Institut für Statistik
Gerhard Tutz, Moritz Berger, Wolfgang Pößnecker Sommersemester 2015

Formelsammlung zu Multivariate Verfahren

Version 07.04.2015

Diese Formelsammlung darf in der Klausur verwendet werden. Eigene Notizen und Ergänzungen dürfen eingefügt, aber 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 von $\mathbf{x}$ mit $E(\mathbf{x}) = \mu$ und Kovarianz $\text{Var}(\mathbf{x}) = \Sigma$
- Bezeichne $\mathbf{X}$ die $(n \times p)$–Datenmatrix $\mathbf{X} = \begin{pmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix} = \begin{pmatrix} x_{11} & \ldots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \ldots & x_{np} \end{pmatrix}$

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:

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

Alternativer Schätzer für Σ:

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

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

$$r_{ij} = \frac{\sum_{s=1}^n (x_{si} - \bar{x}_i)(x_{sj} - \bar{x}_j)}{\sqrt{\sum_{s=1}^n (x_{si} - \bar{x}_i)^2} \sqrt{\sum_{s=1}^n (x_{sj} - \bar{x}_j)^2}}$$

Satz:

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

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

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

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

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 $(n-1)\mathbf{S} = \boldsymbol{\Sigma}_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \sim \mathbf{W}_p(\boldsymbol{\Sigma}, n-1)$.

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

1.2 Schätzung im Mehr–Stichprobenfall

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

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

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:

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

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

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

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

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

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

die Streuung zwischen den Gruppen, wobei

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

den großen Mittelwert über alle Klassen bezeichnet, mit

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

für die Gesamtstreuungsmatrix

$$\mathbf{T} = \sum_{k=1}^{g}\sum_{i=1}^{n_k}(\mathbf{x}_i^{(k)} - \bar{\mathbf{x}})(\mathbf{x}_i^{(k)} - \bar{\mathbf{x}})'$$

gilt die folgende Zerlegung

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

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

Daten:

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

Hypothesen:

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

Teststatistik:

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

H_0 ablehnen, wenn gilt:

$$T^2 > \chi^2(p; 1-\alpha)$$

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

Daten:

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

Hypothesen:

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

Teststatistik:

$$T_s = \frac{n(n-p)}{p(n-1)}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0),$$
$$T_s \overset{H_0}{\sim} F(p, n-p)$$

H_0 ablehnen, wenn gilt:

$$T_s > F(p, n-p; 1-\alpha)$$

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 vorausgesetzt.

Hypothesen:

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

Teststatistik:

$$\begin{aligned} T^2 &= \frac{n_1 \cdot n_2}{n_1 + n_2} \cdot (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \\ \text{mit } \mathbf{S} &= \frac{1}{n_1 + n_2 - 2} [(n_1 - 1)\mathbf{S}_1 + (n_2 - 1) \cdot \mathbf{S}_2] \\ \text{und } \mathbf{S}_i &= \frac{1}{n_i - 1} \sum\nolimits_{j=1}^{n_i} (\mathbf{x}_j^{(i)} - \bar{\mathbf{x}}^{(i)})(\bar{\mathbf{x}}_j^{(i)} - \bar{\mathbf{x}}^{(i)})^T \end{aligned}$$

H_0 ablehnen:

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

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

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:

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

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

Unbekannte Kovarianzmatrix:

$$\begin{aligned} H_0 \text{ ablehnen} &\Leftrightarrow \frac{(n-p) \cdot n}{(n-1) \cdot p} \cdot \bar{\mathbf{d}}^T \cdot \mathbf{S}_{\mathbf{d}}^{-1} \bar{\mathbf{d}} > F(p; n-p; 1-\alpha) \\ \text{mit } \bar{\mathbf{d}} &= \frac{1}{n} \boldsymbol{\Sigma}_{i=1}^n \mathbf{d}_i \\ \mathbf{S}_d &= \frac{1}{n-1} \sum\nolimits_{i=1}^n (\mathbf{d}_i - \bar{\mathbf{d}})(\mathbf{d}_i - \bar{\mathbf{d}})^T \end{aligned}$$

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

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

2 Diskriminanzanalyse/Klassifikation

2.1 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 $f(\mathbf{x}) = p(1)f(\mathbf{x} \mid 1) + \ldots + p(k)f(\mathbf{x} \mid k)$.

Satz von Bayes:

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

2.2 Bayes–Zuordnung:

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

Fehlklassifikationswahrscheinlichkeiten:
Es lassen sich verschiedene Formen der Fehlklassifikationen durch die Zuordnungsregel δ unterscheiden.

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

$$\begin{aligned}\varepsilon(\mathbf{x}) &= P(\delta(\mathbf{x}) \neq Y \mid \mathbf{x}) = 1 - P(\delta(\mathbf{x}) = Y \mid \mathbf{x}) \\ &= 1 - P(\delta(\mathbf{x}) \mid \mathbf{x}).\end{aligned}$$

- *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

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

Es 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

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

minimiert die Gesamtfehlerrate ε.

Äquivalente Diskriminanzfunktionen:

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

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

2.3 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

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

Mit den Diskrimininanzfunktionen

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

erhält man

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

2.4 Zuordnungsregel unter der Voraussetzung Normalverteilung

Klassenverteilungen:

$$f(\mathbf{x} \mid i) = \frac{1}{(2\pi)^{p/2}(\det \mathbf{\Sigma}_i)^{1/2}} \cdot \exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\}$$

Bayes–Regel:

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

Zusätzlich unabhängige standardisierte Merkmale, $\mathbf{\Sigma}_i = \sigma^2 \mathbf{I}$:

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

Sind die a priori Wahrscheinlichkeiten gleich groß, so reduziert sich die Diskriminanzfunktion auf

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

Klassenweise identische Kovarianzmatrizen; $\mathbf{\Sigma}_i = \mathbf{\Sigma}, \;\; i = 1,\ldots,k$:

$$d_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \ln p(i)$$

$$d_i(\mathbf{x}) = \boldsymbol{\mu}_i^T \mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu}_i + \ln p(i), \quad i = 1,\ldots,k$$

Für k=2 erhält man:

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

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

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

2.5 Fishersche Diskriminanzanalyse

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

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

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

wobei

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

Lösung:

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

mit

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

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:

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

mit

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

3 Clusteranalyse

Objektmenge $\Omega = \{a_1, \ldots, a_n\}$
Zugehörige Merkmalsvektoren $\{x_1, \ldots, x_n\}$
Gesucht: Partition = disjunkte vollständige Zerlegung $C_1, \ldots, C_g$ so daß

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

3.1 Distanzmaße für Objekte:

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

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

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

3.2 Distanzmaße:

Binäre Merkmale

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

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

$h_{ij}(y, z)$ = $|\ \{s \mid (x_{is}, x_{js}) = (y, z)\}\ |$ 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

$$\mathbf{x} = (x_1, \ldots, x_p)\ ,\ \ \mathbf{y} = (y_1, \ldots, y_p)$$

L_q– Metrik

$$d(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{p} |\ x_i - y_i\ |^q\right)^{1/q}$$

$q = 1$ City–Block–Metrik

$q = 2$ Euklidische Metrik

Mahalanobis–Distanz

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

mit empirischer Kovarianzmatrix $\mathbf{S}$

3.3 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
$$\begin{aligned} D(C_r, C_s) &= \frac{1}{n_r n_s} \sum_{a_i \in C_r} \sum_{a_j \in C_s} d(a_i, a_j) \\ \text{mit } n_i &= |\ C_i\ | \end{aligned}$$

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

3.4 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

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

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

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

Die Gesamtstreuung

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

ist zerlegbar in

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

3.5 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

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

3. *Verallgemeinertes Determinantenkriterium*

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

4 Multivariate lineare Regression

4.1 Konzept

Daten:

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

$$\mathbf{y}_i^T = (y_{i1}, \ldots, y_{iq}) \quad q \text{ abhängige Variablen}$$
$$\mathbf{x}_i^T = (1, x_{i1}, \ldots, x_{ip}) \quad p \text{ konstante und fixe Prädiktoren}$$

Klassisches lineares Modell:

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

Matrixdarstellungen:

Vektorielle Form

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

bzw.

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

und als Gesamtmodell

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

wobei

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

Multivariate Formulierung I

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

Multivariate Formulierung II

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

bzw. als Gesamtmodell

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

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

Annahmen:

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

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

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

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

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

4.2 Schätzen

Kleinste Quadrate–Schätzung

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

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

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

Als Vektor

$$\hat{\boldsymbol{\beta}} = \text{BDiag}\{(\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T\} \mathbf{y}_0 \text{ oder } \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

mit

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

Schätzung der Kovarianz:

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

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-p-1} \mathbf{E}^T \mathbf{E}$$

wobei

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

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

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

(2) $\text{Var}(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\boldsymbol{\beta}}_{(j)}) = \sigma_{ij} (\mathbf{X}_0^T \mathbf{X}_0)^{-1}$

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

Weiter gilt mit Normalverteilungsannahme

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

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

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

4.3 Testverfahren

Tests: (Unter Normalverteilungsannahme)

$H_0: \ \mathbf{CB} = \mathbf{\Gamma} \qquad H_1: \ \mathbf{CB} \neq \mathbf{\Gamma} \qquad rg(\mathbf{C}) = s$

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

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

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

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

5 Assoziationsstrukturen

5.1 Marginale, bedingte und partielle Korrelation

Marginale Korrelation von Y und X

$$\rho_{YX} = \frac{\text{cov}(Y,X)}{\sigma_Y \sigma_X}$$

Bedingte Korrelation von Y und X, gegeben $Z = z$

$$\rho_{YX|Z=z} = \frac{\text{cov}(Y,X|Z=z)}{\sqrt{\text{var}(Y|Z=z)}\sqrt{\text{var}(X|Z=z)}}$$

Partielle Korrelation von Y und X nach Elimination des linearen Anteils von Z

$$\rho_{XY\cdot Z} = \frac{\text{cov}(\varepsilon_{Y\cdot Z}, \varepsilon_{X\cdot Z})}{\sqrt{\text{var}(\varepsilon_{Y\cdot Z})}\sqrt{\text{var}(\varepsilon_{X\cdot Z})}} = \frac{\rho_{YX} - \rho_{XZ}\rho_{YZ}}{\sqrt{1-\rho_{YZ}^2}\sqrt{1-\rho_{XZ}^2}},$$

mit $\varepsilon_{Y\cdot Z}$, $\varepsilon_{X\cdot Z}$ als Residuen aus den Regressionsmodellen

$$Y = \beta_0 + \beta Z + \varepsilon_{Y\cdot Z} \quad \text{und} \quad X = \gamma_0 + \gamma Z + \varepsilon_{X\cdot Z}.$$

Für (Y, X, Z) gemeinsam normalverteilt gilt

$$\rho_{YX|Z=z} = \rho_{XY\cdot Z}.$$

5.2 Korrelationskoeffizienten bei vektoriellen Größen

Ausgangspunkt: Zufallsvektoren $\mathbf{y}$, $\mathbf{x}$ und $\mathbf{z}$ mit

$$\boldsymbol{\mu} = \mathbf{E}\begin{pmatrix}\mathbf{y}\\ \mathbf{x}\\ \mathbf{z}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\mu}_\mathbf{y}\\ \boldsymbol{\mu}_\mathbf{x}\\ \boldsymbol{\mu}_\mathbf{z}\end{pmatrix}, \qquad \mathbf{\Sigma} = \text{cov}\begin{pmatrix}\mathbf{y}\\ \mathbf{x}\\ \mathbf{z}\end{pmatrix} = \begin{pmatrix}\mathbf{\Sigma_{yy}} & \mathbf{\Sigma_{yx}} & \mathbf{\Sigma_{yz}}\\ \mathbf{\Sigma_{xy}} & \mathbf{\Sigma_{xx}} & \mathbf{\Sigma_{xz}}\\ \mathbf{\Sigma_{zy}} & \mathbf{\Sigma_{zx}} & \mathbf{\Sigma_{zz}}\end{pmatrix}$$

Marginale Korrelationsmatrix für $\mathbf{y}$ gegeben $\mathbf{x}$:

$$\mathbf{K} = (k_{ij}) \text{ mit } k_{ij} = \sigma_{y_i x_j}/(\sigma_{y_i}\sigma_{x_i}),$$

wobei $\sigma_{y_i x_j}$ Einträge von $\mathbf{\Sigma_{yx}}$.

Matrix der **bedingten Kovarianzen** für $\mathbf{y}$ und $\mathbf{x}$ gegeben $\mathbf{z} = \tilde{\mathbf{z}}$:

$$\mathbf{\Sigma}_{\mathbf{yx}|\mathbf{z}=\tilde{\mathbf{z}}} = (\text{cov}(y_i, x_j|\mathbf{z} = \tilde{\mathbf{z}}))_{ij}.$$

Partielle Korrelation skalarer Merkmale y und x nach Elimination des linearen Anteils von $\mathbf{z}$:

$$\varrho_{xy\cdot\mathbf{z}} = \frac{\text{cov}(\varepsilon_{y\cdot\mathbf{z}}, \varepsilon_{x\cdot\mathbf{z}})}{\sqrt{\text{var}(\varepsilon_{y\cdot\mathbf{z}})}\sqrt{\text{var}(\varepsilon_{x\cdot\mathbf{z}})}} = \frac{\sigma_{yx} - \mathbf{\Sigma}_{y\mathbf{z}}\mathbf{\Sigma}_\mathbf{z}^{-1}\mathbf{\Sigma}_{\mathbf{z}x}}{\sqrt{\sigma_y^2 - \mathbf{\Sigma}_{y\mathbf{z}}\mathbf{\Sigma}_\mathbf{z}^{-1}\mathbf{\Sigma}_{\mathbf{z}y}}\sqrt{\sigma_x^2 - \mathbf{\Sigma}_{x\mathbf{z}}\mathbf{\Sigma}_\mathbf{z}^{-1}\mathbf{\Sigma}_{\mathbf{z}x}}},$$

wobei $\varepsilon_{y\cdot\mathbf{z}} = y - \mathbf{z}^T\boldsymbol{\beta}$ und $\varepsilon_{x\cdot\mathbf{z}} = x - \mathbf{z}^T\boldsymbol{\gamma}$.

5.3 Multiple Korrelation

Ausgangspunkt: lineares Regressionsmodell $y = \mathbf{x}^T\boldsymbol{\beta} + \varepsilon_{y\cdot\mathbf{x}}$, damit ergibt sich

$$\begin{aligned}\text{var}(\mathbf{x}^T\boldsymbol{\beta}) &= \mathbf{\Sigma}_{y\mathbf{x}}\mathbf{\Sigma}_\mathbf{x}^{-1}\mathbf{\Sigma}_{\mathbf{x}y} = \text{cov}(y, \mathbf{x}^T\boldsymbol{\beta})\\ \text{var}(y|\mathbf{x}) &= \text{var}(\varepsilon_{y\cdot\mathbf{x}}) = \sigma_y^2 - \mathbf{\Sigma}_{y\mathbf{x}}\mathbf{\Sigma}_\mathbf{x}^{-1}\mathbf{\Sigma}_{\mathbf{x}y}\end{aligned}$$

und die Zerlegung

$$\text{var}(y) = \text{var}(\mathbf{x}^T\boldsymbol{\beta}) + \text{var}(y|\mathbf{x}).$$

Multipler Korrelationskoeffizient:

$$R = \varrho(y, \mathbf{x}^T\boldsymbol{\beta}) = \frac{\text{cov}(y, \mathbf{x}^T\boldsymbol{\beta})}{\sqrt{\text{var}(y)}\sqrt{\text{var}(\mathbf{x}^T\boldsymbol{\beta})}}.$$

Es gilt

$$R^2 = \frac{\text{var}(\mathbf{x}^T\boldsymbol{\beta})}{\text{var}(y)} = 1 - \frac{\text{var}(y|\mathbf{x})}{\text{var}(y)}.$$

5.4 Kanonische Korrelation

Ausgangspunkt: Zufallsvektoren $\mathbf{y}$ und $\mathbf{x}$ der Länge p bzw. q mit $\mathbf{\Sigma_{xy}} = \text{cov}(\mathbf{x}, \mathbf{y})$.

Betrachte **Maximierungsproblem**

$$\varrho(\mathbf{a}^T\mathbf{x}, \mathbf{b}^T\mathbf{y}) \overset{\mathbf{a},\mathbf{b}}{\to} \max$$
mit der Nebenbedingung $\quad \text{var}(\mathbf{a}^T\mathbf{x}) = \text{var}(\mathbf{b}^T\mathbf{y}) = 1.$

wobei

$$\varrho(\mathbf{a}, \mathbf{b}) = \varrho(\mathbf{a}^T\mathbf{x}, \mathbf{b}^T\mathbf{y}) = \frac{\text{cov}(\mathbf{a}^T\mathbf{x}, \mathbf{b}^T\mathbf{y})}{\sqrt{\text{var}(\mathbf{a}^T\mathbf{x})}\sqrt{\text{var}(\mathbf{b}^T\mathbf{y})}} = \frac{\mathbf{a}^T\mathbf{\Sigma_{xy}}\mathbf{b}}{\sqrt{\mathbf{a}^T\mathbf{\Sigma_{xx}}\mathbf{a}}\sqrt{\mathbf{b}^T\mathbf{\Sigma_{yy}}\mathbf{b}}}.$$

Lösen des Maximierungsproblems für $i = 1, \ldots, s = \min\{p, q\}$ mit

$$\text{cov}(\mathbf{a}_i\mathbf{x}, \mathbf{a}_j\mathbf{x}) = \text{cov}(\mathbf{b}_i\mathbf{y}, \mathbf{b}_j\mathbf{y}) = 0 \quad \text{für} \quad i \neq j.$$

ergibt

$\mathbf{a}_i$, $\mathbf{b}_i$ als die $i$ten *Korrelationsvektoren*,

$u_i = \mathbf{a}_i^T\mathbf{x}$, $v_i = \mathbf{b}_i^T\mathbf{y}$ als $i$te *Korrelationsvariablen* und

$\varrho_i = \varrho(u_i, v_i)$ als $i$te *kanonische Korrelationskoeffizienten*.

Für $\mathbf{u}^T = (u_1, \ldots, u_s)$ und $\mathbf{v}^T = (v_1, \ldots, v_s)$ gilt

$$\text{cor}\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 & \varrho_1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & \varrho_s \\ \varrho_1 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & \varrho_s & 0 & \cdots & 1 \end{pmatrix}.$$

Schärfere Formulierung des Maximierungsproblems:

$$\varrho(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i^T\mathbf{y}) \overset{\mathbf{a}_i,\mathbf{b}_i}{\to} \max$$

mit den Nebenbedingungen

$$\text{var}(\mathbf{a}_i^T\mathbf{x}) = \text{var}(\mathbf{b}_i^T\mathbf{y}) = 1, \quad i = 1, \ldots, s$$
$$\text{cov}(\mathbf{a}_i^T\mathbf{x}, \mathbf{a}_j^T\mathbf{x}) = \text{cov}(\mathbf{b}_i^T\mathbf{y}, \mathbf{b}_j^T\mathbf{y}) = 0, \quad j = 1, \ldots, i-1.$$

Lösungen $(\mathbf{a}_i, \mathbf{b}_i)$ über das **kombinierte Eigenwertproblem**

$$\begin{aligned}(\mathbf{\Sigma_{xy}}\mathbf{\Sigma_{yy}^{-1}}\mathbf{\Sigma_{yx}} - \lambda_i\mathbf{\Sigma_{xx}})\mathbf{a}_i &= 0 \\ (\mathbf{\Sigma_{yx}}\mathbf{\Sigma_{xx}^{-1}}\mathbf{\Sigma_{xy}} - \lambda_i\mathbf{\Sigma_{yy}})\mathbf{b}_i &= 0\end{aligned}$$

6 Hauptkomponentenanalyse

Ausgangsgrößen:

$$x^T = (x_1, \ldots, x_p)$$
$$\mu = \mathbf{E}(\mathbf{x}), \mathbf{\Sigma} = \text{cov}(\mathbf{x})$$

Problem:

$\mathbf{y}_i = \mathbf{a}_i^T\mathbf{x}, i = 1, \ldots, p$ so dass

$$\text{var}(y_i) = \mathbf{a}_i^T\mathbf{\Sigma}\mathbf{a}_i \text{maximal}$$

Restriktionen:

(1) $\mathbf{a}_i^T\mathbf{a}_i = 1$

(2) $\mathbf{a}_i^T\mathbf{a}_j = 0, j = 1, \ldots, i-1$
Äquivalent ist
$\text{cov}(y_i, y_j) = 0, j = 1, \ldots, i-1$

Lösung:

$$\mathbf{y} = \mathbf{\Psi}^T\mathbf{x} \quad \text{mit} \quad \mathbf{y}^T = (y_1, \ldots, y_p)$$

wobei

$\mathbf{\Sigma} = \mathbf{\Psi}\mathbf{\Lambda}\mathbf{\Psi}^T$
$\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_p), \lambda_1 \geq \lambda_2 \geq \ldots \lambda_p$
$\mathbf{\Psi} = (\mathbf{a}_1, \ldots, \mathbf{a}_p)$

Eigenschaften:

(1) $\text{cov}(\mathbf{y}) = \mathbf{\Lambda} \quad (\text{var}(y_i) = \lambda_i, \text{cov}(y_i, y_j) = 0, i \neq j)$

(2) $tr(\mathbf{\Sigma}) = tr(\mathbf{\Lambda})$

7 Faktorenanalyse

Modell:

$$\mathbf{X} - \boldsymbol{\mu} = \boldsymbol{\Gamma}\mathbf{F} + \mathbf{U},$$

wobei

$\mathbf{X}$	$(p \times 1)$ - Vektor der beobachtbaren Größen
$\boldsymbol{\mu}$	Erwartungswert von $\mathbf{X}$
$\mathbf{F}$	$(k \times 1)$ - Vektor der Faktoren
$\boldsymbol{\Gamma}$	$(p \times k)$ - Matrix der Faktorladungen
$\mathbf{U}$	$(p \times 1)$ - Vektor der spezifischen Faktoren

Annahmen:

(1) $E(\mathbf{F}) = 0$

(2) $\mathbf{U} \sim V(0, \boldsymbol{\Psi})$, mit $\boldsymbol{\Psi} = diag(\Psi_1^2, \ldots, \Psi_p^2)$

(3) $cov(\mathbf{F}, \mathbf{U}) = \mathbf{0}$

(4) $cov(\mathbf{F}) = \mathbf{I}$ orthogonales Faktormodell

Fundamentaltheorem für $\mathbf{s} = cov(\mathbf{X})$.

$$\mathbf{s} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \Psi$$

Insbesondere:

$$var(X_i) = \sum_{j=1}^{m} \gamma_{ij}^2 + \Psi_i^2 \ , \ cov(\mathbf{X}, \mathbf{F}) = \boldsymbol{\Gamma}$$

Rotationsproblem (Eindeutigkeit von Γ bei gegebenem k, Ψ)

$$\Gamma'\Psi^{-1}\Gamma = diag(\lambda_1, \ldots, \lambda_k) \ , \ \lambda_1 \geq \lambda_2 \geq, \ldots, \geq \lambda_k.$$