Multiples lineares Regressionsmodell

Das multiple lineare Regressionsmodell hat die Form

$$Y_i = \underbrace{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}}_{\boldsymbol{x}_i^\top = (1, x_{i1}, \ldots, x_{ip})} + \varepsilon_i; i = 1, \ldots, n$$

$x_i^T \beta$

oder in Matrix-Vektor Notation: $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$ mit

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{pmatrix}$$ Designmatrix

$$\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

$Var(\varepsilon) = Var(Y) = \begin{pmatrix} \sigma^2 & & \\ & \sigma^2 & \\ & & \ddots \ \sigma^2 \end{pmatrix} \sigma^2 \cdot I$

↳ Homosked. und unkorrelierte Fehler

Wir nehmen dabei an, dass $\boldsymbol{X} \in \mathbb{R}^{n\times(p+1)}$ eine feste Design-Matrix mit vollem Rang ist und dass $\varepsilon \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. Wir definieren $p' := p + 1$. $=$ rank $rg(X)$

Kleinste Quadrate (KQ) Schätzer

Wir schätzen den Parameter(vektor) β durch

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^{p'}}{\arg\min} (\boldsymbol{Y} - \boldsymbol{X}\beta)^\top (\boldsymbol{Y} - \boldsymbol{X}\beta) \qquad (6)$$

und nennen $\hat{\beta}$ den KQ-Schätzer von β und $\hat{\varepsilon}_i := Y_i - \boldsymbol{x}_i^\top \hat{\beta}$ die Residuen.

Existenz und Berechnung vom KQ Schätzer

Der KQ-Schätzer existiert und ist eindeutig, falls $\underline{\boldsymbol{X}^\top \boldsymbol{X}}$ invertierbar ist. Dieser lässt sich berechnen als

symmetr. Produktsummenmatrix

$(X^T X)\hat{\beta} = X^T Y \iff$ $$\hat{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$$

Durch differenzieren von der Gleichung (2) erhält man $\hat{\beta}$ als Lösung der Normalengleichung

$\hat{Y}^T \hat{\varepsilon} = \hat{\beta}^T X^T \hat{\varepsilon} = 0$ (folgt auch aus ⊥ Proj.) $\Leftarrow$

$$\boldsymbol{X}^\top \hat{\varepsilon} = 0$$

Nur bei Modell mit Intercept

- $\underline{e}^T \hat{\varepsilon} = 0 \Longrightarrow \frac{1}{n} \sum_i \hat{\varepsilon}_i = 0$
- $\bar{\hat{Y}} = \bar{Y}$
- $Y^T \hat{\varepsilon}$

auch gilt immer $E(X^T \varepsilon) = 0$ $E(X^T) E(\varepsilon)$ mit $E(\varepsilon) = 0$

Gauss-Markov-Theorem

Sei das Modell $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$ gegeben mit $\mathbb{E}(\varepsilon) = \boldsymbol{0}$ und $\mathrm{Cov}(\varepsilon) = \sigma^2 \boldsymbol{I}$. Dann ist der KQ-Schätzer $\hat{\beta}$ der beste lineare erwartungstreue Schätzer (best linear unbiased estimator, BLUE) von β.

Das heißt, dass für jeden anderen linearen erwartungstreuen Schätzer $\tilde{\beta}$ von β gilt $\mathbb{V}(\hat{\beta}) \leq \mathbb{V}(\tilde{\beta})$.

Interpretation der Modellparameter

- ceteris paribus: alle anderen x-Variablen bleiben konstant. möglich ist, dass x (bzw. alle x_j im mult. Modell) $= 0$ sind β_0 kann sinnfreie Interpretation haben, wenn es garnicht
- (Theoretische) Interpretation: Steigt x_k um eine Einheit, so steigt Y (ceteris paribus) im Erwartungswert um β_k Einheiten. β_0: $EW(Y)$ wenn alle $x_k = 0$
  $\beta_0 = 0 \Rightarrow E(Y) = \bar{Y} = 0$
- (Empirische) Interpretation: Steigt x_k um eine Einheit, so steigt Y (ceteris paribus) im Durchschnitt um $\hat{\beta}_k$ Einheiten.
- ! β_k charakterisiert den Einfluss von x_k unter Berücksichtigung der übrigen Variablen (Confounder-Korrektur). Das heißt, dass in einem einfachen linearen Regressionsmodell mit $Y_i = \beta_0 + \beta_k' x_{ik} + \varepsilon_i$ wäre im Allgemeinen $\beta_k' \neq \beta_k$.

Eigenschaften des KQ-Schätzers

Gegeben dem multiplen linearen Modell, gilt für den KQ-Schätzer $\hat{\beta}$

$\hat{\beta} \perp \hat{\sigma}^2$

- Erwartungstreue: $\mathbb{E}(\hat{\beta}) = \beta$.
  ! Gilt auch ohne die Annahme $\varepsilon \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, solange $\mathbb{E}(\varepsilon) = \boldsymbol{0}$
- $\mathbb{V}(\hat{\beta}) = \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$.
  ! Gilt auch ohne die Annahme $\varepsilon \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, solange $\mathrm{Cov}(\varepsilon) = \sigma^2 \boldsymbol{I}$
  - $\hat{\beta}$ abh. von ε, da $\hat{\beta} = \beta + (X^T X)^{-1} X^T \varepsilon$
- $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1})$ $\hat{\beta} \perp \hat{\varepsilon}$ gilt aber, auch $\hat{\beta} \perp \varepsilon_{n+1}$ neu gezogen, frische Daten

Hat-Matrix und Residualmatrix

Gegeben dem multiplen linearen Modell mit $\mathrm{rang}(\boldsymbol{X}) = p'$ gilt

$$\hat{\boldsymbol{Y}} := \boldsymbol{X} \underbrace{(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}}_{\hat{\beta}}$$

orthogonale, von Vektor Projektion auf $col(X)$ $$\boldsymbol{P} := \underbrace{\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top}_{n \times n}$$ Hat-Matrix $P \cdot Y = \hat{Y}$ "P setzt Y Hut auf"

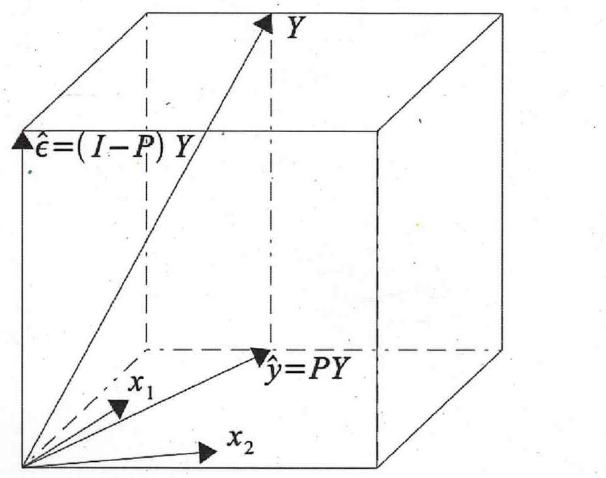
$$\hat{\varepsilon} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}$$ $\stackrel{!}{=} (I - P)\varepsilon$ nach Eins. von Modell $Y = X\beta + \varepsilon$ u. $P(X\beta) = X\beta$

$$\boldsymbol{Q} := \boldsymbol{I} - \boldsymbol{P}$$ $\iff I = (Q + P)$

P, Q beides quadr. $n \times n$ Matrizen

$\boldsymbol{P}$ heißt Hat-Matrix und $\boldsymbol{Q}$ heißt Residualmatrix.

Geometrische Interpretation



Die KQ-Schätzung ist eine orthogonale Projektion von $\boldsymbol{Y}$ auf den von den $\boldsymbol{x}$-Vektoren aufgespannten Unterraum. $col(X)$ ohne Intercept / nach Zentrierung bzw. $col([1, X])$ mit Interc.

Eigenschaften von P und Q

Die Hat-Matrix $\boldsymbol{P}$ und die Residualmatrix $\boldsymbol{Q}$ sind Projektionsmatrizen und zueinander orthogonal: $\hat{Y} = PY$

Als orthog. Projektionsmatrixen sind P, Q beide symmetrisch & idempotent

$$\boldsymbol{P}^\top = \boldsymbol{P} \text{ und } \underline{\boldsymbol{P}^2 = \boldsymbol{P}}$$ doppelte Regression $P(PY)$ führt zu selbem Ergebnis

$$\boldsymbol{Q}^\top = \boldsymbol{Q} \text{ und } \boldsymbol{Q}^2 = \boldsymbol{Q}$$ $PX = X \Rightarrow QX = (I - P)X = 0$

$$\boldsymbol{P}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{P} = 0$$ $\underline{PQ = 0}$, also $P(I - P) = 0$

Logisch auch: $P\hat{Y} = \hat{Y}$, $P\bar{Y} = \bar{Y}$

$(PQ)Y = P(QY) = P\hat{\varepsilon} = 0$

Regression von Residuen liefert $\hat{Y} = 0$!!

Daraus folgt

$$\mathbb{V}(\hat{\boldsymbol{Y}}) = \sigma^2 \boldsymbol{P}$$ $Var(\hat{Y}) = Var(PY) = P\sigma^2 P^T = \sigma^2 P^2 = \sigma^2 P$

$$\mathbb{V}(\hat{\varepsilon}) = \sigma^2 \boldsymbol{Q}, \text{ da } \hat{\varepsilon} = \boldsymbol{Q}\boldsymbol{Y}$$ $Var(\hat{\varepsilon}) = Var(QY) = \ldots$

Schätzer für σ^2

Gegeben dem multiplen linearen Modell, gilt

$$\hat{\sigma}^2 := \frac{\hat{\varepsilon}^\top \hat{\varepsilon}}{n - p'} = \frac{1}{n - p'} \sum_{i=1}^n \hat{\varepsilon}_i^2$$

ist ein erwartungstreuer Schätzer von σ^2.

! Gilt auch ohne die Annahme $\varepsilon \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, solange $\mathbb{E}(\varepsilon) = \boldsymbol{0}$ und $\mathrm{Cov}(\varepsilon) = \sigma^2 \boldsymbol{I}$

Quadratsummenzerlegung

Gegeben sei das multiple lineare Regressionsmodell mit $\text{rang}(\boldsymbol{X}) = p'$. Dann gilt

$$\underbrace{(\boldsymbol{Y}-\overline{\boldsymbol{Y}})^\top(\boldsymbol{Y}-\overline{\boldsymbol{Y}})}_{SST} = \underbrace{(\boldsymbol{Y}-\hat{\boldsymbol{Y}})^\top(\boldsymbol{Y}-\hat{\boldsymbol{Y}})}_{SSE\ \text{RSS}} + \underbrace{(\hat{\boldsymbol{Y}}-\overline{\boldsymbol{Y}})^\top(\hat{\boldsymbol{Y}}-\overline{\boldsymbol{Y}})}_{SSM}.$$

SST(otal): Gesamt-Quadratsumme (korrigiert)
SSE(rror): Fehler-Quadratsumme
SSM(odel): Modell-Quadratsumme

Quadratsummenzerlegung ohne β_0

nicht zentriert

Gegeben sei das multiple lineare Regressionsmodell mit, aber ohne Absolutglied β_0. Dann gilt

$$\underbrace{\boldsymbol{Y}^\top\boldsymbol{Y}}_{SST^*} = \underbrace{(\boldsymbol{Y}-\hat{\boldsymbol{Y}})^\top(\boldsymbol{Y}-\hat{\boldsymbol{Y}})}_{SSE} + \underbrace{\hat{\boldsymbol{Y}}^\top\hat{\boldsymbol{Y}}}_{SSM^*}. \quad = \varepsilon^T\varepsilon + \hat{y}^T\hat{y}$$

$y^Ty = y^TIy = y^T(Q+P)y = y^TQy + y^TPy \underset{\text{Eigensch. von } P, Q}{=} y^TQ^TQy + y^TP^TPy = (Qy)^TQy + (Py)^TPy$

SST^*: Gesamt-Quadratsumme (nicht korrigiert)
SSE: Fehler-Quadratsumme (wie zuvor)
SSM^*: Modell-Quadratsumme (nicht korrigiert)

$\big(E(\varepsilon^T)X\beta = (X\beta)^TE(\varepsilon) = 0\big)$

$E(SST^*) = E(y^Ty) = E[(X\beta+\varepsilon)^T(X\beta+\varepsilon)] = E[(\beta^TX^T+\varepsilon^T)(X\beta+\varepsilon)] = E(\beta^TX^TX\beta) + E(\varepsilon^T\varepsilon)$ $E(\varepsilon)=0$
$= \beta^TX^TX\beta + \sigma^2 n$ $\quad E(\varepsilon_i^2) \overset{!}{=} Var(\varepsilon_i) = \sigma^2$
$(\frac{1}{n}\sum \varepsilon_i^2)\cdot n = \sigma^2 n$
$= E(\sum \varepsilon_i^2) = E(\varepsilon^T\varepsilon)$

Erwartungswerte der Quadratsummen

Gegeben sei das multiple lineare Regressionsmodell mit den üblichen Annahmen. Wir definieren

P_e, Q_e symmetrisch & idempotent genau wie P, Q

$$\boldsymbol{e} = \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix} \text{ und } \boldsymbol{P_e} = \underbrace{\boldsymbol{e}(\boldsymbol{e}^\top\boldsymbol{e})^{-1}\boldsymbol{e}^\top}_{\text{Projektion auf den von } e \text{ aufgespannten 1D-Raum}} \text{ und } \boldsymbol{Q_e} = \boldsymbol{I} - \boldsymbol{P_e}.$$

$= e(n)^{-1}e^T$
$= 1/n \cdot ee^T$

$\hat{=}$ "MW" also Regression auf Konstante $E(y) = \beta_0$

mit $\bar{v} = \frac{1}{n}\sum_{i=1}^n v_i$ ($\vec{v}$ kann auch Spalte v. Matrix sein)

Dann gilt für alle n-dim. Vektoren (!) $\vec{v}$: $P_e v = \bar{v}\cdot e$; $Q_e v = v - \bar{v}e$

P_e: MW bilden $\qquad$ Q_e: Mittelwertbereinigungs-Operator, Zentrierung

$$\boldsymbol{P_e}\boldsymbol{Y} = \overline{\boldsymbol{Y}} = P_e\hat{y} \text{ und } \boldsymbol{Q_e}\boldsymbol{Y} = \boldsymbol{Y} - \overline{\boldsymbol{Y}}, \; Q_e\hat{\varepsilon} = \hat{\varepsilon} \; (\bar{\hat{\varepsilon}} = \bar{\varepsilon} = 0)$$

$P_e\hat{\varepsilon} = 0$

$$\begin{aligned}
\mathbb{E}(SST^*) \; (= E(y^Ty)) &= \underbrace{\sigma^2 n}_{\text{Rauschen}} + \underbrace{\beta^\top\boldsymbol{X}^\top\boldsymbol{X}\beta}_{\text{determin.}}\\
\mathbb{E}(SST) &= \sigma^2(n-1) + \beta^\top(\boldsymbol{Q_e X})^\top(\boldsymbol{Q_e X})\beta\\
\mathbb{E}(SSE) &= \sigma^2(n-p') \quad \sigma^2\cdot \text{trace}(I-H)\\
\mathbb{E}(SSM^*) &= \sigma^2 p' + \beta^\top\boldsymbol{X}^\top\boldsymbol{X}\beta\\
\mathbb{E}(SSM) &= \underbrace{\sigma^2(p'-1)}_{\sigma^2 p} + \beta^\top(\boldsymbol{Q_e X})^\top(\boldsymbol{Q_e X})\beta
\end{aligned}$$

Weil $Q_e x = x - \bar{x}$ folgt, dass 1. Spalte von Q_eX (Intercept...) $= \vec{0}$ ist

Wir können diese Eigenschaften zur Konstruktion von Tests verwenden. Es gilt nämlich unter anderem

$\beta = 0 \implies \mathbb{E}(SST^*) = \sigma^2 n$ $\quad SSM >> E(SSM) \Rightarrow \mathbb{P}(\text{mind. ein } \beta_k \neq 0) \to 1$

$\beta_1 = \dots = \beta_p = 0 \implies \mathbb{E}(SSM) = \sigma^2(p'-1)$

(ohne $\beta_0 = 0$, das wäre deterministische Regression ohne Fehler, $\hat{\varepsilon} = 0$)

Mittlere Quadratsummen

Wir definieren entsprechend der Zahl der Freiheitsgrade die mittleren Quadratsummen als

$$\text{MSE} = \frac{SSE}{n-p'} = \hat{\sigma}^2$$
$$\text{MSM} = \frac{SSM}{p'-1}$$
$$\text{MST} = \frac{SST}{n-1} \quad \neq MSE + MSM \text{ wegen untersch. Nenner}$$
$$\text{MSM}^* = \frac{SSM^*}{p'}$$
$$\text{MST}^* = \frac{SST^*}{n}$$

brauchen wir für lineare Hypothesen und Modellvergleiche (F-Tests).
Nur mittlere Quadratsummen sind untereinander sowie zwischen Modellen vergleichbar durch FG-Korrektur!

R^2

Gegeben sei das multiple lineare Regressionsmodell mit allen Annahmen. Dann definieren wir das Bestimmtheitsmaß R^2 als

$$R^2 = \frac{SSM}{SST} = 1 - \frac{SSE}{SST} = r^2_{\boldsymbol{Y}\hat{\boldsymbol{Y}}} \in [0,1]$$

wobei $r_{\boldsymbol{Y}\hat{\boldsymbol{Y}}}$ der Korrelationskoeffizient zwischen $\boldsymbol{Y}$ und $\hat{\boldsymbol{Y}}$ ist.

Wir interpretieren R^2 als den Anteil der Varianz von $\boldsymbol{Y}$, die durch das Modell erklärt wird. Ein hohes R^2 deutet darauf hin, dass wir unser Modell gut nutzen können, um $\boldsymbol{Y}$ zu erklären. R^2 steigt (oder bleibt konst.) mit steigender Anzahl an Kovariablen, auch wenn diese kaum/keine Erklärungskraft haben. Es ist daher nicht sinnvoll, R^2 zwischen Modellen zu vergleichen, die unterschiedlich viele Kovariablen haben. Hierfür nutzen wir das adjustierten Bestimmtheitsmaß

$$\begin{aligned}
R^2_{\text{adj}} &= \frac{MSM}{MST} = 1 - \frac{MSE}{MST} = 1 - \frac{SSE/(n-1)}{SST/(n-p')}\\
&= 1 - \frac{n-1}{n-p'}(1-R^2)
\end{aligned}$$

Art Penalisierung von zusätzl. Kovariablen

Für $n \gg p'$ gilt $R^2_{\text{adj}} \approx R^2$.

! Bei einem Modell ohne Absolutglied ist R^2 nach obiger Definition nicht sinnvoll, da es dann auch negative Werte annehmen kann. Das kommt daher, dass die Zerlegung $SST = SSE + SSM$ nicht mehr gilt. Stattdessen ist es sinnvoll, R^2 als $\frac{SSM^*}{SST^*}$

Multivariate Normalverteilung

Eine Zufallsvariable $\boldsymbol{Z} \in \mathbb{R}^n$ heißt multivariat normalverteilt mit Erwartungswert $\mu \in \mathbb{R}^n$ und positiv definiter Kovarianzmatrix $\Sigma \in \mathbb{R}^{n\times n}$, wenn ihre Dichte gegeben ist durch

$$f(\boldsymbol{z}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{z}-\mu)^\top\Sigma^{-1}(\boldsymbol{z}-\mu)\right).$$

Wir schreiben $\boldsymbol{Z} \sim \mathcal{N}_n(\mu, \Sigma)$.

Eigenschaften von $\mathcal{N}_n(\mu, \Sigma)$

Sei $\boldsymbol{Z} \sim \mathcal{N}_n(\mu, \Sigma)$ und $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ mit $\text{rang}(\boldsymbol{A}) = m$. Dann gilt

1. $\mathbb{E}(\boldsymbol{Z}) = \mu$
2. $\mathbb{V}(\boldsymbol{Z}) = \Sigma$
3. $\boldsymbol{AZ} \sim \mathcal{N}_m(\boldsymbol{A}\mu, \boldsymbol{A}\Sigma\boldsymbol{A}^\top)$
4. Es existiert eine orthogonale Matrix $\boldsymbol{T} \in \mathbb{R}^{n\times n}$ mit $\boldsymbol{T}\Sigma\boldsymbol{T}^\top = \text{diag}(\lambda_1, \dots, \lambda_n)$, so dass
$$\boldsymbol{TZ} \sim \mathcal{N}_n(\boldsymbol{T}\mu, \text{diag}(\lambda_1, \dots, \lambda_n))$$

Chi-Quadrat Verteilung

Sei $\boldsymbol{Z} \sim \mathcal{N}_n(\mu, \boldsymbol{I})$, so heißt $\boldsymbol{W} = \boldsymbol{Z}^\top\boldsymbol{Z} = \sum_{i=1}^n Z_i^2$ (nicht-zentral) Chi-Quadrat-verteilt und wir schreiben
$$W \sim \chi^2(n, \delta).$$

Wir nennen n die Zahl der Freiheitsgrade und $\delta = \mu^\top\mu$ den Nicht-Zentralitätsparameter. Es gilt
($\delta = 0$ ergibt zentrale $\chi^2(n)$-Verteilung)

$$\mathbb{E}(W) = n + \delta$$
$$\mathbb{V}(W) = 2n + 4\delta$$

Merke: $R^2 < 0$ bedeutet Modell performt schlechter als MW-Vorhersage $\hat{y} = \bar{y}$

t-Verteilung

Seien $Z \sim \mathcal{N}(\delta, 1)$ und $W \sim \chi^2(n, 0)$ unabhängig. Dann heißt $T = \frac{Z}{\sqrt{\frac{W}{n}}}$ (nicht-zentral) t-verteilt mit n Freiheitsgraden und Nicht-Zentralitätsparameter δ und wir schreiben

$$T \sim t(n, \delta).$$

Es gilt

$\mathbb{V}(T)$ existiert nicht!

$$\mathbb{E}(T) = \delta \sqrt{\frac{n}{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \quad \text{für } n > 1$$

F-Verteilung

Sei $W_1 \sim \chi^2(n_1, \delta)$ und $W_2 \sim \chi^2(n_2, 0)$ unabhängig. Dann heißt $X = \frac{W_1/n_1}{W_2/n_2}$ (nicht-zentral) F-verteilt mit n_1 und n_2 Freiheitsgraden und Nicht-Zentralitätsparameter δ und wir schreiben

$\delta = 0$: zentrale F-Verteilung $F(n_1, n_2)$

$$X \sim F(n_1, n_2, \delta).$$ linkssteile / rechtsschiefe Verteilung

Es gilt

$$\mathbb{E}(X) = \frac{n_2 + \frac{n_2 \delta}{n_1}}{n_2 - 2} \quad \text{für } n_2 > 2$$

Satz von Cochran

Unabhängigkeit von Quadratsummen der orthog. Projektionen auf X-Ebene

- $\boldsymbol{Z} \sim \mathcal{N}_n(\mu, \Sigma)$, $\dim(Z) = n$
- $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ mit $\operatorname{rang}(\boldsymbol{A}) = r$ und $\boldsymbol{A}^2 = \boldsymbol{A}$,
- $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ mit $\boldsymbol{B}^2 = \boldsymbol{B}$ und $\boldsymbol{B}^\top = \boldsymbol{B}$,
- $\boldsymbol{C} \in \mathbb{R}^{m \times n}$.

dann gilt

$$\boldsymbol{Z}^\top \boldsymbol{A} \boldsymbol{Z} \sim \chi^2(r, \mu^\top \boldsymbol{A} \mu)$$

$$\boldsymbol{C}\boldsymbol{A} = \boldsymbol{0} \implies \boldsymbol{C}\boldsymbol{Z} \text{ und } \boldsymbol{Z}^\top \boldsymbol{A} \boldsymbol{Z} \text{ sind unabhängig.}$$

$$\boldsymbol{A}\boldsymbol{B} = \boldsymbol{0} \implies \boldsymbol{Z}^\top \boldsymbol{A} \boldsymbol{Z} \text{ und } \boldsymbol{Z}^\top \boldsymbol{B} \boldsymbol{Z} \text{ sind unabhängig.}$$

Verteilung des KQ-Schätzers

Gegeben sei das multiple lineare Regressionsmodell mit $\operatorname{rang}(\boldsymbol{X}) = p'$ und den üblichen Annahmen über ε. Dann gilt für den KQ-Schätzer $\hat{\beta}$:

also auch Erwartungstreue: $\mathbb{E}(\hat{\beta}) = \beta$

$$\hat{\beta} \sim \mathcal{N}_p(\beta, \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1})$$

$$\hat{\Sigma}_{\hat{\beta}} := \hat{\sigma}^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$

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

$\hat{\sigma}^2$ und $\hat{\beta}$ sind unabhängig.

$$\hat{\sigma}^2_{\hat{\beta}_k} := (\hat{\Sigma}_{\hat{\beta}})_{kk}$$

$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2_{\hat{\beta}_k}}} \sim t(n - p', 0)$$

nur $\hat{\beta}$-Vektor ist NV
einzelne β_k sind t-verteilt, wobei t-Verteilung ähnlich wie NV (und für $n \to \infty$ $t \approx$ NV)

Um exakte Tests durchzuführen, ist die Normalverteilungsannahme für ε notwendig. Jedoch gelten einige Eigenschaften auch approximativ ohne diese Annahme. Nehmen wir stattdessen an, dass gilt

$$\lim_{n \to \infty} \frac{1}{n} X^\top X = V, \quad V \text{positive definit.}$$

Dann gilt weiterhin, dass $\hat{\beta}$ und $\hat{\sigma}^2$ consistent (konvergieren für $n \to \infty$ f.s. gegen β, σ^2) sind und

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}_p(\boldsymbol{0}, \sigma^2 V^{-1}).$$

Daraus folgt die für die praxis essenzielle Eigenschaft

$$\hat{\beta} \overset{a}{\sim} \mathcal{N}_p(\beta, \hat{\sigma}^2 \frac{(X^\top X)^{-1}}{n}) \text{ für großes } n.$$

$\Rightarrow$ auch ohne NV der Fehler können wir Inferenz betreiben, wenn n groß genug!

Overall-Test

Gegeben sei das multiple lineare Regressionsmodell mit $\operatorname{rang}(\boldsymbol{X}) = p'$. Dann gilt für die mittleren Quadratsummen $SSM \sim \chi^2(p'-1, \delta = \beta^T \dots / \sigma^2)$, $SSE \sim \chi^2(n-p', 0)$

$$F_0 = \frac{MSM}{MSE} \sim F(p' - 1, n - p', \frac{\beta^\top (\boldsymbol{Q}_e \boldsymbol{X})^\top (\boldsymbol{Q}_e \boldsymbol{X}) \beta}{\sigma^2})$$

Wir können damit den Overall-Test durchführen, um die Hypothese allgemeine Form $AB = 0$ mit $A = ([0, I])$ 1. Spalte Intercept egal

$$H_0^O : \beta_1 = \dots = \beta_{p'} = 0$$

bzw. $R^2 = 0$ zu testen. Wir lehnen H_0^O ab, wenn $F_0 > F_{1-\alpha}(p' - 1, n - p')$.

Allgemeine lineare Hypothese

Es sollen Hypothesen der Form $\boldsymbol{H}_0 : \boldsymbol{A}\beta = \boldsymbol{c}$ getestet werden, wobei $\boldsymbol{A} \in \mathbb{R}^{a \times p'}$ mit $\operatorname{rang}(\boldsymbol{A}) = a$ und $\boldsymbol{c} \in \mathbb{R}^a$.
Wir definieren

alternativ $SSH = \hat{\tilde{\varepsilon}}^T \hat{\tilde{\varepsilon}} - \hat{\varepsilon}^T \hat{\varepsilon}$

$$SSH := (\boldsymbol{A}\hat{\beta} - \boldsymbol{c})^\top (\boldsymbol{A}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{A}^\top)^{-1} (\boldsymbol{A}\hat{\beta} - \boldsymbol{c})$$

$$MSH := \frac{SSH}{a}$$

$$\delta_{SSH} := (\boldsymbol{A}\beta - \boldsymbol{c})^\top (\boldsymbol{A}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{A}^\top)^{-1} (\boldsymbol{A}\beta - \boldsymbol{c}).$$

SSH beschreibt die Quadratsumme der Abweichung von der Hypothese $\boldsymbol{A}\beta = \boldsymbol{c}$.
Es gilt

$$\frac{SSH}{\sigma^2} \sim \chi^2(a, \frac{\delta_{SSH}}{\sigma^2}),$$

$$TF = \frac{MSH}{MSE} \sim F(a, n - p', \frac{\delta_{SSH}}{\sigma^2}).$$

Damit können wir nun die Hypothese $H_0 : \boldsymbol{A}\beta = \boldsymbol{c}$ testen. Wir lehnen H_0 ab, wenn $\frac{MSH}{MSE} > F_{1-\alpha}(a, n - p')$.

Dieses Vorgehen können wir als Wald-Test identifizieren und in diesem Fall entsprich dieser einem Likelihood-Quotienten-Test, ist also optimal.

Der Test vergleicht intuitiv den SSE des Modells mit dem SSE des Modells unter $H_0 : \boldsymbol{A}\beta = \boldsymbol{c}$.

Für $n \to \infty$ gilt $\frac{MSH}{MSE} \to \frac{SSH}{\sigma^2}$, d.h. im Allgemeinen ist die F-Verteilung asymptotisch Chi-Quadrat-verteilt.

Partielle Quadratsummen

Gegeben sei das multiple lineare Regressionsmodell mit $\operatorname{rang}(\boldsymbol{X}) = p'$. Die zu der Hypothese $\boldsymbol{H}_0 : \beta_k = 0$ gehörende Quadratsumme bzgl. des Gesamtmodells heißt partielle Quadratsumme und wird definiert als

$$R(\beta_k | \beta_1, \dots, \beta_{k-1}, \beta_{k+1}, \dots, \beta_{p'}) = SSE(M_{-k}) - SSE$$

wobei M_{-k} das Modell mit $\beta_k = 0$ ist.