Formelsammlung

Erwin Grüner
FB Psychologie

Beschreibung mehrdimensionaler Datensätze

Gegeben: Datenmatrix $\mathbf{X}_{n\times p}$

Mittelwertsvektor
$\bar{\mathbf{x}}_{p\times 1} = \mathbf{X}'_{p\times n}\mathbf{1}_{n\times 1}/n$

Abweichungsmatrix
$\mathbf{A}_{n\times p} = \mathbf{X}_{n\times p} - \mathbf{1}_{n\times 1}\bar{\mathbf{x}}'_{1\times p}$

SSCP-Matrix der Abweichungen
$\mathbf{A}'_{p\times n}\mathbf{A}_{n\times p}$

Kovarianzmatrix
$\mathbf{S}_{p\times p} = \frac{1}{n-1}\mathbf{A}'_{p\times n}\mathbf{A}_{n\times p}$

Korrelationsmatrix
$\mathbf{R}_{p\times p} = \mathbf{S}_{p\times p}/\mathbf{s}_{p\times 1}\mathbf{s}'_{1\times p}$

Lineare Regression

Modell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (\boldsymbol{\beta} = [\beta_0\ \beta_1\ \dots\ \beta_p]' \quad \text{und} \quad \mathbf{x}_0 = \mathbf{1}_n)$$

Zielkriterium

$$RSS = \mathbf{e}'\mathbf{e} = \sum(y_i - \hat{y}_i)^2 \to Min$$

Schätzung der Regressionsparameter
$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
$[b_1\ \dots\ b_p]' = \mathbf{S}_{XX}^{-1}\mathbf{s}_{Xy} \qquad b_0 = \bar{y} - \bar{\mathbf{x}}'\mathbf{b}$

Standardisierte Gewichte
$\mathbf{b}^* = \mathbf{R}_{XX}^{-1}\mathbf{r}_{Xy} \qquad b_i^* = b_i\frac{s_i}{s_y}$

Hat-Matrix
$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

Gefittete Werte
$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} = \mathbf{H}\mathbf{y}$

Standardschätzfehler
$s = \sqrt{\frac{RSS}{n-p-1}}$

Standardfehler der Parameterschätzungen
$SE(\mathbf{b}) = s \cdot \operatorname{diag}(\sqrt{(\mathbf{X}'\mathbf{X})^{-1}})$

Multipler Korrelationskoeffizient
$R^2 = \mathbf{b}^{*\prime}\mathbf{r}_y = \mathbf{r}_y'\mathbf{R}^{-1}\mathbf{r}_y$

Modelltest
$F = \frac{\sum(\hat{y}_i - \bar{y})^2/p}{\sum(y_i - \hat{y}_i)^2/(n-p-1)} = \frac{R^2/p}{(1-R^2)/(n-p-1)}$

Punktweises Konfidenzintervall an der Stelle $\mathbf{x}$
$\mathbf{x}'\boldsymbol{b} \pm t_{n-p-1;1-\alpha/2} \cdot s \cdot \sqrt{\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}$

Konfidenzbereich für die Regressionsgerade
$f(\mathbf{x}) = \mathbf{x}'\boldsymbol{b} \pm \sqrt{pF_{p,n-p-1;1-\alpha}} \cdot s \cdot \sqrt{\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}$

Toleranz- bzw. Vorhersageintervall an der Stelle $\mathbf{x}$

$$\mathbf{x}'\boldsymbol{b} \pm t_{n-p-1;1-\alpha/2} \cdot s \cdot \sqrt{1 + \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}$$

Test für den Vergleich zweier hierarchischer Modelle

$$F = \frac{(R_{u+v}^2 - R_u^2)/v}{(1-R_{u+v}^2)/(n-u-v-1)} \qquad df_1 = v \qquad df_2 = n-u-v-1$$

Mallow's C_p-Statistik
$C_p = (n-p-1)\left(\frac{MSE_p}{MSE_{full}} - 1\right) + (p+1)$

Akaike's Informationskriterium
$AIC = n\ \ln\left(\frac{RSS_p}{n}\right) + 2(p+1)$

Einflussreiche Datenpunkte

$e^{(-i)} = y_i - \hat{y}_i^{(-i)}$ (Deleted Residuals)

$DFFITS_i = \frac{\hat{y}_i - \hat{y}_i^{(-i)}}{(s^{(-i)})^2 h_{ii}}$

$D_i = \frac{(\hat{\mathbf{y}}-\hat{\mathbf{y}}^{(-i)})'(\hat{\mathbf{y}}-\hat{\mathbf{y}}^{(-i)})}{p \cdot s^2}$ (Cook's Abstand)

$DFBETAS_{ki} = \frac{\hat{\beta}_k - \hat{\beta}_k^{(-i)}}{s^{(-i)}\sqrt{c_{kk}}}$

Multivariate multiple Regression

Zielkriterium

$\|\mathbf{Y} - \mathbf{X}\mathcal{B}\|^2 \rightarrow Min$

Schätzung der Regressionsparameter

$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

Modelltest

$\Lambda = \frac{|\mathbf{S}|}{|\mathbf{S}_{XX}||\mathbf{S}_{YY}|}$

Bartlett's Approximation

$V = -[(n-1) - (p+q)/2]\ln\Lambda$ mit $df = q(p-1)$

MANOVA

Hotelling's T^2 für zwei unabhängige Stichproben

$$T^2 = \frac{n_1 n_2}{n_1+n_2}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)'\mathbf{S}^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2) = \frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)'\mathbf{W}^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$$

$$F = \frac{(n_1+n_2-2)-(q-1)}{q(n_1+n_2-2)}T^2 = \frac{n_1+n_2-q-1}{q}\frac{T^2}{n_1+n_2-2} \qquad df_1 = q \qquad df_2 = n_1+n_2-q-1$$

Mahalanobis-Distanz

$$D^2(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_1 - \mathbf{x}_2)'\mathbf{S}^{-1}(\mathbf{x}_1 - \mathbf{x}_2)$$

Wilks' Lambda

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \frac{|\mathbf{W}|}{|\mathbf{B}+\mathbf{W}|} = \frac{1}{|\mathbf{B}\mathbf{W}^{-1}+\mathbf{I}|} = \prod \frac{1}{1+\lambda_i} \qquad (\lambda_i \text{ Eigenwerte von } \mathbf{B}\mathbf{W}^{-1})$$

Bartlett's Chiquadrat

$$\chi^2 = -[(n-1) - (q+m)/2]\ln\Lambda \qquad df = q(m-1)$$

Roy's Kriterium des größten Eigenwerts

$$\frac{\lambda_{max}}{1+\lambda_{max}}$$

Hotelling-Lawley Spur

Summe der Eigenwerte (Spur) von $\mathbf{B}\mathbf{W}^{-1}$

Pillai-Bartlett Spur

Summe der Eigenwerte (Spur) von $\mathbf{B}\mathbf{T}^{-1}$

Test eines linearen Kontrasts

$$\hat{\boldsymbol{\psi}} = \sum c_j \bar{\mathbf{y}}_j \qquad T^2 = \frac{1}{\sum c_j^2/n_j}\hat{\boldsymbol{\psi}}'\mathbf{S}^{-1}\hat{\boldsymbol{\psi}} \qquad F = \frac{n-m-q+1}{(n-m)q}T^2 \qquad df_1 = q \quad df_2 = n-m-q+1$$

Box' M

$$M = (n-q)ln|\mathbf{S}| - \sum(n_j-1)ln|\mathbf{S}_j|$$

Diskriminanzanalyse

Zielkriterium

$\lambda = \frac{\mathbf{v}'\mathbf{B}\mathbf{v}}{\mathbf{v}'\mathbf{W}\mathbf{v}} \rightarrow Max$

Lösungsansatz

$(\mathbf{W}^{-1}\mathbf{B})\mathbf{V} = \mathbf{V}\,\mathrm{diag}(\boldsymbol{\lambda})$ bzw. $(\mathbf{W}^{-1/2}\mathbf{B}\mathbf{W}^{-1/2})\mathbf{Q} = \mathbf{Q}\,\mathrm{diag}(\boldsymbol{\lambda})$ mit $\mathbf{V} = \mathbf{W}^{-1/2}\mathbf{Q}$

Test der Diskriminanzfunktionen

$\Lambda = \prod_{l=j}^{r} \frac{1}{1+\lambda_l} \quad (j = 1, \ldots, r)$ Bartlett: $\chi^2 = [n-1-(q+m)/2] \ln \Lambda$ $df = q(m-1)$

Diskriminanzscores

$\mathbf{F} = \mathbf{X}\mathbf{V}$

Gruppenzentroid

$\bar{\mathbf{f}}^{(g)} = (\bar{\mathbf{x}}^{(g)\prime}\mathbf{V})'$

PCA und FA

Grundgleichung der Faktorenanalyse

$\mathbf{X} = \mathbf{F}\mathcal{A}' + \mathcal{E}$ $X_{ik} = \alpha_{k1}F_{i1} + \alpha_{k2}F_{i2} + \cdots + \alpha_{kr}F_{ir} + \varepsilon_{ik}$

Diagonalisierung von **S** *(Kovarianz- bzw. Korrelationsmatrix)*

$\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}'$ (**V**: normierte Eigenvektoren; **Λ**: Diagonalmatrix der Eigenwerte)

Kovarianzmatrix der Faktoren

$\boldsymbol{\Lambda} = \mathbf{V}'\mathbf{S}\mathbf{V} = \mathbf{A}'\mathbf{A}$

Fundamentaltheorem

$\mathcal{R} = \mathcal{A}\mathcal{A}' + \boldsymbol{\Psi}$ $\mathbf{R}_h = \mathbf{A}\mathbf{A}'$ ($\mathbf{R}_h$: reduzierte Korrelationsmatrix)

Varianz der Variable k

$\sigma_k^2 = h_k^2 + u_k^2 = h_k^2 + b_k^2 + \varepsilon_k^2 = 1$ (h_k^2: Kommunalität; u_k^2: „Uniqueness“)

Faktorladungen

$\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}^{1/2}$ (**V**: normierte Eigenvektoren; **Λ**: Diagonalmatrix der Eigenwerte)

Orthogonale Rotation

$\mathbf{B} = \mathbf{A}\mathbf{T}$ (**T**: orthonormale Transformationsmatrix)

Transformationsmatrix für eine Rotation (zweidimensionaler Fall)

$\mathbf{T} = \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix}$ (φ: Drehwinkel gegen den Uhrzeigersinn)

Transformationsmatrix für eine Spiegelung an der ersten Achse (zweidimensionaler Fall)

$\mathbf{T} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$