

Klausur Statistik

Lösungshinweise

Prüfungsdatum: 20. Januar 2022 – Prüfer: Etschberger, Henle, Henle, Jansen
Studiengang: BW, IM
Punkte: 14, 12, 19, 12, 11, 13, 9 ; Summe der Punkte: 90

Aufgabe 1

14 Punkte

Gegeben ist die Funktion $f : D \rightarrow \mathbb{R}$ mit

$$f(x, y) = 2xy + y + \frac{5}{x} + 1.$$

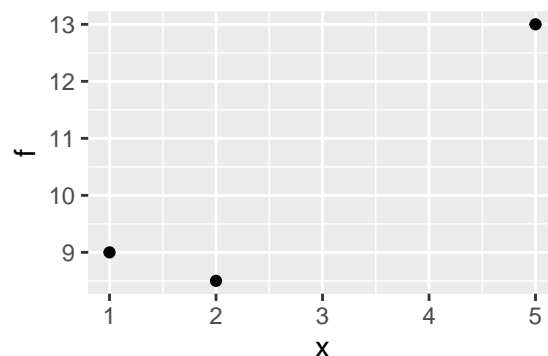
- a) Bestimmen Sie den maximalen Definitionsbereich D_f von f .
- R** b) Geben Sie einen R-Befehl an, der die Funktion unter dem Bezeichner f speichert.
- R** c) Skizzieren Sie die Grafik, die damit durch den folgenden R-Code ausgegeben wird:

```
tibble( x=c(1, 2, 5),
        y=c(1, 1, 1) ) %>%
  mutate(f = f(x, y)) %>%
  ggplot(aes(x, f)) +
  geom_point()
```

- d) Berechnen Sie den Gradient ∇f .
- e) Berechnen Sie alle Nullstellen des Gradienten.
- f) Berechnen Sie die Hessematrix H_f von f .
- g) Entscheiden Sie mittels H_f bzgl. der Nullstellen des Gradienten jeweils, ob es sich um ein Minimum, ein Maximum oder einen Sattelpunkt handelt.

Lösungshinweis:

- a) $D = \{(x, y) \in \mathbb{R}^2 : x \neq 0\}$
- b) `f = function(x, y) {2*x*y + y + 5/x + 1}`
- c) siehe Grafik rechts
- d) $\nabla f = \begin{pmatrix} 2y - \frac{5}{x^2} \\ 2x + 1 \end{pmatrix}$
- e) Einzige Nullstelle: $(x, y) = (-0.1, 10)$.
- f) $H_f = \begin{pmatrix} 10x^{-3} & 2 \\ 2 & 0 \end{pmatrix}$, damit ist $\det H_f = -4$ und der Punkt $(-0.5, 10)$ ist ein Sattelpunkt.



Aufgabe 2

12 Punkte

Bei einer Umfrage an einer Hochschule wurden zufällig ausgewählte Studenten gefragt, wie oft sie bereits während Vorlesungen eingeschlafen sind. Die Umfrage ergab folgende Daten, die folgendermassen in einem R-Vektor gespeichert sind:

```
x = c(6, 0, 0, 0, 3, 1, 0, 0, 0, 1, 0, 0, 2, 0, 0, 1, 1, 1, 2, 1)
```

Geben Sie für die Teilaufgaben a), b), c) jeweils die Ausgabe der angegebenen R-Befehls-Sequenzen an:

R a) `mean(x)`

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

R c) `tibble(x) %>%
 group_by(x) %>%
 summarize(h.j=n()) %>%
 rename(a.j=x) %>%
 arrange(a.j) %>%
 mutate(H.j = cumsum(h.j)) %>%
 mutate(f.j = h.j/sum(h.j),
 F.j = H.j/sum(h.j))`

R d) Geben Sie R-Befehle an, mit welchen ein Balkendiagramm der Daten gezeichnet wird.

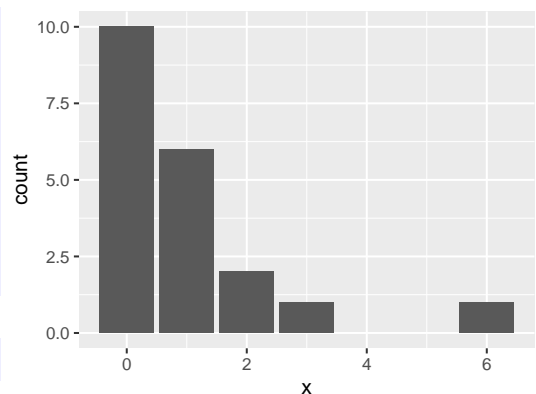
Lösungshinweis:

a) `## [1] 0.95`

b) `## 25% 50% 75%
0.0 0.5 1.0`

c) `## # A tibble: 5 x 5
a.j h.j H.j f.j F.j
<dbl> <int> <int> <dbl> <dbl>
1 0 10 10 0.5 0.5
2 1 6 16 0.3 0.8
3 2 2 18 0.1 0.9
4 3 1 19 0.05 0.95
5 6 1 20 0.05 1`

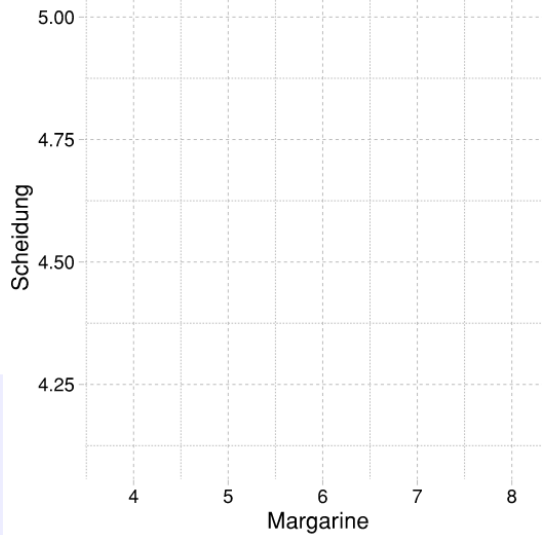
d) `tibble(x) %>% ggplot(aes(x)) + geom_bar()`



In den USA hat sich gezeigt, dass es in den Jahren 2000 bis 2009 im Bundesstaat Maine einen Zusammenhang zwischen der Anzahl der jährlichen Scheidungen pro 1000 Einwohner und dem jährlichen durchschnittlichen pro-Kopf-Konsum von Margarine in Pfund gibt. Folgende Daten wurden dazu erfasst:

Jahr	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Scheidung	5.0	4.7	4.6	4.4	4.3	4.1	4.2	4.2	4.2	4.1
Margarine	8.2	7.0	6.5	5.3	5.2	4.0	4.6	4.5	4.2	3.7

- a) Zeichnen Sie in das nebenstehende Koordinatensystem den Schwerpunkt der Daten ein.
- b) Berechnen Sie den Wert eines geeigneten Zusammenhangsmaßes zwischen den beiden Merkmalen.
- c) Gehen Sie davon aus, dass die Daten in der R-Variable `Maine` mit den beiden Merkmalen `Scheidung` und `Margarine` gespeichert sind. Was gibt folgende R-Sequenz aus?



```
Maine.lm = lm(Scheidung ~ Margarine,
              data=Maine)
Maine.ab = Maine.lm$coefficients
a = Maine.ab[1]
b = Maine.ab[2]
cat("Scheidung = ", a,
    " + ", b,
    " * Margarine")
```

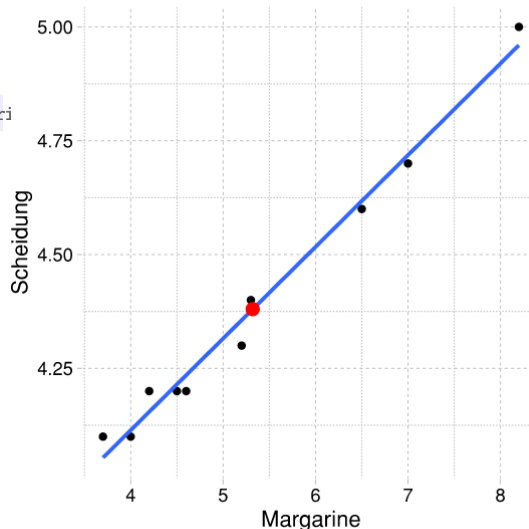
- d) Zeichnen Sie in das Koordinatensystem oben die Ausgabe folgender R-Sequenz ein:

```
Maine %>% ggplot(aes(Margarine, Scheidung)) +
  geom_smooth(method = "lm", se = FALSE)
```

- e) Bestimmen Sie den Determinationskoeffizient des Regressionsmodells aus c) und interpretieren Sie diesen.
- f) Angenommen der pro-Kopf-Margarinekonsum in Maine erreicht in einem Jahr den Wert 10 Pfund. Welche Scheidungsrate erwarten Sie nach diesem Modell?
- g) Erklären Sie in maximal 20 Worten den Unterschied zwischen Kausalität und Korrelation. (Hinweis: ab dem 21. Wort wird Ihre Erklärung in der Bewertung ignoriert.)

Lösungshinweis:

- a) siehe nebenan.
- b) Bravais-Pearson-Korrelationskoeffizient:
 $r = 0.9925585$
- c) `## Scheidung = 3.308626 + 0.201386 * Margari`
- d) siehe nebenan (Regressionsgerade)
- e) $R^2 = r^2 \approx 0.9851723$, das bedeutet ca. 98.5% der Information in den Daten können durch das Modell beschrieben werden, sagt aber nichts über den kausalen Zusammenhang oder die Prognosequalität des Modells aus.
- f) $Scheidung(10) \approx 3.309 + 0.201 \cdot 10 \approx 5.32$
- g) Korrelation misst Stärke des Zusammenhangs in Daten, sagt aber nichts zu ursächlichen Abhängigkeiten, dazu ist Kausalität nötig.



In dieser Aufgabe werden Kugeln aus einer Urne gezogen. Alle Kugeln haben die gleiche Größe und Kugeln einer Farbe sind nicht zu unterscheiden. Zieht man Kugeln aus der Urne, so wird jede Kugel mit der gleichen Wahrscheinlichkeit gezogen.

Hinweis: Alle Aufgabenteile sind unabhängig voneinander.

- a) In einer Urne sind 3 blaue Kugeln, 4 rote Kugeln und 8 schwarze Kugeln.
 - i) Es werden alle Kugeln nacheinander aus der Urne gezogen und in eine Reihe gelegt. Wie viele verschiedene Anordnungen gibt es?
 - ii) Es werden zwei Kugeln gezogen. Wie wahrscheinlich ist es, dass diese beide schwarz sind?
 - iii) Es werden zwei Kugeln gezogen. Wie wahrscheinlich ist es, dass diese die gleiche Farbe haben?
 - iv) Es werden zwei Kugeln gezogen. Sie wissen, dass die erste Kugel nicht blau ist. Wie wahrscheinlich ist es als zweite eine rote Kugel zu ziehen?
- b) In einer zweiten Urne sind 3 rote und eine unbekannte Menge schwarzer Kugeln. Nun werden zwei Kugeln aus der Urne gezogen. Für welche Anzahl(en) an schwarzen Kugeln ist es genauso wahrscheinlich zwei gleichfarbige Kugeln zu ziehen wie zwei unterschiedlich farbige?

Lösungshinweis:

a) i) Mississippi-Formel: $\frac{(3 + 4 + 8)!}{3! \cdot 4! \cdot 8!} = 225\,225$.

ii) Für zwei blaue Kugeln gilt (kann man z.B. mit Hilfe eines Baumdiagramms gut sehen):

$$P = \frac{8}{15} \cdot \frac{7}{14} = \frac{4}{15} = 0,2\bar{6}$$

iii) Analog zum vorherigen Teil, allerdings müssen die blauen und roten Kugeln auch berücksichtigt werden:

$$P = \frac{8}{15} \cdot \frac{7}{14} + \frac{3}{15} \cdot \frac{2}{14} + \frac{4}{15} \cdot \frac{3}{14} = \frac{37}{105} \approx 0.3524$$

iv)

$$P = \underbrace{\frac{4}{15} \cdot \frac{3}{14}}_{\text{erste Kugel rot}} + \underbrace{\frac{8}{15} \cdot \frac{4}{14}}_{\text{erste Kugel schwarz}} = \frac{22}{105} \approx 0.2095$$

b) Die Wahrscheinlichkeit für zwei gleiche Kugeln ist:

$$P(n) = \underbrace{\frac{3}{3+n} \cdot \frac{2}{2+n}}_{\text{zwei rote Kugeln}} + \underbrace{\frac{n}{3+n} \cdot \frac{n-1}{2+n}}_{\text{zwei schwarze Kugeln}} = \frac{n^2 - n + 6}{n^2 + 5n + 6}$$

Nun muss gelten $P(n) = 0.5$ und damit ergibt sich:

$$n^2 - n + 6 = 0,5 \cdot (n^2 + 5n + 6) \Leftrightarrow n^2 - 7n + 6 = 0$$

Die quadratische Gleichung hat die beiden Lösungen $n = 1$ und $n = 6$. Also bei einer oder sechs schwarzen Kugeln.

Anmerkung: Bei zwei bis fünf Kugeln ist die Wahrscheinlichkeit zwei gleiche zu ziehen kleiner als 50% bei null oder mehr als sechs Kugeln ist die Wahrscheinlichkeit zwei gleiche zu ziehen größer als 50%.

Aufgabe 5

11 Punkte

- a) Ein Glücksrad ist in 11 gleich große Abschnitte unterteilt, welche mit den Zahlen von 1 bis 11 durchnummeriert sind.
- Wie groß ist die Wahrscheinlichkeit, dass bei achtmaligem Drehen das Rad zweimal im Feld mit der Zahl 9 stehen bleibt?
 - Wie groß ist die Wahrscheinlichkeit, dass bei neunmaligem Drehen das Rad mindestens einmal in einem Feld mit einer geraden Zahl stehen bleibt?
- b) Jedes Sommersemester schreiben mehrere hundert Studentinnen und Studenten eine Statistik-klausur, aber nur wenige von ihnen erreichen die volle Punktzahl. In den letzten Jahren waren das im Durchschnitt pro Sommersemester nur 0.2 Leute. Wie groß ist die Wahrscheinlichkeit, dass dieses Semester mindestens eine Person volle Punktzahl erzielt?
- R** c) Geben Sie einen R-Befehl an, mit dem man die Lösung von Aufgabe Teilaufgabe b) ausgeben kann.
- d) Gegeben ist die diskrete Zufallsvariable X mit der Wahrscheinlichkeitsfunktion

$$f(x) = \begin{cases} c \cdot x & \text{für } x = 1, 2, 3, 4 \\ 0 & \text{sonst.} \end{cases}$$

Ermitteln Sie die Konstante $c \in \mathbb{R}$ und bestimmen Sie den Wert der zugehörigen Verteilungsfunktion für $x = 3.8$, also $F(3.8)$.

Lösungshinweis:

- a) (i) Binomialverteilung:
 $X \sim B(n = 8, p = \frac{1}{11})$:

$$P(X = 2) = \binom{8}{2} \cdot \left(\frac{1}{11}\right)^2 \cdot \left(\frac{10}{11}\right)^6 \approx 0.1306$$

- (ii) Binomialverteilung:
 $X \sim B(n = 9, p = 5/11)$. Damit:

$$P(X \geq 1) = 1 - P(X = 0) \approx 1 - 0.004 \approx 0.996$$

- b) Poissonverteilung: $X \sim \text{Pois}(\lambda = 0.2)$:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{0.2^0 \cdot e^{-0.2}}{0!} \approx 0.1813$$

- c) `dpois(3, 2.53)`
`## [1] 0.2150034`

- d) Es gilt $c = 0.1$ und

$$F(x) = \begin{cases} 0 & \text{für } x < 1 \\ 0.1 & \text{für } 1 \leq x < 2 \\ 0.3 & \text{für } 2 \leq x < 3 \\ 0.6 & \text{für } 3 \leq x < 4 \\ 1 & \text{für } x \geq 4 \end{cases}$$

also $F(3.8) = 0.6$

Aufgabe 6

13 Punkte

Eine Zufallsvariable Y sei normalverteilt. Von Y sind zwei Tatsachen bekannt:

$$\sigma = 200, \quad \mu = 600$$

Zu berechnen sind damit

- a) $P(Y = \sigma)$,
- b) x , wenn gilt, dass $P(Y \leq x) = 0.11314$,
- c) $P(Y < 400 \vee Y \geq 800)$,
- d) $P(\mu - 3\sigma \leq Y \leq \mu + 3\sigma)$.

- R** e) Geben Sie jeweils einen R-Befehl an, mit dem man das Ergebnis der Teilaufgaben a)-d) ausgeben kann.

Lösungshinweis:

- a) $P(Y = \sigma) = 0$,
- b) $P(Y \leq x) = F(x) = \Phi\left(\frac{x - 600}{200}\right) = 0.11314$,
 $\Leftrightarrow \Phi\left(\frac{600 - x}{200}\right) = 1 - 0.11314 = 0.88686$
 $\Leftrightarrow \frac{600 - x}{200} = 1.21$
 $\Leftrightarrow x = 600 - 200 \cdot 1.21 = 358$
- c) $P(Y < 400 \vee Y \geq 800) = 1 - P(800 < Y < 400) = 1 - (\Phi(1) - \Phi(-1))$,
 $= 1 - (\Phi(1) - 1 + \Phi(1))$
 $= 2(1 - \Phi(1)) \approx 2 \cdot (1 - 0.84134)$
 ≈ 0.3174
- d) $P(\mu - 3\sigma \leq Y \leq \mu + 3\sigma) = 2\Phi(3) - 1 \approx 0.9973002$.

- e)

```
pnorm(200, mean=600, sd=600) - pnorm(200, mean=600, sd=600)
qnorm(0.11314, mean=600, sd=200)
1-(pnorm(800, mean=600, sd=200) - pnorm(400, mean=600, sd=200))
pnorm(600+3*200, mean=600, sd=200) - pnorm(600-3*200, mean=600, sd=200)
```

```
## [1] 0
## [1] 358.0006
## [1] 0.3173105
## [1] 0.9973002
```

- a) Beantworten Sie die folgenden Fragen mit *wahr* oder *falsch*.
- Ein Fehler 1. Art kann auftreten, wenn die Testfunktion einen Wert des Ablehnungsbereiches der H_0 annimmt.
 - Ein Fehler 2. Art kann auftreten, wenn die Testfunktion einen Wert des Ablehnungsbereiches der H_0 annimmt.
 - Wird bei einem Test H_0 verworfen, dann wurde H_1 statistisch nachgewiesen.
 - Der Ablehnungsbereich der Nullhypothese bei einem Test wird durch die Vorgabe der Wahrscheinlichkeit für einen Fehler zweiter Art festgelegt.
 - Führt der Test einer Hypothese zu ihrer Ablehnung, so ist damit bewiesen, dass diese Hypothese falsch ist.
 - Die Wahrscheinlichkeit eines Fehlers 1. Art bei statistischen Tests ist immer kleiner als die Wahrscheinlichkeit eines Fehlers 2. Art.
- b) Getränkehändler Tristo prüft seine Lieferung an Glasflaschen immer sorgfältig. Bei der Überprüfung von 1000 Glasflaschen fällt ihm auf, dass 18 davon zerbrochen sind.
- Berechnen Sie ein Konfidenzintervall zum Konfidenzniveau 95 % für den Anteil zerbrochener Flaschen.
 - Geben Sie einen R-Befehle an, mit dem $x_{1-\alpha/2}$ aus Teilaufgabe b.i) berechnet werden kann.

R

Lösungshinweis:

a) wahr, falsch, wahr, falsch, falsch, falsch

b) KI = [0,0097 ; 0,0263]

c) `qnorm(0.975)`

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