

Probeklausur Statistik

Lösungshinweise

Prüfungsdatum: Juni 2015 – Prüfer:
Studiengang: IM und BW

Aufgabe 1

18 Punkte

20 Studenten werden gefragt, wie viele Stunden sie durchschnittlich pro Tag ihr Smartphone eingeschaltet haben und Nachrichten empfangen sowie senden (Merkmal X). Es ergibt sich folgende Urliste:

2, 1, 3, 7, 10, 2, 5, 10, 2, 11, 9, 4, 14, 9, 0, 5, 3, 3, 5, 3

- Bestimmen Sie das arithmetische Mittel, den Median sowie die Standardabweichung von X .
- R**: Geben Sie R-Befehle an, um die Daten in einer Variable x zu speichern und die gesuchten Größen aus Teilaufgabe a) zu berechnen (Sie dürfen die Liste der Daten abkürzen).
- Erstellen Sie eine Tabelle der kumulierten relativen Häufigkeiten zu allen in der Umfrage genannten Ausprägungen.
- R**: Mit welchem R-Befehl kann man Teilaufgabe c) realisieren?
- R**: Geben Sie in R zur empirischen Verteilungsfunktion F die Werte $F(3,9)$ und $F(11)$ an.
- R**: Die erhobenen Werte sollen den Klassen

Klasse	K_1	K_2	K_3	K_4	K_5
Intervall	[0, 4)	[4, 8)	[8, 10)	[10, 12)	[12, 24]

zugeordnet werden. Geben Sie einen passenden R-Befehl an, so dass ein Histogramm der Daten mit dieser Klassenaufteilung gezeichnet würde.

Lösungshinweis:

- Für das arithmetische Mittel ergibt sich $\bar{x} = 5,4$. Der Median ist $x_{\text{med}} = 4,5$. Die Standardabweichung beträgt $s = 3,7735925$.
- ```
x = c(2, 1, 3, ...) # Eingabe abgekürzt
mean(x) # arithmetisches Mittel
median(x) # Median
sqrt(mean((x - mean(x))^2)) # Standardabweichung
```
- ```
##      0      1      2      3      4      5      7      9     10     11     14
## 0,05 0,10 0,25 0,45 0,50 0,65 0,70 0,80 0,90 0,95 1,00
```
- ```
cumsum(table(x))/length(x)
```

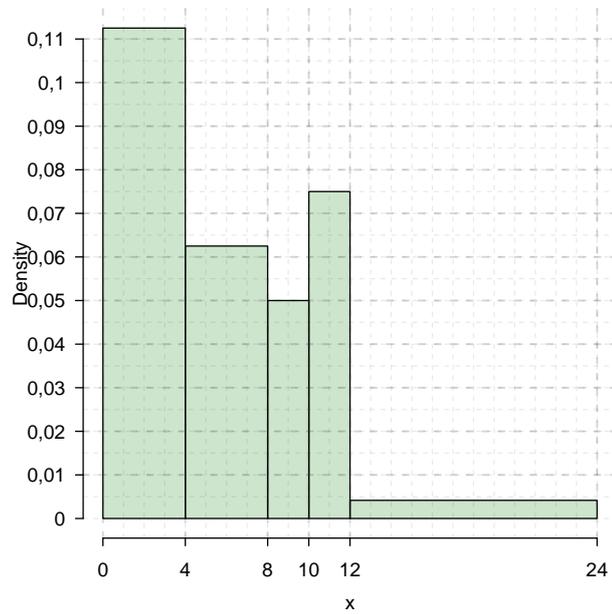
empirical  
cumulative  
distribution  
function

```
e) F.x = ecdf(x)
F.x(c(3.9, 11))
[1] 0,45 0,95
```

$F.x(3.9)$   
 $F.x(11)$

default

```
f) Grenzen = c(0,4,8,10,12,24)
hist(x, breaks=Grenzen, right=FALSE, (include.lowest=TRUE))
```



## Aufgabe 2

17 Punkte

In einem Restaurant wird bei 7 Gästen die Rechnungshöhe  $R$  zusammen mit dem gegebenen Trinkgeld  $T$  erfasst. Es ergeben sich folgende Werte:

| $i$   | 1      | 2     | 3     | 4      | 5      | 6     | 7      |
|-------|--------|-------|-------|--------|--------|-------|--------|
| $R_i$ | 218,92 | 63,56 | 11,74 | 285,72 | 181,66 | 28,33 | 392,99 |
| $T_i$ | 19,08  | 8,44  | 2,26  | 22,28  | 17,34  | 4,67  | 25,01  |

- Berechnen Sie einen geeigneten Korrelationskoeffizienten zwischen Rechnungshöhe und Trinkgeld. Interpretieren Sie das Ergebnis.
- R**: Geben Sie R-Befehle an, um die Lösung von a) zu berechnen (Die Eingabe der Daten dürfen Sie abkürzen).
- Bestimmen Sie die Regressionsparameter eines linearen Modells, in dem die Höhe des Trinkgeldes in Abhängigkeit von der Rechnungshöhe beschrieben wird. Wieviel Trinkgeld würden Sie nach diesem Modell bei einer Rechnung von 200 € erwarten?
- R**: Geben Sie R-Befehle an, um
  - ▶ das Regressionsmodell zu bestimmen,
  - ▶ den Determinationskoeffizienten des Modells zu berechnen,
  - ▶ einen geeigneten Streuplot auszugeben und
  - ▶ die Regressionsgerade in den Streuplot einzuzeichnen.
- Zeichnen Sie die Werte zusammen mit der Regressionsgerade in ein geeignetes Koordinatensystem ein.

### Lösungshinweis:

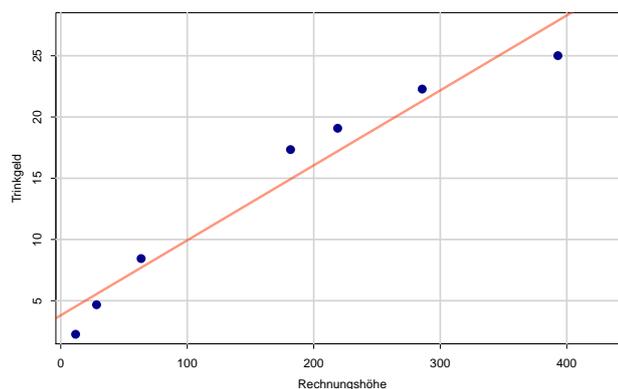
- Bravais-Pearson:  $r = 0,9738711$ . *starker positiver Zusammenhang (je größer Rechnungshöhe, desto höheres Trinkgeld)*
- ```
R = c(218.92, 63.56, ...) # abgekürzt
T = c(19.08, 8.44, ...) # abgekürzt
cor(R, T) # Bravais-Pearson-Korrelationskoeffizient
```

- Modell: $T = 0,0612153 \cdot R + 3,8095985$
Prognose: $T(200) = 16,05$

- ```
Modell = lm(T ~ R) # Regressionsmodell
cor(T, R)^2 # Determinationskoeffizient
plot(T, R) # Streuplot
abline(Modell) # Gerade in Plot
```

*Modell \$ coefficients [1] a  
" [2] b  
y = a + b \* x*

- Streuplot:



### Aufgabe 3

14 Punkte

Johann wohnt bei seinen Eltern. Sein Studium und insbesondere die Begleiterscheinungen durch soziale Aktivitäten haben ihn in letzter Zeit ziemlich gefordert, so dass er eines Tages nach dem Aufwachen überhaupt nicht weiß, welcher Wochentag gerade ist.

Allerdings hat er folgendes beobachtet:

- ▶ Wenn er die Küche seiner Eltern an Sonntagen betritt erklingt mit 20 % Wahrscheinlichkeit Blasmusik aus dem Radio, an anderen Wochentagen nur mit 2 % Wahrscheinlichkeit.
- ▶ Johanns Vater ist am Montag, Dienstag und Mittwoch immer schlecht gelaunt.

- a) Johann öffnet die Küchentüre. Es erklingt Blasmusik. Mit welcher Wahrscheinlichkeit ist gerade Sonntag?
- b) Johann stellt nach einer Minute am Küchentisch fest, dass sein Vater blendender Laune ist. Wie hoch ist jetzt die Wahrscheinlichkeit, dass Sonntag ist?

Lösungshinweis:

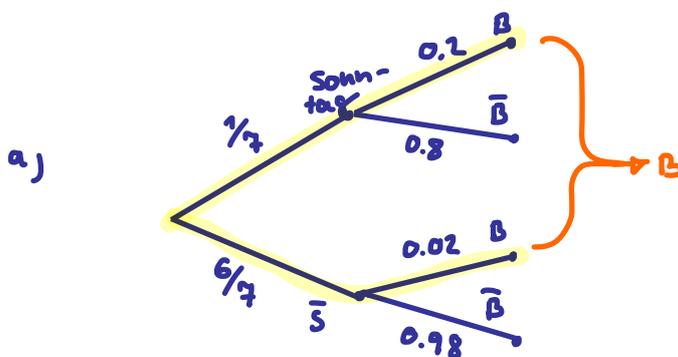
$B \equiv$  „Es erklingt Blasmusik“

$S \equiv$  „Es ist Sonntag“

$$a) P(S|B) = \frac{P(B|S)P(S)}{P(B|S)P(S) + P(B|\bar{S})P(\bar{S})} = \frac{0,2 \cdot 1/7}{0,2 \cdot 1/7 + 0,02 \cdot 6/7} = \frac{20}{20 + 2 \cdot 6} = 5/8 = 62,5\%$$

b) Genauso, nur ist jetzt  $P(S) = 1/4$ , denn Montag, Dienstag, Mittwoch scheiden aus. Also:

$$P(S|B) = \frac{0,2 \cdot 1/4}{0,2 \cdot 1/4 + 0,02 \cdot 3/4} = \frac{20}{20 + 2 \cdot 3} = 10/13 \approx 76,9\%$$



$$P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{\frac{1}{7} \cdot 0,2}{\frac{1}{7} \cdot 0,2 + \frac{6}{7} \cdot 0,02} = \frac{0,2}{0,2 + 6 \cdot 0,02} = \frac{20}{20 + 12} = \frac{5}{8} = 0,625$$

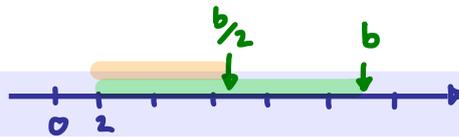
b)

|           |      |           |       |
|-----------|------|-----------|-------|
|           | S    | $\bar{S}$ |       |
| B         | 0.05 | 0.015     | 0.065 |
| $\bar{B}$ | 0.20 | 0.735     | 0.935 |
|           | 1/4  | 3/4       |       |

Handwritten notes:  $0,2 \cdot 1/4$  points to the 0.05 cell;  $0,02 \cdot 3/4$  points to the 0.015 cell.

$$P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{0,05}{0,065} \approx 0,769$$

## Aufgabe 4



$$F\left(\frac{b}{2}\right) = \frac{\frac{b}{2} - 2}{b - 2} = 0.4$$

$$\Leftrightarrow 0.5b - 2 = 0.4b - 0.8$$

$$\Leftrightarrow 0.1b = 1.2 \Rightarrow b = 12$$

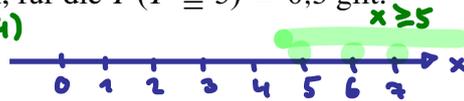
15 Punkte

Gegeben ist eine im Intervall  $[2; b]$  gleichverteilte Zufallsvariable  $X$ .

- 4 a) Bestimmen Sie  $b$  so, dass für die Verteilungsfunktion  $F(x)$  gilt:  $F\left(\frac{b}{2}\right) = 0.4$ .
- 2 b) Berechnen Sie  $F(4)$ .
- 3 c) **R**: Geben Sie einen R-Befehl an, der den gesuchten Wert von b) mit der Verteilungsfunktion der Gleichverteilung ausgibt.

Jetzt ist eine poissonverteilte Zufallsvariable  $Y$  gegeben, für die  $P(Y \geq 5) = 0.3$  gilt.

- 3 d) Bestimmen Sie für  $Y$  den Wert von  $F(4)$ .



Schließlich ist eine poissonverteilte Zufallsvariable  $Z$  mit  $Z \sim P(\lambda = 3)$  gegeben.

- 3 e) **R**: Geben Sie eine Zeile mit R-Befehlen an, die das Ergebnis von  $P(2 \leq Z < 5)$  berechnet.

### Lösungshinweis:

$$a) F\left(\frac{b}{2}\right) = \frac{\frac{b}{2} - 2}{b - 2} = 0.4 \Rightarrow b = 12$$

$$b) F(4) = \frac{4 - 2}{12 - 2} = \frac{2}{10} = 0.2$$

c) `punif(q=4, min = 2, max = 12)`  
## [1] 0,2

$$d) P(Y \geq 5) = 0.3 \Leftrightarrow F(4) = P(Y < 5) = 0.7$$

e) `ppois(q=4, lambda=3) - ppois(q=1, lambda=3)`  
## [1] 0,616115

alternativ:  
`sum(dpois(q=2:4, lambda=3))`

oder (etwas umständlich):  
`dpois(q=2, lambda=3)`  
`+ dpois(q=3, lambda=3)`  
`+ dpois(q=4, lambda=3)`

$d < \text{Verteg.} >$   
Dichte, Wahrsch.-funktion

$p < \text{Verteg.} >$   
Verteilungsfunktion

$q < \text{Verteg.} >$   
Funktionswert

Verteilungen: Binomial, Poisson,  
hypergeom., normal,  
unif, norm

## Aufgabe 5

14 Punkte

Es sei  $(X_1, \dots, X_n)$  eine einfache Stichprobe, wobei für die Stichprobenvariablen gelte:

$$E(X_i) = 1, \quad \text{Var}(X_i) = \sigma^2 \quad \text{für } i = 1, \dots, n$$

- 6 a) Berechnen Sie  $E(X_i^2)$ .  
 (Hinweis: Benutzen Sie den Verschiebungssatz der Varianz).

$$\text{Var}(X_i) = E(X_i^2) - E^2(X_i) \\ \Rightarrow E(X_i^2) = \text{Var}(X_i) + E^2(X_i) = \sigma^2 + 1^2$$

- 8 b) Nun soll die Varianz  $\sigma^2$  der Grundgesamtheit geschätzt werden. Ist die Funktion

$$\hat{\theta} = \frac{1}{n} (X_1^2 + X_2^2 + \dots + X_n^2) - 1$$

ein erwartungstreuer Schätzer für  $\sigma^2$ ?

$$E(\hat{\theta}) = \frac{1}{n} (E(X_1^2) + \dots + E(X_n^2)) - 1 \\ = \frac{1}{n} (\sigma^2 + 1) + (\sigma^2 + 1) + \dots + (\sigma^2 + 1) - 1 \\ = \frac{1}{n} (n\sigma^2 + n \cdot 1) - 1 = \sigma^2 + 1 - 1 = \sigma^2$$

Lösungshinweis:

a)  $\text{Var}(X_i) = \sigma^2 = E(X_i^2) - E^2(X_i) \Leftrightarrow \sigma^2 = E(X_i^2) - 1 \Leftrightarrow E(X_i^2) = \sigma^2 + 1$

b)  $E(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n E(X_i^2) - 1 = \frac{1}{n} \sum_{i=1}^n (\sigma^2 + 1) - 1 = \sigma^2$ , also ist  $\hat{\theta}$  erwartungstreu

$X_i$ : einfache Stichprobe  
 $E[X_i] = \mu, \quad \text{Var}[X_i] = \sigma^2$   
 $E[\bar{X}] = \mu, \quad \text{Var}[\bar{X}] = \frac{\sigma^2}{n}$

## Aufgabe 6

12 Punkte

Ein Getränkehersteller füllt Limonade in 700 ml Flaschen ab. Die Anlage ist auf ein Abfüllvolumen von 702 ml eingestellt. Dem Hersteller ist daran gelegen, dass das Abfüllvolumen in den Flaschen dem  $\mu_0$  eingestellten Volumen entspricht.

Bei einer einfachen Stichprobe von 17 Flaschen betrug das Stichprobenmittel des Abfüllvolumens 701,2 ml bei einer Stichprobenstandardabweichung von 1 ml.

Das Abfüllvolumen in der Grundgesamtheit kann als normalverteilt angesehen werden.

- 8 a) Testen Sie zu einem Signifikanzniveau von 1%, ob das durchschnittliche Abfüllvolumen  $\mu$  in der Grundgesamtheit dem eingestellten Volumen entspricht.
- 4 b) **R**: Angenommen, die Stichprobe ist in der R-Variablen  $x$  gespeichert. Geben Sie einen R-Befehl an, mit dem man
- ▶ ein Konfidenzintervall für den Mittelwert des abgefüllten Volumens in der Grundgesamtheit zum Konfidenzniveau 95 % sowie
  - ▶ das Ergebnis des Tests aus Teilaufgabe a) ablesen kann.

Lösungshinweis:

a)  $H_0 : \mu = \mu_0 = 702$  ml, gegen  $H_1 : \mu \neq \mu_0$ .

$X_i \sim N(\mu, \sigma)$ , mit unbekanntem  $\sigma$ , also gilt  $V = \frac{\bar{X} - \mu_0}{s} \cdot \sqrt{n} \sim t(n - 1)$  mit 16 Freiheitsgraden.  $s = 1$  ml,  $\alpha = 0,01$ .

Damit:  $x_{1-\alpha/2} \approx 2,921$ , also: Verwerfungsbereich

$$B = (-\infty; -2,921) \cup (2,921; \infty)$$

Mit einem Testwert von  $v = \frac{701,2 - 702}{1} \cdot \sqrt{17} \approx -3,298$  wird  $H_0$  damit verworfen. Die durchschnittliche Füllmenge wird also als ungleich 702 ml angenommen.

b) `t.test(x, alternative = 'two.sided', conf.level = 0.95)`