

# Statistik für Informationsmanager

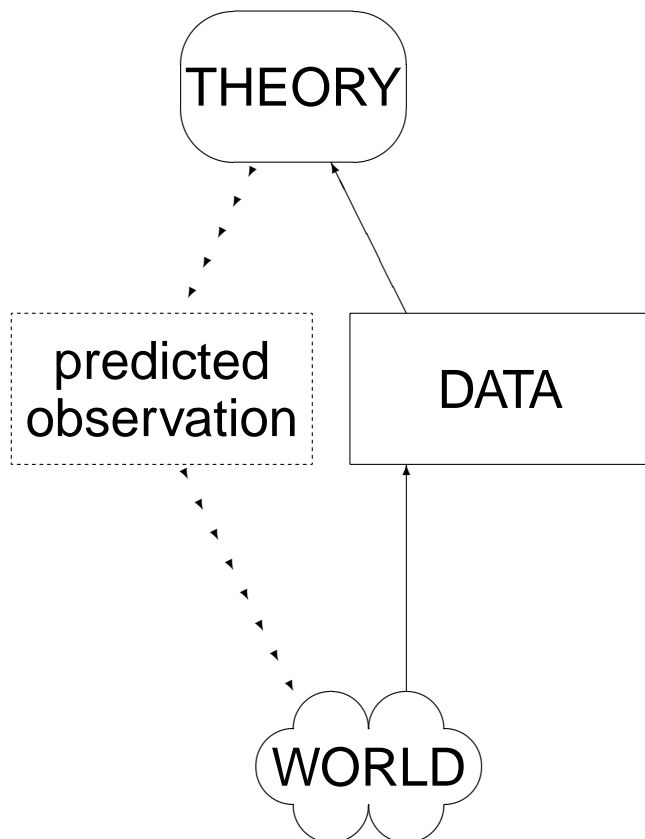
## Klausur vom 9. Juli 2002

Name:

Matrikelnummer:

### 1. Grundlagen

Erinnern Sie sich an die mehrfach gezeigte Abbildung;



Zeichnen Sie ein, wohin

- a. Clusteranalyse
- b. Regressionsanalyse

hingehören! Geben Sie jeweils eine kurze Begründung!

- a. Clusteranalyse gehört dorthin, weil ...
  
- b. Regressionsanalyse gehört dorthin, weil ...

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)

## 2. Skalenniveaus

Geben Sie jeweils für die folgenden Variablen an, ob sie nominal, ordinal, intervall- oder ratioskaliert sind und geben Sie jeweils eine kurze Begründung. (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
Stromverbrauch					
Höhe über dem Meer					
Besoldungs-/Vergütungsgruppe					
Platz eines Landes in der PISA-Studie					
Telefonnummer					
Bewertung eines Aspekts einer Lehrveranstaltung auf einer Skala von 1= „stimme voll zu“ bis 5= „stimme überhaupt nicht zu“					
Fußballweltmeister					
Zinssatz					

## 3. Univariate Statistik

Berechnen Sie (jeweils falls sinnvoll!) den Modus, den Median und den Mittelwert für folgende Variablen:

1. Preis von 1 kg Erdbeeren in zwanzig verschiedenen Geschäften:

0.99	0.99	0.99	1.19	1.19	1.29	1.29	1.29	1.29	1.29
1.29	1.39	1.39	1.39	1.49	1.49	1.59	1.59	1.79	1.99
Mittelwert:									
Median									
Modus									
gegebenenfalls Begründung für nicht berechnete Parameter:									

2. Vergütungsgruppe von 20 Angestellten der Universität:

Ib	IIa	IIa	IVa	VIb	VII	VIII	III	IIa	IVb
Vc	VII	VII	VII	VIII	IX	VIb	Vib	IIa	VII
Mittelwert:									
Median									
Modus									
gegebenenfalls Begründung für nicht berechnete Parameter:									

3. Postleitzahl von 10 Personen:

53111	53456	54222	56068	56075	56016	56076	58022	63222	64747
Mittelwert:									
Median									
Modus									
gegebenenfalls Begründung für nicht berechnete Parameter:									

## 4. Bivariate Statistik

In der ZUMA-Verbraucherumfrage von 1995 wird unter anderem das Vorhandensein von Gärten in den befragten Haushalten untersucht. Aufgeschlüsselt nach alten und neuen Bundesländern und nach Miet- bzw. Eigentumsverhältnissen ergibt sich folgendes Bild:

		kein Garten	Garten am Haus	Garten nicht am Haus	Total
<b>Altes Bundesgebiet</b>					
eigenes Haus	abs.	42	1310	53	<b>1405</b>
	%	3.0	93.2	3.8	<b>100.0</b>
Eigentumswohnung	abs.	144	65	37	<b>246</b>
	%	58.5	26.4	15.0	<b>100.0</b>
Hauptmieter	abs.	826	329	134	<b>1289</b>
	%	64.1	25.5	10.4	<b>100.0</b>
Untermieter	abs.	40	29	10	<b>79</b>
	%	50.6	36.7	12.7	<b>100.0</b>
<b>Total</b>	<b>abs.</b>	<b>1052</b>	<b>1733</b>	<b>234</b>	<b>3019</b>
	<b>%</b>	<b>34.8</b>	<b>57.4</b>	<b>7.8</b>	<b>100.0</b>
<b>Neue Bundesländer</b>					
eigenes Haus	abs.	9	156	15	<b>180</b>
	%	5.0	86.7	8.3	<b>100.0</b>
Eigentumswohnung	abs.	4	2	2	<b>8</b>
	%	50.0	25.0	25.0	<b>100.0</b>
Hauptmieter	abs.	294	84	228	<b>606</b>
	%	48.5	13.9	37.6	<b>100.0</b>
Untermieter	abs.	20	8	15	<b>43</b>
	%	46.5	18.6	34.9	<b>100.0</b>
<b>Total</b>	<b>abs.</b>	<b>327</b>	<b>250</b>	<b>260</b>	<b>837</b>
	<b>%</b>	<b>39.1</b>	<b>29.9</b>	<b>31.1</b>	<b>100.0</b>

Sie haben folgende Statistiken für den Zusammenhang zwischen Eigentumsverhältnissen und dem Gartenbesitz herausbekommen:

	N	Chi-Quadrat	Signifikanz	Phi	Cramer's V	Kontingenzkoeff.
Alte Bundesländer	3019	1419.813	>0.0005	0.686	0.485	0.566
Neue Bundesländer	837	355.092	>0.0005	0.651	0.461	0.546

... und außerdem folgende PRE-Maße:

	N	lambda	Goodman's and Kruskal's tau	Uncertainty coefficient
Alte Bundesländer	3019	0.456	0.373	0.309
Neue Bundesländer	837	0.288	0.204	0.189

Beurteilen Sie:

1. die Unterschiede zwischen den alten und den neuen Bundesländern,
2. die unterschiedliche Stärke des Zusammenhangs zwischen Hausbesitz und Gartenbesitz in den alten und den neuen Bundesländern.

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

Warum ist Chi-Quadrat für die Beziehung zwischen Miet-/Eigentumsverhältnissen und Gartenbesitz im Falle der neuen Bundesländer um so viel niedriger als im Falle der alten Bundesländer, obwohl die beiden Koeffizienten (Phi, Cramer's V, Kontingenzkoeffizient) recht nahe beieinander liegen?

Eine weitere Auswertung desselben Datensatzes fragt nach dem Zusammenhang zwischen der Haushaltsgröße (Anzahl der Personen im Haushalt) einerseits und nach einigen Aussagen zum Konsumverhalten andererseits. Die letzteren waren fünfstufig kodiert (1: stimme überhaupt nicht zu ... 5: stimme voll und ganz zu). Sie haben folgende Koeffizienten erhalten:

Land	Nominal / Chi-Quadrat		Nominal PRE		Ordinal	
	Phi	Cramer's V	Unsicherheitskoeffizient	Lambda	Gamma	Somers' D
Zusammenhang Haushaltsgröße / „neue Produkte sind vielleicht teurer, aber deshalb nicht besser“						
Alte Bundesländer	0.104	0.052	0.004	0.000	+0.008	+0.006
Neue Bundesländer	0.216	0.108	0.020	0.060	-0.168	-0.117
Zusammenhang Haushaltsgröße / „ich genieße lieber das Leben mit meinem Geld, anstatt es zu sparen“						
Alte Bundesländer	0.163	0.081	0.010	0.020	-0.103	-0.078
Neue Bundesländer	0.303	0.151	0.029	0.058	-0.204	-0.154
Zusammenhang Haushaltsgröße / „ich bevorzuge Markenartikel“						
Alte Bundesländer	0.189	0.094	0.013	0.001	-0.165	-0.123
Neue Bundesländer	0.267	0.134	0.027	0.002	-0.223	-0.169

Fragen:

Ist die Verwendung der Maße Gamma und Somers' D für ordinale Variable hier angemessen? Wenn Ja, warum, wenn nein, warum nicht? (gute Begründungen für Antworten, die wir für falsch halten, geben Punkte, fehlende Begründungen auch für richtige Antworten geben einen Punktabzug!)

Ist die Verwendung der nominalen Maße hier überhaupt angemessen? Wenn Ja, warum, wenn nein, warum nicht? (gute Begründungen für Antworten, die wir für falsch halten, geben Punkte, fehlende Begründungen auch für richtige Antworten geben einen Punktabzug!)

Was für Zusatzinformationen brauchen Sie, um zu verstehen, warum in den alten Bundesländern einmal Lambda=0 gemessen wurde? (Möglicherweise brauchen Sie auch keine Zusatzinformationen: Warum ist lambda in den alten Bundesländern gleich Null, obwohl die anderen Maße für nominalen Zusammenhang größer als Null sind?)

Interpretieren Sie die Unterschiede zwischen den alten und neuen Bundesländern! (Hinweis: Ihre Antwort sollte folgende Phrase enthalten: „In den neuen (alten) Bundesländern ist der Zusammenhang ‚Je größer der Haushalt, desto größer (geringer) die Zustimmung zu der Aussage ...‘ geringer (stärker) ausgeprägt als in den alten (neuen) Bundesländern.“)

Interpretieren Sie die Unterschiede zwischen den drei Zusammenhängen in den alten Bundesländern! Entscheiden Sie sich dabei für ein Maß für die Stärke des Zusammenhangs und begründen Sie Ihre Wahl.

Die Verbraucherstudie fragte auch nach Einstellungen zur Lebensführung. Drei Items dazu lauteten:

- A. ich kaufe viele Artikel, die anderen noch unbekannt sind
- B. ich suche nach neuen Produkten meinen Ansprüchen entsprechend
- C. mir liegt immer daran, neue Produkte vor den Bekannten zu besitzen

Die Antwortmöglichkeiten waren jeweils: stimme überhaupt nicht zu, stimme eher nicht zu, bin unentschieden, stimme etwas zu, stimme voll und ganz zu. Die Produkt-Moment-Korrelationskoeffizienten zwischen diesen drei Variablen lauten:

.	A	B	C
A	1.000	+0.391	+0.667
B	+0.391	1.000	+0.464
C	+0.667	+0.464	1.000

Fragen:

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

Wie beurteilen Sie, dass das Variablenpaar A/C relativ hoch miteinander korreliert, während die Paare A/B und B/C nur eine deutlich geringere Korrelation aufweisen?

Welche kausalen Zusammenhänge zwischen den drei Variablen können Sie vermuten? (Lassen sich die Statements A, B und C sinnvoll durch „denn“ oder „weil“ miteinander verknüpfen?) Wenn Sie keinen kausalen Zusammenhang vermuten wollen: Haben Sie dafür Gründe?

## 5. Multivariate Statistik / Multiple Regression

Das International Social Survey Project (ISSP 1992) bietet folgende Fragen, die darauf hin ausgewertet werden können, von welchen anderen Einstellungen die Auffassung gesteuert wird, dass im eigenen Land die Einkommensdifferenzen zu hoch seien (bei sämtlichen Variablen gemessen auf einer Skala von 1: stimme voll zu bis 5: stimme überhaupt nicht zu):

V17 So wie die Verhältnisse in <Land des Befragten> beschaffen sind, haben Leute wie ich und meine Familie eine gute Chance, ihren Lebensstandard zu erhöhen.

V19 Keiner ist dazu bereit, mehr Verantwortung zu übernehmen, wenn er dafür nicht extra bezahlt wird.

V20 Arbeiter bemühen sich nur dann um zusätzliche Fertigkeiten und Qualifikationen, wenn sie damit auch mehr verdienen.

V21 Die sozialen Unterschiede bleiben bestehen, weil sie den Reichen und Mächtigen nützen.

V22 Keiner wird jahrelang studieren, um Rechtsanwalt oder Arzt zu werden, wenn er nicht auch erwartet, viel mehr zu verdienen als ein einfacher Arbeiter.

V23 Große Einkommensunterschiede sind für den Wohlstand <des Landes des Befragten> notwendig.

V24 Der Wirtschaft hohe Gewinne zu erlauben, ist der beste Weg, den Lebensstandard zu erhöhen.

V25 Die Einkommensunterschiede bleiben bestehen, weil sich die einfachen Bürger nicht zusammenschließen, um diese zu beseitigen.

Die Auswertung mit einer multiplen Regression (Begeisterung der Befragten für ausländische Lebensmittel als abhängige, die sieben oben genannten Einstellungen als unabhängige Variable) erbringt u.a. folgende Ergebnisse für die Bundesländer (R und  $R^2$ ):

Land	R	$R^2$
Schweden	0.667	0.445
Norwegen	0.580	0.337
Neuseeland	0.557	0.310
Großbritannien	0.550	0.302
Australien	0.549	0.301
Canada	0.492	0.242
Deutschland (alte Bundesländer)	0.471	0.222
USA	0.462	0.213
Polen	0.454	0.206
Slowenien	0.427	0.183
Österreich	0.417	0.174
Tschechoslowakei	0.384	0.148
Deutschland (neue Bundesländer)	0.380	0.144
Ungarn	0.345	0.119
Russland	0.316	0.100
Bulgarien	0.290	0.084
Italien	0.271	0.074
Philippinen	0.211	0.045

Interpretieren Sie dieses Ergebnis!

Die Einzelauswertung für drei dieser Länder zeigt folgende standardisierten Regressionskoeffizienten:

	D-W	I	S
V17 So wie die Verhältnisse ... sind, haben Leute wie ich ... eine gute Chance, ihren Lebensstandard zu erhöhen.	-0.107	-0.033	-0.029
V19 Keiner ist ... bereit, mehr Verantwortung zu übernehmen, wenn er ... nicht extra bezahlt wird.	0.055	-0.018	0.025
V20 Arbeiter bemühen sich nur dann um zusätzliche Fertigkeiten ..., wenn sie damit auch mehr verdienen.	0.011	0.091	0.002
V21 Die sozialen Unterschiede bleiben bestehen, weil sie den Reichen ... nützen.	0.261	0.168	0.369
V22 Keiner wird jahrelang studieren, ..., wenn er nicht auch erwartet, viel mehr zu verdienen ...	-0.026	0.025	0.088
V23 Große Einkommensunterschiede sind für den Wohlstand ... notwendig.	-0.114	-0.114	-0.314
V24 Der Wirtschaft ... Gewinne ... zu erlauben, ist der beste Weg, den Lebensstandard zu erhöhen.	-0.046	-0.045	-0.093
V25 Die Einkommensunterschiede bleiben bestehen, weil sich die ... Bürger nicht zusammenschließen, um diese zu beseitigen.	0.218	0.099	0.133

Vergleichen Sie den Einfluss der acht unabhängigen Variablen auf die abhängige Variable „Einschätzung der Einkommensdifferenzen im eigenen Land“ in den drei Ländern. Finden Sie Interpretationen für die gefundenen Unterschiede.

## 5. Multivariate Statistik / Faktorenanalyse

Die Befragung von 99 Studierenden in einer Vorlesung bezog sich u.a. auf die folgenden 17 Items zur Bewertung von Aufbau und Struktur der Veranstaltung. Es wurden vier Faktoren extrahiert und rotiert, es ergab sich folgende Ladungsmatrix:

	Faktor			
	1	2	3	4
B2V04 Semester strukturiert	0.708			
B2V03 Zielsetzung der Sitzung bekannt	0.703			
B2V05 Sitzung strukturiert	0.594	0.459		
B2V10 Methodenvielfalt	0.521			
B2V11 hochschuldidaktisch Neues	0.518			
B2V01 Semesterziele klar	0.494			
B2V19 aktueller Forschungsstand	0.487		0.369	
B2V13 Dozent gut vorbereitet		0.718		
B2V12 zu wenig Medien		-0.664		
B2V14 Voraussetzungen der Studierenden berücksichtigt		0.585		
B2V16 Visualisierungen ertragreich		0.417	0.414	-0.378
B2V17 Folien im Internet zugänglich			0.738	
B2V15 zus. Arbeitsmaterial		0.309	0.702	
B2V18 Semesterapparat			0.568	
B2V08 zuwenig Diskussionszeit				0.750
B2V20 hohes Anspruchsniveau			0.417	0.607
B2V02 ausführliche Ankündigung	0.361			-0.390

Koeffizienten kleiner als 0.3 sind zur Erleichterung des Auffindens einer Einfachstruktur weggelassen; außerdem sind die Variablen nach der absoluten Größe ihrer größten Ladungskoeffizienten sortiert.. Interpretieren Sie die vier Faktoren, indem Sie ihnen möglichst prägnante Namen geben. Begründen Sie Ihre Namensgebung.

Faktor 1 beschreibt die Beurteilung der Veranstaltung durch die Studierenden unter dem Gesichtspunkt ...

Faktor 2 ...

Faktor 3 ...

Faktor 4 ...

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



Halten Sie danach die Extraktion von vier Faktoren noch für gerechtfertigt? (Ja / Nein / Kommt darauf an)

Wenn ja: warum?

Wenn nein: warum nicht? Wie viele hätte man statt dessen extrahieren sollen?

(Antworten auf beide Unterfragen werden gewertet! Wenn Sie dezidiert der Auffassung sind, dass nicht vier, sondern zwei oder drei 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 vier Faktoren die richtige Wahl waren, auf „wenn ja, warum?“ besonders ausführlich antworten.)