
Name, Vorname

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

Matrikelnummer



Modulklausur 31821 – Multivariate Verfahren

Datum

Punkte

Note

Termin: 30. September 2016, 9.00 - 11.00 Uhr

Erstprüfer: Univ.-Prof. Dr. H. Singer

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.
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.

Wir wünschen Ihnen viel Erfolg!

Aufgabe 1

(15 Punkte)

Kennzeichnen Sie die folgenden Aussagen zur Varianzanalyse mit **R** für richtig oder **F** für falsch.

- In der Varianzanalyse wird der Zusammenhang zwischen zwei diskreten Merkmalen untersucht.
- In der H_1 - Hypothese der Varianzanalyse wird behauptet, dass alle Mittelwerte der Subpopulationen sich voneinander unterscheiden.
- Mit wachsendem F -Wert bei der Varianzanalyse wird die empirische Varianz zwischen den Gruppen immer größer im Verhältnis zur empirischen Innergruppen-Varianz.
- Die Varianzanalyse kann nur bei gleicher Beobachtungszahl innerhalb der Gruppen angewendet werden.
- Mithilfe der zweifaktoriellen Varianzanalyse können Haupteffekte, aber keine Interaktionseffekte getestet werden.

Hinweis: Für jede korrekte Kennzeichnung werden 3 Punkte vergeben. Jede falsche Kennzeichnung sowie nicht oder unlesbar gekennzeichnete Felder werden mit 0 Punkten bewertet. Die minimale Punktzahl der Aufgabe beträgt 0 Punkte.

Aufgabe 2

(28 Punkte)

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

a) Bestimmen Sie die gepoolte Kovarianzmatrix! (6 P.)

b) Die Inverse der gepoolten Kovarianzmatrix ist

$$\mathbf{S}^{-1} = \begin{bmatrix} 0.39 & -0.14 \\ -0.14 & 0.17 \end{bmatrix}.$$

Testen Sie zum 5%-Niveau, ob die Mittelwerte der beiden Stichproben gleich sind! Gehen Sie davon aus, dass die Kovarianzen beider Gruppen gleich, aber unbekannt sind. (12 P.)

c) Geben Sie die Korrelation zwischen den beiden Kreditarten in der Gruppe mit dem Einkommen unter 50 Tausend Euro im Jahr an. Ist sie zum 5%-Niveau von 0 verschieden? (2 P.)

d) Welcher Test lässt sich mit der Prüfgröße

$$t^2 = 248 \left(\begin{bmatrix} 3.0447 & 5.8104 \end{bmatrix} - \begin{bmatrix} 3 & 5 \end{bmatrix} \right) \mathbf{S}^{-1} \left(\begin{bmatrix} 3.0447 \\ 5.8104 \end{bmatrix} - \begin{bmatrix} 3 \\ 5 \end{bmatrix} \right) = 9.49$$

bei korrekter Wahl von \mathbf{S} durchführen? Geben Sie auch den Namen der Verteilung und die Nullhypothese an. Führen Sie den Test mit dem Signifikanzniveau $\alpha = 5\%$ aus! (8 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
	Andere Schulden (in T Euro)	Korrelation nach Pearson	,562**	1
		Signifikanz (2-seitig)	,000	
		Quadratsummen und Kreuzprodukte	2106,574	5652,378
		Kovarianz	10,074	8,529
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
	Andere Schulden (in T Euro)	Korrelation nach Pearson	,466**	1
		Signifikanz (2-seitig)	,000	
		Quadratsummen und Kreuzprodukte	445,647	1542,310
		Kovarianz	,988	,742
		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

(27 Punkte)

In dieser Aufgabe sollen mit dem Logistischen Modell die Einflussfaktoren für das Bestehen einer Erwerbstätigkeit seitens der Ehefrauen untersucht werden. Abhängige Variable **Erwerbstätig** nimmt den Wert 1 bei Erwerbstätigen und den Wert 0 bei Arbeitslosen an. Als erklärende Variablen stehen zur Verfügung Daten zur **Bildung** (=1 wenn mindestens das Abitur vorliegt, und 0 andernfalls) und Variable **Kinder unter 6 Jahren** mit dem Wert 1, wenn im Haushalt Kind bzw. Kinder unter 6 Jahren wohnen, sonst mit dem Wert 0. In der folgenden Tabelle sind die zu analysierenden Daten zu finden.

Person	Bildung	Kinder unter 6 Jahren	Erwerbstätig
A	1	0	1
B	1	1	1
C	0	1	0
D	0	1	0
E	1	0	1

a) Geben Sie die Datenmatrix in Effekt- und Dummy-Kodierung mit Interaktionseffekten an! (6 P.)

b) Berücksichtigen Sie jedoch in den folgenden Aufgabenteilen nur die unabhängige Variable **Kinder unter 6 Jahren** in Dummy-Kodierung, und die abhängige Variable der Erwerbstätigkeit der Ehefrauen. Bestimmen Sie nun die Likelihoodfunktion und die Log-Likelihood für ein Modell mit Konstante! (10 P.)

c) Bei der Parameterschätzung hat der Optimierungsalgorithmus

$$\hat{\beta} = [\hat{\beta}_0 \ \hat{\beta}_1] = [0.7 \ -0.7]'$$

ergeben (Dummy-Kodierung). Wie hoch ist die Wahrscheinlichkeit, dass eine Frau, die ein fünfjähriges Kind hat, erwerbstätig ist? (4 P.)

d) Die Log-Likelihood hat an der Stelle $\hat{\beta} = [\hat{\beta}_0 \ \hat{\beta}_1] = [0.7 \ -0.7]'$ den Wert -3.3 , die zweite Ableitung nach beiden Variablen ist gleich $\mathbf{H} = \begin{bmatrix} -1.17 & -0.5 \\ -0.5 & -0.5 \end{bmatrix}$. Testen Sie mit der Wald-Statistik zum 5%-Niveau, ob die tatsächlichen Werte $\xi = \beta = [0 \ 0]'$ sind. (7 P.)

Aufgabe 4

(30 Punkte)

In der Datenmatrix sind 15 Städte anhand von drei Merkmalen des Lebensstandards gegenübergestellt. Es stehen Daten über die durchschnittliche Anzahl von Arbeitsstunden im Jahr, das durchschnittliche Gehalt (Gehalt in Zürich als Basiswert) und durchschnittliche Preise (Preise in Zürich als Basiswert) zur Verfügung. Die Daten sowie die durchgeführte Clusteranalyse finden Sie auf den nächsten Seiten.

a) Berechnen Sie die L_2 -Distanz und die City-Block-Metrik zwischen den Städten Düsseldorf und Dublin! (6 P.)

b) Zeigen Sie, dass die Rotationsmatrix

$$\mathbf{B} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

eine orthogonale Matrix ist, indem Sie die Determinante bestimmen und $\mathbf{B}' = \mathbf{B}^{-1}$ zeigen. (6 P.)

c) Rotieren Sie die Datenpunkte 5 und 6 mit der Rotationsmatrix \mathbf{B} und zeigen Sie, dass der City-Block-Abstand sich durch die Rotation nicht geändert hat! (7 P.)

d) Welche Objekte wurden bei der SPSS-Clusteranalyse zuerst zusammengefasst? Wenn Sie eine 4-Cluster-Lösung wählen, wieviele Objekte sind dann im kleinsten Cluster (das mit den wenigsten Objekten)? Geben Sie den Indexwert der Fusionierung für die 4-Cluster-Lösung an! (6 P.)

e) Angenommen, Objekt 4 und 8 bilden ein Cluster. Wenn Sie die quadrierte euklidische Distanz und die Single-Linkage-Methode verwenden, wird dann zuerst Objekt 5 oder Objekt 13 zu diesem Cluster hinzugefügt? (5 P.)

*Hinweis: Die Lösung dieser Aufgabe können Sie **nicht** dem Dendrogramm entnehmen.*

	Stadt	Arbeitsstunden	Preise	Gehalt	var
1	Amsterdam	1714,0	65,6	49,0	
2	Brussels	1708,0	73,8	50,5	
3	Chicago	1924,0	73,9	61,9	
4	Copenhagen	1717,0	91,3	62,9	
5	Dublin	1759,0	76,0	41,4	
6	Dusseldorf	1693,0	78,5	60,2	
7	Luxembourg	1768,0	71,1	71,1	
8	Madrid	1710,0	93,8	50,0	
9	New York	1942,0	83,3	65,8	
10	Oslo	1583,0	115,5	63,7	
11	Paris	1744,0	81,6	45,9	
12	Singapore	2042,0	64,4	16,1	
13	Stockholm	1805,0	111,3	39,2	
14	Vienna	1780,0	78,0	51,3	
15	Zurich	1868,0	100,0	100,0	

Cluster

Zusammenfassung der Fallverarbeitung^a

Fälle					
Gültig		Fehlend		Gesamt	
N	Prozent	N	Prozent	N	Prozent
15	100,0	0	,0	15	100,0

a. Complete Linkage

Näherungsmatrix

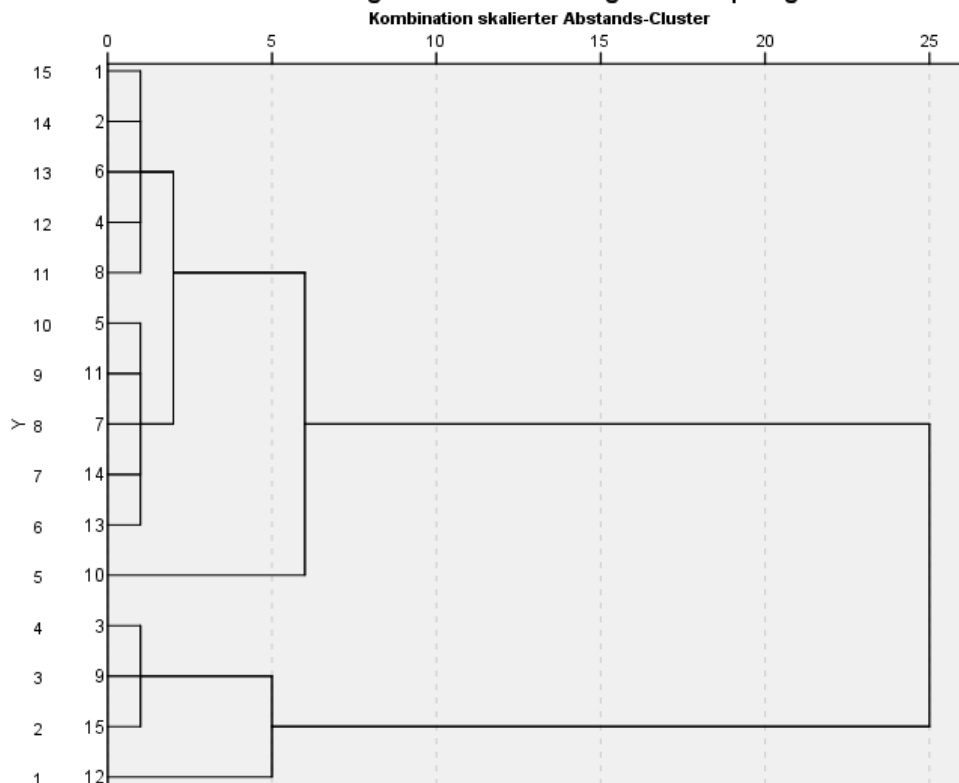
Fall	Quadratisches euklidisches Distanzmaß														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	,00	105,49	44335,30	862,70	2190,92	732,85	3434,66	812,24	52579,53	19867,10	1165,61	108667,85	10465,53	4515,05	27500,36
2	105,49	,00	46785,97	541,01	2688,65	341,18	4031,65	404,25	55080,34	17538,13	1378,00	112827,72	10942,94	5202,28	28736,69
3	44335,30	46785,97	,00	43152,76	27649,66	53385,05	24428,48	46333,62	427,57	118014,80	32715,29	16111,89	16075,05	20865,17	5268,82
4	862,70	541,01	43152,76	,00	2460,34	747,13	3076,28	221,66	50697,41	18542,28	1112,09	108538,85	8705,69	4280,45	24253,10
5	2190,92	2688,65	27649,66	2460,34	,00	?	987,10	2791,80	34137,65	33033,54	276,61	80863,65	3366,93	543,01	15890,96
6	732,85	341,18	53385,05	747,13	?	,00	5798,57	627,13	62055,40	13481,25	2815,10	123944,62	14060,84	7648,46	32671,29
7	3434,66	4031,65	24428,48	3076,28	987,10	5798,57	,00	4324,50	30452,93	36251,12	1321,29	78145,89	4002,65	583,65	11670,42
8	812,24	404,25	46333,62	221,66	2791,80	627,13	4324,50	,00	54183,89	16787,58	1321,65	112237,57	9447,89	5151,33	27502,44
9	52579,53	55080,34	427,57	50697,41	34137,65	62055,40	30452,93	54183,89	,00	129922,25	39602,90	12827,30	20260,56	26482,34	6924,53
10	19867,10	17538,13	118014,80	18542,28	33033,54	13481,25	36251,12	16787,58	129922,25	,00	27387,05	215557,97	49901,89	40369,01	82782,94
11	1165,61	1378,00	32715,29	1112,09	276,61	2815,10	1321,29	1321,65	39602,90	27387,05	,00	89987,88	4647,98	1338,12	18641,37
12	108667,85	112827,72	16111,89	108538,85	80863,65	123944,62	78145,89	112237,57	12827,30	215557,97	89987,88	,00	58902,22	70068,00	38582,57
13	10465,53	10942,94	16075,05	8705,69	3366,93	14060,84	4002,65	9447,89	20260,56	49901,89	4647,98	58902,22	,00	1880,30	7793,33
14	4515,05	5202,28	20865,17	4280,45	543,01	7648,46	583,65	5151,33	26482,34	40369,01	1338,12	70068,00	1880,30	,00	10599,69
15	27500,36	28736,69	5268,82	24253,10	15890,96	32671,29	11670,42	27502,44	6924,53	82782,94	18641,37	38582,57	7793,33	10599,69	,00

Dies ist eine Unähnlichkeitsmatrix

Complete Linkage

Zuordnungsübersicht						
Schritt	Zusammengeführte Cluster		Koeffizienten	Erstes Vorkommen des Clusters		Nächster Schritt
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	2	105,490	0	0	6
2	4	8	221,660	0	0	7
3	5	11	276,610	0	0	8
4	3	9	427,570	0	0	10
5	7	14	583,650	0	0	8
6	1	6	732,850	1	0	7
7	1	4	862,700	6	2	11
8	5	7	1338,120	3	5	9
9	5	13	4647,980	8	0	11
10	3	15	6924,530	4	0	12
11	1	5	14060,840	7	9	13
12	3	12	38582,570	10	0	14
13	1	10	49901,890	11	0	14
14	1	3	215557,970	13	12	0

Dendrogramm mit vollständiger Verknüpfung



LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte

LÖSUNGSBOGEN

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

Punkte