Klausur Statistik
Lösungshinweise

Prüfungsdatum: 5. Juli 2017 – Prüfer: Etschberger, Ivanov, Jansen, Wesp, Wins
Studiengang: IM, BW, Inf und W-Inf
Punkte: 18, 16, 14, 10, 9, 15, 8 ;   Summe der Punkte: 90

Aufgabe 1 — 18 Punkte

Eine Studie bei 100 Absolventen einer Hochschule ergab für das Merkmal $X \mathrel{\hat{=}}$ „Studiendauer in Semestern“ für die Ausprägungen a_i, die absoluten Häufigkeiten h_i, die kumulierten relativen Häufigkeiten u_i sowie dem kumulierten Merkmalssummenanteil v_i mit

$$u_i = \sum_{k=1}^{i} f_k\,, \qquad v_i = \frac{\sum_{k=1}^{i} a_k \cdot h_k}{\sum_{k=1}^{n} a_k \cdot h_k}$$

folgende unvollständige Tabelle:

i	1	2	3	4	5	6	7
a_i	4	6	7	**8**	10	20	40
h_i	10	10	20	**20**	28	8	4
u_i	0.10	0.20	0.40	**0.60**	0.88	**0.96**	**1.00**
v_i	0.04	0.10	0.24	**0.40**	0.68	**0.84**	**1.00**

Außerdem ist das arithmetische Mittel des Merkmals mit $\overline{x} = 10$ bekannt.

a) Füllen Sie die fehlenden Felder in der Tabelle aus.

(*Hinweis: Wenn Sie die Teilaufgabe a) nicht lösen können, rechnen Sie mit den (falschen) Werten* $a_4 = 9.5$, $h_4 = 37$ *weiter*)

b) Bestimmen Sie den Median, das empirische 25 % sowie das 75 % Quantil des Merkmals.

c) Zeichnen Sie einen Boxplot zu diesen Daten.

d) Nehmen Sie an, die Daten des Merkmals X sind im R-Vektor x gespeichert. Schreiben Sie in das folgende Kästchen R-Befehle, mit denen Sie die Teilaufgaben b) und c) lösen können:

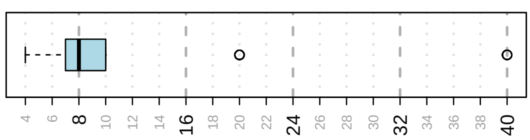
```
quantile(x, probs = c(0.25, 0.5, 0.75), type=2)
boxplot(x)
```

Lösungshinweis:

a) S.o.

b)
```
quantile(x, probs = c(0.25, 0.5, 0.75),
         type=3)
## 25% 50% 75%
##   7   8  10
```

c)
```
boxplot(x)
```



d) s.o.

Aufgabe 2 **16 Punkte**

Für 180 Zugverbindungen wurden zwei Merkmale erhoben:

- der Wochentag der Abfahrt als „*Tag*" mit den beiden Ausprägungen *Montag - Freitag* bzw. *Samstag oder Sonntag*
- die Differenz zur geplanten Abfahrtszeit als „*Verspätung*"

Hieraus ergab sich folgende unvollständige Kontingenztabelle:

Tag	Verspätung			Summe
	pünktlich	1-15 Min.	> 15 Min.	
Montag - Freitag	58	**46**	**16**	**120**
Samstag, Sonntag	**32**	**14**	14	**60**
Summe	**90**	60	**30**	**180**

Es sei zudem bekannt, dass insgesamt $^2/_3$ der Züge an Arbeitstagen (Mo-Fr) abgefahren sind und zudem 50 % aller Züge pünktlich abgefahren sind.

a) Ergänzen Sie die obige Kontingenztabelle sowie deren Randhäufigkeiten.

R b) In R seien mit den beiden Merkmalsvektoren T (Tag) und V (Verspätung) die 180 individuellen Ausprägungen gegeben. Geben Sie R-Befehle an, die die Kontingenztabelle und deren Randhäufigkeiten ausgeben.

c) Zeigen Sie – ohne die Bestimmung eines Zusammenhangsmaßes – dass die beiden Merkmale *Tag* und *Verspätung* nicht unabhängig sind.

d) Aus der Kontingenztabelle ergibt sich $\chi^2 = 5.3$ (Diesen Wert müssen Sie *nicht* nachrechnen). Bestimmen Sie damit den normierten Kontingenzkoeffizienten und interpretieren Sie diesen.

In einer anderen Untersuchung wurde bei 1000 Zugverbindungen die Verspätung zur geplanten Abfahrtszeit (Merkmal: *Abfahrt* [in Minuten]) sowie die Verspätung zur geplanten Ankunftszeit an der Endstation (Merkmal: *Endstation* [in Minuten]) untersucht. Es wird ein linearer Zusammenhang vermutet, der mittels Regression untersucht werden soll. Die erhobenen Daten seien in einem R-Dataframe mit dem Bezeichner *Verspaetungen* gespeichert. Für das Regressionsmodell ergibt sich demnach:

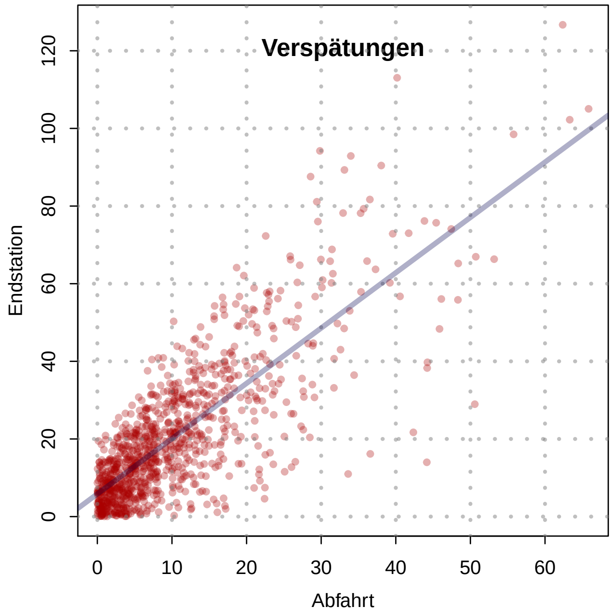
```
lm(Endstation ~ Abfahrt, data=Verspaetungen)

##
## Call:
## lm(formula = Endstation ~ Abfahrt, data = Verspaetungen)
##
## Coefficients:
## (Intercept)      Abfahrt
##       5.840        1.425
```

e) Geben Sie den Funktionsterm der Regressionsgerade an und zeichnen Sie diese in den nebenstehenden Streuplot der Daten ein.

f) Ein Zug fährt mit 5 Minuten Verspätung ab. Wieviel würde sich gemäß dem Regressionsmodell die Ankunftszeit verspäten?

g) Geben Sie R-Befehle an, mit denen Sie ein geeignetes Streudiagramm mit der farbig eingezeichneten Regressionsgeraden erzeugen können.



Lösungshinweis:

a) siehe oben

b)
```
Tabelle = table(T, V)
addmargins(Tabelle)

##        V
## T       puenktlich [1; 15] >15 Sum
##   Mo-Fr         58      46  16 120
##   Sa,So         32      14  14  60
##   Sum           90      60  30 180
```

c) Betrachte z.B. Mo-Fr/pünktlich: $120 \cdot 90/180 = 60 \neq 58$

d) $K = \sqrt{\dfrac{\chi^2}{\chi^2 + n}} = \sqrt{\dfrac{5.3}{5.3 + 180}} = 0.169122$. $K_{\max} = \sqrt{\dfrac{2-1}{2}} = \sqrt{0.5}$.

e) $\hat{y} = 5.8399929 + 1.4245971 \cdot x$

f) $\hat{y}(5) = 12.9629785$

g)
```
V.Reg = lm(Endstation ~ Abfahrt, data=Verspaetungen)
plot(Abfahrt, Endstation)
abline(V.Reg, col="blue")
```

Aufgabe 3 **14 Punkte**

Leonard, Sheldon, Howard und Raj spielen ein Brettspiel. Dazu haben sie insgesamt fünf faire Würfel in Form der fünf platonischen Körper:

- Würfel A ist ein 4-seitiger mit den Zahlen 1 bis 4,
- Würfel B ein 6-seitiger mit den Zahlen 1 bis 6,
- Würfel C ein 8-seitiger mit den Zahlen 1 bis 8,
- Würfel D ein 12-seitiger mit den Zahlen 0 bis 9 sowie je einer weiteren Seite mit den Zahlen 1 und 2,
- Würfel E ein 20-seitiger Würfel, der zweimal die Zahlen 0 bis 9 hat.

a) Leonard würfelt die Würfel in der Reihenfolge von A bis E. Wie viele verschiedene Ausgänge gibt es?

b) Raj würfelt alle Würfel gleichzeitig. Wie groß ist die Wahrscheinlichkeit dafür, dass alle Würfel eine 1 zeigen?

c) Sheldon bekommt zufällig zwei der drei Würfel A, B und C und würfelt diese. Wie groß ist die Wahrscheinlichkeit, dass beide Würfel eine 1 zeigen?

R d) Howard wirft 7 mal Würfel C. Er interessiert sich für die Wahrscheinlichkeit p, dass er mindestens 3 mal eine Zahl wirft, die größer als 4 ist. Geben Sie einen R-Ausdruck an, mit dem p berechnet werden kann.
(Hinweis: Die Wahrscheinlichkeit müssen Sie nicht berechnen.)

Lösungshinweis:

a) $4 \cdot 6 \cdot 8 \cdot 10 \cdot 10 = 19\,200$

b) Man multipliziert die Wahrscheinlichkeiten aller Würfel für eine 1. Es ergibt sich:

$$P = \frac{1}{4} \cdot \frac{1}{6} \cdot \frac{1}{8} \cdot \frac{2}{12} \cdot \frac{2}{20} = \frac{4}{46\,080} = \frac{1}{11\,520} \approx 8{,}68055 \cdot 10^{-5}$$

c) Die Kombinationen AB, AC oder BC werden mit je $1/3$ Wahrscheinlichkeit ausgewählt.

$$P = \frac{1}{3}\left(\frac{1}{4} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{8} + \frac{1}{6} \cdot \frac{1}{8}\right) = \frac{1}{32} = 0{,}03125$$

d) $X =$ „Anzahl größer als 4“; $X \sim B(n = 7, p = 0.5)$

```
1 - pbinom(q = 2, size = 7, prob = 0.5)
## [1] 0.7734375
```

Aufgabe 4 **10 Punkte**

Gegeben sei eine Zufallsvariable X mit $X \sim B(3000; 0.001)$.

a) Berechnen Sie
   (1) $P(X \geqq 3 | X \leqq 1)$
   (2) $P(X \geqq 1 | X \leqq 3)$
   (3) $P(X \leqq 1 | X \leqq 3)$
   (4) $P(X \leqq 3 | X \leqq 1)$

b) Geben Sie eine R-Zeile an, der die Lösung von Teilaufgabe a), (2) ausgibt.

Lösungshinweis:

a) (1) $P(X \geqq 3 | X \leqq 1) = 0$

   (2) $P(X \geqq 1 | X \leqq 3) = \dfrac{P(X \in \{1, 2, 3\})}{P(X \leqq 3)} = \dfrac{0.5975195}{0.6472319} = 0.9231923$

   (3) $P(X \leqq 1 | X \leqq 3) = \dfrac{0.1989989}{0.6472319} = 0.3074615$

   (4) $P(X \leqq 3 | X \leqq 1) = 1$

b)
```r
(sum(dbinom(x=1:3, p=0.001, size=3000)))/pbinom(3, p=0.001, size=3000)

## [1] 0.9231923
```

Aufgabe 5 **9 Punkte**

Für eine einfache Stichprobe X_1, X_2 mit $\mathrm{E}[X_i] = \mu$ und $\mathrm{Var}[X_i] = \sigma^2$ (für $i = 1, 2$) sollen Stichprobenfunktionen zum Schätzen des arithmetischen Mittels der Grundgesamtheit untersucht werden. Zur Diskussion stehen drei Funktionen $\hat{\Theta}_1, \hat{\Theta}_2, \hat{\Theta}_3$:

$$\begin{aligned}
\hat{\Theta}_1 &= 0.5 \cdot X_1 + 0.5 \cdot X_2 \\
\hat{\Theta}_2 &= 0.4 \cdot X_1 + 0.6 \cdot X_2 \\
\hat{\Theta}_3 &= a_1 \cdot X_1 + a_2 \cdot X_2 \quad (a_1, a_2 \geqq 0)
\end{aligned}$$

a) Zeigen Sie, dass $\hat{\Theta}_1$ und $\hat{\Theta}_2$ erwartungstreue Schätzer für μ sind.

b) Welche Schätzfunktion ist wirksamer: $\hat{\Theta}_1$ oder $\hat{\Theta}_2$?

c) Zeigen Sie, dass $\hat{\Theta}_3$ genau dann erwartungstreu ist, wenn $a_1 + a_2 = 1$ gilt.

d) Für welche Werte von a_1, a_2 ist $\hat{\Theta}_3$ am wirksamsten?

Lösungshinweis:

a) $\mathrm{E}[\hat{\Theta}_1] = \mathrm{E}[0.5 \cdot X_1 + 0.5 \cdot X_2] = 0.5 \cdot \mathrm{E}[X_1] + 0.5 \cdot \mathrm{E}[X_2] = 0.5 \cdot \mu + 0.5 \cdot \mu = \mu$

$\mathrm{E}[\hat{\Theta}_2] = \mathrm{E}[0.4 \cdot X_1 + 0.6 \cdot X_2] = 0.4 \cdot \mathrm{E}[X_1] + 0.6 \cdot \mathrm{E}[X_2] = 0.4 \cdot \mu + 0.6 \cdot \mu = \mu$

b) $$\begin{aligned}
\mathrm{Var}[\hat{\Theta}_1] &= \mathrm{Var}[0.5 \cdot X_1 + 0.5 \cdot X_2] = 0.5^2 \cdot \mathrm{Var}[X_1] + 0.5^2 \cdot \mathrm{Var}[X_2] \\
&= 0.25 \cdot \sigma^2 + 0.25 \cdot \sigma^2 = 0.5 \cdot \sigma^2 \\
\mathrm{Var}[\hat{\Theta}_2] &= \mathrm{Var}[0.4 \cdot X_1 + 0.6 \cdot X_2] = 0.4^2 \cdot \mathrm{Var}[X_1] + 0.6^2 \cdot \mathrm{Var}[X_2] \\
&= 0.16 \cdot \sigma^2 + 0.36 \cdot \sigma^2 = 0.52 \cdot \sigma^2
\end{aligned}$$

Damit ist $\hat{\Theta}_1$ wirksamer als $\hat{\Theta}_2$, da $\mathrm{Var}[\hat{\Theta}_1] < \mathrm{Var}[\hat{\Theta}_2]$.

c) $\hat{\Theta}_3$ ist erwartungstreu

$$\begin{aligned}
\Leftrightarrow \mathrm{E}[\hat{\Theta}_1] &= a_1 \cdot \mathrm{E}[X_1] + a_2 \cdot \mathrm{E}[X_2] \\
&= a_1 \cdot \mu + a_2 \cdot \mu \cdot \mu \\
&= (a_1 + a_2) \cdot \mu = \mu \\
\Leftrightarrow a_1 + a_2 &= 1
\end{aligned}$$

d) $\mathrm{Var}[\hat{\Theta}_3] = a_1^2 \cdot \mathrm{Var}[X_1] + a_2^2 \cdot \mathrm{Var}[X_2] = \left(a_1^2 + a_2^2\right) \cdot \sigma^2$

$\hat{\Theta}_3$ ist erwartungstreu, wenn $a_1 + a_2 = 1 \;\Leftrightarrow\; a_2 = 1 - a_1$ und am wirksamsten, wenn

$$a_1^2 + a_2^2 = a_1^2 + (1 - a_1)^2 = a_1^2 + 1 - 2a_1 + a_1^2 = 2a_1^2 - 2a_1 + 1$$

minimal, also die Ableitung der Varianz V nach a_1

$$V'(a_1) = 4a_1 - 2a_1 = 0 \quad \Leftrightarrow \quad a_1 = 0.5$$

Das ist ein globales Minimum. Damit ist $\hat{\Theta}_3$ am wirksamsten, wenn $a_1 = a_2 = 0.5$.

Aufgabe 6 **15 Punkte**

Eine Nagelfabrik möchte die Produktionsgenauigkeit einer Charge mit einer Solllänge von 80 mm überprüfen. Dazu werden 10 Stück Nägel aus der Produktionsanlage entnommen und deren Länge gemessen. Es ergibt sich:

Stichprobenelement Nr.	1	2	3	4	5
Länge des Nagels [in mm]	80.0	80.4	79.9	81.1	80.5
Stichprobenelement Nr.	6	7	8	9	10
Länge des Nagels [in mm]	79.9	80.6	80.7	80.6	80.2

Gehen Sie im Folgenden davon aus, dass

- die Länge der vermessenen Nägeln eine einfache Stichprobe aus einer normalverteilten Grundgesamtheit aller hergestellten Nägeln darstellen und
- die Urliste der Stichprobe in R in der Variable x gespeichert ist.

a) Bestimmen Sie ein Konfidenzintervall für die Länge aller Nägeln in der Produktion zu einem Konfidenzniveau von 95 %.

R b) Schreiben Sie in folgendes Kästchen *einen* R-Befehl, mit dem man das Konfidenzintervall aus Teilaufgabe a) berechnen kann.

```
t.test(x, conf.level = 0.95)
```

c) Wie müsste die Nullhypothese H_0 und die Gegenhypothese H_1 lauten, wenn ein Kunde der Firma mit einem Test statistisch nachweisen möchte, dass die durchschnittliche Nagellänge in der Grundgesamtheit einer Lieferung von 1 Mio. Nägeln höher als 80 cm ist?

d) Führen Sie den Test mittels der obigen Stichprobe zu einem Signifikanzniveau von $\alpha = 1\,\%$ durch.

R e) Schreiben Sie in folgendes Kästchen *einen* R-Befehl, mit dem man den Test aus d) durchführen kann.

```
t.test(x, mu=80, alternative = “greater“)
```

Lösungshinweis:

a) $c = x_{0.975} = 2.262, \overline{x} = 80.39, s = 0.39 \quad \Rightarrow \quad \left[\overline{x} \pm \frac{sc}{\sqrt{n}}\right] = [80.111; 80.669]$

b) siehe oben

c) H_0: Arithm. Mittel der Nagellänge in GG ist 80 mm; H_1: ... ist größer 80 mm

d) $\overline{x} = 80.39,\ s = 0.39,\ t(9) : x_{1-\alpha} = x_{0.99} = 2.821,\ v = \frac{\overline{x} - 80}{s} \cdot \sqrt{10} \approx 3.162.$
$B = (2.821; \infty)$

e) siehe oben

Aufgabe 7 **8 Punkte**

Gegeben ist für $D = \{(x,y) \in \mathbb{R}^2 : x,y > 0\}$ die Funktion $g\colon D \to \mathbb{R}$ mit

$$g(x,y) = x + 4y + \frac{2}{xy}.$$

a) Berechnen Sie die partielle Ableitung $g_x(x,y)$ und tragen Sie das Ergebnis in das leere Feld des Gradienten von g ein:

$$\nabla g(x,y) = \begin{pmatrix} \boxed{1 - \dfrac{2}{x^2 y}} \\ 4 - \dfrac{2}{xy^2} \end{pmatrix}$$

Hinweis: Sollten Sie Teilaufgabe a) nicht gelöst haben, benutzen Sie bitte das Ergebnis

$$g_x(x,y) = 2y\sqrt{x} - 4 \cdot \sqrt{x^{-3}}\,.$$

b) Geben Sie sämtliche kritische Punkte der Funktion g an.

c) Ergänzen Sie die fehlenden Einträge der Hesse-Matrix H_g zu g:

$$H_g(x,y) = \begin{pmatrix} 4\left(yx^3\right)^{-1} & 2(xy)^{-2} \\ \boxed{2(xy)^{-2}} & \boxed{4\left(xy^3\right)^{-1}} \end{pmatrix}$$

d) Setzen Sie alle kritischen Punkte aus Teilaufgabe b) in H_g ein und ermitteln Sie so Lage und Art aller Extremwerte von g.

Lösungshinweis:

a) Siehe oben

b) $x = 2$, $y = 1/2$

c) siehe oben

d) $H(2, 1/2) = \begin{pmatrix} 1 & 2 \\ 2 & 16 \end{pmatrix} \Rightarrow 1 \cdot 16 - 2 \cdot 2 > 0$ und $1 > 0$, also ist der Punkt $(x,y) = (2, 0.5)$ ein Minimum.