

Klausur Statistik

Lösungshinweise

Prüfungsdatum: 3. Juli 2019 – Prüfer: Etschberger, Henle, Jansen, Wesp

Studiengang: BW, IM

Punkte: 15, 17, 15, 15, 15, 13 ; Summe der Punkte: 90

Aufgabe 1

15 Punkte

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

$$f(x, y) = y \cdot (x^2 - 4) + 3.$$

- Bestimmen Sie den Gradienten ∇f .
- Berechnen Sie alle Kandidaten für Extremwerte von f , also alle Punkte mit $\nabla f(x, y) = (0, 0)^T$.
- Bestimmen Sie die Hessematrix $H_f(x, y)$.
- Entscheiden Sie mit Hilfe der Hessematrix für jeden der kritischen Punkte aus Teilaufgabe b), ob es sich um ein lokales Maximum oder Minimum handelt.
- Skizzieren Sie in nebenstehender Grafik $f_y(x, y)$ in Abhängigkeit von x für einen beliebigen konstanten Wert von y .

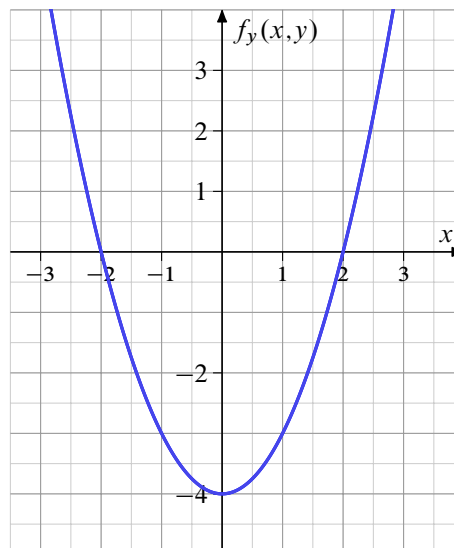


Abbildung 1: Graph von $f_y(x, y)$

Jetzt sei eine zweimal stetig differenzierbare Funktionen $g: \mathbb{R} \rightarrow \mathbb{R}$ sowie die Funktion $h: \mathbb{R}^2 \rightarrow \mathbb{R}$ mit $h(x, y) = y \cdot g(x)$ gegeben.

- Geben Sie den Gradienten $\nabla h(x, y)$ an.

Lösungshinweis:

- $\nabla f(x, y) = \begin{pmatrix} 2xy \\ x^2 - 4 \end{pmatrix}$
- $(x, y) = (\pm 2, 0)$.
- $\begin{pmatrix} 2y & 2x \\ 2x & 0 \end{pmatrix}$
- $\det H_f(\pm 2, 0) = \det \begin{pmatrix} 0 & \pm 4 \\ \pm 4 & 0 \end{pmatrix} = -16$. Also jeweils ein Sattelpunkt.
- Siehe Grafik
- $\nabla h(x, y) = \begin{pmatrix} y \cdot g'(x) \\ g(x) \end{pmatrix}$

Aufgabe 2

17 Punkte

Gegeben ist ein Datensatz in der R-Variable `Daten`. Die Ausgabe der ersten 10 Zeilen ergibt

```
## # A tibble: 957 x 4
##   Alter Geschlecht AusgKomm NoteMathe
##   <dbl> <chr>      <dbl>     <dbl>
## 1     20 Frau         130         5
## 2     19 Mann         420         5
## 3     21 Mann         825         4
## 4     23 Mann         420         1.7
## 5     20 Frau         400         1
## 6     25 Frau         130         1
## 7     20 Frau         200         5
## 8     20 Frau         420         4
## 9     20 Mann         300         1.7
## 10    20 Frau         620         2
## # ... with 947 more rows
```

- R
- Geben Sie R-Kommandos an, die das Geschlecht sowie das Alter aller Objekte dieses Datensatzes sortiert nach `NoteMathe` ausgeben bei denen das Merkmal `AusgKomm` mehr als das empirische 97 %-Quantil beträgt.
 - Wieviele Objekte (Zeilen) werden dabei mindestens ausgegeben?

Das folgende R-Kommando ergibt die Ausgabe in der Tabelle rechts

```
Daten %>%
  mutate(NoteMathe = round(NoteMathe, 0)) %>%
  group_by(Geschlecht, NoteMathe) %>%
  summarize(Anzahl=n()) %>%
  spread(key = NoteMathe,
         value = Anzahl, fill=0)
```

Geschlecht	1	2	3	4	5
Frau	39	101	188	143	124
Mann	29	68	108	77	80

- Zeichnen Sie in ein Koordinatensystem die empirische Verteilungsfunktion der gerundeten Mathenoten aller Frauen des Datensatzes.

Im Folgenden besteht die Urliste aus folgenden Werten des Merkmals `AusKomm`:

$$x = (100, 100, 100, 800, 7000, 100, 800, 800, 100, 100)$$

- Zeichnen Sie eine Lorenzkurve dieser Urliste
- Berechnen Sie den normierten Gini-Koeffizienten.

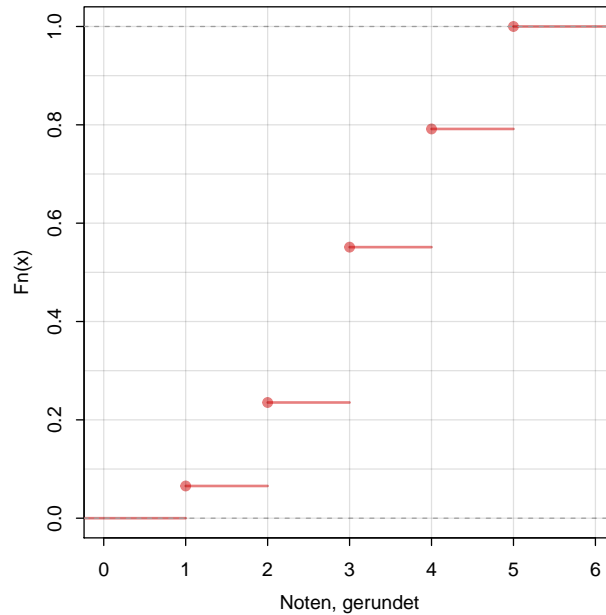
Lösungshinweis:

a) Daten %>%

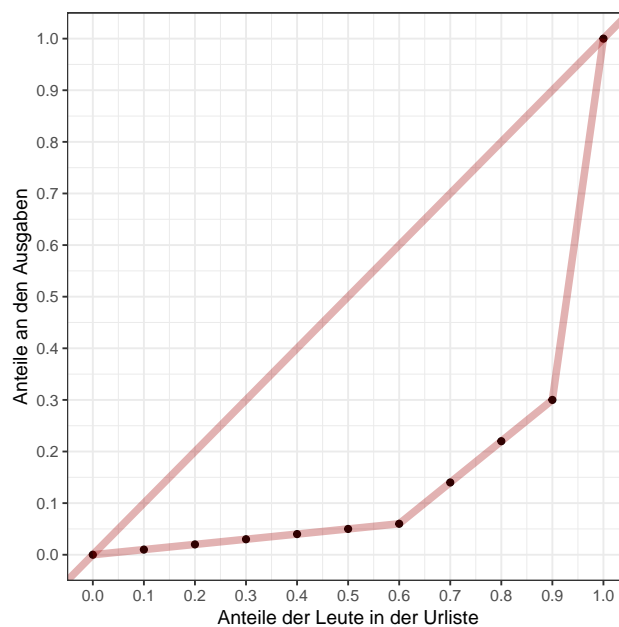
```
filter(AusgKomm > quantile(AusgKomm, probs = 0.97, type=2)) %>%  
arrange(NoteMathe) %>%  
select(Geschlecht, Alter)
```

b) Mindestens $0.03 \cdot 957 = 29$

c) Graph der Empirischen Verteilungsfunktion:



d) Lorenzkurve



e) $G = 363/500 = 0.726$,

$G^* = 121/150 \approx 0.8066667$

Aufgabe 3

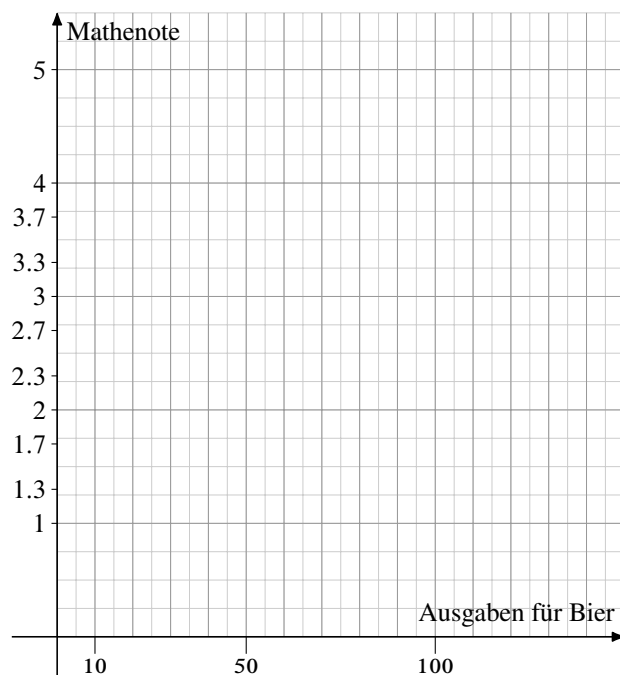
15 Punkte

Walter Weizen ist bei seinem ersten Versuch in der Matheklausur durchgefallen und sucht jetzt nach Gründen dafür. Bisher war er der Meinung, dass Bier während der Prüfungsvorbereitung die Lernleistung steigert. Das will er nun überprüfen. Unter seinen Freunden, die auch mitgeschrieben haben macht er eine Umfrage: 10 Freunde verraten ihm wieviel sie in den zwei Wochen vor der Klausur für Bier ausgegeben haben (in €) sowie die Note in der Prüfung:

Ausgaben für Bier	98	101	118	76	31	69	90	67	140	17
Note in Matheklausur	5.0	5.0	4.0	3.0	1.7	4.0	3.3	2.3	5.0	1.3

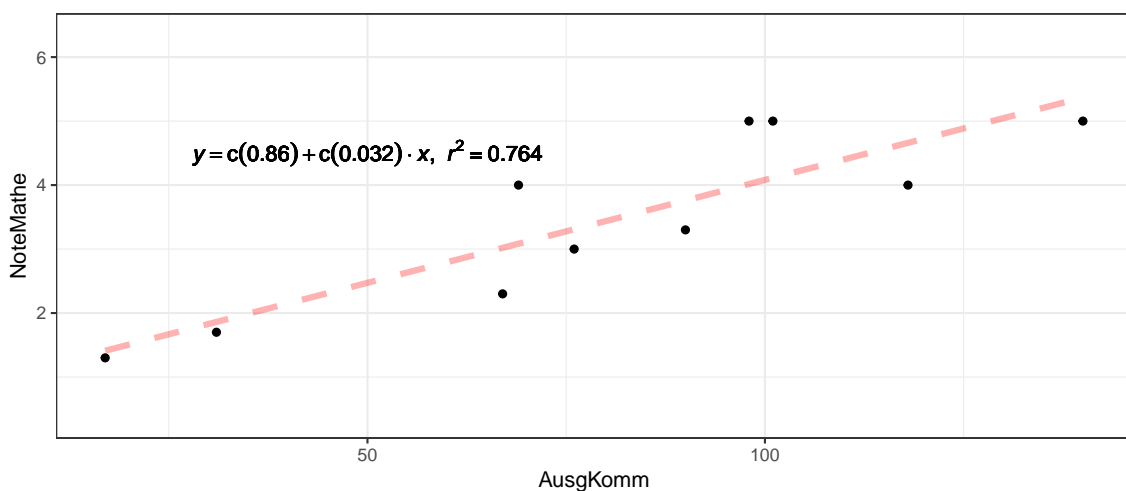
Hinweis: Gehen Sie im folgenden davon aus, dass beide Merkmale metrisch skaliert sind.

- Übertragen Sie die Daten als Streuplot in das nebenstehende Koordinatensystem.
- Berechnen Sie den Wert eines geeigneten Zusammenhangsmaßes zwischen den beiden Merkmalen.
- Geben Sie die Modellgleichung eines auf den Daten basierenden Regressionsmodells an.
- Bestimmen Sie den Determinationskoeffizient des Regressionsmodells und interpretieren Sie diesen.
- Walter hat zur letzten Prüfungszeit 100 € in Bier investiert. Welche Note hätte er anhand des Regressionsmodells eigentlich erwarten können?



- R** f) Gehen Sie davon aus, dass die Daten in der R-Variable `Matheklausur` mit den beiden Merkmalen `Bier` und `Note` gespeichert sind. Geben Sie ein R-Kommando an, mit dem man das Regressionsmodell aufstellen und das Ergebnis ausgeben kann.

Lösungshinweis:



Paula besitzt mehrere Plattenläden in der Stadt und hat dort einige Beobachtungen zu Diebstählen sammeln müssen. Paula bittet Ihre Freundin Franziska bei der Interpretation der Daten um Hilfe, denn Franziska besucht aktuell eine Statistikvorlesung im Rahmen Ihres BWL-Studiums.

Hinweise: Die Aufgabenteile a), b), c) sind jeweils unabhängig voneinander!

Geben Sie die Ergebnisse jeweils mit mindestens drei signifikanten Stellen an.

- a) Im Plattenladen in der Antonstraße beträgt die Wahrscheinlichkeit, dass ein Besucher des Ladens etwas stiehlt $p = 0.005$. Im letzten Monat betraten 1000 Leute den Laden. Wie hoch ist die Wahrscheinlichkeit pro Monat für
1. keinen Ladendiebstahl,
 2. maximal einen Ladendiebstahl?
- b) Heute hat der Ladendetektiv in der Bismarckstraße zwölf Verdächtige festgehalten, unter denen sich zwei Ladendiebe befinden. Wie groß ist die Wahrscheinlichkeit, dass beide Diebe ermittelt werden, wenn der Detektiv zufällig fünf der zwölf Verdächtigen genauer untersucht und einen Dieb durch diese Untersuchung eindeutig identifizieren kann?
- R** c) Mit welchem R-Befehl berechnet man das Ergebnis aus Teilaufgabe b)?
- d) Diebstähle werden in Paulas Plattenladen in der Bismarckstraße selten aufgedeckt. Die Wahrscheinlichkeit für mehr als vier erwischte Ladendiebe innerhalb eines Monats beträgt 42.0%. Wie viele Diebe werden im Durchschnitt pro Monat erwischt?

Lösungshinweis:

- a) X : Anzahl der Diebstähle pro Monat.
 $X \sim P(\lambda)$, dabei ist λ gesucht.
 Es gilt: $P(X > 4) = 1 - P(X \leq 4) = 0.42$. Also: $P(X \leq 4) = 0.58$.
 Mit der Tabelle folgt: $\lambda = 4.25$. Damit: 4.25 erwischte Diebstähle pro Monat.
- b) Hier: Modellierung mit Binomial- oder Poissonverteilung möglich.
 Binomial:

$$P(X = 0) = \binom{1000}{0} \cdot 0.005^0 \cdot (1 - 0.005)^{1000} = 0.006653969$$

$$P(X = 1) = \binom{1000}{1} \cdot 0.005^1 \cdot (1 - 0.005)^{999} = 0.03343703$$

$$P(X \leq 1) = P(X = 0) + P(X = 1) = 0.040091$$

Poisson-Verteilung: $\lambda = n \cdot p = 1000 \cdot 0.005 = 5$

$$P(X = 0) = \frac{\lambda^0}{0!} e^{-\lambda} = 0.006737947$$

$$P(X = 1) = \frac{\lambda^1}{1!} e^{-\lambda} = 0.03368973$$

$$P(X \leq 1) = P(X = 0) + P(X = 1) = 0.04042768$$

c) Hypergeometrische Verteilung: $N = 12, M = 2, n = 5$:

$$p(X = 2) = \frac{\binom{2}{2} \binom{10}{3}}{\binom{12}{5}} = 0.15$$

d) `dhyper(x = 2, m = 2, n = 10, k = 5)`

```
## [1] 0.1515152
```

```
choose(2,2) * choose(10,3) / choose(12,5)
```

```
## [1] 0.1515152
```

Aufgabenteil I:

Gegeben sind Funktion $g, h, k, \ell: \mathbb{R} \rightarrow \mathbb{R}$ sowie ihre Graphen in den folgenden Abbildungen:

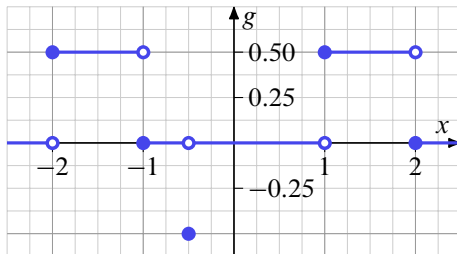


Abbildung 2: Graph von g

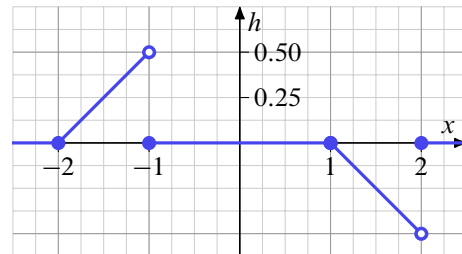


Abbildung 3: Graph von h

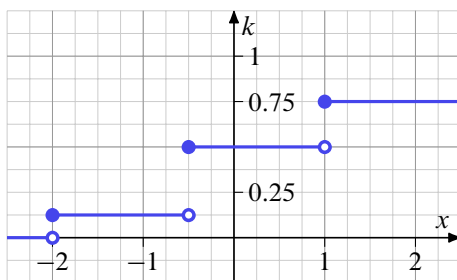


Abbildung 4: Graph von k

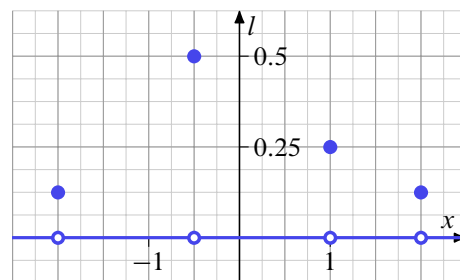


Abbildung 5: Graph von ℓ

Hinweis: Ohne (richtige) Begründung gibt es bei a) und b) keinen Punkt.

- Entscheiden und begründen Sie bei g, h, k, ℓ , ob es sich jeweils um eine Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariable handeln kann.
- Geben Sie auch bei jeder der vier Funktionen an, ob es sich jeweils um eine Verteilungsfunktion einer beliebigen (diskreten oder stetigen) Zufallsvariable handeln kann.

Aufgabenteil II:

Im Folgenden ist zu einer Zufallsvariable X die Dichtefunktion f gegeben mit

$$f(x) = \begin{cases} 1/6 & \text{für } x \in [-101, -100] \cup [-11, -9] \cup [-0.5, 0.5] \cup [1, 3] \\ 0 & \text{sonst.} \end{cases}$$

Ermitteln Sie, falls möglich, die nachfolgenden Wahrscheinlichkeiten.

- | | |
|-------------------------|--|
| c) $P(X \leq 0)$ | f) $P(X = -100.5)$ |
| d) $P(X \geq 3)$ | g) $P(X \leq -3 \mid -20 \leq X \leq 0.5)$ |
| e) $P(0 \leq X \leq 2)$ | h) $P(X \geq 0 \mid X \leq 4)$ |

Lösungshinweis:

- a) f ist eine Dichtefunktion, da das Integral über den Definitionsbereich 1 ergibt. Die Funktion g ist weder das eine noch das andere, da nicht monoton wachsend bzw. negative Massen. Die Funktion h ist ebenfalls weder eine Wahrscheinlichkeitsfunktion noch eine Verteilungsfunktion, da das Integral einen Wert größer als 1 besitzt bzw. $\lim_{x \rightarrow \infty} h(x) = 0.75 \neq 1$ gilt. Bei g handelt es sich um eine diskrete Wahrscheinlichkeitsdichte bzw. Wahrscheinlichkeitsfunktion, da die Summe der Massepunkte 1 ergibt.
- b) $P(X \leq 0) = 1/2 \cdot 1/6 + 2 \cdot 1/6 + 1/6 = 7/12$
- c) $P(X \geq 3) = 0$
- d) $P(0 \leq X \leq 2) = 1/2 \cdot 1/6 + 1 \cdot 1/6 = 3/12$
- e) $P(X = -100.5) = 0$
- f) $P(X \leq -3 \mid -20 \leq X \leq 0.5) = \frac{2 \cdot 1/6}{2 \cdot 1/6 + 1/6} = 2/3$
- g) $P(X \geq 0 \mid X \leq 4) = \frac{1/2 \cdot 1/6 + 2 \cdot 1/6}{1} = 5/12$

Aufgabe 6

13 Punkte

Walter Weizen geht davon aus, dass er die Statistikklausur bestehen wird. In diesem Fall möchte er beim Oktoberfest in München gebührend feiern. Dazu möchte er jetzt schon sicherheitshalber bei einer Online-Privatwohnungs-Vermietungsagentur während der Wies'n-Zeit ein Bett in München reservieren. Um nicht zuviel zu bezahlen, klickt er wahllos 15 Angebote für den fraglichen Zeitraum an und notiert sich die Preise (in €):

450 180 350 380 700 890 540 1750 1500 1000 790 800 1200 1000 700

Hinweis: Gehen Sie im folgenden davon aus, dass es sich bei Walters Daten um eine einfache Stichprobe einer normalverteilten Grundgesamtheit handelt.

Walter möchte mit seinen Daten den Durchschnittspreis aller Angebote schätzen.

- Bestimmen Sie ein Konfidenzintervall zum Konfidenzniveau 99% für den Mittelwert der Grundgesamtheit der Preise aller Angebote.
- Geben Sie R-Befehle an, um das Konfidenzintervall aus Teilaufgabe a) auszugeben.
- Aus sicherer Quelle erfährt Walter, dass in der Grundgesamtheit aller Angebote die Standardabweichung $\sigma = 1214 \text{ €}$ beträgt. Wieviele Angebote müsste er mindestens anklicken, wenn er bei der Berechnung des 99%-Konfidenzintervalls eine Intervallbreite von höchstens 200 € erreichen möchte?

Lösungshinweis:

- a) Ergebnisse:

c	s	s * c / sqrt(n)	Mittel	K.I.L	K.I.R
2.98	431.82	331.91	815.33	483.43	1147.24

- b)

```
x = c(...)  
t.test(x, conf.level=0.99)
```

```
## [1] 483.428 1147.239  
## attr(,"conf.level")  
## [1] 0.99
```

- c) $L = \frac{2\sigma c}{\sqrt{n}} < 200 \Leftrightarrow n > \left(\frac{2\sigma c}{200}\right)^2 \approx 980.82$, also mindestens 981 Angebote.