Multivariate Statistik, Übung 6

Henry Haustein

Aufgabe 1

Es lohnt sich meiner Meinung nach hier, Ideen aus der linearen Algebra vorher sich anzuschauen. Das Produkt zweier Zeilenvektoren xy' ist eine Zahl, man nennt es das Skalarprodukt von x und y und schreibt dafür $\langle x, y\rangle$. Das Skalarprodukt hat unter anderem die Eigenschaft der Additivität, also $\langle x, y\rangle + \langle u, v\rangle = \langle x+u, y+v\rangle$. Mehr brauchen wir nicht, um das Fundamentaltheorem der multivariaten Varianzanalyse in wenigen Zeilen zu beweisen:

$$
\begin{aligned}
\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)' + \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})' &= \sum_{i=1}^{g}\sum_{j=1}^{n_i}\underbrace{(x_{ij}-\bar{x}_i)}_{x}(x_{ij}-\bar{x}_i)' + \underbrace{(\bar{x}_i-\bar{x})}_{y}(\bar{x}_i-\bar{x})' \\
&= \sum_{i=1}^{g}\sum_{j=1}^{n_i} xx' + yy' \\
&= \sum_{i=1}^{g}\sum_{j=1}^{n_i} \langle x, x\rangle + \langle y, y\rangle \\
&= \sum_{i=1}^{g}\sum_{j=1}^{n_i} \langle x+y, x+y\rangle \\
&= \sum_{i=1}^{g}\sum_{j=1}^{n_i} [(x_{ij}-\bar{x}_i) + (\bar{x}_i-\bar{x})]\cdot[(x_{ij}-\bar{x}_i) + (\bar{x}_i-\bar{x})]' \\
&= \sum_{i=1}^{g}\sum_{j=1}^{n_i} (x_{ij}-\bar{x})(x_{ij}-\bar{x})'
\end{aligned}
$$

Aufgabe 2

Bei einer standardisierten Matrix sind die Mittelwerte der Spalten 0.

$$
\begin{aligned}
S &= \frac{1}{n-1}T \\
&= \frac{1}{n-1}(Z'Z - n\underbrace{\bar{z}}_{0}\underbrace{\bar{z}'}_{0}) \\
&= \frac{1}{n-1}Z'Z \\
&= R
\end{aligned}
$$

Aufgabe 3

(a) Der F-Test ist hier ein geeignetes Mittel, wobei wir Körpergröße und Körpergewicht herausrechnen wollen.

(b) Gruppe 1: $(x_1,\ x_2,\ x_3,\ x_4)$
Gruppe 2: (Körpergewicht, Körpergröße)
$H_0 : \mu_{\text{Mitglied},1.2} = \mu_{\text{Nicht-Mitglied},1.2}$ vs. $H_1 : \mu_{\text{Mitglied},1.2} \neq \mu_{\text{Nicht-Mitglied},1.2}$

(c) Wir haben $p = 4$, $q = 2$, $g = 2$, $k = 6$, $n = 10$

$$
\begin{aligned}
\nu_1 &= p(g-1) = 4 \\
\nu_2 &= s\left[(n-1-p) - \frac{q+g}{2}\right] - \frac{q(g-1)-2}{2} = 3 \\
s &= \sqrt{\frac{k^2(g-1)^2-4}{k^2+(g-1)^2-5}} = 1 \\
\Lambda &= \frac{\det(W_{1.2})}{\det(T_{1.2})} = -0.8391879 \\
F &= \frac{1-\Lambda^{\frac{1}{s}}}{\Lambda^{\frac{1}{s}}} = -13.14977
\end{aligned}
$$

Ich bin mir nicht sicher, ob meine berechneten Werte richtig sind, da als Teststatistik der Wert 3.32 rauskommen soll. Ich vermute, dass die gegebenen Matrizen falsch sind, so ist z.B T nicht symmetrisch: $t_{26} \neq t_{62}$.

(d) Die Matrix T_{11} ist eine 4×4-Matrix, T_{22} hat die Dimensionen 2×2. Logischerweise dann die Dimension von T_{12} 4×2. Selbiges gilt für die Partitionen von W.

(e) Der kritische Wert ist $F_{4,3;0.95} = 9.1172$. Man kann also H_0 nicht ablehnen.

Aufgabe 4

Wir bilden wieder 2 Gruppen. Gruppe 1: $(X_1,\ X_2)$ und Gruppe 2: $(X_3,\ X_4)$. Wir testen dann $H_0 : \mu_{\text{Land}_1,1.2} = \mu_{\text{Land}_2,1.2}$ vs. $H_1 : \mu_{\text{Land}_1,1.2} \neq \mu_{\text{Land}_2,1.2}$. Zudem haben wir $p = 2$, $q = 2$, $g = 2$, $k = 4$ und $n = 72$. Damit ergibt sich

$$
\begin{aligned}
\nu_1 &= 2 \\
\nu_2 &= 67 \\
s &= 1 \\
F &= 136.5508
\end{aligned}
$$

Es gilt $F > F_{2,67;0.95} = 3.134$, also lehnen wir H_0 ab. Ein Tempolimit verändert also die Verkehrsunfälle.

```
p = 2
q = 2
g = 2
n = 72
k = p + q
alpha = 0.05

v1 = p*(g-1)
```

```
s = sqrt((k^2 * (g-1)^2 - 4)/(k^2 + (g-1)^2 - 5))
v2 = s*((n-1-p)-(q+g)/2) - (q*(g-1)-2)/2

lambda = 0.197

F = (1-lambda^(1/s))/(lambda^(1/s)) * v2/v1

krit = qf(1-alpha,v1,v2)
```