Lösungsvorschläge für die Übungsaufgaben der Vorlesung “Statistik“

Thomas Rupp
thomas@7t7.de

3. Februar 2002

Aufgabe 41 Für $c =$`qtukey(`$1-\alpha, 7, 70-7$`)` und s^2 die geschätzte Streuung ist $I_{j,l} := (\bar{X}_{j\cdot} - \bar{X}_{l\cdot} \mp c\frac{s}{\sqrt{10}})$ ein $1-\alpha$ Konfidenzintervall für die Differenz der Mittelwerte der beiden Gruppen j und l.
s können wir uns aus der Varianzanalyse durch R besorgen. Nun müssen wir also nurnoch das Konfidenzintervall für alle Paare ausrechnen und prüfen ob die Differenz der geschätzten Erwartungswerte noch im KI liegt. Wie man das in R implementieren kann, zeigt das kleine Programm vom Margret Knape.

Aufgabe 42 Mit Hilfe von R ist diese Aufgabe schnell bewältigt (Augenmass lässt einen signifikanten Unterschied als eher unwahrscheinlich erscheinen):

Parametrischer Test (Varianzanalyse - ANOVA Tafel)

Zuerst speichern wir die Daten in der rechtsstehenden Form in eine Datei *Uhren.dat* ab. Mit dem Befehl `Uhren <- read.table(„Uhren.dat“,header=T)` können wir diese in einer Variablen speichern und anschliessend diese mit `analyse <- aov(Zyklen~Typ, data=Uhren)` der gewünschten Analyse unterziehen.
Mit `print(summary(ergebnis))` wird das Resultat ausgegeben. Uns interessiert hier nur der P-Wert von 0.6167. Ein signifikanter Unterschied ist also nicht festzustellen.

Zyklen	Typ
1.7	*I*
1.9	*I*
⋮	⋮
82.5	*I*
13.6	*II*
⋮	⋮
46.9	*III*

Nichtparametrischer Test (Kruskal-Wallis)

Da wir die Daten schon haben, können wir gleich mit `kruskal.test(Uhren$Zyklen ~Uhren$Typ)` das Resultat ausgeben lassen. Der P-Wert von 0.3405 ist ebenfalls kein Nachweis eines signifikanten Unterschieds.

Aufgabe 43
Das Modell wird durch folgende Matrix beschrieben: $X_{i,j} = \alpha + \beta_i + \gamma_j + \sigma Z_{i,j}$. Diese Matrix können wir in einem höherdimensionalen Raum als Vektor auffassen; einfach die Spalten untereinanderschreiben:

$$\mathfrak{X} = \vec{\mu}_1 + \vec{\mu}_2 + \vec{\mu}_3 + \sigma\mathfrak{Z} = \begin{pmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1J} \\ X_{21} \\ \vdots \\ X_{IJ} \end{pmatrix} = \begin{pmatrix} \alpha \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \alpha \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_1 \\ \beta_2 \\ \beta_2 \\ \vdots \\ \beta_I \end{pmatrix} + \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_J \\ \gamma_1 \\ \vdots \\ \gamma_J \end{pmatrix} + \sigma \begin{pmatrix} Z_{11} \\ Z_{12} \\ \vdots \\ Z_{1J} \\ Z_{21} \\ \vdots \\ Z_{IJ} \end{pmatrix}.$$

Dabei treten die Blöcke $\beta_i, \ldots, \beta_i$ jeweils j-mal und die Sequenzen $\gamma_1, \ldots, \gamma_J$ je I-mal auf und $\mathfrak{X}, \mathfrak{Z}, \vec{\mu}_i \in \mathbb{R}^{IJ}$.

a)
$M_1 = \{\vec{\mu}_1 : \alpha \in \mathbb{R}\}, \dim M_1 = 1$.
$M_2 = \{\vec{\mu}_2 : \sum_{i=1}^{I} \beta_i = 0, \beta_i \in \mathbb{R}\}, \dim M_2 = I - 1$.
$M_3 = \{\vec{\mu}_3 : \sum_{j=1}^{J} \gamma_j = 0, \gamma_j \in \mathbb{R}\}, \dim M_3 = J - 1$.
$M_1 \perp M_2 \Leftrightarrow \langle \vec{\mu}_1, \vec{\mu}_2 \rangle = 0 \Leftrightarrow \sum_{i=1}^{I} \alpha J \beta_i = 0 \Leftrightarrow \alpha J \sum_{i=1}^{I} \beta_i = 0$.
$M_1 \perp M_3 \Leftrightarrow \langle \vec{\mu}_1, \vec{\mu}_3 \rangle = 0 \Leftrightarrow \sum_{j=1}^{J} \alpha I \gamma_J = 0 \Leftrightarrow \alpha I \sum_{j=1}^{J} \gamma_j = 0$.
$M_1 \perp M_3 \Leftrightarrow \langle \vec{\mu}_1, \vec{\mu}_3 \rangle = 0 \Leftrightarrow \sum_{i=1}^{I} \beta_i \sum_{j=1}^{J} \gamma_j = 0$.

b)

z.z ist jeweils $(\mathfrak{X} - \underbrace{\mathcal{P}_{M_i}\mathfrak{X}}_{=\text{Schätzvektor}})\perp M_i$:

i)

$\hat{\alpha} = \bar{X}_{\cdot\cdot} = \frac{1}{IJ}\sum_{i,j} X_{ij}$

$\left\langle \mathfrak{X} - \alpha(1,\dots,1)^\top, a,\dots,a \right\rangle = a[(X_{11}-\hat{\alpha})+\dots+(X_{IJ}-\hat{\alpha})] = a[-IJ\hat{\alpha}+IJ\bar{X}_{\cdot\cdot}] = 0$

ii)

$\hat{\beta}_i = \bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot} = \frac{1}{J}\sum_{j=1}^{J} X_{ij} - \frac{1}{IJ}\sum_{i,j} X_{ij}$

$$\left\langle \mathfrak{X} - (\underbrace{\hat{\beta}_1,\dots,\hat{\beta}_1}_{J \text{ Stück}},\dots,\hat{\beta}_2,\dots,\hat{\beta}_I,\hat{\beta}_I)^\top, (b_1,\dots,b_1,\dots,b_2,\dots,b_I) \right\rangle$$

$$= \sum_{i=1}^{I} b_i[-J\hat{\beta}_i + \sum_{j=1}^{J} X_{ij}] = \sum_{i=1}^{I} J\bar{X}_{\cdot\cdot} b_i = J\bar{X}_{\cdot\cdot}\sum_{i=1}^{I} b_i = 0$$

iii)

$\hat{\gamma}_j = \bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot} = \frac{1}{I}\sum_{i=1}^{I} X_{ij} - \frac{1}{IJ}\sum_{i,j} X_{ij}$

$$\left\langle \mathfrak{X} - (\underbrace{\hat{\gamma}_1,\hat{\gamma}_2,\dots,\hat{\gamma}_J}_{I \text{ mal wiederholt}},\dots,\hat{\gamma}_1,\dots,\hat{\gamma}_J)^\top, (g_1,\dots,g_J,\dots,g_1,\dots,g_J) \right\rangle$$

$$= \sum_{j=1}^{J} g_j[-I\hat{\gamma}_j + I\bar{X}_{\cdot j}] = I\bar{X}_{\cdot\cdot}\sum_{j=1}^{J} g_j = 0$$

iv)

$\hat{\sigma}^2$ ist im linearen Modell allgemein $\frac{1}{n-k}\|\mathcal{P}_M\mathfrak{X}\|^2$, wobei M die Summe der Unterräume der systematischen Komponenten ist. Hier ist also $M = M_1 \oplus M_2 \oplus M_3$ und $n = IJ, k = 1 + (I-1) + (J-1)$.[1]

Also ist

$$\hat{\sigma}^2 = \frac{\sum\limits_{i,j}\left[(X_{ij} - \bar{X}_{\cdot\cdot} - (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}) - (\bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot})\right]^2}{(I-1)(J-1)} = \frac{\sum\limits_{i,j}\left[X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot}\right]^2}{(I-1)(J-1)}.$$

Mit $\hat{\sigma}^2 \sim \chi^2((I-1)(J-1))$.

c)

Unter der Hypothese, dass $\alpha = 0$ ist $\frac{1}{\dim M_1}\|\mathcal{P}_{M_1}\mathfrak{X}\|^2 \sim \chi^2(\dim M1)$ verteilt. Bei den anderen geht es analog. Somit ist

$$F_i := \frac{\frac{\|\mathcal{P}_{M_i}\mathfrak{X}\|^2}{\dim M_i}}{\hat{\sigma}^2} \sim F(\dim M_i, I-1, J-1).$$

Somit sind die Teststatistiken unter der jeweiligen Hypthesen fisherverteilt.

d)

Die gegebene Matrix ist $(X)_{ij} = \begin{pmatrix} 1 & 0 & 2 \\ 3 & 4 & 8 \end{pmatrix}$. Die Merkmale, die wir benötigen, sind dann $\bar{X}_{\cdot\cdot} = \frac{18}{6} = 3, \bar{X}_{1\cdot} = 1, \bar{X}_{2\cdot} = 5, \bar{X}_{\cdot 1} = 2, \bar{X}_{\cdot 2} = 2, \bar{X}_{\cdot 3} = 5$, aus denen wir die Schätzer $\hat{\alpha} = 3, \hat{\beta}_1 = -2, \hat{\beta}_2 = 2, \hat{\gamma}_1 = -1, \hat{\gamma}_2 = -1, \hat{\gamma}_3 = 2$ gwinnen.

[1] $n - k = IJ - (I-1) - (J-1) - 1 = (I-1)(J-1)$

Jetzt können wir die Matrix mit den Residuen aufschreiben

$$R_{ij} = X_{ij} - \alpha - \beta_i - \gamma_j \Rightarrow (R)_{ij} = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 0 & 1 \end{pmatrix}.$$

Damit können wir nun die ANOVA-Tafel aufstellen:

ANOVA-Tafel	# DF	Quadratsumme (QS)	QS/DF	F-Wert	P-Wert
Globales Mittel	1	$6 \cdot 3^2 = 54$	54	27	0.035
Zeilenmittel	1	$3((-2)^2 + 2^2) = 24$	24	12	0.074
Spaltenmittel	2	$2(1+1+4) = 12$	6	3	0.25
Residuen	2	$1+1+1+1 = 4$	2		

Angemerkt sei noch, dass die Residuen-Freiheitsgerade ja gerade $(I-1)(J-1)$ sind, die F-Werte gerade die Quotienten von (Quotient von Gruppen-QS mit Gruppen-# DF) mit (Quotient von Residuen QS mit Residuen-# DF) sind. Die P-Werte beziehen sich dann auf die Wahrscheinlichkeit, dass die dazugehörigen Merkmale gleich sind.
Also ist die Wahrscheinlichkeit eines solchen Ergebnisses unter $\gamma_1 = \gamma_2 = \gamma_3$ gleich 0.25. Unter $\beta_1 = \beta_2$ ist dieses nurnoch 0.074 und unter $\alpha = \beta_i = \gamma_j$ nur 0.035.

Aufgabe 44 Diese haben wir ausgelassen.