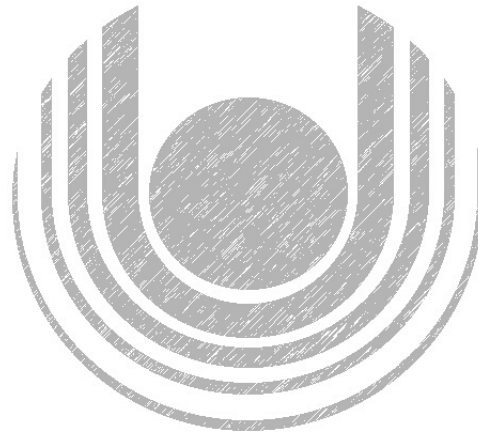

Name, Vorname

--	--	--	--	--	--	--	--

Matrikelnummer



Modulklausur 31821 – Multivariate Verfahren

Datum

Punkte

Note

Termin: 25.09.2020, 17:15 - 19:15 Uhr

Prüfer: Univ.-Prof. Dr. R. Kruse-Becher

Hinweise zur Bearbeitung der Modulklausur 31821

1. Füllen Sie zunächst den **Kopf des Deckblatts** aus!
2. Es können insgesamt 100 Punkte erreicht werden. Bei Erreichen von 50 Punkten ist die Klausur bestanden. **Bitte kontrollieren Sie sofort, ob Sie ein vollständiges Klausurexemplar erhalten haben.**
3. Zugelassen ist Kurseinheit 1 des Moduls 31821 (Kursnr. 00883) mit farblichen Markierungen, kleinen Aufklebern und/oder textbezogenen Anmerkungen. Nicht zugelassen sind selbst ausgedruckte oder kopierte Kursmaterialien.
Die Verwendung eines Taschenrechners ist dann und nur dann erlaubt, wenn dieser einer der drei folgenden Modellreihen angehört:
 - Casio fx86 oder Casio fx87
 - Texas Instruments TI 30 X II
 - Sharp EL 531

Eventuelle Vorgänger- oder Nachfolgemodelle, die nicht in der oben aufgeführten Liste enthalten sind, sind ebenfalls nicht erlaubt. Die Verwendung anderer Taschenrechnermodelle wird als Täuschungsversuch gewertet und mit der Note „nicht ausreichend“ (5,0) sanktioniert.

Ob ein Taschenrechner einer der drei Modellreihen angehört, können Sie selbst überprüfen, indem Sie die vom Hersteller auf dem Rechner angebrachte Modellbezeichnung mit den oben angegebenen Bezeichnungen vergleichen: Bei vollständiger Übereinstimmung ist das Modell erlaubt. Ist die auf dem Rechner angebrachte Modellbezeichnung umfangreicher, enthält aber eine der oben angegebenen Bezeichnungen vollständig, ist das Modell ebenfalls erlaubt. In allen anderen Fällen ist das Modell nicht erlaubt.

4. Bitte benutzen Sie für Ihre Rechnungen nur die beigelegten Lösungsbögen. **Tragen Sie zusätzlich alle Ergebnisse in die vorgesehenen Zeilen ein. Der Rechenweg ist lesbar und nachvollziehbar anzugeben!**
5. Wenn Sie die einzelnen Blätter der Klausur voneinander trennen, **vermerken Sie auf jedem Blatt Ihre Matrikelnummer**. Legen Sie bitte am Ende der Klausur die Blätter wieder zusammen.
6. Vergessen Sie nicht, die Klausur auf der letzten bearbeiteten Seite zu **unterschreiben**.

Wir wünschen Ihnen viel Erfolg!

Aufgabe 1

(20 Punkte)

Nehmen Sie an, dass für die eindimensionalen Zufallsvariablen X und Y folgendes gilt:
 $\text{Var}(X) = 9$, $\text{Var}(Y) = 4$ und $\text{Var}(2X + 4Y) = 52$.

- a) Welchen Wert nimmt der Korrelationskoeffizient zwischen X und Y an? (6 P.)

Betrachten Sie nun die beiden zweidimensionalen Zufallsvektoren

$$\mathbf{x}_1 \sim N \left(\begin{bmatrix} 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right), \quad \mathbf{x}_2 \sim N \left(\begin{bmatrix} 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 & -2 \\ -2 & 5 \end{bmatrix} \right)$$

mit der Kovarianzmatrix $\text{Cov}(\mathbf{x}_1, \mathbf{x}_2) = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$.

Berechnen Sie die folgenden Ausdrücke:

b) $E[\mathbf{x}_2 \mathbf{x}_2']$ und $E[\mathbf{x}_1 \mathbf{x}_2']$ (5 P.)

c) $\text{Var}(\mathbf{x}_1 + \mathbf{x}_2)$ (4 P.)

d) $\text{Cov} \left(\mathbf{x}_1, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right)$ und $\text{Cov} \left(\begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{x}_1, \begin{bmatrix} 1 & 2 \end{bmatrix} \mathbf{x}_2 \right)$ (5 P.)

Aufgabe 2

(20 Punkte)

Auf der folgenden Seite finden Sie die Untersuchung einer Bank, die prüfen möchte, ob sich zwei Gruppen mit verschiedenen Einkommen (unter 50 Tausend Euro im Jahr bzw. über 50 Tausend Euro im Jahr) hinsichtlich ihrer Verschuldung durch Kreditkarten (X) und andere Arten von Krediten (Y) unterscheiden.

Hinweis: Die passenden Werte können Sie den Statistiken auf Seite 3 entnehmen.

- a) Bestimmen Sie die gepoolte Kovarianzmatrix \mathbf{S} mit Hilfe der gruppenspezifischen Kovarianzen! (4 P.)

Hinweis: Falls Sie \mathbf{S} nicht berechnen konnten, verwenden Sie im folgenden die Matrix:

$$\mathbf{S} = \begin{bmatrix} 3,5 & 3 \\ 3 & 8,6 \end{bmatrix}$$

- b) Berechnen Sie Inverse der gepoolten Kovarianzmatrix \mathbf{S} . Testen Sie anschließend zum 5%-Niveau, ob die Mittelwerte der beiden Stichproben gleich sind! Gehen Sie davon aus, dass die Kovarianzen beider Gruppen gleich, aber unbekannt sind. (8 P.)
- c) Geben Sie die Korrelation zwischen X und Y in der Gruppe mit dem Einkommen unter 50 Tausend Euro im Jahr an. Ist der Korrelationskoeffizient zum 5%-Niveau von 0 verschieden? (3 P.)
- d) Führen Sie einen Likelihood-Quotienten-Test für die Hypothese $H_0 : \mathbf{P} = \mathbf{I}$ zum 5%-Niveau durch. Hierbei beschreibt \mathbf{P} die Korrelationsmatrix zwischen X und Y in der Gruppe mit einem Einkommen unter 50 Tausend Euro im Jahr und \mathbf{I} die Einheitsmatrix. (5 P.)

Korrelationen

Deskriptive Statistiken

Einkommen		Mittelwert	Standardabweichung	N
Einkommen über 50 T Euro	Kreditkartenverbindlichkeiten (in T Euro)	3,0447	3,17402	248
	Andere Schulden (in T Euro)	5,8104	4,78374	248
Einkommen unter 50 T euro	Kreditkartenverbindlichkeiten (in T Euro)	,9721	,99406	602
	Andere Schulden (in T Euro)	1,9535	1,60195	602

Korrelationen^{b,c}

Einkommen			Kreditkartenverbindlichkeiten (in T Euro)	Andere Schulden (in T Euro)
Einkommen über 50 T Euro	Kreditkartenverbindlichkeiten (in T Euro)	Korrelation nach Pearson	1	,562**
		Signifikanz (2-seitig)		,000
		Quadratsummen und Kreuzprodukte	2488,372	2106,574
		Kovarianz	10,074	8,529
Andere Schulden (in T Euro)	Kreditkartenverbindlichkeiten (in T Euro)	Korrelation nach Pearson	,562**	1
		Signifikanz (2-seitig)	,000	
		Quadratsummen und Kreuzprodukte	2106,574	5652,378
		Kovarianz	8,529	22,884
Einkommen unter 50 T euro	Kreditkartenverbindlichkeiten (in T Euro)	Korrelation nach Pearson	1	,466**
		Signifikanz (2-seitig)		,000
		Quadratsummen und Kreuzprodukte	593,880	445,647
		Kovarianz	,988	,742
Andere Schulden (in T Euro)	Kreditkartenverbindlichkeiten (in T Euro)	Korrelation nach Pearson	,466**	1
		Signifikanz (2-seitig)	,000	
		Quadratsummen und Kreuzprodukte	445,647	1542,310
		Kovarianz	,742	2,566

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

b. Einkommen=Einkommen über 50 T Euro,;Listenweise N=248

c. Einkommen=Einkommen unter 50 T euro,;Listenweise N=602

Aufgabe 3

(30 Punkte)

Für ein Unternehmen seien nachfolgende Umsätze Y und Werbeausgaben X , jeweils in Mio. Euro gemessen, in den Jahren 2017-2019 bekannt:

<i>Jahr</i>	<i>Umsatz</i>	<i>Werbeausgaben</i>
2017	10	3
2018	12	4
2019	15	5

Betrachten Sie folgendes lineares Regressionsmodell:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad i = 1, 2, 3$$

- Schreiben Sie das Modell in Matrixform auf und schätzen Sie den Parametervektor $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ mit Hilfe der Kleinste-Quadrate-Methode. (6 P.)
- Bestimmen Sie nun die Residuen und entsprechende Fehlervarianz. (4 P.)
- Wie viel Prozent der Varianz wird durch das geschätzte Modell erklärt? Berechnen Sie hierfür den Determinationskoeffizienten R^2 ! (8 P.)
- Stellen Sie die ANOVA-Tabelle auf und prüfen Sie, ob $H_0 : \beta_1 = 0$ zum Niveau $\alpha = 5\%$ gilt. (6 P.)
- Welchen Umsatz erzielt das Unternehmen voraussichtlich, wenn es für die Werbung 7 Mio. Euro ausgibt? Bestimmen Sie einen Schätzer und das Prognoseintervall ($\alpha = 5\%$). (6 P.)

Aufgabe 4

(30 Punkte)

4.1 Definieren Sie die L_q -Distanz und beschreiben Sie die Complete-Linkage-Methode. Nach welchem Kriterium werden die Klassen miteinander fusioniert? (4 P.)

4.2 Gegeben seien die Objekte A, B, C und D, die jeweils durch fünf Merkmale beschrieben werden. Die entsprechende Datenmatrix \mathbf{X} hat dabei folgende Gestalt:

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 & 4 & 3 \\ 3 & 1 & 2 & 5 & 2 \\ 4 & 1 & 5 & 2 & 7 \\ 5 & 2 & 1 & 4 & 1 \end{bmatrix}$$

Zeile 1 enthält dabei die Merkmale von Objekt A, Zeile 2 die Merkmale von Objekt B, etc.

Hinweis: Für zwei positive Zahlen x und y mit $x < y$ gilt $\sqrt{x} < \sqrt{y}$, d.h. Sie müssen die Wurzel nicht explizit ausrechnen, wenn Sie die Distanzen der einzelnen Objekte bestimmen.

- a) Konstruieren Sie agglomerativ eine Hierarchie. Verwenden Sie dabei die Complete-Linkage-Methode, indem Sie zunächst die euklidische Distanzmatrix bilden. Geben Sie stets die Klassen und den Indexwert der Fusion an. (11 P.)
- b) Zeichnen Sie das entsprechende Dendrogramm. (3 P.)
- c) Berechnen Sie die Abstände neu mit Hilfe der City-Block-Metrik. (5 P.)
- d) Führen Sie analog zur a) mit Hilfe der Matrix aus c) eine Clusteranalyse durch. Verwenden Sie dabei die Single-Linkage-Methode. (7 P.)

--	--	--	--	--	--	--	--

Klausur: Kurs 31821

Multivariate Verfahren

Datum: 25.09.2020

Prüfer: Univ.-Prof. Dr. R. Kruse-Becher

Name, Vorname:

Matrikelnummer:

Aufgabe	1	2	3	4		Summe
maximale Punktzahl	20	20	30	30		100
erreichte Punktzahl						

Datum:

Unterschrift des Prüfers:

--	--	--	--	--	--	--

Aufgabe 1 - Rechenweg

Punkte

Aufgabe 1 - Lösungen

(20 Punkte)

Lösung a)

(5 Punkte)

Lösung b)

(4 Punkte)

Lösung c)

(5 Punkte)

Lösung d)

(6 Punkte)

Punkte

--	--	--	--	--	--	--

Aufgabe 2 - Rechenweg

Punkte

Aufgabe 2 - Lösungen

(20 Punkte)

Lösung a)

(4 Punkte)

Lösung b)

(8 Punkte)

Lösung c)

(3 Punkte)

Lösung d)

(5 Punkte)

Punkte


--	--	--	--	--	--	--

Aufgabe 3 - Rechenweg 1

Punkte

--	--	--	--	--	--	--

Aufgabe 3 - Rechenweg 2

Punkte 

Aufgabe 3 - Lösungen

(30 Punkte)

Lösung a)

(6 Punkte)

Lösung b)

(4 Punkte)

Lösung c)

(8 Punkte)

Lösung d)

(6 Punkte)

Lösung e)

(6 Punkte)

Punkte


--	--	--	--	--	--	--

Aufgabe 4 - Rechenweg 1

Punkte

--	--	--	--	--	--	--

Aufgabe 4 - Rechenweg 2

Punkte 

Aufgabe 4 - Lösungen

(30 Punkte)

Lösung a)

(12 Punkte)

Lösung b)

(4 Punkte)

Lösung c)

(6 Punkte)

Lösung d)

(8 Punkte)

Punkte



--	--	--	--	--	--	--

Punkte

--	--	--	--	--	--	--

Punkte

--	--	--	--	--	--	--

Punkte

--	--	--	--	--	--	--

Punkte

--	--	--	--	--	--	--

Punkte