

Bachelorprüfung Data Science: Statistik (21761), Wintersemester 2020/21

Liebe Studierende,

markieren Sie bitte Ihre Antworten auf dem Antwortbogen am Ende des Gehefts in der folgenden Weise: O8O .

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

Bitte füllen Sie folg	gende Angaben deutlich lesbar aus:
Nachname	:
Vorname	
Matrikelnummer	:
Studiengang	:
Raum, Platz	ciille,
Prüfer	: Prof. Dovern
WICHTIG: Bitte k	reuzen Sie Ihre Matrikelnummer auch auf dem Antwortbogen an!
Nachfolgende Ang	gaben sind <u>nur vom Prüfer</u> auszufüllen:
	
	Note:
Unterschrift Prüfer	:



Bitte beachten Sie folgende Hinweise:

- Das Geheft muss zusammen bleiben!
- Die Klausur besteht aus insgesamt 20 Single-Choice-Fragen, von denen 4 R-Bezug haben.
- Verwenden Sie für Ihre Antworten ausschließlich den 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 dort zusätzlich an!
- Verwenden Sie auf dem Antwortbogen bitte einen dunklen Kugelschreiber!
- Bearbeitungszeit: 60 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

VIEI EITOIG!		
	Micht ausfüll	

Bachelorprüfung Data Science: Statistik, WiSe 2020/21

Aufgabe 1

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 16 Teilaufgaben, bei denen jeweils ein Punkt erreicht werden kann. 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 Aussage über statistische Hypothesentests ist allgemein korrekt?
 - **A** Bei der Wahl einer höheren Irrtumswahrscheinlichkeit α , steigt auch die Wahrscheinlichkeit β für einen Fehler 2. Art.
 - **B** Mit wachsendem Stichprobenumfang *n* steigt die Güte eines Hypothesentests an.
 - **C** Mittels des p-Wertes kann man sowohl Aussagen über die statistische Signifikanz als auch über die Größe des Effektes, die sogenannte ökonomische Signifikanz, treffen.
 - D Der p-Wert gibt die Wahrscheinlichkeit an, dass die Nullhypothese richtig ist.
 - E Ist die Irrtumswahrscheinlichkeit α kleiner als der p-Wert, kann die Nullhypothese abgelehnt werden.
- **1.2** Ein befreundeter Börsenmakler erzählt Ihnen, dass sein Portfolio "Top Performance" eine garantierte mittlere Rendite von 5% erzielt. Sie wissen, dass die Rendite durch eine normalverteilte Zufallsvariable X mit unbekanntem Erwartungswert I und bekannter Varianz σ^2 beschrieben werden kann. Sie wollen die Behauptung des Börsenmaklers nun bei einer Irrtumswahrscheinlichkeit von α = 1% mittels eines statistischen Hypothesentests überprüfen.

Sie berechnen dazu folgendes zentrales 99%-Konfidenzintervall für μ :

$$KI_{0.99}$$
 = [2.637%, 5.092%]

Welche Behauptung über die Testentscheidung auf Basis dieses realisierten Konfidenzintervalls ist korrekt?

- **A** Da 5% $\in KI_{0.99}$, wird H_0 auf dem 99%-Niveau nicht verworfen. Die Behauptung des Maklers steht somit nicht im Widerspruch zu den Daten.
- **B** Da 5% $\notin KI_{0.99}$, wird H_0 auf dem 99%-Niveau nicht verworfen. Die Behauptung des Maklers steht somit nicht im Widerspruch zu den Daten.
- **C** Da 5% $\in KI_{0.99}$, wird H_0 auf dem 99%-Niveau verworfen. Die Behauptung des Maklers steht daher im Widerspruch zu den Daten.
- **D** Da 5% $\in KI_{0.99}$, wird H_0 auf dem 99%-Niveau nicht verworfen. Die Behauptung des Maklers steht daher im Widerspruch zu den Daten.
- **E** Da H_0 auf dem 99%-Niveau nicht verworfen wird, hat eine beliebige Erhöhung der Irrtumswahrscheinlichkeit α keinen Einfluss auf die Testentscheidung.

Sie wollen mit einer Befragung von Kommilitoninnen und Kommilitonen überprüfen, ob das Interesse an einer neuen Casting-Show im Fernsehen unabhängig vom Geschlecht ist. Sie befragen dazu n = 200 Personen und erhalten aus den Daten der Stichprobe folgende Kontingenztabelle:

	Männlich (M)	Weiblich (W)	Divers (D)	Σ
Interesse (I)	56	51	1	108
Kein Interesse $(ar{I})$	31	60	1	92
Σ	87	111	2	200

- **1.3** Sie führen einen Unabhängigkeitstest bei einer Irrtumswahrscheinlichkeit von α = 1% durch. Welche kritische Schranke ist bei der Testdurchführung korrekterweise zu wählen?
 - $\mathbf{A} F_{2:3:0.99} = 30.82$
 - **B** $F_{6:200:0.99} = 2.89$
 - **c** $\chi^2_{2;0.99} = 9.21$
 - **D** $t_{6.0.99} = 3.143$
 - $\mathbf{E} \chi^2_{6.0.99} = 16.81$
- 1.4 Durch welche Rechnung erhalten Sie für das Casting-Show-Beispiel von oben die korrekte Teststatistik des Unabhängigkeitstests?
 - **A** $\chi^2 = \frac{(46.98-56)^2}{56} + \frac{(59.94-51)^2}{51} + \frac{(1.08-1)^2}{1} + \frac{(40.02-31)^2}{31}$
 - **B** $\chi^2 = \frac{(56 \cdot 108)^2}{108} + \frac{(51 \cdot 108)^2}{108} + \frac{(1108)^2}{108} + \frac{(3192)^2}{92} + \frac{(60 \cdot 92)^2}{92}$ **C** $\chi^2 = \frac{56 \cdot 46 \cdot 98}{46 \cdot 98} + \frac{51 \cdot 59 \cdot 94}{59 \cdot 94} + \frac{11 \cdot 08}{1 \cdot 08} + \frac{31 \cdot 40 \cdot 02}{40 \cdot 02} + \frac{60 \cdot 51 \cdot 06}{51 \cdot 06}$

 - **E** $\chi^2 = \frac{(87-56)^2}{56} + \frac{(111-51)^2}{51} + \frac{(2-1)^2}{1} + \frac{(87-31)^2}{31} + \frac{(87-31)^2}{31}$
- **1.5** Welche der folgenden Aussagen trifft auf eine einfache Zufallsstichprobe der Größe *n* aus einer Grundgesamtheit der Größe N mit einer Varianz von σ^2 **nicht** zu?
 - **A** Wenn n sehr groß ist im Vergleich zu N, kann man statistische Berechnungen durchführen unter der Annahme, dass die Stichprobenvariablen unabhängig und identisch verteilt sind.
 - B Die Auswahl der Merkmalsträger für die Stichprobe erfolgt "ohne Zurücklegen".
 - C Die einzelnen Ziehungen sind nicht unabhängig voneinander.
 - D Die Auswahlwahrscheinlichkeit für jeden Merkmalsträger aus der Grundgesamtheit beträgt
 - **E** Für $n \ll N$ kann man die Varianz des Stichprobenmittels näherungsweise als $V(\overline{X}) = \frac{\sigma^2}{n}$ berechnen.

- **1.6** Sie möchten eine Studie zum Unterricht an bayerischen Gymnasien durchführen. Dafür befragen Sie alle Schüler an zehn zufällig ausgewählten Gymnasien.
 - Wie nennt man das gewählte Stichprobenverfahren?
 - A Clusterstichprobe bzw. Klumpenstichprobe
 - **B** Einfache Zufallsstichprobe
 - C Geschichtete Stichprobe
 - **D** Quotenstichprobe
 - E Typische Stichprobe
- **1.7** Gegeben sei eine *i.i.d.*-Stichprobe mit n = 20, für die $\sum_{j=1}^{n} x_j = 18$ und $\sum_{j=1}^{n} x_j^2 = 340$ gelte. Berechnen Sie die Stichprobenvarianz auf zwei Nachkommastellen genau.
 - **A** 16.88
 - **B** 18.03
 - **C** 15.31
 - **D** 17.04
 - **E** 16.19
- **1.8** Gegeben seien zwei Schätzer $\hat{\theta}_1$ und $\hat{\theta}_2$, wobei für den mittleren quadratischen Fehler $MQF(\hat{\theta}_1,\theta) > MQF(\hat{\theta}_2,\theta)$ für alle Werte von θ gelte sowie $\lim_{n\to\infty} MQF(\hat{\theta}_1,\theta) = 0$.
 - Welche der folgenden Aussagen ist nicht korrekt?
 - **A** Der Schätzer $\hat{\theta}_2$ ist relativ effizienter als der Schätzer $\hat{\theta}_1$
 - **B** Der Schätzer $\hat{\theta}_1$ ist ein konsistenter Schätzer.
 - **C** Bei $E(\hat{\theta}) \theta \neq 0$ ist der Schätzer $\hat{\theta}$ unverzerrt.
 - **D** Für einen erwartungstreuen Schätzer $\hat{\theta}$ gilt $E(\hat{\theta}) = \theta$.
 - **E** Für $\hat{\theta}_1$ gilt: $\lim_{n\to\infty} E(\hat{\theta}_1) = \theta$
- **1.9** Die Likelihood-Funktion für Beobachtungen $\{X_1, ..., X_n\}$ habe die Form $L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda X_i}$. Wählen Sie die korrekte Loglikelihood-Funktion aus.
 - **A** $ln(L(\lambda)) = ln(\lambda) \lambda \cdot \sum_{i=1}^{n} x_i$
 - **B** $\ln(L(\lambda)) = n \cdot \ln(\lambda) \lambda \sum_{i=1}^{n} x_i$
 - **c** $ln(L(\lambda)) = ln(\lambda) \sum_{i=1}^{n} x_i$
 - **D** $ln(L(\lambda)) = \lambda \lambda \cdot \sum_{i=1}^{n} x_i$
 - $\mathbf{E} \ln(L(\lambda)) = \frac{1}{\lambda} \lambda \cdot \sum_{i=1}^{n} x_i$

- **1.10** Gegeben ist eine Stichprobe der Größe n = 64 einer normalverteilte Zufallsvariable X. Die Varianz σ^2 der Grundgesamtheit sei 16, das wahre Mittel der Grundgesamtheit μ sei unbekannt und wird durch \overline{X} geschätzt. Welche der folgenden Aussagen zum realisierten zweiseitigen 95%-Konfidenzintervall für den tatsächlichen unbekannten Mittelwert ist korrekt?
 - **A** Das realisierte Konfidenzintervall ist $\left[\bar{x} 1.64 \cdot \frac{4}{8}, \bar{x} + 1.64 \cdot \frac{4}{8}\right]$.
 - **B** Das realisierte Konfidenzintervall ist $\left[\bar{x} 1.96 \cdot \frac{16}{64}, \bar{x} + 1.96 \cdot \frac{16}{64}\right]$.
 - C Das realisierte Konfidenzintervall überdeckt den wahren Parameter μ mit einer Wahrscheinlichkeit von 95%.
 - **D** Das realisierte Konfidenzintervall ist $\left[\bar{x} 1.96 \cdot \frac{4}{8}, \bar{x} + 1.96 \cdot \frac{4}{8}\right]$.
 - **E** Das realisierte Konfidenzintervall ist $\left[\bar{x} 1.64 \cdot \frac{16}{64}, \bar{x} + 1.64 \cdot \frac{16}{64}\right]$

Ein Antikörpertests wird 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:

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

- 1.11 Durch welchen Ausdruck und welchen Wert ist die Wahrscheinlichkeit gegeben, bei positivem Testergebnis tatsächlich Antikörper im Blut zu haben? Jichi ausi
 - **A** P(A|T) = 0.84
 - **B** P(T|A) = 0.90
 - **C** P(A|T) = 0.90
 - **D** P(T|A) = 0.84
 - EP(A, T) = 0.126

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

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

Betrachten Sie nun die folgenden Ereignisse:

- A = "Die Summe der Augenzahlen der beiden Würfel beträgt 7"
- B = "Mindestens einer der Würfel zeigt eine 1"

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

- **1.12** Wie groß ist die Wahrscheinlichkeit für P(A|B)
 - **A** P(A|B) = 2/11
 - **B** P(A|B) = 1/18
 - **C** P(A|B) = 6/11
 - **D** P(A|B) = 1/6
 - **E** P(A|B) = 2/36
- 1.13 Sechs Männer und sieben Frauen sind Mitglieder eines Tanzvereins. Für ein Tanzturnier sollen vier Pärchen, jeweils bestehend aus einem Mann und einer Frau, ausgewählt werden. ausfüller Wie viele Paarungen sind in diesem Fall möglich?
 - **A** 12600
 - **B** 525
 - **C** 22
 - **D** 855
 - E 1287

Betrachten Sie die Zufallsvariable Y und die Funktion f(y):

$$f(y) = \begin{cases} \frac{a}{2} \cdot y + 0.25 & \text{für } 0 \le y \le 2\\ 0 & \text{sonst} \end{cases}$$

- **1.14** Für welchen Wert von a ist f(y) eine Dichtefunktion für Y?
 - **A** $a = \frac{1}{2}$
 - **B** $a = \frac{1}{4}$
 - **C** $a = \frac{3}{4}$
 - **D** $a = \frac{1}{3}$
 - **E** $a = \frac{2}{3}$

- **1.15** Die exponentialverteilte Zufallsvariable *X* beschreibt die Lebensdauer eines iPhones in Jahren. Die erwartete Lebensdauer eines iPhones beträgt 4.25 Jahre. Mit welcher Wahrscheinlichkeit beträgt die Lebensdauer eines iPhones mehr als ein Jahr?
 - A 98.570%
 - **B** 1.426%
 - C 23.530%
 - **D** 20.966%
 - E 79.034%
- **1.16** Gegeben Sei eine Poisson-verteilte Zufallsvariable X mit Parameter λ = 4. Wie hoch ist die Wahrscheinlichkeit für $4 \le X \le 6$?
 - **A** 0.456
 - **B** 0.651
 - **C** 0.515
 - **D** 0.261
 - **E** 1.000

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

Aufgabe 2

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 4 Teilaufgaben, bei denen jeweils ein Punkt erreicht werden kann. 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.

Abgesehen von den in den einzelnen Aufgabenteilen explizit erwähnten Datenobjekten haben Sie keine weiteren Datenobjekte (z.B. Dataframes, Values oder Funktionen) abgespeichert. Sie haben das Paket *tidyverse* in Ihrer aktuellen Session bereits aktiviert.

2.1 Welcher der nachfolgeden Befehle gibt es in R **nicht**, um mit einer normalverteilten Zufallsvariable zu arbeiten?

A dnorm()

B pnorm()

C qnorm()

D rnorm()

E xnorm()

2.2 Vervollständigen Sie den nachfolgenden Befehl, um ein Schaubild der Verteilungsfunktion einer normalverteilten Zufallsvariable mit $\mu = 2$ und $\sigma^2 = 9$ zu erstellen:

```
ggplot(data = data.frame(x = -8:12), aes(x = x)) + stat_function(fun = X, args = list(mean = Y, sd = Z))
```

```
A X = pnorm, Y = 2, Z = 3
```

B
$$X = pnorm, Y = 2, Z = 9$$

C
$$X = pnorm, Y = 3, Z = 2$$

D
$$X = dnorm, Y = 2, Z = 3$$

E X = dnorm, Y = 2, Z = 9

2.3 Mit welchem Befehl können Sie eine *i.i.d.*-Stichprobe der Größe n = 4 aus den ganzen Zahlen zwischen 1 und 6 ziehen?

```
A sample(x = seq(from = 1, to = 6, by = 1), size = 4, replace = TRUE)
```

B sample(
$$x = seq(from = 1, to = 6, by = 1), size = 4, replace = FALSE)$$

 \mathbf{C} sample(x = 1:6, size = 4, replace = FALSE)

D sample(x =
$$c(1,2,3,5,6)$$
, size = 4, replace = TRUE)

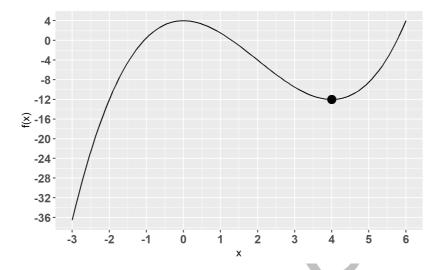
E sample(x = c(1,2,3,4,5,6), size = 4, replace = FALSE)



Betrachten Sie das unten angezeigte Schaubild. Dieses zeigt die Funktion

$$f(x) = 0.5 \cdot x^3 - 3 \cdot x^2 + 4$$

im Bereich $x \in [-3, 6]$. Diese Funktion ist als Objekt unter dem Namen fx in Ihrem Workspace gespeichert.



2.4 Die Funktion hat ein (lokales) Minimum im Punkt (4,-12). Dieses wird im Schaubild durch einen fetten Punkt markiert. Welcher der nachfolgenden Befehle liefert Ihnen als Ergebnis dieses Minimum?

Welcher der nachfolgenden Befehle liefert Ihnen als Ergebnis dieses Minimum?

- A nIm(f = fx, p = -2)
- **B** nlm(f = fx, p = -1)
- **C** nlm(f = -2, p = fx)
- **D** nlm(f = fx, p = 2)
- E nlm(f = 2, p = fx)

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

Musterlösung

Bachelorprüfung Data Science: Statistik, WiSe 2020/21

	2020/21	
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	100
1.15	□A□B□C□D■E	
1.16	■A □B □C □D □E	1510
2.1	□A□B□C□D■E	30
2.2	■A □B □C □D □E	
2.3	■A □B □C □D □E	
2.4	□A □B □C ■D □E	

