

**Einführung in die angewandte Statistik
für Soziologen / Sozialwissenschaftler**

Abschlussklausur WiSe 2003/04

Name: _____

Vorname: _____

Matr. Nr.: _____

Einige Hinweise zur Organisation:

1. Ihnen stehen zwei Zeitstunden für die Bearbeitung der Aufgaben zur Verfügung.
2. Sie können innerhalb der ersten 15 Minuten den Raum endgültig verlassen. Anderenfalls bleiben Sie bitte bis zum Ende der Bearbeitungszeit ruhig auf Ihrem Platz.
3. Bitte nehmen Sie die geheftete Klausur bei der Bearbeitung nicht auseinander.
4. Versuchen Sie, möglichst prägnant und gut lesbar auf die Fragen zu antworten.
5. Wenn der vorgegebene Platz nicht ausreichen sollte, können Sie die leeren Rückseiten benutzen.

Viel Erfolg bei der Bearbeitung der Klausur!

Punkte:	
Note	

Aufgabe 2:

In derselben Befragung wie in Aufgabe 1 wurde darüber hinaus zum einen erhoben, wie zufrieden die Mütter insgesamt mit ihrer Arbeit sind bzw. an ihrem letzten Arbeitsplatz waren und inwieweit sie von ihrer Familie sagen, dass dort alle viel Spaß miteinander haben. Das Ergebnis ist in der nachfolgenden Tabelle enthalten.

Zufriedenheit mit Tätigkeit * Haben viel Spaß miteinander in Familie Kreuztabelle

			Haben viel Spaß miteinander in Familie				Gesamt
			Immer	Häufig	Selten	Nie	
Zufriedenheit mit Tätigkeit	sehr zufrieden	Anzahl	156	150	9	0	315
		% von Zufriedenheit mit Tätigkeit	49,5%	47,6%	2,9%	,0%	100,0%
		% Spaß in der Familie	36,9%	29,7%	22,5%	,0%	32,5%
		% der Gesamtzahl	16,1%	15,5%	,9%	,0%	32,5%
		Standardisierte Residuen	1,6	-1,1	-1,1	-,8	
	eher zufrieden	Anzahl	220	282	18	1	521
		% von Zufriedenheit mit Tätigkeit	42,2%	54,1%	3,5%	,2%	100,0%
		% Spaß in der Familie	52,0%	55,8%	45,0%	50,0%	53,7%
		% der Gesamtzahl	22,7%	29,1%	1,9%	,1%	53,7%
		Standardisierte Residuen	-,5	,7	-,8	-,1	
	eher unzufrieden	Anzahl	40	61	10	0	111
		% von Zufriedenheit mit Tätigkeit	36,0%	55,0%	9,0%	,0%	100,0%
		% Spaß in der Familie	9,5%	12,1%	25,0%	,0%	11,4%
		% der Gesamtzahl	4,1%	6,3%	1,0%	,0%	11,4%
		Standardisierte Residuen	-1,2	,4	2,5	-,5	
	sehr unzufrieden	Anzahl	7	12	3	1	23
% von Zufriedenheit mit Tätigkeit		30,4%	52,2%	13,0%	4,3%	100,0%	
% Spaß in der Familie		1,7%	2,4%	7,5%	50,0%	2,4%	
% der Gesamtzahl		,7%	1,2%	,3%	,1%	2,4%	
Standardisierte Residuen		-1,0	,0	2,1	4,4		
Gesamt	Anzahl	423	505	40	2	970	
	% von Zufriedenheit mit Tätigkeit	43,6%	52,1%	4,1%	,2%	100,0%	
	% Spaß in der Familie	100,0%	100,0%	100,0%	100,0%	100,0%	
	% der Gesamtzahl	43,6%	52,1%	4,1%	,2%	100,0%	

2a. Welche statistischen Zusammenhangsmaße würden Sie heranziehen, um den Zusammenhang der beiden Variablen zu bestimmen. Bitte begründen Sie, welche Koeffizienten Ihnen welche Informationen liefern.

2b. Welchen Wert wird der berechnete Korrelationskoeffizient Ihrer Einschätzung ungefähr annehmen? Inwieweit sind in diesem Zusammenhang die berechneten standardisierten Residuen hilfreich für Ihre Einschätzung?

2c. Bitte interpretieren Sie unter Verwendung der aufgeführten Zellenwerte das zentrale Ergebnis der obigen Kreuztabelle.

Aufgabe 3:

In einer repräsentativen Befragung von 8-bis 10-jährigen Kindern und ihren Müttern wurden u.a. verschiedene Aspekte des familialen Binnenklimas aus Sicht der Kinder (ITEMxK) und aus Sicht der Mütter (ITEMxM) erfragt (Antwortvorgaben: „nie“, „selten“, „häufig“, „immer“).

- Item1 Ich bin gerne mit meiner Familie zusammen.
- Item2 In unserer Familie kommt es zu Reibereien.
- Item3 In unserer Familie haben wir viel Spaß miteinander.

Korrelationen: Kendall-Tau-b

		ITEM1K	ITEM2K	ITEM3K	ITEM1M	ITEM2M	ITEM3M
ITEM1K	r	1,000	-,056	,259**	,187**	-,139**	,192**
	Sig. (2-seitig)		,058	,000	,000	,000	,000
	N	1032	1029	1028	1030	1027	1027
ITEM2K	r	-,056	1,000	-,142**	-,092**	,215**	-,094**
	Sig. (2-seitig)	,058		,000	,002	,000	,001
	N	1029	1030	1026	1028	1025	1025
ITEM3K	r	,259**	-,142**	1,000	,138**	-,161**	,261**
	Sig. (2-seitig)	,000	,000		,000	,000	,000
	N	1028	1026	1029	1027	1024	1024
ITEM1M	r	,187**	-,092**	,138**	1,000	-,219**	,357**
	Sig. (2-seitig)	,000	,002	,000		,000	,000
	N	1030	1028	1027	1040	1037	1037
ITEM2M	r	-,139**	,215**	-,161**	-,219**	1,000	-,292**
	Sig. (2-seitig)	,000	,000	,000	,000		,000
	N	1027	1025	1024	1037	1037	1034
ITEM3M	r	,192**	-,094**	,261**	,357**	-,292**	1,000
	Sig. (2-seitig)	,000	,001	,000	,000	,000	
	N	1027	1025	1024	1037	1034	1037

** . Korrelation ist auf dem Niveau von 0,01 signifikant (2-seitig).

a. Interpretieren Sie die zentralen Ergebnisse der Korrelationsanalyse.

Aufgabe 4:

In einer nicht-repräsentativen Befragung von 18- bis 21-jährigen Jugendlichen ging es u.a. darum, die Vorstellungen von einer eigenen späteren Familie möglichst differenziert zu erfassen. U. a. wurde erfragt, in welchem Alter die Jugendlichen glauben, selbst erstmalig Vater oder Mutter zu werden. Von Interesse war dabei, inwieweit die verschiedenen Teilgruppen sich in dieser Frage unterscheiden. Die Frage wurde mithilfe einer einfachen Varianzanalyse untersucht.

4a. Formulieren Sie ein geeignetes Hypothesenpaar begründen Sie die Art (ein- oder zweiseitig) Ihre Fragestellung.

4b. Worauf beziehen sich die beiden Hypothesen:

auf Grundgesamtheiten

auf Stichproben

Begründung:

4c. Unter welchen Voraussetzungen macht der Einsatz einer einfachen Varianzanalyse Sinn?

4d. Die nachfolgenden Tabellen zeigen die Ergebnisse der einfachen Varianzanalyse.

ONEWAY deskriptive Statistiken

Alter beim ersten Kind

	N	Mittelwert	Standardabweichung	Standardfehler	Minimum	Maximum
Studierende	226	28,48	2,354	,157	22	37
GymnasiastInnen	168	27,78	3,421	,264	3	38
Auszubildende	190	25,51	3,421	,248	2	35
Wehrdienstleistende	185	27,13	3,239	,238	3	40
Zivildienstleistende	156	28,62	3,522	,282	20	55
Gesamt	925	27,50	3,366	,111	2	55

ONEWAY ANOVA

Alter beim ersten Kind

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	1202,668	4	300,667	29,851	,000
Innerhalb der Gruppen	9266,569	920	10,072		
Gesamt	10469,237	924			

Test der Homogenität der Varianzen

Anzahl Kinder

Levene-Statistik	df1	df2	Signifikanz
22,881	4	904	,000

Robuste Testverfahren zur Prüfung auf Gleichheit der Mittelwerte

Anzahl Kinder

	Statistik ^a	df1	df2	Sig.
Welch-Test	12,489	4	432,836	,000
Brown-Forsythe	10,509	4	795,974	,000

a. Asymptotisch F-verteilt

Mehrfachvergleiche

Abhängige Variable: Alter beim ersten Kind

Scheffé-Prozedur

(I) Befragten-Gruppe	(J) Befragten-Gruppe	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	99%-Konfidenzintervall	
					Untergrenze	Obergrenze
Studierende	GymnasiastInnen	,70	,323	,318	-,48	1,88
	Auszubildende	2,97*	,312	,000	1,83	4,11
	Wehrdienstleistende	1,35*	,315	,001	,20	2,50
	Zivildienstleistende	-,13	,330	,997	-1,34	1,07
GymnasiastInnen	Studierende	-,70	,323	,318	-1,88	,48
	Auszubildende	2,27*	,336	,000	1,04	3,50
	Wehrdienstleistende	,65	,338	,450	-,59	1,89
	Zivildienstleistende	-,84	,353	,231	-2,13	,45
Auszubildende	Studierende	-2,97*	,312	,000	-4,11	-1,83
	GymnasiastInnen	-2,27*	,336	,000	-3,50	-1,04
	Wehrdienstleistende	-1,62*	,328	,000	-2,82	-,42
	Zivildienstleistende	-3,10*	,343	,000	-4,36	-1,85
Wehrdienstleistende	Studierende	-1,35*	,315	,001	-2,50	-,20
	GymnasiastInnen	-,65	,338	,450	-1,89	,59
	Auszubildende	1,62*	,328	,000	,42	2,82
	Zivildienstleistende	-1,49*	,345	,001	-2,75	-,22
Zivildienstleistende	Studierende	,13	,330	,997	-1,07	1,34
	GymnasiastInnen	,84	,353	,231	-,45	2,13
	Auszubildende	3,10*	,343	,000	1,85	4,36
	Wehrdienstleistende	1,49*	,345	,001	,22	2,75

*. Die mittlere Differenz ist auf der Stufe .01 signifikant.

Einführung in die Statistik für SozialwissenschaftlerInnen

Klausur am 15.02.2005

Name: _____ **Vorname:** _____

Matrikelnummer: _____

Studienfächer _____

Einige Hinweise zur Organisation:

Ihnen stehen zwei Zeitstunden für die Bearbeitung der Aufgaben zur Verfügung.

Es sind alle Fragen zu beantworten. Sie können innerhalb der ersten 15 Minuten den Raum endgültig verlassen. Anderenfalls bleiben Sie bitte bis zum Ende der Bearbeitungszeit ruhig auf Ihrem Platz.

Bitte nehmen Sie die geheftete Klausur bei der Bearbeitung nicht auseinander.

Versuchen Sie, möglichst prägnant und gut lesbar auf die Fragen zu antworten.

Wenn der vorgegebene Platz nicht ausreichen sollte, können Sie die leeren Rückseiten benutzen.

Viel Erfolg bei der Bearbeitung der Klausur!

Punkte	
Note	

Dr. Anna Brake

Aufgabe 1:

In der Nacht auf den 15. April 1912 sank die als unsinkbar geltende Titanic, nachdem sie auf ihrer Jungfernfahrt im Nordatlantik mit einem Eisberg zusammengestoßen war. Das Unglück forderte 1.503 Todesopfer. Von den insgesamt 1343 Passagieren auf der Titanic waren 337 in der ersten Klasse und 285 in der zweiten Klasse an Bord (die 885 Besatzungsmitglieder werden an dieser Stelle nicht betrachtet). 40% der Passagiere aus der ersten Klasse wurden nach dem Unglück vermisst, von den Passagieren der zweiten Klasse verloren insgesamt 160 ihr Leben und in der dritten Klasse waren dies 541 Passagiere.

- 1 a.) Erstellen Sie auf der Basis der Informationen eine Kreuztabelle, bei der das Merkmal „Passagierklasse“ die Spaltenvariable darstellt. Bestimmen Sie die Zellenhäufigkeiten als Anzahl n und als Zeilenprozente.

n				
%				
n				
%				
n				
%				

- b) Zeigen die Daten einen Zusammenhang zwischen Überlebenschance und Passagierklasse? Beschreiben Sie das Ergebnis.

- c) Welches Assoziationsmaß würden Sie heranziehen, um die Stärke des Zusammenhangs zu bestimmen? Begründung?

b) Welche Information kann der Höhe der standardisierten Residuen entnommen werden?

c) Um die Art des Zusammenhangs näher zu untersuchen, wurde als PRE-Maß Goodman- und Kruskal-Tau berechnet.

Richtungsmaße

		Wert	Asymptotischer Standardfehler	Näherungsweise Signifikanz
Goodman-und-Kruskal-Tau	Geschlecht abhängig	,078	,039	,078
	keine Probleme mit Statistik-Klausur abhängig	,009	,006	,452

Interpretieren Sie bitte die Ergebnisse.

Aufgabe 3:

Für das in Aufgabe 2 beschriebene Antwortverhalten auf einer Zustimmungsskala von „trifft nicht zu“ (= 1) und „trifft zu“ (= 5) wurden die folgende deskriptiven Statistiken berechnet.

Univariate Statistiken

			Statistik	Standardfehler
keine Probleme mit Statistik-Klausur	Mittelwert		2,92	,100
	95% Konfidenzintervall des Mittelwerts	Untergrenze	2,72	
		Obergrenze	3,12	
	erstes Quartil		2,15	
	Median		2,98	
	drittes Quartil		3,720	
	Standardabweichung		1,050	
	Minimum		1	
	Maximum		5	
	Spannweite		4	
	Interquartilbereich		1,57	
	Schiefte		-,173	,230
	Kurtosis		-,591	,457

Bitte beschreiben Sie unter Verwendung der angemessenen statistischen Kennwerte das Antwortverhalten der 110 Studierenden.

Aufgabe 4:

z-Transformation: Welche Bedingungen müssen erfüllt sein, damit eine z-Transformation der Daten sinnvoll durchgeführt werden kann? Was geschieht mit den ursprünglichen Messwerten, wenn diese z-transformiert werden? Beschreiben Sie einen Anwendungsfall, wann eine solche Transformation der Daten benötigt wird.

Aufgabe 5:

Am Lehrstuhl für Soziologie und empirische Sozialforschung wird erwogen, ein neues Softwarepaket für die statistische Datenanalyse anzuschaffen. Dieses Programmpaket wird als sehr anwenderInnenfreundlich beschrieben, ist aber in der Anschaffung mit sehr hohen Kosten verbunden. Hinzu käme, dass neue Lehrbücher angeschafft und MitarbeiterInnenschulungen durchgeführt werden müssten. Daher soll zunächst überprüft werden, ob die Studierenden mit der neuen Software einen größeren Lernerfolg erzielen. Mit zufällig ausgewählten Studierenden wird ein Übungskurs mit der alten, parallel ein Übungskurs mit der neuen Software durchgeführt. Am Ende wird eine benotete Klausur geschrieben. Formulieren Sie ein dieser Situation angemessenes Hypothesenpaar. Worauf beziehen sich Ihre Hypothesen? Entscheiden und begründen Sie, welchen Test Sie zur Überprüfung heranziehen würden. Wie legen Sie im vorliegenden Fall die Irrtumswahrscheinlichkeit α fest? Warum?

Aufgabe 6:

In einer Untersuchung über die statistikbezogenen Einstellungen von Studierenden wurden diese u.a. gebeten einzuschätzen, inwieweit folgende Aussagen auf sie zutreffen (Antwortskala: „trifft nicht zu“ = 1 und „trifft zu“ = 5):

- Ich glaube, dass mir Statistik Spaß machen wird.
- Die Statistik-Klausur wird mir keine Probleme machen.
- Statistik ist leider ein notwendiges Übel.
- Mit der Statistik wird doch sowieso nur rumgetrickst.

Die Auswertung der Daten ergab folgende Korrelationsmatrix.

Korrelationen: Kendall-Tau-b

		dass mir Statistik Spass machen wird	keine Probleme mit Statistik-Klausur	Statistik: ein notwendiges Übel	Statistik: wird nur rumgetrickst
dass mir Statistik Spass machen wird	Korrelationskoeffizient	1,000	,431**	-,482**	,023
	Sig. (2-seitig)	.	,000	,000	,778
	N	109	109	108	108
keine Probleme mit Statistik-Klausur	Korrelationskoeffizient	,431**	1,000	-,367**	-,135
	Sig. (2-seitig)	,000	.	,000	,097
	N	109	110	109	109
Statistik: ein notwendiges Übel	Korrelationskoeffizient	-,482**	-,367**	1,000	,049
	Sig. (2-seitig)	,000	,000	.	,543
	N	108	109	109	109
Statistik: wird nur rumgetrickst	Korrelationskoeffizient	,023	-,135	,049	1,000
	Sig. (2-seitig)	,778	,097	,543	.
	N	108	109	109	109

** . Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Interpretieren Sie die zentralen Ergebnisse der Korrelationsanalyse.

Aufgabe 7:

Verfügen männliche und weibliche Studierende in unterschiedlichem Maß über Geld zur freien Verfügung? An einer Zufallsauswahl von 96 Studierenden wurde diese Frage untersucht. Der t-Test zeigt folgende Ergebnisse.

Kolmogorov-Smirnov-Anpassungstest

		Geld zur freien Verfügung
N		96
Parameter der Normalverteilung ^a	Mittelwert	260,8125
	Standardabweichung	58,98623
Extremste Differenzen	Absolut	,081
	Positiv	,065
	Negativ	-,081
Kolmogorov-Smirnov-Z		,791
Asymptotische Signifikanz (2-seitig)		,560

a. Die zu testende Verteilung ist eine Normalverteilung.

Gruppenstatistiken

Geschlecht		N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Geld zur freien Verfügung	männlich	35	275,1429	51,82388	8,75983
	weiblich	60	252,6667	62,14981	8,02351

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit	
		F	Signifikanz
Geld zur freien Verfügung	Varianzen sind gleich	2,465	,120
	Varianzen sind nicht gleich		

Test bei unabhängigen Stichproben

		T-Test für die Mittelwertgleichheit		
		T	df	Sig. (2-seitig)
Geld zur freien Verfügung	Varianzen sind gleich	1,804	93	,075
	Varianzen sind nicht gleich	1,892	81,801	,062

a) Was genau besagt die in diesem t-Test ausgewiesene „Sig. (2-seitig)“?

Aufgabe 8:

In einem Abschlussbericht finden sich die nachfolgenden Aussagen. Bitte kommentieren Sie diese.

Aussage 1: „Angesichts dieses statistisch höchstsignifikanten Ergebnisses muss an dieser Stelle unbedingt etwas passieren.“

Aussage 2: „Da die Nullhypothese nicht mit der erforderlichen Irrtumswahrscheinlichkeit verworfen werden konnte, ist nachgewiesen, dass sich Männer und Frauen in dieser Frage nicht unterscheiden.“
