Übungen zur Angewandten Stochastik

Blatt 10

1. Aufgabe (2-Faktor-Varianzanalyse) **a)** Nehmen wir an, der Ausgang eines Experiments wird von zwei Faktoren mit $s_1 \geq 2$ bzw. $s_2 \geq 2$ „Stufen“ bestimmt (und die Störgrößen seien unabhängig und normalverteilt mit fester Varianz), d.h. betrachten wir das Modell

$$X_{ijk} = m + \alpha_i + \beta_j + \gamma_{ij} + \sqrt{v}\xi_{ijk}, \quad 1 \leq k \leq \ell,\ 1 \leq i \leq s_1,\ 1 \leq j \leq s_2$$

mit unabhängigen, $\mathcal{N}(0,1)$-verteilten ξ_{ijk}. Hierbei ist $m \in \mathbb{R}$, $\alpha \in \mathbb{R}^{s_1}$ mit $\sum_{i=1}^{s_1} \alpha_i = 0$, $\beta \in \mathbb{R}^{s_2}$ mit $\sum_{j=1}^{s_2} \beta_j = 0$, und $\gamma = (\gamma_{ij}) \in \mathbb{R}^{s_1 \times s_2}$ eine Matrix, in der alle Zeilen- und Spaltensummen $= 0$ sind.
Formulieren Sie dies als normales lineares Modell und finden Sie Teilräume des $\mathbb{R}^{s_1 s_2 \ell}$ und die Projektionen von X auf diese Teilräume, die zu folgenden Hypothesen passen: 1) $\alpha_i \equiv 0$, 2) $\beta_j \equiv 0$, 3) $\gamma_{ij} \equiv 0$. Formulieren Sie entsprechende F-Tests.

b) Es soll untersucht werden, inwieweit ein Medikament in Zusammenwirkung mit Alkohol die Reaktionszeit beeinflusst. Dazu wurden 24 Testpersonen in 6 Gruppen aufgeteilt und folgende Reaktionszeiten gemessen:

	Blutalkoholgehalt (in Promille)		
	0.0	0.5	1.0
mit Medikament	23,21,20,19	22,25,24,25	24,25,22,26
ohne Medikament	22,19,18,20	23,21,24,28	25,28,32,29

Testen Sie anhand dieser Daten (und einer Normalverteilungsannahme) mittels der Methoden aus a) die Hypothesen „das Medikament allein hat keinen Einfluss auf die Reaktionszeit“, „Alkohol allein hat keinen Einfluss auf die Reaktionszeit“ und „es gibt keine Wechselwirkung von Medikament und Alkohol bezüglich der Reaktionszeit“ jeweils zum Niveau $\alpha = 0.05$.

2. Aufgabe (Regression und Prognose) **a)** Seien $X_i = \gamma_0 + \gamma_1 t_i + \sqrt{v}\xi_i$, $i = 1, \dots, n$ mit (unbekannten) $\gamma_0, \gamma_1 \in \mathbb{R}$, $v > 0$, ξ_i u.i.v. gemäß $\mathcal{N}(0,1)$. Seien $\widehat{\gamma}_0, \widehat{\gamma}_1$ die kleinste-Quadrate-Schätzer für γ_0, γ_1, $V^* = (n-2)^{-1} \sum_{i=1}^n (X_i - \widehat{\gamma}_0 - \widehat{\gamma}_1 t_i)^2$ die (empirische) Residuenvarianz.
Nehmen wir an, wir möchten anhand der Beobachtungen von $X_1, \dots, X_n$ den (zufälligen) Wert von X_{n+1} beim Wert t_{n+1} der Regressorvariable vorhersagen. Der offensichtliche Punktschätzer ist $\widehat{\gamma}_0 + \widehat{\gamma}_1 t_{n+1}$. Wir möchten diesen Punktschätzer mit einem Prognoseintervall zum Irrtumsniveau $\alpha \in (0,1)$ ausstatten, d.h. wir suchen $U = U(X_1, \dots, X_n)$ und $O = O(X_1, \dots, X_n)$ so dass

$$\mathbb{P}_{(\gamma_0, \gamma_1, v)}\big(U \leq X_{n+1} \leq O\big) \geq 1 - \alpha$$

für alle γ_0, γ_1, v. Sei $t_{n-2;1-\alpha/2}$ das $(1-\alpha/2)$-Quantil der Student-Verteilung mit $n-2$ Freiheitgraden. Zeigen Sie: Die Wahl

$$\begin{aligned} U &:= \widehat{\gamma}_0 + \widehat{\gamma}_1 t_{n+1} - t_{n-2;1-\alpha/2} \sqrt{V^* \left(1 + \frac{1}{n} + \frac{(t_{n+1} - M(t))^2}{nV(t)}\right)}, \\ O &:= \widehat{\gamma}_0 + \widehat{\gamma}_1 t_{n+1} + t_{n-2;1-\alpha/2} \sqrt{V^* \left(1 + \frac{1}{n} + \frac{(t_{n+1} - M(t))^2}{nV(t)}\right)} \end{aligned}$$

liefert ein Prognoseintervall für X_{n+1} zum Irrtumsniveau α ($M(t) := n^{-1} \sum_{i=1}^n t_i$, $V(t) := n^{-1} \sum_{i=1}^n (t_i - M(t))^2$).
Hinweis: Zeigen Sie, dass $\widehat{\gamma}_0 + \widehat{\gamma}_1 t_{n+1}$ normalverteilt ist mit Mittelwert $\gamma_0 + \gamma_1 t_{n+1}$ und Varianz $v\big(\frac{1}{n} + (t_{n+1} - M(t))^2/(nV(t))\big)$ und unabhängig von V^*. Wie ist dann $X_{n+1} - \widehat{\gamma}_0 - \widehat{\gamma}_1 t_{n+1}$ verteilt, wie

$$\frac{X_{n+1} - \widehat{\gamma}_0 - \widehat{\gamma}_1 t_{n+1}}{\sqrt{V^* \left(1 + \frac{1}{n} + \frac{(t_{n+1} - M(t))^2}{nV(t)}\right)}} \ ?$$

b) Die Datei `800m.dat` enthält die Bestzeiten im 800m-Lauf der Herren für die olympischen Spiele 1920–2004. Finden und plotten Sie die (kleinste-Quadrate-)Regressionsgerade für Jahr gegen Bestzeit. Wie sieht (unter einer Normalverteilungsannahme) Ihre Vorhersage für die Bestzeit 2008 aus? Geben Sie ein Prognoseintervall zum Niveau $\alpha = 0.05$ an.