

Statistik für Informationsmanager

Nachklausur

Name:

Matrikelnummer:

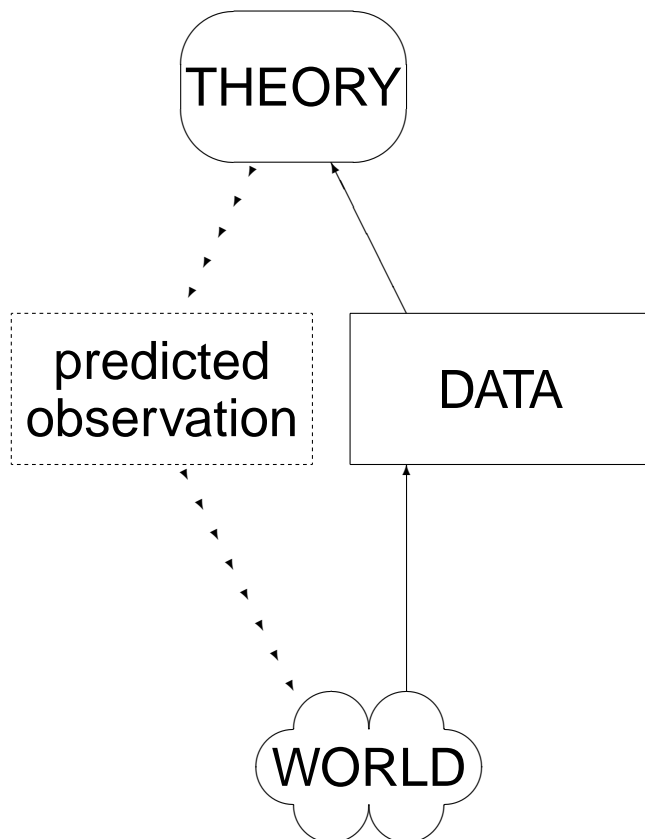
Erlaubte Hilfsmittel: Außer Schreibwerkzeug keine; Antworten bitte im Anschluss an die Fragen, Rückseiten können für Antworten und Entwürfe genutzt werden; Entwürfe bitte anschließend durchstreichen!

1. Grundlagen

Erinnern Sie sich an die mehrfach gezeigte Abbildung;

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

Ja Unterschrift
Nein (keine Unterschrift heißt nein)



Zeichnen Sie ein, wohin

- a. Zeitreihenanalyse
- b. Feldexperimente

hingehören! Geben Sie jeweils eine kurze Begründung!

a. Zeitreihenanalyse gehört dorthin, weil ...

b. Feldexperimente gehören dorthin, weil ...

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
Alter					
DAX					
Höchster Bildungsabschluss					
Diplomnote					
ISBN					
Bewertung der Vorliebe für Politiker auf einer Skala von -5 bis +5					
Bevorzugte Partei					
Rangordnung von sechs Parteien					

3. Univariate Statistik

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

1. Alter der Schüler in einer Schulklasse:

13	14	16	14	13	14	15	16	13	14
14	15	16	14	13	14	13	15	16	15
Mittelwert:									
Median									
Modus									
gegebenenfalls Begründung für nicht berechnete Parameter:									

2. Höchster Bildungsabschluss von 20 Personen (1: ohne Abschluss, 2: Hauptschule, 3: Realschule, 4: Abitur, 5: Hochschulabschluss:

5	5	5	4	4	4	4	4	4	3
3	3	3	2	2	2	2	1	1	1
Mittelwert:									
Median									
Modus									
gegebenenfalls Begründung für nicht berechnete Parameter:									

3. Hauptfach von 10 Personen:

79	79	79	79	79	845	845	845	52	52
Mittelwert:									
Median									
Modus									
gegebenenfalls Begründung für nicht berechnete Parameter:									

4. Bivariate Statistik

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

		Ortsgrösse					Gesamt
		unter 5000	5000 bis unter 20000	20000 bis unter 100000	100000 bis unter 500000	500000 und mehr	
altes Bundesgebiet							
PKW nicht vorhanden	Anzahl	44	123	163	152	185	667
	%	9.4%	15.4%	19.3%	25.5%	30.4%	20.1%
PKW vorhanden	Anzahl	182	344	376	271	309	1482
	%	38.8%	42.9%	44.6%	45.5%	50.7%	44.7%
PKW Zweitwagen vorhanden	Anzahl	207	296	263	155	106	1027
	%	44.1%	37.0%	31.2%	26.1%	17.4%	31.0%
mehr als zwei PKW	Anzahl	36	38	41	17	9	141
	%	7.7%	4.7%	4.9%	2.9%	1.5%	4.3%
insgesamt	Anzahl	469	801	843	595	609	3317
	%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
neue Bundesländer							
PKW nicht vorhanden	Anzahl	40	44	102	44	41	271
	%	18.7%	36.1%	36.2%	34.1%	33.9%	31.2%
PKW vorhanden	Anzahl	90	39	126	58	56	369
	%	42.1%	32.0%	44.7%	45.0%	46.3%	42.5%
PKW Zweitwagen vorhanden	Anzahl	73	35	46	24	23	201
	%	34.1%	28.7%	16.3%	18.6%	19.0%	23.2%
mehr als zwei PKW	Anzahl	11	4	8	3	1	27
	%	5.1%	3.3%	2.8%	2.3%	.8%	3.1%
insgesamt	Anzahl	214	122	282	129	121	868
	%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Sie haben folgende Statistiken für den Zusammenhang zwischen Ortsgröße und dem PKW-Besitz herausbekommen:

	N	Chi-Quadrat	Signifikanz	Cramer's V
Alte Bundesländer	3317	190.298	>0.0005	0.138
Neue Bundesländer	868	44.270	>0.0005	0.130

Beurteilen Sie:

1. die Unterschiede zwischen den alten und den neuen Bundesländern,
2. die Unterschiede nach der Ortsgröße in den alten und den neuen Bundesländern.

Welche Angaben aus der Kreuztabelle haben Sie dazu am besten benutzen können?

Warum ist die Signifikanz der Beziehung zwischen Ortsgröße und PKW-Besitz im Falle der neuen Bundesländer um so viel besser als im Falle der alten Bundesländer, obwohl beide Kontingenzkoeffizienten (Cramer's V) recht nahe beieinander liegen?

Eine weitere Auswertung desselben Datensatzes fragt nach dem Zusammenhang zwischen der Ortsgröße einerseits und nach der Begeisterung für ausländische Spezialitäten bzw. nach der Gleichgültigkeit gegenüber dem Herkunftsland der Lebensmittel andererseits. Die Ortsgröße ist wie oben kodiert, die Begeisterung für ausländische Spezialitäten und die Gleichgültigkeit gegenüber dem Herkunftsland der Lebensmittel war 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
Zusammenhang Ortsgröße / Begeisterung					
Alte Bundesländer	0.070	0.035	0.002	0.000	+0.054
Neue Bundesländer	0.199	0.100	0.014	0.009	+0.140
Zusammenhang Ortsgröße / Gleichgültigkeit					
Alte Bundesländer	0.122	0.061	0.005	0.000	+0.104
Neue Bundesländer	0.196	0.098	0.013	0.016	+0.098

Fragen:

Ist die Verwendung des Maßes Gamma 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 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!

Interpretieren Sie die Unterschiede zwischen dem Ortsgröße-Begeisterung-Zusammenhang und dem Ortsgröße-Gleichgültigkeit-Zusammenhang nach alten und neuen Bundesländern!

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

- A. Traditionelle Lebensführung
- B. Begeisterung für ausländische Spezialitäten
- C. Gleichgültigkeit gegenüber dem Herkunftsland der Lebensmittel

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	-0.285	-0.071
B	-0.285	1.000	+0.251
C	-0.071	+0.251	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 die Variablenpaare A/B und B/C relativ hoch miteinander korrelieren, während das Paar A/C nur eine deutlich geringere Korrelation aufweist?

Welche kausalen Zusammenhänge zwischen den drei Variablen können Sie vermuten? Wenn Sie keinen kausalen Zusammenhang vermuten wollen: Haben Sie dafür Gründe?

5. Multivariate Statistik / Multiple Regression

Die Verbraucherstudie bietet folgende Fragen, die darauf hin ausgewertet werden können, von welchen anderen Einstellungen die Begeisterung der Befragten für ausländische Lebensmittel gesteuert wird:

- Traditionelle Lebensführung
- Innovationsneigung
- Erlebnisorientierung
- Nostalgie
- Mißtrauen gegenüber Neuprodukten
- Convenience-orientiertes Kochen
- Qualitäts-Orientierung

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):

	R	R^2
Rheinland-Pfalz	0.251	0.063
Niedersachsen	0.331	0.110
Brandenburg	0.339	0.115
Baden-Württemberg	0.349	0.122
Berlin (Westteil)	0.393	0.155
Hessen	0.397	0.157
Bremen	0.412	0.169
Mecklenburg	0.419	0.175
Sachsen	0.419	0.176
Sachsen/Anhalt	0.423	0.179
Schleswig-Holstein	0.437	0.191
Nordrhein-Westfalen	0.447	0.200
Bayern	0.459	0.211
Thüringen	0.466	0.217
Hamburg	0.481	0.232
Saarland	0.487	0.237
Berlin (Ostteil)	0.527	0.277

Interpretieren Sie dieses Ergebnis!

Die Einzelauswertung für die beiden Teile Berlins zeigt folgende standardisierten Regressionskoeffizienten:

	Westteil	Ostteil
V58 Traditionelle Lebensführung	-0.122	-0.365
V59 Innovationsneigung	0.184	0.144
V60 Erlebnisorientierung	0.204	0.278
V61 Nostalgie	-0.008	-0.018
V62 Mißtrauen gegenüber Neuprodukten	0.003	0.029
V63 Convenience-orientiertes Kochen	-0.154	0.071
V64 Qualitäts-Orientierung	-0.049	0.111

Vergleichen Sie den Einfluss der sieben unabhängigen Variablen auf die abhängige Variable „Begeisterung der Befragten für ausländische Lebensmittel“ in den beiden Teilen Berlins und zwischen den beiden Teilen Berlins. Finden Sie Interpretationen für die gefundenen Unterschiede.

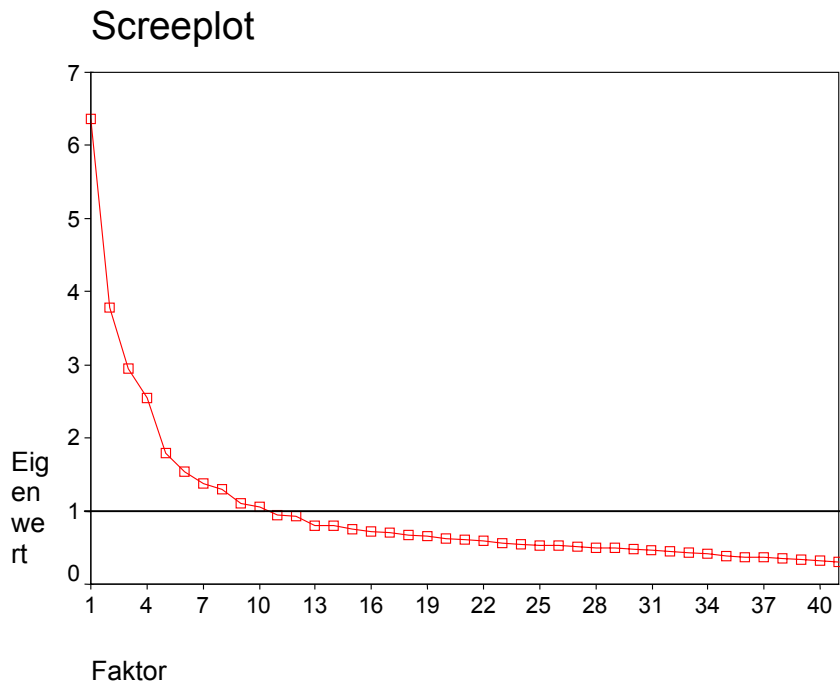
5. Multivariate Statistik / Faktorenanalyse

In der GfK-Verbraucherstichprobe sind zahlreiche Fragen zu Einstellungen zur Ernährung enthalten. Eine Faktorenanalyse ergab folgende Matrix von Ladungskoeffizienten:

	1	2	3	4	5	6
V78 Achten auf fettarme Ernährung	0.730					
V79 Achten auf schonende reizarme Kost	0.716	0.204				
V80 schlankheitsbewusste Ernährung	0.673					
V66 Achten auf Figur	0.654					
V96 Achten auf Kalorien	0.641					
V92 Rücksichtnahme auf Gesundheit bei Ernährung	0.640	0.218				
V102 Vermeidung von gesundheitsschädlicher Ernährung	0.572	0.337				
V72 vollwertige Ernährung	0.464	0.356		-0.205		
V90 regelmäßige Verwendung von Vitamin- und Mineralstoffpräparaten	0.459					0.408
V76 Multivitamin-säfte als wichtige Ergänzung	0.347			0.274		0.207
V103 bei Wahl Kauf von deutschen Lebensmitteln		0.683		0.207		
V83 beste Qualität der deutschen Nahrungsmittel		0.620		0.209		0.235
V81 Kauf von zusatzstofffreien Lebensmitteln	0.433	0.561				
V68 Gleichgültigkeit gegenüber Herkunftsland der Lebensmittel		-0.550	0.262			
V65 Ablehnung von Konservierungsstoffen in Produkten	0.271	0.520		-0.204		
V93 Information über umweltbelastete Lebensmittel- kein Kauf mehr	0.343	0.476				
V104 bei normaler Kost keine Ergänzungspräparate notwendig		0.474		0.271		-0.378
V88 Kauf von ausschließlich frischen Lebensmittel	0.219	0.469			-0.323	
V94 Spaß am Ausprobieren fremdländischer Spezialitäten			0.690	-0.282		-0.251
V82 Vorliebe für ausgefallene Speisen und Gerichte			0.677		-0.270	
V70 Vorzug von edlen Speisen und Getränken	0.203		0.646			0.270
V86 Verwöhnen mit gutem Essen			0.643			
V73 Delikatessen			0.628			0.258
V98 sehr anspruchsvoll beim Essen und Trinken			0.621			0.238
V67 Begeisterung für ausländische Spezialitäten			0.593	-0.257		-0.257
V71 Vorzug von Hausmannskost		0.244	-0.238	0.563		
V91 zuviel Wirbel um Ernährung				0.559		
V87 Vorliebe für herzhaft, deftige Mahlzeiten				0.545		
V77 Überschätzung des Einflusses der Ernährung auf die Gesundheit				0.506		
V97 Kochen von altbewährten Gerichten		0.224	-0.276	0.499	0.206	
V105 Kein Essen und Trinken mehr bei Annahme aller Ernährungstips				0.474		
V99 häufige Verwendung von Getreidekörnern	0.306	0.283		-0.420		
V85 vegetarische Ernährung	0.239	0.261		-0.405		
V84 Vorliebe für schnell kochbare Gerichte					0.803	
V95 Vorzug von Kochen einfacher Gerichte					0.794	
V69 viel Zeit für Kochen			0.275	0.342	-0.585	
V74 kein Kochen ohne Fertigprodukte		-0.229			0.552	
V101 Konserven genauso gut wie Frisches		-0.334		0.303	0.404	
V100 Lebensmittel bekannter Marken besser als die unbekannter Marken		0.266				0.667
V89 kein Vertrauen zu Nahrungsmitteln ohne Markenbezeichnung		0.322				0.653
V75 Preis, nicht Marke, ist für Kauf entscheidend				0.230		-0.571

Koeffizienten kleiner als 0.2 sind zur Erleichterung des Auffindens einer Einfachstruktur weggelassen; außerdem sind die Variablen nach der absoluten Größe ihrer größten Ladungskoeffizienten sortiert.. Interpretieren Sie die sechs Faktoren, indem Sie ihnen möglichst prägnante Namen geben. Begründen Sie Ihre Namensgebung.

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



Halten Sie danach die Extraktion von sechs 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 sechs, sondern drei oder vier oder fünf oder sieben ... 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 sechs Faktoren die richtige Wahl waren, auf „wenn ja, warum?“ besonders ausführlich antworten.)