Klausur Statistik
Lösungshinweise

Prüfungsdatum: 1. Juli 2020 – Prüfer: Etschberger, Henle, Jansen
Studiengang: BW, IM
Punkte: 15, 15, 15, 13, 19, 13 ;   Summe der Punkte: 90

Aufgabe 1 — 15 Punkte

Gegeben ist die reelle Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(x,y) = y(3x - y^2) - x^3$$

a) Bestimmen Sie den Gradienten von f.

b) Bestimmen Sie alle Punkte, an denen der Gradient von f gleich 0 ist.

c) Bestimmen Sie die Hessematrix $H_f(x,y)$.

d) Prüfen Sie, ob es sich bei den in Teilaufgabe b) berechneten kritischen Stellen um Maxima, Minima oder Sattelpunkte handelt.

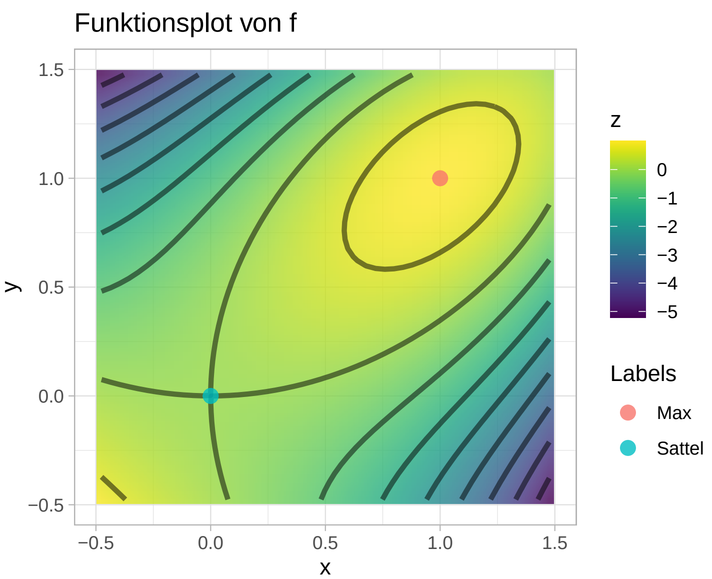
Lösungshinweis:

a) $\nabla f(x,y) = \begin{pmatrix} 3y - 3x^2 \\ 3x - 3y^2 \end{pmatrix}$

b) $(x, y) = (0, 0)$ und $(x, y) = (1, 1)$

c) $H_f(x,y) = \begin{pmatrix} -6x & 3 \\ 3 & -6y \end{pmatrix}$

d) $(x, y) = (0, 0)$: Sattelpunkt,
$(x, y) = (1, 1)$ Maximum,
$f(1, 1) = 1$



Aufgabe 2 — 15 Punkte

Karl hat 9 Personen nach der Anzahl der Gegenstände in ihrer Einkaufstasche gefragt. Leider hat er die Urliste verloren. Er kann sich aber noch an die folgenden Eckdaten erinnern:

$$\begin{aligned} \text{Spannweite} &= 8, & \text{Modus} &= 2 \text{ (eindeutig)}, \\ \text{Median} &= 3, & h(1) &= 1 \\ f(6) &= 2/9, & F(4) &\neq F(3)\,. \end{aligned}$$

Außerdem weiß Karl noch, dass jede befragte Person mindestens einen Gegenstand in ihrer Tasche dabei hatte.

a) Schreiben Sie die sortierte Urliste der Anzahl der Gegenstände in den Einkaufstaschen auf.

Für die Teilaufgaben b) - d) wird ein anderes Merkmal X mit folgender Urliste betrachtet:

$$x = (1, 1, 1, 2, 2, 5, 9)$$

b) Berechnen Sie die kumulierten relativen Häufigkeiten.

c) Geben Sie R-Befehle an, mit denen die empirische Verteilungsfunktion von x ausgegeben werden kann. Vergessen Sie nicht, zunächst die Urliste einem Vektor zuzuweisen.

d) Zeichnen Sie die empirische Verteilungsfunktion.

Lösungshinweis:

a) $x = (1, 2, 2, 2, 3, 4, 6, 6, 9)$

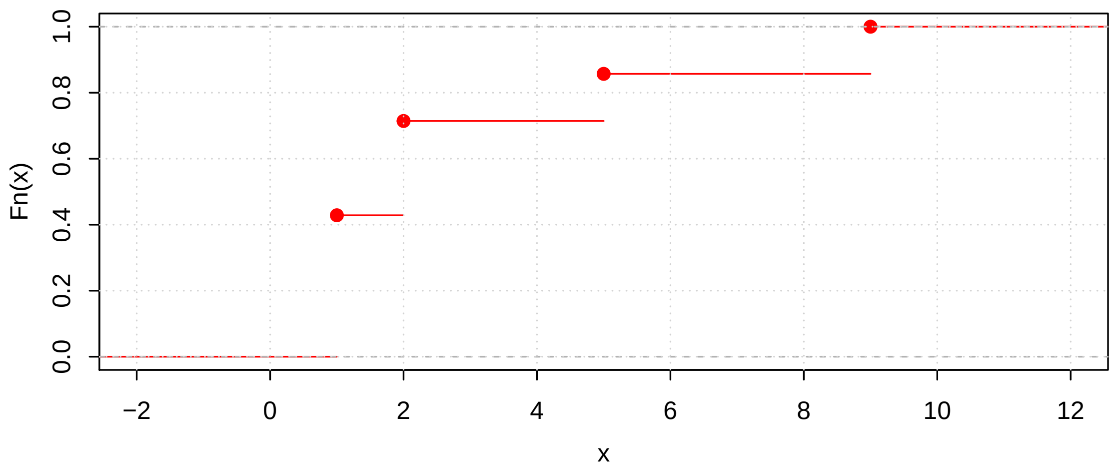
b) Häufigkeitstabelle:

ai	1	2	5	9
hi	3	2	1	1
Fi	$3/7$	$5/7$	$6/7$	1

c) Daten eingeben: `x.b = c(1, 1, 1, 2, 2, 5, 9)`

d) Grafik

```
x.b %>% ecdf %>% plot
```



Aufgabe 3 — 15 Punkte

In einem Park wurden 30 Passanten gefragt, ob Sie einen Hund besitzen (Merkmal X) und wie viele Stunden sie durchschnittlich pro Woche spazieren gehen (Merkmal Y).

Zu diesen Daten ist die folgende unvollständige Kontingenztabelle gegeben:

	Y			
X	[0, 1)	[1, 4.5)	[4.5, 6]	Summe
Ja	0	12	6	18
Nein	6	6	0	12
Summe	6	18	6	30

a) Ergänzen Sie die fehlenden Zahlen in der Kontingenztabelle.

b) Ergänzen Sie auch die beiden folgenden Tabellen

Bei Unabhängigkeit erwartete Häufigkeiten, entspricht $\tilde{h}_{ij}$:

	Y		
X	[0, 1)	[1, 4.5)	[4.5, 6]
Ja	3.6	10.8	3.6
Nein	2.4	7.2	2.4

Quadrierte Residuen, entspricht $\frac{\left(h_{ij}-\tilde{h}_{ij}\right)^2}{\tilde{h}_{ij}}$:

	Y		
X	[0, 1)	[1, 4.5)	[4.5, 6]
Ja	$18/5$	$2/15$	$8/5$
Nein	$27/5$	$1/5$	$12/5$

c) Es ergibt sich $\chi^2 = 40/3$ (diesen Wert müssen Sie nicht nachrechnen). Bestimmen Sie den normierten Kontingenzkoeffizienten K_*.

d) Erklären Sie in maximal 15 Worten allgemein die Bedeutung des normierten Kontigenzkoeffizienten (unabhängig vom Ergebnis der vorherigen Teilaufgaben).

e) Zeichnen Sie ein Histogramm der Dauer der durchschnittlichen wöchentlichen Spaziergänge aller 30 Personen.

R f) Angenommen die konkreten Angaben der Personen zur durchschnittlichen Spaziergangdauer pro Woche sind im R-dataframe `A.df` im Merkmal Y abgelegt. Geben Sie R-Befehle an, mit der das Histogramm aus Aufgabe e) ausgegeben werden kann.

Lösungshinweis:

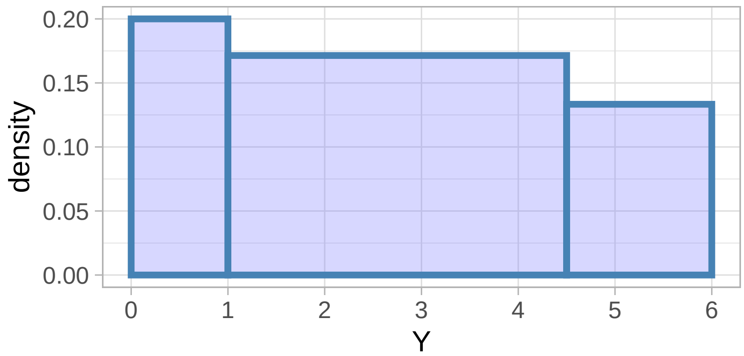
a) s.o.

b) s.o.

c) $K = \sqrt{\frac{40/3}{40/3+30}} \approx 0.5547002$
damit $K_* = K/\sqrt{1/2} \approx 0.7844645$

d) K_* misst die Korrelation zweier Merkmale, 0 bedeutet keine, 1 maximale Korrelation.

e) Histogramm:



f)
```
A.breaks
## [1] 0.0 1.0 4.5 6.0
```

```
hist(A.df$Y, breaks=A.breaks, include.lowest = TRUE, right=FALSE )
```

Aufgabe 4 **13 Punkte**

Gegeben ist für $b > a$ eine gleichverteilte Zufallsvariable X mit der Dichte

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{für } x \in [a, b] \\ 0 & \text{sonst}\,. \end{cases}$$

Es gilt für den Erwartungswert $\mathrm{E}[X] = 12$ und die Varianz $\mathrm{Var}[X] = 12$.

a) Bestimmen Sie die Werte von a, b.

Falls Sie a) nicht lösen können (und nur dann), rechnen Sie bitte mit den (falschen) Werten $a = -3$ und $b = 11$ weiter.

b) Bestimmen Sie $P(2 < X \leqq 15)$.

R c) Geben Sie R-Befehle an, mit denen man das Ergebnis der Teilaufgabe b) ausgeben kann. Weisen Sie dazu zunächst die Variablen a, b zu und verwenden Sie die in R eingebaute Funktion für die Verteilungsfunktion der Gleichverteilung.

d) Zeichnen Sie den Graph der Verteilungsfunktion von X.

Zusätzlich ist jetzt eine normalverteilte (und von X unabhängige) Zufallsvariable Y gegeben mit

$$Y \sim \mathrm{N}\left(\mu = 10, \sigma = \sqrt{5}\right)\,.$$

e) Bestimmen Sie den Erwartungswert sowie die Varianz der Zufallsvariable Z mit

$$Z = \frac{2X - 5Y}{3}\,.$$

Lösungshinweis:

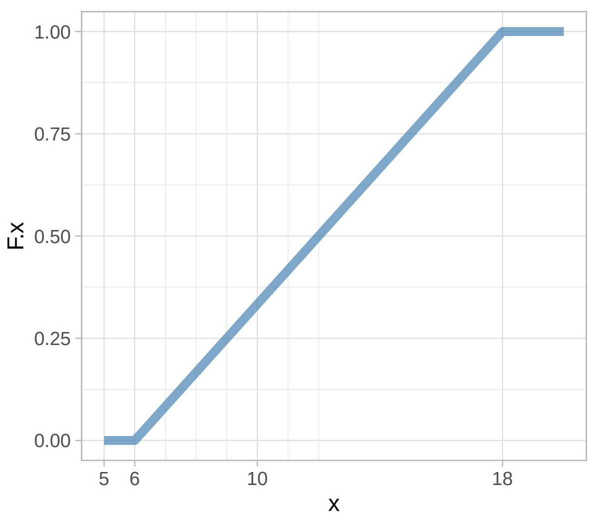
a) Es gilt mit den Formeln für E und Var der Gleichverteilung:

$$\begin{aligned} & \frac{a+b}{2} = 12 \quad \text{und} \quad \frac{(b-a)^2}{12} = 12 \\ \Leftrightarrow \quad & a = 24 - b \quad \text{und} \quad b - a = 12 \\ & \text{(positive Wurzel, weil } b > a) \\ \Rightarrow \quad & 2b - 24 = 12 \\ \Rightarrow \quad & b = 18 \\ & a = 6 \end{aligned}$$

b) da 2 kleiner als a gilt:
$P(2 < X \leqq 15) = \frac{15-6}{18-6} = \frac{3}{4}$.

c)
```r
a = 6; b = 18
punif(15, a, b) - punif(2, a, b)
```

d) Graph zu $F(x)$ (und zu $f(x)$):



e) $\mathrm{E}[Z] = \frac{2}{3} \cdot \mathrm{E}[X] - \frac{5}{3} \cdot \mathrm{E}[Y] = \frac{2}{3} \cdot 12 - \frac{5}{3} \cdot 10 \approx -8.67$
$\mathrm{Var}[Z] = \left(\frac{2}{3}\right)^2 \cdot 12 + \left(\frac{5}{3}\right)^2 \cdot 5 \approx 19.22$

Aufgabe 5 **19 Punkte**

Bei den Konzerten der Boygroup *Bangtan Boys* berichten die Sanitäter, dass bei 97 % der Konzerte höchstens 5 Ohnmachtsanfälle unter den Fans zu verzeichnen sind. Gehen Sie davon aus, dass die Anzahl X der Ohnmachtsanfälle poissonverteilt ist.

a) Wie viele Fans werden durchschnittlich bei einem Konzert ohnmächtig?

b) Bestimmen Sie $\mathrm{Var}[X]$.

Hinweis: Wenn Sie a) nicht lösen können (und nur dann), rechnen Sie bitte mit (dem flaschen Wert) $\lambda = 2$ *weiter.*

c) Wie groß ist die Wahrscheinlichkeit, dass bei einem Konzert 3 bis 8 Fans einen Ohnmachtsanfall haben?

R d) Geben Sie R-Befehle an, die das Ergebnis der Teilaufgabe b) ausgeben.

e) Im Sanitätszelt eines Konzerts stehen 8 Liegen. Wie hoch ist die Wahrscheinlichkeit, dass die Anzahl der Liegen nicht für alle Ohnmächtigen ausreicht.

Hinweis: Nehmen Sie an, alle Ohnmächtigen eines Konzerts verlieren gleichzeitig das Bewusstsein.

f) Wie groß ist die Wahrscheinlichkeit, dass bei den ersten 5 Konzerten der Welttour kein Konzert dabei ist, bei dem mehr als 5 Ohnmachtsanfälle registriert werden?

Hinweis: Gehen Sie davon aus, dass die Anzahl der Ohnmachtsanfälle der einzelnen Konzerte X_i *unabhängig voneinander sind.*

g) Wie viele Leute sind auf einem durchschnittlichen Konzert, wenn die Wahrscheinlichkeit einen Ohnmachtsanfall zu bekommen 0.0005 beträgt?

Lösungshinweis:

a) $P(X \leqq 5) = 0.97$ gilt für $\lambda = 2.3$ (Tabelle).

b) Für die Varianz der Poissonverteilung gilt $\mathrm{Var}[X] = \lambda = 2.3$.

c) $P(3 \leq X \leq 8) = P(X \leq 8) - P(X \leq 2) = 0.4033$

d)
```
ppois(8, lambda=2.3) - ppois(2, lambda=2.3)
```

e) $P(x > 8) = 1 - P(x \leq 8) = 1 - 0.9994 = 0.0006$

f) $P(x \leq 5) = 0.97$ und damit
„Anzahl Konzerte mit höchstens 5 Ohnmächtigen“ $Y \sim B(n = 5, p = 0.97)$:
$P(Y = 5) = 0.97^5 = 0.858734$.

g) $\lambda = n \cdot p$ und damit $n = \frac{\lambda}{p} = \frac{2.3}{0.0005} = 4600$.

Aufgabe 6 **13 Punkte**

Thor ist im Exil und hat das Gefühl, dass er ständig zunimmt. Damit er sich nicht so schlecht fühlt, möchte er wissen, ob es seinen Kollegen auch so geht. Er wählt unter allen seinen Superheldenfreunden (das ist die Grundgesamtheit) zufällig 5 aus (einfache Stichprobe) und befragt sie jeweils nach ihrem Gewicht am 1. Januar und am 1. Juli dieses Jahres. Er erhält die Antworten (siehe Tabelle 1), die er in einem `R`-tibble mit dem Bezeichner `D.Gewicht` speichert.

Name	Januar	Juli
Hulk	230	250
Ironman	86	93
Black Widow	62	58
Loki	61	60
Dr. Strange	78	79

Tabelle 1: Gewicht im Januar und Juli

Daraus berechnet er jeweils die relative Gewichtsveränderung (gerundet auf ganze Prozent) und erhält damit Tabelle 2.

Name	prozentual
Hulk	9
Ironman	8
Black Widow	-6
Loki	-2
Dr. Strange	1

Tabelle 2: prozentuale Gewichtsveränderung

Gehen Sie im Folgenden davon aus, dass die prozentuale Gewichtsänderung in der Grundgesamtheit normalverteilt ist.

R a) Geben Sie R-Befehle an, mit dem aus dem tibble `D.Gewicht` die Tabelle 2 der relativen Gewichtsveränderung berechnet wird.

b) Thor möchte ein Konfidenzintervall für die durchschnittliche prozentuale Gewichtsveränderung aller Superhelden zum Konfidenzniveau 98 % haben. Berechnen Sie dieses anhand der Stichprobe.

R c) Geben Sie *einen* R-Befehl an, mit dem man das Ergebnis aus Teilaufgabe b) bestimmen kann.

d) Thor erfährt, dass die Standardabweichung der prozentualen Gewichtsänderung in der Grundgesamtheit aller Superhelden 2 (Prozentpunkte) beträgt. Wie viele Superhelden müsste er mindestens befragen, damit sein Konfidenzintervall (Niveau 98 %) höchstens 4 Prozentpunkte breit ist?

Lösungshinweis:

a)
```
D.G.proz = D.Gewicht %>%
  mutate(prozentual = round(100 * (Juli/Januar - 1), 0) ) %>%
  select(Name, prozentual)
```

b) Hier σ unbekannt und Stichprobe klein, also t-Verteilung:

$\overline{x} = 2 \qquad s = 6.4420494$

$c = 3.75 \quad \mathrm{KI} = [-8.8, 12.8]$

c)
```
D.G.proz$prozentual %>% t.test(conf.level=0.98)
```

d) Hier σ bekannt, also Normalverteilung:

$n \geqq \left(\frac{2\sigma c}{L}\right)^2 = \left(\frac{2 \cdot 2 \cdot 2.33}{4}\right)^2 \approx 5.4289$, also mindestens 6 müssen befragt werden.