Fachbereich Mathematik
Prof. Dr. J. Lehn
A. Neuenkirch
B. Niese
A. Rößler

TECHNISCHE UNIVERSITÄT DARMSTADT
SS 2006 17.07.2006

Einführung in die Statistik

13. Übung

Gruppenübungen

Aufgabe G40 (Nichtparametrische Testverfahren)
Der Alkoholspiegel von 12 Blutproben wurde jeweils mit zwei verschiedenen Verfahren gemessen. Dabei erhielt man folgende Promillewerte:

Blutprobe	1	2	3	4	5	6	7	8	9	10	11	12
Verfahren A	0.65	0.48	0.93	0.87	1.12	0.58	0.74	0.62	1.04	0.36	0.72	0.51
Verfahren B	0.60	0.31	1.06	0.75	1.15	0.59	0.82	0.60	1.13	0.50	0.68	0.58

Überprüfen Sie die Hypothese, dass die Verfahren als gleichwertig angesehen werden können, mit einem geeigneten nichtparametrischen Testverfahren zum Niveau $\alpha = 0.05$. Warum ist der Zweistichprobentest von Wilcoxon-Mann-Whitney hier nicht geeignet?

Aufgabe G41 (Nichtparametrische Testverfahren)
Von 20 Schülern wurden 10 ausgelost. Sie erhielten während des ganzen Winters eine Multivitamintablette pro Tag. Die übrigen 10 erhielten täglich eine Placebotablette. Gezählt wurden die Tage der Abwesenheit vom Unterricht mit dem Ergebnis:

Versuchsgruppe	0	1	2	4	7	10	15	17	19	26
Kontrollgruppe	3	8	16	18	20	22	23	25	27	28

Es wird angenommen, dass die angegebenen Messwerte $x_1, \ldots, x_{10}, y_1, \ldots, y_{10}$ eine Realisierung unabhängiger Zufallsvariablen $X_1, \ldots, X_{10}, Y_1, \ldots, Y_{10}$ sind und dass $X_1, \ldots, X_{10}$ bzw. $Y_1, \ldots, Y_{10}$ jeweils die gleiche Verteilungsfunktion F bzw. G besitzen. Prüfen Sie die Hypothese $H_0 : F = G$ zum Niveau $\alpha = 0.05$ mit dem

a) Zweistichprobentest von Wilcoxon-Mann-Whitney,

b) Run-Test von Wald und Wolfowitz.

Die zur Bestimmung des kritischen Bereichs jeweils benötigten Quantile bestimmen Sie näherungsweise mit Hilfe von Normalverteilungsquantilen.

Aufgabe G42 (Nichtparametrische Testverfahren)
Drei Landwirte besitzen je 4 ungefähr gleichalte Schweine einer bestimmten Züchtung. Jeder der drei Landwirte benutzt eine andere Mästmethode. Über einen gewissen Zeitraum hinweg wurden nun die Gewichtszunahmen [in kg] bei allen 12 Schweinen festgestellt. Es kann angenommen werden, dass es sich hierbei um eine Realisierung von unabhängigen Zufallsvariablen $X_i, Y_i, Z_i, i = 1, 2, 3, 4$ handelt. Dabei besitzen für $i = 1, 2, 3, 4$ die Zufallsvariablen X_i ebenso wie die Zufallsvariablen Y_i und Z_i jeweils identische Veteilungen. Prüfen Sie mit dem Kruskal-Wallis-Test zum Niveau $\alpha = 0.1$, ob die Hypothese, dass alle drei Mästmethoden gleichwertig sind, verworfen werden muss.

Messreihe	Gewichtszunahme			
x_i	4	11	12	13
y_i	10	14	17	19
z_i	16	18	20	26

Aufgabe G43 (Einfache lineare Regression)
Eine Wohnung wird mit Erdgas beheizt. Der Gasverbrauch hängt von der Außentemperatur ab. Um den Gasverbrauch vorhersagen zu können, wird während einer Heizperiode von Oktober bis Juni für jeden Monat der durchschnittliche tägliche Gasverbrauch (in dm^3) bestimmt und mit der mittlerern Temperatur (in °C) zu diesem Monat verglichen. Es ergaben sich folgende Zahlenwerte:

Monat	Okt.	Nov.	Dez.	Jan.	Feb.	Mär.	Apr.	Mai	Jun.
Temperatur	9.7	3.4	-2.7	-1.9	-1.4	8.0	9.8	13.9	9.7
Verbrauch	147	173	246	241	249	139	127	71	31

a) Berechnen Sie den empirischen Korrelationskoeffizienten zwischen der mittleren Temperatur und dem mittlerern Gasverbrauch.

b) Schätzen Sie die Parameter der Regressionsgeraden.

c) Im Sommer wird eine Wärmeschutzisolierung eingebaut. Im nächsten Februar wird bei einer durchschnittlichen Monatstemperatur von -3.9 °C ein durchschnittlicher Verbrauch von 246 Kubikdezimeter Gas beobachtet. Bestimmen Sie, unter geeigneten Annahmen, aufgrund der Regressionsgleichung ein Prognoseintervall zum Niveau $1-\alpha = 0.95$ für den durchschnittlichen täglichen Gasverbrauch im Februar (basierend auf den Werten vor der Isolierung). Hat sich die Isolierung rentiert?

Benötigte Quantile : $t_{7;0.975} = 2.365$

Aufgabe G44 (Einfache lineare Regression)
Bei einer Stichprobe von 20 Frauen gleichen Alters wurde die Körpergröße x_i (in cm) und das Gewicht y_i (in kg) gemessen. Es wird im folgenden vorausgesetzt, dass $y_1, \dots, y_{20}$ Realisierungen unabhängiger, normalverteilter Zufallsvariablen $Y_1, \dots, Y_{20}$ mit gleicher Varianz σ^2 sind. Y_i habe den Erwartungswert $ax_i + b$ mit unbekannten Parametern $a, b \in \mathbb{R}$. Es ergab sich

$$\sum_{i=1}^{20} x_i = 3365.0 \quad \sum_{i=1}^{20} y_i = 1232.5 \quad \sum_{i=1}^{20} x_i^2 = 566461.25$$

$$\sum_{i=1}^{20} y_i^2 = 76232.812 \qquad \sum_{i=1}^{20} x_i y_i = 207608.12.$$

a) Bestimmen Sie geeignete Schätzwerte für a, b und σ^2.

b) Überprüfen Sie, ob eine Faustregel der Form

$$\text{Gewicht} = \text{Körpergröße - fester Betrag}$$

unrealistisch ist, indem Sie anhand der obigen Daten die Nullhypothese $H_0 : a = 1$ gegen die Alternative $H_1 : a \neq 1$ auf dem Niveau $\alpha = 0.1$ testen.

c) Bestimmen Sie außerdem einen Schätzer für b nach der Maximum-Likelihood-Methode unter der zusätzlichen Voraussetzung, dass $a = 1$ erfüllt ist.

Benötigte Quantile : $t_{18;0.95} = 1.734$

Aufgabe G45 (Einfache Varianzanalyse)

Nochmal die Situation von G42: Drei Landwirte besitzen je 4 ungefähr gleichalte Schweine einer bestimmten Züchtung. Jeder der drei Landwirte benutzt eine andere Mästmethode. Über einen gewissen Zeitraum hinweg wurden nun die Gewichtszunahmen [in kg] bei allen 12 Schweinen festgestellt. Es kann angenommen werden, dass es sich hierbei um eine Realisierung von unabhängigen Zufallsvariablen $X_i, Y_i, Z_i, i = 1, 2, 3, 4$ handelt. Dabei besitzen für $i = 1, 2, 3, 4$ die Zufallsvariablen X_i ebenso wie die Zufallsvariablen Y_i und Z_i jeweils identische Veteilungen. Prüfen Sie nun mit der einfachen Varianzanalyse zum Niveau $\alpha = 0.1$, ob die Hypothese, dass alle drei Mästmethoden gleichwertig sind, verworfen werden muss. Welche zusätzliche Annahme muss nun im Gegensatz zu der Situation in G38 getroffen werden?

Messreihe	Gewichtszunahme			
x_i	4	11	12	13
y_i	10	14	17	19
z_i	16	18	20	26