Vorlesung Statistik II

Florian Wickelmaier

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Sommersemester 2026

Statistischer Hypothesentest

Selektionskriterien

- Unterschiedshypothese vs. Zusammenhangshypothese
- Art der Variablen (stetig, kategorial)
- Anzahl der Stichproben
- (Un-)Abhängigkeit der Stichproben
- Verteilungsannahmen
- Varianzhomogenität vs. -heterogenität

Signifikanz und Effektgröße

Statistische Signifikanz bedeutet nicht notwendig, dass ein bedeutsamer (inhaltlich relevanter) Effekt vorliegt

- Deshalb muss vor der Datenerhebung festgelegt werden, wie groß ein Effekt ist, wenn dieser gerade noch inhaltlich interessant sein soll

Beispiel

- Im Falle eines t-Tests für unabhängige Stichproben von gleicher Größe n und mit identischen Varianzen hängt die Signifikanz unter der Nullhypothese $\mu_x = \mu_y$ ab von der Teststatistik

$$T = \frac{\bar{X} - \bar{Y}}{S} \cdot \sqrt{n}$$

Beispiel

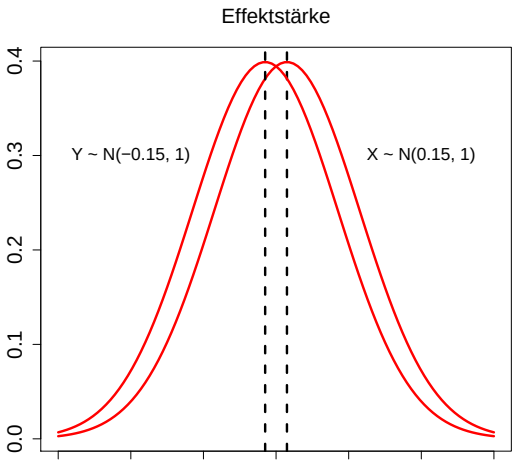
Der Wert der Teststatistik wächst für einen festen Effekt

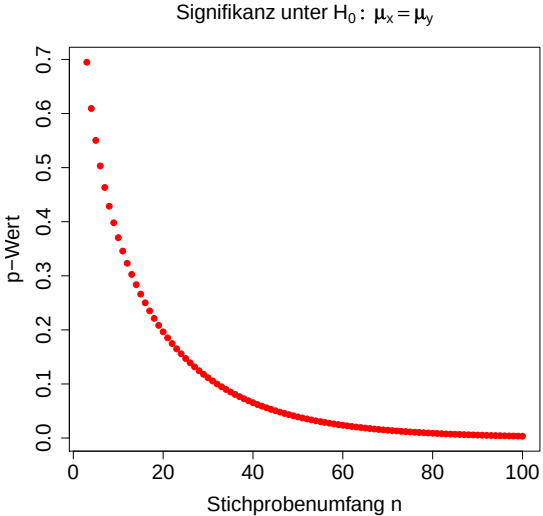
$$d = \frac{\mu_x - \mu_y}{\sigma}$$

(Cohens d) mit zunehmender Stichprobengröße

Obwohl die wachsende Stichprobengröße bei der Bestimmung des kritischen Quantils der t-Verteilung durch die Anzahl der Freiheitsgrade ($df = n - 2$) berücksichtigt wird, nimmt der p-Wert monoton ab (Ausdruck zunehmender Teststärke)

Beispiel


Effektstärke
Y ~ N(−0.15, 1)
X ~ N(0.15, 1)
0.0
0.1
0.2
0.3
0.4
−3
−2
−1
0
1
2
3


Signifikanz unter $H_0: \mu_x = \mu_y$
p−Wert
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
20
40
60
80
100
Stichprobenumfang n

Teststärke (Power)

Die Teststärke steigt ...

- ... mit wachsender Irrtumswahrscheinlichkeit α
- ... mit wachsender Effektgröße
- ... mit wachsendem Stichprobenumfang
- ... mit abnehmender Varianz
- ... bei einem einseitigen Test gegenüber einem zweiseitigen Test (bei einem Ergebnis in Richtung der H_1)

Beispiel

Evaluation von Behandlungseffekten

Unsere Therapieeffekte im Jahr 2008

Skala	Beginn der Therapie		Ende der Therapie		Signifikanztest		Effektstärke
	M	SD	M	SD	t-Wert	p-Wert	d-Wert
Patienten jeglicher Diagnose							
BSI-GSI Gesamtwert (N = 122)	1,32	0,55	0,89	0,69	7,60	< 0,01	0,69
BSI Skala Ängstlichkeit (N = 101)	1,64	0,73	0,99	0,80	8,22	< 0,01	0,85
BSI Skala Aggressivität/Feindseligkeit (N = 89)	1,45	0,61	0,91	0,75	6,47	< 0,01	0,79
Patienten mit depressiver Störung (N = 81)							
Beck-Depressions-Inventar (BDI)	24,4	8,1	13,5	10,5	10,70	< 0,01	1,18
Patienten mit Panikstörung/Agoraphobie							
Agoraphobic Cognitions Questionnaire ACQ (N = 16)	2,17	0,60	1,95	0,61	1,38	n.s.	0,36
Body Sensations Questionnaire BSQ (N = 18)	2,97	0,62	2,16	0,80	4,68	< 0,01	1,14
Mobilitätsinventar MI, Vermeidung in Begleitung (N = 16)	2,49	0,90	1,66	0,50	3,33	< 0,01	1,19
Mobilitätsinventar MI, Vermeidung alleine (N = 12)	2,57	0,91	1,49	0,37	3,98	< 0,01	1,69

nur Therapien, die im Evaluationsjahr 2008 abgeschlossen wurden; nur Patienten mit pathologischem Score bei Therapiebeginn;
Intention-to-Treat-Analyse, d.h. Patienten mit und ohne regulär abgeschlossene Therapie

aus: Jahresbericht 2008, Klinische Psychologie und Psychotherapie, Johannes-Gutenberg-Universität Mainz

Offene Fragen

Welche Anteile der beobachteten Effekte sind statistische Artefakte, die insbesondere durch Regression zur Mitte bedingt sind?

Kann man die durch Regression zur Mitte bedingten Effekte bei der statistischen Entscheidung berücksichtigen?

Verteilungsannahmen

Zur Beantwortung dieser Fragen werden wir uns im Folgenden exemplarisch auf das Vereinfachte Beck-Depressions-Inventar (BDI-V) beziehen

Schmitt, M., Altstötter-Gleich, C., Hinz, A., Maes, J., & Brähler, E. (2006). Normwerte für das Vereinfachte Beck-Depressions-Inventar (BDI-V) in der Allgemeinbevölkerung. *Diagnostica*, **52**, 51-59.

Die Werte liegen im Bereich 0 bis 100 mit den Kennwerten $M = 20.4$ und $SD = 14.2$ für die Normstichprobe ($N = 4494$)

Regression zur Mitte

Als "Regression zur Mitte" (eigentlich genauer "Regression zur größten Dichte") wird ein statistisches Artefakt bezeichnet

- Es werden zwei verbundene Messungen x und y betrachtet
- Liegt der Wert x über (bzw. unter) dem Mittelwert, dann wird der Wert y wahrscheinlich näher zum Mittelwert liegen als x
- Das Ausmaß dieser Regression zur Mitte wird durch eine große Abweichung des Werts x vom Mittelwert und einen geringen Zusammenhang beider Messungen begünstigt

Regression zur Mitte: Voraussetzungen

Die Voraussetzungen für das Auftreten des statistischen Artefakts sind im Beispiel gegeben

- Man hat die verbundenen Messungen x (Pretest) und y (Posttest)
- Die x-Werte werden (zum Teil deutlich) über dem Mittelwert in der Gesamtbevölkerung (Normstichprobe) liegen, was den Effekt begünstigt
- Die Reliabilität wird mit ca. $r = 0.9$ angegeben, ein hoher Zusammenhang, der dem Effekt grundsätzlich eher entgegenwirkt

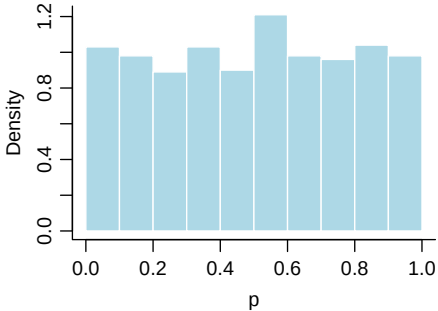
Regression zur Mitte: parametrisches Bootstrapping

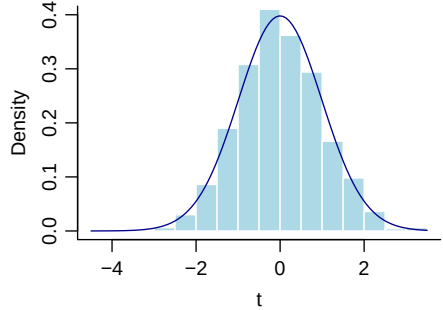
In den folgenden Simulationen der Effekte wird ein Verfahren zur Adjustierung der p-Werte angegeben, das die Regression zur Mitte in Rechnung stellt (Krause & Pinheiro, 2007)

- Simulation einer verbundenen Stichprobe (x, y) mit $N = 100$ aus einer bivariaten Normalverteilung mit den Normkennwerten des BDI-V ($M = 20.4$, $SD = 14.2$ und Korrelation $r = 0.9$)
- Unter der (hier zugrunde liegenden) Nullhypothese sind die für 1000 Simulationen erhaltenen Statistiken des entsprechenden t-Tests für verbundene Stichproben annähernd standardnormalverteilt und die zugehörigen p-Werte annähernd gleichmäßig verteilt auf dem Intervall $(0, 1)$
- Das Erzeugen der Stichprobenverteilung mittels einer auf den Modellannahmen basierenden Simulation wird *parametrisches Bootstrapping* genannt

Regression zur Mitte: Simulation

Stichprobenverteilung ohne Pretest-Einschlusskriterium





Regression zur Mitte: Simulation

Die Gleichverteilung der p-Werte drückt die Tatsache aus, dass bei gegebenem Fehler 1. Art α unter der Gültigkeit der Nullhypothese ein Anteil von $\alpha \cdot 100\%$ der erhaltenen p-Werte erwartet wird, der kleiner oder gleich α ist

- Es gilt $P(p \leq \alpha) = \alpha$ und damit ist die zugehörige Dichtefunktion von p als Ableitung dieser Verteilungsfunktion dann konstant

Den für einen Datensatz mit diesen Verteilungsannahmen resultierenden p-Wert p_0 erhält man demnach als Anteil der p-Werte p_i der Simulationen $i = 1, \ldots, 1000$, für die $p_i \leq p_0$ gilt ($\approx p_0$)

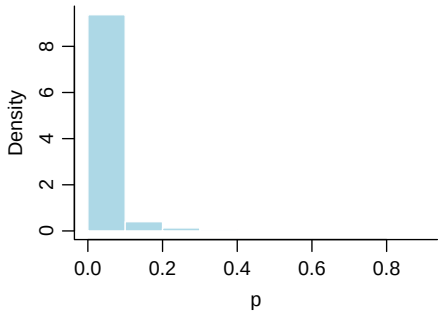
Regression zur Mitte: adjustierter p-Wert

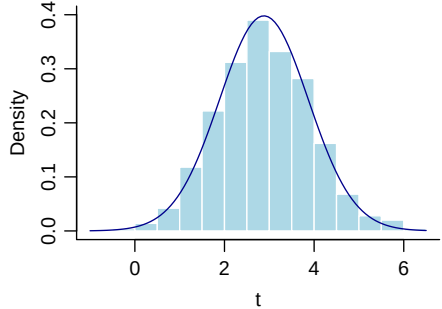
Nachfolgend werden nun die Gegebenheiten einer Behandlungssituation betrachtet (aber ohne Behandlungseffekt!)

- Simulation einer verbundenen Stichprobe (x, y) mit $N = 100$ aus einer bivariaten Normalverteilung mit den Normkennwerten des BDI-V ($M = 20.4$, $SD = 14.2$ und Korrelation $r = 0.9$) und einem Pretest-Wert $x > 30$ (für die klinische Diagnose einer Depression wird der Wert 35 vorgeschlagen)
- Unter diesen Bedingungen weicht die bei 1000 Simulationen erhaltene Verteilung der p-Werte des entsprechenden t-Tests für verbundene Stichproben deutlich von der gleichmäßigen Verteilung auf dem Intervall $(0, 1)$ ab, obwohl H_0: $\mu_x = \mu_y$ gilt

Regression zur Mitte: adjustierter p-Wert

Stichprobenverteilung mit Pretest-Einschlusskriterium ($x > 30$)





Regression zur Mitte: adjustierter p-Wert

Die hohe Wahrscheinlichkeit kleiner p-Werte ist Ausdruck der auf Grund der Regression zur Mitte resultierenden Effekte

Eine Adjustierung p_{adj} des p-Werts p_0 für einen Datensatz mit diesen Verteilungsannahmen erhält man auch hier als Anteil der p-Werte p_i der Simulationen $i = 1, \ldots, 1000$, für die $p_i \leq p_0$ gilt (im Allgemeinen ist hier $p_{\text{adj}} \neq p_0$)

Beispiel

- Für einen Datensatz mit $\bar{x} = 38.72$, $\bar{y} = 37.18$, $s_{x-y} = 6.56$ erhält man $t(99) = 2.35$ und $p_0 = 0.01$ und bestimmt $p_{\text{adj}} = 0.6$ (Bootstrap-p-Wert)

Regression zur Mitte: parametrisches Bootstrapping

```
n <- 100
s <- 14.2; r <- 0.9
p0 <- 0.01                                  # original p-value
mu <- c(20.4, 20.4)                         # mean vector
Sigma <- rbind(c(s^2, r*s*s), c(r*s*s, s^2)) # covariance matrix

cutoff <- 30                                # inclusion criterion
rpt <- round(2 * n / (1 - pnorm(cutoff, mean = mu[1], sd = s)))

pval <- replicate(2000, {
  x <- MASS::mvrnorm(rpt, mu, Sigma)
  x <- x[x[, 1] > cutoff, ]                 # pre > cutoff?
  x <- x[1:n, ]                             # select n
  t.test(x[, 1], x[, 2], paired = TRUE)$p.value
})
p.adj <- mean(pval <= p0)                   # bootstrap p-value
```

Regressionsanalyse

Rückblick

- In Statistik I hatten wir die lineare Regression als Verfahren kennengelernt, bei dem ein Kriterium (Regressand) durch einen Prädiktor (Regressor) vorhergesagt wird
- Die Ausgangsdaten wurden durch eine Folge von Paaren

$$(x_1, y_1), \ldots, (x_n, y_n)$$

gebildet, die Ausprägungen metrischer Merkmale X und Y repräsentieren

- An diese Daten soll möglichst gut eine (Ausgleichs-)Gerade

$$Y = \alpha + \beta \cdot X$$

angepasst werden

Modell

Ausgehend von diesem deskriptiven Regressionsansatz wird ein statistisches Modell entwickelt

- Die deterministische Regressionsfunktion wird durch einen additiven Fehlerterm ergänzt

$$Y = \alpha + \beta \cdot X + \varepsilon$$

- Der Fehler ε wird als Zufallsvariable mit bestimmten Eigenschaften betrachtet
- Bei gegebenem Wert x von X ist damit auch Y eine Zufallsvariable
- Das Merkmal X kann dabei entweder deterministisch sein (d. h. systematisch oder kontrolliert variiert werden) oder ebenfalls eine Zufallsvariable

Modell

Für die Paare

$$(X_1, Y_1), \ldots, (X_n, Y_n)$$

hat man dann das stochastische Modell

$$Y_i = \alpha + \beta \cdot X_i + \varepsilon_i$$

mit den Realisierungen

$$y_i = \alpha + \beta \cdot x_i + \varepsilon_i$$

$(i = 1, \ldots, n)$

Die *Fehler-* oder *Störvariablen* ε_i ...

- ...sind nicht beobachtbar und umfassen den nicht kontrollierten oder nicht systematisch messbaren Einfluss von Messfehlern oder zusätzlich wirksamer Variablen

Fehlervariablen

Die *Fehler-* oder *Störvariablen* ε_i ...

- ... sind unsystematische oder zufällige Komponenten, sodass es vernünftig ist, anzunehmen, dass

$$E(\varepsilon_i) = 0, \text{ für alle } i = 1, \ldots, n$$

Die Werte x_i ...

- ... können deterministisch bestimmt sein (z. B. fest vorgegeben im Rahmen eines geplanten Versuchs)
- ... können sich als Realisierungen der Zufallsvariablen X_i ergeben (die Paare (X_i, Y_i), $i = 1, \ldots, n$, bilden dann eine verbundene Stichprobe)

  Man spricht von einem *stochastischen Regressor*

Grundmodell

Das stochastische Grundmodell

$$Y_i = \alpha + \beta \cdot X_i + \varepsilon_i \text{ mit } E(\varepsilon_i) = 0, \text{ für alle } i = 1, \ldots, n$$

wird durch zusätzliche Annahmen weiter spezifiziert

- Die Zufallsvariablen ε_i, $i = 1, \ldots, n$, werden als stochastisch unabhängig und identisch verteilt angenommen
- Insbesondere sollen die ε_i, $i = 1, \ldots, n$, alle dieselbe Varianz $Var(\varepsilon_i) = \sigma^2$ haben
- Bei stochastischem Regressor sind die obigen Annahmen streng genommen unter der Bedingung $X_i = x_i$ zu betrachten, d. h. für alle $i = 1, \ldots, n$ gilt

$$E(\varepsilon_i \mid X_i = x_i) = 0, \quad Var(\varepsilon_i \mid X_i = x_i) = \sigma^2$$

Folgerungen

Als Folgerung aus den Eigenschaften der Fehlervariablen erhält man

$$E(Y_i) = E(\alpha + \beta \cdot x_i + \varepsilon_i) = \alpha + \beta \cdot x_i$$

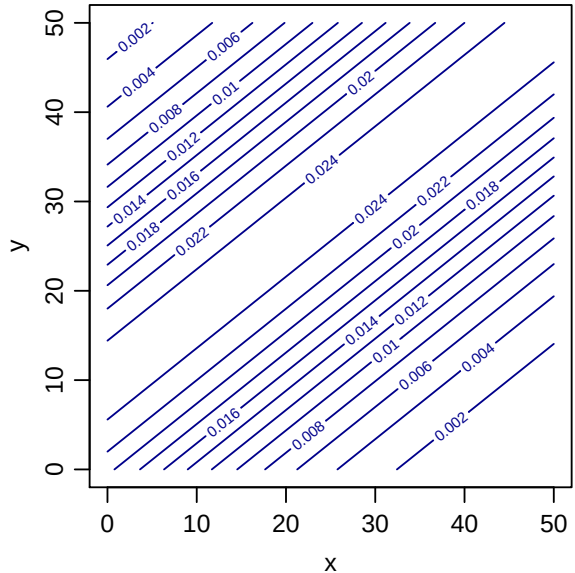
und

$$Var(Y_i) = Var(\alpha + \beta \cdot x_i + \varepsilon_i) = \sigma^2$$

Bei gegebenem x_i überträgt sich auch die stochastische Unabhängigkeit der ε_i auf die Y_i

Die Y_i sind darüber hinaus bis auf die Verschiebung $\alpha + \beta \cdot x_i$ der Erwartungswerte identisch verteilt

Verteilung der Y_i



Homoskedastizität

Die Eigenschaft gleicher Varianz σ^2 der Fehlervariablen ε_i wird auch als *Homoskedastizität* bezeichnet

- Verletzungen dieser Annahme resultieren, wenn die Fehlervarianzen mit größer werdenden x-Werten ebenfalls zunehmen
- Eine derartige *Heteroskedastizität* lässt sich häufig bereits am Streudiagramm der (x_i, y_i)-Werte erkennen
- In diesem Fall sind die Methoden der linearen Regression nur in geeignet modifizierter Form anwendbar

Die dargestellten Annahmen, insbesondere auch die Linearität, sind in Anwendungen mit statistischen Methoden der Modelldiagnose zu überprüfen

Normalverteilungsannahme

Für exakte Aussagen zu Schätzfunktionen und Teststatistiken (auch bei kleinen Stichprobenumfängen n) ist anzunehmen, dass die Fehler- und Kriteriumsvariablen normalverteilt sind, d. h. für $i = 1, \ldots, n$

$$\varepsilon_i \sim N(0, \sigma^2)$$

und

$$Y_i \sim N(\alpha + \beta \cdot x_i, \sigma^2)$$

Diese Normalverteilungsannahme ist in geeigneter Weise zu überprüfen (z. B. mit Normal-Quantil-Plots)

Parameterschätzung

Wie im Rahmen der deskriptiven Statistik wird zur Parameterschätzung das *(gewöhnliche) Kleinst-Quadrat-Kriterium* (ordinary least squares criterion) verwendet

- Die Schätzfunktionen $\hat{\alpha}$ und $\hat{\beta}$ werden so gewählt, dass die Quadratsumme der Residuen

$$Q(\alpha, \beta) = \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$$

  durch $\hat{\alpha}, \hat{\beta}$ minimiert wird

Es ergeben sich die Schätzfunktionen

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\,\bar{X} \text{ und } \hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Parameterschätzung

Mit den vorhergesagten Werten

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$$

hat man die Residuen

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$$

für alle $i = 1, \ldots, n$

Als Schätzfunktion für σ^2 verwendet man dann die mittlere Quadratsumme der Residuen

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2$$

Eigenschaften der Schätzfunktionen

Die angegebenen Schätzfunktionen sind erwartungstreu, d. h. es gilt

$$E(\hat{\alpha}) = \alpha,\ E(\hat{\beta}) = \beta,\ E(\hat{\sigma}^2) = \sigma^2$$

Darüber hinaus hat man

$$Var(\hat{\alpha}) = \sigma^2_{\hat{\alpha}} = \sigma^2 \frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}$$

und

$$Var(\hat{\beta}) = \sigma^2_{\hat{\beta}} = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Eigenschaften der Schätzfunktionen

Gilt für $n \to \infty$

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 \to \infty,$$

so folgt

$$Var(\hat{\alpha}) \to 0 \text{ und } Var(\hat{\beta}) \to 0,$$

sodass die Schätzfunktionen unter dieser Bedingung konsistent sind

- Die Konsistenzbedingung bedeutet, dass die Werte $X_1, \ldots, X_n$ für alle n hinreichend stark um ihr arithmetisches Mittel streuen
- Nur so kommt immer wieder neue Information zur Schätzung von $\hat{\alpha}$ und $\hat{\beta}$ hinzu
- Bei einem stochastischen Regressor mit unabhängig und identisch verteilten Zufallsvariablen $X_1, \ldots, X_n$ ist diese Bedingung erfüllt

Eigenschaften der Schätzfunktionen

Unter der getroffenen Normalverteilungsannahme kann man ableiten, dass

$$\hat{\alpha} \sim N(\alpha, \sigma^2_{\hat{\alpha}}) \quad \text{und} \quad \hat{\beta} \sim N(\beta, \sigma^2_{\hat{\beta}})$$

Daher gilt für die Verteilung der standardisierten Schätzfunktionen

$$\frac{\hat{\alpha} - \alpha}{\hat{\sigma}_{\hat{\alpha}}} \sim t(n-2) \quad \text{und} \quad \frac{\hat{\beta} - \beta}{\hat{\sigma}_{\hat{\beta}}} \sim t(n-2)$$

mit den Schätzern $\hat{\sigma}^2_{\hat{\alpha}}$ und $\hat{\sigma}^2_{\hat{\beta}}$ für die entsprechenden Varianzen

Die standardisierten Schätzfunktionen sind also t-verteilt mit $n-2$ Freiheitsgraden (Anzahl verringert durch zwei geschätzte Parameter)

Konfidenzintervalle

Damit erhält man die $(1-\alpha)$-Konfidenzintervalle

$$\left[\hat{\alpha} - \hat{\sigma}_{\hat{\alpha}}\, t_{1-\alpha/2}(n-2), \hat{\alpha} + \hat{\sigma}_{\hat{\alpha}}\, t_{1-\alpha/2}(n-2)\right]$$

und

$$\left[\hat{\beta} - \hat{\sigma}_{\hat{\beta}}\, t_{1-\alpha/2}(n-2), \hat{\beta} + \hat{\sigma}_{\hat{\beta}}\, t_{1-\alpha/2}(n-2)\right]$$

Für $n > 30$ kann man die t-Quantile der $t(n-2)$-Verteilung durch Quantile der $N(0,1)$-Verteilung ersetzen

Bei Abweichungen von der Normalverteilungsannahme und hinreichend großem Stichprobenumfang n sind die (standardisierten) Schätzfunktionen approximativ t-verteilt bzw. normalverteilt

Statistische Hypothesentests

Waldtest für α

Annahmen

- Unabhängige Stichprobenvariablen (X_i, Y_i), $i = 1, \ldots, n$, mit $Y \mid X$ normalverteilt mit einer von X unabhängigen Varianz σ^2

Hypothesen

1. $H_0\colon\ \alpha = \alpha_0 \qquad H_1\colon\ \alpha \neq \alpha_0$
2. $H_0\colon\ \alpha \geq \alpha_0 \qquad H_1\colon\ \alpha < \alpha_0$
3. $H_0\colon\ \alpha \leq \alpha_0 \qquad H_1\colon\ \alpha > \alpha_0$

Teststatistik

$$T_{\alpha_0} = \frac{\hat{\alpha} - \alpha_0}{\hat{\sigma}_{\hat{\alpha}}}$$

Waldtest für α

Verteilung der Teststatistik unter $\alpha = \alpha_0$

- $T_{\alpha_0} \sim t(n-2)$

Ablehnungsbereich

1. $|T_{\alpha_0}| > t_{1-\alpha/2}(n-2)$
2. $T_{\alpha_0} < -t_{1-\alpha}(n-2)$
3. $T_{\alpha_0} > t_{1-\alpha}(n-2)$

mit dem p-Quantil $t_p(n-2)$ der t-Verteilung mit $df = n-2$ Freiheitsgraden (für $n > 30$ kann man die t-Quantile der $t(n-2)$-Verteilung durch Quantile der $N(0,1)$-Verteilung ersetzen)

Waldtest für β

Annahmen

- Unabhängige Stichprobenvariablen (X_i, Y_i), $i = 1, \ldots, n$, mit $Y \mid X$ normalverteilt mit einer von X unabhängigen Varianz σ^2

Hypothesen

1. $H_0\colon\ \beta = \beta_0 \qquad H_1\colon\ \beta \neq \beta_0$
2. $H_0\colon\ \beta \geq \beta_0 \qquad H_1\colon\ \beta < \beta_0$
3. $H_0\colon\ \beta \leq \beta_0 \qquad H_1\colon\ \beta > \beta_0$

Teststatistik

$$T_{\beta_0} = \frac{\hat{\beta} - \beta_0}{\hat{\sigma}_{\hat{\beta}}}$$

Waldtest für β

Verteilung der Teststatistik unter $\beta = \beta_0$

- $T_{\beta_0} \sim t(n-2)$

Ablehnungsbereich

1. $|T_{\beta_0}| > t_{1-\alpha/2}(n-2)$
2. $T_{\beta_0} < -t_{1-\alpha}(n-2)$
3. $T_{\beta_0} > t_{1-\alpha}(n-2)$

mit dem p-Quantil $t_p(n-2)$ der t-Verteilung mit $df = n-2$ Freiheitsgraden (für $n > 30$ kann man die t-Quantile der $t(n-2)$-Verteilung durch Quantile der $N(0,1)$-Verteilung ersetzen)

Nullmodell

Von besonderem Interesse ist das Hypothesenpaar

$$H_0: \ \beta = \beta_0, \quad H_1: \ \beta \neq \beta_0$$

Dabei bedeutet die Nullhypothese mit $\beta_0 = 0$, dass für alle $i = 1, \ldots, n$ tatsächlich gilt

$$Y_i = \alpha + \varepsilon_i$$

Unter dieser Nullhypothese besitzt das Merkmal X keinerlei Erklärungswert für das Merkmal Y

- Ablehnung von H_0 zugunsten von H_1 bedeutet, dass X zur Vorhersage von Y eingesetzt werden kann

Streuungszerlegung

Analog zur Betrachtung in Statistik I kann man folgende Zerlegung der Gesamtstreuung (SS_Y, *sum of squares total*) durchführen

$$\begin{aligned} SS_Y &= \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \\ &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 \\ &= SS_{\hat{\varepsilon}} + SS_{\hat{Y}} \end{aligned}$$

mit Fehlerstreuung ($SS_{\hat{\varepsilon}}$, *sum of squares residuals*) und der erklärten Streuung ($SS_{\hat{Y}}$, *sum of squares explained*).

Overall-Test

Wie später bei der Varianzanalyse kann man hieraus eine Teststatistik entwickeln

$$F = \frac{\frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{1}}{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}}$$

Man kann zeigen, dass gilt $F = T_{\beta_0}^2$ für $\beta_0 = 0$, sodass auch F für einen Test der Hypothese H_0: $\beta = 0$ verwendet werden kann

- Unter der getroffenen Normalverteilungsannahme gilt unter H_0: $F \sim F(1, n-2)$

Overall-Test

Da der Determinationskoeffizient $R^2 = r_{XY}^2$ den Anteil der durch die Regression erklärten Varianz an der Gesamtvarianz beschreibt, hat man außerdem die Beziehung

$$F = \frac{R^2}{1 - R^2} (n - 2)$$

Diese Gleichung lässt sich auch zur Charakterisierung der Güte der Vorhersage im Rahmen einer *multiplen Regression* einsetzen, bei der mehr als ein Prädiktor verwendet wird