

Statistik für Informationsmanager

Klausur vom 21. Juli 2003

Name:

Matrikelnummer:

Erlaubte Hilfsmittel: Außer Schreibwerkzeug keine; Antworten bitte im Anschluss an die Fragen, Rückseiten können für Antworten und Entwürfe genutzt werden; Entwürfe bitte anschließend durchstreichen!

Mit der elektronischen Veröffentlichung meiner Klausurnote zusammen mit meiner Matrikelnummer bin ich einverstanden:

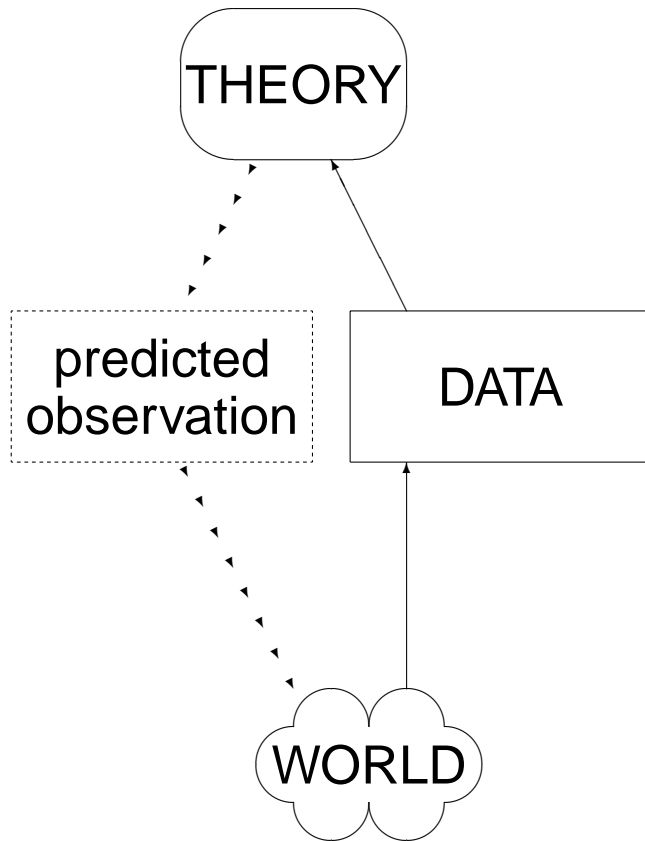
Ja Unterschrift
Nein (keine Unterschrift heißt nein)

Für die Teilaufgaben wird es folgende Punktzahlen geben:

Aufgabe	Maximal erreichbar	Sie haben erreicht
1	$2 \cdot (1 + 2)$	
2	$8 \cdot (3/8 + 3/8)$	
3	$3 \cdot (1 + 1)$	
4.1	$5 \cdot 1$	
4.2	$3 \cdot 1 + 3$	
4.3	$5 \cdot 1$	
5	$4 \cdot 2 + 1$	
6	$5 \cdot 1 + 2 \cdot 2$	
Summe	52	

1. Grundlagen

Erinnern Sie sich an die mehrfach gezeigte Abbildung;



Zeichnen Sie ein, wohin

- a. Amtliche Statistik (1P)
- b. Zeitreihenanalyse (1P)

hingehören! Geben Sie jeweils eine kurze Begründung!

- a. Amtliche Statistik gehört dorthin, weil ... (2P)

- b. Zeitreihenanalyse gehört dorthin, weil ... (2P)

2. Skalenniveaus

Geben Sie jeweils für die folgenden Variablen an, ob sie nominal, ordinal, intervall- oder ratioskaliert sind (jeweils 3/8 P) und geben Sie jeweils eine kurze Begründung (jeweils weitere 3/8 P). (Eine gute Begründung für eine unserer Ansicht nach falsche Einordnung gibt eventuell ebenfalls Punkte, keine Begründung für eine unserer Ansicht nach richtige Einordnung gibt einen Punktabzug!)

	Nominal	Ordinal	Intervall	Ratio	Begründung
ECTS-Punkte für ein Modul eines Studienganges					
Note in einem Leistungsnachweis					
EAN (Europäische Artikelnummer)					
Wechselkurs zwischen zwei Währungen					
Längengrad oder Breitengrad (gilt für beide dasselbe??)					
Listenplatz auf einer Kandidatenliste					
Marktanteil eines Produkts					
Nummer eines Fachbereichs					

3. Univariate Statistik

Berechnen Sie (jeweils falls sinnvoll!) den Modus, den Median und den Mittelwert für folgende Variablen (richtige Berechnung[en] jeweils 1P, richtige Begründung jeweils noch 1P, wenn bei einer Teilaufgabe richtigerweise alle drei Werte berechnet worden sind, bedarf es keiner Begründung, und es gibt trotzdem 2P):

1. Gesamtpunktzahlen von 20 Studierenden bei einer Klausur:

118.2	80.4	133.8	92.8	102.0	90.8	68.4	58.4	55.1	81.9
120.6	41.2	118.6	83.8	105.2	131.5	99.3	83.3	139.0	145.2
Mittelwert:									
Median									
Modus									
gegebenenfalls Begründung für nicht berechnete Parameter:									

2. Gesamtnoten der gleichen Studierenden in der gleichen Klausur:

3.7	5.0	3.0	5.0	4.0	5.0	5.0	5.0	5.0	5.0
3.3	5.0	3.7	5.0	4.0	3.0	5.0	5.0	2.7	2.7
Mittelwert:									
Median									
Modus									
gegebenenfalls Begründung für nicht berechnete Parameter:									

3. Geschlecht von zwanzig Studierenden:

W	M	M	M	W	W	M	M	W	W
M	M	W	W	M	M	M	W	M	W
Mittelwert:									
Median									
Modus									
gegebenenfalls Begründung für nicht berechnete Parameter:									

4. Bivariate Statistik

4.1 Nominale Variable

Bei der Bundestagswahl 2002 war in 125 Wahlkreisen die CDU oder die CSU nach Erststimmen stärker als die SPD, in den übrigen 174 Wahlkreisen war die SPD stärker als die CDU bzw. als die CSU. 1998 hatte die SPD nach diesem Kriterium noch besser abgeschnitten. Untersucht man, welche Partei vor der anderen lag, getrennt nach größeren Regionen in Deutschland, so ergeben sich die nachstehenden beiden Kreuztabellen:

Wahlkreissieger 2002 * NSO Nord/Süd/Ost Kreuztabelle					
		NSO Nord/Süd/Ost			Gesamt
		1.00 Nord	2.00 Süd (BY, BW)	3.00 Ost	
CDU/CSU > SPD	Anzahl	36	73	16	125
	% von NSO Nord/Süd/Ost	23.7 %	90.1 %	24.2 %	41.8 %
SPD > CDU/CSU	Anzahl	116	8	50	174
	% von NSO Nord/Süd/Ost	76.3 %	9.9 %	75.8 %	58.2 %
Gesamt	Anzahl	152	81	66	299
	% von NSO Nord/Süd/Ost	100.0%	100.0%	100.0%	100.0%□

Wahlkreissieger 1998 * NSO Nord/Süd/Ost Kreuztabelle					
		NSO Nord/Süd/Ost			Gesamt
		1.00 Nord	2.00 Süd (BY, BW)	3.00 Ost	
CDU/CSU > SPD	Anzahl	31	64	14	109
	% von NSO Nord/Süd/Ost	20.4 %	79.0 %	21.2 %	36.5 %
SPD > CDU/CSU	Anzahl	121	17	52	190
	% von NSO Nord/Süd/Ost	79.6 %	21.0 %	78.8 %	63.5 %
Gesamt	Anzahl	152	81	66	299
	% von NSO Nord/Süd/Ost	100.0%	100.0%	100.0%	100.0%□

Alle Zusammenhangsmaße sind signifikant von Null verschieden. Sie lauten im einzelnen:

		Wert 2002	Wert 1998
Lambda	Symmetrisch	0.375	0.313
	Siegerpartei abhängig	0.520	0.431
	Nord/Süd/Ost abhängig	0.252	0.224
Goodman-und-Kruskal-Tau	Siegerpartei abhängig	0.357	0.291
	Nord/Süd/Ost abhängig	0.180	0.146
Unsicherheitskoeffizient	Symmetrisch	0.224	0.173
	Siegerpartei abhängig	0.282	0.222
	Nord/Süd/Ost abhängig	0.186	0.141

Beurteilen Sie an Hand der Kreuztabelle die Verteilung der Parteichancen in den verschiedenen Teilen Deutschlands (1P).

Welche Angaben aus der Kreuztabelle — oder welche Koeffizienten — haben Sie dazu am besten benutzen können? (1P)

Bitte bilden Sie einen Satz, mit dem Sie die Bedeutung der Zahl 0.520 (Lambda mit Siegerpartei 2002 als abhängiger Variable) erläutern (1P).

Wie beurteilen Sie, dass die Koeffizienten für 2002 durchgängig größer sind als die für 1998? (1P)

Was können Sie daraus schließen, dass die Koeffizienten für "Siegerpartei abhängig" durchgängig größer sind als die für "Nord/Süd/Ost abhängig"? (1P)

4.2 Ordinale Variable

Aus der Konsumentenbefragung 1995 kann man eine Antwort auf die Frage ableiten, ob Haushalte um so preisbewusster sind, je mehr Kinder sie haben. Aus der Kreuztabelle der Variablen Preisbewusstsein (mit den Ausprägungen 1=nicht preisbewusst, 2=wenig preisbewusst, 3=überwiegend preisbewusst und 4=sehr preisbewusst) und der Variablen Anzahl der Kinder (0 bis 9) haben sich die folgenden Zusammenhangsmaße ergeben:

		Alte Länder	Neue Länder
Somers-d	Symmetrisch	.143	.126
	V19 Anzahl der Kinder unter 18 Jahren abhängig	.120	.106
	V20 Preisbewusstsein abhängig	.178	.154
Kendall-Tau-b		.146	.128
Kendall-Tau-c		.109	.103
Gamma		.268	.214
Anzahl der gültigen Fälle		3711	927

Fragen:

Warum ist Gamma größer als die anderen Koeffizienten? (1P)

Lässt sich aus dem Unterschied zwischen den beiden "asymmetrischen" Somers' D eine Schlussfolgerung ziehen? (0.5P) Wenn Ja, welche, wenn nein, warum nicht? (gute Begründungen für Antworten, die wir für falsch halten, geben weitere 0.5 Punkte, fehlende Begründungen auch für richtige Antworten geben zwar keinen Punktabzug, aber auch keinen Zuschlag!)

Wäre die Verwendung eines nominalen Maßes (Phi, Lambda, ...) hier auch angemessen? (0.5P) Wenn Ja, warum, wenn nein, warum nicht? (wie oben)

Beurteilen Sie die Unterschiede in den Koeffizienten zwischen den neuen und den alten Bundesländern (3 P).

4.3 Metrische Variable

In den letzten beiden Klausuren im Vordiplom Wirtschaftsinformatik, die aus je vier Teilklausuren bestanden haben, habe ich folgende Korrelationen (Produkt-Moment-Korrelationskoeffizienten) zwischen den Punktzahlen für die Teilklausuren, der Gesamtpunktzahl und der Note gefunden:

	Emp. Grdl.	O+M	BWL	KLR	Punkte ges.	Note
Punktzahl in der Teilklausur Empirische Grundlagen	1.000					
Punktzahl in der Teilklausur Organisation und Management	0.318	1.000				
Punktzahl in der Teilklausur Einführung in die Betriebswirtschaftslehre	0.304	0.438	1.000			
Punktzahl in der Teilklausur Kosten- und Leistungsrechnung	0.437	0.315	0.588	1.000		
Punktzahl insgesamt	0.741	0.652	0.768	0.796	1.000	
Note für die gesamte Klausur	-0.735	-0.574	-0.706	-0.677	-0.915	1.000

Fragen:

Ist die Verwendung des Produkt-Moment-Korrelationskoeffizienten hier überhaupt gerechtfertigt? (0.5P) Wenn Ja, warum, wenn nein, warum nicht? (gute Begründungen für Antworten, die wir für falsch halten, geben weitere 0.5 Punkte, fehlende Begründungen auch für richtige Antworten geben zwar keinen Punktabzug, aber auch keinen Zuschlag!)

Beurteilen Sie an Hand der Korrelationskoeffizienten für die vier Teilklausuren die "Verwandtschaftsverhältnisse" zwischen den vier Teilgebieten und prüfen Sie, ob diese Beurteilung mit Ihrem Vorverständnis übereinstimmt! (1P)

Wie erklären Sie sich, dass die Korrelationen der Gesamtpunktzahl mit den Einzelpunktzahl am höchsten sind? (1P)

Wie erklären Sie sich, dass die Korrelationen der Note mit allen anderen Variablen negativ sind? (1P)

Warum konnte ich die Zahlen in der oberen Dreiecksmatrix der Tabelle oben einfach weglassen? (1P)

5. Multivariate Statistik / Multiple Regression

Bei der Vordiplomsklausur Wirtschaftsinformatik vor einigen Wochen wurden die vier Teilklausuren getrennt korrigiert; für jede Teilklausur gab es maximal 50 Punkte — das ergab die vier in die nachstehende Tabelle als unabhängig aufgenommenen Variablen „Organisation und Management“, „Einführung in die BWL“, „Empirische Grundlagen“ und „Kosten- und Leistungsrechnung“. Abhängige Variable ist die Gesamtnote.

Koeffizienten					
	Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
	b	Standardfehler	β		
(Konstante)	6.839	.232		29.460	.000
Punktzahl in der Teilklausur Empirische Grundlagen (x_1)	-.037	.005	-.477	-7.290	.000
Punktzahl in der Teilklausur Organisation und Management (x_2)	-.024	.008	-.205	-3.122	.003
Punktzahl in der Teilklausur Einführung in die Betriebswirtschaftslehre (x_3)	-.035	.007	-.358	-4.751	.000
Punktzahl in der Teilklausur Kosten- und Leistungsrechnung (x_4)	-.018	.007	-.193	-2.560	.014
Abhängige Variable: NOTE (y)					

Wie lässt sich nach dem Ergebnis der Regression die Gesamtnote aus den einzelnen Punktzahlen der Teilklausuren berechnen? Schreiben Sie dazu — unter Verwendung der Symbole y , x_1 , x_2 , x_3 und x_4 sowie der Zahlenwerte von b_0 , b_1 , b_2 , b_3 und b_4 — die Regressionsgleichung auf (2P).

Berechnen Sie mit Hilfe der soeben aufgeschriebenen Regressionsgleichung die geschätzte Gesamtnote von jemand, der in allen vier Teilklausuren 25 Punkte erreicht hat (2P).

Welche Teilklausur(en) hat/haben sich am stärksten auf die Gesamtnote ausgewirkt? (2P)

Kann man bei der kleinen Stichprobe (bei lediglich 50 Teilnehmern) davon ausgehen, dass die Ergebnisse über den Zusammenhang zwischen den Punktzahlen der Teilklausuren und der Gesamtnote auf die Grundgesamtheit übertragbar ist? (2P)

Beschreiben Sie die Grundgesamtheit, aus der diese Stichprobe entnommen ist (1P).

6. Multivariate Statistik / Faktorenanalyse

Ein Teil der Fragen aus dem Verbraucherpanel bezog sich auf Einkaufs- und Ernährungsgewohnheiten. Aus diesen 41 Fragen ließ sich die Faktorladungsmatrix auf der nächsten Seite gewinnen. Koeffizienten kleiner als 0.2 sind zur Erleichterung des Auffindens einer Einfachstruktur weggelassen. Interpretieren Sie die vier Faktoren, indem Sie ihnen möglichst prägnante Namen geben. Begründen Sie Ihre Namensgebung.

Faktor 1 beschreibt die latente Eigenschaft ... (1 P)

Faktor 2 ... (1 P)

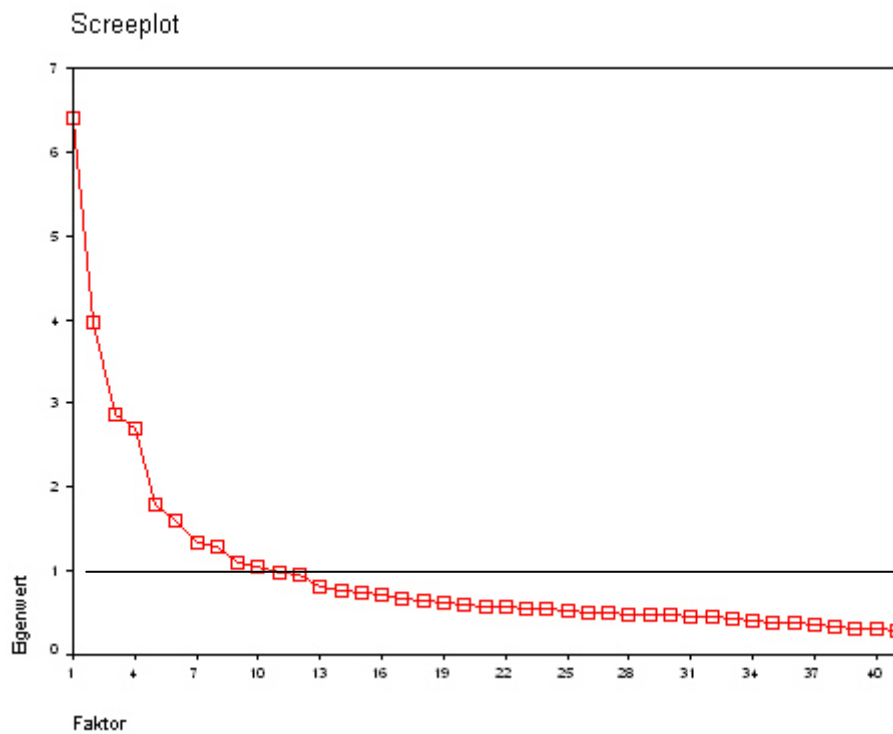
Faktor 3 ... (1 P)

Faktor 4 ... (1 P)

Faktor 5 ... (1 P)

	Faktoren				
	1	2	3	4	5
V78 Achten auf fettarme Ernährung	0.717				
V79 Achten auf schonende reizarme Kost	0.713				
V80 schlankheitsbewusste Ernährung	0.659				
V92 Rücksichtnahme auf Gesundheit bei Ernährung	0.658				
V66 Achten auf Figur	0.655				
V96 Achten auf Kalorien	0.642				
V102 Vermeidung von gesundheitsschädlicher Ernährung	0.611		0.249		
V81 Kauf von zusatzstofffreien Lebensmitteln	0.557		0.382		-0.215
V72 vollwertige Ernährung	0.541				
V93 Information über umweltbelastete Lebensmittel- kein Kauf mehr	0.495		0.312		
V99 häufige Verwendung von Getreidekörnern	0.419			-0.373	
V65 Ablehnung von Konservierungsstoffen in Produkten	0.410		0.374		
V90 regelmäßige Verwendung von Vitamin- und Mineralstoffpräparaten	0.359	0.249			0.333
V70 Vorzug von edlen Speisen und Getränken		0.703			
V73 Delikatessen		0.687			
V98 sehr anspruchsvoll beim Essen und Trinken		0.663	0.215		
V86 Verwöhnen mit gutem Essen		0.651			
V82 Vorliebe für ausgefallene Speisen und Gerichte		0.646	-0.201		-0.263
V94 Spaß am Ausprobieren fremdländischer Spezialitäten		0.581	-0.414	-0.252	
V67 Begeisterung für ausländische Spezialitäten		0.511	-0.378	-0.234	
V103 bei Wahl Kauf von deutschen Lebensmitteln	0.273		0.642		
V83 beste Qualität der deutschen Nahrungsmittel	0.235		0.631		
V100 Lebensmittel bekannter Marken besser als die unbekannter Marken		0.253	0.615		
V68 Gleichgültigkeit gegenüber Herkunftsland der Lebensmittel		0.228	-0.596		
V89 kein Vertrauen zu Nahrungsmitteln ohne Markenbezeichnung			0.596		
V75 Preis, nicht Marke, ist für Kauf entscheidend		-0.215	-0.426	0.214	
V71 Vorzug von Hausmannskost		-0.239	0.208	0.619	
V91 zuviel Wirbel um Ernährung				0.589	
V87 Vorliebe für herzhaft, deftige Mahlzeiten				0.549	
V77 Überschätzung des Einflusses der Ernährung auf die Gesundheit				0.536	
V105 Kein Essen und Trinken mehr bei Annahme aller Ernährungstips				0.531	
V97 Kochen von altbewährten Gerichten		-0.264	0.351	0.506	
V85 vegetarische Ernährung	0.327			-0.407	
V76 Multivitaminsäfte als wichtige Ergänzung	0.270			0.280	0.259
V95 Vorzug von Kochen einfacher Gerichte		-0.231			0.706
V84 Vorliebe für schnell kochbare Gerichte		-0.222			0.701
V74 kein Kochen ohne Fertigprodukte					0.593
V69 viel Zeit für Kochen	0.217	0.336		0.353	-0.553
V101 Konserven genauso gut wie Frisches				0.349	0.497
V88 Kauf von ausschließlich frischen Lebensmittel	0.336		0.355		-0.401
V104 bei normaler Kost keine Ergänzungspräparate notwendig				0.231	-0.293

Hier finden Sie den Scree-Plot zu der vorstehenden Auswertung:



Halten Sie danach die Extraktion von fünf Faktoren noch für gerechtfertigt? (Ja / Nein / Kommt darauf an)
Wenn ja: warum? (2 P)

Wenn nein: warum nicht? Wie viele hätte man statt dessen extrahieren sollen? (2 P)

(Antworten auf beide Unterfragen werden gewertet! Wenn Sie dezidiert der Auffassung sind, dass nicht drei, sondern zwei oder vier oder fünf oder sechs oder sieben oder acht ... Faktoren hätten extrahiert werden sollen, müssen Sie auf „wenn nein: warum nicht?“ besonders ausführlich antworten; ebenso müssen Sie, wenn Sie dezidiert der Auffassung sind, dass drei Faktoren die richtige Wahl waren, auf „wenn ja, warum?“ besonders ausführlich antworten.)