Übung 1: Wiederholung Statistik I

Aufgabe 1

Gegeben sei eine diskrete Zufallsvariable X. Für die möglichen Realisationen x_i von X seien folgende Wahrscheinlichkeiten bekannt:

x_i	0	1	2	3
$f(x_i)$	0,1	$0{,}5 - c$	$0{,}5 - c$	c

a) Bestimmen Sie die Konstante c.

b) Bestimmen Sie die zugehörige Wahrscheinlichkeits- und die zugehörige Verteilungsfunktion und skizzieren Sie diese.

c) Bestimmen Sie den Erwartungswert und die Varianz von X.

d) Bestimmen Sie die Wahrscheinlichkeiten: $P(X \leq 2{,}5)$, $P(1{,}5 < X \leq 3)$, $P(X > 1)$

Aufgabe 2

Gegeben sei die Funktion $f(x)$:

$$f(x) := \begin{cases} cx - 0{,}5 & \text{für } 1 \leq x \leq 3 \\ 0 & \text{sonst} \end{cases}$$

a) Bestimmen Sie die Konstante c so, dass $f(x)$ eine Dichtefunktion der Zufallsvariablen X ist.

b) Bestimmen Sie den Erwartungswert und die Varianz von X.

c) Bestimmen Sie folgende Wahrscheinlichkeiten:

   c1) $P(X = 2)$

   c2) $P(X > 2{,}5)$

   c3) $F_X(2{,}5)$

d) Bestimmen Sie ein Intervall, so dass sich links 6,25% und rechts 19% Wahrscheinlichkeitsmasse befinden.

Aufgabe 3

Ein Student rechnet sich seine Chancen auf seine spätere Wunschstelle aus. Zu Beginn des Studiums glaubt er, dass er das Studium zu 70% erfolgreich beenden wird. Die Wahrscheinlichkeit, die gewünschte Stelle zu erhalten, beträgt 0,8, wenn er sein Studium erfolgreich beendet, und 0,1, wenn er es nicht erfolgreich beendet.

Sind die Ereignisse: „Der Student schließt sein Studium erfolgreich ab“ und „Der Student erhält seine Wunschstelle“ stochastisch unabhängig voneinander? Begründen Sie kurz Ihre Antwort.

Übung 2: Mehrdimensionale Zufallsvariablen I

Aufgabe 1

Gegeben sei für einen diskreten Zufallsvektor $\mathbf{X} = (X_1, X_2)$ folgende Wahrscheinlichkeitstabelle:

X_1 \ X_2	1	3	10
2	0,05	0,03	0,02
4	0,20	0,10	0,05
6	0,20	0,25	0,10

a) Prüfen Sie, ob die Zufallsvariablen X_1 und X_2 stochastisch unabhängig voneinander sind.

b) Bestimmen Sie die folgende Wahrscheinlichkeiten:

b1) $f_{X_1}(4),\ F_{X_1}(4),\ F_{X_2}(8)$

b2) $F_{\mathbf{X}}(5,5),\ F_{\mathbf{X}}(1,10),\ F_{\mathbf{X}}(5,0)$

b3) $P(X_1 \leq 5 \vee X_2 \leq 5)$

b4) $P(X_1 \leq 4 | X_2 = 3),\ P(X_2 = 3 | X_1 = 4)$

c) Geben Sie die Randwahrscheinlichkeitsfunktion sowie die Randverteilungsfunktion von X_2 an.

d) Bestimmen Sie den bedingten Erwartungswert von X_1, gegeben $X_2 = 10$.

Aufgabe 2

Für einen Zufallsvektor $\mathbf{X} = (X_1, X_2)$ sei folgende Dichtefunktion gegeben:

$$f_{\mathbf{X}}(x_1, x_2) := \begin{cases} 0{,}25 & \text{für } 0 \leq x_1 \leq 4 \wedge 0 \leq x_2 \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Bestimmen Sie die Verteilungsfunktion von $\mathbf{X}$ sowie die Wahrscheinlichkeit $P(2 \leq X_1 \leq 3; 0{,}5 \leq X_2 \leq 1)$.

Aufgabe 3

Für eine stetige zweidimensionale Zufallsvariable $\mathbf{X} = (X_1, X_2)$ sei folgende Verteilungsfunktion gegeben:

$$F_{\mathbf{X}}(x_1, x_2) := \begin{cases} (1 - e^{-x_1}) \cdot (1 - e^{-2x_2}) & \text{für } x_1 \geq 0 \wedge x_2 \geq 0 \\ 0 & \text{sonst} \end{cases}$$

a) Bestimmen Sie die Dichtefunktion von $\mathbf{X}$.

b) Bestimmen Sie die Randdichtefunktionen für X_1 bzw. X_2.

c) Sind die Zufallsvariablen X_1 und X_2 stochastisch unabhängig voneinander? Begründen Sie Ihre Antwort.

Übung 3: Mehrdimensionale Zufallsvariablen II

Aufgabe 1

Gegeben sei ein zweidimensionaler Zufallsvektor $\mathbf{X} = (X_1, X_2)$ mit nachfolgender Wahrscheinlichkeitstabelle:

X_1 \ X_2	1	2	5
2	0,1	0,1	0,0
4	0,1	0,3	0,1
8	0,0	0,1	0,2

a) Bestimmen Sie die Varianz-Kovarianz-Matrix von $\mathbf{X}$.

b) Bestimmen Sie den Korrelationskoeffizienten von $\mathbf{X}$.

Aufgabe 2

Bei einem Spiel wird mit einem fairen Würfel zwei Mal gewürfelt. Sie erhalten als Augenzahlen X_1 und X_2.

a) Bestimmen sie die Kovarianz und den Korrelationskoeffizienten von X_1 und der Summe $X_1 + X_2$.

b) Zeigen Sie, dass $X_1 + X_2$ und $X_1 - X_2$ nicht stochastisch unabhängig sind.

c) Bestimmen Sie die Kovarianz $\mathrm{Cov}(X_1 + X_2; X_1 - X_2)$.

Aufgabe 3

Gegeben sei folgende Kontingenztabelle:

X \ Y	-2	0	3	Σ
-1		0,05		0,3
1	c			
Σ	0,15	0,2		1

a) Vervollständigen Sie die Tabelle und geben Sie den Wertebereich von c an.

b) Bestimmen Sie c so, dass Unkorreliertheit zwischen X und Y besteht.

c) Liegt für Ihr gefundenes c stochastische Unabhängigkeit vor? Begründen Sie Ihre Antwort.

Übung 4: Approximationsmöglichkeiten diskreter Verteilungen

Aufgabe 1

Für einen interessierenden Parameter θ sind zwei Schätzer $\hat{\theta}_1$ und $\hat{\theta}_2$ mit folgenden Eigenschaften gegeben:

$$\mathrm{E}(\hat{\theta}_1) = \theta - 3 \qquad \mathrm{E}(\hat{\theta}_2) = \theta - 2 \qquad \mathrm{Var}(\hat{\theta}_1) = 4 \qquad \mathrm{Var}(\hat{\theta}_2) = 9$$

$\hat{\theta}_1$ hat demnach im Vergleich zu $\hat{\theta}_2$ zwar die kleinere Varianz, allerdings den größeren Bias. Welchen der beiden Schätzer würden Sie bevorzugen? Begründen Sie kurz Ihre Antwort.

Aufgabe 2

Eine Klausur besteht aus 48 Multiple-Choice-Aufgaben. Für jede Aufgabe sind 4 Antwortmöglichkeiten vorgegeben, von denen jeweils genau eine richtig ist. Die Klausur wird bestanden, wenn mindestens 19 Aufgaben korrekt angekreuzt wurden.

Bestimmen Sie die Wahrscheinlichkeit, dass die Klausur durch zufälliges Ankreuzen bestanden wird. Geben Sie hierfür zunächst die zu Grunde liegende Verteilung an und approximieren Sie diese anschließend geeignet.

Aufgabe 3

Von 1000 Studierenden eines Fachbereiches bestreiten 300 ihr Studium mit eigenen finanziellen Mitteln.

Bestimmen Sie die Wahrscheinlichkeit, dass unter 50 ausgewählten Studenten 16 bis 35 ihr Studium selbst finanzieren. Geben Sie hierfür zunächst die zu Grunde liegende Verteilung an und approximieren Sie diese anschließend geeignet.

Aufgabe 4

Eine Spedition verspricht, Aufträge, die zwischen 7.00-8.00 Uhr eingehen, noch am selben Tag zu erledigen. Während dieses Zeitraums treffen durchschnittlich 36 Aufträge ein. Die Spedition kann insgesamt 45 derartige Aufträge selbst durchführen. Weitere Aufträge werden fremd vergeben.

Bestimmen Sie die Wahrscheinlichkeit, dass die Spedition alle Aufträge eines beliebigen Tages, die zwischen 7.00-8.00 Uhr eingehen, selbst durchführt. Geben Sie hierfür zunächst die zu Grunde liegende Verteilung an und approximieren Sie diese anschließend geeignet.

Übung 5: Punktschätzer

Aufgabe 1

Für eine stetige Zufallsvariable X sei folgende Dichtefunktion gegeben:

$$f_X(x) := \begin{cases} \frac{1}{a^3} \cdot x^2 + \frac{2}{3a} & \text{für } x \in [0, a] \\ 0 & \text{sonst} \end{cases}$$

Es ergab sich folgende Stichprobe:

i	1	2	3	4	5	6	7	8	9	10
x_i	0	3	4	4	6	9	9	9	10	16

Schätzen Sie den Parameter a mittels Momentenmethode.

Aufgabe 2

Für einen Versicherungsbestand liegen für die Schadenanzahl pro Risiko und Jahr folgende Daten aus der Vergangenheit vor:

Schadenanzahl	0	1	2	3	4	5	6	7
Vorkommen	3	12	23	26	20	11	4	1

Unterstellen Sie, dass die zu Grunde liegende Zufallsvariable einer Binomialverteilung folgt. Bestimmen Sie hierfür die Parameter n und p.

Aufgabe 3

Die stetige Zufallsvariable X habe nachfolgende Dichtefunktion:

$$f_X(x) := \begin{cases} \frac{2}{a} - \frac{2}{a^2} \cdot x & \text{für } x \in [0, a] \\ 0 & \text{sonst} \end{cases}$$

Schätzen Sie den Parameter $a > 0$ mittels Maximum-Likelihood-Methode aus einer Stichprobe mit lediglich $n = 2$ Beobachtungen. Diese Beobachtungen sind:

i	1	2
x_i	0	$\frac{1}{2}$

Aufgabe 4

Ein Polizist überprüft einen Tag lang die Verkehrstauglichkeit von Fahrrädern. Er kontrolliert immer so lange, bis er ein untaugliches Rad findet und wechselt dann seine Position. Die folgenden Daten geben an, wie viele Räder er bei 10 solcher Überprüfungen kontrolliert hatte, bis er auf ein untaugliches Rad stieß.

42 50 40 64 30 36 68 42 46 48

Treffen Sie eine geeignete Verteilungsannahme und bestimmen Sie mittels ML-Methode alle benötigten Parameter.

Übung 6: Konfidenzintervalle

Aufgabe 1

Bei einer Untersuchung der Belastbarkeit $X \sim \mathcal{N}(\mu;\sigma^2)$ von Spielwaren ergab sich bei einer unabhängigen Stichprobe vom Umfang 10:

$$\bar{x} = 34{,}55 \qquad s^2 = 14{,}0539$$

Bestimmen Sie ein nach oben offenes 95%-Konfidenzintervall für die durchschnittliche Belastbarkeit.

Aufgabe 2

Die Zufallsvariable $X :=$ „Effektiv-Füllung von 5l-Ketchup-Eimern" eines Abfüllers sei approximativ normalverteilt. Ein Abnehmer erhält eine zufällig zusammengestellte Lieferung von 20 Eimern, die eine durchschnittliche Füllung von $\bar{x} = 4{,}92$ Litern aufweisen. Unterstellen Sie eine Standardabweichung von $\sigma = 0{,}16$.

a) Bestimmen Sie ein zentrales 90%-Konfidenzintervall für die mittlere Füllung.

b) Welcher Stichprobenumfang ist mindestens erforderlich, damit ein zentrales 90%-Konfidenzintervall für μ höchstens halb so breit wie das unter a) bestimmte Intervall ist?

Aufgabe 3

Für eine bestimmte Glühbirnenart kann die Lebensdauer im Dauerbetrieb aufgefasst werden als eine (näherungsweise) normalverteilte Zufallsvariable X (gemessen in Tagen).

Eine unabhängige Zufallsstichprobe von 20 Birnen ergab:

$$\bar{x} = 132 \text{ Tage} \qquad \sum_{i=1}^{20} x_i^2 = 349020 \text{ Tage}^2$$

a) Bestimmen Sie erwartungstreue Schätzer für den Erwartungswert und die Varianz von X.

b) Bestimmen Sie ein 90%-Konfidenzintervall mit kleinstmöglicher Untergrenze für die Standardabweichung von X.

c) Geben Sie die Wahrscheinlichkeit an, dass die „wahre" Standardabweichung im berechneten Konfidenzintervall liegt. Wie ändert sich Ihre Aussage, wenn Sie ein 90%-Konfidenzintervall bestimmen würden, dass nicht die kleinstmögliche Untergrenze besitzt?

Aufgabe 4

Für die Durchführung eines Entwicklungshilfeprojekts soll in einem Entwicklungsland zunächst der Anteil der Personen ermittelt werden, die unter dem Existenzminimum leben. In einer Studie mit 50 Personen wurden 30 als arm eingestuft.

a) Bestimmen Sie ein geeignetes 90%-Konfidenzintervall für den Anteil der armen Bevölkerung in diesem Entwicklungsland.

b) Bestimmen Sie ein geeignetes 95%-Konfidenzintervall für den Anteil der armen Bevölkerung in diesem Entwicklungsland und vergleichen Sie Ihr Ergebnis mit Aufgabenteil a).

c) In einer weiteren Zufallsstichprobe werden 200 Personen befragt. Auch bei dieser größeren Stichprobe ergab sich ein Anteil von 0,6 an Personen, die als „arm" gelten. Geben Sie ebenfalls ein 95%-Konfidenzintervall an und vergleichen Sie es mit Aufgabenteil b).

Übung 7: Hypothesentests I

Aufgabe 1

Eine Firma verschickt Tee in Holzkisten mit jeweils 10 Teepackungen. Das Gewicht in kg einer Teepackung sei $X \sim \mathcal{N}(6; 0{,}0036)$. Einer großen Lieferung werden 16 Teepackungen entnommen. Dabei ergab sich ein durchschnittliches Gewicht von 5,95 kg.

a) Bestimmen Sie das 95%-Konfidenzintervall für das durchschnittliche Gewicht einer einzelnen Teepackung.

b) Sie vermuten, dass im Mittel nicht der Sollwert abgefüllt wird. Treffen Sie eine Testentscheidung auf dem 5%-Signifikanzniveau und führen Sie gegebenenfalls einen geeigneten Test durch.

Aufgabe 2

Es besteht die Vermutung, dass die handelsüblichen Sahnebecher einer bestimmten Marke, die mit dem Aufdruck 200ml verkauft werden, im Mittel mit weniger Inhalt gefüllt sind.

Um dies zu untersuchen, wird eine Zufallsstichprobe von 120 Bechern gezogen. Diese ergibt für die Füllmenge den Mittelwert $\bar{x} = 194$ml und die Standardabweichung $s = 16$ml.

a) Schätzen Sie den Erwartungswert und die Varianz erwartungstreu.

b) Die durch den Aufdruck gemachte Behauptung soll überprüft werden. Sie möchten dabei eine zu geringe Füllmenge nachweisen. Führen Sie hierfür einen geeigneten Test auf dem 1%-Signifikanzniveau durch.

Aufgabe 3

Der Chef eines Callcenters möchte die Anzahl der Arbeiter optimieren, da er in Zukunft mit weniger Anrufern rechnen muss. Sollte die Anzahl X der Anrufe im Mittel weniger als 10 pro Stunde betragen, müssen leider Mitarbeiter entlassen werden. In einer Stichprobe von 160 zufällig ausgewählten Beobachtungsstunden zählt der Chef insgesamt 1200 eingehende Anrufe bei einer Stichprobenvarianz von $s^2 = 10$.

Sollte der Chef Mitarbeiter entlassen? Begründen Sie Ihre Antwort, indem Sie einen geeigneten Test auf dem 5%-Signifikanzniveau durchführen.

Aufgabe 4

Ein Unternehmen füllt seine Produkte in 0,5l-Flaschen ab. Die Abfüllmaschine arbeitet jedoch nicht exakt. Die Füllmenge X in Litern kann dabei als normalverteilt angesehen werden. Zusätzlich sei aus der Vergangenheit die Standardabweichung $\sigma = 0{,}1$ bekannt. Eine unabhängige Stichprobe vom Umfang 81 ergab:

$$\sum_{i=1}^{81} x_i = 41{,}3$$

Nehmen Sie an, ein Mitarbeiter ist beauftragt worden zu überprüfen, ob die Füllmenge von der Vorgabe abweicht. Allerdings hat sich der Chef noch nicht für ein Signifikanzniveau entschieden. Welche Größe könnte der Mitarbeiter trotzdem bestimmen und seinem Chef übermitteln? Bestimmen Sie diese Größe.

Übung 8: Hypothesentests II

Aufgabe 1

Ein Unternehmen bezieht seit langem von einem bestimmten Lieferanten einen Massenartikel, wobei der Ausschussanteil 5% beträgt. Ein Konkurrenzangebot verspricht bei gleichem Preis einen Ausschussanteil unter 5%. Unter 100 zufällig ausgewählten Artikeln des konkurrierenden Anbieters waren 2 Ausschussteile.

a) Können Sie schließen, dass der Ausschussanteil des Konkurrenzanbieters unter 5% liegt. Formulieren Sie einen statistischen Test als Entscheidungsgrundlage, wobei Sie von der Unabhängigkeit der Artikel ausgehen können. Wie ist auf dem 5%-Signifikanzniveau zu entscheiden?

b) Berechnen Sie den p-Wert des Tests aus Aufgabenteil a). Welches Signifikanzniveau müsste man vorgeben, damit H_0 gerade noch abgelehnt wird?

Aufgabe 2

In einer Versicherungsgesellschaft soll festgestellt werden, ob das Unfallrisiko bei Arbeitsunfällen an verschiedenen Wochentagen gleich groß ist. Dafür wurde folgende Tabelle ermittelt.

Tag	Montag	Dienstag	Mittwoch	Donnerstag	Freitag	Σ
Zahl der Unfälle	26	17	22	24	31	120

Prüfen Sie die Nullhypothese auf dem 5%-Signifikanzniveau, dass das Unfallrisiko an allen Arbeitstagen gleich hoch ist.

Aufgabe 3

Aus einer Grundgesamtheit wird eine uneingeschränkte Zufallsstichprobe vom Umfang $n = 150$ gezogen. Dabei wurde ein stetiges Merkmal erhoben, welches ausschließlich positive reelle Werte annehmen kann. Die einzelnen Merkmalsausprägungen wurden in drei Klassen eingeteilt. Es resultierten nachfolgende absolute Häufigkeiten.

Klasse	i	n_i
$[0; 4{,}351)$	1	90
$[4{,}351; 11{,}070)$	2	50
$[11{,}070; \infty)$	3	10

Überprüfen Sie auf dem 5%-Signifikanzniveau die Nullhypothese, dass die Stichprobe aus einer χ^2-verteilten Grundgesamtheit mit 5 Freiheitsgraden stammt.

Aufgabe 4

In einer Stichprobe von 21 neu emittierten Unternehmensanleihen mit einem Rating von AAA wies die Laufzeit eine empirische Varianz von 40 auf. In einer anderen Stichprobe von 13 neu emittierten Unternehmensanleihen mit einem Rating von CCC betrug die Varianz der Laufzeit nur 12.

Prüfen Sie, ob die Varianz der Laufzeit der AAA-Papiere signifikant größer als die der CCC-Papiere ist. Treffen Sie für Ihren Test eine sinnvolle Verteilungsvoraussetzung und führen Sie diesen auf dem 2,5%-Signifikanzniveau durch.

Übung 9: Hypothesentests III

Aufgabe 1

Bei einer Studie mit 32 Teilnehmern soll untersucht werden, ob Personen mit fleischloser Ernährung am Tag weniger Kalorien zu sich nehmen als Personen, die Fleisch konsumieren. Von den Teilnehmern ernährten sich 12 fleischlos. Sie nahmen im Mittel 1780 Kalorien täglich zu sich. Die übrigen Teilnehmer nahmen täglich im Mittel 1900 Kalorien zu sich. Außerdem ergab sich für die Gruppe der Fleischlosen eine erwartungstreue Standardabweichung von 230. Für die Fleischesser ergab sich eine erwartungstreue Standardabweichung von 250. Bei der Studie wird außerdem angenommen, dass die tägliche Kalorienmenge normalverteilt ist und dass die Varianz bei beiden Gruppen übereinstimmt.

Können Sie schließen, dass Personen mit fleischloser Ernährung am Tag weniger Kalorien zu sich nehmen als Personen, die Fleisch konsumieren? Führen Sie zur Untersuchung einen geeigneten Test auf dem 5%-Signifikanzniveau durch.

Aufgabe 2

Ein Marktforschungsinstitut wurde von einem Einzelhändler beauftragt, das Kaufverhalten in einem Selbstbedienungsmarkt zu untersuchen. Dabei werden zwei Produkte A und B an zwei aufeinander folgenden Wochen je einmal oben im Regal bzw. unten im Regal platziert. Es ergaben sich folgende Verkaufszahlen:

Anordnung	A	B	Σ
1. Woche (A oben; B unten)	146	124	270
2. Woche (B oben; A unten)	84	118	202
Σ	230	242	472

Prüfen Sie mit einem geeigneten Test auf dem 5%-Signifikanzniveau, ob das Kaufverhalten von der Platzierung im Regal abhängt.

Aufgabe 3

In einem Unternehmen wird ein bestimmtes Produkt an drei Produktionsstätten A, B und C hergestellt.

Aus der laufenden Produktion wurden zufällig 800 Produkte ausgewählt. Anhand des Kontrollzettels wird festgestellt, dass 336 bzw. 288 davon an der Produktionsstätte A bzw. B hergestellt wurden. Von den 800 Produkten sind 120 fehlerhaft. 53 bzw. 44 dieser fehlerhaften Produkte stammen von der Produktionsstätte A bzw. B.

Prüfen Sie mit einer Irrtumswahrscheinlichkeit von 0,05 die Hypothese, dass die Fehlerhaftigkeit unabhängig von der Produktionsstätte ist.

Übung 10: Lineare Einfachregression

Aufgabe 1

Zur Schätzung einer Kostenfunktion verwendet ein Unternehmer die Daten von 10 Monaten für die monatliche Produktionsmenge M (in 10000 Stück) und die Herstellungskosten K pro Monat (in 10000 Euro). Aus den Datenpaaren (m_i, k_i) für die Monate $i = 1, \ldots, 10$ ergeben sich folgende Summenwerte:

$$\sum_{i=1}^{10} m_i = 55 \qquad \sum_{i=1}^{10} m_i^2 = 385 \qquad \sum_{i=1}^{10} k_i = 720 \qquad \sum_{i=1}^{10} k_i^2 = 53458 \qquad \sum_{i=1}^{10} m_i k_i = 4323$$

a) Geben Sie ein geeignetes Regressionsmodell an.

b) Schätzen Sie die unbekannten Parameter mit der KQ-Methode und interpretieren Sie den Steigungskoeffizienten.

c) Bestimmen Sie das symmetrische 95%-Konfidenzintervall für den Steigungskoeffizieten.

Aufgabe 2

Der Inhaber eines Eiscafes untersucht die Abhängigkeit des Eisumsatzes im Straßenverkauf V (in 1000 Euro) von der Mittagstemperatur T (in Grad Celsius). Aus seiner Erhebung an 15 Tagen ergaben sich folgende Werte:

$$\frac{1}{15}\sum_{i=1}^{15} v_i = 5{,}3 \qquad \frac{1}{15}\sum_{i=1}^{15} (v_i - \bar{v})^2 = 3{,}24 \qquad \frac{1}{15}\sum_{i=1}^{15} t_i = 22 \qquad \frac{1}{15}\sum_{i=1}^{15} t_i^2 = 494{,}24 \qquad \frac{1}{15}\sum_{i=1}^{15} v_i t_i = 121{,}78$$

a) Geben Sie ein geeignetes Regressionsmodell an.

b) Schätzen Sie die unbekannten Parameter mit der KQ-Methode und interpretieren Sie den Steigungskoeffizienten.

c) Kann auf dem 1%-Signifikanzniveau geschlossen werden, dass der Eisumsatz mit steigenden Temperaturen zunimmt?

Aufgabe 3

Bei einer Verkehrsuntersuchung wurde für 25 Verkehrsbezirke der Individualverkehr erfasst. Als Beschreibung für die Struktur des Verkehrs wurden die Merkmale

X : „Einwohnerzahl in 1000“ und Y : „Zahl der PKW pro Stunde in 1000“

für die 25 Verkehrsbezirke ermittelt. Dabei ergaben sich folgende Werte:

$$\sum_{i=1}^{25} y_i = 79{,}75 \qquad \sum_{i=1}^{25} x_i = 100 \qquad \sum_{i=1}^{25} y_i^2 = 277{,}825 \qquad \sum_{i=1}^{25} x_i^2 = 475 \qquad \sum_{i=1}^{25} y_i x_i = 345{,}5$$

a) Schätzen Sie die unbekannten Parameter mit der KQ-Methode und unterstellen Sie dabei, dass die Zahl der PKW von der Einwohnerzahl abhängt.

b) Testen Sie die Nullhypothese, dass die Steigung der Regressionsgeraden auf dem 10%-Signifikanzniveau höchstens 0,3 beträgt.

Übung 11: Multiple lineare Regression

Aufgabe 1

Mit den Daten von 177 Mietwohnungen einer Stadt wurde versucht, die Determinanten des Mietzinses empirisch zu ermitteln. Mittels KQ-Methode wurde folgendes Modell geschätzt:

$$MIETE = \beta_1 + \beta_2 NWF + \beta_3 Alter + \beta_4 DZENT + \varepsilon$$

Dabei bedeuten die Variablen MIETE: Mietpreis in Euro, NWF: Nettowohnfläche in qm, ALTER: Alter der Wohnung in Jahren, DZENT: Distanz zum Stadtzentrum in km.
EViews liefert folgende Regressionsergebnisse:

Dependent Variable: MIETE
Method: Least Squares
Date: 12/11/02 Time: 08:00
Sample: 1 177
Included observations: 177

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C		144.7024	6.335762	0.0000
NWF	11.40628		7.947048	0.0000
ALTER	-6.469049	0.937492		0.0000
DZENT		20.54634	-3.433036	
R-squared		Mean dependent var		1303.944
Adjusted R-squared		S.D. dependent var		526.3415
S.E. of regression				
Sum squared resid	26928912			

a) Bestimmen Sie die fehlenden Werte des Outputs. Dabei ist die Summe der quadrierten Abweichungen der abhängigen Variablen MIETE von ihrem Mittelwert $48.758.231,3776$.

b) Beurteilen Sie die Signifikanz von $\widehat{\beta}_0$ und $\widehat{\beta}_1$ und interpretieren Sie sie inhaltlich.

c) Bilden Sie das 90%-Konfidenzintervall für das Alter der Wohnung.

d) Testen Sie, ob die Distanz zum Stadtzentrum auf dem 1%-Signifikanzniveau einen signifikanten Einfluss auf den Mietzins hat.

e) Testen Sie die Aussage, dass der Quadratmeter Wohnfläche mindestens 10 Euro kostet.

Aufgabe 2

Gegeben sei folgende Stichprobe mit 5 Beobachtungen für x_1, x_2 und y:

x_1	1,2	3	4,5	5,8	7,2
x_2	0,6	0,75	0,8	0,9	1,4
y	2,6	1,6	4	3	4,9

Bestimmen Sie die Parameter des Modells (mit Intercept):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Nutzen Sie für die Bestimmung:

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} 3,0448 & 0,4025 & -5,1593 \\ 0,4025 & 0,2264 & -1,5565 \\ -5,1593 & -1,5565 & 13,3871 \end{pmatrix}$$

Aufgabe 3

Für eine multiple Regression mit $n = 14$ ergab sich folgender Schätzer für den Vektor $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} -1,0014 \\ -0,2347 \\ -0,001 \end{pmatrix}$$

Für das Bestimmtheitsmaß wurde zusätzlich ein Wert von $0,475$ errechnet. Prüfen Sie auf dem $2,5\%$-Signifikanzniveau, ob mindestens ein Regressor zur Erklärung der Zielvariablen beiträgt.