Skript zur Veranstaltung

LINEARE MODELLE

von

Andreas Handl

Freie Universität Berlin

1.Statistische Modellbildung

Mit Hilfe statistischer Verfahren sollen Gesetzmäßigkeiten und Strukturen in Datensätzen gefunden und beschrieben werden. Statistik dient also der Aufbereitung und Auswertung von Datensätzen. Eine sinnvolle Auswertung von Datensätzen ist nur im Rahmen eines Modells möglich. Ein Modell soll den Mechanismus beschreiben, der die Daten erzeugt hat.

Schauen wir uns dazu ein Beispiel an, das dem Buch Data,Models and Statistical Analysis von Cooper und Weekes entnommen ist.

Ein Arbeitnehmer kann auf zwei unterschiedlichen Wegen zur Arbeit fahren. Er will herausfinden, auf welchem er schneller ist. Naheliegend ist es, jeden der beiden Wege genau einmal zu benutzen und sich dann für den zu entscheiden, bei dem er weniger Zeit brauchte. Nun ist es klar, daß die täglich benötigten Zeiten bei einem Weg nicht gleich sein werden. so daß man diese Streuung bei der Entscheidung berücksichtigen sollte. Dies hat zur Folge, daß man jeden der Wege nicht einmal, sondern mehrmals zurücklegt. Der Arbeitnehmer macht dies und erhält folgende Zeiten (in Min.):

Weg 1: 29.2 34.2 31.1 32.4 28.8 25.5 30.2 29.1 28.1 27.7
Weg 2: 29.6 31.4 32.1 31 28.9 30.4 30.6 31.6 34.9 31.5 30.2 31.8

Bevor wir uns Gedanken machen, was es nun bedeutet, auf einem Weg schneller zu sein, werfen wir einen Blick auf die Daten.
Ein Boxplot liefert ein gutes Bild eines Datensatzes.
Für einen Boxplot benötigt man 5 Größen:

Minimum	$x_{(1)}$
unteres Quartil	$x_{0.25}$
Median	$x_{0.5}$
oberes Quartil	$x_{0.75}$
Maximum	$x_{(n)}$

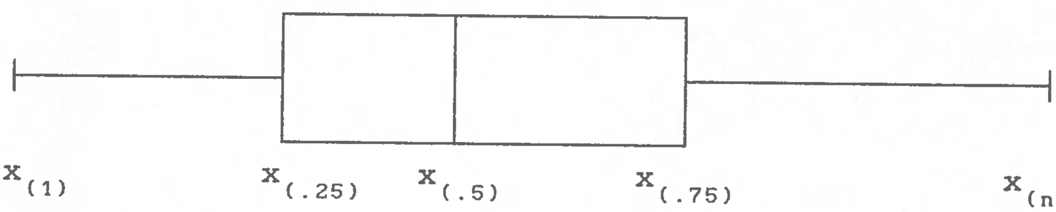
Diese können alle aus dem geordneten Datensatz $x_{(1)},...,x_{(n)}$ bestimmt werden.

Der Median teilt den geordneten Datensatz in zwei gleiche Hälften, d.h.

$$x_{0.5} = \begin{cases} x_{((n+1)/2)} & \text{für n ungerade} \\ 0.5(x_{(n/2)}+x_{(1+n/2)}) & \text{für n gerade} \end{cases}$$

Das untere Quartil ist der Median der unteren Hälfte des geordneten Datensatzes und das obere Quartil der Median der oberen Hälfte des geordneten Datensatzes. Ist der Stichprobenumfang ungerade, so wird der Median sowohl zur unteren als auch zur oberen Hälfte des Datensatzes hinzugenommen.

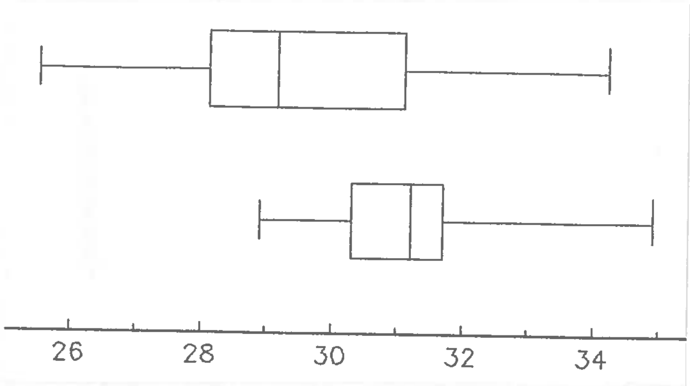
Ein Boxplot wird nun folgendermaßen erstellt:
Man wählt eine Skala, die vom Minimum $x_{(1)}$ bis zum Maximum $x_{(n)}$ geht. Auf dieser Skala markiert man die Quartile und den Median. Nun zeichnet man einen Kasten, der vom unteren bis zum oberen Quartil reicht. In dem Kasten zeichnet man den Median als senkrechten Strich ein. Außerhalb des Kastens zeichnet man von den Quartilen waagerechte Linien bis zu den Extremwerten.



Die 5 Maßzahlen sind für die beiden Datensätze:

	1. Weg	2.Weg
$x_{(1)}$	25.5	28.9
$x_{.25}$	28.1	30.3
$x_{.5}$	29.15	31.2
$x_{.75}$	31.1	31.7
$x_{(n)}$	34.2	34.9

Die Boxplots der beiden Datensätze sind in der folgenden Graphik wiedergegeben.



Die Boxplots zeigen, daß die beiden empirischen Verteilungen mehr oder weniger symmetrisch sind, sich jedoch sowohl hinsichtlich der Lage als auch hinsichtlich der Streuung unterscheiden. Besteht dieser Unterschied auch in den zugrundeliegenden Grundgesamtheiten?

Die "klassische" Antwort auf diese Frage sieht folgendermaßen aus: Wir nehmen an, daß die beiden Grundgesamtheiten normalverteilt sind. Aus diesen beiden Grundgesamtheiten haben wir dann Realisationen von unabhängigen Zufallsvariablen beobachtet.

Unsere Annahmen lauten also :

$X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_n$ sind unabhängige Zufallsvariablen, wobei gilt $X_i \sim N(\mu_1, \sigma_1^2)$,i=1,2,...,m und $Y_j \sim N(\mu_2, \sigma_2^2)$,j=1,2,...,n.

Überprüfen wollen wir nun, auf welchem Weg er eine kürzere Zeit zu erwarten hat,d.h.

$$H_0: \mu_1 = \mu_2 \quad \text{gegen} \quad H_1: \mu_1 \neq \mu_2$$

Seien $\overline{X} = \frac{1}{m} \sum_i X_i$ und $\overline{Y} = \frac{1}{n} \sum_j Y_j$. Sind μ_1 und μ_2 ungleich, so sollte sich dies auch in $\overline{X}$ und $\overline{Y}$ zeigen, da sie erwartungstreue Schätzer sind. Es liegt also nahe den Test auf $\overline{X}$ und $\overline{Y}$ aufzubauen.

Wegen $\overline{X} \sim N(\mu_1, \sigma_1^2/m)$ und $\overline{Y} \sim N(\mu_2, \sigma_2^2/n)$ gilt also

$\overline{X}-\overline{Y}$ $N(\mu_1-\mu_2, (\sigma_1^2/m)+(\sigma_2^2/n))$, und somit bietet sich als Teststatistik an:

$$z = \frac{\overline{X} - \overline{Y}}{\sqrt{(\sigma_1^2/m)+(\sigma_2^2/n)}}$$

In der Praxis sind nun σ_1^2 und σ_2^2 in der Regel unbekannt. Schätzt man diese nun durch die erwartungstreuen Schätzfunktionen

$$S_1^2 = \frac{1}{m-1}\sum_i (X_i - \bar{X})^2 \quad \text{und} \quad S_2^2 = \frac{1}{n-1}\sum_j (Y_j - \bar{Y})^2 ,$$

so ist die finite Verteilung der Teststatistik unbekannt.

Durch eine zusätzliche Annahme können wir dieses Problem aber lösen.

Die Annahme $\sigma_1^2 = \sigma_2^2 = \sigma^2$ liefert nach Einsetzen der erwartungstreuen Schätzfunktion $S^2 = \frac{1}{m+n-2}\left(\sum_i (X_i - \bar{X})^2 + \sum_j (Y_j - \bar{Y})^2\right)$ in die obige Teststatistik

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{m+n}{mn} S^2}} ,$$

die unter H_0 t-verteilt ist mit m+n-2 Freiheitsgraden.

Dies ist der klassische t-Test im Zweistichprobenproblem unabhängiger Stichproben. Zum Signifikanzniveau α wird die Hypothese $H_0: \mu_1 = \mu_2$ abgelehnt, wenn gilt $|t| > t_{m+n-2;1-\alpha/2}$, wobei $t_{m+n-2;1-\alpha/2}$ das $1-\alpha/2$-Quantil der t-Verteilung mit m+n-2 Freiheitsgraden ist.

Für den obigen Datensatz gilt t=-1.79. Da gilt $t_{20;0.975} = 2.086$, wird H_0 zum Niveau $\alpha = 0.05$ nicht abgelehnt. Fraglich ist jedoch, ob die Annahme gerechtfertigt ist, daß die Varianzen in den beiden Grundgesamtheiten identisch sind. Wir wollen uns jedoch mit dieser Frage hier nicht weiter beschäftigen, sondern die beiden Datensätze unter einem anderen Aspekt betrachten.

Wir wählen zunächst eine andere Notation für die den Beobachtungen zugrundeliegenden Zufallsvariablen.Die Zufallsvariablen $X_1, \ldots, X_m$ bezeichnen wir mit $Y_{11}, \ldots, Y_{1n_1}$ und die Zufallsvariablen $Y_1, \ldots, Y_n$ mit $Y_{21}, \ldots, Y_{2n_2}$. Die Beobachtungen $y_{i1}, \ldots, y_{in_i}$ in der i.ten Stichprobe sind Realisationen der Zufallsvariablen $Y_{i1}, \ldots, Y_{in_i}$.

Da diese Beobachtungen aus einer Grundgesamtheit stammen, ist es sinnvoll anzunehmen, daß es einen typischen Wert μ_i gibt, den man für eine Beobachtung erwarten würde. Nun nimmt nicht jede Beobachtung diesen typischen Wert an, so daß wir diese Abweichung berücksichtigen müssen.

Wir unterstellen also für Y_{ij} folgendes Modell:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad ,i=1,2\ j=1,\ldots,n_i\ .$$

Die ε_{ij} bezeichnen wir als Störterme. Wir wollen annehmen, daß die Störterme keinen systematischen Einfluß haben. Im Mittel sollten sich die Störungen also aufheben, d.h. $E(\varepsilon_{ij})=0$.
Weiterhin wollen wir annehmen, daß die Varianzen der Störterme identisch ist, und daß die Störterme unkorreliert sind, d.h.

$$Cov(\varepsilon_{ik},\varepsilon_{jl}) = \begin{cases} \sigma^2 & \text{für } i=j \text{ und } k=l \\ 0 & \text{sonst} \end{cases}$$

Außerdem gehen wir davon aus, daß die Störterme normalverteilt sind.
Wir unterstellen also folgendes Modell:

$$\boxed{\begin{array}{ll} Y_{ij} = \mu_i + \varepsilon_{ij} & i=1,2, j=1,\ldots,n_i \\ \varepsilon_{11},\ldots,\varepsilon_{2n_2} & i.i.d.\ N(0,\sigma^2) \end{array}}$$

Wir können dieses auch in Matrizenschreibweise darstellen:

Seien

$$y = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ \vdots \\ y_{2n_2} \end{pmatrix},\ \varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \vdots \\ \varepsilon_{2n_2} \end{pmatrix},\ \beta = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ und } X = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}$$

Dann gilt für das obige Modell

$$y = X\beta + \varepsilon$$

mit $E(\varepsilon)=0$ und $Cov(\varepsilon)=\sigma^2 I_n$. Die Normalverteilungsannahme des Zufallsvektors y stellen wir dar mit $\varepsilon \sim N_n(0,\sigma^2 I_n)$.(In Kapitel 3 werden wir uns detailliert mit Zufallsvektoren beschäftigen)
Modelle der Form $y=X\beta+\varepsilon$ heißen Lineare Modelle,weil sie linear in dem Parametervektor β sind. Sie nehmen im Rahmen der Statistik eine zentrale Stellung ein.
Die Matrix X nennt man Designmatrix.

Enthält die Designmatrix wie im obigen Beispiel nur Nullen und Einsen, so handelt es sich um ein Modell der Varianzanalyse. Hier soll der Einfluß von einer oder mehreren kategorialen Variablen auf eine stetige Variable untersucht werden. Mit diesen werden wir uns in Kapitel 4 beschäftigen.
Sind die Elemente der Designmatrix in einem gewissen Bereich frei wählbar, so haben wir es mit der Regressionsanalyse.zu tun Hier soll eine der Einfluß einer oder mehrerer stetiger Variablen auf eine stetige Variable beschrieben werden. Mit diesen wwerden wir uns zunächst in den Kapiteln 2 und 3 beschäftigen, wobei wir uns in Kapitel 2 auf eine erklärenden Variable beschränken.
Soll eine stetige Variable sowohl durch kategoriale als auch stetige Variablen erklärt werden, so spricht man von Kovarianzanalyse. Mit dieser werden wir uns nicht beschäftigen.

2. Die Einfachregression

2.1 Das Modell

Als erstes Beispiel eines Linearen Modells werden wir in diesem Abschnitt die Einfachregression betrachten. Das Modell lautet:

$$Y=\beta_0+\beta_1 x+\varepsilon \qquad (2.1)$$
$$\text{mit } E(\varepsilon)=0 \text{ und } Var(\varepsilon)=\sigma^2.$$

Wegen $E(Y)=\beta_0+\beta_1 x$ liegt der lineare Zusammenhang zwischen der zu erklärenden Variablen Y und der erklärenden Variablen x nicht exakt, sondern im Mittel vor.

Unser Modell umfaßt auch Beziehungen der Form

$$y=\beta_0+\beta_1 ln(x^*)+\varepsilon \qquad (2.2) ,$$

da dieses linear in den Parametern ist und für $x=ln(x^*)$ mit Modell (2.1) zusammenfällt.

Die Beziehung $y=\beta_0+\beta_1 ln(x)$ wird auch als psychophysisches Grundgesetz bezeichnet. (siehe dazu Heuser:Lehrbuch der Analysis 1, S.318).

Viele in den Parametern nichtlineare Beziehungen können durch eine geeignete Transformation linearisiert werden wie z.B.

$$y=\beta_0 exp(\beta_1 x) \qquad (2.3)$$

Dabei ist $exp(x)$ die e-Funktion an der Stelle x. Logarithmieren der Gleichung (2.3) führt zu

$$ln(y)=ln(\beta_0)+\beta_1 x \qquad (2.4)$$

Diese Beziehung ist in den Parametern $ln(\beta_0)$ und β_1 linear.Die Gleichungen (2.3) und (2.4) enthalten keinen Störterm. Dieser muß in (2.3) multiplikativ eingehen, um in (2.4) additiv aufzutreten. Außerdem muß er in (2.3) lognormalverteilt sein, um in (2.4) einer Normalverteilung zu folgen. Das Modell (2.3) beschreibt exponentielles Wachstum mit der Wachstumsrate β_1.(siehe dazu Heuser:Lehrbuch der Analysis 1, S.309). Viele Beziehungen können nicht durch Transformationen linearisiert werden wie z.B.

$$y=\frac{\beta_0}{1+\beta_1 exp(\beta_2 x)}$$

Diese bezeichnet man auch als logistisches (autokatalytisches) Modell. Es beschreibt Wachstum mit Sättigungspunkt. In der Ökonomie ist es bekannt unter dem Namen Ertragsgesetz. siehe dazu Heuser:Lehrbuch der Analysis 1, S.312)

2.2 Schätzung der Parameter

Das Modell (2.1) enthält die unbekannten Parameter β_1 und β_2. Um diese und die unbekannte Varianz σ^2 der Störterme zu schätzen, werden n Wertepaare $(x_i, y_i)'$,i=1,2,...,n beobachtet. Wegen (2.1) gilt für dann für die hinter den y_i stehenden Zufallsvariablen Y_i:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , i=1,2,\ldots,n$$

In diesem und im nächsten Abschnitt werden wir die Realisationen y_i der Y_i betrachten. Erst in Abschnitt 2.4 werden wir dann wieder auf die Zufallsvariablen Y_i eingehen.

Von Gauss wurde 1803 und von Legendre 1806 die Methode der Kleinsten Quadrate als Schätzkriterium vorgeschlagen. Dabei werden die Parameter β_0 und β_1 so bestimmt, daß die Summe der quadrierten Abweichungen der y_i von der Geraden minimal wird.

Das Kriterium lautet also

$$\text{Minimiere} \quad S(\beta_0, \beta_1) = \sum_i (y_i - \beta_0 - \beta_1 x_i)^2$$

Notwendige Bedingungen für einen Extremwert sind, daß die partiellen Ableitungen nach β_0 und β_1 Null sind.

Es gilt

$$\frac{\partial}{\partial\beta_0} = (-2) \sum_i (y_i - \beta_0 - \beta_1 x_i) \stackrel{!}{=} 0$$

$$\frac{\partial}{\partial\beta_1} = (-2) \sum_i (y_i - \beta_0 - \beta_1 x_i) x_i \stackrel{!}{=} 0$$

Die Kleinst-Quadrate-Schätzer $\hat\beta_0$ und $\hat\beta_1$ erfüllen also die sogenannten Normalgleichungen

$$\sum_i (y_i - \hat\beta_0 - \hat\beta_1 x_i) = 0 \qquad (2.5)$$

$$\sum_i (y_i - \hat\beta_0 - \hat\beta_1 x_i) x_i = 0 \qquad (2.6)$$

Vereinfacht man diese Gleichungen, so erhält man:

$$\hat{\beta}_0\, n \quad + \hat{\beta}_1 \sum_i x_i = \sum_i y_i \tag{2.7}$$

$$\hat{\beta}_0 \sum_i x_i + \hat{\beta}_1 \sum_i x_i^2 = \sum_i x_i y_i \tag{2.8}$$

Multipliziert man (2.7) mit $\sum_i x_i$, subtrahiert dann von dieser Gleichung das n-fache der Gleichung (2.8) und löst das Ergebnis nach $\hat{\beta}_1$ auf, so erhält man

$$\hat{\beta}_1 = \frac{n\sum_i x_i y_i - \left(\sum_i x_i\right)\left(\sum_i y_i\right)}{n\sum_i x_i^2 \quad - \left(\sum_i x_i\right)^2}$$

Es gilt $\sum_i \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) = \sum_i x_i y_i - \bar{y} \sum_i x_i - \bar{x} \sum_i y_i + \sum_i \bar{x}\bar{y} =$

$$= \sum_i x_i y_i - \frac{1}{n} \sum_i x_i \sum_i y_i - \frac{1}{n} \sum_i x_i \sum_i y_i + \frac{1}{n} \sum_i x_i \sum_i y_i =$$

$$= \sum_i x_i y_i - \frac{1}{n} \sum_i x_i \sum_i y_i = \frac{1}{n}\left(n\sum_i x_i y_i - \sum_i x_i \sum_i y_i\right).$$

Analog erhalten wir

$$\sum_i \left(x_i - \bar{x}\right)^2 = \frac{1}{n}\left(n\sum_i x_i^2 - \left(\sum_i x_i\right)^2\right).$$

Somit gilt

$$\hat{\beta}_1 = \frac{\sum_i \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_i \left(x_i - x\right)^2} \tag{2.9}$$

Mit $S_{xx} = \sum_i \left(x_i - \bar{x} \right)^2$ und $S_{yy} = \sum_i \left(x_i - \bar{x} \right)\left(y_i - \bar{y} \right)$ gilt also

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \tag{2.10}$$

Aus der ersten Normalgleichung folgt sofort

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{2.11}$$

Die geschätzte Regressionsgerade lautet:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{2.12}$$

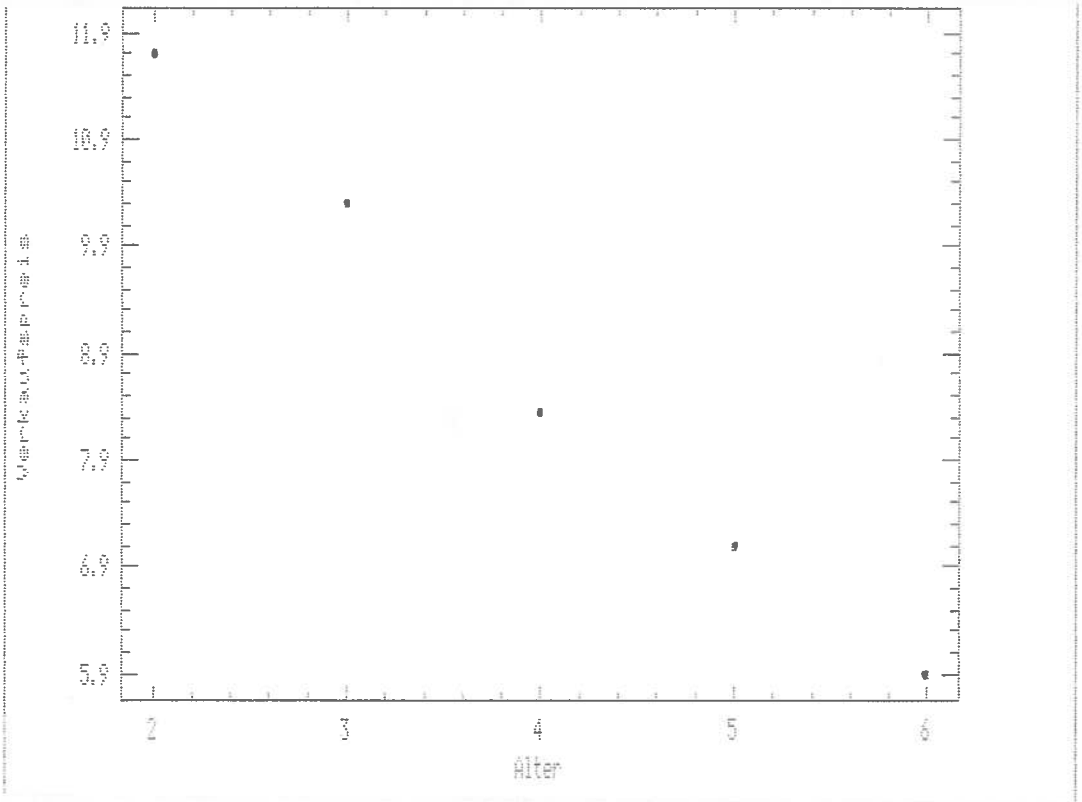
Man kann zeigen, daß die Funktion $S(\beta_0, \beta_1)$ an der Stelle $(\hat{\beta}_0, \hat{\beta}_1)'$ ein Minimum annimmt.(Siehe dazu Kapitel 3)

Beispiel 1

Im Jahre 1984 galten für gebrauchte Pkws eines speziellen Typs folgende Händlerverkaufspreise (in 1000 DM) (aus R.Schlittgen,Einführung in die Statistik)

Alter	2	3	4	5	6
Preis	11.7	10.3	8.35	7.1	5.9

Das Streudiagramm legt nahe, daß zwischen dem Preis y und dem Alter x ein linearer Zusammenhang besteht.



Um die Parameter zu schätzen, erstellen wir folgende Hilfstabelle

i	x_i	y_i	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$
1	2	11.7	-2	3.03	-6.06	4
2	3	10.3	-1	1.63	-1.63	1
3	4	8.35	0	-0.32	0	0
4	5	7.1	1	-1.57	-1.57	1
5	6	5.9	2	-2.77	-5.54	4

Wir erhalten also $\bar{x} = 4$, $\bar{y} = 8.67$, $S_{xy} = -14.8$ und $S_{xx} = 10$.
Somit gilt $\hat{\beta}_1 = -1.48$ und $\hat{\beta}_0 = 14.59$.

Anmerkungen

(a) Sind alle $x_i, i=1,\dots,n$ identisch, so kann der Parameter β_1 nicht geschätzt werden.

(b) Liegen nur zwei Meßpunkte vor, und werden am ersten Punkt n-1 Werte von y und am zweiten Meßpunkt ein Wert von y beobachtet, so geht die geschätzte Regressionsgerade durch den Mittelwert der y_i am ersten Meßpunkt und durch den Wert von y am zweiten Meßpunkt. Dies sieht man folgendermaßen:

Wir setzen $x_1=x_2=\dots=x_{n-1}=x^*$ und bestimmen zunächst den Schätzer von β_1.

Es gilt

$$\sum_i\left(x_i-\bar{x}\right)^2 = \sum_{i=1}^{n-1}\left(x^*-\frac{(n-1)x^*+x_n}{n}\right)^2+\left(x_n-\frac{(n-1)x^*+x_n}{n}\right)^2 =$$

$$= (n-1)\left(\frac{nx^*-nx^*+x^*-x_n}{n}\right)^2+\left(\frac{(n-1)\left(x_n-x^*\right)}{n}\right)^2 =$$

$$= \frac{n-1}{n^2}\left(x^*-x_n\right)^2+\frac{(n-1)^2}{n^2}\left(x^*-x_n\right)^2 =$$

$$= \frac{n-1}{n^2}\left(\left(x^*-x_n\right)^2+(n-1)\left(x^*-x_n\right)^2\right) = \frac{n-1}{n}\left(x^*-x_n\right)^2$$

und

$$\sum_i\left(y_i-\bar{y}\right)\left(x_i-\bar{x}\right) = \sum_i y_i\left(x_i-\bar{x}\right) =$$

$$= \sum_{i=1}^{n-1} y_i\left(x^*-\frac{(n-1)x^*+x_n}{n}\right) + y_n\left(x_n-\frac{(n-1)x^*+x_n}{n}\right) =$$

$$= \sum_{i=1}^{n-1} y_i\frac{x^*-x_n}{n} + y_n\left(\frac{(n-1)\left(x_n-x^*\right)}{n}\right) =$$

$$= \frac{x^*-x_n}{n}\sum_{i=1}^{n-1} y_i - \frac{(n-1)\left(x^*-x_n\right)}{n}y_n =$$

$$= \frac{n-1}{n}\left(x^*-x_n\right)\left(\bar{y}_{n-1}-y_n\right)$$

$$mit \quad \bar{y}_{n-1}= \frac{1}{n-1}\sum_{i=1}^{n-1} y_i \ .$$

Somit gilt

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\bar{y}_{n-1} - y_n}{x - x_n}$$

Setzt man den Schätzer $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ in die geschätzte Regressionsgerade $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ ein, so erhält man

$$\hat{y} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x = \bar{y} + \hat{\beta}_1 (x - \bar{x}) \qquad (2.13)$$

Wir berechnen zunächst den Wert von $\hat{y}$ an der Stelle x^*.
Es gilt

$$x^* - \bar{x} = x^* - \frac{(n-1)x^* + x_n}{n} = \frac{x^* - x_n}{n} \qquad (2.14)$$

Setzt man (2.14) in (2.13) ein, so ergibt sich

$$\hat{y} = \bar{y} + \frac{x^* - x_n}{n} \frac{\bar{y}_{n-1} - y_n}{x^* - x_n} =$$

$$= \frac{1}{n} \sum_{i=1}^{n} y_i + \frac{1}{n} \left(\frac{1}{n-1} \sum_{i=1}^{n-1} y_i - y_n \right) =$$

$$= \frac{1}{n(n-1)} \left((n-1) \sum_{i=1}^{n} y_i - (n-1) y_n + \sum_{i=1}^{n-1} y_i \right) =$$

$$= \frac{1}{n(n-1)} \left((n-1) \sum_{i=1}^{n-1} y_i + \sum_{i=1}^{n-1} y_i \right) = \frac{1}{n-1} \sum_{i=1}^{n-1} y_i = \bar{y}_{n-1}$$

Die geschätzte Regressionsgerade geht also durch den Mittelwert der y_i am ersten Beobachtungspunkt.
Wir berechnen nun den Wert von $\hat{y}$ an der Stelle x_n.
Es gilt $x_n - \bar{x} = \frac{n-1}{n}\left(x_n - x^*\right)$. Setzen wir diese Gleichung in (2.14) ein, so ergibt sich als Wert von $\hat{y}$ an der Stelle x_n:

$$\hat{y} = \bar{y} + \frac{n-1}{n} \left(x_n - x^*\right) \frac{\bar{y}_{n-1} - y_n}{x^* - x_n} = \bar{y} - \frac{n-1}{n} \left(\bar{y}_{n-1} - y_n\right) =$$

$$= \frac{1}{n} \left(\sum_{i=1}^{n} y_i - \sum_{i=1}^{n-1} y_i + (n-1) y_n \right) = \frac{1}{n} \left(y_n + n y_n - y_n \right) = y_n$$

(c) Es gilt

$$\sum_i (x_i-\bar{x})(y_i-\bar{y}) = \sum_i (x_i-\bar{x})y_i + \sum_i (x_i-\bar{x})\bar{y} =$$

$$= \sum_i (x_i-\bar{x})y_i + \bar{y}\sum_i (x_i-\bar{x}) = \sum_i (x_i-\bar{x})y_i, \text{ denn}$$

denn $\sum_i (x_i-\bar{x}) = \sum_i x_i - \sum_i \bar{x} = \sum_i x_i - n\bar{x} = \sum_i x_i - \sum_i x_i = 0$

(d) Die Differenz aus y_i und $\hat{y}_i$ wird als i.tes Residuum e_i bezeichnet. Der Residuenvektor ist dann $e=y-\hat{y}$.
Wegen $e_i = y_i-\hat{y}_i = y_i-\hat{\beta}_0-\hat{\beta}_1 x_i$ können wir die Normalgleichungen (2.5) und (2.6) auch schreiben als

$$\sum_i e_i = 0 \qquad (2.15)$$

$$\sum_i e_i x_i = 0 \qquad (2.16)$$

Multiplizieren wir (2.15) mit $\hat{\beta}_0$ und (2.16) mit $\hat{\beta}_1$ und addieren die so gewonnen Gleichungen, so gilt

$$\sum_i \left(\hat{\beta}_0+\hat{\beta}_1 x_i\right) e_i = 0$$

und somit

$$\sum_i \hat{y}_i e_i = 0 \qquad (2.17)$$

(e) Multiplizieren wir beide Seiten von (2.15) mit $\frac{1}{n}$, so folgt $\bar{e}=0$. Der Mittelwert der Residuen ist also 0.

(f) Mit $e_i=y_i-\hat{y}_i$ folgt aus (2.15)
$\sum_i y_i = \sum_i \hat{y}_i$ und damit $\bar{\hat{y}}=\bar{y}$. Der Mittelwert der $\hat{y}_i$ ist also dem Mittelwert der y_i .

(g) Wegen $\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$ liegt der Punkt $(\bar{x},\bar{y})'$ auf der geschätzten Regressionsgeraden.

2.3 Das Bestimmtheitsmaß

Nachdem die Parameter geschätzt sind, stellt sich die Frage, wie gut die Anpassung ist. Dazu gehen wir aus von

$$y_i - \bar{y} = y_i - \hat{y}_i + \hat{y}_i - \bar{y} \qquad (2.18)$$

Quadrieren wir beide von (2.18) und summieren über i, so gilt

$$\sum_i \left(y_i - \bar{y}\right)^2 = \sum_i \left(y_i - \hat{y}_i\right)^2 + \sum_i \left(\hat{y}_i - \bar{y}\right)^2 + 2\sum_i \left(y_i - \hat{y}_i\right)\left(\hat{y}_i - \bar{y}\right)$$

Nun gilt

$$\sum_i \left(y_i - \hat{y}_i\right)\left(\hat{y}_i - \bar{y}\right) = \sum_i \left(y_i - \hat{y}_i\right)\hat{y}_i - \sum_i \left(y_i - \hat{y}_i\right)\bar{y} =$$

$$= \sum_i e_i \hat{y}_i - \bar{y}\sum_i e_i = 0 \text{ wegen (2.17) und (2.15)}$$

Es gilt also die folgende fundamentale Beziehung

$$\sum_i \left(y_i - \bar{y}\right)^2 = \sum_i \left(\hat{y}_i - \bar{y}\right)^2 + \sum_i \left(y_i - \hat{y}_i\right)^2$$

Die Summe der quadrierten Abweichungen der y_i von ihrem Mittelwert können wir als zerlegen in zwei additive Komponenten. Wir wählen folgende Bezeichnungen:

$$S_{yy} = \sum_i \left(y_i - \bar{y}\right)^2$$

$$SS_R = \sum_i \left(\hat{y}_i - \bar{y}\right)^2$$

$$SS_E = \sum_i \left(y_i - \hat{y}_i\right)^2$$

S_{yy} ist die Summe der quadrierten Abweichungen der y_i von ihrem Mittelwert, SS_R ist die Summe der quadrierten Abweichungen der geschätzten y_i von ihrem Mittelwert, und SS_E ist die Summe der quadrierten Abweichungen der y_i von ihren geschätzten Werten. SS_E ist der Teil der Streuung der y_i, der nicht durch die Regression, d.h. durch die Gerade, erklärt wird, während SS_R der Teil der Streuung der y_i ist, der durch die Regression erklärt wird. SS_E berücksichtigt ja gerade die Abweichungen der y_i von der Geraden.

Wir können diese Größen auch folgendermaßen gewinnen:
Wir gehen aus von den nachstehenden Modellen

Modell I $\quad Y_i=\beta_0+\varepsilon_i \qquad (2.20)$

Modell II $\quad Y_i=\beta_0+\beta_1 x_i+\varepsilon_i \qquad (2.21)$

Im Modell I setzt sich eine Zufallsvariable Y_i additiv aus einem typischen Wert β_0 und einer zufälligen Störung ε_i zusammen. Die Zufallsvariable Y_i hängt im Modell I überhaupt nicht von x_i ab, während im Modell II ein linearer Zusammenhang zwischen Y_i und x_i postuliert wird, der durch eine additive Störung ε_i überlagert wird. Gilt im Modell II $\beta_0=0$, so sind Modell I und Modell II identisch.

Die Schätzungen der Parameter geben wir nun in beiden Modellen für die Realisationen y_i der Y_i an.

Wir schätzen zunächst den Parameter β_0 im Modell I mit der Methode der Kleinsten Quadrate. Es ist zu minimieren

$$S(\beta_0) = \sum_i (y_i-\beta_0)^2$$

Wir differenzieren $S(\beta_0)$ nach β_0

$$\frac{d}{d\beta_0}S(\beta_0) = (-2)\sum_i (y_i-\beta_0) \stackrel{!}{=} 0$$

Der Schätzer $\hat{\beta}_0$ von β_0 muß also folgende Gleichung erfüllen:

$$\sum_i (y_i-\hat{\beta}_0) = 0 \qquad (2.22)$$

Aus (2.22) folgt $\hat{\beta}_0 = \bar{y}$.

Für das Modell I erhalten wir nun die Residuenquadratsumme

$$SS_E^I = S(\hat{\beta}_0) = \sum_i (y_i-\bar{y})^2 = S_{yy} \qquad (2.23)$$

Aus (2.19) erhalten wir als Residuenquadratsumme im Modell II

$$SS_E^{II} = S_{yy}-SS_R = SS_E^I-SS_R \qquad (2.24)$$

Die Beziehung (2.24) zeigt folgendes:

(a) Die Residuenquadratsumme im Modell II ist kleiner gleich der Residuenquadratsumme im Modell I. Der Übergang von Modell I zu Modell II, d.h. die Hinzunahme der Variablen x, führt dazu, daß die Residuenquadratsumme entweder gleich bleibt oder abnimmt.

(b) SS_R ist die Differenz aus SS_E^I und SS_E^{II}, gibt also an, in welchem Maße SS_E^I verringert wird, wenn die Variable x ins Modell aufgenommen wird. Es ist sicher nur dann sinnvoll, von Modell I zu Modell II überzugehen, wenn die Differenz zwischen SS_E^I und SS_E^{II} groß genug ist. Es liegt also nahe, einen Test auf $H_0{:}\beta_1=0$ auf SS_R aufzubauen. Wir kommen auf diesen Test im Abschnitt 2.5 zurück.

Je größer nun SS_R ist, umso mehr wird von der Streuung der y_i durch die Regression erklärt. Ein sinnvolles Maß für die Anpassung ist also

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}} \qquad (2.25)$$

R^2 heißt Bestimmtheitsmaß.

Zwischen dem Bestimmtheitsmaß R^2 und dem Korrelationskoeffizienten r_{xy} besteht folgender Zusammenhang:

$$R^2=r^2_{xy} \qquad (2.26).$$

Dabei ist r_{xy} definiert durch

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}\,S_{yy}}} \qquad (2.27)$$

Um (2.26) zu zeigen, wählen wir eine andere Darstellung für SS_R. Hierzu formen wir zunächst SS_E um.

$$\begin{aligned}
SS_E &= \sum_i e_i^2 = \sum_i (y_i-\hat\beta_0-\hat\beta_1 x_i)^2 = \sum_i (y_i-\bar y+\hat\beta_1\bar x-\hat\beta_1 x)^2 = \\
&= \sum_i\Big((y_i-\bar y)-\hat\beta_1(x_i-\bar x)\Big)^2 = \\
&= \sum_i (y_i-\bar y)^2-2\hat\beta_1\sum_i (y_i-\bar y)(x_i-\bar x)+\hat\beta_1^2\sum_i (x_i-\bar x)^2 = \\
&= S_{yy}-2\hat\beta_1 S_{xy}+\hat\beta_1^2 S_{xx} = S_{yy}-2\hat\beta_1 S_{xy}+\hat\beta_1 S_{xy} = S_{yy}-\hat\beta_1 S_{xy} \qquad (2.28)
\end{aligned}$$

Aus (2.19) folgt somit im Zusammenhang mit (2.28) und (2.10)

$$SS_R = \hat{\beta}_1 S_{xy} = \frac{S_{xy}^2}{S_{xx}} \tag{2.29}$$

Setzen wir nun (2.29) in (2.25) ein, so gilt

$$R^2 = \frac{SS_R}{S_{yy}} = \frac{S_{xy}^2}{S_{xx}S_{yy}} = r_{xy}^2$$

2.4 Eigenschaften der Schätzer

Bisher haben wir die Regressionsbeziehung zwischen den Variablen x und y ausschließlich auf der Ebene der beobachteten Werte betrachtet. Die $y_i, i=1,\ldots,n$ sind aber Realisationen von Zufallsvariablen $Y_i, i=1,\ldots,n$. Die Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ sind somit als Funktionen der y_i ebenfalls Realisationen von Zufallsvariablen und können somit auch auf der Ebene der zugrundeliegenden Zufallsvariablen betrachtet werden. Wir werden im folgenden bei der Notation keinen Unterschied zwischen Schätzern und Schätzfunktionen $\hat{\beta}_0$ und $\hat{\beta}_1$ machen. Es ist immer aus dem Zusammenhang ersichtlich, auf welcher Ebene wir uns befinden. Wir fassen $\hat{\beta}_0$ und $\hat{\beta}_1$ als Schätzfunktionen auf und untersuchen deren Eigenschaften.

Aus dem Regressionsmodell

$$Y_i = \beta_0+\beta_1 x_i+\varepsilon_i \qquad i=1,\ldots,n$$

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

$$\text{und} \quad Cov(\varepsilon_i,\varepsilon_j) = \begin{cases} \sigma^2 & \text{für } i=j \\ 0 & \text{für } i\neq j \end{cases}$$

folgt zunächst

$$E(Y_i) = E(\beta_0+\beta_1 x_i+\varepsilon_i) = \beta_0+\beta_1 x_i+E(\varepsilon_i) = \beta_0+\beta_1 x_i \tag{2.30}$$

und

$$Var(Y_i) = Var(\beta_0+\beta_1 x_i+\varepsilon_i) = Var(\varepsilon_i) = \sigma^2 \tag{2.31}$$

für $i=1,\ldots,n$.

Für $\hat{\beta}_1$ gilt:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_i \left(x_i - \bar{x}\right) Y_i}{\sum_i \left(x_i - \bar{x}\right)^2} = \sum c_i Y_i$$

$$\text{mit } c_i = \frac{\left(x_i - \bar{x}\right)}{\sum_i \left(x_i - \bar{x}\right)^2} = \frac{x_i - \bar{x}}{S_{xx}} \tag{2.32}$$

Es gilt

$$\sum_i c_i = \frac{1}{\sum_i \left(x_i - \bar{x}\right)^2} \sum_i \left(x_i - \bar{x}\right) = 0 \tag{2.33}$$

und

$$\sum_i c_i^2 = \frac{1}{S_{xx}^2} \sum_i \left(x_i - \bar{x}\right)^2 = 1 \tag{2.34}$$

Also folgt

$$E(\hat{\beta}_1) = E(\sum_i c_i Y_i) = \sum_i c_i E(Y_i) = \sum_i c_i (\beta_0 + \beta_1 x_i) =$$

$$= \beta_0 \sum_i c_i + \beta_1 \sum_i c_i x_i = \frac{\beta_1}{S_{xx}} \sum_i (x_i - \bar{x}) x_i = \beta_1$$

$$\text{wegen} \quad S_{xx} = \sum_i (x_i - \bar{x})^2 = \sum_i (x_i - \bar{x}) x_i - \sum_i (x_i - \bar{x}) \bar{x} =$$

$$= \sum_i (x_i - \bar{x}) x_i - \bar{x} \sum_i (x_i - \bar{x}) = \sum_i (x_i - \bar{x}) x_i$$

$\hat{\beta}_1$ ist also eine erwartungstreue Schätzfunktion für β_1.

Da die $\varepsilon_i, i=1,\ldots,n$ unabhängig sind, sind auch die $Y_i, i=1,\ldots,n$ als Funktionen der ε_i unabhängig.

Es gilt also

$$Var(\hat{\beta}_1) = Var(\sum_i c_i Y_i) = \sum_i c_i^2 Var(Y_i) = \sigma^2 \sum_i c_i^2 = \frac{\sigma^2}{S_{xx}} \tag{2.35}$$

wegen (2.31) und (2.34).

Ohne Herleitung geben wir an:

$$E(\hat{\beta}_0) = \beta_0 \qquad (2.36)$$

$$Var(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \qquad (2.37)$$

Außerdem gilt

$$Cov(\hat{\beta}_0,\hat{\beta}_1) = - \frac{\bar{x}\,\sigma^2}{S_{xx}} \qquad (2.38)$$

2.5 Schätzen der Varianz der Störterme

Neben β_0 und β_1 ist auch die Varianz σ^2 der Störterme ein unbekannter Parameter.
Wie wir in Kapitel 3 zeigen, ist

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_i e_i^2 = \sum_i (y_i-\hat{y}_i)^2 \qquad (2.39)$$

Die Beziehung (2.28) ermöglicht die Berechnung von $\hat{\sigma}^2$ ohne die explizite Bestimmung der Residuen.
Für das Datenbeispiel gilt: $S_{yy}=22.078$, $S_{xy}=-14.8$ und $\hat{\beta}_1=-1.48$.
Somit erhalten wir als Schätzer für σ^2: $\hat{\sigma}^2= 0.058$.
Der Schätzer $\hat{\sigma}^2$ ist modellabhängig, da er von den $\hat{y}_i$ abhängt.
Um einen modellunabhängigen Schätzer für σ^2 zu erhalten, muß man an den Meßpunkten x_i mehrere Werte der Variablen Y beobachten. Wir kommen hierauf in Abschnitt 2.8 zurück.

2.6 Hypothesentests und Konfidenzintervalle

Um Tests durchführen zu können oder Konfidenzintervalle aufstellen zu können, benötigen wir die Verteilung der Teststatistiken unter der Nullhypothese. Da die Teststatistiken bei Tests auf β_0 oder β_1 sinnvollerweise auf $\hat{\beta}_0$ oder $\hat{\beta}_1$ beruhen sollten, diese aber von den Y_i abhängen, die wiederum Funktionen der ε_i sind, ist wir zur Herleitung der Verteilung der Teststatistiken unter H_0 die Verteilung der Störterme $\varepsilon_i, i=1,\ldots,n$. Unterstellt man ,daß die Störterme normalverteilt sind, so folgen die Teststatistiken bekannten Verteilungen.

Da bei Normalverteilung Unkorreliertheit und Unabhängigkeit äquivalent sind, unterstellen wir also, daß die Störterme $\varepsilon_1,\ldots,\varepsilon_n$ unabhängig sind und $\varepsilon_i \approx N(0,\sigma^2)$ gilt.

Da gilt $Y_i=\beta_0+\beta_1 x_i+\varepsilon_i$, sind also die Y_i als lineare Funktionen der ε_i ebenfalls normalverteilt mit $E(Y_i)=\beta_0+\beta_1 x_i$ und $Var(Y_i)=\sigma^2$. Sowohl $\hat{\beta}_0$ als auch $\hat{\beta}_1$ sind lineare Funktionen der normalverteilten Y_i und somit ebenfalls normalverteilt.

Die Erwartungswerte und Varianzen von $\hat{\beta}_0$ und $\hat{\beta}_1$ wurden in Abschnitt 2.4 hergeleitet. Es gilt somit

$$\hat{\beta}_0 \approx N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right) \qquad (2.40)$$

$$\hat{\beta}_1 \approx N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right) \qquad (2.41)$$

und natürlich

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \approx N(0,1) \qquad (2.42)$$

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{S_{xx}}}} \approx N(0,1) \qquad (2.43)$$

Unter H_0: $\beta_0 = \beta_{00}$ gilt also

$$\frac{\hat{\beta}_0 - \beta_{00}}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \approx N(0,1) \tag{2.44}$$

und unter H_0: $\beta_1 = \beta_{10}$ gilt also

$$\frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\frac{\sigma^2}{S_{xx}}}} \approx N(0,1) \tag{2.45}$$

Die Statistiken in (2.44) und (2.45) hängen von der Varianz σ^2 der Störterme ab, die in der Regel unbekannt ist. Es liegt nahe, σ^2 durch $\hat{\sigma}^2$ zu schätzen und in die Statistiken in (2.44) und (2.45) einzusetzen. Wir erhalten dann die Teststatistiken

$$T_0 = \frac{\hat{\beta}_0 - \beta_{00}}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \tag{2.46}$$

und

$$T_1 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} \tag{2.47}$$

Wie wir in Kapitel 3 zeigen werden, gilt:

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \approx \chi^2(n-2) \tag{2.48}$$

und sowohl $\hat{\beta}_0$ und $\hat{\sigma}^2$ als auch $\hat{\beta}_1$ und $\hat{\sigma}^2$ sind unabhängig.

Diese Eigenschaften erlauben die Bestimmung der Verteilungen der Teststatistiken (2.46) und (2.47) unter den jeweiligen Nullhypothesen.Wir zeigen dies exemplarisch für (2.47).

$$T_1 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}}} = \frac{\dfrac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\dfrac{\sigma^2}{S_{xx}}}}}{\sqrt{\dfrac{\hat{\sigma}^2}{\sigma^2}}} \qquad (2.49)$$

Die Struktur von T_1 ist:

$$T_1 = \frac{Z}{\sqrt{\dfrac{U}{n-2}}} \quad \text{mit } Z = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\dfrac{\sigma^2}{S_{xx}}}} \quad \text{und } U = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} .$$

Dabei ist Z unter H_0: $\beta_1 = \beta_{10}$ standardnormalverteilt,und U folgt einer Chiquadratverteilung mit n-2 Freiheitsgraden. Außerdem sind Z und U als Funktionen unabhängiger Größen unabhängig. Also ist T_1 unter H_0 t-verteilt mit n-2 Freiheitsgraden.

Aus ähnlichen Überlegungen folgt, daß auch T_0 unter H_0: $\beta_0=\beta_{00}$ t-verteilt ist mit n-2 Freiheitsgraden.

Zum Signifikanzniveau α erhält man also folgende Entscheidungsregeln bei den einzelnen Hypothesen:

$$H_0: \beta_0 = \beta_{00}$$
$$H_1: \beta_0 \neq \beta_{00}$$

H_0 ablehnen, wenn gilt $|T_0| \geq t_{n-2;1-\alpha/2}$, wobei $t_{n-2;1-\alpha/2}$ das $1-\alpha/2$-Quantil einer t-Verteilung mit n-2 Freiheitsgraden ist.

$$H_0: \beta_1 = \beta_{10}$$
$$H_1: \beta_1 \neq \beta_{10}$$

H_0 ablehnen, wenn gilt $|T_1| \geq t_{n-2;1-\alpha/2}$

Ohne Schwierigkeiten erhält man nun auch Konfidenzintervalle.

Konfidenzintervall für β_0 zum Konfidenzniveau 1-α :

$$\left[\hat{\beta}_0 - t_{n-2;1-\alpha/2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)},\ \hat{\beta}_0 + t_{n-2;1-\alpha/2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)}\ \right]$$

Konfidenzintervall für β_0 zum Konfidenzniveau 1-α :

$$\left[\hat{\beta}_1 - t_{n-2;1-\alpha/2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}\ ,\ \hat{\beta}_1 + t_{n-2;1-\alpha/2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}\right]$$

Für unser Beispiel erhalten wir folgende Konfidenzintervalle für β_0 und β_1 zum Konfidenzniveau 1-α=0.95:

β_0 : $\left[13.56, 15.62\right]$ und β_1 : $\left[-1.72, -1.23\right]$.

Wir betrachten nun noch den Test folgender Hypothese genauer:

$$H_0{:}\beta_1=0 \text{ gegen } H_1{:}\beta_1\neq 0$$

Hier wird überprüft, ob dir Variable x überhaupt zur Erklärung von y benötigt wird.

Das Quadrat der Teststatistik T_1 lautet für diese Testproblem:

$$F = T_1^2 = \frac{\hat{\beta}_1^2}{\frac{\hat{\sigma}^2}{S_{xx}}} = \frac{\hat{\beta}_1^2 S_{xx}}{\hat{\sigma}^2} = \frac{SS_R}{\frac{SS_E}{n-2}} \qquad (2.50)$$

da mit (2.10) und (2.28) gilt $SS_R = \hat{\beta}_1 S_{xy} = \hat{\beta}_1^2 S_{xx}$.

Unter $H_0{:}\beta_1=0$ ist die Teststatistik F als Quadrat einer mit n-2 Freiheitsgraden t-verteilten Zufallsvariablen F-verteilt mit 1 und n-2 Freiheitsgraden.

In die Teststatistik F geht also sowohl der durch die Regression erklärte Teil der Streuung der y_i als auch der nicht durch die Regression erklärte Teil der Streuung der y_i ein. Die Nullhypothese $H_0{:}\beta_1=0$ wird abgelehnt, wenn der Zähler von F im Verhältnis zum Nenner groß ist, d.h. wenn zum Niveau α gilt: $F \geq F_{1,n-2;1-\alpha}$, wobei $F_{1,n-2;1-\alpha}$ das 1-α-Quantil einer F-Verteilung mit 1 und n-2 Freiheitsgraden ist.

Sämtliche bei diesem Test benutzten Informationen werden in der sog. ANOVA-Tabelle zusammengestellt. Das Akronym ANOVA steht für Analysis of Variance (Varianzanalyse).

Source	Sum of Squares	df	Mean Square	F	p-value
Regression	SS_R	1	SS_R (=Z)	$\frac{Z}{N}$	
Residual	SS_E	n-2	$\frac{SS_E}{n-2}$ (=N)		
Total	S_{yy}	n-1			

Die Größen sind noch einmal im einzelnen:

$$SS_R = \sum_i (\hat{y}_i - \bar{y})^2 = \hat{\beta}_1 S_{xy}$$

$$SS_E = \sum_i (y_i - \hat{y}_i)^2$$

$$S_{yy} = \sum_i (y_i - \bar{y})^2$$

In der ANOVA-Tabelle ist auch der erwartungtreue Schätzer für σ^2 zu finden, nämlich $\frac{SS_E}{n-2}$. Außerdem kann man aus der ANOVA-Tabelle ohne Schwierigkeiten das Bestimmtheitsmaß $R^2 = \frac{SS_R}{S_{yy}}$ berechnen.

Der p-value (Überschreitungswahrscheinlichkeit) gibt an, wie groß unter H_0 die Wahrscheinlichkeit ist, Werte der Teststatistik zu beobachten, die größer oder gleich dem realisierten sind,falls die Nullhypothese für "große" Werte der Teststatistik abgelehnt wird.. Ist der p-value also "klein", so spricht dies gegen H_0. Die Nullhypothese wird H_0somit zum Niveau α abgelehnt, wenn der p-value kleiner als α ist.

Für das Beispiel erhalten wir folgende Anova Tabelle:

Source	Sum of Squares	df	Mean Square	F	p-value
Regression	21.904	1	21.904	377.7	0.0003
Residual	0.174	3	0.058		
Total	22.078	4			

Da gilt $0.0003<0.05$ lehnen wir zum Niveau $\alpha=0.05$ die Nullhypothese $H_0:\beta_1=0$ ab.
Das Bestimmtheitsmaß R^2 ist 0.9921.

2.7 Ein Konfidenzintervall für den Erwartungswert von Y

Ein Regressionsmodell wird in der Regel aufgestellt, um den Erwartungswert von Y bei einem vorgegebenen Wert x_0 von x zu schätzen. So könnte sich ein Autoverkäufer für den erwarteten Verkaufspreis eines 4 Jahre alten Autos interessieren.
Wir bezeichnen im folgenden den Erwartungswert von Y an der Stelle x_0 mit $E(Y|x_0)$.

Offensichtlich gilt: $E(Y|x_0)=\beta_0+\beta_1 x_0$.

Somit liegt es, nahe folgende Schätzfunktion zu wählen:

$$\hat{Y}_0=\hat{\beta}_0+\hat{\beta}_1 x_0 \qquad (2.51)$$

Da gilt $E\left(\hat{Y}_0\right)=E\left(\hat{\beta}_0+\hat{\beta}_1 x_0\right)=\beta_0+\beta_1 x_0$ ist $\hat{Y}_0$ eine erwartungtreue Schätzfunktion für $E(Y|x_0)$. Die Varianz von $\hat{Y}_0$ kann man sich folgendermaßen herleiten:

$$\begin{aligned}
Var\left(\hat{Y}_0\right) &= Var\left(\hat{\beta}_0+\hat{\beta}_1 x_0\right) = Var\left(\hat{\beta}_0\right) + Var\left(\hat{\beta}_1 x_0\right) + 2\ Cov\left(\hat{\beta}_0,\hat{\beta}_1 x_0\right) = \\
&= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) + x_0^2\frac{\sigma^2}{S_{xx}} + 2x_0 Cov\left(\hat{\beta}_0,\hat{\beta}_1\right) = \\
&= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} + \frac{x_0^2}{S_{xx}}\right) - 2x_0\frac{\bar{x}\,\sigma^2}{S_{xx}} = \\
&= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2 + x_0^2 - 2x_0\bar{x}}{S_{xx}} \right) = \\
&= \sigma^2\left(\frac{1}{n} + \frac{\left(\bar{x} - x_0\right)^2}{S_{xx}}\right) \qquad (2.51)
\end{aligned}$$

$\hat{Y}_0$ ist nun als Linearkombination der normalverteilten Zufallsvariablen $\hat{\beta}_0$ und $\hat{\beta}_1$ normalverteilt. Somit gilt

$$Z = \frac{\hat{Y}_0 - E(Y|x_0)}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\left(\bar{x} - x_0\right)^2}{S_{xx}} \right)}} \approx N(0,1)$$

Schätzen wir nun σ^2 durch die erwartungtreue Schätzfunktion $\hat{\sigma}^2$, so erhalten wir wie in Abschnitt 2.6 eine mit n-2 Freiheitsgraden t-verteilte Zufallsgröße, da $\hat{\beta}_0$ und $\hat{\sigma}^2$ und $\hat{\beta}_1$ und $\hat{\sigma}^2$ unabhängig sind und außerdem gilt $\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \approx \chi^2(n-2)$. Es gilt also

$$T_2 = \frac{\hat{Y}_0 - E(Y|x_0)}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\left(\bar{x} - x_0\right)^2}{S_{xx}}\right)}} \approx t(n-2) \qquad (2.52)$$

Ein Konfidenzintervall zum Konfidenzniveau 1-α für $E(Y|x_0)$ hat also die Grenzen:

$$\hat{Y}_0 - t_{n-2;1-\alpha/2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\left(\bar{x} - x_0\right)^2}{S_{xx}}\right)}$$

$$\hat{Y}_0 + t_{n-2;1-\alpha/2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\left(\bar{x} - x_0\right)^2}{S_{xx}}\right)}$$

Wie man an den obigen Gleichungen sofort erkennen kann, ist das Konfidenzintervall für $x_0=\bar{x}$ am schmalsten. Je weiter x_0 von $\bar{x}$ entfernt ist, um so breiter wird das Konfidenzintervall. In der folgenden Tabelle sind die Konfidenzintervalle zum Konfidenzniveau 1-α=0.95 für $E(Y|x_0)$ an ausgewählten Stellen x_0 angegeben.

x	$\hat{y}_0$	untere Grenze	obere Grenze	Breite
2	11.63	11.04	12.22	1.18
3	10.15	9.73	10.57	0.84
4	8.67	8.33	9.01	0.68
5	7.19	6.77	7.61	0.84
6	5.71	5.12	6.30	1.18

2.8 Ein Test auf Linearität

Eine zentrale Annahme des linearen Modell ist, daß der Zusammenhang zwischen den Variablen x und y linear in den Parametern β_0 und β_1 ist. Diese Annahme kann nicht überprüft werden, wenn an jeder Stelle $x_i, i=1,...,n$, nur ein Wert der abhängigen Variablen beobachtet wurde. Wir gehen deshalb im folgenden davon aus, daß k Meßpunkte $x_1,...,x_k$ vorliegen, und am Meßpunkt x_i n_i Werte der Variablen Y beobachtet wurden. Sei also y_{ij} die j-te Beobachtung von Y an der Stelle $x_i, j=1,...,n_i$. Insgesamt gibt es $n=\sum_i n_i$ Beobachtungen.

Sei nun $\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1 x_i$, wobei $\hat{\beta}_0$ und $\hat{\beta}_1$ die Schätzer von β_0 und β_1 nach der Methode der Kleinsten Quadrate sind. Außerdem sei $\bar{y}_i=\sum_{j=1}^{n_i} y_{ij}$ das arithmetische Mittel der y_{ij} an der Stelle x_i.

Wir können nun das Residuum $e_{ij}=y_{ij}-\hat{y}_i$ in zwei additive Komponenten zerlegen:

$$y_{ij}-\hat{y}_i=y_{ij}-\bar{y}_i+\bar{y}_i-\hat{y}_i \qquad (2.53)$$

Quadrieren wir nun beide Seiten dieser Gleichung und summieren dann über i und j, so ergibt sich:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\hat{y}_i\right)^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\left(y_{ij}-\bar{y}_i\right)+\left(\bar{y}_i-\hat{y}_i\right)\right)^2=$$

$$=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_i\right)^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\bar{y}_i-\hat{y}_i\right)^2+2\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_i\right)\left(\bar{y}_i-\hat{y}_i\right)$$

Nun gilt $\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_i\right)\left(\bar{y}_i-\hat{y}_i\right)=\sum_{i=1}^{k}\left(\bar{y}_i-\hat{y}_i\right)\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_i\right)=0$.

Also gilt $$SS_E = SS_{PE} + SS_{LOF} \qquad (2.54)$$
mit

$$SS_E = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(y_{ij} - \hat{y}_i\right)^2$$

$$SS_{PE} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y}_i\right)^2$$

$$SS_{LOF} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\bar{y}_i - \hat{y}_i\right)^2 = \sum_{i=1}^{k} n_i \left(\bar{y}_i - \hat{y}_i\right)^2$$

LOF steht für Lack of fit. In SS_{LOF} gehen die Abweichungen der Mittelwerte der y_i von der geschätzten Geraden ein. Sind diese Abweichungen groß, so wird der Datensatz schlecht durch eine Gerade beschrieben.

PE steht für Pure Error. In SS_{PE} gehen nur die Abweichungen der y_i von ihrem jeweiligen Mittelwert ein. SS_{PE} ist also ein <u>modellunabhängiges</u> Maß für die Streuung der Beobachtungen.

Mit SS_{PE} und SS_{LOF} können wir einen Test auf die Güte der Anpassung konstruieren. Ist SS_{LOF} im Verhältnis zu SS_{PE} groß, so deutet dies darauf hin, daß das Modell falsch ist.

Da die Y_{ij} unter den Annahmen des linearen Modells unabhängig und identisch mit Erwartungswert 0 und Varianz σ^2 normalverteilt sind, ist

$$\frac{1}{\sigma^2} \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y}_i\right)^2 \sim \chi^2(n_i - 1)$$

Also ist $\frac{1}{\sigma^2} SS_{PE}$ als Summe von k unabhängigen mit jeweils n_i-1 Freiheitsgraden chiquadratverteilten Zufallsvariablen chiquadratverteilt mit $\sum_i (n_i - 1) = n - k$ Freiheitsgraden.

Man kann nun noch zeigen, daß $\frac{1}{\sigma^2} SS_{LOF}$ chiquadratverteilt mit k-2 Freiheitsgraden ist, und daß SS_{LOF} und SS_{PE} unabhängig sind.

Die Teststatistik für den Test auf Linearität ist:

$$F_L = \frac{\frac{SS_{LOF}}{k-2}}{\frac{SS_{PE}}{n-k}}$$

Das Testproblem lautet:

$$H_0:\ E(Y_i)=\beta_0+\beta_1 x_i \qquad \text{für alle } i=1,...,n$$

$$H_1:\ E(Y_i)\neq\beta_0+\beta_1 x_i \qquad \text{für mind. ein } i$$

H_0 wird zum Signifikanzniveau α abgelehnt, wenn gilt $F_L \geq F_{k-2,n-k;1-\alpha}$, wobei $F_{k-2,n-k;1-\alpha}$ das 1-α-Quantil einer F-Verteilung mit k-2 und n-k Freiheitsgraden ist.

Beispiel:

In der folgenden Tabelle sind die von der Arbeitsgemeinschaft deutscher wirtschaftswissenschaftlicher Forschungsinstitute prognostizierten Werte x_i und die realisierten Werte y_i für die Wachstumsrate des realen Bruttosozialprodukts für die Jahre 1980 bis 1989 wiedergegeben.

Jahr	x_i	y_i
1980	2.5	1.5
1981	0.0	0.0
1982	1.0	-1.0
1983	0.0	1.9
1984	2.0	3.3
1985	2.0	1.9
1986	3.0	2.3
1987	3.0	1.6
1988	2.0	3.7
1989	2.0	3.9

Die ANOVA-Tabelle des Datensatzes ist

Source	Sum of Squares	df	Mean Square	F	p-value
Regression	4.1309	1	4.1309	1.89	0.2067
Residual	17.4981	8	2.1873		
Total	21.629	9			

Es gilt also SS_E=17.4981. Wir berechnen zunächst SS_{PE} mit folgender Hilfstabelle, bei der wir nur die Werte von x_i berücksichtigen, die mehrfach vorkommen.

x_i	y_{ij}	$\bar{y}_i$	$\sum_j \left(y_{ij} - \bar{y}_i\right)^2$	df
0	1.9,0	0.95	1.805	1
2	3.9,3.7,1.9,3.3	3.2	2.44	3
3	1.6,2.3	1.95	0.245	1

Wir erhalten somit $SS_{PE} = 4.49$ mit 5 Freiheitsgraden. Somit gilt $SS_{LOF}=17.498-4.49=13.008$ mit 3 Freiheitsgraden. Es ergibt sich $F_L=4.828$. Wegen $F_{3,5;0.95} = 5.41$ wird H_0 nicht abgelehnt.

3. Das lineare Modell in Matrixschreibweise

3.1 Das Modell

Im letzten Kapitel haben wir uns mit der linearen Einfachregression beschäftigt. Die Herleitung der Schätzer, ihrer Erwartungswerte und Varianzen war hier ohne Matrizenrechnung möglich, ohne daß der Überblick verloren ging. Hängt die zu erklärende Variable Y aber von mehreren erklärenden Variablen $x_1,\dots,x_p$ ab, so ist es sinnvoll, das lineare Modell in Matrixschreibweise darzustellen und in diesem Rahmen die Schätzer und deren Eigenschaften herzuleiten.
Wir gehen also aus von folgendem Modell

$$Y = \beta_0+\beta_1x_1+\beta_2x_2+\dots+\beta_px_p+\varepsilon$$

Um die unbekannten Parameter $\beta_0,\beta_1,\dots,\beta_p$ zu schätzen, werden $y_i,x_{i1},x_{i2},\dots,x_{ip}$ $i=1,\dots,n$ beobachtet bzw. gemessen. Für die den Realisationen $y_1,\dots,y_n$ zugrundeliegenden Zufallsvariablen $Y_1,\dots,Y_n$ gilt dann:

$$Y_i = \beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\dots+\beta_px_{ip}+\varepsilon_i \quad ,i=1,..n$$

Für die Störterme $\varepsilon_1,\dots,\varepsilon_n$ machen wir die üblichen Annahmen:

(a) $E(\varepsilon_i)=0$ für $i=1,\dots,n$

(b) $Cov(\varepsilon_i,\varepsilon_j) = \begin{cases} \sigma^2 & \text{für } i=j \\ 0 & \text{für } i\neq j \end{cases}$

(c) Bei der Durchführung von Tests und beim Aufstellen von Konfidenzintervallen unterstellen wir noch, daß die Störterme normalverteilt sind.

Wir schreiben das Modell nun in Matrizenform.
Mit

$$Y=\begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{pmatrix}, \quad \varepsilon=\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{pmatrix}, \quad \beta=\begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_p \end{pmatrix}, \quad X=\begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

erhalten wir das Modell

$$\boxed{Y = X\beta + \varepsilon}$$

In dieser Darstellung sind nun die Komponenten der Vektoren Y und ε univariate Zufallsvariablen. Solche Vektoren bezeichnet man als Zufallsvektoren oder auch mehrdimensionale Zufallsvariablen. Da wir im folgenden immer wieder Eigenschaften von mehrdimensionalen Zufallsvariablen benutzen werden, wollen wir diese im nächsten Abschnitt zunächst näher betrachten.

3.2 Mehrdimensionale Zufallsvariablen

Wie schon im letzten Abschnitt erwähnt wurde, heißt der Vektor $Y=(Y_1,\ldots,Y_n)'$ n-dimensionale Zufallsvariable, wenn $Y_1,\ldots,Y_n$ univariate Zufallsvariablen sind.

Die Matrix $Z=\begin{pmatrix} Z_{11} & Z_{12} \cdots Z_{1m} \\ \vdots & \\ Z_{n1} & Z_{n2} \cdots Z_{nm} \end{pmatrix}$ heißt Zufallsmatrix, wenn die einzelnen Elemente von Z univariate Zufallsvariablen sind.

Für die (n,m)-Matrix Z mit Elementen Z_{ij} schreiben wir im folgenden $Z=\left[\left(Z_{ij}\right)\right]$.

Eine (n,1)-Zufallsmatrix ist eine n-dimensionale Zufallsvariable. Wir werden uns im wesentlichen mit n-dimensionalen Zufallsvariablen beschäftigen. Zunächst wollen wir aber eine Eigenschaft einer Zufallsmatrix betrachten, die wir im folgenden benötigen werden.

$\boxed{\textit{DEF 3.1}}$

Sei Z eine (n,m)-Zufallsmatrix, und es existiere $E\left(Z_{ij}\right)$ für $i=1,\ldots,n, j=1,\ldots,m$. Dann ist $E(Z)=\left[\left(E\left(Z_{ij}\right)\right)\right]$.

$\boxed{\textit{SATZ 3.2}}$

Seien A eine $(1,n)$-Matrix, B eine (m,p)-Matrix und C eine $(1,p)$-Matrix reeller Zahlen. Weiterhin ist Z eine (n,m)-Zufallsmatrix. Dann gilt

$$E(AZB+C)=AE(Z)B+C \tag{3.1}$$

<u>Beweis</u>

Sei $W=AZB+C$. Dann gilt $W_{ij}=\sum_{r=1}^{m}\sum_{s=1}^{n}a_{ir}Z_{rs}b_{sj}+d_{ij}$.

Aufgrund der Eigenschaften des Erwartungswerts univariater Zufallsvariablen gilt also:

$$E\left(W_{ij}\right)=\sum_{r=1}^{m}\sum_{s=1}^{n}a_{ir}E\left(Z_{rs}\right)b_{sj}+d_{ij}\ .$$

Also gilt $E(W)=E(AZB+C)=AE(Z)B+C$. ∎

Ist Y eine k-dimensionale Zufallsvariable mit Erwartungswert $E(Y)$ A eine (n,k)-Matrix reeller Zahlen und b ein n-dimensionaler Spaltenvektor reeller Zahlen, so gilt aufgrund von *SATZ 3.2*:

$$E(AY+b)=AE(Y)+b \tag{3.2}$$

Ein Maß für den linearen Zusammenhang zwischen zwei univariaten Zufallsvariablen Y_1 und Y_2 ist die Kovarianz

$$Cov(Y_1,Y_2)=E\left[\left(Y_1-E\left(Y_1\right)\right)\left(Y_2-E\left(Y_2\right)\right)\right] \tag{3.3}$$

Für $Cov(Y_1,Y_2)$ schreiben wir σ_{12}.

Es gilt

$$Cov(Y_1,Y_2)=Cov(Y_2,Y_1) \tag{3.4}$$

Weiterhin gilt

$$Cov(Y_i,Y_i)=E\text{Ä}(Y_i-E(Y_i))^2\text{Ü}=Var(Y_i)\ ,i=1,2.$$

Für $Var(Y_i)$ schreiben wir σ^2_i, $i=1,2$.

Wir bilden eine Matrix aus den Varianzen und Kovarianzen, die sog. Varianz-Kovarianz-Matrix Σ:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}.$$

Wegen $Cov(Y_1,Y_2)=Cov(Y_2,Y_1)$ ist die Varianz-Kovarianz-Matrix Σ symmetrisch, d.h. $\Sigma=\Sigma'$.
Die Varianz-Kovarianz-Matrix einer k-dimensionalen Zufallsvariablen ist analog aufgebaut. Für Σ schreiben wir auch $Cov(Y)$.
Wir können die Varianz-Kovarianz-Matrix auch folgendermaßen berechnen, wobei wir μ_i für $E\left(Y_i\right)$ schreiben:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} E\left(\left(Y_1-\mu_1\right)\left(Y_1-\mu_1\right)\right) & E\left(\left(Y_1-\mu_1\right)\left(Y_2-\mu_2\right)\right) \\ E\left(\left(Y_2-\mu_2\right)\left(Y_1-\mu_1\right)\right) & E\left(\left(Y_2-\mu_2\right)\left(Y_2-\mu_2\right)\right) \end{pmatrix} =$$

$$= \mathrm{E}\begin{pmatrix} \left(Y_1-\mu_1\right)\left(Y_1-\mu_1\right) & \left(Y_1-\mu_1\right)\left(Y_2-\mu_2\right) \\ \left(Y_2-\mu_2\right)\left(Y_1-\mu_1\right) & \left(Y_2-\mu_2\right)\left(Y_2-\mu_2\right) \end{pmatrix} =$$

$$= \mathrm{E}\left(\begin{pmatrix} Y_1-\mu_1 \\ Y_2-\mu_2 \end{pmatrix}\begin{pmatrix} Y_1-\mu_1 \\ Y_2-\mu_2 \end{pmatrix}'\right) = \mathrm{E}\left(\left(\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}-\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}\right)\left(\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}-\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}\right)'\right)$$

$$= E\left((Y-\mu)(Y-\mu)'\right) = E\left((Y-E(Y))(Y-E(Y))'\right) \qquad (3.5)$$

Offensichtlich gilt diese Beziehung auch für die Varianz-Kovarianz-Matrix einer k-dimensionalen Zufallsvariablen.
Mit *(3.5)* können wir sofort den folgenden Satz beweisen

SATZ 3.3

Sei Y eine k-dimensionale Zufallsvariable mit Erwartungswert $E(Y)=\mu=\left(\mu_1,\ldots,\mu_k\right)'$ und Varianz-Kovarianzmatrix Σ.
Außerdem sei A eine (n,k)-Matrix reeller Zahlen und b ein k-dimensionaler Spaltenvektor.
Dann gilt

$$Cov\left(AY+b\right)=ACov(Y)A' \qquad (3.6)$$

Beweis

Ist A eine (n,k)-Matrix und B eine (k,n)-Matrix, dann gilt $(AB)'=B'A'$.

Mit (3.1) ,(3.2) und (3.5) gilt also

$$\begin{aligned}
Cov(AY+b) &= E\Big((AY+b-E(AY+b))(AY+b-E(AY+b))'\Big) = \\
&= E\Big((AY+b-AE(Y)-b)(AY+b-AE(Y)-b)'\Big) = \\
&= E\Big((AY-AE(Y))(AY-AE(Y))'\Big) = \\
&= E\Big((A(Y-E(Y)))(A(Y-E(Y)))'\Big) = \\
&= E\Big(A(Y-E(Y))(Y-E(Y))'A'\Big) = \\
&= AE\Big((Y-E(Y))(Y-E(Y))'\Big)A' = ACov(Y)A'
\end{aligned}$$

■

Mit Hilfe der Beziehungen (3.2) und (3.6) werden wir im Abschnitt 3.4 die wesentlichen Eigenschaften des Kleinst-Quadrate-Schätzers herleiten Auf die Modellierung der Verteilung der Störterme werden wir dann erst in Abschnitt 3.5 eingehen, so daß wir uns im nächsten Abschnitt gleich der Herleitung des Kleinst-Quadrate-Schätzers widmen werden.

3.3 Schätzen der Parameter

Im Abschnitt 3.1 haben wir das lineare Modell in Matrixschreibweise geschrieben. Es lautet

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

wobei Y und ε n-dimensionale Zufallsvariablen, X eine (n,k)-Matrix mit $k<n$ und $rg(X)=k$ und β ein k-dimensionaler Spaltenvektor sind. Für die ε_i unterstellen wir:

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

$$Cov(\varepsilon_i,\varepsilon_j) = \begin{cases} \sigma^2 & \text{für } i=j \\ 0 & \text{für } i\neq j \end{cases}$$

Für die n-dimensionale Zufallsvariable ε unterstellen wir also

$$E(\varepsilon)=0$$

$$Cov(\varepsilon)=\sigma^2 I_n \quad ,$$

wobei I_n die (n,n)-Einheitsmatrix ist.

Schätzung der Parameter nach der Methode der Kleinsten-Quadrate heißt nun, den Wert für β so zu bestimmen, daß die Summe der quadrierten Störterme minimal wird, d.h.

Minimiere $S(\beta)=\sum_i \varepsilon_i^2=\varepsilon'\varepsilon=(y-X\beta)'(y-X\beta)$

Notwendige Bedingungen für ein lokales Extremum sind, daß die ersten partiellen Ableitungen nach $\beta_i, i=0,\dots,p$ Null werden.

<u>Exkurs</u>

Sei $x\in\mathbb{R}^k$ und $f:\mathbb{R}^k\to\mathbb{R}$ differenzierbar in x. Dann ist der Gradient von f definiert durch

$$grad(f)=\frac{\partial}{\partial x}f(x)=\begin{pmatrix}\frac{\partial}{\partial x_1}f(x)\\ \vdots \\ \frac{\partial}{\partial x_k}f(x)\end{pmatrix}$$

Ist $a\in\mathbb{R}^k$ und $f(x)=a'x=x'a=a_1x_1+a_2x_2+\dots+a_kx_k$, dann gilt

$\frac{\partial}{\partial x_i}f(x)=a_i$ für $i=1,\dots,k$, also

$$grad(a'x)=a \qquad (3.7)$$

Ist A eine symmetrische (k,k)-Matrix, d.h. $A=A'$, und $f(x)=x'Ax$, dann gilt

$$grad(x'Ax)=2Ax \qquad (3.8)$$

Dies sieht man folgendermaßen:

Es gilt $x'Ax=\sum_i\sum_j x_ia_{ij}x_j$. Differenzieren wir diesen Ausdruck nach x_k, so müssen wir von den k^2 Summanden nur die folgenden berücksichtigen: $\sum_{j\neq k}x_ka_{kj}x_j+\sum_{i\neq k}x_ia_{ik}x_k+a_{kk}x_k^2$.

Differenzieren wir diesen Ausdruck nach x_k, so erhalten wir

$$\sum_{j\neq k}a_{kj}x_j+\sum_{i\neq k}x_ia_{ik}+2a_{kk}x_k=\sum_j a_{kj}x_j+\sum_i x_ia_{ik}=$$

$= 2\sum_j a_{kj}x_j$, da wegen $A=A'$ gilt $a_{ik}=a_{ki}$.

Nun gilt aber für die k-te Komponente von $y=Ax$

$y_k=\sum_j a_{kj}x_j$. Somit ist $grad(x'Ax)=2Ax$. ∎

Wir kehren zum ursprünglichen Problem zurück.

$$S(\beta)=(y-X\beta)'(y-X\beta)=(y'-(X\beta)')(y-X\beta)=(y'-\beta'X')(y-X\beta)=$$
$$=y'y-\beta'X'y-y'X\beta+\beta'X'X\beta=y'y-\beta'X'y-(\beta'X'y)'+\beta'X'X\beta=$$
$$=y'y-2\beta'X'y+\beta'X'X\beta\ ,$$

da für einen Skalar a gilt: $a=a'$.

Wir bilden den Gradienten von $S(\beta)$, wobei wir *(3.7)* und *(3.8)* anwenden.

$$grad(S(\beta))=-2X'y+2X'X\beta \tag{3.9}$$

Notwendige Bedingung für einen Extremwert an der Stelle $\hat\beta$ ist also

$$-2X'y+2X'X\hat\beta=0 \tag{3.10}$$

Wir erhalten die sogenannten Normalgleichungen

$$X'X\hat\beta=X'y \tag{3.11}$$

Da gilt $rg(X'X)=rg(X)=k$, und da $X'X$ eine (k,k)-Matrix ist, existiert $(X'X)^{-1}$ und wir können *(3.11)* nach $\hat\beta$ auflösen:

$$\hat\beta=(X'X)^{-1}X'y \tag{3.12}$$

Wir können *(3.11)* auch folgendermaßen schreiben

$$X'(y-X\hat\beta)=0 \tag{3.13}$$

Daß an der Stelle $\hat\beta$ die Funktion $S(\beta)$ auch wirklich ein Minimum aufweist, sieht man folgendermaßen

$$S(\beta)=(y-X\beta)'(y-X\beta)=(y-X\hat\beta+X\hat\beta-X\beta)'(y-X\hat\beta+X\hat\beta-X\beta)=$$
$$=((y-X\hat\beta)+X(\hat\beta-\beta))'((y-X\hat\beta)+X(\hat\beta-\beta))=$$
$$=((y-X\hat\beta)'+(\hat\beta-\beta)'X')((y-X\hat\beta)+X(\hat\beta-\beta))=$$
$$=(y-X\hat\beta)'(y-X\hat\beta)+(y-X\hat\beta)'X(\hat\beta-\beta)+(\hat\beta-\beta)'X'(y-X\hat\beta)+(\hat\beta-\beta)'X'X(\hat\beta-\beta)=$$
$$=(y-X\hat\beta)'(y-X\hat\beta)+(\hat\beta-\beta)'X'X(\hat\beta-\beta) \tag{3.14}$$

da wegen *(3.13)* gilt $(y-X\hat{\beta})'X(\hat{\beta}-\beta)=0$ und $(\hat{\beta}-\beta)'X'(y-X\hat{\beta})=0$.

Der erste Term in *(3.14)* hängt nicht von β ab. Also wird *(3.14)* minimal, wenn $(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)$ möglichst klein wird.

Es gilt $(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)=(X(\hat{\beta}-\beta))'X(\hat{\beta}-\beta)=z'z=\sum_i z_i^2\geq 0$ mit $z=X(\hat{\beta}-\beta)$.

Da X vollen Spaltenrang besitzt, die Spalten von X also linear unabhängig sind, ist $X(\hat{\beta}-\beta)$ genau dann 0, wenn gilt $\hat{\beta}-\beta=0$.

Also ist $\hat{\beta}$ das eindeutige Minimum von $S(\beta)$.

Der Vektor der geschätzten y-Werte ist

$$\hat{y}=X\hat{\beta}=X(X'X)^{-1}X'y=Hy \tag{3.15}$$

mit $H=X(X'X)^{-1}X'$.

Zur Herleitung der Eigenschaften von H benötigen wir folgende Beziehung:

Ist A eine reguläre (k,k)-Matrix,so gilt $(A')^{-1}=(A^{-1})'$,

denn $I_k=AA^{-1}=(AA^{-1})'=(A^{-1})'A'$.

Da gilt $(X'X)'=X'(X')'=X'X$, ist $X'X$ symmetrisch ,und wegen $rg(x)=rg(X'X)=k$, ist $X'X$ regulär. Also gilt

$$\left((X'X)^{-1}\right)'=\left((X'X)'\right)^{-1}=(X'X)^{-1} \tag{3.16}$$

Somit gilt also

$$H'=\left(X(X'X)^{-1}X'\right)'=\left((X')'\left((X'X)^{-1}\right)'X'\right)=X(X'X)^{-1}X'=H \tag{3.17}$$

Also ist H symmetrisch.
Weiterhin gilt

$$H^2=HH=X(X'X)^{-1}X'X(X'X)^{-1}X'=X(X'X)^{-1}X'=H \tag{3.18}$$

Somit ist H idempotent.

Die Residuen sind $e_i=y_i-\hat{y}_i, i=1,\ldots,n$.

Somit erhalten wir den Residuenvektor

$$e=y-\hat{y}=y-X(X'X)^{-1}X'y=(I_n-X(X'X)^{-1}X')y=Py \tag{3.19}$$

mit $P=I_n-X(X'X)^{-1}X'=I_n-H$.

Die Matrix P ist ebenfalls symmetrisch und idempotent, denn

$$P'=(I_n-H)'=I_n'-H'=I_n-H=P \tag{3.20}$$

$$P^2=PP=(I_n-H)(I_n-H)=I_n^2-H-H+H^2=I_n-H-H+H=I_n-H=P \tag{3.21}$$

Wir bestimmen nun noch den Kleinst-Quadrate-Schätzer von β bei der Einfachregression mit Hilfe der Matrix-Darstellung des Modells und des Schätzers.
Bei der Einfachregression lautet die Designmatrix:

$$X=\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{pmatrix} \tag{3.22}$$

Für eine $(2,2)$-Matrix $A=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ gilt $A^{-1}=\frac{1}{ad-bc}\begin{pmatrix} d & -c \\ -c & a \end{pmatrix}$, falls gilt $ad\neq bc$.
Wir erhalten also

$$X'X=\begin{pmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix} \text{ und } (X'X)^{-1}=\frac{1}{k}\begin{pmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & n \end{pmatrix} \text{ mit}$$

$k=n\sum_i x_i^2-\left(\sum_i x_i\right)^2.$

Nun gilt $S_{xx}=\sum_i\left(x_i-\bar{x}\right)^2=\sum_i\left(x_i^2-2x_i\bar{x}+\bar{x}^2\right)=\sum_i x_i^2-2\bar{x}\sum_i x_i+n\bar{x}^2=$

$$=\sum_i x_i^2-\frac{1}{n}\left(\sum_i x_i\right)^2=\frac{1}{n}\left(n\sum_i x_i^2-\left(\sum_i x_i\right)^2\right)$$

Also gilt $k=nS_{xx}$.

Weiterhin erhalten wir $X'y=\begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix}$.

Somit gilt

$$\hat{\beta}=(X'X)^{-1}X'y=\frac{1}{S_{xx}}\begin{pmatrix} \sum_i y_i \sum_i x_i^2-\sum_i x_i y_i \sum_i x_i \\ -\sum_i x_i \sum_i y_i+n\sum_i x_i y_i \end{pmatrix} \tag{3.23}$$

Wir vereinfachen zunächst die zweite Komponente von $\hat{\beta}$.
Es gilt

$$S_{XY} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y}) =$$
$$= \sum_i x_i y_i - \bar{y}\sum_i x_i - \bar{x}\sum_i y_i + n\bar{x}\bar{y} =$$
$$= \sum_i x_i y_i - \frac{1}{n}\sum_i x_i \sum_i y_i =$$
$$= \frac{1}{n}\left(n\sum_i x_i y_i - \sum_i x_i \sum_i y_i\right)$$

Somit ist die zweite Komponente von $\hat{\beta}$ gleich $\dfrac{S_{xy}}{S_{XX}}$,was wir in Kapitel 2 bereits gezeigt haben.

Wir vereinfachen nun noch die erste Komponente von $\hat{\beta}$:

$$\frac{\sum_i y_i \sum_i x_i^2 - \sum_i x_i y_i \sum_i x_i}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2} =$$

$$= \frac{\sum_i y_i \sum_i x_i^2 - \frac{1}{n}\sum_i y_i \left(\sum_i x_i\right)^2 + \frac{1}{n}\sum_i y_i \left(\sum_i x_i\right)^2 - \sum_i x_i y_i \sum_i x_i}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2} =$$

$$= \frac{\frac{1}{n}\sum_i y_i \left(n\sum_i x_i^2 - \left(\sum_i x_i\right)^2\right)}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2} - \frac{\sum_i x_i y_i \sum_i x_i - \frac{1}{n}\sum_i y_i \left(\sum_i x_i\right)^2}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2} =$$

$$= \bar{y} - \frac{\frac{1}{n}\sum_i x_i \left(n\sum_i x_i y_i - \sum_i x_i \sum_i y_i\right)}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2} = \bar{y} - \hat{\beta}_1 \bar{x}$$

3.4 Eigenschaften der Schätzer

Im linearen Modell $Y=X\beta+\varepsilon$ gilt wegen (3.2) und (3.6) unter den Annahmen $E(\varepsilon)=0$ und $Cov(\varepsilon)=\sigma^2 I_n$:

$$E(Y)=E(X\beta+\varepsilon)=X\beta+E(\varepsilon)=X\beta \tag{3.24}$$

$$Cov(Y)=Cov(X\beta+\varepsilon)=Cov(\varepsilon)=\sigma^2 I_n \tag{3.25}$$

Somit gilt für den Kleinst-Quadrate-Schätzer $\hat{\beta}=(X'X)^{-1}X'Y$:

$$E\left(\hat{\beta}\right)=E\left((X'X)^{-1}X'Y\right)=(X'X)^{-1}X'E\left(Y\right)=(X'X)^{-1}X'X\beta=\beta \tag{3.26}$$

Der Kleinst-Quadrate-Schätzer $\hat{\beta}$ ist somit erwartungstreu für β.

Für die Varianz-Kovarianzmatrix von $\hat{\beta}$ gilt wegen (3.6) und (3.25):

$$Cov(\hat{\beta})=Cov\left((X'X)^{-1}X'Y\right)=(X'X)^{-1}X'Cov(Y)\left((X'X)^{-1}X'\right)'=$$

$$=(X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1}=\sigma^2(X'X)^{-1}X'X(X'X)^{-1}=\sigma^2(X'X)^{-1} \tag{3.26}$$

Der Kleinst-Quadrate-Schätzer $\hat{\beta}$ ist in den $Y_1,\dots,Y_n$ linear.
In der Klasse der linearen Schätzer besitzt er folgende Optimalitätseigenschaft:

Satz 3.4

Im linearen Modell $Y=X\beta+\varepsilon$ ist der Kleinst-Quadrate-Schätzer $\hat{\beta}=(X'X)^{-1}X'y$ bester linearer unverzerrter Schätzer (best linear unbiased estimator BLUE), d.h. unter allen Schätzern der Form $\beta^*=AY$ mit $E(\beta^*)=\beta$ besitzt $\hat{\beta}$ minimale Varianz. Es gilt

$$Var(\hat{\beta}_i)\le Var(\beta_i^*),\quad i=0,\dots,p \tag{3.28}$$

und sogar

$$Var(c'\hat{\beta})\le Var(c'\beta^*) \text{ für alle } c\in\mathbb{R}^k \tag{3.29}$$

<u>Beweis</u>

Wir zeigen (3.29), da dann auch (3.28) gilt, wenn man für c die Einheitsvektoren wählt.

Da β^* erwartungstreu für β ist, muß gelten

$E(\beta^*)=E(AY)=AE(Y)=AX\beta=\beta$ für alle $\beta\in\mathbb{R}^k$, also $(AX-I_k)\beta=0$ für alle $\beta\in\mathbb{R}^k$. Dies impliziert $AX=I_k$.

Nun ist

$$Var(c'\beta^*)=Var(c'AY)=c'ACov(Y)(c'A)'=c'A\sigma^2 I_n A'c=\sigma^2 c'AA'c$$

und

$$Var(c'\hat{\beta})=c'Cov(\hat{\beta})c=\sigma^2 c'(X'X)^{-1}c$$

Wir haben also zu zeigen

$$c'AA'c\geq c'(X'X)^{-1}c \qquad \text{für alle } c\in\mathbb{R}^k, \text{d.h.}$$

$$c'(AA'-(X'X)^{-1})c\geq 0 \quad \text{für alle } c\in\mathbb{R}^k.$$

Nun gilt wegen $AX=I_k$

$$AA'-(X'X)^{-1}=AA'-AX(X'X)^{-1}(AX)'=AA'-AX(X'X)^{-1}X'A'=$$

$$=A(I_n-X(X'X)^{-1}X')A'=APA'=AP^2A'=APPA'=APP'A'=AP(AP)',$$

da P wegen (3.20) und (3.21) idempotent und symmetrisch ist.

Mit $W=AP$ gilt also

$$c'WW'c=c'W(c'W)'=z'z=\sum_i z_i^2\geq 0 \qquad \text{mit } z=c'W.$$ ∎

Ist P eine (n,n)-Matrix, so ist die Spur von P gleich der Summe der Hauptdiagonalelemente, d.h. $tr(P)=\sum_i p_{ii}$.

Ist A eine (n,p)-Matrix und B eine (p,n)-Matrix, dann gilt

$$tr(AB)=tr(BA) \qquad (3.30)$$

Dies sieht man folgendermaßen:

Sei $C=AB$. Dann gilt $c_{ii}=\sum_{k=1}^{p} a_{ik}b_{ki}$ und somit $tr(AB)=\sum_{i=1}^{n}\sum_{k=1}^{p} a_{ik}b_{ki}$.

Sei $D=BA$. Dann gilt $d_{kk}=\sum_{i=1}^{n} b_{ki}a_{ik}$ und somit

$$tr(BA)=\sum_{k=1}^{p}\sum_{i=1}^{n} b_{ki}a_{ik}=\sum_{i=1}^{n}\sum_{k=1}^{p} a_{ik}b_{ki}=tr(AB).$$

Diese Beziehung benötigen wir beim Beweis des folgenden Satzes, der eine erwartungstreue Schätzfunktion für σ^2 angibt.

Satz 3.5

Im linearen Modell $Y=X\beta+\varepsilon$ ist $\hat{\sigma}^2=\frac{1}{n-k}\,e'e$ mit $e=y-\hat{y}=y-X(X'X)^{-1}X'y$ eine erwartungstreue Schätzfunktion für σ^2.

Beweis

Es gilt

$$e=y-\hat{y}=y-X(X'X)^{-1}X'y=\left(I_n-X(X'X)^{-1}X'\right)y=\left(I_n-X(X'X)^{-1}X'\right)\left(X\beta+\varepsilon\right)=$$
$$=X\beta+\varepsilon-X(X'X)^{-1}X'X\beta-X(X'X)^{-1}X'\varepsilon=\left(I_n-X(X'X)^{-1}X'\right)\varepsilon=P\varepsilon \qquad (3.31)$$

Dabei ist P wegen *(3.20)* und *(3.21)* eine idempotente und symmetrische Matrix.
Es gilt also

$$e'e=(P\varepsilon)'P\varepsilon=\varepsilon'P'P\varepsilon=\varepsilon'P\varepsilon=\sum_{i=1}^{n}p_{ii}\varepsilon_i^2+\sum_{i\neq j}p_{ij}\varepsilon_i\varepsilon_j.$$

Nun ist $E(\varepsilon_i^2)=Var(\varepsilon_i)=\sigma^2$ für $i=1,\ldots,n$ und $E(\varepsilon_i\varepsilon_j)=Cov(\varepsilon_i,\varepsilon_j)=0$ für $i\neq j$.

Also gilt

$$E(e'e)=\sum_{i=1}^{n}p_{ii}E(\varepsilon_i^2)+\sum_{i\neq j}p_{ij}E(\varepsilon_i\varepsilon_j)=\sigma^2\sum_{i=1}^{n}p_{ii}=\sigma^2 tr(P).$$

Nun gilt wegen *(3.30)*

$$tr(P)=tr\left(I_n-X(X'X)^{-1}X'\right)=tr\left(I_n\right)-tr\left(X(X'X)^{-1}X'\right)=n-tr\left((X'X)^{-1}X'X\right)=$$
$$=n-tr\left(I_k\right)=n-k$$

Also gilt $E(e'e)=(n-k)\sigma^2$ und somit $E\left(\frac{1}{n-k}e'e\right)=\sigma^2$. ∎

3.5 Verteilungen

3.5.1 Die mehrdimensionale Normalverteilung

Die univariate Zufallsvariable Y ist normalverteilt mit den Parametern μ und σ^2, falls ihre Dichte lautet:

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{1}{2}\left(\frac{Y-\mu}{\sigma}\right)^2}=(2\pi\sigma^2)^{-0.5}\ e^{-0.5(Y-\mu)(\sigma^2)^{-1}(Y-\mu)}$$

Die mehrdimensionale Zufallsvariable $Y=(Y_1,\ldots Y_n)'$ heißt n-dimensional normalverteilt, wenn ein n-dimensionaler Vektor $\mu=(\mu_1,\ldots,\mu_n)'$ und eine positiv definite (n,n)-Matrix Σ existiert, so daß die Dichte von Y lautet:

$$f_Y(y)=(2\pi)^{-0.5n}|\Sigma|^{-0.5}\ e^{-0.5(Y-\mu)'\Sigma^{-1}(Y-\mu)} \tag{3.32}$$

Dabei ist $|\Sigma|$ die Determinante von Σ.
Die Parameter μ und Σ haben dieselbe Bedeutung wie bei der univariaten Normalverteilung.

Es gilt $E(Y)=\mu$ und $Cov(Y)=\Sigma$. (Siehe z.B. Seber(1977), Linear Regression Analysis, S.22)

Ist die n-dimensionale Zufallsvariable Y mehrdimensional normalverteilt mit den Parametern μ und Σ, so schreiben wir $Y\sim N_n(\mu,\Sigma)$.

Unterstellen wir also im linearen Modell, daß die Störterme $\varepsilon_1,\ldots,\varepsilon_n$ unabhängig und identisch mit den Parametern 0 und σ^2 normalverteilt sind, so gilt für die n-dimensionale Zufallsvariable $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)'$:

$$\varepsilon\sim N_n\left(0,\sigma^2 I_n\right) \tag{3.33}$$

Bei linearen Transformationen verhält sich die mehrdimensionale Normalverteilung genauso wie die univariate Normalverteilung. Genauer gibt dies der folgende Satz an.

Satz 3.6

Ist $Y=(Y_1,...,Y_n)'$ mehrdimensional normalverteilt mit den Parametern μ und Σ und ist A eine (k,n)-Matrix mit $k \leq n$ und $rg(A)=k$ und b ein k-dimensionaler Spaltenvektor, dann ist $AY+b$ k-dimensional normalverteilt mit $E(Y)=A\mu+b$ und $Cov(Y)=A\Sigma A'$.

Beweis

Auf den Beweis dieses Satzes verzichten wir. Er ist z.B. bei Seber(1977),Linear Regression Analysis auf S.28 zu finden. ∎

Aus *Satz 3.6* folgt sofort, daß jede Randverteilung einer n-dimensional normalverteilten Zufallsvariablen Y eine univariate Normalverteilung ist. Wir wählen für A das Transponierte des i-ten Einheitsvektors und für b den Nullvektor.

Außerdem ist die gemeinsame Verteilung von k Komponenten einer n-dimensionalen Normalverteilung eine k-dimensionale Normalverteilung. Die Zeilen von A bilden in diesem Fall die entsprechenden Einheitsvektoren, und b ist wiederum der Nullvektor.

Ist im linearen Modell $Y=X\beta+\varepsilon$ der Vektor ε n-dimensional normalverteilt mit Parametern 0 und $\sigma^2 I_n$, d.h. $\varepsilon \sim N_n\left(0,\sigma^2 I_n\right)$, so folgt aus *Satz 3.6* sofort

$$Y \sim N_n\left(X\beta,\sigma^2 I_n\right) \qquad (3.34)$$

In der Notation von *Satz 3.6* ist $A=I_n$ und $b=X\beta$, und da I_n vollen Zeilenrang besitzt, ist Y n-dimensional normalverteilt. Die Parameter kann man ohne Schwierigkeiten bestimmen.

Die Verteilung von $\hat{\beta}=(X'X)^{-1}X'y$ unter den Annahmen des linearen Modells gibt der folgende *Satz* an

$\boxed{\textit{Satz 3.7}}$

Im linearen Modell $Y=X\beta+\varepsilon$ mit $\varepsilon\sim N_n\left(0,\sigma^2 I_n\right)$ gilt für den Kleinst-Quadrate-Schätzer $\hat{\beta}=(X'X)^{-1}X'Y$:

$$\hat{\beta}\sim N_k\left(\beta,\sigma^2 (X'X)^{-1}\right) \tag{3.35}$$

<u>Beweis</u>

In der Notation von *Satz 3.6* gilt $\hat{\beta}=AY+b$ mit $A=(X'X)^{-1}X'$ und $b=0$. Die Matrix A ist eine (k,n)-Matrix mit $rg(A)=k$, da X den Rang k besitzt und somit auch $X'X$, $(X'X)^{-1}$ und $(X'X)^{-1}X'$, ist $\hat{\beta}$ aufgrund von *Satz 3.6* k-dimensional normalverteilt .

Wegen *(3.26)* und *(3.27)* gilt $E\left(\hat{\beta}\right)=\beta$ und $Cov\left(\hat{\beta}\right)=\sigma^2 (X'X)^{-1}$. ∎

Ist die Varianz σ^2 der Störterme bekannt, so kann man ohne Schwierigkeiten Tests auf einzelne Komponenten von β durchführen. In der Regel ist σ^2 aber unbekannt und muß durch $\hat{\sigma}^2=\frac{1}{n-k}\,e'e$ geschätzt werden. Im Beweis von *Satz 3.6* wurde gezeigt, daß gilt $e'e=\varepsilon'P\varepsilon$, wobei P eine idempotente und symmetrische Matrix ist. $e'e$ ist somit eine quadratische Form in der n-dimensional normalverteilten Zufallsvariablen ε, wobei die zugehörige Matrix P idempotent und symmetrisch ist. Mit der Verteilung von quadratischen Formen werden wir uns im nächsten Abschnitt näher beschäftigen.

3.5.2 Verteilung quadratischer Formen

Ist P eine symmetrische (n,n)-Matrix und w ein n-dimensionaler Spaltenvektor, dann heißt

$$w'Pw=\sum_i \sum_j w_i p_{ij} w_j \qquad (3.36)$$

quadratische Form.

Die Herleitung der Verteilung von quadratischen Formen ,bei denen w eine n-dimensional normalverteilte Zufallsvariable ist, beruht im wesentlichen auf folgendem Satz:

$\boxed{\textit{Satz 3.8}}$

Zu jeder symmetrischen (n,n)-Matrix P existiert eine orthogonale (n,n)-Matrix T und eine Diagonalmatrix $\Lambda=diag(\lambda_1,...,\lambda_n)$, so daß gilt

$$P=T\Lambda T' \qquad (3.37)$$

Dabei sind die Spalten von T die standardisierten Eigenvektoren von P und die Hauptdiagonalelemente von Λ die zugehörigen Eigenwerte.

Beweis

siehe Mardia,Kent,Bibby(1979), Multivariate Analysis, S.469-470 ∎

Anmerkungen

1. Eine (n,n)-Matrix T heißt orthogonal, wenn gilt $T^{-1}=T'$.
   Für eine orthogonale Matrix T gilt somit $TT'=T'T=I_n$.

   Ist w eine n-dimensionaler Spaltenvektor und T eine orthogonale (n,n)-Matrix, dann gilt $(Tw)'Tw=w'T'Tw=w'w$. Die Multiplikation mit einer orthogonalen Matrix ändert also die Länge eines Vektors nicht(Drehung,Spiegelung).

2. Das in *Satz 3.8* erwähnte Eigenwertproblem ist:
   Gegeben sei eine (n,n)-Matrix P. Gesucht ist ein n-dimensionaler Vektor $t\neq 0$ und eine Zahl λ, so daß gilt:

$$Pt=\lambda t \tag{3.38}$$

   (3.38) kann man umformen zu

$$(P-\lambda I_n)t=0 \tag{3.39}$$

   (3.39) hat genau dann nichttriviale Lösungen, wenn gilt

$$|P-\lambda I_n|=0 \tag{3.40}$$

   (3.40) ist ein Polynom n-ten Grades in λ, das n (nicht notwendigerweise verschiedene) Nullstellen besitzt.
   Diese bezeichnet man als Eigenwerte von P. Zu einem Eigenwert λ kann man dann mit Hilfe von Gleichung *(3.39)* die zugehörigen Eigenvektoren bestimmen. Bei einer symmetrischen Matrix P sind nun alle Eigenwerte reell und die zu unterschiedlichen Eigenwerten gehörenden Eigenvektoren sind orthogonal. Die Vektoren sind standardisiert, wenn ihre Länge gleich 1 ist,d.h. $t't=1$.

3. Ist P eine symmetrische (n,n)-Matrix, so existiert aufgrund von *Satz 3.8* eine orthogonale Matrix T und eine Diagonalmatrix T, so daß gilt $P=T\Lambda T'$. Eine quadratische Form in P kann nun folgendermaßen vereinfacht werden:

$$w'Pw=w'T\Lambda T'w=(T'w)'\Lambda T'w=z'\Lambda z=\sum_i \lambda_i z_i^2 \quad \text{mit } z=T'w \qquad (3.41).$$

   Man bezeichnet die Transformation $z=T'w$ auch als Hauptachsentransformation, da jede quadratische Form in einer symmetrischen Matrix eine Ellipse bzw. ein Ellipsoid beschreibt, und durch die Transformation $z=T'w$ das Koordinatensystem in Richtung der Hauptachsen der Ellipse bzw. des Ellipsoids gedreht wird.

Die Eigenwerte einer idempotenten Matrix P sind entweder 0 oder 1. Dies gilt aufgrund von

Satz 3.9

Ist P eine symmetrische und idempotente Matrix, so sind die Eigenwerte 0 oder 1.

Beweis

Sei P idempotent und λ ein Eigenwert von P, dann existiert ein Vektor $t \neq 0$, so daß gilt

$$Pt=\lambda t \tag{3.42}$$

Multiplizieren wir beide Seiten von (3.42) mit P, so erhalten wir

$$P^2 t=P\lambda t=\lambda Pt=\lambda^2 t \tag{3.43}$$

Aufgrund der Idempotenz von P gilt aber auch

$$P^2 t=Pt=\lambda t \tag{3.44}$$

Aus (3.43) und (3.44) folgt $\lambda^2 t=\lambda t$ und somit $\lambda(1-\lambda)t=0$. Da gilt $t \neq 0$, muß also $\lambda=0$ oder $\lambda=1$ gelten. ∎

Die Diagonalmatrix in (3.37) hat bei einer idempotenten und symmetrischen Matrix also eine sehr schöne Struktur. Sie enthält nur Einsen und Nullen. es stellt sich die Frage, wieviele der Elemente den Wert 1 annehmen.

Satz 3.10

Der Rang einer symmetrischen Matrix P ist gleich der Anzahl der Eigenwerte, die ungleich 0 sind.

Beweis

Wir gehen aus von (3.37), d.h. $P=T\Lambda T'$ und multiplizieren beide Seiten von links mit T' und von rechts mit T und erhalten wegen $T'T=TT'=I_n$:

$$T'PT=\Lambda \tag{3.45}$$

Da der Rang einer Matrix sich nicht ändert, wenn man sie mit einer regulären Matrix multipliziert, gilt also $rg(P)=rg(T'PT)=rg(\Lambda)$. Der Rang der Diagonalmatrix Λ mit den Eigenwerten von P ist aber gerade der Anzahl der Hauptdiagonalelemente, die ungleich Null sind. ∎

Aus den *Sätzen 3.8,3.9* und *3.10* folgt, daß zu jeder idempotenten und symmetrischen (n,n)-Matrix P mit $rg(P)=r$ eine Orthogonalmatrix T und ein Diagonalmatrix E_r mit r Einsen und $n-r$ Nullen auf der Hauptdiagonalen existiert, so daß gilt

$$P=TE_rT' \tag{3.46}$$

$\boxed{\textit{Satz 3.11}}$

Sei $Y\sim N_n\left(\mu,\sigma^2 I_n\right)$ und P eine (n,n)-Matrix ,die symmetrisch und idempotent mit $rg(P)=r$ ist, dann ist $\frac{(Y-\mu)'P(Y-\mu)}{\sigma^2}\sim\chi^2(r)$

Beweis

Ist P symmetrisch und idempotent mit $rg(P)=r$, so sind r Eigenwerte von P gleich 1 und die restlichen $n-r$ gleich 0. Es existiert somit eine Orthogonalmatrix T und eine Diagonalmatrix E_r mit r Einsen und $n-r$ Nullen auf der Hauptdiagonalen, so daß gilt $P=TE_rT'$.
Sei $Z=T'(Y-\mu)$. Da T' vollen Zeilenrang besitzt, ist Z auf Grund von *Satz 3.6* n-dimensional normalverteilt mit

$E(Z)=E(T'(Y-\mu))=T'(E(Y)-\mu)=0$ und

$Cov(Z)=Cov(T'(Y-\mu))=T'Cov(Y)T=T'\sigma^2 I_n T=\sigma^2 I_n$.

Es gilt also $Z\sim N_n\left(0,\sigma^2 I_n\right)$, d.h. $Z_1,\ldots,Z_n$ sind unabhängig und mit den Parametern 0 und σ^2 identisch normalverteilt.

Also sind $\frac{Z_1}{\sigma},\ldots,\frac{Z_n}{\sigma}$ unabhängig und identisch standardnormalverteilt.

Somit gilt

$$\frac{(Y-\mu)'P(Y-\mu)}{\sigma^2}=\frac{(Y-\mu)'TE_rT'(Y-\mu)}{\sigma^2}=\frac{(T'(Y-\mu))'E_rT'(Y-\mu)}{\sigma^2}=$$

$$=\frac{Z'E_rZ}{\sigma^2}=\sum_i\frac{Z_i^2}{\sigma^2}\ .$$

Da die $\frac{Z_i}{\sigma}$ unabhängig und standardnormalverteilt sind, ist $\frac{(Y-\mu)'P(Y-\mu)}{\sigma^2}$ als Summe von r Quadraten von unabhängigen standardnormalverteilten Zufallsvariablen chiquadratverteilt mit r Freiheitsgraden, womit der *Satz* bewiesen ist. ∎

Ist die Varianz-Kovarianzmatrix der in der quadratischen Form vorkommenden Zufallsvariablen keine Diagonalmatrix, dann erhält man unter bestimmten Bedingungen wieder eine Chiquadratverteilung. Hierzu benötigt man aber noch einige Ergebnisse aus der Linearen Algebra.

Satz 3.12

Ist A eine symmetrische (n,n)-Matrix, dann ist die Spur von A gleich der Summe der Eigenwerte von A.

<u>Beweis</u>

Seien $\lambda_1,...,\lambda_n$ die Eigenwerte von A, Λ die Diagonalmatrix mit den Eigenwerten von A und T die Orthogonalmatrix mit den standardisierten Eigenvektoren, dann gilt wegen *(3.30)* und wegen der Orthogonalität von T

$$\sum_i \lambda_i = tr(\Lambda) = tr(T'AT) = tr(TT'A) = tr(A) \qquad \blacksquare$$

Satz 3.13

Der Rang einer idempotenten und symmetrischen Matrix ist gleich ihrer Spur.

<u>Beweis</u>

Eine idempotente und symmetrische Matrix mit Rang r besitzt r Eigenwerte, die gleich 1 sind, die restlichen Eigenwerte sind 0. Also ist die Summe der Eigenwerte auf Grund von *Satz 3.12* gleich r und somit gleich dem Rang. $\blacksquare$

Ist P eine symmetrische (n,n)-Matrix und w ein n-dimensionaler Spaltenvektor, dann heißt die quadratische Form $w'Pw$ positiv definit, wenn $w'Pw>0$ gilt für alle $w\neq 0$.

$\boxed{\textit{Satz 3.14}}$

Die Eigenwerte einer positiv definiten Matrix P sind alle positiv.

<u>Beweis</u>

Da P symmetrisch ist, existiert eine orthogonale Matrix T, so daß gilt $P=T\Lambda T'$. Also gilt

$$w'Pw=w'T\Lambda T'w=(T'w)'\Lambda T'w=z'\Lambda z=\sum_i \lambda_i z_i^2>0 \text{ für alle } z\neq 0 \qquad (3.47)$$

(3.47) muß also auch für die Einheitsvektoren gelten, woraus folgt $\lambda_i>0$ für $i=1,\ldots,n$. ∎

$\boxed{\textit{Satz 3.15}}$

Ist A eine positiv definite (n,n)-Matrix, dann existiert eine reguläre (n,n)-Matrix R, so daß gilt

$$A=RR' \qquad (3.48)$$

<u>Beweis</u>

Da A symmetrisch ist, existiert eine orthogonale Matrix T und eine Diagonalmatrix Λ, so daß gilt $A=T\Lambda T'$. Da A positiv definit ist, sind auf Grund von *Satz 3.14* alle Eigenwerte von A positiv und somit existiert $\Lambda^{0.5}=diag\left(\lambda_1^{0.5},\ldots,\lambda_n^{0.5}\right)$ und ist regulär. Somit gilt $A=T\Lambda T'=T\Lambda^{0.5}\Lambda^{0.5}T'=T\Lambda^{0.5}\left(T\Lambda^{0.5}\right)'= RR'$ mit $R=T\Lambda^{0.5}$. ∎

$\boxed{\textit{Satz 3.16}}$

Ist $Y\sim N_n\left(\mu,\Sigma\right)$, A eine symmetrische (n,n)-Matrix mit $rg(A)=r$ und $A\Sigma A=A$, dann gilt $(Y-\mu)'A(Y-\mu)\sim\chi^2(r)$.

<u>Beweis</u>

Die Varianz-Kovarianzmatrix Σ ist positiv definit. Somit existiert auf Grund von *Satz 3.15* eine reguläre Matrix R, so daß gilt $\Sigma=RR'$.

Wir bilden $Z=R^{-1}(Y-\mu)$. Da R regulär ist, ist auch R^{-1} regulär und auf Grund von *Satz 3.6* ist Z n-dimensionale normalverteilt mit

$E(Z)=E(R^{-1}(Y-\mu)=R^{-1}(E(Y)-\mu)=0$ und

$Cov(R^{-1}(Y-\mu))=R^{-1}Cov(Y)(R^{-1})'=R^{-1}\Sigma(R')^{-1}=R^{-1}RR'(R')^{-1}=I_n$.

Also gilt $Z \sim N_n\left(0, I_n\right)$.

Nun gilt $(Y-\mu)'A(Y-\mu)=(RZ)'ARZ=Z'R'ARZ$.

Aufgrund von *Satz 3.11* ist $Z'R'ARZ$ chiquadratverteilt mit r Freiheitsgraden, wenn $R'AR$ idempotent ist,d.h. wenn gilt $R'ARR'AR=R'AR$. Wegen $RR'=\Sigma$ muß also gelten $R'A\Sigma AR=R'AR$, und somit da R regulär ist $A\Sigma A=A$. ∎

Aus *Satz 3.16* folgt dann sofort

Satz 3.17

Ist $Y \sim N_n\left(\mu, \Sigma\right)$, dann gilt $(Y-\mu)'\Sigma^{-1}(Y-\mu) \sim \chi^2(n)$.

Beweis

In der Notation von *Satz 3.16* ist $A=\Sigma^{-1}$. Da gilt $rg\left(\Sigma^{-1}\right)=n$ und $\Sigma^{-1}\Sigma\Sigma^{-1}=\Sigma^{-1}$, ist $(Y-\mu)'\Sigma^{-1}(Y-\mu)$ chiquadratverteilt mit n Freiheitsgraden. ∎

Wir können nun ohne Schwierigkeiten die Verteilung von einigen im Rahmen des Linearen Modell wichtigen quadratischen Formen bestimmen.

Satz 3.18

Im linearen Modell $Y=X\beta+\varepsilon$ mit $\varepsilon \sim N_n\left(0, \sigma^2 I_n\right)$ gilt

$$\frac{1}{\sigma^2}\left(\hat{\beta}-\beta\right)'X'X\left(\hat{\beta}-\beta\right) \sim \chi^2(k) \tag{3.49}$$

wobei gilt $\hat{\beta}=(X'X)^{-1}X'Y$.

Beweis

Aufgrund von *Satz 3.7* gilt $\hat{\beta} \sim N_k\left(\beta, \sigma^2(X'X)^{-1}\right)$. Die Varianz-Kovarianzmatrix Σ von $\hat{\beta}$ ist $\sigma^2(X'X)^{-1}$. Der Rang von Σ ist k.

Außerdem gilt $\Sigma^{-1}=\sigma^{-2}(X'X)$.

Aufgrund von *Satz 3.17* gilt somit

$$\left(\hat{\beta}-\beta\right)'\sigma^{-2}(X'X)\left(\hat{\beta}-\beta\right)=\frac{1}{\sigma^2}\left(\hat{\beta}-\beta\right)'X'X\left(\hat{\beta}-\beta\right) \sim \chi^2(k)$$ ∎

Der folgende *Satz* gibt nun die Verteilung von $\frac{1}{\sigma^2}\, e'e$ an.

Satz 3.19

Im linearen Modell $Y=X\beta+\varepsilon$ mit $\varepsilon \sim N_n\left(0,\sigma^2 I_n\right)$ gilt

$$\frac{1}{\sigma^2}\; e'e = \frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-k) \tag{3.50}$$

Beweis

Im Beweis von *Satz 3.6* wurde gezeigt $e'e=\varepsilon'P\varepsilon$, wobei die Matrix $P=I_n-X(X'X)^{-1}X'$ idempotent und symmetrisch ist.

Somit gilt $\frac{e'e}{\sigma^2} = \frac{\varepsilon'P\varepsilon}{\sigma^2}$.

Da gilt $\varepsilon \sim N_n\left(0,\sigma^2 I_n\right)$, und P idempotent und symmetrisch ist, ist aufgrund von *Satz 3.11* $\frac{e'e}{\sigma^2}$ chiquadratverteilt, wobei die Anzahl der Freiheitsgrade gleich dem Rang von P ist.

Da aufgrund von *Satz 3.13* der Rang einer idempotenten und symmetrischen Matrix gleich ihrer Spur ist, gilt

$$rg(B)=rg(P)=tr\left(I_n-X(X'X)^{-1}X'\right)=tr\left(I_n\right)-tr\left(X(X'X)^{-1}X'\right)=$$
$$=n-tr\left(X'X(X'X)^{-1}\right)=n-tr\left(I_k\right)=n-k \qquad \blacksquare$$

3.5.3 Unabhängigkeit von Linearformen und quadratischen Formen

Fast alle Teststatistiken, die wir im nächsten Abschnitt herleiten werden, sind Quotienten von quadratischen Formen in normalverteilten Zufallsvariablen. Diese erfüllen die in den *Sätzen 3.11* und *3.16* angegebenen Bedingungen und sind somit chiquadratverteilt. Der Quotient von chiquadratverteilten Zufallsvariablen, die durch die Anzahl ihrer Freiheitsgrade dividiert werden, ist F-verteilt, wenn die chiquadratverteilten Zufallsvariablen unabhängig sind. Wir werden in diesem Abschnitt nun Bedingungen angeben, unter denen quadratische Formen unabhängig sind. Auf die Beweise verzichten wir. Sie sind bei Graybill, S.84-87 zu finden.

$\boxed{\textit{Satz 3.20}}$

Sei $Y \sim N_n\left(\mu, \sigma^2 I_n\right)$, und seien A und B symmetrische (n,n)-Matrizen. $Y'AY$ und $Y'BY$ sind genau dann unabhängig, wenn gilt $AB=0$.

$\boxed{\textit{Satz 3.21}}$

Sei $Y \sim N_n\left(\mu, \sigma^2 I_n\right)$, und sei A eine symmetrische (n,n)-Matrix und B eine (p,n)-Matrix.
$Y'AY$ und BY sind genau dann unabhängig, wenn gilt $BA=0$.

Mit Hilfe von *Satz 3.21* können wir beweisen

$\boxed{\textit{Satz 3.22}}$

Im linearen Modell $Y=X\beta+\varepsilon$ mit $\varepsilon \sim N_n\left(0, \sigma^2 I_n\right)$ sind der Kleinst-Quadrate-Schätzer $\hat{\beta}=(X'X)^{-1}X'Y$ und $e'e=\varepsilon'P\varepsilon$ unabhängig.

<u>Beweis</u>

Aufgrund von *(3.34)* gilt $Y \sim N_n\left(X\beta, \sigma^2 I_n\right)$.

Nun ist $\hat{\beta}=BY$ mit $B=(X'X)^{-1}X'$. Weiterhin gilt

$e'e=(Y-X\hat{\beta})'(Y-X\hat{\beta})=(PY)'PY=Y'P'PY=Y'PPY=Y'PY$, da $P=I_n-X(X'X)^{-1}X'$ idempotent und symmetrisch ist.

Da gilt

$BP=(X'X)^{-1}X'\left(I_n-X(X'X)^{-1}X'\right)=(X'X)^{-1}X'-(X'X)^{-1}X'X(X'X)^{-1}X'=$

$=(X'X)^{-1}X'-(X'X)^{-1}X'=0$, sind $\hat{\beta}$ und $e'e$ aufgrund von *Satz 3.21* unabhängig. ∎

Da stetige Funktionen von unabhängigen Zufallsvariablen ebenfalls unabhängig sind(siehe dazu Mood,Graybill,Boes(1974),Introduction to the Theory of Statistics,S.151), sind auch $\hat{\beta}$ und $\hat{\sigma}^2 = \frac{1}{n-k} e'e$ unabhängig.

3.6 Tests linearer Hypothesen

Ausgangspunkt ist im folgenden ein lineares Modell $Y=X\beta+\varepsilon$, wobei X eine (n,k)-Matrix ist mit $rg(X)=k$, β ein k-dimensionaler Parametervektor ist und $\varepsilon \sim N_n\left(0,\sigma^2 I_n\right)$.

Bisher haben wir uns mit dem Schätzen der unbekannten Parameter β und σ^2 beschäftigt und die Verteilungen der entsprechenden Schätzfunktionen hergeleitet. In diesem Abschnitt wollen wir uns Tests zuwenden.

Alle im Rahmen des Linearen Modells interessierenden Hypothesen lassen sich als Spezialfälle folgender allgemeiner Hypothese auffassen, die auch als lineare Hypothese bezeichnet wird:

$$\boxed{\begin{array}{l} H_0:\ A\beta=c \\ H_1:\ A\beta\neq c \end{array}} \qquad (3.51)$$

Hierbei ist A eine (r,k)-Matrix mit $r\leq k$ und $rg(A)=k$, und c ist ein r-dimensionaler Spaltenvektor. Jede Zeile der Matrix A enthält zusammen mit der korrespondierenden Komponente des Vektors c eine lineare Nebenbedingung für den Parametervektor β. Die Bedingung $rg(A)=k$ gewährleistet, daß diese Nebenbedingungen linear unabhängig sind.

Schauen wir uns einige Beispiele an, bevor wir die Teststatistik für die obige Hypothese herleiten:

(1) $H_0:\beta=\beta^0$ gegen $H_1:\beta\neq\beta^0$.

Hierbei ist β^0 ein spezieller k-dimensionaler Vektor. Es soll also überprüft werden, ob der Parametervektor β einen speziellen Wert β^0 annimmt. Ein wichtiger Spezialfall ist, daß alle Komponenten von β gleich 0 sind.

Die obige Hypothese ist ein Spezialfall der linearen Hypothese *(3.51)* mit $A=I_k$ und $c=\beta^0$.

(2) $H_0:\beta_i=\beta_i^0$ gegen $H_1:\beta_i\neq\beta_i^0$ für ein $i\in 0,1,2,...,p$

Hier soll überprüft werden, ob eine Komponente β_i des Parametervektors β einen speziellen Wert β_i^0 annimmt.

Wichtige Spezialfälle sind, daß β_0 gleich *0* ist,d.h. daß die lineare Funktion durch den Ursprung geht, bzw. daß eines der $\beta_i, i=1,...,p,$ gleich Null ist, d.h. daß die i.te Variable keinen Einfluß auf die zu erklärende Variable Y hat.
Auch diese Hypothese ist ein Spezialfall der linearen Hypothese *(3.51)* mit $A=(0,...,0,1,0,...,0)$, wobei die *1* an der $i+1$-ten Stelle steht. (Wir beginnen mit dem Zählen mit *0*).

Der Vektor c ist in diesem Fall β_i^0.

(3) In Kapitel 1 haben wir gesehen, daß das Zweistichprobenproblem unabhängiger Stichproben ein Spezialfall des Linearen Modells ist. Hier ist der Parametervektor β gleich $\beta=(\mu_1,\mu_2)'$, wobei μ_1 der Erwartungswert der Zufallsvariablen der ersten und μ_2 der Erwartungswert der Zufallsvariablen der zweiten Stichprobe ist. Überprüft werden soll, ob die beiden Erwartungswerte identisch sind,d.h. $H_0:\mu_1=\mu_2$ gegen $H_1:\mu_1\neq\mu_2$.
Schreiben wir die Nullhypothese als $H_0:\mu_1-\mu_2=0$, so sehen wir, daß wir es wieder mit einem Spezialfall der linearen Hypothese *(3.51)* zu tun haben mit A=(1 -1) und c=0.

(4) Soll überprüft werden, ob die Erwartungswerte von drei unabhängigen normalverteilten Grundgesamtheiten identisch sind, so ist der Parametervektor $\beta=(\mu_1,\mu_2,\mu_3)'$, und die zu überprüfende Nullhypothese $H_0:\mu_1=\mu_2=\mu_3$ können wir auch schreiben als $H_0:\mu_1=\mu_2$ *und* $\mu_1=\mu_3$. Sie ist somit ein Spezialfall der linearen Hypothese *(3.51)* mit

$$A=\begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \text{ und } c=\begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Wir leiten nun die Teststatistik für die lineare Hypothese (3.51) her. Dabei gehen wir folgendermaßen vor:
Wir bestimmen den Schätzer $\hat{\beta}_0$ von β unter der durch die lineare Hypothese *(3.51)* beschriebenen Nebenbedingung. Ist die Nullhypothese *(3.51)* wahr, so sollte dieser in "der Nähe" des Kleinst-Quadrate-Schätzers $\hat{\beta}$ liegen. Als Kriterium für die Nähe wählen wir die quadrierte Länge des Differenzenvektors $\hat{y}-\hat{y}_0$, wobei $\hat{y}$ die Schätzung von y ohne die Beschränkung der Nullhypothese und $\hat{y}_0$ die Schätzung von y unter der Nullhypothese *(3.51)* ist. Es wird sich herausstellen, daß dies gerade die Differenz der Residuenquadratsummen unter H_0 und ohne die Beschränkung der Nullhypothese ist. Diese vergleichen wir dann mit der Residuenquadratsumme ohne Beschränkung der Nullhypothese.
Wir benötigen zunächst also den Schätzer von β unter der linearen Hypothese *(3.51)*. Die Nebenbedingung $A\beta=c$ besteht aus den r linearen Nebenbedingungen $a^1\beta=c_1,\ldots,a^r\beta=c_r$, wobei die $a^i, i=1,\ldots,r,$ die Zeilen der Matrix A sind. Wir können A also schreiben als

$$A=\begin{pmatrix} a^1 \\ \vdots \\ a^r \end{pmatrix}.$$

Unser Ziel ist es nun, den Kleinst-Quadrate-Schätzer von β unter der Nebenbedingung *(3.51)* zu bestimmen. Hierzu stellen wir die Lagrangefunktion auf und nutzen die soeben gewonnene Darstellung von A aus:

$$L(\beta,\lambda)=(Y-X\beta)'(Y-X\beta)+2\lambda_1(a^1\beta-c_1)+\ldots+2\lambda_r(a^r\beta-c_r)=$$
$$=(Y-X\beta)'(Y-X\beta)+2\lambda'(A\beta-c)$$

mit $\lambda=(\lambda_1,\ldots,\lambda_r)'$.
Notwendige Bedingungen für einen Extremwert sind, daß die partiellen Ableitungen nach β und λ gleich 0 sind.

Mit Hilfe der Differentiationsregeln *(3.7)* und *(3.8)* erhalten wir

$$\frac{\partial}{\partial\beta} L(\beta,\lambda)=-2X'Y+2X'X\beta+2A'\lambda\overset{!}{=}0 \qquad (3.52)$$

$$\frac{\partial}{\partial\lambda} L(\beta,\lambda)=2(A\beta-c) \overset{!}{=}0 \qquad (3.53)$$

Wir erhalten aus *(3.52)* und *(3.53)* die Gleichungen

$$(X'X)\hat{\beta}_0+A'\hat{\lambda}=X'Y \qquad (3.54)$$

$$A\hat{\beta}_0=c \qquad (3.55)$$

Aus *(3.54)* folgt

$$\hat{\beta}_0=(X'X)^{-1}X'Y-(X'X)^{-1}A'\hat{\lambda}=\hat{\beta}-(X'X)^{-1}A'\hat{\lambda} \qquad (3.56)$$

wobei $\hat{\beta}$ der Kleinst-Quadrate-Schätzer von β ohne die Beschränkung der Nullhypothese *(3.51)* ist.

Multipliziert man *(3.56)* von links mit A, so ergibt sich

$$A\hat{\beta}_0=A\hat{\beta}-A(X'X)^{-1}A'\hat{\lambda} \qquad (3.57)$$

Setzen wir nun in *(3.57)* die Gleichung *(3.55)* ein, so ergibt sich

$$A(X'X)^{-1}A'\hat{\lambda}=A\hat{\beta}-c \qquad (3.58)$$

Die Matrix $A(X'X)^{-1}A'$ ist eine reguläre (r,r)-Matrix, da A vollen Zeilenrang besitzt.Wir können also *(3.58)* von links mit

der Inversen von $A(X'X)^{-1}A'$ multiplizieren und erhalten

$$\hat{\lambda}=\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right) \qquad (3.59)$$

Setzen wir *(3.59)* in *(3.56)* ein, so ergibt sich

$$\hat{\beta}_0=\hat{\beta}-(X'X)^{-1}A'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right) \qquad (3.60)$$

Wir bestimmen nun die Residuenquadratsumme Q_0 unter H_0:

$$Q_0=\left(Y-\hat{Y}_0\right)'\left(Y-\hat{Y}_0\right)=\left(Y-X\hat{\beta}_0\right)'\left(Y-X\hat{\beta}_0\right)=$$

$$=\left(Y-X\hat{\beta}+X\hat{\beta}-X\hat{\beta}_0\right)'\left(Y-X\hat{\beta}+X\hat{\beta}-X\hat{\beta}_0\right)=$$

$$=\left(Y-X\hat{\beta}+X(\hat{\beta}-\hat{\beta}_0)\right)'\left(Y-X\hat{\beta}+X(\hat{\beta}-\hat{\beta}_0)\right)=$$

$$=\left((Y-X\hat{\beta})'+(\hat{\beta}-\hat{\beta}_0)'X'\right)\left((Y-X\hat{\beta})+X(\hat{\beta}-\hat{\beta}_0)\right)=$$

$$=\left(e'+(\hat{\beta}-\hat{\beta}_0)'X'\right)\left(e+X(\hat{\beta}-\hat{\beta}_0)\right)=$$

$$=e'e+e'X(\hat{\beta}-\hat{\beta}_0)+(\hat{\beta}-\hat{\beta}_0)'X'e+(\hat{\beta}-\hat{\beta}_0)'X'X(\hat{\beta}-\hat{\beta}_0)=$$

$$=e'e+(\hat{\beta}-\hat{\beta}_0)'X'X(\hat{\beta}-\hat{\beta}_0) \qquad (3.61)$$

da mit $e=y-X\hat{\beta}$ aufgrund von *(3.13)* gilt $e'X(\hat{\beta}-\hat{\beta}_0)=(\hat{\beta}-\hat{\beta}_0)'X'e=0$.

Nun ist $e'e$ die Residuenquadratsumme ohne die Beschränkung der Nullhypothese *(3.51)*. Wir schreiben $Q_e=e'e$. Also gilt

$$Q_0-Q_e=(\hat{\beta}-\hat{\beta}_0)'X'X(\hat{\beta}-\hat{\beta}_0) \tag{3.62}$$

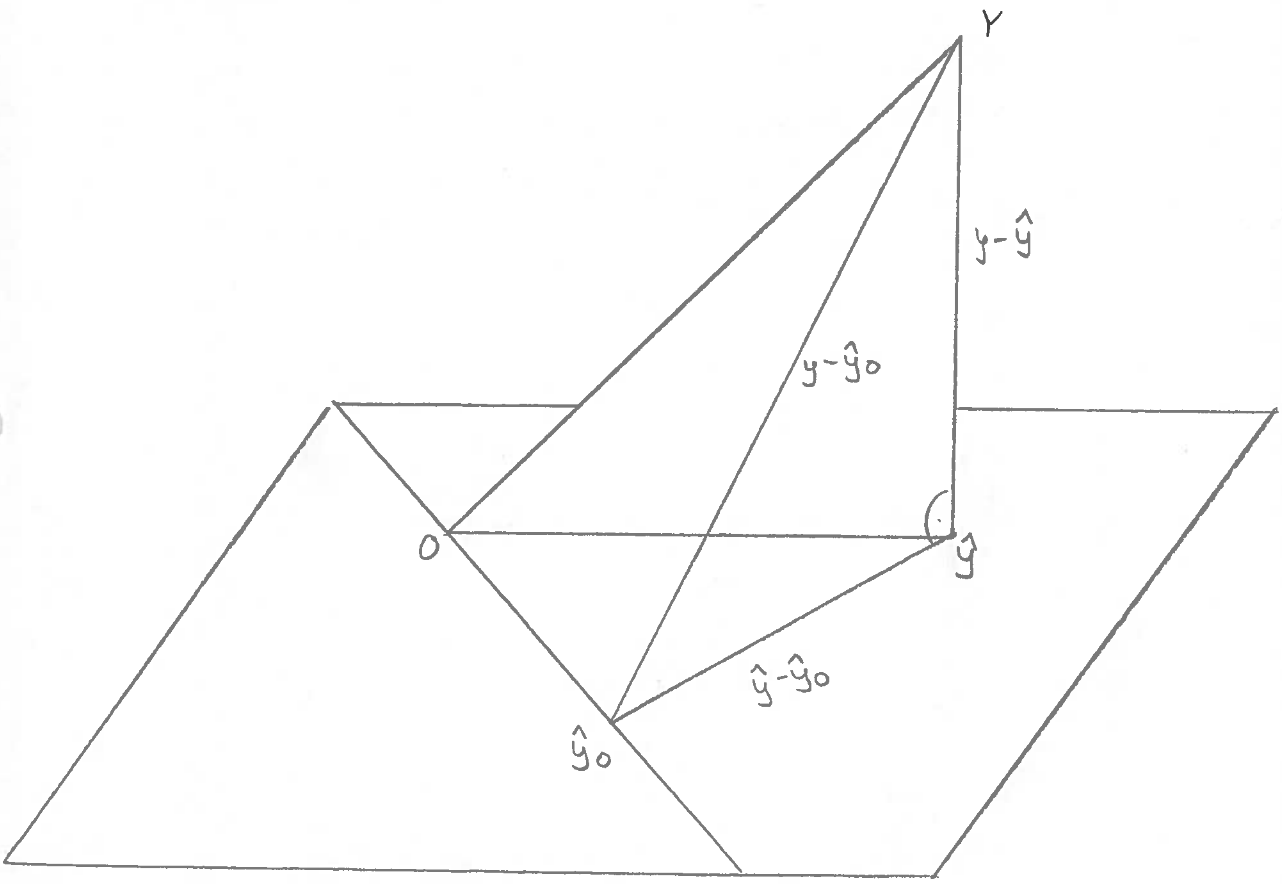
Bevor wir in *(3.62)* $\hat{\beta}_0$ einsetzen, interpretieren wir noch *(3.62)*.

Sei $\hat{y}=X\hat{\beta}$ und $\hat{y}_0=X\hat{\beta}_0$, dann gilt

$$(\hat{y}-\hat{y}_0)'(\hat{y}-\hat{y}_0)=(X\hat{\beta}-X\hat{\beta}_0)'(X\hat{\beta}-X\hat{\beta}_0)=(X(\hat{\beta}-\hat{\beta}_0))'(X(\hat{\beta}-\hat{\beta}_0))=$$

$$=(\hat{\beta}-\hat{\beta}_0)'X'X(\hat{\beta}-\hat{\beta}_0)=Q_0-Q_e$$

Q_0-Q_e ist also die quadrierte Länge des Differenzenvektors aus dem Schätzer von Y unter H_0 und dem Schätzer von Y ohne die Beschränkung von H_0. Die folgende Graphik veranschaulicht den Zusammenhang geometrisch. Je näher $\hat{y}$ an $\hat{y}_0$ ist, um so mehr sprechen die Daten für H_0.



Wir setzen nun in *(3.62)* $\hat{\beta}_0$ aus *(3.60)* ein.

$$Q_0-Q_e=(\hat{\beta}-\hat{\beta}_0)'X'X(\hat{\beta}-\hat{\beta}_0)=$$

$$=\left((X'X)^{-1}A'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)\right)'X'X\left((X'X)^{-1}A'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)\right)$$

$$=\left(A\hat{\beta}-c\right)'\left(A(X'X)^{-1}A'\right)^{-1}A(X'X)^{-1}X'X(X'X)^{-1}A'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)=$$

$$=\left(A\hat{\beta}-c\right)'\left(A(X'X)^{-1}A'\right)^{-1}A(X'X)^{-1}A'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)=$$

$$=\left(A\hat{\beta}-c\right)'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right) \qquad (3.63)$$

Der folgende *Satz* gibt die Verteilung von $\dfrac{Q_0-Q_e}{\sigma^2}$ unter $H_0\!:A\beta=c$ an.

$\boxed{\textit{Satz 3.23}}$

Im linearen Modell $Y=X\beta+\varepsilon$ mit ε $N_n\left(0,\sigma^2 I_n\right)$ gilt unter $H_0\!:A\beta=c$

$$\frac{Q_0-Q_e}{\sigma^2}=\frac{\left(A\hat{\beta}-c\right)'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)}{\sigma^2}\quad \chi^2(r) \qquad (3.64)$$

<u>Beweis</u>

Für $\hat{\beta}=(X'X)^{-1}X'Y$ gilt $\hat{\beta}$ $N_k\left(\beta,\sigma^2(X'X)^{-1}\right)$.

A ist eine (r,k)-Matrix mit $rg(A)=r$.

Also ist aufgrund von *Satz 3.6* $A\hat{\beta}$ r-dimensional normalverteilt mit $E(A\hat{\beta})=AE(\hat{\beta})=A\beta$ und $Cov(A\hat{\beta})=ACov(\hat{\beta})A'=\sigma^2A(X'X)^{-1}A'$.

Es gilt also $A\hat{\beta}$ $N_r\left(A\beta,\sigma^2A(X'X)^{-1}A'\right)$.

Unter H_0 gilt $A\beta=c$.

Also gilt aufgrund von *Satz 3.17*:

$$\left(A\hat{\beta}-c\right)'\sigma^{-2}\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)=\frac{\left(A\hat{\beta}-c\right)'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)}{\sigma^2}\quad \chi^2(r)$$

∎

Der folgende *Satz* zeigt, daß Q_0-Q_e und Q_e unabhängig sind.

$\boxed{\textit{Satz 3.24}}$

Im linearen Modell $Y=X\beta+\varepsilon$ mit $\varepsilon\ N_n\left(0,\sigma^2 I_n\right)$ sind

$Q_0-Q_e=\left(A\hat{\beta}-c\right)'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)$ und $Q_e=e'e$ unabhängig.

<u>Beweis</u>

Siehe Searle(1971),S.111 $\blacksquare$

Unter $H_0:A\beta=c$ gilt somit folgendes:

$$\frac{Q_0-Q_e}{\sigma^2}=\frac{\left(A\hat{\beta}-c\right)'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)}{\sigma^2}\ \chi^2(r)$$

$$\frac{Q_e}{\sigma^2}=\frac{(n-k)\hat{\sigma}^2}{\sigma^2}\ \chi^2(n-k)$$

Q_0-Q_e *und* Q_e sind unabhängig.

Somit ist unter $H_0:A\beta=c$ die Statistik

$$F=\frac{\dfrac{Q_0-Q_e}{r}}{\dfrac{Q_e}{n-k}}=\frac{\dfrac{\left(A\hat{\beta}-c\right)'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)}{r}}{\hat{\sigma}^2}$$

als Quotient von zwei unabhängigen jeweils durch ihre Freiheitsgrade dividierten chiquadratverteilten Zufallsvariablen F-verteilt mit r und $n-k$ Freiheitsgraden.
Die Hypothese $H_0:A\beta=c$ wird zum Niveau α abgelehnt, wenn gilt $F\geq F_{r,n-k;1-\alpha}$, wobei $F_{r,n-k;1-\alpha}$ das 1-α-Quantil einer F-Verteilung mit r und $n-k$ Freiheitsgraden ist.

Wir betrachten nun noch einmal die 4 Spezialfälle:

(1) $H_0: \beta=\beta^0$ gegen $H_1: \beta\neq\beta^0$

Hier ist $A=I_k$ und $c=\beta^0$. Wir erhalten

$$Q_0-Q_e=\left(A\hat{\beta}-c\right)'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)=\left(\hat{\beta}-\beta^0\right)'\left((X'X)^{-1}\right)^{-1}\left(\hat{\beta}-\beta^0\right)=$$
$$=\left(\hat{\beta}-\beta^0\right)'X'X\left(\hat{\beta}-\beta^0\right).$$

Die Teststatistik $F=\dfrac{\dfrac{\left(\hat{\beta}-\beta^0\right)'X'X\left(\hat{\beta}-\beta^0\right)}{k}}{\hat{\sigma}^2}$ ist unter H_0

F-verteilt mit k und $n-k$ Freiheitsgraden.

(2) $H_0: \beta_i=\beta_i^0$ gegen $H_1: \beta_i\neq\beta_i^0$ für ein $i\in 0,1,2,...,p$

Hier ist $A=(0,...,0,1,0,...,0)$, wobei die 1 an der $i+1$-ten Stelle steht, und c ist β_i^0. Wir erhalten

$$Q_0-Q_e=\left(A\hat{\beta}-c\right)'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)=\left(\hat{\beta}_i-\beta_i^0\right)\left(m_{ii}\right)^{-1}\left(\hat{\beta}_i-\beta_i^0\right)=$$

$=\dfrac{\left(\hat{\beta}_i-\beta_i^0\right)^2}{m_{ii}}$, wobei m_{ii} das $i+1$-te Element auf der

Hauptdiagonalen von $(X'X)^{-1}$ ist.

Unter H_0 ist die Teststatistik $F=\dfrac{\left(\hat{\beta}_i-\beta_i^0\right)^2}{m_{ii}\hat{\sigma}^2}$ F-verteilt mit 1 und $n-k$ Freiheitsgraden.

Wir können in diesem Fall auch einen t-Test durchführen, indem wir zur Teststatistik $t=\sqrt{F}=\dfrac{\left(\hat{\beta}_i-\beta_i^0\right)}{m_{ii}^{0.5}\hat{\sigma}}$, die unter H_0 mit $n-k$ Freiheitsgraden t-verteilt ist.

(3) Schauen wir uns das Zweistichprobenproblem als Spezialfall Linearen Modells nun etwas genauer an.
Wir nehmen an, daß die Zufallsvariablen

$Y_{11},\ldots,Y_{1n_1},Y_{21},\ldots Y_{2n_2}$ unabhängig sind und $Y_{i1},\ldots,Y_{in_i}$ identisch mit Parametern $\mu_i, i=1,2,$ und σ^2 normalverteilt sind. In Kapitel 1 haben wir das zugehörige Lineare Modell bereits formuliert. Es lautet:

$$Y_{ij}=\mu_i+\varepsilon_{ij} \qquad i=1,2$$

$$\varepsilon_{11},\ldots,\varepsilon_{2n_2} \quad i.i.d.\ N(0,\sigma^2)$$

Getestet werden soll $H_0:\mu_1=\mu_2$ gegen $H_1:\mu_1\neq\mu_2$.

In Matrixschreibweise lautet das Modell $Y=X\beta+\varepsilon$ mit

$$X=\begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}, \beta=\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad Y=\begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{2n_2} \end{pmatrix}, \quad \varepsilon=\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{2n_2} \end{pmatrix}$$

Nun erhalten wir

$$X'X=\begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}, \quad (X'X)^{-1}=\begin{pmatrix} \frac{1}{n_1} & 0 \\ 0 & \frac{1}{n_2} \end{pmatrix} \text{ und } X'Y=\begin{pmatrix} \sum_i Y_{1i} \\ \sum_j Y_{2j} \end{pmatrix}.$$

Somit gilt

$$\hat{\beta}=(X'X)^{-1}X'Y=\begin{pmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{pmatrix}.$$

Außerdem gilt $e=y-\hat{y}=y-X\hat{\beta}=\begin{pmatrix} Y_{11}-\bar{Y}_1 \\ \vdots \\ Y_{1n_1}-\bar{Y}_1 \\ Y_{21}-\bar{Y}_2 \\ \vdots \\ Y_{2n_2}-\bar{Y}_2 \end{pmatrix}$

Somit folgt

$$e'e=\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_i\right)^2 \text{ und}$$

$$\hat{\sigma}^2=\frac{1}{n_1+n_2-2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_i\right)^2 \qquad (3.65)$$

Die Hypothese $H_0:\mu_1=\mu_2$ ist ein Spezialfall der Linearen Hypothese $H_0:A\beta=c$ mit $A=(1\ \ -1)$ und $c=0$.

Somit gilt $Q_0-Q_e=\left(A\hat{\beta}-c\right)'\left(A(X'X)^{-1}A'\right)^{-1}\left(A\hat{\beta}-c\right)=$

$$=\left(\bar{Y}_1-\bar{Y}_2\right)\left((1\ \ -1)\begin{pmatrix}\frac{1}{n_1} & 0\\ 0 & \frac{1}{n_2}\end{pmatrix}\begin{pmatrix}1\\ -1\end{pmatrix}\right)^{-1}\left(\bar{Y}_1-\bar{Y}_2\right)=$$

$$=\left(\bar{Y}_1-\bar{Y}_2\right)^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}=\left(\bar{Y}_1-\bar{Y}_2\right)^2\left(\frac{n_1+n_2}{n_1\,n_2}\right)^{-1}$$

Wir erhalten somit als Teststatistik

$$F=\frac{\left(\bar{Y}_1-\bar{Y}_2\right)^2}{\frac{n_1+n_2}{n_1\,n_2}\hat{\sigma}^2}$$

Da gilt $rg(A)=1$, ist diese unter $H_0:\mu_1=\mu_2$ F-verteilt mit 1 und n_1+n_2-2 Freiheitsgraden.

$t=F^{0.5}$ ist die bekannte t-Statistik für das Zweistichprobenproblem.

Mit dem Beispiel (4) werden wir uns detailliert in Kapitel 5 beschäftigen.