Institut für Statistik

Gerhard Tutz, Margret Oelker

Sommersemester 2014

Formelsammlung zur Vorlesung Statistik IV für Nebenfachstudierende

Version 25.02.2014

Diese Formelsammlung darf in der Klausur verwendet werden. Sie darf auf den bedruckten Seiten (nicht auf leeren Rückseiten) durch handschriftliche Formeln und Notizen ergänzt werden. Es dürfen keine zusätzlichen Blätter eingeheftet werden.

Inhaltsverzeichnis

1 Multivariate Schätz– und Testprobleme

1.1 Schätzung im Ein–Stichprobenfall

Daten:

- Seien $\mathbf{x}_1, \ldots, \mathbf{x}_n$ unabhängige Wiederholungen einer p-dimensionalen Zufallsvariablen $\mathbf{x}$ mit Erwartungswertvektor $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$ und Kovarianzmatrix $\mathrm{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}$
- Bezeichne $\mathbf{X}$ die $(n \times p)$–Datenmatrix $\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} = \begin{pmatrix} x_{11} & \ldots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \ldots & x_{np} \end{pmatrix}$

Erwartungswertschätzung:

$$\bar{\mathbf{x}} = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^n x_{i1} \\ \vdots \\ \sum_{i=1}^n x_{ip} \end{pmatrix} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{pmatrix};$$

Empirische Kovarianzmatrix:

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top = \frac{1}{n-1} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top - n \bar{\mathbf{x}} \bar{\mathbf{x}}^\top \right)$$

Alternativer Schätzer für $\boldsymbol{\Sigma}$:

$$\hat{\boldsymbol{\Sigma}} = \frac{n-1}{n} \cdot \mathbf{S} \quad \text{„nicht erwartungstreu“}$$

Schätzung der Korrelationsmatrix: $\hat{\mathbf{R}} = (r_{kl})_{p \times p}$

$$r_{kl} = \frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)(x_{il} - \bar{x}_l)}{\sqrt{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2} \sqrt{\sum_{i=1}^n (x_{il} - \bar{x}_l)^2}}$$

Satz:

$\mathbf{x}_1, \ldots, \mathbf{x}_n$ seien unabhängig und identisch ${}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$–verteilt. Dann gilt

a) $\mathrm{E}(\bar{\mathbf{x}}) = \boldsymbol{\mu}, \quad \mathrm{E}(\mathbf{S}) = \boldsymbol{\Sigma}$ Unverzerrtheit

b) $\mathrm{Cov}(\bar{\mathbf{x}}) = \frac{1}{n} \boldsymbol{\Sigma}$

c) $\mathrm{E} \parallel \bar{\mathbf{x}} - \boldsymbol{\mu} \parallel^2 \leq \mathrm{E} \parallel \hat{\boldsymbol{\mu}} - \boldsymbol{\mu} \parallel^2$ für jeden anderen unverzerrten linearen Schätzer $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = a_1 \mathbf{x}_1 + \ldots + a_n \mathbf{x}_n$.

Sind $\mathbf{x}_1 \ldots, \mathbf{x}_n$ zusätzlich normalverteilt, dann gilt außerdem

d) $\bar{\mathbf{x}} \sim N_p \left(\boldsymbol{\mu}, \frac{1}{n} \boldsymbol{\Sigma} \right)$ und $(n-1)\mathbf{S} = \boldsymbol{\Sigma}_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top \sim \mathbf{W}_p(\boldsymbol{\Sigma}, n-1)$

e) $\bar{\mathbf{x}}$ und $\mathbf{S}$ sind unabhängig.

1.2 Schätzung im Mehr–Stichprobenfall

Daten: Grundgesamtheit ist aufgespalten in disjunkte Klassen $\Omega_1, \ldots, \Omega_g$. Separat in den Klassen werden Stichproben gezogen.

$$\begin{array}{lll} \mathbf{x}_1^{(1)} & \ldots & \mathbf{x}_{n_1}^{(1)} \quad (\text{aus } \Omega_1) \\ \vdots & & \\ \mathbf{x}_1^{(g)} & \ldots & \mathbf{x}_{n_g}^{(g)} \quad (\text{aus } \Omega_g) \end{array}$$

Arithmetrisches Mittel in Klasse k:

$$\bar{\mathbf{x}}^{(k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{x}_i^{(k)} \to \boldsymbol{\mu}_k = \begin{pmatrix} \mu_1^{(k)} \\ \vdots \\ \mu_p^{(k)} \end{pmatrix}$$

Empirische Kovarianzmatrix in Klasse k:

$$\mathbf{S}^{(k)} = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (\mathbf{x}_i^{(k)} - \bar{\mathbf{x}}^{(k)})(\mathbf{x}_i^{(k)} - \bar{\mathbf{x}}^{(k)})^\top \to \boldsymbol{\Sigma}_k$$

Häufige Annahme: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \ldots = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$

Wenn Gleichheit gilt, dann nimmt man die gepoolte empirische Kovarianzmatrix:

$$\mathbf{S} = \frac{1}{n-g} \sum_{k=1}^{g} (n_k - 1)\mathbf{S}^{(k)} = \frac{1}{n-g} \sum_{k=1}^{g} \sum_{i=1}^{n_k} (\mathbf{x}_i^{(k)} - \bar{\mathbf{x}}^{(k)})(\mathbf{x}_i^{(k)} - \bar{\mathbf{x}}^{(k)})^\top = \frac{1}{n-g} \mathbf{W}.$$

wobei $\mathbf{W} = \sum_{k=1}^{g} \sum_{i=1}^{n_k} (\mathbf{x}_i^{(k)} - \bar{\mathbf{x}}^{(k)})(\mathbf{x}_i^{(k)} - \bar{\mathbf{x}}^{(k)})^\top$ die Streuung innerhalb der Gruppen bezeichnet. Weiter bezeichnet

$$\mathbf{B} = \sum_{k=1}^{g} n_k (\bar{\mathbf{x}}^{(k)} - \bar{\mathbf{x}})(\bar{\mathbf{x}}^{(k)} - \bar{\mathbf{x}})^\top$$

die Streuung zwischen den Gruppen, mit

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{k=1}^{g} n_k \bar{\mathbf{x}}^{(k)}$$

als großer Mittelwert über alle Klassen bezeichnet, mit

$$n = n_1 + n_2 + \ldots + n_g \, .$$

Für die Gesamtstreuungsmatrix

$$\mathbf{T} = \sum_{k=1}^{g} \sum_{i=1}^{n_k} (\mathbf{x}_i^{(k)} - \bar{\mathbf{x}})(\mathbf{x}_i^{(k)} - \bar{\mathbf{x}})^\top$$

gilt die folgende Zerlegung

$$\mathbf{T} = \mathbf{W} + \mathbf{B}.$$

Es gelten folgende Aussagen:

1. Für $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \ldots = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$ ist $\mathbf{S}$ ein unverzerrter Schätzer.
2. Gilt zusätzlich die Normalverteilungsannahme, so ist

$$\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n-g)$$

   Wishart-verteilt mit $n-g$ Freiheitsgraden und unabhängig von $\mathbf{B}$.

3. Für $\boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_g$ gilt unter Normalverteilungsannahme

$$\mathbf{B} \sim W_p(\boldsymbol{\Sigma}, g-1).$$

1.3 Tests, Konfidenzbereiche für Erwartungswerte im Ein–Stichprobenfall; $\boldsymbol{\Sigma}$ bekannt

Daten:

- $\mathbf{x}_1, \ldots, \mathbf{x}_n$ unabhängige Realisierungen der Stichprobenvariablen
- $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad i = 1, \ldots, n$

Hypothesen:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \ , \ H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

Teststatistik:

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0),$$
$$T^2 \overset{H_0}{\sim} \chi^2(p)$$

Ablehnbereich:

$$T^2 > \chi^2_{1-\alpha}(p)$$

1.4 Tests, Konfidenzbereiche für Erwartungswerte im Ein–Stichprobenfall; $\boldsymbol{\Sigma}$ unbekannt

Daten:

- $\mathbf{x}_1, \ldots, \mathbf{x}_n$ unabhängige Realisierungen der Stichprobenvariablen
- $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mit unbekanntem $\boldsymbol{\Sigma}$

Hypothesen:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \ , \ H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

Teststatistik:

$$T_s = \frac{n(n-p)}{p(n-1)}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0),$$

$$T_s \overset{H_0}{\sim} F(p, n-p)$$

Ablehnbereich:

$$T_s > F_{1-\alpha}(p, n-p)$$

1.5 Tests im Zwei–Stichprobenfall – unabhängige Stichproben

Daten:

- $\mathbf{X}_1 = (\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n_1})^T$ ist eine unabhängige Stichprobe aus einer $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$–verteilten Grundgesamtheit,
- $\mathbf{X}_2 = (\mathbf{x}_{21}, \ldots, \mathbf{x}_{2n_2})^T$ ist eine unabhängige Stichprobe aus einer $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$–verteilten Grundgesamtheit,
- Es werden gleiche Kovarianzmatrizen $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ vorausgesetzt.

Hypothesen:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \ , \ H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

Teststatistik:

$$T^2 = \frac{n_1 \cdot n_2}{n_1 + n_2} \cdot (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top \mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

$$\text{mit } \mathbf{S} = \frac{1}{n_1 + n_2 - 2}\left[(n_1 - 1) \cdot \mathbf{S}_1 + (n_2 - 1) \cdot \mathbf{S}_2\right]$$

$$\text{und } \mathbf{S}_i = \frac{1}{n_i - 1}\sum\nolimits_{j=1}^{n_i} (\mathbf{x}_j^{(i)} - \bar{\mathbf{x}}^{(i)})(\bar{\mathbf{x}}_j^{(i)} - \bar{\mathbf{x}}^{(i)})^\top$$

Ablehnbereich:

$$\frac{n_1+n_2-p-1}{(n_1+n_2-2)\cdot p}\cdot T^2 > F_{1-\alpha}(p; n_1+n_2-p-1)$$

$$\text{weil } \frac{n_1+n_2-p-1}{(n_1+n_2-2)\cdot p}\cdot T^2 \overset{H_0}{\sim} F(p; n_1+n_2-p-1)$$

1.6 Tests im Zwei–Stichprobenfall – verbundene Stichproben

Daten:

$$\begin{pmatrix}\mathbf{x}_i^{(1)}\\ \mathbf{x}_i^{(2)}\end{pmatrix} \sim N_{2p}\left(\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}; \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{22}\end{pmatrix}\right)$$

Hypothesen:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \Leftrightarrow \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}\ ,\ \mathbf{H_1} : \boldsymbol{\mu_1} \neq \boldsymbol{\mu_2}$$

$$\begin{aligned}\mathbf{d}_i &= \mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)} \Rightarrow H_0 \text{ besagt: } E(\mathbf{d}_i) = \mathbf{0}\\ \mathbf{d}_i &\sim N_p(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2; \boldsymbol{\Sigma}_{11} + \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{21})\end{aligned}$$

Unbekannte Kovarianzmatrix:

$$H_0 \text{ ablehnen} \Leftrightarrow \frac{(n-p)\cdot n}{(n-1)\cdot p}\cdot \bar{\mathbf{d}}^\top \cdot \mathbf{S}_d^{-1}\cdot \bar{\mathbf{d}} > F_{1-\alpha}(p; n-p)$$

$$\text{mit } \bar{\mathbf{d}} = \frac{1}{n}\sum_{i=1}^n \mathbf{d}_i$$

$$\mathbf{S}_d = \frac{1}{n-1}\sum_{i=1}^n (\mathbf{d}_i - \bar{\mathbf{d}})(\mathbf{d}_i - \bar{\mathbf{d}})^\top$$

Bekannte Kovarianzmatrix von $\mathbf{d}$ sei $\boldsymbol{\Sigma}_d$:

$$H_0 \text{ ablehnen} \Leftrightarrow n\cdot \bar{\mathbf{d}}^\top \cdot \boldsymbol{\Sigma}_d^{-1}\cdot \bar{\mathbf{d}} > \chi^2_{1-\alpha}(p)$$

2 Grundkonzepte der linearen Regression

2.1 Modellannahmen der Regression

Ein lineares Regressionsmodell hat die Form

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n$$

bzw. in Matrixschreibweise

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

mit

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix},$$

wobei

- $\mathbf{y}$: Zufallsvektor der Zielgröße
- $\mathbf{X}$: feste Design-Matrix (Matrix der Kovariablen)
- $\boldsymbol{\epsilon}$: Zufallsvektor der Fehlerterme
- $\boldsymbol{\beta}$: Vektor der Regressionsparameter der Länge $p + 1$

Allgemeine Annahmen:

$$\begin{aligned} \mathrm{E}(\boldsymbol{\epsilon}) &= \mathbf{0} \\ \mathrm{Cov}(\boldsymbol{\epsilon}) &= \sigma^2 \mathbf{I} \end{aligned}$$

Spezielle Verteilungsannahme:

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

2.2 Der KQ-Schätzer

- Aus dem KQ-Ansatz

  $$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \to \min_{\boldsymbol{\beta}}$$

  ergibt sich durch Nullsetzen der ersten Ableitung nach $\boldsymbol{\beta}$ und, falls $\mathbf{X}^\top\mathbf{X}$ invertierbar ist, anschließendem Lösen des resultierenden Gleichungssystems der KQ-Schätzer

  $$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

- Varianz der Schätzer $\hat{\beta}_j,\ j = 0, 1, \ldots, p$:
  Mit den Diagonalelementen v_j aus $(\mathbf{X}^\top\mathbf{X})^{-1}$ erhält man für bekanntes σ^2 als Varianz von $\hat{\beta}_j$

$$\sigma_j^2 = \text{Var}(\hat{\beta}_j) = \sigma^2 v_j$$

  bzw. für unbekanntes σ^2 die geschätzte Varianz von $\hat{\beta}_j$ gemäß

$$\hat{\sigma}_j^2 = \hat{\sigma}^2 v_j.$$

- Zusammenfassende Darstellung in Vektornotation

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} \quad \text{bzw.} \quad \widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\mathbf{X}^\top\mathbf{X})^{-1}$$

- Der **Vektor der geschätzten Residuen** ist

$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}.$$

Der KQ-Schätzer besitzt folgende Eigenschaften:

1. Ist $\mathbf{X}^\top\mathbf{X}$ invertierbar, so gilt
   $\hat{\boldsymbol{\beta}}$ **existiert** und $\hat{\boldsymbol{\beta}}$ ist **eindeutig**.
2. Der KQ-Schätzer ist **erwartungstreu**, d.h. es gilt

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}.$$

3. Unter der speziellen Verteilungsannahme gilt

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

- Unter der allgemeinen Annahme ist

$$\hat{\sigma}^2 = \frac{1}{n-p-1}\hat{\boldsymbol{\epsilon}}^\top\hat{\boldsymbol{\epsilon}} = \frac{1}{n-p-1}\sum_{i=1}^{n}\hat{\boldsymbol{\epsilon}}_i^2$$

  ein erwartungstreuer Schätzer für σ^2.

- Prognose:

$$\hat{y}_0 = \mathbf{x}_0^\top\hat{\boldsymbol{\beta}}$$

  $(1-\alpha)$-Prognoseintervall für $\hat{y}_0$ unter der speziellen Verteilungsannahme:

$$\left[\hat{y}_0 - t_{1-\frac{\alpha}{2}}\,(n-p-1)\,\hat{\sigma}\sqrt{\mathbf{x}_0^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_0 + 1}\,,\, \hat{y}_0 + t_{1-\frac{\alpha}{2}}\,(n-p-1)\,\hat{\sigma}\sqrt{\mathbf{x}_0^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_0 + 1}\right].$$

2.3 Quadratsummenzerlegung

Gegeben sei ein lineares Modell mit der Designmatrix **X**, die vollen Rang hat. Dann gilt

$$\sum_{i=1}^{n}(y_i - \overline{y}_i)^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \overline{y}_i)^2$$

bzw. in Matrixnotation

$$\underbrace{(\mathbf{y} - \overline{\mathbf{y}})^\top(\mathbf{y} - \overline{\mathbf{y}})}_{SST} = \underbrace{(\mathbf{y} - \hat{\mathbf{y}})^\top(\mathbf{y} - \hat{\mathbf{y}})}_{SSE} + \underbrace{(\hat{\mathbf{y}} - \overline{\mathbf{y}})^\top(\hat{\mathbf{y}} - \overline{\mathbf{y}})}_{SSM}$$

wobei $\hat{y}_i = \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ und $\bar{\mathbf{y}} = (\overline{y}, \ldots, \overline{y})$.

Interpretation:

SST :	Gesamt-Streuung, Gesamt-Quadratsumme
SSE :	Fehler-Quadratsumme, Residuen-Quadratsumme
SSM :	Modell-Quadratsumme

Wichtig bei dieser Zerlegung ist ein Absolutglied in der Regressionsgleichung.

2.4 Kodierung von Kovariablen

Betrachtet wird ein nominales Merkmal P mit K Ausprägungen (faktorielle Kovariable auf K Stufen). Folgende Kodierungen sind möglich:

1. **Einfache Dummy-Kodierung**

$$x_j(P) = x_{P(j)} = \begin{cases} 1, & P = j, \\ 0, & P \neq j, \end{cases} \quad \text{für } j = 1, \ldots, K$$

2. **Effekt-Kodierung**

$$x_j(P) = x_{P(j)} \begin{cases} 1, & P = j, \\ -1, & P = K, \\ 0, & \text{sonst,} \end{cases} \quad \text{für } j = 1, \ldots, K-1$$

2.5 Maße für die Modellgüte

1. Bestimmtheitsmaß:

$$R^2 = \frac{SSM}{SST} = 1 - \frac{SSE}{SST}$$

2. Korrigiertes Bestimmtheitsmaß:

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

3. Akaikes Informationskriterium:

$$AIC = n\ln(SSE) + 2(p+1) - n\ln(n)$$

4. Bayes Informationskriterium

$$BIC = n\ln(SSE) + \ln(n)(p+1) - n\ln(n)$$

3 Multivariate lineare Regression

3.1 Konzept

Daten:

$$(\mathbf{y}_i, \mathbf{x}_i)\ ,\ i = 1, \ldots, n \ \text{ mit}$$

$\mathbf{y}_i^\top = (y_{i1}, \ldots, y_{iq})$ q abhängige Variablen

$\mathbf{x}_i^\top = (1, x_{i1}, \ldots, x_{ip})$ p konstante und fixe Kovariablen (ohne Intercept)

Klassisches lineares Modell:

$$y_{ij} = \mathbf{x}_i^\top \boldsymbol{\beta}_{(j)} + \varepsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots q,$$

Matrixdarstellungen:

Vektorielle Form

$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{iq} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_i^\top & & \\ & \mathbf{x}_i^\top & \\ & & \ddots \\ & & \mathbf{x}_i^\top \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_{(1)} \\ \vdots \\ \boldsymbol{\beta}_{(q)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iq} \end{pmatrix}$$

bzw.

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$$

und als Gesamtmodell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

wobei

$$\boldsymbol{\beta}_{(j)}^\top = (\beta_{0j}, \beta_{1j}, \ldots \beta_{pj}), \quad \boldsymbol{\beta}^\top = (\boldsymbol{\beta}_{(1)}^\top, \ldots, \boldsymbol{\beta}_{(q)}^\top).$$

Multivariate Formulierung I

$$\begin{pmatrix} \mathbf{y}_1^\top \\ \vdots \\ \mathbf{y}_n^\top \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} (\boldsymbol{\beta}_{(1)}, \ldots, \boldsymbol{\beta}_{(q)}) + \begin{pmatrix} \boldsymbol{\epsilon}_1^\top \\ \vdots \\ \boldsymbol{\epsilon}_n^\top \end{pmatrix}$$

Multivariate Formulierung II

$$(\mathbf{y}_{(1)}, \ldots, \mathbf{y}_{(q)}) = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} (\boldsymbol{\beta}_{(1)}, \ldots, \boldsymbol{\beta}_{(q)}) + (\boldsymbol{\epsilon}_{(1)}, \ldots, \boldsymbol{\epsilon}_{(q)})$$

bzw. als Gesamtmodell

$$\mathbf{Y} = \mathbf{X}_0\mathbf{B} + \mathrm{E}$$

$$\text{mit } \mathbf{X}_0 = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix}, \boldsymbol{\beta}_{(j)} = \begin{pmatrix} \beta_{0j} \\ \vdots \\ \beta_{pj} \end{pmatrix}, \mathbf{y}_{(j)} = \begin{pmatrix} y_{1j} \\ \vdots \\ y_{nj} \end{pmatrix}, \boldsymbol{\epsilon}_{(j)} = \begin{pmatrix} \epsilon_{1j} \\ \vdots \\ \epsilon_{nj} \end{pmatrix}.$$

Annahmen:

(1) $\mathrm{E}(\boldsymbol{\epsilon}_i) = \mathbf{0}$

(2) $\mathrm{Cov}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$, $i = 1, \ldots, n$

$\mathrm{Cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = \mathbf{0}$, $i \neq j$

$\Rightarrow \mathrm{Cov}(\boldsymbol{\epsilon}_{(s)}, \boldsymbol{\epsilon}_{(t)}) = \sigma_{st}\,\mathbf{I},\ s \neq t$

(3) Normalverteilung: $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$

3.2 Schätzen

Kleinste–Quadrate–Schätzung

$$\sum_{j=1}^{q} (\mathbf{y}_{(j)} - \mathbf{X}_0\boldsymbol{\beta}_{(j)})^\top (\mathbf{y}_{(j)} - \mathbf{X}_0\boldsymbol{\beta}_{(j)}) \to \min$$

Lösung (voller Rang von $\mathbf{X}$):

$$\hat{\boldsymbol{\beta}}_{(j)} = (\mathbf{X}_0^\top\mathbf{X}_0)^{-1}\mathbf{X}_0^\top\mathbf{y}_{(j)} \quad \text{bzw.} \quad \hat{\mathbf{B}} = (\mathbf{X}_0^\top\mathbf{X}_0)^{-1}\mathbf{X}_0^\top\mathbf{Y}$$

Als Vektor

$$\hat{\boldsymbol{\beta}} = \mathrm{BlockDiag}\{(\mathbf{X}_0^\top\mathbf{X}_0)^{-1}\mathbf{X}_0^\top\}\mathbf{y}_0 \quad \text{oder} \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}.$$

mit

$$\mathbf{y}_0^\top = (\mathbf{y}_{(1)}^\top, \ldots, \mathbf{y}_{(q)}^\top),\ \mathbf{y}^\top = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top).$$

Schätzung der Kovarianz:

$$\hat{\mathbf{\Sigma}} = \frac{1}{n-p-1}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})^\top$$

$$\hat{\mathbf{\Sigma}} = \frac{1}{n-p-1}\mathrm{E}^\top\mathrm{E}$$

wobei

$$\mathrm{E} = \mathbf{Y} - \mathbf{X}_0\hat{\mathbf{B}} \text{ Residuenmatrix.}$$

Gauß–Markov–Theorem (mit Annahmen (1) und (2))

(1) $\mathrm{E}(\hat{\mathbf{B}}) = \mathbf{B}$, $\mathrm{E}(\hat{\mathbf{\Sigma}}) = \mathbf{\Sigma}$

(2) $\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\boldsymbol{\beta}}_{(j)}) = \sigma_{ij}(\mathbf{X}_0^\top\mathbf{X}_0)^{-1}$

(3) $\hat{\mathbf{B}}$ ist BLUE

Weiter gilt mit Normalverteilungsannahme

(4) $\hat{\mathbf{B}}$ ist normalverteilt

(5) $\mathbf{Q}_e = \hat{\mathrm{E}}^\top\hat{\mathrm{E}} \sim W_p(\mathbf{\Sigma}, n-p-1)$

(6) $\hat{\mathbf{B}}$ und $\hat{\mathbf{\Sigma}}$ sind unabhängig.

3.3 Testverfahren

Tests: (Unter Normalverteilungsannahme)
$H_0: \ \mathbf{CB} = \mathbf{\Gamma}$ $\quad$ $H_1: \ \mathbf{CB} \neq \mathbf{\Gamma}$ $\quad$ $\mathrm{rg}(\mathbf{C}) = s$

$$\Lambda = \frac{|\ \mathbf{Q}_e\ |}{|\ \mathbf{Q}_0\ |} \ \sim \ \Lambda(q, n-p-1, s)$$

mit $\mathbf{Q}_e = (\mathbf{Y} - \mathbf{X}_0\hat{\mathbf{B}})^\top(\mathbf{Y} - \mathbf{X}_0\hat{\mathbf{B}})$ ohne Restriktion
$\mathbf{Q}_0 = (\mathbf{Y} - \mathbf{X}_0\hat{\mathbf{B}}_0)^\top(\mathbf{Y} - \mathbf{X}_0\hat{\mathbf{B}}_0)$ mit Restriktion

H_0 ablehnen, wenn $\Lambda < \Lambda_\alpha(q, n-p-1, s)$

4 Binäre Regressionsmodelle

4.1 Grundkonzepte

Basis für das Logit-Modell ist die Binomialverteilung $Y \sim B(n, \pi)$

$$P(Y = y) = \binom{n}{y} \pi^y (1-\pi)^{n-y} \qquad y \in \{0, \ldots, n\}$$

mit

$$E(Y) = n\pi \ , \quad \text{Var}(Y) = n\pi(1-\pi)$$

Wichtig im Logit-Modell sind:

1. Chancen
$$\gamma(\pi) = \frac{\pi}{1-\pi}$$
2. Logarithmierte Chancen (Logits)
$$\text{Logit}(\pi) = \log\left(\gamma(\pi)\right) = \log\left(\frac{\pi}{1-\pi}\right)$$

Für das Logit–Modell gelten folgende Zusammenhänge:

$$\pi(\mathbf{x}) = \frac{\exp(\mathbf{x}^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}^\top \boldsymbol{\beta})}$$

bzw.

$$\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})} = \exp(\mathbf{x}^\top \boldsymbol{\beta})$$

bzw.

$$\log\left(\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\right) = \mathbf{x}^\top \boldsymbol{\beta}$$

Man interpretiert häufig die **Odds** (Chancen) bzw. die **Log-Odds** (Logarithmierte Chancen)

Odds	Log-Odds oder Logits
$\gamma(\pi) = \frac{\pi}{1-\pi}$	$\text{Logit}(\pi) = \log(\gamma(\pi)) = \log(\frac{\pi}{1-\pi})$

In diesem Kontext tritt meist auch der Begriff **Odds-Ratio** auf:
Gegeben sei eine (2x2)-Kontingenztafel mit

		Y		
		1	2	
G	1	n_{11}	n_{12}	n_1
	2	n_{21}	n_{22}	n_2

wobei die Randwerte durch $n_1 = n_{11} + n_{12}, n_2 = n_{21} + n_{22}$ bestimmt sind.
Der **Odds-Ratio** (Chancenverhältniss) ist definiert durch

$$\gamma(1|2) = \frac{\gamma(G=1)}{\gamma(G=2)} = \frac{\pi(1)/(1-\pi(1))}{\pi(2)/(1-\pi(2))}$$

mit $\pi(i) = P(Y=1|G=i),\ i=1,2.$

Für multikategoriale Einflussgrößen gilt:

		Y		
		1	2	
	1	π_{11}	π_{12}	π_1
	2	π_{21}	π_{22}	π_2
A	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	I	π_{I1}	π_{I2}	π_I

Der **Odds-Ratio** von $A=i$ gegenüber $A=j$ ist gegeben durch

$$\gamma(i|j) = \frac{\gamma(A=i)}{\gamma(A=j)} = \frac{\pi(i)/(1-\pi(i))}{\pi(j)/(1-\pi(j))}$$

wobei $\pi(i) = P(Y=1|A=i),\ i=1,\ldots,I.$

4.2 Modelltypen für binäre Regressionsmodelle

Generelle Form des binären Regressionsmodells:

$$\pi(\mathbf{x}) = F(\mathbf{x}^\top\boldsymbol{\beta}) = F(\eta) \quad \text{bzw.} \quad g(\pi(\mathbf{x})) = \mathbf{x}^\top\boldsymbol{\beta} = \eta, \quad \text{mit} \quad \eta = \mathbf{x}^\top\boldsymbol{\beta}$$

Logistisches Modell

$$F(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)} \ , \ g(\pi) = \log\left(\frac{\pi}{1-\pi}\right).$$

Probit–Modell

$$F(\eta) = \phi(\eta) = \int_0^\eta \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \ , \ g(\pi) = \phi^{-1}(\pi)$$

Lineares Modell

$$F(\eta) = \eta \ , \ g(\pi) = \pi$$

Minimum Extremwert–Modell (komplementres log–log Modell)

$$F(\eta) = 1 - \exp(-\exp(\eta)) \ , \ g(\pi) = \log(-\log(1-\pi))$$

Maximum Extremwert–Modell

$$F(\eta) = \exp(-\exp(\eta)) \ , \ g(\pi) = \log(-\log(\pi))$$

Exponentialmodell (komplementäres log–Modell)

$$F(\eta) = 1 - \exp(-\eta) \qquad g(\pi) = -\log(1-\pi)$$

4.3 ML-Schätzung im binären Regressionsmodell

Modell $\pi = F(\mathbf{x}_i^\top\boldsymbol{\beta}) = h(\mathbf{x}_i^\top\boldsymbol{\beta})$ bzw. $g(\pi) = \mathbf{x}_i^\top\boldsymbol{\beta}$

Beobachtungen $(p_i, \mathbf{x}_i)$, $i = 1, \ldots, g$, wobei p_i relative Häufigkeit bei $\mathbf{x}_i$ und g die Anzahl der Gruppen bei gruppierten Daten bezeichnet.

Maximum Likelihood–Schätzung

Log–Likelihoodfunktion

$$\begin{aligned} l(\boldsymbol{\beta}) &= \log L(\boldsymbol{\beta}) \\ &= \sum_{i=1}^{g} \left\{ y_i \log(\pi(\mathbf{x}_i)) + (n_i - y_i)\log(1-\pi(\mathbf{x}_i)) + \log\binom{n_i}{y_i} \right\} \\ &= \sum_{i=1}^{g} n_i\{p_i \log(h(\mathbf{x}_i^\top\boldsymbol{\beta})) + (1-p_i)\log(1-h(\mathbf{x}_i^\top\boldsymbol{\beta}))\} + \log\binom{n_i}{n_i p_i}. \end{aligned}$$

Score–Funktion

$$s(\boldsymbol{\beta})^\top = \left(\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_1}, \ldots, \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_p}\right)$$

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_j} &= \sum_{i=1}^{n} x_{ij}\left\{p_i \frac{h'(\mathbf{x}_i^\top\boldsymbol{\beta})}{h(\mathbf{x}_i^\top\boldsymbol{\beta})} + (1-p_i)\frac{-h'(\mathbf{x}_i^\top\boldsymbol{\beta})}{1-h(\mathbf{x}_i^\top\boldsymbol{\beta})}\right\}, \\
&= \sum_{i=1}^{g} n_i x_{ij} \frac{h'(\mathbf{x}_i^\top\boldsymbol{\beta})}{h(\mathbf{x}_i^\top\boldsymbol{\beta})(1-h(\mathbf{x}_i^\top\boldsymbol{\beta}))}(p_i - h(\mathbf{x}_i^\top\boldsymbol{\beta})), \\
s(\boldsymbol{\beta}) &= \sum_{i=1}^{g} \mathbf{x}_i \frac{n_i h'(\mathbf{x}_i^\top\boldsymbol{\beta})}{h(\mathbf{x}_i^\top\boldsymbol{\beta})(1-h(\mathbf{x}_i^\top\boldsymbol{\beta}))}(p_i - h(\mathbf{x}_i^\top\boldsymbol{\beta}))
\end{aligned}$$

Informationsmatrix

$$\mathbf{F}(\boldsymbol{\beta}) = \mathrm{E}\left(-\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right) = \sum_{i=1}^{g} \frac{h'(\mathbf{x}_i^\top\boldsymbol{\beta})^2}{\mathrm{Var}(p_i)}\mathbf{x}_i\mathbf{x}_i^\top$$

Für die Verteilung des Schätzers gilt: $(n \to \infty)$

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \mathbf{F}(\hat{\boldsymbol{\beta}})^{-1})$$

mit

$$\mathbf{F}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{g} n_i \mathbf{x}_i\mathbf{x}_i^\top \frac{h'(\mathbf{x}_i^\top\hat{\boldsymbol{\beta}})^2}{h(\mathbf{x}_i^\top\hat{\boldsymbol{\beta}})(1-h(\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}))}$$

4.4 Inferenz im binären Regressionsmodell

Ein Maß für die Anpassung ist die **Devianz**

$$\begin{aligned}
D(p_i, \hat{\pi}_i) &= -2\{l(\mathbf{y};\hat{\pi}) - l(\mathbf{y};p)\} \\
&= 2\sum_{i=1}^{g} n_i \left\{p_i \log\left(\tfrac{p_i}{\hat{\pi}_i}\right) + (1-p_i)\log\left(\tfrac{1-p_i}{1-\hat{\pi}_i}\right)\right\}
\end{aligned}$$

wobei p_i die relative Häufigkeit in der i-ten Gruppe ist.

Teststatistiken für lineare Hypothesen

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d} \qquad \mathbf{H_1} : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{d} \qquad \text{mit} \quad \mathbf{rg}(\mathbf{C}) = \mathbf{s}$$

sind:

1. **Likelihood-Ratio-Statistik**

$$\begin{aligned}
\lambda &= -2\{l(\mathbf{y}, \tilde{\boldsymbol{\beta}}) - l(\mathbf{y}, \hat{\boldsymbol{\beta}})\} \\
&= (D(\mathbf{y}, \tilde{\boldsymbol{\beta}}) - D(\mathbf{y}, \hat{\boldsymbol{\beta}}))
\end{aligned}$$

2. **Wald-Statistik**

$$w = (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^\top (\mathbf{C}\mathbf{F}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{C}^\top)^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})$$

3. **Score-Statistik**

$$u = s^\top(\tilde{\boldsymbol{\beta}})\mathbf{F}^{-1}(\tilde{\boldsymbol{\beta}})s(\tilde{\boldsymbol{\beta}})$$

mit restringiertem Schätzer $\tilde{\boldsymbol{\beta}}$ bzw. unrestringiertem Schätzer $\hat{\boldsymbol{\beta}}$

$$\tilde{\boldsymbol{\beta}} = \sup_{\beta:\mathbf{C}\boldsymbol{\beta}=\mathbf{d}} L(\boldsymbol{\beta}) \quad \text{bzw.} \quad \hat{\boldsymbol{\beta}} = \sup_{\boldsymbol{\beta}} L(\boldsymbol{\beta}).$$

Für die Statistiken gilt die asymptotische Verteilung

$$\lambda, w, u \sim \chi^2(s)$$

Anpassungsstatistiken (Goodness of fit) für gruppierte Daten:

1. **Pearson-Statistik**

$$\chi^2_P = \sum_{i=1}^{g} n_i \frac{(p_i - \hat{\pi}_i)^2}{\hat{\pi}_i(1-\hat{\pi}_i)}, \quad \hat{\pi}_i = F(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}})$$

2. **Devianz**

$$D = 2\sum_{i=1}^{g} n_i \left\{ p_i \log\left(\frac{p_i}{\hat{\pi}_i}\right) + (1-p_i)\log\left(\frac{1-p_i}{1-\hat{\pi}_i}\right)\right\}$$

3. **Wald-Statistik**

$$\chi^2_W = \sum_{i=1}^{g} n_i \frac{(g(p_i) - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}})^2}{\left(\frac{\partial g(p_i)}{\partial \pi}\right)^2 p_i(1-p_i)}$$

Für die Statistiken gilt die Verteilungsapproximation $(n_i/n \to \lambda_i)$

$$\chi^2, D, \chi^2_W \overset{(a)}{\sim} \chi^2(g - p_z)$$

5 Mehrkategoriale Regressionsmodelle

5.1 Multinomialverteilung

Hintergrund:
Urnenmodell mit k Typen von Kugeln. Von N Kugeln in der Urne sind N_i vom i-ten Typ $(i = 1, \ldots, k)$. Anteile der Typen an der Gesamtzahl ergeben sich durch $\pi_1 = N_1/N, \ldots, \pi_k = N_k/N$. Gezogen werden insgesamt n Kugeln, wobei das Modell mit Zurücklegen zugrundegelegt ist. Beobachtete Zufallsgröße ist $\mathbf{Y} = (Y_1, \ldots, Y_k)^\top$, wobei

$$Y_i = \text{Anzahl der Kugeln vom Typ } i \text{ bezeichnet.}$$

Multinomialverteilung:

$$\mathbf{Y} = (Y_1, \ldots, Y_k)^\top \sim M(n, \boldsymbol{\pi}), \text{ mit den Parametern}$$

n für die Anzahl gezogener Kugeln
$\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)^\top$ für die Anteile mit $0 \leq \pi_i \leq 1, \sum_{i=1}^k \pi_i = 1$.

Wahrscheinlichkeitsfunktion:

$$f(y_1, \ldots, y_k) = \begin{cases} \frac{n!}{y_1! \ldots y_k!} \pi_1^{y_1} \ldots \pi_k^{y_k} & y_i \in \{0, \ldots, n\}, \sum_i y_i = n \\ 0 & \text{sonst,} \end{cases}$$

Wegen den Restriktionen $\sum_{i=1}^k \pi_i = 1$ und $\sum_{i=1}^k x_i = n$ genügt es, nur $q = k - 1$ Komponenten zu betrachten. Der um eine Komponente reduzierte Vektor $(Y_1, \ldots, Y_q)^\top$ ist wiederum multnominalverteilt mit Wahrscheinlichkeitsvektor $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_q)^\top$; π_k ist implizit durch $\pi_k = 1 - \pi_1 - \ldots - \pi_q$ gegeben.

Binomialverteilung: ergibt sich als Spezialfall der Multinominalverteilung für $k = 2$ bzw. $q = 1$:

$$Y \sim B(n, \pi), \pi \in [0, 1]$$

Eigenschaften von multinomialverteilten Zufallsgrößen:

1. Die Momente sind bestimmt durch

   $\text{Cov}(\mathbf{Y}) = n(\text{diag}(\pi) - \pi\pi^\top)$, d.h.

   $$\text{E}(Y_i) = n\pi_i \qquad \text{Var}(Y_i) = n\pi_i(1-\pi_i) \qquad \text{Cov}(Y_i, Y_j) = -n\pi_i\pi_j, \;\; i \neq j$$

2. Sei $M_1, \ldots, M_r$ eine disjunkte Partition von $\{1, \ldots, k\}$ (also $M_1 \cup \ldots \cup M_r = \{1, \ldots, k\}, \;\; M_i \cap M_j = 0, i \neq j$).

   Wenn gilt $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top \sim M(n, \boldsymbol{\pi})$ dann gilt für die Summen $\tilde{Y}_i = \sum_{j \in M_i} Y_j, i = 1, \ldots, r$, wiederum eine Multinomialverteilung, genauer

   $$\tilde{\mathbf{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_r)^\top \sim M(n, \tilde{\boldsymbol{\pi}} = (\tilde{\pi}_1, \ldots, \tilde{\pi}_r)^\top),$$

   wobei $\tilde{\pi}_i = \sum\limits_{j \in M_i} \pi_j, i = 1, \ldots, r$.

3. Sei $\{i_1, \ldots, i_r\} \subset \{1, \ldots, k\}$ und $\mathbf{Y}^\top = (Y_1, \ldots, Y_k) \sim M(n, \boldsymbol{\pi})$

   dann gilt

   $$(Y_{i1}, \ldots, Y_{ir})^\top \text{ u.d.B. } Y_j = y_j, j \notin \{i_1, \ldots, i_r\}$$
   $$\sim M(\tilde{n}, \tilde{\boldsymbol{\pi}} = (\tilde{\pi}_{i_1}, \ldots, \tilde{\pi}_{i_r})^\top)$$

   mit

   $$\tilde{n} = n - \sum_{j \notin \{i_1, \ldots, i_r\}} Y_j, \;\; \tilde{\pi}_{i_r} = \frac{\pi_{i_r}}{\sum_{j=1}^r \pi_{i_j}}$$

Gelegentlich ist es sinnvoll, die reskalierte Multi– bzw. Binomialverteilung zu betrachten. Ist $\mathbf{Y} = (Y_1, \ldots, Y_k)^\top \sim M(n, \boldsymbol{\pi})$, bezeichnet man $\mathbf{Y}/n = (Y_1/n, \ldots, Y_k/n)^\top$ als **reskalierte Multinomialverteilung**, deren Komponenten Y_i/n jetzt statt der Werte $0, \ldots, n$ die Werte $0, 1/n, \ldots, 1$ annehmen.

5.2 Multinomiales Logit–Modell

Darstellung mit Referenzkategorie k

Verteilungsannahme:

$$\mathbf{y}_i|\mathbf{x}_i \quad \sim \quad M(1, \boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{i,k-1})^\top) \qquad \text{für } i = 1, \ldots, n$$

Strukturannahme:

$$\log\left(\frac{P(Y_i = r \mid \mathbf{x}_i)}{P(Y_i = k \mid \mathbf{x}_i)}\right) = \mathbf{x}_i^\top \boldsymbol{\beta}_r \qquad r = 1, \ldots, k-1,$$

bzw.

$$P(Y_i = r \mid \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta}_r)}{1 + \sum_{s=1}^{k-1} \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_s)} \qquad r = 1, \ldots, k-1,$$

$$P(Y_i = k \mid \mathbf{x}_i) = \frac{1}{1 + \sum_{s=1}^{k-1} \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_s)}$$

Alternative allgemeine Darstellung

$$P(Y_i = r \mid \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta}_r)}{\sum_{s=1}^{k} \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_s)}$$

Drei alternative Identifizierbarkeitsbedingungen:

$\boldsymbol{\beta}_k = (0, \ldots, 0)^\top$	Referenzkategorie k
$\boldsymbol{\beta}_{r_0} = (0, \ldots, 0)^\top$	Referenzkategorie r_0
$\sum_{s=1}^{k} \boldsymbol{\beta}_s = (0, \ldots, 0)^\top$	symmetrische Nebenbedingung

Chancen für $Y_i = r$ gegenüber $Y_i = r_0$ bei Referenzkategorie r_0

$$\gamma_{r,r_0}(\mathbf{x}_i) = \frac{P(Y_i = r \mid \mathbf{x}_i)}{P(Y_i = r_0 \mid \mathbf{x}_i)}$$

Logits bzw. logarithmierte Chancen für r gegenüber r_0

$$\text{Logit}_{r,r_0}(\mathbf{x}_i) = \log\left(\frac{P(Y_i = r \mid \mathbf{x}_i)}{P(Y_i = r_0 \mid \mathbf{x}_i)}\right)$$

6 Diskriminanzanalyse/Klassifikation

6.1 Relevante Größen:

Wichtige Größen für das Klassifikationsproblem sind

- die *a priori–Wahrscheinlichkeiten* $p(r) = P(Y = r), r = 1, \ldots, k,$
- die *a posteriori–Wahrscheinlichkeiten* $P(r \mid \mathbf{x}) = P(Y = r \mid \mathbf{x}), r = 1, \ldots k,$
- die *Verteilung der Merkmale*, gegeben die Klasse, bestimmt durch die Dichten $f(\mathbf{x} \mid 1), \ldots, f(\mathbf{x} \mid k)$.
- die *Mischverteilung* der Population (bzw. die unbedingte Verteilung von $\mathbf{x}$): $f(\mathbf{x}) = p(1)f(\mathbf{x} \mid 1) + \ldots + p(k)f(\mathbf{x} \mid k)$.

Satz von Bayes:

$$P(r \mid \mathbf{x}) = \frac{f(\mathbf{x} \mid r)p(r)}{f(\mathbf{x})} = \frac{f(\mathbf{x} \mid r)p(r)}{\sum_{i=1}^{k} f(\mathbf{x} \mid i)p(i)}$$

6.2 Bayes–Zuordnung:

$$\delta^*(\mathbf{x}) = r \Longleftrightarrow P(r \mid \mathbf{x}) = \max_{i=1,\ldots,k} P(i \mid \mathbf{x})$$

Fehlklassifikationswahrscheinlichkeiten:

- Wahrscheinlichkeit einer Fehlklassifikation, gegeben der feste Merkmalsvektor $\mathbf{x}$

$$\varepsilon(\mathbf{x}) \quad = \quad P(\delta(\mathbf{x}) \neq Y \mid \mathbf{x}) = 1 - P(\delta(\mathbf{x}) = Y \mid \mathbf{x}) = 1 - P(\delta(\mathbf{x}) \mid \mathbf{x}).$$

- Verwechslungswahrscheinlichkeit oder individuelle Fehlerrate

$$\varepsilon_{rs} = P(\delta(\mathbf{x}) = s \mid Y = r) = \int_{\mathbf{x}:\delta(\mathbf{x})=s} f(\mathbf{x} \mid r)d\mathbf{x}, \quad r \neq s$$

- Globale Fehlklassifikationswahrscheinlichkeit oder Gesamt–Fehlerrate

$$\varepsilon = P(\delta(\mathbf{x}) \neq Y)$$

- Wahrscheinlichkeit einer Fehlklassifikation, gegeben das Objekt entstammt der Klasse r

$$\varepsilon_r = P(\delta(\mathbf{x}) \neq r \mid Y = r) = \sum_{s \neq r} \varepsilon_{rs}$$

Für die globale Fehlklassifikationswahrscheinlichkeit gilt der Zusammenhang:

$$\begin{aligned}\varepsilon &= P(\delta(\mathbf{x}) \neq Y) = \sum_{r=1}^{k} P(\delta(\mathbf{x}) \neq r \mid Y = r)p(r) = \sum_{r=1}^{k} \varepsilon_r p(r) \\ &= \sum_{r=1}^{k} \sum_{s \neq r} \varepsilon_{rs} p(r) \\ \varepsilon &= P(\delta(\mathbf{x}) \neq Y) = \int P(\delta(\mathbf{x}) \neq Y \mid \mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \int \varepsilon(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}\end{aligned}$$

Optimalität der Bayes–Zuordnung

Die Bayes–Zuordnung

$$\delta^*(\mathbf{x}) = r \Longleftrightarrow P(r \mid \mathbf{x}) = \max_{i=1,\ldots,k} P(i \mid \mathbf{x})$$

minimiert die Gesamtfehlerrate ε.

Äquivalente Diskriminanzfunktionen für die Bayes-Zuordnung:

a) $d_r(\mathbf{x}) = P(r \mid \mathbf{x})$,

b) $d_r(\mathbf{x}) = f(\mathbf{x} \mid r)p(r)/f(\mathbf{x})$.

c) $d_r(\mathbf{x}) = f(\mathbf{x} \mid r)p(r)$.

d) $d_r(\mathbf{x}) = \log(f(\mathbf{x} \mid r)) + \log(p(r))$.

Maximum–Likelihood–(ML)–Zuordnungsregel

$$\delta_{ML}(\mathbf{x}) = r \Longleftrightarrow f(\mathbf{x} \mid r) = \max_{i=1,\ldots,k} f(\mathbf{x} \mid i)$$

6.3 Kostenoptimale Bayes–Zuordnung:

$$c(i,j) = c_{ij} \quad = \quad \text{Kosten einer Zuordnung eines Objekts aus Klasse } i \text{ in die Klasse } j.$$

Für $I = j$ gelte $c_{ii} = 0$. d.h. die Kosten einer richtigen Zuordnung sind Null, während $c_{ij} \geq 0$ für $i \neq j$.

Bedingtes Risiko, gegeben $\mathbf{x}$

$$r(\mathbf{x}) = \sum_{i=1}^{k} c_{i,\delta(\mathbf{x})} P(i \mid \mathbf{x}).$$

Individuelles Risiko

$$r_{ij} = c_{ij} P(\delta(\mathbf{x}) = j \mid Y = i) = c_{ij} \int_{\mathbf{x}:\delta(\mathbf{x})=j} f(\mathbf{x} \mid i) d\mathbf{x}$$

Risiko gegeben Klasse i

$$r_i = \sum_{j=1}^{k} r_{ij}.$$

Das gesamte Bayes Risiko, d.h. die zu erwartenden Kosten, lassen sich darstellen durch

$$R = E(c(Y, \delta(\mathbf{x}))) = \sum_{i=1}^{k} r_i p(i) = \int r(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}.$$

Bayes–Zuordnung mit Kosten

Bei Vorliegen des Beobachtungsvektors $\mathbf{x}$ wird das Bayes–Risiko minimiert durch die Zuordnung

$$\delta^*(\mathbf{x}) = r \Longleftrightarrow \sum_{i=1}^{k} P(i \mid \mathbf{x}) c_{ir} = \min_{j=1,\ldots,k} \sum_{i=1}^{k} P(i \mid \mathbf{x}) c_{ij}$$

Mit der Diskrimininanzfunktion

$$d_r(\mathbf{x}) = -\sum_{i=1}^{k} P(i \mid \mathbf{x}) c_{ir}$$

erhält man

$$\delta^*(\mathbf{x}) = r \Longleftrightarrow d_r(\mathbf{x}) = \max_{i=1,\ldots,k} d_i(\mathbf{x})$$

6.4 Zuordnungsregel unter der Voraussetzung Normalverteilung

Angenommen werden normalverteilte Merkmale in den Klassen:

$$\mathbf{X}|Y = i \quad \sim \quad N(\boldsymbol{\mu}_i \,,\, \boldsymbol{\Sigma}_i) \quad \text{für} \quad i = 1, \ldots, k$$

$$\text{d.h.} \quad f(\mathbf{x} \mid i) = \frac{1}{(2\pi)^{p/2} (\det \boldsymbol{\Sigma}_i)^{1/2}} \cdot \exp\{-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)\}$$

Diskriminanzfunktionen für die Bayes–Zuordnung:

$$\begin{aligned} d_i(\mathbf{x}) &= \ln(f(\mathbf{x} \mid i)) + \ln(p(i)) \\ d_i(\mathbf{x}) &= -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \ln(\det \boldsymbol{\Sigma}_i) + \ln(p(i)), \quad i = 1, \ldots, k. \end{aligned}$$

Spezielle (reduzierte) Diskriminanzfunktionen

- bei der Annahme $\mathbf{\Sigma}_i = \sigma^2\mathbf{I}$:

$$d_i(\mathbf{x}) = -\frac{\| \mathbf{x} - \boldsymbol{\mu}_i \|^2}{2\sigma^2} + \ln(p(i))$$

- bei der Annahme $\mathbf{\Sigma}_i = \sigma^2\mathbf{I}$ und gleich große a priori Wahrscheinlichkeiten in den Klassen:

$$d_i(\mathbf{x}) = - \| \mathbf{x} - \boldsymbol{\mu}_i \|^2$$

- bei der Annahme klassenweise identischer Kovarianzmatrizen $\mathbf{\Sigma}_i = \mathbf{\Sigma}, \;\; i = 1, \ldots, k$:

$$\begin{aligned} d_i(\mathbf{x}) &= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^\top \mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) + \ln(p(i)) \\ d_i(\mathbf{x}) &= \boldsymbol{\mu}_i^\top \mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^\top \mathbf{\Sigma}^{-1}\boldsymbol{\mu}_i + \ln(p(i)) \end{aligned}$$

- bei der Annahme $k = 2$ und klassenweise identischer Kovarianzmatrizen $\mathbf{\Sigma}_i = \mathbf{\Sigma}, \; i = 1, 2$:

$$d(\mathbf{x}) = d_1(\mathbf{x}) - d_2(\mathbf{x}) = \left[\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)\right]^\top \mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \ln\left(\frac{p(2)}{p(1)}\right)$$

Für die ML–Regel sind die ln–Terme wegzulassen.
Schätzer für $\boldsymbol{\mu}_k$ und $\mathbf{\Sigma}$: $\hat{\boldsymbol{\mu}}_i = \bar{\mathbf{x}}_i$,

$$\hat{\mathbf{\Sigma}} = \mathbf{S} = \frac{1}{N - g}\sum_{i=1}^{k}\sum_{n=1}^{n_i}(\mathbf{x}_{in} - \bar{\mathbf{x}}_i)(\mathbf{x}_{in} - \bar{\mathbf{x}}_i)^\top$$

6.5 Fishersche Diskriminanzanalyse

Daten: $\mathbf{x}_{i1} \ldots \mathbf{x}_{in_i}$ in der Klasse i

Grundprinzip (Zwei–Klassen–Fall):
Projektion $y = \mathbf{a}^\top \mathbf{x}$ mit $\| \mathbf{a} \| = 1$, sodass das folgende Kriterium maximiert wird

$$Q(\mathbf{a}) = \frac{(\bar{y}_1 - \bar{y}_2)^2}{s_1^2 + s_2^2} ,$$

wobei

$$\begin{aligned} \bar{y}_i &= \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{a}^\top \mathbf{x}_{ij} = \mathbf{a}^\top \bar{\mathbf{x}}_i \text{ Klassenmittelpunkt,} \\ s_i^2 &= \sum (\mathbf{a}^\top x_{ij} - \mathbf{a}\bar{x}_i)^2 \text{ Quadratische Abweichungen } i\text{–te Klasse.} \end{aligned}$$

Lösung:

$$\mathbf{a} = \mathbf{W}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

mit

$$\mathbf{W} = (n_1 + n_2 - 2)\mathbf{S} = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^\top .$$

Mehr–Klassen–Fall: Kriterium:

$$Q(\mathbf{a}) = \frac{\sum_{i=1}^{g} n_i (\bar{y}_i - \bar{y})^2}{\sum_{i=1}^{g} s_i^2} \rightarrow \max$$

Resultierende Eigenvektoren:

$$\mathbf{a}_1, \ldots, \mathbf{a}_q \ , \ q = rg(\mathbf{B})$$

Resultierende Eigenwerte:

$$\lambda_1 = \frac{\mathbf{a}_1^\top \mathbf{B} \mathbf{a}_1}{\mathbf{a}_1^\top \mathbf{W} \mathbf{a}_1} \ , \ldots, \lambda_q = \frac{\mathbf{a}_q^\top \mathbf{B} \mathbf{a}_q}{\mathbf{a}_q^\top \mathbf{W} \mathbf{a}_q}$$

mit

$$\begin{aligned} \mathbf{B} &= \sum_{i=1}^{g} n_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^\top \\ \mathbf{W} &= \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^\top . \end{aligned}$$

7 Clusteranalyse

Objektmenge $\Omega = \{a_1, \ldots, a_n\}$
Zugehörige Merkmalsvektoren $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$
Gesucht: Partition = disjunkte vollständige Zerlegung $C_1, \ldots, C_g$, sodass

$$\bigcup_{i=1}^{g} C_i = \{a_1, \ldots, a_n\}, C_i \cap C_j = \emptyset$$

7.1 Distanzmaße für Objekte:

$$d: \ \Omega \times \Omega \to \mathbb{R}_+$$

mit

$$\begin{aligned} d(a_i, a_i) &= 0 \,, \\ d(a_i, a_j) &\geq 0 \,, \\ d(a_i, a_j) &= d(a_j, a_i) \,. \end{aligned}$$

Für metrische Distanzmaße: zusätzlich Dreiecksungleichung

$$d(a_i, a_j) \ \leq \ d(a_i, a_m) \ + \ d(a_m, a_j) \ \text{ für alle } \ a_i, a_j, a_m$$

7.2 Distanzmaße:

Binäre Merkmale

$$\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{ip}) \ \text{ mit } \ x_{ik} \in \{0, 1\}$$

$h_{ij}(y, z)$ bezeichnet die Anzahl der übereinstimmenden Komponenten in $\mathbf{x}_i$ und $\mathbf{x}_j$ mit Ausprägung y in $\mathbf{x}_i$ und z in $\mathbf{x}_j$.

$h_{ij}(y, z)$ = $\mid \{k \mid (x_{ik}, x_{jk}) = (y, z)\} \mid$ für $(y, z) \in \{(0, 0) \,, \ (0, 1) \,, \ (1, 0) \,, \ (1, 1)\}$.

Matching–Koeffizient

$$s_{ij} = \frac{h_{ij}(1, 1) + h_{ij}(0, 0)}{p}$$

Distanz Matching–Koeffizient (Anzahl nicht übereinstimmender Merkmale)

$$d_{ij} = 1 - s_{ij}$$

Metrische Merkmale

$$\mathbf{x}_i = (x_{i1}, \ldots, x_{ip}) \ , \ \mathbf{y}_i = (y_{i1}, \ldots, y_{ip})$$

L_q– Metrik

$$d(\mathbf{x}_i, \mathbf{y}_i) = \left(\sum_{k=1}^{p} \mid x_{ik} - y_{ik} \mid^q \right)^{1/q}$$

$q = 1$ City–Block–Metrik

$q = 2$ Euklidische Metrik

Mahalanobis–Distanz

$$d(\mathbf{x}_i, \mathbf{y}_i) = (\mathbf{x}_i - \mathbf{y}_i)^\top \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{y}_i)$$

mit empirischer Kovarianzmatrix $\mathbf{S}$

7.3 Distanzmaße für Klassen:

Form: $D(C_r, C_s)$ wobei $C_r, C_s \subset \Omega$

1. Single Linkage–Verfahren
$$D(C_r, C_s) = \min_{\substack{a_i \in C_r \\ a_j \in C_s}} d(a_i, a_j)$$

2. Complete Linkage–Verfahren
$$D(C_r, C_s) = \max_{\substack{a_i \in C_r \\ a_j \in C_s}} d(a_i, a_j)$$

3. Average Linkage–Verfahren
$$\begin{aligned} D(C_r, C_s) &= \frac{1}{n_r n_s} \sum_{a_i \in C_r} \sum_{a_j \in C_s} d(a_i, a_j) \\ \text{mit } n_i &= \mid C_i \mid \end{aligned}$$

4. Zentroid–Verfahren
$$D(C_r, C_s) = \parallel \bar{\mathbf{x}}_r - \bar{\mathbf{x}}_s \parallel^2 \qquad \text{mit } \bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{\mathbf{x}_j \in C_i} \mathbf{x}_j$$

7.4 Einige Grundbegriffe:

Mittelwert in Klasse C_k

$$\bar{\mathbf{x}}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{x}_i \text{ mit } n_k = \mid C_k \mid$$

Streuung in Klasse C_k

$$\mathbf{W}(C_k) = \sum_{\mathbf{x}_i \in C_k} (\mathbf{x}_i - \bar{\mathbf{x}}_k)(\mathbf{x}_i - \bar{\mathbf{x}}_k)^\top$$

Streuung der Partition $\mathbb{C}$ innerhalb der Klassen

$$\mathbf{W}(\mathbb{C}) = \sum_{k=1}^{g} \mathbf{W}(C_k)$$

Streuung der Partition $\mathbb{C}$ zwischen den Klassen

$$\mathbf{B}(\mathbb{C}) = \sum_{k=1}^{g} n_k(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^\top \text{ mit } \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$$

Die Gesamtstreuung

$$\mathbf{T} = \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$$

ist zerlegbar in

$$\mathbf{T} = \mathbf{W}(\mathbb{C}) + \mathbf{B}(\mathbb{C})$$

7.5 Gütekriterien: quantitative Merkmale

Partition $\mathbb{C} = \{C_1, \ldots, C_g\}$

1. *Varianzkriterium*

$$H(\mathbb{C}) = \sum_{k=1}^{g} \sum_{\mathbf{x}_i \in C_k} \| \mathbf{x}_i - \bar{\mathbf{x}}_k \|^2$$

Minimum–Distanz–Eigenschaft

$$\| \mathbf{x}_i - \bar{\mathbf{x}}_k \| \leq \| \mathbf{x}_i - \bar{\mathbf{x}}_j \| \quad \text{für } \mathbf{x}_i \in C_k$$

2. *Determinantenkriterium*

$$H(\mathbb{C}) = |\mathbf{W}(\mathbb{C})|$$

Minimum–Distanz–Eigenschaft

$$(\mathbf{x}_i - \bar{\mathbf{x}}_k)^\top \mathbf{W}(\mathbb{C})^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}_k) \leq (\mathbf{x}_i - \bar{\mathbf{x}}_j)^\top \mathbf{W}(\mathbb{C})^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}_j) \text{ für } \mathbf{x}_i \in C_k$$

3. *Verallgemeinertes Determinantenkriterium*

$$H(\mathbb{C}) = \sum_{k=1}^{g} n_k \ln \left(|\frac{1}{n_k} \mathbf{W}(C_k)| \right)$$

8 Verweildauer-Modelle

8.1 Grundlegende Definitionen für stetige Zeit $T \geq t$

Hazardrate, Intensitäts-, Risikofunktion

Die Hazardrate ist die infinitesimale Rate für einen Ausfall zum Zeitpunkt t, gegeben das Überleben bis zum Zeitpunkt t.

$$\lambda(t) \;=\; \lim_{\Delta t \to 0, \Delta t > 0} \frac{1}{\Delta t} P(t \leq T \leq t + \Delta t | T \geq t)$$

Kumulative Hazardrate

Das Integral wird als kumulative Hazardrate bezeichnet

$$\Lambda(t) \;=\; \int_0^t \lambda(u) \mathrm{d}u$$

Survivorfunktion

Die Survivorfunktion ist die unbedingte Wahrscheinlichkeit, bis zum Zeitpunkt t zu überleben.

$$S(t) \;=\; 1 - P(T < t) = 1 - F(t)$$

Zusammenhänge

$$\begin{aligned} \lambda(t) &= \frac{f(t)}{S(t)} \\ S(t) &= \exp\left(- \int_0^t \lambda(u) \mathrm{d}u \right) \end{aligned}$$

Cox-Modell

$$\lambda(t|\mathbf{x}) \;=\; \lambda_0(t) \exp\left(\mathbf{x}^\top \boldsymbol{\beta}\right)$$

8.2 Grundbegriffe bei diskreter Zeit T

$T \in \{1, 2, \ldots\}$ aus einer Unterteilung in Intervalle $[a_{j-1}, \ldots, a_j), j = 1, \ldots, k$.

Diskreter Hazard

$$\lambda(t) \;=\; P(T = t | T \geq t)$$

Wahrscheinlichkeit

$$P(T = t) \;=\; \lambda(t) \prod_{i=1}^{t-1} (1 - \lambda(t|\mathbf{x}))$$

Modelle

$$\lambda(t|\mathbf{x}) \;=\; F(\mathbf{x}^\top \boldsymbol{\beta})$$

wobei F eine streng monoton wachsende, differenzierbare Verteilungsfunktion ist, z.B. die logistische: $F(\eta) = \exp(\eta)/(1 + \exp(\eta))$

9 Verteilungstabellen

9.1 Standardnormalverteilung

Tabelliert sind die Werte der Verteilungsfunktion $\Phi(z) = P(Z \leq z)$ für $z \geq 0$.
Ablesebeispiel: $\Phi(1.75) = 0.9599$
Funktionswerte für negative Argumente: $\Phi(-z) = 1 - \Phi(z)$
Die z-Quantile ergeben sich genau umgekehrt.
Beispielsweise ist $z(0.9599) = 1.75$ und $z(0.9750) = 1.96$.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

9.2 Students t-Verteilung

Tabelliert sind die Quantile für n Freiheitsgrade.
Für das Quantil $t_{1-\alpha}(n)$ gilt $F(t_{1-\alpha}(n)) = 1 - \alpha$.
Links vom Quantil $t_{1-\alpha}(n)$ liegt die Wahrscheinlichkeitsmasse $1 - \alpha$.
Ablesebeispiel: $t_{0.99}(20) = 2.528$
Die Quantile für $0 < 1 - \alpha < 0.5$ erhält man aus $t_\alpha(n) = -t_{1-\alpha}(n)$

Approximation für $n > 30$:

$$t_\alpha(n) \approx z_\alpha \quad (z_\alpha \text{ ist das } (\alpha)\text{-Quantil der Standardnormalverteilung})$$

n	0.6	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.3249	1.3764	3.0777	6.3138	12.706	31.821	63.657	318.31	636.62
2	0.2887	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248	22.327	31.599
3	0.2767	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409	10.215	12.924
4	0.2707	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.2672	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.2648	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.2632	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	0.2619	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	0.2610	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.2602	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	0.2596	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
12	0.2590	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	0.2586	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	0.2582	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	0.2579	0.8662	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728
16	0.2576	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150
17	0.2573	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	0.2571	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216
19	0.2569	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794	3.8834
20	0.2567	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8495
21	0.2566	0.8591	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272	3.8193
22	0.2564	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7921
23	0.2563	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676
24	0.2562	0.8569	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454
25	0.2561	0.8562	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	0.2560	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7066
27	0.2559	0.8551	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6896
28	0.2558	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	0.2557	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594
30	0.2556	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
∞	0.2533	0.8416	1.2816	1.6449	1.9600	2.3263	2.5758	3.0903	3.2906

9.3 χ^2-Verteilung

Tabelliert sind die Quantile für n Freiheitsgrade.
Für das Quantil $\chi^2_{1-\alpha}(n)$ gilt $F(\chi^2_{1-\alpha}(n)) = 1 - \alpha$.
Links vom Quantil $\chi^2_{1-\alpha}(n)$ liegt die Wahrscheinlichkeitsmasse $1 - \alpha$.
Ablesebeispiel: $\chi^2_{0.95}(10) = 18.307$

Approximation für $n > 30$:

$$\chi^2_\alpha(n) \approx \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2 \quad (z_\alpha \text{ ist das } \alpha\text{-Quantil der Standardnormalverteilung})$$

n	0.01	0.025	0.05	0.1	0.5	0.9	0.95	0.975	0.99
1	0.0002	0.0010	0.0039	0.0158	0.4549	2.7055	3.8415	5.0239	6.6349
2	0.0201	0.0506	0.1026	0.2107	1.3863	4.6052	5.9915	7.3778	9.2103
3	0.1148	0.2158	0.3518	0.5844	2.3660	6.2514	7.8147	9.3484	11.345
4	0.2971	0.4844	0.7107	1.0636	3.3567	7.7794	9.4877	11.143	13.277
5	0.5543	0.8312	1.1455	1.6103	4.3515	9.2364	11.070	12.833	15.086
6	0.8721	1.2373	1.6354	2.2041	5.3481	10.645	12.592	14.449	16.812
7	1.2390	1.6899	2.1674	2.8331	6.3458	12.017	14.067	16.013	18.475
8	1.6465	2.1797	2.7326	3.4895	7.3441	13.362	15.507	17.535	20.090
9	2.0879	2.7004	3.3251	4.1682	8.3428	14.684	16.919	19.023	21.666
10	2.5582	3.2470	3.9403	4.8652	9.3418	15.987	18.307	20.483	23.209
11	3.0535	3.8157	4.5748	5.5778	10.341	17.275	19.675	21.920	24.725
12	3.5706	4.4038	5.2260	6.3038	11.340	18.549	21.026	23.337	26.217
13	4.1069	5.0088	5.8919	7.0415	12.340	19.812	22.362	24.736	27.688
14	4.6604	5.6287	6.5706	7.7895	13.339	21.064	23.685	26.119	29.141
15	5.2293	6.2621	7.2609	8.5468	14.339	22.307	24.996	27.488	30.578
16	5.8122	6.9077	7.9616	9.3122	15.338	23.542	26.296	28.845	32.000
17	6.4078	7.5642	8.6718	10.085	16.338	24.769	27.587	30.191	33.409
18	7.0149	8.2307	9.3905	10.865	17.338	25.989	28.869	31.526	34.805
19	7.6327	8.9065	10.117	11.651	18.338	27.204	30.144	32.852	36.191
20	8.2604	9.5908	10.851	12.443	19.337	28.412	31.410	34.170	37.566
21	8.8972	10.283	11.591	13.240	20.337	29.615	32.671	35.479	38.932
22	9.5425	10.982	12.338	14.041	21.337	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	22.337	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	23.337	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	24.337	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	25.336	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	26.336	36.741	40.113	43.195	46.963
28	13.565	15.308	16.928	18.939	27.336	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	28.336	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	29.336	40.256	43.773	46.979	50.892