Beispielsammlung Statistik

Übungsbeispiele für die Vorlesung und Übung Statistik

Institut für Operations und Information Systems

WS 2021

Liebe Studierende!

Diese Beispielsammlung beinhaltet Übungsbeispiele zu sämtlichen relevanten Kapiteln der Lehrveranstaltungen aus Statistik. Die Übungsaufgaben sind so gewählt, dass sie Ihnen einen Überblick über den Stoff der Statistik geben und so die Lehrveranstaltungen ergänzen. Mit den Erläuterungen aus der Vorlesung und den Übungen sollten Sie die Aufgaben lösen können. Sie finden zu allen Beispielen die Lösungen im Anschluss an die Angabe, um Ihnen eine Kontrolle Ihres Lösungsweges zu ermöglichen.

Einige Beispiele sind mit V (Video) markiert. Diese finden sich komplett durchgerechnet und erklärt in Form von Videos auf der Moodle-Seite der Vorlesung Statistik (z.B. finden Sie Beispiel 2.16 V unter „Übungsbeispiele Blatt 2“ im Video „Blatt 2_Beispiel 16“). Sie können sich jederzeit zu dieser Vorlesung anmelden und haben dadurch Zugriff auf diese Videos.

Thomas Ebner, Oktober 2021

Übungen zu Blatt 1

1.1 V Wir unterscheiden verschiedene Arten von Merkmalen. Nennen Sie für jedes Erhebungsmerkmal eine zulässige Ausprägung, geben Sie jeweils eine (sachlogisch) plausible Skalierung an und bestimmen Sie, ob das Merkmal stetig oder diskret ist:

Merkmal	Ausprägung	Skalenniveau	diskret/stetig
Lebensalter (vollendete Jahre)			
Geschlecht			
Nationalität			
Monatliche Miete			
Familienstand			
Beruf			
Unternehmensrating			
Akademischer Grad			
Einkommen			
Postleitzahl			
Schuhgröße			
Erwerbslosenanteil			

Lösung:

Merkmal	Ausprägung	Skalenniveau	diskret/stetig
Lebensalter (vollendete Jahre)	$\mathbb{N}$, z. B.: 34	metrisch (absolut)	diskret
Geschlecht	z.B.: w	nominal	diskret
Nationalität	AUT, D	nominal	diskret
Monatliche Miete	$\mathbb{R}_{++}$, z. B.: 512.-	metrisch (Verhältnis)	(quasi-) stetig
Familienstand	ledig, verheiratet	nominal	diskret
Beruf	Tischler, Arzt	nominal	diskret
Unternehmensrating	AAA, AA, AB, ...	ordinal	diskret
Akademischer Grad	Bakk, Mag, DI, Dr	nominal (ordinal)	diskret
Einkommen	$\mathbb{R}_{+}$, z. B.: 1200.-	metrisch (Verhältnis)	(quasi-) stetig
Postleitzahl	8010, 8020	nominal	diskret
Schuhgröße	38, 42	metrisch (Verhältnis)(ordinal)	diskret
Erwerbslosenanteil	5,9%	metrisch (Verhältnis)	(quasi-) stetig

1.2 $\boxed{\text{V}}$ Gegeben sind die Daten einiger Wohnungen. Dabei sind die Mieten in Euro und die Wohnfläche in m^2 angegeben. Bei der Lage bedeuten 0: „schlecht", 1: „mittelmäßig" und 2: „gut. Versuchen Sie, die gegebenen Daten sinnvoll zu verdichten.

Nr.	Miete	Wohnfläche	Zimmer	Baujahr	Stadtbezirk	Lage
1	831,28	92	3	1966	16	0
2	621,79	65	2	1998	24	1
3	740,71	70	3	1967	24	0
4	586,11	78	3	1989	16	0
5	423,43	68	3	1982	19	2

Die Daten zu dieser Aufgabe finden Sie als Excel-Datei auf der Webseite des Instituts unter dem Namen „Datensätze" im Tabellenblatt „Wohnungen".

Lösung: Arithmetisches Mittel Miete: $640,66$; Arithmetisches Mittel Wohnfläche: $74,6$; Arithmetisches Mittel Zimmer: $2,8$; Modalwert Zimmer: 3; Modalwert Stadtbezirk: 16 bzw. 24; Modalwert Lage: 0; Median Lage: 0

1.3 $\boxed{\text{V}}$ Bei einunddreißig Läufern, die in zwei aufeinanderfolgenden Jahren am Welschlauf im Halbmarathon teilgenommen haben, wurden einige Merkmale erhoben: Alter, Herkunft, Beruf, Gewicht, Laufzeit 2009, Laufzeit 2010, Einkommen. Man beachte: das Merkmal „Herkunft" besitzt drei Ausprägungen: Steiermark (ST), restliches Österreich (AUT), Ausland (GAST); das Merkmal „Beruf" ist in drei Kategorien geteilt: Angestellt (A), Selbständig (S), Auszubildender (AZUBI); das Merkmal „Einkommen" ist ebenfalls in drei Kategorien unterteilt: gering (g), mittel (m), hoch (h). Die Daten finden Sie im Tabellenblatt „Welschlauf". Beantworten Sie die folgenden Fragen.

a) Wie viel Prozent der Läufer gaben „Mittleres Einkommen" an?

b) Erstellen Sie eine (Kontingenz-)Tabelle, aus der man entnehmen kann, wie viele Teilnehmer je nach Herkunft in die drei Einkommensklassen fallen!

c) Welcher Anteil (welcher Prozentsatz) der Selbständigen kam aus dem Ausland?

d) Welcher Anteil (welcher Prozentsatz) der Ausländer ist selbständig?

e) Vergleichen Sie das Durchschnittsgewicht der Steirer mit jenem der anderen Teilnehmer.

f) Berechnen Sie die Varianz für das Merkmal „Körpergewicht"

i. als mittlere Summe der quadrierten Abweichungen,

ii. mit Hilfe des Verschiebungssatzes und

iii. mit einer Excel-Formel.

Lösung: a) 35%; b) Excel; c) 30%; d) 50%; e) $\overline{x}_{ST} = 65,765$ kg; $\overline{x}_{\text{Nicht}ST} = 65,357$ kg; f) $Var\,(\text{Gewicht}) = 82,114$

1.4 $\boxed{\text{V}}$ Die 60 Studierenden einer Statistikübung wurden gebeten anzugeben, an wie vielen Tagen der Vorwoche sie zur Universität gefahren sind. Die Antworten wurden in der nachfolgenden Tabelle zusammengefasst. Die Daten finden Sie im Tabellenblatt „Unifahrten"

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

a) Wie ist das Merkmal skaliert?

b) Wie lautet die (ursprüngliche) Tabelle der absoluten und der relativen Häufigkeiten?

c) Stellen Sie die Daten in einem Stabdiagramm dar.

d) Stellen Sie die Daten in einem Kreisdiagramm dar.

e) Geben Sie die kumulierten relativen Häufigkeiten an.

f) Bestimmen Sie den geeigneten Lageparameter.

g) Zeichnen Sie die empirische Verteilungsfunktion.

Lösung:

a) kardinal, verhältnis skaliert

b)

x_i	1	2	3	4	5
Absolute Häufigkeit H_i	15	6	18	9	12
Relative Häufigkeit h_i	0,25	0,1	0,3	0,15	0,2
$F(x)$	0,25	0,35	0,65	0,8	1

c) Excel

d) Excel

e) Tabelle

f) $\overline{x} = 2,95$

g) Excel

1.5 Bei der Mitarbeit von Studierenden während des Semesters wurde in einer Statistikgruppe folgende Punkteverteilung beobachtet:

Punkte	0	1	2	3	4	5
Häufigkeit	3	3	6	12	24	12

a) Geben Sie die relativen und die kumulierten relativen Häufigkeiten an.

b) Stellen Sie die Daten durch ein Stabdiagramm und durch die empirische Verteilungsfunktion dar.

c) Berechnen Sie Median und Mittelwert.

Lösung:

a)

x_i	0	1	2	3	4	5
Absolute Häufigkeit H_i	3	3	6	12	24	12
Relative Häufigkeit h_i	0,05	0,05	0,1	0,2	0,4	0,2
$F(x)$	0,05	0,1	0,2	0,4	0,8	1

b) Excel

c) Median = 4, Mittelwert = 3,45

1.6 $\boxed{\text{V}}$ Betrachten Sie die Daten des Welschlaufs von Beispiel 3.

a) Berechnen Sie für das Merkmal „Körpergewicht“ Modalwert, Median, Mittelwert, Varianz, Standardabweichung, Variationskoeffizient, Spannweite, 60-%-Quantil, Quartilsabstand, Schiefe und Wölbung und interpretieren Sie all diese Zahlen! Welche dieser Kennzahlen sind aussagekräftig?

b) Wie verändert eine falsche Zahl, wenn etwa statt „58“ einmal der Wert „580“ eingegeben worden ist, die Kennzahlen Median, Mittelwert, Quartilsabstand und Standardabweichung?

c) Zeichnen und interpretieren Sie einen Boxplot zu diesen Daten.

Lösung: Excel

1.7 An einer Registrierkasse eines Supermarktes wurde bei 8 KundInnen jeweils die Bedienungszeit (in Sekunden) gemessen:

Kunde	1	2	3	4	5	6	7	8
Bedienungszeit (in s)	36	48	17	35	67	44	28	68

a) Berechnen Sie die mittlere Bedienungszeit.

b) Berechnen Sie die Varianz, sowie den Variationskoeffizienten der Bedienungszeiten.

c) Bestimmen Sie den Median sowie das untere und obere Quartil der Bedienungszeiten. Zeichnen Sie einen Boxplot der Bedienungszeiten.

Lösung: a) $\overline{x} = 42,875$; b) $Var(X) = 280,109$; $V = 0,390$; c) Median = 40, Q1 (Excel) = 33,25; Q3(Excel) = 52,75 bzw. Q1 = 31,5; Q3 = 57,5

1.8 $\boxed{\text{V}}$ In der nachstehenden Tabelle sind die Mietpreise (in €/m^2) von Zwei-Zimmer-Mietwohnungen in vergleichbarer Wohnlage, die im ersten Quartal auf dem Wohnungsmarkt einer Kleinstadt angeboten wurden, angegeben:

Mietpreis (in €/m^2) [von, bis[	$[4, 6[$	$[6, 9[$	$[9, 12[$	$[12, 18]$
Anzahl der Wohnungen	8	30	42	24

a) Berechnen sie soweit möglich Mittelwert und Standardabweichung.

b) Aus einer weiteren Erhebung mit 80 Zwei-Zimmer-Mietwohnungen ergab sich ein Mittelwert von 7,70 €/m2. Wie groß ist – soweit berechenbar – der mittlere Mietpreis in allen Wohnungen?

c) Kann man auch die Standardabweichung aller Wohnungen bestimmen, wenn man weiß, dass die neuen 80 Beobachtungen eine Standardabweichung von 3 aufweisen?

Lösung: a) $\overline{x}_1 = 1025$; $\sigma_1 = 3,088$; b) $\overline{x}_{ges} = 9,141$; c) $\sigma_{ges} = 10,9$

1.9 $\boxed{\text{V}}$ Ein Anleger kauft am 15.1., am 15.4. und am 15.7. eines Jahres jeweils Aktien eines Ölkonzerns. Der Kurs der Aktie betrug an diesen Börsetagen:

Börsentag i	15.1.	15.4.	15.7.
Kurs (€/Aktie) x_i	200	250	400

Wie hoch ist der durchschnittliche Kaufpreis pro Aktie, wenn der Anleger:

a) am 15.1. und am 15.4. jeweils 40 Aktien und am 15.7. noch einmal 20 Aktien kauft?

b) am 15.1. und am 15.4. für jeweils 10.000 € und am 15.7. für 4.000 € Aktien kauft?

Lösung: a) gewogenes Mittel: 260 €/Stück; 240 €/Stück

1.10 In einem Betrieb wurden die Bruttojahreslöhne der Angestellten ermittelt, und das Ergebnis in der folgenden Tabelle zusammengefasst:

Lohnklasse in 1000 €	$[10, 25[$	$[25, 30[$	$[30, 40[$	$[40, 60[$	$[60, 90[$
Anzahl der Angestellten	40	20	60	50	30

a) Zeichnen Sie das zugehörige Histogramm!

b) Zeichnen Sie die approximierende Verteilungsfunktion und markieren Sie in der Zeichnung den Median sowie das erste und dritte Quartil!

Lösung:

a)

Lohnklasse in 1000 €	$[10, 25[$	$[25, 30[$	$[30, 40[$	$[40, 60[$	$[60, 90[$
Anzahl der Angestellten	40	20	60	50	30
h_i	0,2	0,1	0,3	0,25	0,15
kumuliert	0,2	0,3	0,6	0,85	1
Besetzungsdichte	0,013	0,02	0,03	0,0125	0,005

b) Graphik

1.11 Schwermetallwägungen in Rindernieren ergaben folgende (klassierte) Werte für den Cadmiumgehalt (in mg/kg):

mg Cadmium	$[0,2;0,3[$	$[0,3;0,6[$	$[0,6;1,2[$	$[1,2;1,8]$
Klassenhäufigkeit	10	20	30	40

a) Wie groß ist und was bedeutet die kumulierte relative Häufigkeit an der Stelle 1,2?

b) Berechnen sie soweit wie möglich Mittelwert und Standardabweichung.

c) Zeichnen Sie die approximierende Verteilungsfunktion und markieren Sie in der Zeichnung den Median.

d) Aus einer weiteren Erhebung mit 81 gemessenen Cadmiumwerten ergab sich ein Mittelwert von 0,326 mg/kg. Wie groß ist – soweit berechenbar – der Mittlere Gehalt an Cd in allen 181 Proben? Kann man auch die Varianz aller 181 Proben bestimmen, wenn man weiss, dass die neuen 81 Messungen eine Varianz von 0,2 aufweisen?

Lösung:

a) $F(1,2) = 0,6$

mg Cadmium	$[0,2;0,3[$	$[0,3;0,6[$	$[0,6;1,2[$	$[1,2;1,8]$
Klassenhäufigkeit	10	20	30	40
h_i	0,1	0,2	0,3	0,4
kumuliert	0,1	0,3	0,6	1

b) $\hat{x} = 0,985$; $\hat{\sigma} = 0,469$

c) Tabelle

d) $\overline{x}_{ges} = 0,690$; $\sigma_{ges} = 0,564$

1.12 $\boxed{\text{V}}$ In einer bestimmten Branche konkurrieren zehn Unternehmen miteinander. Nach ihrem Umsatz lassen sich diese in drei Klassen einteilen: fünf kleine, vier mittlere und ein großes Unternehmen. Bei den kleinen Unternehmen beträgt der Umsatz im Durchschnitt 0,6 Mio. € und bei den mittleren Unternehmen 1,5 Mio. €. Das große Unternehmen erzielt einen Umsatz von 6 Mio. €. Zeichnen Sie die zugehörige Lorenzkurve und berechnen Sie den Gini-Koeffizienten.

Lösung: $Gini = 0,42$

1.13 $\boxed{\text{V}}$ Betrachten Sie die Daten des Welschlaufs aus Beispiel 3. Berechnen Sie eine geeignete Kennzahl, um eine Aussage über die Abhängigkeit zwischen den Merkmalen „Herkunft“ und „Einkommen“ treffen zu können!

Lösung: $c_{korr} = 0,547$

1.14 $\boxed{\text{V}}$ In Deutschland wird das Abitur zentral gestellt und mit Punkten von 0 bis 15 bewertet. Acht Abiturienten erreichten in den Fächern „Deutsch“ und „Mathematik“ die folgenden Punktezahlen:

Deutsch	13	14	8	10	15	1	12	11
Mathematik	15	8	1	7	9	4	5	10

a) Berechnen Sie den Pearson-Korrelationskoeffizienten

b) Berechnen Sie den Pearson-Korrelationskoeffizienten der Ränge

c) Geben Sie die Gleichung der Regressionsgeraden zur Beschreibung der Abhängigkeit der Punkte in Mathematik von den Punkten in Deutsch an.

Lösung: a) $\rho = 0,565$; b) $\rho_{\text{Ränge}} = 0,667$; c) $y = 1,707 + 0,54x$

1.15 In einer kleinen Goldgräberstadt wurde erhoben, wie der Besitz von Goldbarren verteilt ist. Dabei ergaben sich die Besitzverhältnisse wie folgt:

Anzahl Goldbarren	0	2	3	5	10
Anzahl Personen	80	60	30	20	10

a) Zeichnen Sie die Lorenzkurve.

b) Berechnen Sie den Gini-Koeffizienten.

Lösung: $Gini = 0,577$

1.16 $\boxed{\text{V}}$ Die Bevölkerung Indiens von 1951 bis 2011 wird gemäß „Census of India“ in folgender Tabelle dargestellt:

Jahr	1951	1961	1971	1981	1991	2001	2011
Bev. (in Mio.)	361	439	548	683	846	1.029	1.211

a) Erstellen Sie ein Streudiagramm, das die Bevölkerung Indiens im Zeitverlauf darstellt.

b) Um wie viel Prozent ist die Bevölkerung Indiens jeweils in den Zehnjahresabständen gewachsen?

c) Um wie viel Prozent ist die Bevölkerung Indiens in den Jahren 1951 bis 2011 insgesamt gewachsen, um wie viel Prozent im Jahresdurchschnitt? Geeignete Mittelwertbildung!

d) Erstellen Sie ein Streudiagramm, das die Wachstumsraten der Bevölkerung Indiens im Zeitverlauf darstellt.

e) Prognostizieren sie auch mit Hilfe der jährlichen Wachstumsraten die Bevölkerungszahlen für die Jahre 2021 und 2031.

Lösung:

a) Excel

b)

Jahr	1951	1961	1971	1981	1991	2001	2011
Wachstum in %		21,61	24,83	24,64	23,87	21,63	17,69

c) $\overline{x}_{geom} = 2,04$

d) Excel

e) $x\,(2021) = 1481,666$; $x\,(2023) = 1812,827$

1.17 Um eine möglicherweise vorhandene Abhängigkeit zwischen höchstem Schulabschluss und regionaler Herkunft aufzudecken, wurden dreißigjährige Personen befragt und man erhielt folgende Ergebnisse in Form einer Tabelle:

	Schulabschluss		
Herkunft	Universität oder FH	Matura	Hauptschule
Wien, Linz, Graz	46	78	53
Andere Bezirkshauptstädte	40	39	79
Sonstige Gemeinden	24	72	68

Berechnen Sie eine geeignete Kennzahl, um eine Aussage über die Abhängigkeit treffen zu können!

Lösung: $0,2686$

1.18 In einem Betrieb wurden für 7 Lehrlinge jeweils der Notenschnitt des Abschlusszeugnisses aus der Schule und eine im Betrieb ermittelte Leistungskennzahl (Je höher, desto besser!) miteinander verglichen:

Lehrling	1	2	3	4	5	6	7
Notenschnitt	2,6	2,2	1,9	2,4	3,3	1,3	1,8
Leistungskennzahl	61	79	82	72	66	89	71

Bestimmen Sie unterschiedliche Kennzahlen um den Zusammenhang zwischen den beiden Merkmalen zu beschreiben! (Skalierung, Interpretation!)

Lösung: $-0,784$; $+0,75$; $-0,75$

1.19 Gegeben ist folgender Datensatz zu COVID-19 (Zahlen in Tausend; Stand März 2021):

Land	USA	BRA	GB	IT	SWE	AT
Infizierte	28.831	10.517	4.171	2.908	657	455
Tote	517	254	123	98	13	8,5

a) Bestimmen Sie die Gleichung der Regressionsgeraden zur Beschreibung der Abhängigkeit der Anzahl der Toten von der Anzahl der Infizierten.

b) Schätzen Sie damit den Wert für die Anzahl der Toten bei fünf Millionen Infizierten.

Lösung: a) $y = 30,22 + 0,018x$; b) $y\,(5000) = 117,746$

1.20 Für acht Unternehmen desselben Wirtschaftszweiges soll untersucht werden, welcher Zusammenhang zwischen Umsatz und Beschäftigtenzahl besteht. Im Jahr 2018 wurden folgende Zahlen festgestellt:

Beschäftigte (in Tsd.)	0,5	0,7	1,3	1,8	1,6	1,4	1,8	2,2
Umsatz (in Mio. Euro)	50	120	260	220	180	160	210	280

a) Bestimmen Sie den (Pearson-) Korrelationskoeffizienten und interpretieren Sie ihn.

b) Bestimmen und zeichnen Sie die Regressionsgerade zur Beschreibung der Abhängigkeit des Umsatzes von der Anzahl der Beschäftigten.

Lösung: a) 0,855; b) $y = 26,849 + 111,965x$

Übungen zu Blatt 2

2.1 $\boxed{\text{V}}$ Die nach dem französischen Mathematiker Pierre-Simon de Laplace benannten Laplace-Experimente beruhen auf der Annahme, dass bei einem derartigen Zufallsexperiment nur endlich viele Ausgänge möglich sind, die alle mit gleicher Wahrscheinlichkeit auftreten. Spricht man im Falle eines Würfels von einem fairen, idealen oder Laplace-Würfel, so nimmt man an, dass bei einem Wurf die sechs Seitenflächen des Würfels – 1, 2, 3, 4, 5, 6 – jeweils mit der Wahrscheinlichkeit $\frac{1}{6}$ auftreten. Es wird nun mit einem fairen Würfel gewürfelt. Man betrachtet die folgenden Ereignisse:

A= „Zahl kleiner drei“ $\quad B$ = „gerade Zahl“
C = „Zahl größer als fünf“ $\quad D = \{2, 3, 4, 5\}$

a) Bestimmen Sie die Wahrscheinlichkeit für die Ereignisse A, B, C und D

b) Bestimmen Sie die Wahrscheinlichkeiten $P(A \cup C)$, $P(B \cap D)$

c) Zeigen Sie: die bedingte Wahrscheinlichkeit $P(B \mid A)$ ist gleich $P(B)$. Sind B und A abhängig?

Lösung:

a) $A = \{1, 2\}$, $B = \{2, 4, 6\}$, $C = \{6\}$, $D = \{2, 3, 4, 5\}$
$P(A) = \frac{1}{3}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{6}$, $P(D) = \frac{2}{3}$

b) $A \cup C = \{1, 2, 6\}$, $P(A \cup C) = \frac{1}{2}$
$B \cap D = \{2, 4\}$, $P(B \cap D) = \frac{1}{3}$

c) $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{1/6}{1/3} = \frac{1}{2} = P(B)$
nein, wegen $P(B|A) = P(B)$ gilt A und B sind unabhängig!

2.2 $\boxed{\text{V}}$ Ein zweistufiges Experiment wird wie folgt durchgeführt: Zuerst wird ein fairer Würfel einmal geworfen. Dann werden so viele Münzwürfe mit einer ebenfalls fairen Münze (Kopf oder Zahl treten mit jeweils 50 % Wahrscheinlichkeit auf) durchgeführt, wie die Augenzahl ergab, höchstens aber drei Würfe.

a) Stellen Sie dieses Experiment durch einen geeigneten Wahrscheinlichkeitsbaum dar und geben Sie für alle möglichen Ergebnisse bzw. Abläufe deren Wahrscheinlichkeiten an!

b) Mit welcher Wahrscheinlichkeit erhält man dann insgesamt genau zweimal „Kopf“?

c) Mit welcher Wahrscheinlichkeit wird eine Zahl größer als zwei geworfen, und man erhält genau zweimal „Kopf“?

d) Mit welcher Wahrscheinlichkeit erhält man, nachdem eine Zahl größer als zwei geworfen worden ist, genau zweimal „Kopf“?

e) Mit welcher Wahrscheinlichkeit wurde eine Zahl größer als zwei geworfen, wenn bekannt ist, dass genau zweimal „Kopf“ geworfen wurde.

f) Mit welcher Wahrscheinlichkeit wird die Münze genau zweimal geworfen und ergibt zweimal „Kopf“?

Lösung: a) Baum; b) 0,292; c) 0,25; d) 0,375; e) 0,857; f) 0,042

2.3 Eine Münze wird dreimal geworfen. „Kopf“ und „Zahl“ sind gleich wahrscheinlich. Allerdings bleibt die (ziemlich dicke) Münze mit einer Wahrscheinlichkeit von ein Prozent auf der Kante stehen. Mit welcher Wahrscheinlichkeit erhält man

a) genau einmal Zahl

b) kein einziges Mal Kopf?

Lösung:

a) $P(K) = P(Z),\ P(Ka) = 0,01$

$$P(K) = \frac{1 - 0,01}{2} = \frac{0,99}{2} = 0,495$$

$$P\,(\text{genau einmal Zahl}) = 0,379$$

b) $P\,(\text{kein einziges Mal Kopf}) = 0,129$

2.4 Man würfelt hintereinander mit zwei fairen regelmäßigen Würfeln. Folgende Ereignisse werden betrachtet:

A: Die Augensumme ist gerade

B: Die erste gewürfelte Zahl ist gerade

C: Die Augensumme beträgt sieben

D: Es ist keine Sechs dabei

Die folgende Tabelle zeigt alle 36 möglichen Ergebnisse des Experimentes:

$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$
$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$
$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$(3,5)$	$(3,6)$
$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$	$(4,5)$	$(4,6)$
$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$	$(5,6)$
$(6,1)$	$(6,2)$	$(6,3)$	$(6,4)$	$(6,5)$	$(6,6)$

a) Geben Sie die Wahrscheinlichkeiten für die Ereignisse A, B, C und D an!

b) Bestimmen Sie die Wahrscheinlichkeit für $P(C \cap D)$.

c) Bestimmen Sie die Wahrscheinlichkeit für $P(C \mid D)$.

d) Bestimmen Sie die Wahrscheinlichkeit für $P(D \mid C)$.

e) Ändert das Ergebnis (B oder $\overline{B}$) des ersten Wurfes die Wahrscheinlichkeit für A?

Lösung:

a) $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $C = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$, $P(C) = \frac{6}{36} = \frac{1}{6}$, $P(D) = \frac{25}{36}$

b) $C \cap D = \{(2,5),(3,4),(4,3),(5,2)\}$, $P(C \cap D) = \frac{4}{36} = \frac{1}{9}$

c) $P(C \mid D) = \frac{4}{25}$

d) $P(D \mid C) = \frac{4}{6} = \frac{2}{3}$

e) $P(A) = P(A \mid B) = P(A \mid \overline{B})$
$\frac{1}{2} = \frac{9}{18} = \frac{9}{18}$, nein.

2.5 $\boxed{\text{V}}$ Nach der letzten Wahl in einer steirischen Gemeinde wurde die Wahlbeteiligung analysiert. Dabei wurde die wahlberechtigte Bevölkerung in drei Gruppen eingeteilt:

Gruppe 1	Wahlberechtigte unter 35 Jahre	30 % der Wahlberechtigten
Gruppe 2	Wahlberechtigte von 35 bis 65 Jahre	45 % der Wahlberechtigten
Gruppe 3	Wahlberechtigte über 65 Jahre	25 % der Wahlberechtigten

Die Wahlbeteiligung betrug in Gruppe 1: 87 %, in Gruppe 2: 82 %. Insgesamt beteiligten sich 79,25 %.

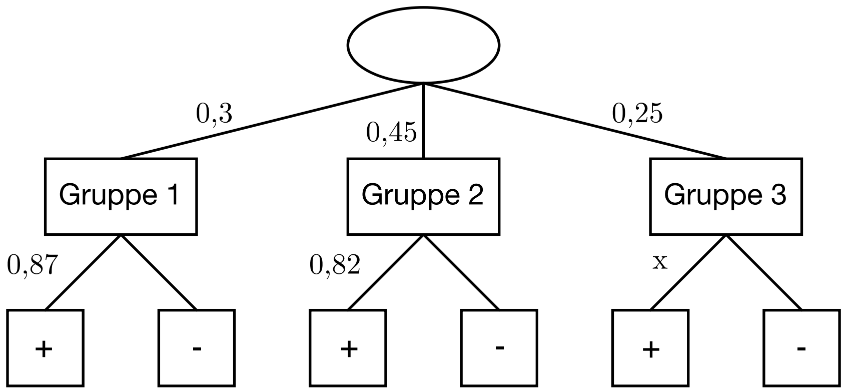
a) Wie groß war die Wahlbeteiligung in Gruppe 3?

b) Wie groß ist der Prozentsatz der Wahlberechtigten, die unter 35 Jahre alt sind und nicht zur Wahl gingen?

c) Mit welcher Wahrscheinlichkeit stammte ein abgegebener Stimmzettel von einer mindestens 35 Jahre alten Person?

Lösung:

a)



a) $P\,(+\,|\text{Gruppe3}) = 0,65$

b) $P(\text{Gruppe 1} \wedge -) = 0,3 \cdot 0,13 = 0,039 = 3,9\ \%$

c) $P(\text{Gruppe 2 oder Gruppe 3}|+) = 0,671 = 67,1\ \%$

2.6 $\boxed{\text{V}}$ Ein fairer Würfel wird einmal geworfen.

a) Bestimmen Sie Erwartungswert und Varianz für die Augenzahl.

b) Bestimmen Sie den Erwartungswert, wenn Sie wissen, dass die geworfene Augenzahl

i. gerade

ii. ungerade ist.

Lösung: a) $E\,(X) = 3,5$; $Var\,(X) = 2,917$; b) $E\,(X\,|\text{gerade}) = 4$; $E\,(X\,|\text{ungerade}) = 3$

2.7 Die Zuverlässigkeit eines bestimmten COVID-19-Schnelltests ist durch folgende Angaben gekennzeichnet: 90 % der Corona-infizierten Personen werden als solche entdeckt, 98 % der Corona-freien Personen werden als solche erkannt. Aus einer großen Bevölkerung, von der 0,5 % Corona-infiziert sind, wird nun eine zufällig herausgegriffene Person getestet.

a) Wie groß ist die Wahrscheinlichkeit, dass diese Person Corona-infiziert ist, wenn sie als Corona-unverdächtig eingestuft worden ist, also ein falsch-negatives Testergebnis hat?

b) Wie groß ist die Wahrscheinlichkeit, dass diese Person Corona-frei ist, wenn sie als Corona-verdächtig eingestuft worden ist, also ein falsch-positives Testergebnis hat?

Lösung: a) 0,00051; b) 0,816

2.8 V Unter den Personen, die Privatkonkurs anmelden, gibt es 45 % Arbeiter, 40 % Angestellte und sonst Beamte. In der Gruppe der Arbeiter beträgt die Wahrscheinlichkeit 30 %, dass es zu einem Ausfall eines Kredits kommt, der mit weniger als 10.000 € aushaftet, 45 % für Kredite ab 10.000 € und unter 25.000 € und der Rest für Kredite, die mit mindestens 25.000 € aushaften. In der Gruppe der Angestellten betragen die Wahrscheinlichkeiten entsprechend 15 %, 50 % und 35 %, in der Gruppe der Beamten 20 %, 35 % und 45 %. Es gab keine ausfallenden Kredite, die mit mehr als 100.000 € aushafteten. Die Kredite sind in jeder Klasse gleichverteilt.

a) Wie lautet die bivariate Verteilung der Berufsgruppen und der Höhe der Kreditausfälle?

b) Wie lautet die Verteilung der ausfallenden Kredite in der Gruppe der Angestellten?

c) Wie hoch ist die erwartete Höhe eines ausfallenden Kredits?

d) Wie hoch ist die erwartete Höhe eines ausfallenden Kredits in der Gruppe der Angestellten?

Lösung: a) Video; b) Video; c) 29.087,5 €; d) 31.375.- €

2.9 $\boxed{\text{V}}$ In einer Schachtel befinden sich 12 Kugeln, 4 davon sind rot, 8 sind schwarz. Es werden zufällig 3 Kugeln entnommen. Wie groß ist die Wahrscheinlichkeit, zufällig genau 0, 1, 2 oder 3 rote Kugeln zu ziehen, wenn jede gezogene Kugel

a) zurückgelegt und sofort wieder untergemischt wird?

b) nicht mehr zurückgelegt wird?

Lösung:

a) $P(X = 0) = 0.296$
   $P(X = 1) = 0.444$
   $P(X = 2) = 0.222$
   $P(X = 3) = 0.037$

b) $P(X = 0) = 0.255$
   $P(X = 1) = 0.509$
   $P(X = 2) = 0.218$
   $P(X = 3) = 0.018$

2.10 Ein Geldautomat in einem Einkaufszentrum wird an einem Samstag besonders stark frequentiert. Die Zufallsgröße X zählt die in der Warteschlange vor dem Automaten stehenden Kunden. Die Wahrscheinlichkeit, dass genau k Kunden warten, kann der folgenden Tabelle entnommen werden:

k	0	1	2	3
$P(X = k)$	0,3	0,4	0,2	0,1

a) Bestimmen Sie Erwartungswert, Varianz und Standardabweichung der Zufallsvariablen X.

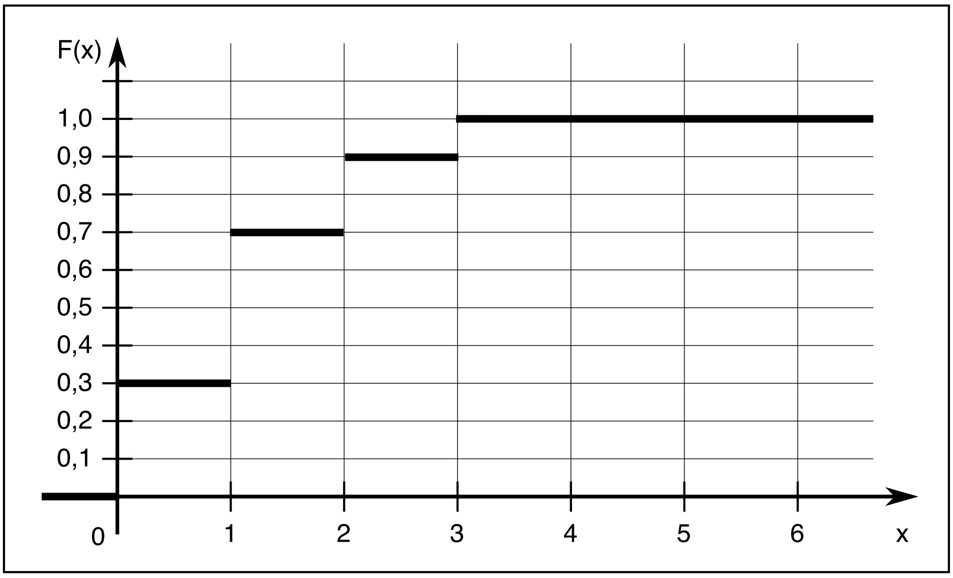
b) Skizzieren Sie den Verlauf der Verteilungsfunktion.

c) Wie groß ist die Wahrscheinlichkeit, dass weniger als zwei Kunden vor dem Geldautomaten warten?

Lösung:

a) $E(X) = 1,1$
   $\sigma(X) = \sqrt{0,89} = 0,943$

b)



c) $P(X < 2) = P(X = 0) + P(X = 1) = 0,7$

2.11 Aus 32 gut durchmischten Karten, in denen die Farben Herz, Karo, Pik und Kreuz auftreten, wird dreimal hintereinander eine Karte gezogen, wobei die gezogenen Karten nicht zurückgelegt werden. Wie groß ist die Wahrscheinlichkeit, dass

a) Dabei genau zwei Herz-Karten

b) Mindestens eine Herz-Karte gezogen wird?

Lösung: a) 0,135; b) 0,592

2.12 V X ist verteilt nach $B(20; 0,8)$. Erstellen Sie eine Tabelle zu einer Zufallsgröße X, die folgende Wahrscheinlichkeiten enthält: k, $P(X = k)$ und $F(k)$.

Berechnen Sie $P(X \in [14, 18])$.

Lösung: Excel; $P(X \in [14, 18]) = 0,844$

2.13 V In einem Casino befinden sich 20 Spielautomaten. Jeder ist auf eine Gewinnwahrscheinlichkeit von 20 % eingestellt.

a) Wie groß ist die Wahrscheinlichkeit, an einem dieser Automaten bei den nächsten fünf Spielen nie zu gewinnen?

b) Mit welcher Wahrscheinlichkeit gewinnt man an einem Automaten mehr als zwei der nächsten fünf Spiele?

c) Mit welcher Wahrscheinlichkeit gewinnt man an mindestens drei der 20 Automaten mehr als zwei der nächsten fünf Spiele, wenn an jedem genau fünf Spiele gespielt werden?

Lösung:

a) $X \sim B(5; 0.2)$, Binomialverteilung; $P(X = 0) = 0,328$; $P(X > 2) = 0,058$

b) $Y \sim B(20; 0,058)$, Binomialverteilung; $P(Y \geq 3) = 0,106$

2.14 $\boxed{\text{V}}$ Die Wahrscheinlichkeit, dass ein Röntgengerät innerhalb einer Woche ausfällt, beträgt ein Prozent. Ein Ausfall führt zu einer Unterbrechung der Untersuchungen für mehrere Tage. Die Zufallsvariable X gibt die Anzahl der Wochen an, bis es zum ersten Ausfall des Gerätes nach seiner Inbetriebnahme kommt.

a) Zeichnen Sie einen Wahrscheinlichkeitsbaum und stellen Sie die Verteilung dieser Zufallsgröße soweit wie möglich in einer Tabelle dar.

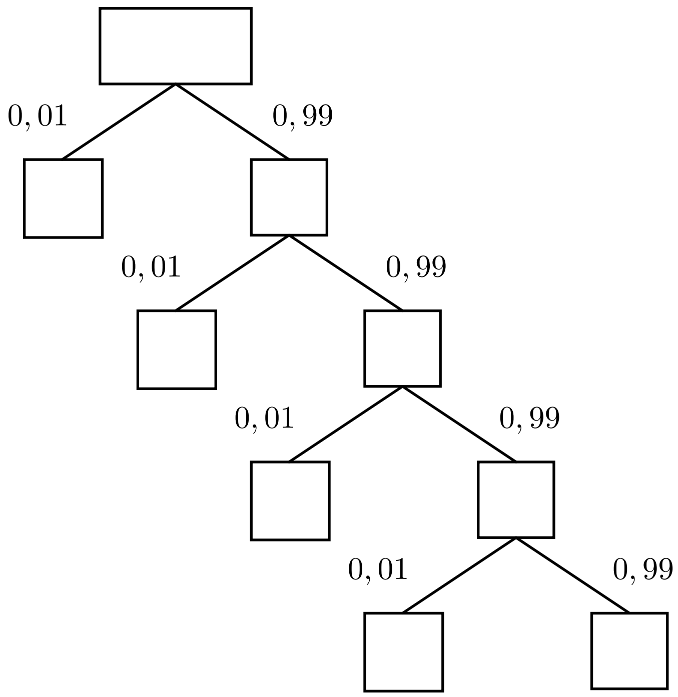
b) Wie groß ist die Wahrscheinlichkeit, dass es innerhalb von vier Wochen bereits zu einem Ausfall kommt?

c) Wie groß ist die Wahrscheinlichkeit, dass es mehr als ein Jahr lang keine Ausfälle gibt (1 Jahr = 52 Wochen)?

d) Wie viele Wochen sind bis zum ersten Ausfall bei diesem Gerät zu erwarten?

Lösung:

a) geometrische Verteilung



k	1	2	3	4	n
$P(X=k)$	0,01	$0,99 \cdot 0,01$	$0,99^2 \cdot 0,01$	$0,99^3 \cdot 0,01$	$0,99^{n-1} \cdot 0,01$

b) $P = 0,0394$

c) $0,593$

d) $E(X) = 100$

2.15 $\boxed{\text{V}}$ Eine statistische Analyse der Konsultationen, die im Rahmen der wöchentlich angebotenen Sprechstunden von Studierenden wahrgenommenen wurden, ergab, dass es ein vergleichsweise seltenes Ereignis ist, dass ein Studierender zur Sprechstunde erscheint. Die Anzahl X der Studierenden, die im Verlauf einer Sprechstunde zu einer Konsultation erscheinen, kann hinreichend genau mit Hilfe einer Poissonverteilung mit Erwartungswert 1 beschrieben werden.

a) Erstellen Sie eine Tabelle für die Wahrscheinlichkeiten von 0 bis 2 zur Sprechstunde erscheinenden Studierenden!

b) Mit welcher Wahrscheinlichkeit werden im Verlauf einer Sprechstunde zwei oder mehr Studierende erscheinen?

c) An wie vielen der 15 Sprechstunden eines Semesters ist daher (im Mittel) kein Studierender in der Sprechstunde zu erwarten?

Lösung:

a)

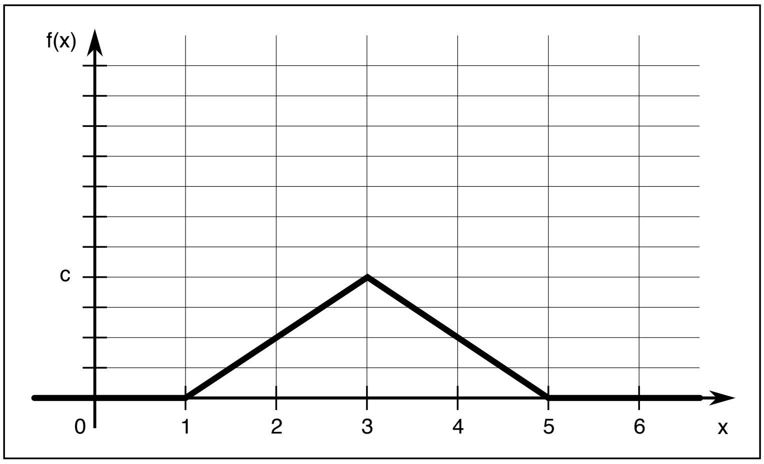
k	0	1	2
$P(X = k)$	0,368	0,368	0,184

b) $P(X \geq 2) = 0,264$

c) $Y \sim B(15; 0,368)$

$E(Y) = 5,518$

2.16 $\boxed{\text{V}}$ Der folgende Graph stellt die Dichtefunktion einer Zufallsgröße dar.



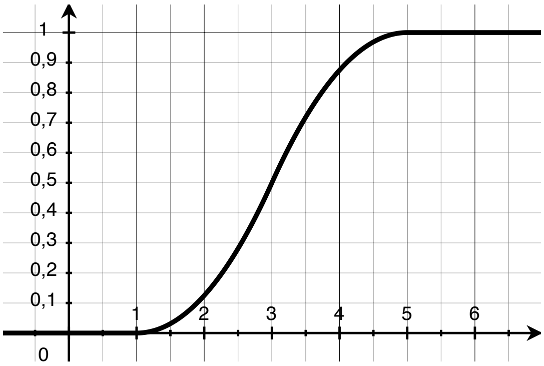
a) Ermitteln Sie die Wahrscheinlichkeiten für die Intervalle: $[1; 2,5]$, $]-\infty\,; 3]$ und $[4\,; 6]$. Hinweis: c ermitteln!

b) Skizzieren Sie den ungefähren Verlauf der Verteilungsfunktion.

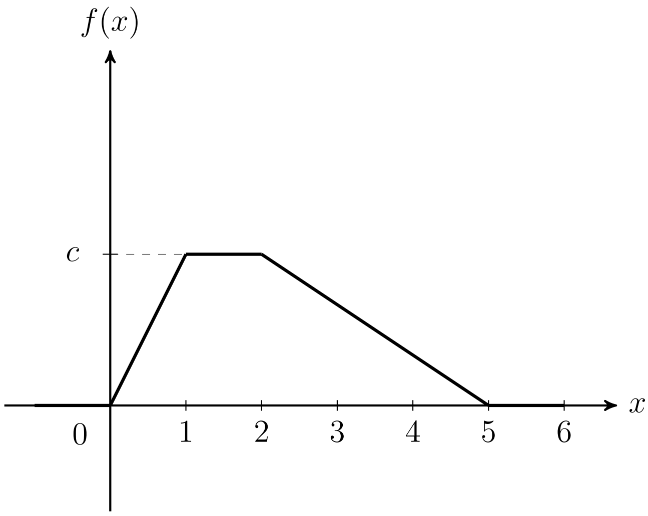
Lösung:

a) $c = \dfrac{1}{2}$; $P([1; 2,5]) = \dfrac{9}{32}$; $P(]-\infty; 3]) = \dfrac{1}{2}$; $P([4; 6]) = \dfrac{1}{8}$

b)



2.17 Der folgende Graph stellt die Dichtefunktion einer Zufallsgröße dar.



a) Ermitteln Sie die Wahrscheinlichkeiten für die Intervalle: $[1; 6]$ und $]-\infty\,; 3]$. Hinweis: c ermitteln.

b) Skizzieren Sie den ungefähren Verlauf der Verteilungsfunktion. Hinweis: Grad des Polynoms der Teilstücke überlegen.

Lösung: a) $c = \dfrac{1}{3}$; $P([1;6]) = \dfrac{5}{6}$; $P(]-\infty;3]) = \dfrac{7}{9}$; b) Graphik

2.18 Angenommen die Zeit, die ein Student benötigt, um in den Video-Chat einzusteigen, ist stetig gleichverteilt zwischen 2 und 7 Minuten.

a) Wie viel Zeit benötigt er dafür im Durchschnitt?

b) Wie stark streut die Zeit?

c) Wie groß ist die Wahrscheinlichkeit, dass der Student zwischen 3 und 5 Minuten benötigt?

d) Wie groß ist die Wahrscheinlichkeit, dass er genau 5 Minuten benötigt?

Lösung: a) 4,5; b) 1,443; c) 0,4; d) 0

2.19 Die in Minuten gemessene Wartezeit an einer Theaterkasse kann als eine exponentialverteilte Zufallsgröße aufgefasst werden. Es wird angenommen, dass die durchschnittliche Wartezeit 8 Minuten beträgt.

a) Mit welcher Wahrscheinlichkeit wartet ein Theaterbesucher länger als fünf Minuten, aber höchstens zwölf Minuten?

b) Mit welcher Wahrscheinlichkeit wartet ein Theaterbesucher genau zehn Minuten?

Lösung:

Bezeichnet man mit X die Wartezeit, so ergibt sich

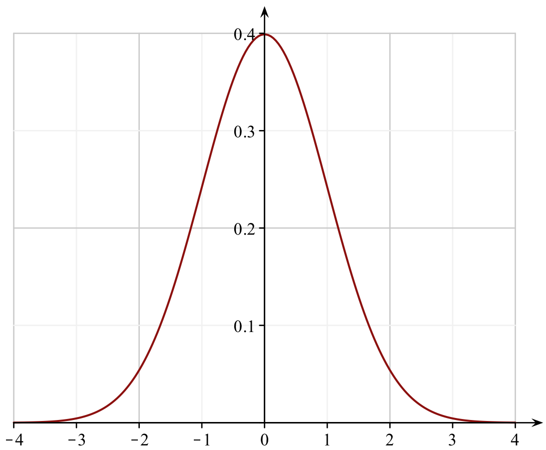
aus $E\left[X\right] = \dfrac{1}{\lambda} = 8$ für $\lambda = \dfrac{1}{8} = 0,125$;

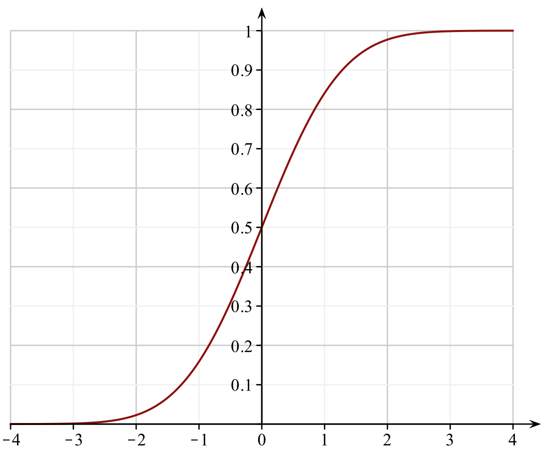
a) $P\left(5 < X \leq 12\right) = F(12) - F(5) = 0,312$

b) $P\left(X = 10\right) = 0$

ACHTUNG!! Die Exponentialverteilung ist eine stetige Verteilung, daher =EXPON.VERT(10;0,08;0) liefert nicht den Wert der Verteilungsfunktion an der Stelle 10, sondern den Wert der Dichte an dieser Stelle!!

2.20 $\boxed{\text{V}}$ Nachstehend finden Sie Dichtefunktion und Verteilungsfunktion einer standardnormalverteilten Zufallsgröße.





Kennzeichnen Sie in diesen beiden Zeichnungen und bestimmen Sie:

a) die Werte $\Phi\left(1,5\right)$, $\Phi\left(-1\right)$,

b) $P\left(Z \in \left]-\infty; 2\right]\right)$, $P\left(Z > 0,5\right)$,

c) den Wert $z_{0,8}$ und $z_{0,2}$.

Lösung:

a) $\Phi\left(1,5\right) = P\left(Z \leq 1,5\right) = 0,933$; $\Phi\left(-1\right) = P\left(Z \leq -1\right) = 0,159$

b) $P\left(X \in \left]-\infty; 2\right]\right) = \Phi\left(2\right) = P\left(Z \leq 2\right) = 0,977$; $P(Z > 0,5) = 1 - \Phi(0,5) = 0,309$

c) $z_{0,8} = 0,842$; $z_{0,2} = -0,842$

2.21 Kennzeichnen Sie in der Dichtefunktion und der Verteilungsfunktion einer standardnormalverteilten Zufallsgröße und bestimmen Sie mit Hilfe von Tabellen:

a) $P(X \in [-1,5\,;\,1,5])$

b) einen Wert c, für den gilt: $P(X < c) = 0,75$

c) den Wert $z_{0,15}$

Lösung:

a) $P(X \in [-1,5;1,5]) = 0,866$

b) $P(X < c) = 0,75$; $c = z_{0,75} = 0674$

c) $z_{0,15} = -1,036$

2.22 V Eine Zufallsgröße unterliegt einer Normalverteilung N(2, 3)

a) Bestimmen Sie dafür die Wahrscheinlichkeit des Intervalls $]-1;3]$.

b) Bestimmen Sie zu dieser Verteilung das $0,9$-Quantil und das $0,2$-Quantil.

c) Bestimmen Sie die Zahl d derart, dass $P(X > d) = 0,3$

Lösung:

a) $P(X \in [-1\,;\,3]) =$0,472

b) $x_{0,9} = 5,845$; $x_{0,2} = -0,525$

c) $d = x_{0,7} = 3,573$

2.23 Eine weitere Zufallsgröße X ist N(5, 4)-verteilt.

a) Bestimmen Sie dafür die Wahrscheinlichkeit des Intervalls $]2;6]$.

b) Mit welcher Wahrscheinlichkeit nimmt diese Zufallsgröße einen Wert unter 4 an?

c) Bestimmen Sie die Zahl c derart, dass $P(X \in [5-c\,;\,5+c]) = 0,9$.

Veranschaulichen Sie Ihre Berechnungen anhand an einer Skizze!

Lösung:

a) $P(X \in]2;6]) = 0,372$

b) $P(X < 4) = 0,401$

c) $c = 6,579$

2.24 V Man berechne die Wahrscheinlichkeit $P(X \in [10;15])$ folgender Verteilungen:

a) $H(300;126;30)$

b) $N(12;3)$

c) Exponentialverteilung mit Parameter 0,1.

Lösung: a) 0,758; b) 0,589; c) 0,145

2.25 In einer Winzerei werden Weinflaschen mit einem Sollinhalt von 0,75 l bei einer Standardabweichung von 0,02 l abgefüllt. Wie groß ist unter Annahme der Normalverteilung und Unabhängigkeit der Flascheninhalte die Wahrscheinlichkeit dafür, dass ein Kunde, der 18 Flaschen kauft, weniger als 13,55 l erhält?

Lösung: 0,722

2.26 Die Länge von Schrauben der Größe $M8 \times 40$, die in einer bestimmten Fabrik hergestellt werden, kann als normalverteilt mit dem Erwartungswert 40 mm und der Standardabweichung $0,2$ mm angesehen werden.

a) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig herausgegriffene Schraube länger als 40,4 mm ist?

b) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig herausgegriffene Schraube kürzer als 40 mm ist?

c) Welche Schraubenlänge wird von nur 5% aller Schrauben überschritten?

d) Im Rahmen der Qualitätskontrolle werden der laufenden Produktion 20 Schrauben zufällig entnommen. Wie groß ist die Wahrscheinlichkeit, dass mindestens 5 dieser Schrauben kürzer als 40 mm sind?

Lösung: a) $P(X \geq 40,4) = 0,023$; b) $P(X < 40) = 0,5$; c) 40,329 mm; d) 0,994

2.27 $\boxed{\text{V}}$ Das Gewicht von Flugzeugpassagieren setzt sich aus dem Körpergewicht und dem Gewicht des Gepäcks zusammen. Das Körpergewicht eines Passagiers besitzt einen Erwartungswert von 65 kg, das Gewicht eines Gepäckstücks einen Erwartungswert von 15 kg. Die beiden Standardabweichungen betragen 8 kg für das Körpergewicht und 3,2 kg für das Gewicht eines Gepäckstücks. Es wird die Unabhängigkeit aller Zufallsgrößen vorausgesetzt!

a) Wie ist das Gesamtgewicht (Körpergewicht + Gepäck) von 180 Passagieren eines Airbus A320-200 verteilt? Bestimmen Sie die Parameter dieser Verteilung!

b) Mit welcher Wahrscheinlichkeit liegt dieses Gesamtgewicht über 14.500 kg?

c) Mit welcher Wahrscheinlichkeit weicht dieses Gesamtgewicht um höchstens 1 % vom Erwartungswert ab?

Lösung:

a) zentraler Grenzwertsatz:

$K \sim N\left(180 \cdot 65; \sqrt{180 \cdot 8^2}\right)$, also $K \sim N\,(11700; 107,331)$,

$G \sim N\left(180 \cdot 15; \sqrt{180 \cdot 3,2^2}\right)$, also $G \sim N\,(2700; 42,933)$;

für das Gesamtgewicht S der 180 Passagiere gilt dann $S = K{+}G \sim N\left(\mu_K + \mu_G; \sqrt{\sigma_K^2 + \sigma_G^2}\right)$,

d. h. $S \sim N\left(11700 + 2700; \sqrt{107,331^2 + 42,933^2}\right)$ also $S \sim N\,(14400; 115,599)$;

b) $P\,(S \geq 14500) = 0,194$

c) $P\,(14400 - 144 \leq S \leq 14400 + 144) = P\,(14256 \leq S \leq 14544) = 0,787$

Übungen zu Blatt 3

3.1 Im Auftrag eines Einzelhandelsunternehmens soll für die durchschnittliche Abfüllmenge einer Flaschenabfüllanlage, mit der 750 ml Weinflaschen gefüllt werden, Konfidenzintervalle bestimmt werden. Die Abfüllmenge X wird dabei als normalverteilt mit einer Standardabweichung von 10 ml angesehen. Es werden zehn auf dieser Anlage abgefüllte Flaschen zufällig ausgewählt und die Füllmenge kontrolliert. Die Stichprobe lieferte die folgenden Werte (Angaben in ml):

759	762	746	741	778	737	752	753	757	750

a) Schätzen Sie die zu erwartende mittlere Füllmenge.

b) Berechnen und interpretieren Sie das (zweiseitige) 90-%-Konfidenzintervall für die mittlere Füllmenge.

c) Wie groß muss der Stichprobenumfang mindestens sein, damit die Länge des 0,9-Konfidenzintervalls höchstens 2 ml beträgt?

d) Angenommen man möchte mit nur 80 Messungen erreichen, dass das Konfidenzintervall zum Konfidenzniveau $1 - \alpha$ höchstens 2 ml breit ist. Welches Konfidenzniveau würde sich daraus ergeben? Würden Sie sich für dieses Konfidenzniveau bei einer statistischen Untersuchung entscheiden? Begründen Sie!

e) Was ändert sich in Aufgabenteil b), wenn die Standardabweichung, mit der die Maschine arbeitet, nicht bekannt ist?

Lösung:

a) $\bar{x} = 753,5$

b) $[748,299;\ 758,701]$

c) 271

d) $0,6266$, nein viel zu gering

e) $[746,766;\ 760,233]$

3.2 $\boxed{\text{V}}$ Zur Beschreibung der wirtschaftlichen und sozialen Lage von BWL-Studierenden einer Universität wurden 201 Studenten zufällig ausgewählt und befragt. Die befragten Studenten gaben ihre zeitliche Gesamtbelastung durch Studium und Erwerbstätigkeit während der Vorlesungszeit mit durchschnittlich 42,8 Stunden pro Woche an; die Standardabweichung der erhobenen Daten betrug dabei 11,35 Stunden. Die zeitliche Belastung wird als normalverteilt angenommen.

a) Bestimmen Sie das zweiseitige 95-%-Konfidenzintervall für die durchschnittliche Gesamtbelastung μ der BWL-Studierenden!

b) Bestimmen Sie das nach unten begrenzte 95-%-Konfidenzintervall für μ.

Lösung:

a) $[41,221;\ 44,379]$

b) $[41,477;\ \infty[$

3.3 $\boxed{\text{V}}$ Für die Durchführung eines Entwicklungshilfeprojekts soll in einem Entwicklungsland zunächst der Anteil der Personen ermittelt werden, die unter dem Existenzminimum leben. In einer Pilotstudie mit $n = 50$ Personen wurden 30 als "arm" (d. h. als unter dem Existenzminimum lebend) eingestuft.

a) Schätzen Sie aus obigen Angaben den Anteil der Armen in diesem Land.

b) Berechnen Sie ein näherungsweises (zweiseitiges) 90-%-Konfidenzintervall für den Anteil der armen Bevölkerung in diesem Entwicklungsland.

c) Berechnen Sie ein (zweiseitiges) 95-%-Konfidenzintervall für den Anteil der Armen, und vergleichen Sie es mit dem in b) berechneten.

d) In einer weiteren Zufallsstichprobe werden $n = 200$ Personen befragt. Bei dieser größeren Stichprobe wurden 120 Personen als unter dem Existenzminimum lebend eingestuft. Geben Sie ebenfalls ein (zweiseitiges) 95-%-Konfidenzintervall an, und vergleichen Sie es mit dem in c) berechneten. Womit lässt sich der Unterschied erklären?

e) Bestimmen Sie den notwendigen Stichprobenumfang, damit der geschätzte Anteil der Armen in der Bevölkerung mit 90 % Sicherheitswahrscheinlichkeit um weniger als 5 Prozentpunkte vom wahren Wert abweicht, wenn Sie über keine Vorabinformationen für den zu schätzenden Anteil verfügen.

Lösung:

a) $\hat{p} = 0,6$

b) $[0,486;\ 0,714]$

c) $[0,464;\ 0,736]$

d) $[0,532;\ 0,668]$

e) 271

3.4 Die öffentliche Meinung bezüglich der Maßnahmen zur Bewältigung der COVID-19-Krise in Österreich soll im Rahmen einer Umfrage erhoben werden. 380 zufällig ausgewählte Personen wurden unter anderem danach befragt, ob sie mit den gesundheitsspolitischen Entscheidungen der Bundesregierung zufrieden seien. 120 Befragte beantworteten diese Frage mit einem: „Ja."

a) Bestimmen Sie das nach oben begrenzte 99-%-Konfidenzintervall für den Anteil der Personen, die mit der Bundesregierung zufrieden sind.

b) Wie groß kann die Länge eines zweiseitigen Konfidenzintervalls bei einem Konfidenzniveau von 0,99 höchstens werden, wenn noch keine Informationen über den Stichprobenanteil vorliegen?

Lösung: a) $[0;\ 0,371]$; b) Maximale Länge: 0,132

3.5 V Die Schadenshöhe X eines Einzelschadens in einer Nicht-Lebensversicherung ist normalverteilt $N(\mu; 50)$ in Euro, Folgende Stichprobe wurde gezogen:

341	487	375	528	520	574	619	864	599

a) Testen Sie zum Niveau $\alpha = 0,1$ die Nullhypothese

$H_0: \ \mu = 520$ gegen $H_1: \ \mu \neq 520$

b) Innerhalb welcher Grenzen muss der beobachtete Mittelwert liegen, damit die Nullhypothese akzeptiert wird?

Lösung:

a) $t_0 = 1,513;\ K =]-\infty; -1,645[\cup]1,645; \infty[;\ t_0 \notin K;\ H_0$ beibehalten

b) $[492,583;\ 547,417]$

3.6 V Die Schrittlänge bei der Hosenproduktion sei in Zentimeter $N(\mu; 0,2)$-verteilt. Es wurden gleichartige Hosen vermessen.

85,1	85,4	85,3	85,0	84,9	85,2	85,4

a) Testen Sie zum Niveau $\alpha = 0,01$ die Hypothese $H_0: \mu \leq 85$ cm.

b) Wie ist zu den Niveaus $\alpha = 0,4$ (bzw. $\alpha = 0,0005$) zu entscheiden?

c) Zu welchem Signifikanzniveau gehört der realisierte Mittelwert, d. h. für welches α ist dieser die Grenze des kritischen Bereiches für den Mittelwert?

d) Der wahre Erwartungswert μ_{wahr} sei nun 85,1, die Nullhypothese ist also offensichtlich falsch. Mit welcher Wahrscheinlichkeit wird dann die H_0 abgelehnt, d. h. erkennt der Test die H_0 als falsch?

Lösung:

a) $t_0 = 2,457;\ K =]2,326; \infty[;\ t_0 \in K;\ H_0$ verwerfen

b) $t_0 = 2,457;\ K =]1,751; \infty[;\ t_0 \in K;\ H_0$ verwerfen; $K =]3,291; \infty[;\ t_0 \notin K;\ H_0$ beibehalten

c) $p - Value = 0,007$

d) 0,159

3.7 Ein Marmeladeproduzent vermutet, dass die Abfüllmaschine die Abfüllmenge von Marmelade – sie sollte 250 g betragen – bei der Abfüllung nicht präzise genug einhält. Die Füllmenge wird als normalverteilte Zufallsgröße betrachtet. Ein Angestellter prüft bei 40 Gläsern die Abfüllmenge und errechnet einen Mittelwert von $\overline{x} = 249,8g$. Die Standardabweichung ist bekannt und beträgt $\sigma = 0,7g$. Prüfen Sie, ob die Vermutung des Produzenten signifikant bestätigt werden kann! Formulieren Sie die Hypothesen und testen Sie zum Niveau $\alpha = 5\%$.

Lösung: $t_0 = -1,807$; $K =]-\infty; -1,96[\cup]1,96; \infty[$; $t_0 \notin K$; H_0 beibehalten

3.8 Ein Unternehmen hat seine Internetseite neu gestaltet und seine Produktinformationen deutlich verbessert. Um den Erfolg des neuen Webauftritts zu messen, wurde danach an 21 zufällig ausgewählten Tagen jeweils die Anzahl der Zugriffe auf die Internetseiten registriert. Aus diesen $n = 21$ Beobachtungswerten, die als normalverteilt angesehen werden können, wurde das arithmetische Mittel $\overline{x} = 5877$ und die Standardabweichung $s = 387,9$ berechnet.

Lässt sich die Behauptung der Marketingabteilung, dass nunmehr im Mittel mehr als 5700 Zugriffe pro Tag auf die Internetseiten erfolgen, signifikant bestätigen? Formulieren Sie die Hypothesen und testen Sie zum Niveau $\alpha = 5$ %!

Lösung: $t_0 = 1,207$; $K =]1,645; \infty[$; $t_0 \notin K$; H_0 beibehalten

3.9 Der Produzent einer Massenware behauptet, dass unter den von ihm produzierten Stücken höchstens 15% Ausschuss ist. Ein Kunde hält diesen Anteil für höher. In einer Stichprobe vom Umfang 20 findet er 6 Ausschussstücke vor. Kann er die Behauptung des Produzenten widerlegen? Testen sie mit $\alpha = 5$ %.

Lösung: $t_0 = 6$; $K = \{7, 8, \dots 20\}$; $t_0 \notin K$; H_0 beibehalten

3.10 $\boxed{\text{V}}$ Der Gourmethändler Exquisit möchte sichergehen, keine Trüffelkartoffel zu kaufen, bei denen der Ausschussanteil über 10 % liegt und er möchte mittels Binomialtest an Hand von 40 zufällig ausgewählten Kartoffeln seine Entscheidung über „Kaufen oder Nicht kaufen" treffen. Da er aber weiß, dass in der Statistik „nichts sicher ist", gibt er sich mit einer Sicherheit von 90 Prozent zufrieden.

a) Wie muss er bei Anwendung des Tests dazu Null- und Gegenhypothese wählen?
b) Bis zu welcher Anzahl schlechter Kartoffel ist er bereit, zu kaufen?
c) In der Stichprobe waren zwei schlechte Kartoffel. Wird der Händler kaufen?
d) Mit welcher Wahrscheinlichkeit wird ein Verkäufer, dessen Schlechtanteil seiner Kartoffel 15 % beträgt, diese dennoch an Exquisit verkaufen können?
e) Mit welcher Wahrscheinlichkeit wird ein Verkäufer, dessen Schlechtanteil seiner Kartoffel 8 % beträgt, diese dennoch an Exquisit nicht verkaufen können?

Lösung:

a) $H_1 : p < 0,1$ gegen $H_0 : p \geq 0,1$
b) B(40; 0,1)
$F(c) \leq \alpha < F(c+1)$

k	$F(k)$
0	0,0148
1	0,0805
2	0,2228

$\underline{\underline{c = 1}}$ $K = \{0, 1\}$
c) $t_0 = 2 \notin K$ H_1 nicht bestätigt, nein
d) $X \sim B(40; 0,15)$
$P(X \leq 1) = 0,012$ (1,21 %)
e) $X \sim B(40; 0,08)$
$P = 0,841$

3.11 $\boxed{\text{V}}$ Betrachten Sie das Datenblatt zum Welschlauf. Einunddreißig Läufer, die in zwei aufeinanderfolgenden Jahren am Welschlauf im Halbmarathon teilgenommen haben, wurden zufällig aus der Grundgesamtheit aller Läufer ausgewählt, und es wurden einige Merkmale erhoben. Verwenden Sie die Daten aus dem Datenblatt. Setzen Sie bei der Beantwortung der folgenden Fragen Normalverteilung voraus.

a) Bestimmen Sie ein 90-Prozent-Konfidenzintervall für den Erwartungswert des Gewichts der Läufer.

b) Bestimmen Sie ein 90-Prozent-Konfidenzintervall für die Standardabweichung des Gewichts der Läufer.

c) Lässt sich mit einer Sicherheit von 90 Prozent „bestätigen", dass der Erwartungswert dieses Gewichtes unter 70 kg liegt?

d) Lässt sich zum Signifikanzniveau $\alpha = 0,1$ zeigen, dass die Standardabweichung des Gewichtes weniger als 10 beträgt (H_1)?

e) Testen Sie zum Signifikanzniveau $\alpha = 0,1$, ob die Normalverteilungsannahme gerechtfertigt ist.

Lösung:

a) $[62,773;\ 68,389]$

b) $[7,626;\ 11,733]$

c) $t_0 = -2,671;\ K =]-\infty; -1,310[;\ t_0 \in K; H_0$ verwerfen, Ja

d) $t_0 = 25,455;\ K = [0;\ 20,599[;\ t_0 \notin K;\ H_0$ beibehalten

e) $t_0 = 4,202;\ K =]4,605; \infty[;\ t_0 \notin K;\ H_0$ beibehalten

3.12 $\boxed{\text{V}}$ In sechs bzw. sieben Gemeinden im Burgenland bzw. in der Steiermark wurden die Pro-Kopf-Ausgaben der Gemeinde für Straßenreinigung/Streudienst ermittelt. Folgende Datenzeilen liegen vor: (Angaben in €)

Burgenland (X)	213	229	291	274	237	243	
Steiermark (Y)	298	272	259	296	290	304	285

Es soll getestet werden, ob, und wenn ja, welcher Unterschied in den Erwartungswerten der zugrundeliegenden Normalverteilungen vorliegt.

a) Ist die Voraussetzung für den t-Test gegeben? D. h.: Kann man „Gleichheit der Varianzen" akzeptieren?

b) Lässt sich die Behauptung „Der Erwartungswert von X ist kleiner" signifikant nachweisen?

Beachten Sie immer die Wahl von H_0 und H_1 und testen Sie unter Annahme normalverteilter Grundgesamtheiten zum Niveau $\alpha = 0,1$.

Lösung:

a) $t_0 = 3,384$; $K = [0; 0,202[\cup]4,39; \infty[$; $t_0 \notin K$, H_0 beibehalten, ja.

b) $t_0 = 3,019$; $K =]1,363; \infty[$; $t_0 \in K$; H_1 bestätigt, ja

3.13 Untersucht wird die Gewinnzunahme bei Firmen, die auf unterschiedliche Art ihre Werbung betreiben. Eine Marketingfirma hat nach ihren Recherchen folgendes Ergebnis für die Gewinnzunahme erhalten:

Radiowerbung (X)	135	63	129	82	85	100	82	132
TV_Werbung (Y)	125	133	130	101	119	124	106	132

a) Ist die Voraussetzung für den Zweistichproben-t-Test gegeben?

b) Lässt sich mit einem geeigneten Test die Behauptung: „TV-Werbung führt zu höherem Gewinn als Radio-Werbung“, signifikant nachweisen? Beachten Sie immer die Wahl von H_0 und H_1 und testen Sie unter Annahme normalverteilter Grundgesamtheiten zum Niveau $\alpha = 0,05$.

Lösung:

a) $t_0 = 5,327$; $K = [0; 0,2[\cup]4,995; \infty[$; $t_0 \in K$; H_1 bestätigt, Nein! Daher Welch-Test

b) $t_0 = 1,906$; $K =]1,833; \infty[$, $t_0 \in K$; H_1 bestätigt, ja

3.14 $\boxed{\text{V}}$ Betrachten Sie das Datenblatt zum Welschlauf. Fassen Sie die erhobenen Gewichte der Steirer (erste Datenreihe) und der anderen Teilnehmer (zweite Datenreihe) als zwei (unabhängige) Stichproben auf und testen Sie folgende Hypothesen (Testniveau jeweils $\alpha = 0,1$) unter Annahme normalverteilter Merkmale und Varianzhomogenität:

a) $H_0 : \mu_x = \mu_y$ (d. h.: H_1 : die beiden Erwartungswerte sind verschieden)

b) $H_1 : \mu_x > \mu_y$

Lösung:

a) $t_0 = -0,121$; $K =]-\infty; -1,699[\cup]1,699; \infty[$; H_0 beibehalten

b) $t_0 = -0,121$; $K =]-\infty; -1,311[$, $t_0 \in K$; H_0 beibehalten

3.15 Ein Arzneimittelhersteller möchte die Wirksamkeit eines neuen, blutdrucksenkenden Medikaments testen. Dabei wird bei einer Gruppe von Personen der Blutdruck vor und nach dem Verabreichen des Medikaments gemessen.

Testperson	A	B	C	D	E
Vor der Einnahme	155	170	162	145	158
Nach der Einnahme	158	152	138	145	148

Lässt sich die Behauptung des Herstellers: „Nach der Einnahme des Medikaments ist der Blutdruck niedriger als vor der Einnahme“, bestätigen ($\alpha = 0,1$)? Gehen Sie von normalverteilten Werten aus. Formulieren Sie die Hypothesen und führen Sie den entsprechenden Test durch!

Lösung:

$t_0 = -1,906$; $K =]-\infty; -1,533[$, $t_0 \in K$; H_0 verwerfen, H_1 bestätigt

3.16 $\boxed{\text{V}}$ Innerhalb einer Woche wurden im LKH Feldbach sechs männliche Kinder geboren, deren Gewicht bei der Geburt und genau 12 Tage danach gemessen wurde:

Geburtsgeweicht	3245	3671	2678	2398	3723	3678
Gewicht nach 12 Tagen	3257	3670	2749	2428	3759	3711

a) Lässt sich mit einer Sicherheit von 90 % und unter Annahme normalverteilter Grundgesamtheiten zeigen, dass innerhalb der ersten zwölf Lebenstage eine Gewichtszunahme vorliegt?

b) Kann man zeigen, dass die Säuglinge nach zwölf Tagen um mindestens 20 g schwerer sind als bei der Geburt?

Lösung:

a) $t_0 = 3,0112$; $K =]1,476; \infty[$; $t_0 \in K$, H_1 bestätigt, ja.

b) $t_0 = 1,015$; $t_0 \notin K$, H_1 nicht bestätigt, nein.

3.17 $\boxed{\text{V}}$ Aufgrund einer Fusion und des daraus resultierenden Rationalisierungsdrucks muss sich der Geschäftsführer der Parkfrei GmbH, die die Parkplätze auf Falschparker kontrolliert, von einem der beiden Parkwächter A oder B trennen. Er überlegt jenem Parkwächter zu kündigen, der weniger Falschparker abstraft. Keinesfalls möchte er aber einem Parkwächter aufgrund dieses Entscheidungskriteriums kündigen, wenn ein möglicher Unterschied in den Zahlen nur zufällig entstanden ist. Für diesen Fall möchte er sich ein anderes Entscheidungskriterium überlegen.

Der Geschäftsführer analysiert die Anzahl der von A bzw. B verteilten Parkstrafen in einer bestimmten Parkzone an jeweils 10 verschiedenen Tagen des vergangenen Jahres. Die Stichprobe liefert untenstehendes Resultat.

A:	10	12	18	8	10	15	19	17	11	9
B:	6	8	11	13	11	13	7	9	12	2

Der Geschäftsführer überlegt, B zu kündigen, da dieser augenscheinlich weniger Parkstrafen verteilt hat.

a) Welcher Test ist anzuwenden? Begründen Sie ausführlich!

b) Formulieren Sie Null- und Alternativhypothese des Geschäftsführers!

c) Führen Sie einen geeigneten Test zum Niveau $\alpha = 0,05$ durch!

d) Wie lautet die Testentscheidung? Interpretieren Sie das Resultat!

e) Bestimmen Sie den p-value zu der vorliegenden Stichprobe.

Lösung:

a) Wilcoxon-Rangsummentest auf Unterschied der Lage
nicht parametrisch, 2 Stichproben, einseitig, unverbunden

b) $H_0 : F_B \geq F_A$ gegen $H_1 : F_B < F_A$ bzw. $H_0 : c \geq 0$ gegen $H_1 : c < 0$; (wobei $c = F_B - F_A$)

c) $t_0 = 1,625$; $K =]1,645; \infty[$,

d) $t_0 \notin K$, H_1 nicht bestätigt, anderes Entscheidungskriterium

e) $\Phi(1,625) = 0,9484 = \gamma = 1 - \alpha$, $\alpha^* = 0,052$

3.18 In fünf bzw. sieben Gemeinden in Vorarlberg bzw. in der Steiermark wurden die Pro-Kopf-Ausgaben der Gemeinde für Straßenreinigung/Streudienst ermittelt. Folgende Daten liegen vor: (Angaben in €):

Vorarlberg (X)	245	205	301	291	261	241	278
Steiermark (Y)	298	272	259	274	290	304	285

Lässt sich mit einem geeigneten Test die Behauptung „Die Ausgaben sind in den beiden Ländern unterschiedlich“ signifikant nachweisen? Testen Sie ohne Annahme normalverteilter Grundgesamtheiten zum Niveau $\alpha = 0,1$.

- Wilcoxon-Rangsummentest auf Unterschied der Lage;
- $H_0 : F_B = F_A$ gegen $H_1 : F_B \neq F_A$ bzw. $H_0 : c = 0$ gegen $H_1 : c \neq 0$; (wobei $c = F_B - F_A$)
- $t_0 = -1,214$; $K =]-\infty; -1,645[\cup]1,645; \infty[$; $t_0 \notin K$, H_1 nicht bestätigt

3.19 $\boxed{\text{V}}$ Im Rahmen der Weiterentwicklung von elektronischen ABS-Systemen wurde die Geschwindigkeit von PKWs sowohl an den Vorderreifen als auch an den Hinterreifen derselben acht PKWs gemessen:

Fahrzeugnummer	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Messwert hinten	101	141	109	130	152	126	102	122
Messwert vorne	103	144	110	129	150	126	104	122

a) Testen Sie zum Niveau $\alpha = 0,1$, ohne Annahme von Normalverteilungen, die Nullhypothese „Die beiden Messungen liefern dieselben Ergebnisse“.

b) Ist die am Hinterreifen gemessene Geschwindigkeit signifikant niedriger ($\alpha = 0,1$)?

Lösung:

a) $t_0 = 1,048$; $K =]-\infty; -1,645[\cup]1,645; \infty[$; $t_0 \notin K$, H_0 beibehalten

b) $t_0 = 1,048$; $K =]1,282; \infty[$; $t_0 \notin K$, H_1 nicht bestätigt, nein

3.20 $\boxed{\text{V}}$ Betrachten Sie zum Datenblatt des Welschlaufs die Laufzeiten 2009 und 2010 aller ausgewählten Läufer.

a) Führen Sie den zweiseitigen t-Test für die Differenz der beiden Erwartungswerte unter Annahme von Normalverteilungen zu α = 0,1 durch.

b) Lässt sich (signifikant zu $\alpha = 0,1$) unter Annahme von Normalverteilungen zeigen, dass die Läufer 2009 schneller sind?

c) Lässt sich ohne Annahme von Normalverteilung (nichtparametrisch, signifikant zu α = 0,1) zeigen, dass die Läufer 2010 schneller sind?

Lösung:

a) $t_0 = -1,968$; $K =]-\infty; -1,697[\cup]1,697; \infty[$; $t_0 \in K$, H_1 bestätigt, ja.

b) $t_0 = -1,968$; $K =]1,310; \infty[$; $t_0 \notin K$, H_1 nicht bestätigt, nein

c) $t_0 = -1,892$; $K =]-\infty; -1,282[$; $t_0 \in K$, H_1 bestätigt, ja.

3.21 An einem viel befahrenen Straßenstück wurde die Lärmbelastung an 8 Tagen gemessen und man erhielt folgende Daten (Werte in dB):

77	82	69	68	59	61	66	68

Nun wurden an derselben Stelle auch nach Montage einer Lärmschutzeinrichtung an 6 Tagen Messungen durchgeführt (Werte in dB):

66	57	58	68	62	59

Lässt sich aus diesen Daten mit einer Signifikanz von 5%, d. h. zum Testniveau $\alpha = 0,05$ zeigen, dass die Lärmbelastung nach Montage der Lärmschutzeinrichtung geringer ist, wenn man davon ausgeht, dass die Daten keiner Normalverteilung unterliegen?

Lösung: $t_0 = 1,936$; $K =]1,645; \infty[$; $t_0 \in K$, H_1 bestätigt, ja.

3.22 Der Blutdruck kann alternativ am Oberarm oder am Handgelenk gemessen werden. Ein Hersteller von Messgeräten behauptet, dass eine Messung am Handgelenk ebenso zuverlässig wie am Oberarm ist, beide Messarten also dieselben Ergebnisse liefern. Zur Überprüfung wurde an acht Personen der Blutdruck jeweils am Oberarm sowie am Handgelenk gemessen. Die Stichprobe lieferte folgendes Resultat:

Person	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Messwert Oberarm	119	141	109	130	134	126	131	102
Messwert Handgelenk	123	144	110	130	132	126	134	109

Testen Sie zum Niveau $\alpha = 0,1$, ohne Annahme von Normalverteilungen, die Nullhypothese „Die beiden Geräte liefern dieselben Ergebnisse“.

Lösung: $t_0 = 1,782$; $K =]-\infty; -1,645[\cup]1,645; \infty[$; $t_0 \in K$, H_0 verwerfen, H_1 bestätigt.

3.23 In der Weststeiermark wurden bei mehreren Anbietern Kürbiskerne auf Qualität geprüft. Dabei wurde jeder Probe aus den drei Anbaugebieten Stainz, Preding und Deutschlandsberg eine der drei vorgegebenen Qualitätsstufen (Sehr gut, Gut, Durchschnittlich) zugeordnet.

	Anbaugebiete		
Qualität	Stainz	Deutschlandsberg	Preding
Sehr Gut	14	14	6
Gut	10	16	10
Durchschnittlich	9	13	8

Testen Sie zum Niveau $\alpha = 0,05$, ob Qualität und Herkunft als voneinander unabhängig betrachtet werden können.

Lösung: $t_0 = 1,987$; $K =]9,488;\infty[$; $t_0 \notin K$ H_0 beibehalten

3.24 Von einem Würfel wird behauptet, dass es sich um einen fairen Würfel handelt, d.h. jede Augenzahl tritt mit der gleichen Wahrscheinlichkeit auf. Um diese Behauptung zu überprüfen, wird der Würfel 60 mal geworfen. Man erhielt dabei die Augenzahlen 1 bis 6 mit den angegebenen Häufigkeiten:

1	**2**	**3**	**4**	**5**	**6**
15	11	7	6	14	7

Testen Sie mit Hilfe eines geeigneten Verfahrens auf einem Signifikanzniveau von $\alpha = 0,05$, ob dieses Ergebnis gegen die Behauptung spricht, dass der Würfel fair ist.

Lösung: $t_0 = 7,600$; $K =]11,07;\infty[$, $t_0 \notin K$ H_0 beibehalten

3.25 Man prüfe, ob die Zufallsgröße X = „Anzahl der täglichen Übertragungsfehler in einem firmeninternen Kommunikationsnetz“ einer Poissonverteilung mit Erwartungswert 2 unterliegen kann! (D. h. man erwartet im Durchschnitt zwei Fehler pro Tag.) An den 254 Arbeitstagen des Jahres 2014 wurden dabei folgende Anzahlen von Fehlern festgestellt:

Fehleranzahl x	**0**	**1**	**2**	**3**	**vier oder mehr**
Tage mit x Fehlern	26	61	78	34	55

a) Welcher Test ist durchzuführen? Formulieren Sie Null – und Gegenhypothese.

b) Testen Sie zum Niveau $\alpha = 0,1$ und interpretieren Sie Ihr Ergebnis genau!

Lösung:

a) Chiquadrat-Anpassungstest

H_0 :

k	0	1	2	3	≥ 4
p_k	0,1353	0,2707	0,2707	0,1804	0,1429

H_1: X unterliegt keiner Poisson-Verteilung mit $\lambda = 2$

b) $n = 254$; $\alpha = 0,1$

n_k	26	61	78	34	55
$n \cdot p_k$	34,38	68,75	68,75	45,83	36,3

$n \cdot p_k > 5 \quad \forall k$

$t_0 = 16,8594$; $K =]7,779; \infty[$; $t_0 \in K$ H_0 verwerfen, X unterliegt keiner Poisson-Verteilung mit $\lambda = 2$.

Übungen zu Blatt 4

4.0 Ein Bankangestellter möchte überprüfen, ob es einen Zusammenhang zwischen Jahreseinkommen und Sparleistung seiner Kunden gibt. Dazu liegen folgende Daten vor:

Kunde	**1**	**2**	**3**	**4**	**5**	**6**
Einkommen [in 1.000 €/Jahr]	12	32	20	15	26	36
Sparleistung [in 1.000 €/Jahr]	2	7	4	3	5	9

a) Berechnen Sie den Pearson-Korrelationskoeffizienten.

b) Bestimmen Sie die Koeffizienten der Regressionsgeraden zur Beschreibung der Abhängigkeit der Sparleistung vom Jahreseinkommen.

c) Berechnen Sie die Residuen.

d) Berechnen Sie die Standardabweichung des Störterms.

e) Ist das Merkmal „Jahreseinkommen“ signifikant? ($\alpha = 0,01$)

f) Können Sie beweisen, dass die Regressionsgerade nicht durch den Ursprung geht? ($\alpha = 0,1$)

Lösung:

a) $\rho = 0,985;$

b) $\hat{y} = -1,3499 + 0,2702 \cdot x$

c) Video

d) $\hat{\sigma} = 0,509$

e) $t_0 = 11,291;\ K =]-\infty; -4,604[\cup]4,604; \infty[;\ t_0 \in K,\ H_1$ bestätigt, ja.

f) $t_0 = -2,252;\ K =]-\infty; -2,132[\cup]2,132; \infty[;\ t_0 \in K,\ H_1$ bestätigt, ja.

4.1 $\boxed{\text{V}}$ Zahlreiche deutsche Städte erstellen sogenannte Mietspiegel, um Mietern, Vermietern, Mietberatungsstellen und Sachverständigen eine „objektive“ Entscheidungshilfe in Mietfragen zur Verfügung zu stellen. Bei der Erstellung von Mietspiegeln wird aus der Gesamtheit aller in Frage kommenden Wohnungen eine repräsentative Zufallsstichprobe gezogen. Der vorliegende Datensatz ist ein Ausschnitt aus dem Mietspiegel München des Jahres 2003 und enthält die Daten von 68 Wohnungen.

Die Merkmale sind: Wohnung ID, monatliche Nettomiete in Euro, Wohnfläche in m^2, Anzahl Zimmer, Baujahr, wobei unterjähriger Bau als Nachkommastelle angezeigt wird, Stadtbezirk und Gute Lage (ja = 1, nein = 0). Die Daten finden Sie als Excel-Datei auf der Webseite des Instituts unter dem Namen „Datensaetze.xlsx“ im Tabellenblatt „Mietspiegel“.

a) Bestimmen Sie die Koeffizienten der Regressionsgeraden zur Beschreibung der Abhängigkeit der Nettomiete von der Wohnfläche. Wie viel kostet dann durchschnittlich eine Wohnung mehr, die um einen Quadratmeter größer ist?

b) Ist das Merkmal Wohnfläche signifikant?

c) Bestimmen Sie die Koeffizienten der Regressionsgeraden zur Beschreibung der Abhängigkeit der Nettomiete von der Anzahl der Zimmer. Wie viel kostet dann durchschnittlich eine Wohnung mehr, die um ein Zimmer größer ist?

d) Ist das Merkmal Anzahl Zimmer signifikant?

e) Bestimmen Sie nun die Koeffizienten der Regressionsgeraden zur Beschreibung der Abhängigkeit der Nettomiete von der Wohnfläche und der Anzahl der Zimmer. Wie viel kostet dann durchschnittlich eine Wohnung mehr, die bei gleicher Anzahl der Zimmer um einen Quadratmeter größer ist bzw. die bei gleicher Anzahl an Quadratmetern um ein Zimmer größer ist?

f) Sind die Merkmale signifikant? Welches der drei Modelle erklärt am besten?

g) Bestimmen Sie nun die Koeffizienten der Regressionsgeraden zur Beschreibung der Abhängigkeit der Nettomiete von der Wohnfläche, dem Baujahr und der Guten Lage. Wie viel kostet dann durchschnittlich eine Wohnung mehr, die um einen Quadratmeter größer ist oder eine gute Lage hat und die jeweiligen anderen Variablen unverändert bleiben?

Lösung:

a) $\hat{y} = -213,325 + 12,887 \cdot x$; Um 12,887 €

b) p-Value = 1,22361E-16; ja

c) $\hat{y} = 162,234 + 221,630 \cdot x$; Um 221,630 €

d) p-Vaue = 4,48889E-06; ja

e) Nettomiete $= -202,589 + 13,253 \cdot$ Wohnflaeche $- 14,358 \cdot$ AnzahlZimmer

f) Anzahl Zimmer ist nicht signifikant. Das erste Modell hat das größte angepasste R^2

g) Nettomiete $= -2233,114 + 12,582 \cdot$ Wohnflaeche $+ 1,018 \cdot$ Baujahr $+ 55,077 \cdot$ Gute Lage

4.2 $\boxed{\text{V}}$ In vielen europäischen Ländern bietet die Eurotax-Liste eine Orientierungshilfe beim Kauf und Verkauf von Gebrauchtwagen. Sie enthält eine Auflistung des von KFZ-Händlern durchschnittlich erzielten Preises für ein bestimmtes Automodell mit einigen zusätzlichen Informationen.

Für 601 Autos der beliebten österreichischen Marke „Stonk Modell E“ liegen die Transaktionsnummer (ID), Daten über die Anzahl an gefahrenen Kilometern (KILOMETER), das Alter (ALTER) in Jahren seit der Erstzulassung, die Motorleistung (LEISTUNG) in KW, die Antriebsart der Verbrennungsmotoren ANTRIEB („Benzin“, „Diesel“) und den Verkaufspreis (PREIS) in Euro vor. Die Daten finden Sie im Tabellenblatt „Gebrauchtwagen“.

Führen Sie eine Regressionsanalyse zur Erklärung des Preises durch!

a) Welche der Merkmale KILOMETER, ALTER, LEISTUNG und ANTRIEB sind signifikant ($\alpha = 0.01$)?
b) Wie sind die einzelnen Parameter zu interpretieren?
c) Um wie viel kostet ein Diesel-Motor mehr als ein Benziner?
d) Ein um wieviel höherer/niedrigerer Preis ist im Mittel zu erwarten, wenn statt eines Alters von 3 Jahren ein Alter von 6 Jahren betrachtet wird und alle anderen Variablen gleich bleiben?
e) Wie groß ist das Bestimmtheitsmaß und wie ist es zu interpretieren?
f) Wie groß ist die Standardabweichung des Störterms? Wie groß ist die Standardabweichung des Merkmals ALTER?
g) (Führen Sie einen simultanen F-Test $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ durch! Wie ist das Ergebnis zu interpretieren?)
h) Prognostizieren Sie den Preis für ein Auto, das 4 Jahre alt ist, 85.000 Kilometer gefahren ist, 60 KW hat und mit Benzin betrieben wird.

Lösung:

a) Alle bis auf „Diesel“
b) Skriptum
c) Um 591,963 €
d) Um 6439,080 niedrigerer Preis
e) 0,851
f) 3419,213; 110,075
g) (Mindestens ein Koeffizient ist von Null verschieden)
h) 20.109,97 €

4.3 $\boxed{\text{V}}$ Zur Erklärung der Diamantenpreise wurde eine Erhebung in Singapur durchgeführt. Dabei wurden je Diamant folgende Merkmale erfasst: Diamant ID, Gewicht in Karat, Farbe (Klassen D, E, F, G, H, I), Klarheit (Klassen IF, VS1, VS2, VVS1, VVS2) und der Preis (in Singapur-Dollar). Die Daten sind der Datei „Diamant.xlsx“ zu entnehmen. Führen Sie eine Regressionsanalyse für die abhängige Variable Preis durch!

a) Wie lauten die Koeffizienten und wie sind die Ergebnisse zu interpretieren?

b) Berechnen Sie die Residuen und stellen Sie diese graphisch dar! Sind Modellannahmen verletzt?

Lösung: Excel

4.4 Bei einunddreißig Läufern, die in zwei aufeinanderfolgenden Jahren am Welschlauf im Halbmarathon teilgenommen haben, wurden einige Merkmale erhoben: Alter, Herkunft, Beruf, Gewicht, Laufzeit 2009, Laufzeit 2010, Einkommen, Geschlecht. Das Merkmal „Herkunft“ besitzt drei Ausprägungen: Steiermark (ST), restliches Österreich (AUT), Ausland (GAST); das Merkmal „Beruf“ ist in drei Kategorien geteilt: Angestellt (A), Selbständig (S), Auszubildender (AZUBI); das Merkmal „Einkommen“ ist ebenfalls in drei Kategorien unterteilt: gering (g), mittel (m), hoch (h); das Merkmal „Geschlecht“ besitzt die Ausprägungen weiblich (W) und männlich (M). Die Daten zu dieser Aufgabe finden Sie als EXCEL-Tabellenblatt unter dem Namen „Datenblatt_erweitert“. Es soll nun untersucht werden, ob und wie sich die Variable „Laufzeit 2010“ durch die Variablen „Gewicht“ und „Alter“ erklären lässt. Führen Sie eine Regressionsanalyse durch!

a) Welche der Merkmale „Gewicht“ und „Alter“ sind signifikant ($\alpha = 0,05$)?

b) Wie sind die einzelnen Parameter zu interpretieren?

c) Um wie viel mehr/weniger Minuten sind im Mittel zu erwarten, wenn statt eines Alters von 50 Jahren ein Alter von 60 Jahren betrachtet wird und alle anderen Variablen gleich bleiben?

d) Wie groß ist das Bestimmtheitsmaß und wie ist es zu interpretieren?

e) Wie groß ist die Standardabweichung des Störterms? Wie groß ist die Standardabweichung des Merkmals „Alter“?

f) Nehmen Sie nun zusätzlich das Geschlecht als erklärendes Merkmal in das Modell auf. Um wie viel schneller/langsamer sind Frauen als Männer?

Lösung:

a) Gewicht

b) Skriptum

c) Um 2,323 Minuten mehr

d) 0,491

e) 13,336

f) Um 3,899 Minuten