Ausgearbeiteter PO für Statistik & Wahrscheinlichkeitstheorie (VO: Prof. K. Felsenstein, WS 2005/06)

schriftliche Prüfung 20. Juni 2006

1. (6 Punkte)

Zu folgenden Daten

25,6 41,2 31,8 7,9 29,1 31,4 35,1 48,7

a) soll ein Boxplot gezeichnet werden. Man berechne alle für den Boxplot benötigten Kenngrößen.

Bildung der Ordnungsstatistik. (Skriptum Seite 7)

i	$x_{(i)}$	$\lvert x_i - x_{med}\rvert$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^4$
1	7,9	23,7	-23,49	551,66	304331,69
2	25,9	5,7	-5,49	30,11	906,77
3	29,1	2,5	-2,29	5,23	27,38
4	31,4	0,2	0,01	0,00	0,00
5	31,8	0,2	0,41	0,17	0,03
6	35,1	3,5	3,71	13,78	189,96
7	41,2	9,6	9,81	96,29	9270,83
8	48,7	17,1	17,31	299,72	89833,67

Ermittlung der Quartile:

$$q_{0.25}=\frac{x_{(2)}+x_{(3)}}{2}=\frac{25{,}9+29{,}1}{2}=27{,}5$$

$$q_{0.5}=\frac{x_{(4)}+x_{(5)}}{2}=\frac{31{,}4+31{,}8}{2}=31{,}6$$

$$q_{0.75}=\frac{x_{(6)}+x_{(7)}}{2}=\frac{35{,}1+41{,}2}{2}=38{,}15$$

Ermittlung des Interquartilen Abstands und der Intervallgrenzen außerhalb der Box:

$$q_{0,75}-q_{0,25}=38{,}15-27{,}5=10{,}65$$

$$1{,}5\cdot(q_{0,75}-q_{0,25})=15{,}975$$

$$[q_{0,25}-15{,}975\,;q_{0,25}]\ [q_{0,75};q_{0,75}+15{,}975]$$

$$[11{,}525\,;27{,}5]\ [38{,}15\,;54{,}125]$$

Zeichenen des Boxplots für die errechneten Werte (Skriptum Seite 9):

b) Man berechne die MAD und die Kurtosis der Stichprobe (Skriptum Seite 10).

$$MAD=\frac{1}{n}\cdot\sum_{i=1}^{n}\lvert x_i-\bar{x}\rvert=7{,}81$$

$$s=\sqrt{\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}\right)}=11{,}9342$$

$$Kurtosis = \left(\frac{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^4}{s^4} \right) - 3 = -0{,}5069$$

2. (6 Punkte)

Die Funktion

$$f(x)=c\cdot e^{-2(x-1)} \quad \text{für} \quad x\geq 1$$

ist die Dichte einer Stochastischen Größe X auf dem Bereich [1,∞)newline

{ } =

i) Man berechne die Konstante c .

Erster Schritt ist die Ermittlung der Verteilungsfunktion durch das Integrieren der Dichtefunktion mit den Grenzen 1, x.

$$F(x)=\int_1^x f(x)dx=\int_1^x c\cdot e^{-2(x-1)}dx$$

$$=c\cdot\int_1^x e^{-2x+2}dx=c\cdot e^2\cdot\int_1^x e^{-2x}dx$$

$$=c\cdot e^2\cdot\left(-\frac{1}{2}\cdot e^{-2x}\right)\Big|_1^x$$

$$=c\cdot e^2\cdot\left(-\frac{1}{2}\cdot e^{-2x}+\frac{1}{2}\cdot e^{-2}\right)$$

$$=c\cdot\left(\frac{1}{2}-\frac{1}{2}\cdot e^{-2x+2}\right)$$

Um c zu ermitteln greift man auf die grundlegenden Eigenschaften der Verteilungsfunktion zurück. (Skriptum Seite 40, Satz 3.3)

$$\lim_{x\to\infty} F(x)=1=c\cdot\left(\frac{1}{2}-\frac{1}{2}\cdot\frac{1}{e^{2\infty+2}}\right)$$

$$1=c\cdot\frac{1}{2}$$

$$c=2$$

ii) Es soll der Erwartungswert und der
iii) Median dieser Verteilung berechnet werden.

Der Erwartungswert ist definiert durch die Formel (wobei E als Ersatz für das übliche Symbol des Erwartungswerts verwendet wird):

$$E(X)=\int_1^\infty x\cdot f(x)dx=c\cdot e^2\cdot\int_1^\infty x\cdot e^{-2x}dx$$

Das Integral lässt sich durch partielle Integration lösen:

$$E(X)=c\cdot e^2\cdot\int_1^\infty x\cdot e^{-2x}\,dx$$

$$=c\cdot e^2\cdot\left(-\frac{1}{2}\cdot e^{-2x}\cdot x+\frac{1}{2}\cdot\int_1^\infty e^{-2x}\,dx\right)$$

$$=c\cdot e^2\cdot-\frac{1}{2}\cdot\left(e^{-2x}\cdot x+\frac{1}{2}e^{-2x}\right)\Big|_1^\infty$$

$$=c\cdot e^2\cdot-\frac{1}{2}\cdot\left(e^{-2x}\cdot\left(x+\frac{1}{2}\right)\right)\Big|_1^\infty$$

$$=c\cdot e^2\cdot-\frac{1}{2}\cdot\left(0-e^{-2}\cdot\left(1+\frac{1}{2}\right)\right)$$

$$=c\cdot\frac{1}{2}\cdot\left(1+\frac{1}{2}\right)=c\cdot\frac{3}{4}$$

$$E(X)=1{,}5$$

$NR:$

$$\int u'\cdot v=u\cdot v-\int u\cdot v'$$

$$u'=e^{-2x}\quad u=-\frac{1}{2}\cdot e^{-2x}$$

$$v=x\quad v'=1$$

Der Median lässt sich über die Verteilungsfunktion ermitteln:

$$F(x_{med})=0{,}5=1-e^{-2x_{med}+2}$$

$$0{,}5=e^{-2x_{med}+2}$$

$$\ln(0{,}5)=\ln(e^{-2x_{med}+2})$$

$$\ln(0{,}5)=-2x_{med}+2$$

$$x_{med}=-\frac{\ln(0{,}5)-2}{2}$$

$$x_{med}=1{,}34657$$

3. (6 Punkte)

Eine Partei befragt ihre Mitglieder über die Zufriedenheit mit der Parteiführung. Das Ergebnis der Befragung war bei Männern und Frauen:

	zufrieden	nicht zufrieden	
Frauen	1174	204	1378
Männer	2608	422	3030
	3782	626	4408

Ist die Zufriedenheit mit der Parteiführung bei beiden Geschlechtern gleich oder bestehen signifikante Unterschiede? ($\alpha=0.1$)

Das Beispiel lässt sich über einen Test mit Kontingenztafeln lösen (Skriptum Seite 102). Über die Randhäufigkeiten und die Summe aller Häufigkeiten wird die Teststatistik der Vierfeldertafel errechnet und mit dem entsprechenden Wert aus der χ^2 -Tabelle verglichen.

$$\hat{\chi}^2=\frac{n\cdot(H_{11}H_{22}-H_{12}H_{21})^2}{H_{1.}H_{2.}H_{.1}H_{.2}}$$

$$\hat{\chi}^2=\frac{4408\cdot(1174\cdot 422-204\cdot 2608)^2}{1378\cdot 3030\cdot 3782\cdot 626}$$

$$\hat{\chi}^2=0{,}59746$$

$$\chi^2_{1;1-\alpha}=\chi^2_{1;0{,}9}=2{,}706$$

Die Unabhängigkeitsannahme (Hyptohese) wir man dann verwerfen wenn gilt, dass $\hat{\chi}^2>\chi^2_{1,1-\alpha}$.
In diesen Beispiel hält die Hypothese, man kann also annehmen, dass die Zufriedenheit bei beiden Geschlechtern sehr ähnlich ist.

4. (6 Punkte)

Für die Umsätze (X) mit Photofilmen (in Mio. €) wird die Trendfunktion für die Zeit t

$$X = a \cdot t^b$$

angenommen. Die Umsätze der letzten Jahre waren

Jahr t	2001	2002	2003	2004	2005
Umsatz X	29,90	25,76	18,86	16,33	14,26

Man prognostiziere den Umsatz mit Photofilmen für heuer.

Über die logarithmischen Rechenregeln kann man die Trendfunktion zu einer linearen Funktion umformen (Skriptum Seite 108).

$$X = a \cdot t^b$$
$$\ln X = \ln(a) + b \ln(t)$$
$$\tilde{X} = \ln(X) \qquad \tilde{t} = \ln(t)$$
$$\theta_1 = \ln(a) \qquad \theta_2 = b$$
$$\tilde{X} = \theta_1 + \theta_2 \tilde{t}$$

Die Regressiongerade wird nun für die transformierten Werte ermittelt.

$\tilde{t}_i = \ln(t_i)$	7,60140	7,60190	7,60240	7,60290	7,60340
$\tilde{x}_i = \ln(x_i)$	3,39786	3,24882	2,93704	2,79300	2,65746

Die Parameter der Regressionsgeraden werden über die folgenden Formeln berechnet:

$$\hat{\theta}_1 = \frac{\bar{\tilde{x}}_n \sum_{i=1}^{n} (\tilde{z}_i^2) - \bar{\tilde{z}}_n \sum_{i=1}^{n} \tilde{x}_i \tilde{z}_i}{\sum_{i=1}^{n} (\tilde{z}_i^2) - n \bar{\tilde{z}}_n^2}$$

$$\hat{\theta}_1 = 2952{,}08761$$

$$\hat{\theta}_2 = \frac{\sum_{i=1}^{n} \tilde{x}_i \tilde{z}_i - \bar{\tilde{z}}_n \sum_{i=1}^{n} \tilde{x}_i \tilde{z}_i}{\sum_{i=1}^{n} (\tilde{z}_i^2) - n \bar{\tilde{z}}_n^2}$$

$$\hat{\theta}_2 = -387{,}91436$$

Für die Prognose wird der Eingansparameter transformiert, über Geradengleichung die Prognose errechnet und rücktransformiert.

$$
\begin{aligned}
&t = 2006 \\
&\tilde{t} = \ln(t) \\
&\tilde{t} = 7{,}60389 \\
&\hat{X} = \hat{\theta}_1 + \hat{\theta}_2 \cdot \tilde{t} \\
&\hat{X} = 2952{,}08761 - 387{,}91439 \cdot 7{,}60389 \\
&\hat{X} = 2{,}426168 \\
&X = e^{\hat{X}} \\
&X = 11{,}31543
\end{aligned}
$$

Das Ergebnis hält sich in etwa an den Abwärtstrend der in der Tabelle ersichtlich ist.