                            # gewünschter Powerfunktionswert
i          = 1                                              # Indexinitialisierung
n_min      = NaN                                            # minimales n Initialisierung
while(pi_n[i] < beta){                                      # Solange \pi(\delta_\mu*,n) < \beta
    n_min = n[i]                                            # Aufnahme des minimal nötigen ns
    i     = i + 1                                           # und Erhöhung des Indexes
}
cat("Minimal nötiges n =", ceiling(n_min))                  # Ausgabe
```

```
Minimal nötiges n = 17
```

(7) Analyse der Powerfunktion

Praktisches Vorgehen



Zweistichproben-T^2-Tests

Zweistichproben-T^2-Tests **bei unabhängigen Stichproben**

Anwendungsszenario

- **Zwei Stichproben** experimenteller Einheiten.
- Annahme unabhängiger und identisch nach $N(\mu_1, \Sigma_1)$ und $N(\mu_2, \Sigma_2)$ verteilter Daten.
- $\mu_1, \Sigma_1, \mu_2, \Sigma_2$ unbekannt.
- Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ_1 und μ_2 beabsichtigt.

Anwendungsbeispiele

- Gruppenvergleich von BDI und Glukokortikoid Daten
    - $\mu_1 \neq \mu_2$ als Evidenz für multivariate Gruppenunterschiede
- Gruppenvergleich von Testdaten bei erfolgreichen vs. nicht-erfolgreichem Studienabschluss
    - $\mu_1 \neq \mu_2$ als Evidenz für multivariate Gruppenunterschiede

Mögliche Modellszenarien

- Annahme identischer Kovarianzmatrixparameter
- Annahme eines bekannten Varianzverhältnisses
- Keine Annahmen zu Varianzen

Mögliche Hypothesenszenarien

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

Hier betrachtetes Modellszenario

- Annahme identischer Kovarianzmatrixparameter
- Annahme eines bekannten Varianzverhältnisses
- Keine Annahmen zu Varianzen

Hier betrachtetes Hypothesenszenario

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

Zweistichproben-T^2-Tests

Anwendungsbeispiel

Gruppe	VP	IQ Test Score	Math Test Score
1	1	54	44
1	2	60	20
1	3	67	36
1	4	41	39
1	5	66	57
1	6	51	28
1	7	51	46
1	8	37	46
1	9	57	54
1	10	47	12
1	11	50	67
1	12	42	63
1	13	60	64
1	14	36	64
1	15	60	71
2	1	71	41
2	2	65	28
2	3	67	76
2	4	68	54
2	5	75	33
2	6	71	82
2	7	68	64
2	8	63	72
2	9	48	54
2	10	53	86
2	11	62	71
2	12	69	25
2	13	67	72
2	14	74	92
2	15	76	75

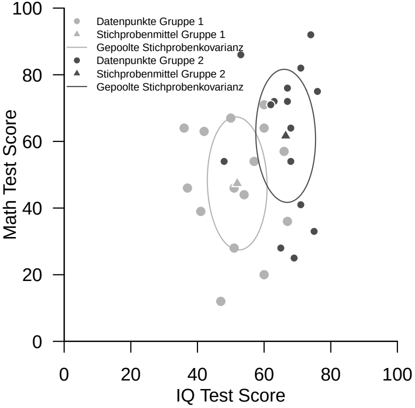
Unterschiede zwischen den (IQ Test Score, Math Test Score) Erwartungswertparametern zwischen Gruppen?

Anwendungsbeispiel

```
# Deskriptivstatistik
library(foreign)                                                          # R Paket
library(ellipse)                                                          # R Paket
library(matlib)                                                           # R Paket
fname     = file.path(getwd(), "10_Daten", "studienerfolg.sav")           # Dateiname
D         = read.spss(fname, to.data.frame = T)                           # Dateneinlesen
m         = 2                                                             # Datendimensionalität
n         = 15                                                            # Datenpunkte pro Gruppe
n_1       = n                                                             # Datenpunkte Gruppe 1
n_2       = n                                                             # Datenpunkte Gruppe 2
j_n       = matrix(rep(1,n), nrow = n)                                    # 1_n
I_n       = diag(n)                                                       # I_n
J_n       = matrix(rep(1,n^2), nrow = n)                                  # 1_{nn}
Y_1       = t(as.matrix(D[D$Gruppe == "ungenügend", 2:3]))                # Daten Gruppe 1
Y_2       = t(as.matrix(D[D$Gruppe == "gut", 2:3]))                       # Daten Gruppe 2
Y_1_bar   = (1/n)*(Y_1 %*% j_n)                                           # Stichprobenmittel
Y_2_bar   = (1/n)*(Y_2 %*% j_n)                                           # Stichprobenmittel
C_1       = (1/(n_1-1))*(Y_1 %*% (I_n-(1/n)*J_n) %*% t(Y_1))              # Stichprobenkovarianzmatrix
C_2       = (1/(n_2-1))*(Y_2 %*% (I_n-(1/n)*J_n) %*% t(Y_2))              # Stichprobenkovarianzmatrix
C         = ((n_1-1)*C_1+(n_2-1)*C_2)/(n_1+n_2-2)                         # Gepoolte Stichprobenkovarianzmatrix
D         = t(Y_1_bar - Y_2_bar) %*% inv(C) %*% (Y_1_bar - Y_2_bar)       # Mahalanobis Distanz

# Ausgabe
cat("Y_1_bar  =", Y_1_bar,
    "\nY_2_bar  =", Y_2_bar,
    "\nD        =", D)
```

```
Y_1_bar  = 51.9 47.4
Y_2_bar  = 66.5 61.7
D        = 3.33
```

Anwendungsbeispiel



Gliederung (vgl. (12) Hypothesentests)

(1) Statistisches Modell in klassischer Form

(2) Statistisches Modell in generativer Form

(3) Testhypothesen, Teststatistik, Test

(4) Analyse der Teststatistik

(5) Analyse der Testgütefunktion

(6) Testumfangkontrolle

(7) p-Werte

Zur Wiederholung des univariaten Falls, siehe (14) Zweistichproben-T-Tests

Zweistichproben-T^2-Tests

(1) Statistisches Modell in klassischer Form

$Y_{11}, ..., Y_{1n_1} \sim N(\mu_1, \Sigma)$ sei eine Stichprobe eines multivariaten Normalverteilungsmodells mit unbekanntem Erwartungswertparameter $\mu_1 \in \mathbb{R}^m$ und unbekanntem Kovarianzmatrixparameter $\Sigma \in \mathbb{R}^{m \times m}$ p.d. $Y_{21}, ..., Y_{2n_2} \sim N(\mu_2, \Sigma)$ sei eine weitere Stichprobe eines multivariaten Normalverteilungsmodells mit unbekanntem Erwartungswertparameter $\mu_2 \in \mathbb{R}^m$ und unbekanntem Kovarianzmatrixparameter $\Sigma \in \mathbb{R}^{m \times m}$ p.d. Die Kovarianzmatrixparameter beider Stichproben werden also als identisch vorausgesetzt. Der Parameter von Interesse ist (μ_1, μ_2), der Parameterraum des Modells ist $\Theta := \mathbb{R}^m \times \mathbb{R}^m$.

(2) Statistisches Modell in generativer Form

Es sei

$$Y_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \Sigma) \text{ für } i = 1, 2 \text{ und } j = 1, ..., n_i \tag{38}$$

wobei

- i die Stichproben indiziert,
- j die experimentellen Einheiten indiziert,
- n_i die Stichprobengrößen sind,
- Y_{ij} beobachtbare Zufallsvariablen sind,
- $\mu_i \in \mathbb{R}^m$ feste Erwartungswertparameter der Stichprobenvariablen sind
- $\Sigma \in \mathbb{R}^{m \times m}$ p.d. der identische Kovarianzmatrixparameter über Stichproben, und
- ε_{ij} unabhängige multivariat-normalverteilte nicht-beobachtbare Zufallsvariablen sind.

Der Zusammenhang zwischen klassischer und generativer Modellform ergibt sich analog zum Einstichprobenfall.

(3) Testhypothesen, Teststatistik, Test

Wir betrachten die einfache Nullhypothese

$$H_0 : \mu_1 = \mu_2 \Leftrightarrow \Theta_0 := \{(\mu_1, \mu_2) \in \mathbb{R}^m \times \mathbb{R}^m | \mu_1 = \mu_2\} \tag{39}$$

und die zusammengesetzte Alternativhypothese

$$H_1 : \mu_1 \neq \mu_2 \Leftrightarrow \Theta_1 := \{(\mu_1, \mu_2) \in \mathbb{R}^m \times \mathbb{R}^m | \mu_1 \neq \mu_2\} \tag{40}$$

Weiterhin betrachten die *Zweistichproben-T^2-Teststatistik*

$$T^2 := \frac{n_1 n_2}{n_1 + n_2} \left(\bar{Y}_1 - \bar{Y}_2\right)^T C^{-1} \left(\bar{Y}_1 - \bar{Y}_2\right) \tag{41}$$

wobei für $i = 1, 2$ und respektiven Stichprobenkovarianzmatrizen C_1 und C_2

$$\bar{Y}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \text{ und } C := \frac{(n_1 - 1)C_1 + (n_2 - 1)C_2}{n_1 + n_2 - 2} \tag{42}$$

die Stichprobenmittel und die *gepoolte Stichprobenkovarianzmatrix*, respektive, bezeichnen.

- T^2 ist die mit den Stichprobenumfängen skalierte Mahalanobis Distanz von $\bar{Y}_1$ und $\bar{Y}_2$ hinsichtlich C.
- Größere T^2 Werte ergeben sich für größere Abstände, geringere Kovarianzen und größere Stichprobenumfänge.

Schließlich definieren wir den kritischen Wert-basierten Test

$$\phi(X) := 1_{\{T^2 \geq k\}}. \tag{43}$$

(4) Analyse der Teststatistik

Theorem (Verteilung der Zweistichproben-T^2 Statistik)

Für $i = 1, 2$ seien $Y_{i1}, ..., Y_{in_i} \sim N(\mu_i, \Sigma)$ mit $\mu_i \in \mathbb{R}^m$ und $\Sigma \in \mathbb{R}^{m \times m}$ p.d.,,

$$\nu := \frac{n_1 + n_2 - m - 1}{m(n_1 + n_2 - 2)} \tag{44}$$

und die Zweistichproben-T^2-Teststatistik sei definiert als

$$T^2 := \frac{n_1 n_2}{n_1 + n_2}(\bar{Y}_1 - \bar{Y}_2)^T C^{-1}(\bar{Y}_1 - \bar{Y}_2) \tag{45}$$

mit den Stichprobenmitteln $\bar{Y}_i$ und der gepoolten Stichprobenkovarianzmatrix C. Dann gilt

$$\nu T^2 \sim f(\delta, m, n_1 + n_2 - m - 1), \tag{46}$$

wobei $f(\delta, m, n_1 + n_2 - m - 1)$ die nichtzentrale f-Verteilung mit Nichtzentralitätsparameter

$$\delta := \frac{n_1 n_2}{n_1 + n_2}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) \tag{47}$$

sowie mit Freiheitsgradparametern m und $n_1 + n_2 - m - 1$ bezeichnet.

Bemerkungen

- Für einen Beweis verweisen wir auf Anderson (2003).
- Für $\mu_1 = \mu_2$ und damit $\delta = 0$ entspricht $f(\delta, m, n_1 + n_2 - m - 1)$ der f-Verteilung $f(m, n_1 + n_2 - m - 1)$.

(4) Analyse der Teststatistik

Theorem (WDF und KVF der Zweistichproben-T^2 Statistik)

Im Zweistichproben-T^2-Testszenario sei

$$\nu := \frac{n_1 + n_2 - m - 1}{m(n_1 + n_2 - 2)} \tag{48}$$

Dann ist eine WDF der Zweistichproben-T^2 Statistik gegeben durch

$$p_{T^2} : \mathbb{R}_{\geq 0} \to \mathbb{R}, t^2 \mapsto p_{T^2}(t^2) := \nu f(\nu t^2; \delta, m, n_1 + n_2 - m - 1) \tag{49}$$

und eine KVF der Zweistichproben-T^2 Statistik ist gegeben durch

$$P_{T^2} : \mathbb{R}_{\geq 0} \to \mathbb{R}, t^2 \mapsto P_{T^2}(t^2) := Ft(\nu t^2; \delta, m, n_1 + n_2 - m - 1) \tag{50}$$

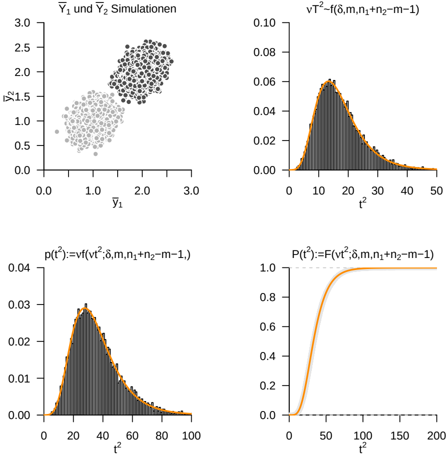
Bemerkung

- Der Beweis erfolgt in Analogie zum Einstichproben-T^2-Testszenario.

(4) Analyse der Teststatistik

```
# Modellparameter
m     = 2                                                   # Dimensionalität der Zufallsvektoren/Daten
n     = 15                                                  # Anzahl Datenpunkte pro Stichprobe
n_1   = n                                                   # Anzahl Datenpunkte Stichprobe 1
n_2   = n                                                   # Anzahl Datenpunkte Stichprobe 1
mu_1  = matrix(c(1,1) , nrow = 2)                           # wahrer, aber unbekannter, Erwartungswertparameter
mu_2  = matrix(c(2,2) , nrow = 2)                           # wahrer, aber unbekannter, Erwartungswertparameter
Sigma = matrix(c(0.4,0.1, 0.1,0.4), nrow = 2, byrow = TRUE)      # wahrer, aber unbekannter, Kovarianzmatrixparameter

# Simulation
library(MASS)                                               # R Paket für multivariate Normalverteilungen
library(matlib)                                             # R Paket für Matrizenrechnung
nsim  = 1e4                                                 # Anzahl Simulationen/Datensatzrealisierungen
Y1b   = matrix(rep(NaN,m*nsim), nrow = 2)                   # Stichprobenmittelarray
Y2b   = matrix(rep(NaN,m*nsim), nrow = 2)                   # Stichprobenmittelarray
T2    = rep(NaN,nsim)                                       # T2 Statistik Array
j_n   = matrix(rep(1,n), nrow = n)                          # 1_n
I_n   = diag(n)                                             # I_n
J_n   = matrix(rep(1,n^2), nrow = n)                        # 1_{nn}
for(s in 1:nsim){                                           # Simulationsiterationen
    Y_1       = t(mvrnorm(n,mu_1,Sigma))                    # Y_1j \sim N(\mu_1,\Sigma), j = 1,...,n_1
    Y_2       = t(mvrnorm(n,mu_2,Sigma))                    # Y_2j \sim N(\mu_2,\Sigma), j = 1,...,n_2
    Y_1_bar   = (1/n)*(Y_1 %*% j_n)                         # Stichprobenmittel
    Y_2_bar   = (1/n)*(Y_2 %*% j_n)                         # Stichprobenmittel
    C_1       = (1/(n_1-1))*(Y_1 %*% (I_n-(1/n)*J_n) %*% t(Y_1)) # Stichprobenkovarianzmatrix
    C_2       = (1/(n_2-1))*(Y_2 %*% (I_n-(1/n)*J_n) %*% t(Y_2)) # Stichprobenkovarianzmatrix

    # Gepoolte Stichprobenkovarianzmatrix und T2 Statistik
    C         = ((n_1-1)*C_1+(n_2-1)*C_2)/(n_1+n_2-2)
    T2[s]     = (((n_1*n_2)/(n_1+n_2))*t(Y_1_bar-Y_2_bar) %*% inv(C) %*% (Y_1_bar-Y_2_bar))

    # Stichprobenmittel für Visualisierung
    Y1b[,s]   = Y_1_bar
    Y2b[,s]   = Y_2_bar
}
```

(4) Analyse der Teststatistik



(5) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

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

$$q_\phi : \mathbb{R}^m \times \mathbb{R}^m \to [0,1], \mu \mapsto q_\phi(\mu_1, \mu_2) := 1 - F(\nu k; \delta_{\mu_{1,2}}, m, n_1 + n_2 - m - 1), \quad (51)$$

wobei

$$\nu := \frac{n_1 + n_2 - m - 1}{m(n_1 + n_2 - 2)}, \quad \delta_{\mu_{1,2}} := \frac{n_1 n_2}{n_1 + n_2}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) \quad (52)$$

und $F(\cdot; \delta_{\mu_{1,2}}, m, n_1 + n_2 - m - 1)$ die KVF der nichtzentralen f-Verteilung mit Freiheitsgradparametern m und $n_1 + n_2 - m - 1$ sowie mit Nichtzentralitätsparameter $\delta_{\mu_{1,2}}$ bezeichnet.

Bemerkungen

- q_ϕ kann zur Bestimmung kritischer Werte für einen erwünschten Testumfang genutzt werden.
- q_ϕ kann zur Bestimmung der Testpower genutzt werden.

(5) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Tests im vorliegenden Testszenario ist definiert als

$$q_\phi : \mathbb{R}^m \times \mathbb{R}^m \to [0,1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) := \mathbb{P}_{\mu_1,\mu_2}(\phi = 1) \tag{53}$$

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

$$\nu := \frac{n_1 + n_2 - m - 1}{(m(n_1 + n_2 - 2)} \text{ und } \delta_{\mu_{1,2}} := \frac{n_1 n_2}{n_1 + n_2}(\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2) \tag{54}$$

gilt, dass

$$\nu T^2 \sim f(\delta_{\mu_{1,2}}, m, n_1 + n_2 - m - 1). \tag{55}$$

Der Ablehnungsbereich des betrachteten Tests ergibt sich als $A :=]k, \infty[$ Also ergibt sich

$$\begin{aligned}
q_\phi(\mu) &= \mathbb{P}_\mu(\phi = 1) \\
&= \mathbb{P}_\mu\left(T^2 \in \,]k, \infty[\right) \\
&= \mathbb{P}_\mu\left(T^2 > k\right) \\
&= 1 - \mathbb{P}_\mu\left(T^2 \le k\right) \\
&= 1 - F\left(\nu k; \delta_{\mu_{1,2}}, m, n_1 + n_2 - m - 1\right)
\end{aligned} \tag{56}$$

(5) Analyse der Testgütefunktion

Beispiele

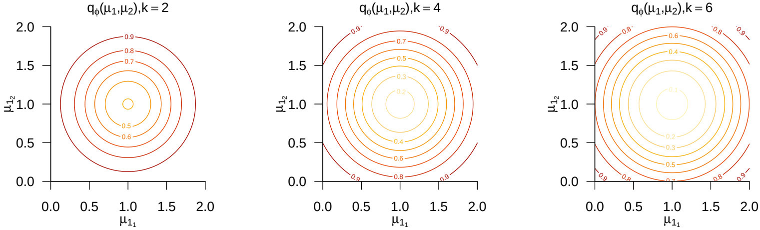
$$m := 2, n_1 := 15, n_2 := 15, \mu_2 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

```
# Modellparameter
library(matlib)                                                # R Paket
m           = 2                                                # m
n_1         = 15                                               # n_1
n_2         = 15                                               # n_2
nu          = (n_1+n_2-m-1)/(m*(n_1+n_2-2))                    # \nu
Sigma       = diag(m)                                          # \Sigma = I_2
iSigma      = inv(Sigma)                                       # \Sigma^{-1}
mu_2        = matrix(c(1,1), nrow = 2)                         # \mu_2

# Testparameter
k_all       = c(2,4,6)                                         # k <-> \phi
n_k         = length(k_all)                                    # Anzahl k Werte/Tests

# q_\phi(\mu) Evaluation
mu_min      = 0                                                # \mu_1 Minimum
mu_max      = 2                                                # \mu_1 Maximum
mu_res      = 1e3                                              # \mu_1 Auflösung
mu_i        = seq(mu_min,mu_max,len = mu_res)                  # mu_1i
q_phi       = array(dim = c(mu_res, mu_res, length(k_all)))    # q_\phi Array
for(k in 1:n_k){
    for(i in 1:mu_res){
        for(j in 1:mu_res){
            mu_1         = matrix(c(mu_i[i],mu_i[j]), nrow = 2)      # \mu_1
            delta        = (((n_1*n_2)/(n_1+n_2))*                   # \delta_{\mu_{1,2}}
                           t(mu_1 - mu_2) %*% iSigma %*% (mu_1-mu_2))
            q_phi[i,j,k] = 1-pf(nu*k_all[k],m,n_1+n_2-m-1,delta)}}}  # q_\phi(\mu)
```

(5) Analyse der Testgütefunktion

Beispiele

$$m := 2, n_1 := 15, n_2 := 15, \mu_2 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



(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

$$k_{\alpha_0} := \frac{1}{\nu} F^{-1} \left(1 - \alpha_0; m, n_1 + n_2 - m - 1\right) \tag{57}$$

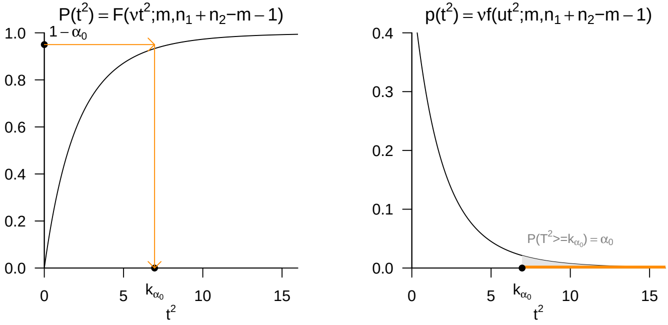
wobei $F^{-1}(\cdot; m, n_1 + n_2 - m - 1)$ die inverse KVF der f-Verteilung mit Freiheitsgradparametern m und $n_1 + n_2 - m - 1$ ist.

Bemerkung

- Der Beweis erfolgt analog zum Einstichprobenszenario.

(6) Testumfangkontrolle

Wahl von $k_{\alpha_0} := \frac{1}{\nu} F^{-1}(1-\alpha_0; m, n_1+n_2-m-1)$ mit $m = 2, n_1 = 15, n_2 = 15, \alpha_0 := 0.05$



(6) Testumfangkontrolle

```r
# Modellparameter bei Zutreffen von H0
m     = 2                                                # Dimensionalität der Zufallsvektoren/Dat
n     = 15                                               # Anzahl Datenpunkte pro Stichprobe
n_1   = n                                                # Anzahl Datenpunkte Stichprobe 1
n_2   = n                                                # Anzahl Datenpunkte Stichprobe 1
mu_1  = matrix(c(1,1) , nrow = 2)                        # wahrer, aber unbekannter, Erwartungswertparameter
mu_2  = mu_1                                             # wahrer, aber unbekannter, Erwartungswertparameter
Sigma = matrix(c(0.4,0.1, 0.1,0.4), nrow = 2, byrow = TRUE)  # wahrer, aber unbekannter, Kovarianzmatrixparameter

# Testparameter
nu          = (n_1+n_2-m-1)/(m*(n_1+n_2-2))              # \nu
alpha_0     = 0.05                                       # Signifikanzlevel
k_alpha_0   = (1/nu)*qf(1-alpha_0,m,n_1+n_2-m-1)         # kritischer Wert

# Simulation
library(MASS)                                            # R Paket für multivariate Normalverteilungen
library(matlib)                                          # R Paket für Matrizenrechnung
nsim  = 1e4                                              # Anzahl Simulationen/Datensatzrealisierungen
Y1b   = matrix(rep(NaN,m*nsim), nrow = 2)                # Stichprobenmittelarray
Y2b   = matrix(rep(NaN,m*nsim), nrow = 2)                # Stichprobenmittelarray
phi   = rep(NaN,nsim)                                    # phi Array
j_n   = matrix(rep(1,n), nrow = n)                       # 1_n
I_n   = diag(n)                                          # I_n
J_n   = matrix(rep(1,n^2), nrow = n)                     # 1_{nn}
for(s in 1:nsim){                                        # Simulationsiterationen
    Y_1     = t(mvrnorm(n,mu_1,Sigma))                   # Y_1j \sim N(\mu_1,\Sigma), j = 1,...,n_1
    Y_2     = t(mvrnorm(n,mu_2,Sigma))                   # Y_2j \sim N(\mu_2,\Sigma), j = 1,...,n_2
    Y_1_bar = (1/n)*(Y_1 %*% j_n)                        # Stichprobenmittel
    Y_2_bar = (1/n)*(Y_2 %*% j_n)                        # Stichprobenmittel
    C_1     = (1/(n_1-1))*(Y_1 %*% (I_n-(1/n)*J_n) %*% t(Y_1))   # Stichprobenkovarianzmatrix
    C_2     = (1/(n_2-1))*(Y_2 %*% (I_n-(1/n)*J_n) %*% t(Y_2))   # Stichprobenkovarianzmatrix
    C       = ((n_1-1)*C_1+(n_2-1)*C_2)/(n_1+n_2-2)      # Gepoolte Stichprobenkovarianzmatrix
    T2      = (((n_1*n_2)/(n_1+n_2))*                    # T^2 Statistik
              t(Y_1_bar-Y_2_bar) %*% inv(C) %*% (Y_1_bar-Y_2_bar))
    if(T2 > k_alpha_0){phi[s] = 1} else {phi[s] = 0}     # Evaluation von phi
}
cat("\nKritischer Wert              =", k_alpha_0,
    "\nGeschätzter Testumfang alpha =", mean(phi))
```

```
Kritischer Wert              = 6.96
Geschätzter Testumfang alpha = 0.0476
```

(6) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass zwei vorliegende Datensätze $y_{11}, ..., y_{1n_1}$ und $y_{21}, ..., y_{2n_2}$ Realisationen von m-dimensionalen Zufallsvektoren

$$Y_{11}, ..., Y_{1n_1} \sim N(\mu_1, \Sigma) \text{ und } Y_{21}, ..., Y_{2n_2} \sim N(\mu_2, \Sigma) \quad (58)$$

  mit unbekannten Parametern μ_1, μ_2, Σ sind.
- Man möchte entscheiden, ob eher $H_0 : \mu_1 = \mu_2$ oder $H_1 : \mu_1 \neq \mu_2$ zutrifft.
- Man wählt ein Signifikanzlevel α_0 und bestimmt den zugehörigen kritischen Wert k_{α_0}.
- Anhand von $m, n_1, n_2, \bar{Y}_1, \bar{Y}_2$ und der gepoolten Stichprobenstandardabweichung C berechnet man die Realisierung der Zweistichproben-T^2-Teststatistik t^2.
- Wenn t^2 größer als k_{α_0} ist lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab.
- Die oben entwickelte Theorie des Zweistichproben-T-^{2}Tests garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

Zweistichproben-T^2-Tests

(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 vorliegenden Wert der Teststatistik ablehnen würde.
- Bei $T^2 = t^2$ würde H_0 für jedes α_0 mit $t^2 \geq \frac{1}{\nu}F^{-1}(1-\alpha_0; m, n_1+n_2-m-1)$ abgelehnt werden. Für diese α_0 gilt, wie analog im Eintsichprobenstzenario gezeigt

$$\alpha_0 \geq \mathbb{P}\left(T^2 \geq t^2\right) \tag{59}$$

- Das kleinste $\alpha_0 \in [0,1]$ mit $\alpha_0 \geq \mathbb{P}\left(T^2 \geq t^2\right)$ ist dann $\alpha_0 = \mathbb{P}(T^2 \geq t^2)$, also folgt

$$\text{p-Wert } = \mathbb{P}\left(T^2 \geq t^2\right) = 1 - F(\nu t^2; m, n_1 + n_2 - m - 1) \tag{60}$$

- Zum Beispiel ergibt sich bei
  - $m = 2, n_1 = 15$ und $n_2 = 15$ der p-Wert für $t^2 = 5$ zu 0.11
  - $m = 2, n_1 = 15$ und $n_2 = 15$ der p-Wert für $t^2 = 7$ zu 0.05
  - $m = 2, n_1 = 99$ und $n_2 = 99$ der p-Wert für $t^2 = 5$ zu 0.09
  - $m = 4, n_1 = 15$ und $n_2 = 15$ der p-Wert für $t^2 = 7$ zu 0.21

Anwendungsbeispiel

```
# R Pakete
library(foreign)                                                  # Dateneinlesen
library(matlib)                                                   # Matrixalgebra

# Datenpräprozessierung
fname     = file.path(getwd(), "10_Daten", "studienerfolg.sav")   # Dateiname
D         = read.spss(fname, to.data.frame = T)                   # Dateneinlesen
Y_1       = t(as.matrix(D[D$Gruppe == "ungenügend", 2:3]))        # Daten Gruppe 1
Y_2       = t(as.matrix(D[D$Gruppe == "gut", 2:3]))               # Daten Gruppe 2

# Testparameter
m         = 2                                                     # Datendimensionalität
n         = 15                                                    # Datenpunkte pro Gruppe
n_1       = n                                                     # Datenpunkte Gruppe 1
n_2       = n                                                     # Datenpunkte Gruppe 2
nu        = (n_1+n_2-m-1)/(m*(n_1+n_2-2))                         # \nu
alpha_0   = 0.05                                                  # Signifikanzlevel
k_alpha_0 = (1/nu)*qf(1-alpha_0,m,n_1+n_2-m-1)                    # kritischer Wert

# Testevaluation
j_n      = matrix(rep(1,n), nrow = n)                             # 1_n
I_n      = diag(n)                                                # I_n
J_n      = matrix(rep(1,n^2), nrow = n)                           # 1_{nn}
Y_1_bar  = (1/n)*(Y_1 %*% j_n)                                    # Stichprobenmittel
Y_2_bar  = (1/n)*(Y_2 %*% j_n)                                    # Stichprobenmittel
C_1      = (1/(n_1-1))*(Y_1 %*% (I_n-(1/n)*J_n) %*% t(Y_1))       # Stichprobenkovarianzmatrix
C_2      = (1/(n_2-1))*(Y_2 %*% (I_n-(1/n)*J_n) %*% t(Y_2))       # Stichprobenkovarianzmatrix
C        = ((n_1-1)*C_1+(n_2-1)*C_2)/(n_1+n_2-2)                  # Gepoolte Stichprobenkovarianzmatrix
T2       = (((n_1*n_2)/(n_1+n_2))*                                # T^2 Statistik
            t(Y_1_bar-Y_2_bar) %*% inv(C) %*% (Y_1_bar-Y_2_bar))
if(T2 > k_alpha_0){phi = 1} else {phi = 0}                        # Test 1_{T^2 > k_alpha_0}
p         = 1 - pf(nu*T2,m,n_1+n_2-m-1)                           # p-Wert
```

Anwendungsbeispiel

```
# Ausgabe
cat("Y_1_bar = ", Y_1_bar,
    "\nY_2_bar = ", Y_2_bar,
    "\nC       = ", C,
    "\nT^2     = ", T2,
    "\nalpha_0 = ", alpha_0,
    "\nk       = ", k_alpha_0,
    "\nphi     = ", phi,
    "\np       = ", p)
```

```
Y_1_bar =  51.9 47.4
Y_2_bar =  66.5 61.7
C       =  78.7 -8.94 -8.94 390
T^2     =  25
alpha_0 =  0.05
k       =  6.96
phi     =  1
p       =  0.000181
```

Anwendungsbeispiel mit MVTests::TwoSamplesHT2()

```
library(MVTests)                                # R Pakete
Y     = D[c(1:15, 31:45), 2:3]                  # Dataframe von Interesse
G     = c(rep(1,15), rep(2,15))                 # Gruppenindikatoren
phi   = TwoSamplesHT2(Y,G, alpha_0)             # Zweistichproben-T^2-Test

# Ausgabe
cat("Y_1_bar = " , phi$Descriptive1[1,],
    "\nY_2_bar = ",  phi$Descriptive2[1,],
    "\nT^2     = ", phi$HT2,
    "\nalpha_0 = ", phi$alpha,
    "\nk       = " ,phi$F,
    "\np       = ", phi$p.value)
```

```
Y_1_bar =  51.9 47.4
Y_2_bar =  66.5 61.7
T^2     =  25
alpha_0 =  0.05
k       =  12.1
p       =  0.000181
```

Vorbemerkungen

Einstichproben-T^2-Tests

Zweistichproben-T^2-Tests

Univariates vs. multivariates Testen

Selbstkontrollfragen

Univariates vs. multivariates Testen

Gegeben sei ein Anwendungsszenario mit n Beobachtungen von m Variablen

Bei Durchführung von m univariaten Tests entsteht ein multiples Testproblem

- Induktion multipler Typ I und Typ II Fehlerraten (cf. Ostwald et al. (2019))
- Familywise Error Rate (FWER) $= \mathbb{P}(\geq 1$ Typ I Fehler$)$
- Bei unabhängigen Variablen bietet zur FWER Kontrolle die *Bonferroni Korrektur* an.
- Bei Durchführung eines multivariaten Tests entsteht kein multiples Testproblem.

Multivariate Tests beziehen Variablenkovarianzen explizit mit ein.

- Bei Durchführung von m univariaten Tests werden Variablenkorrelationen aktiv ignoriert.

Sollten m univariate Tests oder ein multivariater Test durchgeführt werden?

- Je nachdem, ob die Daten als multi- oder univariate Realisierung konzipiert werden.
- Je nachdem, welche geometrische Form des Annahmebereiches gewünscht ist.
- Prinzipiell sollte im wissenschaftlichen Diskurs überhaupt nicht getestet werden.

Wir betrachten in der Folge Simulationsszenarien mit $m := 2$.

Univariates vs. multivariates Testen

Univariate vs. multivariate Einstichproben-T$^{(2)}$-Tests mit $\Sigma = \sigma^2 I_2$

```r
# R Pakete
library(MASS)                                                       # R Paket für multivariate Normalverteilungen
library(matlib)                                                     # R Paket für Matrizenrechnung

# Modellparameter
m     = 2                                                           # Dimensionalität der Zufallsvektoren/Daten
n     = 15                                                          # Anzahl der Datenpunkte
mu    = matrix(c(0,0) , nrow = 2)                                   # Erwartungswertparameter
Sigma = matrix(c(1,0,0,1), nrow = 2, byrow = TRUE)                  # Kovarianzmatrixparameter

# Testparameter
alpha = 0.05                                                        # Signifikanzlevel
nu    = (n-m)/(m*(n-1))                                             # T^2-Test Parameter
k_T2  = (1/nu)*qf(1-alpha, m, n-m)                                  # T^2-Test kritischer Wert
k_Tu  = qt(1-(1/2)*alpha  ,n-1)                                     # T-Test kritischer Wert unkorrigiert
k_Tc  = qt(1-(1/2)*alpha/m,n-1)                                     # T-Test kritischer Wert Bonferonni korrigiert

# Simulation der Testumfangkontrolle
nsim  = 1e4                                                         # Anzahl Simulation
phi   = matrix(rep(NaN,nsim*5), nrow = 5)                           # Testentscheidungsarray
j_n   = matrix(rep(1,n), nrow = n)                                  # 1_n
I_n   = diag(n)                                                     # I_n
J_n   = matrix(rep(1,n^2), nrow = n)                                # 1_{nn}
for(s in 1:nsim){                                                   # Simulationsiterationen
    Y        = t(mvrnorm(n,mu,Sigma))                               # Y_i \sim N(\mu,\Sigma), i = 1,...,n
    Y_bar    = (1/n)*(Y %*% j_n)                                    # Stichprobenmittel
    C        = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))           # Stichprobenkovarianzmatrix
    phi[1,s] =  n*t(Y_bar) %*% inv(C) %*% (Y_bar)  > k_T2           # T^2-Test mit \mu_0
    for(i in 1:m){                                                  # T-Test Iterationen
      y_bar      = Y_bar[i]                                         # Stichprobenmittel
      sigma_hat  = sqrt(C[i,i])                                     # Stichprobenstandardabweichung
      phi[i+1,s] = abs(sqrt(n)*y_bar/sigma_hat)  > k_Tu             # Unkorrigierter T-Test mit \mu_0 = 0
      phi[i+3,s] = abs(sqrt(n)*y_bar/sigma_hat)  > k_Tc}}           # Korrigierter   T-Test mit \mu_0 = 0
```

Univariate vs. multivariate Einstichproben-T$^{(2)}$-Tests mit $\Sigma = \sigma^2 I_2$

```
Kritischer Wert T^2-Test                                         = 8.2
Kritischer Wert T-Test                                           = 2.14
Kritischer Wert T-Test Bonferroni                                = 2.51

Geschätzte Typ I Fehlerwahrscheinlichkeit T^2-Test               = 0.049

Geschätzte Typ I Fehlerwahrscheinlichkeit T-Test 1 unkorrigiert = 0.0474
Geschätzte Typ I Fehlerwahrscheinlichkeit T-Test 2 unkorrigiert = 0.0523
Geschätzte FWER T-Tests unkorrigiert                             = 0.0978

Geschätzte Typ I Fehlerwahrscheinlichkeit T-Test 1 Bonferroni   = 0.0245
Geschätzte Typ I Fehlerwahrscheinlichkeit T-Test 2 Bonferroni   = 0.0264
Geschätzte FWER T-Tests Bonferroni                               = 0.0506
```

Univariate vs. multivariate Einstichproben-T$^{(2)}$-Tests mit $\Sigma \neq \sigma^2 I_2$

```
Kritischer Wert T^2-Test                                        = 8.2
Kritischer Wert T-Test                                          = 2.14
Kritischer Wert T-Test Bonferroni                               = 2.51

Geschätzte Typ I Fehlerwahrscheinlichkeit T^2-Test              = 0.0445

Geschätzte Typ I Fehlerwahrscheinlichkeit T-Test 1 unkorrigiert = 0.0471
Geschätzte Typ I Fehlerwahrscheinlichkeit T-Test 2 unkorrigiert = 0.0465
Geschätzte FWER T-Tests unkorrigiert                            = 0.0675

Geschätzte Typ I Fehlerwahrscheinlichkeit T-Test 1 Bonferroni   = 0.0214
Geschätzte Typ I Fehlerwahrscheinlichkeit T-Test 2 Bonferroni   = 0.0226
Geschätzte FWER T-Tests Bonferroni                              = 0.0226
```

- Kovariabilität von Variablen reduziert die FWER.
- Die Bonferroni FWER Korrektur wird *konservativ*, also $\mathbb{P}(\geq 1$ Typ I Fehler$) < \alpha_0$.

Univariate vs. multivariate Einstichproben-Z$^{(2)}$-Tests

Zur Visualisierung von Stichprobenmittel und Testentscheidung bieten sich (nur) Z^2-Test an.

Z^2-Test $\approx$ T^2-Test mit als bekannt vorausgesetzter Kovarianzmatrix bei $Y_i \sim N(\mu, \Sigma)$.

- Einstichproben-Z^2-Teststatistik:

$$Z^2 := n(\bar{Y} - \mu_0)^T \Sigma^{-1} (\bar{Y} - \mu_0) \tag{61}$$

- Verteilung der Einstichproben-Z^2-Teststatistik bei $H_0 : \mu = \mu_0$:

$$Z^2 \sim \chi^2(m) \tag{62}$$

- Kritischer Wert für Testumfangkontrolle:

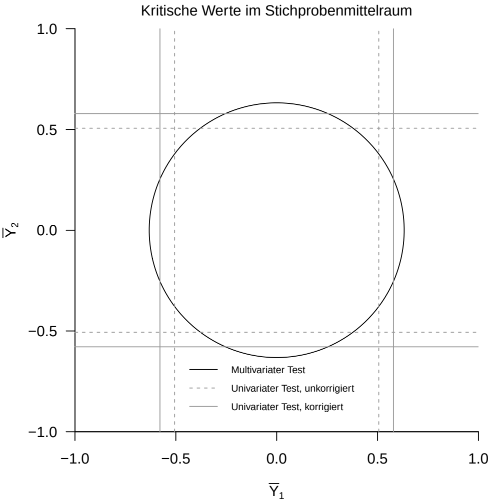
$$k_{\alpha_0} := \Xi^{2^{-1}}(1 - \alpha_0; m) \tag{63}$$

wobei $\Xi^{2^{-1}}$ die inversen KVF der χ^2 Verteilung bezeichnet.

Univariate vs. multivariate Einstichproben-Z$^{(2)}$-Tests mit $\Sigma = \sigma^2 I_2$

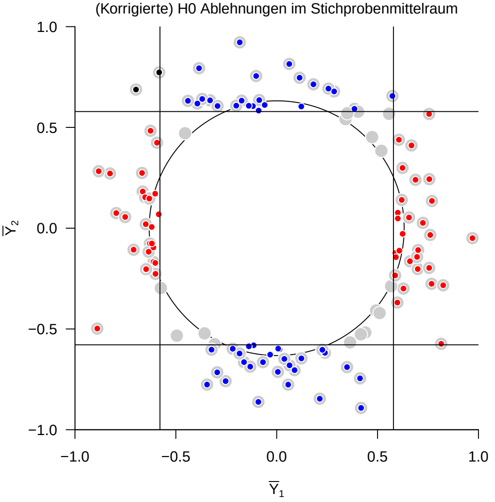
```
# R Pakete
library(MASS)                                          # R Paket für multivariate Normalverteilungen
library(matlib)                                        # R Paket für Matrizenrechnung

# Modellparameter
m     = 2                                              # Dimensionalität der Zufallsvektoren/Daten
n     = 15                                             # Anzahl der Datenpunkte
mu    = matrix(c(0,0) , nrow = 2)                      # Erwartungswertparameter
Sigma = matrix(c(1,0,0,1), nrow = 2, byrow = TRUE)     # Kovarianzmatrixparameter

# Testparameter
alpha = 0.05                                           # Signifikanzlevel
k_Z2  = qchisq(1-alpha, m)                             # Z^2-Test kritischer Wert
k_Zu  = qnorm(1-(1/2)*alpha)                           # Z-Test kritischer Wert unkorrigiert
k_Zc  = qnorm(1-(1/2)*alpha/m)                         # Z-Test kritischer Wert Bonferonni korrigiert

# Simulation der Testumfangkontrolle
nsim  = 2e3                                            # Anzahl Simulation
YB    = matrix(rep(NaN,nsim*2), nrow  = 2)             # Stichprobenmittelarray
phi   = matrix(rep(NaN,nsim*5), nrow  = 5)             # Testentscheidungsarray
j_n   = matrix(rep(1,n), nrow = n)                     # 1_n
I_n   = diag(n)                                        # I_n
J_n   = matrix(rep(1,n^2), nrow = n)                   # 1_{nn}
for(s in 1:nsim){                                      # Simulationsiterationen
    Y        = t(mvrnorm(n,mu,Sigma))                  # Y_i \sim N(\mu,\Sigma), i = 1,...,n
    YB[,s]   = (1/n)*(Y %*% j_n)                       # Stichprobenmittel
    phi[1,s] = (n*t(YB[,s])%*%inv(Sigma)%*%YB[,s]) > k_Z2  # T^2-Test mit \mu_0
    for(i in 1:m){                                     # T-Test Iterationen
      y_bar      = YB[i,s]                             # Stichprobenmittel
      sigma      = sqrt(Sigma[i,i])                    # Stichprobenstandardabweichung
      phi[i+1,s] = abs(sqrt(n)*y_bar/sigma)  > k_Zu    # Unkorrigierter T-Test mit \mu_0 = 0
      phi[i+3,s] = abs(sqrt(n)*y_bar/sigma)  > k_Zc}}  # Korrigierter   T-Test mit \mu_0 = 0
```

Univariate vs. multivariate Einstichproben-Z$^{(2)}$-Tests mit $\Sigma = \sigma^2 I_2$



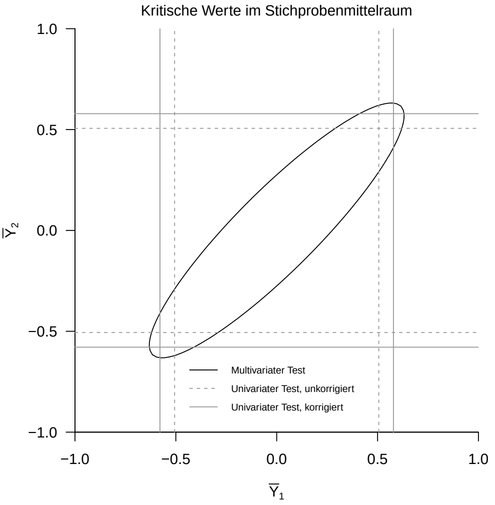
Univariates vs. multivariates Testen

Univariate vs. multivariate Einstichproben-Z$^{(2)}$-Tests mit $\Sigma = \sigma^2 I_2$



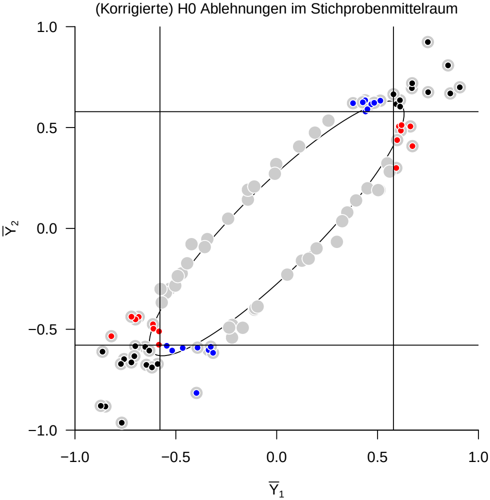
● $\phi(Y) = 1$, ● $\phi(Y_1) = 1$, ● $\phi(Y_2) = 1$, ● $\phi(Y_1) = 1$ und $\phi(Y_2) = 1$

Univariate vs. multivariate Einstichproben-Z$^{(2)}$-Tests mit $\Sigma \neq \sigma^2 I_2$



Univariates vs. multivariates Testen

Univariate vs. multivariate Einstichproben-Z$^{(2)}$-Tests mit $\Sigma = \sigma^2 I_2$



● $\phi(Y) = 1$, ● $\phi(Y_1) = 1$, ● $\phi(Y_2) = 1$, ● $\phi(Y_1) = 1$ und $\phi(Y_2) = 1$

Vorbemerkungen

Einstichproben-T^2-Tests

Zweistichproben-T^2-Tests

Univariates vs. multivariates Testen

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie das Modell und die Standardprobleme der Frequentistischen Inferenz.
2. Erläutern Sie die Standardannahmen Frequentistischer Inferenz.
3. Definieren Sie den Begriff der Mahalanobis Distanz.
4. Erläutern Sie den Unterschied zwischen einer f-Verteilung und einer nichtzentralen f-Verteilung.
5. Beschreiben Sie das Anwendungsszenario für einen Einstichproben-T^2-Test.
6. Geben Sie das statistische Modell eines Einstichproben-T^2-Test in klassischer Form an.
7. Geben Sie das statistische Modell eines Einstichproben-T^2-Test in generativer Form an.
8. Erläutern Sie das statistische Modell eines Einstichproben-T^2-Test in generativer Form.
9. Definieren Sie die Testhypothesen eines Einstichproben-T^2-Test mit einfacher Nullhypothese und zusammengesetzter Alternativhypothese.
10. Definieren Sie die Einstichproben-T^2-Teststatistik.
11. Erläutern Sie, wann die Einstichproben-T^2-Teststatistik hohe Werte annimmt.
12. Geben Sie die WDF und KVF der Einstichproben-T^2-Teststatistik an.
13. Geben Sie den kritischen Wert für einen Level-α_0-Einstichproben-T^2-Testmit einfacher Nullhypothese und zusammengesetzter Alternativhypothese und Testumfang α_0 an.
14. Erläutern Sie das praktische Vorgehen bei der Durchführung eines Einstichproben-T^2-Tests.
15. Definieren Sie den Begriff des p-Wertes.
16. Geben Sie den p-Wert für einen Einstichproben-T^2-Test mit einfacher Nullhypothese und zusammengesetzter Alternativhypothese an und erläutern Sie die Komponenten des entsprechenden Ausdrucks.
17. Definieren Sie die Powerfunktion eines Einstichproben-T^2-Test mit einfacher Nullhypothese und zusammengesetzter Alternativhypothese und erläutern Sie ihre Komponenten.

References

Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd ed. Wiley Series in Probability and Statistics. Hoboken, N.J: Wiley-Interscience.

Hotelling, Harold. 1931. "The Generalization of Student's Ratio." *The Annals of Mathematical Statistics* 2 (3): 360–78. https://doi.org/10.1214/aoms/1177732979.

Ostwald, Dirk, Sebastian Schneider, Rasmus Bruckner, and Lilla Horvath. 2019. "Power, Positive Predictive Value, and Sample Size Calculations for Random Field Theory-Based fMRI Inference." *BioRxiv: Doi.org/10.1101/613331*, April. https://doi.org/10.1101/613331.