Klausur Statistik: Lösungshinweise

Prüfungsdatum: 30. Juni 2014 – Prüfer: Etschberger, Jansen, Nebel
Studiengang: IM und BW

Aufgabe 1 — 12 Punkte

Eine Bäckerei registriert an neun aufeinander folgenden Tagen die Anzahl der eingehenden Großbestellungen.

a) Geben Sie die sortierte Urliste für den Fall an, dass folgende Informationen bekannt sind:

$$\bar{x} = \frac{16}{3}\,, \quad x_{\text{Med}} = 3\,, \quad x_{\text{Mod}} = 2\,, \quad h(2) = 3\,, \quad \text{Spannweite} = 8 \quad \text{und} \quad F(9) = \frac{8}{9}$$

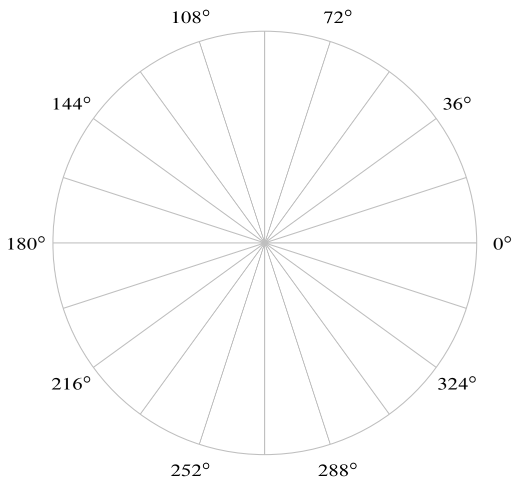
Gehen Sie für die nachfolgenden Teilaufgaben von dieser, nicht notwendigerweise mit der Lösung zu Teil a) überein stimmenden, sortierten Urliste aus:

$$2\,,2\,,3\,,3\,,3\,,4\,,5\,,8\,,8\,,10$$

b) Geben Sie die Häufigkeitsverteilung der Großbestellungen an.

c) Zeichnen Sie das Kreissektorendiagramm der Großbestellungen (Winkel müssen angegeben werden). Benutzen Sie dafür die Zeichnung rechts.

d) Zeichnen Sie die empirische Verteilungsfunktion der Großbestellungen.



Lösungshinweis:

a) Median=3, darunter 4 Werte, davon 3 Stück gleich 2. Keine nichtganzzahligen Werte, 1 als Ausprägung geht nicht, sonst wäre F(9)=1 (Spannweite=8, also max=10). Also Urliste:

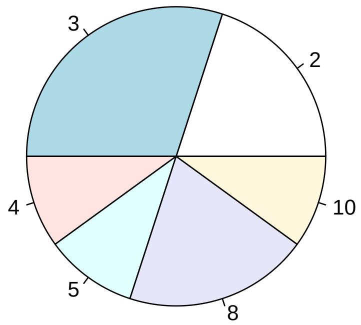
$$2,2,2,3,3,x_6,x_7,9,10$$

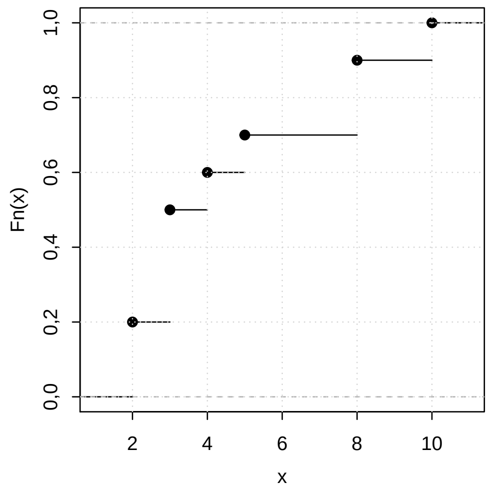
Summe = $\frac{16}{3} * 9 = 48$, d.h. $x_6 + x_7 = 17$, also $x_6 = 8, x_7 = 9$.

b)

2	3	4	5	8	10
2	3	1	1	2	1

c) und d)





Aufgabe 2 — 8 Punkte

In einer Studie mit $n = 100$ (Studie 1) wurde zu den beiden Merkmalen X (Besitz eines Smartphones) und Y (Besitz einer Flatrate) folgende unvollständige auf relativen Häufigkeiten basierende Kontingenztabelle erstellt:

X \ Y	ja	nein	$h_{i.}$
ja	0,4		
nein		0,1	
$h_{.j}$	0,7		

a) Ergänzen Sie sämtliche fehlenden Einträge inklusive Randhäufigkeiten.

b) Ermitteln Sie den normierten Kontingenzkoeffizienten und interpretieren Sie damit die Stärke des Zusammenhangs.

c) In einer anderen Studie (Studie 2) wurden auf Basis von $n = 1000$ Beobachtungen dieselben Merkmale erfasst. Der aus diesen Daten ermittelte, nicht normierte Kontingenzkoeffizient beträgt $K = 0{,}1$. Welche Studie ermittelt einen größeren Zusammenhang zwischen den beiden Merkmalen?

Lösungshinweis:

Aufgabe 3 — 9 Punkte

Johann belauscht eine Feier aus einem Nebenraum. Er zählt beim Anstoßen der Sektgläser mit und überlegt, wie viele Leute beim Sektempfang im Raum sein müssten, wenn jeder mit jedem genau einmal anstößt:

a) Wie oft müssten die Gläser jeweils in einem Raum mit 4, 5 bzw. 6 Leuten klingeln?

b) Wie of müsste bei n Leuten im Raum insgesamt angestoßen werden?

c) Johann zählt bei der Feier 113 mal Gläserklingeln. Es scheint dabei etwas nicht zu stimmen. Anscheinend haben einige Leute nicht miteinander angestoßen. Wieviel Leute sind mindestens im Raum, wenn jede Person höchstens einmal mit jeder anderen Person anstößt?

Lösungshinweis:

a)

Personen	Anstossen
4	6
5	10
6	15

b) Es wird $\binom{n}{2}$ mal angestoßen.

c)

Personen	Anstossen
15	105
16	120

Also: mindestens 16 Personen sind im Raum.

Aufgabe 4 — 10 Punkte

Die Anzahl der pro Jahr erwischten und öffentlich bekannt gemachten prominenten Steuersünderfälle (mit mehreren Millionen hinterzogenen Euro) sei eine poissonverteilte Zufallsvariable X. Die Wahrscheinlichkeit, dass in einem Jahr kein oder ein solcher Fall bekannt wird beträgt 0,0611.

a) Bestimmen Sie den Parameter λ der Verteilung.

b) Geben Sie eine anschauliche Bedeutung für den Parameter λ.

c) Mit welcher Wahrscheinlichkeit werden in einem Jahr genau zwei Prominente in dieser Kategorie erwischt?

d) Wie groß ist die Wahrscheinlichkeit mindestens 5 solche Fälle pro Jahr zu haben?

Lösungshinweis:

a) $P(X \leqq 1) = 0{,}0611 \quad \Leftrightarrow \quad \lambda = 4{,}5$ (siehe Tabelle).

b) Erwartungswert gleich 4,5, auf lange Sicht durchschnittlich 4,5 Fälle pro Jahr.

c) $P(X = 2) = 0{,}1125$

d) $P(X \geqq 5) = 1 - P(x \leqq 4) = 0{,}4679$

Aufgabe 5 — 12 Punkte

In einem Glücksspiel kann man einen Euro setzen. Es gibt genau drei Möglichkeiten (A, B, C) für den Ausgang des Spiels: Bei A bekommt man den eingesetzten Euro wieder, bei B ist der Euro verloren und bei C erhält man 10 Euro. Mit X als der Zufallsvariable für den Gewinn gilt also:

	A	B	C
x	0	-1	9
$P(X = x)$	p_A	p_B	p_C

a) Es gelte $p_A = 0{,}3$ und $p_B = 0{,}65$. Wie groß ist der Erwartungswert des Gewinns?

b) Es gelte $p_A = 0{,}3$. Wie groß müssen p_B und p_C sein, so dass das Spiel fair, also $\mathrm{E}[X] = 0$ ist?

c) Jetzt seien p_A, p_B, p_C unbekannt. Berechnen Sie $\mathrm{E}[X]$ und $\mathrm{E}[X^2]$ in Abhängigkeit von p_A, p_B, p_C.

d) p_C soll mit einer einfachen Stichprobe vom Umfang n geschätzt werden. Zeigen Sie, dass die Stichprobenfunktion

$$\hat{P}_C = \frac{1}{90n} \sum_{i=1}^{n} \left(X_i + X_i^2\right)$$

eine erwartungstreue Schätzfunktion für p_C ist.

Lösungshinweis:

a) $\mathrm{E}[X] = 0 \cdot 0{,}3 + (-1) \cdot 0{,}65 + 9 \cdot 0{,}05 = -0{,}2$

b) $0 = 0 \cdot 0{,}3 + (-1) \cdot p_B + 9 \cdot (0{,}7 - p_B) \Leftrightarrow 10p_B = 6{,}3 \Leftrightarrow p_B = 0{,}63,\ p_C = 0{,}07$

c) $\mathrm{E}[X] = -p_B + 9p_C$

$\mathrm{E}[X^2] = p_B + 81p_C$

d)
$$\begin{aligned}
\mathrm{E}[\hat{P}_C] &= \mathrm{E}\left[\frac{1}{90n} \sum_{i=1}^{n} \left(X_i + X_i^2\right)\right] = \frac{1}{90n} \sum_{i=1}^{n} \left(\mathrm{E}[X_i] + \mathrm{E}[X_i^2]\right) \\
&= \frac{1}{90n} \sum_{i=1}^{n} (-p_B + 9p_C + p_B + 81p_C) \\
&= \frac{1}{90n} \sum_{i=1}^{n} (90p_C) = p_C
\end{aligned}$$

Aufgabe 6 — 9 Punkte

US-Banken verwenden zum automatischen Bearbeiten von Schecks Lesegeräte. Zu einem geringen Prozentsatz treten jedoch Probleme auf. Der folgende Datensatz steht zur Verfügung:

bearbeitete Schecks	1040
davon nicht lesbar	14

a) Berechnen Sie ein Konfidenzintervall zum Konfidenzniveau 95 % für den Anteil der nicht lesbaren Schecks.

b) Die Herstellerfirma des Geräts gibt an, dass der Anteil an nicht lesbaren Schecks maximal 1 % sein soll. Das soll anhand des Datensatzes getestet werden. Formulieren Sie H_0 und H_1.

c) Kann anhand der Daten nachgewiesen werden, dass die Maschine schlechter arbeitet? Das Signifikanzniveau betrage 5%.

Aufgabe 7 — 10 Punkte

Es ist jeweils genau eine Antwort pro Aufgabenteil richtig. Für jede richtige Antwort gibt es 2 Punkte, jede falsch beantwortete gibt −0,5 Punkte, eine nicht beantwortete Frage 0 Punkte. Man kann nicht weniger als 0 Punkte in der Aufgabe bekommen. Tragen Sie ihre Lösungsbuchstaben in die folgenden Kästchen ein. Bitte schreiben sie deutlich!

Aufgabenteil	1	2	3	4	5
Lösung					

Aufgabenteil 1: Arithmetik

Was liefert die folgende R-Eingabe?

```
1:4 + 1:2 * 1:2^2
```

(A) `[1] 2 18 4 20`

(B) `[1] 2 6 6 12 6 14 10 20`

(C) `[1] 2 6 6 12`

(D) `[1] 1 4 9 16 5 12 21 32`

(E) `[1] 2 8 12 24 10 24 28 48`

Aufgabenteil 2: Kombinatorik

Sie wollen die Anzahl der möglichen '6 aus 49' im Lotto ermitteln. Welche der folgenden Zeilen liefert das richtige Ergebnis?

(A) `lotto(6)`

(B) `binom(n = 49, k = 6)`

(C) `choose(49,6)`

(D) `sample(1:49, 6, replace = FALSE)`

(E) `choice(6,49)`

Aufgabenteil 3: Matrizen

Gegeben sei eine Matrix `M`, die Sie mit Hilfe der Funktionen `rbind()` und `cbind()` erweitern.

```
M <- matrix(seq(2, 8, by = 2),
            nrow = 2, byrow = TRUE)
M <- cbind(M,c(9,9))
M <- rbind(1:3, M)
```

Wie sieht die 2. Spalte aus, d.h. welches Ergebnis liefert die Ausgabe `M[,2]` ?

(A) `[1] 2 4 8`

(B) `[1] 1 2 3`

(C) `[1] 2 6 9`

(D) `[1] 3 9 9`

(E) `[1] 4 8 9`

Aufgabenteil 4: Funktionen

Gegeben ist folgende Funktion:

```
neueFunktion <- function(x, diff = TRUE){
 if(diff) {
    rval <- sqrt(sum(x^2))
 } else {
    rval <- max(x)
 }
 return(rval)
}
```

Welche Ausgabe erwarten Sie für die beiden Aufrufe

```
 neueFunktion(c(6,8))
 neueFunktion(c(1,3,3,10), FALSE)
```

(A) `[1] 8` und `[1] 10`

(B) `[1] 10` und `[1] 11`

(C) `[1] 14` und `[1] 11`

(D) `[1] 10` und `[1] 10`

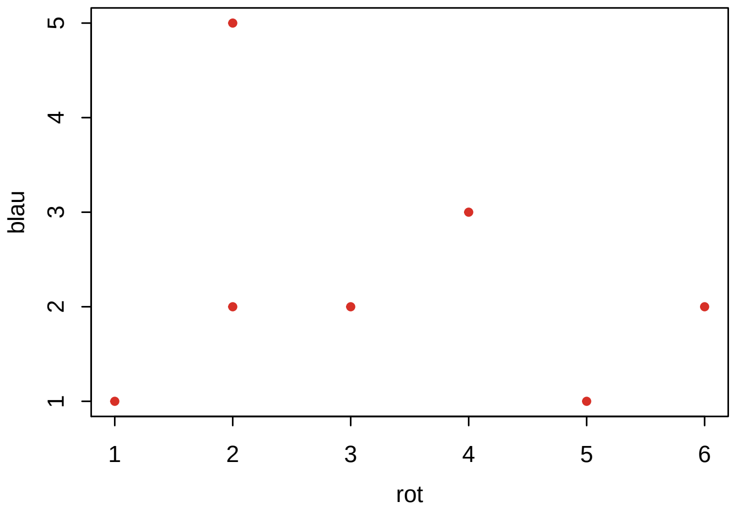
(E) `[1] 14` und `[1] 17`

Aufgabenteil 5: Graphik

Sie haben sieben mal gleichzeitig mit zwei sechsseitigen Würfeln gewürfelt und das Ergebnis in die folgende Tabelle geschrieben:

Wurf	1	2	3	4	5	6	7
roter Würfel	4	5	6	2	1	3	2
blauer Würfel	3	1	2	5	1	2	2

Sie wollen das Ergebnis graphisch darstellen und ein Streudiagramm erstellen. Welche Zeile ergibt die unten abgebildete Graphik?



(A) `plot(c(4,5,6,2,1,3,2), c(3,1,2,5,1,2,2),`
`xlab = "rot", ylab = "blau", col = "Blue", pch = 20)`

(B) `plot(c(3,1,2,5,1,2,2), c(4,5,6,2,1,3,2), xlab = "rot", ylab = "blau", col = 1`

(C) `plot.points(c(4,5,6,2,1,3,2), c(3,1,2,5,1,2,2), xaxis = c("rot"), yaxis = "blau") )`

(D) `scatterplot(c(4,5,6,2,1,3,2), c(3,1,2,5,1,2,2), color = "Blue", axis = c("rot", "blau"))`

(E) `plot([4,5,6,2,1,3,2],[3,1,2,5,1,2,2], xlab = "rot", ylab = "blau", col = "Blue")`