

Prüfungsklausur STATISTIK II

Name, Vorname: _____

Matrikel-Nr.: _____

Die Klausur enthält zwei Typen von Aufgaben:

Teil 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.

Teil 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. Folgende Aussage ist falsch:

- A) Die empirische Varianz stellt einen erwartungstreuen Schätzer für die Varianz der Grundgesamtheit dar, wenn μ bekannt ist.
- B) Die empirische Varianz stellt einen erwartungstreuen Schätzer für die Varianz der Grundgesamtheit dar, wenn die Grundgesamtheit normalverteilt ist.
- C) Die Punktschätzung für die Varianz mit der Momentenmethode ist erwartungstreu.
- D) Die Punktschätzung für die Varianz mit der Momentenmethode ist konsistent.
- 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		

2. Eine Lieferung soll angenommen werden, wenn mindestens 90 Prozent der Produkte einwandfrei sind. Wenn Sie sich entscheiden, die Lieferung zu akzeptieren und es sich später herausstellt, dass die Anforderung nicht erfüllt war,

- A) haben Sie einen α -Fehler begangen.
- B) haben Sie einen β -Fehler begangen.
- C) haben Sie eine richtige Nullhypothese verworfen.
- D) haben Sie eine falsche Nullhypothese verworfen.
- E) Die Antworten A) bis D) sind falsch.

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

3. Ein p-Wert von 0,035 bedeutet, dass

- A) der Koeffizient in einem Koeffiziententest zum 5%-Niveau signifikant von Null verschieden ist.
- B) die Nullhypothese zum 10%-Signifikanzniveau angenommen wird.
- C) die Nullhypothese zum 5%-Signifikanzniveau abgelehnt wird.
- D) die Nullhypothese zum 1%-Signifikanzniveau abgelehnt wird.
- 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		

4. Zensus 2011

- A) ist eine traditionelle Volkszählung, bei der alle Einwohnerinnen und Einwohner befragt werden.
- B) ist eine registergestützte Volkszählung, bei der vor allem Melderegister und Register der Bundesagentur für Arbeit genutzt werden.
- C) ist eine Erhebung, die ermittelt, wie viele Menschen in einem Land leben, wie sie wohnen und arbeiten.
- D) stellt eine Mischung aus Registerauswertung und Wohnungszählung dar.
- 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	X	

5. Mehr als 90 Prozent aller Haushalte in Deutschland kaufen mindestens einmal im Jahr Bio-Produkte. Wie hoch ist die Wahrscheinlichkeit, dass unter 10 befragten Haushalten höchstens 5 Haushalte Bio-Produkte kaufen?

- A) 0,5.
B) 0,1563.
C) 0,0016.
D) 0,8437.
E) Die Antworten A) bis D) sind falsch.

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

6. Folgende Aussage trifft zu:

- A) Mikrozensus wird in Deutschland als eine repräsentative Klumpenstichprobe erhoben.
B) Mikrozensus wird in Deutschland als eine Quotenstichprobe erhoben.
C) Am Mikrozensus nehmen insgesamt 820 000 Haushalte mit 1 800 500 Personen teil.
D) Mikrozensus ist eine jährlich durchgeführte Befragung von einem Prozent aller Haushalte in Deutschland.
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. Die Maximum-Likelihood Schätzer für μ und σ einer Normalverteilung sind

- A) konsistent.
- B) nicht asymptotisch erwartungstreu.
- C) nicht suffizient.
- D) die Momentenschätzer.
- 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	

8. Welche der folgenden Aussagen zur Intervallschätzung treffen zu?

- A) Je größer das Intervall, desto größer ist die Genauigkeit bzw. desto kleiner ist der maximale Schätzfehler.
- B) Je kleiner der maximale Schätzfehler, desto länger das Schätzintervall.
- C) Je länger das Intervall, desto kleiner die Vertrauenswahrscheinlichkeit.
- D) Je kleiner das Intervall, desto größer ist die Genauigkeit bzw. desto kleiner ist der maximale Schätzfehler.
- 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. Ein Moving-Average-Prozess erster Ordnung bzw. MA(1)-Prozess $Y_t = \alpha_0 + \epsilon_t + \alpha_1 \epsilon_{t-1}$
- A) kann in Abhängigkeit von der absoluten Größe von α_1 entweder stationär oder nicht-stationär sein.
 - B) hat einen Erwartungswert von Null oder $E(\epsilon_t)$.
 - C) hat einen konstanten Erwartungswert, eine konstante Varianz und die Kovarianz hängt lediglich von $j = 1$ ab.
 - D) ist grundsätzlich streng stationär.
 - 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. Bei einem autoregressiven Prozess zweiter Ordnung bzw. AR(2)-Prozess $Y_t = \beta_0 + 0,5Y_{t-1} - 0,7Y_{t-2} + \epsilon_t$
- A) bricht die Autokorrelationsfunktion nach einem Lag von 2 ab.
 - B) stirbt die partielle Autokorrelationsfunktion langsam aus.
 - C) zeigt die Autokorrelationsfunktion ein abnehmendes oszillierendes Verhalten auf.
 - D) bricht die partielle Autokorrelationsfunktion bei einem Lag von 2 ab.
 - 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	X	

11. Beim statistischen Testen

- A) gibt der p-Wert das Signifikanzniveau an, zu dem die Nullhypothese gerade noch verworfen werden kann.
- B) wird die Nullhypothese verworfen, wenn der p-Wert größer ist als das Signifikanzniveau.
- C) ordnet die Gütefunktion den Werten aus der Alternativhypothese die Wahrscheinlichkeiten zu, mit denen die Nullhypothese bei Vorliegen der betreffenden Werte verworfen wird.
- D) liegt ein Fehler 2. Art vor, wenn die Alternativhypothese nicht angenommen wird, obwohl sie richtig.
- 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	

12. Aus den Klausuren in Statistik I und II wird jeweils eine Stichprobe vom Umfang 20 gezogen und folgendes festgestellt:

$$\bar{x}_1 = \bar{x}_2 = 48 \text{ Punkte mit } s_1 = 3 \text{ und } s_2 = 5.$$

Es soll nun geprüft werden, ob die Punktzahlen in Statistik I homogener sind, als die in Statistik II. Die Testgröße des hierbei anzuwendenden Tests beträgt

- A) 1,56.
- B) 1,67.
- C) 2,78.
- D) 0,6.
- 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

Anmerkung: Hier ist C oder E richtig, je nachdem welche in der Aufgabenstellung gegebene Standardabweichung man in den Zähler und in den Nenner der Testgröße eingefügt hat.

Teil B (36 Punkte)

1. Im folgenden Regressionsmodell werden anhand einer Stichprobe mit 81 Beobachtungen die Faktoren untersucht, die die Höhe der LehrerInnenlöhne in der Schweiz am Ende des 19. Jahrhunderts beeinflussen könnten. Der Lohn (in Franken) wird dabei durch zwei exogene Variablen erklärt: die Anzahl Dienstjahre der Lehrpersonen (DJAHRE) und die Gemeindegröße (EINWOHNER). Die Verwendung der ersten Variablen kann durch eine ökonomische Produktivitätstheorie begründet werden: je erfahrener der Arbeitnehmer, desto produktiver ist er. Bei der zweiten exogenen Variablen können folgende Zusammenhänge vermutet werden: einerseits haben größere Gemeinden mehr Steuereinnahmen und können es sich daher leisten, höhere Löhne zu zahlen. Andererseits haben größere Gemeinden eher Wirtschaftszweige, die gebildete Wirtschaftskräfte benötigen und daher ein höheres Interesse an Bildung haben.

Gegeben sei folgender TSP-Output für die OLS-Schätzung des Modells für alle 81 Beobachtungen:

```
Method of estimation = Ordinary Least Squares

Dependent variable: LOHN
Current sample: 1 to 81
Number of observations: 81

      Mean of dep. var. = 1363.21      LM het. test = 3.18541 [.074]
      Std. dev. of dep. var. = 445.956  Durbin-Watson = 1.14816 [<.000]
Sum of squared residuals = .814927E+07 Jarque-Bera test = 4.29937 [.117]
      Variance of residuals = 104478.   Ramsey's RESET2 = 1.45683 [.231]
Std. error of regression = 323.230     F (zero slopes) = 37.1411 [.000]
      R-squared = .487793               Schwarz B.I.C. = 588.045
      Adjusted R-squared = .474660     Log likelihood = -581.453

Variable      Estimated      Standard
Coefficient   Error
C              949.057        68.7599      13.8025      [.000]
DJAHRE        14.6949         3.17534      4.62781      [.000]
EINWOHNER     .021447         .261082E-02  8.21457      [.000]
```

Hinweis: der Standardfehler von $\beta_2 = 0.261082E-2$ entspricht dem Wert 0.00261082.

- a) Interpretieren Sie den TSP-Output. Lässt sich das theoretische Modell bestätigen? Gehen Sie bei der Interpretation auf die Bedeutung, Plausibilität und Signifikanz der Regressionskoeffizienten sowie auf die Güte des Gesamtmodells (R^2 , F-Wert) ein.

(3 Punkte)

- b) Testen Sie den Koeffizienten β_1 gegen Null. Geben Sie dabei ausführlich alle Rechenschritte der Testprozedur an (Hypothesenaufstellung, Berechnung der Teststatistik und des kritischen Wertes sowie die Testentscheidung). Signifikanzniveau $\alpha = 0,05$.

(3 Punkte)

- c) Gegeben sei die $(X'X)^{-1}$ -Matrix und der $X'y$ -Vektor aller weiblichen Lehrerinnen:

$$(X'X)^{-1} = \begin{pmatrix} 0.097801 & -0.0051932 & -0.00000162653 \\ -0.0051932 & 0.00056954 & 6.93443D - 09 \\ -0.0000016 & 6.93443D - 09 & 1.27389D - 10 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 34590 \\ 336225 \\ 577969000 \end{pmatrix}$$

Schätzen Sie den β_0 -Koeffizienten für alle untersuchten Frauen. Was könnte der Unterschied zwischen der frauenspezifischen Konstanten und der Konstanten, die für den ganzen Datensatz ermittelt wurde, für die Modellspezifikation bedeuten?

(2 Punkte)

- d) Welche Annahmen der linearen Regression werden für die Spezifikation der Störgröße getroffen? Lässt sich das Problem der Autokorrelation beim Lohnmodell feststellen? (Hinweis: benutzen Sie für die Antwort den TSP-Output).

(1 Punkt)

(9 Punkte)

2.

- a) Ein Intelligenztest bei 15 VWL-Studenten ergab einen durchschnittlichen Intelligenzquotienten von 108 bei einer Standardabweichung von 10. Das Merkmal Intelligenzquotient ist in der Grundgesamtheit normalverteilt. Berechnen Sie das 95%-Konfidenzintervall für den IQ-Mittelwert aller VWL-Studenten.

(2 Punkte)

- b) Bei 120 von insgesamt 200 Statistik-Studenten im Grundstudium lag der durchschnittliche IQ bei 107 mit einer Standardabweichung von 15. Welchen Höchstwert für den IQ aller Statistik-Studenten würde man mit einer Konfidenzwahrscheinlichkeit von 99% erwarten?

(2 Punkte)

- c) Was ändert sich an der Lösung zu der Teilaufgabe b), wenn nicht 120, sondern alle 200 Statistik-Studenten untersucht werden?

(1 Punkt)

- d) Wie groß müsste der Stichprobenumfang bei $\sigma_X = 15$ gewählt werden, damit man ein 95%- Konfidenzintervall der Länge 6 erhält?

(2 Punkte)

- e) Der Frauenanteil unter 20 zufällig ausgewählten Studierenden lag bei 45%. Schätzen Sie das 95%-Konfidenzintervall für den Frauenanteil aller Studierenden.

(2 Punkte)

(9 Punkte)

3. In der finanzwirtschaftlichen Diskussion um die beste Anlagekategorie wird argumentiert, dass langfristig Aktienanlagen einer Investition in Anleihen überlegen ist. Als Anlageberater möchten Sie in einem Verkaufsgespräch mit einem wohlhabenden Kunden dieses Argument statistisch widerlegen. Hierzu ziehen Sie die Jahresrenditen von Aktien (DAX) und Anleihen (REXP) in den Jahren 1994 bis 2003 aus der unten stehenden Tabelle heran.

Jahr	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Aktien (in %)	-7	7	28	47	18	39	-8	-20	-44	37
Anleihen (in %)	-3	17	8	7	11	-2	7	6	9	4

- a) Kann man bei einer Irrtumswahrscheinlichkeit von $\alpha = 0,05$ davon ausgehen, dass langfristig mit Aktien gegenüber Anleihen eine höhere Durchschnittsrendite zu erzielen ist? (Berechnen Sie die Durchschnittsrenditen als arithmetisches Mittel.)

(4 Punkte)

- b) Des weiteren empfehlen Sie Ihrem Kunden aufgrund der geringen Korrelation von $r = -0,16$ ein Portfolio aus beiden Anlagekategorien zu bilden, da bei einem geringen bzw. negativen Korrelationskoeffizient das Risiko im Portfolio gesenkt werden kann. Überprüfen Sie, ob der Korrelationskoeffizient signifikant von Null verschieden ist. ($\alpha = 0,10$)

(1 Punkt)

- c) Im betrachteten Zeitraum war in den ersten 5 Jahren in 20% der Monate eine negative Rendite am deutschen und in 40% der Monate eine negative Rendite am französischen Aktienmarkt zu beobachten. Sind die Anteile von Monaten mit negativen Aktienrenditen zwischen den beiden Aktienmärkten signifikant verschieden? ($\alpha = 0,01$)

(1 Punkt)

- d) In einem letzten Schritt möchten Sie für Ihren Kunden überprüfen, ob die Renditebewegung der beiden Anlagekategorien unabhängig voneinander sind. Dabei unterscheiden Sie zwischen sinkenden (-), annähernd gleichbleibenden (0) und steigenden (+) Aktien (X)- und Anleihenrenditen (Y). Die Häufigkeiten der beobachteten Renditebewegungen aus 100 zufällig ausgewählten Monaten sind der nachstehenden Tabelle zu entnehmen.

y_j	-	0	+	$x_{i.}$
x_i				
-	5	15	20	40
0	10	10	10	30
+	5	5	20	30
$y_{.j}$	20	30	50	100

Kann man davon ausgehen, dass die Renditebewegungen unabhängig voneinander sind ($\alpha = 0,10$)?

(3 Punkte)

(9 Punkte)

4.

- a) Bestimmen Sie mit der Maximum-Likelihood-Methode einen Schätzer für den Parameter λ einer exponentialverteilten Zufallsgröße. Schätzen Sie dabei den Parameter λ aus einer gegebenen Stichprobe von $n = 100$ als exponentialverteilt vorausgesetzten Variablen (Hinweis: $\sum_{i=1}^n x_i = 99$).

(4 Punkte)

- b) Vergleichen Sie die erhaltene Schätzung mit einer Schätzung nach der Momentenmethode.

(1 Punkt)

- c) Zur Schätzung eines unbekannten Anteilswertes p ist die Punktschätzung durch den Stichprobenanteils wert sehr gut geeignet. Zeigen Sie unter Annahme der Unabhängigkeit der Einzelstichproben die Eigenschaften der Varianz des Schätzers auf.

(2 Punkte)

- d) Für ein Merkmal X ist der Erwartungswert μ bekannt, während die Varianz unbekannt ist. Die n Zufallsvariablen seien voneinander unabhängig und identisch verteilt. Folgende Schätzfunktionen werden für die Varianz vorgeschlagen:

$$g_1 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$
$$g_2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)^2$$

Überprüfen Sie die beiden Schätzfunktionen auf Erwartungstreue und berechnen Sie gegebenenfalls den Bias. Zeigen Sie, welche Schätzfunktion besser ist.

(2 Punkte)

(9 Punkte)