

**Bachelorprüfung Statistik (RUW-2170), Wintersemester 2020/21**

Liebe Studierende,

markieren Sie bitte bei den Single-Choice-Fragen Ihre Antwort auf dem Antwortbogen am Ende des Gehefts in der folgenden Weise:  $\bigcirc \otimes \bigcirc$  .

Wenn Sie eine Antwort korrigieren möchten, füllen Sie bitte die **falsch** markierte Antwort vollständig aus, ungefähr so:  $\bigcirc \bullet \otimes$  .

Die Fünfecke beziehen sich auf den Freitextteil und werden nur von der Korrektorin bzw. dem Korrektor ausgefüllt; wenn Sie ein Fünfeck selbst markieren, erhalten Sie für die betreffende Frage 0 Punkte.

Bitte füllen Sie folgende Angaben deutlich lesbar aus:

**Nachname** : \_\_\_\_\_

**Vorname** : \_\_\_\_\_

**Matrikelnummer** : \_\_\_\_\_

**Studiengang** : \_\_\_\_\_

**Raum, Platz** : \_\_\_\_\_

**Prüfer** : Prof. Dovern

**WICHTIG: Bitte kreuzen Sie Ihre Matrikelnummer auch auf dem Antwortbogen an!**

\_\_\_\_\_

\_\_\_\_\_

Nachfolgende Angaben sind nur vom Prüfer auszufüllen:

Aufgaben 1+2: \_\_\_\_\_ Teilnote: \_\_\_\_\_

Aufgabe 3: \_\_\_\_\_ Teilnote: \_\_\_\_\_

\_\_\_\_\_ Gesamtnote: \_\_\_\_\_

Unterschrift Prüfer:

**Bitte beachten Sie folgende Hinweise:**

- Das Geheft **muss** zusammen bleiben!
- Die Klausur besteht aus einem **Single-Choice** und einem **Freitextteil**.
- **Single-Choice-Teil (Aufgaben 1 und 2)**
  - Der Single-Choice-Teil umfasst 27 Single-Choice-Fragen.
  - Verwenden Sie für Ihre Antworten zu den Single-Choice-Fragen ausschließlich den Single-Choice-Antwortbogen am Ende des Gehefts. **Einträge in der Aufgabenstellung werden nicht gewertet!**
  - Beschriften Sie den Antwortbogen deutlich lesbar mit Ihrem Namen und Ihrer Matrikelnummer, und kreuzen Sie Ihre Matrikelnummer zusätzlich an!
  - Verwenden Sie auf dem Antwortbogen bitte einen **dunklen Kugelschreiber!**
- **Freitextteil (Aufgabe 3)**
  - Der Freitextteil umfasst 12 offene Aufgaben, die in den Lösungsfeldern in **diesem Geheft** zu beantworten sind.
  - Schreiben Sie Ihre Freitextantworten **lesbar**.
- Bearbeitungszeit: 120 Minuten
- **Erlaubte Hilfsmittel:**
  - Nicht-programmierbarer Taschenrechner
  - Die vom Lehrstuhl offiziell herausgegebene Formelsammlung, 2. bis 4. Auflage, ohne weitere Eintragungen oder Markierungen, mit Ausnahme von farblichen Hinterlegungen von Textpassagen und/oder Formeln bzw. unbeschriebenen Post-Its
  - Cheat Sheet für Basics in R, das über StudOn bereitgestellt wurde, ohne weitere Eintragungen oder Markierungen, mit Ausnahme von farblichen Hinterlegungen von Textpassagen und/oder Befehlen

**Viel Erfolg!**

---

---

## Bachelorprüfung Statistik, WiSe 2020/21

---

**Aufgabe 1: Single-Choice-Fragen**

---

**Bitte vergessen Sie nicht, Ihre Antworten auf den Antwortbogen zu übertragen und dort auch Ihren Namen, Vornamen sowie Ihre Matrikelnummer anzugeben.**

**Hinweis:** Aufgabe 1 besteht aus 18 Teilaufgaben, bei denen jeweils 3 Punkte erreicht werden können. Jede Frage bietet mehrere Antwortmöglichkeiten, von denen **jeweils nur eine korrekt ist**. Kreuzen Sie jeweils die korrekte Antwort **auf dem Antwortbogen** an. Beachten Sie, dass es **keinen Punktabzug für falsch beantwortete Fragen** gibt.

- 1.1** Welche Behauptung über Konfidenzintervalle ist allgemein korrekt?
- A** Das Konfidenzniveau gibt die Wahrscheinlichkeit an, mit der der wahre Parameterwert in einem realisierten Konfidenzintervall liegt.
  - B** Der zu schätzende Parameter stellt eine Konstante dar; die Grenzen eines Konfidenzintervalls sind dagegen ex ante Zufallsvariablen.
  - C** Schwankungs- und Konfidenzintervalle unterscheiden sich in ihrer inhaltlichen Interpretation nicht, da beide Intervalle feste Grenzen haben.
  - D** Für eine Irrtumswahrscheinlichkeit von  $\alpha = 0.05$  ist der Anteil an realisierten Konfidenzintervallen, die den wahren Parameterwert einschließen, mit  $\alpha \cdot 100\%$  sehr klein.
  - E** Konfidenzintervalle können nur bei der Schätzung von Maßzahlen der Lage verwendet werden.

Ihr Freund ist begeisterter Basketballfan und trainiert sogar eine Kreisligamannschaft. Er behauptet, dass die Sprintfähigkeit seines gesamten Teams sehr homogen ist, so dass die Varianz der normalverteilten Sprintgeschwindigkeit der Spieler  $X$  höchstens 2.25 sei.

Sie sind vom Gegenteil überzeugt und wollen dies mit einem Hypothesentest überprüfen. Dafür betrachten Sie die 12 Sprintgeschwindigkeiten der Spieler, die unabhängig und identisch normalverteilt sind.

1.2 Wie lauten die korrekten Null- und Alternativhypothesen?

- A  $H_0: \sigma^2 = 2.25, H_1: \sigma^2 \neq 2.25$
- B  $H_0: \sigma^2 \geq 2.25, H_1: \sigma^2 < 2.25$
- C  $H_0: \sigma^2 \leq 2.25, H_1: \sigma^2 > 2.25$
- D  $H_0: \sigma^2 \neq 2.25, H_1: \sigma^2 = 2.25$
- E  $H_0: \sigma^2 \geq 2.25, H_1: \sigma^2 \neq 2.25$

1.3 Sie notieren sich in der nächsten Trainingseinheit die 12 Sprintgeschwindigkeiten der Spieler und erhalten als Stichprobenvarianz  $\hat{\sigma}^2 = 2.1818$ . Wie lautet die Teststatistik für den obigen Test?

- A 11.6363
- B 7.2213
- C 10.8320
- D 25.4034
- E 10.6666

Bei einer Erhebung wurden 16 Bachelor-Absolventen,  $i = 1, \dots, 16$ , gefragt, wie viele Semester sie gebraucht haben, um ihr Studium abzuschließen (Merkmal  $B$ ). Die Ergebnisse liegen in der folgenden Urliste vor:

$i:$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$b_i:$	6	6	6	6	6	7	7	7	7	7	8	8	8	9	9	10

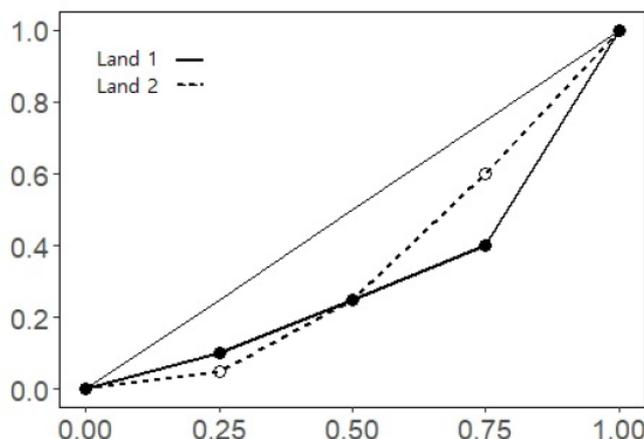
1.4 Welche Werte nehmen die beobachtete Spannweite  $s_M$  und das obere Quartil  $x_{0.75}$  an?

- A  $s_M = 6$  und  $x_{0.75} = 8$
- B  $s_M = 4$  und  $x_{0.75} = 8$
- C  $s_M = 4$  und  $x_{0.75} = 7$
- D  $s_M = 6$  und  $x_{0.75} = 7$
- E  $s_M = 5$  und  $x_{0.75} = 8$

1.5 Gegeben sei eine Urliste mit Beobachtungen  $x_i$  für  $i = 1, \dots, n$  sowie  $y_i = a + bx_i$ , wobei  $a$  und  $b$  reelle Zahlen sind. Welche der folgenden Aussagen ist allgemein korrekt?

- A  $\bar{x} = a + b\bar{y}$
- B  $\bar{y} = \frac{a}{n} + \frac{b}{n}\bar{x}$
- C  $\bar{y} = a + b\bar{x}$
- D  $\bar{y} = b^2\bar{x}$
- E  $\bar{x} = a - b\bar{y}$

Das nachfolgende Schaubild zeigt die Lorenzkurven der Vermögensverteilungen in zwei Ländern. Im Folgenden sei  $G_i$  der Gini-Koeffizient für Land  $i$ ,  $i = 1, 2$ .



1.6 Welche Aussage über die dargestellten Vermögensverteilungen ist korrekt?

- A Die ärmsten 25% der Personen in Land 1 besitzen weniger Vermögensanteile als jene in Land 2.
- B  $G_1 > G_2$
- C  $G_1 = G_2$
- D Die reichsten 25% der Personen in Land 2 besitzen mehr Vermögensanteile als jene im Land 1.
- E  $G_1 < G_2$

Eine zufällige Stichprobe von  $n = 100$  Kunden bzw. Kundinnen eines Investmentfondsanbieters wird nach dem Anlegertyp, d.h. Privatanleger (P) oder institutioneller Anleger (I), sowie nach dem bevorzugten Anlagentyp sortiert.

Die einzelnen Investoren verfolgen unterschiedliche Investitionsstrategien. Sie verteilen ihre Vermögen auf drei Anlageklassen: Aktien (A), Bargeld (B) und Rohstoffe (R).

In der folgenden Kontingenztafel fehlt eine Häufigkeit.

	Aktien (A)	Bargeld (B)	Rohstoffe (R)
Privatanleger (P)	26	?	11
Institutioneller Anleger (I)	24	13	9

1.7 Wie groß ist der Anteil der Investoren, die Privatanleger sind und Bargeld als bevorzugten Anlagentyp haben?

- A  $h(P|B) = 0.566$
- B  $h(P, B) = 0.170$
- C  $h(B|P) = 0.314$
- D  $h(B, P) = 0.130$
- E  $h(P|B) = 0.170$

- 1.8** Wie groß ist der Anteil der Privatanleger unter denjenigen, die ihr Vermögen in Rohstoffe investieren?
- A**  $h(P, R) = 0.110$
  - B**  $h(P|R) = 0.550$
  - C**  $h(R|P) = 0.204$
  - D**  $h(P|R) = 0.450$
  - E**  $h(R|P) = 0.196$

Zwei Würfel werden gewürfelt. Die Ergebnismenge  $E$  kann wie folgt geschrieben werden:

$$E = \{ (1, 1), (1, 2), \dots, (6, 6) \}$$

Betrachten Sie nun das folgende Ereignis:

$A =$  "Mindestens einer der Würfel zeigt eine 1"

*Hinweis: Jedes Ergebnis aus  $E$  ist mit einer Wahrscheinlichkeit von  $1/36$  gleich wahrscheinlich.*

- 1.9** Wie groß ist die Wahrscheinlichkeit  $P(A)$ ?
- A**  $P(A) = 11/36$
  - B**  $P(A) = 1/18$
  - C**  $P(A) = 7/36$
  - D**  $P(A) = 1/3$
  - E**  $P(A) = 1/6$

- 1.10** Eine Urne enthält zehn von 1 bis 10 durchnummerierte Kugeln. Wie viele Variationen gibt es, wenn vier Kugeln ohne Zurücklegen gezogen wird?
- A** 10000
  - B** 5040
  - C** 210
  - D** 288
  - E** 1048576

Betrachten Sie die Zufallsvariable  $Y$  mit folgender Dichtefunktion:

$$f(y) = \begin{cases} \frac{1}{4} \cdot y + \frac{1}{4} & \text{für } 0 \leq y \leq 2 \\ 0 & \text{sonst} \end{cases}$$

**1.11** Welchen Wert hat der Erwartungswert von  $Y$ ?

- A 1.167
- B 0.667
- C 0.500
- D 1.500
- E 2.500

Eine Maschine befüllt leere Flaschen mit einer Flüssigkeit. Der Inhalt einer Flasche in Millilitern kann als normalverteilte Zufallsvariable mit Erwartungswert  $500\text{ml}$  und einer Standardabweichung von  $10\text{ml}$  angesehen werden.

**1.12** Wie hoch ist die Wahrscheinlichkeit dafür, dass eine Flasche mindestens  $499\text{ml}$  aber höchstens  $501\text{ml}$  Inhalt hat?

- A 7.96%
- B 9.94%
- C 11.92%
- D 13.57%
- E 5.98%

**1.13** Die Pearson-Korrelation zwischen zwei Merkmalen A und B beträgt 0.4 bei einer Kovarianz von 0.5. Die Varianz von Merkmal B ist 2. Wie hoch ist die Varianz von Merkmal A?

- A 1.28
- B 1.13
- C 0.63
- D 0.78
- E 0.88

**1.14** Eine *i.i.d.*-Stichprobe vom Umfang  $n = 14$  wird aus einer normalverteilten Grundgesamtheit entnommen, welche durch die Zufallsvariable  $X$  mit Erwartungswert  $\mu_X = 60$  und Standardabweichung  $\sigma_X = 7$  charakterisiert wird. Welcher Verteilung folgt das Stichprobenmittel?

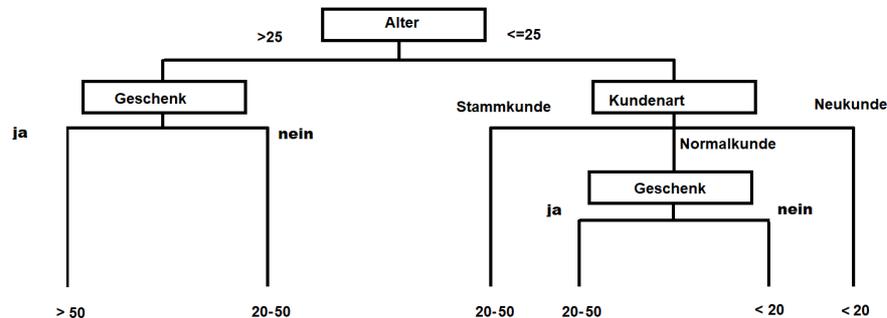
- A  $\bar{X} \sim \mathcal{N}(60, 49)$
- B  $\bar{X} \sim \mathcal{N}(60, 7)$
- C  $\bar{X} \sim \mathcal{N}(60, 3.5)$
- D  $\bar{X} \sim \mathcal{N}(60, 0.5)$
- E  $\bar{X} \sim \mathcal{N}(60, 98)$

Mittels einer statistischen Untersuchung möchte die Bundesregierung einen Einblick über die Qualität der betrieblichen Gesundheitsförderung in Unternehmen gewinnen. Gemäß der Empfehlung 2003/361/EG der Europäischen Union ist die Unternehmensgröße in Abhängigkeit von der Anzahl der Beschäftigten wie folgt definiert:

Typ	Anzahl Beschäftigte
Kleinstunternehmen	$< 10$
Kleine Unternehmen	$< 50$
Mittlere Unternehmen	$< 250$
Große Unternehmen	$\geq 250$

- 1.15** Es wird in jeder Größenklasse eine einfache Zufallsstichprobe gezogen. Anschließend werden die Daten zur Auswertung zusammengeführt. Wie nennt man dieses Stichprobenverfahren?
- A** Clusterstichprobe bzw. Klumpenstichprobe
  - B** Einfache Zufallsstichprobe
  - C** Geschichtete Stichprobe
  - D** Quotenstichprobe
  - E** Typische Stichprobe
- 1.16** Gegeben sei eine unabhängig und identisch verteilte Zufallsstichprobe. Welche der folgenden Aussagen ist **nicht** korrekt?
- A** Der Zentrale Grenzwertsatz besagt, dass das Stichprobenmittel für große  $n$  approximativ normalverteilt ist.
  - B** Für sehr große  $n$  ist die Verteilung des Stichprobenmittels gemäß des Zentralen Grenzwertsatzes symmetrisch.
  - C** Das schwache Gesetz der großen Zahlen besagt, dass die Varianz des Mittelwertschätzers für  $n \rightarrow \infty$  gegen die Populationsvarianz konvergiert.
  - D** Der Zentrale Grenzwertsatz besagt, dass die Summe der Stichprobenvariablen asymptotisch normalverteilt ist.
  - E** Gemäß des Hauptsatzes der Statistik liegt die empirische Verteilungsfunktion bei einem steigenden Stichprobenumfang immer näher an der theoretischen Verteilungsfunktion.

Der Betreiber eines Online-Weinhandels möchte vorhersagen, für welche Weinkategorie sich seine Kunden und Kundinnen beim nächsten Kauf entscheiden. Dabei gibt es drei Preiskategorien (unter 20 Euro, 20-50 Euro und über 50 Euro). Die Vorhersage basiert auf den Merkmalen "Alter des Käufers", "Weinflasche soll als Geschenk verpackt werden" und "Kundenart" (Neukunde, Normalkunde oder Stammkunde). Zur Prognose verwendet er den folgenden Entscheidungsbaum:



### 1.17 Welche der folgenden Aussagen ist **falsch**?

- A Für einen 28 Jahre alten Käufer, der den Wein mit Geschenkverpackung kauft, wird ein Preis von über 50 Euro prognostiziert.
- B Für einen 20 Jahre alten Neukunden wird ein Preis von unter 20 Euro prognostiziert.
- C Für einen 60 Jahre alten Käufer, der den Wein nicht geschenkverpackt kauft, wird ein Preis von 20-50 Euro prognostiziert.
- D Für einen 33 Jahre alten Käufer, der den Wein nicht geschenkverpackt kauft, wird ein Preis von 20-50 Euro prognostiziert.
- E Für einen 18 Jahre alten Normalkunden, der den Wein mit Geschenkverpackung kauft, wird ein Preis von unter 20 Euro prognostiziert.

### 1.18 Welches der folgenden Merkmale ist **nicht** ordinalskaliert?

- A Militärischer Rang (z.B. General, Leutnant oder Feldwebel)
- B Produktzufriedenheit (z.B. zufrieden oder unzufrieden)
- C Lawinengefahr (z.B. groß, erheblich oder gering)
- D Anzahl der Geschwister (z.B. ein, zwei oder drei Geschwister)
- E Schärfegrad von Lebensmitteln (z.B. scharf oder mild)

**Bitte vergessen Sie nicht, Ihre Antworten auf den Antwortbogen zu übertragen und dort auch Ihren Namen, Vornamen sowie Ihre Matrikelnummer anzugeben.**

---

## Aufgabe 2: Single-Choice-Fragen

---

**Bitte vergessen Sie nicht, Ihre Antworten auf den Antwortbogen zu übertragen und dort auch Ihren Namen, Vornamen sowie Ihre Matrikelnummer anzugeben.**

**Hinweis:** Aufgabe 2 besteht aus 9 Teilaufgaben, bei denen jeweils 3 Punkte erreicht werden können. Jede Frage bietet mehrere Antwortmöglichkeiten, von denen **jeweils nur eine korrekt ist**. Kreuzen Sie jeweils die korrekte Antwort **auf dem Antwortbogen** an. Beachten Sie, dass es **keinen Punktabzug für falsch beantwortete Fragen** gibt.

Gehen Sie von fünf Merkmalsträgern aus, für die Beobachtungen für zwei Merkmale  $X$  und  $Y$  in der folgenden Tabelle vorliegen:

$i$	1	2	3	4	5
$x_i$	4	5	0	8	2
$y_i$	3	0	7	2	3

**2.1** Betrachten Sie die beiden Clusterschwerpunkte  $C^A = (1, 5)$  und  $C^B = (4, 3)$ . Die Distanz einer Beobachtung zu einem beliebigen Schwerpunkt  $C = (c_X, c_Y)$  ist gegeben durch die euklidische Distanz:

$$d_i = \sqrt{(x_i - c_X)^2 + (y_i - c_Y)^2},$$

wobei  $c_X$  und  $c_Y$  die x- bzw. y-Koordinaten eines Schwerpunktes bezeichnen. Im Rahmen einer K-means-Clustering-Analyse sollen die Clusterzugehörigkeiten der einzelnen Beobachtungen bestimmt werden.

Welche der folgenden Aussagen ist korrekt?

- A** Die Beobachtung  $i = 1$  gehört zu Cluster B.
- B** Die Beobachtung  $i = 5$  gehört zu Cluster A.
- C** Die Beobachtung  $i = 2$  gehört zu Cluster A.
- D** Die Beobachtung  $i = 4$  gehört zu Cluster A.
- E** Keine der anderen Aussagen ist korrekt.

Sie möchten die Korrelation zwischen dem Alter und der Platzierung in einem Wettkampf mit einem geeigneten Maß berechnen. Ihnen stehen die folgenden Beobachtungen für fünf Sportler zur Verfügung:

$i$	1	2	3	4	5
Platzierung	1	3	5	4	2
Alter	18	22	28	25	19

## 2.2 Welche der folgenden Aussagen ist korrekt?

- A** Der Korrelationskoeffizient nach Pearson ist ein geeignetes Maß, um den Zusammenhang im vorliegenden Beispiel zu berechnen.
- B** Die Berechnung des Zusammenhangs mit dem Korrelationskoeffizienten nach Pearson würde im vorliegenden Beispiel einen negativen Zusammenhang anzeigen.
- C** Die Berechnung des Zusammenhangs mit dem Rangkorrelationskoeffizienten nach Spearman würde im vorliegenden Beispiel keinen Zusammenhang anzeigen.
- D** Der Rangkorrelationskoeffizient nach Spearman ist kein geeignetes Maß, um den Zusammenhang im vorliegenden Beispiel zu berechnen.
- E** Keine der anderen Antworten ist korrekt.

## 2.3 Es ist bekannt, dass 40% aller Menschen die Blutgruppe Null besitzen. Nach einem Aufruf zur Blutspende melden sich unabhängig voneinander 10 Studierende im Krankenhaus zur Blutspende. Welchem Verteilungsmodell folgt die Zufallsvariable $X$ "Anzahl der Studierenden mit Blutgruppe Null"?

- A**  $X \sim \mathcal{B} (n=10, p=0.4)$
- B**  $X \sim \mathcal{P} (n=10, p=0.4)$
- C**  $X \sim \mathcal{P} (\lambda=4)$
- D**  $X \sim \mathcal{G} (p=0.4)$
- E**  $X \sim \mathcal{DR} (M=4, N=10)$

Ein Antikörpertest wird verwendet, um festzustellen, ob ein Patient Antikörper gegen ein Virus im Blut hat. Betrachten Sie folgende Ereignisse:

$D$ : "Der Antikörpertest des Patienten fällt positiv aus."

$A$ : "Der Patient hat Antikörper im Blut."

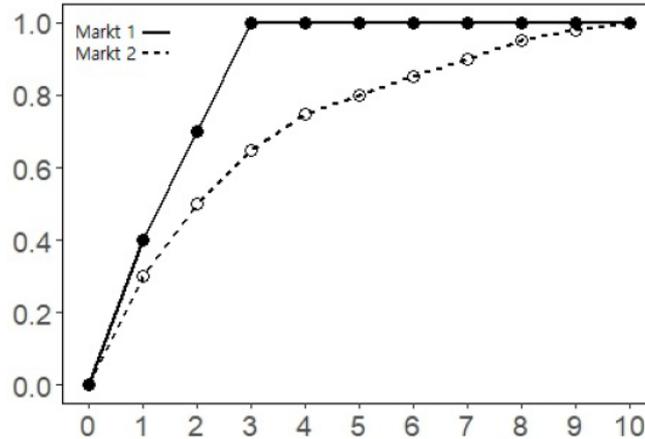
Es sind folgende Wahrscheinlichkeiten bekannt:

$$P(A) = 0.14, \quad P(D) = 0.15, \quad P(D \cap A) = 0.126$$

## 2.4 Durch welchen Ausdruck und durch welche Wert ist die Wahrscheinlichkeit gegeben, bei negativem Testergebnis tatsächlich keine Antikörper im Blut zu haben?

- A**  $P(\bar{D}|\bar{A}) = 0.73$
- B**  $P(\bar{A}|\bar{D}) = 0.98$
- C**  $P(\bar{D}|\bar{A}) = 0.84$
- D**  $P(\bar{A}|\bar{D}) = 0.44$
- E**  $1 - P(D \cap A) = 0.87$

Das nachfolgende Schaubild zeigt die Konzentrationsraten für zwei Märkte. Im Folgenden sei  $H_i$  der Herfindahl-Index für Markt  $i$ ,  $i = 1, 2$ .



2.5 Welche Aussage ist korrekt?

- A Die Firma mit dem größten Marktanteil in Markt 2 hat einen größeren Marktanteil als die Firma mit dem größten Marktanteil in Markt 1.
- B  $H_1 > H_2$
- C Die Firma mit dem größten Marktanteil in Markt 2 hat einen Marktanteil von 40%.
- D Der Marktanteil der Firma mit dem zweitgrößten Marktanteil in Markt 1 beträgt 70%.
- E  $H_1 < H_2$

2.6 Die Zufallsvariablen  $A$  und  $B$  seien bivariat normalverteilt. Sie möchten die beiden Zufallsvariablen auf Unkorreliertheit prüfen und stellen dazu folgende Hypothesen auf:  $H_0: \rho_{AB} = 0$  gegen  $H_1: \rho_{AB} \neq 0$ . Aus einer Stichprobe vom Umfang  $n = 50$  ermitteln Sie einen Korrelationskoeffizienten nach Bravais-Pearson von  $r_{AB} = 0.28$ .

Bestimmen Sie für den Korrelationstest den korrekten Wert der Teststatistik und fällen Sie die richtige Testentscheidung, wenn  $t_{48; 0.995} = 2.682$  die kritische Schranke darstellt.

- A  $t = 2.0207 \rightarrow H_0$  nicht ablehnen, da  $|t| < 2.682$
- B  $t = 2.2862 \rightarrow H_0$  nicht ablehnen, da  $|t| < 2.682$
- C  $t = 2.0207 \rightarrow H_0$  ablehnen, da  $|t| > 2.682$
- D  $t = 2.2862 \rightarrow H_0$  ablehnen, da  $|t| < 2.682$
- E  $t = 3.8188 \rightarrow H_0$  ablehnen, da  $|t| > 2.682$

2.7 Sie erheben eine Stichprobe vom Umfang  $n = 100$  aus einer normalverteilten Grundgesamtheit. Die Varianz der Grundgesamtheit betrage  $\sigma^2 = 36$ , das Stichprobenmittel nimmt den Wert  $\bar{x} = 10$  an. Berechnen Sie das zentrale Konfidenzintervall für den Erwartungswert  $\mu$  der Grundgesamtheit zum 95%-Niveau.

- A [2.9440, 17.0560]
- B [9.8824, 10.1176]
- C [9.0131, 10.9869]
- D [8.8240, 11.1760]
- E [6.7333, 13.2667]

**2.8** Welche Aussage über einen  $\chi^2$ -Anpassungstest ist allgemein korrekt?

- A** Da die Dichtefunktion der  $\chi^2$ -Verteilung stets symmetrisch ist, kann man für die Verteilung der Prüfgröße unter Gültigkeit der Nullhypothese auch die  $t$ -Verteilung heranziehen.
- B** Die Konstruktion der Prüfgröße basiert auf dem Vergleich zwischen den empirisch beobachteten und den unter  $H_0$  erwarteten Häufigkeiten.
- C** Kleine Werte für die Teststatistik sprechen stark für eine Ablehnung der Nullhypothese.
- D** Ein steigender Stichprobenumfang  $n$  wirkt sich nachteilig auf die Erfüllung der Approximationsvoraussetzung  $n \cdot p_i^0 \geq 5$  aus.
- E** Ein  $\chi^2$ -Anpassungstest kann nur bei bekannten Parametern angewendet werden.

In einem Haushalt wurde der tägliche Wasserverbrauch in Litern,  $X$  an  $n = 200$  Tagen gemessen. Das Ergebnis ist in der folgenden Tabelle dargestellt.

$i$	$(x_{i-1}^*, x_i^*]$	$m_i$	$h_i$	$\hat{F}(x^*)$
1	(50, 100]	75	0.125	0.125
2	(100, 150]	125	0.325	0.450
3	(150, 200]	175	0.375	0.825
4	(200, 250]	225	0.175	1

**2.9** Welche Werte haben das arithmetische Mittel und der Median des Wasserverbrauchs?

- A**  $\bar{x} = 150$  und  $\tilde{x} = 156.67$
- B**  $\bar{x} = 150$  und  $\tilde{x} = 145.00$
- C**  $\bar{x} = 155$  und  $\tilde{x} = 145.00$
- D**  $\bar{x} = 155$  und  $\tilde{x} = 156.67$
- E**  $\bar{x} = 145$  und  $\tilde{x} = 220.33$

**Bitte vergessen Sie nicht, Ihre Antworten auf den Antwortbogen zu übertragen und dort auch Ihren Namen, Vornamen sowie Ihre Matrikelnummer anzugeben.**

---

### Aufgabe 3: Freitextaufgaben

---

**Hinweis:** Aufgabe 3 besteht aus 12 Teilaufgaben, bei denen insgesamt 39 Punkte erreicht werden können. Verwenden Sie für die Lösung der Aufgaben die durch die Linien begrenzten Lösungsfelder direkt unter dem jeweiligen Aufgabentext. **Nehmen Sie für diese Aufgabe keine Markierungen auf dem Antwortbogen vor.** Falls nötig, runden Sie Ihre Ergebnisse auf **vier Nachkommastellen**.

Die Anzahl der Gäste, die innerhalb einer Stunde in dem Restaurant von Besitzerin Lisa erscheinen, kann durch die Poisson-verteilte Zufallsvariable  $L$  mit  $\lambda = 4$  dargestellt werden.

- 3.1** Berechnen Sie die Wahrscheinlichkeit dafür, dass innerhalb einer Stunde mindestens zwei, aber höchstens fünf Gäste das Restaurant betreten. (3 Punkte)

*Hinweis:* Sie können hierfür die Tabellen der Formelsammlung verwenden.

---

MUSTER  
Nicht ausfüllen!

**3.2** Es gilt folgender allgemeiner Zusammenhang: Wenn die Anzahl der Ereignisse pro Zeiteinheit Poisson-verteilt ist mit Parameter  $\lambda$ , dann ist die Wartezeit zwischen zwei solchen Ereignissen exponentialverteilt mit dem gleichen Parameter  $\lambda$ .

Benutzen Sie diese Information, um die Wahrscheinlichkeit zu berechnen, mit der die Wartezeit  $Z$  in Stunden zwischen der Ankunft von zwei Gästen in Lisas Restaurant mehr als eine halbe Stunde beträgt. (4 Punkte)

---

---

MUSTER  
Nicht ausfüllen!

**3.3** Verwenden Sie die Wahrscheinlichkeitsfunktion der Poissonverteilung, um die Wahrscheinlichkeit zu bestimmen, mit der innerhalb eines Arbeitstages mit 8 Stunden genau 34 Gäste das Restaurant von Lisa besuchen. Definieren Sie hierfür zunächst eine neue Zufallsvariable  $G$ , welche die Anzahl der Gäste in Lisas Restaurant pro Arbeitstag misst. (3 Punkte)

*Hinweis:* Beachten Sie, dass Ihnen die Verteilungstabellen aus der Formelsammlung in diesem Fall nicht weiterhelfen.

---

MUSTER  
Nicht ausfüllen!

---

**3.4** Gehen Sie in den Aufgaben 3.4 und 3.5 – unabhängig von Ihren Überlegungen in Aufgabenteil 3.3 – davon aus, dass die Anzahl der Gäste in Lisas Restaurant pro Arbeitstag dargestellt werden kann durch  $G \sim \mathcal{P}(\lambda=30)$ .

Da es in Lisas Restaurant nur ein fixes Menü gibt, betragen die Ausgaben jedes Restaurantbesuchers für Speise und Getränke exakt 40 Euro. Darüber hinaus hat Lisa pro Arbeitstag Ausgaben in Höhe von 500 Euro. Der tägliche Profit von Lisa in Euro,  $P$ , kann folglich durch die Zufallsvariable

$$P = a + b \cdot G$$

dargestellt werden.

Bestimmen Sie  $a$  und  $b$  und berechnen Sie den erwarteten Tagesprofit von Lisa. (5 Punkte)

---

MUSTER  
Nicht ausfüllen!

---

**3.5** Berechnen Sie die Standardabweichung des Tagesprofites von Lisa. (5 Punkte)

---

MUSTER  
Nicht ausfüllen!

---

**3.6** Mark, der Besitzer eines konkurrierenden Restaurants, geht davon aus, dass die stündliche Anzahl seiner Gäste ebenfalls Poisson-verteilt ist und durch die Zufallsvariable  $X$  dargestellt werden kann.

Da Mark nur selten vor Ort ist, kennt er den Parameter  $\lambda$  nicht, d.h.  $X \sim \mathcal{P}(\lambda = ?)$ . Um festzustellen, ob sein Restaurant im Mittel beliebter oder weniger beliebt ist als das von Lisa, möchte er den Wert von  $\lambda$  auf Basis einer *i.i.d.*-Stichprobe,  $X_1, \dots, X_n$ , schätzen.

Stellen Sie **allgemein** die Likelihoodfunktion im Falle einer Poissonverteilung auf. Nennen Sie die Annahme, die Sie zum Aufstellen der Likelihoodfunktion benötigen. (3 Punkte)

---

---

**3.7** Erklären Sie **in einem Satz**, wieso man anstatt der Likelihoodfunktion typischerweise die Log-Likelihoodfunktion maximiert. (1 Punkt)

---

---

**3.8** Zeigen Sie, dass die Log-Likelihoodfunktion einer Poisson-verteilten Zufallsvariable die folgende Form hat:

$$\ln(L(\lambda)) = \ln(\lambda) \sum_{i=1}^n X_i - n\lambda - \sum_{i=1}^n \ln(X_i!)$$

(4 Punkte)

---

---

**3.9** Maximieren Sie die Log-Likelihoodfunktion und leiten Sie den Maximum-Likelihood-Schätzer für  $\lambda$  her. (3 Punkte)

---

---

MUSTER  
Nicht ausfüllen!

**3.10** Gehen Sie in den nachfolgenden Aufgaben – unabhängig von Ihrem Ergebnis in

Aufgabenteil 3.9 – davon aus, dass  $\hat{\lambda}_{ML} = \frac{1}{n} \sum_{i=1}^n X_i$ .

Zeigen Sie, dass der Maximum-Likelihood-Schätzer ein erwartungstreuer Schätzer für den unbekannt Parameter ist. (4 Punkte)

---

MUSTER  
Nicht ausfüllen!

---

**3.11** Mark entschließt sich dazu, sein Restaurant einen Tag lang zu besuchen. Dabei beobachtet er die folgende Anzahl an Gästen über die acht Arbeitsstunden hinweg:  $\{5, 4, 2, 4, 5, 5, 3, 4\}$ .

Berechnen Sie auf Basis der Stichprobenrealisation den konkreten Schätzwert von  $\lambda$ . Wie interpretieren Sie Ihr Ergebnis hinsichtlich der durchschnittlichen Beliebtheit der Restaurants von Lisa und Mark? (3 Punkte)

---

---

**3.12** Nennen Sie den Test, mit dem Sie überprüfen können, ob die Stichprobe tatsächlich aus einer Poisson-verteilten Grundgesamtheit stammt. (1 Punkt)

---

---

MUSTER  
Nicht ausfüllen!

**Musterlösung**

Bachelorprüfung Statistik, WiSe 2020/21

1.1	<input type="checkbox"/> A	<input checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.2	<input type="checkbox"/> A	<input type="checkbox"/> B	<input checked="" type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.3	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input checked="" type="checkbox"/> E
1.4	<input type="checkbox"/> A	<input checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.5	<input type="checkbox"/> A	<input type="checkbox"/> B	<input checked="" type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.6	<input type="checkbox"/> A	<input checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.7	<input type="checkbox"/> A	<input checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.8	<input type="checkbox"/> A	<input checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.9	<input checked="" type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.10	<input type="checkbox"/> A	<input checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.11	<input checked="" type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.12	<input checked="" type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.13	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input checked="" type="checkbox"/> D	<input type="checkbox"/> E
1.14	<input type="checkbox"/> A	<input type="checkbox"/> B	<input checked="" type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.15	<input type="checkbox"/> A	<input type="checkbox"/> B	<input checked="" type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.16	<input type="checkbox"/> A	<input type="checkbox"/> B	<input checked="" type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.17	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input checked="" type="checkbox"/> E
1.18	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input checked="" type="checkbox"/> D	<input type="checkbox"/> E
2.1	<input checked="" type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
2.2	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input checked="" type="checkbox"/> E
2.3	<input checked="" type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
2.4	<input type="checkbox"/> A	<input checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
2.5	<input type="checkbox"/> A	<input checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
2.6	<input checked="" type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
2.7	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input checked="" type="checkbox"/> D	<input type="checkbox"/> E
2.8	<input type="checkbox"/> A	<input checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
2.9	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input checked="" type="checkbox"/> D	<input type="checkbox"/> E

**3.1**

$$P(2 \leq L \leq 5) = P(L \leq 5) - P(L \leq 1) = F(5) - F(1) = 0.7851 - 0.0916 = 0.6935$$

3 Punkte. Wahrscheinlichkeit umformulieren (2 TP) und ausrechnen (1 TP). Alternative Lösung über Wahrscheinlichkeitsfunktion möglich.

**3.2**

$$L \sim \mathcal{P}(\lambda = 4) \Rightarrow Z \sim \mathcal{E}(\lambda = 4)$$

$$P(Z > 0.5) = 1 - F(0.5) = 1 - (1 - e^{-4 \cdot 0.5}) = e^{-2} = 0.1353$$

4 Punkte. Je 1 TP Abzug für jeden falschen bzw. fehlenden Schritt.

**3.3**

$$G \sim \mathcal{P}(\lambda = 8 \cdot 4 = 32)$$

$$\text{Es folgt } P(G = 34) = \frac{\lambda^g e^{-\lambda}}{g!} = \frac{32^{34} e^{-32}}{34!} = 0.0642$$

3 Punkte. Je 1 TP Abzug für jeden falschen bzw. fehlenden Schritt.

**3.4**

$$G \sim \mathcal{P}(\lambda = 30) \Rightarrow E(G) = 30 \text{ und } V(G) = 30$$

$$E(P) = E(a + b \cdot G) = E(-500 + 40 \cdot G) = -500 + 40 \cdot E(G) = -500 + 40 \cdot 30 = 700$$

5 Punkte. Korrekte Wahl von a und b (1 TP), Umformung des Erwartungswertes (2 TP), Einsetzen von E(X) (1 TP), Endergebnis (1 TP).

**3.5**

$$V(P) = V(-500 + 40 \cdot X) = 40^2 \cdot V(G) = 40^2 \cdot 30 = 48000$$

$$\sqrt{V(P)} = \sqrt{48000} = 219.089$$

5 Punkte. Einsetzen in Varianz (1 TP), Umformung inkl. 40 quadrieren (2 TP), Varianz berechnen (1 TP), Standardabweichung berechnen (1 TP).



**3.6**

$$L(\lambda) = \prod_{i=1}^n f(X_i; \lambda) = \prod_{i=1}^n \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}$$

unter Berücksichtigung der Tatsache, dass die Stichprobenvariablen unabhängig sind.

3 Punkte. Je 1 TP Abzug für jeden falschen bzw. fehlenden Schritt.

**3.7**

Die Log-Likelihoodfunktion ist meist einfacher zu optimieren als die Likelihoodfunktion und die Lage des Maximums ändert sich dadurch nicht.

1 Punkt.

**3.8**

$$\begin{aligned} \ln(L(\lambda)) &= \ln\left(\prod_{i=1}^n \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}\right) = \sum_{i=1}^n \ln\left(\frac{\lambda^{X_i} e^{-\lambda}}{X_i!}\right) \\ &= \sum_{i=1}^n (\ln(\lambda^{X_i}) + \ln(e^{-\lambda}) - \ln(X_i!)) \\ &= \sum_{i=1}^n (X_i \ln(\lambda) - \lambda - \ln(X_i!)) \\ &= \ln(\lambda) \sum_{i=1}^n X_i - n\lambda - \sum_{i=1}^n \ln(X_i!) \end{aligned}$$

4 Punkte. Logarithmieren (1 TP), als Summe ausdrücken (1 TP), Umformung (2 TP).

**3.9**

$$\frac{\partial \ln(L(\lambda))}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^n X_i - n \stackrel{!}{=} 0 \Rightarrow \hat{\lambda}_{ML} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

3 Punkte. Ableiten (1 TP), gleich Null setzen (1 TP), nach  $\hat{\lambda}$  auflösen (1 TP).

**3.10**

$$E(\hat{\lambda}_{ML}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n \cdot \lambda = \lambda$$

4 Punkte. 1 TP pro Rechenschritt.

**3.11**

$$\hat{\lambda}_{ML} = \bar{x} = \frac{1}{8} \cdot (5+4+2+4+5+5+3+4) = 4$$

Da  $L \sim \mathcal{P}(\lambda=4)$  und  $X \sim \mathcal{P}(\hat{\lambda}_{ML}=4)$  haben die Restaurants von Lisa und Mark im Mittel gleich viele Besucher und scheinen somit gleich beliebt zu sein.

3 Punkte. Rechnung (2 TP) und Interpretation (1 TP).

**3.12**

Mit einem  $\chi^2$ -Anpassungstest.

1 Punkt.