

Übungsaufgaben zu Kapitel 1

1.1. Beschreiben Sie in eigenen Worten die folgenden Begriffe:

- | | |
|-------------------------|------------------------|
| (a) Population | (f) Merkmalsausprägung |
| (b) Stichprobe | (g) Messung |
| (c) Beobachtungseinheit | (h) kategoriale Daten |
| (d) Inferenzstatistik | (i) metrische Daten |
| (e) Variable | (j) Skalenniveau |

1.2. Beantworten Sie bitte die folgenden Fragen für die untenstehenden drei Beispiele *i) - iii)*:

- Was ist im jeweiligen Beispiel die Population und was die Stichprobe ?
- Wurde eine geeignete Stichprobe ausgewählt ? (Wenn Sie diese Frage verneinen, geben Sie Gründe für Ihre Antwort an, und beschreiben Sie, wie die Population und die Beobachtungseinheiten definiert sein müssten, und welche Stichprobe erhoben werden sollte.)
- Welche und wie viele Variablen wurden erhoben ?
- Welche Ausprägungen haben die erhobenen Variablen ?
- Geben Sie für jede Variable an, ob sie kategorial oder metrisch ist ?

i) In Beispiel 3 (Kapitel ??) wollte der Herausgeber eines akademischen Buchverlages wissen, ob zwischen der Anzahl vergebener Freixemplare und dem erzielten Erlös eines bestimmten Buches ein Zusammenhang besteht. Zu diesem Zweck untersuchte er Daten aller Vertreter, wie sie in folgender Tabelle wiedergegeben sind.

Nummer des Repräsentanten	Bruttoerlös in US Dollar	Anzahl vergebener Freixemplare
1305	2086,00	106
1307	63093,00	337
1327	41017,00	182
⋮	⋮	⋮

ii) Ein Bürgermeister einer Kleinstadt möchte den Bedarf für Kindergärten erheben. Dazu lässt er 50 Frauen an einem Samstag vormittags in der Fußgängerzone befragen wieviele Kinder sie haben, wie alt sie sind, ob sie ganztags, halbtags oder gar nicht arbeiten, und ob sie glauben, dass ein Bedarf für mehr Kindergärten besteht.

iii) Ein Hersteller von Computerchips behauptet, dass in seiner Produktion weniger als 1 Prozent defekte Chips anfallen. Im Zuge einer Qualitätskontrolle werden aus der Produktion eines Tages 100 Chips zufällig ausgewählt und überprüft.

1.3. Nennen Sie Beispiele für Situationen, in denen eine Vollerhebung und in denen eine Stichprobenerhebung sinnvoll ist.

1.4. Welche der folgenden Variablen sind kategorial, welche metrisch ?

- | | |
|--------------------------------|--|
| (a) Sozialversicherungsnummer | (f) Anzahl der Geschwister |
| (b) Geburtsdatum | (g) Schulnoten |
| (c) Geschlecht | (h) Studienrichtung |
| (d) Körpergröße | (i) Einkommen |
| (e) Hausnummer der Wohnadresse | (j) Entfernung zwischen Wohn- und Studienort |

Übungsaufgaben zu Kapitel 2

- 2.1. Die folgende Tabelle gibt an wieviele Wochenzeitschriften in den einzelnen österreichischen Bundesländern 1996 erschienen sind (Quelle: Statistisches Jahrbuch für Österreich 1997)

B	K	N	O	S	ST	T	V	W
6	9	50	17	10	13	23	10	15

- (a) Berechnen Sie die relativen Häufigkeiten
 (b) Geben Sie ein grafische Darstellung
- 2.2. Beim Kolloquium aus Statistik für Wirtschaftspädagogik im Wintersemester 1998 gab es folgende Noten:

3	4	3	5	4	4	1	1	3	2	2	1	1	1	2	2	3	5	3	4
3	1	1	3	3	1	1	4	4	3	5	3	3	4	1	3	5	3	1	5
4	5	2	2	5	3	1	1	4	1	4	1	1	1	5	4	5	4	4	5
1	5	2	2	2	1	1	4	4	4	1	5	5	2	4	2	4	3	1	1
4	4	4	2	1	3														

- (a) Erstellen Sie eine Häufigkeitsverteilung
 (b) Berechnen Sie die relativen und kumulierten relativen Häufigkeiten
 (c) Geben Sie ein grafische Darstellung
 (d) Wie hoch ist der Prozentsatz jener, die mindestens mit „gut“ abgeschnitten haben ?
- 2.3. Ein Marktforschungsinstitut möchte testen, ob ein neues Cola von Konsumenten positiv aufgenommen wird. Zu diesem Zweck bekommt jede von 200 Testpersonen vier nummerierte, aber sonst unetikettierte, neutrale Flaschen mit Cola. Zwei enthalten die Produkte der Marktführer (C und P), eine enthält Cola eines weniger bekannten Herstellers (A) und eine das neue Cola (N). Die 200 Testpersonen sollten angeben, welches der vier Getränke ihnen am besten geschmeckt hat. Die Häufigkeiten waren

A	C	P	N
42	56	43	59

- (a) Prüfen Sie die Fragestellung, ob alle vier Getränke gleich gut schmecken.
 (b) Geben Sie ein grafische Darstellung der beobachteten und der erwarteten Häufigkeiten.
 (c) Interpretieren Sie das Ergebnis.
- 2.4. Ein Computerprogramm soll gleichverteilte Zufallszahlen erzeugen, und zwar die natürlichen Zahlen 0 bis 9. Der Programmierer erhält nach einem Testlauf von 10000 Versuchen folgende Häufigkeiten:

0	1	2	3	4	5	6	7	8	9
956	998	1043	1059	968	985	1087	1042	967	895

- (a) Prüfen Sie die Fragestellung, ob das Programm funktioniert.
 (b) Geben Sie ein grafische Darstellung der beobachteten und der erwarteten Häufigkeiten.
 (c) Interpretieren Sie das Ergebnis.

- 2.5. Die Produkte A, B und C sollen nach den Angaben eines Marktforschungsinstitutes Marktanteile von 45%, 25% und 17% haben. Der Hersteller von B bezweifelt diese Angaben, da er weniger, als nach diesem Ergebnis zu erwarten gewesen wäre, verkauft hat. Seine eigene repräsentative Untersuchung ergab die folgende Häufigkeitsverteilung verkaufter Produkte:

A	B	C	sonstige
242	110	98	50

- (a) Prüfen Sie die Fragestellung, ob die Angaben des Marktforschungsinstitutes stimmen.
- (b) Geben Sie ein grafische Darstellung der beobachteten und der erwarteten Häufigkeiten.
- (c) Interpretieren Sie das Ergebnis.
- 2.6. In zwei Drittel aller US Bundesstaaten wurde zwischen 1985 und 1990 die Anlegepflicht von Sicherheitsgurten eingeführt. Dass dadurch die Anzahl tödlicher Verletzungen sinken würde, war relativ klar. Allerdings wusste man wenig über die Auswirkungen auf nicht tödliche Verletzungen. Man wusste, dass in der Vergangenheit 49% aller Unfälle mit Totalschäden an Fahrzeugen ohne Verletzungen von Personen abliefen, bei 39% war ambulante Versorgung notwendig, bei weiteren 8% mussten Personen in ein Krankenhaus eingeliefert werden und bei 1% gab es tödliche Verletzungen. In einer Studie in Virginia (Quelle: Iman, A Data-based Approach to Statistics, 1994) wurde nach Einführung der Gurtenanlegepflicht folgende Daten von 525 Unfällen erhoben:

keine Verletzungen	ambulante Behandlung	stationäre Behandlung	tödliche Verletzungen	unbekannt
316	154	34	9	12

- (a) Prüfen Sie die Fragestellung, ob die Einführung der Gurtenanlegepflicht Auswirkungen auf die Art von Verletzungen hatte.
- (b) Geben Sie ein grafische Darstellung der beobachteten und der erwarteten Häufigkeiten.
- (c) Interpretieren Sie das Ergebnis.
- 2.7. Bei einer Befragung von 296 repräsentativ ausgewählten Studenten wurden unter anderem folgende Angaben gemacht: 42% wohnen bei ihren Eltern, 19% leben in einer festen Gemeinschaft mit einem Partner, 51% müssen nebenbei arbeiten, 11% haben Kinder.
- Berechnen Sie 95%-Konfidenzintervalle für die jeweiligen Anteile.
- 2.8. In einer Online Umfrage gab es nach 60 abgegebenen Meinungen folgenden Zwischenstand zu der Frage „Wie glauben Sie wird die künftige Regierung aussehen?“

SPÖ-ÖVP	FPÖ-ÖVP	ÖVP-FPÖ	SPÖ-ÖVP	nur SPÖ
61%	12%	14%	4%	9%

- (a) Berechnen Sie 95%-Konfidenzintervalle für die jeweiligen Anteile.
- (b) Kann man auf Grund der Daten schliessen, dass eine Mehrheit an das Zustandekommen einer SPÖ-ÖVP Regierung glaubt ?

Übungsaufgaben zu Kapitel 3

3.1. Die folgende Tabelle stellt die Anzahl von Beschäftigten in den Hauptberufsgruppen in den USA für die Jahre 1979 und 1989 dar.

	Anzahl der Beschäftigten (in Tausend)	
	1979	1989
Industrie (Produktion)	26461	25326
Service (Privat)	47416	65318
Regierung	15947	17769
Summe	89824	108413

- (a) Berechnen Sie für jedes Jahr die relativen Häufigkeiten dieser Hauptberufsgruppen.
- (b) Zeichnen Sie gruppierte Balkendiagramme.

3.2. An 25 Teilnehmern eines Statistikproseminars an der WU wurden die Variablen *Geschlecht* mit den Ausprägungen 0 (weiblich) und 1 (männlich) und *Note im Mathematikproseminar* erhoben:

Geschlecht:	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	0	0	1	0		
Note:	4	4	4	2	3	4	3	3	3	1	4	4	3	2	4	3	4	3	4	3	4	4	4	3	2

- (a) Erstellen Sie die zweidimensionale Häufigkeitsverteilung (Kreuzklassifikation) für die Variablen *Geschlecht* und *Note*.
- (b) Bestimmen Sie die relativen Häufigkeiten für die drei Informationsarten: Gemeinsame Information, Randinformation und bedingte Information.
- (c) Erstellen Sie grafische Darstellungen für die 3 Informationsarten.

3.3. Die folgende Tafel enthält Daten aus dem „1984 General Social Survey“ (USA). Die beiden Variablen sind *Einkommen* und *Berufszufriedenheit*.

Einkommen in US\$	Berufszufriedenheit			
	sehr unzufrieden	ein wenig unzufrieden	mäßig zufrieden	sehr zufrieden
< 6000	20	24	80	82
6000-15000	22	38	104	125
15000-25000	13	28	81	113
> 25000	7	18	54	92

- (a) Bestimmen Sie die Prozentsätze für die drei Informationsarten: Gemeinsame Information, Randinformation und bedingte Information.
- (b) Erstellen Sie grafische Darstellungen für die 3 Informationsarten.

3.4. Sozialbeihilfenempfänger können bei Gericht Berufung einlegen, wenn Sie sich ungerecht behandelt fühlen. Dabei besteht die Möglichkeit, dass sie sich bei der Anhörung durch einen Anwalt vertreten lassen oder selbst Beschwerde einlegen. Das Resultat der Verhandlung kann entweder eine Erhöhung, Verminderung oder überhaupt keine Änderung der Betragshöhe der Sozialbeihilfe ergeben. In der folgenden Tabelle werden 320 Fälle angeführt:

Art der Vertretung	Ergebnis der Verhandlung		
	Erhöhung	keine Änderung	Verminderung
Anwalt	59	108	17
selbst	70	63	3

- Geben Sie eine geeignete grafische Darstellung zur Illustration dieser Daten.
- Bestimmen Sie die „Chancen“ (odds) auf Erhöhung, keine Änderung bzw. Verminderung, wenn man durch einen Anwalt vertreten wird.
- Um wie viel Mal ist die „Chance“ höher eine Erhöhung gegenüber keiner Änderung zu erzielen, wenn man sich selbst vertritt?

3.5. Jemand beobachtet das Verhalten von Personen bei der Entnahme von Sonntagszeitungen aus Zeitungsständern und ordnet dieses Verhalten drei Gruppen zu:

- Ungeniertes Klauen: Entnahme der Zeitung ohne auch nur angedeutetes Zahlen
- Genierliches Klauen: Beobachtung der Umgebung, Einwerfen einiger Münzen (meist der 10-Groschen-Kategorie), rasches Verlassen des Tatortes
- Ehrlich währt am längsten: Demonstratives Einwerfen der Münzen, erster Überblick über die Zeitung beim Zeitungständer und Blickkontrolle des Verhaltens anderer Passanten

Die Beobachtungen mehrerer Sonntage, bei denen auch das *Alter* der „Käufer“ grob den Kategorien *jung*, *mittel* und *alt* zugeordnet wurde, sind in folgender Tabelle zusammengefasst:

Alter	Entnahmeverhalten		
	Ungeniert	Genierlich	Ehrlich
jung	30	12	8
mittel	21	45	24
alt	9	23	28

Die interessierende Frage ist, ob das „Kaufverhalten“ in den Altersgruppen gleich ist?

- Berechnen Sie geeignete relative Häufigkeiten.
- Geben Sie eine geeignete grafische Darstellung zur Illustration dieser Daten.
- Prüfen Sie die Fragestellung mittels χ^2 -Test bei einem Sicherheitsniveau von 95% und interpretieren Sie das Ergebnis.

3.6. Um die Wirkung der Erdbeschaffenheit auf das Wachstum neuer Hybridpflanzen zu erforschen, wurden Schößlinge in 3 verschiedene Erdarten eingepflanzt und ihr Wachstum in 3 Kategorien klassifiziert.

Wachstum	Erdsorten		
	Ton	Sand	Lehm
schlecht	16	8	14
mittelmäßig	31	16	21
gut	18	36	25

Die Fragestellung ist, ob die drei Erdarten zu unterschiedlichem Wachstum führen.

- Berechnen Sie geeignete relative Häufigkeiten.
- Geben Sie eine geeignete grafische Darstellung zur Illustration dieser Daten.
- Prüfen Sie die Fragestellung mittels χ^2 -Test bei einem Sicherheitsniveau von 95% und interpretieren Sie das Ergebnis.
- Wäre das Ergebnis anders ausgefallen, hätte man ein Sicherheitsniveau von 90% bzw. 99% zugrunde gelegt?

3.7. Die folgenden Daten von 49 amerikanischen Collegestudenten enthalten das Geschlecht und das Fach in dem eine Abschlussprüfung gemacht wurde. (Die Angaben bedeuten: W ... weiblich , M ... männlich , sowie S ... Soziologie, B ... Biologie, P ... Physik und G... Geisteswissenschaften.)

Nr.	Geschlecht	Fach	Nr.	Geschlecht	Fach	Nr.	Geschlecht	Fach
1	W	S	18	M	B	35	M	S
2	M	P	19	W	S	36	W	B
3	M	S	20	M	S	37	M	B
4	M	B	21	M	S	38	M	S
5	W	P	22	W	B	39	W	P
6	W	B	23	M	S	40	M	S
7	M	G	24	M	G	41	M	B
8	M	B	25	M	S	42	M	B
9	M	S	26	M	B	43	W	S
10	W	B	27	M	P	44	M	P
11	M	G	28	M	B	45	M	S
12	M	S	29	W	P	46	M	S
13	M	P	30	W	B	47	W	S
14	W	P	31	W	B	48	M	B
15	M	S	32	M	B	49	M	S
16	M	B	33	M	G			
17	M	P	34	M	B			

Unterscheiden sich weibliche von männlichen Studenten bezüglich der Auswahl des Abschlussfaches ? Erstellen Sie eine geeignete Kontingenztafel, geben Sie eine grafische Darstellung, testen Sie die Fragestellung und kommentieren Sie das Ergebnis.

3.8. Ein Marktforscher glaubt, dass es in den vier Städten A,B,C und D keinen Unterschied in den Präferenzen für bestimmte Fernsehprogramme gibt. Um die Vermutung zu testen, werden unabhängige Stichproben von 150, 200, 250 und 200 Fernsehteilnehmern aus den vier Städten ausgewählt und befragt: 'Welchen Typ von Fernsehprogramm ziehen Sie vor'. Das Ergebnis ist in folgender Tabelle enthalten:

Programmtyp	Stadt			
	A	B	C	D
Filme	50	70	85	60
Theater	45	50	58	40
Sport	35	50	72	57
wiss. Send.	20	30	35	33

Ist die Vermutung des Marktforschers richtig ?

3.9. Im Übungsbeispiel 3 ging es um darum, ob es eine Beziehung zwischen der Höhe des Einkommens und Berufszufriedenheit gibt. Stellen Sie die Daten grafisch dar, prüfen Sie die Vermutung, dass diese beiden Variablen voneinander unabhängig sind und interpretieren Sie das Ergebnis.

3.10. Bei 205 Ehepaaren wurde die Körpergrösse der beiden Partner jeweils in die Kategorien *gross*, *mittel* und *klein* eingeteilt. Das Ergebnis der Kreuzklassifikation in Prozent ist folgendes:

		Ehefrau			gesamt
		gross	mittel	klein	
Ehemann	gross	8,8%	13,7%	6,8%	29,3%
	mittel	9,8%	24,9%	13,7%	48,3%
	klein	5,9%	12,2%	4,4%	22,4%
gesamt		24,4%	50,7%	24,9%	100,0%

Die Fragestellung lautet, ob die Körpergröße bei Ehepaaren voneinander unabhängig ist. Geben Sie eine geeignete grafische Darstellung, prüfen Sie Fragestellung mittels χ^2 -Test und interpretieren Sie das Ergebnis.

3.11. Um das Heiratsalter von Frauen in zwei ethnischen Gruppen A und B zu untersuchen, wurde jeweils eine Stichprobe von 100 verheirateten Frauen ausgewählt. Im Zuge der Untersuchung antworteten die Frauen auf die Frage „Heirateten Sie zum ersten Mal bevor Sie 19 Jahre alt waren“:

	Gruppe	
	A	B
ja	62	53
nein	38	47

- (a) Bestimmen Sie die „Chance“ (odds) vor 19 zu heiraten für beide Gruppen
- (b) Bestimmen Sie das „Verhältnis der Chancen“ (odds ratio) vor 19 zu heiraten für die Gruppe A gegenüber B.
- (c) Errechnen Sie ein 95%-Konfidenzintervall für das odds ratio und interpretieren Sie das Ergebnis.
- (d) Ergibt der χ^2 -Test ein anderes Ergebnis ?

3.12. Die folgende Tafel enthält Straßenverkehrsunfälle die 1988 vom Department of Highway Safety and Motor Vehicles in Florida erhoben wurden. Die beiden Variable sind *Verletzung* mit den beiden Ausprägungen 'tödlich' und 'nicht tödlich' und *Verwendung von Sicherheitsgurten* mit den beiden Ausprägungen 'ja' und 'nein'.

Gurt verwendet	Verletzung	
	tödlich	nicht tödlich
nein	1601	162527
ja	510	412368

- (a) Bestimmen Sie das „Risiko“ (odds) tödlich zu verunglücken für die beiden Gruppen
- (b) Prüfen Sie die Vermutung, dass die Verwendung von Sicherheitsgurten zu einer Reduktion des Anteils tödlicher Verletzungen bei Verkehrsunfälle führt mittels odds ratio.

Übungsaufgaben zu Kapitel 4

4.1. Die folgende Tabelle zeigt das Alter aller US Präsidenten bei ihrer Inauguration.

	Name	Alter bei Amtsantritt	Name	Alter bei Amtsantritt	
1.	Washington	57	23.	B. Harrison	55
2.	J. Adams	61	24.	Cleveland	55
3.	Jefferson	57	25.	McKinley	54
4.	Madison	58	26.	T. Roosevelt	42
5.	Monroe	58	27.	Taft	51
6.	J. Q. Adams	57	28.	Wilson	56
7.	Jackson	61	29.	Harding	55
8.	Van Buren	54	30.	Coolidge	51
9.	W. H. Harrison	68	31.	Hoover	54
10.	Tyler	51	32.	F. D. Roosevelt	51
11.	Polk	49	33.	Truman	60
12.	Taylor	64	34.	Eisenhower	62
13.	Fillmore	50	35.	Kennedy	43
14.	Pierce	58	36.	L. B. Johnson	55
15.	Buchanan	65	37.	Nixon	56
16.	Lincoln	52	38.	Ford	61
17.	A. Johnson	56	39.	Carter	52
18.	Grant	46	40.	Reagan	69
19.	Hayes	54	41.	Bush	64
20.	Garfield	49	42.	Clinton	46
21.	Arthur	50	42.	Bush jr	54
22.	Cleveland	47			

- (a) Berechnen Sie den Mittelwert und die Standardabweichung des Inaugurationsalters.
 (b) Stellen Sie diese Daten in einem Histogramm dar.
 (c) Berechnen Sie den Median, Q_1 und Q_3 unter Verwendung eines Stem-and-Leaf Plots.
 (d) Verwenden Sie einen Box-Plot zur graphischen Darstellung.
 (e) Kommentieren Sie das Ergebnis. Welche Darstellungsweise (numerisch und graphisch) erscheint Ihnen sinnvoll ?
- 4.2. Der folgende Stem-and-leaf Plot (*Stemwidth* od. *Stammweite* ist 10) zeigt die Ergebnisse, die Studenten eines Soziologielehrganges bei der Endklausur erzielt haben.

Stem-and-Leaf Display der Ergebnisse	
2	48
3	155
4	002
5	03368
6	0124479
7	22355689
8	004577
9	0025

- (a) Berechnen Sie Mittelwert und Varianz.
 (b) Zeichnen Sie ein Histogramm.
 (c) Berechnen Sie den Median und zeichnen Sie diesen, wie auch den Mittelwert in das Histogramm ein.
 (d) Berechnen Sie die Quartile Q_1 und Q_3 .
 (e) Welcher Anteil der Studenten erzielte eine Punkteanzahl von höchstens 70 ? Mindestens 80 ?

- 4.3. Die folgenden Daten entstammen Aufzeichnungen über Zahnpastaverbrauch und geben Auskunft über die Verwendungsdauer (in Wochen) einer Tube Zahnpasta in Haushalten:

(Beachten Sie bitte, daß in englischer Schreibweise das Dezimalzeichen durch eine Punkt und nicht, wie im Deutschen, durch ein Komma dargestellt wird. In vielen international verwendeten Computerprogrammen finden Sie daher die in diesem Beispiel gewählte Schreibweise. Überdies wird oft die Null vor dem Komma weggelassen. Die erste Zahl in diesem Beispiel muß daher als 0,74 gelesen werden. Achten Sie daher stets auf den Zusammenhang in dem Zahlen dargestellt werden.)

.74 .45 .80 .95 .84 .82 .78 .82 .89 .75 .76 .81
 .85 .75 .89 .76 .89 .99 .71 .77 .55 .85 .77 .87

- (a) Stellen Sie diese Daten in einem Stem-and-Leaf Plot dar.
 (b) Berechnen Sie die relative Häufigkeit jenes Verbrauchs, der .80 nicht übersteigt.
 (c) Berechnen Sie den Mittelwert und die Standardabweichung.
 (d) Berechnen Sie den Median und die Quartile.
- 4.4. In einer medizinischen Studie wurden bei 43 Personen die Blutfettwerte untersucht. Man erhielt folgende Daten:

239 212 249 227 218 310 281 330 226 233
 223 161 195 233 249 284 245 174 154 256
 196 299 210 301 199 258 205 195 227 244
 355 234 195 179 357 282 265 286 286 176
 195 163 297

- (a) Überlegen Sie sich eine geeignete Gruppierung dieser Daten.
 (b) Zeichnen Sie ein Histogramm und beschreiben Sie die Form der Verteilung.
 (c) Berechnen Sie die Intervalle $\bar{x} \pm s$, sowie $\bar{x} \pm 2s$,
 (d) Finden die jeweiligen Anteile der Beobachtungen, die innerhalb dieser Intervalle liegen.
- 4.5. Gegeben sind sind 26 Messungen über sauren Regen in Wisconsin.

3.58 3.80 4.01 4.05 4.12 4.18
 4.20 4.30 4.32 4.35 4.50 4.52
 4.57 4.58 4.60 4.61 4.61 4.65
 4.72 4.73 4.78 4.79 5.07 5.40
 5.41 5.48

- (a) Stellen Sie diese Daten in einem Stem-and-Leaf Plot dar.
 (b) Berechnen Sie den Mittelwert und die Standardabweichung.
 (c) Berechnen Sie den Median und die Quartile.
 (d) Zeichnen Sie einen Box-Plot.
- 4.6. Bestimmen Sie für eine standardisierte Normalverteilung, $Z \sim N(0, 1)$ die folgenden Wahrscheinlichkeiten:
 (a) $P(Z \leq 0.44)$, (b) $P(Z < -1.28)$, (c) $P(Z \geq -0.67)$, (d) $P(Z > 1.2)$,
 (e) $P(-0.33 < Z < 1.64)$, (f) $P(1.1 \leq Z \leq 1.96)$, (g) $P(-3 < Z < -0.12)$.
- 4.7. Berechnen Sie für eine standardisierte Normalverteilung, $Z \sim N(0, 1)$ den Wert z , sodaß (a) die Fläche links davon 0.0869 beträgt, (b) die Fläche links davon 0.12 beträgt, (c) die Fläche rechts davon 0.2578 beträgt, (d) die Fläche rechts davon 0.25 beträgt.

- 4.8. Die Variable X sei normalverteilt mit $\mu = 100$ und $\sigma = 8$. Berechnen Sie:
(a) $P(X < 107)$, (b) $P(X < 97)$, (c) $P(X > 110)$, (d) $P(X > 90)$,
(e) $P(95 < X \leq 106)$, (f) $P(103 < X \leq 114)$, (g) $P(88 < X \leq 100)$.
- 4.9. Angenommen die Dauer X (Tage) der problemlosen Benützung eines neuen Staubsaugers sei normalverteilt, $X \sim N(530, 10000)$.
- (a) Wie groß ist die Wahrscheinlichkeit, daß man den Staubsauger mindestens 2 Jahre problemlos benutzen kann?
- (b) Die Firma möchte die Garantiedauer so ansetzen, daß nicht mehr als 10% der Staubsauger ein Reparaturservice während der Garantiedauer benötigen. Für wie lange sollte die Garantieleistung gewährt werden?
- 4.10. Die Punktzahl bei einem Aufnahmetest an einem College folge einer Normalverteilung mit $\mu = 500$ and $\sigma = 200$. Das College beschließt nur die 15% Besten aufzunehmen. Wieviele Punkte müßten Sie mindestens erreichen, damit Sie an diesem College studieren könnten?
- 4.11. Für folgende Daten soll grafisch, mittels eines QQ-Plots, untersucht werden, ob sie einer Normalverteilung folgen:
- 3 4 5 11 25
- 4.12. Von der Zufallsvariablen X mit $\sigma^2 = 36$ wird eine Stichprobe vom Umfang 36 gezogen: $\bar{x} = 108$. Es ist ein 95%-Konfidenzintervall für den Mittelwert anzugeben.
- 4.13. Wie groß muß der Stichprobenumfang in der vorigen Aufgabe gewählt werden, damit das Konfidenzintervall eine Länge von höchstens 2 hat?
- 4.14. Die Reißfestigkeit X eines bestimmten Verpackungsmaterials in kg sei normalverteilt, mit $X \sim N(\mu, 49)$. Berechnen Sie ein 95%-Konfidenzintervall für die durchschnittliche Reißfestigkeit, wenn der Mittelwert von 40 Proben mit $\bar{x} = 52.3$ errechnet wurde.
- 4.15. In einem Baumarkt wurde über einen Zeitraum von 20 Wochen die Anzahl verkaufter Rollen Maschendrahtzaun aufgezeichnet. Es ergab sich eine mittlere Verkaufszahl von 32 Rollen pro Woche. Die Standardabweichung wurde mit 5 errechnet. Bestimmen Sie ein 95%-Konfidenzintervall für die mittlere Anzahl Rollen pro Woche.
- 4.16. Es gelte $X \sim N(\mu, 4)$. Eine Stichprobe vom Umfang 25 ergab $\bar{x} = 14.7$. Es ist die Hypothese $H_0 : \mu = 14$ mit $\alpha = 0.05$ gegen die Alternativen (a) $H_A : \mu > 14$, (b) $H_A : \mu < 14$ und (c) $H_A : \mu \neq 14$ zu testen.
- 4.17. Aus 25 Schadensmeldungen bei einer Versicherung wurden eine mittlere Schadenssumme von $\bar{x} = 21200$ berechnet (σ sei bekannt, $\sigma = 1800$). (a) Führen Sie einen Test von $H_0 : \mu = 19500$ gegen die zweiseitige Alternativhypothese mit $\alpha = 0.05$ durch, (b) könnte ein Test gegen die einseitige Alternativhypothese $H_A : \mu > 19500$ (bei gleichem α) das Schicksal von H_0 ändern?
- 4.18. Aus der Erfahrung sei bekannt, daß die Therapiedauer für eine Krankheit im Durchschnitt 15 Tage (bei einer Standardabweichung von 3 Tagen) beträgt. Von einer neuen Therapie wird erwartet, daß sie die durchschnittliche Therapiedauer reduziert. Um diese Vermutung zu überprüfen, wurden 70 Patienten mit der neuen Therapie behandelt. Es ergab sich ein Mittelwert von 12,3 Tagen. Formulieren Sie eine geeignete Null- und eine Alternativhypothese und überprüfen Sie diese.

Übungsaufgaben zu Kapitel 5

- 5.1. Die Siegerzeiten aus den 400-Meter freestyle" Schwimmwettbewerben der Männer in Olympischen Spielen (1956 - 1976) sind in der folgenden Tabelle dargestellt:

	Siegerzeiten					
Jahr	1956	1960	1964	1968	1972	1976
Zeit	4:27.30	4:18.30	4:12.20	4:09.00	4:00.27	3:51.93

Zeichnen Sie ein Streudiagramm (x -Achse: Jahr).

- 5.2. Würden Sie eine positive, negative oder (fast) Nullkorrelation bei den folgenden Beispielen annehmen? Begründen Sie Ihre Antworten.
- Das Gewicht eines Autos und der durchschnittliche Bezinverbrauch (Liter/100km)
 - Intelligenzwerte von Ehefrauen und Ehemännern.
 - Das Alter eines Flugzeuges und jener Zeit, die das Flugzeug einsatzbereit ist (d.h. wenn es nicht repariert oder gewartet werden muß).
 - Aktienpreise von IBM und General Motors.
 - Aktienpreise von IBM und Microsoft.
 - Die Temperatur bei einem Baseballspiel und Bierverkauf.

- 5.3. Berechnen Sie den Korrelationskoeffizienten $r_{x,y}$ für folgende Werte und verwenden dafür Sie die beiden vereinfachten Formeln.

x :	9	3	5	0	8
y :	4	2	3	1	3

- 5.4. Der Teergehalt von Zigaretten wird oft durch die folgende Methode geprüft: Eine motorisierte Rauchmaschine" (smoking machine) macht pro Minute einen Zug, der jeweils zwei Sekunden dauert, bis eine bestimmte Länge des Zigarettenstummels erreicht ist. Der gesamte Teergehalt wird durch eine Laboranalyse des in der Maschine gesammelten Rauches festgestellt. Natürlich wird dieser Prozeß mehrmals mit mehreren Zigaretten verschiedener Marken wiederholt, um den durchschnittlichen Teergehalt festzustellen. In der folgenden Tabelle sind Werte des durchschnittlichen Teergehaltes und die durchschnittliche Anzahl von Zügen, sechs unterschiedlicher Filterzigarettenmarken angegeben.

durchschnittlicher Teergehalt (mg):	12.2	14.3	15.7	12.6	13.5	14.0
durchschnittliche Anzahl v. Zügen:	8.5	9.9	10.7	9.0	9.3	9.5

- Zeichnen Sie ein Streudiagramm.
- Berechnen Sie $r_{x,y}$.

5.5. Identifizieren Sie die erklärende Variable X und die Responsevariable Y in jeder der folgenden Situationen:

- Das Verkehrsministerium möchte das Verhältnis zwischen Straßenunebenheiten und Benzinverbrauch untersuchen.
- Ein Händler, der seine Waren bei Fußballspielen verkauft, möchte die Verkaufszahlen auf die Anzahl von Siegen des Heimteams beziehen.
- Ein Soziologe möchte die Anzahl von Wochenenden, die ein Student zuhause verbringt im Verhältnis zur Entfernung zwischen Wohn- und Studienort untersuchen.

5.6. Gegeben seien folgende Meßwerte:

$x:$	0	3	5	8	9
$y:$	1	2	4	3	5

- Zeichnen Sie ein Streudiagramm.
 - Zeichnen Sie aufgrund visueller Betrachtung eine gerade Linie in das Streudiagramm ein, die Ihrer Meinung nach gut zu den Daten passen würde.
 - Berechnen Sie die Parameter a und b und zeichnen Sie die tatsächliche Regressionsgerade.
 - Vergleichen Sie Ihre Linie mit der berechneten Regressionsgerade.
- 5.7. Ein Teil einer Studie befaßte sich mit den psychologischen Zusammenhängen von Erfolg bei Athleten. Untersuchungen bei den Sportlern eines US-Ringerteams ergaben Folgendes:

Ärger/Wut:	6	7	5	21	13	5	13	14
Kraft/Energie:	28	23	29	22	20	19	28	19

- Stellen Sie diese Werte in einem Streudiagramm dar.
 - Berechnen Sie $r_{x,y}$.
 - Berechnen Sie Regressionsgerade.
 - Versuchen Sie den Energiewert \hat{y} vorausszusagen, wenn der Ärgerwert $x = 8$ beträgt.
- 5.8. Aus den Berechnungen einer Datenreihe von 20 Wertepaaren (x_i, y_i) erhält man:
 $\sum x = 156$, $\sum y = 1178$, $\sum x^2 = 1262$, $\sum y^2 = 69390$, $\sum xy = 9203$
 Berechnen Sie den Korrelationskoeffizienten und bestimmen Sie die Regressionsgerade.
- 5.9. Müllaufkommen 1988 nach Bundesländer in Tonnen (Haus- und Sperrmüll und hausmüllähnliche Abfälle aus Gewerbe und Industrie):

Bundesland	Müllaufkommen	Bevölkerung
Burgenland	81.000	274.000
Kärnten	217.000	552.000
Niederösterr.	366.000	1,481.000
Oberösterr.	270.000	1,340.000
Salzburg	165.000	484.000
Steiermark	330.000	1,185.000
Tirol	266.000	630.000
Vorarlberg	99.000	333.000
Wien	548.000	1,533.000

- Zeichnen Sie ein Streudiagramm für diese Daten. Welche Variable ist als unabhängig und welche als abhängig anzusehen?
- Bestimmen Sie die Regressionsgerade und zeichnen Sie diese in das Streudiagramm ein. Wie würden Sie das Müllaufkommen für eine Region mit 1,700.000 Einwohner prognostizieren?
- Berechnen Sie den Korrelationskoeffizienten.
- Wie groß ist der Erklärungsgrad des Modells?

Übungsaufgaben zu Kapitel 6

- 6.1. Ungefähr 20 Millionen Amerikaner sind selbständig. Die meisten davon führen ein Ein-Mann-Unternehmen. Ein Viertel von ihnen benützt eigene Computer in ihrem Geschäft. Eine Marktforschungsfirma, CIIC, wollte wissen, ob Einzelunternehmer, die eigene Computer verwenden, erfolgreicher sind als solche ohne Computer. Sie befragten 150 Einzelunternehmer bezüglich ihrer Computerverwendung und ihres jährlichen Einkommens. Kann man aus folgenden Ergebnissen schließen, dass Einzelunternehmer, die einen eigenen Computer benutzen, mehr verdienen als solche, die es nicht tun? Interpretieren Sie das Ergebnis.

(Daten: <http://statistik.wu-wien.ac.at/mathstat/hatz/beispiele/einzelunt.htm>)

Gruppenstatistiken

Computerbenutzer	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
jährl. Einkommen ja	61	48245,90	10079,11	1290,50
nein	89	69932,58	7959,84	843,74

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit		
		F	Signifikanz	T	df	Sig. (2-seitig)
jährl. Einkommen	Varianzen sind gleich	4,686	,032	-14,692	148	,000
	Varianzen sind nicht gleich			-14,065	108,7	,000

- 6.2. Seit der Entdeckung von Krankheitserregern versuchen Eltern ihren Kindern beizubringen, sich die Hände zu waschen. Der Hausverstand sagt einem schließlich, daß dadurch die Ausbreitung von Infektionskrankheiten verringert wird. Eine im *University of California at Berkeley Wellness Letter* (Vol. 13, Issue 6, März 1997) veröffentlichte Studie scheint die Ansicht der Eltern zu bestätigen. In Michigan wurde eine Gruppe von Kindern untersucht. Es wurde aufgezeichnet, ob sie sich ihre Hände mehr als viermal während eines Schultages waschen, sowie die Anzahl der Krankentage aufgrund von Magen- und Darmerkrankungen. Im Median waren Kinder, die sich öfters die Hände wuschen, 2 Tage, jene, die dies nicht taten, 3 Tage erkrankt. Kann man aus den unten angeführten Ergebnissen schließen, daß Kinder, die ihre Hände mindestens viermal täglich waschen, weniger oft krank sind als andere? Interpretieren Sie das Ergebnis.

(Daten: <http://statistik.wu-wien.ac.at/mathstat/hatz/beispiele/infekt.htm>)

Statistik für Test

	Tage Magen/Darmerkrankung
Mann-Whitney-U	2534,000
Wilcoxon-W	7584,000
Z	-6,134
Asymptotische Signifikanz (2-seitig)	,000

a. Gruppenvariable: Hände waschen mehrmals

6.3. Viele Kinder lernen das Lesen nach der sogenannten Ganzheitsmethode. Diese Methode geht davon aus, daß es ein natürlicher und effizienter Weg ist, ganze Wörter im Zusammenhang zu betrachten. Schüler lernen, Wörter wiederzuerkennen, die sie bereits gesehen haben. Für die letzte Generation war dies die vorherrschende Unterrichtsmethode in ganz Nordamerika. Es ersetzte die phonetische Methode, nach der den Kindern beigebracht wurde, einzelne Buchstaben auszusprechen, um Wörter zu formen. Die Ganzheitsmethode wurde ohne wissenschaftliche Fundierung eingesetzt und ist in letzter Zeit heftig kritisiert worden. Eine neue Studie könnte die Frage entscheiden, welche Methode verwendet werden sollte. Barbara Forman, eine Erziehungspsychologin an der Universität von Houston, beschrieb ihr Experiment beim jährlichen Meeting der "Amerikanischen Vereinigung für Fortschritt der Wissenschaft". Auswählt wurden 375 Schüler aus Schulen in Houston, die einen niedrigen Lernerfolg aufwiesen. Sie wurden in 3 Gruppen eingeteilt. Eine wurde entsprechend der Ganzheitsmethode unterrichtet, die zweite Gruppe nach der phonetischen Methode, und die dritte Gruppe wurde mit einer Mischung aus beiden Methoden unterrichtet. Am Ende des Semesters wurden die Schüler aufgefordert, Wörter aus einer Liste von 50 Wörtern zu lesen. Interpretieren Sie die folgenden Ergebnisse dieser Studie.

(Daten: <http://statistik.wu-wien.ac.at/mathstat/hatz/beispiele/lesen.htm>)

Ränge

Leserlernmethode	N	Mittlerer Rang
WOERTER ganzheitlich	125	123,28
phonetisch	125	155,41
gemischt	125	285,30
Gesamt	375	

Statistik für Test^{a,b}

	WOERTER
Chi-Quadrat	158,698
df	2
Asymptotische Signifikanz	,000

a. Kruskal-Wallis-Test

b. Gruppenvariable: Leserlernmethode

6.4. In einer Studie untersuchten Chasnoff et.al. (1989) die Körpergröße von Neugeborenen, deren Mütter während der Schwangerschaft Kokain konsumiert hatten. Zum Vergleich wurde auch eine Gruppe Babies von Müttern, die kein Kokain zu sich genommen hatten, untersucht. Interpretieren Sie die folgenden Ergebnisse.

Deskriptive Statistik

Größe des Neugeborenen

	N	Mittelwert	Standardabweichung
keine Drogen	39	51,1000	2,8975
erste 3 Monate	19	49,3000	2,5024
durchgehend	36	48,0000	3,6010
Gesamt	94	49,5489	3,3871

ANOVA

Größe des Neugeborenen

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	181,375	2	90,687	9,319	,000
Innerhalb der Gruppen	885,580	91	9,732		
Gesamt	1066,955	93			

6.5. In welcher Form beeinflusst das Geschlecht die Beurteilung von Unterricht? Verschiedene Studien beschäftigten sich in letzter Zeit mit dieser Frage. Für eine Untersuchung wurden mehrere weibliche und männliche Professor(inn)en desselben Instituts und mit ähnlichem "background" ausgewählt. Aus allen Beurteilungen ihrer Lehrtätigkeit wurde eine Zufallsstichprobe gezogen. Die Beurteilungen beruhten auf einem Punkte-Score, der die Beantwortung mehrerer Fragen miteinbezog. Die Punkte reichten von 0 bis 10. Das Geschlecht der Studenten (SEXSTUD) wurde ebenso erfaßt wie das der Professor(inn)en (SEXPROF). Beantworten Sie anhand der unten angeführten Ergebnisse folgende Fragen:

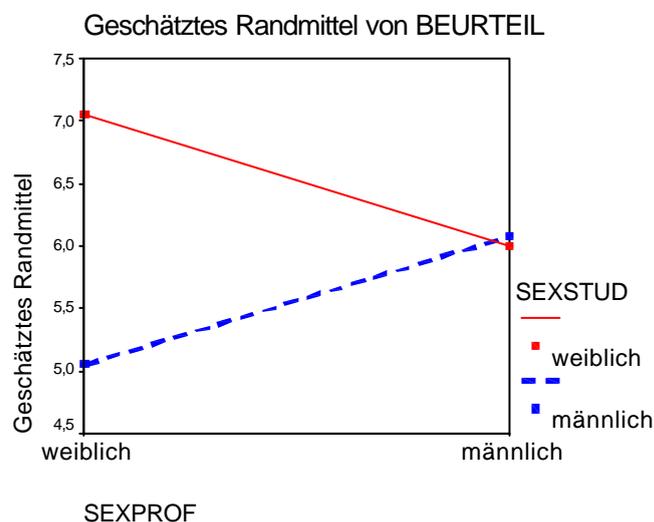
- (a) Werden weibliche und männliche Professor(inn)en gleich beurteilt ?
- (b) Urteilen männliche und weibliche Student(inn)en gleich ?
- (c) Kann man daraus schließen, daß männliche Studenten Professorinnen anders beurteilen als Professoren ?
- (d) Gibt es Hinweise, daß männliche Studenten männliche Professoren höher einschätzen, als es die weiblichen Studenten tun ?

		SEXSTUD		Gesamt wert
		weiblich	männlich	Mittelwert
		Mittelwert	Mittelwert	
SEXPROF	weiblich	7,06	5,05	6,06
	männlich	6,00	6,08	6,04
Gruppen-Gesamtwert		6.53	5.57	6.05

Tests der Zwischensubjekteffekte

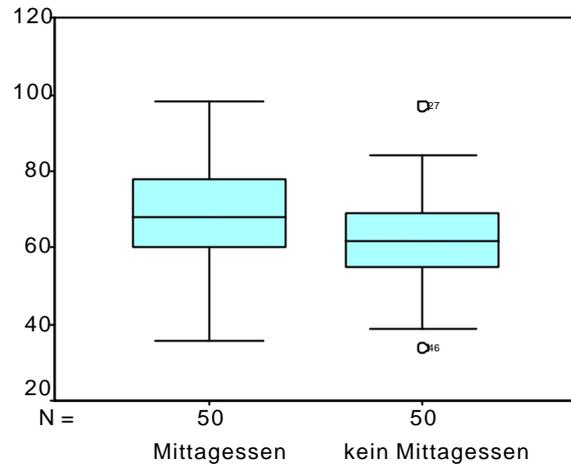
Abhängige Variable: BEURTEIL

Quelle	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	3	67,449	63,937	,000
Konstanter Term	1	14628,903	13867,254	,000
SEXPROF	1	2,250E-02	,021	,884
SEXSTUD	1	93,122	88,274	,000
SEXPROF * SEXSTUD	1	109,202	103,517	,000
Fehler	396	1,055		
Gesamt	400			
Korrigierte Gesamtvariation	399			



Bei den folgenden Beispielen sollen Sie angeben, welche Analysemethode Sie verwenden würden:

- 6.6. Eine Professorin hat festgestellt, daß die Prüfungsergebnisse von Studenten, die ihre Arbeit um 13 Uhr schreiben, besser sind als die Noten von Studenten, die sie um 16 Uhr schreiben. Sie ist der Meinung, daß sich der Unterschied durch Energiestufen erklären läßt, die durch den Zeitpunkt der letzten Mahlzeit beeinflußt werden. Kurz vor einer Prüfung zu essen würde also eine bessere Note produzieren. Um ihre These zu überprüfen, wählte sie 100 Studenten aus, die um 12:30 einen Test schrieben. Die eine Hälfte nahm vorher ein Mittagessen ein, die andere Hälfte verzichtete darauf.



(Daten: <http://statistik.wu-wien.ac.at/mathstat/hatz/beispiele/testerg.htm>)

- 6.7. Dreißig amerikanische Magazine wurden nach dem Ausbildungsgrad ihrer Leser geordnet. Danach wurden je drei Magazine zufällig aus dem obersten, mittleren und untersten Drittel gezogen, nämlich 1. *Scientific American*, 2. *Fortune*, 3. *The New Yorker* (oberste Gruppe); 4. *Sports Illustrated*, 5. *Newsweek*, 6. *People* (mittlere Gruppe); 7. *National Enquirer*, 8. *Grit*, 9. *True Confessions* (unterste Gruppe). Aus jedem dieser neun Magazine wurden zufällig je sechs Anzeigen gezogen von denen die obigen Variablen bestimmt wurden. Folgende Variablen waren definiert: 1. Zahl der Wörter in der Anzeige, 2. Zahl der Sätze in der Anzeige, 3. Zahl der Wörter mit mehr als drei Silben in der Anzeige. Mittels welche Methoden können folgende Fragen beantwortet werden:

- Gibt es charakteristische Unterschiede in der Lesbarkeit der Anzeigen zwischen den einzelnen Magazinen?
- Gibt es charakteristische Unterschiede in der Lesbarkeit der Anzeigen zwischen den einzelnen Magazingruppen?
- Bei welcher der drei Variablen sind diese Unterschiede am stärksten ausgeprägt?

(Daten: <http://statistik.wu-wien.ac.at/mathstat/beispiele/magadsdat.html>)

- 6.8. Eine Studie wurde durchgeführt, um zu erforschen, ob bestimmte Trainingsmethoden (Computertutorial bzw. praktische Übungen) bzw. Softwarepakete (MS Word, Wordperfect, Ami Pro) die in einer Berufsschule angeboten werden, effizienter sind als andere. Sechs Gruppen (aus der Kombination Trainingsmethode und Softwarepaket) zu je 40 StudentInnen hatten ein Trainingsprogramm absolviert. Von ihnen wurde die Anzahl der pro Minute getippten Wörter aufgezeichnet. Welche Frage lassen sich auf Grund dieser Untersuchung beantworten? Welche Methode würden Sie zur Analyse verwenden?

(Daten: <http://statistik.wu-wien.ac.at/mathstat/beispiele/software.htm>)

- 6.9. Ein Verkehrsplaner möchte wissen, ob unterschiedlich viel Verkehr an zwei verschiedenen Kreuzungen anfällt. Der Verkehrsplaner hat einen Mitarbeiter, der für ihn Zählungen durchführt. Jeden Tag wird eine Münze geworfen um zu entscheiden, an welcher der beiden Kreuzungen gezählt wird. Nach 20 Tagen war neun Mal bei Kreuzung 1 und elf Mal an Kreuzung 2 gezählt worden. Eine grafische Darstellung (welche würden Sie verwenden) ergab, dass in beiden Datensätzen outlier (Ausreisser) vorkamen. Welche Methode würde Sie zur Untersuchung der Frage: "Unterscheiden sich die beiden Kreuzung bezüglich des Verkehrsaufkommens?" verwenden?