

# Statistik für Informationsmanager

## Nachklausur vom 13. Januar 2006

**Name:**

**Matrikelnummer:**

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

Einen Freiversuch habe ich vor Beginn der Klausur — nicht — angemeldet (diese Bemerkung ist rechtlich unverbindlich, es gilt nur die vorherige Anmeldung).

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

Für die Teilaufgaben wird es folgende Punktzahlen geben:

Aufgabe	Maximal erreichbar	Mittelwert	Minimum	Maximum	Standardabweichung
1	39	28	11	39	9.4
2	22	18	13	23	3.1
3	6	3.8	2	6	1.1
4.1	4	3.6	0	4	1.1
4.2	4	4.4	2	6	1.1
4.3	3	1.8	0	3	1.4
4.4	3	1.4	0	3	1.2
5.1	2	1.9	1.5	2	0.2
5.2	4	4	4	4	0
5.3	8	6.2	0	8	2.7
5.4	4	3.3	0	5	1.5
5.5	7	1.4	0	5	1.7
6.1	4	4	4	4	0
6.2	4	3.8	3	4	0.4
6.3	8	7.8	7	8	0.4
6.4	4	3.9	2	5	0.6
7	20	5.8	0	10	2.7
Summe	146	102.8	77	120	12.9

Im Hinblick darauf, dass es bei den Aufgaben [1 und] 7 außerordentlich schwierig sein wird, jeweils die volle Punktzahl zu erreichen, genügen zum Bestehen diesmal 59 Punkte, exzellente Antworten auf die Fragen 1 und 7 würden also allein schon ausreichen; für eine 1.0 sind mindestens 128 Punkte erforderlich:

59..62→4.0, 63..71→3.7, 72..79→3.3, 80..86→3.0, 87..95→2.7, 96..103→2.3, 104..111→2.0, 112..119→1.7, 120..127→1.3, 128..146→1.0!

# 1. Grundlagen

Sie haben in der Vorlesung mehrere multivariate Verfahren kennen gelernt. Jedes von ihnen erfüllt im Forschungsprozess eine oder auch mehrere spezifische Funktionen, die in der unten stehenden Übersicht aufgeführt sind. Markieren Sie in der Übersicht, welchem Verfahren welche Funktionen zukommen und geben Sie zu jeder Markierung einen kurzen Kommentar.

Funktion	Verfahren			
	Multiple Regression	Faktorenanalyse	Clusteranalyse	Diskriminanzanalyse
Hypothesenbildung		1	2	
Hypothesenüberprüfung	3			4
Skalenkonstruktion		5		
Konstruktvalidierung		6		
Vorhersage	7			8
Gruppenbildung			9	
Zuordnung zu Gruppen				10
Segmentierung			11	
Klassifikation				12
Datenreduktion		13		

Meine Kreuzchen begründe ich wie folgt (für jedes gut begründete Kreuzchen gibt es 3 Punkte, für jedes richtige, aber nicht begründete Kreuzchen gibt es einen Punkt, für jedes falsche Kreuzchen ohne oder mit nicht überzeugender Begründung wird ein Punkt abgezogen — maximal kann es 39 Punkte geben! Am besten nummerieren Sie die Kreuzchen durch und benutzen die Nummern für Ihre Begründungen):

1 Hypothesen über den Zusammenhang zwischen Variablen, über latente Variable,

2 Hypothesen über Segmente, Gruppen etc.

3 Überprüfung von Hypothesen über Zusammenhänge zwischen Variablen

4 Überprüfung von Hypothesen über Zusammenhänge zwischen einer Gruppierungs- und mehreren metrischen Variablen

5 Konstruktion einer oder mehrerer Skalen für latente Variable

6 Überprüfung, ob Items einer Fragebatterie zu einer einzigen Dimension gehören

7 Vorhersage von künftigen Messwerten der abhängigen Variablen bei einem zusätzlichen Fall, wenn dessen Werte in den unabhängigen Variablen bekannt sind

8 Vorhersage der Gruppenzugehörigkeit bei einem zusätzlichen Fall, wenn dessen Werte in den unabhängigen Variablen bekannt sind

9 Auffinden von Gruppen, Clustern

10 wie 8

11 wie 9

12 wie 10

13 Bilden weniger Skalen aus vielen Indikatoren

Es gab auch für andere Kreuzchen Punkte, wenn die Begründungen einigermaßen überzeugend waren.

## 2. Skalenniveaus

Geben Sie jeweils für die folgenden Variablen an, ob sie nominal, ordinal, intervall- oder ratioskaliert sind (jeweils ein Punkt) und geben Sie jeweils eine kurze Begründung (jeweils ein weiterer Punkt). (Eine gute Begründung für eine unserer Ansicht nach falsche Einordnung gibt eventuell ebenfalls Punkte, falsch gesetzte Kreuzchen ohne oder mit nicht überzeugender Begründung führen zu einem Punktabzug! Maximal sind 22 Punkte möglich, denn in einigen Fällen lassen sich zwei Kreuzchen pro Zeile gut begründen.)

	Nominal	Ordinal	Intervall	Ratio	Begründung
Aktienkurs				X	prozentualer Anstieg ist sinnvoll definierbar
Zählerstand am Strom- oder Gaszähler bzw. an der Tanksäule (Vorsicht!)			X	X	Tanksäule für den einzelnen Tankvorgang: ratio, angezeigte Menge und angezeigter Preis sind proportional; alle anderen Fälle: nur intervall, denn von der Differenz zweier Zählerstände hängt z.B. der Preis ab
Geburtenrate					prozentualer Anstieg ist sinnvoll definierbar
Dienstgrad beim Militär		X			gleiche „Abstände“ bedeuten nicht automatisch gleiche Unterschiede im Kompetenzumfang
Bezeichnung eines Flugzeugtyps (A310, A320, A321, A340, A380, ..., Boeing 737, 747, ..., DC3, DC6, DC8, DC9, ...)	X				innerhalb einer Linie möglicherweise (nach Größe etc.) ordinal, sonst nur Typennamen
Listenplatz bei einer Kandidatenaufstellung		X			gleiche Abstände bedeuten nicht automatisch gleiche Popularitätsdifferenz
Hausnummer	X	X			in vielen Städten gilt: je höher die Nummer, desto weiter von der Stadtmitte entfernt; sonst nur nominal
Examensnote			X		prozentualer Anstieg ist nicht sinnvoll definierbar

## 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. Monatseinkommen in Westdeutschland, Eurobarometer 2003, 2275 Befragte mit gültigen Angaben:

<751€	751..875€	876..1000€	1001..1125€	1126..1150€	1151..1375€	1376..1500€	1501..1750€	1751..2000€	2001..2250€	2251..2500€	>2500€
5.4%	3.3%	6.9%	5.3%	4.4%	5.0%	7.2%	8.2%	10.6%	10.3%	8.0%	25.5%
Mittelwert:	8.14, wenn man die Gruppen von 1 bis 13 numeriert, siehe unten										
Median	9, d.h. der Median liegt innerhalb der 9. Gruppe, also zwischen 1750 und 2000 — die Person, für die gilt, dass eine Hälfte der Befragten mehr und die andere weniger verdient als sie, muss in Gruppe 9 liegen, denn unterhalb von Gruppe 9 liegen 45.7 % und oberhalb liegen 43.8 %										
Modus	12										
gegebenenfalls Begründung für nicht berechnete Parameter:	Der Mittelwert ist bei dieser Klasseneinteilung nicht sinnvoll, die Klassenbreiten sind zu unterschiedlich (750, 125, 250, unbestimmt).										

2. Links-rechts-Einstellung auf einer Skala von 1 („ganz links“) bis 10 („ganz rechts“) — Angaben von 12083 repräsentativ ausgewählten Befragten aus den damaligen EU-Staaten im Eurobarometer 2003

1	2	3	4	5	6	7	8	9	10
4.7%	5.0%	11.9%	13.1%	29.3%	12.8%	10.2%	7.8%	2.7%	2.4%
Mittelwert:	5.09, wenn man die Variable als intervallskaliert betrachtet								
Median	5, unterhalb von Gruppe 5 liegen 34.7 %, oberhalb liegen 35.9 %, also muss der mediane Befragte in Gruppe 5 liegen.								
Modus	5								
gegebenenfalls Begründung für nicht berechnete Parameter:									

3. Region, in der die in Österreich Befragten des Eurobarometer 2003 wohnen:

Vorarlberg	Tirol	Salzburg	Oberösterreich	Steiermark	Kärnten	Niederösterreich	Burgenland	Wien
15	29	22	60	53	25	68	12	70
Mittelwert:	Mittlere Befragtenzahl pro Bundesland macht keinen Sinn							
Median	keine Ordnung							
Modus	Wien							
gegebenenfalls Begründung für nicht berechnete Parameter:	einige fanden 68>70 oder hat den letzten Eintrag nicht gesehen							

### 4. Bivariate Statistik

Im Eurobarometer 2003 wurden die Befragten nach ihrem Einkommen befragt, das dann einheitlich in vier Kategorien codiert wurde (nach Quartilen). Außerdem haben die Interviewer aufgezeichnet, ob die Befragten auf dem Land, in einer kleineren oder in einer größeren Stadt wohnten. Hier sind die Antwortverteilungen in Griechenland und in Deutschland:

D29 INCOME HH QUANTILES (HARMONISED) \* D25 TYPE OF COMMUNITY Crosstabulation

		D25 TYPE OF COMMUNITY			Total	
		1.rural area or village	2.small or middle sized town	3.large town		
D29 INCOME HH QUANTILES (HARMONISED)	--	Count	244	237	230	711
		% within D25 TYPE OF COMMUNITY	26.8%	21.4%	31.2%	25.8%
	-	Count	191	270	172	633
		% within D25 TYPE OF COMMUNITY	21.0%	24.4%	23.3%	23.0%
	+	Count	247	336	160	743
		% within D25 TYPE OF COMMUNITY	27.2%	30.4%	21.7%	27.0%
++	Count	227	264	176	667	
	% within D25 TYPE OF COMMUNITY	25.0%	23.8%	23.8%	24.2%	
Total	Count	909	1107	738	2754	
	% within D25 TYPE OF COMMUNITY	100.0%	100.0%	100.0%	100.0%	

a. Member State = DEUTSCHLAND

D29 INCOME HH QUANTILES (HARMONISED) \* D25 TYPE OF COMMUNITY Crosstabulation

		D25 TYPE OF COMMUNITY			Total	
		1.rural area or village	2.small or middle sized town	3.large town		
D29 INCOME HH QUANTILES (HARMONISED)	--	Count	32	10	12	54
		% within D25 TYPE OF COMMUNITY	37.6%	23.8%	10.7%	22.6%
	-	Count	32	5	14	51
		% within D25 TYPE OF COMMUNITY	37.6%	11.9%	12.5%	21.3%
	+	Count	16	15	39	70
		% within D25 TYPE OF COMMUNITY	18.8%	35.7%	34.8%	29.3%
++	Count	5	12	47	64	
	% within D25 TYPE OF COMMUNITY	5.9%	28.6%	42.0%	26.8%	
Total	Count	85	42	112	239	
	% within D25 TYPE OF COMMUNITY	100.0%	100.0%	100.0%	100.0%	

a. Member State = ELLAS

Land	$\Phi$	$\gamma$	Sig. $\chi^2$	N
EU 15	0.049	+0.033	<0.0005	10532
Belgien	0.131	-0.027	0.666	236
Dänemark	0.119	-0.105	0.856	186
<b>Deutschland</b>	<b>0.108</b>	<b>-0.045</b>	<b>&lt;0.0005</b>	<b>2754</b>
<b>Griechenland</b>	<b>0.503</b>	<b>+0.562</b>	<b>&lt;0.0005</b>	<b>239</b>
Italien	0.145	-0.015	<0.0005	1388
Spanien	0.126	+0.042	0.006	1141
Frankreich	0.219	+0.229	<0.0005	1788
Irland	0.109	+0.080	0.997	50
Luxemburg	0.500	+0.263	0.868	10
Niederlande	0.180	-0.137	0.024	451
Portugal	0.149	-0.147	0.391	284
Großbritannien	0.081	-0.081	0.218	1277
Finnland	0.171	+0.143	0.492	186
Schweden	0.107	-0.085	0.739	308
Österreich	0.135	-0.019	0.634	237

1. Handelt es sich bei den beiden Variablen um nominal oder um ordinal skalierte Variable? (4 Punkte)  
ordinal

2. Beschreiben Sie den Zusammenhang zwischen den beiden Variablen an Hand der beiden Kreuztabellen für Deutschland und Griechenland. Benutzen Sie dazu auch die beiden Zeilen mit den Assoziationsmaßen und der Signifikanzbetrachtung. (4 Punkte)

In Griechenland ist das Einkommen umso höher, je städtischer die Wohnregion ist, in Deutschland ist der Zusammenhang umgekehrt (Einkommen umso höher, je weniger städtisch die Wohnregion ist), aber kaum stark ausgeprägt: der kleinste Anteil armer Leute findet sich in Mittelstädten.

3. In Portugal sind sowohl  $\Phi$  als auch  $\gamma$  größer als in Italien. Warum ist der Zusammenhang in Italien signifikant und in Portugal nicht? (3 Punkte)

Weil in Italien die Stichprobe fast fünfmal so groß ist.

4. In Vergleich der beiden  $\gamma$  ist der Unterschied zwischen Griechenland und den Niederlanden am größten. Erklären Sie, warum dieses Ergebnis plausibel sein könnte. (3 Punkte)

In den Niederlanden ist offenbar noch ausgeprägter die Armut in den großen Städten konzentriert als in Deutschland.

### 5. Multivariate Statistik / Multiple Regression

Für alle (bisher 66, davon 63 mit vollständigen Daten in der Datenbank) Absolventen des Bachelor-Studiengangs Informationsmanagement liegt außer den Noten für einzelne Leistungsnachweise auch die Gesamtnote vor. Sie kann als abhängige Variable in einer Regression auf einige wichtige Klausuren der ersten beiden Semester untersucht werden, etwa um Studienanfängern frühzeitig ihre Erfolgsaussichten zu verdeutlichen. Die folgende Übersicht gibt die wichtigsten Koeffizienten wieder:

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.784 <sup>a</sup>	.615	.588	.2762

a. Predictors: (Constant), Inf1Note, BWLNote, MathNote, StatNote

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7.068	4	1.767	23.161	.000 <sup>a</sup>
	Residual	4.425	58	.076		
	Total	11.493	62			

a. Predictors: (Constant), Inf1Note, BWLNote, MathNote, StatNote  
b. Dependent Variable: Minvonpnote

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.781	.157		4.973	.000
	BWLNote	.077	.040	.168	1.916	.060
	MathNote	.088	.040	.204	2.233	.029
	StatNote	.243	.058	.389	4.180	.000
	Inf1Note	.160	.048	.324	3.345	.001

a. Dependent Variable: Minvonpnote

5.1 Was bedeuten die beiden Zahlen **0.781** und **0.077** in den ersten beiden Zeilen der Koeffiziententabelle? (2P).

0.781 wäre die beste Schätzung für jemand, der in allen vier Fächern eine (praktisch unmögliche) 0.0 erhalten hat. Um 0.077 unterscheiden sich im Mittel zwei Studenten in ihrer Gesamtnote, die in BWL um (praktisch mögliche) 1.0 auseinander gelegen haben.

5.2 Berechnen Sie die zu erwartende Gesamtnote eines/r Studierenden, der/die in Informatik I für Informationsmanager (InfNote) eine 2.0 und in den anderen Klausuren jeweils eine 3.0 erreicht hat. (4 P).

$$0.781 + 0.160 \cdot 2 + (0.077 + 0.088 + 0.243) \cdot 3 = 2.325$$

5.3 Aus welcher der vier Noten lässt sich am besten, aus welcher am wenigsten gut auf die spätere Gesamtnote schließen? (2 Punkte) Wie können Sie sich dieses Ergebnis erklären? (noch einmal 6 Punkte)

Aus der Statistiknote: alle einschlägigen Koeffizienten sind hier am höchsten, und das Signifikanzniveau ist < 0.0005, eine Note Differenz in dieser Klausur hat im Mittel fast eine Viertelnote Unterschied auf dem Abschlusszeugnis zur Folge! Vielleicht liegt das daran, dass für das Bestehen dieser Klausur sowohl einiges mathematisches Verständnis, allgemeines Weltwissen und sprachliche Kompetenz erforderlich sind.

5.4 Drücken Sie in einem ganzen Satz aus, was die Zahl **0.029** (dritte Zeile von unten, letzte Spalte) bzw. ihr Kehrwert 34.5 bedeutet! (4 P)

Nur in 2.9 % aller (bzw. in 2 aus 69) gleich großen Zufallsstichproben aus der Grundgesamtheit aller potentiellen Teilnehmer an diesen Klausuren und Studierenden in diesem Studiengang würde man ein  $|B| > 0.077$  bzw.  $|Beta| > 0.168$  oder ein  $|t| > 2.233$  finden, wenn in der Grundgesamtheit die entsprechenden Größen gerade 0 wären. (Absolutbeträge, es handelt sich um einen zweiseitigen Test! Das Weglassen des Hinweises auf den Absolutbetrag wurde aber nicht gewertet.)

5.5 Die Varianz der Gesamtnote beträgt 0.18533025 (die Standardabweichung ist 0.4305),  $R^2$  (die „Varianzreduktion“) ist (wie man oben sieht) 0.615. Was können Sie mit diesen drei Zahlen anfangen? (7 Punkte)

Die Kenntnis aller vier Einzelklausuren reduziert die Varianz des Schätzfehlers bei einer Schätzung der Gesamtnote aus eben diesen Einzelangaben um 61.5 %, wenn man sie mit der Varianz des Schätzfehlers bei einer gänzlich uninformierten Schätzung vergleicht. Die Varianz des informierten Schätzfehlers ist also nur 0.07135215, die Standardabweichung entsprechend 0.26711823.

## 6. Multivariate Statistik / Faktorenanalyse

Im Eurobarometer 2003 wurden neben den Fragen zur Internetnutzung auch mehrere Fragen zum Thema Datenschutz gestellt. In der Tabelle unten finden Sie die Faktorladungen der Einzelantworten auf diese Fragebatterie. Die Fragen wurden folgendermaßen eingeleitet: „Für jede der folgenden Aussagen sagen Sie mir bitte ob Sie eher zustimmen oder ablehnen!“

Da es sich nur um wenige Variable handelt, sind diesmal die Koeffizienten kleiner als 0.11 nicht weggelassen, außerdem sind die Variablen auch nicht nach ihren höchsten Faktorladungen sortiert. Interpretieren Sie die beiden Faktoren, indem Sie ihnen möglichst prägnante Namen geben. Begründen Sie Ihre Namensgebung.

Rotated Component Matrix	Component	
	1	2
Man sollte sich darüber informieren, warum Organisationen personenbezogene Daten sammeln und ob sie sie an andere Organisationen weitergeben.	.578	.053
Das Niveau der Schutzes personenbezogener Daten in (Ihrem Land) ist hoch.	-.048	.816
Das Bewusstsein für den Schutz personenbezogener Daten in (Ihrem Land) ist gering.	.539	-.096
Sie sind besorgt, wenn Sie im Internet persönliche Informationen wie Name, Adresse, geburtsdatum und Geschlecht hinterlassen.	.572	-.127
Die Gesetzgebung (Ihres Landes) wird mit der wachsenden Zahl von Leuten fertig, die personenbezogene Informationen im Internet hinterlassen.	.010	.763
Organisation in (Ihrem Land) die personenbezogene Informationen verarbeiten, sollten diese ohne Ihre Zustimmung nicht an eine ähnliche Organisation außerhalb der EU weitergeben dürfen.	.675	.133
Die meisten Länder außerhalb der EU habe keine Datenschutzgesetze, die so effektiv sind wie die der EU.	.516	.381
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. a Rotation converged in 3 iterations.		

6.1 Faktor 1 beschreibt die latente Eigenschaft (Dimension, Einstellung der Befragten) ... (4 P)

Mangelhaftes Vertrauen in die Sorgfalt Betroffener bei datenschutzrelevanten Aktivitäten.

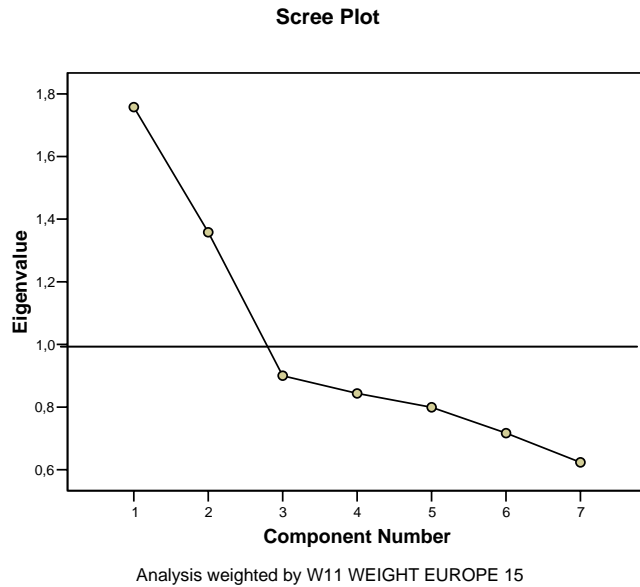
6.2 Faktor 2 ... (4 P)

Vertrauen der Befragten in die nationale oder supranationale Gesetzgebung

6.3 Wie beurteilen Sie die Einfachstruktur? Besteht Grund zu der Annahme, dass es sich hier um ein Artefakt handeln könnte (dass die Einfachstruktur das Ergebnis einer ungeschickten Fragebogenkonstruktion sein könnte)? (8 P)

Bei vier der sieben Observablen liegt eine der Ladungen unter 0.1, bei drei sogar unter 0.055, bei zwei weiteren nur knapp über 0.1. Bei der siebenten Observablen ist es nicht so erstaunlich, dass die Ladung auf beide Faktoren ähnlich hoch ist, denn sie bezieht sich nicht so ausdrücklich auf bestimmte Gesetze.

6.4 Auf der nächsten Seite finden Sie den Scree-Plot zu der vorstehenden Auswertung:



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

Eindeutig ja ...

Wenn ja: warum? (2 P)

Zwei der Punkte in der Graphik liegen deutlich sowohl über der Geraden durch (0,1) (Kaiser-Kriterium), haben also einen deutlich höheren Varianzbeitrag als eine einzelne Observable. Außerdem liegen sie auch noch deutlich oberhalb der Kurve, die von den anderen fünf Punkten gebildet wird. Hinzu kommt, dass die Einfachstruktur befriedigend war, die Faktoren relativ überzeugend interpretiert werden konnten und die beiden Faktoren knapp die Hälfte der Gesamtvarianz ( $1.758 + 1.358 = 0.44513 \cdot 7$ ).

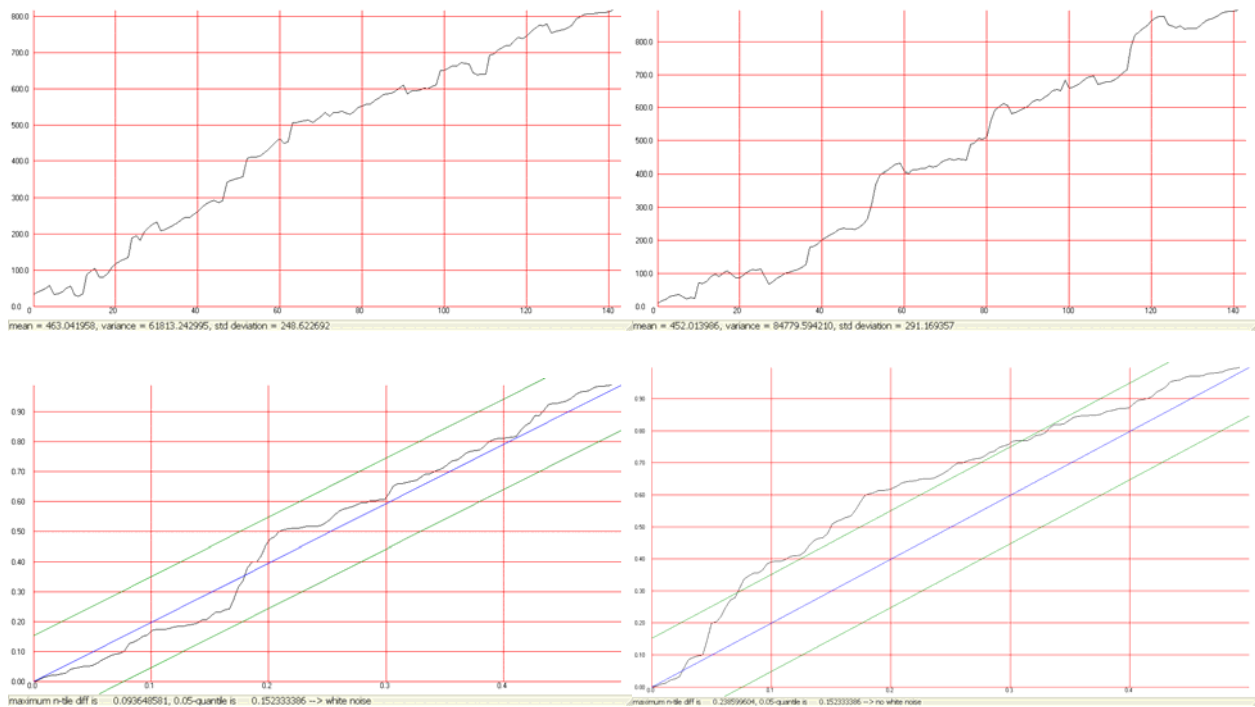
Wenn nein: warum nicht? Wie viele hätte man statt dessen extrahieren sollen? (2 P)

Nur einen Faktor oder gar drei zu extrahieren, kann man nicht begründen (es gibt natürlich trotzdem maximal 4 Punkte!).

(Antworten auf beide Unterfragen werden gewertet! Wenn Sie dezidiert der Auffassung sind, dass nicht zwei, sondern einer oder drei 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.)

## 7. Zeitreihenanalyse

In einem Gesellschaftsspiel erzielten im Laufe von 143 Runden die beiden Spieler die (kumulierten) Punktzahlen, die aus den beiden oberen Abbildungen ersichtlich sind; in den beiden Abbildungen darunter finden Sie die kumulierten Periodogramme der jeweils einmal differenzierten kumulierten Punktzahlen (d.h. die kumulierten Periodogramme der Zeitreihen der jeweils in einer Runde neu hinzugewonnenen Punkte). Die beiden äußeren Geraden in den Periodogramm-Graphiken begrenzen jeweils den 95%-Konfidenzbereich des kumulierten Periodogramms weißen Rauschens.



Was können Sie den Graphiken entnehmen? Was sagen sie über den Spielverlauf aus? In welchem Maße beherrscht der Zufall das Spiel? Haben die Punkte aus früheren Runden Einfluss auf die Punkte aus späteren Runden? Was kann man aus der Lage der Kurven der kumulierten Periodogramme in den Konfidenzregionen abzuleiten? (Für diese Aufgabe gibt es maximal 20 Punkte)

Meine Frau und ich haben einen ganzen Abend lang Tridom gespielt, um die beiden Zeitreihen zu erzeugen ...

Meine Frau hat einen Random Walk hingelegt, bei mir muss es irgendeine verborgene Systematik geben, jedenfalls sind die Punkte, die ich in den einzelnen Runden erzielt habe, nicht ganz unabhängig vom Spielverlauf. Das kumulierte Periodogramm der Einzelrunden-Punkte meiner Frau liegt ganz innerhalb des 95%-Konfidenzintervall, d.h. es weicht vom Erwartungswert des kumulierten Periodogramm eines echten white-noise Prozesses weniger ab als weit mehr als 5 % kumulierte Periodogramme einzelner Realisierungen von white-noise-Prozessen. Bei mir liegen Teile der Kurve außerhalb des 95 %-Konfidenzbandes, was bedeutet, dass dieser Prozess sich so stark von einem white-noise-Prozess unterscheidet, wie das für weit weniger als 5 % aller gleich langen Realisierungen eines white-noise-Prozess gelten kann. Man kann ferner den oberen Graphiken entnehmen, dass im Mittel die Punktzugewinne in den einzelnen Runden positiv sind, dass aber auch Punktverluste vorkommen. Die Varianz der einzelnen Gewinne und Verluste ist bei meiner Frau geringer als bei mir, dafür habe ich insgesamt etwas besser abgeschnitten.

Maximal konnten leider nur 10 Punkte für eine Interpretation der beiden oberen Graphiken vergeben werden, die kumulierten Periodogramme hat niemand als solche zu interpretieren vermocht. Man beachte: auch wenn das absichtlich schwer zu erkennen war: Die waagerechte Skala der oberen Graphiken läuft von 0 bis zur Rundenzahl, die der unteren aber nur von 0 bis 0.5 (der maximalen Frequenz) — oben haben wir ein „Geräusch“, unten sein Spektrum. Mein Spielverlauf ist offenbar der eines ARIMA(2,1,0)-Prozesses, d.h. meine Gewinne und Verluste wirken noch bis zu zwei Runden nach: wenn ich zweimal hintereinander gute Punktzahlen erreichte habe, ist die Wahrscheinlichkeit hoch, dass ich auch in der folgenden Runde eine hohe Punktzahl erreiche (aber das war dem Periodogramm denn doch nicht zu entnehmen ...).