

Bachelorprüfung Statistik (RUW-2172), Wintersemester 2019/2020

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 die Textfragen 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

MUSTER
Nicht ausfüllen!

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**.
- Der Single-Choice-Teil (Aufgaben 1 und 2) umfasst 25 Single-Choice-Fragen.
- Der Freitextteil (Aufgabe 3) umfasst offene Aufgaben, die in den Lösungsfeldern in diesem Geheft zu beantworten sind.
- 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!** Nutzen Sie bitte einen **dunklen Kugelschreiber!**
- Beschriften Sie den Antwortbogen deutlich lesbar mit Ihrem Namen und Ihrer Matrikelnummer!
- 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

Viel Erfolg!

MUSTER
Nicht ausfüllen!

Bachelorprüfung Statistik, WS 2019/2020

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.

- 1.1** Welcher der folgenden Zusammenhänge gilt bei linksschief verteilten Daten?
- A Arithmetisches Mittel < Modus < Median
 - B Arithmetisches Mittel < Median < Modus
 - C Median < Modus < Arithmetisches Mittel
 - D Median < Arithmetisches Mittel < Modus
 - E Modus < Arithmetisches Mittel < Median
- 1.2** Die zufällige Abweichung der Anzeige einer Feinwaage vom wahren Gewicht kann durch eine normalverteilte Zufallsvariable X mit bekanntem Erwartungswert $\mu = 0$ mg und unbekannter Varianz σ^2 dargestellt werden. Der Hersteller garantiert die Genauigkeit der Feinwaage, indem er behauptet, dass $\sigma \leq \sigma_0 = 0.005$ mg sei. Sie bezweifeln diese Aussage und erheben eine Stichprobe vom Umfang $n = 10$. Nehmen Sie an, dass die Messungen X_1, \dots, X_{10} unabhängig und identisch verteilt sind. Nach der Erhebung erhalten Sie einen Schätzwert von $\hat{\sigma} = 0.007$. Wie lautet der realisierte Wert der Prüfgröße T für den Hypothesentest, den Sie durchführen müssen?
- A $t = 12.60$
 - B $t = 19.60$
 - C $t = 17.64$
 - D $t = 158.76$
 - E $t = 4.59$
- 1.3** Gegeben sind zwei Zufallsvariablen X und Y . Die Varianz von Y beträgt 41.5 und die Korrelation zwischen X und Y beträgt 0.94. Wenn die Kovarianz zwischen X und Y 12.4 beträgt, wie hoch ist dann die Varianz von X ?
- A 4.19
 - B 2.06
 - C 0.32
 - D 0.10
 - E 1.86

1.4 Ordnen Sie die folgenden Verteilungen absteigend nach der Höhe ihrer Wölbung (Kurtosis):

$$N(0, 1), N(0, 2), t_5, t_7$$

- A $K(t_5) > K(t_7) > K(N(0, 1)) = K(N(0, 2))$
- B $K(t_5) > K(t_7) > K(N(0, 2)) > K(N(0, 1))$
- C $K(N(0, 1)) > K(N(0, 2)) > K(t_7) > K(t_5)$
- D $K(t_7) > K(t_5) > K(N(0, 2)) > K(N(0, 1))$
- E $K(t_7) > K(N(0, 1)) = K(N(0, 2)) > K(t_5)$

1.5 Es seien A und B beliebige Ereignisse. Welche der folgenden Aussagen ist allgemein richtig?

- A $P(A \setminus B) = P(A) - P(B) + P(A \cap B)$
- B $B \subset A \Rightarrow P(B) \geq P(A)$
- C $P(A \cap B) = 1 - P(\bar{A} \cup \bar{B})$
- D $P(A \cup B) = P(A) + P(B) + P(A \cap B)$
- E $P(A \cap \bar{B}) = P(A) - P(A \cup \bar{B})$

1.6 Bei einer Lotterie werden nacheinander 4 aus 10 Ziffern (ohne Zurücklegen) gezogen ("4 aus 10"). Wie hoch ist die Wahrscheinlichkeit dafür, bei einem zufälligen Tipp genau 2 Richtige zu haben?

- A 0.029
- B 0.429
- C 0.400
- D 0.071
- E 0.486

1.7 Es sei X eine diskrete Zufallsvariable, die nur die Werte 0, 1, 2, 3 und 4 annehmen kann. Es bezeichne F die zugehörige Verteilungsfunktion. Dann gilt:

- A $F(2.5) = F(3)$
- B $F(2.5) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$
- C $F(2.5) = P(X=0) + P(X=1) + P(X=2) + P(X=2.5)$
- D $F(2.5) = 1 - P(X > 3)$
- E $F(2.5) = P(X=2)$

500 zufällig ausgewählten Männern und 500 zufällig ausgewählten Frauen über 18 Jahren wurde die Frage "Welcher der beiden folgenden Aussagen stimmen Sie eher zu?" gestellt. Die Antwortmöglichkeiten lauten:

Antwort A: Die Grenze von 0.5 Promille beim Autofahren sollte beibehalten werden.

Antwort B: Die Grenze sollte auf 0.3 Promille gesenkt werden.

Die Auswertung der Antworten ergab folgende Kontingenztabelle:

	Antwort A	Antwort B	"weiß nicht"	Σ
Männer	325	150	25	500
Frauen	136	315	49	500
Σ	461	465	74	1000

1.8 Sie wollen anhand von einem Unabhängigkeitstest testen, ob das Antwortverhalten unabhängig vom Geschlecht ist. Mit welcher Rechnung erhalten Sie den korrekten Wert der Teststatistik?

A $\chi^2 = \frac{(325-230.5)^2}{230.5} + \frac{(150-232.5)^2}{232.5} + \frac{(25-37)^2}{37} + \frac{(136-230.5)^2}{230.5} + \frac{(315-232.5)^2}{232.5} + \frac{(49-37)^2}{37}$

B $\chi^2 = \frac{(325-230.5)^2}{325} + \frac{(150-232.5)^2}{150} + \frac{(25-37)^2}{25} + \frac{(136-230.5)^2}{136} + \frac{(315-232.5)^2}{315} + \frac{(49-37)^2}{49}$

C $\chi^2 = \frac{325-230.5}{230.5^2} + \frac{150-232.5}{232.5^2} + \frac{25-37}{37^2} + \frac{136-230.5}{230.5^2} + \frac{315-232.5}{232.5^2} + \frac{49-37}{37^2}$

D $\chi^2 = \left(\frac{325-230.5}{230.5}\right)^2 + \left(\frac{150-232.5}{232.5}\right)^2 + \left(\frac{25-37}{37}\right)^2 + \left(\frac{136-230.5}{230.5}\right)^2 + \left(\frac{315-232.5}{232.5}\right)^2 + \left(\frac{49-37}{37}\right)^2$

E $\chi^2 = \frac{325^2-230.5}{230.5} + \frac{150^2-232.5}{232.5} + \frac{25^2-37}{37} + \frac{136^2-230.5}{230.5} + \frac{315^2-232.5}{232.5} + \frac{49^2-37}{37}$

1.9 Als Ernesto feststellt, dass sein Pizzaservice gut läuft, eröffnet er eine zweite Filiale. Es sei X der Tagesumsatz der ersten Filiale und Y der Tagesumsatz am zweiten Standort (beide in Euro). Nehmen Sie an, dass X und Y normalverteilt sind, mit Erwartungswerten μ_X und μ_Y . Gehen Sie davon aus, dass die Varianzen in den beiden Grundgesamtheiten gleich sind. Eine Analyse der Umsätze in beiden Filialen anhand von zwei unabhängigen Stichproben mit Umfängen $n_X = 50$ und $n_Y = 40$ ergab:

$$\bar{X}_{50} = 530.9 \text{ und } \bar{Y}_{40} = 617.9$$

Die gepoolte Stichprobenstandardabweichung aus beiden Datensätzen beträgt $\hat{\sigma} = 248.3$.

Ernesto möchte testen, ob es signifikante Unterschiede in den Umsätzen der beiden Filialen gibt. Dazu testet er die Nullhypothese $H_0: \mu_X = \mu_Y$ auf dem 5%-Niveau mit einem Zwei-Stichproben- t -Test. Wie lauten der korrekte Wert der Prüfgröße und die korrekte Testentscheidung wenn $t_{0.975}(88) \approx 1.9867$?

A $t = -17.4314 \Rightarrow H_0$ ablehnen, da $|t| > 1.9867$.

B $t = -38.9426 \Rightarrow H_0$ ablehnen, da $|t| > 1.9867$.

C $t = -1.6517 \Rightarrow H_0$ nicht ablehnen, da $|t| < 1.9867$.

D $t = -1.6517 \Rightarrow H_0$ ablehnen, da $|t| < 1.9867$.

E $t = -38.9426 \Rightarrow H_0$ nicht ablehnen, da $|t| > 1.9867$.

- 1.10 Welche Behauptung über Konfidenzintervalle der Form $\left[\bar{X}_n \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$ ist **nicht** korrekt?
- A Die Länge eines Konfidenzintervalls nimmt mit wachsendem Stichprobenumfang ab.
 - B Je größer das Konfidenzniveau gewählt wird, desto breiter ist das Konfidenzintervall.
 - C Die Länge eines Konfidenzintervalls ist umso größer, je größer der zu schätzende Parameter ist.
 - D Je größer die Varianz in der Grundgesamtheit, desto breiter ist das Konfidenzintervall.
 - E Die Länge des Konfidenzintervalls ist ex ante eine Zufallsvariable.

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 Merkmalsausprägung ergeben:

	$n(\bullet)$	$E(Y \bullet)$
$X_1 = -1$	5	0.50
$X_1 = 0$	5	0.40
$X_1 = 1$	5	0.80
$X_2 = 1$	5	0.10
$X_2 = 3$	6	0.50
$X_2 = 5$	4	0.00
$X_3 = \text{"niedrig"}$	10	0.80
$X_3 = \text{"hoch"}$	5	0.80
$X_4 = 0.1$	10	0.70
$X_4 = 0.2$	5	0.80

- 1.11 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

- 1.12 Die Verteilung einer Zufallsvariable X sei durch eine geometrische Verteilung gegeben:

$$P(X = x) = (1-p)^x p, \quad x = 0, 1, 2, \dots$$

Welcher Schätzer des Parameters, \hat{p} , ergibt sich mit der Methode der Momente?

Hinweis: \bar{X} bezeichnet das Stichprobenmittel.

- A $\hat{p} = \frac{1}{\bar{X}+1}$
- B $\hat{p} = \frac{1}{\bar{X}-1}$
- C $\hat{p} = \bar{X}$
- D $\hat{p} = \frac{1}{1-\bar{X}}$
- E $\hat{p} = \frac{1}{\bar{X}}$

Im Folgenden sind Charakteristika dreier Datensätze A, B und C gegeben. Gemeinsam ergeben A, B und C einen gepoolten Datensatz Z.

	Datensatz A	Datensatz B	Datensatz C
Mittelwert \bar{x}_i	12	25	-4
Varianz s_i^2	12	8	27
Beobachtungen n_i	130	50	90

1.13 Was ist das arithmetische Mittel des gepoolten Datensatzes Z?

- A 11.74
- B 9.07
- C 11
- D 13.67
- E 33

1.14 Die Varianz des gepoolten Datensatzes beträgt 124.33. Bestimmen Sie die Varianz zwischen den Datensätzen gemäß folgender Varianzzerlegung:

$$s^2 = \sum_{i=1}^k \frac{n_i}{n} s_i^2 + \sum_{i=1}^k \frac{n_i}{n} (\bar{x}_i - \bar{x})^2$$

- A 16.26
- B 47.00
- C 108.07
- D 124.32
- E 75.95

Im Folgenden sind die Umsätze der fünf Online-Versandhändler eines Landes dargestellt (in Mio. Euro).

Online-Versandhändler	1	2	3	4	5
Umsatz	60	78	78	43	41

1.15 Berechnen Sie die Konzentrationsrate 2. Ordnung. Die Konzentrationsrate h -ter Ordnung ist gegeben durch:

$$CR_h = \sum_{i=1}^h \frac{x_{(n+1)}}{n\bar{x}}$$

- A 0.52
- B 0.72
- C 1.93
- D 0.16
- E 0.41

Gegeben sind die relativen Häufigkeiten, dass ein/-e Student/-in (M/W) eine Bachelorarbeit in den Bereichen Wirtschaftsinformatik (WI), VWL bzw. BWL anfertigt. Außerdem ist bekannt, dass insgesamt 60% der Studierenden weiblich sind.

	WI	VWL	BWL
M	0.04	0.20	0.16
W	0.30	0.12	?

- 1.16** Wie groß ist der Anteil der Studierenden, die weiblich sind und gleichzeitig ihre Bachelorarbeit im Bereich BWL geschrieben haben, $h(W, BWL)$? Und wie groß ist der Anteil der weiblichen Studierenden unter denjenigen, die die Bachelorarbeit im Bereich WI geschrieben haben, $h(W|WI)$?
- A** $h(W, BWL) = 0.20$, $h(W|WI) = 0.50$
B $h(W, BWL) = 0.18$, $h(W|W) = 0.118$
C $h(W, BWL) = 0.18$, $h(W|WI) = 0.882$
D $h(W, BWL) = 0.20$, $h(W|WI) = 0.882$
E $h(W, BWL) = 0.18$, $h(W|W) = 0.018$

Gegeben sei die folgende Dichtefunktion f einer stetigen Zufallsvariablen X :

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

- 1.17** Welches ist die korrespondierende Verteilungsfunktion $F(x)$ auf dem Intervall $[0, 1]$?
- A** $2x - x$
B $\frac{1}{2}x - x^2$
C $x - x^2$
D $2x - x^3$
E $2x - x^2$

Die Zufallsvariable X sei exponentialverteilt mit dem Parameter $\lambda = 2$ und folgender Dichtefunktion:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

- 1.18** Was ist $E(X)$, gegeben, dass $X > 2$ bekannt ist, d.h. $E(X|X > 2)$?
- A** $\frac{1}{2}$
B $\frac{5}{2}$
C $\frac{1}{4}$
D $\frac{1}{e^2}$
E 2

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

MUSTER
Nicht ausfüllen!

Aufgabe 2: Single-Choice-Fragen zu R

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 Zufallsvariable X_i gebe an, ob in einem Basketballspiel beim i -ten Freiwurf getroffen wird ($X_i = 1$ bei Treffer, $X_i = 0$ bei Fehlwurf). Es gelte $P(X_i = 1) = 0.7$. Es werden fünf Freiwürfe ausgeführt. $X = \sum_{i=1}^5 X_i$ bezeichne die Summe dieser fünf voneinander unabhängigen Zufallsvariablen.

Mit welchem Befehl können Sie **nicht** die Wahrscheinlichkeit berechnen, dass die Anzahl der Treffer höchstens drei ist, d.h. $P(X \leq 3)$?

- A** `pbinom(3, size=5, prob=0.7)`
- B** `dbinom(0, size=5, prob=0.7)+dbinom(1, size=5, prob=0.7)+dbinom(2, size=5, prob=0.7)+dbinom(3, size=5, prob=0.7)`
- C** `1-(dbinom(4, size=5, prob=0.7)+dbinom(5, size=5, prob=0.7))`
- D** `pbinom(4, size=5, prob=0.7)-dbinom(4, size=5, prob=0.7)`
- E** `1-pbinom(4, size=5, prob=0.7)`

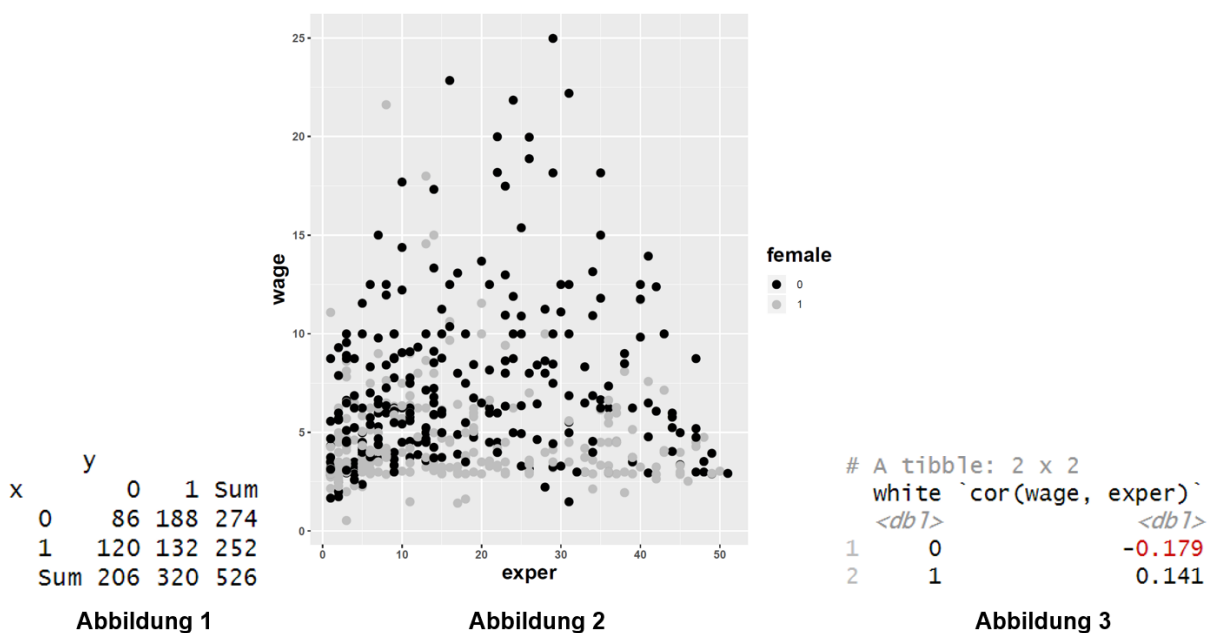
Gehen Sie von folgendem Workspace in R Studio aus.

Der Dataframe `df_wages` enthält Informationen von 526 befragten Personen zu fünf Merkmalen:

1. Stundenlohn in US-Dollar (`wage`)
2. Ehestatus (`married`; =1 für verheiratete Personen, 0 sonst)
3. Geschlecht (`female`; =1 für Frauen, 0 sonst)
4. Hautfarbe (`white`; =1 für weiße Personen, 0 sonst)
5. Arbeitserfahrung in Jahren (`exper`)

Für jede Person liegen Ihnen vollständige Informationen zu allen Merkmalen vor (d.h. es gibt keine NA's). Es gibt keine weiteren Spalten im Dataframe und Sie haben auch sonst keine Datenobjekte (z.B. Values oder Funktionen) abgespeichert. Sie haben das Paket *tidyverse* in Ihrer aktuellen Session bereits aktiviert.

Im Rahmen Ihrer Analyse sind die folgenden Tabellen / Grafiken entstanden:



2.2 Welcher Befehl zeigt Ihnen die Anzahl weißer bzw. nicht-weißer Personen an?

- A `mean(df_wages$white)`
- B `table(df_wages$white)`
- C `length(df_wages$white)`
- D `median(df_wages$white)`
- E `unique(df_wages$white)`

2.3 Betrachten Sie den folgenden Befehl:

```
ggplot(data = X, aes(x = exper)) + Y(Z)
```

Für welche Wahl von **X**, **Y** und **Z** erhalten Sie ein Histogramm für die Arbeitserfahrung mit einer Klassenbreite von fünf Jahren?

- A **X**: `df_wages`, **Y**: `geom_bar`, **Z**: `binwidth = 5`
- B **X**: `df_wages`, **Y**: `geom_histogram`, **Z**: `binwidth = 10`
- C **X**: `df_WAGES`, **Y**: `geom_histogram`, **Z**: `binwidth = 5`
- D **X**: `df_wages`, **Y**: `geom_histogram`, **Z**: `bins = 5`
- E **X**: `df_wages`, **Y**: `geom_histogram`, **Z**: `binwidth = 5`

2.4 Sie führen den Befehl

```
addmargins(table(x = df_wages$female, y = df_wages$married))
```

aus und erhalten den Output aus **Abbildung 1**.

Wie hoch ist der Anteil der verheirateten Personen unter den weiblichen Befragten?

- A 0.7460
- B 0.4125
- C 0.5238
- D 0.4762
- E 0.3413

2.5 Welcher Befehl liefert Ihnen **nicht** das arithmetische Mittel der Arbeitserfahrung?

- A `mean(df_wages$exper)`
- B `sum(df_wages$exper)/length(df_wages$exper)`
- C `sum(df_wages$exper)/NROW(df_wages$exper)`
- D `summary(df_wages$exper)`
- E `total(df_wages$exper)/n(df_wages$exper)`

2.6 Welcher Befehl liefert Ihnen **nicht** den Medianlohn der weiblichen Befragten?

- A `df_wages %>% group_by(female) %>% summarize(median(wage))`
- B `median(df_wages$wage[df_wages$female==1])`
- C `df_wages %>% filter(female==1) %>% summarize(median(wage))`
- D `quantile(df_wages$wage[df_wages$female==1], probs = 0.5, type=1)`
- E `median(df_wages$wage) if female==1`

2.7 Betrachten Sie das Streudiagramm für die Stundenlöhne und die Arbeitserfahrung aus **Abbildung 2**.

Vervollständigen Sie den Befehl

```
ggplot(data = W, aes(x = X, y = Y)) + geom_point(aes(color = Z))
```

so, dass das in **Abbildung 2** gezeigte Streudiagramm erstellt wird.

- A **W**: df_wages, **X**: exper, **Y**: female, **Z**: wage
- B **W**: df_wages, **X**: exper, **Y**: wage, **Z**: female
- C **W**: df_wages, **X**: wage, **Y**: exper, **Z**: female
- D **W**: df_wages, **X**: exper, **Y**: wage, **Z**: married
- E **W**: df_WAGES, **X**: exper, **Y**: wage, **Z**: female

2.8 Sie führen den Befehl

```
df_wages %>% group_by(white) %>% summarize(cor(wage, exper))
```

aus und erhalten den Output aus **Abbildung 3**.

Wie lautet die korrekte Interpretation der Ergebnisse?

- A** Es gibt einen positiven linearen Zusammenhang zwischen der Arbeitserfahrung und den Stundenlöhnen.
- B** Es gibt einen negativen linearen Zusammenhang zwischen der Arbeitserfahrung und den Stundenlöhnen.
- C** Unter den weißen Befragten gibt es einen positiven linearen Zusammenhang zwischen der Arbeitserfahrung und den Stundenlöhnen. Unter den restlichen Befragten ist der Zusammenhang negativ.
- D** Unter den weißen Befragten gibt es einen negativen linearen Zusammenhang zwischen der Arbeitserfahrung und den Stundenlöhnen. Unter den restlichen Befragten ist der Zusammenhang positiv.
- E** Keine der Aussagen lässt sich aus den Werten ableiten.

2.9 Betrachten Sie den folgenden Befehl:

```
df_wages %>% summarize(mean(married),
                        (length(wage)-1)/length(wage) * var(wage),
                        cor(married, wage))
```

Welche deskriptiven Statistiken liefert dieser Befehl?

- A** Den Anteil der nicht-verheirateten Befragten, die deskriptive Varianz der Stundenlöhne und den Pearson-Korrelationskoeffizient zwischen dem Ehestatus und den Stundenlöhnen.
- B** Den Anteil der verheirateten Befragten, die deskriptive Varianz der Arbeitserfahrung und den Pearson-Korrelationskoeffizient zwischen dem Ehestatus und den Stundenlöhnen.
- C** Den Anteil der verheirateten Befragten, die deskriptive Varianz der Stundenlöhne und den Spearman-Korrelationskoeffizient zwischen dem Ehestatus und den Stundenlöhnen.
- D** Den Anteil der verheirateten Befragten, die deskriptive Varianz der Stundenlöhne und den Pearson-Korrelationskoeffizient zwischen dem Ehestatus und den Stundenlöhnen.
- E** Den Anteil der verheirateten Befragten, die Stichprobenvarianz der Stundenlöhne und den Pearson-Korrelationskoeffizient zwischen dem Ehestatus und den Stundenlöhnen.

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 Zufallsvariablen X und Y sind die täglichen Renditen zweier Aktien, angegeben in Prozent. Nehmen Sie an, dass die Renditen der Aktien stochastisch unabhängig sind, so dass $\text{Cov}(X, Y) = 0$ gilt. Ein Portfolio aus beiden Aktien hat die Rendite Z , gegeben durch

$$Z = \alpha X + (1 - \alpha) Y.$$

Dabei ist $\alpha \in [0, 1]$ der Anteil des Kapitals, der in die erste Aktie (Zufallsvariable X) investiert ist.

- 3.1** Leiten Sie einen Ausdruck für den Erwartungswert der Portfoliorendite Z , $E(Z)$, her. (3 Punkte)
-

MUSTER
Nicht ausfüllen!

3.2 Leiten Sie einen Ausdruck für die Varianz der Portfoliorendite Z , $Var(Z)$ her. (3 Punkte)

3.3 Welcher Verteilungsfamilie folgt die Portfoliorendite Z unter der Annahme, dass X und Y unabhängig normalverteilt sind? (3 Punkte)

MUSTER
Nicht ausfüllen!

- 3.4** Sie wollen testen, ob die Korrelation der Renditen beider Aktien statistisch signifikant verschieden von null ist. Konkret betrachten Sie folgendes Hypothesenpaar:

$$H_0: \rho = 0, \text{ mit } H_1: \rho \neq 0$$

Aus $n = 252$ Beobachtungen erhalten Sie einen Korrelationskoeffizienten nach Pearson von $r_{XY} = 0.5$. Geben Sie die relevante Prüfgröße allgemein an und berechnen Sie dann ihren realisierten Wert.

(3 Punkte)

MUSTER
Nicht ausfüllen!

Börsennotierte Unternehmen müssen bestimmte geschäftsrelevante Meldungen veröffentlichen. Sei die Zufallsvariable M die Anzahl der an einem Handelstag veröffentlichten Meldungen, z.B. Gewinnwarnungen.

Die nachfolgende Tabelle zeigt die Häufigkeit n_i , mit der m_i Meldungen an einem der $n = 252$ Handelstage des Jahres 2018 veröffentlicht wurden.

Sie wollen die Nullhypothese testen, dass die beobachteten Häufigkeiten aus einer Poissonverteilung stammen, d.h. dass

$$H_0: M \text{ ist } \mathcal{P}(\cdot)\text{-verteilt, mit } H_1: M \text{ ist nicht } \mathcal{P}(\cdot)\text{-verteilt.}$$

Der Parameter λ der Poissonverteilung soll dabei aus den Daten geschätzt werden. Die Spalte \hat{p}_i^0 gibt die aus einer Poissonverteilung mit $\hat{\lambda} = 1.34$ erhaltenen Wahrscheinlichkeiten an.

i	m_i	n_i	\hat{p}_i^0	$n\hat{p}_i^0$
1	0	74	0.2618	65.9851
2	1	83	0.3509	88.4200
3	2	55	0.2351	59.2414
4	3	22	0.1050	26.4612
5	4	18	0.0352	8.8645

3.5 Nennen Sie ein geeignetes Verfahren zur Bestimmung des Parameters λ und schätzen Sie den Parameter λ aus den Daten des Jahres 2018 mit einer Methode Ihrer Wahl. (4 Punkte)

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

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

MUSTER
Nicht ausfüllen!

- 3.8** 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) einen Wert von $\hat{\lambda} = 1.34$ für den geschätzten Parameter und eine kritische Schranke von 9.71. (4 Punkte)
-

MUSTER
Nicht ausfüllen!

Die folgende Tabelle zeigt die täglichen Renditen in Prozent des deutschen Aktienindex (DAX, Zufallsvariable X) für die ersten 91 Tage des Jahres 2018 (Stichprobenumfang $n = 91$). Die letzte Zeile gibt dabei das arithmetische Mittel der entsprechenden Spalte an.

i	Datum	x_i	x_i^2
1	2018-01-02	-0.3580	0.1282
2	2018-01-03	0.8299	0.6887
3	2018-01-04	1.4615	2.1361
4	2018-01-05	1.1524	1.3281
...			
88	2018-05-08	-0.2775	0.0770
89	2018-05-09	0.2389	0.0571
90	2018-05-10	0.6166	0.3802
91	2018-05-11	-0.1661	0.0276
$\frac{1}{n} \sum_{i=1}^{91}$		0.0122	1.0152

3.9 Berechnen Sie die Stichprobenvarianz aus den gegebenen Daten. (3 Punkte)

MUSTER
Nicht ausfüllen!

3.10 Sie wollen die Nullhypothese testen, dass die durchschnittliche tägliche Rendite des DAX in den ersten 91 Tagen des Jahres 2018 gleich null war, d.h. dass

$$H_0: \mu = 0, \text{ mit } H_1: \mu \neq 0.$$

Die Varianz der Grundgesamtheit sei unbekannt und die tägliche Renditen seien normalverteilt. Geben Sie die theoretische Prüfgröße inklusive Verteilung unter der Nullhypothese an.
(3 Punkte)

3.11 Wie lautet der Ablehnbereich mit konkreten Werten? Verwenden Sie eine Irrtumswahrscheinlichkeit von $\alpha = 1\%$. (3 Punkte)

MUSTER
Nicht ausfüllen!

3.12 Berechnen Sie den Wert der Teststatistik, treffen Sie eine Testentscheidung und begründen Sie Ihre Entscheidung.

Nehmen Sie dabei (unabhängig von Ihren Ergebnissen aus den vorherigen Teilaufgaben) an, dass die Stichprobenvarianz einen Wert von 1.03 annimmt und dass der kritische Bereich durch $|T| > 2.05$ definiert ist.

(4 Punkte)

MUSTER
Nicht ausfüllen!

Musterlösung

Bachelorprüfung Statistik, WS 2019/2020

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 checked="" type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.4	<input checked="" type="checkbox"/> A	<input 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 type="checkbox"/> B	<input checked="" type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.8	<input checked="" type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.9	<input type="checkbox"/> A	<input type="checkbox"/> B	<input checked="" type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.10	<input type="checkbox"/> A	<input type="checkbox"/> B	<input checked="" type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
1.11	<input type="checkbox"/> A	<input checked="" 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 checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input 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 checked="" type="checkbox"/> A	<input type="checkbox"/> B	<input 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 checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
2.1	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input checked="" type="checkbox"/> E
2.2	<input type="checkbox"/> A	<input checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
2.3	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input checked="" type="checkbox"/> E
2.4	<input type="checkbox"/> A	<input type="checkbox"/> B	<input checked="" type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
2.5	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input checked="" type="checkbox"/> E
2.6	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input checked="" type="checkbox"/> E
2.7	<input type="checkbox"/> A	<input checked="" type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> D	<input type="checkbox"/> E
2.8	<input type="checkbox"/> A	<input type="checkbox"/> B	<input checked="" 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

$$E[Z] = E[\alpha X + (1-\alpha)Y] \text{ (Gl. 1)}$$

$$= E[\alpha X] + E[(1-\alpha)Y] \text{ (Gl. 2)}$$

$$= \alpha E[X] + (1-\alpha)E[Y]$$

3 Punkte. Falls nur Ansatz erkennbar (erste Teilgleichung): 1 Punkt. Falls nur bis einschließlich zweiter Zwischenschritt richtig (zweite Teilgleichung): 2 Punkte

3.2

$$\text{Var}[Z] = \text{Var}[\alpha X + (1-\alpha)Y] \text{ (Gl. 1)}$$

$$= \text{Var}[\alpha X] + \text{Var}[(1-\alpha)Y] \text{ (Gl. 2)}$$

$$= \alpha^2 \text{Var}[X] + (1-\alpha)^2 \text{Var}[Y] \text{ (Gl. 3)}$$

3 Punkte. Falls nur Ansatz erkennbar (erste Teilgleichung): 1 Punkt. Falls nur bis einschließlich zweiter Zwischenschritt richtig (zweite Teilgleichung): 2 Punkte. Falls Quadrate vergessen in Gl. 3, 1 TP Abzug.

3.3

Normalverteilung

3 Punkte (keine TP)

3.4

$$T = \sqrt{n-2} \frac{r_{xy}}{\sqrt{1-r_{xy}^2}}$$

$$t = \sqrt{252-2} \frac{0.5}{\sqrt{1-0.5^2}} = 9.1287$$

3 Punkte. 1 TP für Formel, 2 TP für Zahlenwert (falls richtig eingesetzt, aber nicht korrekt gerechnet: 1 Punkt)

3.5

Methode der Momente:

$$\hat{\lambda} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^k m_i n_i = 1.3135$$

4 Punkte. 1 TP für Nennung einer Methode, 1 TP für Formel, 2 TP für Ergebnis (falls richtig eingesetzt, aber falsches Endergebnis: 1 TP)



3.6

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i^0)^2}{np_i^0} \sim \chi_{k-2}^2$$

3 Punkte. 2 TP für TS, 1 TP für asy. Verteilung

3.7

$$\chi^2 > \chi_{0.95;3}^2 = 7.81$$

3 Punkte. 1 TP für Identifikation des kritischen Bereichs (Größer-/Kleinerzeichen), 2 TP für richtigen Wert des Quantils. Falls richtiges Quantil als $\chi_{0.95;3}^2$ angegeben, aber falscher Wert aus Tabelle: 1 TP Abzug.

3.8

$$\chi^2 = 11.7763 > \chi_{0.95;3}^2 = (9.71) 7.81$$

Die Nullhypothese kann bei einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$ abgelehnt werden.

4 Punkte. Wert der Teststatistik (2 TP), Testentscheidung (1 TP), Begründung (1 TP)

3.9

$$\hat{\sigma}^2 = \frac{n}{n-1} (\overline{X^2} - \bar{X}^2) = 1.0263$$

3 Punkte. Falls richtiger Ansatz, aber falsches Ergebnis: 2 Punkte. Falls Ansatz+Ergebnis, allerdings ohne Korrekturfaktor: 2 Punkte

3.10

$$T = \sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \sim t_{n-1}$$

3 Punkte. 2 TP für TS, 1 TP für Verteilung

3.11

$$|T| > t_{1-\alpha/2; n-1}$$

$$|T| > 2.632$$

3 Punkte. 1 TP für Identifikation des kritischen Bereichs (Vorzeichen/Größer-/Kleinerzeichen), 2 TP für richtigen Wert des Quantils. Falls richtiges Quantil als $t_{1-\alpha/2; n-1}$ angegeben, aber falscher Wert aus Tabelle: 1 TP Abzug.

3.12

$$t = \frac{\sqrt{91} \cdot 0.0122}{\sqrt{1.03}} = 0.1147$$

Die Nullhypothese kann bei einer Irrtumswahrscheinlichkeit von 1% nicht abgelehnt werden, da $-2.05 < 0.1147 < 2.05$.

4 Punkte. Wert der Teststatistik: 2 TP, Testentscheidung: 1 TP, Begründung: 1 TP

