Lehrstuhl für Ökonometrie und Statistik

Statistik I
Sommersemester 2014

Prof. Dr. Wolf-Dieter Heller
Frieder Conrad
Hartwig Senska

Aufgabensammlung zur Vorlesung Statistik I

Inhaltsverzeichnis

1 Informationen zu Vorlesung und Übungsbetrieb

Vorlesung: Die Vorlesung findet

- freitags, 8:00 - 9:30 Uhr im Audimax, Geb. 30.95, sowie
- freitags, 14:00 - 15:30 Uhr im Gerthsen, Geb. 30.21

statt.

Tutorien: Begleitend zur Vorlesung werden Tutorien angeboten. Ziel der Tutorien ist die Besprechung einer Auswahl von Aufgaben aus dieser Aufgabensammlung. Die Lösungen zu den Aufgaben aus dieser Aufgabensammlung stehen auf der Homepage des Lehrstuhls zum Download bereit. Zu den Tutorien ist keine Anmeldung nötig.

PC-Praktikum: Ergänzend zur Vorlesung wird zu den folgenden Terminen ein PC-gestütztes Praktikum (3 Lektionen à 2 Stunden) im Raum CIP-I des CIP-Pool der Fakultät für Wirtschaftswissenschaften (Sockelgeschoss von Geb. 11.40) angeboten.

- 1. Durchgang: 12., 13. und 14. Mai 2014, Anmeldung ab dem 5. Mai 2014
- 2. Durchgang: 10., 11. und 16. Juni 2014, Anmeldung ab dem 3. Juni 2014
- 3. Durchgang: 30. Juni, 1. und 2. Juli 2014, Anmeldung ab dem 23. Juni 2014

Die Zeiten an den einzelnen Tagen sind jeweils 08:00 - 10:00 Uhr, 10:00 - 12:00 Uhr oder 12:00 - 14:00 Uhr an allen Montags- und Mittwochsterminen sowie 14:00 - 16:00 Uhr, 16:00 - 18:00 Uhr oder 18:00 - 20:00 Uhr an allen Dienstagsterminen. Eine Anmeldung zu den einzelnen Lektionen ist über den Ilias-Kurs der Vorlesung möglich. Für die erfolgreiche Teilnahme an mindestens fünf der insgesamt sechs Lektionen des PC-Praktikums in Statistik I/II erhält man einen Schein.

Für eine Teilnahme an diesem PC-Praktikum werden Grundkenntnisse im Umgang mit Microsoft Excel vorausgesetzt.

Klausur: Die Klausur zur Vorlesung Statistik I findet am 19. Juli von 8:00 - 10:00 Uhr statt.

Die Anmeldung zur Klausur erfolgt über die Selbstbedienungsfunktion des Studierendenportals. Angehörige anderer Fakultäten, die einen Statistik I Schein erwerben möchten, melden sich mit einem Zettel an, auf dem Name, Matrikelnummer und Studienfach vermerkt sind. Die Anmeldung erfolgt durch die Abgabe dieses Zettels im Sekretariat des Lehrstuhls (Zi. 209 Geb. 20.12). Die Hörsaalverteilungen sind zu gegebener Zeit auf der Prüfungs-Homepage des Lehrstuhls abrufbar.

Die Nachklausur Statistik I findet voraussichtlich in der ersten Novemberhälfte 2014 statt. Der genaue Termin wird rechtzeitig bekannt gegeben.

Beide Klausuren werden ohne Hilfsmittel (mit Ausnahme der vom Lehrstuhl zugelassenen Taschenrechner) geschrieben. Den Klausurexemplaren werden als Hilfestellung eine kleine Formelsammlung sowie die benötigten Tabellen beigelegt. Die Hilfestellungen werden rechtzeitig vor dem Klausurtermin bekannt gegeben. In der Woche vor der Hauptklausur findet eine Fragestunde statt.

Die Ergebnisse aller Klausuren werden lediglich auf der Homepage des Lehrstuhl veröffentlicht und können dort abgefragt werden. Benötigt wird dazu die während der Klausur ausgehändigte PIN!

Ansprechpartner: Inhaltliche Fragen können in den Tutorien und im Ilias-Forum der Veranstaltung gestellt werden. Ihre direkten Ansprechpartner sind Frieder Conrad für Fragen zum Vorlesungs- und Übungsbetrieb und Hartwig Senska für Fragen zur Klausurorganisation. Beide erreichen Sie per Email unter frieder.conrad@kit.edu bzw. hartwig.senska@kit.edu. Für das Rechnerpraktikum ist Jonas Müller zuständig. Die aktuellen Sprechstunden aller Lehrstuhlmitarbeiter entnehmen Sie bitte der Homepage des Lehrstuhls.

Sonstige Informationen: Weitere Informationen werden in der Vorlesung, in den Tutorien, im Ilias-Kurs der Veranstaltung sowie zusätzlich auf den Lehrstuhl-WWW-Seiten (http://statistik.econ.kit.edu) bekannt gegeben.

2 Übungsaufgaben

Aufgabe 1:

Geben Sie für die folgenden Fragestellungen die statistischen Einheiten und Massen an, und grenzen Sie die Massen sachlich, räumlich und zeitlich ab.

(a) Es soll der Anteil der Pkw des Typs A an den im März 2008 in der Bundesrepublik neu zugelassenen Pkw ermittelt werden.

(b) Es soll die durchschnittliche Studiendauer an deutschen Universitäten ermittelt werden.

Aufgabe 2:

(a) Welche der folgenden Beispiele sind Bestands- und welche Ereignismassen?

- Studierende an einer Universität
- Todesfälle in einer Gemeinde
- Personenkraftwagen der Deutsche Post AG
- Maschinenausfälle in einer Fabrik
- Anmeldungen in einem Einwohnermeldeamt
- Wartende Postkunden vor einem Postschalter
- Leichte Verkehrsunfälle in der Bundesrepublik

(b) Zu den Bestandsmassen aus (a) gebe man jeweils die korrespondierenden Ereignismassen, zu den Ereignismassen die zugehörigen Bestandsmassen an.

(c) Geben Sie ein einfaches Beispiel an, bei dem *einer* Bestandsmasse *mehrere* korrespondierende Ereignismassen zugeordnet sind.

Aufgabe 3:

Interpretieren Sie die abbgebildete grafische Darstellung (Quelle: Deutsche Bank Research). Gehen Sie dabei insbesondere auf die folgenden Fragen ein.

(a) Welche statistische(n) Masse(n) ist(sind) involviert? Handelt es sich dabei um Bestands- oder Ereignismassen? Geben Sie jeweils die sachliche, räumliche und zeitliche Abgrenzung an.

(b) Welches Merkmal (welche Merkmale) wurden bei dieser Untersuchung beobachtet? Welcher Art ist dieses Merkmal (sind diese Merkmale)? Ist es (sind sie) häufbar?

(c) Um was für Zahlenangaben handelt es sich dabei? Wie werden sie in der Statistik bezeichnet?

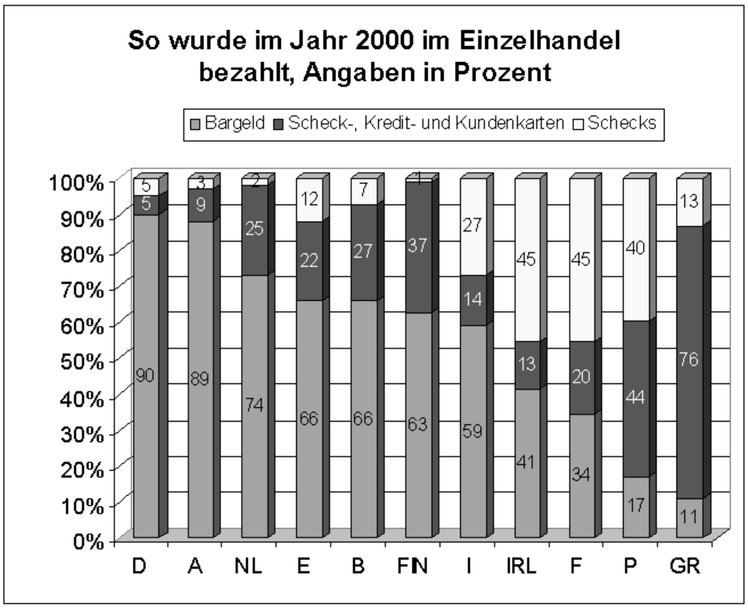


Abbildung 1: Einkäufe: Die Deutschen zahlen bar.

Aufgabe 4:

(a) Welche der folgenden Merkmale sind *diskret*, welche sind *stetig*?

- Geschwindigkeit eines Pkws
- Hörerzahl der Vorlesung „Statistik I“
- Anzahl der Mitarbeiter eines Betriebes
- Einkommen
- Zeit für die Beschleunigung eines Pkws von 0 auf 100 km/h
- Klausurpunkte

(b) Welche Probleme ergeben sich bei der Messung *stetiger* Merkmale?

Aufgabe 5:

(a) Erläutern Sie das Prinzip der mengen- und zahlenmäßigen Fortschreibung anhand der Pkw-Zulassungen in der Bundesrepublik im Monat April dieses Jahres.

(b) Überlegen Sie sich, wann eine Fortschreibung sinnvoll ist.

Aufgabe 6:

Für Merkmalsarten sind auch folgende Bezeichnungen üblich

(a) Nominalskala,

(b) Ordinalskala,

(c) Intervallskala

(d) Verhältnisskala und

(e) Absolutskala

Welche anderen Bezeichnungen sind hierfür üblich. Geben Sie jeweils ein Merkmal als Beispiel an und nennen Sie eine statistische Masse, bei der dieses Merkmal beobachtet werden kann. Welche Anforderungen müssen Skalentransformationen erfüllen, damit die wesentlichen Eigenschaften der jeweiligen Skala erhalten bleiben? Veranschaulichen Sie sich anhand von jeweils mindestens einem Beispiel, welche Folgen eine Missachtung des Skalentyps bei der Transformation haben kann.

Aufgabe 7:

(a) Geben Sie bei den folgenden Merkmalen an, welcher Art sie sind. Welche der Merkmale sind häufbar?

- Unfallursache
- Datum des Semesterendes
- Hubraum eines Motors
- erlernter Beruf
- Windstärke
- Telefonnummer eines Faxgerätes
- Kinderzahl
- Körpergröße

(b) Im Studiengang Wirtschaftsingenieurwesen wird die Möglichkeit geschaffen, im Hauptstudium freiwillig eines oder mehrere der Fächer Statistik, Informatik und Operations Research zusätzlich zu den Pflichtfächern zu belegen. Erläutern Sie, wie das häufbare Merkmal „zusätzliche Fächer“ der statistischen Einheiten „Studierende des Wirtschaftsingenieurwesens“ auf ein nicht häufbares zurückgeführt werden kann.

(c) Machen Sie sich an mindestens zwei Beispielen den Unterschied in der Verwendung der Begriffe „Merkmalsausprägung“ und „Merkmalswert“ deutlich.

Aufgabe 8:

Bei der 2001. Vorführung von „Müllers Büro“ in der Karlsruher Schauburg wurde das Alter von 50 Studierenden ermittelt:

38, 23, 17, 34, 43, 22, 21, 26, 19, 25, 29, 18, 22, 24, 20, 16, 25, 24, 22, 21, 26, 28, 45, 22, 18, 23, 27, 21, 34, 42, 21, 62, 18, 17, 24, 36, 30, 20, 25, 21, 37, 29, 33, 19, 26, 28, 22, 41, 24, 26.

(a) Erstellen Sie für das Merkmal „Alter“ eine geordnete Urliste.

(b) Durch weiteres Nachfragen stellt man fest, dass 23 der 50 Zuschauer an der Universität studieren. Diese gaben als Studienfach an (C$\hat{=}$ Ciw, W$\hat{=}$ WiWi, M$\hat{=}$ Machbau, E$\hat{=}$ Etech, I$\hat{=}$ Info, Ch$\hat{=}$ Chemie, B$\hat{=}$ Biologie, A$\hat{=}$ Architektur, Ma$\hat{=}$ Mathematik):

M, C, B, W, A, Ma, M, Ch, B, M, I, Ma, I, E, A, M, I, E, C, B, C, Ch, I

Sortieren Sie die Urliste, und geben Sie für das Merkmal „Studienfach“ eine Häufigkeitsverteilung an.

(c) Die folgende Stiel-und-Blatt-Darstellung gibt die Werte des Merkmals „Alter“ für die 23 Studierenden wieder. Fassen Sie die Daten mittels einer geeigneten Klassierung zusammen (5 Klassen).

1|
1| 6 7 8 9 9
2| 0 1 2 3 3 4
2| 5 5 6
3| 0 3 4
3| 6 7 8
4| 1 2
4| 5

Aufgabe 9:

(a) Von 2000 im Rahmen einer Umfrage befragten Personen haben 50 das Auto X als Wagen des Jahres gewählt.

Benennen Sie den Merkmalswert und bestimmen Sie seine absolute und relative Häufigkeit.

(b) Geben Sie eine allgemeingültige Unter- und Obergrenze für relative Häufigkeiten an.

(c) Welchen Wert nimmt die Summe der relativen Häufigkeiten bei einer Erhebung normalerweise an? Gibt es Ausnahmen?

Aufgabe 10:

Die folgenden Daten bzgl. der Entwicklung der Studierendenzahlen in Westdeutschland können Sie dem Kompendium „Zahlen 1996“ des Instituts der Deutschen Wirtschaft Köln entnehmen:

Studenten in Westdeutschland nach Hochschulart in 1000			
	1960	1975	1994
Insgesamt	291.1	840.8	1676.1
Uni und PH	238.4	680.2	1253.4
Kunsthochschulen	8.5	15.4	24.5
Fachhochschulen	44.2	145.2	398.2

(a) Stellen Sie für das Jahr 1994 die Aufteilung der Studierenden auf die verschiedenen Hochschularten mit Hilfe eines Stabdiagramms dar. Was müssen Sie bei Verwendung eines Flächen- bzw. Volumendiagramms beachten?

(b) Veranschaulichen Sie sich die Daten der Tabelle mit Hilfe eines kombinierten Flächen-/Kreissektorendiagramms. Kennen Sie Alternativen zu diesem Diagramm?

Aufgabe 11:

(a) Eine Erhebung des Merkmals „Körpergröße" in einer Gemeinde ergab folgende Messwerte:

$[0.30, 0.60)$	$[0.60, 0.90)$	$[0.90, 1.20)$	$[1.20, 1.50)$	$[1.50, 1.80)$	$[1.80, 2.10)$
500	3580	4760	3800	10128	2780

Zeichnen Sie das Histogramm und die Summenhäufigkeitsfunktion.

(b) Zu einer detaillierteren Analyse der Daten werden diese wie folgt umgruppiert:

$[0.30, 0.60)$	$[0.60, 0.75)$	$[0.75, 0.90)$	$[0.90, 1.05)$	$[1.05, 1.20)$	$[1.20, 1.50)$
500	2500	1080	1000	3760	3800

$[1.50, 1.65)$	$[1.65, 1.80)$	$[1.80, 2.00)$
2000	8128	2780

Zeichnen Sie das Histogramm für diese Häufigkeitstabelle.

(c) Zeichnen Sie in die beiden Histogramme aus (a) und (b) jeweils ein Häufigkeitspolygon ein. Welche Schwierigkeiten ergeben sich beim Erstellen und Interpretieren der Diagramme?

(d) Berechnen Sie mit Hilfe der beiden Häufigkeitstabellen den Anteil der Merkmalswerte in der Urliste,

- die kleiner oder gleich 0.7 sind,
- im abgeschlossenen Intervall $[1.4, 1.6]$ liegen bzw.
- größer oder gleich 1.65 sind,

und vergleichen Sie die Ergebnisse.

Aufgabe 12:

Im Rahmen einer Umfrage unter Gästen der Fast-Food-Kette „Crazy Cow" wurde bei 50 Personen das Körpergewicht (in [kg]) gemessen:

$62, 72, 83, 98, 52, 84, 94, 95, 68, 65, 85, 50, 72, 59, 76, 61, 60, 72, 93, 47$

$61, 58, 75, 65, 79, 81, 55, 63, 53, 50, 62, 86, 50, 74, 94, 70, 88, 48, 79, 76$

$70, 47, 72, 58, 83, 74, 85, 79, 75, 76$

(a) Berechnen Sie aus den Daten der Urliste alle Ihnen aus der Vorlesung bekannten Lage- und Streuungsparameter.

(b) Klassieren Sie die Daten sinnvoll, und zeichnen Sie ein Histogramm.

(c) Zeichnen Sie die empirische Verteilungsfunktion für den Datensatz sowie die Summenhäufigkeitsfunktion für die klassierten Daten. Wie läßt sich die Summenhäufigkeitsfunktion interpretieren?

(d) Berechnen Sie nun aus den klassierten Daten analog zu (b) die Lageparameter und die Varianz. Welche Annahmen treffen Sie dabei?

(e) Vergleichen Sie die Ergebnisse aus (d) mit denen aus (a). Was läßt sich über die Verteilung innerhalb der Klassen aussagen?

Aufgabe 13:

Im folgenden sollen Sie sich die Bedeutung der verschiedenen in der Vorlesung betrachteten Lageparameter noch einmal anhand eines Beispiels verdeutlichen:

Situation: *Ein Werk mit 1000 Beschäftigten, das in Form einer Werkstraße angelegt ist (vgl. Skizze), soll eine neue Kantine erhalten. Ihr Standort soll möglichst „optimal“ gewählt werden, wobei die Geschäftsleitung sich bzgl. der zu verwendenden Zielfunktion noch nicht ganz einig werden konnte.*

	Pforte	Halle 1	Halle 2	Halle 3
Anzahl Beschäftigte	3	200	300	497
Abstand von der Pforte	0	100	600	900

Im folgenden soll m den noch zu bestimmenden Abstand der Kantine und x_i den Abstand des Arbeitsplatzes des i'ten Mitarbeiters von der Pforte bezeichnen.

(a) Die Geschäftsleitung beschließt, die Kantine in einer der drei Hallen einzurichten. Zeigen Sie, dass der Modalwert x_M der Entfernungen x_i der Arbeitsplätze von der Pforte den optimalen Standort der Kantine angibt, wenn dieser die „Gesamtrüstzeiten des Kantinengangs“ aller Mitarbeiter minimieren soll (muss ein Mitarbeiter für den Besuch der Kantine die Halle, in der er arbeitet, verlassen, so entsteht durch Umziehen etc. ein Verlust von w Zeiteinheiten).

(b) Zeigen Sie, dass das arithmetische Mittel der Wegstrecken x_i der Mitarbeiter zur Pforte die Summe der quadrierten Wegstrecken $\sum(x_i - m)^2$ minimiert (Sie können sich das Quadrieren als eine Form der Gewichtung der Wegstrecken vorstellen, d.h. lange Wege werden im Vergleich zu relativ kurzen überproportional bestraft).

(c) Zeigen Sie, dass der Median x_z der Wegstrecken zur Kantine die Summe der Wegstrecken aller Beschäftigten, d.h. $\sum |x_i - m|$, minimiert.

(d) Erklären Sie basierend auf den Ergebnissen der Aufgabenteile (b) und (c), warum die mittlere absolute Abweichung mit Hilfe des Medians und die Varianz mit Hilfe des arithmetischen Mittels berechnet wird.

Aufgabe 14:

Berechnen Sie zu den Gewichtsangaben der 50 Personen aus Übung 12 die α-Quantile für $\alpha = 20\%$ und $\alpha = 60\%$. Stellen Sie die Verteilung der Gewichtsangaben mit Hilfe eines Boxplots dar.

Aufgabe 15:

An 21 Personen – anhand ihres Ernährungsverhaltens in drei Gruppen unterteilt – wurden folgende Messwerte des Cholesterolgehaltes im Blut (in [mg/100ml]) festgestellt:

Gruppe	Cholesterolwerte									
I	403	311	269	336	259					
II	312	222	302	420	540	386	353	210	286	290
III	403	244	353	235	319	260				

Vergleichen Sie die Cholesterol-Verteilungen der drei Gruppen mit Hilfe von Boxplots.

Aufgabe 16:

Die Quartile für den Alkoholkonsum von 28 Männern lauten:

25%-Quantil:	5	g/Tag
Median	12	g/Tag
75%-Quantil:	12	g/Tag

Die angegebenen Quantile sind gleichzeitig auch gemessene Werte des zugrunde liegenden Datensatzes.

(a) Welche der Quantile $q_{0.15}$, $q_{0.25}$, $q_{0.40}$, $q_{0.50}$, $q_{0.70}$, $q_{0.75}$, $q_{0.90}$ sind für $n = 28$ notwendigerweise identisch mit tatsächlich gemessenen Größen, welche nur zufälligerweise im hier angegebenen Datensatz? Begründung!

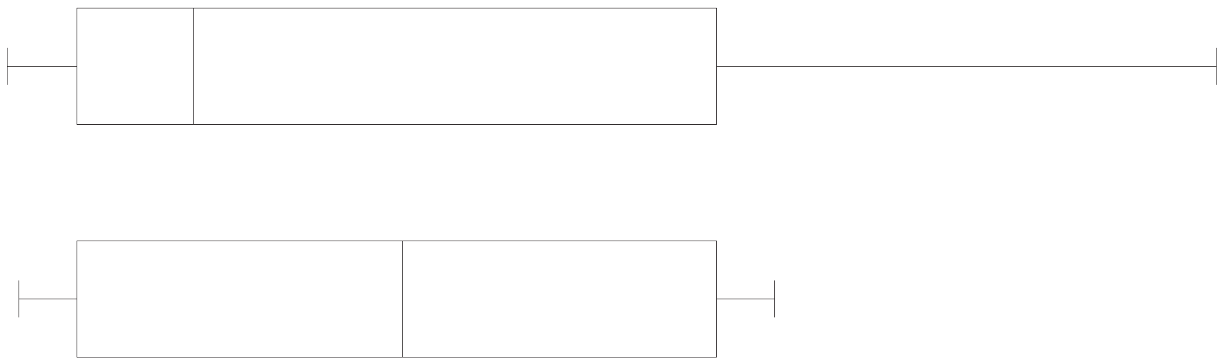
(b) Schätzen Sie mit Hilfe der obigen Angaben ab, wie viele der 28 Männer 5 g bzw. 12 g pro Tag zu sich nehmen. Begründung!

g Alkohol/Tag	minimal mögliche Anzahl	maximal mögliche Anzahl
5		
12		

(c) Zeichnen Sie unter Verwendung der zusätzlichen Angabe, dass im Datensatz die Messwerte 0, 13, 17, 18, 19, 17, 28 vorkommen, einen Boxplot.

Aufgabe 17:

Gegeben sind die folgenden beiden Boxplots:



Was kann man über die zugehörigen Datenverteilungen sagen? Stellen Sie diese in groben Zügen durch Punktediagramme und Histogramme dar.

Aufgabe 18:

In der Physik gibt das Potential eines Punktes x die Spannung zwischen x und einem vorab gewählten Referenzpunkt x_0 an.

(a) Um was für Merkmale (Skalentypen) handelt es sich bei der elektrischen Spannung bzw. dem elektrischen Potential.

(b) Sind die Variationskoeffizienten der Verteilungen der Merkmale Spannung und elektrisches Potential invariant gegenüber Skalentransformationen?

Aufgabe 19:

Betrachtet werden 5 Straßenhändler, die an einem bestimmten Tag auf der Kaiserstraße zwischen 16 und 17 Uhr Modeschmuck verkaufen.

(a) Nehmen Sie an, dass nur ein Händler in dieser Zeit überhaupt etwas verkauft. Wie sehen die Lorenzkurve und der Gini-Koeffizient aus?

(b) Nehmen Sie nun an, der Gesamtumsatz der Händler betrage 300 DM. Betrachten Sie die folgenden 5 Verteilungen:

Händler i / Verteilung	1	2	3	4	5
I	60	60	60	60	60
II	40	30	60	70	100
III	100	100	100	0	0
IV	200	25	25	25	25
V	10	10	260	10	10

Geben Sie zu jeder Verteilung die Lorenzkurve und den Gini-Koeffizienten an. Diskutieren Sie anhand der Ergebnisse die Aussagekraft des Gini-Koeffizienten.

(c) Verdeutlichen Sie sich die Aussagekraft des Herfindahl-Indexes. Interpretieren Sie dazu den Anteil des Merkmals x_i an der Merkmalssumme, d.h. $q_i = \frac{x_i}{\sum_{i=1}^n x_i}$, als neuen Merkmalswert, und zeigen Sie anschließend, dass $H = \sum_{i=1}^n q_i^2$ sich aus der Varianz der q_i berechnen lässt. Was ist der Minimal- und Maximalwert von H?

(d) Bestimmen Sie für jede der Verteilungen den Herfindahl-Index und den Konzentrationskoeffizienten CR_3.

Aufgabe 20:

In einem Unternehmen wurde im Jahr 2001 die folgende Verteilung der Monatsgehälter von 20 Mitarbeitern festgestellt:

Klasse	Monatsgehalt in DM			Anzahl Mitarbeiter
	von	bis	unter	
1	0	-	2000	4
2	2000	-	4000	6
3	4000	-	6000	4
4	6000	-	8000	3
5	8000	-	10000	2
6	über		10000	1

(a) Zeichnen Sie die Lorenzkurve, und berechnen Sie für die klassierten Daten den Ginikoeffizienten der Gehaltskonzentration auf die Mitarbeiter.

(b) Wie ändern sich Lorenzkurve und Ginikoeffizient, wenn der Gesetzgeber die Einkommen folgendermaßen besteuert:

(i) Kopfsteuer von 500,– DM pro Gehaltsempfänger und Monat

(ii) Proportionalsteuer von 30% auf das monatliche Bruttogehalt

(iii) progressive Besteuerung mit den entsprechenden Steuersätzen der Einkommensklassen:

Klasse	1	2	3	4	5	6
Durchschnittssteuersatz in %	6	13	20	27	34	41

(c) Welche Umverteilungseffekte bewirken demnach die unterschiedlichen Besteuerungsarten im Hinblick auf die Verteilung des gesamten Volkseinkommens?

Bemerkung: Der Duchschnittssteuersatz für die Klasse $i, i = 1, \ldots, 6$ gibt denjenigen Steuersatz an, mit dem ein Mitarbeiter der entsprechenden Lohnklasse sein *gesamtes* Einkommen versteuern muss.

Aufgabe 21:

Aus dem statistischen Jahrbuch entnimmt man für 1975 die folgenden Zahlen für die Verteilung des Umsatzes auf die Umsatzgrößenklassen der Betriebe der verarbeitenden Industrie:

Klasse	Umsatz von ... bis unter ... DM	Anzahl der Unternehmen	Umsatz in Mill. DM	Beschäftigte
1	unter 1 Mill. DM	7.047	4.577	119.746
2	1 Mill. - 2 Mill.	8 167	11 887	209 079
3	2 Mill. - 5 Mill.	10 594	34 163	485 871
4	5 Mill. - 10 Mill.	6 237	44 102	552 409
5	10 Mill. - 25 Mill.	5 238	81 856	928 103
6	25 Mill. - 50 Mill.	2 167	82 669	776 854
7	50 Mill. - 100 Mill.	1 194	82 669	813 254
8	100 Mill. und mehr	948	484 975	3.578.868

(a) Geben Sie die Lorenzkurve und den Gini-Koeffizienten für die Umsatzkonzentration auf die Unternehmensanzahl an. Wie müßten Sie vorgehen, wenn der Umsatz der einzelnen Klassen nicht bekannt wäre?

(b) Wie würden sich die Lorenzkurve und der Gini-Koeffizient ändern, wenn die Klassen paarweise (Klasse 1 und 2, Klasse 3 und 4 usw.) zusammengefasst würden?

Aufgabe 22:

In einem 24-Stunden-Dauertest wurde die Leuchtdauer von 24 Glühbirnen getestet. Hierbei ergaben sich folgende Messwerte (in [h]):

1.0	1.2	1.4	1.7	2.2	2.5	3.1	3.5	4.4
5.1	6.1	7.5	8.9	10.7	12.6	15.4	18.3	22.1

6 Birnen brannten am Ende des 24 Stunden-Tests noch immer.

Erläutern Sie, inwiefern bei der Erstellung eines Boxplots bzw. eines Histogramms auf Basis der gegebenen Messwerte Probleme entstehen. Welche Lösungsansätze fallen Ihnen ein?

Aufgabe 23:

Eine Befragung von 20 Personen, welcher Partei sie bei der nächsten Bundestagswahl ihre Stimme geben würden, ergab folgendes Ergebnis:

Person	Geschlecht	Partei
1	m	SPD
2	m	CDU
3	w	Grüne
4	m	SPD
5	w	CDU
6	w	CDU
7	w	FDP
8	m	SPD
9	m	CDU
10	w	SPD

Person	Geschlecht	Partei
11	m	Grüne
12	m	CDU
13	m	SPD
14	w	CDU
15	w	CDU
16	m	Grüne
17	m	SPD
18	w	CDU
19	w	SPD
20	m	FDP

(a) Erstellen Sie eine Kontingenztabelle der absoluten Häufigkeiten, und bestimmen Sie die Randhäufigkeiten.

(b) Stellen Sie die Kontingenztabelle der relativen Häufigkeiten auf und geben Sie die Randverteilungen an.

(c) Bestimmen Sie die bedingten Häufigkeitsverteilungen des Wahlverhaltens für Frauen und Männer.

(d) Stellen Sie die Daten aus (b) geeignet graphisch dar.

Aufgabe 24:

Der Karlsruher Wiwi-Student John Highway fährt freitags grundsätzlich mit dem Auto von Karlsruhe in seine hessische Heimat, wo er das (verlängerte) Wochenende verbringt. Je nach Motivation fährt er dann montags, dienstags oder mittwochs nach Karlsruhe zurück, um sich dem Studienstress zu unterziehen. Da er häufig in den Stau kommt, hat er sich bei seinen letzten 60 Fahrten nach Karlsruhe notiert, an welchem Wochentag er gefahren ist, und ob er in einen Stau gekommen ist oder nicht. Hier sind seine Ergebnisse:

- Ein Drittel der Fahrten wurde montags gemacht, ein weiteres Drittel mittwochs.
- Insgesamt gab es genauso viele Fahrten mit Stau wie ohne Stau.
- Montags war in 60% der Fälle Stau.
- Mittwochs war sechsmal kein Stau.

(a) Erstellen Sie die Kontingenztabellen der absoluten und relativen Häufigkeiten.

(b) Berechnen Sie die bedingten Häufigkeiten für das (Nicht-)Auftreten eines Staus an den drei genannten Wochentagen.

(c) Was meinen Sie: Hat der Wochentag, an dem die Fahrt unternommen wird, einen Einfluss auf das Stauaufkommen?

Aufgabe 25:

Bei einer Untersuchung von Familienstand und Geschlecht bei 50 Personen hat sich herausgestellt, dass die Merkmale unabhängig sind. Für die Randverteilungen gilt:

$$\begin{array}{lll} h(\text{weiblich}) = 30, & h(\text{männlich}) = 20, & h(\text{ledig}) = 5, \\ h(\text{verheiratet}) = 30, & h(\text{geschieden}) = 10, & h(\text{verwitwet}) = 5. \end{array}$$

Wie lautet die gemeinsame Häufigkeitsverteilung?

Aufgabe 26:

Gegeben seien die Daten aus Aufgabe 23.

(a) Prüfen Sie

- mittels der Definition der Unabhängigkeit,
- mit Hilfe der bedingten Verteilungen und
- mittels des (korrigierten) Kontingenzkoeffizienten,

ob das Merkmal Geschlecht einen Einfluss auf die Parteipräferenz hat.

(b) Können Sie die Unabhängigkeit auch mit Hilfe eines gestapelten Balkendiagramms überprüfen?

(c) Erstellen Sie ein kombiniertes Flächen-/Kreissektorendiagramm. Eignet sich dieses Diagramm ebenfalls zur Überprüfung der Unabhängigkeit?

Aufgabe 27:

Bestimmen Sie für die beiden folgenden Häufigkeitstabellen den Kontingenzkoeffizienten.

b \ a	a_1	a_2	a_3	a_4
b_1	20	12	8	4
b_2	10	6	4	2
b_3	5	3	2	1

b \ a	a_1	a_2	a_3
b_1	0	0	20
b_2	10	0	0
b_3	0	15	0

Aufgabe 28:

Im folgenden sind die Ergebnisse einer Umfrage unter 28 Personen wiedergegeben.

Körpergröße	180	179	162	170	180	184	170	160	169	172	182	166	158	164
Gewicht	72	98	52	84	94	95	85	50	59	60	93	47	75	65

Körpergröße	170	170	160	168	179	173	182	178	166	181	176	173	170	161
Gewicht	79	55	53	86	94	70	88	79	76	70	72	79	75	76

(a) Klassieren Sie die Personen anhand des Merkmals „Körpergröße". Verwenden Sie dazu die Klasseneinteilung $[150, 160), [160, 170), \ldots, [180, 190)$. Berechnen Sie anschließend für jede der Klassen die Quartile, das arithmetische Mittel und die Varianz der zugehörigen Gewichtsangaben.

(b) Stellen Sie die Ergebnisse aus (a) grafisch dar, indem Sie für die Körpergewichte der Personen in den einzelnen Klassen Boxplots erstellen. Zeichnen Sie die Boxplots in ein Diagramm (Körpergröße auf der Abszissen-, Körpergewicht auf der Ordinatenachse) jeweils senkrecht über der entsprechenden Klassenmitte des Merkmals „Körpergröße" ein. Welche Informationen entnehmen Sie der Grafik?

(c) Klassieren Sie die Personen nun zusätzlich nach dem Merkmal „Gewicht" anhand der Klassen $[45, 65), [65, 85)$ und $[85, 105)$. Stellen Sie eine Korrelationstabelle der absoluten Häufigkeiten auf, und berechnen Sie anschließend den Kontingenzkoeffizienten.

(d) Erstellen Sie ein dreidimensionales Histogramm für die klassierten Daten.

Aufgabe 29:

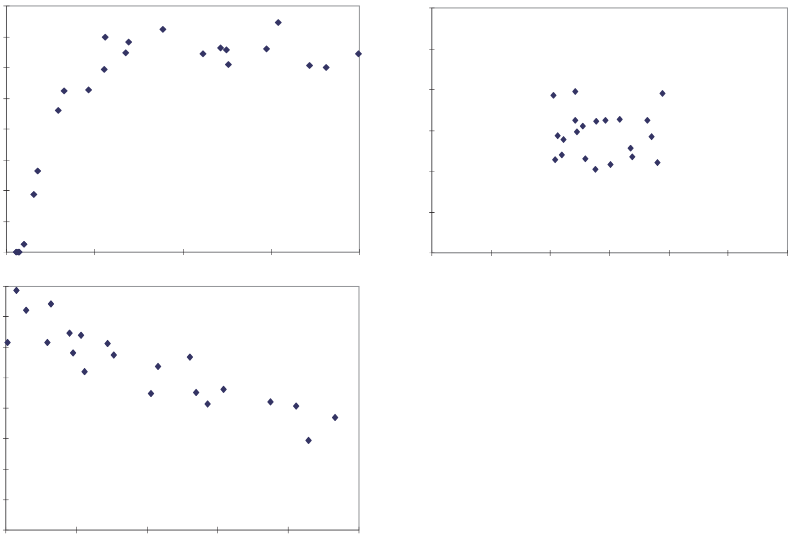
Bei einer Mathematik- und einer Statistik-Klausur wurden von 11 Wiwi-Studenten folgende Punktzahlen erreicht:

Student	1	2	3	4	5	6	7	8	9	10	11
Mathematik	38	47	44	51	35	29	22	14	12	19	9
Statistik	39	34	31	48	46	23	17	12	16	28	10

(a) Zeichnen Sie ein Streudiagramm.

(b) Nehmen Sie an, dass zwischen den Ergebnissen in Mathematik (x) und denen in Statistik (y) der funktionale Zusammenhang $y = f(x) + \epsilon$ besteht, wobei $f(x)$ eine beliebige Funktion von x und ϵ ein Störterm ist.
Welche Form könnte f hier haben?
Wie groß wird der Einfluss des Störterms sein?

(c) Beantworten Sie die Fragen aus (b) für die folgenden Streudiagramme.



Aufgabe 30:

Die folgende Tabelle gibt für die Jahre 1975 - 1994 die Aufwendungen der deutschen Wirtschaft für FuE (Merkmal x) und die Anzahl der Patentanmeldungen (Merkmal y) an.

Jahr	FuE-Aufwendungen in Mrd. DM x_i	Patentanmeldungen in Tausend y_i
1975	14.54	64.595
1985	39.55	45.213
1989	50.81	41.244
1991	55.12	41.799
1992	56.93	43.663
1993	56.76	45.380
1994	56.25	49.011

(a) Zeichnen Sie das Streudiagramm.

(b) Bestimmen Sie die Regressionsgerade $y = a \cdot x + b$.

(c) Berechnen Sie den Korrelationskoeffizienten zwischen x und y.

(d) Welchen Wert für die Anzahl der Patentanmeldungen würden Sie voraussagen, wenn Sie wüssten, dass im Jahr 1998 50 bzw. 60 Mrd. DM für Forschung und Entwicklung aufgewendet werden sollen?

(e) Wiederholen Sie die Aufgabenteile (b)–(d) unter Weglassung des ersten Wertepaares. Wie sind die veränderten Ergebnisse zu interpretieren?

Aufgabe 31:

Von der Controlling-Abteilung eines Konsumgüterherstellers wurde der Zusammenhang zwischen dem Werbebudget w und dem Umsatz u eines Produktes X im Jahr 2001 untersucht. Man erhielt für die 12 Monate folgende Größen:

- Werbeausgaben: insgesamt 120 000 DM im untersuchten Jahr.
- Der durchschnittliche Umsatz des Produktes X in DM beträgt das zehnfache des durchschnittlichen Werbeetats pro Monat.
- Standardabweichung der monatlichen Werbeausgaben im Untersuchungszeitraum: 9 000 DM.
- Standardabweichung der Umsatzdaten: 50 000 DM.
- Als Maß für den Zusammenhang zwischen Werbeausgaben und Umsatz wurde ein Korrelationskoeffizient von 0.90 ermittelt.

(a) Berechnen Sie die Regressionsgerade für den Fall eines einfachen linearen Zusammenhangs der beiden Merkmale auf Monatsebene in der Form $u = m \cdot w + b$.

(b) Welche Maßnahmen würden Sie als Werbemanager ergreifen, falls Sie diesem ermittelten Zusammenhang ohne Einschränkungen vertrauen könnten und das Produkt mit zunehmendem Umsatz den Gewinn des Unternehmens steigert? Warum wäre ein solches Vertrauen unrealistisch?

(c) Skizzieren Sie eine realistischere Funktionsbeziehung, die nicht die Nachteile des Modells in (a) besitzt, und begründen Sie Ihre Modellierung.

Aufgabe 32:

Für 80 Gemeinden verschiedener Größe werden das monatliche Müllaufkommen (= Merkmal y) und die Anzahl der zu Monatsanfang gemeldeten Einwohner (= Merkmal x) einer linearen Regressionsanalyse unterzogen.

(a) Der Analytiker berechnet eine Regressionsgerade, die durch 79 der beobachteten Werte verläuft. Lediglich der Beobachtungswert (x_{80}, y_{80}) liegt nicht auf ihr. Muss zwangsläufig ein Rechenfehler vorliegen, oder ist ein derartiger Befund bei empirischen Daten möglich?

(b) Wären (genau) zwei von Null verschiedene Residuen möglich? Wenn ja, wie?

Aufgabe 33:

Aus der Produktionswirtschaft ist das Erfahrungskurvenkonzept bekannt, das den zeitlichen Zusammenhang zwischen Produktionsmenge und Stückkosten beschreibt. Ein Wirtschaftswissenschaftler hat dies folgendermaßen formuliert: „*Mit jeder Verdoppelung der im Zeitablauf kumulierten Produktionsmenge x fallen die realen Stückkosten K potentiell um 20-30%*“. Es wird also der folgende funktionale Zusammenhang unterstellt:

$$K_t = a x_t^{-b} \qquad (t = 1, \ldots, n)$$

(a) Berechnen Sie Schätzwerte für die Parameter a und b des Erfahrungskurvenkonzeptes mit Hilfe des Kleinste-Quadrate-Ansatzes (KQ-Ansatz), indem Sie eine Linearisierung vornehmen.

(b) Entsprechen die über den Umweg einer Linearisierung gewonnenen Schätzer für a und b denen, die man bei einer direkten KQ-Lösung des Ausgangsproblems erhalten würde? Versuchen Sie, Ihre Antwort zu begründen!

Aufgabe 34:

Passen Sie an die folgenden Daten mit Hilfe des KQ-Ansatzes das Modell $y = mx^2 + b$ an, und berechnen Sie den Wert des Bestimmtheitsmaßes.

x	0.497	2.648	0.247	3.3	3.936	3.505	0.881	2.835	0.025	1.383
y	3.435	15.052	-0.810	24.110	34.402	23.956	5.0323	16.211	0.595	2.735

Aufgabe 35:

Im folgenden sind noch einmal die Klausurergebnisse aus Übung 29 gegeben:

Student	1	2	3	4	5	6	7	8	9	10	11
Mathematik	38	47	44	51	35	29	22	14	12	19	9
Statistik	39	34	31	48	46	23	17	12	16	28	10

(a) Berechnen Sie den Rangkorrelationskoeffizienten von Spearman.

(b) Bei Student 5 erhöht sich die Anzahl der Punkte in der Mathematikklausur nachträglich um 3 auf 38 Punkte. Welche Auswirkung hat diese Änderung auf den Wert des Rangkorrelationskoeffizienten?

Aufgabe 36:

Gegeben sei die folgende Aussage: *In einem x, y-Koordinatensystem befinde sich eine Punktewolke mit Schwerpunkt im Ursprung. Dann gilt: Die Gerade g aus einer linearen Regression von y auf x und die Gerade g' aus einer linearen Regression von x auf y sind nicht identisch, außer x und y sind vollständig miteinander korreliert ($|r| = 1$).*
Führen Sie einen rechnerischen Nachweis für diese Aussage, und überlegen Sie sich die Ursache für den angesprochenen Unterschied.

Aufgabe 37:

(a) Welche der folgenden Aussagen sind richtig, welche falsch? (Begründung!)

(1) Je größer der Anstieg der Regressionsgeraden, desto größer ist der Korrelationskoeffizient.

(2) Ist der Anstieg der Regressionsgeraden positiv, so ist auch der Korrelationskoeffizient positiv.

(3) Das Bestimmtheitsmaß wird mit wachsender Anzahl der Beobachtungen größer.

(4) Ist das Bestimmtheitsmaß gleich 1, so sind alle Residuen gleich 0.

(5) Ist der Korrelationskoeffizient gleich 0, so liegen alle Beobachtungswerte auf einer waagerechten Geraden.

(6) Ist der Korrelationskoeffizient gleich 0, so sind die Merkmale X und Y unabhängig.

(b) Gegeben sei das Modell $y = b + \frac{m}{x}$. Wie ist x zu transformieren, damit eine lineare Regression durchgeführt werden kann?

(c) Überlegen Sie sich ein Beispiel für zwei Merkmale, die unkorreliert aber nicht unabhängig sind.

Aufgabe 38:

Die folgende Zeitreihe beschreibt monatliche Einfuhrmengen eines Industriezweiges in die Bundesrepublik Deutschland:

Monat / Jahr	J	F	M	A	M	J	J	A	S	O	N	D
1954	21	29	38	32	31	24	34	54	63	52	46	38
1955	26	34	40	36	24	28	43	69	63	60	46	42
1956	34	41	46	37	35	37	42	61	47	42	35	37

(a) Berechnen Sie „sinnvolle“ gleitende Durchschnitte für diese Zeitreihenwerte, und vergleichen Sie diese mit den gleitenden Durchschnitten der Ordnung 7.

(b) Zeichnen Sie die zugehörigen Zeitreihen in ein Koordinatensystem.

(c) Bestimmen Sie die saisonbereinigten Werte für einen additiven Ansatz mit konstanter Saisonfigur.

Aufgabe 39:

Ermitteln Sie die 12 Saisonindexziffern für folgende Zeitreihe der Arbeitslosenzahl eines Landes (die Werte sind bereits um die glatte Komponente bereinigt) unter der Voraussetzung einer variablen Saisonfigur.

Monat / Jahr	J	F	M	A	M	J	J	A	S	O	N	D
1976	1.7	1.7	1.1	1.2	0.8	0.7	0.8	0.8	0.9	1.0	1.1	1.4
1977	1.8	1.6	1.5	0.9	0.9	0.6	0.6	0.8	1.0	0.9	1.2	1.3
1978	1.6	1.5	1.3	0.9	0.7	0.8	0.7	0.8	0.8	1.1	1.0	1.2

Aufgabe 40:

(a) Man stelle den Laspeyres- und Paasche-Mengenindex als gewichtetes arithmetisches Mittel von Mengenmeßzahlen dar und interpretiere die Gewichte.

(b) Es gilt für ein Gut i: Umsatz = Preis x Menge.
Läßt sich diese Gleichung auch auf die entsprechenden Indices übertragen?

(c) Ein Unternehmer stelle 3 Produkte A, B und C her; in den Jahren 2010 und 2011 werden folgende Mengen zu folgenden Preisen abgesetzt:

	Preise je Stück		**abgesetzte Mengen**	
	2010	2011	2010	2011
Produkt A	20	21	100	110
Produkt B	45	44	95	115
Produkt C	53	58	80	85

Zeigen Sie, dass die folgende Indentitäten gelten:

$$W_{0,n} = P^L_{0,n} \cdot Q^P_{0,n} \qquad\qquad W_{0,n} = P^P_{0,n} \cdot Q^L_{0,n}.$$

Aufgabe 41:

(a) Erläutern Sie die folgenden Begriffe, und geben Sie jeweils Beispiele an:

- Grundraum Ω
- σ-Algebra $A(\Omega)$
- Wahrscheinlichkeitsmaß P

Welche Bedeutung haben die einzelnen Komponenten für den *Wahrscheinlichkeitsraum*?

(b) Bilden Sie alle σ-Algebren für den Grundraum $\Omega := \{a, b, c\}$.

(c) Gegeben sei der Grundraum $\Omega := \{a, b, c, d\}$. Zeigen Sie, dass $A(\Omega) := \{\emptyset, \{a, d\}, \{b, c\}, \Omega\}$ eine σ-Algebra auf Ω ist.

Aufgabe 42:

Betrachten Sie eine vereinfachte Variante eines Roulette-Spiels für einen Spieler mit den folgenden Spielregeln: In jeder Runde kann der Spieler auf eine der Zahlen $1, \ldots, 36$ (nicht aber auf die 0!) setzen. Er gewinnt, falls die von ihm gesetzte Zahl mit der durch das Roulette-Rad ausgespielten übereinstimmt.

(a) Beschreiben Sie den Wahrscheinlichkeitsraum.

(b) Sei X die erste und Y die zweite Ziffer der gespielten Zahl (für die Zahlen $0, \ldots, 9$ wird X auf 0 gesetzt). Berechnen Sie die Wahrscheinlichkeit dafür, dass eine Zahl

- aus $\{1, \ldots, 36\}$,

- mit $X = 3$,
- mit $X = 2$ oder $X = 3$,
- aus $\{25, \ldots, 28\}$,
- mit $X \geq 2$ und $Y \leq 4$ bzw.
- mit $X + Y = 4$

gewinnt.

Aufgabe 43:

Bei zahlreichen Brettspielen (u.a. „Mensch ärgere dich nicht“) gilt folgende Vereinbarung für das Würfeln:

Würfelt ein Spieler eine der Zahlen 1 bis 5, so kommt anschließend der nächste Spieler dran. Würfelt er hingegen eine 6, so darf er noch mal würfeln (so lange, bis er keine 6 mehr würfelt).

Formulieren Sie den zugehörigen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}(\Omega), P)$.

Aufgabe 44:

Gegeben ist die folgende Summenhäufigkeitstabelle für die Einkommensverteilung einer Gruppe von Berufstätigen (Einkommen in [k€]):

rechte Klassengrenzen	10	20	25	30	40	60	85
Summenhäufigkeit (in [%])	5	20	40	65	85	90	100

Die rechte Klassengrenze gehöre dabei jeweils zur Klasse.

(a) Bestimmen und zeichnen Sie die Summenhäufigkeitsfunktion. Erstellen Sie ein Histogramm.

(b) Definieren Sie auf Grundlage der Summenhäufigkeitsfunktion ein Wahrscheinlichkeitsmaß P für die Höhe des Einkommens einer zufällig aus der Gruppe der Berufstätigen herausgegriffenen Person.

(c) Bestimmen Sie mit Hilfe von P die Wahrscheinlichkeiten für folgende Ereignisse: Das Einkommen einer zufällig herausgegriffenen Person ist:

- in dem Intervall $(20; 30]$.
- in dem Intervall $[20; 30)$.
- gleich 55.
- mindestens 75.
- geringer als 25 oder größer als 54.
- in sämtlichen Intervallen A_i der Form $A_i := [20 + \frac{5}{i}; 30 + \frac{15}{i}]$, $i \in \mathbb{N}$.

Veranschaulichen Sie sich Ihre Lösungen anhand des Histogramms sowie des Graphen der Summenhäufigkeitsfunktion.

Aufgabe 45:

Ein Händler will zu Silvester 25 Feuerwerkskörper, die ihm aus früheren Jahren übriggeblieben sind, loswerden. Er verspricht einem daran Interessierten, daß mindestens 60 % davon noch funktionsfähig sind. Dieser verlangt, 5 der 25 Feuerwerkskörper sofort ausprobieren zu dürfen, und ist bereit, die restlichen 20 zu kaufen, wenn mindestens 3 der 5 geprüften funktionieren. Der Händler ist mit einem Test einverstanden, will jedoch nur 3 Feuerwerkskörper dafür zur Verfügung stellen.

(a) Stellen Sie den Wahrscheinlichkeitsraum für den Test von 3 Feuerwerkskörpern auf.

(b) Sei X eine Zufallsvariable, die die Anzahl der funktionsfähigen Feuerwerkskörper in der Stichprobe angibt. Bestimmen Sie die Wahrscheinlichkeitsverteilung von X.

(c) Wie groß ist die Wahrscheinlichkeit, daß das Geschäft zustande kommt, wenn der Händler sich auf den Test von 5 Feuerwerkskörpern einläßt und tatsächlich

- 60 %
- 80 %
- 20 %

der 25 Feuerwerkskörper noch funktionsfähig sind?

Aufgabe 46:

Im Wareneingang einer Unternehmung werden Transistoren auf ihre Funktionsfähigkeit hin untersucht. Bei einer Warenpartie von $N = 100$ Teilen wird eine Stichprobe vom Umfang $n = 10$ gezogen. Aus langjähriger Erfahrung weiss man, dass im Mittel 3% der Transistoren fehlerhaft sind. Die Warenpartie wird abgelehnt, wenn mindestens 1 Transistor in der Stichprobe defekt ist.

(a) Berechnen Sie die Wahrscheinlichkeit, dass die Warenpartie abgelehnt wird, wenn die Stichprobe ohne Zurücklegen gezogen wird.

(b) Ein Mitarbeiter schlägt vor, die Stichprobe mit Zurücklegen zu ziehen, weil dies den Rechenaufwand vermindere. Berechnen Sie für diesen Fall die Wahrscheinlichkeit, die Warenpartie abzulehnen, und vergleichen Sie das Ergebnis mit dem aus (a).

(c) Ein weiterer Mitarbeiter schlägt vor, den Umfang der Warenpartie auf $N = 1000$ zu erhöhen. Überprüfen Sie, ob für diese Warenpartie der Unterschied zwischen Ziehen ohne Zurücklegen und mit Zurücklegen ins Gewicht fällt.

Aufgabe 47:

(a) Für welche Parameter $a, b \in \mathbb{R}$ sind auf $\Omega = \{0, 1, \ldots, k, \ldots, n\}$ die Werte

$$\alpha_k = \binom{n}{k} a^k b^{n-k}$$

für ein Wahrscheinlichkeitsmaß brauchbar? Interpretieren Sie dieses Wahrscheinlichkeitsmaß im Hinblick auf praktische Anwendungen. Skizzieren Sie die zugehörige Verteilungsfunktion.

(b) Geben Sie Bedingungen für die Parameter $p_1, \ldots, p_m \in \mathbb{R}$ an, so dass auf der Menge Ω mit

$$\Omega \quad = \quad \{(k_1, \ldots, k_m) \,|\, k_i \in \{0, \ldots, n\}, \sum_{i=1}^{m} k_i = n\}$$

durch die Werte

$$\alpha_{k_1,\ldots,k_m} \quad = \quad \binom{n}{k_1 \ldots k_m} p_1^{k_1} \ldots p_m^{k_m}$$

ein Wahrscheinlichkeitsmaß definiert wird. Interpretieren Sie dieses Wahrscheinlichkeitsmaß im Hinblick auf praktische Anwendungen.

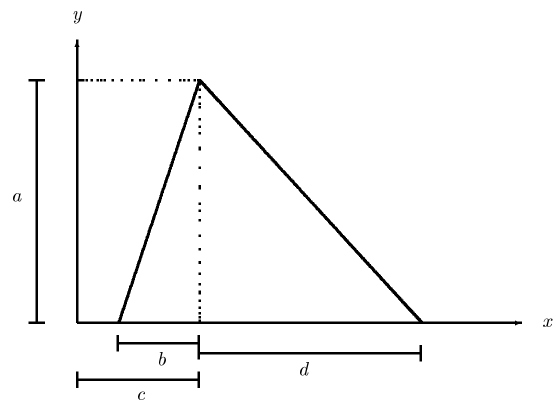
(c) Leiten Sie aus der geometrischen Reihe mit den Gliedern $(\frac{1}{2})^n$ eine Wahrscheinlichkeitsverteilung auf $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ ab. Verallgemeinern Sie Ihr Ergebnis für die Glieder a^n.

Aufgabe 48:

$F(x_i) \;=\; 1 - 2^{-x_i}$ für $x_i \;=\; 1, 2, \ldots$ gebe die Werte der Verteilungsfunktion einer diskreten Zufallsvariable an sämtlichen Sprungstellen an. Geben Sie die Verteilungsfunktion auf $\mathbb{R}$ an. Stellen Sie fest, welche Werte die Zufallsvariable mit positiver Wahrscheinlichkeit annimmt und berechnen Sie die Wahrscheinlichkeitsverteilung der Zufallsvariable.

Aufgabe 49:

(a) Bestimmen Sie für die folgende geometrische Figur zu den Parameter a, b und c den Parameter d derart, dass es sich um eine Dichtefunktion handelt. Berechnen Sie eine zugehörige Verteilungsfunktion. Ist diese eindeutig?



(b) Die Zufallsvariable Y habe die folgende Dichtefunktion:

$$f_Y(y) = \begin{cases} \alpha \cdot (1 - (y-1)^2) & , y \in [0;1) \\ \alpha \cdot \sqrt{2-y} & , y \in [1;2] \\ \beta & , \text{sonst} \end{cases}$$

Bestimmen Sie die Konstanten α und β, und berechnen Sie die Verteilungsfunktion.

(c) Bestimmen Sie für die Zufallsvariable Y aus Aufgabenteil (b) den Median sowie den Erwartungswert.

Aufgabe 50:

(a) Berechnen Sie für die folgende diskrete Verteilung den Modalwert, den Median, den Erwartungswert, die Varianz, das 2. Moment sowie das 0.85-Quantil.

x	-2	-1	0	1	2	3	4
$P(X = x)$	0.1	0.1	0.4	0.1	0.05	0.1	0.15

(b) Bestimmen Sie für die Pareto-Verteilung, die durch die Dichtefunktion

$$f_X(x) = \begin{cases} k\,x^{-n-1} & \text{für } x \geq c \\ 0 & \text{sonst} \end{cases}$$

charakterisiert wird ($c, n, k > 0$) den Parameter k in Abhängigkeit von c und n, die Verteilungsfunktion sowie den Modalwert, den Erwartungswert, die Varianz und das 0.85-Quantil.

Aufgabe 51:

Gegeben ist folgende Funktionenfamilie ($a > 0$):

$$f_k(x) \;=\; \begin{cases} k\,(a - |x|) & \text{für } -a < x < a \\ 0 & \text{sonst} \end{cases}$$

(a) Bestimmen Sie k in Abhängigkeit von a, so dass $f_k(x)$ eine Dichtefunktion wird. Skizzieren Sie diese.

(b) Bestimmen Sie die zugehörige Verteilungsfunktion.

Aufgabe 52:

Gegeben ist die Dichtefunktion der Cauchy-Verteilung

$$f_X(x) \;=\; \frac{1}{\pi\,(1\;+\;x^2)}, \quad x \in \mathbb{R}$$

(a) Untersuchen Sie, ob die Cauchy-Verteilung einen Erwartungswert besitzt. Was folgt aus Ihrem Ergebnis für die Varianz der Verteilung?

(b) Skizzieren Sie den Graphen der Dichtefunktion.

Aufgabe 53:

(a) Geben Sie mit Hilfe einer Tabelle für die Standardnormalverteilung $N(0,1)$ an:

(i) $P((-\infty, 1.50))$, $P((-\infty, -2.05))$, $P((1.65, \infty))$, $P((-0.78, 1.75))$.

(ii) c mit $P([c, \infty)) = 0.85$.

(iii) c mit $P([-c, c]) = 0.95$.

(b) Berechnen Sie mit Hilfe der Funktionen "NORMVERT" und "NORMINV" aus Excel für die Normalverteilung $N(5, 9)$:

(i) $P((-\infty, 2))$, $P((-\infty, 6))$, $P((2, 7))$, $P((4.5, \infty))$

(ii) c mit $P((-\infty, c]) = 0.84$.

(iii) c mit $P([5-c, 5+c]) = 0.57$.

3 Literatur zur Vorlesung

3.1 Lehrbücher und Nachschlagewerke zur Statistik

- Bamberg, G., Baur, F. und Krapp, M.: Statistik, 15. überarb. Auflage. Oldenbourg, München 2009, ISBN 978-3486590883.
- Hartung, J., Elpelt, B. und Klösener, K.-H.: Statistik, 15., überarb. und erw. Aufl., Oldenbourg, München 2009, ISBN 978-3486590289.
- Rinne, H.: Taschenbuch der Statistik, 4. überarb. u. erw. Aufl., Harri Deutsch, Frankfurt a. M. 2008, ISBN 978-3817118274,

3.2 Bücher zu Statistik mit Excel

- Anderson, D., Sweeney D.J. und Williams, T.A.: Essentials of Modern Business Statistics with Microsoft Excel. 4th edition, South Western Educational Publishing, 2008, ISBN 978-0324590067.
- McClave, J.T. und Sincich, T.: Statistics, 9th edition. Prentice Hall, 2003 ISBN 0-13-065598-8.
- Kohler, H.: Statistics for business and Economics, Thomson Learning, 2002, ISBN 0-03-029731-1.
- Black, K. und Eldridge D.: Business and Economic Statistics. Thomson Learning, 2002, ISBN 0-324-01726-X.
- Zwerenz, K.: Statistik verstehen mit Excel, 2. verb. Aufl., Oldenbourg 2007, ISBN 978-3486585919

3.3 Bücher zur Deskriptiven Statistik

- Bol, G.: Deskriptive Statistik, 6. überarb. Aufl., Oldenbourg, München 2004, ISBN 978-3486576122.
- V. d. Lippe, P.: Deskriptive Statistik, 7. Aufl., Oldenbourg, München 2006, ISBN 978-3486578638.
- Mosler, K. und Schmid, F.: Beschreibende Statistik und Wirtschaftsstatistik, 4. akt. und verb. Aufl., Springer, Berlin 2009, ISBN 978-3642015564.
- Schwarze, J.: Grundlagen der Statistik, Beschreibende Verfahren, 11. vollst. überarbeitete Aufl., NWB, Herne 2009, ISBN 978-3482594816.

3.4 Bücher zur Wahrscheinlichkeitstheorie

- Bol, G.: Wahrscheinlichkeitstheorie, 6. überarb. Aufl., Oldenbourg, München etc., 2007, ISBN 978-3486584356.
- Henze, N.: Stochastik für Einsteiger, 8. erw. Auflage, Vieweg + Teubner, Wiesbaden, 2009, ISBN 978-3834808158.

- Mosler, K. und Schmid, F.: Wahrscheinlichkeitsrechnung und schließende Statistik, 4. verb. Aufl., Springer, Berlin 2010, ISBN 978-3642150098.
- Schwarze, J.: Grundlagen der Statistik 2: Wahrscheinlichkeitsrechnung und induktive Statistik, 9. vollst. überarb. Aufl., NWB, Herne 2009, ISBN 978-3482568695

3.5 Lernsoftware

- Dialekt-Projekt: Statistik interaktiv, deskriptive Statistik. Springer, Berlin etc., 2001, ISBN 978-3540149118.
- Härdle, W., Rönz, B.: MMStat. Springer, Berlin 2001, ISBN 3-540-14893-0. http://mars.wiwi.hu-berlin.de/mediawiki/mmstat_de/index.php/Hauptseite
- Mittag, H.-J. und Stemann, D.: Statistik, beschreibende Statistik und explorative Datenanalyse. Hanser Fachbuchverlag, München 2004, ISBN 978-3446227521.
- Velleman, P. : Activ Stats for Excel. Pearson, London 2003, ISBN 978-0201782455.

4 Tabelle zur Standardnormalverteilung

Standardnormalverteilung: $\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5078	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8600	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9065	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9807	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9867	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9959	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9993	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
4.0	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
4.4	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999