Technische Universität Berlin

FAKULTÄT VII
WIRTSCHAFT UND
MANAGEMENT

Institut für Volkswirtschaftslehre
und Wirtschaftsrecht

Fachgebiet
Ökonometrie und Wirtschaftsstatistik

Konrad Urbanski, ORat

AUFGABENSAMMLUNG

STATISTIK

Inhalt

Version 8.0

A

Häufigkeiten

Aufgabe A 1:

A: Um die Auswirkung der Regelstudienzeit zu demonstrieren, wurden die Studienzeiten von 2oo Wirtschaftsingenieuren erhoben, die in den vergangenen **4** Semestern ihr Studium erfolgreich abgeschlossen haben (fiktive Daten, damit sich's leichter rechnen läßt). Es ergaben sich folgende Daten:

Semesterzahl	1o	11	12	13	14	15
rel. Häufigkeiten	o,1	o,1	o,4	o,2	o,15	o,o5

1) Wie heißt die statistische Größe (Merkmal) und wie ist sie skaliert ?
2) Bestimmen Sie die absoluten Häufigkeiten !
3) Bestimmen Sie die absoluten kumulierten Häufigkeiten !
4) Skizzieren sie die Häufigkeits- und Verteilungsfunktion der statistischen Größe !
5) Wieviele Semester höchstens benötigen die 1o% schnellsten Studenten ?
6) Wieviele Semester mindestens benötigen die 8o% langsamsten Studenten ?
7) Geben sie die Semesterzahl an, die genau 2o% der Studenten benötigen !

B: Von nun an soll nur noch das Merkmal Y: "Semesterzahl" mit den Ausprägungen

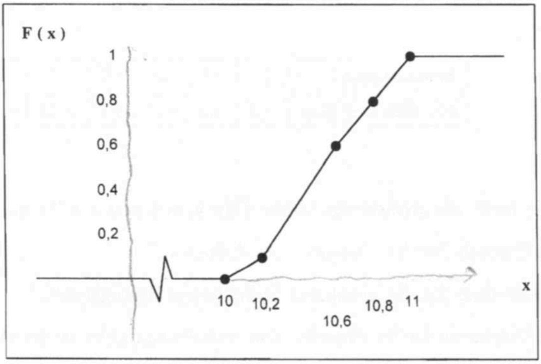
"klein" (weniger als 12 Semester)
"mittel" (genau 12 Semester)
"groß" (mehr als 12 Semester)

betrachtet werden.

1) Welche Skalierungsart liegt jetzt vor ?
2) Stellen Sie die Häufigkeiten graphisch dar !
3) Ist es sinnvoll, bei einem nominal skalierten Merkmal eine Verteilungsfunktion anzugeben ? Begründung !

Aufgabe A 2

Ein Sportverein hat sich in seiner Leichtathletikabteilung einen Schwerpunkt in der Förderung des 100 m – Laufs gesetzt.
Nach einem Jahr intensivsten Trainings wurden die Zeiten der 20 Läufer des Vereins gemessen. Dabei ergab sich folgende Verteilungsfunktion :



1) Zeichnen Sie das zur Verteilungsfunktion gehörende Histogramm !

2) Welche Zeit höchstens benötigen die 80 % schnellsten Läufer ?

3) Welche Annahme über die Verteilung der Meßwerte innerhalb der Gruppe wurde bei der Ermittlung der Verteilungsfunktion gemacht ?

4) Lesen Sie aus der Verteilungsfunktion ab, welche Zeit die 5 besten Läufer höchstens benötigen ?

Aufgabe A 3

Tennis–As Boris B. blickt wieder einmal auf eine völlig mißglückte Saison zurück. Insgesamt nahm er an 40 Turnieren in aller Welt teil. Jedes Turnier ging über sechs Runden :

1. Runde
2. Runde
3. Runde ≅ Achtelfinale
4. Runde ≅ Viertelfinale
5. Runde ≅ Halbfinale
6. Runde ≅ Finale

Gespielt wurde in jeder Runde im K.O.–System, d.h. jeder Spieler, der seine Partie verlor, schied aus.
Die Bilanz seiner Turnierergebnisse läßt bei Boris keine Freude aufkommen. :

Zwar stand er zweimal im Finale und wurde sechsmal erst im Halbfinale besiegt, zehnmal verlor er jedoch schon in der ersten Runde und sechzehnmal kam für ihn in der 2. Runde das „ Aus “. In der dritten Runde schied er bei keinem Turnier aus.

1) Wie heißt das statistische Merkmal und wie ist es skaliert ?

2) Bestimmen Sie tabellarisch die absoluten und relativen Häufigkeiten sowie die absoluten und relativen kumulierten Häufigkeiten !

3) Zeichnen Sie die empirische (relative) Verteilungsfunktion !

4) Bei wieviel Prozent der Turniere schied Boris vor dem Achtelfinale aus ?

5) Bei wieviel Prozent der Turniere erreichte Boris mindestens das Halbfinale ?

6) Bei wie vielen Turnieren spielte Boris noch in der zweiten Runde ?

7) In welcher Runde spätestens schied Boris bei 80 % der Turniere aus ?

8) In welcher Runde schied Boris bei genau 40 % der Turniere aus ?

9) Interpretieren Sie – bezogen auf den Aufgabentext – <u>kurz</u> den Funktionswert $F(x)$ an der Stelle $x = 7$!

Aufgabe A 4:

Ein Universitätspräsident gibt eine statistische Untersuchung in Auftrag, mit der festgestellt werden soll, wieviel Zeit die Studenten für ihr Studium aufwenden. Die Befragung von 100 Studenten der Technischen Lernanstalten Berlins brachte folgendes Ergebnis:

Keiner der Befragten lernt 12 und mehr Stunden am Tag. Bei 22 Studenten wurde festgestellt, daß sie sich mindestens 6 Stunden am Tag mit ihrem Studium beschäftigen. *Weniger* als 3 Stunden am Tag investieren 30% der Befragten in ihr Studium. Die Masse der Studenten (65%) beschäftigt sich zwischen 3 und 8 Stunden täglich mit Studienangelegenheiten.

1) Wie heißt das untersuchte Merkmal und wie ist es skaliert ?

2) Bestimmen Sie tabellarisch die absoluten, die relativen und die relativen kumulierten Häufigkeiten. Teilen Sie zu diesem Zweck die Daten in *vier* Gruppen ein !

3) Stellen Sie die absoluten Häufigkeiten graphisch dar !
   Für welche Darstellungsform haben Sie sich entschieden ? Begründen Sie kurz Ihre Wahl !

4) Stellen Sie die relativen kumulierten Häufigkeiten graphisch dar !
   Auf welcher Annahme über die Verteilung der Daten innerhalb der einzelnen Gruppen ist Ihre Darstellung begründet ?

5) Ermitteln Sie graphisch, wieviele Stunden höchstens genau die Hälfte der Befragten täglich sich mit dem Studium beschäftigt !

6) Wieviele der Befragten arbeiten mindestens 5 Stunden täglich für ihr Studium ?

Aufgabe A 5

Der Eisverkäufer Hardy möchte seinen Kunden täglich frisches Eis verkaufen. Damit er immer die richtigen Mengen vorrätig hat, und in Zukunft besser planen kann, zählt er an einem Tag bei 200 Kunden den Konsum an Eiskugeln :

- Kein Kunde hat 2 oder mehr als 6 Kugeln verlangt.
- Je 20 % der Kunden wollte 1 Kugel oder mindestens 5 Kugeln.
- Höchstens 3 Kugeln verlangten 45 % der Kunden.
- die Summe der Anteile der Kunden, die 4 bzw. 6 Kugeln wollten, war zehnmal so hoch wie der Anteil der Kunden, die 5 Kugeln verlangten.

1) Wie lautet die untersuchte statistische Größe und wie ist sie skaliert ?

2) Bestimmen Sie tabellarisch die absoluten Häufigkeiten sowie die absoluten kumulierten Häufigkeiten !

3) Stellen Sie die absoluten Häufigkeiten graphisch dar !

4) Stellen Sie die relativen kumulierten Häufigkeiten graphisch dar !
Wie nennt man diese graphische Darstellung ?

5) Wieviel Prozent der Kunden haben höchstens 5 Kugeln Eis verlangt ?

6) Wie viele Kugeln Eis mindestens kauften 80 % der Kunden ?

7) Wie viele Kugeln Eis kauften genau 35 % der Kunden ?

B

Empirische Kennzahlen

Aufgabe B 1

Herr Maier besitzt einen Gartenzwerg–Großhandel mit drei Filialen : Berlin, New York und Flensburg. Am Ende des Geschäftsjahres möchte er einen Überblick über die Geschäftslage erhalten und fordert deshalb in allen drei Filialen Informationen über die innerhalb des letzten Jahres eingegangenen Aufträge an.

A

Seine Berliner Filiale übermittelt ihm folgende Informationen :

Auftragshöhe in Tsd. € von . . . unter	Anzahl der Aufträge
0 – 20	**15**
20 – 50	**30**
50 – 150	**45**
150 – 300	**10**

1) Bestimmen Sie die relativen kumulierten Häufigkeiten !
2) Ermitteln Sie <u>rechnerisch</u> das untere und obere Quartil sowie den Median der gruppierten Daten. Erstellen Sie einen Box–Plot !
3) Berechnen Sie die durchschnittliche Auftragshöhe (arithm. Mittel) !
4) Von welcher Annahme gehen Sie bei dieser Berechnung aus ?
5) Wie weit kann das arithm. Mittel der in der letzten Gruppe zusammengefaßten Originaldaten maximal von jenem Wert abweichen, den Sie als Mittelwert dieser Gruppe angenommen haben ?

B

Die New Yorker Filiale antwortet auf Herrn Maiers Anfrage kurz und bündig :

$$\textbf{95 Aufträge} \quad ; \quad \bar{x} = 60.000\ \$ \quad ; \quad s^2 = 16.000.000\ \$^2$$

Um besser vergleichen zu können, möchte Herr Maier diese Angaben in € kennen. Berechnen Sie die entsprechenden Werte, falls folgender Dollarkurs gilt :

$$1\ \$ = 1{,}5\ €$$

C

Aus Flensburg erhält Herr Maier folgende Daten :

2.000 € | **12.000 €** | **17.000 €** | **12.000 €** | **200.000 €**

1) Welchen Mittelwert bzw. welche Mittelwerte können Sie in diesem Fall sinnvoll angeben ? Begründung !

Ein Fernschreiben korrigiert die Höhe des letzten Auftrages auf **2.000 €** .

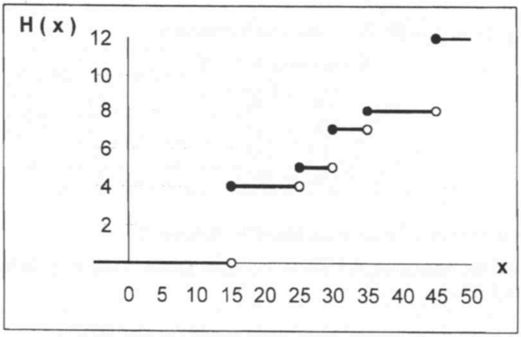
2) Berechnen Sie aus den korrigierten Daten das arithm. Mittel und die Standardabweichung !
3) Vergleichen Sie die Streuung der Auftragshöhe in New York und Flensburg mit Hilfe einer geeigneten Maßzahl !
4) Berechnen Sie die durchschnittliche Auftragshöhe (arithm. Mittel) für das gesamte Unternehmen !

<u>Ergänzung</u> :

5) Berechnen Sie die (emp.) Varianz der Auftragshöhe [in 1000 €]^2 für das gesamte Unternehmen !
6) Welcher Anteil (in %) der Gesamtvarianz wird durch Berücksichtigung der Filialen untereinander verursacht ?

Aufgabe B 2

Der Produktionsleiter einer Firma, die Werkzeuge herstellt, steht jede Woche vor dem Problem, einen Maschinenbelegungsplan erarbeiten zu müssen. Dazu benötigt er alle Aufträge mit den jeweiligen Bestellmengen. Der Computer der Firma liefert ihm die Daten dieser Woche in folgender graphischen Darstellung :



1) Wie heißt das hier untersuchte statistische Merkmal und wie ist es skaliert ?

2) Geben Sie tabellarisch die absoluten und relativen Häufigkeiten an !

3) Geben Sie – ohne zu rechnen – den Wert des arithmetischen Mittels und des Medians an ! Begründen Sie kurz Ihre Antwort !

4) Bestimmen Sie die empirische Varianz s^2 !

Teilen Sie die Bestellmengen in folgende drei Gruppen ein :

0 – 20 | 20 – 40 | 40 – 60

5) Berechnen Sie das arithmetische Mittel und die empirische Varianz aus den gruppierten Daten !

6) Wie erklären Sie sich etwaige Abweichungen der Parameterwerte berechnet aus den Originaldaten bzw. den gruppierten Daten ?

Der Produktionsleiter will die Fertigung der bestellten Werkzeuge auf die drei Maschinen A , B und C verteilen.

- Vom Bestellvolumen dieser Woche soll Maschine A genau 40 % fertigen.

- Die von Maschine B produzierte Menge hängt aus technischen Gründen von der Produktionsmenge der Maschine A ab und beträgt nur 25 % der von Maschine A gefertigten Menge.

- Der Rest des Bestellvolumens soll von Maschine C gefertigt werden.

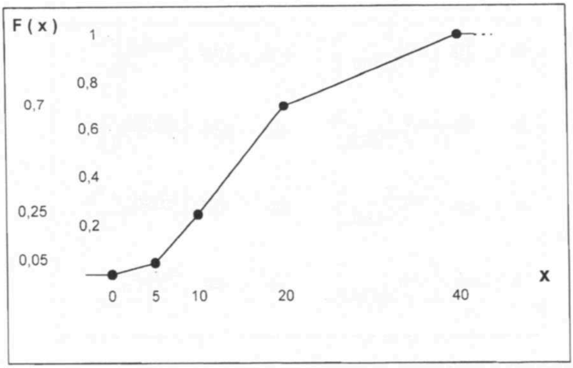
7) Wie viele Werkzeuge werden von Maschine A gefertigt ?

8) Wieviel Prozent des gesamten Bestellvolumens wird von Maschine B gefertigt ?

9) Wie viele Werkzeuge werden von Maschine C gefertigt ?

Aufgabe B 3

Im Rahmen einer medizinischen Untersuchung wurden 200 Personen befragt, wie viele Zigaretten sie pro Tag rauchen.
Das Ergebnis dieser Untersuchung ist in folgender Graphik dargestellt :



1) Wie nennt man diese graphische Darstellung ?
   Was wurde über die Verteilung der Daten angenommen ?

2) Zeichnen Sie die absolute Häufigkeitsverteilung !
   Welches Prinzip ist dabei zu beachten ?

3) Bestimmen Sie das arithmetische Mittel und den Median aus den Daten !

4) Interpretieren Sie beide Werte hinsichtlich der Verteilungsform der Daten !

5) Geben Sie an, wie hoch der Anteil derjenigen Raucher ist, die mindestens 20 Zigaretten pro Tag rauchen !

6) Angenommen, bei einer nochmaligen Durchsicht der Daten hat man festgestellt, daß die Häufigkeiten der 3. Gruppe und der 4. Gruppe vertauscht wurden.
   Ändert sich dadurch der Wert des Medians ? Begründung !

Aufgabe B 4

In einer Walzabteilung bedienen 4 Arbeiter (A , B , C und D) vier unterschiedlich moderne Maschinen. Sie benötigen jeweils folgende Durchschnittszeiten zum Walzen eines Stückes Blech :

A :	**20** $\left[\text{sec}/\text{Stück}\right]$	$\Leftrightarrow$	**180**	$\left[\text{Stück}/\text{h}\right]$
B :	**30** $\left[\text{sec}/\text{Stück}\right]$	$\Leftrightarrow$	**120**	$\left[\text{Stück}/\text{h}\right]$
C :	**60** $\left[\text{sec}/\text{Stück}\right]$	$\Leftrightarrow$	**60**	$\left[\text{Stück}/\text{h}\right]$
D :	**60** $\left[\text{sec}/\text{Stück}\right]$	$\Leftrightarrow$	**60**	$\left[\text{Stück}/\text{h}\right]$

1) Die Arbeiter arbeiten gleichlange.

   a) Welche Durchschnittszeit pro Stück wird in der Abteilung benötigt ?

   b) Wieviel Stück pro Stunde werden im Durchschnitt je Maschine gewalzt ?

2) Den Arbeitern sind folgende Stückkontingente vorgegeben :

A :	**1.000**	Stück
B :	**500**	Stück
C :	**300**	Stück
D :	**200**	Stück

   a) Welche Durchschnittszeit pro Stück wird in der Abteilung benötigt ?

   b) Wieviel Stück pro Stunde werden im Durchschnitt je Maschine gewalzt ?

Aufgabe B 5:

Bei einer Untersuchung von 78 Daten wurde ein Durchschnitt von 2.813 bestimmt.
Zu diesen 78 Daten kam nachträglich noch eine Angabe in Höhe von 2.100 hinzu.

Wie groß ist nun der Durchschnitt, wenn es sich dabei um:

1) das arithmetische Mittel

2) das harmonische Mittel

gehandelt hätte?

Aufgabe B 6

A

Bert ist ein Computerfreak. Damit er im Computerbereich nicht den Anschluß verliert, möchte er ständig über alle Entwicklungen informiert sein. Zu diesem Zweck hat er seinen Rechner – rund um die Uhr – mit einem sog. „ Internet–Server " verbunden, wofür er allerdings saftige Telefongebühren bezahlen muß. Aufgrund einer Tarifänderung der Telefongesellschaft befürchtet Bert nun, daß er sich diesen Luxus nicht mehr leisten kann, was genau dann der Fall ist, wenn er für die Telefonverbindung durchschnittlich mehr als **6 €** pro Stunde zahlen muß.

Neuer Tarif für Telefongespräche

Tarif	Uhrzeit von ... bis unter	Sprechzeit in Sekunden pro Takteinheit à **8** Cent
Nachttarif	$1.^{00}$ - $6.^{00}$	120
Vormittagstarif	$6.^{00}$ - $12.^{00}$	40
Nachmittagstarif	$12.^{00}$ - $19.^{00}$	30
Abendtarif	$19.^{00}$ - $1.^{00}$	60

Berechnen Sie für den neuen Tarif, wieviel die Telefonverbindung durchschnittlich in **€ pro Stunde** kostet !

Kann sich also Bert in Zukunft sein Hobby noch leisten ?

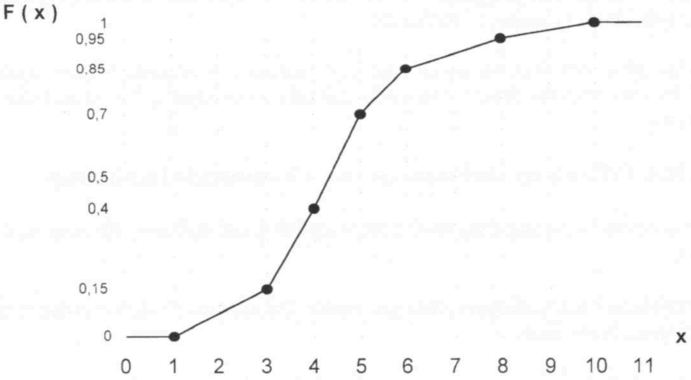
B

Ein Flugzeug soll eine Strecke von 9000 km mit einer durchschnittlichen Geschwindigkeit von 750 $^{km}/_{h}$ fliegen. Es legt die ersten 3000 km mit einer Geschwindigkeit von 250 $^{km}/_{h}$ zurück.

Wie schnell muß das Flugzeug auf der restlichen Strecke fliegen, damit die gewünschte durchschnittliche Geschwindigkeit erreicht wird ?

Aufgabe B 7

Für das monatliche Bruttoeinkommen [in 1000 €] der 2000 Angestellten eines Unternehmens erhielt man folgende (empirische) Verteilungsfunktion F (x) :



1) Wie heißt das hier untersuchte statistische Merkmal und wie ist es skaliert ?

2) Bestimmen Sie tabellarisch die absolute und die relative Häufigkeitsverteilung dieses statistischen Merkmals !

3) Stellen Sie die absolute Häufigkeitsverteilung graphisch dar !

4) Wie groß ist nach dieser Darstellung der Anteil der Angestellten, die monatlich genau 6.400 € verdienen ?

5) Ermitteln Sie rechnerisch , wieviel Prozent der Angestellten monatlich mehr als 5.800 € verdienen !

6) Wie hoch ist das (geschätzte) monatliche Bruttoeinkommen aller Angestellten dieses Unternehmens ?

7) Wie hoch ist das durchschnittliche monatliche Bruttoeinkommen in diesem Unternehmen
   a) nach dem arithmetischen Mittel ?
   b) nach dem Median (rechnerische Lösung !) ?

8) Berechnen Sie den (empirischen) Quartilskoeffizienten der Schiefe !

9) Beurteilen Sie die Schiefe der vorliegenden Häufigkeitsverteilung
   a) nach der „ Lageregel " !
   b) nach dem (empirischen) Quartilskoeffizienten der Schiefe !

Aufgabe B 8

Richtig oder falsch ?

1) Man spricht vom „ Prinzip der Flächentreue " , wenn bei der graphischen Darstellung von absoluten Häufigkeiten gruppierter Daten der Flächeninhalt eines Blockes proportional ist zu der darzustellenden absoluten Häufigkeit.

2) Die Häufigkeitsverteilungen zweier kardinal skalierter Merkmale müssen auch dann nicht identisch sein, falls die beiden Merkmale dasselbe arithmetische Mittel und dieselbe Varianz aufweisen.

3) Statistische Größen lassen sich hinsichtlich ihres Wertebereichs klassifizieren.

4) Die empirische Verteilungsfunktion ist nicht nur für Werte definiert, die auch beobachtet wurden.

5) Der empirische Quartilsabstand eines geordneten Datensatzes ist ein Streuungsmaß, das ausreißerempfindlich ist.

6) Die Sprunghöhen einer empirischen (relativen) Verteilungsfunktion entsprechen den relativen kumulierten Häufigkeiten.

7) Der Variationskoeffizient ist eine dimensionslose Maßzahl.

8) Falls gilt : $h(a_j) = n$, so hat das zugrunde liegende statistische Merkmal nur eine Realisationsmöglichkeit.

9) Eine Eisenbahn fährt 97 km mit einer Geschwindigkeit von 64 $^{km}/_h$ und weitere 97 km mit einer Geschwindigkeit von 192 $^{km}/_h$. Dann beträgt die Durchschnittsgeschwindigkeit für die Gesamtstrecke 96 $^{km}/_h$.

10) Ein statistisches Merkmal heißt stetig, wenn sein Wertebereich abzählbar unendlich ist.

Aufgabe B 9

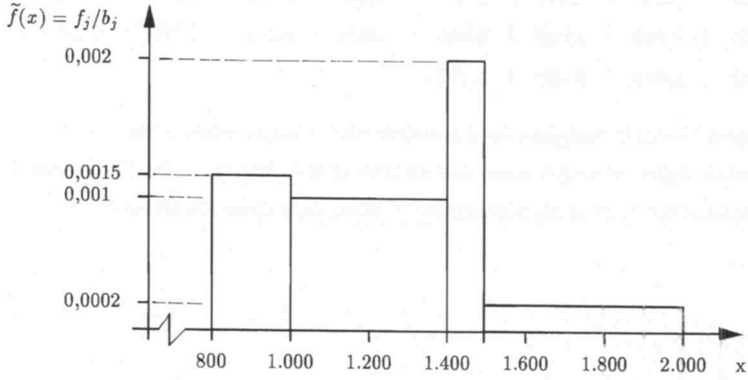
Bei 20 Haushalten wurden die monatlichen Bruttolöhne des Haushaltsvorstandes festgestellt. Es ergaben sich folgende Werte (in €) :

2.400	**3.300**	**2.400**	**2.300**	**2.500**	**1.700**	**2.100**	**3.100**
2.900	**1.800**	**3.900**	**2.100**	**2.900**	**1.500**	**2.700**	**2.200**
2.000	**2.500**	**2.600**	**2.200**				

1) Geben Sie den monatlichen Durchschnittsbruttoverdienst (arthm. Mittel) an !
2) Berechnen Sie die emp. Varianz und die Standardabweichung der 20 Bruttolöhne !
3) Ermitteln Sie den Variationskoeffizient ! Wozu dient dieser Koeffizient ?

Aufgabe B 1o:

In Metropolis wurden **4.000** Bewohner mit **5o m²** Wohnfläche nach ihrer Miethöhe befragt. Um eine bessere Übersicht zu erhalten, wurden die Mietbeträge in Klassen eingeteilt. Es ergab sich folgendes Histogramm:



1) Bestimmen Sie <u>rechnerisch</u> die Werte der empirischen Verteilungsfunktion an den Stellen

   a) x = 1.000

   b) x = 1.3oo

   c) x = 1.95o

2) Zeichnen Sie die empirische Verteilungsfunktion !

3) Berechnen Sie das arithmetische Mittel !

4) Bestimmen Sie <u>rechnerisch</u>

   a) $x_{0,25}$

   b) $\widetilde{x} = x_{0,5}$

   c) $x_{0,75}$

   und erstellen Sie mit Hilfe der "5–Zahlen–Zusammenfassung" ein Boxplot !

5) Berechnen Sie den empirischen Quartilskoeffizienten $g_{0,25}$ der Schiefe !

6) Welche Aussage bezüglich der Schiefe der vorliegenden Verteilung läßt sich aus

   a) dem errechneten Schiefekoeffizienten $g_{0,25}$

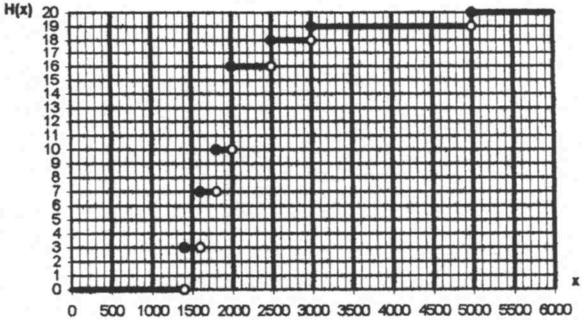
   b) der Lageregel (Vergleich von Median und arithmetischem Mittel)

   ableiten ?

7) Geben Sie <u>kurz</u> eine Erklärung für die unterschiedlichen Ergebnisse, die Sie unter 6) erhalten haben (sollten) !

Aufgabe B 11 :

Manni ist Chefredakteur für die Rubrik Fahrzeugtest der Automobilzeitschrift „Vollgas !“. In der Jubiläumsausgabe möchte er seiner Leserschaft eine statistische Auswertung über die Hubraumgröße [in cm³] der bisher getesteten PKWs präsentieren. Hierfür läßt er sich folgende (absolute) Verteilungsfunktion von seinem Praktikanten anfertigen :



1) Wie lautet das statistische Merkmal und wie ist es skaliert ?

2) Bestimmen Sie tabellarisch aus der Verteilungsfunktion die absoluten Häufigkeiten !

3) Berechnen Sie das arithmetische Mittel und die (empirische) Standardabweichung !
[**Rechenhilfe** : Die Summe der quadrierten Beobachtungswerte beträgt 96.340.000 cm⁶]

4) Transformieren Sie die unter 3) berechneten Parameterwerte in Liter (1 Liter ≅ 1 dm³) !

5) Manni möchte die Daten in drei Gruppen einteilen :

Mittelklasse	:	von 1400 cm³ bis unter 2000 cm³
obere Mittelklasse	:	von 2000 cm³ bis unter 3000 cm³
Oberklasse	:	von 3000 cm³ bis unter 6000 cm³

Bestimmen Sie tabellarisch die absoluten und die relativen Häufigkeiten der gruppierten Daten !

6) Berechnen Sie das arithmetische Mittel der gruppierten Daten !

7) Stellen Sie die alsolute Häufigkeitsverteilung der gruppierten Daten graphisch dar !

8) Bestimmen Sie tabellarisch die absoluten kumulierten und die relativen kumulierten Häufigkeiten der gruppierten Daten und stellen Sie die relative Verteilungsfunktion graphisch dar ! Welche Annahme über die Verteilung der Daten innerhalb der einzelnen Gruppen wird dabei getroffen ?

9) Ermitteln Sie <u>rechnerisch</u> das untere und obere Quartil sowie den Median aus den gruppierten Daten !

10) Erstellen Sie einen Box–Plot für die gruppierten Daten !

C

Konzentrationsanalyse

Aufgabe C 1:

Im heutigen Profifußball spielen die Finanzen eines Vereins inzwischen eine so große Rolle, daß eine Fußball-Liga als ein Wettbewerbsmarkt angesehen werden kann. Große Vereine werden finanziell immer mächtiger, während kleine Vereine aufgrund ihrer Schuldenberge um das Überleben kämpfen.

Die Untersuchung einer Fußball-Liga bezüglich des Gewinns der einzelnen Vereine nach Saisonende brachte folgendes Ergebnis:

Die **18** Mannschaften der **1.** Division wurden in folgenden Klassen aufgeteilt:

Gewinn in der letzten Saison (in Mio.) von ... bis unter	
0 – 2 2 – 10 10 – 12	Es ergab sich eine statistische Gleichverteilung !

Die **2o** Mannschaften der **2.** Division wurden in zwei Klassen aufgeteilt:

Gewinn in der letzten Saison (in Mio.) von ... bis unter	
0 – 2 2 – 4	Es ergab sich eine Verteilung im Verhältnis 3:1 !

1) Ermitteln Sie tabellarisch die Lorenzsche Konzentrationsverteilung für beide Divisionen und zeichnen Sie die zugehörigen Lorenzkurven !

2) Wie hoch ist der Anteil am gesamten Gewinnvolumen in der 2. Division, den die 5o% gewinnstärksten Vereine besitzen ?

3) Wieviel Pozent gewinnstarke Vereine der 1. Division besitzen 3o% des gesamten Gewinnvolumens in dieser Division ?

4) Welchen Anteil am gesamten Gewinnvolumen in der 2. Division haben diejenigen Vereine, die von ihrem finanziellen Erfolg her gesehen die "mittleren 5o%" ausmachen ?

5) Der Wert des Herfindahlmaßes für die 1. Division beträgt $H_{(1)} = \frac{1}{15}$.

   a) Berechnen Sie den Wert des Herfindahlmaßes für die 2. Division: $H_{(2)}$!

   b) Vergleichen Sie beide Divisionen bzgl. der absoluten Konzentration nach Herfindahl !

   c) In welcher Division ist die absolute Konzentration größer, wenn man nur den Merkmalseffekt berücksichtigt ?

Aufgabe C 2:

A: An einer Börse werden 200 Aktien gehandelt. Für das vergangene Jahr wurde der durchschnittliche Tagesumsatz jeder Aktie registriert. Diese Daten wurden in drei Gruppen zusammengefaßt:

Durchschnittlicher Tagesumsatz pro Aktie (in Mio. Taler) von ... bis unter
0 – 2
2 – 4
4 – 26

Auf die zweite und dritte Gruppe entfallen gleich viele Aktien; auch entfallen auf die erste Gruppe genauso viele Aktien wie auf die anderen beiden Gruppen zusammen.

1) Ermitteln Sie die Lorenzsche Konzentrationsverteilung und zeichnen Sie die Lorenzkurve !

2) Die Entwicklung des Kursniveaus an der Börse soll zukünftig in kompakter Form beschrieben werden. Dazu soll täglich ein Duchschnittskurs aus Aktienkursen gebildet werden (Aktienindex). Um den Rechenaufwand zu beschränken, sollen dabei nur diejenigen umsatzstärksten Aktien berücksichtigt werden, auf die 60% des Umsatzes entfallen. Wieviele Aktien sind das ?

B: Die Börse ist aufgeteilt in zwei Segmente: Amtlicher Handel (I) und Geregelter Markt (II). In jedem Segment gibt es mehrere Makler, auf die sich der Umsatz des jeweiligen Segments wie folgt verteilt:

Segment I	
Makler	Umsatzanteil
A	20%
B	40%
C	40%

Segment II	
Makler	Umsatzanteil
D	10%
E	20%
F	30%
G	40%

1) In welchem Segment herrscht eine höhere Konzentration nach Herfindahl ? Auf welchen Effekt ist dies zurückzuführen ? Begründung !

2) Man hatte es an der Börse schon längere Zeit erzählt — nun ist es soweit : Maklerin E und Makler F heiraten ! In Zukunft legen Sie ihre Geschäfte zusammen. Ermitteln Sie die Veränderung des Herfindahl–Maßes am Geregelten Markt !

Aufgabe C 3

Richtig oder falsch ?

1) Auf den Märkten A und B gebe es gleichviele Marktteilnehmer. Unterscheiden sich die Werte des Herfindahlindex beider Märkte, so liegt dies ausschließlich am Merkmalseffekt.

2) Fusionieren auf einem Markt zwei beliebige Anbieter mit den Marktanteilen p_1 und p_2, so wird der Herfindahlindex um den Faktor $2 \cdot (p_1 + p_2)$ größer.

3) Der Herfindalindex ist eine dimensionslose Maßzahl.

4) Die Konzentration nach Herfindahl verändert sich nicht, falls die Zahl der Anbieter an einem Markt konstant bleibt, die Marktanteile der Anbieter sich jedoch in Richtung „ gleicher Marktanteil für jeden " verschieben.

5) Hat für zwei Märkte der Ginikoeffizient den gleichen Wert, so ist die Lorenzsche Konzentration auf beiden Märkten inhaltlich als gleich zu bewerten.

6) Ist die Varianz einer statistischen Größe gleich Null, so ist der betreffende Markt im Lorenzschen Sinne konzentriert.

7) Auf einem Neuwagenmarkt herrsche eine Umsatzkonzentration mit einem Ginikoeffizienten gleich Null. Angenommen, es kommt ein neuer Anbieter dazu, dessen Umsatz derselbe ist wie der jedes anderen Anbieters. Dann ändert sich der Ginikoeffizient nicht, während der Wert des Herfindahlindex kleiner wird.

8) Auf den Märkten A und B gebe es gleichviele Marktteilnehmer. Unterscheiden sich die Herfindahlmaße der beiden Märkte in ihrem Wert, so liegt das nur am Anzahleffekt.

9) Auf einem Markt gebe es n Anbieter $(n > 1)$. Hat das Konzentrationsmaß nach Herfindahl den Wert $1/n$, so hat der Gini-Koeffizient den Wert Null.

10) Folgt das untersuchte statistische Merkmal einer statistischen Gleichverteilung, so ist im Rahmen einer Konzentrationsmessung nach Lorenz der betreffende Markt nicht konzentriert.

Aufgabe C 4

Die Diplom–Kauffrau Natalie Wood–Pecker hat ihre Liebe zur Tourismusbranche entdeckt und in Nordamerika $n_{NA} = 5$ Landschaftsparks ausgesucht, in denen sie büromüde Manager an Wochenenden dem Überlebenstraining aussetzen möchte. Ein besonderes Risiko stellen dabei die wilden Bären dar. Daher möchte sie sich mit Hilfe einer Lorenzschen Konzentrationsanalyse einen Überblick über die Bärenkonzentration in diesen Parks verschaffen. Von der Forstverwaltung erhält sie folgende Daten über die Bärenzahl :

Park Nr.	1	2	3	4	5
Bärenzahl	50	100	250	800	800

1) Wie heißt die hier untersuchte statistische Größe ? Was ist hier der Merkmalsträger ?

2) Erstellen Sie tabellarisch die Lorenzsche Konzentrationsverteilung für die Bärenkonzentration und zeichnen Sie die Lorenzkurve [10 cm x 10 cm] !

3) Wieviel Prozent des gesamten Bärenbestandes sind in den 60 % bärenärmsten Parks zu finden ?

Von einem Freund erhält Natalie folgende Lorenzsche Konzentrationsverteilung der Bärenkonzentration aus $n_{SK} = 10$ skandinavischen Parks, die einen gesamten Bärenbestand von 2.400 Bären haben :

u_ℓ	0,4	0,7	1,0
v_ℓ	0,2	0,5	1,0

4) Zeichnen Sie die zugehörige Lorenzkurve in die obige Graphik ein !
Ist die Konzentration nach Lorenz in Nordamerika oder in Skandinavien höher ?
Kann diese Frage aus der Graphik eindeutig beantwortet werden ? Begründung !

5) Berechnen Sie die Herfindahlindices für die Bärenkonzentration in Nordamerika (H_{NA}) und in Skandinavien (H_{SK}) !

6) Wo ist die Konzentration nach Herfindahl größer ?
Auf welchen Effekt (welche Effekte) ist dies zurückzuführen ?
Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

Aufgabe C 5

Die Unternehmen A und B sind Konkurrenten auf einem bestimmten Markt. Eine Untersuchung bezüglich des Umsatzes [in Mio €] für die einzelnen vom jeweiligen Unternehmen angebotenen Produkte ergab folgendes Ergebnis :

A

Umsatz in Mio € von...bis unter	h_j	f_j	u_ℓ	m_j	$m_j \cdot h_j$	$\frac{m_j \cdot h_j}{MV_A}$	v_ℓ
0 - ?	?	?	?	10	?	?	?
? - ?	?	0,5	1,0	90	?	?	1,0
	n_A = ?				MV_A = ?		

B

Umsatz in Mio € von...bis unter	h_j	f_j	u_ℓ	m_j	$m_j \cdot h_j$	$\frac{m_j \cdot h_j}{MV_B}$	v_ℓ
0 - ?	?	?	?	50	?	?	?
? - ?	?	0,1	1,0	450	?	?	1,0
	n_B = ?				$MV_B = 900$		

Variationskoeffizient $V_A = 0{,}8$

Herfindahlindex $H_A = \dfrac{41}{350}$

Ginikoeffizient $G_B = 0{,}4$

Marktvolumen $MV_B = 900$

1) Wie lautet die hier untersuchte statistische Einheit ?

2) Wie lautet das hier untersuchte statistische Merkmal ?

3) Berechnen Sie die absoluten Häufigkeiten

   a) für Unternehmen A $\left[\text{unter Zuhilfenahme von } V_A \text{ und } H_A \text{, wobei gilt : } H_A = \dfrac{V_A^2 + 1}{n_A}\right]$!

   b) für Unternehmen B [unter Zuhilfenahme von MV_B] !

4) Bestimmen Sie für beide Unternehmen tabellarisch die Lorenzsche Konzentrationsverteilung und zeichnen Sie die Lorenzkurven für beide Unternehmen in ein Diagramm !

5) Wie hoch ist der Anteil am Gesamtumsatz in Unternehmen A , den die 40 % umsatz**stärksten** Produkte bei Unternehmen A besitzen ?

6) Wie hoch ist der Anteil umsatz**starker** Produkte in Unternehmen B , die 60 % des gesamten Umsatzes bei Unternehmen B erwirtschaften ?

7) Geben Sie den Ginikoeffizienten für Unternehmen A an !

8) Vergleichen Sie die Lorenzsche Konzentration bei beiden Unternehmen.
   Erläutern Sie, warum der alleinige Vergleich der Ginikoeffizienten dafür nicht ausreichend ist, indem Sie kurz die inhaltlichen Sachverhalte beschreiben !

9) Berechnen Sie den Herfindahlindex für Unternehmen B !

10) Vergleichen Sie beide Unternehmen bezüglich der absoluten Konzentration nach Herfindahl !

11) Auf welchen (welche) Effekt (Effekte) ist dies zurückzuführen ?
    Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

Aufgabe C 6:

Zwei Märkte **A** und **B** sollen hinsichtlich des Umsatzes verglichen werden.

Markt A	
Unternehmen Nr.	Marktanteil
1	o,2
2	o,3
3	o,5

Markt B	
Unternehmen Nr.	Marktanteil
1	o,4
2	o,6

1) Zeichnen Sie die Lorenzkurven für beide Märkte in ein Diagramm !

2) Bestimmen Sie für beide Märkte den Herfindahl-Index !

3) Welcher Markt ist laut Herfindahl-Index stärker konzentriert, und auf welchem Effekt ist dies zurückzuführen ?
Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

4) Untenehmen Nr. **1** und Nr. **2** fusionieren auf **Markt A**. Wie groß ist die Konzentration auf Markt A nach dieser Fusion

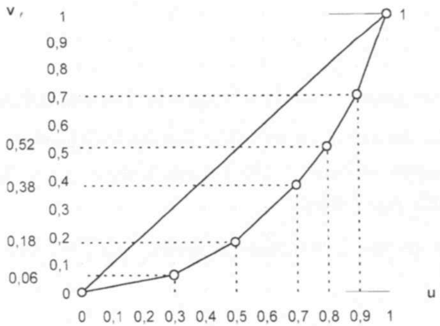
   a) nach dem Herfindahl-Index ?

   b) nach dem Gini-Koeffizienten ?

5) Wie ist der Umsatz auf Markt A nach der Fusion (statistisch) verteilt und wie groß ist die empirische Varianz des Umsatzes ?

Aufgabe C 7

In einem Bundesland hat die Polizei in 100 Landkreisen die „ Anzahl der Verkehrsdelikte im letzten Jahr " ermittelt und – bei Null beginnend – in Gruppen aufgeteilt. Außerdem wurde berechnet, daß in jedem Landkreis durchschnittlich (arithm. Mittel) 100 Verkehrsdelikte registriert wurden. Aus den Daten wurde zuletzt folgende Graphik erstellt :



1) Um was für eine Darstellung handelt es sich hier ?
2) Ermitteln Sie tabellarisch die absolute und die relative Häufigkeitsverteilung !
3) Stellen Sie die relative Verteilungsfunktion graphisch dar !
4) Wie viele Verkehrsdelikte mindestens hatten die 30 % Landkreise mit den meisten Verkehrsdelikten ?
5) Wieviel Prozent am gesamten Verkehrsdeliktvolumen hatten die 30 % Landkreise mit den meisten Verkehrsdelikten ?

6) Berechnen Sie die (emp.) Varianz !
7) Berechnen Sie den Variationskoeffizienten
8) Berechnen Sie den (emp.) Quartilskoeffizienten der Schiefe !
9) Beurteilen Sie die Schiefe der vorliegenden Häufigkeitsverteilung
   a) nach der „ Lageregel " !
   b) nach dem (emp.) Quartilskoeffizienten der Schiefe !

Aufgabe C 8:

A: Das Herfindahlsche Konzentrationsmaß für die Kapitalkonzentration in einem Land ergab $H = 0,02$. Zwei Unternehmen, deren Anteil am Gesamtkapital 0,1% und 1,5% betrug, fusionieren. Geben Sie an, welchen Wert H nach diesem Unternehmenszusammenschluß hat !

B:
1) Welche Effekte gehen in das Herfindahlsche Konzentrationsmaß ein ?
2) Die Konzentrationsmessung mit Hilfe des Herfindahlschen Maßes für die Elektrobranche ergab in Land A (80 Unternehmen): $H_A = 0,6$, in Land B (100 Unternehmen): $H_B = 0,48$.

In welchem Land ist die Konzentration größer, und auf welchen Effekt ist dies zurückzuführen ?

C: Im folgenden sind die Anteile am Nominalkapital von zwei Gesellschaften wiedergegeben:

I) **Phönix Gummi**

Großaktionäre

Firestone	25%
Schulze Erben	25%
Münchner Rück	25%
Sonstige	25%

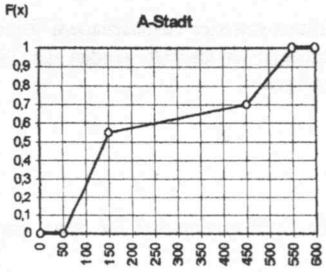
II) **Duisburger Kupferhütte**

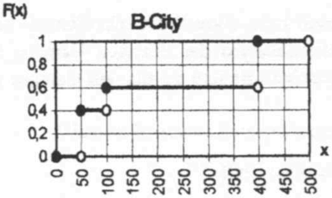
Großaktionäre

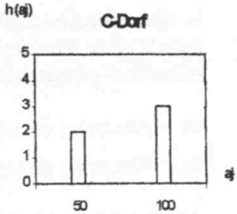
BASF	$33\frac{1}{3}$ %
Bayer	$33\frac{1}{3}$ %
Hoechst	$33\frac{1}{3}$ %

1) Bestimmen Sie für beide Gesellschaften den Gini-Koeffizienten !
2) Berechnen Sie für beide Gesellschaften das Herfindahlsche Konzentrationsmaß H !
3) Die Münchner Rück verkauft ihren Anteil an Firestone.
   Berechnen Sie H für diese neue Situation !

Aufgabe C 9 :







In A−Stadt, B−City und C−Dorf wurden die Betreiber von Imbißbuden nach ihren durchschnittlichen Tagesumsätzen in DM befragt.

Aus den Angaben der Befragten in A−Stadt ($n_A = 20$) , in B−City ($n_B = 10$) und C−Dorf ($n_C = 5$) erhielt man die oben aufgeführten graphischen Darstellungen.

A :

1) Ermitteln Sie tabellarisch die Lorenzsche Konzentrationsverteilung für
   a) A−Stadt
   b) B−City
   c) C−Dorf

2) Zeichnen Sie die zugehörigen Lorenzkurven in **ein** Diagramm !
   In welchem Ort herrscht die größte Konzentration nach Lorenz ?

3) Berechnen Sie den Gini−Koeffizienten für die Konzentrationsverteilung von A−Stadt !

B :

1) Berechnen Sie den Herfindahlindex für
   a) B−City
   b) C−Dorf

2) In welchem Ort ist die absolute Konzentration nach Herfindahl höher ?
   Auf welchen Effekt (welche Effekte) ist dies zurückzuführen ? Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

Aufgabe C 10 :

In Klein–Kleckersdorf kauft jeder Einwohner sein Gemüse bei einem von **vier** Gemüsebauern. Zum Zweck einer Konzentrationsanalyse der dortigen Gemüse kaufenden Kundschaft wurden die 28 Einwohner gefragt, bei welchem Bauern sie sich mit Gemüse versorgen.

1) Wie lautet das untersuchte statistische Merkmal ?

2) Wie lautet die statistische Einheit ?

Angenommen, auf dem Kleckersdorfer Gemüsemarkt herrscht Nullkonzentration im Sinne der Lorenzschen Konzentrationsanalyse.

3) Was für eine Verteilung des statistischen Merkmals liegt dann vor ?

4) Wieviele Kunden hat dann jeder der vier Bauern ?

5) Welchen Wert hat der Ginikoeffizient G ?

Angenommen, auf dem Kleckersdorfer Gemüsemarkt herrscht eine größtmögliche Konzentration im Sinne der Lorenzschen Konzentrationsanalyse.

6) Skizzieren Sie die Lorenzkurve !

7) Welchen Wert hat nun der Ginikoeffizient G ?

Für den Klein–Kleckersdorfer Gemüsemarkt ergibt sich ein Herfindahlindex von $H_K = 0{,}41$.
Im Nachbardorf Poppelshausen wurden die 33 Einwohner ebenfalls gefragt, bei welchem der **drei** dort ansässigen Bauern sie ihr Gemüse kaufen. Die Summe der <u>quadrierten</u> Kundenzahl jedes Bauern ergibt den Wert 509 $[= \sum_{i=1}^{3} x_i^2]$.

8) Berechnen Sie den Herfindahlindex H_P für den Gemüsemarkt von Poppelshausen !

9) Auf welchem Gemüsemarkt herrscht die größere Konzentration nach Herfindahl, und auf welchen Effekt (welche Effekte) ist dies zurückzuführen ?
Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

Aufgabe C 11

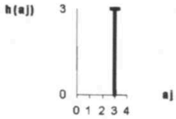
A

Der Markt für Speiseeis in Norddeutschland (ND) bzw. Süddeutschland (SD) sei in einer bestimmten Periode unter den Speiseeisanbietern bzgl. des Merkmals

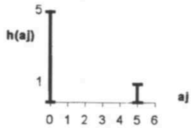
X : „ Verkaufszahlen [in Tsd. Stück] “

folgendermaßen verteilt :

Norddeutschland



Süddeutschland



1) Skizzieren Sie die Lorenzkurven für beide Märkte !

2) Bestimmen Sie die Werte der Ginikoeffizienten G_{ND} und G_{SD} !

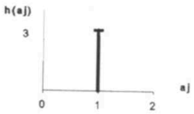
B

Der Speiseeismarkt in Norditalien (NI) bzw. Süditalien (SI) ist in einer bestimmten Periode unter den Speiseeisanbietern bzgl. des Merkmals

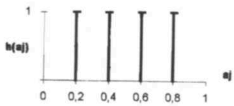
X : „ Verkaufszahlen [in Tsd. Stück] “

folgendermaßen verteilt :

Norditalien



Süditalien



1) Berechnen Sie die Werte der Herfindahlindices H_{NI} und H_{SI} ! Wo ist die Konzentration höher ?

2) Auf welchen Effekt (welche Effekte) ist dies zurückzuführen ?
Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

C

Über zwei weitere Märkte (A und B) sei folgendes bekannt :

- In B herrscht eine Einpunktverteilung bzgl. des betrachteten Merkmals.
- In B gibt es nur halb so viele Anbieter wie in A .
- In A ist die Standardabweichung des betrachteten Merkmals doppelt so groß wie das arithmetische Mittel.

1) Bestimmen Sie (in Abhängigkeit von $\mathbf{n_B}$) die Herfindahlindices H_A und H_B !

2) Auf welchen Effekt (welche Effekte) ist dies zurückzuführen ?
Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

D

Mehrdimensionale Daten

Aufgabe D 1

Bei einer Untersuchung wurden 200 Vater / Sohn – Paare nach ihrem Wahlverhalten bei der letzten Wahl befragt.

Erstellen Sie mit Hilfe der folgenden Angaben eine Kontingenztafel :

- die Söhne entschieden sich für die zur Wahl stehenden Parteien **A** , **B** und **C** im Verhältnis 2 : 1 : 2 .
- die Hälfte der Väter, die Partei B wählten, muß der Tatsache ins Auge sehen, daß ihre Söhne die Partei A bevorzugen.
- 5 % der befragten Söhne entschieden sich – wie ihre Väter – für die Partei A .
- die Partei B konnte 40 % aller Stimmen auf sich vereinigen.
- 15 % aller Väter gaben der Partei C ihre Stimme.
- 20 Söhne, deren Väter die Partei A wählten, entschieden sich für die Partei B .
- nur 30 Väter wählten dieselbe Partei wie ihre Söhne.

Aufgabe D 2:

Die Marketing–Abteilung einer Firma hat im Zusammenhang mit der Entwicklung einer Marketing–Strategie für ihr neuestes Produkt eine Untersuchung in der Hoffnung durchgeführt, Hinweise über die erfolgversprechenste Verpackungsfarbe zu erhalten. Hierzu wurde von dem Produkt insgesamt 6oo Stück auf drei verschiedenen Testmärkten angeboten.

Mehrere Mitarbeiter haben bisher folgende Informationen über die auf den Testmärkten verkauften Stücke zusammengetragen:

- die auf dem Testmarkt I verkaufte Stückzahl verhielt sich bezüglich ihrer Verpackungsfarbe (schwarz, rot, weiß) im Verhältnis 1 : 3 : 1 .
- 3o% der auf dem Testmarkt II verkauften Stücke waren schwarz verpackt.
- 25% aller angebotenen Stücke wurden nicht verkauft.
- es wurden ebenso viele schwarz verpackte wie weiß verpackte Stücke verkauft.
- auf dem Testmarkt III wurden 175 Stücke verkauft.
- Die auf dem Testmarkt II verkaufte Stückzahl macht 75% der auf allen Testmärkten insgesamt verkauften rot verpackten Stücke aus.
- auf dem Testmarkt I wurden 5o Stück weniger als auf dem Testmarkt III verkauft.
- 35% der verkauften *rot* verpackten Stücke wurden auf dem Testmarkt III verkauft.

1) Stellen Sie die aus der Untersuchung erhaltenen Daten in einer Kontingenztafel tabellarisch dar !

2) Ermitteln Sie die Verkaufsanteile der einzelnen Verpackungsfarben auf den verschiedenen Testmärkten !

3) Für welche Verpackungsfarbe wird sich die Marketing–Abteilung aufgrund der unter 2) erhaltenen Ergebnisse voraussichtlich entscheiden ? Begründung !

Aufgabe D 3:

Es wurde eine Umfrage durchgeführt, bei der die befragten Personen Auskunft über ihre Tierliebe (ja/nein) geben sollten. Die Befragten wurden in drei Kategorien aufgeteilt (Kinder, Jugendliche, Erwachsene).

Die Umfrage brachte folgende Ergebnisse:

1. Von den 4oo befragten Kindern sind 8o% tierlieb.
2. Es wurden 2oo Erwachsene befragt.
3. Die Zahl der tierliebenden Erwachsenen war um 14o geringer als die Zahl der tierliebenden Kinder.
4. insgesamt waren 8o% (= 8oo Personen) tierlieb.

1) Wie hoch ist die Zahl der nicht tierliebenden Erwachsenen ?

2) Wieviel Prozent der Befragten sind nicht erwachsen ?

3) Wie groß ist der Anteil der tierliebenden Jugendlichen unter den Befragten ?

4) Wie groß ist der Anteil der Jugendlichen unter den Tierfreunden ?

5) Haben die Kinder mehr Tierliebe als die Erwachsenen ?
Begründung !

Aufgabe D 4:

Bei einer Untersuchung des Gesundheitsamtes in B–Stadt ergab sich folgendes Bild über Infektionen mit dem Erreger Z in den Risikogruppen A, B und C:

		b_j	
		infiziert	nicht infiziert
	A	18	42
a_i	B	5	31
	C	17	37

1) Ist für einen beliebigen Angehörigen einer dieser Risikogruppen die Häufigkeit $f(a_i/b_j)$ interessanter als die Häufigkeit $f(b_j/a_i)$?
   Kurze Begründung !

2) Ist der Infektionsgrad für die Gruppe B in dieser Untersuchung am kleinsten ?

3) Ist die Aussage richtig, daß bei der Gruppe B die wenigsten "Nichtinfizierten" gefunden wurden ?

4) Ist die Aussage richtig, daß bei der Gruppe C der Anteil der "Nichtinfizierten" am kleinsten ist ?

Aufgabe D 5

A

Der Obstzüchterverein „ Birnwiese " möchte nach intensiver Forschungsarbeit zwei neue Apfelsorten „Grüner Bitterer" [G] und „Kleiner Saurer" [K] auf den Markt bringen. Zum Testen ihrer Marktchancen liefert der Verein die neuen Sorten zusammen mit den zwei schon etablierten Sorten „Edler Gelber" [E] und „Rotbäckchen" [R] an drei Großküchen (eine Schulküche { **S** }, eine Mensa { **M** } und eine Betriebskantine { **B** }) in einer Gesamtmenge **aller** Apfelsorten von 200 kg.

- ☐ { **S** } erhält von Sorte [R] 10 kg .
- ☐ Von [E] werden insgesamt 70 kg geliefert .
- ☐ Insgesamt werden von [G] 30 kg geliefert .
- ☐ Die gelieferten Mengen an { **S** } und { **M** } zusammen ergeben die gelieferte Menge an { **B** } .
- ☐ Die von [R] an { **M** } gelieferte Menge macht 10 % der gesamten an { **B** } gelieferten Menge aus .
- ☐ Von [E] wird an { **S** } doppelt soviel geliefert wie von [E] an { **M** } .
- ☐ { **B** } erhält viermal so viel von [E] wie { **M** } von [E] erhält.
- ☐ An { **M** } wird insgesamt doppelt so viel geliefert wie von [E] an { **S** } .
- ☐ Von [G] wird an { **S** } viermal soviel geliefert wie von [G] jeweils an { **M** } und { **B** } geliefert wird.
- ☐ Von [G] wird an { **S** } halb soviel geliefert wie von [R] an { **B** } .

1) Wie heißen die hier untersuchten statistischen Merkmale und wie sind sie skaliert ?

2) Stellen Sie die **gemeinsame** Häufigkeitsverteilung dieser beiden Merkmale tabellarisch dar !

3) Sind die beiden Merkmale bzgl. der vorliegenden Daten unabhängig voneinander ?
Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

4) Welche der drei Großküchen bekommt den kleinsten Anteil neuer Apfelsorten bezogen auf die an sie jeweils gelieferte Gesamtmenge an Äpfeln ?
Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

B

In den beiden neuen Apfelsorten sollen nach Angaben des Obstzüchtervereins kaum Maden vorhanden sein. Der Mensakoch mißtraut diesen Versprechungen und untersucht die Hälfte der an ihn gelieferten Äpfel der **neuen** Sorten auf Maden (6 Äpfel $\cong$ 1 kg). Dabei erhält er folgendes Ergebnis :

Zahl der Maden pro Apfel	rel. Häufigkeit
0	0,2
1	0,4
2	0,3
3	0,1

1) Wie heißt das hier untersuchte statistische Merkmal und wie ist es skaliert ?

2) Bestimmen Sie tabellarisch die absoluten Häufigkeiten sowie die absoluten und die relativen kumulierten Häufigkeiten des statistischen Merkmals !

3) Stellen Sie die (relative) Verteilungsfunktion graphisch dar !

4) Ermitteln Sie die Werte einer sog. „ 5 - Zahlen - Zusammenfassung " und zeichnen Sie den zugehörigen Box - Plot !

Aufgabe D 6

A

Der Sauberkeitsstandard in Jugendherbergen läßt leider oftmals sehr zu wünschen übrig. Eine Gruppe von Statistik-Studenten möchte sich daher im Rahmen einer Studie in vier verschiedenen Regionen (**A , B , C und D**) des Landes ein Bild machen vom Sauberkeitsstandard (**gut , mäßig , schlecht**) einiger dort ansässiger Jugendherbergen. Insgesamt wurden **40** Jugendherbergen mit folgendem Ergebnis überprüft:

- ☐ In Region **D** gibt es 2 Herbergen mit **mäßigem** Standard.
- ☐ Insgesamt wurden 14 Herbergen in Region **C** und 6 Herbergen in Region **A** überprüft.
- ☐ In 12 überprüften Herbergen ist der Standard **mäßig** .
- ☐ In Region **C** haben 8 Herbergen einen **guten** Standard.
- ☐ In Region **C** gibt es doppelt so viele Herbergen mit **mäßigem** wie mit **schlechtem** Standard.
- ☐ Die Herbergen in Region **C** mit **schlechtem** Standard machen 10 % der Herbergen mit **gutem** Standard aller Regionen aus.
- ☐ In der Region **A** gibt es viermal so viele Herbergen mit **mäßigem** Standard wie jeweils Herbergen mit **gutem** bzw. **schlechtem** Standard.
- ☐ Die Zahl der Herbergen mit **mäßigem** Standard in Region **A** ist halb so hoch wie die Zahl der Herbergen mit **gutem** Standard in Region **D** .
- ☐ Es gibt in allen Regionen zusammen viermal so viele Herbergen mit **mäßigem** Standard wie es Herbergen mit **schlechtem** Standard in Region **B** gibt.

1) Wie heißen die hier untersuchten statistischen Merkmale und wie sind sie skaliert ?
2) Stellen Sie die **gemeinsame** Häufigkeitsverteilung dieser beiden Merkmale tabellarisch dar !
3) Sind die beiden Merkmale bzgl. der vorliegenden Daten unabhängig voneinander ?
   Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !
4) In welcher Region ist die Sauberkeit (von Herbergen mit <u>mindestens mäßigem</u> Standard) am größten ?
   Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

B

Nach Bekanntgabe der Ergebnisse dieser Studie wurde von allen Verantwortlichen derjenigen Herbergen, denen ein **schlechter** Sauberkeitsstandard bescheinigt wurde, einhellig versichert, daß man schnellstens für eine gründliche Beseitigung der beklagten Mängel sorgen werde. Nach einiger Zeit wurden daraufhin jeweils 20 Zimmer in <u>jeder</u> der betroffenen Herbergen noch einmal auf ihren Sauberkeitsstandard überprüft :

Zahl der Sauberkeitsmängel pro Zimmer von ... bis unter	rel. Häufigkeit
0 - 2	0,20
2 - 5	0,45
5 - 10	0,25
10 - 12	0,10

1) Wie heißt das hier untersuchte statistische Merkmal und wie ist es skaliert ?
2) Bestimmen Sie tabellarisch die absoluten Häufigkeiten sowie die absoluten und die relativen kumulierten Häufigkeiten des statistischen Merkmals !
3) Stellen Sie die (relative) Verteilungsfunktion graphisch dar !
4) Ermitteln Sie **<u>rechnerisch</u>** die Werte einer sog. „ 5 - Zahlen - Zusammenfassung “ und zeichnen Sie den zugehörigen Box - Plot !

Aufgabe D 7

Bei einer bevorstehenden Wahl streiten neben den etablierten Parteien „ Weiß " [**W**] und „ Lila " [**L**] auch die beiden neuen Parteien „ Blau " [**B**] und „ Türkis " [**T**] um die Gunst der Wähler. Von politischer Seite ist man vor allem daran interessiert, ob sich die sozialen Verhältnisse von Wählern auf das Wahlverhalten auswirkt.
Aus diesem Grund wurden **300** Erstwähler befragt, ob sie sich eher zur „ Unterschicht " { **U** } , zur „ Mittelschicht " { **M** } oder zur „ Oberschicht " { **O** } zugehörig fühlen und welche Partei sie wählen werden.
Das Ergebnis dieser Befragung lautete :

- ☐ 20 % aller Befragten werden [**W**] wählen.
- ☐ Partei [**B**] wird insgesamt 70 Stimmen erhalten.
- ☐ Von allen Befragten fühlten sich 135 zu { **M** } zugehörig.
- ☐ Innerhalb von { **M** } bekam die Partei [**T**] doppelt so viele Stimmen wie alle drei anderen Parteien zusammen.
- ☐ Von allen Wählern aus { **M** } , die <u>nicht</u> [**T**] wählen, werden genau ein Drittel der Partei [**L**] ihre Stimme geben.
- ☐ Es gab genau so viele [**T**] – Wähler aus { **M** } wie es insgesamt Wähler aus { **O** } gab.
- ☐ 40 % aller Wähler aus { **U** } werden Partei [**T**] wählen.
- ☐ Die Partei [**L**] bekommt aus { **U** } und { **M** } zusammen genau soviel Stimmen wie aus { **O** } .
- ☐ 12,5 % aller [**L**] – Wähler stammen aus { **U** } .
- ☐ Die Partei [**W**] bekam aus { **U** } genau soviel Stimmen wie die Partei [**T**] aus { **O** } bekam.
- ☐ Es gab viermal soviel [**W**] – Wähler aus { **O** } wie es [**W**] – Wähler aus { **U** } gab.

1) Wie heißen die hier untersuchten statistischen Merkmale und wie sind sie skaliert ?

2) Stellen Sie die **gemeinsame** Häufigkeitsverteilung dieser beiden Merkmale tabellarisch dar !

3) Sind die beiden Merkmale bzgl. der vorliegenden Daten unabhängig voneinander ?
Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

4) Welche der drei sozialen Schichten hat den größten Wähleranteil, der für die <u>neuen</u> Parteien votiert hat ?
Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

In einer weiteren Untersuchung wurden diejenigen Jungwähler, die die neuen Parteien wählen werden, nach ihren monatlichen Einlommensverhältnissen in DM befragt.

Dabei erhielt man folgendes Ergebnis für die vier aufgeführten Gruppen :

Gruppe	von ... bis unter
1	500 – 1000
2	1000 – 1600
3	1600 – 2400
4	2400 – 3000

- 70 % aus der Unterschicht und 30 % aus der Mittelschicht gehören zur 1. Gruppe.
- 5% aus der Unterschicht, 10 % aus der Mittelschicht und 90 % aus der Oberschicht gehören zur 4. Gruppe.
- Genau 50 Befragte entfallen auf die 2. Gruppe.

5) Wie heißt das hier untersuchte statistische Merkmal und wie ist es skaliert ?

6) Bestimmen Sie tabellarisch die absoluten Häufigkeiten, die relativen und die relativen kumulierten Häufigkeiten des statistischen Merkmals !

7) Stellen Sie die (relative) Verteilungsfunktion graphisch dar !

8) Ermitteln Sie <u>rechnerisch</u> das untere Quartil ($x_{0,25}$), den Median ($x_{0,5}$) und das obere Quartil ($x_{0,75}$) !

E

Korrelations– und Regressionsanalyse

Aufgabe E 1:

Eine Gruppe von Studenten der Naturwissenschaften mußte zu schriftlichen Prüfungen im Form von Klausuren je eine Aufgabe in Physik und Mathematik lösen. Von 1o zufällig ausgewählten Studenten sind die Ergebnisse (in Punkten) in folgender Tabelle zuammengefaßt:

Student	1	2	3	4	5	6	7	8	9	1o
Mathematik	14	8	1o	2o	12	8	12	2o	16	1o
Physik	1o	12	12	8	12	1o	1o	12	16	8

A: Berechnen Sie den Korrelationskoeffizienten nach Bravais-Pearson !

B: Berechnen Sie den Rangkorrelationskoeffizienten nach Kendall !

C: Jede der zwei Kurzklausuren gilt als bestanden, wenn der/die KandidatIn mindestens **12** Punkte erreicht hat.

1) Stellen Sie, um den Zusammenhang zwischen dem Klausurergebnis in Physik (bestanden, nicht bestanden) und dem Klausurergebnis in Mathematik (bestanden, nicht bestanden) zu messen, das Klausurergebnis der 1o StudentInnen in einer Kontingenztafel dar !
2) Geben Sie den χ^2–Wert an !
3) Ermitteln Sie den Kontingenzkoeffizienten K !

D: Geben Sie eine Erklärung für die unterschiedlichen Werte von r, τ und K !

Aufgabe E 2:

15 Lehrlinge, die im Rahmen ihres Berufsschulunterrichts an einer Unterrichtsreihe "Sexualkunde" teilnehmen, wurden aufgefordert, auf der Fünf–Punkte–Skala

1 — 2 — 3 — 4 — 5
○ — ○ — ○ — ○ — ○
unwichtig — sehr wichtig

die Wichtigkeit anzukreuzen, die ihrer Meinung nach der Sexualerziehung generell zukommen müßte. Außerdem sollten die befragten Lehrlinge auf einer Drei–Punkte–Skala

1 — 2 — 3
○ — ○ — ○
gar nichts — viel

ankreuzen, in welchem Ausmaß das im Unterricht verwendete Aufklärungsbuch zu einer vernünftigen Sexualerziehung beitragen kann. Das Ergebnis ist in folgender Häufigkeitstabelle erfaßt:

		Einschätzung der Wichtigkeit				
		1	2	3	4	5
Einschätzung des Buches	1	-	4	1	-	-
	2	-	-	5	1	-
	3	2	-	1	-	1

Berechnen Sie, um den Grad des Zusammenhangs der beiden Einschätzungen anzugeben, den Rangkorrelationskoeffizienten nach Kendall !

Aufgabe E 3:

A: Bei 9 ExamenskandidatInnen wurden die Merkmale "Schrifliche Noten im Fach VWL" und "Mündliche Noten im Fach VWL" erhoben, wobei nur ganze Noten von 1.o bis 5.o vergebeñ wurden. Das Ergebnis der Erhebung ist in folgener Matrix dargestellt:

		Mündliche Noten				
		1.o	2.o	3.o	4.o	5.o
schriftliche Noten	1.o	1	-	-	-	-
	2.o	1	1	1	-	-
	3.o	1	-	-	1	1
	4.o	-	-	-	1	-
	5.o	-	-	-	1	-

P20
Q4
$\tau^{a} = 0.533$

Berechnen Sie einen geeigneten Koeffizienten als Maß für die Stärke des Zusammenhangs beider Merkmale !

B: Bei 1o ExamenskandidatInnen wurden die Merkmale "Vorbereitungszeit (in Wochen)" und "Anzahl der abgelegten Prüfungen" erhoben. Das Ergebnis der Erhebung ist in folgender Matrix dargestell:

		Vorbereitungszeit				
		1	2	3	4	5
Anzahl der Prüfungen	1	-	-	-	1	1
	2	2	1	-	-	-
	3	-	-	-	-	-
	4	2	1	-	-	-
	5	-	-	-	1	1

$r = 0$

Berechnen Sie einen geeigneten Koeffizienten als Maß für die Stärke des linearen Zusammenhangs beider Merkmale !

Aufgabe E 4:

An der Universität A sind die für die Statistik I bzw. II – Klausur – Verantwortlichen daran interessiert, mittels eines Koeffizienten zu schätzen, wie stark der formale Zusammenhang zwischen den beiden Klausurergebnissen (gemessen in Punktwerten) ist.

Zu diesem Zweck werden aus der Gesamtzahl der StudentInnen, die diese Klausuren geschrieben haben, 5 zufällig ausgewählt. In folgender Tabelle sind ihre erarbeiteten Punktwerte aufgeführt:

StudentIn	1	2	3	4	5
Stat.I	20	30	50	70	80
Stat.II	70	40	10	40	70

1) Zeichnen Sie die Daten in ein Streudiagramm !
2) Berechnen Sie den für dieses Problem adäquaten Koeffizienten !
3) Interpretieren Sie das Ergebnis !
4) Was mißt dieser Koeffizient und was mißt er nicht ?

Weisen Sie den (kardinalen) Daten Rangziffern zu.

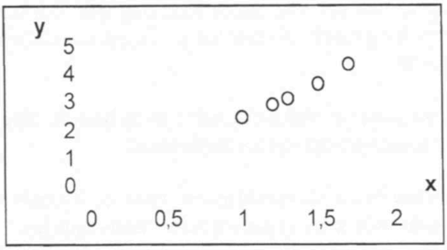
1) Berechnen Sie den adäquaten Koeffizienten !
2) Interpretieren Sie das Ergebnis !

Aufgabe E 5

Ein Student soll im Rahmen seiner Diplomarbeit für ein Unternehmen eine Regressionsanalyse hinsichtlich des Umsatzes der 10 Firmenprodukte in Abhängigkeit von den Werbekosten durchführen. Zu diesem Zweck erhält er folgendes Datenmaterial :

Werbekosten [in Mio €]	1	1,2	1,3	1,5	1,7
Umsatz [in Mio €]	2,5	2,94	3,19	3,75	4,39

Es ergibt sich folgender Datenplot :

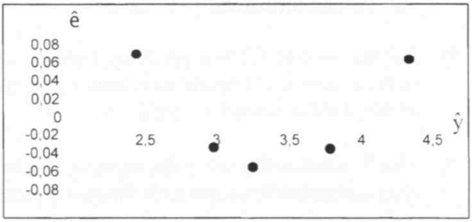


Der Student entscheidet sich aufgrund des Datenplots für eine lineare Einfachregressionsanalyse.
Diese Analyse liefert für das Bestimmtheitsmaß den Wert :

$$R^2 = 0.9936$$

1) Kann sich der Student durch den hohen Wert des Bestimmtheitsmaßes in der Wahl seines linearen Regressionsansatzes bestätigt fühlen ? Begründen Sie <u>kurz</u> Ihre Antwort !

Der betreuende Professor empfiehlt dem Studenten, sich den Residuenplot genau anzusehen :



2) Worüber gibt ein Residuenplot Aufschluß ? Interpretieren Sie den nebenstehenden Residuenplot !

Der Professor schlägt folgenden <u>nicht-linearen</u> Regressionsansatz vor :

$$\boxed{\hat{y} = \hat{c} + \hat{d} \cdot z \quad \text{mit } \mathbf{z} = \mathbf{x}^2}$$

3) Bestimmen Sie für die transformierten Daten (z , y) die Regressionsparameter $\hat{c}$ und $\hat{d}$ nach der Methode der kleinsten Quadrate !

<u>Hilfe</u> :

$$\bar{z} = 1{,}854 \; ; \; \bar{y} = 3{,}354 \; ; \; \sum_{i=1}^{5} z_i \cdot y_i = 33{,}2493 \; ; \; \sum_{i=1}^{5} z_i^2 = 19{,}3443 \; ; \; \bar{z}^2 = 3{,}437316$$

Als Schätzwerte $\hat{y}_i$ erhält man :

$\hat{y}_i$	2,5	2,94	3,19	3,75	4,39

4) Geben Sie den Wert des Bestimmtheitsmaßes R^2 an !

5) Skizzieren Sie für die transformierten Daten den Residuenplot !

Aufgabe E 6

Richtig oder falsch ?

1) $r = 0{,}8$ bedeutet, daß in 80 von 100 Fällen die Werte der zwei betrachteten Merkmale übereinstimmen.

2) r ist eine dimensionslose Maßzahl.

3) Innerhalb der Regressionsrechnung läßt sich mit Hilfe des Bestimmtheitsmaßes untersuchen, ob der gewählte Ansatz für die Regressionsfunktion (z.B. linear) als erfüllt angesehen werden kann.

4) Der Korrelationskoeffizient r nach Bravais–Pearson verändert nicht seinen Wert, falls man die Ausgangsdaten linear transformiert.

5) Unter einem standardisierten Datensatz versteht man einen Datensatz, den man einer speziellen nicht – linearen Transformation unterzogen hat.

6) Der Korrelationskoeffizient nach Kendall ist eine Maßzahl für die Stärke des linearen Zusammenhangs der betrachteten Rangreihen.

7) Der Wert des Korrelationskoeffizienten r nach Bravais–Pearson kann nicht gleich dem Wert des Bestimmtheitsmaßes R^2 sein.

8) Bei der linearen Einfachregression bedeutet ein Wert des Bestimmtheitsmaßes $R^2 = 0$, daß Null Prozent der Gesamtabweichung (in y–Richtung) durch die lineare Regressionsfunktion erklärt werden kann, d.h. es gilt : $\hat{y} = \bar{y}$.

9) Eine Lineartransformation des einen und / oder des anderen Merkmals verändert den Wert des Korrelationskoeffizienten nach Bravais–Pearson.

10) Bei der linearen Einfachregressionsanalyse verläuft eine nach der Methode der kleinsten Quadrate bestimmte Regressionsgerade ($\hat{y} = \hat{a} + \hat{b} \cdot x$) stets durch die Punkte ($\bar{x}$, $\bar{y}$) und (0 , $\hat{b}$) .

Aufgabe E 7:

Um die Abhängigkeit des Weizenertrages von der Düngemittelmenge zu schätzen, will Jungbauer Ferdinand eine lineare Einfachregressionsanalyse durchführen lassen. Für die statistische Untersuchung stellt er fünf gleichgroße Ackerbauflächen zur Verfügung. Die Untersuchung brachte folgendes Ergebnis:

Düngemittel / Fläche in kg	1	3	5	9	12
Ertrag / Fläche in kg	1	5	5	3	1

1) Schätzen Sie die Parameter a und b nach der Methode der kleinsten Quadrate und zeichnen Sie die Beobachtungswerte und die Regressionsgerade in ein Koordinatensystem ein !

2) Zeigen Sie mit Hilfe eines mathematischen Beweises, welche Beziehung zwischen den Werten $\bar{\hat{y}}$ und $\bar{y}$ besteht !

3) a) Bestimmen Sie den geschätzten Weizenertrag, wenn Ferdinand 4 kg Düngemittel verwenden würde !

   b) Wie hoch ist die geschätzte Zu- bzw. Abnahme des Weizenertrages in g, falls die Düngemittelmenge um 1 kg erhöht wird ?

   c) Welcher geschätzte Weizenertrag würde sich bei einem Düngemitteleinsatz von 4o kg ergeben ?
   Ist diese Schätzung sinnvoll ? Begründung !

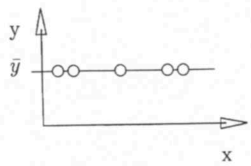
4) Berechnen und interpretieren Sie das Bestimmtheitsmaß R^2 !
   Rechenhilfe: $\sum_i \hat{y}_i^2 = 46{,}25$!

5) Welche Beziehung besteht zwischen den Größen: y_i ; $\hat{y}_i$; $\bar{y}$, wenn gilt:

   a) $R^2 = 0$

   b) $R^2 = 1$?

6) Das Ergebnis einer linearen Einfachregressionsanalyse ist in folgender Abbildung dargestellt:



   a) Welchen Wert hat in diesem Fall R^2 ?

   b) Nehmen Sie zu der inhaltlichen Bedeutung dieses Falls Stellung !

7) Erläutern Sie kurz den Unterschied in der Datenerfassung bei der Regressionsanalyse und bei der Korrelationsanalyse !

Aufgabe E 8

Zahnärzte vertreten die Auffassung, daß der Zahnverfall (Anzahl der von Karies befallenen Zähne) vom täglichen Zeitaufwand für die Mundhygiene (Zähneputzen, Munddusche etc.) abhängig ist. Untersuchen Sie diese Frage mit Hilfe einer linearen Einfachregressionsanalyse. Bei 10 Probanden erhielt man folgende Daten :

Patient Nr.	1	2	3	4	5	6	7	8	9	10
Anzahl der von Karies befallenen Zähne	9	5	4	2	6	2	2	1	5	4
tägl. Mundhygiene in Minuten	1	0	10	12	0	1	15	3	6	2

1) Stellen Sie das Ergebnis in einem Koordinatensystem dar !

2) Schätzen Sie die Regressionsparameter a und b mit Hilfe der Methode der kleinsten Quadrate !

3) Zeichen Sie die geschätzte Regressionsgerade in das Koordinatensystem ein !

4) Mit wie vielen von Karies befallenen Zähnen ist bei einem Patienten zu rechnen, der für die tägliche Mundhygiene 5 Minuten aufwendet ?

5) Errechnen Sie das Bestimmtheitsmaß R^2 !

6) Interpretieren Sie den unter 5) erhaltenen Wert !

7) Kritisieren Sie den hier gewählten Ansatz der linearen Einfachregression und geben Sie noch zwei weitere Einflußgrößen an, die Ihrer Meinung nach auf den „ Zahnverfall “ wirken !

Rechenhilfe : $\sum_{i=1}^{n} \hat{y}_i^2 = 170,8$

Aufgabe E 9:

In einem Windkanal wird der Prototyp einer Windkraftanlage getestet. Die Abhängigkeit der erzeugten Leistung (in $\frac{kw}{h}$) von der Windgeschwindigkeit (in $\frac{m}{sec.}$) soll mit Hilfe einer linearen Einfachregressionsanalyse untersucht werden. Folgende Meßwerte wurden ermittelt:

Windgeschwindigkeit (in $\frac{m}{sec.}$)	2	3	4	6	10
Leistung (in $\frac{kw}{h}$)	50	80	80	90	100

1) Beschreiben Sie den Vorgang der Datenerfassung, d.h. wie zum Zweck der Regressionsanalyse diese Daten gesammelt wurden, und erläutern Sie den Unterschied zur Datenerfassung bei der Korrelationsanalyse !

2) Zeichnen Sie die Beobachtungswerte in ein Koordinatensystem ein !

3) Bestimmen Sie mit der Methode der kleinsten Quadrate die Parameter der Regressionsgeraden !
   Zeichnen Sie die Regressionsgerade in das Koordinatensystem ein !

4) Berechnen Sie das Stichprobenbestimmtheitsmaß R^2 und interpretieren Sie das erhaltene Ergebnis !

5) Zeigen Sie mit Hilfe eines Beweisganges, daß das arithmetische Mittel der beobachteten y-Werte gleich dem arithmetischen Mittel der durch die Regressionsgerade geschätzten y-Werte ist !

Aufgabe E 10

Ein Mieterverein möchte die Abhängigkeit der monatlichen Heizkostenvorauszahlung von der jeweiligen Wohnfläche untersuchen.

Führen Sie mit den folgenden Daten eine lineare Einfachregressionsanalyse durch :

Monatl. Heizkostenvorauszahlung (in €)	80	100	90	120	110
Wohnfläche (in m^2)	50	60	70	80	90

1) Beschreiben Sie den Vorgang der Datenerfassung, d.h. wie zum Zweck der Regressionsanalyse die oben stehenden Daten hätten erfaßt werden müssen !

2) Tragen Sie die Datenpaare in ein Koordinatensystem ein !

3) Bestimmen Sie mit der Methode der kleinsten Quadrate die Parameter der Regressionsgeraden und zeichnen Sie diese in das Koordinatensystem ein !

4) Berechnen Sie das Bestimmtheitsmaß R^2 und interpretieren Sie den erhaltenen Wert !

5) Was versteht man unter der „ Schwerpunkteigenschaft “ der Regressionsgeraden ?

Aufgabe E 11

Ein Marktforschungsunternehmen hat im Auftrag einer Lebensmittelhandelskette Daten erhoben, um die Faktoren zu ergründen, die die Höhe der Lebensmittelausgaben eines Haushaltes bestimmen.

A

Die Marktforscher untersuchen zunächst den Zusammenhang zwischen Monatsmiete (in €) und Lebensmittelausgaben pro Monat (in €).

1) Geben Sie ein geeignetes Zusammenhangsmaß an !

2) Angenommen, die Berechnung dieser Maßzahl ergibt einen Wert von 0,66 .
Interpretieren Sie dieses Ergebnis !

3) Angenommen, bei der Berechnung des Zusammenhangs zwischen Monatseinkommen (in €) und Lebensmittelausgaben pro Monat (in €) ergibt einen Wert von 0,15 .
Interpretieren Sie dieses Ergebnis !

B

Zum Zweck einer anschließenden Regressionsanalyse werden die Lebensmittelausgaben pro Monat (in €) und die Anzahl der Personen pro Haushalt untersucht.

1) Beschreiben Sie den Vorgang der Datenerfassung !

Man erhält folgende Daten :

Lebensmittelausgaben $\left(\text{in } {}^{100\,€}\!/_{\text{Monat}}\right)$	4	6	6	7	9	10
Anzahl der Personen pro Haushalt	1	2	2	3	3	4

Man unterstellt, daß zwischen der Anzahl der Personen im Haushalt und den monatlichen Lebensmittelausgaben folgender Zusammenhang besteht :

$$y = a + b \cdot x$$

2) Schätzen Sie die Regressionskoeffizienten und zeichnen Sie die Wertepaare sowie die geschätzte Regressionsgerade in ein Koordinatensystem ein !

3) Interpretieren Sie den Wert $\hat{b}$ <u>inhaltlich</u> !

4) Berechnen Sie das Bestimmtheitsmaß R^2 !

5) Was bedeutet es für die geschätzte Regressionsgerade, wenn gilt : $R^2 = 1$?

6) Bestimmen Sie mit Hilfe eines formalen Rechenganges die Formel, die sich nach der Methode der kleinsten Quadrate für $\hat{b}$ ergibt, falls folgender Zusammenhang unterstellt wird :

$$y = b \cdot x$$

Aufgabe E 12:

A : Ein Psychologe untersucht in einem Betrieb verschiedene Arbeitsgruppen danach, ob ein Zusammenhang zwischen der Attraktivität einer Arbeitsgruppe (Zusammenhalt und Wohlbefinden ihrer Mitglieder) und der Streuung ihrer Arbeitsleistung besteht. Die Gruppenattraktivität wurde mit Hilfe einer Ordinalskala gemessen (von “1 = sehr hoch” bis “6 = sehr niedrig”), die Streuung der Leistung bezieht sich auf die Zahlen der von einer Arbeitsgruppe an verschiedenen Tagen fertiggestellten Produkte. Die Untersuchung brachte folgendes Ergebnis:

Arbeisgruppe	A	B	C	D	E	F	G
Gruppenattraktivität	1	2	2	3	4	5	6
Leistungsvarianz	32	2o	4o	32	64	68	6o

1) Was ist bei dieser Untersuchung der einzelne Merkmalsträger ?
2) Berechnen Sie einen geeigneten Korrelationskoeffizienten !
3) Interpretieren Sie inhaltlich das errechnete Vorzeichen dieses Koeffizienten !
4) Gibt der errechnete Wert dieses Koeffizienten Aufschluß über einen “Ursache–Wirkungs–Zusammenhang” zwischen den beiden untersuchten Merkmalen ? (Kurze Begründung !)

B : Die folgende Tabelle gibt für die Monate des ersten Halbjahres 19.. jeweils die Durchschnittszahl der täglich in Berlin fahrenden PKW's und die Zahl der Verkehrsunfälle, an denen PKW's beteiligt waren, an:

Monat	Jan.	Febr.	März	Apr.	Mai	Jun.
Verkehrsaufkommen (in Mio. PKW pro Tag)	1,4	1,4	1,6	1,8	1,8	1,6
Unfälle (inTsd.)	16	16	15	14	15	14

1) Berechnen Sie den Korrelationskoeffizienten nach Bravais–Pearson !
(Hinweis: Wenn Sie die Daten linear auf einstellige Zahlen transformieren, läßt es sich besser rechnen !)

Die Verkehrsverwaltung gibt folgende Pressemitteilung heraus:
“Diese Zahlen beweisen, daß zwischen dem PKW–Aufkommen und der Zahl der Unfälle mit PKW–Beteiligung nicht nur kein positiver Zusammenhang besteht, sondern mehr Autos sogar zu weniger Unfällen führen !”

2) Nehmen Sie kurz Stellung zu dieser Behauptung !
Welche dritte (latent wirkende) Variable könnte sowohl das Verkehrsaufkommen als auch die Zahl der Unfälle beeinflussen ?
Wie nennt man eine derartige Korrelation ?
3) Erläutern Sie kurz den Unterschied bei der Datenerfassung zum Zweck einer Korrelations– bzw. Regressionsanalyse ?

Aufgabe E 13:

A: Zum Zweck einer linearen Einfachregressionsanalyse mit Y als abhängiger und X als unabhängiger Variablen sei folgender Datensatz erhoben worden:

x_i	1	2	3	4	5
y_i	2	1	4	3	5

1) Zeichnen Sie die Punkte in ein Koordinatensystem ein !
2) Bestimmen Sie die Regressionskoeffizienten der Regressionsgeraden mit Hilfe der Methode der kleinsten Quadrate und zeichnen Sie die Regressionsgerade in das Koordinatensystem ein !
3) Berechnen Sie das Stichprobenbestimmtheitsmaß R^2 und interpretieren Sie den erhaltenen Wert !
4) Wie lautet das Optimalitätskriterium formal, nach dem Sie unter 2) die Regressionsgerade berechnet haben ?
5) Erläutern Sie kurz, warum man mit Hilfe der unter 2) berechneten Regressionsgeraden keine Schätzungen außerhalb des Stützbereiches vornehmen sollte!
6) Erläutern Sie kurz folgende Eigenschaften der unter 2) berechneten Regressionsgeraden:
   a) $\sum\limits_{i=1}^{n} \widehat{u}_i = 0$
   b) $\bar{y} = \widehat{a} + \widehat{b}\bar{x}$
   c) $\bar{y} = \bar{\widehat{y}}$

B: Zum Zweck einer linearen Korrelationsanalyse werden die beiden Merkmale X und Y untersucht.

1) Welche Skalierungsart müssen die beiden Untersuchungsmerkmale besitzen, wenn man den Korrelationskoeffizienten nach Bravais–Pearson zwischen ihnen berechnen will ?
2) Geben Sie ein Beispiel für eine Scheinkorrelation zwischen zwei Untersuchungsmerkmalen an !
3) Erläutern Sie kurz den Unterschied bei der Datenerfassung zum Zwecke einer Regressions– bzw. Korrelationsanalyse !

Aufgabe E 14 :

A

Bei der Produktion von bestimmten Teilen wird eine Maschine A zur Materialprüfung dieser Teile eingesetzt. Die benötigte **„ Prüfzeit [in sek] “ : Y_A** hängt dabei von der **„ Länge [in cm] “ : X_A** des hergestellten Teils ab. Aus der Produktion wurden **n = 5** Teile unterschiedlicher Länge ausgewählt, dann wurde die Prüfzeit gemessen und notiert :

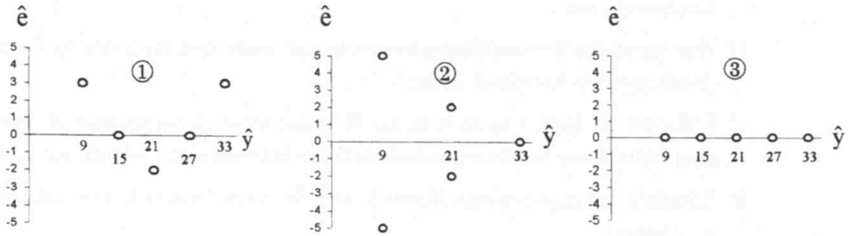
x_A	1	2	3	4	5
y_A	9	15	21	27	33

Eine lineare Einfachregression $\hat{y}_A = \hat{a} + \hat{b} \cdot x_A$ ergab folgende Regressionskoeffizienten :

$$\hat{a} = \mathbf{3} \quad \text{und} \quad \hat{b} = \mathbf{6}$$

1) Geben Sie den Wert des Bestimmtheitsmaßes R^2 an !

2) Welche der folgenden Abbildungen ① , ② bzw. ③ stellt den zugehörigen Residuenplot für die durchgeführte Regressionsanalyse dar ?



B

Eine neu angeschaffte moderne Maschine B soll bessere (≅ schnellere) Ergebnisse bringen. Die Prüfzeiten liegen bei doppelt so langen Teilen jetzt nur noch bei einem Drittel der Zeiten, welche die alte Maschine A benötigt.

Wie lauten die Regressionskoeffizienten $\hat{c}$ und $\hat{d}$ einer linearen Einfachregressionsanalyse $\hat{y}_B = \hat{c} + \hat{d} \cdot x_B$?

C

Angenommen, die **n = 5** Teile wären ausgewählt und hinsichtlich ihrer „ Länge “ und der benötigten „ Prüfzeit “ mit Hilfe einer **Korrelationsanalyse** untersucht worden.

Welchen Wert hätte dann der Korrelationskoeffizient r nach Bravais – Pearson ergeben

1) im Fall A ?

2) im Fall B ?

D

Angenommen, die **n = 5** Teile wären ausgewählt und hinsichtlich ihrer „ Länge “ und der benötigten „ Prüfzeit “ mit Hilfe einer **Rangkorrelationsanalyse** untersucht worden.

Welchen Wert hätte dann der Rangkorrelationskoeffizient τ nach Kendall ergeben

1) im Fall A ?

2) im Fall B ?

Aufgabe E 15

A

Ein Golfspieler benötigt für die 54 Löcher von drei Golfplätzen die folgende Anzahl von Schlägen (X) :

(Die jeweils 18 Löcher der drei Golfplätze sind hier von 1 bis 54 durchnummeriert)

Loch Nr. i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
x_i	3	7	6	4	8	6	2	4	10	3	7	4	4	3	5	3	2	6

Loch Nr. i	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
x_i	9	5	6	3	6	8	2	4	3	3	7	4	4	5	6	4	7	6

Loch Nr. i	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
x_i	2	5	4	3	8	5	2	3	9	4	10	4	6	5	3	4	7	4

Seine Tochter, die auch Golf spielt, benötigt für jedes Loch genau zwei Schläge mehr als der Vater. Die Anzahl der von der Tochter benötigten Schläge sei Y .

1) Geben Sie unter diesen Voraussetzungen die lineare Regressionsfunktion $\hat{y}_i = \hat{a} + \hat{b} \cdot x_i$ an, d.h. bestimmen Sie $\hat{a}$ und $\hat{b}$!

2) Geben Sie den Wert des Korrelationskoeffizienten r nach Bravais–Pearson zwischen X und Y an ! Interpretieren Sie den erhaltenen Wert inhaltlich !

B

1) Zeigen Sie mit Hilfe eines formalen Beweisganges, daß bei einer linearen Regressionsanalyse nach der Methode der kleinsten Quadrate mit

$$\mathbf{\hat{y}_i = y_i + \hat{e}_i = \hat{a} + \hat{b} \cdot x_i \quad (i = 1, \ldots, n)}$$

stets gilt :

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$$

2) Können die folgenden Zahlen die beobachteten Residuen einer linearen Regressionsanalyse nach der Methode der kleinsten Quadrate sein ?

–5 ; 2 ; 0 ; 1 ; –4

Begründen Sie kurz Ihre Antwort !

F

Kombinatorik und Ereignisalgebra

Aufgabe F 1

In der Unfallstation eines Krankenhauses arbeiten drei Ärzte : N , O und P . Da die Aufteilung der Wochenenddienste (Samstag und Sonntag) große Schwierigkeiten bereitet, entscheiden sich die drei Ärzte für ein Zufallsexperiment, um die Aufteilung vorzunehmen.
Es werden drei Zettel mit den Anfangsbuchstaben ihrer Namen (N , O und P) in eine Urne getan. Für die Aufteilung werden dann nach dem Zufallsprinzip aus der Urne zwei Zettel gezogen.

Geben Sie die möglichen Ausgänge (2 – Tupel) dieses Zufallsexperiments an, wenn

1) mit der Aufteilung festgelegt werden soll, an welchem Tag ein Arzt Dienst hat (1. Element des 2 – Tupels steht für Samstag), und es möglich sein soll, daß ein Arzt an beiden Tagen Dienst (d.h. Doppel–Dienst) hat !

2) Doppel–Dienst möglich ist, aber nicht bestimmt werden soll, an welchem Tag ein Arzt Dienst hat !

3) kein Doppel–Dienst möglich ist, aber bestimmt werden soll, an welchem Tag ein Arzt Dienst hat !

4) kein Doppel–Dienst möglich ist und nicht bestimmt werden soll, an welchem Tag ein Arzt Dienst hat !

Aufgabe F 2

A

Zu Beginn eines Semesters möchte eine Studentengruppe an 5 Tagen jeweils einen Orientierungsrundgang durch die Universität für Erstsemestler anbieten, der jeweils von einem Mitglied der Gruppe durchgeführt werden soll. Es haben sich 5 willige Mitglieder zur Verfügung gestellt.

1) Wie viele verschiedene Möglichkeiten gibt es, jedem Mitglied einen Wochentag (Montag bis Freitag) zuzuordnen ?

2) Da Klaus erkrankt ist, wird Karl zweimal einen Rundgang leiten.
   Wie viele verschiedene Möglichkeiten für die namentliche Belegung der Rundgänge gibt es ?

3) Karl besteht plötzlich darauf, daß er nicht an zwei aufeinanderfolgenden Tagen einen Rundgang leiten will.
   Wie viele Möglichkeiten für die namentliche Belegung der Rundgänge gibt es jetzt ?

B

In zwei Urnen I und II befinden sich jeweils 6 verschiedenfarbige Kugeln, wobei die farbliche Zusammensetzung der Kugeln in beiden Urnen gleich ist. Es werden zufällig aus jeder Urne 2 Kugeln ohne Zurücklegen gezogen.

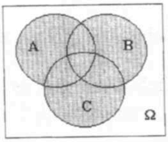
Wie viele Möglichkeiten gibt es,

1) zwei gleichfarbige Paare von Kugeln zu erhalten ?
2) vier verschiedenfarbige Kugeln zu erhalten ?
3) genau ein Paar von gleichfarbigen Kugeln zu erhalten ?

Aufgabe F 3

Richtig oder falsch ?

1) Seien A, B und $C \subset \Omega$ Ereignisse. Das Ereignis : „ Mindestens zwei dieser Ereignisse treten ein " läßt sich folgendermaßen graphisch darstellen :



2) Aus n Objekten sollen r Elemente ausgewählt werden. Dann kann die Zahl der Auswahlmöglichkeiten bei einer Variation mit Wiederholung nicht kleiner sein als bei einer Variation ohne Wiederholung.

3) Das sichere Ereignis besitzt die Wahrscheinlichkeit Eins. Ein Ereignis, das die Wahrscheinlichkeit Eins besitzt, ist aber nur „ fast sicher " .

4) Zwei Ereignisse, die keine gemeinsamen Elemente haben, sind disjunkt.

5) Verwendet man den objektiven Wahrscheinlichkeitsbegriff, so wird der Einzelfall betrachtet, nicht aber die Serie („ long run ") .

6) Für zwei Ereignisse A und B gilt :

$$P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B) .$$

7) Sind drei Ereignisse A , B und C paarweise stochastisch unabhängig, dann gilt auch :

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) .$$

8) Sind drei Ereignisse A , B und C vollständig disjunkt [d.h. $(A \cap B \cap C) = \varnothing$], dann sind sie auch paarweise disjunkt [d.h. $(A \cap B) = \varnothing$ bzw. $(A \cap C) = \varnothing$ bzw. $(B \cap C) = \varnothing$] .

9) Seien A und B Ereignisse (mit : $0 < P(A) < 1$ und $0 < P(B) < 1$) . Sei $P(A \cap B) = 0$. Dann sind A und B stochastisch unabhängig.

10) Die Wahrscheinlichkeit beim zweimaligen Werfen eines nicht–idealen Würfels höchstens eine „ 6 " zu würfeln, ist gleich der Wahrscheinlichkeit, mindestens einmal keine „ 6 " zu würfeln.

Aufgabe F 4

A

In einer Spielshow werden drei verschiedene Sonderpreise verlost. Für die Verlosung stehen drei Urnen bereit, die jeweils sechs Lose mit den Namen der sechs Spielteilnehmer enthalten. Aus jeder Urne wird jeweils ein Los gezogen.

1) Wie viele Möglichkeiten gibt es, aus den drei Urnen jeweils ein Los zu ziehen ?

2) Wie viele Möglichkeiten gibt es, aus den drei Urnen jeweils ein Los zu ziehen, wobei kein Name mehrfach gezogen wird ?

B

Nach der Spielshow werden die sechs (auf Bequemlichkeit bedachten Kandidaten) in zwei unterschiedlich bequemen Autos in ihr Hotel gebracht. In einem Fahrzeug können bis zu vier Kandidaten Platz nehmen.

Wie viele Möglichkeiten gibt es, die sechs Kandidaten in den beiden Autos unterzubringen ?

C

Der Verlag „ Schöne Kinderwelt “ möchte ein Märchenbuch herausbringen. Der Verlag besitzt die Rechte an 6 Märchen der Gebrüder Grimm und 4 Märchen von Hans Christian Andersen sowie die Rechte an 2 Märchen von anderen Autoren.
Aus Kostengründen sollen nur 7 Märchen ausgewählt werden.

1) Wie viele Möglichkeiten gibt es, die 7 Märchen auszuwählen ?

2) Wie viele Möglichkeiten bleiben, falls 3 Märchen der Gebrüder Grimm, 3 von Hans Christian Andersen und 1 Märchen von anderen Autoren ausgewählt werden sollen ?

3) Wie viele Möglichkeiten gibt es, die 7 ausgewählten Märchen in einem Buch anzuordnen ?

Aufgabe F 5

A

Bei einer Lotterie gibt es 10 Gewinner. Für diese stehen folgende Autos als Hauptpreis zur Verfügung :

5 Autos der Firma „ Audi “ | 3 Autos der Firma „ BMW “ | 2 Autos der Firma „ VW “

Wie viele Möglichkeiten gibt es, die Autos auf die Gewinner zu verteilen ?

B

Bei einer anderen Lotterie stellen sechs Firmen jeweils ein Auto zur Verfügung, wobei sich die Autos in Preis und Qualität erheblich voneinander unterscheiden.
Nur drei Gewinner haben ihren Gewinnanspruch angemeldet.

Wie viele Möglichkeiten gibt es, die Autos auf die drei Gewinner zu verteilen, wenn jeder jeweils nur ein Auto bekommen soll ?

Aufgabe F 6

Von drei Ereignissen $A, B, C \subset \Omega$ trete ein :

1) nur A
2) genau eines
3) höchstens eines
4) mindestens eines
5) genau zwei
6) mindestens zwei
7) mindestens eines nicht
8) mindestens zwei nicht

Stellen Sie die gesuchten Ereignisse mit Hilfe der Ereignisoperationen formal dar und zeichnen Sie jeweils das zugehörige Venn–Diagramm !

Aufgabe F 7

In einer Kleinstadt gibt es nur die beiden Tageszeitungen Z_1 und Z_2.
60 % der Bewohner lesen Z_1 und 80 % der Bewohner lesen Z_2.
Keine der beiden Zeitungen lesen 10 % der Bewohner.

Eine Person wird zufällig ausgewählt. Wie groß ist die Wahrscheinlichkeit, daß diese Person

1) beide Zeitungen liest ?

2) Z_1 liest, aber nicht Z_2 ?

3) Z_2 liest, wenn sie nicht Z_1 – Leser ist ?

4) höchstens eine Zeitung liest ?

5) Z_1 nicht liest ?

Aufgabe F 8:

In einer Altbauwohnung mit Außenwänden und veralteter elektrischer Anlage kommt es vor, daß der Strom ausfällt bzw. die Wasserzufuhr einfriert. Beide Umstände treten zwar unabhängig voneinander auf, sind jedoch von der Jahreszeit abhängig. So friert natürlich das Wasser nur ein, wenn es Winter ist, und zwar mit 80%iger Wahrscheinlichkeit. Der Strom fällt aber selbst wenn es nicht Winter ist, mit 40%iger Wahrscheinlichkeit aus, und zwar mit der gleichen Wahrscheinlichkeit, mit der der Strom, wenn es Winter ist, nicht ausfällt.

Gehen Sie davon aus, daß die Winterzeit 30% der gesamten Jahresszeit ausmacht.

1) Formalisieren Sie mit Hilfe von Ereignissymbolen die im Text genannten Wahrscheinlichkeitsaussagen !

Für die Lösung der folgenden Aufgabenteile muß der formale Lösungsweg deutlich ersichtlich sein !

2) Mit welcher Wahrscheinlichkeit muß man in dieser Wohnung

   a) mit dem Einfrieren der Wasserzufuhr rechnen ?
   b) mit einem Stromausfall rechnen ?
   c) mit dem Einfrieren der Wasserzufuhr und einem Stromausfall rechnen ?
   d) zusätzlich mit dem Einfrieren der Wasserzufuhr rechnen, wenn bereits der Stom ausgefallen ist ?
   e) zusätzlich mit einem Stromausfall rechnen, wenn bereits die Waserzufuhr eingefroren ist ?
   f) mit mindestens einem von beiden Mißständen rechnen ?
   g) mit höchstens einem von beiden Mißständen rechnen ?

Aufgabe F 9:

An einem kleinen Grenzübergang teilen sich zwei Zollbeamte den Dienst genau zur Hälfte. Wenn der 1. Zollbeamte Dienst hat, kontrolliert er mit 2o%iger Wahrscheinlichkeit Pässe und durchsucht unabhängig von einer Paßkontrolle mit 1o%iger Wahrscheinlichkeit Autos. Wenn der 2. Zollbeamte Dienst hat, kontrolliert dieser mit 7o%iger Wahrscheinlichkeit Pässe und durchsucht wiederum unabhängig von einer Paßkontrolle mit 4o%iger Wahrscheinlichkeit Autos.

1) Definieren Sie aufgrund des Textes sinnvolle Ereignisse und fomalisieren Sie damit die angegebenen Wahrscheinlichkeiten !

2) Sie fahren zu einem beliebigen Zeitpunkt mit einem Auto zu diesem Grenzübergang. Mit welcher Wahrscheinlichkeit müssen Sie

   a) mit einer Duchsuchung Ihres Autos rechnen ?

   b) mit einer Paßkontrolle rechnen ?

   c) mit einer Durchsuchung Ihres Autos und einer Paßkontrolle rechnen ?

   d) zusätzlich mit einer Duchsuchung Ihres Autos rechnen, wenn Ihr Paß bereits kontrolliert wurde ?

   e) zusätzlich mit einer Paßkontrolle rechnen, wenn Ihr Auto bereits durchsucht wurde ?

   f) mit genau einer von beiden Maßnahmen ?

Aufgabe F 1o:

Ein Mann sucht seinen Schirm, den er mit Wahrscheinlichkeit o,6 in einem Gebäude hat stehenlassen — und zwar, wenn es stimmt, mit gleichgroßer Wahrscheinlichkeit in einer der 6 Etagen.

1) Wie groß ist die Wahrscheinlichkeit, daß der Schirm im Gebäude und in der 6. Etage des Gebäudes ist ?

2) Wie groß ist die Wahrscheinlichkeit, daß der Schirm in der 6. Etage ist, wenn man gar nicht weiß, ob er im Gebäudes ist ?

3) Wie groß ist die Wahrscheinlichkeit, daß der Schirm nicht in den ersten 5 Etagen ist, wenn man weiß, daß er nicht in der 6. Etage ist ?

4) Wie groß ist die Wahrscheinlichkeit, daß der Schirm nicht in den ersten 5 Etagen ist ?

5) In den ersten 5 Etagen ist der Schirm nicht !
Wie groß ist die Wahrscheinlichkeit, daß der Schirm im Gebäude ist ?

Achtung: Benutzen Sie für Ihre Lösung folgende Ereignisdefinitionen:

G: "Schirm ist im Gebäude"

F: "Schirm ist in den ersten 5 Etagen "

S: "Schirm ist in der 6. Etage"

Aufgabe F 11

Das Straßenverkehrsamt führt eine Untersuchung zum Aggressionsverhalten von Autofahrern gegenüber Fahrradfahrern durch und kommt zu folgendem Ergebnis :

Auf eine Behinderung durch einen Fahrradfahrer reagiert ein männlicher Autofahrer mit 30 %–iger Wahrscheinlichkeit durch „ Hupen " und – unabhängig davon – mit 20 %–iger Wahrscheinlichkeit durch „ Beschimpfung " .

Auf eine Behinderung durch einen Fahrradfahrer reagiert ein weiblicher Autofahrer mit 15 %–iger Wahrscheinlichkeit durch „ Hupen " und – unabhängig davon – mit 10 %–iger Wahrscheinlichkeit durch „ Beschimpfung " .

Einem Mann am Steuer bzw. einer Frau am Steuer zu begegnen, kann als gleichwahrscheinlich angesehen werden.

Achtung : Im folgenden Rechengang muß der Lösungsweg in Ereignisschreibweise klar ersichtlich sein !

1) Definieren Sie entsprechende Ereignisse und ordnen Sie diesen die im Text genannten Wahrscheinlichkeitsangaben zu !

Angenommen, Sie behindern als Radfahrer eine Person am Steuer eines Autos.

Mit welcher Wahrscheinlichkeit

2) wird diese Person hupen ?
3) wird diese Person Sie beschimpfen ?
4) wird diese Person hupen und Sie beschimpfen ?
5) wird diese Person Sie auch noch beschimpfen, wenn sie bereits hupt ?

Aufgabe F 12:

Stellen Sie für die Prävalenz, Sensivität und Spezifität eine Übersicht dar, die die relative Häufigkeit und die Wahrscheinlichkeit für $n \to \infty$ beinhaltet !

Aufgabe F 13:

Die UNOCIF führt im Jahre 1993 eine Analyse über eine Erbkrankheit namens DNA–Verifiose (K), die meist tödlich verläuft, durch. Hierzu wurden weltweit 1.200.000 Personen auf Störungen ihrer DNA und der damit verbundenen DNA–Verifiose untersucht.
Ziel dieser Analyse war es, festzustellen, ob der von Prof. Dr. Gieser entwickelte Test mit sehr hoher Genauigkeit die Krankheit erkennt.

20.000 Personen wiesen die Symptome dieser Krankheit auf. Bei 5.350 Personen ist der Test negativ ausgefallen.

Der Test hatte eine Spezifität von 78% $\big($d.h. $P(T^-|\bar{K}) = 0{,}78\big)$.

1) Wie groß ist die Sensivität des Tests ?
2) Wie groß ist der positive Voraussagewert des Tests ?
3) Was besagen überhaupt Spezifität und Sensivität (verbal und formal) ?
4) Was ist die Prävalenz und wie läßt sie sich schätzen ?
5) Wann wäre der Test denn überhaupt ideal ?
6) Welcher fundamentale Satz der Statistik findet bei der Berechnung der sog. Voraussagewerte Anwendung ?
7) Woraus setzt sich die Test–Effizienz zusammen ?
8) Was versteht man formal unter "Testvalidität" und "Resultatvalidität" ?

Aufgabe F 14

A : Bei einem Pferderennen starten genau sieben Pferde.

1) Wie viele Möglichkeiten gibt es, die **sieben** Pferde auf die vorhandenen **zehn** Startboxen aufzuteilen ?

2) Wie viele mögliche Zieleinläufe gibt es auf den Plätzen **eins** bis **sieben** ?

3) Wie viele mögliche „Einlaufwetten" gibt es, d.h. wie viele Möglichkeiten gibt es für die Verteilung der Ränge **eins** bis **drei** ?

B : Heute ist in Hoppewald ein großer Renntag mit drei angesetzten Rennen. In den einzelnen Rennen gehen folgende Pferde an den Start :

Rennen I : A , B und C

Rennen II : D , E , F , G und H

Rennen III : A , B , C , D , E , F , G und H

Angenommen, es gelten folgende Siegwahrscheinlichkeiten für die einzelnen Pferde, wobei die Ausgänge der drei Rennen voneinander unabhängig sind :

Rennen	Pferde	Siegwahrscheinlichkeit
I	A	20 %
I	B	40 %
II	D	10 %
II	E	20 %
II	F	30 %
II	G	40 %

Rennen	Pferde	Siegwahrscheinlichkeit
III	A	5 %
III	B	10 %
III	C	5 %
III	D	5 %
III	E	10 %
III	F	10 %
III	G	5 %

Achtung : Im folgenden muß der formale Lösungsweg in Ereignisschreibweise klar ersichtlich sein !

1) Wie groß ist die Wahrscheinlichkeit, daß Pferd C in

a) genau einem
b) beiden
c) mindestens einem
d) höchstens einem
e) keinem

seiner Rennen gewinnt ?

Die Pferde C und H gehören einem bestimmten Rennstall.

2) Wie groß ist die Wahrscheinlichkeit, daß dieser Rennstall heute einen Sieger feiern kann, d.h. in mindestens einem Rennen den Sieger stellt ?

Aufgabe F 15 :

A : Beim Skat wird ein ideales Kartenspiel (≅ 32 Karten) auf drei Spieler und den „Skat" verteilt. Jeder Spieler erhält 10 Karten, die restlichen 2 Karten kommen in den „Skat". Die vier vorhandenen Buben haben beim Skatspiel ein ganz besonders wichtige Bedeutung.

Wie groß ist die Wahrscheinlichkeit, daß

1) zwei Buben im Skat liegen ?

2) genau ein Bube im Skat liegt ?

3) Kreuz-Bube im Skat liegt ?

B : Zwei ideale Würfel werden geworfen. Der eine Würfel ist **rot**, der andere ist **weiß**.

Folgende Ereignisse seien definiert :

A : „Die Augenzahl des **weißen** Würfels ist *gerade*"

B : „Die Augenzahl des **roten** Würfels ist *ungerade*"

C : „Die Augensumme beider Würfe ist *gerade*"

D : „Die Augensumme beider Würfe ist *ungerade*"

1) Bestimmen Sie die Wahrscheinlichkeiten der Ereignisse A , B , C und D !

2) Aus welchen Elementen von Ω setzen sich folgende Ereignisse zusammen :

a) $A \cap B$; b) $A \cap C$; c) $B \cap C$?

3) Bestimmen Sie mit Hilfe des Wahrscheinlichkeitsbegriffs von Laplace die Wahrscheinlichkeiten der Ereignisse

a) $A \cap B$; b) $A \cap C$; c) $B \cap C$!

4) Sind die Ereignisse A , B und C [**paarweise**] stochastisch unabhängig ?
Begründen Sie Ihre jeweilige Antwort mit Hilfe eines formalen Rechenganges !

5) Bestimmen Sie die Wahrscheinlichkeit des Ereignisses : $(A \cap B \cap C)$!

6) Sind die Ereignisse A , B und C [**vollständig**] stochastisch unabhängig, d.h. gilt

$$P(A \cap B \cap C) = P(A)\cdot P(B)\cdot P(C) \quad ?$$

7) Eine der folgenden Aussagen ist richtig. Geben Sie an, welche Aussage richtig ist :

a) „**Vollständig** stochastisch unabhängige Ereignisse sind immer auch **paarweise** stochastisch unabhängig"
b) „**Paarweise** stochastisch unabhängige Ereignisse sind immer auch **vollständig** stochastisch unabhängig"

C : Bei einem Tennismatch gewinnt Spieler A gegen Spieler B einen einzelnen Satz mit der Wahrscheinlichkeit p . Das gesamte Match gewinnt derjenige Spieler, der zuerst zwei Sätze gewonnen hat.

Sei G_i : „Spieler A gewinnt den i-ten Satz" $(i = 1 , 2 , 3)$

1) Geben Sie in Abhängigkeit von p die Wahrscheinlichkeit dafür an, daß Spieler A das Match gewinnt !

2) Wie groß ist die unter 1) bestimmte Wahrscheinlichkeit, falls gilt :

a) $p = \frac{1}{2}$; b) $p = \frac{3}{4}$?

Aufgabe F 16

Eine Familie war heute beim Picknick : Vater Martin, Mutter Silke, die Kinder Anja und Dirk sowie Opa Arnold. Auf dem Nachhauseweg bemerken die Kinder plötzlich, daß der Opa nicht mehr da ist. Es gibt drei Möglichkeiten :

A : Opa ist schon zu Hause und sitzt gemütlich in seinem Sessel

B : Opa ist noch auf dem Picknickplatz und flirtet mit jungen Mädchen

C : Opa ist in den nahegelegenen Wald gegangen und sucht Pilze

Aufgrund der Gewohnheiten des Opas kennt man die Wahrscheinlichkeiten für das Eintreten der Ereignisse A , B und C :

$$P(A) = 15\,\%; \quad P(B) = 80\,\% \quad ; \quad P(C) = 5\,\%$$

Anja wird zurück zum Picknickplatz und Dirk zum Waldrand geschickt, um den Opa zu suchen. Wenn Opa auf dem Picknickplatz ist, findet ihn Anja mit 90 %–iger Wahrscheinlichkeit, läuft er aber im Wald herum, wird ihn Dirk mit einer Wahrscheinlichkeit von nur 50 % finden.

[<u>Achtung</u> : Im folgenden muß der Lösungsweg in Ereignisschreibweise klar ersichtlich sein !]

1) Wie groß ist die Wahrscheinlichkeit, daß Anja den Opa findet ?

2) Wie groß ist die Wahrscheinlichkeit, daß Dirk den Opa findet ?

3) Wie groß ist die Wahrscheinlichkeit, daß eines der Kinder den Opa finden wird ?

4) Wie groß ist die Wahrscheinlichkeit dafür, den Opa bei der Rückkehr zu Hause in seinem Sessel sitzend anzutreffen, falls die Kinder ihn nicht finden sollten ?

Aufgabe F 17

Ein Blumenhändler hat ein Sortiment von Tulpenzwiebeln. Von diesen Zwiebeln ergeben **20 %** gelbe Tulpen, der Rest ergibt rote Tulpen. **60 %** der Zwiebeln ergeben Tulpen mit glatten Blättern, der Rest ergibt Tulpen mit spitzen Blättern.

Wenn eine Tulpe gelb ist, dann hat sie in **10 %** der Fälle glatte Blätter.

Achtung : In Ihrem Rechengang muß der Lösungsweg in Ereignisschreibweise klar ersichtlich sein !

Hinweis : Benutzen Sie bitte für Ihre Lösung folgende Ereignisdefinitionen :

R : „ Die Tulpe wird rot "

G : „ Die Tulpe wird glatte Blätter haben "

Berechnen Sie die Wahrscheinlichkeit, daß

1) eine Tulpe rote Farbe und glatte Blätter haben wird !
2) eine Tulpe rot wird, falls man weiß, daß sie glatte Blätter haben wird !
3) eine Tulpe glatte Blätter haben wird, falls man weiß, daß sie rot wird !
4) eine Tulpe rot wird, falls man weiß, daß sie spitzblättrig wird !
5) eine Tulpe rot oder glattblättrig wird, falls man weiß, daß sie rot und glattblättrig wird !
6) eine Tulpe gelb oder glattblättrig wird, falls man weiß, daß sie rot und spitzblättrig wird !
7) eine Tulpe rot oder glattblättrig wird, falls man weiß, daß sie rot und spitzblättrig wird !
8) eine Tulpe rot und glattblättrig wird, falls man weiß, daß sie rot oder glattblättrig wird !

Aufgabe F 18

Mastermind

4 – stellige Farbkombination
6 mögliche Farben
Farben dürfen mehrfach vorkommen

Der Spieler **S** (Super–Hirn) will die von **C** (Computer) gesteckte Farbkombination – bei optimaler Ausnützung aller bisherigen Informationen – erraten.

Spielzug	Farbkombination von **S**				Antwort von **C**
1	b	b	b	b	
2	r	r	r	r	☺ ☺
3	r	r	g	g	☺ 😐
4	r	w	r	w	☺ 😐
5	s	r	r	p	😐 😐 😐
6	r	p	p	r	☺ ☺

☺ : Farbe und Platz stimmen 😐 : Farbe stimmt, Platz stimmt nicht

[Farben : r ≅ rot ; w ≅ weiß ; b ≅ blau ; s ≅ schwarz ; g ≅ gelb ; p ≅ pink]

Bestimmen Sie die sechs Wahrscheinlichkeiten π_i (i = 1 , 2 , 3 , 4 , 5 , 6) , daß Spieler **S** **vor** dem i – ten Spielzug die richtige Farbkombination rät, falls er die Information über den vergangenen Spielverlauf optimal ausnützt !

G

Diskrete Zufallsvariablen und spezielle Verteilungsmodelle

Aufgabe G 1

Ein Bauer besitzt einen Schweinestall mit 10 Abteilen, in denen jeweils 2 Schweine stehen. Es besteht der Verdacht, daß einige Schweine von einer bestimmten, aber nicht übertragbaren Krankheit befallen sind, die durch einen Bluttest nachgewiesen werden kann.
Der Tierarzt macht den Vorschlag, bei jedem Schwein für ein Honorar von 10 € (pro Schwein) einen Bluttest durchzuführen.
Das bei diesem Vorschlag resultierende Gesamthonorar von 200 € erscheint dem Bauern jedoch viel zu hoch.
Der Bauer schlägt daher seinerseits vor, daß der Arzt das von den Tieren eines Abteils entnommene Blut mischt und diese Mischung für 10 € analysiert. Nur wenn hierbei ein Befund eintritt, soll der Arzt bei jedem Schwein des betreffenden Abteils einen Einzeltest durchführen.

Ob die Methode des Bauern von Vorteil ist (also weniger als 200 € kostet), hängt von der Wahrscheinlichkeit π ab, mit der ein Schwein von dieser Krankheit befallen wird.

1) Überlegen Sie zunächst *intuitiv* , für welche Werte von π (eher kleine oder eher große) die vom Bauern vorgeschlagene Methode kostengünstiger ist !

2) Entscheiden Sie nun *mathematisch*, für welche Werte von π die vom Bauern vorgeschlagene Methode kostengünstiger als die vom Arzt vorgeschlagene Methode ist, indem Sie den Erwartungswert des Arzthonorars (beim Vorschlag des Bauern) mit dem Fixbetrag des Arzthonorars (beim Vorschlag des Arztes) vergleichen !

   Rechenhilfe : $\sqrt{0,5} \approx 0,71$

Angenommen, Sie wären nicht Volks– oder Betriebswirt, sondern Landwirt und hätten dieses Problem zu lösen. Sicherlich hätten Sie dem Tierarzt dann einen Vorschlag unterbreitet, der etwas günstiger und auch für eine höhere Erkrankungswahrscheinlichkeit π für Sie vorteilhafter ist als die vom Tierarzt vorgeschlagene Methode !

3) Modifizieren Sie die oben beschriebene Methode des Bauern derart !
   Für welche Werte von π ist Ihre modifizierte Methode nun kostengünstiger als die vom Tierarzt vorgeschlagene Methode ?

   Begründen Sie Ihren Vorschlag mit Hilfe eines formalen Rechenganges !

   Rechenhilfe : $\sqrt{1,25} \approx 1,118$

Aufgabe G 2

Ein Losverkäufer verkauft („ unendlich viele “) Lose mit den gleichverteilten Ziffern 0 bis 9 .
Es gilt folgender Auszahlungsplan :

Beim Erscheinen der Ziffer 0 , 6 oder 9 wird nichts ausgezahlt.

Beim Erscheinen einer geraden Ziffer (außer 0 oder 6) wird 4,50 € ausgezahlt.

Beim Erscheinen einer ungeraden Ziffer (außer 9) wird 3 € ausgezahlt.

1) Berechnen Sie den Erwartungswert des Auszahlungsbetrages pro Ziehung !
2) Mit welcher Wahrscheinlichkeit erhält ein Loskäufer beim Kauf von 3 Losen eine Auszahlung von genau 3 € ?
3) Mit welcher Wahrscheinlichkeit erhält ein Loskäufer beim Kauf von 3 Losen eine Auszahlung von mindestens 4 € ?
4) Der Losverkäufer hat den Lospreis so festgesetzt, daß sein Tagesgewinn bei 1.000 verkauften Losen einen Erwartungswert von 450 € besitzt.
   Wie hoch ist dieser Lospreis ?

Aufgabe G 3

Jeder Flugzeugmotor versage bei einem Flug unabhängig von den anderen Motoren mit einer Wahrscheinlichkeit von $(1 - \pi) \cdot 100\,\%$.

Sei

X : „ Anzahl der funktionierenden Motoren eines **zwei**motorigen Flugzeugs “

Y : „ Anzahl der funktionierenden Motoren eines **drei**motorigen Flugzeugs “

Z : „ Anzahl der funktionierenden Motoren eines **vier**motorigen Flugzeugs “

1) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsfunktion von X !

2) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsfunktion von Y !

3) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsfunktion von Z !

A

Ein Flugzeug kann sich in der Luft halten (und auch sicher landen), wenn mindestens **zwei Drittel** der Motoren funktionieren.

Sind Flugzeuge mit zwei, drei oder vier Motoren zuverlässiger, d.h. für welchen Flugzeugtyp ist die Wahrscheinlichkeit eines Absturzes am geringsten, falls für die Ausfallwahrscheinlichkeit eines Motors gilt :

a) 70 % b) 20 % c) 1 %

B

Ein Flugzeug kann sich in der Luft halten (und auch sicher landen), wenn mindestens **die Hälfte** der Motoren funktioniert.

Sind Flugzeuge mit zwei, drei oder vier Motoren zuverlässiger, d.h. für welchen Flugzeugtyp ist die Wahrscheinlichkeit eines Absturzes am geringsten, falls für die Ausfallwahrscheinlichkeit eines Motors gilt :

a) 70 % b) 20 % c) 1 %

Aufgabe G 4

In einer Spielbank eines kleines Kurortes wird „ Mini–Roulette “ gespielt. Dieses Roulette ist in vier gleich große Sektoren aufgeteilt. Die Kugel kann bei einem Durchlauf nur in einen dieser Sektoren fallen. In jedem Sektor ist eine Zahl aufgedruckt :

„ – 1 “	„ – 2 “	„ + 1 “	„ + 2 “

Es gelten folgende Spielregeln :

⇨ Fällt die Kugel in den Sektor „ + 1 “ oder „ + 2 “ , so steht das Ergebnis fest, Sie bekommen den entsprechenden Betrag von der Bank in € ausgezahlt und das Spiel ist beendet.

⇨ Fällt die Kugel in den Sektor „ – 1 “ oder „ – 2 “ , so wird das Roulette noch einmal in Gang gesetzt. Die Ergebnisse <u>beider</u> Durchgänge werden addiert und bilden das Ergebnis :

- Ist das Ergebnis positiv, so erhalten Sie diesen Betrag von der Bank in € ausgezahlt und das Spiel ist beendet.

- Ist das Ergebnis negativ, so müssen Sie diesen Betrag an die Bank in € bezahlen und das Spiel ist beendet.

- Ist das Ergebnis gleich Null, so haben Sie weder Gewinn noch Verlust und das Spiel endet unentschieden.

1) Stellen Sie tabellarisch die Wahrscheinlichkeitsfunktion der Zufallsvariablen

X : „ Gewinn pro Spiel (in €) “

auf !

2) Ist dieses Spiel „ fair “ ?
Begründung !

3) Berechnen Sie Var (X) !

Aufgabe G 5:

Ein aus 4 Karten (Pik As, Herz As, Kreuz Dame und Karo Bube) bestehendes Kartenspiel wird gemischt und an zwei Spieler verdeckt verteilt, so daß jeder Spieler zwei Karten bekommt.

Hinweis: Notieren Sie die 6 Kartenverteilungen, die ein Spieler erhalten kann !

A : Angenommen, einer der beiden Spieler verkündet:

" Ich habe ein As ".

Wie groß ist die Wahrscheinlichkeit, daß dieser Spieler auch das zweite As in der Hand hält ?

B : Angenommen, einer der beiden Spieler verkündet:

" Ich habe ein Pik As ".

Wie groß ist die Wahrscheinlichkeit, daß dieser Spieler auch das zweite As in der Hand hält ?

C : Bestimmen Sie die erwartete Zahl der Asse, die ein Spieler bei der Kartenverteilung erhält !

Aufgabe G 6

Falls in der Sendung „ Einer wird gewinnen " zwei Kandidaten punktgleich sind, wird der Sieger durch Würfelwurf ermittelt, d.h. es findet ein Würfeldurchgang statt, bei dem jeder Kandidat mit einem idealen Würfel einmal würfelt. Derjenige von beiden, der die höhere Augenzahl würfelt, wird zum Sieger erklärt. Würfeln beide die gleiche Augenzahl, findet ein zweiter (dritter usw.) Würfeldurchgang statt, bis der Sieger feststeht.

Bestimmen Sie die Wahrscheinlichkeit, daß bis zur Entscheidung

1) genau 4

2) mindestens 4

3) höchstens 4

Würfeldurchgänge notwendig sind !

Aufgabe G 7

Die beiden Statistik–Studenten Otto und Paul haben sich folgendes Zufallsexperiment ausgedacht :
Eine ideale Münze (Zahl oder Wappen) wird dreimal hintereinander geworfen.

Spiel A Man gewinnt so viel €, wie bei diesem Experiment „ Zahl " geworfen wird !
Sei X : „ Gewinn bei Spiel A " .

Spiel B Hier wird nur das Ergebnis des ersten Wurfes berücksichtigt.

- Zeigt der 1. Wurf „ Wappen " , so gewinnt man 3 € .
- Zeigt der 1. Wurf „ Zahl " , hat man weder Gewinn noch Verlust.

Sei Y : „ Gewinn bei Spiel B " .

1) Geben Sie die Wahrscheinlichkeitsfunktion der Zufallsvariablen X an !
2) Geben Sie die Wahrscheinlichkeitsfunktion der Zufallsvariablen Y an !

Paul entscheidet sich für Spiel A und Otto entscheidet sich für Spiel B !

3) Berechnen Sie den Erwartungswert und die Varianz des Gewinns von Paul !
4) Berechnen Sie den Erwartungswert und die Varianz des Gewinns von Otto !
5) Für welches Spiel würden Sie sich entscheiden ?
Begründen Sie Ihre Antwort !
6) Bestimmen Sie die Wahrscheinlichkeitsfunktion der Zufallsvariablen $X \cdot Y$ und berechnen Sie $E(X \cdot Y)$!
7) Berechnen Sie $Cov(X, Y)$!
8) Bestimmen Sie die Wahrscheinlichkeitsfunktion der Zufallsvariablen

Z : „ Gemeinsamer Gewinn von Otto und Paul " !

9) Berechnen Sie den Erwartungswert und die Varianz des gemeinsamen Gewinns von Otto und Paul !
10) Sind X und Y stochastisch unabhängig ?
Begründung !

Aufgabe G 8

Eine Gruppe von Umweltschützern fährt auf das Meer hinaus und sammelt treibende Giftfässer ein. Sei

X : „ Anzahl der am ersten Tag in einer Region gefundenen Fässer “

mit

x	3	4	5
P (X = x)	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

1) Mit welcher Wahrscheinlichkeit werden die Umweltschützer am ersten Tag mindestens vier Fässer in einer neuen Region finden ?

2) Wie viele Fässer findet man durchschnittlich bei der Suche am ersten Tag in einer neuen Region ?
Berechnen Sie die Varianz von X !

Fahren die Umweltschützer einen zweiten Tag in dieselbe Region, verändern sich die Wahrscheinlichkeiten für das Auffinden von Fässern.
Sei

Y : „ Anzahl der am zweiten Tag in derselben Region gefundenen Fässer “

mit folgenden <u>bedingten</u> Wahrscheinlichkeiten :

$$P(Y=0/X=3)=\frac{2}{4} \qquad P(Y=0/X=4)=\frac{1}{2} \qquad P(Y=0/X=5)=\frac{3}{4}$$

$$P(Y=1/X=3)=\frac{1}{4} \qquad P(Y=1/X=4)=\frac{1}{2} \qquad P(Y=1/X=5)=\frac{1}{4}$$

$$P(Y=2/X=3)=\frac{1}{4} \qquad P(Y=2/X=4)=0 \qquad P(Y=2/X=5)=0$$

3) Stellen Sie die gemeinsame Wahrscheinlichkeitsfunktion von X und Y in einer Tabelle dar !

4) Berechnen Sie die Kovarianz von X und Y !

5) Sind X und Y stochastisch unabhängig ? Begründung !

Die Umweltschützer bekommen eine Art „ Fundprämie “ für die gefundenen Fässer. Ein unbekannter Spender zahlt ihnen 5 € für jedes gefundene Faß. Zusätzlich wird jede Fahrt mit 10 € honoriert.

6) Welchen Erlös in € können die Umweltschützer erwarten, wenn sie in einer neuen Region einmal am ersten Tag und einmal am zweiten Tag auf das Meer fahren ?

Aufgabe G 9:

Die zweidimensionale Zufallsvariable (X, Y) besitze die folgende Verteilung (Wahrscheinlichkeitsfunktion):

		Y		
		1	2	3
X	1	o,1	o,3	o,2
	2	o,1	o,1	o,2

Bestimmen Sie

1) den Erwartungswert und die Varianz von X und Y !

2) den Erwartungswert und die Varianz sowie die Verteilung der Summe $X + Y$!

3) den Erwartungswert und die Varianz sowie die Verteilung des Produktes $X \cdot Y$!

4) Berechnen Sie $Cov(X, Y)$!

5) Interpretieren Sie 4).

Aufgabe G 1o:

Kerzen einer bestimmten Sorte werden in einem Geschäft als Einzel- und als Doppelpackung angeboten. Die Zahl der pro Tag verkauften Einzelpackungen (X) und die Zahl der pro Tag verkauften Doppelpackungen (Y) seien unabhängige Zufallsgrößen mit folgenden Wahrscheinlichkeitsverteilungen:

x	0	1	2	3
P (X = x)	o,1	o,4	o,3	o,2

y	0	1	2
P (Y = y)	o,4	o,3	o,3

1) Bestimmen Sie den Erwartungswert und die Varianz von X !
   Es gilt: $E(Y) = 0{,}9$ und $Var(Y) = 0{,}69$.
   Sei Z: "Anzahl der pro Tag verkauften Kerzen".

2) Bestimmen Sie den Erwartungswert und die Varianz von Z !

3) Bestimmen Sie die Wahrscheinlichkeitsverteilung von Z !

Aufgabe G 11

Nach dem vergeblichen Versuch, das Vordiplom in Statistik zu erwerben, macht sich Student Emil selbständig. Er will sich als Glücksspieler eine sichere Existenz aufbauen. Ihm werden die beiden folgenden Spiele angeboten :

Spiel 1 Es wird mit zwei idealen Würfeln gewürfelt.
Das Produkt der obenliegenden Augenzahlen wird in € ausgezahlt.

Spiel 2 Es wird mit zwei idealen Würfeln gewürfelt.
Die 8 – fache Summe der obenliegenden Augenzahlen wird in € ausgezahlt.

Sei

Z_1 : „ Auszahlung bei Spiel 1 “

Z_2 : „ Auszahlung bei Spiel 2 “

G_1 : „ Gewinn bei Spiel 1 “

G_2 : „ Gewinn bei Spiel 2 “

1) Welchen Einsatz in € muß Emil bei **Spiel 1** bezahlen, damit dieses Spiel „ fair “ ist ?
2) Welchen Einsatz in € muß Emil bei **Spiel 2** bezahlen, damit dieses Spiel „ fair “ ist ?

Es gilt :

$$E\left(Z_1^2\right) \approx 230 \quad \text{und} \quad E\left(Z_2^2\right) \approx 3509$$

3) Wie hoch ist die Standardabweichung des Gewinns bei **Spiel 1** ?
4) Wie hoch ist die Standardabweichung des Gewinns bei **Spiel 2** ?

Aufgabe G 12

Es wird folgendes Glücksspiel betrachtet :

Ein idealer Würfel wird einmal geworfen. Der Spieler kann Spielmarken auf die Ereignisse

$$A = \{ 1 , 3 , 5 \} \quad \text{bzw.} \quad B = \{ 5 , 6 \} \quad \text{bzw.} \quad C = \{ 6 \}$$

setzen. Für eine auf A (bzw. B , bzw. C) gesetzte Marke erhält er 1 € (bzw. 2 € , bzw. 6 €), falls A (bzw. B , bzw. C) eintritt. Ansonsten erhält er nichts.

Sei X_A (bzw. X_B , bzw. X_C) die Auszahlung, die ein Spieler erhält, der eine Marke auf A (bzw. B , bzw. C) gesetzt hat.

1) Geben Sie tabellarisch die Wahrscheinlichkeitsfunktionen von X_A , X_B und X_C an !

2) Bestimmen Sie tabellarisch die gemeinsame Wahrscheinlichkeitsfunktion von

   (X_A , X_B) bzw. (X_A , X_C) bzw. (X_B , X_C)

   Welche dieser Variablenpaare sind stochastisch unabhängig ?
   Welche dieser Variablenpaare sind korreliert ?
   Erklären Sie die Vorzeichen der jeweiligen Kovarianz <u>inhaltlich</u> !

3) Spieler 1 setzt jeweils 3 Marken auf A , B und C.

   Wie groß sind Erwartungswert und Varianz der Auszahlung ?

4) Spieler 2 setzt 1 Marke auf A , 6 Marken auf B und 2 Marken auf C.

   Wie groß sind Erwartungswert und Varianz der Auszahlung ?

5) Spieler 3 möchte seine 9 Marken so plazieren, daß der Erwartungswert der Auszahlung 6,50 € beträgt und die <u>Varianz minimal</u> wird.

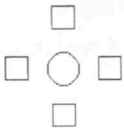
   Wie muß er die Marken setzen ?

<u>**Hilfe**</u> : $\mathbf{(a + b + c)^2 = a^2 + b^2 + c^2 + 2 \cdot a \cdot b + 2 \cdot a \cdot c + 2 \cdot b \cdot c}$

: Dieser Punkt ist etwas knifflig !

Aufgabe G 13

An einem runden Tisch stehen vier Stühle für eine Bridgepartie, an der vier Personen teilnehmen.



An der Bridgepartie nehmen zwei Frauen, die Röcke tragen, und zwei Männer, die Smokings tragen, teil.

Sei

X : „ Anzahl der männlichen Sitznachbarn eines beliebigen [≅ ■] Bridgespielers "

Y : „ Anzahl der Röcke, die dieser beliebige [≅ ■] Bridgespieler trägt "

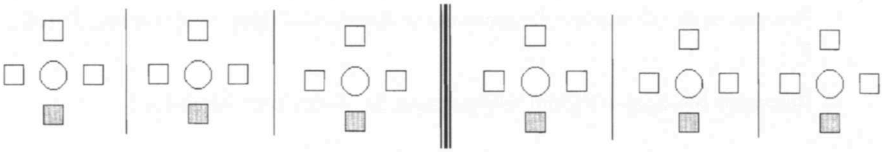
1) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsfunktion von X !

2) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsfunktion von Y !

3) Bestimmen Sie tabellarisch die gemeinsame Wahrscheinlichkeitsfunktion von X und Y !

4) Berechnen Sie Cov (X , Y) !

Hilfe : Es gibt **6** unterschiedliche Sitzkonstellationen :



Aufgabe G 14:

Die Fußballsaison geht zu Ende. Der 1.FC Berlin muß noch zwei Spiele bestreiten, zunächst ein Heimspiel und zum Schluß ein Auswärtsspiel.

In jedem Spiel gibt es für einen Sieg zwei Punkte, für ein Unentschieden einen Punkt und für eine Niederlage keinen Punkt.

Der Trainer orakelt:
"In dem Heimspiel holen wir mit 50%-iger Wahrscheinlichkeit beide Punkte und mit der Wahrscheinlichkeit $\frac{1}{3}$ einen Punkt. In dem Auswärtsspiel ist Sieg, Niederlage und Unentschieden gleichwahrscheinlich.

Es ist fast unmöglich, daß wir in den beiden Spielen überhaupt keinen Punkt mehr holen bzw. beide Spiele unentschieden ausgehen.

Sollten wir im Heimspiel jedoch beide Punkte holen, beträgt die Wahrscheinlichkeit, daß wir im letzten Spiel mindestens einen Punkt holen, 60% .

Wenn wir aber im Heimspiel verlieren sollten, ist die Wahrscheinlichkeit immerhin 40%, daß die Mannschaft sich aufrafft und das Auswärtsspiel gewinnt."

Sei X: "Anzahl der gewonnenen Punkte im Heimspiel"

Sei Y: "Anzahl der gewonnenen Punkte im Auswärtsspiel"

1) Bestimmen Sie tabellarisch die gemeinsame Wahrscheinlichkeitsfunktion von X und Y !

2) Berechnen Sie $E(X+Y)$ und interpretieren Sie diesen Wert inhaltlich !

Aufgabe G 15

In einer Urne befinden sich drei mit den Zahlen 1 bis 3 nummerierte Kugeln. Folgendes Zufallsexperiment wird durchgeführt :

In der ersten Ziehung wird eine Kugel zufällig ausgewählt und deren Zahl festgestellt. Vor der zweiten Ziehung werden zunächst alle Kugeln aus der Urne entfernt, die eine größere Zahl haben. Danach wird die in der ersten Ziehung gezogene Kugel in die Urne zurückgelegt, und es wird (in der zweiten Ziehung) wiederum eine Kugel ausgewählt.

Sei

X : „ Zahl der ersten gezogenen Kugel "

Y : „ Zahl der zweiten gezogenen Kugel "

1) Bestimmen Sie tabellarisch die gemeinsame Wahrscheinlichkeitsfunktion von X und Y !

2) Berechnen Sie Cov (X , Y) !

3) Sind X und Y stochastisch unabhängig ? Begründung !

Angenommen, das oben beschriebene Zufallsexperiment bildet die Basis für ein Glücksspiel, bei dem die Summe der Zahlen auf den gezogenen Kugeln in € ausgezahlt wird.

4) Welchen Einsatz in € müßte man für einen solchen Spieldurchgang verlangen, damit dieses Spiel als „ fair " gelten kann ?

Aufgabe G 16:

Drei Studenten wollen sich nacheinander von einem Professor in Statistik prüfen lassen, wobei das Bestehen bzw. Nichtbestehen der Prüfung für die drei Studenten als unabhängig voneinander angesehen werden kann. Die Wahrscheinlichkeit, die Prüfung zu bestehen, sei für den ersten Prüfling 80%, für den zweiten Prüfling 70% und für den dritten Prüfling 60%.

Sei: X: "Anzahl der erfolgreichen Studenten unter den ersten beiden Prüflingen"

Sei: Y: "Anzahl der erfolgreichen Studenten unter den letzten beiden Prüflingen"

Sei: Z: "Anzahl der insgesamt erfolgreichen Studenten unter allen drei Prüflingen"

1) Bestimmen Sie tabellarisch die gemeinsame Wahrscheinlichkeitsfunktion der Zufallgröße X und Y !

2) Sind die Zufallsgrößen X und Y stochastisch unabhängig ? Begründung !

3) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsfunktion der Zufallgröße Z !

Aufgabe G 17

Der Besitzer eines Obstladens hat folgende Probleme, die Sie für ihn lösen sollen :

Die Zufallsvariable X sei die täglich nachgefragte Menge an Erdbeerkörbchen und besitze folgende Wahrscheinlichkeitsverteilung :

x	1	2	3
P (X = x)	0,3	0,4	0,3

Der Besitzer des Obstladens bezieht täglich die Erdbeerkörbchen vom Fruchthof nach folgendem Zufallsprinzip :

> In einer Urne befinden sich 2 weiße und 3 grüne Kugeln. Er zieht zufällig (und ohne Zurücklegen) 3 Kugeln aus dieser Urne. Falls 2 von den gezogenen Kugeln grün sind, bestellt er 2 Erdbeerkörbchen, in allen anderen Fällen bestellt er 3 Erdbeerkörbchen.

Die nachgefragte Menge an Erdbeerkörbchen sei unabhängig von der Angebotsmenge (= Bestellmenge).

1) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsfunktion für die Angebotsmenge (Y) und errechnen Sie E (Y) und Var (Y) !

2) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsfunktion der Zufallsvariablen Z = Y – X und interpretieren Sie die Realisationsmöglichkeiten von Z !

3) Berechnen Sie E (Z) und interpretieren Sie diesen Wert !

4) Berechnen Sie Var (Z) !

5) Stellen Sie die gemeinsame Wahrscheinlichkeitsfunktion von X und Y tabellarisch dar !

6) Bestimmen Sie die Kovarianz Cov (X , Y) !

7) Bestimmen Sie die <u>bedingte</u> Wahrscheinlichkeitsfunktion von X , falls gilt : y = 3 !

Falls die Nachfrage nicht befriedigt werden kann, möge das den Besitzer 1 € pro Körbchen kosten (Image–Verlust). Können Erdbeerkörbchen jedoch nicht verkauft werden, kostet das den Besitzer 2 € pro Körbchen.

8) Wie hoch sind die zu erwartenden Kosten ?

Aufgabe G 18

Die Firma Kallocks bringt eine neue Sorte Kornflakes auf den Markt. Als Kaufanreiz wird den Packungen ein halbes Jahr lang ein Sammelcoupon beigelegt. Für jeweils 4 solcher Coupons kann man sich dann nach einem halben Jahr ein Poster zuschicken lassen.
Nun ist leider die Maschine, die für die Verteilung der Coupons auf die einzelnen Packungen vorgesehen ist, mit einem Fehler behaftet. Dieser Fehler bewirkt, daß 25 % der Packungen ohne Coupon zur Auslieferung kommen.
Fritzchen ist scharf auf diese Poster und kauft daher jede Woche von seinem Taschengeld eine Packung Kornflakes von dieser neuen Sorte.

Sei

X : „ Anzahl der von Fritzchen gekauften Packungen mit Coupon innerhalb eines halben Jahres ($\cong$ 26 Wochen) "

1) Wie ist die Zufallsvariable X verteilt ? [Verteilungstyp und –parameter !]

2) Wie groß ist die Wahrscheinlichkeit, daß Fritzchen Coupons für

   a) genau 3 Poster
   b) höchstens 4 Poster
   c) genau 6 Poster
   d) höchstens 1 Poster

   in einem halben Jahr zusammenbekommt ?

3) Mit wie vielen Packungen mit Coupon kann Fritzchen in dem halben Jahr rechnen ?

Aufgabe G 19:

Von allen Fluggästen, die Plätze reservieren, erscheinen 1o% nicht. Die Fluggesellschaft weiß dies und verkauft 29 Flugkarten für 26 verfügbare Plätze.

Wie groß ist die Wahrscheinlichkeit, daß alle Fluggäste Platz bekommen ?

Berechne diese Wahrscheinlichkeit !

Aufgabe G 2o:

Zwischen 2 und 4 Uhr nachmittags ist die durchschnittliche Anzahl der Telefongespräche, die die Vermittlung einer Firma pro Minute empfängt, gleich 2,5 .

1) Wie ist die Z.V. X:"Anzahl der in dieser Zeit empfangenen Telefonate pro Minute" verteilt ?

2) Wie groß ist die Wkt, daß während einer bestimmten Minute in dieser Zeit

   a) kein,

   b) weniger als drei,

   c) vier oder mehr

Telefonate empfangen werden ?

Aufgabe G 21

A

Einem Prüfling A wird ein Gesamtkatalog mit 10 Zetteln vorgelegt, auf denen je eine Prüfungsfrage steht. Der Prüfling weiß, daß der zuständige Prüfer von diesen 10 Fragen 6 Fragen so schwer gemacht hat, daß kein Prüfling sie beantworten könnte.
Von den 10 Fragen darf der Prüfling nun selbst 3 Fragen für seine Prüfung zufällig auswählen.

Wie groß ist die Wahrscheinlichkeit, daß der Prüfling

1) alle 3 Fragen richtig beantwortet ?

2) 3 beantwortbare Fragen zieht ?

3) mindestens eine beantwortbare Frage zieht ?

B

Einem Prüfling B werden 12 Fragen zum Beantworten vorgelegt. Die Wahrscheinlichkeit, daß der Prüfling auf eine solche Frage eine richtige Antwort gibt, sei 75 % .
Die Prüfung gilt als bestanden, falls der Prüfling mindestens zwei Drittel der Fragen richtig beantworten kann.

Wie groß ist die Wahrscheinlichkeit, daß der Prüfling

1) genau 4 Fragen richtig beantwortet ?

2) mindestens 7 Fragen richtig beantwortet ?

3) höchstens 9 Fragen richtig beantwortet ?

4) mehr als 2 , aber höchstens 8 Fragen richtig beantwortet ?

5) die Prüfung besteht ?

C

Einem Prüfling C werden 60 Fragen zum Beantworten vorgelegt. Die Wahrscheinlichkeit, daß der Prüfling auf eine solche Frage eine falsche Antwort gibt, sei 5 % .
Die Prüfung gilt als bestanden, falls der Prüfling mindestens 80 % der Fragen richtig beantworten kann.
Die Prüfungsleistung wird mit „ ausgezeichnet " bewertet, falls der Prüfling höchstens 3 Fragen nicht richtig beantwortet.

Wie groß approximativ ist die Wahrscheinlichkeit, daß der Prüfling

1) die Prüfung besteht ?

2) die Prüfung besteht und als Prüfungsleistung „ ausgezeichnet " erhält ?

3) die Prüfung besteht, aber als Prüfungsleistung nicht „ ausgezeichnet " erhält ?

Aufgabe G 22:

Ein Beamter im Umweltamt ist zuständig für die Betriebe der chemischen Industrie. Jedes Unternehmen erstellt täglich einen Bericht für das Umweltamt, wobei alle Betriebe voneinander unabhängig arbeiten. Der Beamte teilt die Tagesberichte in positive (keine Zwischenfälle) und negative (ein oder mehrere Zwischenfälle) Tagesberichte ein.

A: Es werden 25 zufällig ausgewählte Betriebe einer betimmten Region betrachtet. Die Wahrscheinlichkeit, daß ein positiver Tagesbericht erstellt wird, ist für jeden Betrieb gleich 90% .

1) Wie ist die Zufallsgröße X: "Anzahl der pro Tag erstellten negativen Tagesberichte" verteilt ? (Verteilungstyp und Verteilungsparameter !)
2) Wie groß ist die Wahrscheinlichkeit, daß an einem Tag unter den Tagesberichten mindestens einer negativ ist ?
3) Wie groß ist die Wahrscheinlichkeit, daß an einem Tag mehr als 18 , aber weniger als 22 positive Tagesberichte erstellt werden ?
4) Wie groß ist die Wahrscheinlichkeit, daß an einem Tag mindestens 22 positive Tagesberichte erstellt werden ?
5) Der Beamte denkt über die anfallende Arbeit der nächsten Woche nach. Wieviele negative Berichte kann er für Montag erwarten ?

B: Durch strenge Gesetze wurden die 25 Betriebe gezwungen, Schutz– und Sicherheitsmaßnahmen zu verstärken. Die Wahrscheinlichkeit, für einen negativen Tagesbericht sank daraufhin bei allen Betrieben auf 1% .

Im Umweltamt wird überlegt, ob die Tagesberichte für einen Monat (= 30 Tage) zusammengefaßt werden sollen.

1) Wie ist die Zufallsgröße Y: "Anzahl der pro Monat erstellten negativen Tagesberichte" exakt verteilt ? (Verteilungstyp und Verteilungsparameter !)
2) Wieviele negative Tagesberichte können im Umweltamt für April (30 Tage) erwartet werden ?
3) Durch welche <u>diskrete</u> Verteilung kann die Verteilung von Y approximiert werden? (Überprüfen Sie die Approximationsvoraussetzungen und geben Sie Verteilungstyp und Verteilungsparameter an !)
4) Mit welcher Wahrscheinlichkeit werden für April weniger als 8 negative Tagesberichte erstellt ?
5) Mit welcher Wahrscheinlichkeit werden für April mehr als 744 positive Tagesberichte erstellt ?

Aufgabe G 23

Am Ende einer Pflichtveranstaltung muß man sich entweder bei Prof. Lockerleicht oder bei Prof. Ätzend mündlich prüfen lassen. Auf dem Prüfungsamt muß jeder Kandidat ein Los aus einer Urne mit 100 Losen ziehen, wobei 50 Lose mit „ Lockerleicht “ und 50 Lose mit „ Ätzend “ beschriftet sind. Nach Ziehung eines Loses schreibt sich Herr Schmidt vom Prüfungsamt den Namen des Prüfers auf und legt das Los sofort zurück in die Urne.

A

1) Wie groß ist die Wahrscheinlichkeit , daß ein Kandidat von Prof. Lockerleicht geprüft wird ?
2) Um welches spezielles Wahrscheinlichkeitsmodell handelt es sich ?

B

Herr Schmidt stellt zwei Tage vor Ende der Anmeldefrist fest, daß sich noch 8 Studenten anmelden werden.

1) Wie ist die Zufallsvariable X : „ Anzahl der für Prof. Lockerleicht anfallenden Prüfungen bzgl. der 8 verbleibenden Studenten “ verteilt ?
2) Wie viele von den noch nicht angemeldeten Prüflingen werden vermutlich von Prof. Lockerleicht geprüft werden ?
3) Wie groß ist die Wahrscheinlichkeit, daß von den noch nicht angemeldeten Prüflingen
   a) genau die Hälfte
   b) mindestens die Hälfte
   c) mindestens 75 %

   von Prof. Lockerleicht geprüft werden ?

C

Da der Prüfungsausschuß zu der Auffassung gekommen ist, daß die Prüfungen bei Prof. Lockerleicht zu locker und zu leicht sind, wurde nun für den nächsten Prüfungszeitraum, in welchem 50 Kandidaten geprüft werden müssen, die Zusammensetzung der Lose in der Urne geändert. Nunmehr sind von den 100 Losen genau 95 Lose mit „ Ätzend “ beschriftet.

1) Wie ist die Zufallsvariable Y : „ Anzahl der für Prof. Lockerleicht anfallenden Prüfungen “ exakt verteilt ?
2) Durch welche andere Verteilung kann die Verteilung von Y approximiert werden ? Begründung !
3) Wie groß ist approximativ die Wahrscheinlichkeit, daß nun von Prof. Lockerleicht
   a) höchstens 20 % der Kandidaten geprüft werden ?
   b) mindestens 20 % der Kandidaten geprüft werden ?
   c) mindestens 4 % , aber höchstens 12 % der Kandidaten geprüft werden ?

Aufgabe G 24

Innerhalb der Jagdsaison geht Obelix an jedem Tag der Woche zur Wildschweinjagd auf eine Lichtung im Wald. Auf diese Lichtung kommen zufällig über den Tag verteilt Wildschweine, um dort zu fressen. Obelix weiß, daß sich – solange es hell ist (≅ zwölf Stunden) – insgesamt neun Stunden lang dort Wildschweine aufhalten. Obelix geht bei Helligkeit zu einem beliebigen Zeitpunkt auf die Jagd. Findet er auf der Lichtung keine Wildschweine vor, versteckt ber sich hinter einem Baum und wartet.

1) Wie groß ist die Wahrscheinlichkeit, daß sich zum Zeitpunkt seines Eintreffens auf der Lichtung Wildschweine dort aufhalten ?

Sei

X : „ Anzahl der Tage pro Woche, an denen Obelix bei seinem Erscheinen auf der Lichtung Wildschweine vorfindet "

2) Wie ist die Zufallsvariable X verteilt ? [Verteilungstyp und –parameter !]

3) An durchschnittlich wie vielen Tagen der Woche wird Obelix auf das Erscheinen von Wildschweinen warten müssen ?

4) Wie groß ist die Wahrscheinlichkeit, daß Obelix innerhalb einer Woche mindestens an drei, aber höchstens an sechs Tagen bei seinem Erscheinen bereits Wildschweine vorfindet ?

Obelix ist ein guter Jäger. Wenn er hinter einem Wildschwein herläuft, bekommt er es in 90 % aller Fälle zu fassen. Hat er ein Wildschwein erlegt, geht er zurück ins Dorf und beendet die Jagd für diesen Tag.

5) Wie groß ist die Wahrscheinlichkeit, daß für Obelix erst der dritte Versuch zum ersehnten Jagderfolg führt ?

Es kann vorkommen, daß ein erlegtes Wildschwein nicht schmeckt, weil es kurz vor seinem Ende eine bestimmte Sorte Waldbeeren gefressen hat. Da Obelix ein Jäger mit langjähriger Erfahrung ist, weiß er, daß dies in den letzten 2.000 Jagdtagen, an denen er jeweils ein Wildschwein erlegt hat, in 0,2 % aller Fälle vorgekommen ist.

Sei

W : „ Anzahl der erlegten Wildschweine einer Jagdsaison (≅ 200 Tage im Jahr), die nicht schmecken "

6) Wie ist die Zufallsvariable W verteilt ? [Verteilungstyp und –parameter !]

7) Berechnen Sie die Wahrscheinlichkeit, daß innerhalb einer Jagdsaison weniger als drei erlegte Wildschweine nicht schmecken !

Aufgabe G 25

In einer 5 – Tage – Woche fährt Jim Panse einmal täglich „ schwarz " (, d.h. ohne Fahrgeld zu entrichten) mit der U–Bahn. Dabei läuft er Gefahr, beim Verlassen des Bahnhofs kontrolliert zu werden. Die Wahrscheinlichkeit für eine solche Kontrolle beträgt (unabhängig von Fahrt zu Fahrt) 10 % .

1) Wie groß ist die Wahrscheinlichkeit, in einer beliebigen Woche

   a) am Freitag zum drittenmal kontrolliert zu werden ?

   b) bis zum Freitag (inkl.) dreimal kontrolliert zu werden ?

   c) am Montag kontrolliert zu werden ?

   d) am Dienstag zum zweitenmal kontrolliert zu werden ?

   e) gar nicht kontrolliert zu werden ?

   f) Donnerstag zum erstenmal kontrolliert zu werden ?

   g) nur Mittwoch und Donnerstag kontrolliert zu werden ?

2) Wie oft wird Jim Panse durchschnittlich pro Woche kontrolliert ?

3) Wie oft wird Jim Panse im Mittel fahren müssen, bis er zum fünftenmal kontrolliert wird ?

Aufgabe G 26

In einem Studienfach schreiben die Studenten am Ende des Semesters eine Klausur. Die dabei erreichten Leistungen ($\cong$ Noten) seien voneinander unabhängig.

Sei

$\mathbf{X} = (X_1 , X_2 , X_3 , X_4 , X_5)$

mit X_i : „ Anzahl der Studenten, die Note i bekommen “ $(i = 1 , 2 , 3 , 4 , 5)$

- Die Wahrscheinlichkeit, die Note „ 1 “ zu erhalten, beträgt 5 % .
- Die Wahrscheinlichkeit, die Note „ 2 “ zu erhalten, beträgt 20 % .
- Die Wahrscheinlichkeit, die Note „ 3 “ zu erhalten, beträgt 30 % .
- Die Wahrscheinlichkeit, die Klausur überhaupt zu bestehen, beträgt 70 % .

An der Klausur nehmen 30 Studenten teil.

1) Wie groß ist die Wahrscheinlichkeit, daß jede Note gleich oft vorkommt ?

2) Wie viele Studenten werden im Mittel durchfallen ?

Aufgabe G 27 :

In einer Lostrommel befinden sich **10** Lose : **4** *Nieten* und **6** *Gewinnlose*, wobei genau **2** dieser Gewinnlose sogar *Hauptgewinne* sind, d.h. auf 2 Gewinnlosen steht „ Hauptgewinn “.
Aus dieser Lostrommel sollen nun nacheinander (und ohne Zurücklegen) **zwei** Lose gezogen werden.

1) Wie groß ist die Wahrscheinlichkeit, daß das <u>zweite</u> gezogene Los ein Hauptgewinn ist ?

Sei

X : „ Anzahl der gezogenen *Gewinnlose* “

Y : „ Anzahl der gezogenen *Hauptgewinne* “

2) Bestimmen Sie tabellarisch die gemeinsame Wahrscheinlichkeitsverteilung von X und Y !

3) Berechnen Sie die Kovarianz von X und Y !

4) Sind X und Y stochastisch unabhängig ? Begründen Sie Ihre Antwort <u>formal</u> !

5) Angenommen, es werden genau **zwei** *Gewinnlose* gezogen.
Bestimmen Sie die Wahrscheinlichkeit,

a) keinen *Hauptgewinn*

b) einen *Hauptgewinn*

c) zwei Hauptgewinne

zu ziehen, d.h. bestimmen Sie **P (Y = y / X = 2)** !

6) Bestimmen Sie die <u>erwartete</u> Anzahl der gezogenen *Hauptgewinne*, falls genau **zwei** *Gewinnlose* gezogen werden sollten, d.h. bestimmen Sie **E (Y / X = 2)** !

Aufgabe G 28:

In einer großen Kiste liegen 3 verschiedene Sorten von Schrauben (1mm, 2mm, 3mm Durchmesser). Nehmen Sie an, daß die Schraubentypen in der Kiste einer Gleichverteilung folgen.

Sei $\mathbf{X} = (X_1, X_2, X_3)$ mit X_i: "Anzahl der Schrauben mit Durchmesser i" $i = 1,2,3$.

Es werden 3 Schrauben zufällig gezogen.

1) Stellen Sie alle möglichen Ausgänge (x_1, x_2, x_3) auf.

2) Berechnen Sie die Wahrscheinlichkeit, daß man

   a) nur 3mm Schrauben erhält.

   b) zwei 1mm und eine 3mm Schraube erhält.

   c) drei verschiedene Schraubentypen erhält.

Sei **Y:** "Anzahl der verschiedenen Schraubentypen, die man erhält".

3) Stellen Sie die Wahrscheinlichkeitsfunktion von Y auf.

4) Wieviele verschiedene Schraubentypen wird man erwarten. Berechnen Sie $E(Y)$.

Sei **Z:** "Anzahl der Schrauben mit einem Durchmesser von 2mm".

5) Geben Sie die Verteilung von Z an (Typ und Parameter).

6) Berechnen Sie $E(Z)$ und $Var(Z)$.

Sei **W:** "Anzahl der Schrauben mit einem Durchmesser von 3mm".

7) Geben Sie $Cov(W, Z)$ an.

Aufgabe G 29

Telefonverkäufer Jim wählt heute aus seiner sehr umfangreichen Kundenkartei zufällig 20 Kunden aus, um ihnen ein neues Produkt anzubieten. Aus Erfahrung weiß er, daß (unabhängig von Kunde zu Kunde) mit 10 %–iger Wahrscheinlichkeit ein Kunde ein solches Angebot annimmt.

A

Wie groß ist die Wahrscheinlichkeit, daß von den 20 Kunden

1) kein Kunde sich für das neue Produkt entscheiden wird ?

2) mehr als 3 Kunden sich für das neue Produkt entscheiden werden ?

3) höchstens 4 Kunden sich für das neue Produkt entscheiden werden ?

4) mehr als 2 , aber höchstens 6 Kunden sich für das neue Produkt entscheiden werden ?

Angenommen, Jim will heute seine Arbeit genau dann beenden, wenn er 4 Kunden das neue Produkt verkauft hat.

5) Wie groß ist die Wahrscheinlichkeit, daß Jim zur Erreichung dieses Ziels genau 9 Kunden anrufen muß ?

6) Wieviel Zeit wird Jim heute voraussichtlich für seine Arbeit benötigen, wenn jedes Telefonat mit einem Kunden 15 Minuten dauert ?

B

Das neue Produkt kostet 100 € .

Wie viele Kunden muß Jim mindestens anrufen, damit er bei einer 15 %–igen Umsatzprovision mit 50 € Umsatzprovision rechnen kann ?

Aufgabe G 30

In Bärlin sieht die Parkraumbewirtschaftungsordnung vor, daß für das Parken im Citybereich ein Parkschein erworben werden muß. Dieser kostet je angefangene Stunde 1 €. Wird durch einen Ordnungshüter festgestellt, daß für ein Fahrzeug kein gültiger Parkschein vorhanden ist, so muß eine einmalige Strafe in Höhe von 40 € gezahlt werden.
In diesem Citybereich hat eine Politesse Dienst, die unregelmäßig, aber im Durchschnitt alle 3 Stunden den Bereich kontrolliert und gegebenenfalls einen Strafzettel ausstellt.

Anton löst während seines 4–stündigen Aufenthaltes keinen Parkschein und hofft auf sein Glück.

Bert, der ebenfalls 4 Stunden dort parkt, versucht sein Risiko (auf einen Strafzettel) zu verringern, indem er nur für die ersten zwei Stunden einen Parkschein erwirbt.

1) Wie groß ist die Wahrscheinlichkeit, daß Anton einen Strafzettel erhält ?

2) Wie groß ist die Wahrscheinlichkeit, daß Bert einen Strafzettel erhält ?

3) Wie hoch sind die erwarteten Kosten für das 4–stündige Parken in diesem Bereich

   a) für Anton ?

   b) für Bert ?

4) Bei welchem Parkscheinpreis **p** je Stunde sind beide Vorgehensweisen (von den erwarteten Kosten) als gleichwertig anzusehen ?

Aufgabe G 31

Ein Katzenliebhaber hat **2** schwarze und **3** weiße Katzenbabys. Unter seinen **5** Katzenbabys sind **3** männliche Katzen (Kater). ***Bodo*** ist der einzige weiße Kater.
Nun möchte der Katzenliebhaber **3** Katzenbabys zufällig auswählen und seiner Freundin zum Geburtstag schenken.

Sei

X : „ Anzahl der ausgewählten schwarzen Katzenbabys "

Y : „ Anzahl der ausgewählten männlichen Katzenbabys (Kater) "

1) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsverteilung von X !
2) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsverteilung von Y !
3) Bestimmen Sie tabellarisch die gemeinsame Wahrscheinlichkeitsverteilung von X und Y !
4) Sind X und Y stochastisch unabhängig ?
   Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

Wie groß ist die Wahrscheinlichkeit, daß ***Bodo***

5) als einziges weißes Katzenbaby auf dem Geburtstagstisch der Freundin landet ?
6) als einziger Kater auf dem Geburtstagstisch der Freundin landet ?
7) als einziger weißer Kater auf dem Geburtstagstisch der Freundin landet ?

Aufgabe G 32

Richtig oder falsch ?

1) Der Erwartungswert einer poissonverteilten Zufallsvariablen kann keine negativen Werte annehmen.

2) Die Summe von n ($n > 1$) unabhängigen Bernoulli-verteilten Zufallsvariablen ist binomialverteilt mit n und π .

3) Eine $B(n;\pi)$ – verteilte Zufallsvariable hat $n+1$ Realisationsmöglichkeiten.

4) Die Zufallsvariable X : „ Anzahl der geworfenen Wappen beim gleichzeitigen Werfen von drei beliebigen Münzen " ist binomialverteilt.

5) Die Zufallsvariable X : „ Anzahl der gezogenen Kugeln beim gleichzeitigen Ziehen aus einer Urne mit roten und nicht–roten Kugeln " ist hypergeometrisch verteilt.

6) Verwendet man den „ Klassischen Wahrscheinlichkeitsbegriff nach Laplace " , so muß die Ergebnismenge des zugrundeliegenden Zufallsexperiments endlich sein.

7) Auf der Basis des Zufallsexperiments „ Werfen zweier idealer Würfel " wird die Zufallsvariable X : „ Augensumme beider Würfe " mit dem Wertebereich $\{ x \mid 2 \leq x \leq 12 \}$ definiert. Als Elementarereignisse bezeichnet man die einelementigen Mengen $\{2\}, \{3\}, \ldots, \{11\}, \{12\}$ dieses Wertebereichs.

8) In einer Urne sind $N = 10$ Kugeln, von denen $M = 6$ rot sind. Es werden $n = 5$ Kugeln zufällig und ohne Zurücklegen aus dieser Urne gezogen. Dann ist die Zufallsvariable X : „ Anzahl der gezogenen roten Kugeln " hypergeometrisch verteilt mit dem Wertebereich $\{0, 1, 2, 3, 4, 5\}$.

9) Die geometrische Verteilung ist ein Spezialfall der Binomialverteilung.

10) Es gilt : $\operatorname{Var}(X) = E[X \cdot (X-1)] + E(X) - [E(X)]^2$.

H

Stetige Zufallsvariablen und spezielle Verteilungsmodelle

Aufgabe H 1

A

Ein Taschenrechner wird mit zwei unterschiedlichen Batterien betrieben. Er arbeitet, solange beide Batterien funktionieren.
Die Lebensdauer X der Batterie A sei normalverteilt mit $\mu_X = 30$ Std. und $\sigma_X = 3$ Std.
Die Lebensdauer Y der Batterie B sei normalverteilt mit $\mu_Y = 35$ Std. und $\sigma_Y = 4$ Std.
X und Y seien unabhängige Zufallsvariablen.

1) Wie groß ist die Wahrscheinlichkeit, daß die Batterie A eine Lebensdauer zwischen 15 und 27 Stunden hat ?

2) Welche Lebensdauer wird bei der Batterie B mit einer Wahrscheinlichkeit von 85,3141 % nicht überschritten ?

3) Mit welcher Wahrscheinlichkeit arbeitet der Taschenrechner noch nach 24 Stunden ?

4) Wie groß ist die Wahrscheinlichkeit, daß die Lebensdauer der Batterie A geringer ist als die Lebensdauer der Batterie B ?

B

Bei einem elektronischen Bauteil kann man 48 Ausfälle pro Tag ($\cong$ 24 Std.) erwarten.
Die Ausfälle erfolgen rein zufällig und unabhängig voneinander.

1) Wie ist Y : „ Zeit in Stunden zwischen zwei Ausfällen " verteilt ?
[Verteilungstyp und –parameter !]

2) Wie groß ist die Wahrscheinlichkeit, daß bis zum nächsten Ausfall mehr als 2 Stunden vergehen ?

3) Verbalisieren Sie an diesem Beispiel folgende Formel : $\int_1^2 2 \cdot e^{-2 \cdot y}\, dy$!

4) Nehmen Sie an, daß ein elektronisches System aus zwei dieser Bauteile besteht, welche unabhängig voneinander funktionieren. Das System fällt aus, sobald ein Bauteil nicht mehr funktioniert.

Wie groß ist die Wahrscheinlichkeit, daß das System mehr als 2 Stunden funktioniert ?

Aufgabe H 2:

Das Ausborgen von Büchern ist vor allem in großen Bibliotheken mit Büchermagazin sehr zeitaufwendig. Also hat Student Bücherwurm die Arbeitsweise der Zentralbibliothek der TUB genau studiert, um besser planen zu können.

A: Die meisten Bücher der Zentralbibliothek kann man nicht sofort mitnehmen. Bücherwurm füllt also einen Leihschein aus und wirft ihn in den Bestellkasten. Er weiß, daß die Kästen regelmäßig zur vollen Stunde geleert und die Leihscheine ins Magazin gebracht werden.

1) Bücherwurm wirft den Leihschein zu einem zufällig ausgewählten Zeitpunkt, an dem er sich gerade im Hauptgebäude befindet, in den Kasten. Wie ist die Zufallsgröße X: “Wartezeit bis zur nächsten Leerung des Bestellkastens” verteilt ? (Verteilungstyp und –parameter !)

2) Wie lange wird ein Leihschein durchschnittlich im Kasten liegen, bis er ins Magazin wandert ?

B: Auch das Heraussuchen der Bücher aus dem Magazin dauert seine Zeit. Bücherwurm geht davon aus, daß die Zufallsgröße Y: “Wartezeit zwischen dem Leeren des Kastens und dem Ankommen des Buches in der Buchausgabe” normalverteilt ist mit einem Mittelwert von 2,5 Std. und einer Standardabweichung von o,5 Std. !

1) Wie groß ist die Wahrscheinlichkeit, daß die Bibliothek — von der Leerung des Bestellkastens ab gerechnet — länger als 3,5 Std. braucht, bis das Buch an der Ausgabe ankommt ?

2) Wie lange muß Bücherwurm an der Buchausgabe nach dem Einwerfen des Leihscheines durchschnittlich auf das Buch warten ?

C: In der Bibliotheksverwaltung weiß man, daß auch so manchem Entleiher Bücher abhanden kommen. So melden sich während der Öffnungszeiten (8 Stunden) an einem Tag durchschnittlich 6 Kunden mit einer Verlustanzeige.

1) Wie ist die Zufallsgröße Z: “Wartezeit auf den nächsten mit einer Verlustanzeige ankommenden Kunden” verteilt ? (Verteilungstyp und –parameter !)

2) Wieviele Minuten vergehen durchschnittlich, bis die erste Verlustanzeige von einem Kunden aufgegeben wird ?

3) Wie groß ist die Wahrscheinlichkeit, daß

   a) innerhalb der ersten drei Stunden, nachdem die Bibliothek geöffnet wurde, kein Kunde mit einer Verlustanzeige kommt ?

   b) nach höchstens vier Stunden der erste Verlust angezeigt wird, nachdem in den ersten drei Stunden kein Verlust gemeldet wurde ?

Aufgabe H 3

Eine Maschine produziert Stahlstifte. Leider ist der Durchmesser der Stifte produktionsbedingten Schwankungen unterworfen.

X_1 : „ Durchmesser eines Stiftes “ sei normalverteilt mit $\mu_1 = 6$ mm und $\sigma_1 = 0{,}4$ mm .

1) Wie groß ist die Wahrscheinlichkeit, daß der Durchmesser eines Stiftes um mehr als 2 % vom Sollwert 6 mm abweicht ?

2) Wie groß ist die Wahrscheinlichkeit, daß der Durchmesser eines Stiftes genau 6 mm beträgt ?

3) Welcher Wert des Durchmessers wird mit einer Wahrscheinlichkeit von 85 % nicht überschritten ?

Eine zweite Maschine, die unabhängig von der ersten Maschine arbeitet, bohrt Löcher in ein Werkstück, in die die Stahlstifte eingesetzt werden sollen. Auch der Durchmesser der Bohrlöcher ist Schwankungen unterworfen.

X_2 : „ Durchmesser eines Bohrloches “ sei normalverteilt mit $\mu_2 = 6{,}05$ mm und $\sigma_2 = 0{,}3$ mm .

4) Wie groß ist die Wahrscheinlichkeit, daß der Durchmesser eines Bohrloches kleiner als 6 mm ist ?

5) Wie ist die Zufallsvariable $Y = X_2 - X_1$ verteilt ? [Verteilungstyp und –parameter !]

6) Formalisieren Sie die Aussage : „ Stift paßt (in das Bohrloch) ! “

7) Wie groß ist die Wahrscheinlichkeit, daß ein Stift nicht in ein Bohrloch paßt ?

Aufgabe H 4

Die Telefonzentrale einer Feuerwache empfängt in einer Stunde durchschnittlich 0,5 Alarmmeldungen.

1) Wie groß ist die Wahrscheinlichkeit, daß während der 6–stündigen Dienstzeit einer Feuerwehrmannschaft

   a) kein Alarm

   b) mindestens dreimal Alarm

   c) höchstens siebenmal Alarm

   gegeben wird ?

2) Wie groß ist die Wahrscheinlichkeit, daß die Feuerwehrmannschaft

   a) innerhalb der ersten Dienststunde den ersten Alarm bekommt ?

   b) länger als 2 Stunden auf den ersten Alarm warten muß ?

   c) ausgerechnet in der letzten Dienststunde zum ersten Alarm „ ausrücken " muß, nachdem in der gesamten 5–stündigen Dienstzeit zuvor kein Alarm gekommen ist ?

3) Der diensthabende Oberbrandmeister erklärt seiner Feuerwehrmannschaft, daß mit 95 %–iger Wahrscheinlichkeit der erste Alarm noch in die Dienstzeit dieser Mannschaft fallen wird !

   Muß die Mannschaft tatsächlich weniger als 6 Stunden warten, um mit 95 %–iger Wahrscheinlichkeit den ersten Alarm zu bekommen ?

Aufgabe H 5

A

Bäcker Backfrisch vertreibt in seinem Laden auch hochfeine Marzipanschweine
Sei

X : „ Gewicht eines Marzipanschweines “

normalverteilt mit $E (X) = 150$ g und $Var (X) = 16\ g^2$.

Max, der davon ausgeht, daß das Gewicht der einzelnen Marzipanschweine unabhängig voneinander ist, geht in den Laden und kauft 4 Marzipanschweine.
Sei

Y : „ Gewicht von 4 Marzipanschweinen “

1) Wie ist die Zufallsvariable Y verteilt ? [Verteilungstyp und –parameter !]

2) Wie groß ist die Wahrscheinlichkeit, daß das Gewicht der 4 Marzipanschweine

   a) genau 600 g beträgt ?

   b) nicht mehr als 1 % vom Erwartungswert abweicht ?

B

Verkäuferin Mona weiß aus Erfahrung, daß während der Mittagszeit (13.00 – 15.00 Uhr) durchschnittlich nur alle 15 Minuten ein Kunde den Laden betritt.

1) Wie ist die Zufallsvariable

   Z_1 : „ Wartezeit bis zum Eintreffen des nächsten Kunden in der Mittagszeit “

   verteilt ? [Verteilungstyp und –parameter !]

2) Wie groß ist die Wahrscheinlichkeit, daß Mona in der Mittagszeit mehr als 30 Minuten auf den ersten Kunden warten muß ?

3) Mona wartet nun schon 30 Minuten vergeblich.

   Wie groß ist die Wahrscheinlichkeit, daß auch in den nächsten 15 Minuten kein Kunde den Laden betritt ?

C

Mona erhält regelmäßig alle 3 Stunden frische Ware. Sie hat heute leider ihre Uhr vergessen und bittet ihren Freund Leonardo um statistischen Rat.

1) Wie ist die Zufallsvariable

   Z_2 : „ Wartezeit auf die frische Ware “

   verteilt ? [Verteilungstyp und –parameter !]

2) Mona hat sich bereits eine Stunde angeregt mit Leonardo unterhalten, ohne daß frische Ware eingetroffen ist.

   Wie groß ist die Wahrscheinlichkeit, daß Mona innerhalb der nächsten halben Stunde frische Ware erhält ?

Aufgabe H 6

Student Erwin finanziert sein Studium, indem er Schreibarbeiten übernimmt.

A

Mitten bei seiner Schreibarbeit ist plötzlich die Farbbandkassette (Inhalt : 20 m Band) seiner Schreibmaschine leergeschrieben. Zum Glück besitzt sein Nachbar eine baugleiche Schreibmaschine mit gleichem Farbband. Da dieser Nachbar zur Zeit in Urlaub ist, entfernt Erwin von dieser Maschine die Farbbandkassette unbesehen, installiert diese in seiner eigenen Schreibmaschine und tippt den angefangenen Schreibauftrag weiter.

1) Wie groß ist die Wahrscheinlichkeit, daß die Farbbandkassette

   a) ganz voll ist ?

   b) ganz leer ist ?

   c) noch mindestens die bis zum Ende des Schreibauftrages notwendigen 20 cm Band enthält ?

B

Erwin tippt sehr korrekt. Auf 10 Zeilen ($\cong$ 1,5 m Farbband) geschriebenen Text rechnet er aus Erfahrung mit durchschnittlich nur $3/8$ Tippfehler.

1) Wie groß ist der zu erwartende Abstand zwischen zwei Tippfehlern (in cm) ?

2) Wie groß ist die Wahrscheinlichkeit, daß er bei den letzten 20 cm Farbband keinen Tippfehler macht ?

3) Wie groß ist die Wahrscheinlichkeit, daß die nächsten 4 zu schreibenden Zeilen fehlerfrei bleiben, falls er bei den letzten 4 geschriebenen Zeilen keinen Fehler gemacht haben sollte ?

Aufgabe H 7

A

Frank Furt ist neu in Berlin. Da seine finanziellen Mittel beschränkt sind, kann er sich kein Auto leisten und muß mit der BVG fahren. Dabei spielt er mit dem Gedanken, „ schwarz “ zu fahren (d.h. ohne Fahrgeld zu entrichten). Er muß im Monat 20 mal die U–Bahn benutzen. Wird man beim Schwarzfahren erwischt, muß man eine saftige Strafe bezahlen, zusätzlich einen Fahrschein nachlösen und kann dann die Fahrt fortsetzen.
Die Wahrscheinlichkeit, bei einer beliebigen Fahrt erwischt zu werden, sei 10 % .
Die Kontrollen erfolgen unabhängig voneinander.

1) Wie ist die Zufallsvariable „ Anzahl der Fahrten pro Monat, bei denen Frank erwischt wird “ verteilt ? [Verteilungstyp und –parameter !]

2) Wie groß ist die Wahrscheinlichkeit, daß Frank pro Monat
   a) genau 2 mal
   b) mehr als 4 mal
   c) gar nicht
   d) mindestens 3 mal, aber höchstens 5 mal

   erwischt wird ?

3) Wie oft wird Frank durchschnittlich pro Monat erwischt ?

Sollte Frank erwischt werden, muß er eine Strafe von 60 € bezahlen. Zusätzlich muß er natürlich einen Fahrschein kaufen, der 4 € kostet. Da Frank möglichst billig fahren möchte, sucht er nach einer geeigneten Strategie. Dazu vergleicht er die monatlichen Fixkosten, falls er nie schwarzfährt, mit den zu erwartenden Kosten, falls er immer schwarzfährt.

4) Führen Sie diesen Vergleich durch ! Welche Strategie sollte man Frank (demnach) vorschlagen ?

5) Wie groß müßte die Wahrscheinlichkeit (für das „ Erwischtwerden “) sein, so daß es egal wäre, ob Frank immer einen Fahrschein kauft bzw. immer schwarzfährt ?

B

Der Bruder von Frank hat ähnliche Probleme, muß aber im Monat 60 mal die U–Bahn benutzen. Auch hier möge die Wahrscheinlichkeit, bei einer beliebigen Fahrt erwischt zu werden, 10 % betragen, wobei die Kontrollen unabhängig voneinander erfolgen.

1) Wie ist die Zufallsvariable „ Anzahl der Fahrten pro Monat, bei denen Frank´s Bruder erwischt wird “ verteilt ? [Verteilungstyp und –parameter !]

2) Wie groß ist <u>approximativ</u> die Wahrscheinlichkeit, daß Frank´s Bruder pro Monat
   a) mehr als 2 mal
   b) höchstens 5 mal
   c) gar nicht
   d) mindestens 3 mal, aber weniger als 5 mal

   erwischt wird ?

3) Ab welcher Höhe des Fahrpreises wird er sich dafür entscheiden, immer schwarz zu fahren ?

Aufgabe H 8:

Emil Langfinger hat sich auf das Stehlen von Autoradios spezialisiert. Er weiß aus Erfahrung, daß sich in einem Auto mit 5o%–iger Wahrscheinlichkeit ein Autoradio befindet. Langfinger geht nachts in ein Parkhaus, in dem sich 1oo Autos befinden (Zufallsstichprobe).

1) Wie ist die Zufallsgröße X: "Anzahl der Autoradios im Parkhaus" verteilt ? (Verteilungstyp und –parameter!)

2) Geben Sie die Formel an, mit der man die Wahrscheinlichkeit dafür berechnen kann, daß sich höchstens 6o Autoradios im Parkhaus befinden !

3) Berechnen Sie approximativ die Wahrscheinlichkeit dafür, daß sich höchstens 6o Autoradios im Parkhaus befinden !

4) Berechnen Sie approximativ die Wahrscheinlichkeit dafür, daß sich mindestens 4o Autoradios im Parkhaus befinden !

Langfinger weiß, daß ein Wachmann in genau einstündigem Abstand in das Parkhaus kommt, um einen Kontrollgang zu machen. Langfinger geht nachts zu einem beliebigen Zeitpunkt in das Parkhaus.

5) Wie ist die Zufallsgröße Y: "Wartezeit bis zum Auftauchen des Wachmannes" verteilt ? (Verteilungstyp und –parameter!)

6) Wie lange wird Langfinger durchschnittlich auf das Erscheinen des Wachmannes warten müssen ?

7) Wie groß ist die Wahrscheinlichkeit dafür, daß Langfinger weniger als 2o Minuten auf das Erscheinen des Wachmannes warten muß ?

8) Langfinger hat nun schon 3o Minuten vergeblich auf das Erscheinen des Wachmannes gewartet. Wie groß ist die Wahrscheinlichkeit dafür, daß der Wachmann innerhalb der nächsten 2o Minuten auftaucht ?

Um es den Leuten wie Langfinger schwerer zu machen, beschließt der Geschäftsführer des Parkhauses, daß die Kontrollgänge nicht mehr in exakt einstündigem Abstand vonstatten gehen sollen. In Zukunft hat der Wachmann die Anweisung, seine Kontrollgänge unabhängig voneinander in durchschnittlich einstündigem Abstand durchzuführen.

9) Wie ist die Zufallsgröße Y: "Wartezeit bis zum Auftauchen des Wachmannes" bei der zukünftigen Regelung verteilt ?
(Verteilungstyp und –parameter!)

Aufgabe H 9:

Willi möchte heute in seinem Stammlokal gemütlich drei Stunden verbringen.

A: Da er keinen Parkplatz findet, parkt er kurz entschlossen in einer Halteverbotszone, was bekanntermaßen verboten ist und von der Polizei mit einem Bußgeld bestraft wird. Nach Aussage des Wirtes kommt regelmäßig alle 4 Stunden eine Polizeistreife vorbei, die den falsch geparkten Autos einen Strafzettel unter den Scheibenwischer klemmt.

1) Wie ist die Zufallsgröße X: "Wartezeit bis zum Eintreffen der Polizeistreife" verteilt ?
(Verteilungstyp und –parameter !)

2) Wie groß ist die Wahrscheinlichkeit, daß Willi Glück hat und keinen Strafzettel bekommt ?

3) Nach 2 Stunden verläßt Willi kurz das Lokal, um nachzusehen, ob schon ein Strafzettel unter dem Scheibenwischer seines Autos klemmt. Da dies nicht der Fall ist, geht er beruhigt zurück in das Lokal.
Wie groß ist die Wahrscheinlichkeit, daß die Polizei in der verbleibenden letzten Stunde, die Willi noch im Lokal verbringt, doch noch einen Strafzettel unter den Scheibenwischer von Willis Auto klemmt ?

B: Willis Lokalbesuch fällt heute genau in die 1o–stündige Dienstzeit der Kellnerin Susi. Sie ist bekannt dafür, daß sie trotz aller Hektik, die im Lokal herrscht, durchschnittlich nur ein Glas (und nie mehrere Gläser gleichzeitig) in ihrer Dienstzeit fallen läßt, wobei Mißgeschicke dieser Art unabhängig voneinander geschehen.
Wie groß ist die Wahrscheinlichkeit, daß während Willis Aufenthalt

1) Susi bereits in der ersten Stunde 2 Gläser fallen läßt ?

2) Susi erst nach zweieinhalb Stunden ein Glas aus der Hand fällt ?

3) kein Glas zu Bruch geht ?

Willi ist schon 2 Stunden im Lokal, ohne daß es Scherben gegeben hat.

4) Wie groß ist die Wahrscheinlichkeit, daß Susi in der letzten Stunde von Willis Anwesenheit im Lokal doch noch ein Glas fallen läßt ?

C: Der Wirt des Lokals will sich heute einen Spaß mit seinen Gästen machen und hat 5 von den 2o frisch zubereiteten Frikadellen mit Tabasco (einem besonders scharfen Gewürz) gefüllt. Willi hat sowohl Mut als auch Hunger auf 3 Frikadellen, die er sich selbst aussuchen darf.

1) Wie groß ist die Wahrscheinlichkeit, daß Willi kein Faß Bier bestellen muß (keine Frikadelle mit Tabasco ißt) ?

2) Willi hat schon 2 Frikadellen mit Genuß (und ohne Tabasco) verspeist.
Wie groß ist die Wahrscheinlichkeit, daß auch in der letzten von ihm ausgewählten Frikadelle kein Tabasco ist ?

Aufgabe H 10

Um die Erfolgschancen einer neuen Zeitschrift „ Ultra " zu überprüfen, läßt ein Verlag zunächst nur einer kleinen Zahl von 30 zufällig ausgewählten Personen per Post einen Werbecoupon zukommen. Wird der Coupon an den Verlag zurückgeschickt, erhält man ein kostenloses Probeexemplar der neuen Zeitschrift zugesandt. Seitens des Verlages ist man eher skeptisch und glaubt, daß 80 % der Coupons nicht an den Verlag zurückgeschickt werden.

1) Wie groß ist die Wahrscheinlichkeit, daß die dritte angeschriebene Person die erste ist, die den Coupon an den Verlag zurückschickt ?

2) Wie groß ist die Wahrscheinlichkeit, daß

   a) an mehr als 10 der angeschriebenen Personen ein Probeexemplar geschickt werden kann ?

   b) an mehr als 4 , aber weniger als 10 Personen ein Probeexemplar geschickt werden kann ?

   c) mehr als 22 der angeschriebenen Personen kein Probeexemplar erhalten werden ?

Aufgrund dieser Überprüfung ist man beim Verlag überzeugt, daß der Anteil der Personen, die solche Werbecoupons zurückschicken, tatsächlich nur 20 % beträgt. Für die eigentliche Werbeaktion in größerem Rahmen wählt der Verlag nunmehr 10.000 Personen zufällig aus und schickt diesen Personen jeweils einen Werbecoupon.

3) Wie groß ist approximativ die Wahrscheinlichkeit, daß der Verlag bei dieser Aktion

   a) höchstens 1.920 Probeexemplare verschicken kann ?

   b) zwischen 1.960 und 2.040 Probeexemplare verschicken kann ?

   c) weniger als 1.950 oder mehr als 2.070 Probeexemplare verschicken kann ?

Der Verlag will dem Chefredakteur der neuen Zeitschrift 4 Assistenten zur Seite stellen, die sich dieser zufällig per Losverfahren aus 10 Personen auswählen darf. Leider haben nur 6 dieser 10 Personen einschlägige Erfahrungen in der Zeitungsbranche.

4) Wie groß ist die Wahrscheinlichkeit, daß

   a) keiner

   b) zwei

   der ausgewählten Assistenten Erfahrungen in der Zeitungsbranche haben ?

Aufgabe H 11:

Die lieben Verwandten haben ihren alljährlichen Besuch bei der Familie Furchtsam angekündigt. Die Besuchsdauer wird 12 Tage betragen.

A: Mutter Furchtsam sieht dem Besuch mit Sorge entgegen. Die Verwandten zertrümmern nämlich unabhänig von Tag zu Tag täglich mit einer Wahrscheinlichkeit von 30% Geschirr.

1) Wie groß ist die Wahrscheinlichkeit, daß während der Besuchsdauer an genau 3 Tagen Geschirr kaputtgeht ?

Falls Cousin Kurt auch anreisen sollte, erhöht sich die tägliche "Trümmerwahrscheinlichkeit"sogar auf 60%.

2) Mit welcher Wahrscheinlichkeit gibt es jetzt öfter als 3 Tage Scherben ?

B: Tochter Ursel hat ein anderes Problem. Der zu Besuch weilende liebe Onkel weckt sie nämlich durchschnittlich viermal während der achtstündigen Nachtruhe durch kurze und laute, jedoch völlig unregelmäßig ausgestoßene Schreie.

1) Wie groß ist die Wahrscheinlichkeit, daß Ursel an einem Samstagmorgen zwischen der Heimkehr aus der Disco um 5 Uhr und dem Klingeln des Weckers um 8 Uhr durch das Geschrei des Onkels im Schlaf gestört wird ?
2) Wie wahrscheinlich ist es, daß der Onkel in 9 Nächten (á 8 Stunden) mehr als 24 mal schreit ?

C: Während der Besuchsdauer wird traditionsgemäß an jedem Abend Oligopoly gespielt. Die Wahrscheinlichkeit für einen Streit zwischen den ehrgeizigen Familienvätern beträgt allabendlich 20%, wobei man sich von Abend zu Abend unabhängig voneinander streitet.

1) Am wievielten Abend hat man zum erstenmal Streit zu erwarten ?
2) Wie wahrscheinlich ist es, daß es am 4. oder 5. Abend Streit gibt, wenn es an den ersten drei Abenden keinen Streit gab ?

D: Für das Kaffeekränzchen muß Mutter Furchtsam neben den 6 normalen auch 2 wertvolle altchinesiche Tassen herausgeben.

1) Mit welcher Wahrscheinlichkeit werden die beiden China–Tassen heil bleiben, falls beim Kaffeekränzchen drei Tassen zu Bruch gehen sollten ?

Aufgabe H 12:

A: In einer Autofabrik werden Karosserieteile von einem Industrie–Roboter geschweißt. Der Verlauf der Schweißnaht wird von elektronischen Fühlern gesteuert, die unabhängig voneinander mit einer Wahrscheinlichkeit von 9o% einwandfrei arbeiten. Am Arm dieses Roboters sind 4 derartige Fühler angebracht.

1) Wie ist die Zufallsgröße X: "Anzahl der nicht einwandfrei arbeitenden Fühler am Roboterarm" verteilt ?
(Verteilungstyp und –parameter!)

2) Wieviele Fühler arbeiten im Mittel nicht einwandfrei ?

3) Wie groß ist die Wahrscheinlichkeit, daß der Verlauf der Schweißnaht korrekt gesteuert wird, falls dazu notwendig ist, daß

   a) alle 4 Fühler einwandfrei arbeiten ?

   b) mindestens 2 Fühler einwandfrei arbeiten ?

   c) 2 bestimmte Fühler einwandfrei arbeiten ?

B: Zur Lackierung der Autokarosserie werden 2 Maschinen eingesetzt, die unabhängig voneinander durchschnittlich alle 1o Stunden einen Defekt haben. Ein Arbeiter beginnt seinen Schichtdienst und stellt fest, daß beide Maschinen einwandfrei arbeiten.

Sei T_1 : "Wartezeit auf den nächsten Defekt von Maschine 1"
T_2 : "Wartezeit auf den nächsten Defekt von Maschine 2"

1) Wie sind die Zufallsgrößen T_1 und T_2 verteilt ?
(Verteilungstyp und –parameter!)

2) Wie groß ist die Wahrscheinlichkeit, daß der Arbeiter während seines 8 Stunden dauernden Schichtdienstes keine Unterbrechung des Lackierungsprozesses erlebt, falls

   a) der Prozeß nur dann weiterlaufen kann, wenn beide Maschinen funktionieren ?

   b) der Prozeß nur dann weiterlaufen kann, solange mindestens eine Maschine funktioniert ?

   (Gehen Sie davon aus, daß mit der Behebung eines Defekts bis zum Schichtende gewartet wird !)

3) Es sind bereits 6 Stunden der Schicht vergangen, ohne daß bei einer der beiden Maschinen ein Defekt aufgetreten ist.
Wie groß ist die Wahrscheinlichkeit, daß die Maschine 1 bis zum Schichtende nicht defekt wird ?

Aufgabe H 13

Alfred wohnt in C. und will Berti in D. besuchen. Dazu muß er mit dem Bus zum Hafen fahren. Die Bushaltestelle ist genau vor seiner Haustür. Alfred weiß, daß der Bus nach Fahrplan fährt, aber nicht zu welcher Zeit. Er ist sich aber sicher, daß die Wahrscheinlichkeit, weniger als 5 Minuten zu warten, $^1/_3$ beträgt. Die Busfahrt zum Hafen dauert 30 Minuten. Dort muß er auf die Fähre warten, die leider nach keinem Fahrplan fährt. Alfred weiß aus Erfahrung, daß die Wartezeit auf die Fähre normalverteilt ist mit $\mu = 0{,}5$ Stunden und $\sigma^2 = 0{,}25$ Stunden ^{2}. Die Überfahrt dauert 5 Stunden. Im Hafen von D. angekommen muß Alfred ein Taxi nehmen, um zum Haus von Berti zu gelangen. Das Taxiwesen in D. ist als völlig unstrukturiert bekannt ; Taxistände sind unbekannt und ein abgestimmtes Verhalten der Taxifahrer untereinander ist keineswegs zu erwarten. Also muß man in den engen, einspurigen Hafengassen ein zufällig vorbeifahrendes Taxi durch heftiges Winken zum Halten bringen. Nach Auskunft von Berti kann man mit durchschnittlich 24 Taxis innerhalb von 4 Stunden rechnen. Wenn Alfred bei Berti ankommt, wird er sich ärgern, daß der Fahrpreis so hoch ist (1 € pro Minute), d.h. daß er 50 € bezahlen muß.

Folgende Zufallsvariablen seien definiert :

X : „ Wartezeit auf den Bus “

Y : „ Wartezeit auf die Fähre “

W : „ Anzahl der vorbeifahrenden Taxis pro Stunde “

Z : „ Wartezeit auf ein Taxi in Stunden “

R : „ Reisezeit von Haus zu Haus “

1) Geben Sie für die Zufallsvariablen X , Y , W und Z jeweils den Verteilungstyp und die Verteilungsparameter an !

2) Wie groß ist die erwartete Reisezeit von Haus zu Haus ?

3) Alfred steht schon 10 Minuten an der Bushaltestelle.
   Wie groß ist die Wahrscheinlichkeit, daß er weniger als noch 2 Minuten auf den Bus warten muß ?

4) Alfred steht schon 10 Minuten am D.–Hafen und wartet auf ein Taxi.
   Wie groß ist die Wahrscheinlichkeit, daß er weniger als noch 2 Minuten auf ein Taxi warten muß ?

5) Angenommen, Alfred würde sich am D.–Hafen von Berti abholen lassen wollen.
   Alfred würde um 8^{00} Uhr in C.–Hafen ankommen und Berti anrufen.
   Für welche Uhrzeit müßte er Berti zum D.–Hafen bestellen, damit Berti mit einer Wahrscheinlichkeit von **93,9429 %** nicht auf ihn warten müßte ?

Aufgabe H 14:

In Berlin soll eine Umfrage über die Zustimmung der Bevölkerung zu Olympia 2.000 durchgeführt werden. Das Institut “Info” wird beauftragt, Fragebögen zu entwickeln und zu verteilen. Der Chef von “Info” betraut zwei seiner Mitarbeiter (Xaver und Yvonne) mit der Verteilung der Fragebögen : nachmittags, wenn auf dem Ku–Damm dichtes Gedränge herrscht, sollen in einer bestimmten Zeitspanne jeweils soviele Fragebögen wie möglich verteilt werden. Da sich Xaver und Yvonne bei ihrer Arbeit nicht überanstrengen möchten, wählen sie folgende unterschiedliche Vorgehensweisen:
Xaver sucht sich regelmäßig alle 10 Minuten eine Person heraus, der er einen Fragebogen in die Hand drückt.
Yvonne dagegen verteilt pro Stunde durchschnittlich 6 Bögen in unregelmäßigen Abständen.

1) Wie ist die Zufallsgröße
   X: “Wartezeit bis zur Verteilung eines Fragebogens durch Xaver” verteilt ?
   (Verteilungstyp und –parameter !)

2) Wie ist die Zufallsgröße
   Y: “Wartezeit bis zur Verteilung eines Fragebogens durch Yvonne” verteilt ?
   (Verteilungstyp und –parameter !)

Der Chef kommt nachmittags zu einem beliebigen Zeitpunkt zum Ku–Damm, um seine Mitarbeiter heimlich bei der Arbeit zu beobachten.

3) Wie lange wird er durchschnittlich warten müssen, bis er die Verteilung eines Fragebogens durch Xaver beobachten kann ?

4) Wie groß ist die Wahrscheinlichkeit, daß er nicht länger als 5 Minuten warten muß, bis er die Verteilung eines Fragebogens

   a) durch Xaver

   a) durch Yvonne

   beobachten kann ?

5) Der Chef wartet schon 10 Minuten, ohne daß Yvonne in dieser Zeit einen Fragebogen verteilt hat. Wie groß ist die Wahrscheinlichkeit, daß er nicht mehr als weitere 10 Minuten auf die Verteilung eines Fragebogens von Yvonne warten muß ?

Der Chef beschließt, einen weiteren Mitarbeiter (Zorro) zur Unterstützung von Xaver und Yvonne zum Ku–Damm zu schicken. Da Zorro jedoch Hunde und Frauen liebt, verteilt er Fragebögen ausschließlich an Frauen in Begleitung eines Hundes. Man kann davon ausgehen, daß nachmittags auf dem Ku–Damm die Wartezeit auf eine Frau in Begleitung eines Hundes normalverteilt ist ($\mu = 12$; $\sigma = 4$ Minuten).

6) Wie groß ist die Wahrscheinlichkeit, daß Zorro weniger als 8 Minuten warten muß, bis eine Frau mit Hund vorbeikommt ?

7) Zorro möchte in 10 Minuten Feierabend machen, aber noch seinen letzten Fragebogen loswerden. Wie groß ist die Wahrscheinlichkeit, daß er seinen letzten Fragebogen mit nach Hause nehmen muß ?

Aufgabe H 15 :

In einer Fabrik gibt es zwei Maschinen.

A : Maschine A produziert Spezialfässer für hochgiftige Chemikalien. Der Durchmesser eines Fasses in cm (X_A) sei eine normalverteilte Zufallsvariable :

$$X_A \text{ ist } N(\mu_A ; \sigma^2_A) \sim N(50 ; 4)$$

Aufgrund der hochgiftigen Chemikalien werden die Fässer beim Transport an speziellen Halterungen befestigt, die eine Abweichung von nur ± 1 cm vom Normdurchmesser zulassen.

1) Wie groß ist die Wahrscheinlichkeit, daß ein Faß in eine solche Halterung paßt ?
2) Wie groß ist die Wahrscheinlichkeit, daß ein Faß genau den Normdurchmesser besitzt ?
3) Wie groß ist die Wahrscheinlichkeit, daß der Durchmesser eines Fasses kleiner als 49 cm ist ?

B : Maschine B stellt unabhängig von Maschine A Deckel für diese Fässer her. Der Durchmesser eines Deckels in cm (X_B) sei eine normalverteilte Zufallsvariable :

$$X_B \text{ ist } N(\mu_B ; \sigma^2_B) \sim N(53 ; 5)$$

Damit die Fässer luftdicht verschlossen sind, darf der Durchmesser eines Deckels nicht kleiner sein als der Durchmesser eines Fasses.

1) Wie groß ist die Wahrscheinlichkeit, daß aus einem verschlossenen Faß giftige Gase ausströmen können ?
2) Wie groß ist die Wahrscheinlichkeit, daß ein Faß und sein Deckel den gleichen Durchmesser besitzen ?

C : Damit ein Deckel möglichst zu einem Faß paßt, muß jedes Faß mit seinem Deckel zwei Kontrollen durchlaufen. Die Kontrollzeit in Minuten von Kontrolleur K_1 für ein Faß mit Deckel (Y_1) sei eine stetig gleichverteilte Zufallsvariable :

$$Y_1 \text{ ist stetig gleichverteilt in } [a ; b] \sim [1 ; 7]$$

Nach Beendigung der Kontrolle durch K_1 erfolgt davon unabhängig eine zweite Kontrolle durch Kontrolleur K_2. Die Kontrollzeit in Minuten von Kontrolleur K_2 für ein Faß mit Deckel (Y_2) sei eine exponentialverteilte Zufallsvariable :

$$Y_2 \text{ ist } E(\lambda) \sim E(1)$$

1) Bestimmen Sie den Erwartungswert und die Varianz für die **gesamte** Kontrollzeit für ein Faß (mit Deckel) !
2) Wie groß ist die Wahrscheinlichkeit, daß die Gesamtkontrolle für ein Faß mindestens 2 Minuten, aber höchstens 8 Minuten dauert ?

Pro Tag werden 49 Fässer mit Deckel hergestellt und durchlaufen unabhängig voneinander den Kontrollprozeß.

3) Wie groß ist die Wahrscheinlichkeit, daß die Gesamtkontrolle für eine solche Tagesproduktion länger als 3 Stunden und 37 Minuten dauert ?

Aufgabe H 16

In einem Restaurant herrschen merkwürdige Zustände. Es gibt nur einen Koch und nur einen Kellner. Der Kellner kommt regelmäßig alle 15 Minuten an den Tisch, um die Wünsche der Gäste entgegenzunehmen. Er notiert den Essenswunsch jedes einzelnen Gastes auf einem gesonderten Zettel. Ist die Bestellung aufgenommen, wirft der Kellner diese Zettel in eine Kiste. Der Koch zieht dann zufällig einen Zettel aus der Kiste heraus, kocht und serviert dann dieses Essen. Die bestellten Essen kommen in vollkommen unregelmäßigen Abständen (durchschnittlich 15 Minuten nach der Bestellung) beim Gast an.

Sei

X : „Wartezeit bis zum Erscheinen des Kellners"

Y : „Anzahl der pro Minute servierten Essen"

Z : „Wartezeit bis das Essen serviert wird (≅ Zeitraum zwischen Bestellung und Servierung)"

1) Wie ist die Zufallsvariable X verteilt ? [Verteilungstyp und –parameter !]

2) Wie ist die Zufallsvariable Y verteilt ? [Verteilungstyp und –parameter !]

3) Wie ist die Zufallsvariable Z verteilt ? [Verteilungstyp und –parameter !]

4) Wie lange wird man im Durchschnitt auf das Erscheinen des Kellners warten müssen ?

5) Angenommen, man wartet schon 10 Minuten vergeblich auf den Kellner.
Wie groß ist die Wahrscheinlichkeit, daß man innerhalb der nächsten 2 Minuten endlich sein Essen bestellen kann ?

6) Angenommen, die Bestellung ist endlich erfolgt.
Wie groß ist die Wahrscheinlichkeit, daß nach genau 15 Minuten das Essen serviert wird ?

7) Angenommen, die Bestellung ist endlich erfolgt.
Wie groß ist die Wahrscheinlichkeit, daß man weniger als 15 Minuten auf sein Essen warten muß ?

8) Angenommen, man hat sein Essen bestellt und wartet bereits länger als 15 Minuten vergeblich auf das Essen.
Wie groß ist die Wahrscheinlichkeit, nach weiteren 5 Minuten immer noch hungrig am Tisch zu sitzen ?

9) Genau 30 Minuten nachdem Sie das Restaurant betreten und am Tisch Platz genommen haben, sind Sie mit Ihren Eltern vor dem Restaurant verabredet.
Wie groß ist die Wahrscheinlichkeit, daß Ihre Eltern nicht auf Sie warten müssen, wenn der Kellner nach 5 Minuten die Bestellung aufgenommen hat und Sie 5 Minuten für den Verzehr Ihres Essens benötigen ?

10) Angenommen, Sie wären mit Ihren Eltern in dieses Restaurant zum Essen gegangen.
Wieviel Zeit wird durchschnittlich vergehen, bis das Essen bestellt und **alle** ihr Essen serviert bekommen hätten ?

11) Wie ändert sich das Ergebnis unter 10) , wenn die „Wartezeit bis das Essen serviert wird" normalverteilt wäre mit $\mu = 15$ Minuten und $\sigma^2 = 25$ Minuten ?

Aufgabe H 17

A

Student Statistix fährt jeden Tag in etwa zur gleichen Zeit mit dem Bus oder einem Taxi zur Universität. Wegen seiner bevorstehenden Statistikklausur ist er heute derart aufgeregt, daß er seine Uhr nicht mitgenommen hat. An der Bushaltestelle angekommen fällt ihm ein, daß der Bus regelmäßig alle 20 Minuten fährt. Taxis hingegen fahren völlig regellos an dieser Stelle vorbei, zu dieser Tageszeit jedoch beträgt die durchschnittliche Wartezeit auf ein solches Ereignis ungefähr 15 Minuten. Weiterhin kann man davon ausgehen, daß Taxis und Busse sich in ihrem Fahrverhalten nicht beeinflussen.

1) Wie groß ist die Wahrscheinlichkeit, daß Statistix heute genau 13 Minuten auf den Bus warten muß ?

2) Wie groß ist die Wahrscheinlichkeit, daß Statistix heute noch höchstens 5 Minuten auf eine Fahrgelegenheit warten muß ?

3) Wie groß müßte die durchschnittliche Anzahl der vorbeifahrenden Taxis sein, wenn die Wahrscheinlichkeit, mehr als 12 Minuten auf ein Taxi zu warten, 25 % betragen würde ?

4) Wie groß ist die Wahrscheinlichkeit dafür, daß spätestens 15 Minuten nach Statistix Ankunft an der Bushaltestelle der Bus kommt und in diesen 15 Minuten kein einziges Taxi vorbeifährt ?

5) In welchem Zeitabstand müßte der Bus fahren, wenn die Wahrscheinlichkeit, mehr als 4 Minuten auf den Bus zu warten, genau 60 % beträgt ?

Angenommen, die Wartezeit auf ein an dieser Stelle vorbeifahrendes Taxi ist normalverteilt mit einem Erwartungswert von 18 Minuten bei einer Standardabweichung von 2 Minuten.

6) Wie groß ist dann die Wahrscheinlichkeit, mehr als 16 Minuten, aber weniger als 24 Minuten auf ein Taxi warten zu müssen ?

B

Die Zufallsvariable X sei stetig gleichverteilt in $[\,a\,;\,b\,]$

Zeigen Sie mit Hilfe eines formalen Beweisganges, daß gilt : $E(X) = \dfrac{a+b}{2}$!

I

Auswahlverfahren, Grenzwertsätze und Testverteilungen

Aufgabe I 1:

In der Statistik unterscheidet man zwischen zufälliger und nicht–zufälliger Auswahl von Elementen aus einer Grundgesamtheit.

1) Erläutern Sie, was man unter einer Zufallsstichprobe versteht !

2) Erläutern Sie, was man unter einer uneingeschränkten Zufallsstichprobe versteht !

3) Erläutern Sie, was man unter einer einfachen Zufallsstichprobe versteht !

4) Geben Sie zwei Auswahltechniken an, bei denen die Auswahl nicht–zufällig erfolgt !

Aufgabe I 2:

Erläutern Sie den Zusammenhang zwischen den Begriffen:

— Grundgesamtheit — Stichprobenvariable — Stichprobenergebnis
— Stichprobenfunktion !

Aufgabe I 3:

Im Jahre 5o v. Chr. steht täglich der Legionär Kugelrund auf einem Feld vor dem befestigten Römerlager Wache. Jeden Vormittag kommen die beiden Gallier Asterix und Obelix einmal am Lager vorbei. In diesem Augenblick beginnt sich das Lagertor zu schließen, was genau 18 sek. dauert; danach kommt niemand mehr in das Lager hinein. Legionär Kugelrund läuft beim Erscheinen der Gallier sofort in Richtung Lagertor. Er benötigt zum Erreichen des Lagertores durchschnittlich 14 sek. bei einer Varianz von 4 sek^2. Erreicht Kugelrund das Lager nicht vor Toresschluß, so muß er draußenbleiben, wird von den Galliern verprügelt und muß anschließend ins Lazarett. Läuft er vor lauter Angst aber so schnell, das das Tor sich erst später als 8 sek. nach seiner Ankunft im Lager vollständig geschlossen hat, dann ist er total erschöpft und verschläft den Rest des Tages.

Wie groß ist die Wahrscheinlichkeit dafür, daß Kugelrund nach dem nächsten Auftauchen von Asterix und Obelix wohlbehalten mit seinem Freund im Lager Karten spielen kann ?

Aufgabe I 4:

Sei X: "Körpergewicht eines Koalbären in kg" mit $E(X) = 10$ und $Var(X) = 0{,}25$.

Sei Y: "Tragfähigkeit eines Astes in kg, den ein Koalabär besteigt" mit $E(Y) = 12$ und $Var(Y) = 0{,}75$.

Ist das Körpergewicht größer als die Tragfähigkeit des bestiegenen Astes, so fällt der Koalabär zu Boden und bricht sich das Genick. Eine weitere Gefahr geht von einer bestimmten Art von Giftschlange aus, die auf den Ast schleicht und den Bären tötet. Die Schlange schleicht genau dann auf den Ast, falls die Tragfähigkeit es zuläßt. Das ist genau dann der Fall, wenn die Tragfähigkeit des Astes das Körpergewicht des Koalabären um mehr als 4 kg übersteigt. Koalabären suchen sich stets unabhängig von ihrem Körpergewicht einen Ast zum Schlafen.

1) Wie groß ist die Wahrscheinlichkeit, daß ein Koalabär ungestört auf einem Ast ausschlafen kann ?

2) Warum ist die unter 1) formulierte Fragestellung mit den gegebenen Angaben nicht lösbar, falls sich ein Koalabär bei der Suche nach einem Ast zum Ausschlafen irgendwie an seinem Körpergewicht orientieren würde ?

Aufgabe I 5:

Fritz Flink ist Leichtathlet und steht vor einem Qualifikationswettkampf im 4oo m–Lauf. Aufgrund seines Trainingszustandes ist der Erwartungswert für seine Laufzeit 5o Sekunden mit einer Varianz von 3 Sekunden^2. Sepp Sprint, sein Vereinskollege und Wettkampfrivale, kann aufgrund einer Verletzung nicht starten. Er hält den alten Bezirksrekord von 46,54 Sekunden, den er nicht an Fritz verlieren möchte. Bei einer Zeit von über 53,46 Sekunden wäre Fritz jedoch nicht für die Landesmeisterschaft qualifiziert, was Sepp im Hinblick auf das Ansehen des Vereins peinlich wäre.

1) Wie groß ist die Wahrscheinlichkeit, daß keine von Sepps Befürchtungen eintritt ?

2) Wie groß ist die Wahrscheinlichkeit, daß mindestens eine von Sepps Befürchtungen eintritt, wenn man Sepps Trainer glauben darf, der die Leistungen von Fritz für gleichverteilt hält ?

Rechenhilfe: $\sqrt{3} = 1,73$

Aufgabe I 6:

Student Fridolin plant, mit seiner viersitzigen Cessna (Kleinflugzeug) zum Bodensee zu fliegen. Um Geld zu sparen, möchte er sich von der Mitflugzentrale drei Personen vermitteln lassen, die sich die Flugkosten mit ihm teilen.

Dabei darf das maximale Abfluggewicht der Cessna von 1.2oo kg nicht überschritten werden. Fridolin weiß, daß das Leergewicht seines Flugzeuges 65o kg beträgt und er für die Strecke 2oo kg Flugbenzin benötigt. Er selbst wiegt 67 kg, sein Gepäck wiegt 1o kg.

Bei seinen Mitfliegern geht er von einem mittleren Körpergewicht von 75 kg bei einer Standardabweichung von o,7 kg pro Person aus. Beim Gepäckgewicht seiner Mitflieger geht er von einem mittleren Gewicht von 12 kg bei einer Standardabweichung von 2,2 kg pro Person aus.

Die Passagier- und Gepäckgewichte seiner Mitflieger seien voneinander unabhängig. Für den Fall, daß zwischen dem tatsächlichen Abfluggewicht und dem maximal zulässigen Abfluggewicht noch für mehr als 24 kg Platz übrig bleibt, soll Fridolion noch ein Paket Expreßfracht mitnehmen.

Wie groß ist die Wahrscheinlichkeit, daß das Flugzeug weder das maximale Abfluggewicht überschreitet noch daß Fridolin sich an die Expreßfrachtfirma wenden muß ?

Aufgabe I 7:

Sie haben die Zufallsgröße Y, die sich als Summe von Zufallsgröße $X_i (i = 1, \ldots, n)$ darstellen läßt. Über die Verteilung X_i sei nichts bekannt.

1) Unter welchen Bedingungen an die einzelnen X_i und n können Sie Y als approximativ normalverteilt ansehen ?

2) Um welchen fundamentalen Satz der Statistik handelt es sich dabei ?

3) Wenn die unter 1), genannten Bedingungen erfüllt sind, wie ist dann Y (approximativ) verteilt, falls

   a) die X_i nicht identisch verteilt sind ?

   b) die X_i identisch verteilt sind ?

Aufgabe I 8

Aus einer Grundgesamtheit von **N = 3** Personen mit den Lebensaltern 20 , 22 und **24** Jahren wird eine Zufallsstichprobe vom Umfang **n = 2** (mit Zurücklegen) gezogen.

Sei X : „ Lebensalter [in Jahren] "

1) Berechnen Sie das arithmetische Mittel μ (d.h. das mittlere Lebensalter der Personen in der Grundgesamtheit) und die Varianz σ^2 (d.h. die Streuung des Lebensalters der Personen in der Grundgesamtheit) !

2) Auf wie viele Arten kann sich die theoretische Stichprobe (X_1 , X_2) realisieren ?

3) Listen Sie alle möglichen Stichprobenergebnisse (x_1 , x_2) auf !

4) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsverteilungen folgender Stichprobenfunktionen $f(X_1 , X_2)$:

   a) $Y = \sum_{i=1}^{2} X_i$

      Bestimmen Sie $E(Y)$ und $Var(Y)$!

   b) $\bar{X} = \frac{1}{2} \cdot \sum_{i=1}^{2} X_i$

      Bestimmen Sie $E(\bar{X})$ und $Var(\bar{X})$!

   c) $Z^2 = \frac{1}{2} \cdot \sum_{i=1}^{2} \left(X_i - \bar{X} \right)^2$

      Bestimmen Sie $E(Z^2)$!

   d) $S^2 = \frac{1}{2-1} \cdot \sum_{i=1}^{2} \left(X_i - \bar{X} \right)^2$

      Bestimmen Sie $E(S^2)$!

5) Überprüfen Sie anhand der unter 4) erhaltenen Ergebnisse folgende Beziehungen :

   a) $E(Y) = n \cdot \mu$ und $Var(Y) = n \cdot \sigma^2$

   b) $E(\bar{X}) = \mu$ und $Var(\bar{X}) = \frac{\sigma^2}{n}$

   c) $E(Z^2) = \frac{n-1}{n} \cdot \sigma^2$

   d) $E(S^2) = \sigma^2$

Aufgabe I 9

A

Geben Sie bei folgenden Verteilungsmöglichkeiten von X_i $(i = 1, 2, \ldots, n)$ an

- welche Verteilung
- welchen Erwartungswert
- welche Varianz

die Stichprobenfunktion $\hat{\Theta}_n = \sum_{i=1}^{n} X_i$ besitzt. Die X_i seien unabhängig.

1) $X_i \sim$ Bernoulli–verteilt mit π

2) $X_i \sim N(\mu ; \sigma^2)$

3) $X_i \sim$ Einpunkt–verteilt mit μ

B

Bei einem Spielautomaten sei der Gewinn pro Spiel normalverteilt mit $\mu = 0$ € und $\sigma = 1$ €.

Wie groß ist die Wahrscheinlichkeit, daß der Gesamtgewinn eines Abends (mit 16 unabhängigen Spielen) über 16 € liegt ?

C

Der Anteil der Studentinnen an allen Studierenden einer Hochschule betrage 40 % . Das Studentenwerk zieht für eine Erhebung eine einfache Stichprobe vom Umfang $n = 30$.

Wie groß ist die Wahrscheinlichkeit dafür, daß in dieser Stichprobe weniger als 30 % oder mehr als 50 % Studentinnen sind ?

Aufgabe I 10

Bert sorgt sich in seinem Gemüseladen angesichts steigender Gemüsepreise um seine leicht verderblichen Tomaten. Er weiß aus Erfahrung, daß in seinem Laden täglich im Mittel 100 kg Tomaten bei einer Varianz von 24,2 [kg^2] nachgefragt werden. Mann kann davon ausgehen, daß die von Tag zu Tag nachgefragten Mengen unabhängig voneinander sind.

1) Wie groß ist die Wahrscheinlichkeit, daß die pro Woche (Montag bis Freitag) nachgefragte Menge an Tomaten um höchstens 33 kg von der erwarteten Nachfrage abweicht ?

Der Verkaufspreis für 1 kg Tomaten beträgt 4 €.

2) In welchen symmetrischen Grenzen um den Erwartungswert (des wöchentlichen Umsatzes) wird mit mindestens 84 % – iger Wahrscheinlichkeit sein wöchentlicher Umsatz liegen ?

Bert muß wöchentlich im vorhinein seine Tomatenbestellung beim Großhändler aufgeben. Dabei bestellt er immer die erwartete wöchentliche Nachfragemenge.
Sei

Z : „ Differenz zwischen nachgefragter und bestellter Menge an Tomaten pro Woche [in kg] “

3)

a) Berechnen Sie E (Z) und Var (Z) !

b) Erläutern Sie inhaltlich, was $Z > 0$ bzw. $Z < 0$ für Berts Gemüseladen (ökonomisch) bedeutet !

c) Wie groß ist die Wahrscheinlichkeit, daß die Differenz zwischen nachgefragter und bestellter wöchentlicher Menge an Tomaten um höchstens 16,5 kg von ihrem Erwartungswert abweicht ?

Aufgabe I 11:

Der Medizinstudent Quacksalber muß sein Praktikum an einer Unfallambulanz absolvieren. Er erfährt, daß in dieser Unfallstation die Anzahl der pro Tag eingelieferten Patienten eine Zufallsgröße ist, wobei

an	Werktagen	im Mittel	2o	Zufallsgröße X;	$Var(X)$	$= 25$
an	Samstagen	im Mittel	15	Zufallsgröße Y;	$Var(Y)$	$= 9$
an	Sonntagen	im Mittel	11	Zufallsgröße Z;	$Var(Z)$	$= 16$

Patienten eingeliefert werden. Man kann davon ausgehen, daß die Anzahl der an verschiedenen Tagen eingelieferten Patienten voneinander unabhängig sind.

Sei <u>W</u> : "Anzahl der innerhalb einer Woche (ohne Feiertag) eingelieferten Patienten".

1) Wieviele Patienten werden im Mittel innerhalb einer Woche in dieser Unfallstation eingeliefert ?

Quacksalbe's Praktikum dauert genau 6 Wochen (ohne Feiertag).

Sei <u>S</u> : "Anzahl der während der Praktikumszeit eingelieferten Patienten".

2) Wie ist die Zufallsgröße S verteilt (Verteilungstyp und -parameter) ?

3) Wie groß ist die Wahrscheinlichkeit, daß während der Praktikumszeit höchstens 75o Patienten eingeliefert werden ?

4) Wie groß ist die Wahrscheinlichkeit, daß während der Praktikumszeit zwischen 72o und 81o Patienten eingeliefert werden ?

Sei <u>T</u> : "Durchschnittliche Anzahl während des Praktikums pro Tag eingelieferte Patienten".

5) Wie ist die Zufallsgröße T verteilt (Verteilungstyp und -parameter) ? Begründen Sie Ihre Antwort ! (<u>Hinweis</u>: Rechnen Sie mit Brüchen !)

6) Mit wieviel Patienten kann man pro Tag im Mittel rechnen ?

7) Wie groß ist die Wahrscheinlichkeit, daß pro Tag die durchschnittliche Anzahl der eingelieferten Patienten höchstens 17 beträgt ?

Aufgabe I 12:

Die Verantwortlichen eines Fußballverbandes sind daran interessiert, einen Überblick zu bekommen, wieviele Tore innerhalb einer Spielsaison ($\hat{=}$ 245 Spiele) in der betreffenden Fußball–Liga geschossen werden. Aus Erfahrung weiß man, daß für jedes einzelne Spiel folgende Verteilung gilt:

X: "Anzahl der geschossenen Tore pro Spiel"	Wahrscheinlichkeit
0	o,2
1	o,2
2	o,1
3	o,4
mehr als 4	0

1) Berechnen Sie den Erwartungswert und die Varianz der Zufallsgröße

   X: "Anzahl der geschossenen Tore pro Spiel".

2) Durch welche Verteilung läßt sich die Verteilung der Zufallsgröße

   Y: "Anzahl der geschossenen Tore pro Spielsaison"

   approximieren ? Begründen Sie Ihre Antwort ausführlich !

3) Berechnen Sie den Erwartungswert und die Varianz der Zufallsgröße Y !

4) Geben Sie die Wahrscheinlichkeit dafür an, daß pro Spiel <u>mehr</u> als **drei** Tore geschossen werden !

5) Geben Sie die Wahrscheinlichkeit dafür an, daß pro Spielsaison <u>mehr</u> als 49o Tore geschossen werden !

6) Geben Sie die Wahrscheinlichkeit dafür an, daß pro Spiel <u>weniger</u> als **drei** Tore geschossen werden !

7) Geben Sie die Wahrscheinlichkeit dafür an, daß pro Spielsaison weniger als 448 Tore geschossen werden !

Aufgabe I 13:

Bei einer Kabinenbahn fährt 121 mal am Tag eine viersitzige Kabine von der Tal- zur Bergstation. Aus Erfahrung weiß man, daß die Kabine pro Fahrt folgendermaßen besetzt ist:

Sie ist nie leer, mit der Wkt von o,4 ist sie voll besetzt, und die Wkt, daß in der Kabine drei Personen sitzen, beträgt o,3 , und die Wkt, daß in der Kabine zwei Personen sitzen beträgt o,2 .

Die Besetzungszahl bei einer Fahrt ist unabhängig von den Besetzungszahlen bei den anderen Fahrten.

1) Durch welche spezielle Verteilung läßt sich die Verteilung von X: "Anzahl der täglich beförderten Personen" approximieren ? Begründung !

2) Wie groß, ist die Wkt, daß an einem Tag mehr als 374 Personen nach oben befördert werden ?

3) Wie beurteilen Sie die Modellannahme der Unabhängigkeit in diesem Beispiel ?

Aufgabe I 14:

Die Verantwortlichen eines Eishockey–Verbandes wissen aus Erfahrung, daß bei 60% der Spiele 5 Pucks, bei 30% der Spiele 6 Pucks und bei 10% der Spiele 7 Pucks in den Taschen der Souvenir–Jäger verschwinden. Pro Saison werden 80 Spiele ausgetragen.

1) Wie groß ist der Erwartungswert und die Varianz des Puckschwundes pro Saison ?

2) Durch welche Verteilung läßt sich die Verteilung des Puckschwundes pro Saison approximieren ? Begründung !

3) Wie groß ist die Wahrscheinlichkeit, daß pro Saison mehr als 452 Pucks verschwinden ?

4) Wie groß ist die Wahrscheinlichkeit, daß pro Saison höchstens 431 Pucks verschwinden ?

5) Wie groß ist die Wahrscheinlichkeit, daß pro Spiel höchstens 3 Pucks verschwinden ?

6) Wieviele Pucks müssen pro Saison eingekauft werden, wenn mit einer Wahrscheinlichkeit von 95% der Bedarf an Pucks in einer Saison gedeckt werden soll ?

Aufgabe I 15

Richtig oder falsch ?

1) Eine gleichgewichtete Stichprobe ist immer auch eine uneingeschränkte Stichprobe.

2) Das (schwache) Gesetz der großen Zahlen besagt, daß die Wahrscheinlichkeit, mit der $\overline{X}_n$ in ein (beliebig klein) vorgegebenes Intervall $[\mu - c ; \mu + c]$ fällt, mit wachsender Anzahl der Versuche gegen Null konvergiert.

3) Eine Stichprobenfunktion gibt an, nach welcher Vorschrift Elemente der Grundgesamtheit in die Stichprobe gelangen.

4) Die sog. „ nicht - bewußte " Auswahl ist ein zufälliges Auswahlverfahren.

5) Der **allgemeine** Fall eines zweistufigen Auswahlverfahren läßt sich folgendermaßen charakterisieren : **Teil**erhebung im 1. und 2. Auswahlvorgang.

6) Wird eine einfache Stichprobe gezogen, haben alle Elemente der Grundgesamtheit eine gleichgroße (und von Null verschiedene) Wahrscheinlichkeit, in die Stichprobe zu gelangen.

7) Bei einem proportional geschichteten Auswahlverfahren hat jede Stichprobe $(X_1 , \dots , X_n)$ die gleiche Wahrscheinlichkeit realisiert zu werden.

8) Die „ Auswahl aufs Geratewohl " ist eine nicht-zufällige Auswahltechnik.

9) Die Tschebyscheffsche Ungleichung läßt sich auch dann anwenden, wenn die betrachtete Zufallsvariable normalverteilt ist.

10) Sei $(X_1 , \dots , X_n)$ eine einfache Stichprobe mit $E (X_i) = \mu$ und $Var (X_i) = \sigma^2$.
Nach dem (schwachen) Gesetz der großen Zahlen gilt :
„ Für großes n nimmt $\overline{X}_n$ mit hoher Wahrscheinlichkeit Werte nahe bei μ an. "

Aufgabe I 16

A

Erläutern Sie (<u>in einem Satz</u>), was man unter einer

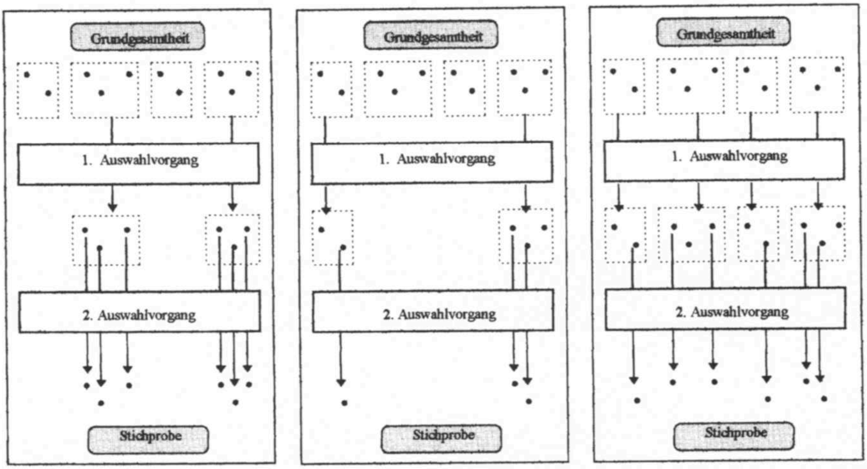
1) Zufallsauswahl
2) uneingeschränkten Zufallsauswahl
3) gleichgewichteten Zufallsauswahl
4) einfachen Zufallsauswahl

versteht !

B

Innerhalb der Theorie der ***eingeschränkten Zufallsauswahlen*** haben Sie die drei in den Graphiken **A , B** und **C** veranschaulichten Verfahren kennengelernt :

A **B** **C**



1) Charakterisieren Sie **kurz** alle drei Verfahren bzgl. des 1. und 2. Auswahlvorganges !
2) Wie nennt man das Verfahren **A** ?
3) Wie nennt man das Verfahren **B** ?
4) Wie nennt man das Verfahren **C** ?

J

Schätzprinzipien

Aufgabe J 1

Familie B. (Vater, Mutter und fünf Kinder) wollen in einem Möbelgeschäft, das mit dem netten Slogan „ *Schraubst Du noch oder wohnst Du schon* " wirbt, einen Schrank kaufen, der an die 306 cm lange Wohnzimmerwand passen soll. Sie finden auch einen Schrank, der ihnen gefällt. Leider stehen keine Längenmaße dran, und ein Verkäufer ist auch nicht in Sicht. Herr B. hat aber ein kleines Holzlineal (Länge etwa 10 cm) zur Hand und schlägt vor, damit die Länge des Schrankes auszumessen (Lineal anlegen, Endpunkt notieren, Lineal wieder anlegen etc.). Frau B. erklärt sich zwar grundsätzlich mit diesem Verfahren einverstanden, mißtraut aber den handwerklichen Fähigkeiten ihres Mannes und will auch selbst eine solche Messung durchführen. Daraufhin will natürlich auch jedes der fünf Kinder nicht untätig bleiben. Es soll in folgender Reihenfolge gemessen werden : Vater, Mutter, jüngstes Kind, . . . , ältestes Kind.

Herr B. überlegt sich, daß jedes Familienmitglied somit eine konkrete Längenmessung x_i ($i = 1, ..., 7$) erzeugt, die als Realisation einer Zufallsvariablen X_i aufgefaßt werden kann.

1) Durch welche Wahrscheinlichkeitsverteilung kann man die Verteilung von X approximieren ? Begründung !

Um die wahre (unbekannte) Länge μ des Schrankes zu schätzen, zieht Herr B. folgende drei Schätzfunktionen alternativ ins Kalkül :

a) $\hat{\Theta}_1 = X_2$ („ *Mutter – Schätzfunktion* ")

b) $\hat{\Theta}_2 = \overline{X}$ („ *gleichgewichtete Familienschätzfunktion* ")

c) $\hat{\Theta}_3 = \frac{1}{5} \cdot \left[\frac{1}{n-1} \cdot \sum_{i=1}^{n-1} X_i \right] + \frac{4}{5} \cdot X_n$ („ *ungleichgewichtete Familienschätzfunktion* ")

2) Geben Sie für jede der drei Schätzfunktionen eine inhaltlichen Erklärung, die den Vater zu dieser Wahl bewogen haben könnte !

3) Was besagt (bzw. besagt nicht) die Eigenschaft : $E(\hat{\Theta}) = \mu$ für unser Schrankproblem ?

4) Vergleichen Sie alle drei Schätzfunktionen hinsichtlich ihrer Streuung, d.h. bestimmen Sie : $Var(\hat{\Theta}_1)$, $Var(\hat{\Theta}_2)$ und $Var(\hat{\Theta}_3)$!

Angenommen, bei den Messungen der einzelnen Familienmitglieder ergaben sich folgende Werte :

$x_1 = 303 \mid x_2 = 308 \mid x_3 = 305 \mid x_4 = 304 \mid x_5 = 303 \mid x_6 = 304 \mid x_7 = 301$,

d.h. wir erhalten folgende drei Schätzwerte :

$$\hat{\vartheta}_1 = \mathbf{308} \quad \text{bzw.} \quad \hat{\vartheta}_2 = \mathbf{304} \quad \text{bzw.} \quad \hat{\vartheta}_3 = \mathbf{301{,}7}$$

5) Kann Herr B. aufgrund einer dieser Schätzwerte sicher sein, daß der Schrank wirklich paßt (bzw. nicht paßt) ?

Aufgabe J 2 :

Bei einer bestimmten Stoffwechselkrankheit weiß man, daß der sog. „ Genotyp " bestimmend für die Krankheit ist. Das „ Allel " [bedeutet : die einander entsprechenden Erbanlagen übereinstimmender Chromosomen] **A** sei dominant über **a** . Daher sind die Träger des Genotyps **aa** krank, die Träger der Genotypen **aA** oder **AA** sind jedoch nicht krank.

Nach den Gesetzen der Genetik sind die Wahrscheinlichkeiten für die Genotypen bei Neugeborenen wie folgt verteilt :

Genotyp	**aa**	**aA** (bzw. **Aa**)	**AA**
Wahrscheinlichkeit	ϑ^2	$2 \cdot \vartheta \cdot (1 - \vartheta)$	$(1 - \vartheta)^2$

Dabei ist ϑ die Wahrscheinlichkeit für das Auftreten des Allels **a** in der Bevölkerung. Dann ist die Wahrscheinlichkeit, daß ein Neugeborenes diese Krankheit hat, gleich ϑ^2.

Ein Genetiker möchte die Wahrscheinlichkeit ϑ schätzen. Dazu läßt er sich von einer Klinik die Anzahl **n** der in einem bestimmten Zeitraum geborenen Kinder angeben, sowie die Anzahl **x** derjenigen unter diesen Neugeborenen, bei denen diese Krankheit festgestellt wurde.

Sei

X : „ Anzahl der von dieser Krankheit betroffenen Neugeborenen bei einer einfachen Stichprobe vom Umfang n "

1) Wie ist die Zufallsvariable X verteilt ? [Verteilungstyp und -parameter !]

2) Stellen Sie die Maximum-Likelihood-Schätzfunktion $L[\,\vartheta \mid (x_1, \ldots, x_n)\,]$ auf !

3) Bestimmen Sie mit Hilfe eines formalen Rechenganges allgemein den Maximum-Likelihood-Schätzwert für ϑ !

Angenommen, in der Klinik sind im betreffenden Zeitraum **1842** Kinder geboren worden, von denen **35** an dieser Stoffwechselkrankheit leiden.

4) Bestimmen Sie mit Hilfe dieser Angaben konkret den Maximum-Likelihood-Schätzwert für ϑ !

Aufgabe J 3

A

Eine Grundgesamtheit hat den Mittelwert μ und die Varianz σ^2. Sei (X_1 , X_2 , X_3) eine (theoretische) Stichprobe aus dieser Grundgesamtheit.

Folgende drei Stichprobenfunktionen sind gegeben :

$$\hat{\Theta}_1 = \frac{1}{3} \cdot \left(X_1 + X_2 + X_3 \right)$$

$$\hat{\Theta}_2 = \frac{1}{4} \cdot \left(2 \cdot X_1 + 2 \cdot X_3 \right)$$

$$\hat{\Theta}_3 = \frac{1}{3} \cdot \left(2 \cdot X_1 + X_2 \right)$$

1) Welche dieser Stichprobenfunktionen sind erwartungstreue Schätzfunktionen für μ ?

2) Welcher Stichprobenfunktion würden Sie nach dem Kriterium der Wirksamkeit den Vorzug geben ? Begründung !

B

Eine Lieferung von $N = 1.000$ Lampen soll mit Hilfe einer Zufallsstichprobe vom Umfang $n = 20$ (mit Zurücklegen) untersucht werden.

Mit der Zufallsvariablen Y : „ Anzahl der defekten Lampen in der Stichprobe “ soll die Zahl δ der defekten Lampen in der Lieferung geschätzt werden.

1) Geben Sie eine erwartungstreue Schätzfunktion $\hat{\Theta} = f(Y)$ für δ an und zeigen Sie, daß gilt :

$$E(\hat{\Theta}) = \delta \ !$$

2) In der vorliegenden Stichprobe sei die Zahl der defekten Lampen gleich 3 .

   Wie hoch schätzen Sie demnach die Anzahl der defekten Lampen in der Lieferung ?

Aufgabe J 4

G. ist ein begabter Lehrling seines Meisters B. und möchte ein eigenes Parfüm entwickeln. Dafür braucht er Zimtzöl (**Z**) , Rosmarinöl (**R**) , Sandelholzöl (**S**) und Jasminöl (**J**) . Von jeder dieser vier Essenzen hat sein Meister viele Fläschchen in der Werkstatt, leider jedoch unbeschriftet.

Die Wahrscheinlichkeit, ein Fläschchen mit **R** herauszuziehen beträgt **q** .
G. weiß, daß die Wahrscheinlichkeit ein Fläschchen mit **S** herauszuziehen nur ein Drittel so groß ist wie die Wahrscheinlichkeit, daß man ein Fläschchen mit **R** <u>oder</u> **J** herauszieht. Die Wahrscheinlichkeit, daß man ein Fläschchen mit **S** <u>oder</u> **R** in die Hände bekommt, ist gleich der Wahrscheinlichkeit, ein Fläschchen mit **J** zu erhalten.

Mit Hilfe einer einfachen Stichprobe vom Umfang n will G. nun den Parameter **q** schätzen. In der Stichprobe befanden sich :

z	Fläschchen mit Zimtöl
r	Fläschchen mit Rosmarinöl
s	Fläschchen mit Sandelholzöl
j	Fläschchen mit Jasminöl

Bestimmen Sie (allgemein) den Maximum–Likelihood–Schätzwert für **q** !

Empfehlung : Verwenden Sie die logarithmierte Maximum–Likelihood–Funktion !

Aufgabe J 5

Die Firma ABC mit $N = 5.000$ Mitarbeitern soll bestreikt werden. Die unbekannte Gesamt**anzahl** δ der tatsächlich streikwilligen Mitarbeiter soll mit Hilfe einer einfachen Stichprobe vom Umfang $n = 25$ geschätzt werden.

Sei

X_i : „ Anzahl der streikwilligen Mitarbeiter bei der i–ten befragten Person "

mit $E(X_i) = \pi$ und $Var(X_i) = \pi \cdot (1 - \pi)$

Zur Schätzung von π werden folgende Stichprobenfunktionen betrachtet :

$$\hat{\Theta}_1 = \frac{N}{n} \cdot \sum_{i=1}^{n} X_i$$

$$\hat{\Theta}_2 = \frac{N}{2} \cdot (X_1 + X_n)$$

1) Sind $\hat{\Theta}_1$ und $\hat{\Theta}_2$ ***erwartungstreue*** Schätzfunktionen für δ ?

2) Welche der beiden Stichprobenfunktionen ist ***wirksamer*** ?

3) Sind $\hat{\Theta}_1$ und $\hat{\Theta}_2$ ***konsistente*** Schätzfunktionen für δ ?

4) Welche der beiden Stichprobenfunktionen ist ***suffizient*** ?

Von den 25 befragten Personen wollen nur die ersten 6 und die letzten 2 befragten Personen nicht am Streik teilnehmen.

5) Bestimmen Sie auf der Basis des vorliegenden Stichprobenergebnisses die Schätzwerte

   a) $\hat{\vartheta}_1$

   b) $\hat{\vartheta}_2$

6) Welcher von beiden Werten ist der „ bessere " Schätzwert für δ ?
   Begründen Sie Ihre Antwort !

Aufgabe J 6:

Die Firma "Längenmaß AG" beabsichtigt ein neues Meßgerät auf den Markt zu bringen. Da es sich um ein mechanisches Gerät handelt, werden n ($n \geq 1$) Mitarbeiter zu Testzwecken gebeten, den (unbekannten) Radius r einer vorgelegten Kreisscheibe zu messen. Der dabei auftretende Meßfehler U ist eine Zufallsvariable und wird als beliebig verteilt angenommen mit dem Erwartungswert $E(U) = 0$ und der konstanten, aber unbekannten Varianz $Var(U) = \sigma^2$.

Sei $(R_1, \ldots, R_n)$ eine einfache Stichprobe mit
R_i : "Radius bei der Messung durch i–ten Mitarbeiter"
$R_i = r + U_i$

A: 1) a) Bestimmen Sie auf der Basis sämtlicher Stichprobenvariablen $(R_1, \ldots, R_n)$ eine erwartungstreue Schätzfunktion $\widehat{\Theta}_1$ für den unbekannten Lageparameter $\vartheta_1 = r$.

b) Ist $\widehat{\Theta}_1$ konsistent ? (Formale Begründung)

2) Von Interesse sei nun die unbekannte Fläche der Kreisscheibe, d.h. $\vartheta_2 = r^2 \cdot \pi$.

a) Bestimmen Sie unter Beachtung des unter 1) erhaltenen Ergebnisses eine Schätzfunktion $\widehat{\Theta}_2$ für ϑ_2.

b) Untersuchen Sie diese Schätzfunktion auf Erwartungstreue bzw. asymptotische Erwartungstreue !

Hinweis: $Var(\bar{R}) = E\left[\bar{R} - E(\bar{R})\right]^2 = E(\bar{R}^2) - \left[E(\bar{R})\right]^2$

3) Konstruieren Sie mit Hilfe der Schätzfunktion $\widehat{\Theta}_2$ eine erwartungstreue Schätzfunktion $\widehat{\Theta}_3$ für ϑ_2.

Die Stichprobe ergab folgende n = 1o Meßwerte:

(11, 13, 12, 12, 12, 4, 12, 11, 11, 12)

4) Berechnen Sie unter Verwendung der Schätzfunktion $\widehat{\Theta}_1$ einen Schätzwert $\widehat{\vartheta}_1$ für $\vartheta_1 = r$.

5) Berechnen Sie jeweils den Schätzwert für $\vartheta_2 = r^2 \cdot \pi$ unter Verwendung der Schätzfunktion

a) $\widehat{\Theta}_2$;

b) $\widehat{\Theta}_3$.

Rechenhilfe: $\pi = 3,14$

B: Angenommen durch ein elektronisches Meßgerät würde festgestellt, daß der Radius die wahre Länge $r = 12$ besitzt.

1) Vergleichen Sie die Werte der wahren Parameter ϑ_1 und ϑ_2 mit den Schätzwerten $\widehat{\vartheta}_1, \widehat{\vartheta}_2$ und $\widehat{\vartheta}_3$.

2) Worin könnten die Unterschiede begründet sein ?

Aufgabe J 7

Ein Spielautomat besitzt folgende Wahrscheinlichkeitsverteilung für

X : „ Gewinn pro Spiel [in €] "

x	-1	0	$+1$
$P(X=x)$	π	π	$(1-2\cdot\pi)$

Der Produzent dieser Spielautomaten beauftragt einen Statistiker, eine Schätzung für π durchzuführen, um einen Anhaltspunkt dafür zu haben, ob sich der Wert von π seit Inbetriebnahme des Spielautomaten geändert hat.

Sei

V : „ Anzahl der **verlorenen** Spiele bei n Spielrunden "
U : „ Anzahl der **unentschiedenen** Spiele bei n Spielrunden "
G : „ Anzahl der **gewonnenen** Spiele bei n Spielrunden "

X : „ Gewinn in € pro Spiel "

Dann sind v , u und g die Realisationen der Zufallsvariablen V , U und G (mit $v+u+g=n$)

1) Der Statistiker zieht eine Stichprobe vom Umfang $n = 6$, d.h. er betätigt den Spielautomaten genau 6 mal. Die (theor.) Stichprobe $(X_1, \dots, X_n)$ hat sich folgendermaßen realisiert :

$$(-1, +1, -1, 0, +1, +1)$$

Verbalisieren Sie dieses Stichprobenergebnis !

2) Geben Sie folgende Wahrscheinlichkeiten an : $P(X=0)$; $P(X=-1)$; $P(X=+1)$!

3) Wie würden Sie die Wahrscheinlichkeiten für den Gewinn X pro Spiel aufgrund der obigen Stichprobe bestimmen, wenn Sie überhaupt keine Information über die Wahrscheinlichkeitsverteilung von X hätten ?

4) Bestimmen Sie

a) $E(V)$; b) $E(U)$; c) $E(G)$!

5) Ermitteln Sie nach der ***Maximum–Likelihood–Methode*** den Schätzwert für π !

Wie lautet die nach der ***Maximum–Likelihood–Methode*** ermittelte Schätzfunktion $\hat{\Theta}_{ML}$ für π ?

6) Ermitteln Sie nach der ***Methode der kleinsten Quadrate*** den Schätzwert für π !

Wie lautet die nach der ***Methode der kleinsten Quadrate*** ermittelte Schätzfunktion $\hat{\Theta}_{KQ}$ für π ?

7) Prüfen Sie, ob die Funktionen $\hat{\Theta}_{ML}$ bzw. $\hat{\Theta}_{KQ}$ unverzerrte Schätzfunktionen für π sind !

8) Bestimmen Sie

a) $Var(V)$; b) $Var(U)$; c) $Var(G)$!

9) Bestimmen Sie die Varianz der Schätzfunktionen $\hat{\Theta}_{ML}$ und $\hat{\Theta}_{KQ}$!

Hinweis : $\mathbf{Cov(X_i, X_j) = -n\cdot\pi_i\cdot\pi_j \quad (i \neq j)}$

10) Welche der beiden Schätzfunktionen $\hat{\Theta}_{ML}$ und $\hat{\Theta}_{KQ}$ ist wirksamer ?

11) Prüfen Sie, ob die Funktionen $\hat{\Theta}_{ML}$ bzw. $\hat{\Theta}_{KQ}$ konsistente Schätzfunktionen für π sind !

Aufgabe J 8:

Sie haben drei Eigenschaften kennengelernt, die eine Schätzfunktion haben kann:

a) *Erwartungstreue* b) *Konsistenz* c) *Wirksamkeit*

1) Beschreiben Sie kurz, was jede dieser Eigenschaften aussagt, und formalisieren Sie diese Eigenschaften !

2) Kritisieren Sie jede dieser Eigenschaften !

3) Stellen Sie diese Eigenschaften in Zusammenhang mit dem M.S.E. !

4) Sie kennen als eine Methode der Punktschätzung die Maximum-Likelihood-Methode. Stellen Sie diese Methode in Zusammenhang mit den oben genannten drei Eigenschaften !

Aufgabe J 9:

Otto N. wartet auf dem Finanzamt. Er weiß, daß die Zeit T , die ein Klient im Zimmer der Beamten verbringt, exponentialverteilt ist

$$\left(\text{Dichte}: \; f(t) = \lambda \cdot e^{-\lambda \cdot t}, \quad E(T) = \frac{1}{\lambda}\right).$$

Wenn Frau Hurtig Dienst hat, beträgt die durchschnittliche Zeit für die Abfertigung einer Person 3 Minuten, falls Herr Lasch Dienst hat, beträgt die durchschnittliche Zeit für die Abfertigung einer Person dagegen 5 Minuten.

Otto N. registriert, daß die drei vor ihm wartenden Personen nach 1 Minute, 5 Minuten und 3 Minuten das Zimmer der Beamten wieder verlassen.

1) Stellen Sie die Likelihoodfunktion $L(\lambda)$ für dieses Stichprobenergebnis auf !

2) Ermitteln Sie den Maximum-Likelihood-Schätzwert für λ !   Wer hat Ihrer Meinung nach Dienst ?

Rechenhilfe: $a^n \cdot a^m = a^{n+m}$; $(e^{-bx})' = -b \cdot e^{-bx}$

Aufgabe J 1o:

Im Winter hat so manches Auto seine Startprobleme. Statistik-Student Karl weiß, daß X: "Anzahl der Startversuche bis sein Auto (endlich) anspringt" geometrisch verteilt ist, d.h.

$$P(X = x) \;=\; (1-\pi)^{x-1} \cdot \pi \,,$$

wobei π die unbekannte Wahrscheinlichkeit dafür ist, daß der Motor bei einem Startversuch anspringt. Um π mit der Maximum–Likelihood–Methode zu schätzen, führt Karl unter gleichen Witterungsbedingungen zwei voneinander unabhängige Startversuchsreihen durch. Bei der ersten Versuchsreihe springt der Motor beim 7. Startversuch ($x_1 = 7$) und bei der zweiten Versuchsreihe springt der Motor beim 3. Startversuch ($x_2 = 3$) an.

1) Stellen Sie allgemein (π unbekannt) die Maximum–Likelihood–Funktion

   $$L(\pi) \;=\; P\left[(X_1 = 7) \,\cap\, (X_2 = 3)\right]$$

   auf !

2) Bestimmen Sie für den unbekannten Parameter π den Maximum–Likelihood–Schätzwert $\widehat{\pi}$ zur obigen Stichprobenrealisation !

Aufgabe J 11:

Sie interessieren sich für die Unfallhäufigkeit an einem Verkehrsknotenpunkt. Sowohl die Polizei als auch die Versicherungsgesellschaften haben diesen Sachverhalt bereits in der Vergangenheit untersucht, so daß Sie auf diese Unterlagen zurückgreifen können:

Versicherungsges. :

x	0	1	2	$x \neq 0,1,2$
$P(X=x)$	0,7	0,1	0,2	0

Polizei :

x	0	1	2	$x \neq 0,1,2$
$P(X=x)$	0,1	0,4	0,5	0

1) Wie nennt man ein solches "Zurückgreifen" auf vorhandene Daten ?

Leider unterscheiden sich beide Wahrscheinlichkeitsverteilungen doch recht erheblich voneinander. Um zwischen beiden Verteilungen entscheiden zu können, stellen Sie sich selbst an 5 zufällig ausgewählten Stichtagen an diese Straßenkreuzung und registrieren die Unfälle. Die Stichprobe ergab folgendes Ergebnis:

$$(0,\ 2,\ 0,\ 2,\ 1)$$

2) Wie lautet Ihre Entscheidung nach dem Maximum–Likelihood–Prinzip ?

3) Wie lautet Ihre Entscheidung nach der Methode der kleinsten Quadrate ?

Aufgabe J 12 :

Angenommen, der Sicherheitsstandard von Reaktoren in Atomkraftwerken innerhalb einer bestimmten Periode kann in drei Sicherheitsstufen angesetzt werden :

0 (≅ kein Zwischenfall) ; 1 (≅ ein Zwischenfall) ; 2 (≅ zwei Zwischenfälle)

Um die Anteile π_j der zur Stufe j (j = 0, 1, 2) gehörenden Reaktoren zu schätzen, werden n Reaktoren zufällig ausgewählt (einfache Stichprobe) und ihre Sicherheitsstufen notiert.

Sei

X_i : „ Sicherheitsstufe des i – ten Reaktors " (i = 1, ... , n)

π_j = Anteil der zur Sicherheitsstufe j (j = 0, 1, 2) gehörenden Reaktoren

= Wahrscheinlichkeit, daß ein Reaktor zur Sicherheitsstufe j gehört

1) Bestimmen Sie tabellarisch (in Abhängigkeit von π_j) die Wahrscheinlichkeitsfunktion von $\mathbf{X_i}$!

2) Bestimmen Sie $E(\mathbf{X_i})$!

3) Bestimmen Sie $E(\mathbf{X_i}^2)$!

Hinweis (falls erforderlich) :
$\mathbf{E(X^2) = E[X \cdot (X-1)] + E(X) = \sum_x x \cdot (x-1) \cdot P(X=x) \quad + \quad E(X)}$

Die Anteile π_j sind unbekannt und sollen geschätzt werden.

Folgende drei Schätzfunktion stehen zur Wahl :

$$\hat{\Theta}_1 = \frac{1}{2 \cdot n} \cdot \sum_{i=1}^{n} (X_i^2 - X_i)$$

$$\hat{\Theta}_2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (1 - \frac{3}{2} \cdot X_i + \frac{1}{2} \cdot X_i^2)$$

$$\hat{\Theta}_3 = \frac{1}{n} \cdot \sum_{i=1}^{n} (2 \cdot X_i - X_i^2)$$

4) Welche der drei Schätzfunktionen ist (sind) **erwartungstreu** bzgl.

a) π_0

b) π_1

c) π_2 ?

Aufgabe J 13:

A: Die Partei der Einsiedler (PdE) bezieht ihre Wähler vornehmlich aus kleinen Ortschaften mit bis zu 2o Wahlberechtigten. Hier ist man überzeugt mit durchschnittlich jeder fünften Stimme rechnen zu können.

Während der Spitzenkandidat der PdE in der Stimmenzahl für seine Partei eine Binominalverteilung zu erkennen glaubt, halten die Befürworter eines Wahlbündnisses jede Stimmabgabe zugunsten der PdE eher für ein poissonverteiltes Ereignis.

Um nun die Strategie im Superwahljahr festzulegen, wurden in vier repräsentativ ausgewählten Ortschaften jeweils 2o Wähler nach ihrer Zustimmung zur PdE befragt. Das Umfrageergebnis lautet:

$$(3, 3, 0, 1o).$$

1) Welches der beiden Verteilungsmodelle ist nach dem Maximum–Likelihood–Prinzip aufgrund des vorliegenden Stichprobenergebnisses plausibler ?

B: Sei X ein in der Grundgesamtheit durch folgende Verteilungsfunktion gekennzeichnetes Merkmal:

$$F(x) = \begin{cases} 1 & \text{für} \quad x \geq b \\ \pi & \text{für} \quad a \leq x < b \\ 0 & \text{sonst.} \end{cases}$$

Die Realisation einer einfachen Stichprobe vom Umfang n ergab:

$$x_i \geq b \quad \text{für alle} \quad i = 1, \ldots, n$$

1) Bestimmen Sie $P(X \geq b)$!
2) Stellen Sie (in Abhängigkeit von x) die Verteilungsfunktion $F(x)$ graphisch dar!
3) Stellen Sie die Likelihoodfunktion $L\big(\pi/(x_1, \ldots, x_n)\big)$ auf !
4) Skizzieren Sie die Likelihoodfunktion für n = 3 und
   a) $\pi = $ o,6 ;
   b) $\pi = $ o,2 ;
   c) $\pi = $ o,o1 !
5) Wie lautet der Maximum–Likelihood–Schätzwert für π ?
6) Geben Sie die so geschätzte Verteilungsfunktion $\widehat{F}(x)$ vollständig an !
7) Welches Verteilungsmodell ergibt sich somit für das Merkmal X ?
8) Bestimmen Sie $E(X)$ und $Var(X)$!

Aufgabe J 14 :

Der Statistikstudent Anton Gambler ist ein begeisterter Glücksspieler, der gern Spielcasinos besucht. Er will bei seinem nächsten Casinobesuch im Spielerparadies Bad Pankow die Spielautomaten beobachten und sich für jeden Automaten die Anzahl der Spiele notieren, die benötigt werden bis der Hauptgewinn kommt.

Sei

X : „Anzahl der Spiele an einem Automaten bis der Hauptgewinn kommt“

Dabei geht Anton von folgender Wahrscheinlichkeitsverteilung der Zufallsvariablen X aus :

$$P(X=x) \;=\; (1-\pi)^{x-1}\cdot\pi \qquad \text{mit} \qquad E(X)=\frac{1}{\pi}$$

$\pi :=$ Wahrscheinlichkeit für den Hauptgewinn pro Spiel

Anton will den Parameter π schätzen, indem er n Automaten beobachtet und folgende Realisation der einfachen Stichprobe $(X_1, X_2, \ldots, X_n)$ erhält : $(x_1, x_2, \ldots, x_n)$.

1) Stellen Sie die Maximum-Likelihood-Funktion $L[\pi \mid (x_1, x_2, \ldots, x_n)]$ auf!

2) Stellen Sie die **logarithmierte** Maximum-Likelihood-Funktion $\mathbf{ln}\, L[\pi \mid (x_1, x_2, \ldots, x_n)]$ auf!

3) Bestimmen Sie mit Hilfe eines formalen Rechenganges <u>allgemein</u> den Maximum-Likelihood-Schätzwert für π !

4) Stellen Sie die Kleinste-Quadrate-Funktion $Q[\pi \mid (x_1, x_2, \ldots, x_n)]$ auf!

5) Bestimmen Sie mit Hilfe eines formalen Rechenganges <u>allgemein</u> den Kleinste-Quadrate-Schätzwert für π !

Angenommen, bei der Beobachtung von $n = 3$ Automaten hätte sich die einfache Stichprobe (X_1, X_2, X_3) wie folgt realisiert : $(637, 427, 541)$.

6) Wie lautet <u>konkret</u> der Maximum-Likelihood-Schätzwert für π aufgrund des vorliegenden Stichprobenergebnisses ?

7) Wie lautet <u>konkret</u> der Kleinste-Quadrate-Schätzwert für π aufgrund des vorliegenden Stichprobenergebnisses ?

<u>Hilfe</u> : $\ln(a \bullet b) = \ln a + \ln b$

$\ln(a/b) = \ln a - \ln b$

$\ln a^b = b \bullet \ln a$

<u>Hilfe</u> :

$$\frac{d(\ln u)}{dx} = \frac{1}{u}\cdot\frac{du}{dx} \quad \text{mit} \quad u = f(x)$$

Aufgabe J 15

Richtig oder falsch ?

1) Das „Mean Square Error - Konzept“ betrachtet den mittleren tatsächlichen Schätzfehler einer Schätzfunktion.

2) Vor der Beobachtung ordnet ein festes Stichprobenergebnis (x_1 , ... x_n) den möglichen ϑ - Werten eine Likelihood zu, nach der Beobachtung ordnet ein ϑ - Wert den möglichen Werten von ($X_1, \ldots, X_n$) eine Wahrscheinlichkeit zu.

3) Das „ Likelihood - Prinzip “ ist nicht frequentistisch, d.h. die Bewertung der ϑ - Werte durch die Likelihoodfunktion richtet sich allein nach der einen beobachteten Stichprobe. Nach der Beobachtung wird nicht mehr berücksichtigt, was sonst noch alles hätte beobachtet werden können.

4) Eine Likelihoodfunktion $L(\vartheta)$ kann auch negative Werte annehmen.

5) Anhand des **„** *Mean Square Error - Konzeptes* **“** lassen sich Gütekriterien für Stichprobenfunktionen herleiten.

6) Eine Schätzfunktion, die weder erwartungstreu noch asymptotisch erwartungstreu ist, ist auch nicht konsistent.

7) Die Maximum-Likelihood-Methode besagt, daß zu einem festen Stichprobenergebnis ($x_1, \ldots, x_n$) derjenige Schätzwert für den unbekannten (zu schätzenden) Parameter ϑ zu wählen ist, unter dem im nachhinein die Wahrscheinlichkeit für das Eintreten dieses Stichprobenergebnisses am größten ist.

8) Von zwei konsistenten Schätzfunktionen $\hat{\Theta}_1$ und $\hat{\Theta}_2$ ist diejenige wirksamer zum Schätzen des unbekannten Parameters ϑ, die die kleinere Varianz besitzt.

9) Die Gütekriterien „ Suffizienz “ und „ Robustheit “ lassen sich aus dem „ *Mean Square Error – Konzept* “ herleiten.

10) Eine erwartungstreue Stichprobenfunktion heißt konsistent, falls ihre Varianz mit wachsendem Stichprobenumfang gegen Null konvergiert.

K

Intervallschätzung

Aufgabe K 1:

A: Bei einem Test wurden 49 zufälllig ausgewählte PKW's des gleichen Typs mit der gleichen Kraftstoffmenge ausgestattet. Mit dieser Füllung legten sie im Durchschnitt 5o km zurück. (Standardabweichung der Grundgesamtheit ist mit 7 km als bekannt vorausgesetzt.)

1) Geben Sie explizit das Konfidenzintervall $[V_u; V_o]$ für die durchschnittliche Kilometerleistung μ dieses PKW–Typs zum Konfidenzniveau $1 - \alpha$ an !
2) Bestimmen Sie das Schätzintervall für $\mu\,(1 - \alpha = 95\%)$!
3) Wie groß müßte der Stichprobenumfang n gewählt werden, wenn bei gleichem Konfidenzniveau das Schätzintervall für μ eine Breite von höchstens 2 km aufweisen soll?

B: Einige Zuschauer dieser Testverantaltung werden zufällig von einem Sportreporter ausgewählt und nach ihrer Zugehörigkeit zum ADAC befragt. Unter den 2oo befragten Personen befanden sich 4o Mitglieder des ADAC.

1) Geben Sie explizit das approximative Konfidenzintervall $[V_u; V_o]$ für den unbekannten Anteil π der ADAC–Mitglieder bei dieser Veranstaltung zum Konfidenzniveau $1 - \alpha$ an !
2) Bestimmen Sie das Schätzintervall für $\pi\,(1 - \alpha = 99\%)$!
(Approximativ und exakt !)

C: Ein geschäftstüchtiger Automatenaufsteller hat am Rande der Zuschauertribühne einen Kaffee–Automaten aufgestellt, der die Plastikbecher zu je o,2 l gegen Einwurf einer Münze mit Kaffee füllt. Man kann davon ausgehen, daß die Füllmenge annähernd normalverteilt ist. Eine Zufallsstichprobe vom Umfang $n = 5$ ergab folgende Werte in l :

$$0,18; \quad 0,25; \quad 0,12; \quad 0,20; \quad 0,25; \; .$$

1) Geben Sie explizit das Konfidenzintervall $[V_u; V_o]$ für die durchschnittliche Füllmenge μ dieses Kaffee–Automaten zum Konfidenzniveau $1 - \alpha$ an !
2) Bestimmen Sie das Schätzintervall für $\mu\,(1 - \alpha = 95\%)$!

Aufgabe K 2

Eine Maschine produziert Fahrradschläuche mit einem Durchmesser von 40 cm. Man weiß aus Erfahrung, daß der Durchmesser der produzierten Fahrradschläuche vom Zufall abhängt und $N(\mu ; \sigma^2)$ – verteilt ist. Um zu kontrollieren, ob die Maschine noch richtig eingestellt ist, werden 25 Fahrradschläuche zufällig ausgewählt (einfache Stichprobe). Aus den 25 Meßwerten wurde ein mittlerer Durchmesser (arithm. Mittel) von 41 cm bei einer Standardabweichung von 3 cm ermittelt.

1) Geben Sie explizit das Konfidenzintervall $[V_U ; V_O]$ für μ zum Niveau $1 - \alpha$ an !

2) Bestimmen Sie das Schätzintervall für μ ($1 - \alpha = 90\,\%$) !

3) Angenommen, Sie können davon ausgehen, daß die Stichprobenstandardabweichung garantiert kleiner oder gleich 5 cm ist.
   Wie groß müssen Sie dann den Stichprobenumfang wählen, damit das Schätzintervall maximal 1 cm breit wird ?

Aufgabe K 3

Nach Aufhebung der Mietpreisbindung möchte der Wohnungssenator dieser Stadt Aufschluß über die Auswirkungen dieser Maßnahme erhalten.

A

Es soll der durchschnittliche Mietpreis einer 80 m^2 – Altbauwohnung geschätzt werden. Dazu läßt sich der Senator vom Wohnungsamt 36 derartige Wohnungen zufällig auswählen (einfache Stichprobe) und die zur Zeit dafür bezahlte Miete mitteilen.

1) Geben Sie zum Konfidenzniveau $1 - \alpha = 98$ % explizit das Konfidenzintervall für den wahren durchschnittlichen Mietpreis einer 80 m^2 – Altbauwohnung an !

Aus der Stichprobe wurden folgende Werte berechnet : $\bar{x} = 950$ € ; $s = 180$ €.

2) Mit welcher Formal muß s berechnet worden sein ?

3) Bestimmen Sie das Schätzintervall !

4) Vor Aufhebung der Mietpreisbindung betrug der wahre durchschnittliche Mietpreis für diesen Wohnungstyp 875 €. Wenn Sie zum Vergleich das unter 3) errechnete Ergebnis heranziehen, können Sie dann sicher sagen, ob sich der wahre durchschnittliche Mietpreis verändert hat ? Begründung !

B

Es soll der Anteil an Altbauwohnungen ermittelt werden, die länger als 3 Monate leerstehen. Dazu wird eine einfache Stichprobe vom Umfang $n = 24$ gezogen und gezählt, wie viele dieser Altbauwohnungen seit mehr als 3 Monaten leerstehen.

1) Geben Sie zum Konfidenzniveau $1 - \alpha = 95$ % explizit das approximative Konfidenzintervall für den gesuchten Anteil an !

In der Stichprobe befanden sich 6 Wohnungen, die seit mehr als 3 Monaten leerstehen.

2) Geben Sie das Schätzintervall für den gesuchten Anteil an, und zwar

   - approximativ (rechnerisch)
   - exakt (Nomogramm)

3) Wie hätte das Konfidenzniveau $1 - \alpha$ gewählt werden müssen (größer oder kleiner), um (bei derselben Stichprobe) ein <u>kürzeres</u> Schätzintervall zu erhalten ? Begründen Sie Ihre Antwort !

Aufgabe K 4

Ein Schweinemäster möchte Informationen über das Gewicht seiner ausgewachsenen Mastschweine gewinnen. Dazu bittet er den gerade auf seinem Hof Urlaub machenden Studenten Statistix um Mithilfe.
Sie ziehen eine (einfache) Stichprobe von $n = 6$ Schweinen und wiegen diese.

Man kann davon ausgehen, daß X : „ Gewicht eines Schweins “ eine $N(\mu ; \sigma^2)$ – verteilte Zufallsvariable ist.

1) Geben Sie explizit das Konfidenzintervall für den tatsächlichen Gewichtsdurchschnittswert μ zum Konfidenzniveau $1 - \alpha$ an !

Es ergaben sich folgende 6 Meßwerte (in kg) :

100	104	98	100	101	97

2) Führen Sie mit Hilfe dieser Daten erwartungstreue Punktschätzungen für $E(X) = \mu$ und $Var(X) = \sigma^2$ durch !

3) Bestimmen Sie das Schätzintervall für μ $(1 - \alpha = 95\,\%)$!

4) Interpretieren Sie das Schätzergebnis !

5) Welche Möglichkeiten haben Sie bei der Versuchsplanung, Einfluß auf die Länge des Schätzintervalls zu nehmen ?

6) Wie würde sich das Schätzintervall verändern, wenn – unter sonst gleichen Bedingungen – $\sigma^2 = 6$ als bekannt vorausgesetzt werden könnte ?
Begründung !

Aufgabe K 5:

1) Welchen Einfluß haben

   a) Stichprobenumfang,

   b) Varianz der Schätzfunktion,

   c) Konfidenzniveau

   auf die Länge eines Konfidenzintervalls ?

	richtig	falsch
2) Sei $1 - \alpha = 99\%$ und $[0,8;\ 1,2]$ ein Schätzintervall für μ . Dann liegt μ mit einer Wahrscheinlichkeit von 99% in diesem Intervall.	○	○

	richtig	falsch
3) Sei $[V_u;\ V_o]$ ein Konfidenzintervall für μ zum Konfidenzniveau $1 - \alpha = 99\%$. Dann überdeckt das Intervall mit 99%iger Wahrscheinlichkeit den unbekannten Parameter μ.	○	○

Aufgabe K 6

Die Studenten Anton und Berta sind befreundet und planen, nach ihrem Studium eine gemeinsame Weltreise zu machen. Beide spielen an mehreren Abenden in der Woche Karten . . . um Geld ! Sie verabreden, ab sofort die Gewinne, die jeder erzielt, in eine gemeinsame Urlaubskasse einzulegen. Die Reise werden sie jedoch nur dann antreten können, wenn zu ihrem bereits ersparten Geld ein Gemeinschaftsgewinn für alle kommenden Spielabende von mindestens 1.100 € zusammenkommt.
Um sich das Geld nicht gegenseitig abzunehmen, spielen Sie zwar an denselben Abenden, jedoch in verschiedenen Spielrunden :
Anton spielt in einer Männerrunde Skat, Berta spielt in einer Damenrunde Bridge.
Sei :

X : „ Gewinn von Anton pro Spielabend “ mit $E(X) = \mu_X$; $Var(X) = \sigma_X^2 = \mathbf{64}$ €²

Y : „ Gewinn von Berta pro Spielabend “ mit $E(Y) = \mu_Y$; $Var(Y) = \sigma_Y^2 = \mathbf{36}$ €²

Z : „ Gemeinschaftsgewinn pro Spielabend “

1) Wie groß ist die Wahrscheinlichkeit, daß der Gemeinschaftsgewinn pro Spielabend um höchstens ± 17 € von seinem erwarteten Wert abweicht ?

Da Anton und Berta noch viel für die bevorstehenden Prüfungen lernen müssen, haben sie leider nur noch bis zu ihrer Reise Zeit für **27** Spielabende.

Sowohl Anton als auch Berta sehen ihre erzielten Gewinne bei den letzten **81** Spielabenden jeweils als Ergebnisse einer einfachen Stichprobe an.

Anton erzielte dabei einen Gesamtgewinn von **1.782** € , während Berta nur einen Gesamtgewinn von **1.539** € verbuchen konnte.

2) Bestimmen Sie auf einem Konfidenzniveau von $1 - \alpha = 95{,}45\ \%$ ein Schätzintervall für
   a) die zu erwartende zukünftige Gewinnsumme von Anton !
   b) die zu erwartende zukünftige Gewinnsumme von Berta !
   c) den zu erwartenden zukünftigen Gemeinschaftsgewinn von Anton und Berta !

3) Werden Anton und Berta aufgrund des unter 2c) erhaltenen Ergebnisses ihre Weltreise nach dem Studium antreten können ?
   Begründen Sie kurz Ihre Antwort !

L

Testtheorie
[parametrisch]

Aufgabe L 1

Ein Unternehmen stellt Spezialgefrierschränke her, die zur Konservierung bestimmter Güter verwendet werden. Die Soll–Kühltemperatur beträgt für derartige Gefrierschränke $-25\ ^0C$.
Da man weiß, daß die tiefgefrorenen Güter bei höheren Temperaturen leicht verderben und da der potentielle Kundenstamm nicht sehr groß ist, würde ein mangelhaftes Produkt, das also nicht tief genug kühlt, das Schlimmste – nämlich den Ruin der Firma – bedeuten. Aus Gründen der Vorsicht soll nun die Kühlleistung der Gefrierschränke an 100 zufällig aus der Produktion ausgewählten Gefrierschränken (einfache Stichprobe) auf einem Signifikanzniveau von 2,275 % getestet werden, um zu entscheiden, ob die Produktion weiterlaufen kann oder eine Konstruktionsänderung an den Geräten vorgenommen werden muß.
Aus Erfahrung weiß man, daß die erreichte Kühltemperatur eines solchen Gefrierschrankes normalverteilt ist mit einer Standardabweichung von $2\ ^0C$.

1) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

2) Geben Sie die Prüfgröße formal und verbal sowie ihre Verteilung unter H_O an !

3) Wie lautet die Testfunktion und wie ist diese unter H_O verteilt ?

4) Bestimmen Sie den Verwerfungsbereich für diesen Test !

5) Bestimmen Sie die Werte der Gütefunktion, falls die tatsächliche mittlere Kühltemperatur

   a) $-24{,}8\ ^0C$ bzw. b) $-25{,}4\ ^0C$ bzw. c) $-29{,}0\ ^0C$ beträgt !

6) Skizzieren Sie die Gütefunktion !

7)

   a) Die Zufallsstichprobe ergab eine mittlere Kühltemperatur pro Gerät von $-26\ ^0C$ bei einer Standardabweichung von $1{,}5\ ^0C$.

      a_1) Wie lautet die Testentscheidung ?

      a_2) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

      a_3) Ist der Firma bei der Testentscheidung ein Fehler unterlaufen ? Wenn ja, um welchen Fehler handelt es sich ?

   b) Die Zufallsstichprobe ergab eine mittlere Kühltemperatur pro Gerät von $-25{,}3\ ^0C$ bei einer Standardabweichung von $1{,}8\ ^0C$.

      b_1) Wie lautet die Testentscheidung ?

      b_2) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

      b_3) Welcher Fehler kann bei dieser Entscheidung unterlaufen sein ?

      b_4) Wie groß ist die Wahrscheinlichkeit, daß dieser Fehler hier unterlaufen ist ?

      b_5) Wie groß ist die Wahrscheinlichkeit, bei diesem Testverfahren diesen Fehler zu begehen, falls μ tatsächlich $-29\ ^0C$ beträgt ?

8) Warum reicht es beim einseitigen Test aus, daß man unter der Nullhypothese nur den Fall $\mu = \mu_O$ betrachtet ?

Aufgabe L 2:

Für den Durchmesser von Wellen ist ein Sollwert von 2oo mm vorgeschrieben. Außerdem ist bekannt, daß die Produktion normalverteilt ist mit einer Standardabweichung von 5 mm. Um die Produktion zu kontrollieren, zog der Kontrolleur K_1 eine Zufallsstichprobe von n=1oo. Das arithmetische Mittel aus diesen 1oo Stück ergab eine Abweichung vom Sollwert von +o,4 mm.

Es soll die Hypothese $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ getestet werden auf einem Signifikanzniveau von 5%.

1) Geben Sie den Verwerfungsbereich B an !

2) Wie entscheidet sich K_1 ?

3) Welchen Fehler könnte K_1 bei seiner Entscheidung begangen haben ?

4) Ein zweiter Kontrolleur K_2 berechnet aus einer zweiten Zufallsstichprobe von n=1oo ein $\bar{x} = 2o2$ mm. Wie entscheidet sich K_2 (bei gleichen Entscheidungskriterien) ?

5) Welchen Fehler könnte K_2 bei seiner Entscheidung begangen haben ?

Aufgabe L 3

Seit geraumer Zeit beklagen Umweltschützer die Verschmutzung der Seen durch die Abwässer der Haushalte, insbesondere durch den Phosphatgehalt der Waschmittel. So greifen sie auch eine bestimmte Firma an, da sie glauben, daß der zulässige Durchschnittswert von 18 g pro Packung überschritten wird. Die Firma bestreitet dies energisch und verspricht den Umweltschützern, das Produkt vom Markt zu nehmen, falls sich statistisch zeigen läßt, das der mittlere Phosphatgehalt ihres Produkts tatsächlich zu hoch ist.
Die Firma will selbst diesen Test durchführen und schlägt den Umweltschützern eine Irrtumswahrscheinlichkeit von 0,001 vor, da man ja dann mit hoher Sicherkeit ein richtiges Testergebnis bekommen würde. Die Umweltschützern akzeptieren dies, da ihnen die Argumentation völlig einleuchtet.
Die Varianz des Phosphatgehalts pro Packung wird mit 36 g^2 als bekannt vorausgesetzt.

1) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

2) Geben Sie die Prüfgröße formal und verbal an !

3) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?

4) Bestimmen Sie den Ablehnbereich für diesen Test !

In der (einfachen) Stichprobe vom Umfang $n = 36$ ergab sich ein durchschnittlicher Phosphatgehalt pro Packung von 20 g bei einer Standardabweichung von 5,9 g .

5) Wie lautet die Testentscheidung ?

Die Firma gibt daraufhin folgende Pressemitteilung heraus :

> „ Mit Hilfe eines Tests konnte unsere Firma statistisch belegen, daß der mittlere Phosphatgehalt in unseren Waschmittelpaketen den Richtwert von 18 g nicht überschreitet. Für diesen Test wurden unter den Augen der Umweltschützer 36 Waschmittelpakete zufällig ausgewählt und untersucht. Um ein Fehlentscheidung bei diesem Test fast vollständig auszuschließen, haben wir nach Absprache mit den Umweltschützern den Test so durchgeführt, daß eine Fehlentscheidung nur mit 0,1 %–iger Wahrscheinlichkeit vorkommen kann. Damit hat unsere Firma wieder einmal bewiesen, daß sie zu den umweltbewußten Herstellern von . . . etc. . . . etc. . . . ! “

6) Nehmen Sie ausführlich Stellung zu dieser Presseerklärung !

7) Hat die Firma bei der Testentscheidung einen Fehler begangen ?

8) Berechnen Sie die Wahrscheinlichkeit, bei diesem Test eine Fehlentscheidung zu begehen, falls der wahre mittlere Phosphatgehalt pro Packung den Wert 21,1 g haben sollte !

9) Ist der Verlauf der Gütefunktion dieses Tests anhängig vom

   a) Stichprobenergebnis ?

   b) Stichprobenumfang ?

   Begründen Sie kurz Ihre Antworten !

Aufgabe L 4

Der Glühlampenhersteller Marso hat eine neuartige Energiespar–Glühlampe entwickelt, die nun auf den Markt gebracht werden soll. Wenn der mittlere Energieverbrauch dieser Glühlampen unterhalb von 12 Watt liegt, vergibt das Umweltministerium ein begehrtes, da verkaufswirksames Zertifikat, welches die Glühlampen als besonders umweltfreundliches Produkt einstuft.

Die Entscheidung für die Erteilung dieses Zertifikats soll mit Hilfe eines statistischen Tests ($\alpha = 5\ \%$) herbeigeführt werden. Zu diesem Zweck sollen 20 Glühlampen zufällig ausgewählt (einfache Stichprobe) und jeweils deren Energieverbrauch gemessen werden.

Aus Erfahrung weiß man, daß der Energieverbrauch von Energiespar–Glühlampen als normalverteilt angesehen werden kann mit einer Varianz von 0,8 Watt^2 .

Die Testdurchführung soll zweifach, nämlich

in **Fall I** aus Sicht und Interessenlage des <u>Glühlampenherstellers Marso</u>

und

in **Fall II** aus Sicht und Interessenlage des <u>Umweltministeriums</u>

durchgeführt werden.

Beantworten Sie zunächst für **Fall I** und danach für **Fall II** folgende Fragen :

1) Wie heißt das (bei Ausschöpfung aller vorhandenen Informationen) zu verwendende Testverfahren ?

2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – <u>falls nötig</u> – mit Hilfe einer Risikoüberlegung !

3) Wie lautet formal und verbal die Prüfgröße des Test und deren Verteilung unter H_0 ?

4) Wie lautet die Testfunktion des Tests und deren Verteilung unter H_0 ?

5) Wie lautet der Verwerfungsbereich des Tests ?

6) Wie groß ist bei diesem Test die Wahrscheinlichkeit einer Fehlentscheidung, wenn der wahre mittlere Energieverbrauch der neuartigen Glühlampen 12,2 Watt beträgt ?

7) Bei welchem wahren mittleren Energieverbrauch beträgt die Wahrscheinlichkeit einer richtigen Testentscheidung 85 % ?

Das Stichprobenergebnis ergab einen mittleren Energieverbrauch der neuartigen Glühlampen von 12 Watt bei einer Standardabweichung von 0,49 Watt.

8) Wie lautet die Testentscheidung ?

9) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

Aufgabe L 5:

In einer Behörde mit sehr vielen Mitarbeitern, hat der Behördenvorsteher den Verdacht, daß an einem normalen Arbeitstag höchstens 7o% der Arbeitnehmer wirklich an ihren Arbeitsplätzen sind. Er möchte nun einerseits diesen Mißstand bei der Senatsinnenverwaltung zur Sprache bringen, möchte aber auf der anderen Seite auf jeden Fall das Risiko so klein wie möglich halten, Ärger mit der einflußreichen Personalvertretung zu bekommen, der sich zwangsläufig bei einer ungerechtfertigten voreiligen Meldung ergeben würde.

Stellen Sie aus der Sicht des Behördenvorstehers die Hypothesen für einen Test auf den wahren Anteil der an ihrem Arbeitsplatz befindlichen Arbeitnehmer auf !
Begründen Sie Ihre Wahl !

Aufgabe L 6:

1) Skizzieren Sie für die folgenden Hypothesen den Annahmebereich und kennzeichnen Sie α !

   a) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$
   b) $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$
   c) $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$

2) Erläutern Sie die folgenden in der Testtheorie auftretenden Wahrscheinlichkeiten:

   a) α

   b) $1 - \alpha$

   c) β

   d) $1 - \beta$

3) Wenn man bei einem Test $\alpha = 0$ wählt, kann man niemals einen Fehler begehen.

richtig	falsch
◯	◯

4) Ein einseitiger Test wird — wenn möglich — so aufgebaut, daß die Hypothese, die statistisch zu beweisen ist, als Nullhypothese formuliert wird.

richtig	falsch
◯	◯

5) In den Hypothesen eines statistischen Parametertests wird eine Annahme über die bekannten Verteilungsparameter der Grundgesamtheit formuliert.

richtig	falsch
◯	◯

Aufgabe L 7:

A: Welche der folgenden Aussagen sind richtig bzw. falsch ?

1) Die Gütefunktion ist eine Funktion des zu testenden Parameters.

2) Wenn bei einem Test auf μ in Wahrheit die Gegenhypothese richtig ist, so ist die Wahrscheinlichkeit für die Verwerfung der Nullhypothese gleich $1 - \beta_{(\mu)}$.

3) Bei der Durchführung eines statistischen Tests kann man stets den Fehler 1.Art und den Fehler 2.Art begehen.

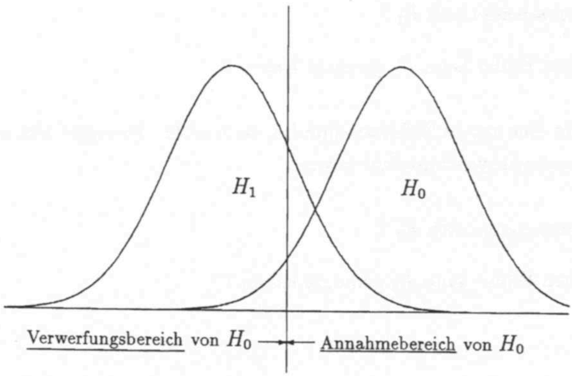
4) Die Gütefunktion eines Tests läßt sich erst berechnen, wenn die Testentscheidung vorliegt.

B: Die Festlegung der Hypothese H_0 und H_1 beim Signifikanztest muß abhängig vom Stichprobenergebnis erfolgen.

richtig	falsch
○	○

C: Sie führen einen Test ($H_0 : \mu \geq \mu_0$) bei normalverteilter Grundgesamtheit durch.

Skizzieren Sie, welche Flächenstücke den Wahrscheinlichkeit α; β; $1 - \alpha$; $1 - \beta$ entsprechen !



Aufgabe L 8:

Ein Supermarkt hatte bisher Hähnchen mit einem Durchschnittsgewicht von 1.4oo g zu einem bestimmten Preis bezogen. Ein Händler macht nun ein Angebot, Hähnchen von gleichem Durchschnittsgewicht zu einem günstigerem Stückpreis liefern zu können. Die Einkäufer E_1 und E_2 des Supermarkts, die beide wissen, daß das Hähnchengewicht normalverteilt ist, vermuten, daß der günstige Preis durch ein zu geringes Durchschnittsgewicht zustande kommt. E_1 wiegt daraufhin 25 zufällig ausgewählte Hähnchen ab. Dabei stellt sich heraus, daß das arithmetische Mittel um -9 g vom Sollgewicht abwich und die Standardabweichung sich zu 5o g aus der Stichprobe ergab. $(\alpha = 0{,}05)$

1) Die Einkäufer stellen folgende Hypothese auf:

$$H_0 : \mu \geq \mu_0\,(= 1.400)$$
$$H_1 : \mu < \mu_0\,(= 1.400)$$

   Welches Risiko wird bei dieser Hypothesenformulierung klein gehalten ?

2) Geben Sie diejenige Stichprobenfunktion, die sich zur Prüfung der aufgestellten Hypothesen eignet, verbal an !

3) Geben Sie Ihre Verteilung und Parameter unter der Annahme an, daß H_0 richtig ist!

4) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?

5) Ermitteln Sie den Annahmebereich und den Verwerfungsbereich B !

6) Wie entscheidet sich E_1 ?

7) Welchen Fehler kann E_1 gemacht haben ?

E_2 entnimmt eine zweite Zufallsstichprobe von n = 25 . Es ergibt sich ein $\bar{x} = 1.381$ g bei gleicher Standardabweichung wie zuvor.

8) Wie entscheidet sich E_2 ?

9) Welchen Fehler kann E_2 gemacht haben ?

Aufgabe L 9:

Kein Produktionsprozeß ist vollkommen. Deshalb muß man darauf bedacht sein, daß die vorgeschriebenen Sollwerte der Produkte möglichst gut eingehalten werden. Unter Verwendung einer geeigneten Prüfgröße testet man hierzu die Hypothese $H_0 : \mu = \mu_0$, daß ein bestimmter Sollwert eingehalten ist, gegen die Alternativhypothese $H_1 : \mu \neq \mu_0$. Behält man bei einem solchen Test $\mu = \mu_0$ bei, so läßt man den Produktionsvorgang weiterlaufen. Führt der Test zur Ablehnung, so stoppt man den Prozeß.

Eine Füllmaschine füllt Konserven mit Erbsen. Der Hersteller vermutet, daß die Füllgewichte seiner Dosen normalverteilt sind. der Sollwert sei $\mu_0 = 300$ g . Ein Kontrolleur nimmt aus der laufenden Produktion eine Zufallsstichprobe von n = 100 und errechnet sich aus den Stichprobenwerten ein $\bar{x} = 304$ g bei einer Standardabweichung von 20 g .

1) Muß aufgrund des Ergebnisses der Stichprobe der Produktionsprozeß gestoppt werden ? ($\alpha = 0{,}05$)

Ein Abnehmer hat die Vermutung, daß die Dosen aus dieser Produktion zu leicht sind und möchte diesen Sachverhalt statistisch beweisen.

2) Welche Hypothesenformulierung wählt er ?

Aufgabe L 1o:

Bei einer normalverteilten Grundgesamtheit wird ein einseitiger Mittelwerttest mit folgender Hypothesenformulierung duchgeführt:

$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$

1) Warum muß die Hypothesenformulierung unabhängig vom Stichprobenergebnis erfolgen ?

2) Warum muß das Signifikanzniveau α vor der Testdurchführung festgelegt werden ?

3) Geben Sie die vier Entscheidungssituationen, die bei diesem Test möglich sind, in der Form ("..." / ...) an und diskutieren Sie deren inhaltliche Bedeutung !

4) Warum spricht man bei der Ablehnung von H_0 von einem statistischen Beweis der gewählten Gegenhypothese H_1 ?

5) Warum spricht man bei einer Beibehaltung von H_0 nicht von einem statistischen Beweis dieser H_0–Hypothese ?

6) Mit welchem "Sammelbegriff" kann man die Wahrscheinlichkeiten bezeichnen, aus denen eine Gütefunktion $g(\mu)$ besteht ?

7) Warum ist (bei einem Signifikanzniveau von $\alpha = 0{,}05$) die Formulierung:
"Die Hypothese H_0 ist, falls sie abgelehnt wird, mit der Wahrscheinlichkeit von 95% unzutreffend" nicht haltbar ?

8) Wie müßte eine "saubere" Interpretation des Signifikanzniveau α unter 7) lauten ?

Aufgabe L 11

Der Zigarettenkonzern TAB will den Zigarettenpreis erhöhen, obwohl dieser vor einem Jahr schon einmal angehoben wurde. Prokurist L. liegt das Wohl des Konzerns am Herzen. Er meint, daß der Zigarettenkonsum pro Tag bei neuerlichen Preiserhöhungen abnehmen werde. Er weiß, daß der tägliche Zigarettenkonsum pro Raucher vor der letzten Preiserhöhung einem erwarteten Wert von 16 Stück entsprach.
Mit Hilfe einer zufälligen Befragung (einfache Stichprobe) unter $n = 100$ Rauchern will er statistisch nachweisen, daß sich der durchschnittliche Konsum verringert hat ($\alpha = 1\,\%$).

1) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

2) Geben Sie die Prüfgröße dieses Tests verbal an !

3) Geben Sie Verteilungstyp und Verteilungsparameter dieser Prüfgröße unter H_0 an !

4) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?

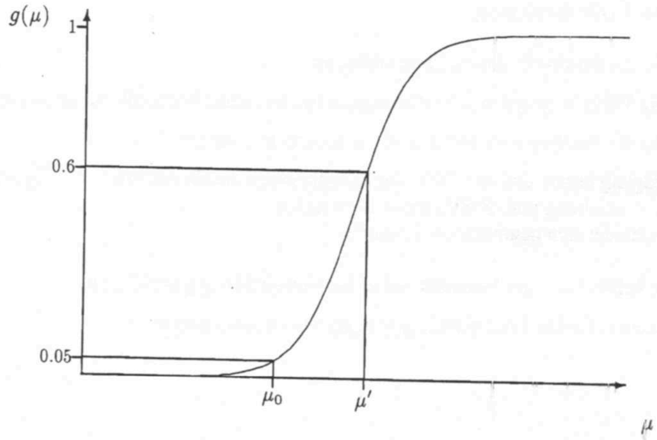
5) Die Befragung ergab, daß die 100 Versuchspersonen durchschnittlich 15 Zigaretten bei einer Standardabweichung von 5 Stück pro Tag rauchen.
   Wie entscheidet sich demnach der Prokurist ?

6) Welchen Fehler kann der Prokurist bei seiner Entscheidung gemacht haben ?

7) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

Aufgabe L 12:

Der Verkaufsschlager der Edumm–Käserei ist ihr 1.ooo–g–Kugelkäse. In letzter Zeit häufen sich die Beschwerden der Kunden, da ihrer Meinung nach der Käse zu leicht sei. Die Käserei führt daraufhin (basierend auf einer einfachen Stichprobe von n = 2o) einen statistischen Test auf das wahre mittlere Gewicht dieser Sorte Kugelkäse durch. Dieser Test besitze folgende Gütefunktion:



1) Welche Hypothesen liegen diesem Test zugrunde ?

2) Angenommen, die Hypothesen basieren auf einer Risikoüberlegung. Erläutern Sie inhaltlich dasjenige Fehlentscheidungsrisiko, das bei diesem Test als weniger schwerwiegend eingeschätzt wird !

3) Wie groß ist an der Stelle $\mu = \mu'$ die Wahrscheinlichkeit, sich falsch zu entscheiden ?

4) Nennen Sie das Testergebnis, falls der Wert der Prüfgröße nicht in den Ablehnbereich fällt !
   Interpretieren Sie dieses Testergebnis inhaltlich und statistisch exakt !

5) Erläutern Sie kurz, worin bei Verwendung eines weniger trennscharfen Tests der Nachteil läge !

Aufgabe L 13

Die Verantwortlichen einer kleinen Partei sind sich einig darüber, daß man bei den bevorstehenden Wahlen „ ums Überleben " kämpfen wird. Würde man mindestens 10 % der abgegebenen Wählerstimmen erhalten, wäre der erneute Einzug in das Parlament gesichert.

Der Spitzenkandidat dieser Partei ist zwar einerseits vom bisherigen Wahlkampf ermüdet, meint aber andererseits, daß vielleicht die Anstrengungen noch nicht ausreichen und die letzten Reserven zur Erlangung der Wählergunst noch einmal mobilisiert werden müßten.

Er will seine Entscheidung vom Ergebnis eines statistischen Tests mit geringer Irrtumswahrscheinlichkeit ($\alpha = 0{,}01$) abhängig machen basierend auf einer repräsentativen Blitzumfrage unter 30 wahlberechtigten Personen (einfache Stichprobe).

1) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – <u>falls nötig</u> – mit Hilfe einer Risikoüberlegung !

2) Verbalisieren Sie die für diesen Test geeignete Stichprobenfunktion !

3) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?

4) Bestimmen Sie den Annahme – bzw. Ablehnbereich für diesen Test !

Die Blitzumfrage ergab, daß von den 30 befragten Personen genau 23 dieser Partei ihre Stimme nicht geben werden.

5) Wie lautet die Testentscheidung ?

6) Was bedeutet diese Entscheidung für das weitere Verhalten des Spitzenkandidaten ?

7)
   a) Wenn die Blitzumfrage unter <u>20</u> Personen – unter sonst gleichen Bedingungen – dasselbe Ergebnis geliefert hätte, wie würde nun die Testentscheidung lauten und wie würde sich der Spitzenkandidat bis zur Wahl verhalten ?

   b) Wenn die Blitzumfrage unter <u>300</u> Personen – unter sonst gleichen Bedingungen – dasselbe Ergebnis geliefert hätte, wie würde nun die Testentscheidung lauten und wie würde sich der Spitzenkandidat bis zur Wahl verhalten ?

Aufgabe L 14

Die beiden Schwergewichtsboxer Jim Knockout und Bill Uppercut werden aufgrund von Computer–Ranglisten als gleichstarke weltbeste Boxer eingeschätzt. Eine bekannte Firma für Hühneraugenpflaster will dem weltbesten Boxer einen Werbevertrag über 5 Mill. € pro Jahr anbieten. Der Chef dieser Firma glaubt, daß Jim Knockout, der mit seinem gefürchteten „ rechten Hammer " schon viele K.O.–Siege errungen hat, der bessere Boxer ist, zumal dieser stets barfuß boxt, so daß wegen der Werbewirkung auch die Hühneraugenpflaster besser zur Geltung kämen.

Um seine These statistisch zu untermauern, organisiert der Firmenchef 11 Schaukämpfe zwischen den beiden Boxern, wobei es in jedem dieser Kämpfe stets einen Sieger geben soll, ein Unentschieden also nicht möglich ist.

1) Formulieren Sie die Hypothesen für diesen Test ($\alpha = 0{,}05$) ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

2) Geben Sie die Prüfgröße für diesen Test an !

3) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?

4) Bestimmen Sie Annahme – und Verwerfungsbereich für diesen Test !

5) Wie lautet die Testentscheidung, wenn Jim Knockout genau 3 Kämpfe verliert ?

6) Kann diese Testentscheidung falsch sein ? Wenn ja, um welchen Fehler handelt es sich ?

7) Die Firma beabsichtigt, das Testergebnis in kurzer und verständlicher Form in einer Presseerklärung zu veröffentlichen.

   Formulieren Sie diese Presseerklärung !

Aufgabe L 15

In einem bekannten Wintersportort wird alljährlich ein großes Skirennen für Gäste veranstaltet, bei dem alle Gäste teilnehmen können. Diesmal soll der Slalom an einem erst kürzlich erschlossenen Hang stattfinden. Eine Gruppe von Skilehrern bekommt den Auftrag, einen Slalom abzustecken, der für alle Gäste befahrbar ist. Es ist beabsichtigt, daß mehr als 90 % („ Sicherheitsfaktor ") der Gäste heil durchs Ziel kommen sollen. Die bisherigen Erfahrungen haben gezeigt, daß man den Schwierigkeitsgrad solcher Hänge genau überprüfen sollte.
Es soll ein Test auf einem Signifikanzniveau von 0,1 durchgeführt werden auf der Basis der Ergebnisse von 22 zufällig ausgewählten Gastskiläufern, die den abgesteckten Slalom probeweise durchfahren sollen (einfache Stichprobe).

1) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

2) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?

3) Bestimmen Sie den Verwerfungsbereich für diesen Test !

4) Wie groß ist die exakte Wahrscheinlichkeit (schlimmstenfalls), sich bei diesem Test für H_1 zu entscheiden ?

5) Von den 22 Testläufern schied nur ein Läufer aus. Wie lautet die Testentscheidung ?

6) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

7) Bestimmen Sie die Werte der Gütefunktion, falls für den „ Sicherheitsfaktor " der Slalompiste in Wahrheit gilt :

   a) 100 %

   b) 95 %

   c) 90 %

   d) 85 %

   e) 80 %

   f) 75 %

   g) 70 %

   h) 65 %

8) Skizzieren Sie den Verlauf der Gütefunktion unter Verwendung der unter 7) ermittelten Werte !

Aufgabe L 16

A

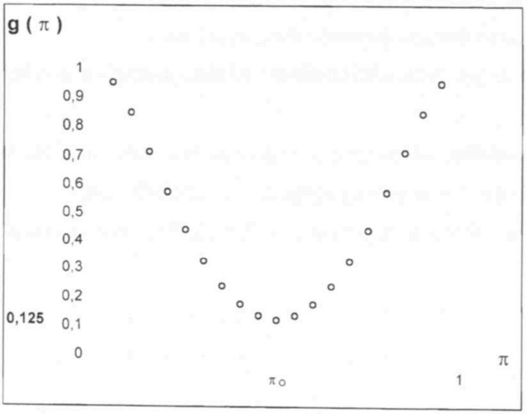
Bei einem statistischen Test auf π basierend auf einer einfachen Stichprobe vom Umfang $n = 25$ sei der Annahmebereich des Tests : $\{ y \,/\, y \geq 6 \}$. Das vorgegebene Signifikanzniveau sei $\alpha = 10\,\%$.

1) Rekonstruieren Sie die Hypothesen für diesen Test einschließlich des Sollwertes π_0 !

2) Geben Sie die exakte Irrtumswahrscheinlichkeit α an !

B

Ein statistischer Test auf π basierend auf einer einfachen Stichprobe vom Umfang $n = 7$ habe folgende Gütefunktion :



1) Geben Sie die Hypothesen für diesen Test einschließlich des Sollwertes π_0 !

2) Bestimmen Sie (numerisch) den Annahme– und Verwerfungsbereich dieses Tests !

Aufgabe L 17

Der in den neuen Bundesländern nicht sehr beliebte Minister Williams plant für das Frühjahr eine Reise in den Osten Deutschlands, wo er auf Kundgebungen als Redner auftreten will. Williams hält seine Reise aber nur dann für lohnenswert, wenn er damit rechnen kann, bei mindestens 80 % seiner Kundgebungen nicht mit Eiern beworfen zu werden.
Sie als Staatssekretär für statistische Angelegenheiten wollen dem Minister mit Hilfe eines Binomialtests ($\alpha = 5\,\%$; $n = 30$) bei seiner Entscheidung, ob er die Reise antreten soll oder nicht, behilflich sein.
Obwohl nun der Minister ein Typ ist, der schnell beleidigt ist und im übrigen eine Lebensmittelallergie gegen Eier hat, steht jedoch für ihn die Gefahr eines Popularitätsverlustes im Falle des Absagens seiner Reise deutlich im Vordergrund. Daher möchte er das Risiko, sich fälschlicherweise gegen die Reise zu entscheiden, unbedingt so klein wie möglich halten.

1) Stellen Sie die Hypothesen für diesen Test auf ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

2) Wie lautet die Testfunktion formal und verbal ? Wie ist diese unter H_0 verteilt ?

3) Bestimmen Sie den Verwerfungsbereich für diesen Test !

4) Wie groß ist ungünstigstenfalls die exakte Wahrscheinlichkeit, sich fälschlicherweise für H_1 zu entscheiden ?

5) Bestimmen Sie die Wahrscheinlichkeit für eine Fehlentscheidung bei diesem Test, falls es in Wahrheit bei

   a) 90 %

   b) 70 %

   der Kundgebungen nicht zu Eierwürfen kommen sollte !

6) Bestimmen Sie die Wahrscheinlichkeit, die Reise berechtigterweise abzusagen, falls es in Wahrheit bei 65 % der Kundgebungen zu Eierwürfen kommen sollte !

7) **Um zu einer Testentscheidung zu gelangen, benötigt man selbstverständlich Datenmaterial !**
   Beschreiben Sie kurz, welche Probleme inhaltlicher Art bei der Datenbeschaffung hier in diesem Beispiel auftreten !

Aufgabe L 18

Ein Hersteller von Überraschungseiern wirbt damit, daß in jedem fünften produzierten Ei eine Spielzeugfigur sei. Wird diese Quote unterschritten, sind Umsatzeinbußen zu befürchten. Wird diese Quote überschritten, werden die Produktionskosten zu hoch, da diese aufwendig bemalten Spielzeugfiguren sehr teuer sind. Daher möchte er beide Umstände möglichst vermeiden.
In Berlin betreibt der Hersteller eine veraltete Produktionsanlage, die er genau dann abschalten will, wenn sich mit Hilfe eines statistischen Tests ($\alpha = 0{,}1$) absichern läßt, daß die vorgegebene Norm nicht eingehalten wird.
Zu diesem Zweck werden 24 Überraschungseier aus derselben Produktionsreihe ausgewählt (einfache Stichprobe) und dahingehend überprüft, ob ein ausgewähltes Ei eine Spielzeugfigur enthält oder nicht.

1) Wie heißt das hier zu verwendende Testverfahren ?

2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

3) Wie lautet die Testfunktion formal und verbal ? Wie ist diese unter H_0 verteilt ?

4) Bestimmen Sie den Annahme– und Verwerfungsbereich dieses Tests !

5) Wie groß exakt ist die Wahrscheinlichkeit, sich bei diesem Test fälschlicherweise für H_1 zu entscheiden ?

6) Bestimmen Sie die Werte der Gütefunktion und skizzieren Sie ihren Verlauf, wenn in Wahrheit der Spielfigurenanteil in den produzierten Überraschungseiern

   a) 10 %

   b) 25 %

   c) 30 %

   d) 60 % beträgt !

Die Überprüfung der 24 ausgewählten Überraschungseier ergab einen Eieranteil ohne Spielfigur von $^5/_6$.

7) Wie lautet die Testentscheidung ?

8) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

Aufgabe L 19:

Auf zwei Maschinen Nr. 1 und Nr. 2 wird Tee abgepackt. Auf Stichprobenbasis soll geprüft werden, ob die Maschine Nr. 1 mit einem größeren durchschnittlichen Füllgewicht arbeitet als die Maschine Nr. 2 $(\alpha = 0{,}01)$.

A: Aus der Vergangenheit sei bekannt, daß die Füllgewichte der beiden Maschinen annähernd normalverteilt sind mit $\sigma_1^2 = 36\,g^2$ und $\sigma_2^2 = 16\,g^2$.
Eine Zufallsstichprobe vom Umfang $n_1 = 12$ aus der Produktion der Maschinen Nr.1 liefert ein durschschnittliches Füllgewicht von $\bar{x}_1 = 130$ g. Eine Zufallsstichprobe aus der Produktion der Maschine Nr. 2 vom Umfang $n_2 = 10$ ergibt ein durchschnittliches Füllgewicht von $\bar{x}_2 = 127$ g.

B: Aus der Vergangenheit sei bekannt, daß die Füllgewichte der beiden Maschinen annähernd normalverteilt sind und die Varianzen der Füllgewichte als gleich angenommen werden können. Eine Zufallsstichprobe vom Umfang $n_1 = 12$ aus der Produktion der Maschine Nr. 1 liefert ein durchschnittliches Füllgewicht von $\bar{x}_1 = 130$ g bei einer Standardabweichung von $s_1 = 2{,}2$ g. Eine Zufallsstichprobe aus der Produktion der Maschine Nr. 2 vom Umfang $n_2 = 10$ ergibt ein durchschnittliches Füllgewicht von $\bar{x}_2 = 127$ g bei einer Standardabweichung von $s_2 = 1{,}8$ g.

C: Eine Zufallsstichprobe vom Umfang $n_1 = 60$ aus der Produktion der Maschine Nr.1 liefert ein durchschnittliches Füllgewicht von $\bar{x}_1 = 130$ g bei einer Standardabweichung von $s_1 = 2{,}2$ g. Eine Zufallsstichprobe aus der Produktion der Maschine Nr.2 vom Umfang $n_2 = 40$ ergibt ein durchschnittliches Füllgewicht von $\bar{x}_2 = 127$ g bei einer Standardabweichung von $s_2 = 1{,}8$ g.

Führen Sie für jede dieser drei Modellsituationen den Test durch !

Aufgabe L 20

Von der Pharma–Industrie wurde ein neues Präparat gegen das – bisher als unheilbar geltende und weltweit verbreitete – „ Nasenjucken " entwickelt, wobei vom Hersteller versprochen wird, daß bei höchstens 65 % der mit diesem Präparat behandelten Patienten kein Heilerfolg eintritt.

Da Sie als behandelnder Arzt es mit der Kostendämpfung im Gesundheitswesen sehr genau nehmen, wollen Sie dieses (sehr teure) Präparat in Zukunft ihren Patienten nur dann verordnen, wenn man den Angaben des Herstellers trauen kann und die Krankenkassen nicht zur Kasse gebeten werden für die Bezahlung eines Präparats, dessen Heilungsquote minimal ist.

Sie beschließen, dieses Präparat an Ihren 19 Patienten, die an „ Nasenjucken " leiden, auszuprobieren (einfache Stichprobe) und mit diesen Zahlen die Angaben des Herstellers zu testen ($\alpha = 0{,}01$). Von dieser Testentscheidung wollen Sie die Einführung des Präparats abhängig machen.

1) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

2) Geben Sie die Prüfgröße für diesen Test formal und an !

3) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?

4) Bestimmen Sie den Verwerfungsbereich für diesen Test !

Da man in der Regel erst nach einigen Tagen über den Heilerfolg von medizinischen Präparaten konkret etwas sagen kann, stellen Sie sich in der Zeit, in der sie auf die Untersuchungsergebnisse warten, folgende Fragen :

Wie groß ist die Wahrscheinlichkeit, daß Sie sich aufgrund des Testergebnisses

5) nicht für die Einführung des Präparats entscheiden, falls die Heilungsquote in Wahrheit sogar 50 % betragen sollte ?

6) für die Einführung des Präparats entscheiden, falls die Heilungsquote in Wahrheit nur 10 % betragen sollte ?

7) nicht für die Einführung des Präparats entscheiden, falls die Heilungsquote in Wahrheit 35 % betragen sollte ?

8) für die Einführung des Präparats entscheiden, falls die Heilungsquote in Wahrheit 40 % betragen sollte ?

Aufgabe L 21

In einem Automobilwerk wird die Qualität zweier besonders günstig vom Hersteller angebotener Lieferungen von Autoreifen auf Stichprobenbasis überprüft.

Die 100 zufällig ausgewählten Reifen (einfache Stichprobe) aus der Lieferung A erreichten eine durchschnittliche Laufleistung von 40.000 km pro Reifen bei einer Standardabweichung von 2.000 km.

Die 180 zufällig ausgewählten Reifen (einfache Stichprobe) aus der Lieferung B erreichten eine durchschnittliche Laufleistung von 51.680 km pro Reifen bei einer Standardabweichung von 4.000 km.

Um den Preis zu drücken, will das Automobilwerk dem Hersteller der Reifen statistisch nachweisen, daß die Lieferung A hinsichtlich der Laufleistung signifikant schlechter ist als die Lieferung B ($\alpha = 0{,}01$).

1) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

2) Geben Sie die Prüfgröße formal und verbal sowie ihre Verteilung unter H_O an !

3) Wie lautet die Testfunktion und wie ist diese unter H_O verteilt ?

4) Bestimmen Sie den Verwerfungsbereich für diesen Test !

5) Wie lautet die Testentscheidung ?

6) Welcher Fehler kann bei der Entscheidung unterlaufen sein ?

7) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

8) **Lesen Sie den gesamten Aufgabentext noch einmal sorgfältig duch !**

   Welchen Vorwurf kann der Reifenhersteller dem Automobilwerk hinsichtlich der Versuchsplanung berechtigterweise machen ?

Aufgabe L 22:

Der Soziologe Dr. T. Fau beschäftigt sich seit längerer Zeit mit dem Fernsehverhalten von Kindern. Er glaubt, daß der tägliche Fersehkonsum von Kindern in einem Zusammenhang mit dem sozialen Status der Eltern steht. Seiner Auffassung nach ist der mittlere Fernsehkonsum (in Min.) pro Tag bei Kindern, die der unteren sozialen Schicht zuzuordnen sind, höher als bei Kindern aus der Mittelschicht. Er möchte nun versuchen, seine Auffassung mit Hilfe eines statistischen Tests ($\alpha = 0{,}01$) zu untermauern. Zu diesem Zweck will er zwei einfache Stichproben von $n_U = 72$ bzw. $n_M = 98$ von Kindern der unteren bzw. mittleren sozialen Schicht ziehen.

1) Stellen Sie die Hypothesen für diesen Test auf !
   Begründen Sie — falls nötig — Ihre Wahl mit Hilfe einer Risikoüberlegung !

2) Geben Sie die Prüfgröße formal und verbal an !
   Wie ist diese unter H_0 verteilt ?

3) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?

4) Bestimmen Sie den Verwerfungsbereich für diesen Test !

Dr. T. Fau ermittelte bei 170 ausgewählten Kindern zwischen 6 und 10 Jahren, wieviel Zeit sie durchschnittlich pro Tag vor dem Fernsehgerät verbringen:
Die 72 aus der unteren sozialen Schicht ausgewählten Kinder (einfache Stichprobe) sahen durchschnittlich 142 Minuten pro Tag fern bei einer Standardabweichung von 24 Minuten. Die übrigen aus der Mittelschicht ausgewählten 98 Kinder (einfache Stichprobe) saßen im Mittel 134 Minuten pro Tag vor dem Fernsehgerät bei einer Standardabweichung von 28 Minuten.

5) Wie lautet die Testentscheidung ?

6) Welchen Fehler könnte Dr. T. Fau bei seiner Testentscheidung begangen haben ?
   Könnte die Wahrscheinlichkeit für eine derartige Fehlentscheidung zum Beispiel 94,9% betragen ?
   Begründen Sie kurz Ihre Antwort !

7) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

8) Wie groß hätte — unter sonst gleichen Bedingungen — das Signifikanzniveau sein müssen, damit es zu einer anderen Testentscheidung gekommen wäre ?

Aufgabe L 23

Der berühmte Parfümeur B. (Hersteller von Parfümen) besitzt viele kleine Geschäfte in Paris, in denen nicht nur Parfüm, sondern auch zwei Sorten Seife angeboten werden, nämlich „ Rosenseife " und „ Nelkenseife " . B. macht sich Gedanken über seine Blütenvorräte für den Winter. Sein finanzieller Etat erlaubt ihm jedoch nur, entweder eine Lieferung Rosenblüten oder eine Lieferung Nelkenblüten einzukaufen. Seine Entscheidung will er vom Ergebnis eines statistischen Tests ($\alpha = 0{,}1$) abhängig machen. Läßt sich statistisch nachweisen, daß in einem bestimmten Zeitraum von der Rosenseife durchschnittlich mehr als 50 Stück (zusätzlich) verkauft werden als von der Nelkenseife, wird B. nur eine Lieferung Rosenblüten nachkaufen, andernfalls muß er bei der Bank einen Kredit aufnehmen, um auch noch Nelkenblüten für den Winter zu bestellen.
Als Datenbasis dienen ihm die Verkaufszahlen von Rosen– und Nelkenseifen in jedem von 10 zufällig ausgewählten Geschäften.

Sei

X : „ Anzahl der verkauften Stück Rosenseife in einem Geschäft " mit $X \sim N(\mu_X ; \sigma_X^2)$

Y : „ Anzahl der verkauften Stück Nelkenseife in einem Geschäft " mit $Y \sim N(\mu_Y ; \sigma_Y^2)$

1) Wie heißt das zu verwendende Testverfahren ?
2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !
3) Wie lautet die Prüfgröße formal und verbal ?
4) Wie ist die Prüfgröße unter H_0 verteilt ?
5) Wie lautet die Testfunktion ? Wie ist diese unter H_0 verteilt ?
6) Bestimmen Sie den Verwerfungsbereich für diesen Test !

Die Stichprobe brachte folgendes Ergebnis :

Geschäft Nr.		1	2	3	4	5	6	7	8	9	10
Verkaufszahlen in Stück	x_i	210	200	185	280	190	220	155	250	150	230
	y_i	130	110	125	200	150	180	165	160	50	170

Hilfe :

$$\sum_{i=1}^{10} (x_i - y_i)^2 = 49.500$$

7) Wie lautet die Testentscheidung ?
8) Interpretieren Sie das Ergebnis inhaltlich und statistisch exakt !
9) Welche Einkaufsstrategie wird B. aufgrund des Testergebnisses wählen ?
10) Nennen Sie zwei Gründe, die bei einem Test auf Vergleich zweier Mittelwerte ($\mu_X \leftrightarrow \mu_Y$) für die Wahl einer verbundenen Stichprobe sprechen !

Aufgabe L 24:

Auf einer landwirtschaftlichen Versuchsanlage werden zufällig 1o Felder ausgewählt, um ein neues Düngemittel für den Kartoffelanbau zu testen. Nachdem jedes Versuchsfeld halbiert wurde, wird in der ersten Hälfte das herkömmliche Düngemittel und in der zweiten Hälfte das neue Mittel eingesetzt. Die jeweiligen Ernteerträge, die in der unten stehenden Tabelle zusammengestellt sind, sollen als Realisationen normalverteilter Zufallsvariablen X und Y angesehen werden.

Feld i	1	2	3	4	5	6	7	8	9	1o
x_i	7.1	6.4	6.8	8.8	7.2	9.1	7.4	5.2	5.1	5.9
y_i	7.3	5.1	8.6	9.8	7.9	8.o	9.2	8.5	6.4	7.2

Ernteerträge x_i und y_i, $i = 1, \ldots, 1o$, in $\frac{kg}{m^2}$ von 1o Versuchsfeldern nach Einsatz von zwei verschiedenen Düngemitteln.

Wurden die durchschnittlichen Ernteerträge bei einer Irrtumswahrscheinlichkeit von o,o5 durch das zweite, neue Düngemittel signifikant gegenüber dem herkömmlichen Düngemittel gesteigert ?

Aufgabe L 25

Ein Verbraucherverband will einen statistischen Test auf die wahre mittlere Füllmenge von 250 g – Fruchtgummitüten des Herstellers „ Hariflo“ durchführen, da sich in letzter Zeit die Klagen der Verbraucher häuften, daß die Füllmenge dieser Tüten zu niedrig sei.

Die Abfüllmenge einer solchen Tüte kann als normalverteilt angesehen werden : $N(\mu\,;\,\sigma^2)$ mit $\sigma^2 = 25\ g^2$.

Der Testaufbau soll zwei Forderungen erfüllen :

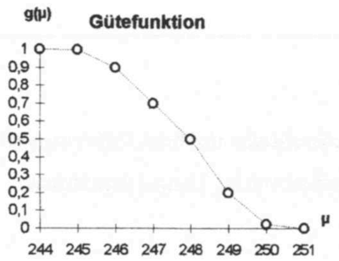
❶ Das Risiko, den Hersteller unberechtigterweise zu verklagen, soll so klein wie möglich gehalten werden und darf den Wert 2,275 % nicht übersteigen.

❷ Die Entscheidungskonstellation

[„*Die mittlere Füllmenge ist zu niedrig*“ / die mittlere Füllmenge beträgt tatsächlich 245 g]

soll eine Wahrscheinlichkeit von 99,865 % besitzen.

Die Skizze der Gütefunktion sieht folgendermaßen aus :



1) Wie lauten die Hypothesen für diesen Test ?

2) Übertragen Sie die Skizze der Gütefunktion in Ihren Lösungsbogen und tragen Sie zusätzlich in die Skizze ein :

   a) α ; b) β ; c) $1-\alpha$; d) $1-\beta$; e) H_O ; f) H_1 !

3) Wie groß muß der Stichprobenumfang **n** der einfachen Stichprobe mindestens sein, damit der Test die beiden Forderungen ❶ und ❷ erfüllt ?

Aufgabe L 26

Die Zeitschrift „ Guten Appetit " möchte mit Hilfe eines statistischen Tests überprüfen, ob in der Kleinstadt Posemuckel im Durchschnitt genauso viel Butter (X) wie Margarine (Y) [gemessen in g pro Woche und Einwohner] auf das Brot geschmiert wird, wobei man davon ausgehen kann, daß Butter und Margarine „ Substitutionsgüter " sind. In der Redaktion der Zeitschrift ist man sich nicht sicher, ob ein *Zweistichprobentest* oder ein *Differenzentest* bei dieser Fragestellung das sinnvollere statistische Verfahren ist.

Es werden folgende Annahmen gemacht :

- $X \sim N(\mu_X \,;\, \sigma^2_X)$ und $Y \sim N(\mu_Y \,;\, \sigma^2_Y)$
- $\sigma^2_X = \sigma^2_Y = \sigma^2$ **unbekannt**

n_X bzw. n_Y seien die jeweiligen Stichprobenumfänge beim ***Zweistichprobentest***

n_D sei der Stichprobenumfang beim ***Differenzentest***

α_Z bzw. α_D seien die Signifikanzniveaus beim ***Zweistichprobentest*** bzw. ***Differenzentest***

- $n_X = n_Y = n_D < 16$
- $\alpha_Z = \alpha_D$

A :

1) Welches Vorzeichen hat **Cov (X , Y)** ?
2) Vergleichen Sie **Var (X − Y)** beim *Zweistichprobentest* und beim *Differenzentest* !
3) Welches Testverfahren sollte demnach durchgeführt werden ? Kurze Begründung !

B :

1) Wie ist die **Testfunktion** unter H_O beim *Zweistichprobentest* bzw. beim *Differenzentest* verteilt ?
2) Vergleichen Sie den **Annahmebereich** beim *Zweistichprobentest* und beim *Differenzentest* !
3) Welches Testverfahren sollte demnach durchgeführt werden ? Kurze Begründung !

Aufgabe L 27 :

Auch im Gesundheitswesen muß gespart werden. Daher wird jede Neueinführung eines medizinischen Präparats nicht nur nach medizinischen, sondern auch verstärkt nach finanziellen Gesichtspunkten geprüft. Diesmal geht es um ein neuentwickeltes Präparat gegen Prüfungsstreß.

1) Formulieren Sie für jede der drei folgenden Testsituationen (A , B und C) die Hypothesen ! Begründen Sie Ihre Wahl – falls nötig – in Form einer Risikoüberlegung !

2) Treffen Sie die Entscheidung über das zu verwendende Signifikanzniveau α („ eher groß “ oder „ eher klein “) ! Begründen Sie kurz Ihre Entscheidung !

3) Skizzieren Sie die Gütefunktion für jeden Test !

A :

Die Pharmaindustrie möchte statistisch nachweisen, daß das neue Präparat zu einem „Heilerfolg “ von über $\pi_0 = 60\,\%$ führt.

B :

Die Krankenkassen wissen noch nicht, ob sie das neue (und teurere) Präparat übernehmen werden. Von ihrer Seite soll getestet werden, ob die mittlere Wirkungsdauer μ_{neu} des neuen Präparats größer ist als die mittlere Wirkungsdauer μ_{alt} des bisher verordneten (billigeren) Präparats.

Obwohl den Patienten kein Präparat vorenthalten werden sollte, das ihnen (besser) helfen könnte, ist für die Krankenkassen jedoch aus finanzieller Sicht viel wichtiger, daß nicht ein neues (und teureres) Präparat bezahlt wird, das letztendlich doch nicht besser ist als das alte Präparat.

C :

Die Ärzteschaft befürchtet, daß das neue Präparat mehr als den zulässigen Grenzwert von $\mu_0 = 0{,}0001$ mg des Wirkstoffes „ Toxin “ enthält, der in Verdacht steht, unangenehme Nebenwirkungen auszulösen. Bevor die Ärzteschaft sich für das Präparat entscheidet, will sie sich auf jeden Fall mit Hilfe eines entsprechenden Tests dagegen absichern.

M

Testtheorie
[nichtparametrisch]

Aufgabe M 1

Einer im Berlin–Verkehr tätigen Fluggesellschaft wird gelegentlich schlampiges Gepäck–Handling vorgeworfen. Gewöhnlich verteidigt sich der Direktor der Fluggesellschaft gegenüber Journalisten mit dem Hinweis, daß es sich bei diesen Vorkommnissen um „ seltene Ereignisse “ handle. Als der Direktor durch Zufall erfährt, daß es eine statistische Verteilung gibt, die gerade „ Verteilung der seltenen Ereignisse “ heißt, wittert er eine Chance, seine Behauptung statistisch zu untermauern. Dies soll durch einen statistischen Test geschehen, bei dem die Behauptung „ Nichtmitnahme von Passagiergepäck mit einem Flugzeug dieser Gesellschaft unterliegt der Verteilung der seltenen Ereignisse “ überprüft werden soll ($\alpha = 1\,\%$) .
Von der Vertretung der Airline in Berlin läßt er sich die Beschwerdeübersicht der letzten 1.000 Flüge vorlegen und betrachtet dies als einfache Stichprobe.
Danach wurde in 460 Fällen ohne Beanstandung abgefertigt. In 350 Fällen wurde das Gepäck jeweils eines Passagiers nicht befördert. Für 135 Flüge war dies bei zwei Fluggästen, für 40 Flüge bei drei Passagieren und für 15 Flüge bei vier Berlin–Reisenden der Fall. In keinem Fall hatten mehr als vier Fluggäste diesbezüglich Grund zur Beanstandung.

1) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

2) Wie lautet die Testfunktion und wie ist diese unter H_O verteilt ?

3) Bestimmen Sie den Verwerfungsbereich für diesen Test !

4) Welche Schätzfunktion muß nach dem Maximum–Likelihood–Prinzip verwendet werden, falls zur Durchführung dieses Tests ein Parameter geschätzt werden muß ?

5) Wie lautet die Testentscheidung ?

   Hinweis : Runden Sie die Werte aus der Verteilungstabelle auf drei Nachkommastellen !

6) Nach Kenntnisnahme des Testergebnisses gibt der Direktor folgende Presseerklärung :

   „ Vorwürfe widerlegt ! Mittels eines statistischen Tests auf der Grundlage einer Stichprobe vom Umfang 1.000 konnte bei einer Irrtumswahrscheinlichkeit von nur 1 % statistisch bewiesen werden, daß fehlerhaftes Gepäck–Handling nur selten vorkommt ! “

   Nehmen Sie Stellung zu dieser Behauptung !
   Begründen Sie Ihre Stellungnahme !

Aufgabe M 2:

Drei unterscheidbare Münzen wurden insgesamt 240 mal geworfen, und jedesmal wurde die erscheinende Anzahl von "Kopf" beobachtet.

Die Ergebnisse sind im folgenden zusammengefaßt:

0	mal Kopf	:	24
1	mal Kopf	:	108
2	mal Kopf	:	85
3	mal Kopf	:	23

Testen Sie die Hypothese, daß es sich bei den drei Münzen um ideale Münzen handelt ($\alpha = 0{,}05$) !

Aufgabe M 3

In einer Fußball–Liga wurden für eine Saison die Torerfolge pro Spiel in folgender Häufigkeitstabelle zusammengefaßt :

Torerfolge pro Spiel	0	1	2	3	4	5	6	7	8	9	> 9
Häufigkeit	18	24	56	63	61	39	26	6	5	2	0

Es soll mit einem statistischen Testverfahren überprüft werden, ob die Torerfolge pro Spiel einer Poissonverteilung mit $\lambda = 3{,}4$ folgen ($\alpha = 0{,}1$) !

1) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

2) Wie lautet die Testfunktion und wie ist diese unter H_O verteilt ?

3) Bestimmen Sie den Verwerfungsbereich für diesen Test !

   Hinweis : Runden Sie die Tabellenwerte auf zwei Nachkommastellen !

4) Wie lautet die Testentscheidung ?
   Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

Aufgabe M 4

Ein Taxiunternehmen will überprüfen, ob sich eine Abhängigkeit zwischen der Wetterlage und der Geschäftslage nachweisen läßt. Dazu werden einige Tage des vergangenen Jahres zufällig ausgewählt (einfache Stichprobe). Folgende Angaben stehen zur Verfügung :

- Von insgesamt 20 Regentagen gab es ebensoviel mit guter wie mit schlechter Geschäftslage.
- 5 Sonnentage brachten ein normales Geschäft.
- An 15 Tagen war das Geschäft schlecht.
- Ein gutes Geschäft konnte an 15 Sonnentagen beobachtet werden.
- Das Geschäft lief an 15 Tagen normal.

1) Stellen Sie das Ergebnis in einer Kontingenztafel dar !
2) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !
3) Ist die Anwendung der χ^2 – Verteilung hier gerechtfertigt ? Begründung !
4) Wie lautet die Testentscheidung, falls gilt :

   a) $\alpha = 1\,\%$

   b) $\alpha = 5\,\%$?

5) Welcher Fehler kann im Fall a) bzw. b) unterlaufen sein ?

Aufgabe M 5

Der Mieterverein von Bärenhausen kämpfte vor einiger Zeit gegen die Aufhebung der Mietpreisbindung für Altbauwohnungen. Alles, was er jedoch erreichen konnte, war eine Einigung darüber, daß der Mietpreis jährlich um maximal 5 % angehoben werden darf. Ein Jahr nach Inkrafttreten des Gesetzes veröffentlicht der Verein der „ Baulöwen und Großgrundbesitzer " folgende Meldung :

> „ Entgegen allen Befürchtungen der Mieter lag die durchschnittliche Mietpreissteigerung in den Altbauwohnungen im letzten Jahr nur bei 2,5 % . Ferner ist festzustellen, daß die Höhe der einzelnen Mietpreissteigerungen einer Gleichverteilung innerhalb des vereinbarten Bereichs folgt. "

Der Mieterverein bezweifelt diese Zahlen und führt deshalb selbst eine Erhebung vom Umfang n = 100 (einfache Stichprobe) durch. Die Erhebung brachte folgendes Ergebnis :

Mietpreissteigerung in %	h_i
0 – 1	0
1 – 2	0
2 – 3	10
3 – 4	10
4 – 5	40
> 5	40

1) Mit welchem statistischen Testverfahren kann der Mieterverein die Verteilungsannahme der Gegenseite überprüfen (α = 0,5 %) ?

2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

3) Wie lautet die Testfunktion ?

4) Wie ist die Testfunktion unter H_0 verteilt ?

5) Bestimmen Sie den Verwerfungsbereich für diesen Test !

6) Wie lautet die Testentscheidung ?

7) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

Aufgabe M 6

A

Der Generaldirektor einer Ferienhotelkette ist durch intensive Studien seiner Unterlagen zu der Auffassung gelangt, daß die Aufenthaltsdauer der Gäste seiner Hotels folgende Verteilung besitzt, die sich von Hotel zu Hotel nur durch den jeweiligen Wert des Parameters π unterscheidet :

x	1	2	3	4
$P(X=x)$	π	$1-4\cdot\pi$	$2\cdot\pi$	π

X : „ Aufenthaltsdauer (in Wochen) "

Nach Eröffnung eines neuen Hotels möchte er für dieses den Wert π schätzen. Aus den Geschäftsunterlagen läßt er $n = 100$ Buchungen zufällig auswählen (einfache Stichprobe) .

Aufenthaltsdauer	Häufigkeit
1	20
2	40
3	20
4	20

1) Bestimmen Sie auf der Basis dieser Stichprobe den Maximum – Likelihood – Schätzwert für den unbekannten Parameter π !
   Empfehlung : Verwenden Sie die logarithmierte Likelihood – Funktion !

B

Der Manager dieses Hotels zweifelt an der Auffassung des Generaldirektors über die Verteilung der Aufenthaltsdauer. Mit Hilfe eines statistischen Tests ($\alpha = 0{,}1$) möchte er diesem Zweifel nachgehen. Dabei plant er, auf die obige Stichprobe zurückzugreifen.

1) Welches Testverfahren kann er verwenden ?

2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

3) Wie lautet die Testfunktion ? Wie ist diese unter H_O verteilt ?

4) Bestimmen Sie den Verwerfungsbereich für diesen Test !

5) Wie lautet die Testentscheidung ?

Die Sekretärin des Hotelmanagers hätte den Test mit einer geringeren Irrtumswahrscheinlichkeit ($\alpha = 0{,}01$) durchgeführt, um dem Generaldirektor mit besser abgesicherten Ergebnissen entgegentreten zu können.

6)
   a) Wie würde in diesem Fall die Testentscheidung lauten ?
   b) Wie würde sich ein (eventueller) Unterschied zu Punkt 5) begründen lassen ?

Aufgabe M 7

Ein Versicherungsunternehmen hatte von 1989 bis 2004 die folgenden Schadenssummen in Millionen € zu begleichen :

0	1	1	2	4	4	5	6	10	12	14	19	22	36	51	75

Die Versicherungsgesellschaft kalkuliert ihre Pläne derzeit unter der Annahme, daß der Median der zufälligen jährlichen Schadenssumme bei 5 Millionen € liegt. Um die Prämie erhöhen zu können, will die Gesellschaft mit Hilfe eines statistischen Tests versuchen nachzuweisen, daß der Median größer ist als 5 Millionen € .

1) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

2)
   a) Treffen Sie Ihre Entscheidung mit Hilfe des ***Vorzeichentests*** ($\alpha = 0.1$) !
   b) Welche Voraussetzungen müssen für diesen Test erfüllt sein ?

3)
   a) Treffen Sie Ihre Entscheidung mit Hilfe des ***Vorzeichenrangtests von Wilcoxon*** ($\alpha = 0.1$) !
   b) Welche Voraussetzungen müssen für diesen Test erfüllt sein ?

4) Begründen Sie, warum man bei beiden Testverfahren zu unterschiedlichen Testergebnissen kommt !

5) Welchen Test würden Sie durchführen, falls man annehmen könnte, daß die Verteilung der Schadenssumme einer Normalverteilung folgt ? Kurze Begründung !

Aufgabe M 8

Ein Mediziner will einen Artikel für eine Fachzeitschrift schreiben. Sein derzeitiges Forschungsgebiet beschäftigt sich damit, ob es einen Zusammenhang zwischen dem Körpergewicht und dem Cholesterinwert gibt. Da er von einem derartigen Zusammenhang überzeugt ist, möchte er seine These mit Hilfe eines statistischen Tests ($\alpha = 0{,}01$) untermauern. Zu diesem Zweck hat er 500 zufällig ausgewählte Personen befragt. Leider jedoch ist seine Datenauswertung in einem unbeobachteten Augenblick in die Pfoten seines cholesterinkranken und übergewichtigen Schäferhundes geraten, so daß nur noch einige Daten übriggeblieben sind :

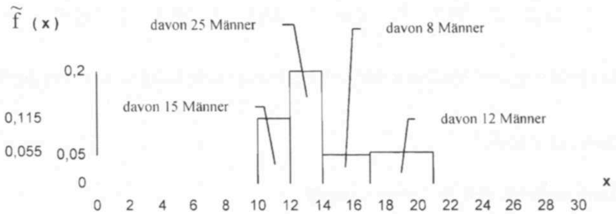
Cholesterin / Gewicht	normal	hoch	sehr hoch	
normal	115	**57,8**	10	
bis 20 kg über normal	**44,8**	85	**60,8**	
mehr als 20 kg über normal	5			
			190	

$\tilde{h}_{ij}$
h_{ij}

1) Ermitteln Sie die verlorengegangenen Daten und vervollständigen Sie die Kontingenztafel !
2) Welches Testverfahren sollte hier verwendet werden ?
3) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !
4) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?
5) Bestimmen Sie den Annahme– und Verwerfungsbereich für diesen Test !
6) Wie lautet die Testentscheidung ?
7) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

Aufgabe M 9

Der Verein der emanzipierten Männer Berlins will selbst einen Test ($\alpha = 0{,}1$) durchführen, um zu überprüfen, ob es einen Unterschied zwischen Männern und Frauen hinsichtlich der Eitelkeit gibt. Von 60 Männern und 40 Frauen (jeweils einfache Stichproben) werden die Umkleidezeiten in Bekleidungsgeschäften gemessen. Über die Wahrscheinlichkeitsverteilung von Umkleidezeiten wird keine Annahme getroffen.
Aus Gründen der Übersichtlichkeit wurden die Daten in Gruppen zusammengefaßt.
Es ergab sich folgendes Histogramm :



mit $\tilde{f}(x) = f_j / b_j$

1) Wie heißt das hier zu verwendende „ verteilungsfreie “ Testverfahren ?

2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – <u>falls nötig</u> – mit Hilfe einer Risikoüberlegung !

3) Wie lautet die Testfunktion und wie ist diese unter H_O verteilt ?

4) Bestimmen Sie den Annahme– und Verwerfungsbereich für diesen Test !

5) Wie lautet die Testentscheidung ?

6) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

Aufgabe M 1o:

Eine Reifenfirma hat einen neuen Reifentyp entwickelt, dessen Lebensdauer (in 1.ooo km) als symmetrisch verteilt angenommen wird. Auf Basis einer einfachen Stichprobe vom Umfang $n = 21$ soll getestet werden, ob die Hypothese $\tilde{\mu} \geq \tilde{\mu}_0 (= 33)$ statistisch widerlegt werden kann ($\alpha =$ o,o5).
Es ergaben sich folgende (bereits geordnete) Werte:

23,2 / 23,8 / 26,o / 26,9 / 27,4 / 28,o / 3o,3
3o,5 / 3o,9 / 31,o / 31,3 / 32,8 / 33,o / 33,1
33,9 / 34,3 / 34,7 / 34,9 / 35,1 / 35,1 / 42,6

Führen Sie zur Überprüfung des Medians der Lebensdauerverteilung des neuen Reifentyps

1) den Vorzeichentest durch !

2) den Vorzeichenrangtest von Wilcoxon durch

   a) exakt

   b) approximativ

Aufgabe M 11

A

Um zu prüfen, ob ein Medikament gegen Grippe wirkt, wurden von 326 Kranken genau 163 mit diesem Medikament behandelt (Gruppe A), die andere Hälfte der Kranken (Gruppe B) erhielt ein Medikament ohne Wirkstoff.

Bei Gruppe A zeigte sich bei 105 und bei Gruppe B bei 88 eine baldige Besserung.

Testen Sie die Hypothese, daß beide Medikamente die gleiche Wirkung haben ($\alpha = 0{,}05$) !

B

Major v. Drill (verarmter Adel) hat die Absicht, sich durch eine Heirat zu sanieren. Dabei ergeben zwei gleichlautende Anzeigen in den Zeitungen A und B genau $n_A = 50$ bzw. $n_B = 40$ Zuschriften, die als unabhängige Stichproben aufgefaßt werden können. Aus den Zuschriften ging hervor, daß die Anteile $f_A = 1/2$ und $f_B = 1/8$ der heiratswilligen Damen begütert waren.

Unzufrieden mit dem Erfolg dieser Aktion erwägt der Major eine neue Anzeige, in der seine Vorzüge und Absichten etwas deutlicher hervorgehoben werden sollen.

1) Läßt sich statistisch nachweisen, daß die Wahl der Zeitung (A oder B) von Bedeutung ist ($\alpha = 0{,}025$) ?

2) Zu welcher Entscheidung würde der Major gelangen, falls sich folgende Anteile ergeben hätten :

$$f_A = 0{,}4 \text{ und } f_B = 0{,}25 \quad ?$$

C

Um eine ausgeschriebene Mitarbeiterstelle bewerben sich insgesamt 40 Interessenten. Die Bewerber werden einem psychologischen Test unterzogen und zusätzlich vom Personalchef beurteilt.
Die möglichen Resultate beider Beurteilungen sind „ geeignet " und „ ungeeignet ".

Vom Personalchef als geeignet wurden 22 Personen eingestuft, von denen wiederum 16 auch vom Test als geeignet beurteilt wurden.
Von den 12 aus dem Test als ungeeignet hervorgegangenen Bewerbern sind genau 6 auch beim Personalchef durchgefallen.

Sind beide Beurteilungsverfahren als identisch anzusehen ($\alpha = 0{,}05$) ?

Aufgabe M 12

Richtig oder falsch ?

1) Fällt bei einem Parametertest der Wert der Testfunktion in den Ablehnbereich, kann auf keinen Fall ein Fehler 2.Art begangen worden sein.

2) Bei einem Einstichproben – Gaußtest auf den Parameter μ ist die Wahrscheinlichkeit für den Fehler 1.Art am größten, falls gilt : $\mu = \mu_0$.

3) Ein „ Niveau – α – Test " heißt **konservativ** , falls der Test die vorgegebene Wahrscheinlichkeit für den Fehler 2. Art nicht voll ausschöpft.

4) Wird bei einem Chi ^{2} – Anpassungstest H_0 verworfen, so ist signifikant untermauert, daß die untersuchte Zufallsvariable nicht poissonverteilt ist.

5) Der McNemar–Test gehört zu den sog. „ nichtparametrischen " Testverfahren und untersucht mit Hilfe eines „ Vorher–Nachher–Vergleichs " die Wechselwirkung von zwei Zufallsvariablen anhand zweier abhängiger Stichproben.

6) Aus der Gütefunktion eines Tests kann man ablesen, ob man eine falsche Entscheidung getroffen hat.

7) Ein „ Niveau – α – Test " heißt **gleichmäßig bester** ($\cong$ trennschärfster) Test, falls kein anderer „ Niveau – α – Test " auf dem gesamten Hypothesenbereich H_1 eine steilere Gütefunktion besitzt.

8) Bei den sog. „ verteilungsfreien" Tests benötigt man keine Testfunktion und folglich auch keine Verteilung derselben, um den Test durchzuführen.

9) Parametrische Tests sind dadurch gekennzeichnet, daß die theoretische Verteilung des zugrundeliegenden Merkmals durch einzelne Parameter eindeutig charakterisiert werden kann. Somit ist es möglich, den Test ausschließlich auf den Wert eines solchen Parameters zu beziehen.

10) Beim Einstichproben–Gaußtest auf μ ist die Wahrscheinlichkeit, den Fehler 2.Art zu begehen, begrenzbar, da der Verteilungstyp der Prüfgröße unter H_1 bekannt ist und der Annahme– und Ablehnbereich nicht vom Stichprobenergebnis abhängen.

Aufgabe M 13 :

Die Meisterschaft im Triathlon steht bevor. Der ehrgeizige Trainer der Athleten weiß zwar, daß Dopingmittel verboten sind, versucht aber die Leistung seiner Sportler mit einem Magnesiumpräparat zu steigern. Er glaubt, daß Magnesium die Übersäuerung der Muskeln vermindert und somit die sportliche Leistung steigert. Magnesium zählt zwar nicht zu den Dopingmitteln, kann aber in zu hoher oder zu niedriger Konzentration im Blut zu Schäden am menschlichen Organismus führen. Man kann davon ausgehen, daß die Magnesiumkonzentration im Blut symmetrisch verteilt ist, wobei der mittlere Normwert (ohne zusätzliche Einnahme von Magnesium) bei 2,4 mval pro Liter Blut liegt.
An 12 ausgewählten Sportlern (einfache Stichprobe) will er mit Hilfe statistischer Tests überprüfen, ob es sich statistisch absichern läßt, daß durch die zusätzliche Einnahme von Magnesium

A : eine Leistungssteigerung eintritt [gemessen in : „Benötigte Zeit beim 10.000 Meter-Lauf"]

B : ein Risiko für den menschlichen Organismus besteht [gemessen in : „Magnesiumkonzentration im Blut"]

Nach der zusätzlichen Einnahme von Magnesium erhielt man folgende Daten von den 12 Sportlern :

Sportler Nr. i	Benötigte Zeit auf 10.000 m	Magnesiumkonzentration im Blut [mval / l]
1	schneller (als sonst)	2,3
2	schneller (als sonst)	2,7
3	gleich	2,4
4	langsamer (als sonst)	2,9
5	schneller (als sonst)	2,3
6	schneller (als sonst)	2,7
7	langsamer (als sonst)	2,4
8	schneller (als sonst)	2,5
9	schneller (als sonst)	2,6
10	schneller (als sonst)	2,8
11	schneller (als sonst)	2,7
12	langsamer (als sonst)	2,2

Führen Sie für **Fall A** und **Fall B** jeweils einen statistischen Test auf einem Signifikanzniveau von 10 % durch, der den Informationsgehalt der Daten bestmöglich ausschöpft !

Beantworten Sie in beiden Fällen folgende Fragen :

1) Wie heißt das verwendete Testverfahren ?
2) Wie lauten die Hypothesen des Tests ?
3) Wie lautet die Testfunktion des Tests ?
4) Wie ist die Testfunktion unter H_0 verteilt ?
5) Wie lauten Annahme- und Verwerfungsbereich des Tests ?
6) Wie lautet die Testentscheidung ?
7) Wie ist das Testergebnis inhaltlich (und statistisch exakt) zu interpretieren ?

Würden Sie dem Trainer aufgrund der Testergebnisse raten, seinen Sportlern bei der bevorstehenden Meisterschaft zusätzlich Magnesium zu verabreichen ? Begründen Sie **kurz** Ihre Antwort !

Aufgabe M 14

Der engere Kader einer Spitzenmannschaft im Fußball bestehe aus 16 Spielern. Da in der jüngsten Vergangenheit der Erfolg auf dem Rasen ausblieb (d.h. es wurde häufig verloren), ging das Präsidium nach folgendem Plan vor :

Fall A : Zuerst wurden die Spieler im Training beobachtet, und ihre Einsatzbereitschaft und Motivation wurde in einer Rangskala von 1 bis 6 beurteilt, wobei nur eine Bewertung von „ 1 " bzw. „ 2 " auf ein profihaftes Verhalten hindeuten.

Dann wurde der Trainer entlassen und durch einen neuen Trainer ersetzt.

Drittens wurde die Einsatzbereitschaft und Motivation der Spieler im Training unter Leitung des neuen Trainers noch einmal beurteilt :

Spieler	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
alter Trainer	2	3	3	3	1	4	2	2	5	2	3	4	3	1	6	3
neuer Trainer	2	2	1	3	2	2	1	3	2	4	1	2	2	3	6	2

Um dem Vorwurf entgegenzuwirken, daß man im Fall A nach nicht objektiv nachvollziehbaren Kriterien urteile, beschließt das Präsidium folgendes zusätzliches Vorgehen :

Fall B : Zum Trainingsprogramm fast jeden Trainers gehört das Durchführen von „Kurz-Sprints" über die Distanz von 30 Meter, um die Antrittsschnelligkeit der Spieler zu testen. Die erreichte Laufzeit aller 16 Spieler beim alten Trainer war symmetrisch um den zentralen Mittelwert von 4,3 Sekunden verteilt. Die Ergebnisse aller Spieler unter Leitung des neuen Trainers werden vom Präsidium notiert und lauten :

Spieler	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
erreichte Zeit	4,8	3,6	4,4	4,4	4,0	5,0	4,3	3,5	4,5	4,7	4,5	4,1	3,2	4,8	4,4	3,5

Das Präsidium will nun mit Hilfe statistischer Testverfahren überprüfen, ob durch den Trainerwechsel

A : eine Motivationsänderung bei den Spielern eingetreten ist.

B : ein signifikanter Leistungsschub (im Sinne höherer Sprintgeschwindigkeiten) zu verzeichnen ist.

Führen Sie für den **Fall A** und für den **Fall B** jeweils einen statistischen Test auf einem Signifikanzniveau von 5 % durch ! Beantworten Sie in beiden Fällen folgende Fragen :

1) Wie heißt das verwendete Testverfahren ?
2) Wie lauten die Hypothesen des Tests ?
3) Wie lautet die Testfunktion ?
4) Wie ist die Testfunktion unter H_O verteilt ?
5) Wie lauten Annahme- und Verwerfungsbereich des Tests ?
6) Wie lautet die Testentscheidung ?
7) Wie ist das Testergebnis inhaltlich und statistisch exakt zu interpretieren ?

Können Sie aufgrund der Testergebnisse dem Präsidium raten, öfter einmal den Trainer zu wechseln ?

Aufgabe M 15

A

In einem Dorf finden demnächst Bürgermeisterwahlen statt. Es bewerben sich zwei Kandidaten für dieses Amt : der bisherige Bürgermeister und der jüngst zugereiste Bäckermeister. Da die wahlberechtigten Bürger letzteren noch nicht gut genug kennen, hält dieser im Dorfgasthaus eine längere Wahlrede.

Die **60** anwesenden wahlberechtigten Bürger wurden vor und nach dieser Rede befragt, ob sie den Bäckermeister wählen würden [+] oder dem amtierenden Bürgermeister bei der Wahl ihre Stimme geben wollen [-] .

Von den **44** Bürgern, die vor der Rede gegen den Bäckermeister waren, wollen ihm nach der Rede genau **21** ihre Stimme geben. Nach seiner Rede wollen ihn jedoch noch immerhin **33** Bürger nicht wählen.

Prüfen Sie mit Hilfe eines statistischen Testverfahrens ($\alpha = 2{,}5$ %), ob sich die Wahlchancen des Bäckermeisters durch seine Rede signifikant verändert haben !

1) Wie lautet das hier zu verwendende Testverfahren ?

2) Wie lauten die Hypothesen dieses Tests ? Begründen Sie - <u>falls nötig</u> - Ihre Wahl mit Hilfe einer Risikoüberlegung !

3) Wie lautet die Testfunktion ?

4) Wie ist diese Testfunktion unter H_O verteilt ?

5) Bestimmen Sie den Annahme- und Verwerfungsbereich dieses Tests !

6) Wie lautet die Testentscheidung ?

7) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

B

Erläutern Sie <u>kurz</u>, was man unter einer „ **verbundenen Stichprobe** " versteht !

C

Erläutern Sie <u>kurz</u>, wodurch sich sog. „ **nichtparametrische Testverfahren** " von den parametrischen Testverfahren unterscheiden !

Aufgabe M 16 :

Der Öko–Landwirt Peter Hagen aus Klein–Kleckersdorf besitzt zwölf Milchkühe. Die frisch gemolkene Milch wird von den Schülern der Dorfschule in der Frühstückspause getrunken. Da die Milch den Kindern sehr gut schmeckt und die Milchmenge leider nur für ein einziges Glas pro Schulkind reicht, möchte er wissen, ob er die Milchproduktion durch den Einsatz des Kraftfutters „ Milch–Max " erhöhen soll. Er befürchtet allerdings, daß der gesteigerte Milchertrag zu einer Verschlechterung der Milchqualität führen könnte. Er hält vor allem den Calcium–Gehalt der Milch für besonders wichtig, da für einen gesunden Knochenbau bei heranwachsenden Kindern eine ausreichende Versorgung mit Calcium gewährleistet sein muß, d.h. der mittlere Calcium–Wert von 120 g (pro 100 ml Milch) sollte nicht unterschritten werden.
Man kann davon ausgehen, daß der Calcium–Gehalt der Milch symmetrisch um den Normwert von 120 g (pro 100 ml Milch) verteilt ist.

Peter Hagen füttert probeweise seine zwölf Milchkühe mit dem Kraftfutter „ Milch–Max " und ermittelt jeweils den Calcium–Gehalt der Milch. Diese Untersuchung brachte folgendes Ergebnis :

Milchkuh Nr.	1	2	3	4	5	6	7	8	9	10	11	12
Calcium–Gehalt in g pro 100 ml	119,8	122,7	117,2	121,3	119,6	119,3	120,0	123,4	118,6	117,0	119,3	120,3

Entscheiden Sie mit Hilfe eines statistischen Tests ($\alpha = 5\ \%$), der den Informationsgehalt der Daten bestmöglich ausschöpft, ob Peter Hagen seine Milchkühe weiterhin mit „ Milch–Max " füttern soll, wenn das Risiko, daß die Kinder mehr Milch bei verminderter Qualität trinken werden, so klein wie möglich gehalten werden soll.

1) Wie lautet das hier zu verwendende Testverfahren ?
2) Wie lauten die Hypothesen ? Begründen Sie Ihre Wahl – falls nötig – in Form einer Risikoüberlegung !
3) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?
4) Bestimmen Sie den Verwerfungsbereich für diesen Test !
5) Wie lautet die Testentscheidung ?
6) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

Aufgabe M 17 :

A

An der Volkshochschule in Hintertupfingen müssen beim Abschlußdiplom in „ Statistik für den Hausgebrauch “ eine mündliche und eine schriftliche Prüfung abgelegt werden. Die Prüfungen werden jeweils mit einer Note bewertet. Im folgenden sind die Prüfungsergebnisse von zufällig ausgewählten Prüflingen dargestellt :

Mündlich / Schriftlich	sehr gut, gut	befriedigend, ausreichend	nicht bestanden	Σ
sehr gut, gut	**55**	**70**	**25**	**150**
befriedigend, ausreichend	**45**	**75**	**50**	**170**
nicht bestanden	**25**	**55**	**50**	**130**
Σ	**125**	**200**	**125**	**450**

Prüfen Sie mit Hilfe eines geeigneten statistischen Testverfahrens ($\alpha = 5\ \%$) , ob die Ergebnisse der schriftlichen und der mündlichen Prüfung voneinander **unabhängig** sind !

1) Wie lautet das hier zu verwendende Testverfahren ?
2) Wie lauten die Hypothesen für diesen Testen ? Begründen Sie Ihre Wahl – falls nötig – in Form einer Risikoüberlegung !
3) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?
4) Bestimmen Sie den Verwerfungsbereich !

Die Testfunktion hat den Wert $v \approx 20{,}26$.

5) Wie lautet die Testentscheidung ? Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

B

Angenommen, die Ergebnisse der schriftlichen und mündlichen Prüfungen sollen nur noch nach „ bestanden “ und „ nicht bestanden “ unterschieden werden.

Prüfen Sie mit Hilfe eines Mc Nemar – Tests ($\alpha = 5\ \%$) , ob die Beurteilungen der schriftlichen Prüfungen als **identisch** mit denen der mündlichen Prüfungen angesehen werden können !

1) Stellen Sie die Kontingenztafel auf !
2) Wie lauten die Hypothesen für diesen Testen ? Begründen Sie Ihre Wahl – falls nötig – in Form einer Risikoüberlegung !
3) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?
4) Bestimmen Sie den Verwerfungsbereich !
5) Wie lautet die Testentscheidung ? Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

N

Klausuraufgaben

Aufgabe N 1 **[Statistik I]**

A

Student Statistix hat bei der neuen Billigfluggesellschaft NCB („ Never – Come – Back ") einen Flug von Berlin nach Teneriffa gebucht. Aus informierten Kreisen erfährt Statistix, daß diese Gesellschaft nur die ältesten und schlechtesten Maschinen besitzt. Außerdem sind von den 10 Piloten dieser Gesellschaft 8 Piloten absolute Anfänger, und nur 2 Piloten weisen eine langjährige Flugpraxis auf.

Für Statistix Flug werden die Cockpitmannschaften aus diesen 10 Piloten ausgelost :

Aus einer Urne, in der sich Zettel mit den Namen der 10 Piloten befinden, wird im 1. Zug der Flugkapitän und im 2. Zug der Copilot gezogen.

Sei

X : „ Anzahl der ausgewählten erfahrenen Piloten beim ersten Zug "

Y : „ Anzahl der ausgewählten erfahrenen Piloten bei beiden Zügen "

1) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsfunktion von X !
2) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsfunktion von Y !
3) Bestimmen Sie tabellarisch die gemeinsame Wahrscheinlichkeitsverteilung von X und Y !

Angenommen, als Flugkapitän wird ein Anfänger ausgewählt !

4) Wie groß ist dann die Wahrscheinlichkeit, daß wenigstens der Copilot ein erfahrener Pilot ist ?

Angenommen, als Flugkapitän wird ein erfahrener Pilot ausgewählt !

5) Wie groß ist dann die Wahrscheinlichkeit, daß leider der Copilot ein Anfänger ist ?
6) Sind die beiden Zufallsvariablen X und Y stochastisch unabhängig ?
   Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

B

Sei die Zufallsvariable

Z : „ Flugzeit von Berlin nach Teneriffa "

normalverteilt mit $\mu = 5$ und $\sigma^2 = 1$.

Wie groß ist die Wahrscheinlichkeit, daß ein Flug von Berlin nach Teneriffa

1) weniger als 4,5 Stunden
2) mehr als 6 Stunden
3) nicht weniger als 4 , aber auch nicht mehr als 7 Stunden

dauert ?

C

Angenommen, die Zufallsvariable

T : „ Flugzeit von Teneriffa nach Berlin "

sei exponentialverteilt mit einem Erwartungswert von 5 Stunden.

Wie groß ist die Wahrscheinlichkeit, daß ein Flug vonTeneriffa nach Berlin

1) weniger als 4,5 Stunden
2) mehr als 6 Stunden
3) nicht weniger als 4 , aber auch nicht mehr als 7 Stunden

dauert ?

Aufgabe N 2 **[Statistik I]**

A

Der Sicherheitschef der kleinen Fluggesellschaft RGI („ Runter – Geht's – Immer ") weiß, daß alle Maschinen seiner Gesellschaft mehr oder weniger schrottreif sind. Daher wurden alle 20 Passagiere und die 5 Besatzungsmitglieder der gerade gestarteten Maschine nach Hunulolo mit Fallschirmen ausgerüstet, die leider auch nicht auf dem neuesten Sicherheitsstand sind und sich jeweils nur mit 80 %–iger Wahrscheinlichkeit öffnen.

Wie groß ist im Falle eines Absturzes dieser Maschine die Wahrscheinlichkeit dafür, daß sich

1) genau 5 Fallschirme nicht öffnen ?

2) mindestens 4 Fallschirme nicht öffnen ?

3) mehr als 19 , aber weniger als 23 Fallschirme öffnen ?

B

Als um 12 Uhr der Funkkontakt zu diesem Flugzeug abreißt, eilt kurze Zeit später die Sekretärin des Sicherheitschefs in dessen Büro :

„ Zuerst die gute Nachricht : Nach Augenzeugenberichten haben sich alle Fallschirme geöffnet !

Nun die schlechte Nachricht : Die Maschine ist über dem Haifischdreieck abgestürzt ! "

Schreckensbleich kramt sich der Sicherheitschef eine Statistik aus seinem Schreibtisch hervor, wonach die Haie in diesem Gebiet völlig unregelmäßig und unabhängig voneinander auftauchen. Durchschnittlich lassen sich dort 2 Haie in 10 Stunden blicken.
Von amtlicher Seite erfährt man, daß die Bergung der Verunglückten eine Stunde dauern wird.

1) Wie ist die Zufallsvariable

   Y : „ Anzahl der innerhalb der Bergungszeit auftauchenden Haie "

   verteilt ? [Verteilungstyp und –parameter !]

Wie groß ist die Wahrscheinlichkeit dafür, daß innerhalb der Bergungszeit

2) kein Hai auftaucht ?

3) mehr als 2 Haie auftauchen ?

4) mindestens ein Hai, aber weniger als drei Haie auftauchen ?

Plötzlich stürzt wieder die Sekretärin in das Büro :

„ Zwei neue schlechte Nachrichten : Erstens haben Sie sich in Ihrer Aufregung verlesen, Chef ! Leider lassen sich dort durchschnittlich 10 Haie in 2 Stunden blicken ! Und zweitens wird die Bergung der Verunglückten wegen der Wetterverhältnisse nun leider 3,2 Stunden dauern ! "

Wie groß ist die Wahrscheinlichkeit dafür, daß innerhalb der nun geltenden Bergungszeit

5) kein Hai auftaucht ?

6) mehr als 2 Haie auftauchen ?

7) mindestens ein Hai, aber weniger als drei Haie auftauchen ?

C

Eine andere Maschine mit 10 Wissenschaftlern an Bord, die in Hunulolo Wasserproben auf Verunreinigungen durch Öl untersuchen sollen, ist zum Glück sicher in Hunulolo gelandet. Die Wissenschaftler haben ihre Arbeit bereits aufgenommen und erfahren von den Kollegen vor Ort, daß nur **1 %** aller Wasserproben (unabhängig voneinander) verunreinigt sein sollen. Daher wird für die weitere Arbeit aus Kostengründen der Vorschlag gemacht, Sammelproben aus jeweils **8** Einzelproben zu bilden und zunächst eine solche Sammelprobe auf Verunreinigungen durch Öl zu untersuchen. Wird in der Sammelprobe kein Öl festgestellt, so kann man sich die Untersuchung der 8 Einzelproben sparen. Andernfalls muß jede der 8 Einzelproben gesondert untersucht werden.

Sei

X : „ Anzahl der verunreinigten Proben innerhalb einer Sammelprobe "

1) Wie ist die Zufallsvariable X verteilt ? [Verteilungstyp und –parameter !]

2) Wie groß ist die Wahrscheinlichkeit, in einer Sammelprobe Öl zu finden ?

Sei

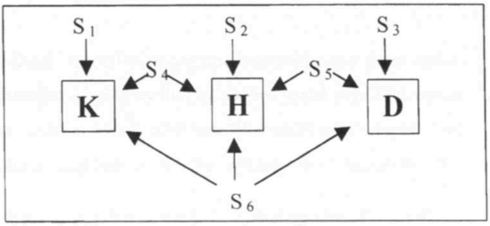
Y : „ Anzahl der Untersuchungen, die für eine Sammelprobe notwendig sind "

3) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsfunktion der Zufallsvariablen Y !

4) Wie viele Untersuchungen werden durchschnittlich für eine Sammelprobe benötigt ?

5) Wieviel Prozent der Untersuchungen können also durch Bildung von Sammelproben im Durchschnitt eingespart werden ?

Aufgabe N 3 **[Statistik I]**



Die Schließanlage eines Hauses besteht aus drei Schlössern :

- Ein Schloß ist in der **Haustür**
- Ein Schloß ist in der **Kellertür**
- Ein Schloß ist in der **Dachbodentür**

Der außerhalb dieses Hauses wohnende Hausverwalter besitzt **6** (gleich aussehende) Schlüssel für die Schließanlage dieses Hauses, die er in einer Schublade aufbewahrt :

- unter diesen (6 Schlüsseln) sind **3** Schlüssel, von denen jeweils einer in genau eines der drei vorhandenen Schlösser paßt.
- der vierte Schlüssel paßt sowohl zur **Haus**tür als auch zur **Keller**tür.
- der fünfte Schlüssel paßt sowohl zur **Haus**tür als auch zur **Dachboden**tür.
- der sechste Schlüssel paßt zu allen drei Schlössern.

Für einen Kontrollgang nimmt der Hausverwalter zufällig **3** von den sechs Schlüsseln aus der Schublade heraus.

1) Wie viele Arten ist dies kombinatorisch möglich ?
2) Wie ist die Zufallsvariable

   Y : „ Anzahl der aus der Schublade entnommenen Schlüssel, die zur **Haus**tür passen "

   verteilt ? [Verteilungtyp und –parameter !]
3) Geben Sie tabellarisch die Wahrscheinlichkeitsfunktion von Y an !
4) Wie ist die Zufallsvariable

   X : „ Anzahl der aus der Schublade entnommenen Schlüssel, die zur **Keller**tür passen "

   verteilt ? [Verteilungtyp und –parameter !]
5) Geben Sie tabellarisch die Wahrscheinlichkeitsfunktion von X an !
6) Geben Sie tabellarisch die gemeinsame Wahrscheinlichkeitsverteilung von X und Y an !
7) Bestimmen Sie die Wahrscheinlichkeit dafür, daß

   a) der Verwalter mit den entnommenen Schlüsseln <u>nicht</u> die **Haus**tür aufschließen kann !

   b) der Verwalter mit den entnommenen Schlüsseln <u>nicht</u> die **Dachboden**tür aufschließen kann !

   c) mindestens zwei der entnommenen Schlüssel zur **Keller**tür passen, falls zwei der entnommenen Schlüssel in das Schloß der **Haus**tür passen.

Aufgabe N 4 [Statistik I]

Anton ist ein begeisterter und sehr guter Golfspieler. Ein Golfplatz besteht aus 18 Löchern, d.h. es gibt 18 Bahnen unterschiedlicher Länge, wobei der Ball auf jeder einzelnen Bahn mit möglichst wenig Schlägen „ eingelocht " werden soll. Wegen der unterschiedlichen Länge (≅ Entfernung vom Abschlag bis zum Loch) der Bahnen, gibt es 3 sog. „ Par–Vorgaben " , die angeben, mit wieviel Schlägen ein guter Spieler den Ball einlochen sollte :

Par 5	: Sehr lange Bahn	[mit 5 Schlägen sollte eingelocht werden]
Par 4	: Mittlere Bahn	[mit 4 Schlägen sollte eingelocht werden]
Par 3	: Kurze Bahn	[mit 3 Schlägen sollte eingelocht werden]

Spielt man eine Bahn mit einem Schlag <u>weniger</u> als vorgegeben, so hat man ein sog. „ Birdie " erzielt.

Spielt man eine Bahn mit <u>genau</u> den vorgegebenen Schlägen, so hat man ein sog. „ Par " erzielt.

Spielt man eine Bahn mit einem Schlag <u>mehr</u> als vorgegeben, so hat man ein sog. „ Bogie " erzielt.

Es kommt praktisch nie vor, daß Anton etwas anderes spielt als „ Birdie " , „ Par " oder „ Bogie " .

Da Anton in seiner Karriere schon auf unzähligen Golfkursen gespielt hat, weiß er aus Erfahrung, daß auf Golfkursen der Anteil der „ Par 3 " – Löcher ca. 25 % , der Anteil der „ Par 4 " – Löcher ca. 55 % und der Anteil der „ Par 5 " – Löcher ca. 20 % beträgt.

Anton kennt die Wahrscheinlichkeit, daß er ein Birdie spielt ganz genau. Sie beträgt auf einer Par 3 – Bahn 40 % , auf einer Par 4 – Bahn 50 % und auf einer Par 5 – Bahn 12,5 % .

1) Zeigen Sie mit Hilfe eines formalen Rechenganges in Ereignisschreibweise, daß die Wahrscheinlichkeit, daß Anton auf einer beliebigen Golfbahn ein <u>Birdie</u> spielt, genau 40 % beträgt !

Die Wahrscheinlichkeit, daß Anton auf einer beliebigen Golfbahn ein Bogie spielt, beträgt 15 % .

Wie groß ist die Wahrscheinlichkeit, daß Anton auf einem beliebigen Golfplatz

2) genau 8 Birdies spielt ?
3) mindestens 5 Bogies spielt ?
4) höchstens 5 mal kein Par erzielt ?
5) mindestens 11 mal keinen Birdie erzielt ?
6) mindestens 6 mal, aber weniger als 10 mal kein Bogie spielt ?
7) erst auf der 6. Bahn sein erstes Birdie spielt ?
8) auf der 7. Bahn seinen dritten Bogie spielt ?
9) 11 Birdies und 5 Bogies spielt ?

10) Wieviel Birdies wird Anton durchschnittlich auf einem beliebigen Golfplatz spielen ?
11) Wieviel Pars wird Anton durchschnittlich auf einem beliebigen Golfplatz spielen ?
12) Wieviel Bogies wird Anton durchschnittlich auf einem beliebigen Golfplatz spielen ?

Aufgabe N 5 **[Statistik I]**

Ein Buchhändler möchte seinen Buchhandel aus Altersgründen verkaufen. Er weiß, daß die potentiellen Käufer Informationen über die Zahl der monatlich von seinen Stammkunden gekauften Bücher haben wollen. Er gibt ihnen folgende Hinweise über das Kaufverhalten seiner in fünf Gruppen eingeteilten Stammkunden :

- ☐ Jeder seiner 200 Stammkunden kauft mehr als ein Buch, aber weniger als 22 Bücher pro Monat.
- ☐ Von den 40 % Stammkunden (der 1. und 2. Gruppe) , die weniger als 8 Bücher im Monat kaufen, erwirbt die Hälfte aber mindestens 4 Bücher im Monat.
- ☐ Von den 20 Stammkunden der 3. Gruppe kauft monatlich jeder mindestens 8 , aber weniger als 10 Bücher.
- ☐ Die restlichen 100 Stammkunden fallen in die 4. bzw. 5. Gruppe und kaufen monatlich weniger als 22 Bücher.
- ☐ Die Stammkunden der 5. Gruppe kaufen monatlich mindestens 14 Bücher.
- ☐ Die Zahl der Stammkunden der 3. und 5. Gruppe ist zusammen fünfmal so hoch wie die Zahl der Stammkunden, die monatlich mindestens 10 , aber weniger als 14 Bücher kaufen (4. Gruppe) .

1) Wie lautet das statistische Merkmal und wie ist es sklaiert ?
2) Geben Sie tabellarisch die absolute und die relative Häufigkeitsverteilung an !
3) Stellen Sie die absolute Häufigkeitsverteilung graphisch dar !
4) Stellen Sie die (relative) Verteilungsfunktion graphisch dar !
   Welche Annahme treffen Sie zu diesem Zweck über die Datenverteilung innerhalb der einzelnen Gruppen ?
5) Bestimmen Sie <u>rechnerisch</u> den Median sowie das untere und obere Quartil !
6) Erstellen Sie eine sog. „ 5 – Zahlen – Zusammenfassung “ und zeichnen Sie den zugehörigen Boxplot !
7) Berechnen Sie das arithmetische Mittel aus den Daten !
8) Berechnen Sie den (empirischen) Quartilskoeffizienten der Schiefe !
   Welche Aussage läßt sich aufgrund seines Wertes über die Schiefe der Häufigkeitsverteilung treffen ?
9) Wieviel Prozent der Stammkunden kaufen monatlich mehr als 18 Bücher ?
10) Wie viele Bücher höchstens kaufen die 70 % weniger kauffreudigen Stammkunden im Monat ?
11) Beschreiben Sie <u>kurz</u> , was <u>inhaltlich</u> der „ mittlere Datenkörper “ angibt !

Aufgabe N 6 **[Statistik I]**

Die Brüder Peter und Hans haben sich dem Kegelsport verschrieben, bei dem es bekanntermaßen darum geht, mit einer Holzkugel möglichst viele von 9 aufgestellten Kegeln abzuräumen. Um ihr Hobby in Zukunft als Leistungssport zu betreiben, wollen sie einem Kegelverein beitreten, wofür sie allerdings folgende Aufnahmeprüfung erfolgreich bestehen müssen :

Bei 64 Kegelversuchen müssen mindestens 40 mal mindestens 7 Kegel geworfen werden.

A

Peter geht davon aus, daß seine „ Kegelerfolge " von Versuch zu Versuch unabhängig voneinander sind und daß die Wahrscheinlichkeit 80 % beträgt, in einem beliebigen Versuch mindestens 7 Kegel abzuräumen.

1) Wie ist die Zufallsvariable

   X : „ Anzahl der Versuche, bei denen Peter mindestens 7 Kegel abräumt "

   verteilt ? [Verteilungstyp und –parameter !]

2) Berechnen Sie E (X) und Var (X) !

3) Wie groß ist die Wahrscheinlichkeit, daß Peter

   a) die Aufnahmeprüfung erfolgreich meistert ?

   b) bereits in seinem 11. Versuch zum achtenmal höchstens 2 Kegel stehenläßt ?

B

Auch Bruder Hans geht davon aus, daß seine „ Kegelerfolge " von Versuch zu Versuch unabhängig voneinander sind. Allerdings beträgt für ihn die Wahrscheinlichkeit nur 75 % , in einem beliebigen Versuch mindestens 7 Kegel abzuräumen. Dafür beträgt die Wahrscheinlichkeit, daß er alle Neune wirft (also alle Kegel abräumt) immerhin 5 % .

1) Wie ist die Zufallsvariable

   Z : „ Anzahl der Versuche, bei denen Hans alle Neune wirft "

   verteilt ? [Verteilungstyp und –parameter !]

2) Berechnen E (Z) und Var (Z) !

3) Wie groß ist die Wahrscheinlichkeit, daß Hans

   a) mindestens siebenmal alle Neune wirft ?

   b) mindestens dreimal, aber weniger als achtmal alle Neune wirft ?

   c) in seinen ersten 14 Versuchen nur zweimal weniger als 7 Kegel abräumt und sogar viermal alle Neune wirft ?

4) Wie oft müßte Hans *mindestens* werfen, damit er mit einer Wahrscheinlichkeit von *mindestens* 99 % *mindestens* einmal alle Neune wirft ?

   Hinweis : A : „ Bei n Würfen keinmal alle Neune " ⇨ $P(A) = ?$

   $\overline{A}$: „ Bei n Würfen mindestens einmal alle Neune " ⇨ $P(\overline{A}) = 1 - P(A)$

   gesucht : $P(\overline{A}) \geq 0{,}99$

Aufgabe N 7 **[Statistik I]**

Im Land Fantasia haben **50 %** der Bevölkerung eine lange Nase, **30 %** der Bevölkerung haben kurze Beine und **55 %** der Bevölkerung lügen nie.

- Langnasige, kurzbeinige Lügner gibt es nicht.
- **15 %** der Bevölkerung sind kurznasige, kurzbeinige Lügner.
- **5 %** der Bevölkerung sind kurznasige, kurzbeinige Nichtlügner.
- **20 %** der Bevölkerung sind kurznasige, langbeinige Lügner.

Achtung : In Ihrem Rechengang muß der Lösungsweg in Ereignisschreibweise klar ersichtlich sein !

Hinweis : Benutzen Sie bitte für Ihre Lösung folgende Ereignisdefinitionen :

N : „ Person hat eine lange Nase "
B : „ Person hat kurze Beine "
L : „ Person ist ein Lügner "

Mit welcher Wahrscheinlichkeit ist eine beliebige Person dieser Bevölkerung ein Lügner, falls sie

1) kurze Beine hat ?

2) kurze Beine und eine kurze Nase hat ?

3) eine lange Nase hat ?

Aufgabe N 8 **[Statistik I]**

Beim Tennisspielen ist der Aufschlag von besonders großer Bedeutung.
Ein „ Aufschlagversuch “ besteht aus einem 1. Aufschlag und (eventuell) einem 2. Aufschlag :
Ist nämlich der 1. Aufschlag erfolgreich, so ist der Ball im Spiel, und der 2. Aufschlag muß natürlich nicht mehr durchgeführt werden. Nur wenn der 1. Aufschlag nicht erfolgreich war, wird der 2. Aufschlag durchgeführt. Wenn auch der 2. Aufschlag nicht erfolgreich ist, hat man einen sogenannten „ Doppelfehler “ begangen.
Eine Stärke der Spitzenspielerin Doris Decker ist ihr Aufschlag. So ist in einem Aufschlagversuch jeder ihrer 1. Aufschläge mit 80 %–iger Wahrscheinlichkeit erfolgreich. Dieselbe hohe Erfolgsquote hat sie auch bei jedem ihrer 2. Aufschläge, falls ihr 1. Aufschlag nicht erfolgreich war.

Im folgenden muß der formale Lösungsweg klar ersichtlich sein !

1) Wie groß ist die Wahrscheinlichkeit, daß Doris bei einem Aufschlagversuch nur beim 2. Aufschlag Erfolg hat ?

2) Wie groß ist die Wahrscheinlichkeit, daß Doris bei einem Aufschlagversuch beim 2. Aufschlag Erfolg hat ?

3) Wie groß ist die Wahrscheinlichkeit, daß Doris bei einem Aufschlagversuch mindestens mit einem der beiden Aufschläge erfolgreich ist ?

4) Wie groß ist die Wahrscheinlichkeit, daß Doris bei einem Aufschlagversuch einen Doppelfehler begeht ?

5) Sind die beiden Ereignisse „ 1. Aufschlag ist erfolgreich “ und „ 2. Aufschlag ist erfolgreich “ stochastisch unabhängig ? Begründen Sie Ihre Antwort mit Hilfe eines formalen Rechenganges !

Wie groß ist die Wahrscheinlichkeit, daß bei 24 Versuchen der 1. Aufschlag

6) höchstens 14 man erfolgreich ist ?

7) mindestens 16 mal erfolgreich ist ?

8) mindestens 13 mal, aber höchstens 17 mal erfolgreich ist ?

Wie groß ist die Wahrscheinlichkeit, daß Doris genau im 5. Versuch

9) zum viertenmal Erfolg mit dem 1. Aufschlag hat ?

Wie groß ist die Wahrscheinlichkeit, daß Doris bei 50 Versuchen

10) mehr als 4 Doppelfehler macht ?

11) weniger als 9 Doppelfehler macht ?

12) mehr als 6 , aber höchstens 11 Doppelfehler macht ?

Aufgabe N 9 [Statistik II]

Statistix ist begeistert von einem Geschicklichkeitsspiel, bei dem man durch Ausbalancieren der Spielfläche eine Kugel durch ein Labyrinth von einem Startpunkt zu einem Zielpunkt bringen muß, ohne daß die Kugel dabei unterwegs in eines der vielen in der Spielfläche angebrachten Löcher fällt.
Falls die Kugel vor Erreichen des Zielpunktes in ein solches Loch fällt, ist der Versuch gescheitert und somit nicht erfolgreich. Dann kann einer neuer Versuch vom Startpunkt aus begonnen werden.
Sei

X : „ Anzahl der Versuche bis die Kugel erfolgreich zum Ziel gebracht wird "

Statistix geht davon aus, daß die Zufallsvariable X folgende Wahrscheinlichkeitsverteilung besitzt :

$$P(X=x)=(1-\pi)^{x-1}\cdot\pi \quad [x=1,2,\ldots] \qquad \text{mit} \quad E(X)=\mu=\frac{1}{\pi} \text{ und } Var(X)=\sigma^2=\frac{1-\pi}{\pi^2}$$

Statistix möchte nun den Parameter μ (d.h. die zu erwartende Anzahl der Versuche bis er die Kugel erfolgreich zum Ziel gebracht hat) schätzen. Er glaubt, daß $\overline{X}$ eine geeignete Schätzfunktion sein könnte, ist sich seiner Sache aber nicht sicher. Daher will er mit der Maximum–Likelihood–Methode seine Vermutung überprüfen.

1) Stellen Sie die Maximum–Likelihood–Funktion in Abhängigkeit von π (**nicht** in Abhängigkeit von μ !) auf !

2) Stellen Sie die logarithmierte Maximum–Likelihood–Funktion in Abhängigkeit von π auf !

3) Bestimmen Sie mit Hilfe eines formalen Rechenganges allgemein den Maximum–Likelihood–Schätzwert für π , wenn sich die theoretische (einfache) Stichprobe ($X_1, X_2, \ldots, X_n$) zum Stichprobenergebnis ($x_1, x_2, \ldots, x_n$) realisiert hat !

4) Leiten Sie daraus den Maximum–Likelihood–Schätzwert für μ ab !
5) Wie lautet demnach die Maximum–Likelihood–Funktion $\hat{\Theta}_\mu$ für μ ?
6) Prüfen Sie die Maximum–Likelihood–Funktion $\hat{\Theta}_\mu$ auf Erwartungstreue !
7) Prüfen Sie die Maximum–Likelihood–Funktion $\hat{\Theta}_\mu$ auf Konsistenz !
8) Geben Sie Verteilungstyp und –parameter der Schätzfunktion $\hat{\Theta}_\mu$ für $n > 30$ an !
9) Zeigen Sie mit Hilfe eines formalen Rechenganges, daß gilt : $Var(\hat{\Theta}_\mu) \overset{!}{=} \frac{\mu^2-\mu}{n}$!
10) Leiten Sie daraus eine Schätzfunktion $\hat{\Theta}_1$ für $\frac{\mu^2-\mu}{n}$ ab !
11) Zeigen Sie mit Hilfe eines formalen Beweisganges, daß es sich bei $\hat{\Theta}_1$ um keine erwartungstreue Schätzfunktion für $\frac{\mu^2-\mu}{n}$ handelt , da gilt : $E(\hat{\Theta}_1)=\frac{n+1}{n}\cdot\left[\frac{\mu^2-\mu}{n}\right]$

    Hilfe : $\mathbf{E(Z^2) = Var(Z) + [E(Z)]^2}$

12) Modifizieren Sie $\hat{\Theta}_1$ so, daß Sie nun eine erwartungstreue Schätzfunktion $\hat{\Theta}_2$ für $\frac{\mu^2-\mu}{n}$ erhalten !
13) Geben Sie nunmehr das (zweiseitig symmetrische) Konfidenzintervall für μ an !

Statistix nimmt nun an $n = 35$ aufeinanderfolgenden Tagen sein Geschicklichkeitsspiel einmal in die Hand und spielt so lange, bis er endlich erfolgreich ist. Im Durchschnitt benötigte er 9 Versuche, bis er die Kugel erfolgreich ins Ziel gebracht hat.

14) Berechnen Sie das Schätzintervall für μ (d.h. die zu erwartende Anzahl der Versuche bis er die Kugel erfolgreich zum Ziel gebracht hat) zum Konfidenzniveau $1-\alpha = 95\,\%$!

Aufgabe N 10 **[Statistik II]**

A

Unter Schulkindern ist die Verwendung von sog. „ Tinten–Killern " sehr beliebt. Dabei enthalten diese Tinten–Killer laut Herstellerangaben 2 ml eines Lösungsmittels, das bei höherer Dosierung leider sehr gesundheitsschädlich ist. Die Gesundheitsministerin hat aufgrund sich häufender Beschwerden die Vermutung, daß der Hersteller den Sollwert von 2 ml überschreitet. Sie zieht daher ihren Praktikanten zu Rat, der ihr die Überprüfung ihrer Vermutung mit Hilfe eines statistischen Tests ($\alpha = 5\ \%$) vorschlägt.
Anhand von 100 zufällig aus dem Handel ausgewählten Tinten–Killern (einfache Stichprobe) soll überprüft werden, ob der Hersteller den Sollwert einhält.

Der Praktikant weiß, daß ein unberechtigter Vorwurf gegenüber dem Hersteller seine Chefin in Zugzwang bringen und ihn letztlich seinen Job kosten könnte.

Andererseits erscheint es der Gesundheitsministerin, die selbst Mutter zweier Kinder ist, unverantwortlich, Kinder einer zu hohen Dosierung des Lösungsmittels auszusetzen.

1) Wie heißt das hier zu verwendende parametrische Testverfahren ?

Fall Praktikant	Fall Ministerin
2a) Wie lauten die Hypothesen für diesen Test, wenn der **Praktikant**, der seinen Job behalten will, diesen Test durchführt ? Begründen Sie Ihre Wahl – falls nötig mit Hilfe einer Risikoüberlegung !	2b) Wie lauten die Hypothesen für diesen Test, wenn die verantwortungsbewußte **Ministerin** diesen Test durchführt ? Begründen Sie Ihre Wahl – falls nötig mit Hilfe einer Risikoüberlegung !
3a) Wie lautet die Prüfgröße und wie ist diese unter H_0 verteilt ?	3b) Wie lautet die Prüfgröße und wie ist diese unter H_0 verteilt ?
4a) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?	4b) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?
5a) Bestimmen Sie den Ablehnbereich dieses Tests !	5b) Bestimmen Sie den Ablehnbereich dieses Tests !

Die Stichprobe ergab eine durchschnittliche Menge von 2,05 ml des Lösungsmittels pro Tinten–Killer bei einer Standardabweichung von 0,7 ml .

6a) Wie lautet die Testentscheidung ?	6b) Wie lautet die Testentscheidung ?
7a) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !	7b) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !
8a) Welcher Fehler kann bei der Testentscheidung unterlaufen sein ?	8b) Welcher Fehler ist bei der Testentscheidung unterlaufen ?

B

Student Statistix feiert mit seinen Freunden die bestandene Statistik–Klausur. Am nächsten Morgen erwacht er mit fürchterlichen Kopfschmerzen und kann nun nicht an einem für ihn sehr wichtigen Bewerbungsgespräch teilnehmen. Er ist darüber sehr wütend und bezweifelt, daß der Bierhersteller des von ihm auf der Feier konsumierten Bieres den Sollwert von 5 % Alkoholgehalt je Flasche tatsächlich einhält. Mit Hilfe eines statistischen Tests ($\alpha = 2{,}5\ \%$) will er versuchen nachzuweisen, daß der wahre mittlere Alkoholgehalt je Flasche höher ist als 5 % . Gelingt ihm dieser statistische Nachweis, so will er den Hersteller verklagen.
Statistix geht davon aus, daß der Alkoholgehalt je Flasche eine normalverteilte Zufallsvariable ist mit einer Standardabweichung von 0,9 % .
Um den Test durchzuführen will er bei verschiedenen Einzelhändlern 9 Flaschen des diesbezüglichen Bieres erwerben (einfache Stichprobe) und in einem Labor den Alkoholgehalt je Flasche feststellen lassen.

Er stellt folgende Hypothesen auf :

$$H_0: \mu \leq \mu_0 \ (= 5\,\%)$$

$$H_1: \mu > \mu_0 \ (= 5\,\%)$$

1) Wie heißt das hier zu verwendende Testverfahren ?

2) Wie lautet die Prüfgröße für diesen Test und wie ist diese unter H_0 verteilt ?

3) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?

4) Bestimmen Sie den Annahme– und Verwerfungsbereich für diesen Test !

5) Wie groß ist die Wahrscheinlichkeit für eine richtige Testentscheidung, wenn der wahre mittlere Alkoholgehalt 5,5 % beträgt ?

6) Wie groß ist die Wahrscheinlichkeit für eine falsche Testentscheidung, wenn der wahre mittlere Alkoholgehalt 6 % beträgt ?

7) Welchen Wert müßte der wahre mittlere Alkoholgehalt haben, damit die Wahrscheinlichkeit für eine richtige Entscheidung 90,32 % beträgt ?

8) Skizzieren Sie – unter Verwendung zuvor errechneter Ergebnisse – die Gütefunktion !

9) Die unter **5)** errechnete Wahrscheinlichkeit empfindet Statistix als viel zu gering.

   Wie viele Flaschen Bier hätte Statistix mindestens untersuchen lassen müssen, damit diese Wahrscheinlichkeit mindestens 69,1462 % beträgt ?

Bei der Untersuchung im Labor ergaben sich folgende Werte :

5,75 %	4,5 %	5,35 %	6 %	5,5 %	4,75 %	5,45 %	5,65 %	6,55 %

10) Wie lautet die Testentscheidung ?

11) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt ! Wird Statistix den Hersteller verklagen ?

Aufgabe N 11 [Statistik II]

A

Herr Mai ist Inhaber einer Stoffkäfer–Produktion. Ein Stoffkäfer besteht aus zwei Fühlern, sechs Beinen und einem Körper.
Maschine F stellt die **Fühler** her und hat im Mittel $\mu_F = 15$ Ausfälle pro Jahr bei einer Varianz von $\sigma_F^2 = 16$.
Maschine B stellt die **Beine** her und hat im Mittel $\mu_B = 10$ Ausfälle pro Jahr bei einer Varianz von $\sigma_B^2 = 4$.
Maschine K produziert den **Körper** her und hat im Mittel $\mu_K = 25$ Ausfälle pro Jahr bei einer Varianz von $\sigma_K^2 = 16$.
X_F (Anzahl der Ausfälle einer Maschine vom Typ F pro Jahr) und X_B (Anzahl der Ausfälle einer Maschine vom Typ B pro Jahr) seien normalverteilt.
X_K (Anzahl der Ausfälle einer Maschine vom Typ K pro Jahr) sei beliebig verteilt.

In einer Fabrik hat Herr Mai zwei Maschinen vom Typ F , sechs Maschinen vom Typ B und eine Maschine vom Typ K aufgestellt, um komplette Stoffkäfer zu produzieren.

In Europa besitzt Herr Mai jeweils eine Fabrik in Berlin, Paris, Rom und London.

Man kann davon ausgehen, daß Maschinen eines bestimmten Typs genauso unabhängig voneinander arbeiten wie Maschinen verschiedener Typen unabhängig voneinander arbeiten.

1) Geben Sie für die Zufallsvariablen X_F, X_B und X_K jeweils den Verteilungstyp und die Verteilungsparameter an !

2) Wie viele Ausfälle von Maschinen pro Jahr gibt es im Mittel <u>in einer Fabrik</u> ?

3) Wie groß ist die Wahrscheinlichkeit, daß eine Maschine vom Typ F mehr als 11 Ausfälle pro Jahr hat ?

4) Weniger als wie viele Ausfälle pro Jahr hat eine Maschine vom Typ B mit einer Wahrscheinlichkeit von 99,865 % ?

5) Wie groß ist die Wahrscheinlichkeit, daß eine Maschine vom Typ K zwischen 20 und 30 Ausfälle pro Jahr hat ?

Sei G : „ Anzahl der Ausfälle von Maschinen in Europa insgesamt pro Jahr "

6) Wie ist die Zufallsvariable G verteilt ? [Verteilungstyp und –parameter !]

7) Wie groß ist die Wahrscheinlichkeit, daß in Europa insgesamt pro Jahr zwischen 494 und 511 Maschinen ausfallen ? [**<u>Hinweis</u> :** Runden Sie σ_G auf eine ganze Zahl auf !]

B

Am Samstag nach der Statistik–Klausur soll bei Susi eine Party von 22 Uhr bis 4 Uhr früh mit sicherlich reichlichem Alkoholkonsum stattfinden. Von vergangenen Partys ist Susi bekannt, daß durchschnittlich alle zwei Stunden einem Partygast übel wird und daß (davon unabhängig) durchschnittlich alle vier Stunden ein Partygast Streit sucht.
Susi geht davon aus, daß die Zufallsvariablen

$X_Ü$: „ Anzahl der Partygäste, denen auf der Party übel wird [pro Stunde] "

bzw.

X_S : „ Anzahl der Partygäste, die auf der Party Streit suchen [pro Stunde] "

poissonverteilt sind.

1) Geben Sie für die beiden Zufallsvariablen $X_Ü$ und X_S (den Verteilungstyp und) die Verteilungsparameter an !

2) Nennen Sie die drei Eigenschaften eines Poissonprozesses und problematisieren Sie <u>kurz</u>, inwiefern jede dieser Eigenschaften im vorliegenden Fall verletzt sein könnte !

Gehen Sie im folgenden davon aus, daß diese Eigenschaften <u>nicht</u> verletzt sind !

Sei

$T_Ü$: „ Wartezeit bis dem nächsten Partygast übel wird "

bzw.

T_S : „ Wartezeit bis der nächste Partygast Streit sucht "

3) Wie sind die Zufallsvariablen $T_Ü$ und T_S verteilt ? [Verteilungstyp und –parameter !]

„ Übelkeit “ und „ Streit “ sind schlimme Vorfälle, die Susi auf ihren Partys natürlich gar nicht gern sieht.

4) Wie groß ist die Wahrscheinlichkeit, daß die Party ein voller Erfolg wird (d.h. daß nichts dergleichen passiert) ?

5) Wie groß ist die Wahrscheinlichkeit, daß die Party zumindest ein halber Erfolg wird (d.h. daß Susi wenigstens von einem dieser Vorfälle verschont bleibt) ?

Angenommen, daß sich bis Mitternacht kein schlimmer Vorfall ereignet hat.

6) Wie groß ist die Wahrscheinlichkeit, daß bis zum Ende der Party kein Partygast Streit sucht ?

Um 4 Uhr früh wird die Party (leider nicht von Susi, sondern) von der Polizei beendet. Die Polizei testet den Alkoholspiegel von 49 zufällig ausgewählten Partygästen (einfache Stichprobe).

Dabei ergab sich ein durchschnittlicher Alkoholspiegel von 2 Promille pro Partygast.

Man kann davon ausgehen, daß der Alkoholspiegel eines Partygastes normalverteilt ist mit einer Standardabweichung von 0,25 Promille.

7) Geben Sie explizit das (zweiseitige und symmetrische) Konfidenzintervall für den wahren mittleren Alkoholspiegel μ eines Partygastes zum Konfidenzniveau $1 - \alpha$ an !

8) Bestimmen Sie das Schätzintervall $[v_U ; v_O]$ für μ zum Konfidenzniveau von 95 % !

9) Interpretieren Sie das erhaltene Schätzergebnis inhaltlich und statistisch exakt !

10) Mit welcher Wahrscheinlichkeit liegt der wahre mittlere Alkoholspiegel μ eines Partygastes in diesem Intervall ?

Aufgabe N 12

[Statistik II]

Ein Statistik – Student fährt jeden Tag mit dem Bus zur Universität.

Der Student hat die Vermutung, daß die Busse völlig unregelmäßig verkehren und die Wartezeit auf den Bus durchschnittlich 3 Minuten beträgt.

Der Student möchte nun die Vermutung mit Hilfe eines statistischen Tests zum Signifikanzniveau von 5 % überprüfen. Für die letzten 30 Tage hat er folgende Wartezeiten auf den Bus festgestellt und notiert (einfache Stichprobe) :

Kategorie	Wartezeit auf den Bus [in Min]	beobachtete Häufigkeit
1	0 – 2	12
2	2 – 4	6
3	4 – 6	5
4	6 – 8	4
5	8 – 10	3
6	> 10	0
		30

1) Wie heißt das zu verwendende Testverfahren ?
2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Hypothesenwahl – falls nötig – mit Hilfe einer Risikoüberlegung !
3) Ermitteln Sie die zu erwartenden Häufigkeiten für die Wartezeiten auf den Bus !
4) Wie lautet die Testfunktion, und wie ist diese unter H_0 verteilt ?
5) Bestimmen Sie den Verwerfungsbereich für diesen Test !
6) Wie lautet die Testentscheidung ?
7) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

Angenommen, die durchschnittliche Wartezeit auf den Bus $\vartheta_2 \left[= \frac{1}{\lambda} \right]$ sei nicht bekannt.

8) Geben Sie die Maximum–Likelihood–Funktion zum Schätzen von ϑ_2 an !
9) Geben Sie die logarithmierte Maximum–Likelihood–Funktion zum Schätzen von ϑ_2 an !
10) Leiten Sie mit Hilfe der Maximum–Likelihood–Methode die Schätzfunktion $\hat{\Theta}_2$ für ϑ_2 her !
11) Prüfen Sie, ob die erhaltene Schätzfunktion konsistent ist !
12) Bestimmen Sie mit Hilfe der angegebenen Häufigkeitsverteilung konkret den Maximum–Likelihood–Schätzwert $\hat{\vartheta}_2$ für ϑ_2 !

Aufgabe N 13 [Statistik II]

A

In einem kleinen Land herrschte König Leopold. Jedes Jahr zu seinem Geburtstag ließ er sich von seinen Untertanen mit Vorführungen unterhalten.
Gefiel ihm die Vorstellung sehr, so entlohnte er den Künstler mit 50 Gulden. War er mit der Vorstellung zufrieden, gab er dem Künstler immerhin noch 10 Gulden.
War er jedoch gelangweilt, wurde er wütend und der Künstler mußte das Land verlassen. Fühlte er sich sogar persönlich beleidigt, wurde der Künstler sofort in den Kerker geworfen.

Angenommen, die „ Entlohnung " für die Vorstellung folge der angegebenen Wahrscheinlichkeitsfunktion :

x	50 Gulden	10 Gulden	Land verlassen	Kerker
$P(X=x)$	$1-3\cdot\pi$	$1{,}5\cdot\pi$	$0{,}5\cdot\pi$	π

Angenommen, Sie seien einer der Untertanen, die beim nächsten Geburtstag etwas vorführen sollen. Sie möchten mit der Maximum–Likelihood–Methode die Wahrscheinlichkeit schätzen, in den Kerker geworfen zu werden.
Als Datenbasis (einfache Stichprobe) ziehen Sie alle in dieser Weise begangenen Geburtstagsfeiern heran, wobei der König **a** mal sehr zufrieden war, **b** mal zufrieden war, **c** mal gelangweilt war und **d** mal beleidigt war.

Hinweis : $\mathbf{a + b + c + d = n}$

1) Stellen Sie die Maximum–Likelihood–Funktion auf !
2) Stellen Sie die logarithmierte Maximum–Likelihood–Funktion auf !
3) Ermitteln Sie den Maximum–Likelihood–Schätzwert für π (Wahrscheinlichkeit in den Kerker geworfen zu werden) !
4) Wie lautet demnach die Maximum–Likelihood–Schätzfunktion für π (Wahrscheinlichkeit in den Kerker geworfen zu werden) ?

B

Im Verein der emanzipierten Frauen Berlins ist man der Meinung, daß Männer mindestens so eitel seien wie Frauen. Um dies zu untermauern, will man mit Hilfe eines statistischen Tests ($\alpha = 0{,}13$) anhand der Umkleidezeiten von 20 zufällig ausgewählten Männer (einfache Stichprobe) und von 20 zufällig ausgewählten Frauen (einfache Stichprobe) versuchen nachzuweisen, daß Männer beim Einkaufen in Bekleidungsgeschäften durchschnittlich mehr als 2 Minuten länger in der Umkleidekabine benötigen als Frauen.
Umkleidezeiten können als normalverteilt angesehen werden.
Sowohl bei den Männern als auch bei den Frauen beträgt die Standardabweichung der durchschnittlich in der Kabine verbrachten Umkleidezeit 3 Minuten.

1) Wie lautet das hier zu verwendende Testverfahren ?
2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !
3) Geben Sie die Prüfgröße des Test formal und verbal an !
4) Wie ist die Prüfgröße unter H_0 verteilt ? [Verteilungstyp und –parameter !]
5) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?
6) Bestimmen Sie den Annahme– und Verwerfungsbereich für diesen Test !

Die Datenanalyse brachte folgendes Ergebnis :

- Die Männer brauchten durchschnittlich 18 Minuten bei einer Standardabweichung von 4 Minuten.
- Die Frauen brauchten durchschnittlich 11 Minuten bei einer Standardabweichung von 4 Minuten.

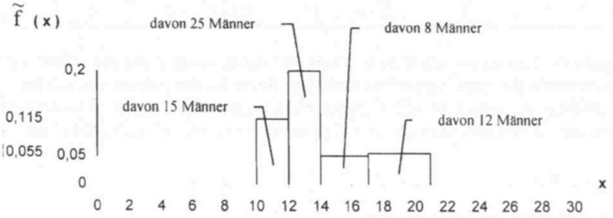
7) Wie lautet die Testentscheidung ?

8) Welcher Fehler kann bei der Entscheidung unterlaufen sein ?

9) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

C

Der Verein der emanzipierten Männer Berlins will selbst einen Test ($\alpha = 0{,}1$) durchführen, um zu überprüfen, ob es einen Unterschied zwischen Männern und Frauen hinsichtlich der Eitelkeit gibt.
Von 60 Männern und 40 Frauen (jeweils einfache Stichproben) werden die Umkleidezeiten in Bekleidungsgeschäften gemessen. Über die Wahrscheinlichkeitsverteilung von Umkleidezeiten wird keine Annahme getroffen.
Aus Gründen der Übersichtlichkeit wurden die Daten in Gruppen zusammengefaßt.
Es ergab sich folgendes Histogramm :



mit $\tilde{f}(x) = f_j / b_j$

1) Wie heißt das hier zu verwendende „ verteilungsfreie “ Testverfahren ?

2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – <u>falls nötig</u> – mit Hilfe einer Risikoüberlegung !

3) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?

4) Bestimmen Sie den Annahme– und Verwerfungsbereich für diesen Test !

5) Wie lautet die Testentscheidung ?

6) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

D

Im Schwimmbad „ Erlebniswelt “ gibt es einen 10–Meter–Sprungturm. Die Jungen, die von diesem Turm springen haben Angst, beim Sprung ihre Badehose zu verlieren. Ein solches Mißgeschick passiert durchschnittlich 3 „ Turmspringern “ pro Stunde.
Angenommen, Sie beobachten das Geschehen die nächste halbe Stunde lang. Danach verlassen Sie das Schwimmbad.

1) Wie groß ist die Wahrscheinlichkeit, höchstens eine halbe Stunde auf ein solches Mißgeschick warten zu müssen ?

2) Angenommen, Sie haben bereits 10 Minuten vergeblich gewartet. Wie groß ist die Wahrscheinlichkeit, daß in der verbleibenden Wartezeit ein solches Mißgeschick passiert ?

3) Wie viele Minuten müßten Sie höchstens warten, damit Sie mit 30 %–iger Wahrscheinlichkeit ein solches Mißgeschick bestaunen können ?

Aufgabe N 14 [Statistik II]

A

Herr B. ist Besitzer einer Baumschule und hat sich auf die sehr mühevolle und (starke Rückenschmerzen verursachende) Aufzucht von Gingkobäumen spezialisiert. Nach einem langen Arbeitsleben lautet daher sein Motto : „ Gesundheit ist wichtiger als Geldverdienen " !
Düngemittellieferant D. bietet ihm nun ein neues und teures Produkt an, welches das Wachstum dieser Bäume deutlich beschleunigt und die Aufzucht daher erheblich erleichtert. D. wirbt damit, daß es bei mehr als 75 % der mit diesem Mittel gedüngten Bäume zu einem Düngeerfolg kommt. B. möchte vor Abschluß eines langfristigen Liefervertrages diese Aussage mit Hilfe eines statistischen Tests ($\alpha = 0{,}05$) überprüfen. Zu diesem Zweck behandelt er 30 Bäume (einfache Stichprobe) mit diesem Mittel und wartet ab, ob sich ein Düngeerfolg einstellt.

1) Wie heißt das hier zu verwendende Testverfahren ?

2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

3) Wie lautet die Testfunktion formal und verbal ?

4) Wie ist diese unter H_0 verteilt ? [Verteilungtyp und –parameter !]

5) Bestimmen Sie den Annahme– und Verwerfungsbereich für diesen Test !

6) Wie groß ist (schlimmstenfalls) die Wahrscheinlichkeit, sich fälschlicherweise für H_1 zu entscheiden ?

7) Bestimmen Sie die Wahrscheinlichkeit für eine Fehlentscheidung bei diesem Test, falls es in Wahrheit bei

   a) 90 %

   b) 60 %

   der Bäume zu einem Düngeerfolg kommt !

8) Bestimmen Sie die Wahrscheinlichkeit für eine richtige Entscheidung bei diesem Test, falls es in Wahrheit bei

   a) 30 %

   b) 40 %

   c) 50 %

   der Bäume zu einem Düngeerfolg kommt !

9) Skizzieren Sie die Gütefunktion dieses Test (unter Verwendung Ihrer zuvor ermittelten Ergebnisse) !

Im Langzeitversuch hat sich gezeigt, daß 18 der mit dem Düngemittel behandelten Bäume einen Düngeerfolg aufweisen.

10) Wie lautet die Testentscheidung ?

11) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !
Wird B. den Liefervertrag abschließen ?

B

Der Baumschulbesitzer B. hat eine Riesenmenge von Rosensetzlingen in seiner Baumschule. Der Angestellte A. der Baumschule hat die (weißen, gelben und roten) Rosensetzlinge leider nicht beschriftet. B. ist darüber sehr verärgert, da er seinen Kunden zum bevorstehenden Valentinstag rote Rosen zum Verkauf anbieten möchte. Um den Anteil der roten Rosen zu schätzen, bringt er daher 60 Setzlinge (einfache Stichprobe) ins Treibhaus und wartet, bis diese erblühen.

1) Geben Sie das (approximative) Konfidenzintervall für den unbekannten Anteil π (der roten Rosen) zum Konfidenzniveau $1 - \alpha$ explizit an !

Bei den Setzlingen im Treibhaus stellt sich heraus, daß die Zahl der roten Rosen genauso groß ist wie die Zahl der weißen und gelben Rosen zusammengenommen.

2) Bestimmen Sie (approximativ) das Schätzintervall für π zum Konfidenzniveau von 99% !

3) B. benötigt für den Valentinstag in seinem Angebotssortiment 60 % rote Rosen. Kann er aufgrund seines unter 2) erhaltenen Ergebnisses (mit 99 %–iger Wahrscheinlichkeit) sicher sein, daß er derart viele rote Rosen anbieten kann ? (Kurze Begründung !)

4) Wie viele Rosensetzlinge hätte B. ins Treibhaus bringen müssen, damit das Schätzintervall eine Länge von 0,258 nicht überschreitet ?

C

Tragen Sie in Ihrem Lösungsbogen den (die) fehlenden Begriff(e) ein :

1) Eine Auswahl, die nicht zufällig und nicht bewußt erfolgt, nennt man . . . . . . !

2) Eine Auswahl heißt Zufallsauswahl, wenn jedes Element der Grundgesamtheit . . . . . . Chance hat, in die Auswahl zu gelangen !

3) Wenn nur Elemente der Grundgesamtheit in die Auswahl gelangen sollen, die von wesentlicher Bedeutung sind, nennt man ein solches Auswahlverfahren . . . . . . !

Aufgabe N 15 [Statistik II]

A

Smutje will Angelurlaub an der Ostsee machen und dort jeden Tag mit einem Fischkutter zum Angeln in See stechen. Aus Erfahrung weiß man, daß bei der täglichen Angeltour im Mittel 20 Kilogramm Fisch gefangen werden bei einer Standardabweichung von 2 Kilogramm.

1) Wie groß ist die Wahrscheinlichkeit, daß Smutje bei einer Aufenthaltsdauer von 16 Tagen im Mittel zwischen 19 und 21 Kilogramm Fisch pro Tag fangen wird ?

2) Wie groß ist die Wahrscheinlichkeit, daß Smutje bei einer Aufenthaltsdauer von 36 Tagen im Mittel zwischen 19 und 21 Kilogramm Fisch pro Tag fangen wird ?

3) In welchen symmetrischen Grenzen um den Erwartungswert wird mit 90 %–iger Wahrscheinlichkeit die von Smutje im Mittel pro Tag gefangene Menge Fisch (in Kilogramm) bei einer Aufenthaltsdauer von 16 Tagen liegen ?

4) In welchen symmetrischen Grenzen um den Erwartungswert wird mit 90 %–iger Wahrscheinlichkeit die von Smutje im Mittel pro Tag gefangene Menge Fisch (in Kilogramm) bei einer Aufenthaltsdauer von 36 Tagen liegen ?

B

Jeweils am Abend vorher hören die Anglerkollegen im Radio den Wetterbericht mit der Vorhersage für den kommenden Tag. Da Smutje schon oft an der Ostsee war, glaubt er, am Sonnenuntergang erkennen zu können, ob das Wetter am nächsten Tag gut oder schlecht wird und verzichtet daher auf den Wetterbericht.
Mit Hilfe eines statistischen Tests ($\alpha = 5\ \%$) soll nun überprüft werden, ob die tägliche Vorhersage („ gut “ oder „ schlecht “) im Radio mit der täglichen Vorhersage („ gut “ oder „ schlecht “) von Smutje identisch ist.

1) Wie heißt das zu verwendende Testverfahren ?

2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

3) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?

4) Bestimmen Sie den Annahme– und Verwerfungsbereich des Tests !

An 22 von insgesamt 36 Tagen haben sowohl der Wetterbericht als auch Smutje „ gutes “ Wetter vorausgesehen. An 2 Tagen wurde im Radio „ schlechtes “ Wetter und von Smutje „ gutes “ Wetter angesagt, während an 4 Tagen beide Parteien „ schlechtes “ Wetter vorhergesehen haben.

5) Wie lautet die Testentscheidung ?

6) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

C

Kneipenbesitzer Hans hat im Anglerdorf in seinem Lokal einen Spielautomaten aufgestellt. Wenn man dort eine 1 €–Münze einwirft, erhält man (mit einer bestimmten Wahrscheinlichkeit) eine bestimmte Anzahl von 1 €–Münzen ausgezahlt. Die Automat ist so programmiert, daß die einzelnen Spieldurchgänge unabhängig voneinander sind und die möglichen Auszahlungsbeträge (0 € , 1 € und ϑ €) folgende Wahrscheinlichkeitsverteilung besitzen :

x	0	1	ϑ
P (X = x)	0,1	0,8	0,1

Im Automaten gibt es eine Anzeige, an der man die insgesamt ausgezahlte Eurosumme sowie die Anzahl der Spieldurchgänge ablesen kann. Diese Anzeige wird von Hans vor Geschäftsöffnung immer auf Null zurückgesetzt.
Hans kennt leider den Auszahlungsbetrag ϑ nicht und möchte diesen schätzen.

Aufgabe N 16 [Statistik II]

A

Ein Mediziner will einen Artikel für eine Fachzeitschrift schreiben. Sein derzeitiges Forschungsgebiet beschäftigt sich damit, ob es einen Zusammenhang zwischen dem Körpergewicht und dem Cholesterinwert gibt. Da er von einem derartigen Zusammenhang überzeugt ist, möchte er seine These mit Hilfe eines statistischen Tests ($\alpha = 0{,}01$) untermauern. Zu diesem Zweck hat er 500 zufällig ausgewählte Personen befragt. Leider jedoch ist seine Datenauswertung in einem unbeobachteten Augenblick in die Pfoten seines cholesterinkranken und übergewichtigen Schäferhundes geraten, so daß nur noch einige Daten übriggeblieben sind :

Gewicht \ Cholesterin	normal	hoch	sehr hoch	
normal	115	**57,8**	10	
bis 20 kg über normal	**44,8**	85	**60,8**	
mehr als 20 kg über normal	5			
			190	

$\tilde{h}_{ij}$
h_{ij}

1) Ermitteln Sie die verlorengegangenen Daten und vervollständigen Sie die Kontingenztafel !
2) Welches Testverfahren sollte hier verwendet werden ?
3) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !
4) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?
5) Bestimmen Sie den Annahme– und Verwerfungsbereich für diesen Test !
6) Wie lautet die Testentscheidung ?
7) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

B

Der Artikel des Mediziners fand großen Anklang. Daher möchte er als nächstes die Ursachen von zu hohem Körpergewicht erforschen. Er glaubt, daß der „ übermäßige Konsum von Süßigkeiten " und die „ mangelnde sportliche Aktivität " zwei wichtige Ursachen sind. Auch hier möchte er seine These, daß diese beiden Einflüsse gleichwirksam sind, mit Hilfe eines statistischen Tests ($\alpha = 0{,}01$) untermauern. Zu diesem Zweck wählt er aus der Gruppe der „ Übergewichtigen " (d.h. derjenigen Personen, die über dem Normalgewicht liegen) der ersten Untersuchung zufällig 10 % aus und befragt diese Personen hinsichtlich der beiden genannten Kriterien erneut :

- bei 15 Befragten hatten sowohl übermäßiger Süßigkeitenkonsum als auch mangelnde sportliche Aktivität einen Einfluß auf ihr Übergewicht.
- bei 7 Befragten hatten weder übermäßiger Süßigkeitenkonsum noch mangelnde sportliche Aktivität einen Einfluß auf ihr Übergewicht.
- bei 12 Befragten hatte mangelnde sportliche Aktivität keinen Einfluß auf ihr Übergewicht.

1) Formulieren Sie die Hypothesen für einen **McNemar–Test** ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !
2) Wie lautet die Testfunktion und wie ist diese unter H_0 verteilt ?
3) Bestimmen Sie den Annahme– und Verwerfungsbereich für diesen Test !
4) Wie lautet die Testentscheidung ?
5) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

C

Jeden Tag sind die Schüler eines Internats über die Verspätung des Frühstücks verärgert, weil die zuständige Bäckerei die frischen Brötchen zu spät liefert. Die Internatsschüler wollen nun ihrem Ärger durch einen angekündigten Schulstreik Luft machen und die Internatsleitung mit statistischen Fakten konfrontieren. Zur Schätzung der wahren mittleren Verspätung des Frühstücks werden die letzten 35 Tage (einfache Stichprobe) ausgewählt, und die Differenz zwischen tatsächlicher und planmäßiger Frühstückszeit [in Minuten] ermittelt. Nur wenn die wahre mittlere Verspätung 15 Minuten (oder mehr) betragen sollte, will die Internatsleitung – im Sinne der Schüler – den Brötchenlieferanten wechseln. Ansonsten werden den Schülern harte Sanktionen in Form von verschärftem Statistikunterricht angedroht.

1) Geben Sie explizit das Konfidenzintervall für die wahre mittlere Verspätung μ zum Konfidenzniveau 95 % an !

Die Stichprobendaten ergaben ein arithmetisches Mittel von 18 Minuten bei einer Varianz von 65 Minuten^2 .

2) Berechnen Sie das Schätzintervall für μ !

3) Wer hat Recht ? Müssen die Schüler in Zukunft nachsitzen und Statistik „ büffeln " ?

4) Interpretieren Sie dieses Schätzergebnis inhaltlich und statistisch exakt !

D

Die Bäckerei, die dem Internat die Brötchen liefert hat 32 Angestellten und benötigt 2 Säcke Mehl à 50 kg pro Tag für die zu backenden Brötchen.
Sei

X : „ Anzahl der täglich (für die Internatsschüler) gebackenen Brötchen "

mit $E(X) = \mu$ und $Var(X) = \sigma^2 = 25$

Wie groß ist die Wahrscheinlichkeit, daß die gebackene Brötchenzahl pro Tag um höchstens 12 Brötchen von der zu erwartenden Anzahl der täglich gebackenen Brötchen abweicht ?

E

Sei T eine exponentialverteilte Zufallsvariable, d.h. es gilt : $T \sim E(\lambda)$.

Zeigen Sie mit Hilfe eines formalen Beweisganges, daß gilt :

$$P\left(T \le s+t / T > s\right) \overset{!}{=} P\left(T \le t\right)$$

V

Verschiedenes

ABT. MOGELKUNDE

Ein altes Sprichwort behauptet: „Zahlen lügen nicht". – Stimmt. Aber umso mehr lügen die Leute, die sie benutzen! Denn was man so alles mit Zahlen anfangen kann, zeigt . . .

DER GROSSE MAD-LEHRGANG ÜBER DIE KUNST DER STATISTIK

ZEICHNUNGEN: PAUL COKER   TEXT: DICK DE BARTOLO

DAS ERGEBNIS EINER UMFRAGE HÄNGT DAVON AB, WO DIESE GEMACHT WURDE

Hier zum Beispiel eine eindrucksvolle Anzeige: Und nun ein Blick, woher die Umfrage stammt:

DIE FORMULIERUNG EINER FRAGE BEEINFLUSST DIE ANTWORT

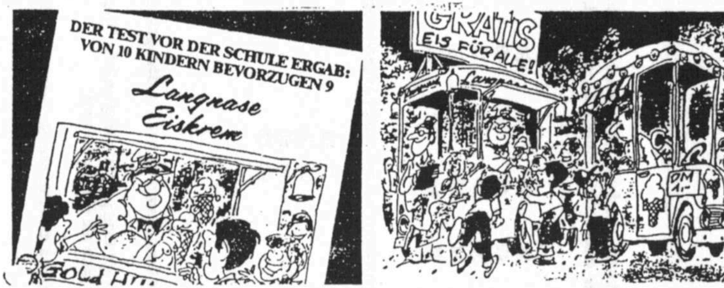
Nehmen wir als Beispiel die nächsten beiden Umfragen. Ganz sicher würden Sie einmal „JA" und einmal „NEIN" antworten – obwohl sie genau denselben Inhalt haben!

UND NATÜRLICH IST ES AUCH ENTSCHEIDEND, WEN MAN FRAGT

Nehmen wir an, Sie wollen bei der Umfrage „Planen Sie den Kauf eines Videorecorders" möglichst viele JA-Antworten erzielen. Welche der beiden Gruppen würden Sie* befragen?

*Sehen Sie, wie rasch Sie lernen!

ES GIBT IMMER EINEN TRICK, UM DER STATISTIK NACHZUHELFEN



SELBST HOFFNUNGSLOSE FÄLLE SIND STATISTISCH VERWERTBAR

AUCH FÜR EINE STATISTIK GILT: KLEINE GESCHENKE WIRKEN WUNDER

V2

Um den Schwierigkeitsgrad einer Abituraufgabe durch die Bearbeitungszeit zu untersuchen, wird einer Abiturklasse (50 Schüler) eine ähnliche Aufgabe zur Bearbeitung vorgelegt, Die Schüler benötigten folgende Bearbeitungszeit (in min.):

19	27	72	68	35	20	30	43	42	33
28	37	34	47	49	30	40	19	38	43
25	49	25	32	29	23	32	28	27	18
21	26	31	32	25	39	34	38	41	36
51	31	36	48	46	49	45	52	61	57

Dieses Datenmaterial soll gruppiert ausgewertet werden, wobei folgende vier Gruppen gebildet werden sollen:

0 - 20
20 - 30
30 - 60
60 - 80

Einfluß des Startpunktes und der Gruppenbreite auf das Histogramm

Um den Einfluß des Startpunktes x_0 zur Gruppeneinteilung und der Gruppenbreite selbst auf das Histogramm zu verdeutlichen, hier noch ein kleines Beispiel:

Die nachfolgenden Histogramme stellen die Häufigkeitsverteilung der Bearbeitungszeit aus obigen Daten dar.

Abbildung 1: Histogramme für die Abituraufgabe bei unterschiedlichen Startpunkten x_0= 0, 4, 8, 12, 16 mit konstanter Gruppenbreite b_j=10.

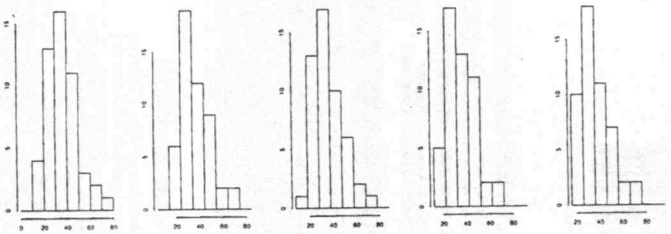
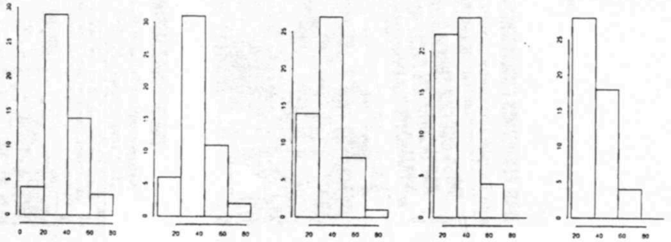


Abbildung 2: Histogramme für die Abituraufgabe bei unterschiedlichen Startpunkten x_0= 0, 4, 8, 12, 16 mit konstanter Gruppenbreite b_j=20.



Einfluß des Startpunktes und der Gruppenbreite auf das Histogramm

Um den Einfluß des Startpunktes x_0 zur Gruppeneinteilung und der Gruppenbreite selbst auf das Histogramm zu verdeutlichen, hier noch ein kleines Beispiel:

Die nachfolgenden Histogramme stellen die Häufigkeitsverteilung der Bearbeitungszeit einer Arbituraufgabe dar ,an der 50 Schüler teilnahmen.

Die Schüler benötigten folgende Bearbeitungszeit (in min.):

19	27	72	68	35	20	30	43	42	33
28	37	34	47	49	30	40	19	38	43
25	49	25	32	29	23	32	28	27	18
21	26	31	32	25	39	34	38	41	36
51	31	36	48	46	49	45	52	61	57

Abbildung 1: Histogramme für die Abiturdaten bei unterschiedlichen Startpunkten x_0= 0, 4, 8, 12, 16 mit konstanter Gruppenbreite b_j=10.

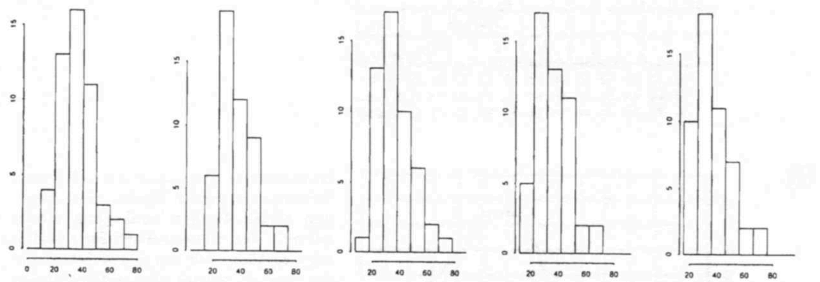
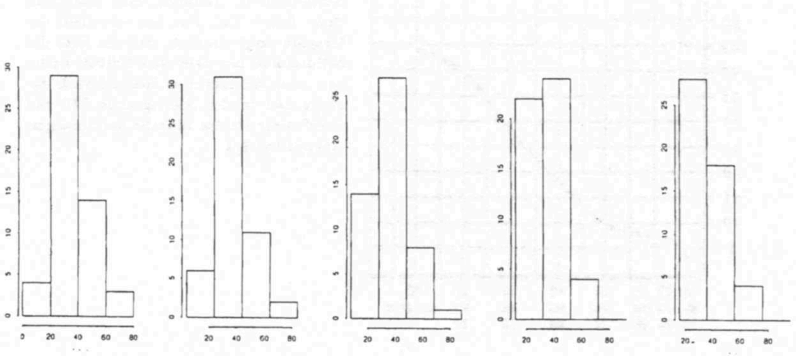


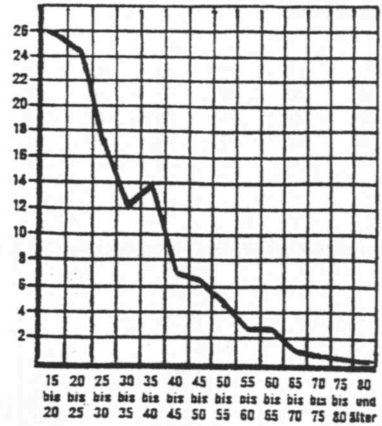
Abbildung 2: Histogramme für die Abiturdaten bei unterschiedlichen Startpunkten x_0= 0, 4, 8, 12, 16 mit konstanter Gruppenbreite b_j=20.



Darstellungen der Selbstmordneigung in Abhängigkeit vom Alter

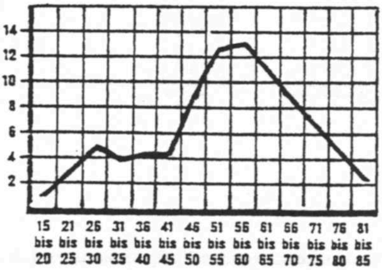
Die folgenden Graphiken entstammen ein und demselben Datensatz:

A)



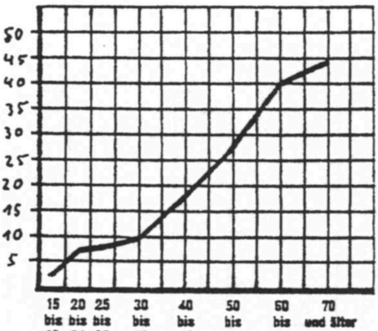
Selbstmord ist eine Krankheit der Jugend und mimmt mit wachsendem Alter an Bedeutung immer mehr ab. Der Kurvenverlauf der Graphik zeigt deutlich, daß der Anteil der Selbstmörder an den gesamten Todesfällen einer Altersklasse von 25 Prozent unter den jungen Menschen (15-25 Jahre) auf weniger als ein Prozent absinkt, sobald man sich dem 70. Lebensjahr nähert.

B)



Selbstmorde erreichen im mittleren Lebensalter einen Gipfel, sind dagegen unter Jungen und ganz Alten selten. Der Kurvenverlauf der Graphik zeigt deutlich, daß von je 100 Selbstmorden rund ein Viertel allein auf das Altersjahrzehnt vom 51. zum 60. Lebensjahr entfällt, auf die Siebzigjährigen (71-80) dagegen nur etwa 10 Prozent, auf die Gruppe 21-30 Jahre nur 8 Prozent.

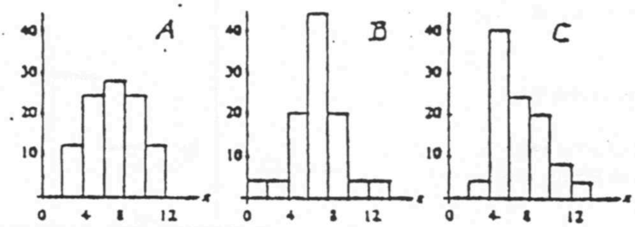
C)



Selbstmorde nehmen mit wachsendem Alter zu. Der Kurvenverlauf der Graphik zeigt deutlich, daß die Zahl der Selbstmorde pro Jahr und 100 000 Personen einer Altersgruppe kontinuierlich ansteigt und in der Altersgruppe "70 und älter" mehr als das 10-fache der Teenager-Gruppe beträgt

Es werden drei Verteilungen eines kardinal skalierten Merkmals betrachtet :

von....bis unter	A	B	C
0-2	0	4	0
2-4	12	4	4
4-6	24	20	40
6-8	28	44	24
8-10	24	20	20
10-12	12	4	8
12-14	0	4	4
	100	100	100



$\bar{x}_A = \bar{x}_B = \bar{x}_C = 7$ $\qquad$ $s^2_A = s^2_B = s^2_C = 5.76$

Lage und Streuung der drei Verteilungen sind - gemessen an den Maßzahlen arithmetisches Mittel und empirische Varianz - gleich, jedoch ist ihre **Form** jeweils verschieden:

- die Verteilungen A und B sind symmetrisch, die Verteilung C ist asymmetrisch.
- Verteilung B verläuft in der Mitte stärker zugespitzt bzw. steiler als Verteilung A.

Es liegt also nahe sogenannte **Formmaßzahlen** der Schiefe und der Steilheit(Wölbung,Kurtosis,Exzeß) zu suchen.

	A	B	C
Schiefe g_1	0	0	0.694
Steilheit g_2	-.0917	+0.935	-0.222

$g_1 = 0$ ➡ Verteilung symmetrisch

$g_1 < 0$ ➡ Verteilung rechtssteil (linksschief)

$g_1 > 0$ ➡ Verteilung linkssteil (rechtsschief)

$g_2 = 0$ ➡ Exzeß genau wie bei der N.V.

$g_2 < 0$ ➡ Exzeß kleiner als bei der N.V.

$g_2 > 0$ ➡ Exzeß größer als bei der N.V.

V5a

Quantile

für sortierte Daten

$x_{(1)}, x_{(2)}, \dots, x_{(n)}$

Die empirische Verteilungsfunktion ist definiert als : $F(x) = \sum_{a_j \le x} f(a_j)$

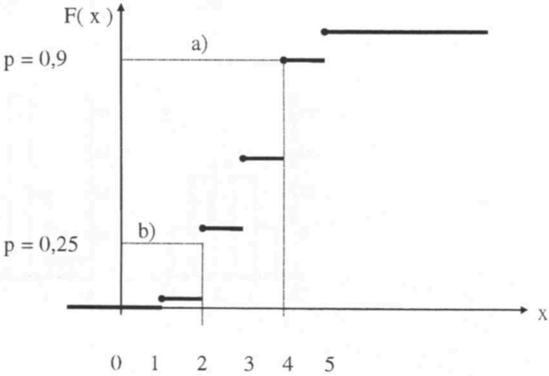
Man kann einen Anteil p vorgeben und fragen, bei welchem x die kumulierte Häufigkeit diesen Anteil erreicht. Da bei ungruppierten Daten die Verteilungsfunktion Sprünge aufweist, kann es vorkommen, daß der vorgegebene Anteil p gerade **über**troffen wird.

1) Es gibt mehr als einen Wert

mit F(x) = p ➡ a)

2) Es gibt keinen Wert

mit F(x) = p ➡ b)
[p wird übertroffen]



Fall a) : Man nimmt den kleinsten der infrage kommenden Werte an, also **x = 4**
Fall b) : Man nimmt den Wert, bei dem der Anteil p gerade **über**schritten wird, also **x = 2**

Für jeden Anteil p mit $0 < p < 1$ ist das empirische p - Quantil x_p des Datensatzes $x_1, x_2, \dots, x_n$ der **kleinste x - Wert**, für den gilt $\mathbf{F(x) \geq p}$

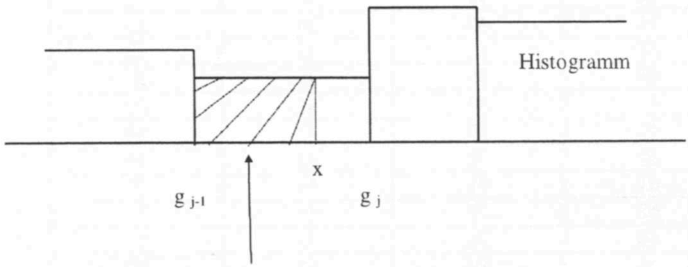
V5b

für gruppierte Daten

Da hier die Verteilungsfunktion keine Sprünge aufweist, gibt es zu jedem p einen Wert x_p mit $F(x) = p$!

1) **graphisch** aus der Verteilungsfunktion
2) **rechnerisch** durch Auflösen der Definitionsgleichung von F(x) nach x

Exkurs :



mit der „ Häufigkeitsdichte " $\tilde{f}(x)$

$$(x - g_{j-1}) \bullet \tilde{f}(x) = (x - g_{j-1}) \bullet \frac{f_j}{b_j} \quad ; \quad \tilde{f}(x) \cong \text{„ Blockhöhe "} \cong \ell_j$$

F(x) bei gruppierten Daten mit den Gruppengrenzen $g_0, g_1, \ldots, g_k$ für $j = 1, \ldots, k$ Gruppen ist definiert als

$$F(x) = \begin{cases} 0 & \text{für} \quad x < g_0 \\ F(g_{j-1}) + (x - g_{j-1}) \bullet \dfrac{f_j}{b_j} & \text{für} \quad g_{j-1} \le x < g_j \\ 1 & \text{für} \quad x \ge g_k \end{cases}$$

Das p - Quantil x_p ($0 < p \le 1$), falls $F(g_{j-1}) < p \le F(g_j)$, ergibt sich zu

$$F(x) = p = F(g_{j-1}) + (x_p - g_{j-1}) \bullet \frac{f_j}{b_j}$$

$$\Rightarrow \quad x_p \bullet \frac{f_j}{b_j} = g_{j-1} \bullet \frac{f_j}{b_j} + p - F(g_{j-1})$$

$$\Rightarrow \quad \boxed{x_p \approx g_{j-1} + [\, p - F(g_{j-1})\,] \bullet \frac{b_j}{f_j}}$$

QQ-Plot

Eine Klausur, an der 4 Schulklassen mit jeweils 20 Schülern teilgenommen haben, führte zu folgenden Ergebnis:

Klasse A	Klasse B	Klasse C	Klasse D
7	17	0	7
10	20	3	9
21	31	14	15
35	45	28	20
42	52	35	22
46	56	39	27
47	57	40	30
49	59	42	50
52	62	45	55
61	71	54	70
62	72	55	72
66	76	59	75
70	80	63	78
70	80	63	80
73	83	66	80
75	85	68	85
79	89	72	87
80	90	73	88
82	92	75	89
86	96	79	90

Von diesem Klausurergebnis sind in den Abbildungen 1-4 QQ-Plots dargestellt

Abbildung 1: Die Quantile der Häufigkeitsverteilung A wurden gegeneinander geplottet.

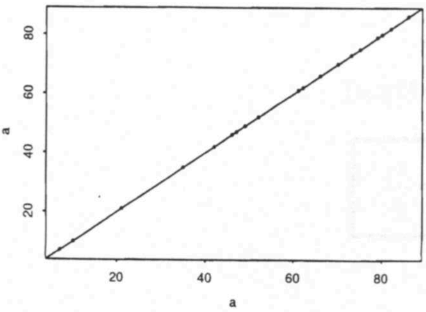


Abbildung 2: Die Quantile der Häufigkeitsverteilung A wurden gegen die Quantile der Häufigkeitsverteilung B geplottet.

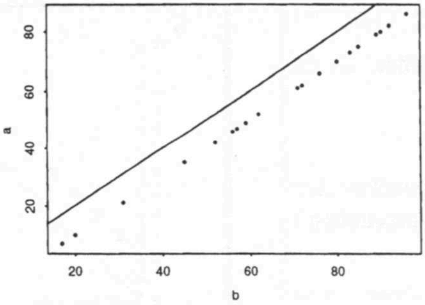


Abbildung 3: Die Quantile der Häufigkeitsverteilung A wurden gegen die Quantile der Häufigkeitsverteilung C geplottet.

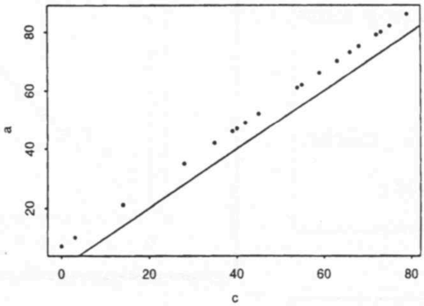
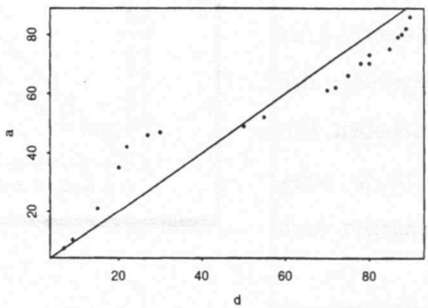


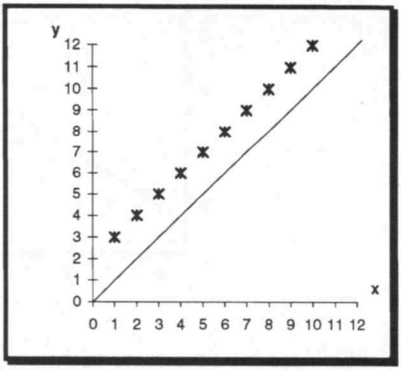
Abbildung 4: Die Quantile der Häufigkeitsverteilung A wurden gegen die Quantile der Häufigkeitsverteilung D geplottet.



QQ – Diagramme

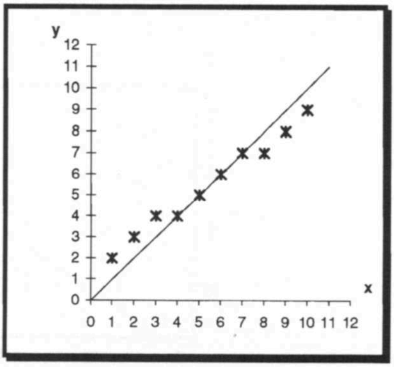
Die Quantile y_p sind **größer** als die Quantile x_p.

Der y – Datensatz ist gegenüber dem x – Datensatz nach **rechts verschoben** !



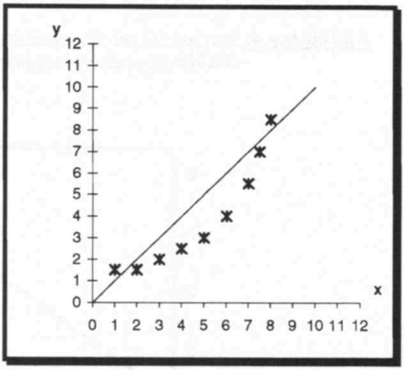
Für kleine p sind die Quantile y_p größer als die x_p. Für größere p kehrt sich die Situation um.

Die y – Daten sind **stärker konzentriert** als die x – Daten !



Für kleine <u>und</u> große p sind die y – Daten **größer** als die x – Daten.

Der zentrale Datenkörper (≅ 90 %) des y – Datenbereichs ist gegenüber den x – Daten nach **links verschoben**. Der äußere Datenkörper (≅ 10 %) des y – Datenbereichs ist dagegen nach **rechts verschoben** !

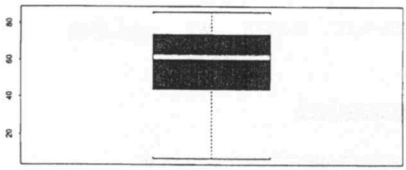


Box-Plot

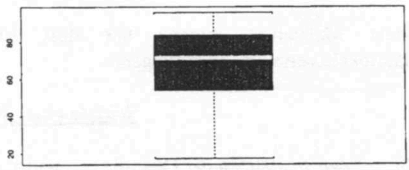
Eine weitere Klausur, an der wiederum 4 Schulklassen mit jeweils 20 Schülern teilgenommen haben, wo maximal 100 Punkte erreichbar waren, führte zu folgenden Ergebnis:

Klasse A	Klasse B	Klasse C	Klasse D
7	17	0	12
10	20	1	12
21	31	9	13
35	45	14	21
42	52	22	27
46	56	26	29
47	57	27	30
49	59	28	53
52	62	34	54
61	71	34	55
62	72	45	55
66	76	46	62
70	80	55	65
70	80	55	68
73	83	74	70
75	85	78	74
79	89	86	76
80	90	93	99
82	92	99	100
86	96	99	100

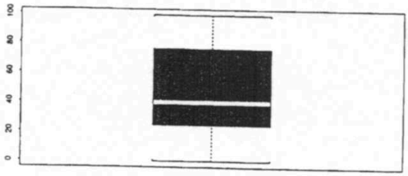
Box-Plot für Klasse A



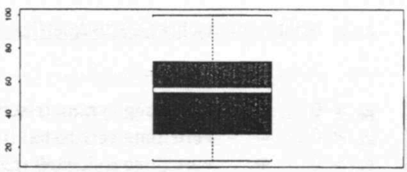
Box-Plot für Klasse B



Box-Plot für Klasse C



Box-Plot für Klasse D



Momente einer Häufigkeitsverteilung

Das r-te Moment bzgl. einer Zahl c eines Datensatzes ($x_1, x_2, \dots, x_n$) ist definiert durch

$$m_r(c) = \frac{1}{n} \sum_{i=1}^{n} (x_i - c)^r \quad ; \quad r = 1, 2, \dots$$

Für c = 0 erhält man die sog. Anfangsmomente

$$m_r(0) = \frac{1}{n} \sum_{i=1}^{n} x_i^r \qquad mit \quad m_1(0) = \overline{x}$$

Momente bzgl. $\overline{x}$ heißen zentrale Momente.
Die zentralen Momente bezeichnet man mit

$$m_r = m_r(\overline{x}) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^r$$

Dabei gilt : $m_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 = s^2$

Bei symmetrischen Verteilungen verschwinden alle zentralen Momente ungerader Ordnung (mit ungeradem r).

Schiefe und Exzeß (Formmaßzahlen)

Beide Maßzahlen können nur dort sinnvoll verwandt werden, wo eingipflige Häufigkeitsverteilungen vorliegen.

Schiefe einer Häufigkeitsverteilung

Ein auf den Abständen der Quantile x_p und x_{1-p} zu $\tilde{x}$ aufgebautes Schiefemaß ist der

Empirische p – Quantilskoeffizient der Schiefe :

$$g_p = \frac{(x_{1-p} - \tilde{x}) - (\tilde{x} - x_p)}{x_{1-p} - x_p} \quad (0 < p < 0{,}5)$$

$$\boxed{\begin{array}{c} -1 \le g_p \le +1 \\ g_p = -1 \Rightarrow x_{1-p} = \tilde{x} \\ g_p = +1 \Rightarrow x_p = \tilde{x} \end{array}}$$

$g_{0,25}$ heißt „empirischer Quartilskoeffizient der Schiefe"

$g_p = 0$ ➡ Verteilung **symmetrisch**
$g_p < 0$ ➡ Verteilung **rechtssteil** (linksschief)
$g_p > 0$ ➡ Verteilung **linkssteil** (rechtsschief)

Bei nichtsymmetrischen Verteilungen könnte man das **dritte zentrale Moment m_3** als Maß für die Abweichung von der Symmetrie, also als Maß für die „ *Schiefe* " benutzen.
Diese Größe ist jedoch maßstabsabhängig. Division durch s^3 ergibt die vom Maßstab unabhängige **Schiefe** :

$$g_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^3}{s^3} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{s}\right)^3$$

$$g_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^3}{\sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\right]}^{\,3}} \quad \Leftrightarrow \quad \text{Urliste}$$

$$g_1 = \frac{\frac{1}{n}\sum_{j=1}^{k}\left(a_j - \overline{x}\right)^3 * h\left(a_j\right)}{\sqrt{\left[\frac{1}{n}\sum_{j=1}^{k}\left(a_j - \overline{x}\right)^2 * h\left(a_j\right)\right]}^{\,3}} \quad \Leftrightarrow \quad \text{sortierte Daten}$$

$$g_1 = \frac{\frac{1}{n}\sum_{j=1}^{k}\left(m_j - \overline{x}\right)^3 * h_j}{\sqrt{\left[\frac{1}{n}\sum_{j=1}^{k}\left(m_j - \overline{x}\right)^2 * h_j\right]}^{\,3}} \quad \Leftrightarrow \quad \text{gruppierte Daten}$$

$g_1 = 0$ ➡ Verteilung **symmetrisch**
$g_1 < 0$ ➡ Verteilung **rechtssteil** (linksschief)
$g_1 > 0$ ➡ Verteilung **linkssteil** (rechtsschief)

Da g_p und g_1 auf unterschiedlichen Konzepten beruhen, sind sie nicht miteinander vergleichbar !

V12

Exzeß (Wölbung, Steilheit) einer Häufigkeitsverteilung

Der Exzeß g_2 gibt an, ob (bei gleicher Varianz) das absolute Maximum der eingipfligen Häufigkeitsverteilung größer ist als bei der Dichte der Normalverteilung.
Der (theor.) Exzeß einer Normalverteilung ist gerade Null !

$$g_2 = \underbrace{\frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^4}{s^4}}_{\text{Kurtosis}} - 3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{s}\right)^4 - 3$$

Normalverteilung hat eine Kurtosis von 3 !

$$g_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^4}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\right]^2} - 3 \quad \Leftrightarrow \quad \text{Urliste}$$

$$g_2 = \frac{\frac{1}{n}\sum_{j=1}^{k}\left(a_j - \overline{x}\right)^4 * h(a_j)}{\left[\frac{1}{n}\sum_{j=1}^{k}\left(a_j - \overline{x}\right)^2 * h(a_j)\right]^2} - 3 \quad \Leftrightarrow \quad \text{sortierte Daten}$$

$$g_2 = \frac{\frac{1}{n}\sum_{j=1}^{k}\left(m_j - \overline{x}\right)^4 * h_j}{\left[\frac{1}{n}\sum_{j=1}^{k}\left(m_j - \overline{x}\right)^2 * h_j\right]^2} - 3 \quad \Leftrightarrow \quad \text{gruppierte Daten}$$

$g_2 = 0$ ➡ Exzeß genau wie bei der N.V.
$g_2 < 0$ ➡ Exzeß **kleiner** als bei der N.V. d.h. „ *platykurtisch* " [platys $\cong$ **flach**]
$g_2 > 0$ ➡ Exzeß **größer** als bei der N.V. d.h. „ *leptokurtisch* " [leptos $\cong$ **schmal**]

Fazit :

Sind Schiefe und Exzeß einer Häufigkeitsverteilung wesentlich von Null verschieden, so ist dies ein Hinweis darauf, daß die Verteilung des Untersuchungsmerkmals in der Ggh. wesentlich von einer Normalverteilung abweicht.

Ein neues Medikament zur Behandlung von Bluthochdruck soll erprobt werden.

X: 'Erfolg der Behandlung' a_1 = Besserung, a_2 = keine Besserung
Y: 'Behandlungsart' b_1 = Standardbehandlung, b_2 = neues Medikament

	b_1	b_2	
a_1	10	100	110
a_2	100	500	600
	110	600	710

Wie würden Sie das Medikament aufgrund dieser Daten beurteilen ? (Warum ?)

An einer zweiten Klinik wird der gleiche Versuch mit anderen Patienten durchgeführt:

	b_1	b_2	
a_1	100	60	160
a_2	50	20	70
	150	80	230

Wie würden Sie das Medikament nach den Daten dieses Versuches beurteilen ?

Zur Gesamtdarstellung werden die Ergebnisse beider Versuche zusammengefaßt:

	b_1	b_2	
a_1	110	160	270
a_2	150	520	670
	260	680	940

Wie würden Sie das Medikament aufgrund dieser Dartstellung beurteilen ?

Herfindahl

Im Land A bzw. Land B beherrscht bezgl. des Marktes der (s) .... folgende Konzentrationsverteilung :

Land A ($n_A = 4$)

i	p_i
1	0.06
2	0.09
3	0.21
4	0.64
	100 %

Land B ($n_B = 3$)

i	p_i
1	0.12
2	0.20
3	0.68
	100 %

$H_A = 0.06^2 + ... + 0.64^2 = \underline{0.4654}$ $\quad$ $H_B = 0.12^2 + ... + 0.68^2 = \underline{0.5168}$

$$\boxed{H_A < H_B}$$

Eleminierung des Anzahleffekts E_A

	i	p_i
G_1	1	0.02
	2	0.02
	3	0.02
G_2	4	0.03
	5	0.03
	6	0.03
G_3	7	0.07
	8	0.07
	9	0.07
G4	10	0.213
	11	0.213
	12	0.213
		100 %

	i	p_i
G_1	1	0.03
	2	0.03
	3	0.03
	4	0.03
G_2	5	0.05
	6	0.05
	7	0.05
	8	0.05
G_3	9	0.17
	10	0.17
	11	0.17
	12	0.17
		100 %

$$H'_A = 3 * 0.02^2 + ... + 3 * 0.213^2 = 3\left(\frac{0.06}{3}\right)^2 + ... + 3\left(\frac{0.64}{3}\right)^2$$

$$= \frac{1}{3}(0.06^2 + ... + 0.64^2 = \frac{1}{3} * H_A = \underline{0.1551}$$

$$H'_B = 4 * 0.03^2 + ... + 4 * 0.17^2 = 4\left(\frac{0.12}{4}\right)^2 + ... + 4 * \frac{0.68}{4})$$

$$= \frac{1}{4}(0.12^2 + ... + 0.68^2) = \frac{1}{4} * H_B = \underline{0.1292}$$

$$H'_A = \frac{1}{n_B} * H_A \qquad H'_B = \frac{1}{n_A} * H_B$$

$$\boxed{H'_A > H'_B}$$

Interpretation von : $\boxed{H_A < H_B}$

[1] $n_A = n_B$

➪ ausschließlich am Merkmalseffekt

[2] $n_A > n_B$

- $H_A' > H_B'$

  ➪ Anzahleffekt wirkt stärker als Merkmalseffekt

- $H_A' < H_B'$

  ➪ Anzahl- und Merkmalseffekt wirken in gleicher Richtung (Land B)

- $H_A' = H_B'$

  ➪ ausschließlich am Anzahleffekt

[3] $n_A < n_B$

- $H_A' > H_B'$

  ➪ kann nicht vorkommen !!!

- $H_A' < H_B'$

  ➪ ausschließlich am Merkmalseffekt

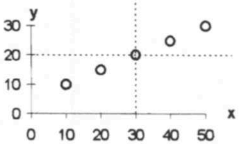
- $H_A' = H_B'$

  ➪ kann nicht vorkommen !!!

1 A

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})\cdot(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
10	10	− 20	− 10	200	400	100
20	15	− 10	− 5	50	100	25
30	20	0	0	0	0	0
40	25	10	5	50	100	25
50	30	20	10	200	400	100
150	100			500	1000	250

$\bar{x} = 30$ $\bar{y} = 20$



$s_y^2 = 50$

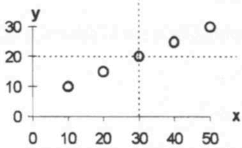
$s_x^2 = 200$

$s_{xy} = \underline{+100}$

2 A

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})\cdot(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
10	10	− 20	− 10	200	400	100
20	15	− 10	− 5	50	100	25
30	20	0	0	0	0	0
40	25	10	5	50	100	25
50	30	20	10	200	400	100
150	100			500	1000	250

$\bar{x} = 30$ $\bar{y} = 20$



$s_y^2 = 50$

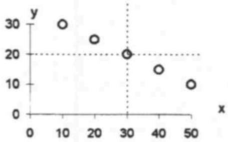
$s_x^2 = 200$

$s_{xy} = \underline{+100}$

1 B

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
10	30	− 20	10	− 200	400	100
20	25	− 10	5	− 50	100	25
30	20	0	0	0	0	0
40	15	10	− 5	− 50	100	25
50	10	20	−10	− 200	400	100
150	100			− 500	1000	250

$\bar{x} = 30$ $\bar{y} = 20$



$s_y^2 = 50$

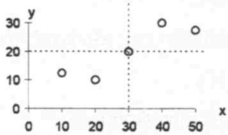
$s_x^2 = 200$

$s_{xy} = \underline{-100}$

2 B

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
10	12,5	− 20	− 7,5	150	400	56,25
20	10	− 10	− 10	100	100	100
30	20	0	0	0	0	0
40	30	10	10	100	100	100
50	27,5	20	7,5	150	400	56,25
150	100			500	1000	312,5

$\bar{x} = 30$ $\bar{y} = 20$



$s_y^2 = \underline{62{,}5}$

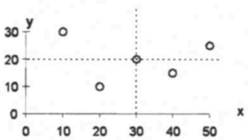
$s_x^2 = 200$

$s_{xy} = \underline{+100}$

1 C

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
10	30	− 20	10	− 200	400	100
20	10	− 10	− 10	100	100	100
30	20	0	0	0	0	0
40	15	10	− 5	− 50	100	25
50	25	20	5	100	400	25
150	100			− 50	1000	250

$\bar{x} = 30$ $\bar{y} = 20$



$s_y^2 = 50$

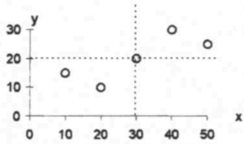
$s_x^2 = 200$

$s_{xy} = \underline{-10}$

2 C

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
10	15	− 20	− 5	100	400	25
20	10	− 10	− 10	100	100	100
30	20	0	0	0	0	0
40	30	10	10	100	100	100
50	25	20	5	100	400	25
150	100			400	1000	250

$\bar{x} = 30$ $\bar{y} = 20$

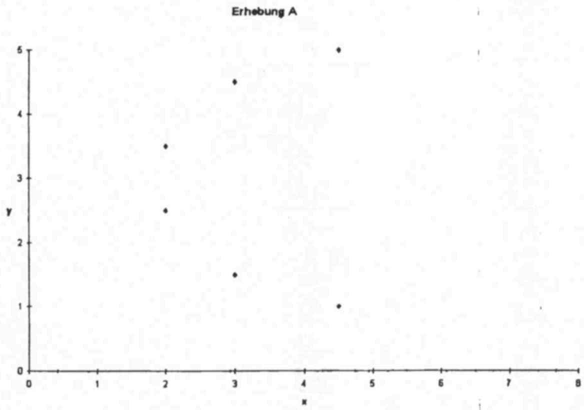


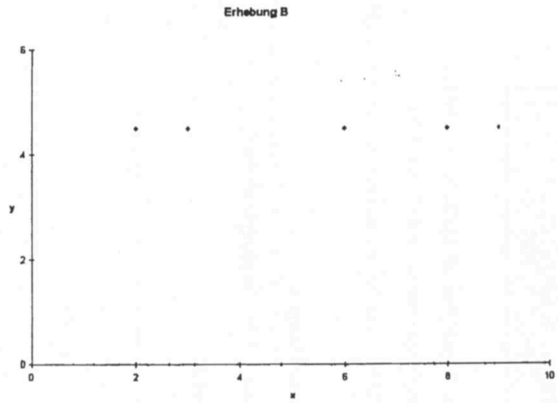
$s_y^2 = 50$

$s_x^2 = 200$

$s_{xy} = \underline{+80}$

Es liegen die Ergebnisse zweier Erhebungen A und B der Merkmale X und Y in Form von Streudiagrammen vor:





1. Welchen Wert hat die Kovarianz s_{xy} in Fall B ?

2. Welchen Wert hat der Korrelationskoeffizient $r = \dfrac{s_{xy}}{s_x * s_y}$ in Fall A und in Fall B ?

V17a

Herleitung : Bestimmtheitsmaß

Ziel : Die für einen Punkt (x_i , y_i) gemachte Aussage soll verallgemeinert werden :

$$\underbrace{\sum_{i=1}^{n}(y_i-\bar{y})^2}_{SQT}=\underbrace{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}_{SQE}+\underbrace{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}_{SQR}$$

SQT	:	Gesamtabweichungsquadratsumme in y-Richtung
SQE	:	Durch die Regression erklärter Teil von SQT
SQR	:	Trotz der Regression unerklärt bleibender Teil von SQT

Zum Beweis benutzen wir folgende Zusammenhänge :

① : $\sum x_i \cdot y_i - \hat{b} \cdot \sum x_i^2 - \hat{a} \cdot \sum x_i = 0$ [**II.** Normalgleichung]

② : $y_i - \hat{y}_i = \hat{e}_i$ [„ Residuum "]

③ : $\hat{a} = \bar{y} - \hat{b}\,\bar{x}$

Lemma 1 : $\sum y_i = \sum \hat{y}_i \Leftrightarrow \bar{y} = \bar{\hat{y}}$

Bew. : $\hat{y}_i = \hat{a} + \hat{b} \cdot x_i \overset{3}{=} \bar{y} - \hat{b} \cdot \bar{x} + \hat{b} \cdot x_i$

⇨ $\sum \hat{y}_i = \sum \bar{y} - \hat{b} \cdot \left(\sum \bar{x} - \sum x_i\right) = n \cdot \bar{y} - \underbrace{\hat{b} \cdot (n \cdot \bar{x} - n \cdot \bar{x})}_{=0} = n \cdot \bar{y} = n \cdot \frac{1}{n} \sum y_i = \underline{\sum y_i}$

w.s.m.

Lemma 2 : $\frac{1}{n} \cdot \sum \hat{e}_i = 0 \Leftrightarrow \bar{\hat{e}} = 0$

Bew. : $\frac{1}{n} \cdot \sum \hat{e}_i = \frac{1}{n} \cdot \left[\sum (y_i - \hat{y}_i)\right] \overset{L1}{=} \frac{1}{n} \cdot \left(\underbrace{\sum y_i - \sum \hat{y}_i}_{=0}\right) = \underline{0}$ w.s.m.

V17b

Lemma 3 :
$$\sum x_i \cdot \hat{e}_i = 0 \Leftrightarrow \sum x_i \cdot \hat{e}_i - \underbrace{n \cdot \overline{x} \cdot \overline{\hat{e}}}_{=0} = 0$$

$$\Leftrightarrow \quad \sum (x_i - \overline{x}) \cdot (\hat{e}_i - \overline{\hat{e}}) = 0$$

Bew. : ① $$\sum x_i \cdot y_i - \hat{b} \cdot \sum x_i^2 - \hat{a} \cdot \sum x_i = 0$$

$$\Rightarrow \quad \sum \left(x_i \cdot y_i - \hat{b} \cdot x_i^2 - \hat{a} \cdot x_i\right) = 0$$

$$\Rightarrow \quad \sum x_i \left(y_i - \hat{b} \cdot x_i - \hat{a}\right) = 0$$

$$\Rightarrow \quad \sum x_i \left[y_i - \left(\hat{a} + \hat{b} \cdot x_i\right)\right] = 0$$

$$\Rightarrow \quad \sum x_i \cdot (y_i - \hat{y}_i) = 0$$

② $$\Rightarrow \quad \sum x_i \cdot \hat{e}_i = 0$$ w.s.m.

Lemma 4 :
$$\sum \hat{y}_i \cdot \hat{e}_i = 0 \quad \Leftrightarrow \quad \sum \hat{y}_i \cdot \hat{e}_i - \underbrace{n \cdot \overline{\hat{y}} \cdot \overline{\hat{e}}}_{=0} = 0$$

$$\Leftrightarrow \quad \sum (\hat{y}_i - \overline{\hat{y}}) \cdot (\hat{e}_i - \overline{\hat{e}}) = 0$$

Bew. : $$\hat{y}_i = \hat{a} + \hat{b} \cdot x_i$$

$$\Rightarrow \quad \hat{y}_i \cdot \hat{e}_i = \hat{a} \cdot \hat{e}_i + \hat{b} \cdot x_i \cdot \hat{e}_i$$

$$\Rightarrow \quad \sum \hat{y}_i \cdot \hat{e}_i \overset{L2\ ;\ L3}{=} \hat{a} \cdot \sum \hat{e}_i + \hat{b} \cdot \sum x_i \cdot \hat{e}_i$$

$$\Rightarrow \quad \sum \hat{y}_i \cdot \hat{e}_i = 0$$ w.s.m.

Lineare Einfachregession und Residuenplots

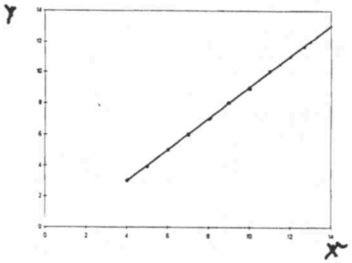
Im folgenden wird jeweils eine Zielgröße Y in Abhängigkeit von einer Einflußgröße X betrachtet.

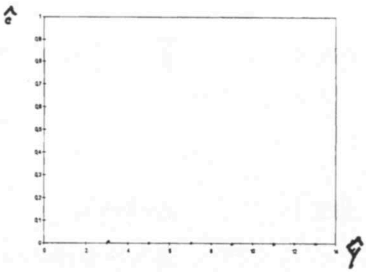
Die Graphiken der linken Seite stellen jeweils ein Streudiagramm der y-Werte gegen die entsprechenden x-Werte mit dazugehöriger Regressionsgeraden dar; die rechten Graphiken beinhalten die entsprechenden Residuenplots.

1. Versuchen Sie anhand der Streudiagramme Aussagen bezüglich des Bestimmtheitsmaßes R^2 zu treffen!

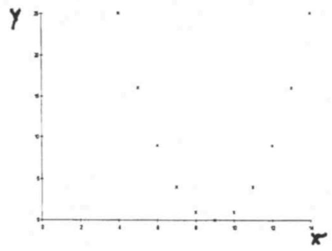
2. Interpretieren Sie die Residuenplots !

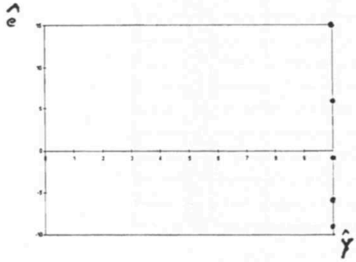
x_i	y_i
10	9
8	7
13	12
9	8
11	10
14	13
6	5
4	3
12	11
7	6
5	4



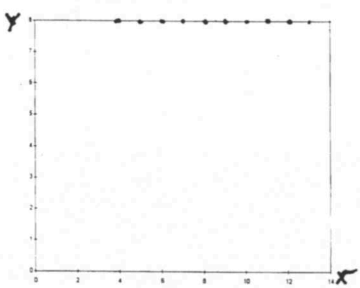


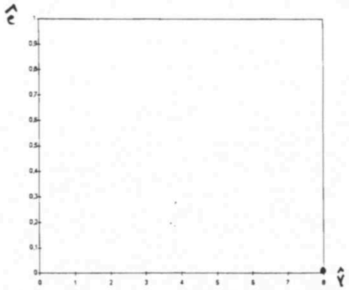
x_i	y_i
10	1
8	1
13	16
9	0
11	4
14	25
6	9
4	25
12	9
7	4
5	16





x_i	y_i
10	8
8	8
13	8
9	8
11	8
14	8
6	8
4	8
12	8
7	8
5	8





Im folgenden wird wiederum jeweils eine Zielgröße Y in Abhängigkeit von einer Einflußgröße X betrachtet.

Die Graphiken der linken Seite stellen jeweils ein Streudiagramm der y-Werte gegen die entsprechenden x-Werte mit dazugehöriger Regressionsgeraden dar; die rechten Graphiken beinhalten die entsprechenden Residuenplots.

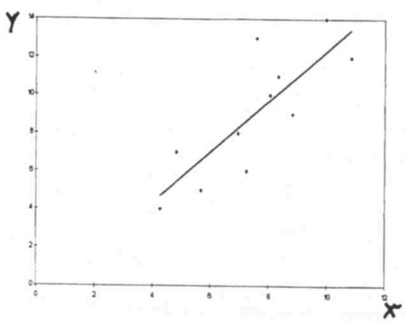
1. Interpretieren Sie Die Streudiagramme !

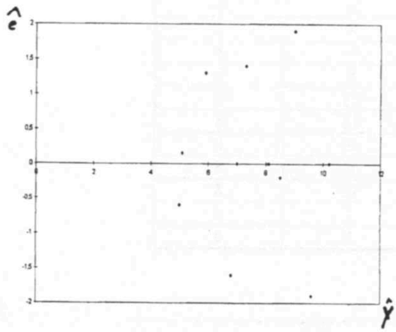
2. Interpretieren Sie die Residuenplots !

3. Nehmen Sie Stellung zu der Aussage:' Der Wert des Bestimmtheitsmaßes gibt Auskunft, ob ein linearer Regressionsansatz gerechtvertigt ist ' !

x_i	y_i
10	8.04
8	6.95
13	7.58
9	8.81
11	8.33
14	9.96
6	7.24
4	4.26
12	10.84
7	4.82
5	5.68

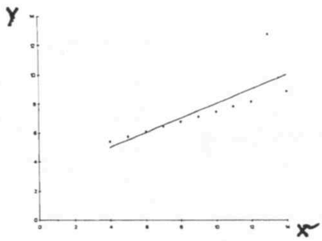
$\hat{y}_i = 3 + 0.5 * x_i$ $\qquad R^2 = 0.67$

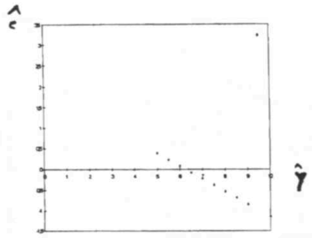




x_i	y_i
10	7.46
8	6.77
13	12.74
9	7.11
11	7.81
14	8.84
6	6.08
4	5.39
12	8.15
7	6.42
5	5.73

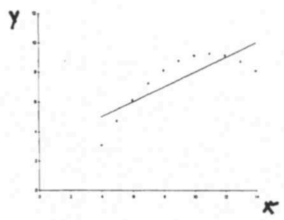
$\hat{y}_i = 3 + 0.5 * x_i$ $\qquad\qquad$ $R^2 = 0.67$

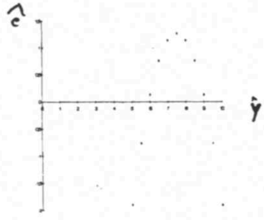




x_i	y_i
10	9.14
8	8.14
13	8.74
9	8.77
11	9.26
14	8.1
6	6.13
4	3.1
12	9.13
7	7.26
5	4.74

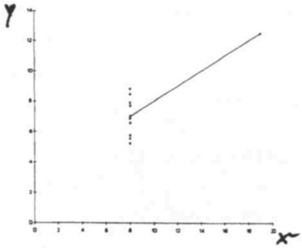
$\hat{y}_i = 3 + 0.5 * x_i$ $\qquad\qquad$ $R^2 = 0.67$

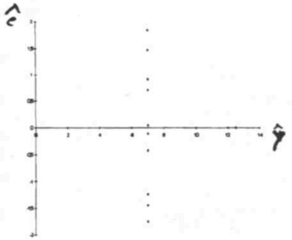




x_i	y_i
8	6.58
8	5.76
8	7.71
8	8.84
8	8.47
8	7.04
8	5.25
8	5.56
8	7.91
8	6.89
19	12.5

$\hat{y}_i = 3 + 0.5 * x_i$ $R^2 = 0.67$





Aufgaben zur Kombinatorik

1) Wie viele verschiedene Tipreihen gibt es bei der 11–er Wette beim Toto ?

2) Wie viele verschiedene Wörter lassen sich aus allen Buchstaben der folgenden Wörter bilden :

   a) dort b) gelesen c) Ruderregatta ?

3) Auf wie viele Arten können 8 Türme auf ein Schachbrett gestellt werden, so daß sie sich nicht schlagen können ?

4) Auf wie viele Arten kann man 5 Hotelgäste auf 10 freie Einzelzimmer verteilen, die sich qualitativ unterscheiden ?

5) Wie groß ist die Anzahl der unterschiedlichen Blattverteilungen beim Skatspiel, wobei jeder der drei Spieler 10 Karten erhält und 2 Karten den Skat bilden ?

6) In einer Projektgruppe von 15 Teilnehmern sollen Untergruppen gebildet werden, wobei jeweils 2 oder 3 Studenten eine Untergruppe bilden. Wieviele verschiedene Aufteilungen sind möglich ?

7) In einer Behörde seien 4 ehrenamtliche Aufgaben zu erfüllen, und es gebe 3 dafür infrage kommende Personen. Wieviele Möglichkeiten der Besetzung der Ehrenämter gibt es, wenn auch mehrere Ämter von einer Person besetzt werden können ?

8) Man bestimme die Anzahl derjenigen vierziffrigen Zahlen, deren Ziffern alle verschieden sind !

9) Ein Autonummernschild besteht aus ein oder zwei Buchstaben gefolgt von 1, 2, 3 oder 4 Ziffern. Wie viele Autos können theoretisch zugelassen werden ?

10) Herr Meyer hat seinen Schlüssel für sein Schließfach am Bahnhof verloren. Die Schließfachnummer hat er leider vergessen. Er erinnert sich allerdings daran, daß es sich um eine vierstellige Zahl handelt, bei der zwei Ziffern gleich sind und daß als Ziffern die 3 , 5 und 7 vorkommen. Wie viele Schließfächer müssen gesperrt werden ?

11) Am Ende einer Vorlesung wurden 68 Scheine ausgestellt und 61 davon verteilt. Im nächsten Semester kommen 2 Studenten möchten ihre Scheine abholen. Der zerstreute Dozent greift blind in den Ordner und händigt jedem der beiden Studenten einen Schein aus. Wie viele Möglichkeiten gibt es, daß

   a) **nicht** jeder der beiden Studenten seinen eigenen Schein bekommt ?

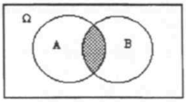
   b) jeder der beiden Studenten **nicht** seinen eigenen Schein bekommt ?

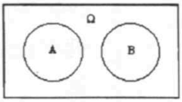
Mengentheoretische Operationen lassen sich in die „ Ereignistheoretische “ Sprache übersetzen :

Mit $A, B \subset \Omega$ gilt :

---

1) **Durchschnitt** : $A \cap B$





Menge aller Elemente von Ω , die

➨ sowohl zu A als auch zu B gehören

besser :

➨ zu A <u>und gleichzeitig</u> zu B gehören

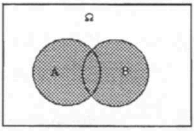
Zwei Ereignisse, die nicht gemeinsam auftreten können, für die also gilt : $A \cap B = \varnothing$, nennt man

<u>disjunkt</u> oder **<u>unvereinbar</u>** oder **<u>elementefremd</u>**

> **<u>Vorsicht</u>** :
>
> Der Gedankengang : „ Zwei Ereignisse, die keine gemeinsamen Elemente haben, haben somit nichts miteinander zu tun, sind also unabhängig “ ➨ **<u>ist falsch</u>** ! ! !
>
> Das Gegenteil stimmt : Wenn eines von beiden eintritt, kann gerade deshalb das andere nicht eintreten !

---

2) **Vereinigung** : $A \cup B$



Menge aller Elemente von Ω , die

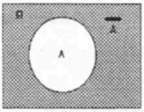
➨ zu A <u>oder</u> zu B <u>oder zu beiden</u> gehören

oder :

➨ mindestens zu einem der beiden Ereignisse gehören

---

3) **Komplementärereignis** : $\overline{A}$



Menge aller Elemente von Ω , die

➨ <u>nicht</u> zu A gehören

<u>Regeln</u> :

$A \cap B = B \cap A$	$A \cap (B \cap C) = (A \cap B) \cap C$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$A \cup B = B \cup A$	$A \cup (B \cup C) = (A \cup B) \cup C$	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

$\overline{A \cup B} = \overline{A} \cap \overline{B}$ ➨ „ <u>beide nicht</u> “ oder „ <u>weder A noch B</u> “

$\overline{A \cap B} = \overline{A} \cup \overline{B}$ ➨ „ <u>höchstens eines von beiden</u> “

$A = (A \cap B) \cup (A \cap \overline{B})$ ➨ disjunkte Zerlegung

Nur für Profis

POKER

Ein Pokerspiel enthält 52 Karten :

2 , 3 , ... , König , As jeweils in den Farben ♠ , ♥ , ♦ , ♣

Sie erhalten 5 Karten : Ihr Pokerblatt

A : Wie viele verschiedene Pokerblätter gibt es ?

B : Wie viele Möglichkeiten gibt es, ein Pokerblatt zu bekommen mit :

❶ „ one pair “ ? ➪ **1.098.240**

❷ „ two pairs “ ? ➪ **123.552**

❸ „ three of a kind “ ? ➪ **54.912**

❹ „ straight “ ? ➪ **10.200**

❺ „ flush “ ? ➪ **5.108**

❻ „ full house “ ? ➪ **3.744**

❼ „ four of a kind “ ? ➪ **624**

❽ „ straight flush “ ? ➪ **40**

---

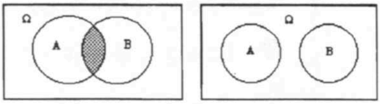
❾ „ no pairs “ ? ➪ **1.302.540**

Mengentheoretische Operationen lassen sich in die „ Ereignistheoretische “ Sprache übersetzen :

Mit $A , B \subset \Omega$ gilt :

---

1) **Durchschnitt** : $A \cap B$



Menge aller Elemente von Ω , die

➨ sowohl zu A als auch zu B gehören

besser :

➨ zu A <u>und gleichzeitig</u> zu B gehören

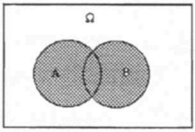
Zwei Ereignisse, die nicht gemeinsam auftreten können, für die also gilt : $A \cap B = \varnothing$, nennt man

<u>disjunkt</u> oder **<u>unvereinbar</u>** oder **<u>elementefremd</u>**

> **<u>Vorsicht</u>** :
>
> Der Gedankengang : „ Zwei Ereignisse, die keine gemeinsamen Elemente haben, haben somit nichts miteinander zu tun, sind also unabhängig “ ➨ **<u>ist falsch</u>** ! ! !
>
> Das Gegenteil stimmt : Wenn eines von beiden eintritt, kann gerade deshalb das andere nicht eintreten !

---

2) **Vereinigung** : $A \cup B$



Menge aller Elemente von Ω , die

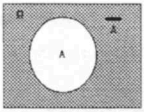
➨ zu A <u>oder</u> zu B <u>oder zu beiden</u> gehören

oder :

➨ mindestens zu einem der beiden Ereignisse gehören

---

3) **Komplementärereignis** : $\overline{A}$



Menge aller Elemente von Ω , die

➨ <u>nicht</u> zu A gehören

<u>Regeln</u> :

$$\boxed{\begin{array}{l} A \cap B = B \cap A \\ A \cup B = B \cup A \end{array}} \quad \boxed{\begin{array}{l} A \cap (B \cap C) = (A \cap B) \cap C \\ A \cup (B \cup C) = (A \cup B) \cup C \end{array}} \quad \boxed{\begin{array}{l} A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \end{array}}$$

$\overline{A \cup B} = \overline{A} \cap \overline{B}$ ➨ „ <u>beide nicht</u> “ oder „ <u>weder A noch B</u> “

$\overline{A \cap B} = \overline{A} \cup \overline{B}$ ➨ „ <u>höchstens eines von beiden</u> “

$A = (A \cap B) \cup (A \cap \overline{B})$ ➨ disjunkte Zerlegung

Ereignisalgebra

Eine Ereignisalgebra ist ein Modell zur Beschreibung eines Zufallsexperiments :

- Einem Zufallsexperiment wird eine Menge Ω ($\Omega \neq \varnothing$) zugeordnet
- Die Elemente von Ω heißen Ergebnisse ω , Ω heißt Ergebnismenge
- Bezüglich Ω wird ein System $\mathscr{A}$ von Teilmengen A , B , C , ... von Ω betrachtet. Diese Teilmengen heißen Ereignisse und sollen folgende Forderungen erfüllen :

❶ $\Omega \in \mathscr{A}$, $\varnothing \in \mathscr{A}$

❷ $A \in \mathscr{A} \Rightarrow \overline{A} \in \mathscr{A}$

❸ $A, B, \ldots \in \mathscr{A} \Rightarrow A \cup B \cup \ldots \in \mathscr{A}$ und $A \cap B \cap \ldots \in \mathscr{A}$

Erfüllt $\mathscr{A}$ diese Bedingungen, so heißt $\mathscr{A}$ **Ereignisalgebra** (auf Ω) .

Diese Forderungen garantieren, daß mit gegebenen Ereignissen A , B , ... sämtliche mengentheoretischen Verknüpfungen auch zu $\mathscr{A}$ gehören, also ebenfalls Ereignisse sind.

Axiomensystem von Kolmogoroff (1933)

Sei Ω eine Ergebnismenge, $\mathscr{A}$ eine Ereignisalgebra auf Ω und P eine reellwertige Funktion mit $\mathscr{A}$ als Definitionsbereich und dem [0 ; 1] als Wertebereich

$$P : \mathscr{A} \longrightarrow [0;1]$$

Dann heißt P **Wahrscheinlichkeitsfunktion** und P (A) heißt **Wahrscheinlichkeit** des Ereignisses A, falls folgende Axiome gelten :

Axiom 1 (Positivität)

$$P(A) \geq 0 \quad \forall A \in \mathscr{A}$$

Axiom 2 (Normierung)

$$P(\Omega) = 1$$

Axiom 3 (Additivität)

Für abzählbar (unendlich) viele, paarweise disjunkte Ereignisse $A_1, A_2, \ldots, A_n, \ldots$ gilt stets :

$$P(A_1 \cup A_2 \cup \ldots \cup A_n \cup \ldots) = P(A_1) + P(A_2) + \ldots + P(A_n) + \ldots \text{ mit } (A_i \cap A_j) = \varnothing \text{ für } i \neq j$$

Zusammenfassung

① Behandlung der Wahrscheinlichkeitsrechnung auf der Grundlage der Kolmogoroffschen Axiome

② Auffassung der Wahrscheinlichkeit kann subjektiv oder objektiv sein

③ Rechenregeln sind für Subjektivisten und Objektivisten dieselben

④ Numerische Bestimmung von Wahrscheinlichkeiten erfolgt

- a – posteriori (frequentistisch)

oder

- a – priori (nach Laplace)

oder

- subjektiv

Aufgaben zur Wahrscheinlichkeitsrechnung

I . Hausschlüssel-Problem

Der zerstreute Student Statistix hat mit 90 %-iger Wahrscheinlichkeit seinen Hausschlüssel dabei.

Wenn er ihn wirklich dabei hat, steckt er ihn immer wahllos mit gleicher Wkt ($^1/_{11}$) in eine seiner 11 Taschen.

Zu Hause angekommen, sucht er seinen Schlüssel. In den ersten 10 Taschen ist der Schlüssel nicht !

Wie groß ist die Wkt, daß er in der 11.ten Tasche steckt ?

II .Goldmünzen-Problem

In drei gleichartigen Kästchen A , B und C befinden sich je zwei Schubladen.

In jeder Schublade liegt entweder ein Goldstück (G) oder ein Silberstück (S).

Die Kästchen sind folgendermaßen bestückt :

A	B	C
G	G	S
G	S	S

1) Ein Kästchen wird zufällig ausgewählt.
   Eine der beiden Schubladen wird geöffnet. Es liegt ein Goldstück drin.

   Wie groß ist die Wkt, daß in beiden Schubladen ein Goldstück liegt ?

2) Ein Kästchen wird zufällig ausgewählt.
   Anhand des Gewichts muß mindestens ein Goldstück drin sein.

   Wie groß ist die Wkt, daß in beiden Schubladen ein Goldstück liegt ?

III . Drei Türen-Problem

Der Spielleiter einer Game-Show hat sich für die Vergabe des großen Preises an den Hauptgewinner des Abends folgendes Spiel ausgedacht :

Spiell : Ich habe hier drei Schachteln : Box A , Box B und Box C . Eine dieser Boxen enthält die Schlüssel für eine komplett eingerichtete Traumvilla, die anderen Boxen sind leer. Wählen Sie eine Box aus !

Kand : Ich nehme Box B !

Spiell : Ich biete Ihnen 1000 € , die Sie sicher mit nach Hause nehmen können, Sie geben mir Ihre Box B zurück und das Spiel ist vorbei.

Kand : Nein, danke !

Spiell : Wie wär's mit 2000 € ?

Kand : Nein !

Spiell : Bedenken Sie, daß Ihre Chance gar nichts zu gewinnen immerhin $^2/_3$ beträgt. Ich biete 5000 € !

Kand : Ich behalte meine Box B !

Spiell : Ich werde Ihnen jetzt einen Gefallen tun und eine der beiden übrigen Boxen öffnen. (Er öffnet Box A, sie ist leer.) Die Chance, daß in Ihrer Box B die Schlüssel sind, beträgt jetzt sogar 50 % ! Ich biete 10.000 € !

Kand : Herr Spielleiter, wissen Sie, in welcher Box die Schlüssel sind ?

Spiell : Ja, warum ?

Kand : Ach, nur so. Ich nehme Ihre 10.000 € **nicht**, möchte aber meine Box B gegen die übrige Box C tauschen. Darf ich ?

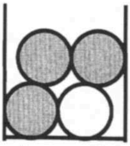
Spiell : Na gut, ich versteh' zwar nicht warum, aber von mir aus !

Wie groß ist die Chance des Kandidaten, die Traumvilla zu gewinnen ?

IV . Gameshow–Problem

Eine Gruppe aus drei Studenten ist das glückliche Gewinnerteam des Abends. Leider steht nur ein Preis (eine tolle und billige Wohnung in der City von Berlin) zur Verfügung. Der Spielleiter bietet folgendes Spiel an :

In einer Urne befinden sich vier Kugeln, drei schwarze Kugeln (≅ Nieten) und eine weiße Kugel (Gewinn). Der Reihe nach darf jeder einmal eine Kugel aus der Urne ziehen (Ziehen ohne Zurücklegen). Wer die weiße Kugel zieht hat den Preis gewonnen.



Die Spieler sollen untereinander ausmachen, wer als erster, als zweiter bzw. als dritter in die Urne greift.

Würden Sie lieber als erster, zweiter oder dritter ziehen ?

V . Flugzeug–Problem

Ein Propellerflugzeug kann sich sicher in der Luft halten, wenn der Hauptmotor in der Mitte oder beide Nebenmotoren funktionieren.
Alle Motoren arbeiten unabhängig voneinander.
Der Hauptmotor funktioniert mit einer Wkeit von 90 % .
Jeder der beiden Nebenmotoren funktioniert mit einer Wkeit von 80 % .

1) Wie groß ist die Wkeit, daß das Flugzeug abstürzt ?

2) Der Hauptmotor fällt aus !
Wie groß ist die Wkeit, daß das Flugzeug abstürzt ?

3) Mindestens ein Nebenmotor fällt aus !
Wie groß ist die Wkeit, daß das Flugzeug abstürzt ?

4) Beide Nebenmotoren fallen aus !
Wie groß ist die Wkeit, daß das Flugzeug abstürzt ?

5) Das Flugzeug stürzt ab !
Wie groß ist die Wkeit, daß der Hauptmotor ausgefallen ist ?

6) Das Flugzeug stürzt ab !
Wie groß ist die Wkeit, daß mindestens ein Nebenmotor ausgefallen ist ?

7) Das Flugzeug stürzt ab !
Wie groß ist die Wkeit, daß beide Nebenmotoren ausgefallen sind ?

VI . Teilungs-Problem

Gerecht teilen – aber wie ?

Zwei Spieler tragen ein Glücksspiel aus, das sich über mehrere Runden erstreckt, in denen die Gewinnchancen jeweils 50 : 50 sind. Den gesamten Gewinn soll derjenige Spieler bekommen, der als Erster vier Runden für sich entscheidet.
Als der Spielstand 3 : 2 erreicht ist, muß das Match leider vorzeitig angebrochen werden. Man einigt sich darauf, den Gesamtgewinn dem Spielstand entsprechend fair und gerecht zu teilen.
Aber welches Teilungsverhältnis ist fair ?

A

Man könnte den Gesamtgewinn im Verhältnis der gewonnenen Runden, im vorliegenden Fall also **3 : 2** , teilen, d.h. Spieler **A** , der drei Spiele gewonnen hat, bekäme **60 %** und Spieler **B** , der nur zwei Spiele gewonnen hat, bekäme **40 %** des Gesamtgewinns. Das entspräche der kaufmännisch üblichen Verfahrensweise, bei der gemeinsam erwirtschaftete Erlöse geteilt werden. Bei dieser Verfahrensweise wird also nur die Vergangenheit bewertet.

B

Man könnte sich jedoch auch an den noch fehlenden Siegen orientieren. So muß Spieler **A** nur noch einen, Spieler **B** jedoch noch zwei Siege bis zum Matchgewinn erreichen. Das führt zu einem Teilungsverhältnis von **2 : 1** , d.h. Spieler **A** , der drei Spiele gewonnen hat, bekäme **66,67 %** und Spieler **B** , der nur zwei Spiele gewonnen hat, bekäme **33,33 %** des Gesamtgewinns. Bei dieser Verfahrensweise wird also die Zukunft bewertet.

C

Sowohl **Fermat** (1601 – 1665) als auch **Pascal** (1623 – 1662) lösten dieses Teilungproblem mit zwei verschiedenen, allgemein anwendbaren Verfahren, deren Resultate aber übereinstimmten [Briefwechsel im Jahre 1654]. Dieser Teilungsplan orientiert sich an den Spielchancen, wie sie sich bei einer fiktiven Fortsetzung des Spiels ergeben. So führt eine weitere Runde entweder zum Spielstand 4 : 2 oder 3 : 3 . Im ersten Fall gewinnt Spieler **A** alles, während es im zweiten Fall zweifellos gerecht ist, den Gesamtgewinn zu halbieren. Eine gerechte Teilung müßte daher im Verhältnis **3 : 1** vorgenommen werden, d.h. Spieler **A** , der drei Spiele gewonnen hat, bekäme **75 %** und Spieler **B** , der nur zwei Spiele gewonnen hat, bekäme **25 %** des Gesamtgewinns. Auch bei dieser Verfahrensweise wird also die Zukunft bewertet.

Bei der zuvor beschriebenen Idee wird also *umgekehrt zur Chronologie des Spiels* vorgegangen. Dabei handelt es sich bei den beiden Anteilen (75 % und 25 %) um nichts anderes als die Gewinnwahrscheinlichkeiten, die beide Spieler besitzen, d.h. der Gesamtgewinn wird im Verhältnis der beiden Gewinnwahrscheinlichkeiten geteilt. Diese Gewinnwahrscheinlichkeiten können nun direkt berechnet werden. Auch hier wird davon ausgegangen, daß noch weitere fiktive Runden gespielt werden, allerdings diesmal gerade genau so viele, wie notwendig sind, damit das Match auf jeden Fall entschieden wird. Daher werden im vorliegenden Beispiel noch genau **zwei** fiktive Runden gespielt und zwar selbst dann, wenn die erste Runde bereits das Match entscheiden sollte. Bei diesen zwei Runden sind insgesamt **4** verschiedene Spielverläufe möglich, die alle untereinander gleichmöglich sind und daher die Wahrscheinlichkeit ¼ besitzen :

nächste Runde	übernächste Runde
A gewinnt	**A** gewinnt
A gewinnt	**B** gewinnt
B gewinnt	**A** gewinnt
B gewinnt	**B** gewinnt

Nur im letzten der zusammengestellten Fälle gewinnt Spieler **B** das Match. Seine Gewinnwahrscheinlichkeit beträgt daher nur ¼ , während Spieler **A** mit einer Wahrscheinlichkeit von ¾ gewinnt. Eine gerechte Teilung müßte daher im Verhältnis **3 : 1** vorgenommen werden, d.h. Spieler **A** , der drei Spiele gewonnen hat, bekäme **75 %** und Spieler **B** , der nur zwei Spiele gewonnen hat, bekäme **25 %** des Gesamtgewinns.

V30a

Beweise : Rechenregeln ⇨ Erwartungswert und theor. Varianz

Bew. : ❶

$$E(X+a)=\sum_x (x+a)\cdot P(X=x)=\sum_x \left[x\cdot P(X=x)+a\cdot P(X=x)\right]=\sum_x x\cdot P(X=x)+\sum_x a\cdot P(X=x)$$
$$=E(X)+a\cdot\sum_x P(X=x)=E(X)+a$$

Bew. : ❷

$$E(a\cdot X)=\sum_x a\cdot x\cdot P(X=x)=a\cdot\sum_x x\cdot P(X=x)=a\cdot E(X)$$

Bew. : ❸

$$E(X\pm Y)=\sum_{x\pm y}(x\pm y)\cdot P[(X=x)\cap(Y=y)]=\sum_x\sum_y (x\pm y)\cdot P[(X=x)\cap(Y=y)]$$
$$=\sum_x\sum_y x\cdot P[(X=x)\cap(Y=y)]\pm\sum_y\sum_x y\cdot P[(X=x)\cap(Y=y)]$$
$$=\sum_x x\cdot\underbrace{\sum_y P[(X=x)\cap(Y=y)]}_{=P(X=x)}\pm\sum_y y\cdot\underbrace{\sum_x P[(X=x)\cap(Y=y)]}_{=P(Y=y)}$$
$$=\sum_x x\cdot P(X=x)\pm\sum_y y\cdot P(Y=y)=E(X)\pm E(Y)$$

Bew. : ①

$$\mathrm{Var}(X+a)=E[X+a-E(X+a)]^2=E[X+a-E(X)-a]^2=E[X-E(X)]^2=\mathrm{Var}(X)$$

Bew. : ②

$$\mathrm{Var}(a\cdot X)=E[a\cdot X-E(a\cdot X)]^2=E[a\cdot X-a\cdot E(X)]^2$$
$$=E\left\{a^2\cdot X^2-2\cdot a^2\cdot E(X)\cdot X+a^2\cdot[E(X)]^2\right\}$$
$$=a^2\cdot\left\{E\left(X^2\right)-2\cdot E(X)\cdot E(X)+[E(X)]^2\right\}$$
$$=a^2\cdot\left\{E\left(X^2\right)-2\cdot[E(X)]^2+[E(X)]^2\right\}$$
$$=a^2\cdot\left\{E\left(X^2\right)-[E(X)]^2\right\}=a^2\cdot\mathrm{Var}(X)$$

Bew. : ③

$$\mathrm{Var}(X+Y)=E[(X+Y)-E(X+Y)]^2=E[(X+Y)^2]-[E(X+Y)]^2$$
$$=E[X^2+2XY+Y^2]-[E(X)+E(Y)]^2$$
$$=E(X^2)+2E(XY)+E(Y^2)-[E(X)]^2-2E(X)E(Y)-[E(Y)]^2$$
$$=E(X^2)-[E(X)]^2+E(Y^2)-[E(Y)]^2+2[E(XY)-E(X)E(Y)]$$
$$=\quad \mathrm{Var}(X)\quad+\quad\mathrm{Var}(Y)\quad+\quad 2\,\mathrm{Cov}(X,Y)$$

Das Zugabteil-Problem

Jeder von zwei Reisenden (Herr A und Herr B) setzen sich zufällig und unabhängig voneinander in eines von **3** Zugabteilen.

Die Zufallsvariable X bezeichne die Zahl der von den Reisenden besetzten Abteile.

Die Zufallsvariable Y_i ($i = 1,2,3$) bezeichne die Zahl der Reisenden im i - ten Abteil.

1) Bestimmen Sie die gemeinsame Wahrscheinlichkeitsfunktion der Zufallsvariablen (X , Y_3) !

2) Berechnen Sie : $\mathrm{Cov}(X , Y_3)$!

Das Fahrstuhl-Problem

In einem Wohnhaus fahren unabhängig voneinander zwei kleine Fahrstühle. Jeder Fahrstuhl kann maximal **3** Personen befördern und ist zu jedem Zeitpunkt mit einer Wahrscheinlichkeit von $^2/_3$ besetzt (d.h. es befindet sich mindestens eine Person darin). Falls ein Fahrstuhl besetzt ist, folgt die „ Anzahl der sich in diesem Fahrstuhl befindenden Personen " einer statistischen Gleichverteilung.

1) Bestimmen Sie die Wahrscheinlichkeitsfunktion der Zufallsvariablen

   X : „ Anzahl der besetzten Fahrstühle " !

2) Bestimmen Sie die Wahrscheinlichkeitsfunktion der Zufallsvariablen

   Y : „ Anzahl der Fahrgäste **eines** Fahrstuhls " !

3) Bestimmen Sie die Wahrscheinlichkeitsfunktion der Zufallsvariablen

   Z : „ Anzahl der Fahrgäste in **beiden** Fahrstühlen " !

4) Bestimmen Sie die gemeinsame Wahrscheinlichkeitsfunktion der Zufallsvariablen (X , Z) !

Das Tennisball-Problem

Lucy geht heute mit ihrem Ballkarton unter dem Arm in den Tennisclub.
In ihrem Karton befinden sich **5** Tennisbälle, von denen **2** noch **un**gespielt und **3** bereits gespielt sind. Für ein Tennisspiel nimmt Lucy zufällig **3** Tennisbälle aus dem Karton heraus und legt sie nach dem Spiel in den Karton zurück.
Nach zwei Stunden hat Lucy das Tennisspiel leider verloren und beschließt, noch ein wenig an der Ballwand zu üben. Zu diesem Zweck entnimmt sie zufällig einen Ball aus dem Karton.
Sei

X : „Anzahl der **un**gespielten Bälle, die Lucy für ihr Tennisspiel aus dem Karton entnimmt"

Y : „Anzahl der **un**gespielten Bälle, die Lucy zum Üben aus dem Karton entnimmt"

1) Bestimmen Sie die gemeinsame Wahrscheinlichkeitsfunktion von X und Y !

2) Berechnen Sie : $\mathrm{Cov}(X , Y)$!

3) Wie groß ist die Wahrscheinlichkeit dafür, daß sich auf Lucys Heimweg vom Tennisclub noch mindestens ein **un**gespielter Ball in ihrem Karton befindet !

Schiefe einer Wahrscheinlichkeitsverteilung

Ein auf den Abständen der Quantile x_p und x_{1-p} zum Median $\tilde{\mu}$ aufgebautes Schiefemaß ist der

theoretische p – Quantilskoeffizient der Schiefe

$$\gamma_p = \frac{(x_{1-p} - \tilde{\mu}) - (\tilde{\mu} - x_p)}{x_{1-p} - x_p} \qquad (0 < p < 0{,}5)$$

$\gamma_{0,25}$ heißt „ *Quartilskoeffizient der Schiefe* "

$\gamma_p = 0$	➡	Verteilung symmetrisch
$\gamma_p < 0$	➡	Verteilung rechtssteil (≅ linksschief)
$\gamma_p > 0$	➡	Verteilung linkssteil (≅ rechtsschief)

$$\gamma_1 = E\left(\frac{X-\mu}{\sigma}\right)^3$$

heißt „ *theor. Momentenkoeffizient der Schiefe* "

Exzeß einer Wahrscheinlichkeitsverteilung

Der Exzeß γ_2 gibt an, ob (bei gleicher Varianz) das abs. Maximum der eingipfligen Wkts-verteilung größer ist als bei der Dichte der Normalverteilung.

Der theor. Exzeß einer Normalverteilung ist gerade Null !

$$\gamma_2 = E\left(\frac{X-\mu}{\sigma}\right)^4 - 3$$

$\gamma_2 = 0$	➡	Exzeß genau wie bei N.V.
$\gamma_2 < 0$	➡	Exzeß **kleiner** als bei N.V. d.h. „ *platykurtisch* " [platys ≅ **flach**]
$\gamma_2 > 0$	➡	Exzeß **größer** als bei N.V. d.h. „ *leptokurtisch* " [leptos ≅ **schmal**]

Fazit :

Sind Schiefe und Exzeß einer Wktsverteilung wesentlich von Null verschieden, so ist dies ein Hinweis darauf, daß die Verteilung der ZV wesentlich von einer Normalverteilung abweicht !

√32a

Beispiele : Binomialverteilung & Hypergeometrische Verteilung

Beispiel 1 :

Die Wkeit, bei einer U-Bahnfahrt kontrolliert zu werden, betrage 10 % .

Wie groß ist die Wkeit, bei 20 Fahrten

1) höchstens 3-mal
2) mehr als 3-mal
3) weniger als 3-mal
4) mindestens 3-mal
5) genau 3-mal
6) mehr als 1-mal, aber weniger als 4-mal

kontrolliert zu werden ?

Beispiel 2 :

Bei der Produktion von Leuchtraketen treten 10 % Ausschuss auf. Für die in 10-er Packungen angebotenen Feuerwerkskörper garantiert der Hersteller eine „ Zündsicherheit " von 80 % .

Mit welcher Wkeit reklamiert ein verärgerter Kunde nach Sylvester eine nicht der Garantie entsprechende Packung ?

Beispiel 3 :

Durch überhöhte Geschwindigkeit werden 80 % aller Verkehrsunfälle verursacht.

Wie groß ist die Wkeit, dass von 20 Verkehrsunfällen

1) mindestens 12
2) weniger als 15
3) mehr als 12, aber höchstens 17

durch überhöhte Geschwindigkeit verursacht werden ?

Beispiel 4 :

Die Männer Anton und Bert sind glühende Verehrer der schönen Susi, die beiden gerade den Laufpass gegeben hat. Zum Abschied schenkt ihr jeder eine Schachtel Pralinen der Marke Sonnentau (Inhalt 15 Stück).
Anton hat in seiner Schachtel **8** Pralinen vergiftet, Bert in seiner Schachtel hingegen nur **3** .
Susi isst aus einer der beiden Schachteln 6 Pralinen und fällt tot um.
Die Polizei stellt fest, dass Susi genau 2 vergiftete Pralinen gegessen haben muss, findet aber auf beiden Schachteln nur Fingerabdrücke von Susi.

Wer ist vermutlich der Mörder ???

V32b

Modellaufgabe : Exponentialverteilung

Jäger Waldmeister weiß aus Erfahrung, dass früh morgens zwischen 4^{00} und 6^{00} Uhr durchschnittlich alle 10 Minuten ein Nasenbär eine Waldwiese aufsucht, die von einem Hochsitz aus eingesehen werden kann. Der Jäger entschließt sich, am nächsten Tag um 4^{00} Uhr früh auf dem Hochsitz zu sein.

1) Wie groß ist die Wkeit, dass der Jäger höchstens 10 Minuten warten muss, bis zum erstenmal ein Nasenbär erscheint ?

Der Jäger kann sich aus dienstlichen Gründen leider nur 40 Minuten auf dem Hochsitz aufhalten und hat nun bereits 10 Minuten erfolglos auf das Erscheinen eines Nasenbären gewartet.

2) Wie groß ist die Wkeit, dass sich vor seinem Abstieg vom Hochsitz noch ein Nasenbär auf der Wiese blicken lässt ?

Modellaufgabe : Normalverteilung

Ein Statistikdozent macht mit seinen Studenten einen Test (Lösen einer Denksportaufgabe).

Die erreichte Zeit in diesem Test sei normalverteilt mit $\mu = 53$ [Min] und $\sigma^2 = 25^2$ [Min ^{2}].

Die besten 10 % der Testpersonen erhalten eine Auszeichnung, die schlechtesten 15 % der Testpersonen bestehen den Test nicht.

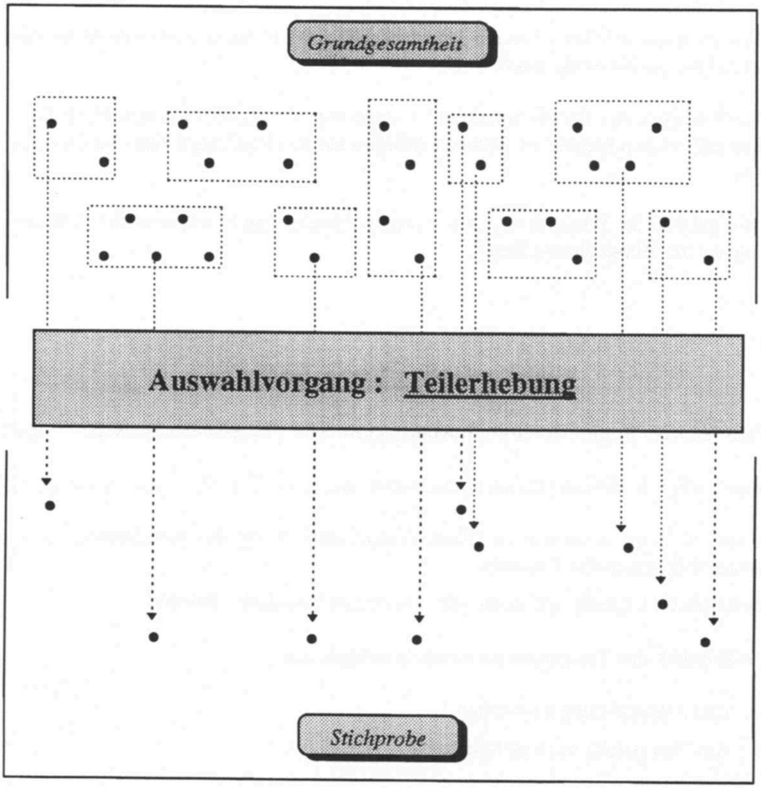
Der Dozent heißt Leopold und ist als sehr „ studentenfreundlich " bekannt.

Welche Zeit darf eine Testperson höchstens benötigen, um

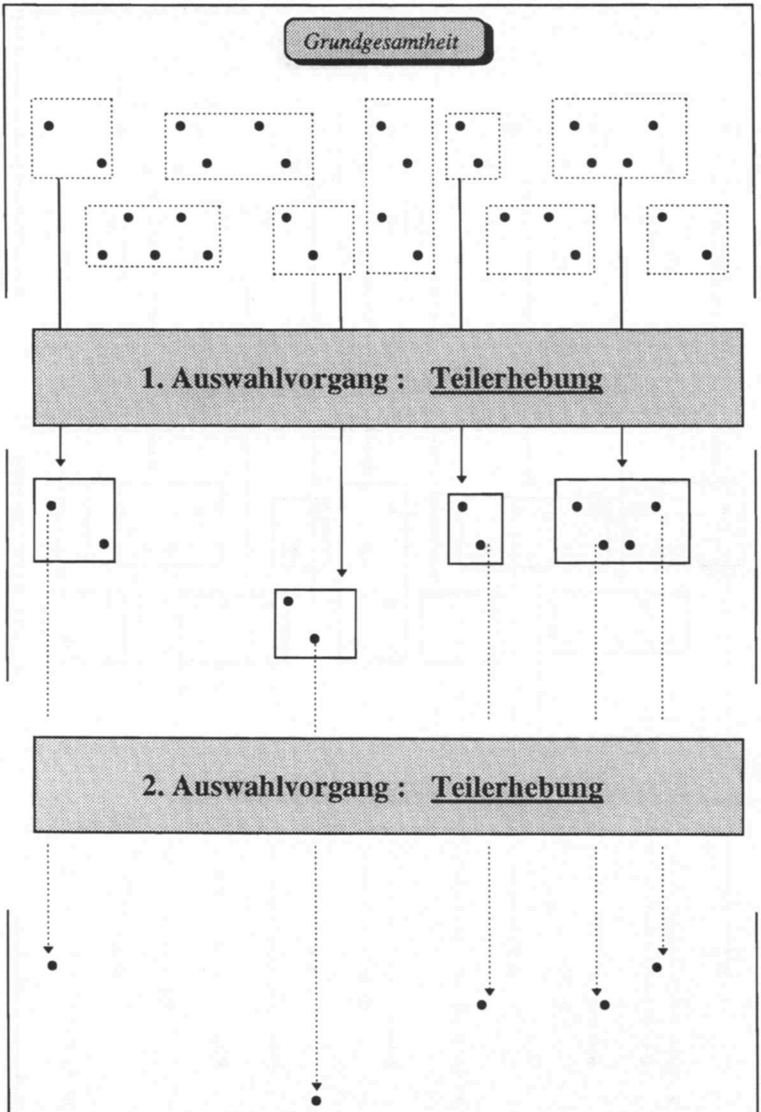
1) eine Auszeichnung zu erhalten ?
2) den Test gerade noch zu bestehen ?

Graphische Darstellung der Auswahlverfahren

Einstufiges Auswahlverfahren

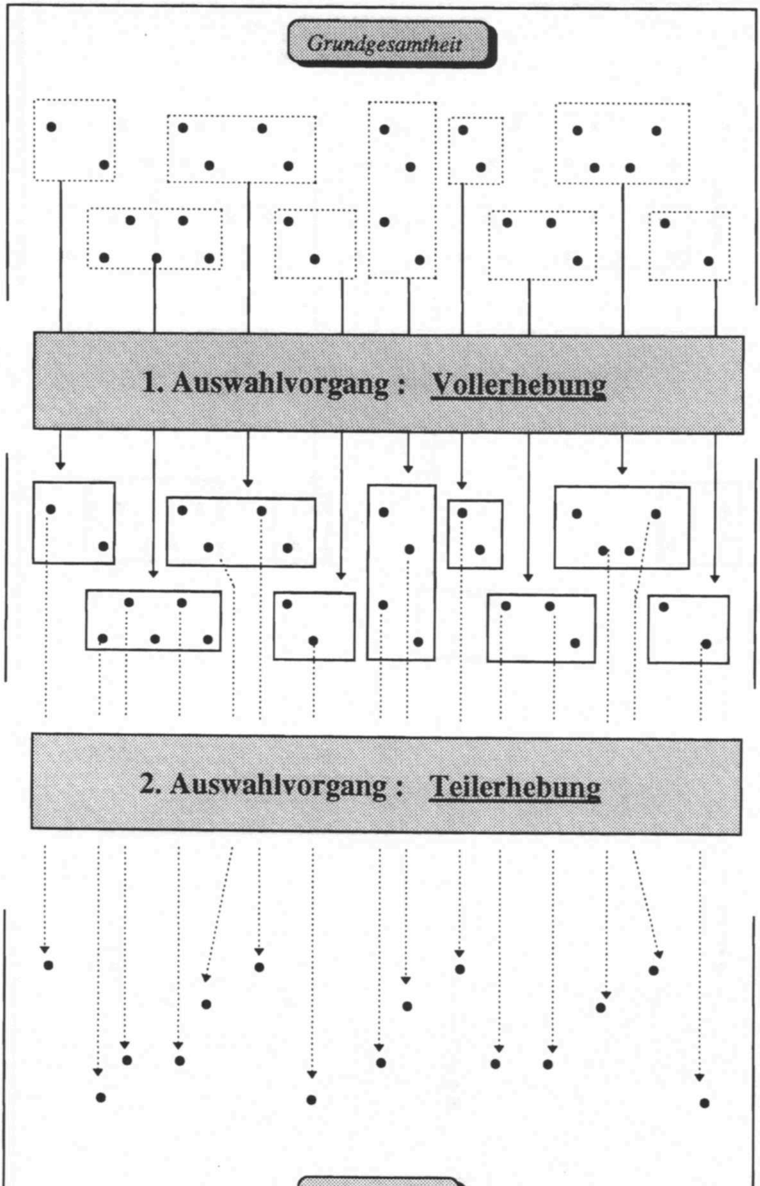


<u>Zweistufiges Auswahlverfahren</u> : Allgemeiner Fall

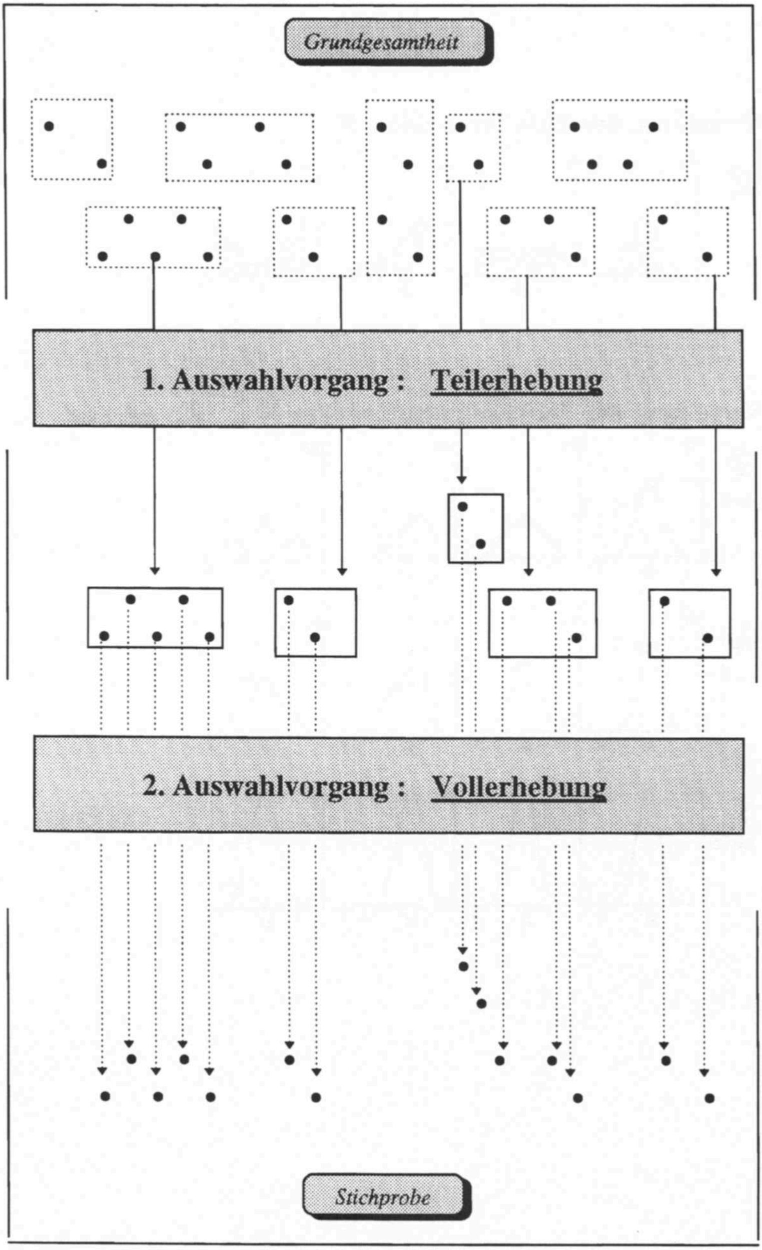


V35

Zweistufiges Auswahlverfahren : **Schichtungsverfahren**

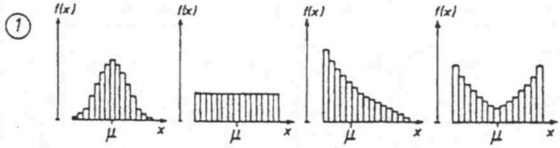


<u>Zweistufiges Auswahlverfahren</u> : Klumpungsverfahren

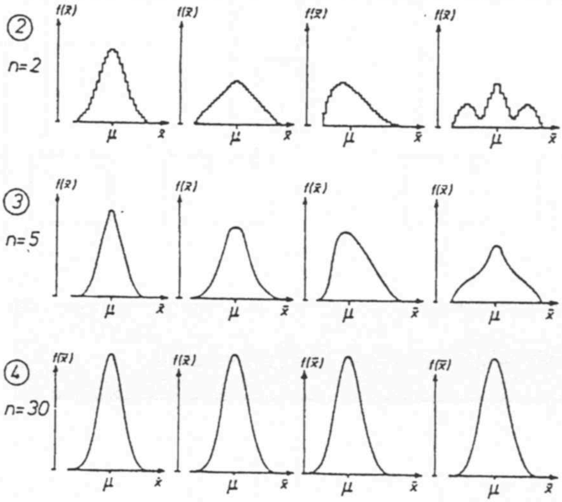


V37

Verteilung der Zufallsvariablen X



Verteilung der Durchschnittsfunktion $\overline{X}$, X_i ist iid



Zur Likelihood – Theorie

Die **Maximum – Likelihood – Methode** besagt, daß zu einem festen Stichprobenergebnis ($x_1, \dots x_n$) derjenige Schätzwert für den unbekannten (zu schätzenden) Parameter ϑ zu wählen ist, unter dem **von vornherein** die Wahrscheinlichkeit für das Eintreten dieses Stichprobenergebnisses am größten ist.

Die Funktion $L[\,\vartheta \mid (x_1, \dots, x_n)\,]$ gibt zu jedem ϑ – Wert an, wie wahrscheinlich es **war**, die Stichprobenwerte ($x_1, \dots, x_n$) zu beobachten, wenn ϑ der wahre Parameter ist.

Einem Parameter ϑ, für den die Beobachtung ($x_1, \dots, x_n$) **von vorherein** wahrscheinlicher war als für einen anderen Parameterwert, wird **im nachhinein** die höhere Glaubwürdigkeit (Plausibilität) zugesprochen, der wahre Parameter der Grundgesamtheit zu sein.

Sei ($X_1, \dots X_n$) eine **einfache** Stichprobe.

$$L\left[\vartheta \mid (x_1, \dots, x_n)\right] = \begin{cases} \prod_{i=1}^{n} P\left(X_i = x_i \mid \vartheta\right) & X_i \ \textbf{diskret} \\ \prod_{i=1}^{n} f\left(x_i \mid \vartheta\right) & X_i \ \textbf{stetig} \end{cases}$$

Ein Parameterwert ϑ_1 ist aufgrund der Beobachtung ($x_1, \dots, x_n$) **plausibler als** ein Parameterwert ϑ_2, wenn gilt : $L[\,\vartheta_1 \mid (x_1, \dots, x_n)\,] > L[\,\vartheta_2 \mid (x_1, \dots, x_n)\,]$

Grundidee der Likelihood – Theorie

Vor der Beobachtung ordnet ein ϑ – Wert den möglichen Werten von ($X_1, \dots, X_n$) eine **Wahrscheinlichkeit** zu, nach der Beobachtung ordnet ein festes Stichprobenergebnis ($x_1, \dots, x_n$) den möglichen ϑ – Werten eine **Likelihood** zu !

Durch die Beobachtung haben sich die Rollen von ϑ und ($x_1, \dots, x_n$) vertauscht !

Intervallschätzung

V39

Beispiel 1

Eine Abfüllmaschine in einer Brauerei füllt 0.5 l -Flaschen ab. Es kann angenommen werden, daß die Füllmenge normalverteilt ist mit einer Standardabweichung von $\sigma = 5$ ml. Eine zufällige Stichprobe von n=4 abgefüllten Flaschen ergab folgende Werte (in ml) :

$x_1 = 502$ $x_2 = 498$ $x_3 = 497$ $x_4 = 501$

1. Bestimmen Sie ein 95%-Konfidenzintervall für die durchschnittliche Abfüllmenge der Maschine !
2. Berechnen Sie das Schätzintervall !
3. Wieviele Flaschen hätten als Stichprobe gezogen werden müssen, um beim Konfidenzniveau 0.95 ein Konfidenzintervall zu erhalten, das höchstens 3 ml lang ist ?

Beispiel 2

Von der Abfüllmaschine aus Beispiel 1 sei nur bekannt, daß die Abfüllmenge normalverteilt ist, allerdings nicht mehr mit welcher Varianz.

1. Bestimmen Sie aufgrund der Stichprobenwerte aus Beispiel 1 ein 95%-Konfidenzintervall für die durchschnittliche Abfüllmenge der Maschine.

Beispiel 3

Und wieder unsere Abfüllmaschine aus Beispiel 1.
Es wurde eine Zufallsstichprobe vom Umfang n=36 gezogen der Produktion entnommen, aus der sich ein $\bar{x} = 498$ und ein $s^2 = 5.8$ ergab.

1. Bestimmen Sie aus diesem Stichprobenergebnis ein 95%-Konfidenzintervall für die durchschnittlich abgefüllte Menge, wenn über die Verteilung der X_i nichts bekannt ist !

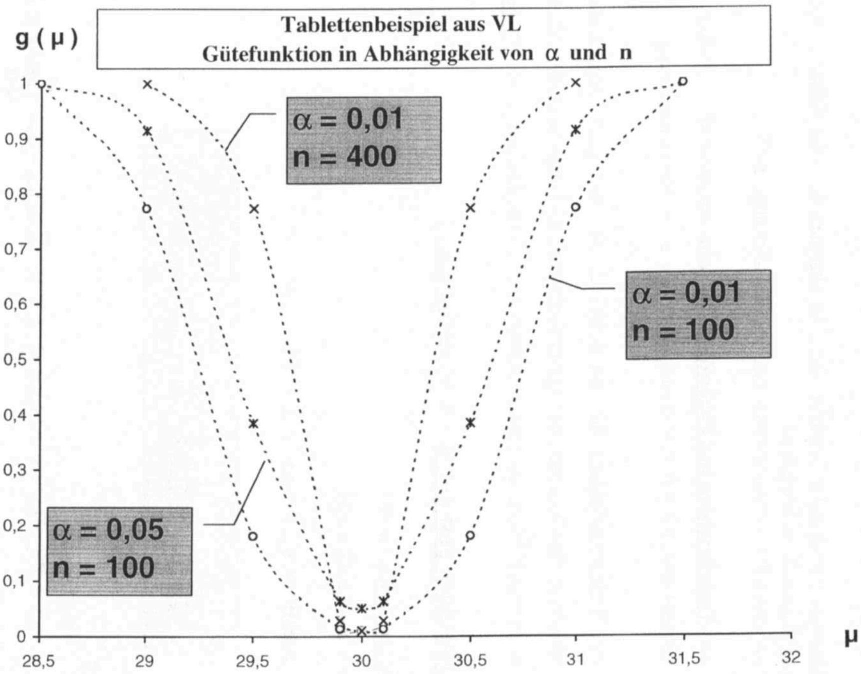
Beispiel 4

Ein Gartenfreund pflanzt im Frühjahr 20 kleine Salatpflänzchen. Im Lauf der Zeit stellt sich heraus, daß 5 davon eingehen und nur 15 überleben.

1. Wie groß ist die Wahrscheinlichkeit, daß Salatpflänzchen dieser Sorte nach dem Setzen eingehen ? (Konfidenzniveau $1 - \alpha = 0.95$)

V 39a

	μ	$g(\mu)$
$\alpha = 0{,}05$, $n = 100$	28,5	1
	29	0,915
	29,5	0,385
	29,9	0,063
	30	0,05
	30,1	0,063
	30,5	0,385
	31	0,915
	31,5	1
$\alpha = 0{,}01$, $n = 100$	28,5	1
	29	0,774
	29,5	0,181
	29,9	0,014
	30	0,01
	30,1	0,014
	30,5	0,181
	31	0,774
	31,5	1
$\alpha = 0{,}01$, $n = 400$	28,5	1
	29	1
	29,5	0,774
	29,9	0,028
	30	0,01
	30,1	0,028
	30,5	0,774
	31	1
	31,5	1



V396

Bemerkungen zum Test

① Mehrmals wurde bisher erwähnt, daß „ im allgemeinen " der Fehler 2. Art (leider) nicht unter Kontrolle zu bringen ist.
Unter speziellen Voraussetzungen gelingt dies allerdings doch :

- der **Verteilungstyp der Prüfgröße** muss nicht nur *unter* H_0 , sondern auch ***unter*** H_1 bekannt sein, wie z.B. beim Gauß - Test auf μ ⇨ Normalverteilung.

- die Wahrscheinlichkeit für den Fehler 2. Art darf nicht vom Stichprobenergebnis abhängen, was bedeutet, der Annahmebereich (und damit auch der Ablehnbereich) darf nicht vom Stichprobenergebnis abhängen, d.h. beim Test auf μ ⇨ **σ^2 bekannt**.

Einseitiger Test auf μ (s. Modellaufgabe)

$H_0 : \mu \geq \mu_0 \; (= 30)$

$H_1 : \mu < \mu_0 \; (= 30)$

$\alpha = 0{,}05$ ⇨ $c_\alpha = 1{,}65$; $n = 100$; $\sigma^2 = 9$

Forderung : $\beta_{(\mu = 29{,}8)} \overset{!}{=} 0{,}25$

$$\beta_{(\mu = 29{,}8)} = P(\,„H_0\text{“} / \mu = 29{,}8 \in H_1\,) = P\left(\overline{X} \geq \mu_0 - c_\alpha \cdot \frac{\sigma}{\sqrt{n}} \;/\; \mu \in H_1\right)$$

$$= P\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \geq \underbrace{\frac{\mu_0 - c_\alpha \cdot \frac{\sigma}{\sqrt{n}} - \mu}{\frac{\sigma}{\sqrt{n}}}}_{c_\beta}\right) \overset{!}{=} 0{,}25$$

V39c

$$\Rightarrow \quad \frac{\mu_o - c_\alpha \cdot \frac{\sigma}{\sqrt{n}} - \mu}{\frac{\sigma}{\sqrt{n}}} = c_\beta \quad \left[= \underline{0{,}68}\right]$$

$$\Rightarrow \quad \mu_o - c_\alpha \cdot \frac{\sigma}{\sqrt{n}} - \mu = c_\beta \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \quad \mu_o - \mu = c_\alpha \cdot \frac{\sigma}{\sqrt{n}} + c_\beta \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \quad \mu_o - \mu = \frac{\sigma}{\sqrt{n}} \cdot (c_\alpha + c_\beta)$$

$$\Rightarrow \quad \sqrt{n} = \left[\frac{\sigma \cdot (c_\alpha + c_\beta)}{\mu_o - \mu}\right]$$

$$\Rightarrow \quad \boxed{n \;\geq\; \left[\frac{\sigma \cdot (c_\alpha + c_\beta)}{\mu_o - \mu}\right]^2}$$

<u>Also</u> :

$$n \;\geq\; \left[\frac{3 \cdot (1{,}65 + 0{,}68)}{30 - 29{,}8}\right]^2 \;=\; 34{,}95^2 \;=\; 1221{,}5$$

$$\boxed{\underline{n \geq 1222}}$$

$$\boxed{\beta(\mu = 29{,}8) \overset{!}{=} \mathbf{0{,}1} \;\Rightarrow \ldots \Rightarrow\; n \geq 1944{,}81 \qquad \Rightarrow \quad \underline{n \geq 1945}}$$

Wie man sieht, hat (auch in der Statistik) alles seinen Preis !

Die aufgestellte Forderung kann zwar erfüllt werden, aber nur durch einen erheblich höheren Datenaufwand [Kosten- und Zeitfaktor] !

V 342

② Daten, die zur Durchführung eines statistischen Tests „ inspirieren “ , dürfen **nicht** zur Testdurchführung verwendet werden, da das Testergebnis praktisch vorher bereits feststeht.

③ Das Ergebnis eines Test auf einem bestimmten Signifikanzniveau α möge lauten : „ H_O “ !
„ H_1 “ , also ein signifikantes Ergebnis wäre einem aber lieber gewesen. Was liegt näher, als mit jeweils immer neuen Daten die Testdurchführung so lange zu wiederholen, bis sich endlich das gewünschte Testergebnis : „ H_1 “ einstellt ! ? !
Keine gute Idee, denn bei Gültigkeit von H_O ist die Entscheidung ja richtig !
Wird der Test aber sehr häufig wiederholt durchgeführt, kommt man fast zwangsläufig (in durchschnittlich $100 \cdot \alpha$ % aller Fälle) zu „ H_1 “ , also der Ablehnung einer richtigen Nullhypothese !

④ Das Ergebnis eines Test auf einem bestimmten Signifikanzniveau α möge lauten : „ H_O “ !
„ H_1 “ , also ein signifikantes Ergebnis wäre einem aber lieber gewesen. Was liegt näher, als „ nachträglich “ α genau um so viel zu erhöhen, dass der Wert der Prüfgröße gerade in den Ablehnbereich fällt ! ? !
Und noch draufgesattelt : Warum eigentlich nicht von vornherein α = 99,99 % ?
Jetzt bekommt man fast immer „ signifikante “ Testergebnisse „ H_1 “ ! ! !

⑤ **„ Empirisches “ α^* - Niveau**

Sei $H_O : \mu \geq \mu_O$
$H_1 : \mu < \mu_O$

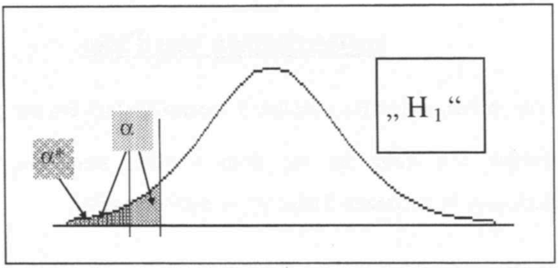
Computer-Programme oder sogar manche Taschenrechner, in denen die Prüfverteilung des Tests programmiert vorliegt, drucken das Testergebnis nicht in Form von „ H_O “ oder „ H_1 “ aus, sondern sagen, auf welchem Niveau der berechnete Wert der Prüfgröße signifikant **wäre** ! Dieses Niveau nennt man **„ empirisches “ α^* - Niveau** (Level attained , p - Value , descriptive Level) .

Fall A : $\alpha^* \leq \alpha \Rightarrow$ **„ H_1 “**

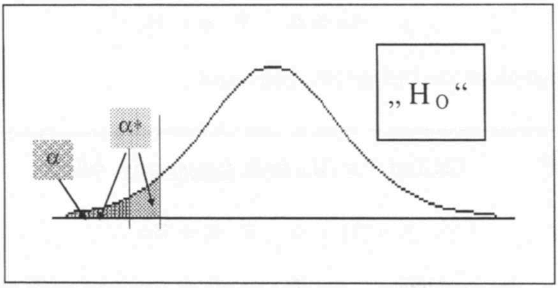
Fall B : $\alpha^* > \alpha \Rightarrow$ **„ H_O “**

V 39e

Fall A



Fall B



Natürlich muss im Sinne einer „ sauberen “ Vorgehensweise das Signifikanzniveau α <u>**vor**</u> der Testdurchführung festgelegt werden, so dass der spätere Erhalt von α^* praktisch gleichbedeutend mit der Testentscheidung ist und nicht nachträglich zur Manipulation in Hinblick auf vielleicht erwünschte „ signifikante “ Testergebnisse dienen darf (s. ④) !

⑥ In manchen (älteren) Statistik-Büchern findet man folgende gefährliche Globalisierung :

~~$\alpha \cong$ **Produzentenrisiko** ; $\beta \cong$ **Konsumentenrisiko**~~

Das mag in speziellen Fällen durchaus der Sachlage entsprechen, grundsätzlich jedoch gilt :

Wer führt den Test durch, und **welche Interessen** leiten ihn ?

- **Der Verbraucherverband als Vertreter der Interessen des Konsumenten** möchte nachweisen, dass im Mittel zu wenig abgefüllt wird [$\cong$ $\mathbf{H_1}$] .
- **Der Abfüller selbst als Produzent** möchte nachweisen, dass er im Mittel mehr als erforderlich abfüllt [$\cong$ $\mathbf{H_1}$] .
- **Die Gewerbeaufsicht** möchte (zur Bestätigung ihrer Daseinsberechtigung) nachweisen, dass im Mittel nicht korrekt abgefüllt wird [$\cong$ $\mathbf{H_1}$] .

Gütekriterien von Tests

Ein Test auf ϑ (H_0 gegen H_1) sei mit T , seine Gütefunktion mit $g_T(\vartheta)$ bezeichnet.

Ein Vergleich von Tests ist nur dann sinnvoll, wenn die betrachteten Tests dasselbe Signifikanzniveau α aufweisen. Daher sei α fest vorgegeben.

Weiterhin seien nur „ **<u>Niveau - α - Tests</u>** “ zugelassen, also Tests, für die gilt :

$$g_T(\vartheta) \leq \alpha \quad \forall \ \vartheta \in H_0$$

Die Menge dieser Tests sei mit $\mathcal{M}_\alpha$ bezeichnet.

❶ Ein Test $T \in \mathcal{M}_\alpha$ heißt **<u>konservativ</u>** , falls gilt :

$$g_T(\vartheta) < \alpha \quad \forall \ \vartheta \in H_0$$

<u>Interpretation</u> : *Der Test schöpft die vorgegebene Wahrscheinlichkeit für den Fehler 1. Art nicht voll aus !*

❷ Ein Test $T^* \in \mathcal{M}_\alpha$ heißt **<u>gleichmäßig bester</u>** (≅ **trennscharfer**) Test, fall gilt :

$$g_{T^*}(\vartheta) \geq g_T(\vartheta) \quad \forall \ \vartheta \in H_1 \quad \text{und} \quad \forall \ T \in \mathcal{M}_\alpha$$

<u>Interpretation</u> : *Kein anderer Test hat auf dem gesamten Hypothesenbereich $\mathcal{H}_1$ eine steilere Gütefunktion !*

❸ Sei $T \in \mathcal{M}_\alpha$. Gibt es ein $\vartheta \in H_0$ mit $g_T(\vartheta) = \alpha$, so heißt T **<u>unverfälscht</u>** , falls gilt :

$$g_T(\vartheta) \geq \alpha \quad \forall \ \vartheta \in H_1$$

<u>Interpretation</u> : *Die Nullhypothese wird, wenn sie nicht zutrifft (d.h. $\vartheta \in \mathcal{H}_1$), mit einer mindestens so hohen Wahrscheinlichkeit abgelehnt wie im Fall ihres Zutreffens (d.h. $\vartheta \in \mathcal{H}_0$) !*

❹ Ein Test $T \in \mathcal{M}_\alpha$ heißt **konsistent**, falls gilt :

$$\lim_{n\to\infty} P\left(\hat{\Theta}_n \in B \;/\; \vartheta \in H_1\right) = 1$$

Interpretation : *Mit wachsendem Stichprobenumfang wird die Ablehnung einer unzutreffenden Nullhypothese immer sicherer !*

Modellaufgabe zu den Zweistichproben-Mittelwert-Tests:

Auf einer Messe stellen zwei Hersteller von Küchenherden unabhängig voneinander ihr neues Modell vor. Beide Herde sind mit Schnellkochplatten ausgerüstet, die nach 3 Minuten Temperaturen von über 100°C erreichen. Die Firman X will nun mit Hilfe eines statistischen Tests beweisen, daß ihre Schnellkochplatten im Mittel heißer werden, als die der Firma Y.
(α = 0.01).

I. Dabei weiß man aus Erfahrung, daß die erreichte Temperatur für beide Grundgesamtheiten (Herdproduktion der Firman X bzw. Y) eine normalverteilte Größe ist mit $\sigma_X^2 = 17$ und $\sigma_Y^2 = 8$. Eine Zufallsstichprobe von $n_X = 4$ ergab eine mittlere Temperatur von 108°C, eine Zufallsstichprobe von $n_Y = 4$ ergab eine mittlere Temperatur von 102 °C.

II. Man weiß aus Erfahrung, daß die erreichte Temperatur eine normalverteilte Größe ist mit einer Varianz von 8. Zufallsstichproben vom Umfang $n_X = 4$ bzw. $n_Y = 4$ ergaben mittlere Temperaturen von 108°C bzw. 102°C.

III. Man weiß aus Erfahrung jedoch nur, daß die erreichte Temperatur eine normalverteilte Größe ist mit einer Varianz σ^2. Zufallsstichproben von $n_X = 4$ bzw. $n_Y = 4$ ergaben eine mittlere Temperatur von 108°C bei einer Varianz von 10 bzw. eine mittlere Temperatur von 102 °C bei einer Varianz von 6.

IV. Man weiß man aus Erfahrung, daß die erreichte Temperatur in beiden Grundgesamtheiten eine normalverteilte Größe ist . Zufallsstichproben vom Umfang $n_X = 400$ bzw. $n_Y = 400$ ergaben eine mittlere Temperatur von 107°C bei einer Varianz von 9 bzw. 106.5 °C bei einer Varianz von 7.

V. Zufallsstichproben vom Umfang $n_X = 400$ bzw. $n_Y = 400$ ergaben eine mittlere Temperatur von 107°C bei einer Varianz von 9 bzw. 106.5 °C bei einer Varianz von 7.

Überleitung : II ⇨ III

pooled variance

$\boxed{\text{II}}$: $V = \dfrac{\overline{X} - \overline{Y}}{\sigma \cdot \sqrt{\dfrac{n_X + n_Y}{n_X \cdot n_Y}}}$

Zur Schätzung der **unbekannten** Varianz σ^2 wird man beide Stichprobenvarianzfunktionen $S_X{}^2$ und $S_Y{}^2$ heranziehen !

$W_X = \dfrac{(n_X - 1) \cdot S_X^2}{\sigma^2}$ ist χ^2 – verteilt mit $f_X = n_X - 1$

$W_Y = \dfrac{(n_Y - 1) \cdot S_Y^2}{\sigma^2}$ ist χ^2 – verteilt mit $f_Y = n_Y - 1$

Da W_X und W_Y unabhängig sind, gilt :

$W = W_X + W_Y = \dfrac{(n_X - 1) \cdot S_X^2 + (n_Y - 1) \cdot S_Y^2}{\sigma^2}$ ist χ^2 – verteilt mit

$f = f_X + f_Y = (n_X - 1) + (n_Y - 1) = \underline{n_X + n_Y - 2}$

Dann gilt : $T = \dfrac{V}{\sqrt{\dfrac{W}{f}}}$ ist t – verteilt mit $f = n_X + n_Y - 2$ und $V \sim N(0 ; 1)$

Also :

$$T = \frac{\dfrac{\overline{X} - \overline{Y}}{\sqrt{\sigma^2 \cdot \dfrac{n_X + n_Y}{n_X \cdot n_Y}}}}{\sqrt{\dfrac{\dfrac{(n_X - 1) \cdot S_X^2 + (n_Y - 1) \cdot S_Y^2}{\sigma^2}}{n_X + n_Y - 2}}} = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma^2 \cdot (n_X + n_Y)}{n_X \cdot n_Y} \cdot \dfrac{(n_X - 1) \cdot S_X^2 + (n_Y - 1) \cdot S_Y^2}{\sigma^2 \cdot (n_X + n_Y - 2)}}}$$

$$= \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{(n_X - 1) \cdot S_X^2 + (n_Y - 1) \cdot S_Y^2}{(n_X + n_Y - 2)}} \cdot \sqrt{\dfrac{n_X + n_Y}{n_X \cdot n_Y}}} = \frac{\overline{X} - \overline{Y}}{S \cdot \sqrt{\dfrac{n_X + n_Y}{n_X \cdot n_Y}}} \quad : \boxed{\text{III}}$$

Übersicht

verbundene oder unabhängige Stichproben

Zweistichproben t–Test	Differenzen t–Test
unabhängige Stichproben	**verbundene** Stichproben
n_1 bzw. n_2 Probanden	n Probanden
⇨ $n_1 + n_2$ Daten	⇨ $2 \cdot n$ Daten

Zum Vergleich : $n_1 + n_2 = 2 \cdot n$

---

❶ **Kostenfaktor**

doppelt so viele Probanden „ – “	halb so viele Probanden „ + “

❷ **Varianzhomogenität : $\sigma_X^2 \overset{?}{=} \sigma_Y^2$**

Entscheidung : ja ⇨ Zweist.–t–Test; nein ⇨ Test von Welch (umständlich, unscharf) „ – “	Entscheidung : entfällt „ + “

❸ **Freiheitsgrade**

$f = n_1 + n_2 - 2 = 2n - 2 = 2 \cdot (n-1)$

➪ Ablehnber. größer
➪ Annahmeber. kleiner
➪ β–Risiko kleiner
➪ Test trennschärfer

„ + “

$f = n - 1$ (halb so groß)

➪ Ablehnber. kleiner
➪ Annahmebereich größer
➪ β–Risiko größer
➪ Test weniger trennscharf

„ – “

❹ $\text{Varianz}(X - Y) = \sigma_X^2 + \sigma_Y^2 - 2 \cdot \sigma_{XY}$

$\text{Var}(X - Y) = \sigma_X^2 + \sigma_Y^2$

„ – “

$\text{Var}(X - Y) = \sigma_X^2 + \sigma_Y^2 \boxed{- 2 \cdot \sigma_{XY}}$

Je stärker X und Y (bei je gleichem Probanden) **positiv korreliert** sind, um so stärker wirkt sich die **Varianzreduzierung** aus !

➪ Test trennschärfer !

„ + “

V46

Herleitung der Prüfverteilung für den Vorzeichenrangtest nach Wilcoxon für n = 5

$n = 5$; $\quad w^{+}_{max} = \dfrac{n\cdot(n+1)}{2} = \sum_{i=1}^{5} i = \underline{15}$; $\quad a(w^{+}) :=$ Anzahl der Möglichkeiten, „ + “ und „ – “ Zeichen den ersten n natürlichen Zahlen so zuzuordnen, daß die Summe der positiven Zahlen gleich w^{+} ist.

① w^{+}	Rangtupel, die zu positiven Differenzen gehören	$a(w^{+})$	$P(W^{+} = w^{+}) = \dfrac{a(w^{+})}{2^{n}}$ ②	$P(W^{+} \leq w^{+})$
0	(–)	1	$\frac{1}{32} = 0{,}03125$	0,03125
1	(1)	1	$\frac{1}{32} = 0{,}03125$	0,06250
2	(2)	1	$\frac{1}{32} = 0{,}03125$	0,09375
3	(1 , 2) | (3)	2	$\frac{2}{32} = 0{,}06250$	0,15625
4	(1 , 3) | (4)	2	$\frac{2}{32} = 0{,}06250$	0,21875
5	(1 , 4) | (2 , 3) | (5) ③	3	$\frac{3}{32} = 0{,}09375$	0,31250
6	(1 , 2 , 3) | (1 , 5) | (2 , 4)	3	$\frac{3}{32} = 0{,}09375$	0,40625
7	(1 , 2 , 4) | (2 , 5) | (3 , 4)	3	$\frac{3}{32} = 0{,}09375$	0,50000

8	(3 , 5) | (1 , 3 , 4) | (1 , 2 , 5)	3	$\frac{3}{32} = 0{,}09375$	0,59375
9	(4 , 5) | (2 , 3 , 4) | (1 , 3 , 5)	3	$\frac{3}{32} = 0{,}09375$	0,68750
10	(1 , 4 , 5) | (2 , 3 , 5) | (1 , 2 , 3 , 4)	3	$\frac{3}{32} = 0{,}09375$	0,78125
11	(2 , 4 , 5) | (1 , 2 , 3 , 5)	2	$\frac{2}{32} = 0{,}06250$	0,84375
12	(3 , 4 , 5) | (1 , 2 , 4 , 5)	2	$\frac{2}{32} = 0{,}06250$	0,90625
13	(1 , 3 , 4 , 5)	1	$\frac{1}{32} = 0{,}03125$	0,93750
14	(2 , 3 , 4 , 5)	1	$\frac{1}{32} = 0{,}03125$	0,96875
15	(1 , 2 , 3 , 4 , 5)	1	$\frac{1}{32} = 0{,}03125$	1,00000

① $w^+ \cong$ Summe der Rangwerte, die sich aus <u>positiven</u> Differenzen $(x_i - \tilde{\mu}_O)$ ergeben

② Jeder Rangplatz hat bzgl. seines Vorzeichens („ + “ oder „ – “) 2 Möglichkeiten

③ <u>Zweit</u>kleinste und <u>viert</u>kleinste Differenz $(x_i - \tilde{\mu}_O)$ ist positiv

V47