

# Deskriptive und Inferenzstatistik

## Musterlösung zur Klausur vom 9. Juli 2008

**Name:**

**Matrikelnummer:**

Erlaubte Hilfsmittel: Außer Schreibwerkzeug diesmal auch Taschenrechner (wird aber nicht nötig sein); Antworten bitte im Anschluss an die Fragen, Rückseiten können für Antworten und Entwürfe genutzt werden; Entwürfe bitte anschließend durchstreichen!

Diese Klausur soll — nicht — als Freiversuch gewertet werden (gilt nur für Studierende nach alter Ordnung, die anderen können das später entscheiden),

.....  
Unterschrift

Für die Studierenden nach der Prüfungsordnung vom 15. März 2007 gilt folgendes: zu den sich aus der Tabelle weiter unten ergebenden Punktzahlen werden die Punktzahlen der beiden besten Übungen aus dem Sommersemester 2008 addiert. Diese Punktzahl (maximal 70 plus zweimal maximal 15, also 100) wird zu der Punktzahl (ebenfalls maximal 100) hinzugerechnet, die im Wintersemester 2008/2009 in der Veranstaltung „Stochastische Prozesse und ihre Anwendungen“ erreicht wird. Auf diese Weise wird sich eine Gesamtpunktzahl von maximal 200 ergeben; 200 Punkte ergeben eine Modulnote von 1.0, die Mindestpunktzahl für eine 4.0 wird höchstens 100 betragen.

Für die Studierenden nach anderen Prüfungsordnungen gilt: Für die beiden besten Übungsaufgaben gibt es die dort erzielten Punkte (max. 30), für die Klausur gibt es bis zu 70 Punkte: 100 Punkte geben eine 1.0. Die Mindestpunktzahl für eine 4.0 wird höchstens 50 betragen.

Aufgabe	Maximal erreichbar		Im Mittel wurden erreicht	
1	5	5	3.96	3.96
2.1	4	14	3.61	11.88
2.2	5		4.65	
2.3	5		3.62	
3.1	2	15	1.63	9.29
3.2	2		1.73	
3.3	2		1.12	
3.4	6		3.99	
3.5	3		0.82	
4.1	3	18	1.93	12.96
4.2	3		2.20	
4.3	3		2.34	
4.4	3		2.09	
4.5	3		2.20	
4.6	3		2.20	
5.1	4	18	1.90	13.73
5.2	8		6.33	
5.3	3		2.68	
5.4	3		2.82	
Summe	70	70		51.82
Übung	30	20	25.485	25.485
Gesamt	100	100		77.305

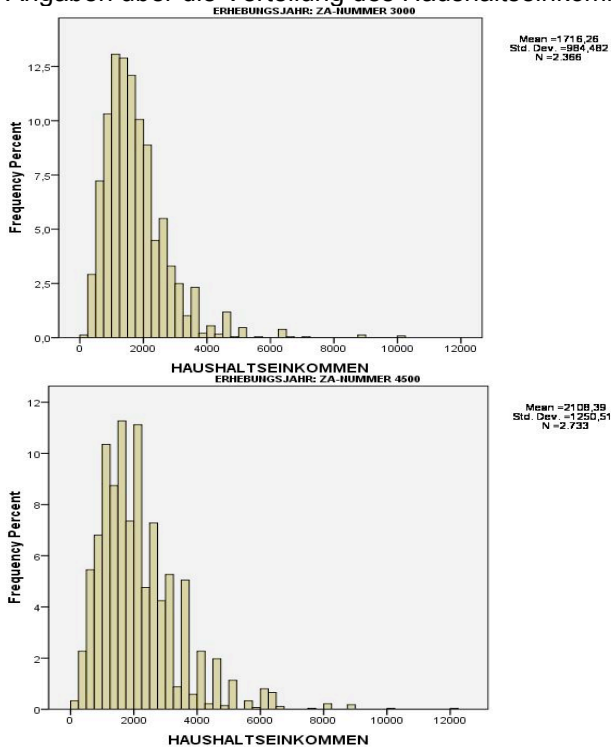
### 1. Skalenniveaus

In den Aufgaben dieser Klausur werden verschiedene Variable benutzt. Bitte geben Sie hier schon zusammenfassend an, welches Skalenniveau nach Ihrer Einschätzung die einzelnen Variablen haben und begründen Sie Ihre Einschätzung. Sie können auch einer Variablen verschiedene Skalenniveaus zuweisen, müssen dann aber auch das begründen.

	Dichotom	Nominal	Ordinal	Interval	Ratio	Begründung
Haushaltseinkommen (siehe Aufgabe 2)					X	numerische Verdopplung ist zugleich materielle Verdopplung, Null ist möglich.
Häufigkeit der Belästigung durch Passivrauchen (siehe Aufgabe 3)			X			Häufigkeit lässt sich abstufen, aber die Stufen sind nicht gleich hoch
Wahlbeteiligung bei der Bundestagswahl 2005 (siehe Aufgabe 4)					X	numerische Verdopplung ist zugleich materielle Verdopplung, Null ist möglich. Und es geht um den Prozentanteil der Wähler an den Wahlberechtigten, nicht um die individuelle Wahlbeteiligung (die wäre dichotom).
Die einzelnen Variablen zum Thema Freizeit aus Aufgabe 5			X			Häufigkeit lässt sich abstufen, aber die Stufen sind nicht gleich hoch, und das gilt für jede einzelne dieser Variablen!
Die Variable „Erhebungsgebiet West Ost“ aus Aufgabe 2.2.	X					Hier spielen nur zwei Himmelsrichtungen eine Rolle.

### 2. Univariate Statistik und Mittelwertvergleiche

Den ALLBUS-Umfragen (Bundesrepublik Deutschland) aus den Jahren 1998 und 2006 sind folgende Angaben über die Verteilung des Haushaltseinkommens (auch 1998 in € umgerechnet) zu entnehmen:

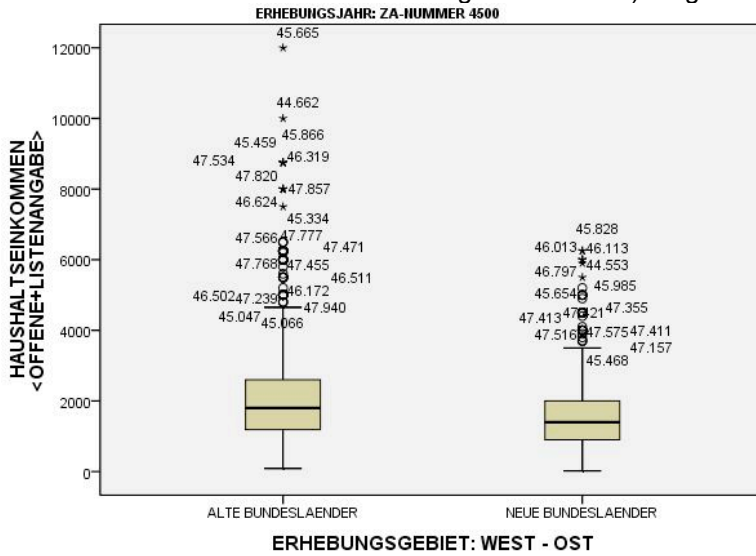


		Haushaltsnettoeinkommen €	
		1998	2006
N	Gültig	2312.06	2763.44
	Fehlend	921.93	611.56
Mittelwert		1771.58 €	1957.43 €
Median		1564.83 €	1700.00 €
Mode		1500.00 € <sup>a</sup>	2000.00 €
Std. Deviation		1022.48 €	1239.60 €
Schiefe		+2.15	1.77
Wölbung (Exzess, Kurtosis)		+9.40	5.63
Perzentile	10	750.00 €	700.00 €
	25	1063.00 €	1100.00 €
	50	1564.83 €	1700.00 €
	75	2150.00 €	2500.00 €
	90	3000.00 €	3500.00 €

2.1 Erläutern Sie, was Sie am Histogramm und an den Parametern erkennen können.

Die Verteilungen sind beide gleichermaßen linkssteil und spitzer als die Normalverteilung (erkennbar sowohl am Histogramm als auch an Schiefe und Wölbung, aber auch daran, dass die Differenz zwischen dem 1. Quartil und dem Median kleiner ist als die zwischen Median und zweiten Quartil, schließlich daran, dass der Mittelwert größer ist als der Median). Zwischen 1998 und 2006 ist das mittlere Haushaltseinkommen deutlich gestiegen. Ob der Unterschied zwischen den beiden Mittelwerten signifikant größer ist als 0, lässt sich aus den angegebenen Parametern nicht ohne weitere Berechnungen ermitteln (die kann man aber fast im Kopf erledigen: der Standardfehler des Mittelwerts 1998 ist ungefähr 21.29 €  $\approx 1022.48/\sqrt{2311}$ , die Differenz zwischen beiden Mittelwerten beträgt ungefähr 8.7 Standardfehler).

2.2 Die folgende Abbildung zeigt vereinfacht die Verteilung der Haushaltseinkommen 1990 in den westlichen und den östlichen Bundesländern. Was entnehmen Sie dieser graphischen Darstellung? Zur Erinnerung: die „Boxes“ umfassen die mittleren beiden Quartile, die „Whiskers“ sind höchstens anderthalb mal so lang wie die „Boxes“, und falls auch jenseits der „Whiskers“ noch „Ausreißer“ vorkommen, sind die durch kleine Kreise bzw. Sternchen (mehr als 1.5 bzw. mehr als 3 Quartilsabstände vom 3. Quartil entfernt; die dabei stehenden Zahlen sind Befragtennummern!) dargestellt.



Das Minimum liegt in beiden Teilen Deutschlands bei 0, alle drei Quartile liegen im Osten deutlich unter denen des Westens, das Maximum im Osten liegt nur etwa halb so hoch wie das im Westen. In beiden Teilen ist die Verteilung linkssteil (wie bei Aufgabe 2.1).

Cases weighted by OST-WEST TRANSFORMATIONSGEWICHT HAUSHALT

2.3 1990 wurden die ALLBUS-Teilnehmer auch danach befragt, ob sie sich vorstellen können, Steuern zu hinterziehen. Wertet man die Frage nach dem Haushaltseinkommen danach aus, ob die Befragten Steuern hinterziehen würden oder nicht, so ergibt sich für 1990 folgendes Bild:

bereit, Steuern zu hinterziehen	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwerts
ja, unter Umständen	589	1711.73 €	1106.55 €	45.60 €
nein, nie	1583	1565.84 €	1235.10 €	31.04 €

und folgende Statistik:

	t-Test für die Gleichheit der Mittelwerte						
	t	df	Sig. (zweiseitig)	Mittelwert-Differenz	Differenz der Standardfehler	95% Confidence Interval der Differenz	
						Lower	Upper
Varianzgleichheit unterstellt	2.516	2170	<b>.012</b>	145.893	57.997	32.159	259.628

Kann man aus dieser Stichprobe schließen, dass potentielle Steuerhinterzieher ein höheres Netto-Haushaltseinkommen hatten? Was können Sie mit der unteren Tabelle anfangen? Was sagt Ihnen die Zahl 0.012?

Ja. Der Unterschied zwischen den beiden Einkommensmittelwerten ist signifikant von Null verschieden. Eine weitere Stichprobe aus der gleichen Grundgesamtheit ergäbe unter der Annahme, dass es in der Grundgesamtheit keinen Unterschied zwischen Ehrlichen und Steuerhinterziehern gibt, nur mit einer Wahrscheinlichkeit von 1.2% einen größeren Mittelwertunterschied als 145.89 € (oder mit 98.8% Wahrscheinlichkeit einen geringeren Unterschied).

### 3. Bivariate Statistik

In der Vorlesung und in der Übung hatten Sie die Beziehung zwischen der empfundenen Belästigung durch Passivrauchen und der Bereitschaft untersucht, Raucher in der Umgebung zu bitten, das Rauchen einzustellen.

		Ich bitte darum, nicht zu rauchen				
		Immer	Oft	Manchmal	Nie	Insgesamt
		Männer				
Ich fühle mich durch Passivrauchen gestört	nie	73	68	300	2933	<b>3374</b>
	% nie	2,2%	2,0%	8,9%	86,9%	<b>100,0%</b>
	selten	144	213	1178	1505	<b>3040</b>
	% selten	4,7%	7,0%	38,8%	49,5%	<b>100,0%</b>
	manchmal	189	478	1150	792	<b>2609</b>
	% manchmal	7,2%	18,3%	44,1%	30,4%	<b>100,0%</b>
	sehr oft	351	339	315	320	<b>1325</b>
	% sehr oft	26,5%	25,6%	23,8%	24,2%	<b>100,0%</b>
	<b>insgesamt</b>	<b>757</b>	<b>1098</b>	<b>2943</b>	<b>5550</b>	<b>10348</b>
	<b>% insgesamt</b>	<b>7,3%</b>	<b>10,6%</b>	<b>28,4%</b>	<b>53,6%</b>	<b>100,0%</b>
		Frauen				
nie	116	102	380	2663	<b>3261</b>	
% nie	3,6%	3,1%	11,7%	81,7%	<b>100,0%</b>	
selten	318	418	1663	1804	<b>4203</b>	
% selten	7,6%	9,9%	39,6%	42,9%	<b>100,0%</b>	
manchmal	410	831	1816	1029	<b>4086</b>	
% manchmal	10,0%	20,3%	44,4%	25,2%	<b>100,0%</b>	
sehr oft	771	694	543	466	<b>2474</b>	
% sehr oft	31,2%	28,1%	21,9%	18,8%	<b>100,0%</b>	
<b>insgesamt</b>	<b>1615</b>	<b>2045</b>	<b>4402</b>	<b>5962</b>	<b>14024</b>	
<b>% insgesamt</b>	<b>11,5%</b>	<b>14,6%</b>	<b>31,4%</b>	<b>42,5%</b>	<b>100,0%</b>	

Hier sind einige Koeffizienten, die SPSS aus der obigen Kreuztabelle berechnet hat (alle sind auf 0.0005-Prozent-Niveau signifikant von Null verschieden):

	Wert	Schätzfehler	Wert	Schätzfehler
	Männer		Frauen	
Gamma	+0.628	0.009	+0.588	0.008
Kendall's tau b	+0.454	0.007	+0.422	0.006
Somers' d	+0.417	0.007	+0.406	0.006
Spearman	+0.507	0.008	+0.478	0.007
Pearson's r	+0.484	0.009	+0.462	0.007

3.1: Wie unterscheiden sich Männer und Frauen in ihrer Bereitschaft, Raucher zu bitten, das Rauchen einzustellen? Woran können Sie das erkennen? **Wenn man die Randverteilungen in den Zeilen „% insgesamt“ miteinander vergleicht, sieht man, dass 53.6 % aller Männer sich niemals beschwerten, aber bloß 42.5 % der Frauen; immer oder oft beschwerten sich nur 17.9% der Männer, aber 26.1% der Frauen. (Viele haben hier schon den Zusammenhang zwischen den beiden Variablen besprochen, danach war hier aber noch nicht gefragt!)**

3.2 Lässt sich aus der Tabelle entnehmen (gegebenenfalls durch ein bisschen Kopfrechnen ☺), ob Männer sich häufiger (oder seltener) belästigt fühlen? Woran können Sie das erkennen? **Dazu musste man ein bisschen Kopfrechnen: 3374 von 10348 Männern (32.6 %) fühlen sich „nie“ durch Passivrauchen gestört, aber nur 3261 von 14024 Frauen (23.3 %) — oder 1325 von 10348 Männern (12.8 %) fühlen sich „sehr oft“ gestört, aber immerhin 2474 von 14024 (17.6 %).**

3.3: Warum hat sowohl bei den Männern als auch bei den Frauen der Koeffizient Gamma den höchsten Zahlenwert? **Weil bei allen anderen Koeffizienten die „verbundenen Paare“ (bei tau und im Nenner, bei den beiden anderen auf andere Weise) mit berücksichtigt werden, bei gamma aber nicht — und bei der kleinen Anzahl von Ausprägungen in beiden Variablen gibt es eben sehr viele Paare von Befragten, die sich in einer oder beiden Variablen nicht unterscheiden.**

3.4: Wie beurteilen Sie den Zusammenhang zwischen den beiden Variablen, und wie beurteilen Sie die Koeffizienten, die den Zusammenhang beschreiben?

Diese Aussage ist ...	eher richtig	eher falsch
3.3.1 Je eher sich jemand durch Passivrauchen belästigt fühlt, desto eher wird er oder sie Raucher bitten, das Rauchen einzustellen.	<b>X</b>	<input type="checkbox"/>
3.3.2 Bei Frauen ist dieser Zusammenhang etwas schwächer ausgeprägt.	<b>X</b>	<input type="checkbox"/>
3.3.3 Dass dieser Zusammenhang bei Frauen etwas schwächer ausgeprägt ist, lässt sich mühelos der Kreuztabelle entnehmen.	<input type="checkbox"/>	<b>X</b>
3.3.4 Dass dieser Zusammenhang bei Frauen etwas schwächer ausgeprägt ist, kann man daran erkennen, dass die jeweiligen Koeffizienten bei den Frauen etwas kleiner sind als bei den Männern.	<b>X</b>	<input type="checkbox"/>
3.3.5 Dass dieser Zusammenhang bei Frauen etwas schwächer ausgeprägt ist, ist sogar statistisch signifikant.	<b>X</b>	<input type="checkbox"/>
3.3.6 Der Unterschied der beiden Gammas beträgt 0.040. Wenn in der Grundgesamtheit der Zusammenhang zwischen den beiden Variablen bei Männern und Frauen gleich stark wäre (in der Gesamtstichprobe ist er übrigens 0.598), müsste der Unterschied zwischen Gamma-Männern und Gamma-Frauen viel kleiner sein.	<b>X</b>	<input type="checkbox"/>

3.5 Als Ergänzung zu 3.3.4 bis 3.3.6: Um wie viele Standardfehler liegen Gamma-Männer und Gamma-Frauen auseinander? **Um ungefähr 4.5 bis 5 (aus der Sicht der Männer 0.040/0.009, aus der Sicht der Frauen 0.040/0.008); die Testgröße liegt also irgendwo zwischen 4.5 und 5 und ist damit signifikant von Null verschieden. Wie man aus der Tabelle in der nächsten Aufgabe sehen kann, ist die Wahrscheinlichkeit, bei einer gleich großen Stichprobe aus der gleichen Grundgesamtheit (bei der man keinen Unterschied zwischen Männern und Frauen annimmt!) mit einer Wahrscheinlichkeit von weniger als 0.0005 ein größerer Unterschied als 0.040 zwischen den beiden geschlechtsspezifischen Koeffizienten zu erwarten ( $t = \pm 4.580 \rightarrow \text{Sig.} < 0.0005$ ). **Wer bei 3.3.3 „eher richtig“ angekreuzt hat, muss ein ungeheures Vermögen haben, mit Zahlen umzugehen (trotzdem gab es dafür nur einen halben Punkt — der geringe Unterschied lässt sich nämlich wirklich nicht — und schon gar nicht mühelos — der Kreuztabelle entnehmen!). — Häufig wurde hier „0.001“ geantwortet, das ist aber die Differenz der Standardfehler, und die hat eigentlich gar keine Bedeutung (jedenfalls nicht die, nach der hier gefragt war!****

### 4. Multivariate Statistik / Multiple Regression

Bei der Betrachtung der Bundestagswahl 2005 in den 299 Wahlkreisen wurde die Wahlbeteiligung als abhängige Variable genommen, einige Variable, die die Sozialstruktur der Wahlkreise beschreiben, wurden zur Erklärung der Unterschiede in der Wahlbeteiligung herangezogen.

Die Regression ergibt folgende Koeffizienten:  $R^2 = 0.499$ ,  $R = 0.706$ .

Koeffizienten

	Unstandardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.
	b	Standardfehler	b*		
(Konstante)	<b>87.059</b>	1.290		67.501	.000
Arbeitslosenquote	<b>-0.265</b>	0.031	-0.466	-8.528	.000
% ohne Hauptschule	-0.213	0.065	-0.162	-3.285	.001
% 18- bis 25jährige	-0.733	0.160	-0.224	-4.580	.000
% mit Abitur	<b>+0.069</b>	0.020	+0.152	+3.505	.001

a. Abhängige Variable: Wahlbeteiligung Bundestagswahl 2005

4.1 Welche Bedeutung haben die beiden Zahlen -0.265 und +0.069 (in Zeile 2 und 5 in der ersten Spalte)? Von zwei Wahlkreisen, die sich nur in der Arbeitslosenquote (und nicht in den drei anderen unabhängigen Variablen) um einen Prozentpunkt unterscheiden, kann man erwarten, dass sie sich in der Wahlbeteiligung um 0.265 Prozentpunkte unterscheiden, und zwar so, dass der Wahlkreis mit der höheren Arbeitslosigkeit die geringere Wahlbeteiligung aufweist. Von zwei Wahlkreisen, die sich nur im Anteil der Personen mit Abitur (und nicht in den drei anderen unabhängigen Variablen) um einen Prozentpunkt unterscheiden, kann man erwarten, dass sie sich in der Wahlbeteiligung um 0.069 Prozentpunkte unterscheiden, und zwar so, dass der Wahlkreis mit der geringeren Abiturientenquote die geringere Wahlbeteiligung aufweist. **Hier wurde häufig von „Statements“ und „Bewertung“ geschrieben, es geht aber um den Prozentanteil (der Wähler an den Wahlberechtigten, der Arbeitslosen an den Erwerbsfähigen etc.), da wird nur gezählt und dividiert, nicht bewertet.**

4.2 Welche (ziemlich fiktive) Bedeutung hat die Zahl 87.059 (erste Zeile, erste Spalte)? Wenn es einen Wahlkreis geben könnte ganz ohne Arbeitslose, ganz ohne Menschen ohne Hauptschulabschluss, ganz ohne 18- bis 25jährige und ganz ohne Menschen mit Abitur, dann müsste man dort eine Wahlbeteiligung von 87.059 % erwarten.

4.3 Wie können Sie sich die unterschiedlichen Vorzeichen der Koeffizienten in der ersten Spalte erklären (einer ist positiv, die meisten sind negativ)? Das könnte zum Beispiel daran liegen, dass Leute mit höherem Bildungsstand, Ältere und Erwerbstätige eher zur Wahl gehen. Das ist aber nicht die einzige mögliche Begründung. Das gleiche Bild könnte sich ergeben, wenn in Wahlkreisen mit vielen höher Gebildeten auch die Arbeitslosen eher zur Wahl gehen. Am sichersten ist die Aussage: In Wahlkreisen, in denen der Bildungsstand höher, die Arbeitslosigkeit geringer und der Anteil von Erstwählern geringer ist, spricht das politische Klima eher dafür, zur Wahl zu gehen als in Wahlkreisen mit entgegengesetzter Sozial- und Bildungsstruktur.

4.4 Was bedeutet die Zahl 0.001 am Ende der letzten Zeile der Tabelle? Im konkreten Fall eigentlich gar nichts, denn die 299 Bundestagswahlkreise sind keine Stichprobe, sondern die Grundgesamtheit, es ist also gar nicht möglich, aus dieser Grundgesamtheit eine andere gleich große Stichprobe zu ziehen. Das kann SPSS aber nicht wissen, also führt es die gleiche Berechnung durch, die es auch durchführen würde, wenn es statt der Wahlkreisdaten die Daten je eines Wahllokals pro Wahlkreis bekommen hätte (dann könnte eine weitere Stichprobe nämlich jeweils ein anderes Wahllokal pro Wahlkreis enthalten). Und wenn das so wäre, dann könnte man sagen: Die Wahrscheinlichkeit, aus einer anderen gleich großen Stichprobe aus dieser Grundgesamtheit **(bei der wir annehmen, dass keine Beziehung zwischen dem Abiturientenanteil und der Wahlbeteiligung besteht)**, könnte man mit einer Wahrscheinlichkeit von 0.001 erwarten, dass der unstandardisierte Regressionskoeffizient größer als +0.069 oder kleiner als -0.069 ist (oder der standardisierte Regressionskoeffizient größer als +0.152 oder kleiner als -0.152 ist). **Wer den in Klammern fett gedruckten Satzteil nicht aufgeschrieben hat, hat 0.5 Punkte abgezogen bekommen.**

4.5 Welche Sozialstrukturvariable trägt zur Erklärung der Wahlbeteiligung am stärksten bei, welche am wenigsten? Am meisten die Arbeitslosenquote, am wenigsten der Anteil der Menschen mit Abitur.

4.6 Begründen Sie, warum Sie in 4.5 gerade diese beiden als unabhängig eingesetzten Variablen genannt haben. Welche Spalte(n) der Tabelle mussten Sie dazu auswerten? In Spalte b\* musste man für die Suche nach dem stärksten (geringsten) Erklärungsbeitrag nach dem größten (kleinsten) Absolutwert suchen. Die Suche nach dem größten (kleinsten) Absolutwert in der Spalte t bzw. nach dem kleinsten (größten) Wert in der Spalte Sig. (diese Spalte ist aber in diesem Beispiel weniger geeignet, weil sich je zwei Werte gar nicht sichtbar unterscheiden!) führt ausnahmsweise zu einem etwas anderen Ergebnis: der Absolutwert des Koeffizienten für „% mit Abitur“ ist etwas kleiner, der Absolutwert der Teststatistik t ist aber etwas höher — der Unterschied zwischen den beiden Absolutwerten ist statistisch in diesem Falle nicht signifikant. Deswegen ist auch die Antwort „% ohne Hauptschulabschluss“ auf Frage 4.5 als richtig zu werten!

### 5. Multivariate Statistik / Faktorenanalyse

In den beiden Tabellen unten finden Sie zwei Faktorenanalysen auf der Basis zweier ALLBUS-Befragungen aus den Jahren 1998 und **2004**. Die Variablen sind die Bewertungen von Aussagen der Art „In meiner Freizeit unternehme ich <die jeweilige Tätigkeit> (1) täglich, (2) mindestens einmal in der Woche, (3) mindestens einmal im Monat, (4) seltener als einmal im Monat, (5) nie). Man kann diese Variablen mit etwas Großzügigkeit als intervallskaliert ansehen, wohl wissend, dass die angegebenen Häufigkeiten eher gleiche Quotienten als gleiche Intervalle haben (aber das trifft für viele Einschätzungen – z.B. auch für die Einschätzung von Lautstärke! – zu.

In beiden Tabellen sind die (im übrigen gleichen) Variablen so geordnet, dass zunächst die kommen, die auf den Faktor 1 den höchsten Ladungskoeffizienten haben (die sind dann absteigend nach diesem Ladungskoeffizienten sortiert), dann folgen die Variablen die auf den zweiten Faktor den höchsten Ladungskoeffizienten haben (wieder nach diesem geordnet) usw. Zur weiteren Verbesserung der Übersichtlichkeit sind die höchsten Ladungskoeffizienten einer Zeile fett gedruckt und die Koeffizienten unter einem Absolutwert von <0.25 weggelassen.

Variablen 1998	Faktoren 1998				Variablen 2006	Faktoren 2004			
	1	2	3	4		1	2	3	4
V360 Videokassetten, DVD anschauen	<b>,692</b>				V361 mit dem Computer beschäftigen	<b>,776</b>			
V379 Kino, Pop-+ Jazzkonzerte, Tanzen	<b>,666</b>	,358			V362 Internet, Online-Dienste nutzen	<b>,767</b>			
V359 LP, CD, Cassetten hören	<b>,614</b>	,288			V379 Kino, Pop- + Jazzkonzerte, Tanzen	<b>,719</b>			
V367 Essen oder Trinken gehen	<b>,557</b>				V360 Videokassetten, DVD anschauen	<b>,632</b>			,269
V378 Besuch von Sportveranstaltungen	<b>,554</b>		,451		V359 LP, CD, Cassetten Hören	<b>,589</b>	,256		,268
V377 Aktive sportliche Betätigung	<b>,540</b>	,315	,313		V367 Essen oder Trinken gehen	<b>,540</b>			
V374 Kirchgang	<b>-,440</b>		,311		V363 Sich privat weiterbilden	<b>,475</b>	,453		
V357 Bücher lesen		<b>,664</b>		,292	V377 Aktive sportliche Betätigung	<b>,465</b>	,378	,279	
V363 Sich privat weiterbilden	,310	<b>,648</b>			V357 Bücher lesen		<b>,714</b>		
V361 mit dem Computer beschäftigen	,374	<b>,605</b>			V380 Klass. Konzerte, Theater etc.		<b>,602</b>		
V362 Internet, Online-Dienste nutzen		<b>,598</b>			V375 Kunst, musische Tätigkeiten		<b>,597</b>		
V380 Klass. Konzerte, Theater etc.		<b>,588</b>			V366 Yoga, Meditation, Autog. Training		<b>,498</b>		
V375 Kunst, musische Tätigkeiten		<b>,519</b>	,329		V371 Ausflüge oder kurze Reisen		<b>,386</b>		,274
V366 Yoga, Meditation, Autog. Training		<b>,435</b>	,298		V358 Zeitschriften lesen		<b>,308</b>		
V373 Ehrenamtliche Tätigkeiten			<b>,641</b>		V373 Ehrenamtliche Tätigkeiten			<b>,641</b>	
V376 Basteln, Reparaturen			<b>,529</b>		V372 Beteiligung in Politik			<b>,533</b>	
V372 Beteiligung in Politik		,303	<b>,493</b>		V374 Kirchgang	<b>-,310</b>		<b>,444</b>	
V370 Gesellschaftsspiele in Familie			<b>,394</b>	,320	V378 Besuch von Sportveranstaltungen	,399		<b>,407</b>	
V368 Besuch Nachbarn, Freunde, Bek.	,295			<b>,570</b>	V364 Einfach nichts tun, faulenzten			<b>-,400</b>	
V365 Spazierengehen, Wandern				<b>,555</b>	V376 Basteln, Reparaturen			<b>,393</b>	,296
V369 Besuch Familie, Verwandtschaft				<b>,550</b>	V368 Besuch Nachbarn, Freunde, Bek.				<b>,598</b>
V358 Zeitschriften lesen		,282		<b>,386</b>	V369 Besuch Familie, Verwandtschaft				<b>,583</b>
V364 Einfach nichts tun, faulenzten			<b>-,352</b>	<b>,367</b>	V370 Gesellschaftsspiele in Familie				<b>,541</b>
V371 Ausflüge oder kurze Reisen	,295	,267	,295	<b>,324</b>	V365 Spazierengehen, Wandern	<b>-,323</b>	,385		<b>,416</b>

5.1 Beurteilen Sie die Unterschiede zwischen den beiden Ladungsmatrizen. Hat sich zwischen den beiden Erhebungsjahren Nennenswertes geändert? Die Faktoren bedeuten in den beiden Jahren nicht ganz genau dasselbe. Sie stimmen aber in folgenden Hinsichten überein: Faktor 1 fasst Vorlieben für intellektuell weniger anspruchsvolle Freizeitbeschäftigungen zusammen (dass die Beschäftigung mit dem Computer und die Nutzung von Online-Diensten 1998 noch nicht dazugehörte, 2006 aber schon, passt durchaus ins Bild!). Der zweite Faktor fasst Vorlieben für intellektuell eher anspruchsvolle Freizeitbeschäftigungen zusammen (1998 war das offenbar bei der Beschäftigung mit Computer und Internet noch der Fall, 2006 eher nicht mehr!). Der dritte Faktor fasst ehrenamtliche und gesellschaftlich relevante Beschäftigung zusammen (und hat eine negative Ladung aufs Nichtstun) — dass die Faktorenanalyse Basteln und Reparaturen auch hierhin platziert, ist nicht so leicht zu interpretieren. Der vierte Faktor schließlich fasst Vorlieben für eher gesellige Beschäftigungen zusammen. Der wesentliche Unterschied ist also die „Umgruppierung“ von computerorientierten Beschäftigungen aus der Gruppe „Bücher lesen“ in die Gruppe „Videokassetten anschauen“. **Der Unterschied wird nicht recht deutlich, wenn man Video/DVD/CD und Computer/Internet unterschiedslos zu Multimedia zusammenfasst: Video/DVD/CD ist fast völlig passiv, Computer/Internet setzen Aktivität voraus! Aber es scheint, dass die Art der Beschäftigung mit Computer und Internet zwischen 1998 und 2004 „passiver“ geworden ist. Die Zuordnung dieser beiden Variablen zum zweiten Faktor erscheint nur aus der Sicht von 2008 sinnvoller (1998 war die Welt noch eine andere!). Und: Die Ladungskoeffizienten haben nichts mit Häufigkeit zu tun! Man könnte, etwas (zu) salopp, formulieren: „Man erkennt, dass die Liebhaber von Popkonzerten im Gegensatz zu 1998 im Jahre 2004 auch im Internet surfen, werden die Internetsurfer kaum noch Bücher lesen.“**

5.2 Geben Sie den Faktoren Namen (Hinweis: die Namen sollten wiederum Tätigkeiten bezeichnen, aber nun in einer etwas größeren Kategorisierung von Freizeitbeschäftigungen; wenn Sie finden, dass ein Faktor aus einem Jahr einem Faktor aus dem anderen Jahr sehr ähnlich ist, können Sie natürlich in der entsprechenden Zelle einfach so etwas wie „=2004-3“ — also „=<Jahr><Faktornummer>“ hinschreiben. Wenn Sie finden, dass man einem Faktor keinen sinnvollen Namen geben kann, müssen Sie dies begründen.):

Faktor	Faktoren 1998	Faktoren 2004
1	Video/DVD/CD und Ähnliches ohne Weiterbildung und ohne Computer; Aktivitäten zur Entspannung	Video/DVD/CD und Ähnliches aber jetzt mit Weiterbildung und Computer
2	Bücher lesen und Ähnliches, einschließlich der Beschäftigung mit Computern	Bücher lesen und Ähnliches, aber nun ohne Beschäftigung mit Computern
3	Ehrenamtliche und andere gesellschaftlich relevante Aktivitäten	Ehrenamtliche und andere gesellschaftlich relevante Aktivitäten
4	Nachbarschaftlich, familiär gesellige Aktivitäten	Nachbarschaftlich, familiär gesellige Aktivitäten

5.3 Wie beurteilen Sie die Einfachstruktur? Sind die Faktoren in den beiden Erhebungen sauber genug getrennt? In beiden Jahren ist die Einfachstruktur als einigermaßen gut zu beurteilen. Soweit einzelne Variablen Ladungen von mehreren Faktoren aufweisen, ist das verständlich (am ausgeprägtesten gilt das für „Ausflüge, kurze Reisen“: diese Aktivität hat Aspekte aller Art, man kann sie zur Entspannung, zur Bildung, oder aus Geselligkeit unternehmen); Kirchengang gilt einerseits als gerade nicht entspannend, andererseits als durchaus gesellschaftlich relevant; Weiterbildung, wenn sie privat ausgeübt wird, kann einerseits entspannend sein, andererseits gehört sie zum Bereich Bildung.

5.4 Hätten Sie ebenfalls vier Faktoren (oder mehr? oder weniger?) berücksichtigt? Warum? (Hier sind auch die Plots der Eigenwerte gegen ihre Ordnungsnummer.) Man hätte sich auch mit einem einzigen Faktor zufrieden geben können (allerdings wäre der kaum interpretierbar gewesen!), denn hinter ihm kommt der auffälligste Knick. Der zweitauffälligste Knick kommt aber nach dem vierten Faktor. Nach dem Kaiser-Kriterium hätte man sieben Faktoren extrahiert (auf den Bildern schlecht zu sehen).

