FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
FACHBEREICH WIRTSCHAFTS-
WISSENSCHAFTEN

Bachelorprüfung Statistik (RUW-2170), Sommersemester 2020

Liebe Studierende,

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

Wenn Sie eine Antwort korrigieren möchten, füllen Sie bitte die **falsch** markierte Antwort vollständig aus, ungefähr so: ○●⊗ .

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 Aufgaben-stellung 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. oder 3. 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, SS 2020

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.

Die monatlichen Mietausgaben von Studierenden in Nürnberg können durch eine normalverteilte Zufallsvariable X mit unbekanntem Erwartungswert μ und unbekannter Standardabweichung σ dargestellt werden. Sie befragen 100 zufällig ausgewählte Studierende zu ihren monatlichen Mietausgaben. Die durchschnittliche Monatsmiete der Befragten liegt bei $\bar{x}_{100} = 320$ Euro bei einer Stichprobenstandardabweichung von $\hat{\sigma} = 50$ Euro.

1.1 Ein Bekannter behauptet, dass μ bei 300 Euro liegt. Sie bezweifeln diese Behauptung und wollen sie zum Signifikanzniveau $\alpha = 5\%$ gegen eine zweiseitige Alternative testen.

Wie lauten die korrekten Null- und Alternativhypothesen?

A $H_0: \mu = 300, H_1: \mu \neq 300$
B $H_0: \mu \neq 300, H_1: \mu = 300$
C $H_0: \mu = 300, H_1: \mu < 300$
D $H_0: \mu \geq 300, H_1: \mu < 300$
E $H_0: \mu \leq 300, H_1: \mu > 300$

1.2 Wie lautet der korrekte Wert der Teststatistik sowie die Verteilung der Prüfgröße unter H_0?

A $T = 4$ und $T \overset{H_0}{\sim} t_{99}$

B $T = -4$ und $T \overset{H_0}{\sim} t_{99}$

C $T = 4$ und $T \overset{H_0}{\sim} t_{100}$

D $T = -4$ und $T \overset{H_0}{\sim} t_{100}$

E $T = 40$ und $T \overset{H_0}{\sim} t_{99}$

1.3 Es sei $KI_{0.95} = [310.08, 329.92]$ das zentrale Konfidenzintervall zum 95%-Niveau für μ. Wie lautet die korrekte Testentscheidung auf Basis des Konfidenzintervalls?

A Da $300 \notin KI_{0.95}$, wird H_0 auf dem 95%-Niveau verworfen. Wir schließen daraus, dass die Behauptung des Bekannten vermutlich nicht stimmt.

B Da $300 \notin KI_{0.95}$, wird H_0 auf dem 95%-Niveau verworfen. Wir schließen daraus, dass die Behauptung des Bekannten vermutlich stimmt.

C Da $300 \notin KI_{0.95}$, wird H_0 auf dem 95%-Niveau nicht verworfen. Wir schließen daraus, dass die Behauptung des Bekannten vermutlich stimmt.

D Da $300 \in KI_{0.95}$, wird H_0 auf dem 95%-Niveau nicht verworfen. Wir schließen daraus, dass die Behauptung des Bekannten vermutlich stimmt.

E Da $300 \in KI_{0.95}$, wird H_0 auf dem 95%-Niveau verworfen. Wir schließen daraus, dass die Behauptung des Bekannten vermutlich nicht stimmt.

1.4 Welcher der folgenden Zusammenhänge gilt gemäß der sogenannten Lageregel üblicherweise bei rechtsschief verteilten Daten?

A Arithmetisches Mittel < Modus < Median
B Modus < Median < Arithmetisches Mittel
C Median < Modus < Arithmetisches Mittel
D Median < Arithmetisches Mittel < Modus
E Modus < Arithmetisches Mittel < Median

1.5 Gegeben sind zwei Zufallsvariablen X und Y. Die Varianz von X beträgt 6.25, die Varianz von Y beträgt 31.36. Wenn der Korrelationskoeffizient nach Pearson zwischen X und Y bei 0.22 liegt, wie hoch ist dann die Kovarianz?

A 3.08
B 6.57
C 9.49
D 63.64
E 43.12

1.6 Zwischen welchem Paar der folgenden Verteilungen ist die Differenz der Wölbungen (Kurtosis) am größten?

$$N(\mu = 0, \sigma^2 = 3),\ N(\mu = 0, \sigma^2 = 5),\ t_5,\ t_{10}$$

A $N(\mu = 0, \sigma^2 = 3)$ und t_5
B $N(\mu = 0, \sigma^2 = 3)$ und t_{10}
C $N(\mu = 0, \sigma^2 = 5)$ und t_{10}
D $N(\mu = 0, \sigma^2 = 3)$ und $N(\mu = 0, \sigma^2 = 5)$
E t_5 und t_{10}

1.7 Es seien A und B beliebige Ereignisse. Welche der folgenden Aussagen ist nicht immer richtig?

A $P(A \cap B) = 1 - P(\overline{A} \cup \overline{B})$
B $P(A) \leq P(B)$, falls $A \subseteq B$
C $P(\overline{A \cap B}) = 1 - P(A \cup B)$
D $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
E $P(A \cap \overline{B}) = P(A) - P(A \cap B)$

1.8 Bei einer Lotterie werden nacheinander 5 aus 10 Ziffern (ohne Zurücklegen) gezogen ("5 aus 10"). Wie hoch ist die Wahrscheinlichkeit dafür, bei einem zufälligen Tipp genau eine richtige Ziffer zu haben?

A 0.238
B 0.099
C 0.040
D 0.020
E 0.036

1.9 Für eine diskrete Zufallsvariable X mit Verteilungsfunktion F und $a \leq b$ gilt allgemein:

A $P(a \leq X < b) = F(b) - P(X = b) - F(a) + P(X = a)$
B $P(a < X \leq b) = F(b) - P(X = b) - F(a)$
C $P(a < X < b) = F(b) - F(a)$
D $P(a \leq X \leq b) = F(b) - F(a)$
E $P(X \geq a) = 1 - F(a)$

Das Merkmal Z misst die Augenzahl, die sich beim Werfen eines fairen 6-seitigen Würfels ergibt. 12-maliges Werfen ergab folgende Urliste:

i	1	2	3	4	5	6	7	8	9	10	11	12
z_i	1	1	1	1	2	2	2	3	4	5	5	5

1.10 Wie groß ist die beobachtete Spannweite in obiger Urliste?

A 6
B 5
C 4
D $\frac{12}{6}$
E 12

1.11 Welcher der folgenden Werte entspricht dem 25%-Quantil?

A 1
B 3
C 2
D 4
E 5

Eine Zielvariable Y soll durch einen Entscheidungsbaum erklärt werden. Dazu liegen die Merkmalsausprägungen der Merkmale X_1, X_2, X_3, X_4 für $n = 15$ Beobachtungen vor. Die Entropie des Merkmals Y ist

$$E(Y) = 0.8.$$

Die folgende Tabelle zeigt die Entropien, die sich bei Partitionierung der Daten in Teildatensätze anhand der angegebenen Merkmalsausprägungen ergeben:

	$n(\bullet)$	$E(Y\,\vert\,\bullet)$
$X_1 = -1$	5	0.50
$X_1 = 0$	5	0.40
$X_1 = 1$	5	0.80
$X_2 = 1$	5	0.80
$X_2 = 3$	6	0.50
$X_2 = 5$	4	0.70
$X_3 =$ "niedrig"	10	0.10
$X_3 =$ "hoch"	5	0.20
$X_4 = 0.1$	10	0.20
$X_4 = 0.2$	5	0.10

1.12 Welches Merkmal ist gemäß der Änderung der Entropie das informativste Merkmal und würde damit für den Wurzelknoten des Entscheidungsbaumes herangezogen?

A X_1
B X_2
C X_3
D Das informativste Merkmal lässt sich nicht eindeutig bestimmen.
E X_4

1.13 Um die Zufriedenheit seiner Kunden und Kundinnen zu analysieren, möchte ein Busunternehmen eine Befragung durchführen. Dazu sollen die Kunden und Kundinnen interviewt werden. Das Unternehmen weiß, dass sich sein Kundenstamm wie folgt zusammensetzt:

30% Vielfahrer/-innen,
30% Wochenendfahrer/-innen und
40% Gelegenheitsfahrer/-innen

Das Unternehmen möchte eine repräsentative Zusammensetzung der Stichprobe herbeiführen, indem 30% der zu Befragenden aus der Gruppe der Vielfahrer/-innen, 30% aus der Gruppe der Wochenendfahrer/-innen und 40% aus der Gruppe der Gelegenheitsfahrer/-innen zufällig ausgewählt werden.

Wie nennt man eine auf diese Weise gewonnene Stichprobe?

A Einfache Zufallsstichprobe
B Auswahl auf's Geratewohl
C Geschichtete Stichprobe
D Cluster-Stichprobe
E Typische Stichprobe

Im Folgenden sind Charakteristika vierer Datensätze X_1, X_2, X_3 und X_4 gegeben. Gemeinsam ergeben X_1, X_2, X_3 und X_4 einen gepoolten Datensatz X.

	Datensatz X_1	Datensatz X_2	Datensatz X_3	Datensatz X_4
Mittelwert $\bar{x}_i$	10	5	-4	0
Beobachtugen n_i	5	18	109	3

1.14 Was ist das arithmetische Mittel, $\bar{x}$, des gepoolten Datensatzes X?

A 11.74
B -2.19
C 2.75
D 13.67
E 0.08

Gegeben sind die relativen Häufigkeiten, mit denen soziale Netzwerke von Nutzern und Nutzerinnen als Lieblingsnetzwerk genannt werden. Außerdem ist bekannt, dass insgesamt 61% der Befragten weiblich sind.

	Facebook (F)	Youtube (Y)	Instagram (I)
Männlich (M)	0.18	0.14	0.07
Weiblich (W)	?	0.22	0.12

1.15 Wie groß ist der Anteil der Befragten, die weiblich sind und Facebook als Lieblingsnetzwerk angegeben haben? Wie groß ist der Anteil der weiblichen Befragten unter denjenigen, die Instagram als Lieblingsnetzwerk genannt haben?

A $h(W,F) = 0.82$, $h(W|I) = 0.632$
B $h(W,F) = 0.27$, $h(I|W) = 0.197$
C $h(W,F) = 0.27$, $h(W|I) = 0.632$
D $h(W,F) = 0.66$, $h(W|I) = 0.308$
E $h(W,F) = 0.82$, $h(I|W) = 0.197$

1.16 Durch welchen Ausdruck ist der Anteil der Männer unter allen Befragten, die entweder Facebook oder Youtube als Lieblingsnetzwerk genannt haben, gegeben?

A $\frac{h(F)+h(Y)}{h(M)}$
B $h(M,F) + h(M,Y)$
C $\frac{h(M,F)+h(M,Y)}{h(F)+h(Y)}$
D $h(M|F) + h(M|Y)$
E $h(F|M) + h(Y|M)$

Gegeben sei folgende Dichtefunktion f einer stetigen Zufallsvariablen X:

$$f(x) = \begin{cases} 1 - 0.5x & \text{für } 0 \leq x \leq 2 \\ 0 & \text{sonst} \end{cases}$$

1.17 Welches ist die korrespondierende Verteilungsfunktion $F(x)$ auf dem Intervall $[0, 2]$?

A $x - \frac{1}{2}x^2$

B $2 - x^2$

C $1 - \frac{1}{4}x^2$

D $x - \frac{1}{4}x^2$

E $2 - 2x^2$

1.18 Welchen Wert nimmt der Erwartungswert von X an?

A $\frac{1}{6}$

B $\frac{1}{3}$

C $\frac{2}{3}$

D $\frac{3}{2}$

E 1

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.

2.1 Die Dauer von Arbeitslosigkeit in Quartalen (Zufallsvariable X) sei exponentialverteilt mit $\lambda = 0.25$. Mit welcher Wahrscheinlichkeit dauert für einen zufällig ausgewählten Fall die Arbeitslosigkeit länger als 2 Quartale, aber höchstens 10 Quartale?

A 0.5244
B 0.4755
C 0.9179
D 0.6886
E 0.3114

Das Okun'sche Gesetz unterstellt einen negativen Zusammenhang zwischen dem Wirtschaftswachstum und der Arbeitslosigkeit. Um die Validität dieses Zusammenhangs für den Euroraum zu überprüfen, haben Sie im Zeitraum 1999-2018 ($n = 20$) jährliche Werte für die Wachstumsrate des Bruttoinlandsprodukts (X) und der Arbeitslosenrate (Y) erhoben. Gehen Sie davon aus, dass beide Merkmale normalverteilt sind. Sie möchten nun auf dem 5%-Niveau testen, ob die beiden Merkmale unkorreliert sind, d.h. $H_0: \rho_{XY} = 0$ mit Gegenhypothese $H_1: \rho_{XY} \neq 0$. Die empirische Korrelation nach Pearson hat einen Wert von $r_{XY} = -0.3$.

2.2 Wie lautet der korrekte Wert der Teststatistik sowie die Verteilung der Prüfgröße unter H_0?

A $T = -1.3342$ und $T \overset{H_0}{\sim} t_{20}$

B $T = -1.3342$ und $T \overset{H_0}{\sim} t_{18}$

C $T = 1.3342$ und $T \overset{H_0}{\sim} t_{18}$

D $T = -5.6607$ und $T \overset{H_0}{\sim} t_{18}$

E $T = -1.3987$ und $T \overset{H_0}{\sim} t_{19}$

2.3 Eine Stichprobe vom Umfang n wird aus einer Poisson-verteilten Grundgesamtheit X gezogen. Die Dichtefunktion von X hat folgende Gestalt:

$$f_X(x) = (\lambda^x e^{-\lambda}) \,/\, (x!)$$

Welche der folgenden Funktionen stellt die entsprechende logarithmierte Likelihoodfunktion von λ dar?

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

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

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

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

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

2.4 Sie erheben eine Stichprobe mit Umfang $n = 71$ aus einer normalverteilten Grundgesamtheit. Das Stichprobenmittel ist gegeben durch $\bar{x}_{71} = 8$. Die Stichprobenstandardabweichung betrage $\hat{\sigma} = 5$. Berechnen Sie das zentrale Konfidenzintervall für μ zum 99%-Niveau.

A $[6.4287, 9.5713]$
B $[6.8370, 9.1630]$
C $[0.1435, 15.8565]$
D $[7.8135, 8.1865]$
E $[6.8168, 9.1832]$

Ein Roulettetisch besteht aus einer drehbaren Scheibe mit 37 Nummernflächen mit Zahlen von 0 bis 36. Mithilfe einer Kugel, die in jedem Spiel auf nur eine Nummerfläche rollen kann, wird die Gewinnzahl ermittelt. Die Zufallsvariable Z beschreibt die Ergebniszahl nach Drehen der Roulettescheibe. Die folgende Tabelle gibt die Ergebnisse von 11 Durchgängen an:

i	1	2	3	4	5	6	7	8	9	10	11
z_i	0	5	14	14	23	25	27	30	31	34	35

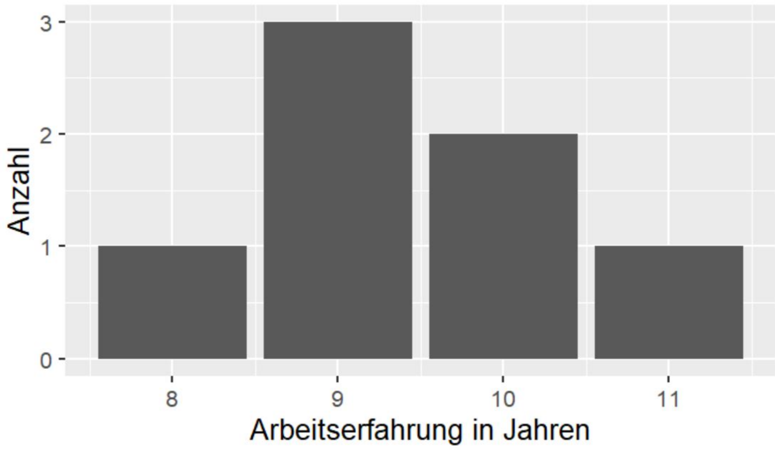
2.5 Was ist der Median dieser Stichprobe?

A 19.73
B 14
C 23
D 24
E 25

2.6 Unabhängig von den konkreten Realisationen von Z, welchen Wert kann der Quartilsabstand maximal annehmen?

A 37
B 18.5
C 35
D 36
E Den Wert des arithmetischen Mittels

Die Zufallsvariable X sei die Arbeitserfahrung in Jahren. Für ein siebenköpfiges Team ergibt sich die folgende Verteilung:



2.7 Was ist der Modus dieser Verteilung?

A 2
B 3
C 8
D 9
E 10

2.8 Es sei X eine Zufallsvariable, die den Jahresumsatz eines Unternehmens in Millionen Euro misst. Der Umsatz sei normalverteilt mit einem Erwartungswert von 22 und einer Varianz von 0.25, d.h., es gilt $X \sim N(22, 0.25)$. Wie groß ist die Wahrscheinlichkeit, dass der Umsatz größer als 20.5 ist?

A 0.6915
B 0.8413
C 0.9332
D 0.9972
E 0.9987

2.9 Seien $X_1,...,X_n$ unabhängig identisch normalverteilte Zufallsvariablen mit Erwartungswert μ und Varianz σ^2. Wodurch sind Erwartungswert und Varianz des arithmetischen Mittels der Zufallsvariablen, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, gegeben?

A $E(\bar{X}_n) = \frac{1}{n}\mu$ und $V(\bar{X}_n) = \frac{1}{n}\sigma^2$
B $E(\bar{X}_n) = n\mu$ und $V(\bar{X}_n) = \frac{1}{n}\sigma^2$
C $E(\bar{X}_n) = \mu$ und $V(\bar{X}_n) = \frac{1}{n^2}\sigma^2$
D $E(\bar{X}_n) = \mu$ und $V(\bar{X}_n) = \frac{1}{n}\sigma^2$
E $E(\bar{X}_n) = \frac{1}{n}\mu$ und $V(\bar{X}_n) = \frac{1}{n^2}\sigma^2$

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**.

Eine Versicherung bietet eine Handyversicherung an. Um die Schadensfälle zu modellieren, betrachtet die Versicherung die Zufallsvariable

X: "Anzahl der Schadensfälle innerhalb eines Monats".

Die Versicherung geht davon aus, dass X Poisson-verteilt ist mit Parameter $\lambda = 7.6$.

3.1 Geben Sie die Wahrscheinlichkeit $P(X = 8)$ an. (3 Punkte)

3.2 Geben Sie alle Zahlen aus der Menge $\{0, -8, 7.6, 4, 100, -1.5\}$ an, die im Träger von X enthalten sind. (3 Punkte)

3.3 Benutzen Sie die Wahrscheinlichkeitsfunktion, um den Wert der Verteilungsfunktion von X an der Stelle 1, $F_X(1)$, zu berechnen und geben Sie dabei Ihren Rechenweg an. (4 Punkte)

3.4 Außerdem betrachten Sie die Zufallsvariable

Y: "Monatliche Gesamtkosten für die Schadensregulierung in Euro",

die durch

$$Y = 5000 + 450 \cdot X$$

gegeben ist. Berechnen Sie den Erwartungswert von Y, $E(Y)$. (3 Punkte)

3.5 Berechnen Sie die Varianz der in der vorherigen Frage gegebenen Zufallsvariable Y, $V(Y)$. (3 Punkte)

Antikörpertests werden verwendet, um festzustellen, ob ein Patient Antikörper gegen ein Virus im Blut hat. Betrachten Sie folgende Ereignisse:

T: "Der Antikörpertest des Patienten fällt positiv aus."
A: "Der Patient hat Antikörper im Blut."

Es sind folgende Wahrscheinlichkeiten bekannt:

$$P(A) = 0.12,\ P(T) = 0.15,\ P(T|A) = 0.9.$$

3.6 Wie hoch ist die Wahrscheinlichkeit, dass ein zufälliger Patient Antikörper im Blut hat und außerdem ein positives Testergebnis erhält? (3 Punkte)

3.7 Nehmen Sie an, der Test fällt positiv aus. Wie hoch ist die Wahrscheinlichkeit, dass der Patient tatsächlich Antikörper im Blut hat? Berechnen Sie dazu die Genauigkeit $P(A|T)$ des Tests. (3 Punkte)

3.8 Berechnen Sie die Trennfähigkeit des Tests, $P(\overline{A}|\overline{T})$, als Funktion von $P(\overline{T}|\overline{A})$. Geben Sie Ihr Ergebnis in der Form $P(\overline{A}|\overline{T}) = c \cdot P(\overline{T}|\overline{A})$ an, wobei c eine aus den oben gegebenen Wahrscheinlichkeiten zu berechnende Konstante darstellt. (4 Punkte)

Mit einer Befragung von jungen Erwachsenen wollen Sie überprüfen, ob Kinder aus Akademiker-Familien häufiger studieren. Dazu soll ein χ^2-Unabhängigkeitstest durchgeführt werden. Sie erfassen mit der Befragung die folgenden Merkmale:

S: "Der/Die Befragte studiert oder hat bereits ein Studium abgeschlossen."
E: "Der/Die Befragte kommt aus einer Akademiker-Familie."

Insgesamt werden $n = 172$ Personen befragt und Sie stellen aus den Daten der einfachen Stichprobe die folgende Kontingenztabelle auf:

	Akademiker-Familie (E)	Keine Akademiker-Familie ($\overline{E}$)	$\sum$
Studium (S)	54	51	105
Kein Studium ($\overline{S}$)	22	45	67
$\sum$	76	96	172

3.9 Geben Sie die Nullhypothese H_0 des χ^2-Unabhängigkeitstests an. (2 Punkte)

3.10 Geben Sie die Prüfgröße für den oben beschriebenen Hypothesentest inklusive asymptotischer Verteilung unter der Nullhypothese an. (3 Punkte)

3.11 Wie lautet der Ablehnbereich mit konkretem Wert? Verwenden Sie eine Irrtumswahrscheinlichkeit von $\alpha = 5\%$. (3 Punkte)

3.12 Berechnen Sie den Wert der Teststatistik, treffen Sie eine Testentscheidung und begründen Sie Ihre Entscheidung. Verwenden Sie dabei eine Irrtumswahrscheinlichkeit von $\alpha = 5\%$.
Verwenden Sie (unabhängig von Ihren Ergebnissen aus den vorherigen Teilaufgaben) eine kritische Schranke von 4.105. (5 Punkte)

Musterlösung

Bachelorprüfung Statistik, SS 2020

1.1	■ A □ B □ C □ D □ E
1.2	■ A □ B □ C □ D □ E
1.3	■ A □ B □ C □ D □ E
1.4	□ A ■ B □ C □ D □ E
1.5	■ A □ B □ C □ D □ E
1.6	■ A □ B □ C □ D □ E
1.7	□ A □ B ■ C □ D □ E
1.8	□ A ■ B □ C □ D □ E
1.9	■ A □ B □ C □ D □ E
1.10	□ A □ B ■ C □ D □ E
1.11	■ A □ B □ C □ D □ E
1.12	□ A □ B ■ C □ D □ E
1.13	□ A □ B ■ C □ D □ E
1.14	□ A ■ B □ C □ D □ E
1.15	□ A □ B ■ C □ D □ E
1.16	□ A □ B ■ C □ D □ E
1.17	□ A □ B □ C ■ D □ E
1.18	□ A □ B ■ C □ D □ E
2.1	■ A □ B □ C □ D □ E
2.2	□ A ■ B □ C □ D □ E
2.3	□ A ■ B □ C □ D □ E
2.4	■ A □ B □ C □ D □ E
2.5	□ A □ B □ C □ D ■ E
2.6	□ A □ B □ C ■ D □ E
2.7	□ A □ B □ C ■ D □ E
2.8	□ A □ B □ C □ D ■ E
2.9	□ A □ B □ C ■ D □ E

3.1

$$P(X=8) = \frac{7.6^8 e^{-7.6}}{8!} = 0.1381$$

3 Punkte. Falls nur korrekte Formel: 1 TP

3.2

0, 4, 100

3 Punkte. Je 1 TP Abzug für jede falsche bzw. fehlende Zahl

3.3

$$F(1) = P(X=0) + P(X=1)$$
$$= \frac{7.6^0 e^{-7.6}}{0!} + \frac{7.6^1 e^{-7.6}}{1!}$$
$$= 0.0005 + 0.0038 = 0.0043$$

4 Punkte. 1 TP für Darstellung als Summe zweier Wahrscheinlichkeiten; 1 TP für Formel der Poissonverteilung; 1 TP für korrekte Berechnung der beiden Einzelwahrscheinlichkeiten; 1 TP für Endergebnis

3.4

$$E(Y) = E(5000 + 450 \cdot X)$$
$$= 5000 + 450 \cdot E(X) = 5000 + 450 \cdot 7.6 = 8420$$

3 Punkte. Umformung des Erwartungswertes (1 TP), Einsetzen von E(X) (1 TP), Endergebnis (1 TP)

3.5

$$V(Y) = V(5000 + 450 \cdot X)$$
$$= 450^2 \cdot V(X) = 450^2 \cdot 7.6 = 1'539'000$$

3 Punkte. Umformung der Varianz (1 TP), Einsetzen von V(X) (1 TP), Endergebnis (1 TP)

3.6

$$P(A, T) = P(T|A) \cdot P(A) = 0.9 \cdot 0.12 = 0.108$$

3 Punkte. Falls nur Ansatz erkennbar (erste Teilgleichung): 2 Punkte.

3.7

$$P(A|T) = \frac{P(T|A)P(A)}{P(T)} = 0.72$$

3 Punkte. Falls nur Ansatz erkennbar (erste Teilgleichung): 2 Punkte.

3.8
$P(\overline{A}\|\overline{T}) =$ $= \frac{1-P(A)}{1-P(T)} \cdot P(\overline{T}\|\overline{A})$ *(Gl. 1)* $= \frac{1-0.12}{1-0.15} \cdot P(\overline{T}\|\overline{A})$ *(Gl. 2)* $= 1.0353 \cdot P(\overline{T}\|\overline{A})$ *(Gl. 3)* *4 Punkte. 2 TP für Ansatz (Gl. 1), je 1 TP für Gl. 2 und 3*
3.9
$H_0: p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$ *Alternativ: Die betrachteten Merkmale sind stochastisch unabhängig.* *2 Punkte.*
3.10
$X^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(N_{ij} - \frac{N_{i\cdot} N_{\cdot j}}{n}\right)^2}{\frac{N_{i\cdot} N_{\cdot j}}{n}} \sim X^2_{(k-1)(l-1)}$ *3 Punkte. 2 TP für TS, 1 TP für asy. Verteilung*
3.11
$X^2 > X^2_{(k-1)(l-1);0.95} = X^2_{1;0.95} = 3.841$ *3 Punkte. 1 TP für Identifikation des kritischen Bereichs (Größer-/Kleinerzeichen), 2 TP für richtigen Wert des Quantils. Falls richtiges Quantil als $X^2_{3;0.95}$ angegeben, aber falscher Wert aus Tabelle: 1 TP Abzug.*
3.12
$X^2 = 5.7332 > X^2_{1;0.95} = (4.105)\ 3.841$ *Die Nullhypothese kann bei einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$ abgelehnt werden.* *5 Punkte. Wert der Teststatistik (3 TP, ggf. Folgefehler), Testentscheidung (1 TP), Begründung (1 TP)*