

Nachholklausur Statistik

Lösungshinweise

Prüfungsdatum: 23. Januar 2019 – Prüfer: Etschberger, Henle, Jansen, Wesp

Studiengang: BW, IM

Punkte: 23, 16, 14, 9, 12, 16 ; Summe der Punkte: 90

Aufgabe 1

23 Punkte

Es wurde eine Untersuchung bei 600 zufällig ausgewählten Personen zum Thema „Einkommen und Fernsehkonsum“ durchgeführt. Dabei wurden folgende Beobachtungskategorien gebildet:

Jahreseinkommen $x \in \{ \text{sehr hoch, hoch, mittel, niedrig, sehr niedrig} \}$

Fernsehkonsum (in h/Tag) $y \in \{ [0, \frac{1}{2}]; (\frac{1}{2}, 1]; (1, 2]; (2, 6] \}$

Es ergab sich die folgende gemeinsame Häufigkeitstabelle:

<i>Fernsehkonsum</i>	<i>Einkommen</i>				
	sehr hoch	hoch	mittel	niedrig	sehr niedrig
$[0, \frac{1}{2}]$	60	25	0	0	5
$(\frac{1}{2}, 1]$	30	40	60	10	15
$(1, 2]$	40	25	70	70	30
$(2, 6]$	0	20	30	20	50

- a) Zeichnen Sie ein Histogramm des Fernsehkonsums *aller* 600 Personen.
- R** b) Geben Sie R-Befehle an, mit denen ein Vektor x definiert wird, der die Daten aus Teilaufgabe a) repräsentiert und das geforderte Histogramm ausgibt.
- c) Warum läßt sich kein Histogramm des Einkommens zeichnen?
- d) Sind in der gegebenen Häufigkeitstabelle die beiden Merkmale Einkommen und Fernsehkonsum unabhängig? Bitte begründen Sie Ihre Antwort, eine Behauptung reicht nicht.

Im Folgenden wird nur die Gruppe der Befragten mit sehr hohem Einkommen betrachtet.

- e) Wie groß ist (näherungsweise) der Mittelwert und die Spannweite des Fernsehkonsums für diese Gruppe?
- f) Wie würden Sie den Vorschlag bewerten, die Standardabweichung des Fernsehkonsums für sehr hohe Einkommen durch die Klassenmitten verbunden mit den jeweiligen Häufigkeiten zu berechnen? Ist das Ergebnis einer solchen Rechnung korrekt, zu klein oder zu groß? Begründen Sie jeweils Ihre Antwort.

Lösungshinweis:

a)

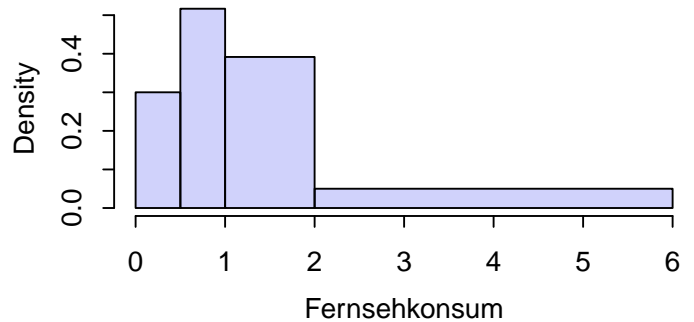
Häufigk. durch Klassenbreite	$\frac{90}{0.5}$	$\frac{155}{0.5}$	$\frac{235}{1}$	$\frac{120}{4}$
Ergebnis	180	310	235	30

R

b)

```

y = c(0.25, 0.75, 1.5, 4)
yF = c(90, 155, 235, 120)
D = rep(y, yF)
hist(D,
      breaks=c(0, .5, 1, 2, 6),
      col="#0010ee30",
      main="",
      xlab="Fernsehkonsum")
  
```



Histogramm aller Fernsehkonsumklassen

c) Einkommen nicht metrisch

d) $\tilde{h}_{[0,1/2],\text{sehr hoch}} = \frac{90 \cdot 130}{600} = 19.5 \neq 60 = h_{[0,1/2],\text{sehr hoch}} \Rightarrow$ nicht unabhängig

e)

Ausprägung	geschätztes \bar{x}	Spannweite
sehr hoch	$\frac{1}{130} \begin{pmatrix} 60 \cdot 0.25 \\ + 30 \cdot 0.75 \\ + 40 \cdot 1.5 \end{pmatrix} = 0.75$	2 (oder 1.25)
hoch	1.398	6 (oder 3.75)
mittel	1.688	5.5 (oder 3.25)
niedrig	1.925	5.5 (oder 3.25)
sehr niedrig	2.575	6 (oder 3.75)

f) Alles ist möglich: Zu klein (echte Verteilung immer außen an Intervallgrenzen), genau richtig, oder zu groß; bei typischer Verteilung Schätzung zu klein

Aufgabe 2

16 Punkte

An einer Hochschule sollen die Studierenden ihre Dozenten bezüglich der Qualität der Lehre beurteilen. Dazu können Sie jeder von 9 Lehrpersonen eine Punktzahl zwischen 0 und 30 zuordnen.

a) Studentin A und Student B vergaben folgende Bewertungen:

Dozent	1	2	3	4	5	6	7	8	9
Punktzahl von A	13	21	10	12	27	25	8	14	18
Punktzahl von B	14	20	6	8	28	27	12	10	15

Gehen Sie im folgenden davon aus, dass die vergebene Punktzahl nicht als metrisches Merkmal angesehen werden kann.

- (1) Berechnen Sie zu den von Studentin A vergebenen Punktzahlen einen geeigneten Lageparameter.
- (2) Berechnen Sie eine geeignete Maßzahl für die Korrelation zwischen den beiden Bewertungsreihen.
- (3) Geben Sie R-Befehle an, um die Daten in R einzulesen und den Korrelationskoeffizienten aus Teilaufgabe (2) zu berechnen.

b) In einer Voruntersuchung wurden 50 Studenten aus vier Studienjahren befragt, und ein Dozent wurde bezüglich der Lehre als schlecht, mittel bzw. gut eingestuft, wenn die erreichte Punktzahl im Intervall $[0; 10]$, $(10; 20]$ bzw. $(20; 30]$ lag. Folgende Häufigkeitstabelle faßt die Ergebnisse für einen bestimmten Dozenten zusammen:

<i>Bewertung</i>	<i>Studienjahr</i>			
	1	2	3	4
schlecht	0	6	3	1
mittel	5	9	1	0
gut	5	5	6	9

- (4) Berechnen Sie den Summanden von χ^2 für das 3. Studienjahr und mittlere Bewertungen.
- (5) Es ergibt sich insgesamt $\chi^2 \approx 17.067$ (muss nicht nachgerechnet werden). Bestimmen Sie den normierten Kontingenzkoeffizienten.
- (6) Bestimmen Sie die bedingte Verteilung der Bewertung für die Befragten aus dem zweiten Studienjahr.

Lösungshinweis:

(1) A sortiert: $x_i = (8, 10, 12, 13, 14, 18, 21, 25, 27) \Rightarrow x_{\text{med}} = 14$

(2) Es ergibt sich

Dozent	1	2	3	4	5	6	7	8	9
A	13	21	10	12	27	25	8	14	18
B	14	20	6	8	28	27	12	10	15
Rang.A	4	7	2	3	9	8	1	5	6
Rang.B	5	7	1	2	9	8	4	3	6

$$r_{\text{SP}} = 1 - \frac{6}{8 \cdot 9 \cdot 10} [(5-4)^2 + (7-7)^2 + \dots + (6-6)^2] = 1 - \frac{2}{15} \approx 0.867$$

(3) A = c(13, 21, 10, 12, 27, 25, 8, 14, 18)

B = c(14, 20, 6, 8, 28, 27, 12, 10, 15)

`cor(rank(A), rank(B))` # Bravais-Pearson über Ränge: Spearman.

[1] 0.8666667

`cor(A, B, method = "spearman")` # Spearman direkt

[1] 0.8666667

`cor(A, B)` # FALSCH!

[1] 0.9278743

(4) $\chi^2 = \dots + \frac{(1-3)^2}{3} + \dots = \dots + \frac{4}{3} + \dots$

(5) $K = \sqrt{\frac{17.067}{50 + 17.067}} \approx 0.5045 \Rightarrow K^* = \sqrt{3/2} \cdot K \approx 0.618$

(6)

schlecht	mittel	gut
0.3	0.45	0.25

Aufgabe 3

14 Punkte

In einer Firma arbeiten 40 % mehr Männer als Frauen. Es wird ein Team aus zwei Beschäftigten zufällig ausgewählt.

In den Teilaufgaben a), b), c) betrage die Anzahl der Mitarbeiterinnen in dieser Firma 30.

- Wie viele Beschäftigte hat die Firma?
- Wie groß ist die Wahrscheinlichkeit, dass beide Ausgewählte Männer sind?
- Wie groß ist die Wahrscheinlichkeit, dass das Team aus einer Frau und einem Mann besteht?

Für die Teilaufgabe d) ist die Anzahl der Mitarbeiterinnen und der Mitarbeiter der Firma unbekannt.

- Wie viele Frauen und wie viele Männer arbeiten in der Firma, wenn die Wahrscheinlichkeit, dass das ausgewählte Zweierteam aus einem Mann und einer Frau besteht 50 % beträgt?

Lösungshinweis:

a) Anzahl Frauen: $F = 30$. Männer: $M = 1.4 \cdot F = 42$. Insgesamt: $F + M = 72$.

b) M_1 : Erster Ausgewählter ist ein Mann, M_2 : Zweiter Ausgewählter ist ein Mann, analog F_1, F_2 für die Frauen.

$$P(M_1 \cap M_2) = \frac{42}{72} \cdot \frac{41}{71} \approx 0.3368545.$$

c) $P((M_1 \cap F_2) \cup (F_1 \cap M_2)) = \frac{42}{72} \cdot \frac{30}{71} + \frac{30}{72} \cdot \frac{42}{71} \approx 0.4929577$

d) $M = 1.4 \cdot F$. Also ist die Anzahl aller Beschäftigten $= F + M = F + 1.4 \cdot F = 2.4 \cdot F$. Damit:

$$\begin{aligned} P((M_1 \cap F_2) \cap (F_1 \cap M_2)) &= \frac{1.4 \cdot F}{2.4 \cdot F} \cdot \frac{F}{2.4 \cdot F - 1} + \frac{F}{2.4 \cdot F} \cdot \frac{1.4 \cdot F}{2.4 \cdot F - 1} \\ &= 2 \cdot \frac{1.4 \cdot F}{2.4 \cdot (2.4 \cdot F - 1)} = 0.5 \end{aligned}$$

$$\Leftrightarrow 2.8F = 1.2 \cdot (2.4F - 1)$$

$$\Leftrightarrow 1.2 = 0.08F$$

$$\Leftrightarrow F = 15, M = 1.4 \cdot 15 = 21.$$

Insgesamt 36 Beschäftigte.

Aufgabe 4

9 Punkte

Gegeben ist eine Zufallsvariable X mit stetiger Verteilung und der zugehörigen Dichtefunktion $f: \mathbb{R} \rightarrow \mathbb{R}$ sowie ihr Graph in Abbildung 1.

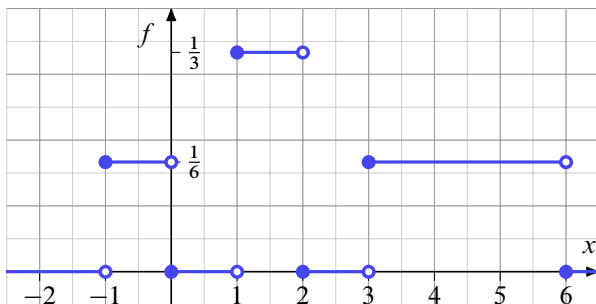


Abbildung 1: Graph von f

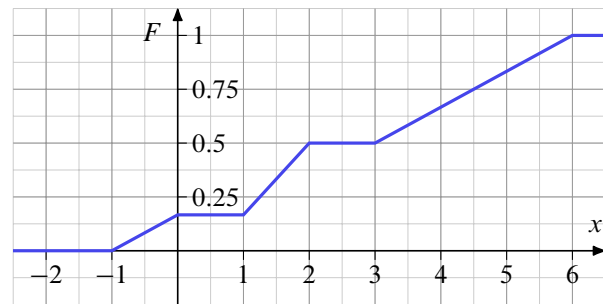


Abbildung 2: Graph von F

a) Zeichnen Sie in Abbildung 2 zu X den Graph der Verteilungsfunktion F .

Ermitteln Sie, falls möglich, die nachfolgenden Wahrscheinlichkeiten:

b) $P(X \leq 0)$

c) $P(X \leq 1)$

d) $P(X < 1)$

e) $P(X \geq 5)$

f) $P(1 \leq X \leq 3)$

g) $P(X \leq 3 \cap X \geq 2)$

h) $P(X \leq 2 \cap X \geq 3)$

i) $P(X \geq 5 | X \leq 6)$

Lösungshinweis:

a) Lösung siehe Zeichnung.

b) $P(X \leq 0) = 1/6$

c) $P(X \leq 1) = 1/6$

d) $P(X < 1) = 1/6$

e) $P(X \geq 5) = 1/6$

f) $P(1 \leq X \leq 3) = F(3) - F(1) = 1/2 - 1/6 = 1/3$

g) $P(X \leq 3 \cap X \geq 2) = F(3) - F(2) = 0.5 - 0.5 = 0$

h) $P(X \leq 2 \cap X \geq 3) = P(\{\}) = 0$

i) $P(X \geq 5 | X \leq 6) = \frac{1 - F(5)}{F(6)} = P(5 \leq X \leq 6) = 1/6$

Aufgabe 5

12 Punkte

In der Woche vor einer großen Wahl werden von einem Meinungsinstitut 1900 Wahlberechtigte befragt, welche Partei sie am kommenden Wochenende wählen werden. 646 der Befragten gaben an, dass sie Partei A wählen würden.

Gehen Sie im folgenden davon aus, dass es bei Umfragen und Wahlen nur gültige Stimmen gibt und sich jeder Wahlberechtigte für genau eine Partei entscheidet.

- a) Bestimmen Sie das 95%-Konfidenzintervall für das Wahlergebnis von Partei A .
- R** b) Geben Sie einen R-Befehl an, mit dem $x_{1-\alpha/2}$ aus Teilaufgabe a) berechnet werden kann.
- c) Wie groß müsste die Stichprobe der Befragten mindestens sein, um das Wahlergebnis einer beliebigen Partei zu einem Konfidenzniveau von 99 % auf plus/minus ein Prozentpunkt vorherzusagen?

Lösungshinweis:

- a) Mit der Normalverteilungsnäherung ergibt sich

$$\begin{aligned}n &= 1900, \bar{x}_A = \frac{646}{1900} = 0.34, \\ \hat{\sigma} &= \sqrt{\bar{x}_A \cdot (1 - \bar{x}_A)} = \sqrt{0.34 \cdot (1 - 0.34)} \approx 0.4737088, \\ c &= x_{0.975} \approx 1.96\end{aligned}$$

Damit ergibt sich: $\left[\bar{x}_A - \frac{\hat{\sigma}c}{\sqrt{n}}, \bar{x}_A + \frac{\hat{\sigma}c}{\sqrt{n}} \right] \approx [0.3187, 0.3613]$.

- R** b) `qnorm(0.975)`

```
## [1] 1.959964
```

- c) Gefordert ist $\hat{\sigma}c/\sqrt{n} \leq 0.01$. Da die Stichprobe nicht bekannt ist, rechnet man mit der höchstmöglichen Varianz $\hat{\sigma} = \sqrt{0.5 \cdot (1 - 0.5)} = 0.5$. Mit $c = x_{0.995} \approx 2.58$ ergibt sich:

$$\frac{\hat{\sigma}c}{\sqrt{n}} \leq 0.01 \Leftrightarrow n \geq \left(\frac{\hat{\sigma}c}{0.01} \right)^2 \approx 16641$$

also sind mindestens 16 641 Befragte nötig.

Aufgabe 6

16 Punkte

Gegeben sind die drei Funktionen $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ sowie $g, h : \mathbb{R} \rightarrow \mathbb{R}$. Es gilt $f(x, y) = g(x) + h(y)$. Bei g, h handelt es sich jeweils um ein Polynom.

In Abbildung 3 sind alle Wertepaare (x, y) eingezeichnet, für die $f_y(x, y) = 0$ gilt (gestrichelte Linien).

- Geben Sie ein Polynom für $h(y)$ an, welches für $f_y(x, y) = h'(y)$ die eingezeichneten Nullstellen besitzt.
- Zeichnen Sie in die Abbildung alle Punkte (x, y) ein mit $f_x(x, y) = 0$, wenn gilt:

$$g(x) = \frac{1}{3}x^3 + \frac{1}{2}x^2 - 6x.$$

- Geben Sie alle kritischen Punkte (x, y) , also die Nullstellen des Gradienten von f an.

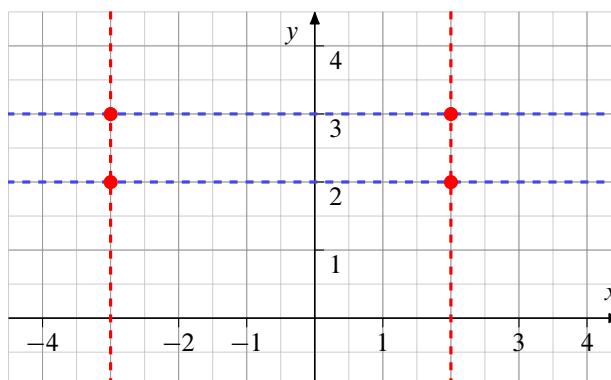


Abbildung 3: Nullstellen von f_y, f_x

- Berechnen Sie die zweiten partiellen Ableitungen $f_{xy}(x, y)$ und $f_{xx}(x, y)$.
- Es gilt $f_{yy}(x, y) = 2y - 5$. Bestimmen Sie Lage und Art aller Extrema von f .
- Angenommen g und h wären beides nach oben geöffnete Normalparabeln. Wie viele Extrema besäße f dann maximal (Begründung) und von welcher Art wären diese jeweils im Falle ihrer Existenz?

Lösungshinweis:

- Zum Beispiel: $h'(y) = (y - 2)(y - 3) = y^2 - 5y + 6$. Damit $h(y) = \frac{1}{3}y^3 - \frac{5}{2}y^2 + 6y + C$.
- $g'(x) = x^2 + x - 6 = (x - 2)(x + 3) = 0$ für $x = -3, 2$. Und $y \in \mathbb{R}$ beliebig. Siehe Zeichnung.
- Kritische Punkte: Schnittpunkte der Linien, also $(x, y) = (-3, 2), (-3, 3), (2, 2), (2, 3)$.
- $f_{xx}(x, y) = 2x + 1$ und $f_{xy}(x, y) = 0$.

$$e) \quad H_f = \begin{pmatrix} 2x + 1 & 0 \\ 0 & 2y - 5 \end{pmatrix}$$

$$H_f(-3, 2) = \begin{pmatrix} -5 & 0 \\ 0 & -1 \end{pmatrix} \rightarrow \text{Maximum} \quad H_f(-3, 3) = \begin{pmatrix} -5 & 0 \\ 0 & 1 \end{pmatrix} \rightarrow \text{Sattel}$$

$$H_f(2, 2) = \begin{pmatrix} 5 & 0 \\ 0 & -1 \end{pmatrix} \rightarrow \text{Sattel} \quad H_f(2, 3) = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix} \rightarrow \text{Minimum}$$

- Eine Normalparabel ist vom Grad 2. Die Ableitungen sind damit vom Grad 1. Damit gibt es genau eine kritische Stelle. Da die Normalparabel nach oben geöffnet ist, sind die Koeffizienten des quadratischen Terms von g, h jeweils positiv. Damit ist $f_{xx}(x, y) > 0$ und $f_{yy}(x, y) > 0$. Es handelt sich um ein Minimum.