

Klausur Modul B2 Inferenzstatistik

Termin: 12.07.2021

Name, Vorname: _____

Matrikelnummer: _____

Bearbeitungshinweise

- Die Klausur besteht aus **30 Aufgaben**. Sie haben zur Bearbeitung **60 Minuten** Zeit.
- Bei jeder Aufgabe sind jeweils **vier Antwortmöglichkeiten** vorgegeben, es trifft **immer genau eine** Antwort zu. Bitte kreuzen Sie bei jeder Aufgabe die zutreffende Antwort an.
- Für jede richtig gelöste Aufgabe erhalten Sie einen Punkt.
- Als Arbeitsunterlagen sind lediglich die ausgegebenen Unterlagen erlaubt. Geben Sie bitte alle Unterlagen wieder ab.
- Es sind **keinerlei** Hilfsmittel erlaubt.
- Vermeiden Sie bitte unnötige Störungen der anderen.
- Jede Kommunikation mit anderen Teilnehmer:innen oder mit Personen außerhalb des Klausorraums während der Klausur gilt als Täuschungsversuch und führt ohne Verwarnung zum Nichtbestehen der Klausur.

Viel Erfolg!

IN DIESER KLAUSURLÖSUNG SIND DIE RICHTIGEN ANTWORTEN MIT EINEM **X** MARKIERT.

1. Es seien $x_1 := 1, x_2 := 2$ und $x_3 := 3$. Welche Aussage trifft dann zu?
- a) $\sum_{i=1}^3 x_i = 3$.
 - X** $\sum_{i=1}^3 x_i = 6$.
 - b) $\sum_{i=1}^3 x_i = 9$.
 - c) $\sum_{i=1}^3 x_i = 12$.
2. f, g_1 und g_2 seien univariate reellwertige Funktionen und es sei $f(x) := g_1(g_2(x))$. Welche Aussage für die Ableitung von f trifft dann zu?
- a) $f'(x) = g_1'(x)g_2(x) + g_1(x)g_2'(x)$.
 - b) $f'(x) = g_1'(x)g_2(x)$.
 - X** $f'(x) = g_1'(g_2(x))g_2'(x)$.
 - c) $f'(x) = g_1'(g_2'(x))$.
3. F sei eine Stammfunktion einer stetigen reellwertigen Funktion f . Welche der folgenden Aussagen wird dann als zweiter Hauptsatz der Differential- und Integralrechnung bezeichnet?
- a) $\int_a^b f(x) dx = F(a) - F(b)$.
 - X** $\int_a^b f(x) dx = F(b) - F(a)$.
 - b) $\int_a^b F(x) dx = f(b) - f(a)$.
 - c) $\int_b^a f(x) dx = F(b) - F(a)$.
4. Welche Aussage zu den fundamentalen Annahmen der Frequentistischen Statistik trifft **nicht** zu?
- X** Wahrscheinlichkeiten werden als Grade subjektiver Sicherheit interpretiert.
 - a) Die Parameter probabilistischer Modelle sind feste, unbekannte Konstanten.
 - b) Statistische Methoden sollten gute langfristige relative Frequenzeigenschaften besitzen.
 - c) Statistische Methoden werden typischerweise anhand ihrer Stichprobenverteilungen bewertet.
5. p sei Wahrscheinlichkeitsdichtefunktion einer Zufallsvariable X . Welche Aussage trifft **nicht** zu?
- a) $p(x) \geq 0$ für alle $x \in \mathbb{R}$.
 - X** $\int_{-\infty}^{\infty} p(x) dx = 0$.
 - b) $\mathbb{P}(a \leq X \leq b) = \int_a^b p(x) dx$ für alle $a, b \in \mathbb{R}, a \leq b$.
 - c) $\mathbb{P}(X = a) = \int_a^a p(x) dx = 0$.

6. Welche Aussage trifft zu? Die kumulative Verteilungsfunktion $P : \mathbb{R} \rightarrow [0, 1]$ einer Zufallsvariable X mit Werten $x \in \mathbb{R}$ ist definiert als

a) $P(x) := \mathbb{P}(X = x)$.

X $P(x) := \mathbb{P}(X \leq x)$.

b) $P(x) := \{x \in \mathbb{R} | P(x) = x\}$.

c) $P(x) := \mathbb{P}(x)$.

7. Welche Aussage trifft zu? Zwei Zufallsvariablen X_1, X_2 mit Ergebnisräumen $\mathcal{X}_1, \mathcal{X}_2$ heißen unabhängig, wenn für alle $S_1 \subseteq \mathcal{X}_1$ und $S_2 \subseteq \mathcal{X}_2$ gilt, dass

a) $\mathbb{P}_X(X_1 \in S_1, X_2 \in S_2) = \mathbb{P}_{X_1}(X_1 \in S_1) + \mathbb{P}_{X_2}(X_2 \in S_2)$.

b) $\mathbb{P}_X(X_1 \in S_1, X_2 \in S_2) = \frac{\mathbb{P}_{X_1}(X_1 \in S_1)}{\mathbb{P}_{X_2}(X_2 \in S_2)}$.

X $\mathbb{P}_X(X_1 \in S_1, X_2 \in S_2) = \mathbb{P}_{X_1}(X_1 \in S_1) \mathbb{P}_{X_2}(X_2 \in S_2)$.

c) $\mathbb{P}_X(X_1 \in S_1, X_2 \in S_2) = \mathbb{P}_{X_1}(X_1 \in S_1) - \mathbb{P}_{X_2}(X_2 \in S_2)$.

8. Es bezeichne $\mathbb{E}(X)$ den Erwartungswert einer Zufallsvariable X . Welche Aussage trifft **nicht** zu?

a) $\mathbb{E}(X) := \int_{-\infty}^{\infty} x p_X(x) dx$, wenn X kontinuierlich mit WDF p_X ist.

b) Der Erwartungswert ist eine skalare Zusammenfassung einer Verteilung.

c) Intuitiv ist $\mathbb{E}(X) \approx \frac{1}{n} \sum_{i=1}^n X_i$ für eine große Zahl n von Kopien X_i von X .

X Es sei $X \sim N(\mu, \sigma^2)$. Dann gilt $\mathbb{E}(X) = \sigma$.

9. Es bezeichne $\mathbb{V}(X)$ die Varianz einer Zufallsvariable X und es sei $a \in \mathbb{R}$ eine Konstante. Dann gilt

X $\mathbb{V}(aX) = a^2 \mathbb{V}(X)$.

a) $\mathbb{V}(aX) = a \mathbb{V}(X)$.

b) $\mathbb{V}(aX) = \mathbb{V}(X)$.

c) $\mathbb{V}(aX) = \mathbb{V}(X)^2$.

10. Es sei X eine Zufallsvariable mit Erwartungswert $\mathbb{E}(X)$ und Varianz $\mathbb{V}(X)$. Die Chebyshev Ungleichung besagt, dass für alle $x \in \mathbb{R}$ gilt, dass

a) $\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}(X)}{x}$.

X $\mathbb{P}(|X - \mathbb{E}(X)| \geq x) \leq \frac{\mathbb{V}(X)}{x^2}$.

b) $\mathbb{P}(X \geq x) \leq \mathbb{E}(X)$.

c) $\mathbb{P}(|X - \mathbb{E}(X)| \leq x) \leq \mathbb{V}(X)$.

11. Es seien $X \sim N(\mu, \sigma^2)$ und $Y := \frac{X-\mu}{\sigma}$. Welche Aussage zur Verteilung von Y trifft dann zu?
- a) $Y \sim N(\mu, \sigma^2)$.
 - b) $Y \sim N(0, \sigma^2)$.
 - c) $Y \sim N(\mu, \sigma^2/n)$.
 - X** $Y \sim N(0, 1)$.
12. $X_1, \dots, X_n \sim N(0, 1)$ seien unabhängig und identisch verteilte standardnormalverteilte Zufallsvariablen und es sei $Y := \sum_{i=1}^n X_i^2$. Welche Aussage zur Verteilung von Y trifft dann zu?
- a) $Y \sim N(0, 1)$.
 - X** $Y \sim \chi^2(n)$.
 - b) $Y \sim t(n-1)$.
 - c) $Y \sim f(n, n-1)$.
13. Welche Aussage zu Statistiken und Schätzern trifft **nicht** zu?
- a) Statistiken sind Abbildungen aus dem Stichprobenraum in einen anderen Raum.
 - b) Das Stichprobenmittel ist eine Statistik.
 - c) Schätzer sind Abbildungen aus dem Stichprobenraum in den Parameterraum.
 - X** Beobachtungen und Statistiken werden mathematisch nie als Zufallsvariablen modelliert.
14. $\mathcal{M}$ sei ein parametrisches statistisches Produktmodell mit WMF oder WDF p_θ , so dass $X_1, \dots, X_n \sim p_\theta$ gilt. Dann ist die Log-Likelihood-Funktion $\ell_n : \Theta \rightarrow \mathbb{R}$ definiert als
- a) $\ell_n(\theta) := \prod_{i=1}^n p_\theta(x_i)$.
 - b) $\ell_n(x_i) := \prod_{i=1}^n p_\theta(x_i)$.
 - X** $\ell_n(\theta) := \ln \prod_{i=1}^n p_\theta(x_i)$.
 - c) $\ell_n(\theta) := \ln \sum_{i=1}^n p_\theta(x_i)$.
15. Welche Aussage zur Cramér-Rao Schranke trifft **nicht** zu?
- a) Die Cramér-Rao-Ungleichung gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an.
 - b) Für $\tau(\theta) = \theta$ gilt $\mathbb{V}_\theta(\hat{\theta}_n) \geq \frac{1}{J(\theta)}$.
 - X** In der Cramér-Rao-Ungleichung bezeichnet $J(\theta)$ die Scorefunktion einer Zufallsvariable.
 - c) Ein erwartungstreu Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuen Schätzer.

16. Welche Aussage zu Maximum Likelihood Schätzern trifft **nicht** zu?
- ☒ Maximum Likelihood Schätzer sind immer erwartungstreu.
 - a) Maximum Likelihood Schätzer sind konsistent.
 - b) Maximum Likelihood Schätzer sind asymptotisch normalverteilt.
 - c) Maximum Likelihood Schätzer sind asymptotisch erwartungstreu.
17. Welche Aussage trifft zu? $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungsmodells, $\bar{X}_n$ bezeichne das Stichprobenmittel und S_n bezeichne die Stichprobenstandardabweichung. Dann ist die T -Statistik definiert als
- a) $T := \sqrt{n} \left(\frac{\bar{X}_n}{S_n} \right)$.
 - ☒ $T := \sqrt{n} \left(\frac{\bar{X}_n - \mu}{S_n} \right)$.
 - b) $T := \sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right)$.
 - c) $T := \frac{\bar{X}_n - \mu}{S_n}$.
18. Es sei $X = X_1, \dots, X_n \sim p_\theta$ eine Stichprobe, $\delta \in]0, 1[$, und $b_l(X)$ und $b_u(X)$ seien zwei Statistiken. Welche Aussage zu einem δ -Konfidenzintervall $C_n := [b_l, b_u]$ trifft dann **nicht** zu?
- a) Es gilt $\mathbb{P}_\theta(C_n \ni \theta) = \delta$ für alle $\theta \in \Theta$.
 - b) C_n ist ein zufälliges Intervall, weil $b_l(X)$ und $b_u(X)$ Zufallsvariablen sind.
 - ☒ C_n ist ein zufälliges Intervall, weil $\theta \in \Theta$ eine Zufallsvariable ist.
 - c) Wird ein Experiment unter gleichen Umständen häufig wiederholt, so überdeckt das zugehörige δ -Konfidenzintervall den wahren, aber unbekannten Parameterwert θ in $\delta \cdot 100\%$ der Fälle.
19. Welche Aussage zur grundlegenden Logik statistischer Hypothesentests trifft **nicht** zu?
- a) Ein vorliegender Datensatz wird als Realisation einer Stichprobe konzeptualisiert.
 - b) Man berechnet basierend auf dem Datensatz eine Teststatistik.
 - ☒ Ist die betrachtete Wahrscheinlichkeit dafür, den beobachteten oder einen extremeren Wert der Teststatistik unter Annahme eines Nullmodells zu observieren groß, so verwirft man die Hypothese, dass das Nullmodell die Daten generiert haben könnte.
 - c) Wie in der frequentistischen Statistik üblich, weiß man nach Durchführung der Testprozedur nicht, ob im vorliegenden Fall das Nullmodell oder ein anderes Modell die Daten generiert hat.
20. Θ sei der Parameterraum eines statistischen Modells. Welche Aussage zu den statistischen Hypothesen Θ_0 und Θ_1 trifft zu?
- a) Es gilt immer $0 \in \Theta_0$.
 - b) Es gilt $\Theta = \Theta_0 \cap \Theta_1$.
 - c) Enthält Θ_1 mehr als ein Element, so heißt Θ_1 einfach.
 - ☒ Bei einer einfachen Hypothese ist die Verteilung der Stichprobe genau festgelegt.

21. Welche Aussage zu Hypothesentests trifft **nicht** zu?

- a) Ein Test ist eine Abbildung aus dem Ergebnisraum $\mathcal{X}$ einer Stichprobe nach $\{0, 1\}$.
- b) Formal hat ein Test die Form $\phi : \mathcal{X} \rightarrow \{0, 1\}, x \mapsto \phi(x)$.
- c) Der Testwert $\phi(x) = 0$ repräsentiert den Vorgang des Nichtablehnens der Nullhypothese.
- X** Weil die Stichprobe ein Zufallsvektor ist, ist ein Test keine Zufallsvariable.

22. Welche Aussage zur Testgütefunktion eines Einstichproben-T-Tests ϕ trifft zu?

- a) Die Testgütefunktion ist definiert als Wahrscheinlichkeit, dass $\phi(X) = 0$.
- X** Die Testgütefunktion kann zur Kontrolle des Testumfangs und zur Berechnung der Power des Tests genutzt werden.
- b) Die Testgütefunktion hängt nicht von den wahren, aber unbekannten Modellparametern ab.
- c) Zur Bestimmung der Testgütefunktion sind die Verteilung der Teststatistik und der kritische Bereich des Tests unerheblich.

23. Welche Aussage zur praktischen Durchführung eines T-Tests trifft **nicht** zu?

- a) Man nimmt an, dass ein vorliegender Datensatz eine Realisation eines Normalverteilungsmodells $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ ist.
- b) Man möchte zum Beispiel entscheiden, ob für ein μ_0 eher $H_0 : \mu = \mu_0$ oder $H_1 : \mu \neq \mu_0$ gilt.
- X** Anhand des Stichprobenmittels und der Stichprobenstandardabweichung berechnet man das Signifikanzniveau α_0 des T-Tests.
- c) Die Theorie des T-Tests garantiert, dass man bei einem Signifikanzniveau von α_0 höchstens in $\alpha_0 \cdot 100$ von 100 Testdurchführungen die Nullhypothese fälschlicherweise ablehnt.

24. Welche Aussage zur Powerfunktion eines zweiseitigen Einstichproben-T-Tests trifft **nicht** zu?

- a) Der Wert der Powerfunktion hängt vom Stichprobenumfang ab.
- b) Der Wert der Powerfunktion hängt von den wahren, aber unbekannten Modellparametern ab.
- c) Der Wert der Powerfunktion hängt vom kritischen Wert des Tests ab.
- X** Der Wert der Powerfunktion hängt nicht vom Signifikanzniveau des Tests ab.

25. Welche Aussage trifft zu? Das Behrens-Fisher Problem bezeichnet den Zweistichproben-T-Test ...

- a) ... bei unabhängigen Stichproben unter Annahme identischer Varianzen.
- X** ... bei unabhängigen Stichproben ohne weitere Varianzannahmen.
- b) ... bei abhängigen Stichproben und einfacher Nullhypothese.
- c) ... bei abhängigen Stichproben und zusammengesetzter Nullhypothese.

26. Es sollen inferentiell die wahren, aber unbekannten Erwartungswertparameter dreier unabhängiger normalverteilter Stichproben verglichen werden. Welches der folgenden Inferenzverfahren sollte sinnvollerweise zur Anwendung kommen?
- a) Ein Einstichproben-T-Test.
 - b) Ein Zweistichproben-T-Test.
 - X** Eine einfaktorielle Varianzanalyse.
 - c) Eine zweifaktorielle Varianzanalyse
27. Es sollen inferentiell die Haupt- und Interaktionseffekte in einem zweifaktoriellen Versuchsplan mit crossed design bei normalverteilten Daten getestet werden. Welches der folgenden Inferenzverfahren sollte sinnvollerweise zur Anwendung kommen?
- a) Ein Einstichproben-T-Test.
 - b) Ein Zweistichproben-T-Test.
 - c) Eine einfaktorielle Varianzanalyse.
 - X** Eine zweifaktorielle Varianzanalyse
28. Für $i = 1, \dots, p$ und $j = 1, \dots, m$ bezeichne X_{ij} die j te Stichprobenvariable der i ten Stichprobengruppe eines einfaktoriellen Varianzanalysemodells. Weiterhin seien $\bar{X}$ der Gesamtmittelwert, $\bar{X}_i$ das i te Stichprobenmittel, SQB die Between Group Sum of Squares und SQW die Within Group Sum of Squares. Welche Aussage trifft **nicht** zu?
- a) Es gilt $\text{SQB} := \sum_{i=1}^p m(\bar{X}_i - \bar{X})^2$.
 - b) SQB quantifiziert die Variabilität der Gruppenmittelwerte um den Gesamtmittelwert.
 - c) Es gilt $\text{SQW} := \sum_{i=1}^p \sum_{j=1}^m (X_{ij} - \bar{X}_i)^2$.
 - X** SQW quantifiziert die Variabilität der Gruppenmittelwerte um den Gesamtmittelwert.
29. Es sei ein 2×2 faktorieller Versuchsplan mit crossed design gegeben, bei dem Posttherapie-BDI-Scores nach einer klassischen Psychotherapie und einer Online-Psychotherapie für zwei Altersklassen mithilfe einer 2×2 ANOVA analysiert werden sollen. Welche Aussage bezüglich der Nomenklatur dieser zweifaktoriellen Varianzanalyse trifft dann **nicht** zu?
- a) Der Faktor Psychotherapie hat die Level (1) Klassisch und (2) Online.
 - b) Der Faktor Altersklasse hat zwei Level.
 - X** Der Faktor Posttherapie-BDI-Scores hat die Level (1) Psychotherapie und (2) Altersklasse.
 - c) Eine Interaktion der Faktoren Psychotherapie und Altersklasse entspricht verschiedenen Therapieeffekten in den verschiedenen Altersklassen.
30. Welche Aussage zum Modell der zweifaktoriellen Varianzanalyse in Effektdarstellung trifft **nicht** zu?
- a) Das Modell besitzt Haupteffekt- und Interaktionsparameter.
 - b) Das Modell definiert Bedingungen für die Summen seiner Parameter.
 - X** Die F-Teststatistiken zum Testen der Haupteffektsparemeteryypothesen sind identisch.
 - c) Das Modell erlaubt die inferentielle Auswertung von Hypothesen bezüglich seiner Haupteffekts- und Interaktionsparameter.

Inferenzstatistik 2021 | Klausurergebnisbericht

Die Klausur zum Modul B2 Inferenzstatistik im Sommersemester 2021 fand am 12.7.2021 von 8.00 - 9.00 Uhr in Hörsaal 6, Gebäude 44 der OVGU mit 50 Teilnehmer:innen statt. Die Klausur bestand aus 30 Multiple Choice Aufgaben mit jeweils vier Antwortmöglichkeiten und jeweils genau einer richtigen Antwort. Die Klausurlösung ist diesem Bericht beigelegt.

Bewertungschema

Die Aufteilung der zugelassenen Noten auf die erreichten Prozentpunkte wurde anhand der empfohlenen Notenskala der Freien Universität Berlin vorgenommen. Diese trifft folgende Zuordnung der erreichten Prozentpunkte zu den zugelassenen Noten anhand von linkshalboffenen Prozentpunktintervallen.

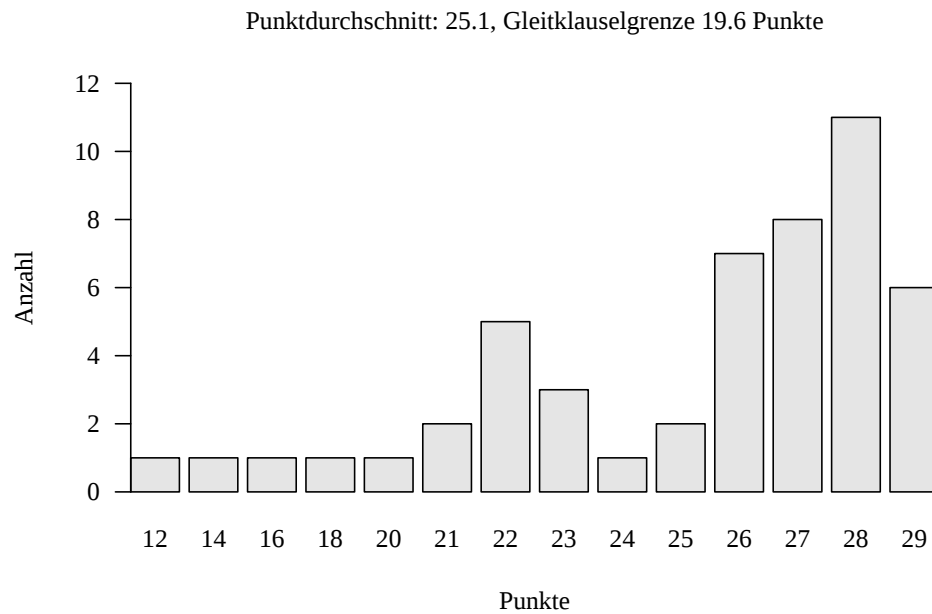
<	>=	Note
100.0	94.9	1.0
94.9	89.5	1.3
89.5	84.3	1.7
84.3	79.0	2.0
79.0	73.7	2.3
73.7	68.2	2.7
68.2	63.1	3.0
63.1	57.9	3.3
57.9	52.6	3.7
52.6	50.0	4.0
50.0	0.0	5.0

Es ergibt sich folgendes Punktenotenschema, wobei < 15 Punkte mit 5.0 bewertet wurden.

Punkte	Prozent	Note
30	100.00	1.0
29	96.67	1.0
28	93.33	1.3
27	90.00	1.3
26	86.67	1.7
25	83.33	2.0
24	80.00	2.0
23	76.67	2.3
22	73.33	2.7
21	70.00	2.7
20	66.67	3.0
19	63.33	3.0
18	60.00	3.3
17	56.67	3.7
16	53.33	3.7
15	50.00	4.0

Ergebnisse

Die nachfolgende Abbildung zeigt die absolute Häufigkeitsverteilung der erzielten Punkte.



Die nachfolgende Abbildung zeigt die absolute Häufigkeitsverteilung der erreichten Noten.

