

# Deskriptive und Inferenzstatistik

## Nachklausur vom 27. Oktober 2008

### Musterlösung

Die Anmerkungen in blau habe ich bereits während der Klausur aufgeschrieben, die in rot sind Reaktionen auf typische Fehler.

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	2.73	2.73
2.1	4	14	1.64	4.50
2.2	5		1.77	
2.3	5		1.09	
3.1	4	15	2.23	4.50
3.2	2		0.36	
3.3	2		0.36	
3.4	3		0.27	
3.5	4		1.27	
4.1	3	18	2.23	12.59
4.2	3		2.36	
4.3	3		2.55	
4.4	3		2.23	
4.5	3		2.41	
4.6	3		0.82	
5.1	4	18	2.09	6.95
5.2	8		0.66	
5.3	3		1.41	
5.4	3		2.00	
Summe	70	70		31.27

### 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.

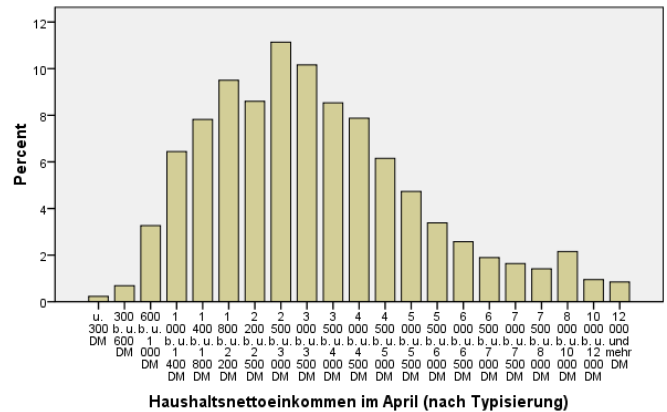
	Dichotom	Nominal	Ordinal	Intervall	Ratio	Begründung
Haushaltseinkommen (siehe Aufgabe 2)			X			Da das Einkommen nur klassifiziert vorliegt, kann man mit den Daten nicht rechnen. Gruppe 8 verdient nicht doppelt so viel wie Gruppe 4.
Die Variable Ausbildung (siehe Aufgabe 3)	X	X				Es gibt genau zwei Ausprägungen.
Wahlbeteiligung bei der Bundestagswahl 2005 (siehe Aufgabe 4)				X	X	Die doppelte Zahl bedeutet eine doppelt so hohe Wahlbeteiligung (auch wenn 2*85% unmöglich ist!).
Die Variable Land (siehe Aufgabe 3)		X				Keine Ordnung ist sinnvoller als irgend eine andere.
Die Variablen, die Gegenstand der Faktorenanalyse in Aufgabe 5 sind	(X)			X		Mit Ausnahme von „WestOst“ (dichotom) gilt für alle anderen das Gleiche wie bei der Wahlbeteiligung. Siehe den zweiten Satz bei Aufgabe 5.

### 2. Univariate Statistik und Vergleiche

2.1 Einer Teilstichprobe aus dem Mikrozensus 1998 ist folgende Verteilung der Haushaltsnettoeinkommen zu entnehmen:

	Frequency	Valid Percent	Cumulative Percent
1: u. 300 DM	25	,2	,2
2: 300 b. u. 600 DM	74	,7	,9
3: 600 b. u. 1 000 DM	351	3,3	4,2
4: 1 000 b. u. 1 400 DM	693	6,4	10,6
5: 1 400 b. u. 1 800 DM	841	7,8	18,4
6: 1 800 b. u. 2 200 DM	1022	9,5	27,9
7: 2 200 b. u. 2 500 DM	925	8,6	36,5
8: 2 500 b. u. 3 000 DM	1198	11,1	47,7
9: 3 000 b. u. 3 500 DM	1093	10,2	57,8
10: 3 500 b. u. 4 000 DM	918	8,5	66,4
11: 4 000 b. u. 4 500 DM	847	7,9	74,3
12: 4 500 b. u. 5 000 DM	662	6,2	80,4
13: 5 000 b. u. 5 500 DM	509	4,7	85,1
14: 5 500 b. u. 6 000 DM	364	3,4	88,5
15: 6 000 b. u. 6 500 DM	277	2,6	91,1
16: 6 500 b. u. 7 000 DM	204	1,9	93,0
17: 7 000 b. u. 7 500 DM	176	1,6	94,6
18: 7 500 b. u. 8 000 DM	152	1,4	96,0
19: 8 000 b. u. 10 000 DM	232	2,2	98,2
20: 10 000 b. u. 12 000 DM	102	,9	99,2
21: 12 000 und mehr DM	91	,8	100,0
Total	10756	100,0	
50: Selbst. Landwirt	89		
99: Angabe fehlt	421		
System	4		
Total	514		
	11270		

Haushaltsnettoeinkommen im April (nach Typisierung)



Geben Sie an, in welcher Einkommensgruppe folgende Maßzahlen liegen:

- 1. Quartil in Gruppe 6: Unterhalb des ersten Quartils haben wir ein Viertel der Fälle nicht der Ausprägungen.
- Median ...in Gruppe 9.....
- 3. Quartil in Gruppe 12

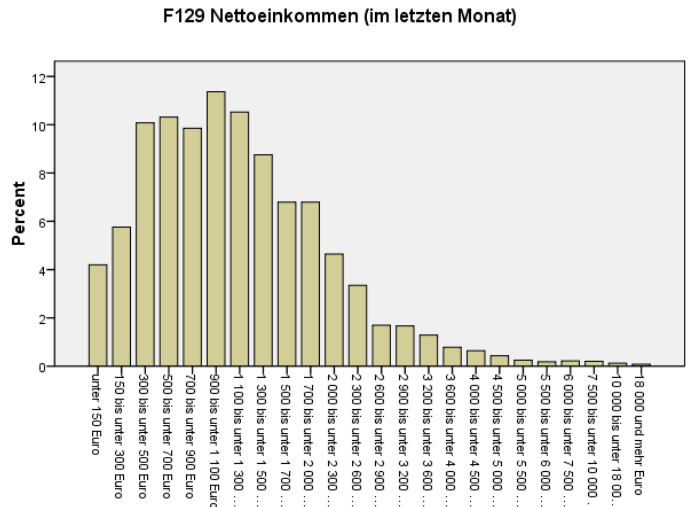
Wieviel Prozent aller Haushalte hatten ein Nettoeinkommen von mehr als 5000 DM?

19,6, bis unter 5000 waren es 80,4 %, also sind es über 5000 19,6 %, dazu brauchte man nur wenige Kopfrechnen.

Wie erklären Sie sich, dass die Einkommensgruppe 19 deutlich mehr Haushalte umfasst als die benachbarten Gruppen 18 und 20? Größer als Gruppe 18 wegen der vierfachen Spannweite, größer als 20, weil höhere Einkommen immer seltener werden. (Und Beamte verdienen längst nicht so viel, wie immer geglaubt wird.)

2.2 Dem Mikrozensus von 2005 lässt sich folgende Verteilung der (individuellen) Nettomonatseinkommen entnehmen:

	Frequency	Valid Percent	Cumulative Percent
1 unter 150 €	15122	4.2	4.2
2 150 bis unter 300 €	20753	5.8	10.0
3 300 bis unter 500 €	36303	10.1	20.0
4 500 bis unter 700 €	37156	10.3	30.4
5 700 bis unter 900 €	35481	9.9	40.2
6 900 bis unter 1 100 €	40935	11.4	51.6
7 1 100 bis unter 1 300 €	37897	10.5	62.1
8 1 300 bis unter 1 500 €	31522	8.8	70.9
9 1 500 bis unter 1 700 €	24463	6.8	77.7
10 1 700 bis unter 2 000 €	24461	6.8	84.4
11 2 000 bis unter 2 300 €	16733	4.6	89.1
12 2 300 bis unter 2 600 €	12055	3.3	92.4
13 2 600 bis unter 2 900 €	6110	1.7	94.1
14 2 900 bis unter 3 200 €	6006	1.7	95.8
15 3 200 bis unter 3 600 €	4647	1.3	97.1
16 3 600 bis unter 4 000 €	2822	0.8	97.9
17 4 000 bis unter 4 500 €	2299	0.6	98.5
18 4 500 bis unter 5 000 €	1558	0.4	98.9
19 5 000 bis unter 5 500 €	893	0.2	99.2
20 5 500 bis unter 6 000 €	653	0.2	99.4
21 6 000 bis unter 7 500 €	787	0.2	99.6
22 7 500 bis unter 10 000 €	717	0.2	99.8
23 10 000 bis unter 18 000 €	444	0.1	99.9
24 18 000 und mehr €	301	0.1	100.0
Total	360118	100.0	
50 Selbstständiger Landwirt	1377		
90 Kein Einkommen	95218		
99 Ohne Angabe	20526		
Total	117121		
	477239		



Die Graphik soll Ihnen lediglich den Vergleich erleichtern, falls Sie die beiden Auswertungen überhaupt für vergleichbar halten.

Lässt sich diese Tabelle mit der in 2.1 vergleichen? Wenn nein: warum nicht?

Einmal geht es um das Haushaltseinkommen, das andere Mal um das Individualeinkommen, und das Haushaltseinkommen ist immer gleich oder größer als das Individualeinkommen.

Wenn Sie meinen: ja, welchen Vergleich ziehen Sie?

Das einzige, was man hier sicher sagen kann, ist, dass beide Variablen eine schiefe Verteilung aufweisen.

2.3 Wo liegen 2005

- das 1. Quartil.....in Gruppe 4.....
- der Median.....in Gruppe 6.....
- das 3. Quartil.....in Gruppe 9.....

der Verteilung der individuellen Nettoeinkommen.

Warum ist es nicht sinnvoll, für die Variablen der Aufgaben 2.1 und 2.2 Mittelwerte, Varianzen, Schiefe und Wölbung zu berechnen? Die Gruppennummern eignen sich nicht für Berechnungen (was soll uns der Mittelwert der Gruppennummern sagen?), allenfalls könnte man annehmen, dass alle Mitglieder einer Einkommensgruppe ein Einkommen haben, das gerade in der Mitte zwischen den Einkommensgrenzen der Gruppe liegt. (Statt „21“ würde man bei allen Befragten der Gruppe 21 „6750€“ einsetzen). Dann käme man auf Parameter, die mit Sicherheit ungenau sind. Die unterschiedlichen Währungseinheiten spielen eine untergeordnete Rolle.

### 3. Bivariate Statistik

Laut Eurobarometer haben von den Befragten (repräsentativ gewichtet) in den einzelnen Ländern unterschiedlich viele im Ausland Schule oder Universität besucht oder sonst eine Ausbildung absolviert:

	nur im Inland ausgebildet		wenigstens teilweise im Ausland ausgebildet		Total
	abs.	%	abs	%	
1 France	2588.070	87.4%	372.208	12.6%	2960.279
2 Belgium	518.554	89.7%	59.838	10.3%	578.392
3 The Netherlands	838.940	94.2%	51.777	5.8%	890.716
4 Germany West	2991.008	86.6%	464.422	13.4%	3455.431
14 Germany East	812.740	94.4%	48.374	5.6%	861.114
5 Italy	2502.259	75.6%	807.610	24.4%	3309.869
6 Luxembourg	13.813	55.9%	10.885	44.1%	24.699
7 Denmark	267.887	90.9%	26.728	9.1%	294.615
8 Ireland	187.304	90.1%	20.523	9.9%	207.827
9 Great Britain	2896.480	92.9%	222.589	7.1%	3119.069
10 Northern Ireland	66.565	75.3%	21.833	24.7%	88.398
11 Greece	546.366	93.6%	37.088	6.4%	583.453
12 Spain	2012.831	83.4%	400.749	16.6%	2413.580
13 Portugal	462.823	85.1%	80.722	14.9%	543.545
16 Finland	267.139	92.8%	20.698	7.2%	287.837
17 Sweden	453.479	91.4%	42.697	8.6%	496.176
18 Austria	395.776	88.1%	53.502	11.9%	449.278
19 Cyprus (Republic)	30.872	83.1%	6.272	16.9%	37.143
20 Czech Republic	537.652	93.3%	38.905	6.7%	576.557
21 Estonia	56.666	95.0%	3.002	5.0%	59.668
23 Latvia	89.312	95.2%	4.476	4.8%	93.788
24 Lithuania	175.314	93.0%	13.268	7.0%	188.582
22 Hungary	550.509	96.2%	21.452	3.8%	571.961
25 Malta	20.593	94.8%	1.127	5.2%	21.720
26 Poland	1984.091	93.3%	142.117	6.7%	2126.207
27 Slovakia	271.863	93.6%	18.473	6.4%	290.336
28 Slovenia	104.751	93.6%	7.165	6.4%	111.917
Total	21643.656	87.8%	2998.502	12.2%	24642.159

Einschlägige statistische Maßzahlen können Sie den folgenden Tabellen entnehmen.

Lambda	symmetrisch	0.014
	Land abhängig	0.016
	Ausbildung abhängig	0.000
Goodman and Kruskal tau	Land abhängig	0.004
	Ausbildung abhängig	0.035
Uncertainty coefficient	symmetrisch	0.011
	Land abhängig	0.006
	Ausbildung abhängig	0.045
Phi		0.188
Cramer's V		0.188
Contingency coefficient		0.188

3.1: Beschreiben Sie die Unterschiede zwischen Ländern und Ländergruppen hinsichtlich der Bereitschaft, wenigstens einen Teil der Ausbildung im Ausland zu absolvieren.

Italien, Spanien und Portugal verzeichnen überdurchschnittlich viele teilweise im Ausland ausgebildete Befragte, ebenso Westdeutschland; in Ostmitteleuropa kommt Ausbildung im Ausland unter den Befragten wesentlich seltener vor, schon weil die älteren Befragten vor 1989 nur sehr eingeschränkte Möglichkeiten zur Mobilität hatten. Um das zu beurteilen, musste man nicht einmal wissen, wann die Befragung durchgeführt wurde — und natürlich konnte man in der DDR oder in Litauen vor 1989 keine solchen Befragungen durchführen! Dass Luxemburger so viel im Ausland studieren, liegt allein schon daran, dass man die meisten Fächer in Luxemburg gar nicht studieren kann — die Universität du Luxembourg wurde erst 2003 gegründet..

3.2 Warum ist Lambda, Ausbildung abhängig exakt gleich Null?

Weil „nur im Inland“ in allen Ländern die Mehrheit hat. Im übrigen hat Lambda mit „verbundenen Paaren“ gar nichts zu tun (das wurde mit Gamma verwechselt).

3.3 Wie stehen Sie dazu, dass bei Goodman und Kruskal tau und beim Uncertainty coefficient jeweils der Wert für „Ausbildung abhängig“ größer ist als der für „Land abhängig“?

Das dürfte daran liegen, dass die Landeszugehörigkeit als abhängige Variable eigentlich sowieso unsinnig ist.

3.4 Ist es überhaupt sinnvoll, einen Koeffizienten für „Land abhängig“ auszurechnen? Wenn ja, warum, wenn nein, warum nicht? (Nur „JA“ oder „NEIN“ ist nicht punktwürdig, aber jede einigermaßen plausible Begründung zählt, Sie können auch für *und* gegen diese beiden Koeffizienten argumentieren.)

JA: Es mag sein, dass Leute mit hoher Bildungsmobilität nach Luxemburg zu europäischen Institutionen gehen, das wäre aber die einzige Begründung für die Abhängigkeit der Landeszugehörigkeit von der Bildung.

NEIN: Die Leute entscheiden über ihren Wohnsitz nicht an Hand ihrer Bildungsbiographie.

Offenbar haben viele missverstanden, dass „Land abhängig“ im SPSS-Output bedeutet: „die Ausprägung der Variablen Land hängt von der Ausprägung der Variablen Ausbildung ab“, oder „die Leute wohnen deswegen im Land X, weil sie einen Teil ihrer Ausbildung im Ausland verbracht haben.“

3.5 Welche Erklärung bieten Sie an für die drastischen Unterschiede zwischen

a) Deutschland West und Deutschland Ost,

Die heute älteren Ostdeutschen waren bis 1989 in ihrer Bildungsmobilität stark eingeschränkt

b) Großbritannien und Nordirland?

Offenbar zählt der Rest des vereinigten Königreichs für diese Umfrage aus der Sicht von Nordirland zum Ausland, aber Großbritannien (=England, Wales und Schottland) und Nordirland sind zueinander nicht mehr Ausland als Sachsen und Schleswig-Holstein. Oder viele Nordiren haben in Irland studiert.

### 4. Multivariate Statistik / Multiple Regression

Bei der Betrachtung der Bundestagswahl 2005 in den 54 Wahlkreisen der „neuen Bundesländer“ (einschließlich „Berlin, Hauptstadt der DDR“) 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:  $R^2 = 0.475$ ,  $R = 0.689$ .

**Coefficients<sup>a</sup>**

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Konstante	87,375	1,966		44,452	,000
Arbeitslosenquote Dez 2004	<b>-,396</b>	,083	<b>-,488</b>	-4,772	,000
ohne Hauptschulabschluss (in %)	<b>-,456</b>	,115	<b>-,405</b>	-3,960	,000

4.1 Welche Bedeutung haben die beiden Zahlen -0.396 und -0.456 (in Zeile 2 und 3 in der ersten Zahlenspalte)?

In einem Wahlkreis mit einer um einen Prozentpunkt höheren Arbeitslosenquote ist im Mittel mit einer um 0.396 Prozentpunkte niedrigeren Wahlbeteiligung zu rechnen.

In einem Wahlkreis mit einem um einen Prozentpunkt höheren Anteil von Einwohnern ohne Hauptschulabschluss (an allen Schulabgängern) ist im Mittel mit einer um 0.456 Prozentpunkte niedrigeren Wahlbeteiligung zu rechnen.

4.2 Welche (ziemlich fiktive) Bedeutung hat die Zahl 87.375 (erste Zeile, erste Spalte)?

Ein Wahlkreis ohne Arbeitslosen, in dem alle Schulabgänger mindestens den Hauptschulabschluss erreicht haben, wäre im Mittel mit einer Wahlbeteiligung von 87.375 zu rechnen.

4.3 Wie können Sie sich erklären, dass die Vorzeichen der Koeffizienten in der ersten und dritten Spalte (zweite und dritte Zeile) beide negativ sind?

Das liegt daran, dass beide Sozialstrukturmerkmale den in 4.1 beschriebenen negativen Einfluss auf die Wahlbeteiligung haben.

4.4 Was bedeutet die Zahl 0.000 am Ende der letzten Zeile (Spalte „Sig.“) der Tabelle?

Wenn es sich bei den hier untersuchten Daten um Stichprobendaten handelte (was nicht der Fall ist), dann könnte man fast mit Sicherheit schließen, dass die Koeffizienten in den Vorspalten in der zugehörigen Grundgesamtheit von Null verschieden sind; die Wahrscheinlichkeit, in einer erneuten, gleich großen Stichprobe Koeffizienten zu ermitteln, deren Absolutwerte größer wären als die hier vorgefunden, wäre kleiner als 0.0005, unterstellt, dass in der Grundgesamtheit die Koeffizienten genau Null wären.

4.5 Welche der beiden Sozialstrukturvariablen trägt zur Erklärung der Wahlbeteiligung stärker bei, welche weniger?

Die Arbeitslosenquote trägt stärker zur Erklärung bei. Ihr „Beta“ ist absolut größer.

4.6 Begründen Sie, warum in der Spalte „B“ der Koeffizient zu „ohne Hauptschulabschluss“ in absoluten Zahlen größer ist, obwohl in der Spalte „Beta“ der Absolutwert des standardisierten Regressionskoeffizienten zu dieser Variablen der kleinere ist ( $|-0.396| < |-0.456|$  gegenüber  $|-0.488| > |-0.405|$ )?

Beta ist gleich B, multipliziert mit der Standardabweichung der erklärenden Variablen und dividiert durch die Standardabweichung der erklärten Variablen. Die erklärte Variable ist in beiden Fällen dieselbe, aber da die Standardabweichung der Arbeitslosenquote (5.21) höher als die des Anteils der Hauptschulabschlosslosen (2.25), ergibt sich das dargestellte Bild.

### 5. Multivariate Statistik / Faktorenanalyse

In der folgenden Tabelle finden Sie das Ergebnis einer Faktorenanalyse. Die Variablen beschreiben unterschiedliche Aspekte der Sozialstruktur der Bundestagswahlkreise um die Zeit der Bundestagswahl 2005. Alle Variablen sind sicher metrisch, lediglich die Variable "WestOst" nimmt nur die beiden Werte 1 für die alte Bundesrepublik und 2 für das früher so genannte „Beitrittsgebiet“ an

In der Tabelle sind die 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). Zur weiteren Verbesserung der Übersichtlichkeit sind die höchsten Ladungskoeffizienten einer Zeile fett gedruckt und die Koeffizienten unter einem Absolutwert von <0.25 weggelassen. Der jeweils höchste Wert in einer Zeile ist auch nicht fett gedruckt.

Rotated Component Matrix<sup>a</sup>

		Faktor	
		1	2
1	Arbeitslosenquote Dez 2004	<b>,924</b>	
2	Arbeitslosenquote Dez 2003	<b>,920</b>	
3	WestOst (siehe oben)	<b>,893</b>	
4	mit Hauptschulabschluss (in %)	<b>-,736</b>	-,363
5	mit Realschulabschluss (in %)	<b>,692</b>	-,390
6	Sozialversicherungspflichtig Beschäftigte am 30.06.2004 in der Landwirtschaft in %	<b>,692</b>	-,285
7	ohne Hauptschulabschluss (in %)	<b>,508</b>	
8	Sozialversicherungspflichtig Beschäftigte am 30.06.2004 im sonstigen Dienstleistungsgewerbe in %	,278	<b>,866</b>
9	Sozialversicherungspflichtig Beschäftigte am 30.06.2004 im Produzierenden Gewerbe in %	-,346	<b>-,837</b>
10	mit allgemeiner Hochschulreife (in %)		<b>,806</b>
11	Zuzüge je 1000 Einwohner	-,366	<b>,674</b>
12	Fortzüge je 1000 Einwohner	-,263	<b>,657</b>
13	Sozialversicherungspflichtig Beschäftigte am 30.06.2004 in Handel und Verkehr in %		<b>,268</b>

5.1 Geben Sie den Faktoren Namen. Wenn Sie finden, dass man einem Faktor keinen sinnvollen Namen geben kann, müssen Sie dies begründen:

Faktor	Faktorbezeichnung, -beschreibung
1	Arbeitslosigkeit und ihre Ursachen (hohe Faktorwerte finden wir in Wahlkreisen mit hoher Arbeitslosigkeit und hohen Anteilen von Schulabgängern ohne Hauptschulabschluss bzw. niedrigen Anteilen von Schulabgängern mit Hauptschulabschluss (die Ladung des Anteils der Realschulabsolventen ergibt sich einfach daraus, dass alle vier Schulabgängeranteile sich zu 100 % addieren). Dieser Interpretation steht nicht entgegen, dass auch der Anteil der in der Landwirtschaft Tätigen auf diesen Faktor lädt: in ländlichen Gebieten kann die Arbeitslosigkeit durchaus höher sein als in eher städtisch geprägten. Im übrigen hat die Sozialversicherungspflicht als solche hier überhaupt keine Bedeutung; dass sie hier erwähnt ist, sagt nur aus, dass die Daten über die Wahlkreise nicht aus einer Volkszählung stammen, sondern von der Deutschen Rentenversicherung oder der Bundesanstalt für Arbeit, d.h. aus der Sozialversicherungsstatistik.
2	Bildungsgrad (hohe Faktorwerte finden wir in Wahlkreisen, die durch Dienstleistung und einen hohen Anteil an Abiturienten geprägt sind, aber einen geringen Anteil an Industrie aufweisen; in diesen Wahlkreisen ist auch die Mobilität hoch)

5.2 Erklären Sie, warum:

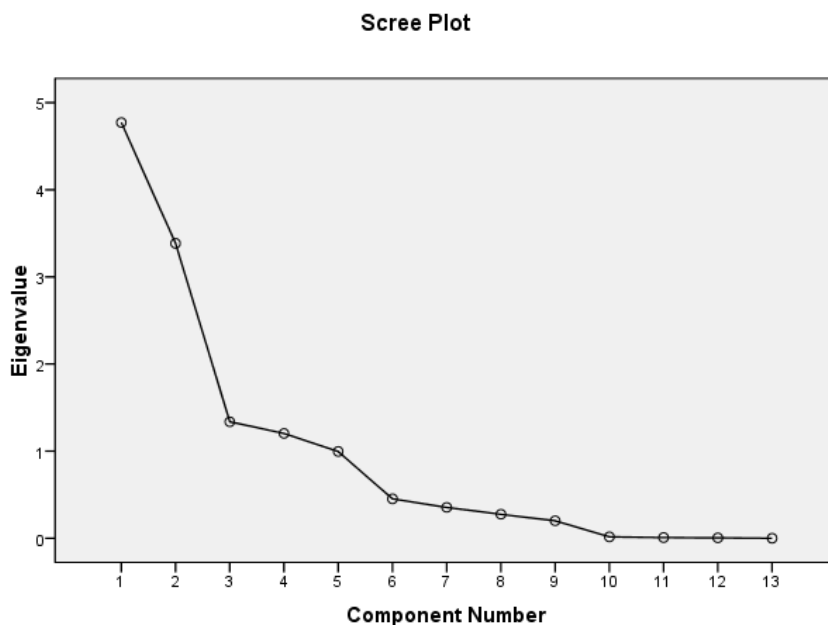
- die Vorzeichen in Spalte 1, Zeile 4 und 5
- die Vorzeichen in Spalte 2, Zeile 8 und 9 unterschiedlich sind!

Für beide Variablenpaare gilt, dass sie jeweils die größten Anteile innerhalb ihrer Gruppe (Anteil der verschiedenen Arten von Schulabgängern bzw. Anteil der verschiedenen Arten der Erwerbstätigkeit nach Wirtschaftszweigen) sind und sich zu einer Summe addieren müssen, die bei 65 bzw. 75 % liegt. Sie müssen also schon aus numerischen Gründen ziemlich hoch negativ korreliert sein (-0.60 bzw. -92)., und also kann man erwarten, dass ihre Ladungen entgegengesetztes Vorzeichen haben.

5.3 Wie beurteilen Sie die Einfachstruktur? Sind die Faktoren in den beiden Erhebungen sauber genug getrennt?

Sauber genug für eine einigermaßen einleuchtende Interpretation, aber doch nur mittelmäßig, denn es gibt in fast jeder Zeile noch eine zweite Ladung nicht zu vernachlässigender Größe.

5.4 Hätten Sie ebenfalls zwei Faktoren (oder mehr? oder weniger?) berücksichtigt? Warum? (Hier ist der Plot der Eigenwerte gegen ihre Ordnungsnummer. Die Punkte gibt es für die Begründungen, nicht für „JA“ oder „5“ oder „7“!)



Zwei Faktoren liegen nahe, weil der Graph ab dem dritten Punkte sehr deutlich flacher verläuft als vorher. Allenfalls hätte man noch fünf Faktoren nehmen können, weil nach den fünften Faktor die zugehörigen Eigenwerte noch einmal deutlich niedriger ausfallen. Auch das Kaiser-Kriterium hätte für fünf Faktoren gesprochen.