

# Nachholklausur Statistik

## Lösungshinweise

Prüfungsdatum: 25. Januar 2021 – Prüfer: Etschberger, Henle, Jansen

Studiengang: BW, IM

Punkte: 15, 15, 15, 15, 15 ; Summe der Punkte: 90

### Aufgabe 1

15 Punkte

Für die Teilaufgaben a), b), c) sei die Funktion  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  gegeben mit

$$f(x, y) = 3xe^{-y^2} - x.$$

- Berechnen Sie den Funktionswert  $f(3, 0)$ .
- Bestimmen Sie  $\nabla f$ , also den Gradient von  $f$ .
- Berechnen Sie alle Nullstellen von  $\nabla f$ .

Für die Teilaufgabe d) wird eine Funktion  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$  betrachtet, deren Hessematrix  $H_g$  folgendermaßen gegeben ist:

$$H_g(x, y) = \begin{pmatrix} 2y & 2x + 2y - 2 \\ 2x + 2y - 2 & 2x - 6y \end{pmatrix}.$$

Zudem sind die beiden einzigen Nullstellen des Gradienten von  $g$  bekannt, die für  $(x, y)$  bei den Punkten:  $(0, 0)$  und  $(2, 0)$  liegen.

- Untersuchen Sie  $g$  auf lokale Extrema/Sattelpunkte.

### Lösungshinweis:

- $f(3, 0) = 6$
- $\nabla f(x, y) = \begin{pmatrix} 3e^{-y^2} - 1 \\ 3xe^{-y^2} \cdot (-2y) \end{pmatrix}$
- Zwei Nullstellen:  
 $(x, y) = (0, \pm \sqrt{\ln 3})$
- Beide kritischen Punkte sind Sattel.

## Aufgabe 2

15 Punkte

10 Studierende werden im 3. Semester befragt, wieviel Stunden Sie im letzten halben Jahr durchschnittlich pro Tag zur Unterhaltung vor dem Bildschirm verbracht haben. Zum Merkmal

$X$ : Anzahl Stunden Bildschirmunterhaltung pro Tag

ergibt sich aus den Umfrageergebnissen folgende unvollständige empirische Verteilungsfunktion:

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0.2 & \text{für } 0 \leq x < 2 \\ 0.4 & \text{für } 2 \leq x < 3 \\ 0.7 & \text{für } 3 \leq x < 5 \\ 0.9 & \text{für } 5 \leq x < 6 \\ 1.0 & \text{für } 6 \leq x \end{cases}$$

Außerdem ist bekannt, dass

- ▶ die Urliste zu  $X$  genau 5 ganzzahlige Ausprägungen enthält,
- ▶ die Spannweite von  $X$  gleich 6,
- ▶ das empirische Quantil  $\tilde{x}_{0,2} = 1$  und
- ▶ die kumulierte absolute Häufigkeit  $H(5) = 9$  beträgt.

a) Ergänzen Sie die fehlenden Angaben von  $F(x)$  oben in den leeren Feldern.

Falls Sie Teilaufgabe a) nicht lösen können (und nur dann) rechnen Sie bitte mit folgendem (falschen) Zwischenergebnis weiter:

$$F(x) = \begin{cases} 0 & \text{für } x < 1.5 \\ 0.30 & \text{für } 1.5 \leq x < 2.7 \\ 0.50 & \text{für } 2.7 \leq x < 3 \\ 0.75 & \text{für } 3 \leq x < 3.5 \\ 0.85 & \text{für } 3.5 \leq x < 10 \\ 1.00 & \text{für } 10 \leq x \end{cases}$$

- b) Skizzieren Sie den Graph der empirischen Verteilungsfunktion und zeichnen Sie den Median von  $X$  ein.
- c) Erstellen Sie eine Tabelle mit den Ausprägungen  $a_i$ , den absoluten Häufigkeiten  $h_i$  sowie den kumulierten absoluten Häufigkeiten  $H_i$  von  $X$ .

Für die Teilaufgaben d) und e) wird nun ein anderes Merkmal  $Y$  betrachtet. Zu diesem Merkmal ist die Urliste  $y = (0, 1, 1, 3, 5)$  gegeben.

d) Geben Sie die Koordinaten der Knickpunkte der Lorenzkurve zu  $Y$  an.

- R** e) Schreiben Sie R-Befehle auf, um das 30%-Quantil  $\tilde{y}_{0,3}$  von  $Y$  zu berechnen. Vergessen Sie dabei nicht zuerst die Urliste einzugeben.

## Lösungshinweis:

a) s.o.

b)

$a_i$	0	2	3	5	6
$h_i$	2	2	3	2	1
$H_i$	2	4	7	9	10

c) Knicke bei (0.2, 0), (0.6, 0.2), (0.8, 0.5)

```
d) x=c(0, 1, 1, 3, 5)
quantile(x, probs=0.3, type=2)
## 30%
## 1
```

### Aufgabe 3

15 Punkte

Maria ist Kundin bei den Streamanbieter *Tenxilf* und *Snidey Plus*. Sie besitzt dort aber keine Flat-Rate, sondern hat in den letzten 4 Jahren jeweils Stundenkontingente mit den Mengen ( $q$ ) in Stunden zu den Preisen ( $p$ ) in Euro pro Stunde gekauft. Darüber führt sie akribisch Buch:

Jahr	1		2		3		4	
	$p_1$	$q_1$	$p_2$	$q_2$	$p_3$	$q_3$	$p_4$	$q_4$
<i>Tenxilf</i>	0.50	300	0.55	240	0.60	260	0.75	200
<i>Snidey Plus</i>	0.60	340	0.55	300	0.65	200	0.80	240

- Berechnen Sie den Preisindex von Fisher und den Preisindex von Lowe jeweils zur Berichtszeit 4 und der Basiszeit 1.
  - Berechnen Sie anhand der Ergebnisse aus a) jeweils die jährliche durchschnittliche Preiserhöhung von Marias Nutzungsprofil.
  - Maria will nun wissen, ob zwischen den Preisen der beiden Anbieter ein Zusammenhang besteht. Berechnen Sie ein geeignetes Zusammenhangsmaß und interpretieren Sie das Ergebnis.
  - Bestimmen Sie den Funktionsterm eines linearen Regressionsmodells, indem der Stundenpreis von *Tenxilf* in Abhängigkeit vom Preis von *Snidey Plus* beschrieben wird.
- R**
- Geben Sie R-Befehle an, mit dem die Daten in Variablen abgelegt werden und das Regressionsmodell aus d) erzeugt und ausgegeben wird.
  - Wie hoch wäre nach dem Modell aus d) der Preis einer Stunde bei *Tenxilf*, wenn *Snidey Plus* den Stundenpreis auf 0.9 € erhöht?

### Lösungshinweis:

$$\text{a) } P_{1,4}^L = \frac{p_4^1 q_1^1 + p_4^2 q_1^2}{p_1^1 q_1^1 + p_1^2 q_1^2} \approx 1.403\,954\,8 \quad P_{1,4}^P = \frac{p_4^1 q_4^1 + p_4^2 q_4^2}{p_1^1 q_4^1 + p_1^2 q_4^2} \approx 1.401\,639\,3$$

$$P_{1,4}^F = \sqrt{P_{1,4}^L \cdot P_{1,4}^P} \approx 1.402\,796\,6$$

$$\text{Lowe: Durchschnittliche Mengen: } \bar{q}^1 \approx 250, \quad \bar{q}^2 \approx 270$$

$$P_{1,4}^{\text{Lo}} = \frac{p_4^1 \bar{q}^1 + p_4^2 \bar{q}^2}{p_1^1 \bar{q}^1 + p_1^2 \bar{q}^2} \approx 1.405\,923\,3$$

- b) In Prozent durchschn. pro Jahr:

$$\text{Fisher: } \left( \sqrt[3]{P_{1,4}^F} - 1 \right) \cdot 100 \approx 11.94$$

$$\text{Lowe: } \left( \sqrt[3]{P_{1,4}^{\text{Lo}}} - 1 \right) \cdot 100 \approx 12.03$$

- c)  $r \approx 0.9285714$

d)  $\hat{y}(x) = -0.0035714 + 0.9285714 \cdot x$

- e) `p.1 = c(0.50, 0.55, 0.60, 0.75)`  
`p.2 = c(0.60, 0.55, 0.65, 0.80)`  
`lm(p.1 ~ p.2)`

f)  $\hat{y}(15) = -0.0035714 + 0.9285714 \cdot 0.9 \approx 0.83$

## Aufgabe 4

15 Punkte

In einem französischen Kartenblatt befinden sich 52 Karten, jeweils 2 bis 10, Bube, Dame, König, Ass in vier Farben: Kreuz, Pik, Herz und Karo.

Arnold und Boris denken über diese Karten nach. Bitte helfen Sie den beiden, indem sie die folgenden Anzahlen und Wahrscheinlichkeiten (auf mindestens 5 Nachkommastellen genau) berechnen:

Jeder der beiden zieht jeweils zufällig 2 Karten, die sogenannte Hand. Dabei spielt die Reihenfolge der gezogenen Karten keine Rolle.

- a) Beachtet man nur die Werte und ignoriert die Farben, wie viele verschiedene Hände kann Arnold ziehen?

Wie wahrscheinlich ist es, dass Arnold

- b) ein beliebiges Paar (zwei Karten mit den gleichen Werten, z.B. (9,9) oder (D,D) ) zieht?  
c) ein Paar zieht, wenn er weiss, dass Boris kein Paar auf der Hand hat?  
d) ein Paar zieht, wenn er weiss, dass Boris ein Paar auf der Hand hat?

Die beiden legen alle Karten zurück und mischen das Spiel. Dann zieht Boris zufällig drei Karten aus dem kompletten Kartenstapel.

- e) Wie wahrscheinlich ist es, dass er nur rote Karten (Herz oder Karo) gezogen hat?

### Lösungshinweis:

- a) Es gibt  $N = \frac{1}{2} \cdot 12 \cdot 13 + 13 = 91$  Möglichkeiten. Die ersten 78 Möglichkeiten sind die bei denen die Karten verschieden sind, sowie 13 bei denen die beiden Karten gleich sind.  
b) Man gibt die Karten hintereinander; welchen Wert die erste Karte hat spielt hierbei keine Rolle, bei der zweiten machen 3 der verbleibenden 51 Karten die Hand zu einem Paar, so dass die Wahrscheinlichkeit

$$P(\text{Paar}) = \frac{3}{51} = \frac{1}{17} = 0,0588235294117647$$

für ein Paar ist.

- c) Hat der Gegner kein Paar, so gibt es zwei Möglichkeiten: die erste Karte ist eine Karte die der Gegner hat, dies geschieht mit einer Wahrscheinlichkeit von  $\frac{6}{50}$  oder nicht, was mit einer Wahrscheinlichkeit von  $\frac{44}{50}$  passiert. Damit ist die Wahrscheinlichkeit für ein Paar:

$$P(\text{Paar} | \text{Gegner kein Paar}) = \frac{6}{50} \cdot \frac{2}{49} + \frac{44}{50} \cdot \frac{3}{49} = \frac{72}{1225} \approx 0,05877571020408162165...$$

- d) Die Rechnung ist analog zum vorherigen, allerdings ist die Wahrscheinlichkeit für eine Karte die der Gegner bereits hat  $\frac{2}{50}$  und damit ergibt sich:

$$P(\text{Paar} | \text{Gegner Paar}) = \frac{2}{50} \cdot \frac{1}{49} + \frac{48}{50} \cdot \frac{3}{49} = \frac{73}{1225} \approx 0,0595918367346...$$

- e)

$$P = \frac{26}{52} \cdot \frac{25}{51} \cdot \frac{24}{50} = \frac{15600}{132600} = 0,11764...$$

## Aufgabe 5

15 Punkte

Maria guckt gerade das Finale der 20. und damit letzten Staffel ihrer Lieblingskrimiserie. Sie hat keine Folge verpasst und jedes mal aufgeschrieben, wie viele Leute jeweils in der aktuellen Folge umgebracht worden sind. Die Zufallsvariable  $X$  beschreibt damit die Anzahl der Morde bei einer zufällig ausgewählten Episode dieser Serie. Die Verteilung von  $X$  sieht folgendermaßen aus:

$x$	0	1	2	3	20
$P(X = x)$	0.1	0.5	0.2	0.19	0.01

Berechnen Sie damit

- den Erwartungswert  $E[X]$ ,
- den Modus von  $X$ ,
- den Median von  $X$ ,
- die Varianz  $\text{Var}[X]$ ,
- die Standardabweichung  $\text{Sta}[X]$ ,
- $P(|X - E[X]| < \text{Sta}[X])$ .
- Warum ist der Median hier kleiner als  $E[X]$ ?
- Tschebyschew hat entdeckt, dass die folgende Ungleichung für alle  $\varepsilon > 0$  gilt:

$$P(|X - E[X]| \geq \varepsilon) \leq \frac{\text{Var}[X]}{\varepsilon^2}$$

Überprüfen Sie die Gültigkeit der Ungleichung für die Mordanzahlverteilung und  $\varepsilon = 18$ .

### Lösungshinweis:

```
tibble(x = c(0,1,2,3,20), p = c(1,5,2,1.9,.1)/10) %>%
  summarize(E.x=sum(p*x),
            Var = sum(x^2*p) - sum(x*p)^2,
            Sta = sqrt(Var),
            Median = 1, # kleinster Wert mit F(x) > 1/2
            Modus = 1)

## # A tibble: 1 x 5
##   E.x  Var  Sta Median Modus
##   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  1.67  4.22  2.05     1     1

# P(|X - 1.67| < 2.02) = P({0, 1,2,3}) = 1-0.01 = 0.99
# Tschebyschew:
# P(|X - 1.67| > 18) = P({20}) = 0.01 < 4.221/(18^2)
```

## Aufgabe 6

15 Punkte

Der Osterhase hat in den letzten Jahrhunderten jeweils am Ostersonntag um 6.30 Uhr die Schneehöhe in cm (Merkmal  $X$ ) direkt vor seinem Bau gemessen. Aus diesen Daten wählt er 10 Messergebnisse zufällig aus. Er erhält damit die Stichprobe:

$k$	1	2	3	4	5	6	7	8	9	10
$x_k$	3	3	4	10	1	0	0	12	2	4

Gehen Sie im folgenden davon aus, dass es sich bei  $x$  um eine einfache Stichprobe einer normalverteilten Grundgesamtheit handelt.

- Bestimmen Sie ein Konfidenzintervall zum Konfidenzniveau 90% für die durchschnittliche Schneehöhe in der Grundgesamtheit.
- Interpretieren Sie das Ergebnis aus Teilaufgabe a).
- Schreiben und zeichnen Sie die komplette Ausgabe folgender R Befehle auf:

R

```
x = c(3, 3, 4, 10, 1, 0, 0, 12, 2, 4)
Konf.Niveau = 0.9
alpha = 1 - Konf.Niveau

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

tibble(x) %>%
  summarize(x.quer = mean(x),
            x.sd = sd(x),
            x.n = length(x),
            x.FG = length(x) - 1 ) %>%
  mutate(Fraktil = qt(1-alpha/2, df=x.FG) ) %>%
  mutate(KI.links = x.quer - Fraktil * x.sd / sqrt(x.n),
         KI.rechts = x.quer + Fraktil * x.sd / sqrt(x.n) )
```

(Hinweis: Die Antwort besteht aus einer Grafik und einer Tabelle.)

## Lösungshinweis:

- KI = [1.558, 6.242]
- Mit einer Wahrscheinlichkeit von 90 % enthält das Intervall den Mittelwert der Schneehöhe in der Grundgesamtheit.

```
c) ## # A tibble: 1 x 7
##   x.quer x.sd   x.n x.FG Fraktil KI.links KI.rechts
##   <dbl> <dbl> <int> <dbl> <dbl>   <dbl>   <dbl>
## 1     3.9  4.04    10     9   1.83    1.56    6.24
```

