Vorlesung: Lineare Modelle

Prof. Dr. Helmut Küchenhoff

Institut für Statistik, LMU München

SoSe 2014

Quadratsummenzerlegung

Gegeben sei das Modell (2.1) mit Design-Matrix X und

$$rg(X) = p + 1 =: p'. \tag{3.1}$$

Dann gilt:

$$\underbrace{(Y - \bar{Y})'(Y - \bar{Y})}_{\text{SST}} = \underbrace{(Y - \hat{Y})'(Y - \hat{Y})}_{\text{SSE}} + \underbrace{(\hat{Y} - \bar{Y})'(\hat{Y} - \bar{Y})}_{\text{SSM}} \tag{3.2}$$

Interpretation:

SST	:	Gesamt-Streuung, (korrigierte) Gesamt-Quadratsumme, "Total"
SSE	:	Fehler-Quadratsumme, "Error"
SSM	:	Modell-Quadratsumme, "Model"

Zerlegung ohne Absolutglied

Die Zerlegung (3.2) setzt ein Absolutglied in der Regression voraus.

Weitere Zerlegung, die nicht notwendig ein Absolutglied in der Regression voraussetzt:

$$\underbrace{Y'Y}_{\text{SST*}} = \underbrace{(Y - \hat{Y})'(Y - \hat{Y})}_{\text{SSE}} + \underbrace{\hat{Y}'\hat{Y}}_{\text{SSM*}} \tag{3.3}$$

SST^*	:	nicht korrigierte Gesamt-Quadratsumme Erfasst auch Abweichungen von $Y = 0$ und nicht nur von $\bar{Y}$.
SSE	:	Fehler-Quadratsumme, wie bei (3.2)
SSM^*	:	nicht korrigierte Modell-Quadratsumme

Nachweis der Quadratsummenzerlegung (3.3)

$$
\begin{aligned}
Y'Y &= Y'IY \\
&= Y'(Q+P)Y \\
&= Y'QY + Y'PY \\
&= Y'Q'QY + Y'P'PY \\
&= \hat{\varepsilon}'\hat{\varepsilon} + \hat{Y}'\hat{Y} \qquad\qquad \text{qed}
\end{aligned}
$$

P und Q Projektionsmatrizen.

Nachweis der Quadratsummenzerlegung (3.2) I

Wir definieren:

$$P_e = e(e'e)^{-1}e', \quad Q_e = I - P_e \text{ mit } \ e = (1, 1, \dots, 1)'. \tag{3.4}$$

P_e entspricht der Regression auf eine Konstante ($E(Y) = \beta_0$)

$$\begin{aligned} P_e Y &= \quad \bar{Y} && \text{(„Mittelwert bilden“)} \\ Q_e Y &= \quad Y - \bar{Y} && \text{(„Mittelwertsbereinigung“)} \end{aligned}$$

$$\begin{aligned} (Y - \bar{Y})'(Y - \bar{Y}) &= Y'Q_e'Q_e Y = Y'Q_e Y \\ &= Y'Q_e(Q + P)Y = Y'Q_e QY + Y'Q_e PY \\ &\overset{(*)}{=} \dots \end{aligned}$$

Nachweis der Quadratsummenzerlegung (3.2) II

$$(*) \quad Q_e QY = Q_e \hat{\varepsilon} \underset{\uparrow}{=} \hat{\varepsilon} = QY$$

„Residuen haben Mittelwert 0"

$$(*) \quad \begin{aligned} Q_e P Q_e P Y &= Q_e P Q_e \hat{Y} = Q_e P(\hat{Y} - \bar{Y}) \\ = Q_e P \hat{Y} - Q_e P \bar{Y} &\underset{\uparrow}{=} Q_e \hat{Y} - Q_e \bar{Y} = Q_e \hat{Y} - 0 \end{aligned}$$

„Regression auf $\hat{Y}$ und $\bar{Y}$ liefert Identität"

$$\begin{aligned} (Y - \bar{Y})'(Y - \bar{Y}) &= Y' Q_e' Q_e Y = Y' Q_e Y \\ &= Y' Q_e (Q + P) Y = Y' Q_e Q Y + Y' Q_e P Y \\ &\overset{(*)}{=} Y' Q Y + Y' Q_e P Q_e P Y \\ &= \hat{\varepsilon}' \hat{\varepsilon} + Y' P' Q_e P Y \\ &= (\hat{Y} - \bar{Y})'(\hat{Y} - \bar{Y}) \end{aligned}$$

Erwartungswerte der Quadratsummen

Wir betrachten das multiple Regressionsmodell (2.1) mit (2.2) bis (2.4) und

$$P_e = e(e'e)^{-1}e', \quad Q_e = I - P_e \text{ mit } \; e = (1, 1, \ldots, 1)'.$$

Dann gilt für die Erwartungswerte der Quadratsummen:

$$\begin{aligned}
E(SST^*) = E(Y'Y) &= \sigma^2 n + \beta' X' X \beta && (3.5)\\
E(SST) = E(Y - \bar{Y})'(Y - \bar{Y}) &= \sigma^2 (n-1) + \beta'(Q_e X)'(Q_e X)\beta && (3.6)\\
E(SSE) = E(\hat{\varepsilon}'\hat{\varepsilon}) &= \sigma^2 (n - p') && (3.7)\\
E(SSM^*) = E(\hat{Y}'\hat{Y}) &= \sigma^2 p' + \beta' X' X \beta && (3.8)\\
E(SSM) = E(\hat{Y} - \bar{Y})'(\hat{Y} - \bar{Y}) &= \sigma^2 p + \beta'(Q_e X)'(Q_e X)\beta && (3.9)
\end{aligned}$$

Bemerkungen

1. Die stochastischen Eigenschaften der Quadratsummen werden zur Konstruktion von Tests bezüglich β genutzt
2. $\beta = 0 \Longrightarrow E(Y'Y) = \sigma^2 n$
3. $\beta_1, ...\beta_p = 0 \Longrightarrow E(SSM) = \sigma^2 p$

Zum Nachweis von 3. benutzt man, dass Q_e der „Mittelwertsbereinigungs- Operator“ ist:

$$Q_e x = (I - P_e)x = x - \bar{x}$$

Die erste Spalte von $Q_e X$ ist also der Nullvektor.

Beweisidee

Allgemein berechnet man den Erwartungswert von quadratischen Formen wie folgt:

$$
\begin{aligned}
E(Y'AY) &= E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} Y_i Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} E(Y_i Y_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} E(Y_i) E(Y_j) + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} Cov(Y_i, Y_j) \\
&= E(Y)' A E(Y) + Sp(AV(Y)) \\
V(Y) &:= \text{Varianz-Kovarianzmatrix von Y}
\end{aligned}
$$

Unter Benutzung von $Sp(AV(Y)) = sp(A) * \sigma^2 = rang(A) * \sigma^2$ für Projektionsmatrizen erhält man obige Identitäten.

Mittlere Quadratsummen

Wir definieren entsprechend der Zahl der Freiheitsgrade die **mittleren Quadratsummen**:

$$MST^* := \frac{SST^*}{n} \tag{3.10}$$

$$MST := \frac{SST}{n-1} \tag{3.11}$$

$$MSE := \frac{SSE}{n-p'} \tag{3.12}$$

$$MSM^* := \frac{SSM^*}{p'} \tag{3.13}$$

$$MSM := \frac{SSM}{p} \tag{3.14}$$

Verteilungsdefinitionen

Normalverteilung

Ein n-dimensionaler Zufallsvektor Z heißt multivariat normalverteilt, falls für seine Dichtefunktion gilt:

$$f_z(z) = \frac{1}{\sqrt{\det \Sigma}\sqrt{(2\pi)^n}} \exp\left[-\frac{1}{2}(z-\mu)'\Sigma^{-1}(z-\mu)\right] \tag{3.15}$$

mit positiv definiter, symmetrischer Matrix Σ.

Bezeichnung: $Z \sim N_n(\mu, \Sigma)$

Eigenschaften der Normalverteilung

Ist $Z \sim N_n(\mu, \Sigma)$, so gilt:

1. **Momente:**
$$\begin{array}{rcl} E(Z) & = & \mu \\ V(Z) & = & \Sigma \end{array}$$

2. **Lineare Transformationen:**
Ist $A : R^n \longrightarrow R^m$ eine lineare Tranformation mit $rg(A) = m$
$$\Longrightarrow AZ \sim N_m(A\mu, A\Sigma A') \qquad (3.16)$$

3. **Orthogonale Transformation in unabhängige Komponenten (Spektralzerlegung)**
Es existiert eine Matrix $T \in R^{n \times n}$ mit $T'T = I$ und $T\Sigma T' = \text{ diag } (\lambda_1, \ldots, \lambda_n)$, so dass
$$TZ \sim N_n(T\mu, \text{ diag } (\lambda_1, \ldots, \lambda_n)) \text{ gilt.} \qquad (3.17)$$

Chi-Quadrat-Verteilung I

Ist $Z \sim N_n(\mu, I)$ so heißt $X = Z'Z$ (nicht-zentral) Chi-Quadrat-verteilt.

Bezeichnung: $X \sim \chi^2(n, \delta)$

n:	Zahl der Freiheitsgrade
$\delta := \mu'\mu$:	Nichtzentralitätsparameter

Im Fall $\delta = 0$ erhält man die zentrale $\chi^2(n)$-Verteilung.

Chi-Quadrat-Verteilung II

Eigenschaften: Ist $X \sim \chi^2(n, \delta)$, so gilt:

1. **Momente:**

$$\begin{aligned} E(X) &= n + \delta \\ V(X) &= 2n + 4\delta \end{aligned}$$

2. **Allgemeiner Bezug zur Normalverteilung**

$$Z \sim N_n(\mu, \Sigma) \qquad \Longrightarrow Z'\Sigma^{-1}Z \sim \chi^2(n, \mu'\Sigma^{-1}\mu) \tag{3.18}$$

t-Verteilung

Seien Z und W voneinander unabhängige Zufallsgrößen mit

$$\begin{aligned} Z &\sim N(\delta, 1), \\ W &\sim \chi^2(n). \end{aligned}$$

Dann heißt $X = \frac{Z}{\sqrt{\frac{W}{n}}}$ (nicht-zentral) t-verteilt.

Bezeichnung: $X \sim t(n, \delta)$

n: Zahl der Freiheitsgrade,
δ: Nicht-Zentralitätsparameter

Erwartungswert: Ist $X \sim t(n, \delta)$,so gilt:

$$E(X) = \delta \sqrt{\frac{n}{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \qquad \text{für } n > 1. \tag{3.19}$$

F-Verteilung

Seien W_1 und W_2 voneinander unabhängige Zufallsgrößen mit

$$\begin{aligned} W_1 &\sim \chi^2(n_1, \delta) \\ W_2 &\sim \chi^2(n_2) \end{aligned}$$

Dann heißt $X = \frac{W_1/n_1}{W_2/n_2}$ (nicht-zentral) F-verteilt.

Bezeichnung: $X \sim F(n_1, n_2, \delta)$

n_1: Zählerfreiheitsgrade

n_2: Nennerfreiheitsgrade

δ: Nichtzentralitätsparameter

Erwartungswert: Ist $X \sim F(n_1, n_2, \delta)$, so gilt:

$$E(X) = \frac{n_2}{n_2 - 2}(1 + \delta/n_1) \text{ für } n_2 > 2. \tag{3.20}$$

Unabhängigkeit von Quadratsummen (Satz von Cochran)

Sei $Z \sim N(\mu, I)$, dim $Z =$ n,
A $\in R^{n \times n}$, $A' = A$, $rg(A) = r$, $A^2 = A$,
$B \in R^{n \times n}, B^2 = B,\; B = B'$,
$C \in R^{m \times n}$.

Dann gilt:

$$Z'AZ \sim \chi^2(r, \mu' A \mu) \qquad (3.21)$$

$$CA = 0 \implies CZ \text{ und } Z'AZ \text{ sind unabh.} \qquad (3.22)$$

$$AB = 0 \implies Z'AZ \text{ und } Z'BZ \text{ sind unabh.} \qquad (3.23)$$

Beweis von (3.21)

Spektralzerlegung von A (3.17)

$A = T\Sigma T'$

$$\Sigma = \underbrace{\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & 0 \end{pmatrix}}_{r} \qquad \text{da } A^2 = A$$

$\Rightarrow Z'AZ = Z'T\Sigma T'Z$

$Var(T'Z) = T'IT = I$

$\Rightarrow Z'AZ = \widetilde{Z}\Sigma Z = \sum_{i=1}^{r} \widetilde{z}_i^2 \sim \chi^2(r, \mu' A\mu)$

Beweis von (3.22) und (3.23)

$$Cov(CZ, Z'A) = CCov(Z, Z')A = CA = 0$$

Damit sind CZ und $Z'A$ unabhängig (NV-Annahme!)
$\Rightarrow$ CZ und $(Z'A)(Z'A)' = Z'AZ$ unabh.

qed

$$Cov(AZ, Z'B) = ACov(Z, Z')B = AB = 0$$

Damit sind AZ und $Z'B$ unabhängig (NV-Annahme!)
$\Rightarrow$ $(AZ)'(AZ) = Z'AZ$ und $(Z'B)(Z'B)' = Z'BZ$ unabh.

qed

Verteilung des KQ-Schätzers unter Normalverteilung

Sei multiples Regressionsmodell (2.1) mit (2.5) und $rg(X) = p'$ gegeben.
Für den KQ-Schätzer $\hat{\beta}$ gilt:

$$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1}) \tag{3.24}$$

$$\hat{\Sigma}_{\hat{\beta}} := \hat{\sigma}^2 (X'X)^{-1} \tag{3.25}$$

$$(n-p')\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p') \tag{3.26}$$

$$\hat{\sigma} \text{ und } \hat{\beta} \quad \text{sind} \quad \text{unabhängig} \tag{3.27}$$

$$\hat{\sigma}_{\hat{\beta}_k} := \sqrt{c_{kk}}\hat{\sigma}, \tag{3.28}$$

$(c_{kk}$ entspr. Diagonalelement von $(X'X)^{-1})$

$$\frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_{\hat{\beta}_k}} \sim t(n-p', 0) \tag{3.29}$$

Beweis

(3.24) (Normalverteilung von β) folgt aus Eigenschaften der NV (3.16)

(3.26) folgt aus (3.21), da $\hat{\sigma}^2 = \frac{1}{n-p'}\hat{\varepsilon}'\hat{\varepsilon} = \frac{1}{n-p'}Y'QY$
$rg(Q) = sp(Q)$

(3.27) folgt aus (3.22):

$$C \in R^{p\times n}, CA = 0 \Longrightarrow CZ \text{ und } Z'AZ \text{ sind unabh.}$$

mit

$$\begin{aligned} &C := (X'X)^{-1}X', \quad A := Q, \quad Z := Y \\ &CA = (X'X)^{-1}X'Q = (X'X)^{-1}X'(I-P) = \\ &(X'X)^{-1}X' - (X'X)^{-1}X'X(X'X)^{-1}X' = 0. \end{aligned}$$

Overall-Tests

Sei Modell (2.1) mit (2.5) und $rg(X) = p'$ gegeben. Dann gilt für die mittlere Quadratsummen:

$$F_O = \frac{MSM}{MSE} \quad \sim \quad F(p, n - p', \sigma^{-2}\beta'(Q_e X)'(Q_e X)\beta) \qquad (3.30)$$

(3.30) folgt aus (3.21) und (3.23) mit

$$\begin{aligned} MSM &= SSM/p = Y'(P - P_e)Y/p \\ MSE &= SSE/(n - p') = Y'QY/(n - p') \\ Q(P - P_e) &= QP - QP_e = 0 \end{aligned}$$

Die Verteilung wird zur Konstruktion des folgenden Tests benutzt:
$H_0^O : \beta_1 = \beta_2 = \cdots = \beta_p = 0$
Lehne H_0^O ab, falls

$$F_O > F_{1-\alpha}(p, n - p') \qquad (3.31)$$

Overall Test mit $\beta_0 = 0$

$$F_O^* = \frac{MSM^*}{MSE} \quad \sim \quad F(p', n - p', \sigma^{-2}\beta'(X'X)\beta) \tag{3.32}$$

$H_0^{O*} : \beta_0 = \beta_1 = \cdots = \beta_p = 0$

Lehne H_0^{O*} ab, falls

$$F_0^* > F_{1-\alpha}(p', n - p') \tag{3.33}$$

$F_{1-\alpha}(p, n - p')$: $(1 - \alpha)$-Quantil der zentralen $F(p, n - p')$–Verteilung.

Dieser Test wird nur in Ausnahmefällen angewendet.

Lineare Hypothesen

Es sollen Hypothesen, die sich mit Hilfe linearer Transformationen von β darstellen lassen, betrachtet werden:

$$A \in R^{a\times(p+1)} \ , c \in R^a,$$

$$A\beta = c \ \text{ mit } rg(A) = a$$

Beispiele:

$$p = 2 \ \ \beta_1 = \beta_2 \quad \leftrightarrow \quad A = \begin{pmatrix} 0 & 1 & -1 \end{pmatrix}, c = 0$$

$$p = 4 \ \ \beta_3 = \beta_4 = 0 \quad \leftrightarrow \quad A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, c = 0$$

$$p = 4 \ \ \beta_2 = 3 \ \ \beta_3 + \beta_4 = 1 \quad \leftrightarrow \quad A = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}, c = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

Allgemeine lineare Hypothese, Wald-Test

Sei das Modell (2.1) mit (2.5) und $A \in R^{a \times p'}$, $rg(A) = a, c \in R^a$ gegeben.

$$\begin{aligned}
V(A\hat{\beta} - c) &= \sigma^2 A(X'X)^{-1}A' && (3.34)\\
SSH &:= (A\hat{\beta} - c)'(A(X'X)^{-1}A')^{-1}(A\hat{\beta} - c) && (3.35)\\
\sigma^{-2}SSH &\sim \chi^2(a, \sigma^{-2}(A\beta - c)'(A(X'X)^{-1}A')^{-1}(A\beta - c)) && (3.36)\\
MSH &:= \frac{SSH}{a} && (3.37)
\end{aligned}$$

$$\frac{MSH}{MSE} \sim F(a, n - p', \sigma^{-2}(A\beta - c)'(A(X'X)^{-1}A')^{-1}(A\beta - c)) \quad (3.38)$$

SSH: Quadratsumme, die die Abweichung von der Hypothese $A\beta = c$ beschreibt.

Test nach Wald: $H_0 : A\beta = c$. Lehne H_0 ab, falls:

$$TF = MSH/MSE > F_{1-\alpha}(a, n - p') \quad (3.39)$$

Overall-Test und zweiseitiger t-Test

Die Overall-Tests aus (3.31) und (3.33) sind spezielle Wald-Tests.

Nachweis von (3.33) erfolgt direkt, da hier A = I ist.

Der Nachweis von (3.31) erfolgt später.

Der mit Hilfe von (3.29) konstruierte zweiseitige Test auf $\beta_k = \beta_k^0$ ist ebenfalls ein Wald-Test.

Der Nachweis erfolgt direkt unter Benutzung von $A = (0 \cdots 1 \cdots 0)$

Lineare Hypothese und Modell mit Restriktion

Wir betrachten das lineare Modell mit der linearen Restriktion $A\beta = c$.
Alle Schätzungen aus diesem Modell werden mit $\hat{}$ gekennzeichnet.

Wir lösen das Minimierungsproblem:

$$(y - X\beta)'(y - X\beta) \rightarrow \min \quad \text{unter} \quad A\beta = c$$

mit der Lagrange-Methode:

$$S(\beta, \lambda) = (y - X\beta)'(Y - X\beta) + 2\lambda'(A\beta - c)$$

λ ist der Vektor der Lagrange-Multiplikationen.

Berechnung der Lösung $\hat{\hat{\beta}}$

$$
\begin{aligned}
\frac{\partial S}{\partial \beta} &= 0 \Leftrightarrow -2X'Y + 2X'X\hat{\hat{\beta}} + 2A'\lambda = 0 \\
\frac{\partial S}{\partial \lambda} &= 0 \Leftrightarrow A\hat{\hat{\beta}} = c \\
A'\lambda &= X'Y - X'X\hat{\hat{\beta}} \quad | \cdot (X'X)^{-1} \\
(X'X)^{-1}A'\lambda &= \hat{\beta} - \hat{\hat{\beta}} \quad | \cdot A \ (*) \\
\Rightarrow A(X'X)^{-1}A'\lambda &= A\hat{\beta} - c \quad (A\hat{\hat{\beta}} = c) \\
\Rightarrow \lambda &= (A(X'X)^{-1}A')^{-1}(A\hat{\beta} - c) \\
(*) \Rightarrow \hat{\hat{\beta}} &= \hat{\beta} - (X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - c) \\
w &:= (X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - c)
\end{aligned}
$$

Darstellung von SSH

$$
\begin{aligned}
\hat{\hat{Y}} &= X\hat{\hat{\beta}} = \hat{Y} - Xw \\
\hat{\hat{\varepsilon}} &= Y - \hat{\hat{Y}} = \hat{\varepsilon} + Xw \\
\Rightarrow \hat{\hat{\varepsilon}}'\hat{\hat{\varepsilon}} &= \hat{\varepsilon}'\hat{\varepsilon} + w'X'Xw \;\; (\text{da } \; X'\hat{\varepsilon} = 0) \\
&= \hat{\varepsilon}'\hat{\varepsilon} + (A\hat{\beta} - c)'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - c) \\
\Rightarrow SSH &= \hat{\hat{\varepsilon}}'\hat{\hat{\varepsilon}} - \hat{\varepsilon}'\hat{\varepsilon}
\end{aligned}
$$

Damit haben wir eine andere Möglichkeit, SSH aus den Residuenquadratsummen zu berechnen.

Lineare Hypothese (alternative Durchführung)

Sei das Modell (2.1) mit (2.5) und $A \in R^{a \times p'}$, $rg(A) = a, c \in R^a$ gegeben.

$$H_0 : A\beta = c$$

Berechne SSE des Modells und des Modells unter H_0:

$$
\begin{aligned}
SSE &= \hat{\varepsilon}' \hat{\varepsilon}' \\
SSE(H_0) &:= \hat{\hat{\varepsilon}}' \hat{\hat{\varepsilon}} && (3.40) \\
SSH &= SSE(H_0) - SSE && (3.41) \\
MSH &= SSH/a \\
TF &:= MSH/MSE = \frac{(SSE(H_0) - SSE)/a}{SSE/(n-p')} && (3.42) \\
TF &\sim F(a, n-p') \text{ unter } H_0 && (3.43)
\end{aligned}
$$

Das Vorgehen entspricht dem Wald- Test (3.39).

Spezialfall: Overall - Test

Overall- Test

$$H_0^O : \beta_1 = \beta_2 = \ldots = \beta_p = 0$$

Das Modell unter H_0 ist $y = \beta_0$

$$\begin{aligned} SSE(H_0) &= (Y - \hat{\hat{\beta}}_0)'(Y - \hat{\hat{\beta}}_0) = SST \\ SSH &= SSE(H_0) - SSE = SSM \end{aligned}$$

Damit entspricht hier der Wald -Test dem Test (3.31)

Weglassen von Einflussgrößen

Lineare Hypothese:

$$H_0 : \beta_k = \beta_{k+1} = \ldots = \beta_p = 0$$

Das Modell unter H_0 ist M1:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_{k-1} x_{k-1}$$

$$\begin{aligned} SSE(H_0) &= SSE(M1) \\ SSH &= SSE(M1) - SSE \end{aligned}$$

Vergleich der SSE der beiden Modelle.

Beispiel: Welche Faktoren stehen in Zusammenhang mit den Fehlern bei einem Lesetest

Daten von Christa Kieferle (Pädagogik, LMU)
Daten von 180 Kindern aus den 8 Klassen (3. und 4. Klassen Grundschule)

Zielgröße:
Anzahl der Fehler bei einem Lesetest

potentielle Einflussgrößen:
Geschlecht, Jahrgang, Leseförderzeit, sonstiges lesen (1= oft,.. 5= fast nie), Gameboy (1= oft,.. 5= fast nie), Jahrgang.

Modell

$$Y = \beta_0 + \beta_1 * GE + \beta_2 * JG + \beta_3 * LZ + \beta_4 * WOL + \beta_5 * WOG + \beta_6 * WOTV$$

Y:	Anzahl der Fehler
GE:	Indikator für männlich (= 1 für männlich, 0 sonst)
JG:	Indikator für Klassenstufe (=1 für 3. Klasse, 0 für 4.Klasse)
LZ:	Lesezeit in der Schule
WOL:	Wie oft wird sonst gelesen
WOG:	Wie oft wird Gameboy gespielt
WOTV:	Wie oft TV

Interpretation der Parameter, z.B β_1 : Geschlechtsunterschied

Annahme: Bei WOL ist der durchschnittliche Unterschied in der Fehlerzahl zwischen den Stufen 1-5 gleich groß (Änderung um 1 entspricht β_4

Auswertung

1. Schätzung des Gesamtmodells
2. Gibt es einen Zusammenhang zwischen den Variablen WOL, WOG, WOTV und Y ?
3. Ist der Einfluss von WOG und WOTV auf die Fehlerzahl gleich ?
4. Gibt es einen Zusammenhang von Lesezeit und WOL und der Fehlerzahl ?
5. Wie lässt die Assoziation zur Lesezeit quantifizieren?

Fragen 2-4 lassen sich als lineare Hypothesen in dem Modell formulieren. Fragen 1 und 5 betreffen die Schätzung des Modells.

Reparametrisierung des Modells unter linearer Restriktion

Wir betrachten das Modell (2.1) mit (2.5) unter der linearen Restriktion $A\beta = c$ mit $A \in R^{a \times p'}$, $rg(A) = a, c \in R^a$.

Dann gibt es Matrizen $B \in R^{p' \times (p'-a)}$ und $d \in R^{p'}$ mit:

$$V = Z\gamma + \varepsilon \tag{3.44}$$

$$V := Y - Xd \tag{3.45}$$

$$Z := XB \tag{3.46}$$

Das Modell ist das reparametrisierte Modell mit:

Zielgröße:	$V = Y - Xd \in R^n$
Design-Matrix:	$Z,\ Z \in R^{n \times (p'-a)};\ \ rgZ = p' - a$
Parameter:	$\gamma \in R^{p'-a}$
Störterm :	ε stimmt mit dem aus dem Grundmodell überein!

Zusammenhang Reparametrisierung und Modell unter linearer Restriktion

Es gilt:

$$\hat{\hat{\beta}} = B\hat{\gamma} + d \tag{3.47}$$

$$SSH = \hat{\hat{\varepsilon}}'\hat{\hat{\varepsilon}} - \hat{\varepsilon}'\hat{\varepsilon} \tag{3.48}$$

mit
- $\hat{\hat{\beta}}$: KQ-Schätzer unter Restriktion $A\beta = c$
- $\hat{\gamma}$: KQ-Schätzer aus Modell (3.44)
- $\hat{\hat{\varepsilon}}$: Residuenvektor aus Modell (3.44) $=$ Residuenvektor aus KQ-Schätzung unter Restriktion.

ML-Schätzung

Die maximierte Likelihood des Modells ist:

$$(2\pi)^{-n/2} \cdot \hat{\sigma}^{-n} \cdot exp\left[-\sum_{i=1}^{n} \frac{\hat{\varepsilon}_i^2}{2\hat{\sigma}^2}\right]$$

Es folgt (siehe Nachweis ML = KQ aus Kapitel 1)

$$\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^2}{n} \tag{3.49}$$

$$MaxL = C * \hat{\sigma}^n = C * (SSE/n)^{(n/2)} \tag{3.50}$$

Likelihood Quotienten-Test

Grundidee des Likelihood-Quotienten-Tests:
Vergleiche (Bilde den Quotienten) maximierte Likelihood des Modells unter H_0 mit maximierter Likelihood ohne H_0

Wir betrachten also den ML - Schätzer mit und ohne die Restriktion $A\beta = c$:

$$\begin{array}{lcl} \hat{\hat{\varepsilon}} & : & \text{Residuen unter dem Modell mit } H_0 \\ \hat{\varepsilon} & : & \text{Residuen unter dem Modell ohne Einschränkung} \end{array}$$

Die LQ- Teststatistik lautet dann:

$$\tau_{LQ} = \left(\frac{\hat{\hat{\sigma}}}{\hat{\sigma}} \right)^{-n} = \left(\frac{\hat{\hat{\varepsilon}}' \hat{\hat{\varepsilon}}}{\hat{\varepsilon}' \hat{\varepsilon}} \right)^{-n/2} \tag{3.51}$$

Wald-Test ist Likelihood-Quotienten-Test

Sei das Modell (2.1) mit (2.5) und $A \in R^{a \times p'}, rg(A) = a, c \in R^a$ gegeben.

$$H_0 : A\beta = c \;\text{ gegen }\; H_1 : A\beta \neq c$$

Dann ist der Wald-Test zu dem Likelihood-Quotienten-Test äquivalent, d. h. :
Die Testgröße des LQ-Tests ist eine streng monotone Funktion der Testgröße des Wald-Tests:

$$\tau_{LQ} \quad = \quad g\left[\frac{MSH}{MSE}\right] \tag{3.52}$$

g streng monoton

Beweis

Damit ist τ_{LQ} eine monotone Funktion der Testgröße des Wald-Tests:

$$\frac{MSH}{MSE} = \frac{(\hat{\hat{\varepsilon}}'\hat{\hat{\varepsilon}} - \hat{\varepsilon}'\hat{\varepsilon})/a}{\hat{\varepsilon}'\hat{\varepsilon}/(n-p')} =: \tau_W$$

$$\Rightarrow \left[\tau_W \cdot \frac{a}{(n-p)'} + 1\right]^{-n/2} = \tau_{LQ}$$

Sequentielle und Partielle Quadratsummen

Gegeben sei das Modell (2.1). Wir betrachten nun Teilmodelle, die durch Nullrestriktionen von Komponenten des Vektors β entstehen und deren Residuenquadratsummen.

$$R(\beta_{i1}, \ldots, \beta_{ik} | \beta_{j1}, \ldots, \beta_{jl}) = SSH = SSE(M1) - SSE(M2) \qquad (3.53)$$

M2: Modell, das die Parameter $\beta_{i1}, \ldots, \beta_{ik}, \beta_{j1}, \ldots, \beta_{jl}$ enthält.
M1: Modell, das die Parameter $\beta_{j1}, \ldots, \beta_{jl}$ enthält.
SSH: Hypothesenquadratsumme zur Hypothese
($\beta_{i1} = \ldots = \beta_{ik} = 0$) im Modell M2.

Partielle Quadratsummen

Die zu der Hypothese $\beta_i = 0$ gehörigen Quadratsummen bzgl. des Gesamtmodells heißen **partielle Quadratsummen**:

$$R(\beta_i|\beta_0, \dots \beta_{i-1}, \beta_{i+1}, \dots, \beta_p) = SSE(M_{-i}) - SSE \tag{3.54}$$

M_{-k}: Modell mit $\beta_k = 0$.

Beachte: Die partiellen Quadratsummen führen im Allgemeinen zu keiner Zerlegung von SST, d.h.

$$SST \neq \sum_{k=1}^{p} R(\beta_k|\beta_0, \dots, \beta_{k-1}, \beta_{k+1}, \dots, \beta_p) + SSE \tag{3.55}$$

Sequentielle Quadratsummen

Wir betrachten die Folge von Modellen:

$$M_0: \quad Y = \beta_0 + \varepsilon \tag{3.56}$$

$$M_1: \quad Y = \beta_0 + \beta_1 x_1 + \varepsilon \tag{3.57}$$

$$\dots$$

$$M_p: \quad Y = X\beta + \varepsilon \tag{3.58}$$

$$R(\beta_k|\beta_0, \dots, \beta_{k-1}) = SSE(M_{k-1}) - SSE(M_k) \tag{3.59}$$

heißen **sequentielle Quadratsummen** und es gilt:

$$SST = \sum_{k=1}^{p} R(\beta_k|\beta_0, \dots, \beta_{k-1}) + SSE \tag{3.60}$$

Bereinigung

Wir betrachten folgende Zerlegung des Modells (2.1):

$$
\begin{aligned}
Y &= X\beta + \varepsilon && (3.61)\\
Y &= (X_1 X_2)(\beta_1' \beta_2')' + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon && (3.62)\\
\hat\beta &= (X'X)^{-1}X'Y = (\hat\beta_1', \hat\beta_2')' && (3.63)\\
Q_2 &:= I - X_2(X_2'X_2)^{-1}X_2' && (3.64)\\
Y^* &:= Q_2 Y && (3.65)\\
X_1^* &:= Q_2 X_1 && (3.66)\\
\Longrightarrow \quad \hat\beta_1 &= (X_1^{*\prime}X_1^*)^{-1}X_1^{*\prime}Y^* && (3.67)
\end{aligned}
$$

Y^*, X_1^* : von X_2 bereinigte Variablen

Beweis von (3.67) I

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon \qquad X = (X_1 \ X_2) \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

KQ:
$Y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{\varepsilon}$

$Q_2 = I - P_2 = I - X_2(X_2'X)^{-1}X_2'$

Es gilt: $Q_2 Q = (I - P_2)(I - P) = I - P_2 - P + \underbrace{P_2 P}_{P_2} = I - P = Q$

(Residuen von Regression auf $(Y_1 \ Y_2)$ und Residuen bzgl. Y_2)

Analog: $QQ_2 = Q$

Beweis von (3.67) II

$$
\begin{aligned}
Q_2 Y &= Q_2 X_1 \hat{\beta}_1 + \underbrace{Q_2 X_2}_{0} \hat{\beta}_2 + \underbrace{\underbrace{Q_2 \hat{\varepsilon}}_{Q_2 Q Y}}_{Q} \\
\Rightarrow Q_2 Y &= Q_2 X_1 \hat{\beta}_1 + \hat{\varepsilon} \\
Y^* &= X_1^* \hat{\beta}_1 + \hat{\varepsilon} \qquad \text{Bereinigte Variablen}
\end{aligned}
$$

$$X_1^{*\prime} \hat{\varepsilon} = (Q_2 X_1)' \hat{\varepsilon} = X_1' Q_2 Q \hat{\varepsilon} = X_1' \hat{\varepsilon} = 0$$

$\Rightarrow \hat{\beta}_1$ ist KQ-Lösung der Regression von Y* auf X*.

Beispiele zur Bereinigung

Mittelwerts-Bereinigung beim einfachen linearen Regressionsmodell

$$\begin{aligned} X_2 &= (1 \ldots 1)' \\ X_1 &= (x_1 \ldots x_n)' \\ Y^* &= Y - \bar{y} \\ X_1^* &= x - \bar{x} \\ (X'^* X^*)^{-1} X'^* Y^* &= S_x^{-2} S_{xy} \end{aligned}$$

Geschlechtseffekt, Trendbereinigung

Beispiel: Zusammenhang zwischen Fehleranzahl bei Test zu starken Verben

Zielgröße :	Anzahl der Fehler bei Test
Einflussgrößen :	Geschlecht Alter Leseverhalten, Fernsehverhalten, etc.

Regression auf binäre Variable Geschlecht entspricht Mittelwertsschätzung

Bereinigung nach Geschlecht: Abziehen des jeweiligen Gruppenmittelwertes

Bonferroni-Konfidenzintervalle

Gegeben sei das Modell (2.1) mit der Normalverteilungsannahme (2.5). Dann sind für die Parameter $\beta_{i1}, \ldots, \beta_{ik}$

$$\hat{\beta}_{il} \pm \hat{\sigma}_{\hat{\beta}_{il}} t_{1-\alpha/2k}(n-p'), \;\; l = 1, \ldots, k \tag{3.68}$$

simultane Konfidenzintervalle zum Sicherheitsniveau $1-\alpha$:

$$P(\text{es gibt l} : |\beta_{il} - \hat{\beta}_{il}| > \hat{\sigma}_{\beta_{il}} t_{1-\alpha/2k}(n-p')) \leq \alpha \tag{3.69}$$

Die Parameter β_{il} können durch Linearkombinationen $\gamma_l = a'\beta$ mit den entsprechenden geschätzten Standardabweichungen ersetzt werden.

Konfidenzellipsoide

Sei das Modell (2.1) mit NV-Annahme (2.5) gegeben.
Dann ist

$$\left\{\beta \,|\, (\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta}) \leq p'\hat{\sigma}^2 F_{1-\alpha}(p', n - p')\right\} \tag{3.70}$$

eine Konfidenzregion für β.

Entsprechendes gilt für lineare Transformationen $\gamma = A\beta$:

$$\left\{\gamma \,|\, (\gamma - \hat{\gamma})' \widehat{V(\hat{\gamma})}^{-1} (\gamma - \hat{\gamma}) < \dim \gamma \cdot F_{1-\alpha}(\dim(\gamma), n - p'))\right\}$$

ist Konfidenzregion zum Sicherheitsniveau $1 - \alpha$.

Konfidenzintervalle nach Scheffe

Sei Modell (2.1) mit NV-Annahme (2.5) gegeben.
Dann sind

$$\hat{\beta}_j \pm \sqrt{p' F_{1-\alpha}(p', n - p')}\, \hat{\sigma}_{\hat{\beta}_j} \tag{3.71}$$

für die Parameter β_j und

$$\hat{\gamma} \pm \sqrt{p' F_{1-\alpha}(p', n - p')}\, \hat{\sigma}_{\hat{\gamma}}$$

für beliebige Linearkombinationen $\gamma = a'\beta$ simultane Konfidenzintervalle.

Bemerkungen

- Die KIs nach Scheffe sind bei einer Vielzahl von Parametern sinnvoll
- Die KIs nach Scheffe sind z.B zur Bestimmung von simultanen Konfidenzregionen für Y geeignet
- Es gibt in Analogie zu den KIs einen entsprechenden multiplen Test
- Allgemeine Strategie für KI siehe Hothorn, Bretz und Westfall: Simultaneous inference in general parametric models. Biometrical Journal, 2008; R-package „multcomp“

Das Gauss-Markov-Theorem

Sei das Modell

$$\begin{aligned} Y &= X\beta + \varepsilon, \quad \text{rg } X = p' \\ E(\varepsilon) &= 0 \\ V(\varepsilon) &= \sigma^2 I \end{aligned}$$

gegeben.
Dann ist der KQ-Schätzer $\hat{\beta}$ unter den erwartungsteuen linearen Schätzern derjenige mit der kleinsten Varianz: $\hat{\beta}$ ist BLUE-Schätzer (**b**est **l**inear **u**nbiased **e**stimator).

Ist $\tilde{\beta}$ ein weiterer Schätzer von β mit $E(\tilde{\beta}) = \beta$ und $\tilde{\beta} = CY$, so gilt:

$$V(\tilde{\beta}) \geq V(\hat{\beta})$$

$V(\tilde{\beta}) \geq V(\hat{\beta}) \Leftrightarrow V(\tilde{\beta}) = V(\hat{\beta}) + M$ mit M positiv semidefinit.

Konsistenz des KQ–Schätzers I

Sei das Modell (2.1) - (2.4) gegeben. Wir betrachten nun das Modell mit steigendem Stichprobenumfang n. Da die Einflussgrößen fest sind, gehen wir von einer gegebenen Folge x_n der Einflussgrößen aus.

Sei zu jedem $n > p'$

X_n	:	Designmatrix, die aus den ersten n Beobachtungen besteht
$\hat{\beta}^{(n)}$	:	KQ- Schätzer aus den ersten n Beobachtungen

Vor.: X_n hat vollen Rang für alle $n \geq p$

$$\lim_{n \to \infty} (X_n' X_n)^{-1} = 0.$$

Konsistenz des KQ–Schätzers II

Dann folgt die schwache Konsistenz des KQ-Schätzers (Konvergenz in Wahrscheinlichkeit):

$$\hat{\beta}^{(n)} \xrightarrow{P} \beta. \tag{3.72}$$

Sind die Störgrößen zusätzlich identisch verteilt, so folgt die starke Konsistenz (fast sichere Konvergenz)

$$\hat{\beta}^{(n)} \xrightarrow{\text{f.s.}} \beta. \tag{3.73}$$

Asymptotische Normalität des KQ–Schätzers

Sei das Modell (2.1) - (2.4) gegeben.

Sei zu jedem $n > p'$

X_n	:	Designmatrix, die aus den ersten n Beobachtungen besteht.
$\hat{\beta}^{(n)}$	:	KQ- Schätzer aus den ersten n Beobachtungen

Vor.: X_n hat vollen Rang für alle $n \geq p$

$$\lim_{n\to\infty} \max_{1\leq i\leq n} x_i'(X_n'X_n)^{-1}x_i = 0.$$

Dann folgt die asymptotische Normalität von β

$$\sigma^{-1}(X_n'X_n)^{1/2}(\hat{\beta}^{(n)} - \beta) \xrightarrow{d} N(0, I). \qquad (3.74)$$