

Probeklausur STATISTIK II

Name, Vorname: _____

Matrikel-Nr.: _____

Die Klausur enthält zwei Typen von Aufgaben:

T e i l A besteht aus Fragen mit mehreren vorgegebenen Antwortvorschlägen, von denen mindestens eine Antwort richtig ist und von denen mehrere Antworten richtig sein können. Kreuzen Sie alle richtigen Antworten an. Sind alle Kreuze richtig, erhalten Sie für die Aufgabe 2 Punkte. Jede Abweichung ergibt 1 Punkt Abzug. Es werden keine negativen Punktezahlen vergeben, Sie erhalten also für jede Aufgabe mindestens 0 Punkte. Wenn Sie keine Antwort ankreuzen, gilt die Aufgabe als nicht bearbeitet und Sie erhalten 0 Punkte.

T e i l B enthält ausführlich zu lösende Aufgaben. Nur mit der Darstellung der einzelnen Rechenschritte kann die volle Punktzahl erreicht werden.

Zulässige Hilfsmittel: Nicht programmierbarer Taschenrechner, Lehrbuch von Schira, eine handschriftlich von Ihnen selbst beschriebene Seite ("Spickzettel", kann auf beiden Seiten beschrieben sein), Formelblatt zur Intervallschätzung.

Teil A enthält 12 Aufgaben und Teil B enthält 4 Aufgaben. Bitte überprüfen Sie die Vollständigkeit Ihres Exemplars.

Die maximal zu erreichende Punktzahl ist 60, davon können maximal 24 Punkte in Teil A und maximal 36 Punkte in Teil B erreicht werden. Mit mindestens 24 erreichten Punkten bestehen Sie die Klausur.

Auswertung - Teil A

Aufgabe	1	2	3	4	5	6	7	8	9	10	11	12
Erreichte Punktzahl												

Auswertung - Teil B

Aufgabe	1	2	3	4
Erreichte Punktzahl				

Erreichte Gesamtpunktzahl

Teil A (24 Punkte)

1. Bei den Metadaten handelt es sich um
- A) Merkmalsausprägungen eines einzelnen Merkmalsträgers
 - B) Zusammenfassung von Mikrodaten über mehrere diskrete Merkmale
 - C) Zusammenfassung von Makrodaten über mehrere diskrete Merkmale
 - D) Daten, die Informationen über andere Daten enthalten
 - E) Die Antworten A) bis D) sind falsch.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
			X	

2. Ausgelöste Statistik
- A) wird von jemandem betrieben, der an den Ergebnissen interessiert ist
 - B) wird von jemandem betrieben, der kein Interesse an den Ergebnissen hat
 - C) ist Statistik, die eine staatliche Institution über ihre eigene Arbeit erhebt, aufbereitet und veröffentlicht
 - D) wird von den statistischen Ämtern erhoben, aufbereitet und veröffentlicht
 - E) Die Antworten A) bis D) sind falsch.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
	X		X	

3. Ein Jäger trifft sein Ziel mit einer Wahrscheinlichkeit von 45%. Mit welcher Wahrscheinlichkeit erzielt er bei 18 Schüssen mehr als 3 Treffer?

- A) 0,0198
- B) 0,01198
- C) 0,98802
- D) 0,45
- E) Die Antworten A) bis D) sind falsch.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
		X		

4. Eine Telefonauskunft wird durchschnittlich 5 mal pro Stunde angewählt. Mit welcher Wahrscheinlichkeit erhält die Auskunft in einer Stunde 10 und mehr Anrufe?

- A) 0,0137
- B) 0,9863
- C) 0,9945
- D) 0,0055
- E) Die Antworten A) bis D) sind falsch.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
				X

5. Der Korrekturfaktor

- A) bewirkt bei relativ kleinen Stichproben, dass die Varianz beim Ziehen ohne Zurücklegen kleiner ist als beim Modell mit Zurücklegen
- B) kann entfallen, wenn die Stichprobe aus einer sehr großen Grundgesamtheit gezogen wird
- C) strebt gegen Null, wenn N größer wird
- D) wird zu Eins bei einer Totalerhebung
- E) Die Antworten A) bis D) sind falsch.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
X	X			

6. Die unten stehende Tabelle enthält Informationen über die Häufigkeitsverteilung der Merkmale psychische Lage und Geschlecht bei 106 untersuchten Personen. Prüfen Sie mit einem passenden Test, ob die beiden Merkmale voneinander unabhängig sind (Signifikanzniveau = 0,05).

Geschlecht	Psychische Lage		gesamt
	labil	stabil	
weiblich	34	10	44
männlich	25	37	62
gesamt	59	47	106

- A) H_0 wird verworfen
- B) H_0 wird angenommen
- C) Die Teststatistik beträgt 15,4
- D) Die Teststatistik beträgt 14,2
- E) Die Antworten A) bis D) sind falsch.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
X			X	

7. Für den Stichprobenmittelwert \bar{x} sagt das schwache Gesetz der grossen Zahlen aus, dass

- A) $P(|\bar{x} - \mu| < \epsilon) \rightarrow 1$
- B) $P(|\bar{x} - \mu| > \epsilon) \rightarrow 1$
- C) $P(|\bar{x} - \mu| < \epsilon) \rightarrow \theta$
- D) $P(|\bar{x} - \hat{\mu}| < \epsilon) \rightarrow 0$
- E) Die Antworten A) bis D) sind falsch.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
X				

8. Die Verteilungsfunktion der standardisierten Summe unabhängiger Zufallsvariablen konvergiert mit zunehmender Anzahl n gegen:

- A) $N(\mu, \sigma)$
- B) $N(\mu, \sigma^2)$
- C) $N(0, 1)$
- D) $N(1, 0)$
- E) Die Antworten A) bis D) sind falsch.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
		X		

9. Bei der Schätzung von $y = \alpha + \beta x + u$ mit Hilfe der KQ-Methode wird

- A) der Abstand zwischen x_i und y_i minimiert
- B) u minimiert
- C) die Summe der e_i^2 minimiert
- D) die Likelihood Funktion $f(y | x)$ minimiert.
- E) Die Antworten A) bis D) sind falsch.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
		X		

10. Ein erwartungstreuer Schätzer $\hat{\beta}$ hat die Eigenschaft, dass

- A) β für kleine Stichprobenumfänge $n < 30$ t-verteilt ist
- B) $E(\hat{\beta}) = \beta$
- C) β konsistent ist
- D) $\lim_{n \rightarrow \infty} \hat{\beta} = \beta$
- E) Die Antworten A) bis D) sind falsch.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
	X			

11. Bei einem p-Wert von 0,07 wird die Nullhypothese

- A) zum 5% Signifikanzniveau abgelehnt
- B) zum 12% Signifikanzniveau angenommen
- C) der p-Wert muss immer im Verhältnis zum t-Wert gesehen werden
- D) zum 10% Signifikanzniveau abgelehnt
- E) Die Antworten A) bis D) sind falsch.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
			X	

12. In einem Internet Forum wird behauptet, Tagesrenditen mit einer t-Verteilung anzunähern sei gefährlich, da im Falle von 2 Freiheitsgraden die Varianz der Renditen nicht mehr definiert sei. Diese Aussage ist

- A) richtig
- B) unprofessionell, da heute niemand mehr die t-Verteilung benutzt
- C) falsch, da die t-Verteilung wie die Normalverteilung symmetrisch ist
- D) unnötig, da es keinen Grund gibt, warum Renditen exakt t-verteilt sein sollten
- E) Die Antworten A) bis D) sind falsch.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
X			X	

Teil B (36 Punkte)

1. Die folgende Tabelle enthält die in der Vorrunde der EM 2008 erzielten Punkte aller teilnehmenden Fußballmannschaften:

Gruppe A	Gruppe B	Gruppe C	Gruppe D
6	9	9	9
6	6	4	6
3	1	2	3
3	1	1	0

Quelle: <http://de.euro2008.uefa.com>

- a) Prüfen Sie, ob die Gruppen A bis D den gleichen Mittelwert haben (Signifikanzniveau = 0,05).
(4 Punkte)
- b) Betrachten Sie nun die Gruppen A und C. Kann die Hypothese bestätigt werden, Gruppe A sei homogener in der Punktverteilung als Gruppe C? (Signifikanzniveau = 0,05)
(4 Punkte)
- c) Angenommen, die variable Punktzahl sei in der Grundgesamtheit normalverteilt mit einem Mittelwert von 4 und einer Standardabweichung von 1 Punkt. Wie hoch ist der p-Wert, wenn alle 16 Beobachtungen zur Berechnung der Teststatistik herangezogen werden?
(1 Punkt)
-
- (9 Punkte)

2.

- a) Diskutieren Sie folgendes Statement des Leiters einer Risikomanagement Abteilung: "Für unsere tägliche Aktienrenditen können wir problemlos eine Normalverteilung annehmen, da wir sehr lange Zeitreihen haben und die Renditen dadurch in Verbindung mit dem Grenzwertsatz normalverteilt sind."
(1 Punkt)
- b) Ein neu auf den Markt gekommenes Mineralwasser wird zur Analyse in ein Chemielabor eingeschickt. Da die Qualität von mehreren Umweltbedingungen abhängt und nicht konstant ist, werden dem Labor 20 Stichproben zugeschickt. Dabei wird im Durchschnitt ein Natriumgehalt von 8,6 mg/l bei einer Standardabweichung von 2,4 mg/l festgestellt. Berechnen Sie das 95% Konfidenzintervall für den Erwartungswert der Grundgesamtheit.
(2 Punkte)

- c) Bei einer grösseren Stichprobe vom Umfang $n = 60$ wird ein Natriumgehalt von 8,1 mg/l bei einer Standardabweichung von 2,2 mg/l festgestellt. Berechnen Sie das Konfidenzintervall, in dem der Erwartungswert mit 99% Wahrscheinlichkeit höchstens liegt. Warum ist das Intervall kleiner geworden?

(2 Punkte)

- d) Der amerikanische Ökonom Steven Levitt ist durch seine ungewöhnlichen Fragestellungen bekannt geworden. Ein Artikel aus dem Jahr 2000 ("An Economic Analysis of a Drug-Selling Gang's Finances." *Quarterly Journal of Economics*, 2000, 115(3), pp. 755-789) untersucht die Verteilung der Bezahlung unter Gang Mitgliedern. In einer Stichprobe vom Umfang 15 sei festgestellt worden, dass nur 3 der 15 zufällig ausgewählten Mitgliedern eine dauerhafte Bezahlung über einem herkömmlichen Arbeitsmarktlohn bei ähnlichem Risiko haben. Schätzen Sie das 95% Konfidenzintervall für den Anteilswert der gut verdienenden Gangmitglieder für die Grundgesamtheit aller Gangs.

(3 Punkte)

- e) Eine größere Stichprobe vom Umfang 200 ergäbe, dass 46 Mitglieder der gutverdienenden Schicht angehören. Schätzen Sie das 90% Konfidenzintervall.

(1 Punkte)

(9 Punkte)

3. Am 31.12.2007 hatte Deutschland nach vorläufigen Ergebnissen rund 82.218.000 Einwohner. Darunter waren rund 40.274.000 männlich (Quelle: Statistisches Bundesamt). Wie hoch ist die Wahrscheinlichkeit, dass

- a) unter 100 zufällig ausgewählten Personen höchstens 50 Frauen sind

(2 Punkte)

- b) mindestens 40 und höchstens 60 Frauen sind?

(2 Punkte)

- c) unter 8 zufällig ausgewählten Personen mehr als die Hälfte Frauen sind?

(2 Punkte)

- d) Wieviel Prozent der Wahrscheinlichkeitsmasse der männlichen Einwohner liegt in einem Zwei-Sigma-Intervall $\mu \pm 2\sigma$, wenn 8 Personen zufällig ausgewählt werden?

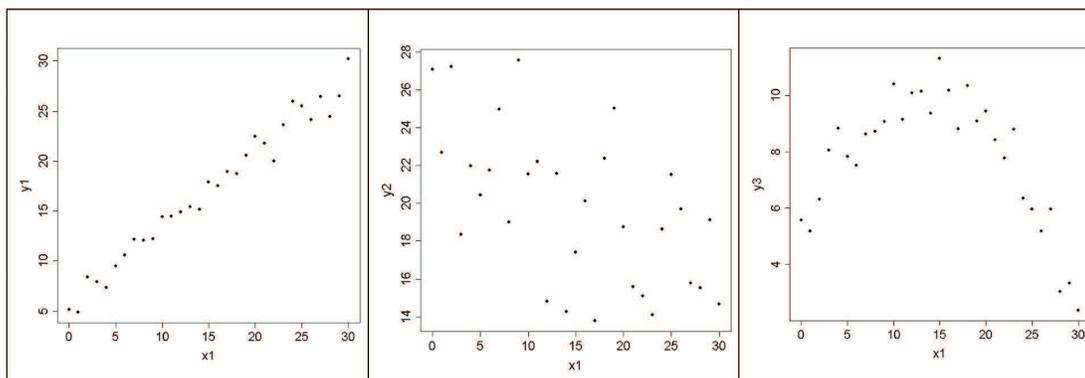
(3 Punkte)

(9 Punkte)

4.

- a) Zeichnen Sie den ungefähren Verlauf der Regressionsgeraden in die folgenden drei Schaubilder ein. Geben Sie auch ihre subjektive Einschätzung zur Konstanten, zum Steigungsparameter und zum R^2 dieser Schätzung an

(2 Punkte)



- b) Welche Probleme treten im dritten Schaubild auf? Kommentieren Sie kurz, ob und warum in diesem Fall die Anwendung einer linearen Regression angebracht ist oder nicht.

(2 Punkte)

- c) Während eines Praktikums untersucht ein Student für den amerikanischen Markt, welche makroökonomischen Variablen einen Einfluss auf die Wohnimmobilienpreise haben. Dabei schätzt er folgende Funktion mit Hilfe der KQ-Methode:

$$RP_t = \alpha + \beta_1 rgdp_t + \beta_2 constr_t + \beta_3 long_t + \epsilon_t \quad t = 1 \dots T$$

RP = reale Hauspreise

rgdp = reales BIP

constr = Kosten für Material und Arbeit im Baugewerbe

long = 10 Jahres Staatsanleihen

Schreiben Sie die Gleichung in Matrixnotation in Abhängigkeit von t (RP_t) sowie in vollständiger Matrixnotation unabhängig von t (\mathbf{RP}).

(1 Punkt)

- d) Da der Student im letzten Semester Hausaufgabenwettbewerbsgewinner war, programmiert er den KQ-Schätzer $\hat{\beta} = (X'X)^{-1}X'Y$ natürlich selbst und erhält:

$$(X'X)^{-1} = \begin{pmatrix} 5.82 & -3.22 & -5.05 & -7.96 \\ -0.01 & 5.73 & 1.77 & 3.55 \\ -0.05 & 1.77 & 4.73 & -7.48 \\ -0.01 & 3.55 & -7.48 & 7.96 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} -8,35242 \\ 0,10117 \\ -0,0083125 \\ -0,067195 \end{pmatrix} \quad \hat{\beta} = \begin{pmatrix} -48.36 \\ ??? \\ 1.06 \\ -0.03 \end{pmatrix}$$

Berechnen Sie den fehlenden Parameter von *rgdp* und interpretieren Sie kurz die ökonomische Aussage aller 4 Parameter.

(4 Punkte)

(9 Punkte)