



**KLAUSUR: MULTIVARIATE ANALYSIS**      **WS 2006/07**  
**PRÜFER: PROF. DR. B. ERICHSON**      **PRÜFUNGSNUMMER: 2624**

Als Hilfsmittel sind zugelassen: Taschenrechner lt. Aushang des Prüfungsamtes.  
 Die folgenden Aufgaben sind alle zu bearbeiten. (Die Sollbearbeitungszeit ist für jede Aufgabe angegeben. Wobei ein Punkt einer Minute entspricht) Der Klausuraufgabenbogen umfaßt zwei Seiten.

**Aufgabe 1 (18 Punkte)**

Ein Umfrage bezüglich des Biermarktes wurde mittels Faktorenanalyse ausgewertet.

- a) Erläutern Sie die Aufgabenstellung der Faktorenanalyse! Was sind Faktoren und welche kausale Struktur wird zwischen ihnen und den Ausgangsvariablen zugrunde gelegt?  
 b) Definieren Sie kurz die Begriffe Eigenwert, Kommunalität und Faktorladung (inhaltlich)! Berechnen Sie den Eigenwert des zweiten Faktors sowie die Kommunalität der Variable Kosten!

Komponentenmatrix

	Komponente	
	1	2
Herstellungsland	,746	,594
Kosten	,771	,556
Kalorien	,842	-,500
Alkohol-Volumenprozente	,806	-,559

Erklärte Gesamtvarianz

Komponente	Anfängliche Eigenwerte
	Gesamt
1	2,511
2	?
3	,189
4	,076

- c) Was besagt das Kaiser-Kriterium? Erklären Sie es kurz! Wie viele Faktoren würden Sie auf dieser Basis im Beispiel extrahieren? Wie hoch ist der Varianzerklärungsanteil des zweiten Faktors?  
 d) Wie werden Faktoren extrahiert, d.h. welche Bedingungen müssen sie erfüllen?  
 e) Erläutern Sie kurz das Problem der Interpretation der Faktoren? Wie würde sich die Interpretierbarkeit verbessern lassen? Was bedeutet in diesem Zusammenhang eine „Einfachstruktur“?  
 f) Aus der Matrixdarstellung  $Z=A \cdot P$  läßt sich für jede Zeile die Gleichung  $z_j = a_{j1} \cdot p_1 + a_{j2} \cdot p_2$  schreiben. Identifizieren Sie abhängige und unabhängige Variablen sowie Regressionskoeffizienten!

**Aufgabe 2 (12 Punkte)**

- a) Experimentellen Auswertungen wird häufig folgendes Modell zugrunde gelegt:  
 $y_{ghk} = \mu + \alpha_g + \beta_h + (\alpha\beta)_{gh} + \varepsilon_{ghk}$ . Welches Verfahren wird dadurch beschrieben? Was stellt in der Formel:  $(\alpha\beta)_{gh}$  dar? Wie baut sich das Grundtheorem auf, dass dem Verfahren zu Grunde liegt?  
 b) Existiert ein signifikanter Einfluss der unabhängigen Variable Platzierung auf die abhängige Variable Absatzmenge Cola? ( $\alpha = 0,05$ )  
 c) Berechnen Sie das Bestimmtheitsmaß  $R^2$ ?

Tests der Zwischensubjekteffekte  
 Abhängige Variable: Absatzmenge Cola

Quelle	Quadratsumme	df
verpack	12401,422	2
platzierung	431,211	1
verpack * platzierung	753,689	2
Fehler	5628,133	84
Gesamtvariation	19214,456	89

F-Tabelle 5 % Irrtumswahrscheinlichkeit

K-J-1	J=1	J=2	J=3	J=4	J=5
1	161,0	200,00	216,00	225,00	230,00
2	18,50	19,00	19,20	19,20	19,30
3	10,10	9,55	9,28	9,12	9,01
4	7,71	6,94	6,59	6,39	6,26
5	6,61	5,79	5,41	5,19	5,05
6	5,99	5,14	4,76	4,53	4,39
7	5,59	4,74	4,35	4,12	3,97
8	5,32	4,46	4,07	3,84	3,69
9	5,12	4,26	3,86	3,63	3,48
10	4,96	4,10	3,71	3,48	3,33



d) Wie groß ist die experimentelle Wirkung der Verpackungsart „Flasche“?

Verpackungsart	Absatzmenge/Tag								$\bar{Y}_g$
	1	2	3	4	5	6	7	8	
Dose	39	40	46	45	68	65	63	59	53
Flasche	39	35	36	37	59	57	54	56	47
TetraPack	32	31	19	39	28	20	35	36	30

**Aufgabe 3 (18 Punkte)**

- a) Stellen Sie die Regressionsgleichung aus folgenden Outputs auf! Treffen Sie Aussagen über die Wichtigkeiten der unabhängigen Variablen! Hat die Variable Vertrieb einen signifikanten Einfluss auf die Absatzmenge? Begründen Sie Ihre Antwort statistisch! (Hilfe: tabellarischer Wert = 2,447)
- b) Was besagt R<sup>2</sup>? Berechnen Sie es! Wie lässt sich R<sup>2</sup> unter Umständen vergrößern?
- c) Ein alternatives Gütekriterium ist der F-Test. Welchen Vorteil hat dieser Test gegenüber R<sup>2</sup>? Berechnen Sie den empirischen F-Wert!

Koeffizienten(a)				ANOVA(b)				Deskriptive Statistiken	
Modell		Nicht standardisierte Koeffizienten		Modell		Quadratsumme	df		Standardabweichung
		B	Standardfehler						
1	(Konstante)	-6,866	673,205	1	Regression	?	3	Absatzmenge	449,22842
	Preis	9,927	38,164		Residuen	135226,900	6	Preis	1,54748
	Plakate	,655	,103		Gesamt	1816255,600	9	Plakate	544,28853
	Vertrieb	11,085	4,428					Vertrieb	13,98571

a Einflussvariablen : (Konstante), Vertrieb, Plakate, Preis  
b Abhängige Variable: Absatzmenge

a Abhängige Variable: Absatzmenge

d) Für 20 Beobachtungen wurde folgende multiple Regression geschätzt:

Variable	Koeffizienten	Standardfehler
X <sub>1</sub>	28.4	11.2
X <sub>2</sub>	4.0	1.5
X <sub>3</sub>	12.7	14.1
X <sub>4</sub>	0.84	0.76

Konstruieren Sie ein 95% Konfidenzintervall für  $\beta_3$  (t-Tabelle: t=2,13). Interpretieren Sie die Prognoseeignung der Variable x<sub>3</sub>! Berechnen Sie den t-Wert von x<sub>3</sub>! Ist die Variable x<sub>3</sub> statistisch signifikant?

**Aufgabe 4 (12 Punkte)**

- a) Erläutern Sie kurz die Grundidee der Clusteranalyse!
- b) Beschreiben Sie den Single-Linkage-Algorithmus und berechnen Sie darauf basierend die Distanzmatrix nach dem 1. Fusionierungsschritt!

	Snickers	Mars	Bounty	Milky Way
Mars	5			
Bounty	3	5		
Milky Way	55	25	43	
KitKat	74	40	58	10

c) Distanzen lassen sich generell mit folgender Formel berechnen:  $d_{k..} = \left[ \sum_{j=1}^r |x_{kj} - x_{lj}|^r \right]^{1/r}$   
Welche Distanzmaße ergeben sich für r=1 und r=2?

d) Mittels folgender Formel lässt sich die Güte einer Clusterlösung prüfen. Interpretieren Sie diese Formel! Welcher Wert ergibt sich, wenn ein Cluster in sich vollkommen homogen ist?

$$V_g = \sum_{k=1}^{K_g} \sum_{j=1}^J (x_{kj} - \bar{x}_{jg})^2$$