

Nachholklausur 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. Folgende Aussage trifft zu:

- A) Die Log-Normalverteilung ist eine Verteilung, deren Mittelwert und Standardabweichung normalverteilt sind.
- B) Die Zufallsvariablen einer Log-Normalverteilung sind normalverteilt, wenn sie vorher logarithmiert werden.
- C) e hoch die Zufallsvariablen einer Log-Normalverteilung sind normalverteilt.
- D) Die Log-Normalverteilung ist eine Verteilung, deren Mittelwert und Standardabweichung log-normalverteilt sind.
- 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. Die Gamma-Funktion $\Gamma(x) = \int_{z=0}^{\infty} z^{x-1} e^{-z} dz$

- A) ist die stetige Version der Fakultäten-Funktion.
- B) ist die Gamma-Verteilung.
- C) kommt in der Binomialverteilung zum Einsatz.
- D) kommt in der Gamma-Verteilung zum Einsatz.
- 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. Wenn bei einem $AR(1)$ Prozess $y_t = \mu + \alpha y_{t-1} + \epsilon_t$ der Koeffizient α 0.8 beträgt,
- A) steigt der Prozess über die Zeit exponentiell an.
 - B) steigt der Prozess gleichmäßig um 1 an.
 - C) steigt der Prozess um das 0.8 fache von y_{t-1} an, wenn y_{t-1} positiv war. (unter der Annahme $\epsilon_t = 0$)
 - D) sinkt der Prozess um das 0.8 fache von y_{t-1} , wenn y_{t-1} negativ war. (unter der Annahme $\epsilon_t = 0$)
 - E) Die Antworten A) bis D) sind falsch.

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

4. damit eine Funktion eine Wahrscheinlichkeitsdichte darstellen kann, muss

- A) $\int_{-\infty}^{+\infty} f(x)dx \leq 1$ sein.
- B) $\int_{-\infty}^{+\infty} f(x)dx = 1$ sein.
- C) $f(x) \geq 0$ sein.
- D) die Zufallsvariable diskret sein.
- 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		

5. Das Schwache Gesetz der großen Zahlen besagt, dass

- A) $P(|\bar{X}_n - \mu| \geq \epsilon) \rightarrow 0$
- B) $P(|\bar{X}_n - \mu| \leq \epsilon) \rightarrow 1$
- C) $P(|\bar{X}_n - \mu| \geq \epsilon) = 0$
- D) $\bar{X}_n = \mu$
- 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. Der Zentrale Grenzwertsatz besagt,

- A) dass mit steigender Anzahl an Beobachtungen X_n für $n \rightarrow \infty$ die Zufallsvariable X gegen eine Normalverteilung strebt.
- B) dass beide Ränder der Normalverteilung für $x \rightarrow +\infty$ und $x \rightarrow -\infty$ gegen Null streben.
- C) die Summe aus unabhängigen und identisch verteilten Zufallsvariablen gegen eine t-Verteilung strebt (solange $n \leq 30$).
- D) die Summe aus unabhängigen und identisch verteilten Zufallsvariablen gegen eine Normalverteilung strebt.
- E) Die Antworten A) bis D) sind falsch.

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

7. Die Likelihood Funktion $L(x_1, x_2, \dots, x_n | \theta)$ gibt an

- A) wie wahrscheinlich ein Parameter θ unter Berücksichtigung der beobachteten x-Werte $x_1, x_2, \dots, x_n$ ist.
- B) für welchen Parameter θ das Auftreten der x-Werte maximal wahrscheinlich ist.
- C) wie wahrscheinlich das Auftreten der x-Werte für einen gegebenen Parameter θ ist.
- D) für welchen Parameter θ das Auftreten der x-Werte am unwahrscheinlichsten ist.
- 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. Ein t-Test auf Signifikanz eines Mittelwertes wird benötigt um zu testen,

- A) ob ein beobachtbarer Mittelwert von 0 abweicht.
- B) ob der wahre aber unbekannte Mittelwert von 0 abweicht.
- C) ob die Abweichung des beobachtbaren Mittelwertes von Null groß genug ist um daraus schließen zu können, dass der wahre Mittelwert $\neq 0$ ist.
- D) ob der wahre aber unbekannte Mittelwert um 0 schwankt.
- 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		

9. Bei der Stichprobenvarianz geht ein Freiheitsgrad verloren

- A) weil die Stichprobenvarianz vom Stichprobenmittelwert abhängt, welcher wiederum selbst eine Schätzung darstellt.
- B) weil durch das Abziehen des Stichprobenmittelwertes eine Beobachtung verloren geht.
- C) weil bei $n - 1$ Beobachtungen die letzte Beobachtung nicht mehr frei ist in dem Sinne, dass mit Hilfe des Mittelwertes die n -te Beobachtung berechnet werden kann.
- D) weil die Varianz um einen konstanten Wert schwankt, welcher in diesem Sinne also nicht frei ist.
- 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		

10. Bei einem Regressionsoutput einer Statistik-Software gibt ein großer F-Wert mit einem dazugehörigen p-Wert von 0.02 an,

- A) dass das Modell die abhängige Variable ausreichend erklärt.
- B) dass jeder einzelne Parameter der exogenen Variablen signifikant ist.
- C) dass mindestens eine Variable einen signifikanten Einfluss hat.
- D) dass Heteroskedastie in den Störtermen vorliegt.
- 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. Ein α -Fehler ist schwerwiegender als ein β -Fehler, weil

- A) lieber eine Nullhypothese annehmen, die falsch ist, als eine Nullhypothese ablehnen, die richtig ist.
- B) lieber eine Nullhypothese ablehnen, die richtig ist, als eine Nullhypothese annehmen, die falsch ist.
- C) der α -Fehler immer wahrscheinlicher ist als ein β -Fehler.
- D) ein β -Fehler nicht zu Verzerrungen der Parameter führt, ein α -Fehler hingegen schon.
- E) Die Antworten A) bis D) sind falsch.

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

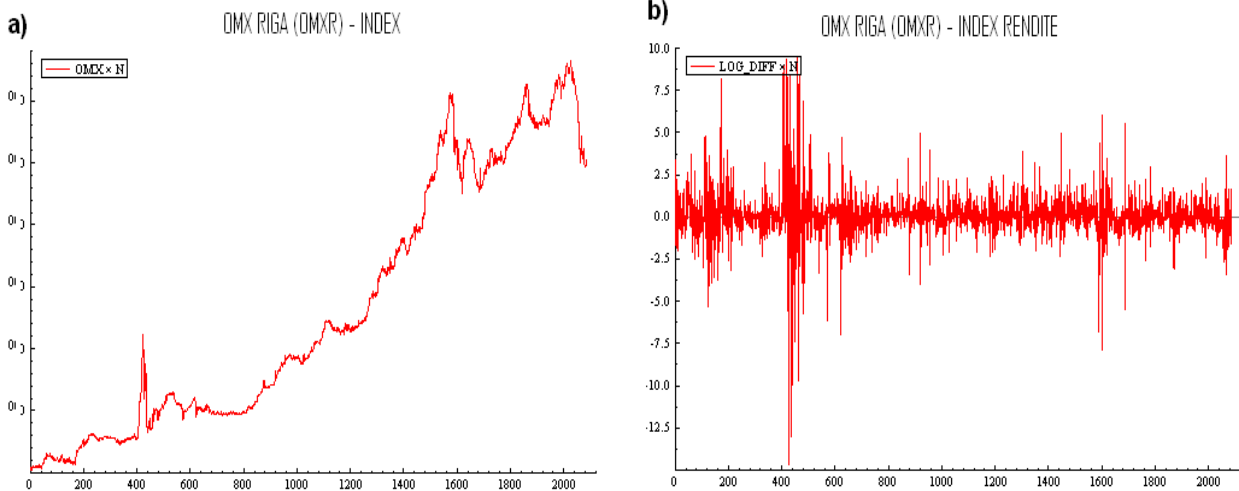
12. $P(t_u \leq t \leq t_o) = 0.95$ bei 7 Freiheitsgraden. Dann ist t_o

- A) -3.124
- B) 2.660
- C) 2.333
- D) 2.365
- E) Die Antworten A) bis D) sind falsch.

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

Teil B (36 Punkte)

1. Mit der Methode der Kleinsten Quadrate wurde der aktuelle Kurswert des Lettischen Aktienindexes (OMX) auf seinen um einen Tag zurückliegenden Wert regressiert (OMX(-1)). TSP generierte hierbei folgenden Output: (Anmerkung: Grafik b zeigt die Renditen des OMX Indexes, welche sich als Differenz aus den logarithmierten Kurswerten ermitteln lassen.)



Equation 1

=====

Method of estimation = Ordinary Least Squares

Dependent variable: OMX
Current sample: 2 to 2086
Number of observations: 2085

Variable	Estimated Coefficient	Standard Error	t-statistic	P-value
C	0.3708	.221270		[.094]
OMX(-1)	1.0200	.531184E-03	1881.89	[.000]

a)

- Für diese Regression wurde angenommen, dass der Verlauf des Aktienindex durch seine Vergangenheitswerte beschrieben wird. Wie nennt man einen solchen Prozess ? (1P)
- Wie lautet die dem Prozess zugrunde liegende theoretische Gleichung, wie sieht ihr Schätzwert aus ? (1P)
- Berechnen sie anhand der oben angegebenen Informationen die Test-Statistik (t-statistic) der Variablen C.(1P)
- Ist der Prozess mittelwertstationär ? (Hinweis: Betrachten sie die geschätzten Koeffizienten) (1P)

(4 Punkte)

- b) In einem zweiten Schritt wurden die Renditen des Index der gleichen Analyse unterzogen und folgender Output generiert:

Equation 2

=====

Method of estimation = Ordinary Least Squares

Dependent variable: LOG_DIFF

Current sample: 3 to 2086

Number of observations: 2084

Variable	Estimated Coefficient	Standard Error	t-statistic	P-value
C	.073535	.033031	2.22621	[.026]
LOG_DIFF(-1)	.099856	.021782	4.58442	[.000]

- Wie lautet hier die Schätzgleichung? Ist dieser Prozess mittelwertstationär ? (2P)

- Die schwache Finanzmarkteffizienz besagt, dass vergangene Renditen keine Informationen für die zukünftigen Renditen liefern. Trifft diese These bei einem Signifikanzniveau von 5% auch auf den lettischen Markt zu ? (2P)

(3 Punkte)

- c) Erklären Sie verbal, was man unter einem Random Walk ohne Drift versteht.

(1 Punkt)

(9 Punkte)

2.

- a) Um die Einstellung der Bürger hinsichtlich der Bankenkrise zu untersuchen, werden in der Freiburger Innenstadt in einer Stichprobe 400 Passanten befragt. Im Durchschnitt geben 45% an, dass die Kredite von in Mitleidenschaft gezogenen Banken durch Steuergelder unterlegt sein sollen. 55% sind dagegen. Berechnen Sie das 95%-Konfidenzintervall für den wahren Anteilswert der zustimmenden Passanten. Wie repräsentativ ist diese Stichprobe für die Einstellung aller Bürger in Deutschland?

(2 Punkte)

- b) In einer Befragung unter 20 Banken wird festgestellt, dass das Verhältnis von Fremdkapital zu Eigenkapital im Durchschnitt 13 beträgt, mit einer Standardabweichung von 2. Berechnen Sie das 90%-Konfidenzintervall in dem sich dieses Verhältnis bei allen Banken befinden wird (mit Zurücklegen).

(2 Punkte)

- c) Um eine robustere Schätzung des wahren Mittelwertes zu erhalten, wird in einer zweiten Befragung vom Umfang 100 ein Mittelwert von 13.5 bei einer Standardabweichung von 1.8 festgestellt. Berechnen Sie das Intervall, in dem der wahre Mittelwert mit 90% Wahrscheinlichkeit höchstens liegt (mit Zurücklegen).

(2 Punkte)

(6 Punkte)

3.

- a) Das lineare Regressionsmodell wird in Summennotation im univariaten Fall oft folgendermaßen dargestellt: $y_i = \alpha + \beta x_i + u_i$.
In Matrixnotation lautet die Schreibweise hingegen: $Y = X\beta + u$.
Gehen Sie kurz darauf ein, wieso die Konstante α nicht in der Matrixnotation auftaucht und weshalb man im Gegensatz zur Summennotation nicht βX schreiben kann.

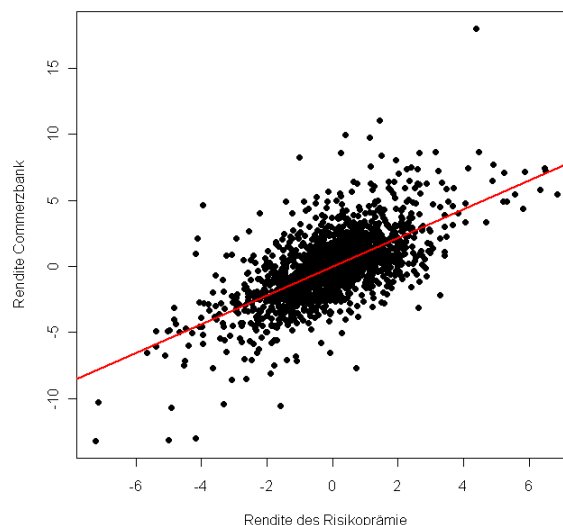
(2 Punkte)

- b) Für den Vektor der Störterme u gelten die Annahmen, dass die Störterme untereinander unabhängig sind, d.h. $E(uu') = \sigma^2 I$, und dass Homoskedastizität gilt. Kommentieren Sie in diesem Sinne die folgenden 3 Matrizen im Hinblick auf Annahmeverletzungen.

$$A = \begin{pmatrix} \sigma^2 & 0.3 & -0.12 \\ 0.3 & \sigma^2 & 0.28 \\ -0.12 & 0.28 & \sigma^2 \end{pmatrix} \quad B = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix} \quad C = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$$

(2 Punkte)

- c) Das Capital Asset Pricing Model ist ein Gleichgewichtsmodell zur Beschreibung von Aktienrenditen. Dabei wird im folgenden die Renditezeitreihe der Commerzbank-Aktie der letzten Jahre durch die Risikoprämie erklärt: $R_{comm} = \alpha + \beta R_{Prem}$.



Als Trainee Equity Analyst einer Bank haben Sie folgende Werte zusammengetragen: $n = 2239$ Beobachtungen, $\mu_{Comm} = -0.17$, $\mu_{Prem} = 0.0009$, $\sigma_{Comm}^2 = 5.39$, $\sum (R_{Prem} - \bar{R}_{Prem})^2 = 4331.288$. Der Korrelationskoeffizient ρ beträgt 0.6535. Berechnen Sie β und α .

(4 Punkt)

- d) Der Regression aus Aufgabenteil c) wird nun zusätzlich die Rendite der Siemens-Aktie als exogene Variable beigelegt. In Matrixnotation erhalten Sie:

$$(X'X)^{-1} = \begin{pmatrix} 4.466292e-04 & 3.731439e-07 & -4.663067e-07 \\ 3.731439e-07 & 5.603454e-04 & -2.636778e-04 \\ -4.663067e-07 & -2.636778e-04 & 2.110917e-04 \end{pmatrix} \quad X'Y = \begin{pmatrix} -39.06524 \\ 4725.63245 \\ 5894.25654 \end{pmatrix}$$

Berechnen Sie den Koeffizientenvektor β . Hat das Hinzuziehen der Siemensrendite ihrer Meinung nach einen wesentlichen Beitrag zur Erklärung des Modells geleistet?

(Hinweis: $4.466292e-04 = \frac{4.466292}{10000} = 0.0004466292$)

(4 Punkte)

(12 Punkte)

4.

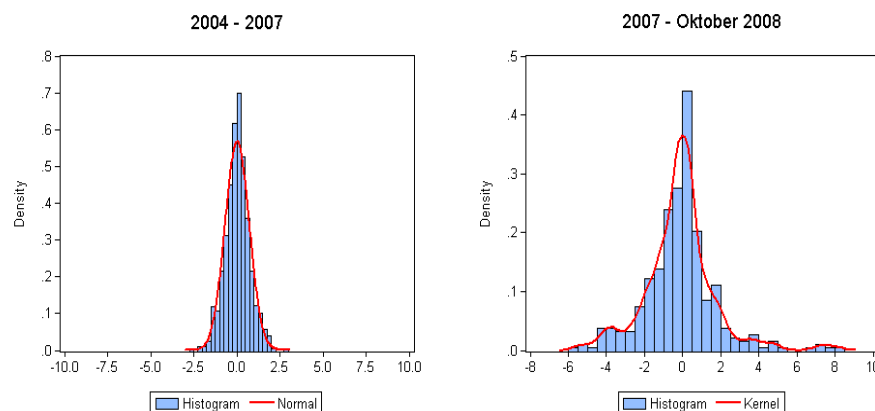
- a) Die Wahrscheinlichkeit, bei einer zufällig ausgewählten deutschen Landesbank einen sehr hohen Anteil an amerikanischen Subprime-Krediten im Portfolio zu finden, beträgt 0.75. Wie hoch ist die Wahrscheinlichkeit, bei den 12 deutschen Landesbanken zufällig 3 Banken herauszusuchen, bei denen der Anteil eher gering ist?

(1 Punkt)

- b) Wie hoch ist die Wahrscheinlichkeit mindestens 2 Banken mit geringem Anteil an den Subprime-Krediten zu finden?

(1 Punkt)

- c) Die folgenden Abbildungen zeigen die Verteilung der Renditen amerikanischer Banken von 2004 bis 2007 und von 2007 bis Oktober 2008. Die durchschnittliche Tagesrendite im Zeitraum vor der Krise betrug $\mu = 0.04$ bei einer Standardabweichung von $\sigma = 0,7$. In der Periode seit 2007 beträgt die durchschnittliche Tagesrendite -0.13 und die Volatilität 1.88.



Wie hoch war die Wahrscheinlichkeit im Zeitraum 2004 - 2007 (vor der Krise) einen Tagesverlust zu erzielen, der mehr als 2% beträgt (Annahme einer Normalverteilung der Renditen)? Vergleichen Sie die Wahrscheinlichkeit mit derjenigen während der Krise.

(3 Punkte)

- d) Der 5% Value-at-risk gibt diejenige negative Rendite an, welche nur mit 5% Wahrscheinlichkeit noch unterschritten wird: $P(X < VaR) = 0.05$. Berechnen Sie den VaR für beide Zeiträume und kommentieren Sie kurz.

(2 Punkte)

- e) Während die Renditen im Zeitraum vor der Krise noch annähernd normalverteilt waren, weichen sie im Krisenzeitraum deutlich von der Normalverteilung ab. An welchen 2 Merkmalen erkennen Sie graphisch, dass die Renditen nicht mehr normalverteilt sind? Welche Probleme treten dadurch auf, wenn man den VaR wie in Aufgabenteil d) berechnet?

(2 Punkte)

(9 Punkte)