Statistik für Informatiker **SS 2017**

Prof. Dr. Matthias Birkner
M.Sc. Frederik Klement

Extrablatt:Lösung

Falls Sie einen Fehler entdeckt haben oder eine Verständnisfrage zur Lösung besitzen, schreiben Sie mir bitte eine Mail an *fklement@uni-mainz.de*.

Aufgabe 1

- $A_1 \cap A_{10}$
- $A_2^{\complement} \cup A_3^{\complement}$
- $A_1 \cap A_2 \cap A_3^{\complement} \cap A_4 \cap A_5^{\complement}$.
- $(A_1 \cap A_2^{\complement} \cap A_3 \cap A_4 \cap A_5^{\complement})^{\complement}$
- $\bigcap_{i=1}^{10} A_i^{\complement}$
- $\bigcup_{i=1}^{10} A_i$
- $\bigcap_{i=1}^{6} A_i^{\complement} \cap A_7$
- $\bigcap_{i=1}^{5} A_{2i-1} \cup \bigcap_{i=1}^{5} A_{2i}^{\complement}$

Aufgabe 2

a) i)

$$\begin{aligned} P(X_1 + X_2 = 4) =& P(X_1 = 1, X_2) \\ &+ P(X_2 = 2, X_1 = 2) \\ &+ P(X_2, X_1 = 3) \\ &= \frac{3}{36} = \frac{1}{12} \end{aligned}$$

ii) $P(Y = 8) = \frac{5}{36}$

iii) $P(Y = 10) = \frac{1}{12}$

b) $P(Y \text{ ist ungerade })$

c) $P(Y \text{ ungerade}, X_1 = 3) = P(X_1 = 3, X_2 \in \{2, 4, 6\}) = \frac{1}{12}$

d) $P(Y \text{ ungerade}, X_1 = 3) = \frac{1}{12}$, $P(X_1 = 3)P(Y \text{ ungerade}) = \frac{1}{6}\frac{1}{2} = \frac{1}{12}$. Somit unabhängig.

e) $E[X_1 + X_2] = E[X_1] + E[X_2] = 7$.
$Var[X_1 + X_2] = 2Var[X_1] = 2(E[X_1^2] - E[X_1]^2) = 5.8\bar{3}$.

Aufgabe 3

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[\mathbf{1}_{U_i=i}] = \sum_{i=1}^{n} P[U_i = i] = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

$$\begin{aligned}\mathbf{E}[X^2] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{E}[\mathbf{1}_{U_i=i}\mathbf{1}_{U_j=j}] &= \sum_{i\neq j} \mathbf{E}[\mathbf{1}_{U_i=i}\mathbf{1}_{U_j=j}] + \sum_{i=1}^{n} \mathbf{E}[\mathbf{1}_{U_i=i}] \\ &= \sum_{i\neq j} P[U_i = i, U_j = j] + \sum_{i=1}^{n} P[U_i = i] \\ = \sum_{i\neq j} \frac{1}{n(n-1)} + \sum_{i=1}^{n} \frac{1}{n} &= (n^2 - n)\frac{1}{n(n-1)} + 1 = 2\end{aligned}$$

Damit $Var[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = 1$.

Aufgabe 4:

a) Es können soviele unterschiedliche Dreieckskonstellationen gebildet werden, wie es 3-elementige Untermengen von $\{1, 2, ..., n\}$ gibt, also $\binom{n}{3}$.

b) Wir schreiben $A_{\{i,j\}}$ für das Ereignis, dass Spieler i und j eine Verbindung eingehen. Dann

$$P(A_{\{1,2\}} \cap A_{\{2,3\}} \cap A_{\{1,3\}}) \overset{unab.}{=} P(A_{\{1,2\}})^3 = p^3.$$

c) Sei D die Menge aller möglichen Dreieckskonstellationen und sei B_w für $w := \{i, j, k\} \in D$ das Ereignis, dass i, j, k eine Dreieckskonstellation eingehen.

$$E[N] = \sum_{w\in D} E[I_{B_w}] \overset{b)}{=} \sum_{w\in D} p^3 \overset{a)}{=} \binom{n}{3} p^3.$$

Aufgabe 5:
Wir denken uns, dass die Plätze in der Abordnung für Auschuss i die Nummern $(i-1)6+1, ..., 6i$ besitzen. Wenn nun X_i die Platznummer der i-ten Person ist, dann ist der Wertebereich von $X = (X_1, X_2, ..., X_{30})$ gegeben durch

$$S := \{x \in \{1, 2, ..., 30\}^{30} : x_i \neq x_j \text{ für } i \neq j\}$$

Es gibt 30! verschiedene Möglichkeiten die 30 Personen auf die 5 Abordnungen aufzuteilen, denn $|S| = 30!$. Somit besitzt eine bestimmte Belegung die Wahrscheinlichkeit $\frac{1}{30!}$. Im folgenden Zählen wir wieviele günstige Belegungen es jeweils gibt.

a) Es gibt $\binom{11}{3}\binom{19}{3}$ Möglichkeiten 3 Männer und 3 Frauen auszuwählen. Sie sechs Plätze des Finanzausschüsse können mit diesen auf 6! Wege belegt werden. Die restlichen Abgeordneten können auf $(30-6)!$ Möglichkeiten angeordnet werden. Also

$$P[A_1] = \frac{\binom{11}{3}\binom{19}{3}6!(30-6)!}{30!} = \frac{\binom{11}{3}\binom{19}{3}}{\binom{30}{6}}$$

b) Da die Fraktion nur 11 Frauen besitzt, können nur maximal 3 Abordnungen paritätisch besetzt sein. Es gibt $\binom{5}{3}$ Möglichkeiten drei paritätisch zu besetzende Abordnungen auszuwählen. Es gibt $\binom{11}{3}\binom{19}{3}6!$ Möglichkeiten die ersten Abordnungen, $\binom{8}{3}\binom{16}{3}6!$ Möglichkeiten die zweiten Abordnungen und $\binom{5}{3}\binom{13}{3}6!$ Möglichkeiten die dritte Abordnungen zu besetzen. Die restlichen Abgeordneten werden auf $(30-18)!$ verschiedenen Möglichkeiten angeordnet.

$$P[A_2] := \binom{5}{3}\frac{\binom{11}{3}\binom{19}{3}6!\binom{8}{3}\binom{16}{3}6!\binom{5}{3}\binom{13}{3}6!(30-18)!}{30!}$$
$$= \binom{5}{3}\frac{\binom{11}{3,3,3,2}\binom{19}{3,3,3,10}}{\binom{30!}{6,6,6,12}}.$$

c) $A_3^{\complement} :=$ {„Es gibt mindestens einen Abordnungen der nur mit Frauen belegt ist.“} Es gibt nur 11 Frauen, also maximal einen Ausschuss der nur aus Frauen besteht. Sei $B_i :=$ { „Ausschuss i ist nur mit Frauen belegt.“} ($B_i \cap B_j = \emptyset$ für $i \neq j$). Für $P[B_i]$, es gibt $\binom{11}{6}$ Möglichkeiten sechs Frauen auszuwählen und 6! Möglichkeiten mit diesen den Abordnung zu belegen. Für die Anordnung der restlichen Abgeordneten gibt es 24! Möglichkeiten, daher

$$P[B_i] = \frac{\binom{11}{6}6!24!}{30!} = \frac{\binom{11}{6}}{\binom{30}{24}}.$$

$P[A_3] = P[A_3^{\complement}] = 1 - P[\cup_{i\in\{1,2,3,4,5\}} B_i] = 1 - \sum_{i=1}^{5} P[B_i] = 1 - 5\frac{\binom{11}{6}}{\binom{30}{24}}$.

Aufgabe 6

a) Die Verteilungsfunktion F_X von X ist gegeben durch $F_X(x) = (1 - e^{-\lambda x})\mathbf{1}_{[0,\infty)}(x)$. Die ZV Y kann nur Werte aus $[1,\infty)$ annehmen, daher $F_Y(y) = 0$ für $y < 1$. Für $y \geq 1$ gilt:

$$F_Y(y) := P(Y \leq y) = P(e^{\frac{X}{2}} \leq y) = P(X \leq 2ln(y)) \overset{ln(y)\geq 0}{=} (1 - y^{-2\lambda}).$$

Zusammen ergibt dies $(1 - y^{-2\lambda})\mathbf{1}_{[1,\infty)}(y)$.

b) Da F_Y stetig ist und wir sofort sehen, dass diese bis auf die Stelle $y = 1$ eine differenzierbare Funktion ist, erhalten wir die Dichte auf $y \neq 1$ durch differenzieren (für $y = 1$ können wir einen beliebigen Wert setzen, da dies den Wert des Integrals $\int_a^b f_y(y)dy$ nicht verändert.) Somit $f_Y(y) = 2\lambda y^{-2\lambda-1}\mathbf{1}_{[1,\infty)}(y)$.

c)
$$\mathbf{E}[Y] = \int_0^\infty e^{\frac{x}{2}} e^{-\lambda x}\lambda dx = \int_0^\infty \lambda e^{(\frac{1}{2}-\lambda)x}dx. = \frac{\lambda}{\lambda - \frac{1}{2}}.$$

Alternative Rechnung:

$$\mathbf{E}[Y] = \int_1^\infty y \cdot 2\lambda y^{-2\lambda-1}dy = \int_1^\infty 2\lambda y^{-2\lambda}dy = \frac{2\lambda}{-2\lambda+1}[y^{-2\lambda+1}]_1^\infty = \frac{\lambda}{\lambda - \frac{1}{2}}.$$

Aufgabe 7

a) Für die Verteilungsfunktion von $|X|$ gilt $F_{|X|}(x) = 0$ für $x < 0$ und für $x > 0$:

$$F_{|X|}(x) = P(|X| \leq x) = P(X \in [0,x)) + P(X \in (-x,0)) =$$
$$F_X(x) - F_X(0) + F_X(0) - F_X(-x) = F_X(x) - F_X(-x).$$

Durchs Ableiten erhalten wir die Dichte $\tilde{\phi}(x) = 2\phi(x)\mathbf{1}_{(0,\infty)}$, wobei wir $\tilde{\phi}(0) = 0$ gesetzt haben (Dies hat keinen Effekt auf das Integral von $\tilde{phi}$, daher ist dies okay).

b) Es sei $f(x) = x^2$ und es sei

$$g(x) := f^{-1}(x) = \sqrt{x}; \quad g'(x) = \frac{1}{2}x^{-\frac{1}{2}}.$$

Da f auf ganz $\mathbb{R}$ kein Diffeomorphismus ist aber auf $(0, \infty)$, wenden wir die Transformationsformel auf $|X|$ statt auf X an. Somit ist die Dichte ψ von X^2 auf gegeben durch

$$\psi(x) = g'(x)\tilde{\phi}(g(x)) = \frac{1}{2}x^{-\frac{1}{2}}\frac{2}{\sqrt{2\pi}}e^{-\frac{x}{2}}\mathbf{1}_{[0,\infty)}(x) = \frac{e^{-\frac{x}{2}}}{\sqrt{2\pi x}}\mathbf{1}_{[0,\infty)}(x).$$

Aufgabe 8:
Die Chebyshev-Ungleichung gibt uns für $p \in (0, 1)$:

$$P(|\hat{p} - p| \geq \delta) \leq \frac{Var(p)}{\delta^2} = \frac{\frac{1}{n^2}n(1-p)p}{\delta^2} = \frac{(1-p)p}{n\delta^2}$$

Der Ausdruck $p(1 - p)$ nimmt für $p = \frac{1}{2}$ sein Maximum an, welches $\frac{1}{4}$ ist, somit

$$P(|\hat{p} - p| \geq \delta) \leq \frac{1}{4n\delta^2}.$$

Wir suchen basierend auf unserer Ungleichung das kleinste n, so dass

$$P(|\hat{p} - p| \geq \frac{1}{10}) \leq \frac{1}{100},$$

Mit unserer Ungleicung ergibt sich

$$\frac{1}{4n}\left(\frac{1}{10}\right)^{-1} \leq \frac{1}{100} \Leftrightarrow \frac{1}{4}10^4 \leq n$$

Also $n = 2500$.

Aufgabe 9:

a) Sei U :=„Katze überlebt. “, F := „Katze wurde gefüttert.“
$P[U] = P[U|F]P[F] + P[U|F^\complement]P[F^\complement] = \frac{99}{100}\frac{11}{12} + \frac{3}{100}\frac{1}{12} = \frac{1089+3}{1200} = \frac{1092}{1200}$.
Somit $P[U^\complement] = 1 - P[U] = \frac{108}{1200}$

b) $P[F^\complement|U^\complement] = P[F^\complement \cap U^\complement]/P[U^\complement] = \frac{P[U^\complement|F^\complement]P[F^\complement]}{P[U^\complement]} = \frac{1200}{108}\frac{97}{100}\frac{1}{12} = \frac{97}{108}$.

Aufgabe 10:

a)

$$\rho(x, \theta) = \prod_{i=1}^{2n+1} e^{-2|x_i - \theta|} = e^{-2\sum_{i=1}^{2n+1}|x_i-\theta|}.$$

b) Für den Maximum-Likelihood-Schätzer müssen wir das Maximum von ρ finden. Da der Logarithmus eine strikt steigende Funktion ist, nimmt ρ sein Maximum an der Stelle an, an der auch $\log(\rho) = -2\sum_{i=1}^{2n+1}|x_i - \theta|$ sein Maximum annimmt (also das Minimum von $\sum_{i=1}^{2n+1}|x_i - \theta|$). Wir zeigen nun, dass sich das Maximum beim Median x_{med} befindet. Sei dazu $x_1 \leq x_2 \leq ... \leq x_{2n+1}$, dann $x_{med} = x_{n+1}$.

Für $k \in \{1, 2, ..., n\}$ betrachten wir den k-ten Wert von links und den k-ten Wert von rechts, dann gilt für alle $\theta \in \mathbb{R}$

$$|\theta - x_k| + |\theta - x_{2n+1-k}| \geq x_{2n+1-k} - x_k$$

und für x_{med} gilt: $|x_{med} - x_k| + |x_{med} - x_{2n+1-k}| = x_{2n-k} - x_k$. Somit gilt für alle $\theta \in \mathbb{R}$:

$$\sum_{i=1}^{2n+1} |x_i - \theta| \geq \sum_{k=1}^{n} x_{2n+1-k} - x_k + |x_{med} - \theta| \geq \sum_{k=1}^{n} x_{2n+1-k} - x_k = \sum_{i=1}^{2n+1} |x_i - x_{med}|.$$

Somit ist der Maximum-Liklihood-Schätzer x_{med}.

Aufgabe 11:

Es seien $x_1, x_2, ..., x_n$ unabhängige Stichproben einer Bernoulliverteilung mit unbekannten Parameter p und sei $s := \sum_{i=1}^{n} x_i$. Wir nehmen an, dass der Parameter p als a priori Verteilung die Betaverteilung $\beta_{k,l}$ mit $n \in \mathbb{N}$ und $m \in \mathbb{N}$ besitzt.

a) Bestimmen Sie die A-posteriori Verteilung von p.

b) Bestimmen Sie den Bayes-Schätzer für p.

a) Sei α die a-priori Verteilung von p, somit $\alpha(p) = \frac{p^{k-1}(1-p)^{l-1}}{B(k,l)}$. Dann gilt für $m \in \{1, 2, ..., n\}$

$$\rho(m, x)\alpha(p) = \binom{n}{m} x^m (1-x)^{n-m} \frac{x^{k-1}(1-x)^{l-1}}{B(k, l)} = \binom{n}{m} \frac{x^{k+m-1}(1-x)^{l+n-m-1}}{B(k, l)},$$

$$P(S = m) = \int_0^1 \rho(m, x)\alpha(p)dp = \binom{n}{m} \frac{B(k + m, l + n - m)}{B(k, l)}.$$

Zusammen ergibt dies die a-posterori Verteilung von p:

$$P(p \in dx|S = m) = \frac{\rho(m, x)}{P(S = m)} = \frac{x^{k+m-1}(1-x)^{l+n-m-1}}{B(k + m, l + n - m)}$$

Somit gilt bedingt auf $S = k$, dass $p \sim \beta_{k+m,l+n-m}$.

b) Der Bayes-Schätzer ist dann gegeben durch

$$\hat{p}_B(m) = \mathbf{E}[p|S = m] = \frac{k + m}{k + l + n},$$

hierbei haben wir das Resultat von Seite 42 der Folien *Schaetzprinzipien_kompakt.pdf.*

Aufgabe 12:

a) $\overline{x} = \frac{5{,}3+7{,}7+8{,}5+6{,}2+7{,}3}{5} = 7$.

b) Bei fünf Messwerten ist der Median der drittgrößte (bzw. -kleinste) Wert. Hier ist das 7,3.

c) $s^2 = \frac{1}{4}\left((5{,}3 - 7)^2 + (7{,}7 - 7)^2 + (8{,}5 - 7)^2 + (6{,}2 - 7)^2 + (7{,}3 - 7)^2\right) = 1{,}59$, also $s \approx 1{,}26$.

d) Das 95%-Konfidenzintervall ist gegeben durch

$$\left[\overline{x} - t_{4;0{,}975} \cdot \frac{s}{\sqrt{5}}, \overline{x} + t_{4;0{,}975} \cdot \frac{s}{\sqrt{5}}\right] \approx [5{,}44\ ,\ 8{,}56]\,.$$

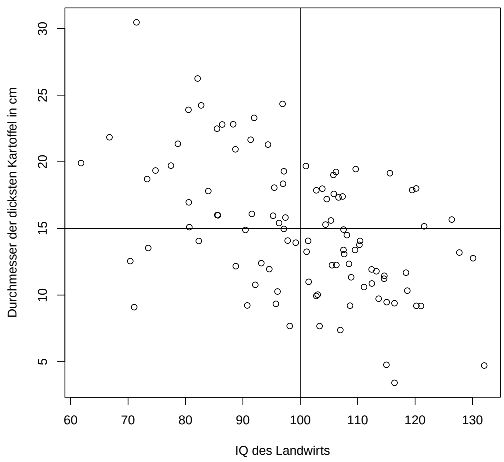
(Achtung: Rundet man s genauer, so verschieben sich die gerundeten Intervallgrenzen jeweils um 0,01 nach außen.)

e) Wenn uns die wahre Standardabweichung bekannt ist, müssen wir sie nicht mehr durch s schätzen und können die Quantile der Normalverteilung verwenden (anstatt der Quantile der t-Verteilung). Das 95%-Konfidenzintervall ist dann

$$\left[\overline{x} - z_{0{,}975} \cdot \frac{1}{\sqrt{5}}, \overline{x} + z_{0{,}975} \cdot \frac{1}{\sqrt{5}}\right] \approx [6{,}12\ ,\ 7{,}88]\,.$$

Aufgabe 13:

a) Das Diagramm unterstützt die These, da Landwirte mit unterdurchschnittlichem IQ eher überdurchschnittlich dicke Kartoffeln ernten. Man sieht das gut, wenn man die Durchschnittswerte für den Landwirts-IQ und die Kartoffeldicke in das Diagramm einträgt:

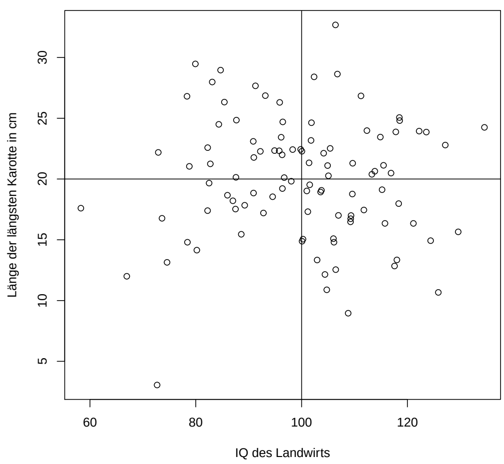


Die Sektoren links oben und rechts unten enthalten nämlich besonders viele Datenpunkte.

b) Der Durchschnittsdurchmesser beträgt $15cm$.

c) Die Standardabweichung der Durchmesser beträgt $5cm$ (zwischen $10cm$ und $20cm$ befinden sich rund $\frac{2}{3}$ der Datenpunkte).

d) Die Daten sind am ehesten unkorreliert. Zur Verdeutlichung sind im folgenden Diagramm wieder die Mittelwerte eingezeichnet:



Aufgabe 14:

a) Wir haben zwei Stichproben x und y vorliegen, zwischen denen keine direkte Verbindung besteht (man kann die Tiere der beiden Gruppen nicht sinnvoll einander „zuordnen"). Ein geeigneter t-Test ist somit der *ungepaarte* Zwei-Stichproben-t-Test. Wir nehmen gleiche Varianzen an und verwenden deshalb die Formel

$$t = \frac{\overline{x} - \overline{y}}{s\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \text{ mit } s = \frac{(n_x - 1)\, s_x^2 + (n_y - 1)\, s_y^2}{n_x + n_y - 2}.$$

Man berechnet nun $n_x = n_y = 5$, $\overline{x} = 31{,}24$, $\overline{y} = 29{,}68$, $s_x^2 \approx 0{,}9$ und $s_y^2 \approx 1{,}1$. Somit ist $t \approx 2{,}46$. Um den Test abzuschließen, müssen wir diesen Wert noch mit dem 0,975-Quantil der t-Verteilung in 8 $(= n_x + n_y - 2)$ Freiheitsgraden vergleichen. Da $t \approx 2{,}46 > 2{,}31 \approx t_{8;0{,}975}$, können wir die Nullhypothese verwerfen.

b) Im Fall, dass die Daten von nur fünf Kühen kommen, können wir jeweils zwei Daten einander zuordnen (nämlich jeweils „Milchmenge Kuh i bei Mozartmusik“ und „Milchmenge Kuh i bei Biebermusik“ $(i = 1, \ldots, 5)$. Wir wollen also einen gepaarten t-Test durchführen. Wir bestimmen zunächst die Differenzdaten $z_i := x_i - y_i$:

Nummer	1	2	3	4	5
Differenz	2,2	-1,4	4,0	0,9	2,1

Wir berechnen $\overline{z} = 1{,}56$ und $s_z \approx 1{,}99$. Die t-Statistik ist dann $t = \frac{\overline{z}\cdot\sqrt{5}}{s_z} \approx 1{,}75$. Diesen Wert müssen wir nun mit dem 0,975-Quantil der t-Verteilung in 4 $(= n_z - 1)$ Freiheitsgraden vergleichen: $t \approx 1{,}75 < 2{,}78 \approx t_{4;0{,}975}$, wir könnten in diesem Fall also die Nullhypothese nicht verwerfen.

Aufgabe 15:

a) $\bar{x} = 40, \bar{y} = 20.9$.

b) Die Steigung der Regressionsgerade ist gegeben durch $\beta_1 = \frac{cov_{x,y}}{\sigma_x^2}$ mit

$$cov_{x,y} = \frac{1}{n}\sum_{i=1}^{6}(x_i - \hat{x})(y_i - \hat{y}) = 142\frac{2}{3};\ \sigma^2 = \frac{1}{n}\sum_{i=1}^{6}(x_i - \hat{x})^2 = 291\frac{2}{3}$$

also $\beta_1 = 0.49$ und $\beta_0 = \hat{y} - \beta_1\hat{x} = 1.33$.

c) Mit der Regressionsgeraden ergibt sich der Wert 38.2 für $x = 75$.

Aufgabe 16:

a) Bei einem fairen Würfel würde jede Zahl mit Wahrscheinlichkeit $\frac{1}{6}$ fallen. Bei 100 Würfen erhält man also $E_1 = \ldots = E_6 = \frac{1}{6} \cdot 100 = \frac{50}{3}$. Die Chi-Quadrat-Statistik berechnet man mit der Formel $X^2 = \frac{\sum((O_i - E_i)^2)}{E_i}$. Für den ersten Würfel erhalten wir also

$$X^2 = \frac{(7 - \frac{50}{3})^2}{\frac{50}{3}} + \frac{(5 - \frac{50}{3})^2}{\frac{50}{3}} + \frac{(13 - \frac{50}{3})^2}{\frac{50}{3}} + \frac{(25 - \frac{50}{3})^2}{\frac{50}{3}} + \frac{(23 - \frac{50}{3})^2}{\frac{50}{3}} + \frac{(11 - \frac{50}{3})^2}{\frac{50}{3}} = 27{,}56,$$

und genauso für den zweiten Würfel $X^2 = 95{,}6$. Beide Werte müssen wir mit dem 0,95-Quantil der Chi-Quadrat-Verteilung in $(6 - 1) = 5$ Freiheitsgraden vergleichen. $x_{5;0{,}95} \approx 11{,}07$, somit ist der Tabellenwert deutlich kleiner als die χ^2-Statistiken, die wir von beiden Würfeln erhalten haben. Wir können die Nullhypothese in beiden Fällen also verwerfen.

b) Wir fügen zunächst die beiden Tabellen zusammen:

Augenzahl	1	2	3	4	5	6
Würfel 1	7	5	13	25	23	27
Würfel 2	2	3	10	49	25	11

Wir wollen die $E_{i,j}$ berechnen (exemplarisch am Beispiel $E_{1,1}$): Von 200 Würfen wurden 100 mit Würfel 1 ausgeführt, anteilig also die Hälfte. Von allen gewürfelten Einsen müssten also unter der Annahme der Homogenität auch die Hälfte mit Würfel 1 gewürfelt worden sein, das sind $\frac{1}{2} \cdot (7+2) = 4{,}5$. Also ist $E_{1,1} = 4{,}5$. Insgesamt erhalten wir die folgenden $E_{i,j}$:

Augenzahl	1	2	3	4	5	6
Würfel 1	4,5	4	11,5	37	24	19
Würfel 2	4,5	4	11,5	37	24	19

Wieder berechnen wir die Chi-Quadrat-Statistik: $X^2 = \sum_{i,j} \frac{(O_{i,j}-E_{i,j})^2}{E_{i,j}} \approx 18{,}27$. Wir vergleichen diesen Wert mit dem 95%-Quantil der χ^2-Verteilung in $(2-1)\cdot(6-1) = 5$ Freiheitsgraden. Da $X^2 \approx 18{,}27 > 11{,}07 \approx x_{5;0{,}95}$, können wir auch hier die Nullhypothese verwerfen.

Aufgabe 17:
Wir sortieren die Daten nach Größe (kursivgeschriebene Werte sind Daten von Kunden ohne Karte):

Einkaufswert	10,65	30,27	37,55	45,10	*45,21*	*46,12*	66,23	*70,69*	*77,56*	176,55
Rang	1	2	3	4	5	6	7	8	9	10

Ist die Nullhypothese wahr, so haben Kunden ohne Karte typischerweise eine eher kleine Rangsumme, wir verwerfen deshalb, wenn die Rangsummenstatistik der Kunden *ohne* Karte groß ist. Für die Kunden ohne Karte ergibt sich die Rangsummenstatistik $W = 5+6+8+9-(1+2+3+4) = 18$. Wir vergleichen diesen Wert mit dem 95%-Quantil der Wilcoxonverteilung mit $n = 4$ und $m = 6$ (wir testen einseitig, denn aus der Aufgabenstellung ergibt sich als Nullhypothese „Ausgabe mit Karte $\geq$ Ausgabe ohne Karte“).
Es ist $W = 18 < 20 = x_{4,6}$, wir können die Nullhypothese also nicht verwerfen.

Aufgabe 18:
Da die Boxplots aus jeweils 1000 Stichproben erstellt würden, können wir davon ausgehen, dass der Median und die Streuung im Boxplots in etwa dem Median und der Streuung der zugrunde liegenden Verteilung entsprechen.
Betrachten wir die Generierung der Datensätze fällt auf, dass sich diese nur durch den Erwartungswert, welcher bei der Normalverteilung identisch ist mit dem Median, und der Varianz unterscheiden. Hieraus kann man leicht folgern: a)=3,b)=1,c)=2,d)=4.

Aufgabe 19:

a) Aus dem Text entnehmen wir die Vermutung, dass es je nach Wochentag die Wahrscheinlichkeiten für die Ausfälle unterschiedlich sind. Um diese These zu stütze, müssen wir die Hypothese entkräften, dass an jedem Tag die Ausfallswahrscheinlichkeit identisch ist. Wir nutzen daher einen χ^2-Test mit der Nullhypothese

$$H_0 : \forall k \in \mathbb{N}_0 \forall (t, \tilde{t}) \in \{Mo, Di, Mi, Do, Fr\}^2 : p_{k,t} = p_{k,\tilde{t}},$$

wobei $p_{k,t}$ die Wahrscheinlichkeit ist, dass am Tag t genau k Rechner ausfallen.

b) Die 2.Ausgabe ist richtig. Der p-Wert ist 0.09094. Somit kann die H_0 zum Niveau 10% und 15% abgelehnt werden, aber nicht zum Niveau 5%.

Aufgabe 20:

```
irrfahrt<-function(n){
  steps<-2*rbinom(n,size=1,prob=0.5)-1
  pfad<-c(0,cumsum(steps))
  return(pfad)
}

check<-function(n,m){
  O<-max(c(n,m))
  U<-min(c(n,m))
  pfad<-irrfahrt(O)
  ergebnis<-0
  if((pfad[U]<0 & pfad[O]>=0) | (pfad[U]>=0 & pfad[O]<0)){
    ergebnis<-1
  }
  return(ergebnis)
}
```

Aufgabe 21:

```
#a)
simChi2<-function(n){
  x<-rnorm(n)
  S<-sum(x^2)
  return(S)
}
#b)
m<-10^4; n<-10
sim<-replicate(m,simChi2(n))
hist(sim,freq = FALSE)
#c)
densityChi2<-function(x,n){
  x^{n/2-1}*exp(-x/2)/gamma(n/2)/2^(n/2)
}
t<-seq(0,40,0.1)
points(t,densityChi2(t,10))
```