1. Aufgabe: Eine sächsische Molkerei füllt Milch in Tetrapacks ab. Die Füllmenge ist normalverteilt mit Standardabweichung $\sigma = 1\, ml$. In einer Stichprobe von 12 Tetrapacks wurden folgende Füllmengen gemessen:

499,7	501,3	499,1	500,6	500,2	499,5
500,0	499,1	500,5	499,6	498,7	501,7

a) Bestimmen Sie zum Konfidenzniveau von 99 % ein zentrales Konfidenzintervall für die erwartete Füllmenge.

b) Wie groß muss n mindestens sein, damit die Länge des 99 % Konfidenzintervalles höchstens $0,5\, ml$ ist?

c) Mit Statgraphics wurde aus den Daten noch folgendes Testergebnis erzielt:

Tests for Normality for Füllmenge

Test	*Statistic*	*P-Value*
Shapiro-Wilk W	0,963671	0,779243

Welche Hypothesen werden getestet und wie lautet die Testentscheidung bei $\alpha = 0,1$?

Lösung:

$$1 - \alpha = 0,99 \quad \Longrightarrow \quad \alpha = 0,01 \quad \Longrightarrow \quad z_{1-\frac{\alpha}{2}} = z_{0,995} = 2,5758$$

$\sigma = 1\, ml$, $n = 12$ und $\overline{x} = 500\, ml$.

a)

$$\begin{aligned} \overline{x} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < \quad & \mu \quad < \overline{x} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \\ 500 - \frac{1}{\sqrt{12}} \cdot 2,5758 < \quad & \mu \quad < 500 + \frac{1}{\sqrt{12}} \cdot 2,5758 \\ 499,26 < \quad & \mu \quad < 500,74 \end{aligned}$$

b)

$$l = 2 \cdot d = 0,5 \qquad \Longrightarrow \qquad d = 0,25.$$

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{d} \right)^2 \cdot \sigma^2 = \left(\frac{2,5758}{0,25} \right)^2 \cdot 1^2 = \underline{106,156}$$

Der Stichprobenumfang muss mindestens 107 betragen.

c) X - zufällige Füllmenge
H_0 : „Füllmenge X ist normalverteilt."
H_A : „Füllmenge X ist nicht normalverteilt."

$$P = 0,779243 > 0,1 = \alpha \quad \Longrightarrow \quad H_0 \text{ wird angenommen.}$$

Die Verteilung der zufälligen Füllmenge X unterscheidet sich nicht signifikant von der einer Normalverteilung.

---

2. Aufgabe: Im Rahmen der statistischen Qualitätskontrolle wird ein Posten mit höchstens 1,7 % Ausschussanteil als gut angesehen. Ein Posten mit mehr als 5% Ausschussanteil ist hingegen ein schlechter Posten. Der Produzent fordert, dass ein guter Posten höchstens mit Wahrscheinlichkeit von 0,01 abgelehnt wird. Der Konsument fordert, dass ein schlechter Posten höchstens mit Wahrscheinlichkeit 0,05 angenommen wird.

a) Prüfen Sie, ob für die Annahmezahl $c = 14$ ein n bestimmbar ist, so dass die Forderungen des Produzenten und auch die des Konsumenten erfüllt sind. Falls ein n bestimmbar ist, dann geben Sie alle möglichen (n, c)-Pläne an.

   Hinweis: Nutzen Sie die Poissonapproximation.

b) Für die Annahmezahl $c = 16$ gibt es folgende Pläne, welche beide Forderungen (die vom Konsument und die vom Produzent) erfüllen:

$$\begin{aligned}(n,c) &= (487,16)\\ (n,c) &= (488,16)\\ &\vdots\\ (n,c) &= (523,16)\end{aligned}$$

   Welchen dieser Pläne würde der Konsument bevorzugen, um sein Risiko kleinstmöglich zu halten? Begründen Sie kurz!

Lösung:

$$p_\alpha = 0,017 \qquad p_\beta = 0,05$$
$$\alpha = 0,01 \qquad \beta = 0,05$$

a) Poisson-Approximation: c=14

$$\frac{\chi^2_{2(c+1);1-\beta}}{2p_\beta} \le n \le \frac{\chi^2_{2(c+1);\alpha}}{2p_\alpha} \implies \frac{\chi^2_{30;0,95}}{0,1} \le n \le \frac{\chi^2_{30;0,01}}{0,034}$$
$$\implies \frac{43,77}{0,1} \le n \le \frac{14,95}{0,034} \implies 437,7 \le n \le 439,7$$

   Mögliche Pläne sind: $(n,c) = (438,14)$ und $(n,c) = (439,14)$.

b) Der Konsument würde den Plan $(n,c) = (523,16)$ wählen. Das Konsumentenrisiko ist bei diesen Plan am geringsten unter allen möglichen Plänen. Da bei der festen Annahmezahl $c = 16$ mit stiegenden Stichprobenumfang n die Wahrscheinlichkeit der Ablehnung einer schlechten Lieferung immer größer wird.

---

3. Aufgabe: In einer Studie soll der Erfolg von 3 verschiedenen Düngemitteln verglichen werden. Diese wurden auf 22 Erdbeerfeldern verwendet. Dabei wurden folgende Erträge in Kilogramm erzielt. (Für die einzelnen Düngemittel sind die Erträge schon der Größe nach geordnet.)

Düngemittel 1	44	45	55	61	63	67	78		
Düngemittel 2	62	68	75	79	80	81	86	90	92
Düngemittel 3	72	85	89	100	101	121			

Die Erträge sind **nicht** normalverteilt. Testen Sie zum Niveau $\alpha = 0,01$, ob die erwarteten Erträge der 3 Düngemittel gleich sind oder sich signifikant voneinander unterscheiden.

Lösung: Bestimmung der Ränge:

Rang(D_1)	1	2	3	4	6	7	11			$r_{1.} = 34$
Rang(D_2)	5	8	10	12	13	14	16	18	19	$r_{2.} = 115$
Rang(D_3)	9	15	17	20	21	22				$r_{3.} = 104$

Bindungskorrektur: Es liegen keine Bindungen vor also ist $B = 1$.

$$\begin{aligned} t &= \frac{1}{B}\left[\frac{12}{n(n+1)}\sum_{i=1}^{p}\frac{1}{n_i}r_{i.}^2 - 3(n+1)\right] \\ &= \frac{1}{1}\left[\frac{12}{22\cdot 23}\left(\frac{1}{7}34^2 + \frac{1}{9}115^2 + \frac{1}{6}104^2\right) - 3\cdot 23\right] \\ &= \underline{12,52} \end{aligned}$$

$$K = \{T \mid T > \chi^2_{p-1,1-\alpha}\} = \{T \mid T > \chi^2_{2,0.99} = 9,21\}$$

$T = 12,52 > 9,21 \quad \Longrightarrow \quad H_0$ wird abgelehnt, d.h. die erwarteten Erträge unterscheiden sich sich signifikant bei den unterschiedlichen Düngemitteln.

---

4. Aufgabe: Für 29 PKWs wurden die Merkmale

X_1 - Alter,
X_2 - Leistung und
X_3 - Verbrauch erfasst.

Aus der Stichprobe erhält man folgende Schätzung der Korrelationen:

$$r_{X_1,X_2} = -0,18 \quad ; \quad r_{X_1,X_3} = 0,39 \quad \text{und} \quad r_{X_2,X_3} = 0,51.$$

a) Schätzen Sie die partielle Korrelation ($\rho_{X_2,X_3|X_1}$) zwischen Leistung und Verbrauch unter Partialisierung des Alters.

b) Testen Sie, ob die partielle Korrelation zwischen Leistung und Verbrauch unter Partialisierung des Alter signifikant ($\alpha = 0,01$) größer als 0 ist.

Lösung:

a)

$$\begin{aligned} r_{X_2,X_3|X_1} &= \frac{r_{X_2,X_3} - r_{X_2,X_1} r_{X_3,X_1}}{\sqrt{(1-r^2_{X_2,X_1})(1-r^2_{X_3,X_1})}} \\ &= \frac{0,51 - (-0,18)\cdot 0,39}{\sqrt{(1-(-0,18)^2)\cdot(1-0,39^2)}} = \underline{0,64} \end{aligned}$$

b)

$$H_0:\ \rho_{X_2,X_3|X_1} \leq 0 \quad \text{gegen} \quad H_A:\ \rho_{X_2,X_3|X_1} > 0.$$

$$\begin{aligned} K &= \{t \mid t \geq t_{n-3,1-\alpha}\} \qquad n = 29, \alpha = 0,01 \Longrightarrow t_{26,0.99} = 2,48 \\ &= \{t \mid t \geq 2,48\} \end{aligned}$$

$$\begin{aligned} t &= \frac{r_{(X,Y)|Z}\cdot\sqrt{n-3}}{\sqrt{1-r^2_{(X,Y)|Z}}} \\ &= \frac{0,64\cdot\sqrt{26}}{\sqrt{1-0,64^2}} = 4,25 > 2,48 \quad t \in K \quad H_0 \text{ wird abgelehnt.} \end{aligned}$$

D.h. die partielle Korrelation zwischen der Leistung X_2 und den Verbrauch X_3 unter Partialisierung des Alters X_1 ist signifikant größer als 0.

---

5. Aufgabe: Für 29 PKWs wurden die Merkmale **Preis** Y (in €), gefahrene **Kilometer** X_1, **Leistung** X_2 (in Ps) und **Alter** X_3 (in Jahren) erfasst. Aus den Daten erhält man das folgende Ergebnis.

Multiple Regression - Preis

Dependent variable: Preis
Independent variables:
Kilometer
Leistung
Alter

		Standard	T	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	16646,2	771,963	21,5635	0,0000
Kilometer	-0,0157827	0,00253198	-6,23334	0,0000
Leistung	5,92975	5,19015	1,1425	0,2641
Alter	-907,708	44,4003	-20,4437	0,0000

a) Wie lautet die geschätzte Regressionsfunktion?

Lösung:

$$\begin{aligned} \hat{y}(\underline{x}) &= \hat{a}_1 - \hat{a}_2 x_1 + \hat{a}_3 x_2 + \hat{a}_4 x_3 \\ &= 16646,2 - 0,0157827 x_1 + 5,92975 x_2 - 907,708 x_3 \end{aligned}$$

b) Welche der drei Einflussgrößen würden Sie am ehesten aus dem Modell entfernen? Begründen Sie Ihre Entscheidung kurz.

Lösung:

Egal ob man das Signifikanzniveau $\alpha = 0,05$ oder $\alpha = 0,01$ wählt, man sieht bei den t-Tests für die einzelenen Parameter immer, dass der Parameter a_3 nicht signifikant von 0 versieden ist und die anderen 3 Parameter hingen schon. Darum kann die Variable Leistung am ehesten aus den Modell entfernt werden.

$H_0: a_1 = 0$ gegen $H_1: a_1 \neq 0$ $p = 0,0000 < \alpha$ $\Longrightarrow H_0$ wird abgelehnt.
$H_0: a_2 = 0$ gegen $H_1: a_2 \neq 0$ $p = 0,0000 < \alpha$ $\Longrightarrow H_0$ wird abgelehnt.
$H_0: a_3 = 0$ gegen $H_1: a_3 \neq 0$ $p = 0,2641 > \alpha$ $\Longrightarrow H_0$ wird angenommen.
$H_0: a_4 = 0$ gegen $H_1: a_4 \neq 0$ $p = 0,0000 < \alpha$ $\Longrightarrow H_0$ wird abgelehnt.

c) Im Folgenden wurden für zwei weitere Regressionsmodelle die Ergebnisse mit Statgraphics erstellt.

Modell 1:

Simple Regression - Preis vs. Kilometer

Dependent variable: Preis
Independent variable: Kilometer
Linear model: Y = a + b*X

Coefficients

	Least Squares	*Standard*	*T*	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
Intercept	15966,7	1626,16	9,81863	0,0000
Slope	-0,0460363	0,00863454	-5,33165	0,0000

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	1,93448E8	1	1,93448E8	28,43	0,0000
Residual	1,8374E8	27	6,8052E6		
Total (Corr.)	3,77188E8	28			

R-squared = 51,2868 percent
Standard Error of Est. = 2608,68

Modell 2:

Simple Regression - Preis vs. Alter

Dependent variable: Preis
Independent variable: Alter
Linear model: Y = a + b*X

Coefficients

	Least Squares	*Standard*	*T*	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
Intercept	15775,4	450,764	34,9971	0,0000
Slope	-1071,45	54,8556	-19,5323	0,0000

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	3,52258E8	1	3,52258E8	381,51	0,0000
Residual	2,49299E7	27	923328,		
Total (Corr.)	3,77188E8	28			

i) Bestimmen Sie für das Modell 2 das Bestimmtheitsmaß.

Lösung:

$$B = \frac{SSE}{SST} = \frac{3,52258 \cdot 10^8}{3,77188 \cdot 10^8} = \underline{0,9339}$$

ii) Bestimmen Sie für das Modell 2 die Schätzung für die Standardabweichung des Fehlers.

Lösung:

$$\hat{\sigma}^2 = s^2_{Rest} = \frac{1}{n-2} SSR = \frac{1}{29-2} 2,49299 \cdot 10^7 = \underline{923329,6}$$
$$\Longrightarrow \hat{\sigma} = s_{Rest} = \underline{960,9}$$

---

iii) Welches der beiden Modelle würden Sie dem anderen Modell vorziehen? Begründen Sie Ihre Wahl kurz.

Lösung:
Beide Modelle haben 2 Parameter, welche beide signifikant von 0 verschieden sind.
Das Modell 2 hat aber das deutlich bessere Bestimmtheitsmaß(0,9339) im Vergleich zum Modell 1 (0,512868). Beim Modell 2 werden 93,39% durch die geschätze Regressionsfunktion erklärt und beim Modell 1 nur 51,29%. Das Modell 2 ist also besser und man würde es den Modell 1 vorziehen.
Dieser Vorteil spiegelt sich auch in der kleineren geschätzten Standardabweichung des Fehlers wieder. Beim Modell 2 ist diese 960,9 und beim Modell 1 mit 2608,68 deutlich größer.

---

6. Aufgabe:

a) Bei einer Mittelwertkarte mit Sollwert $a = 10$ ist die Gütefunktion an der Stelle $\mu = 11$ gleich 0,25, d.h. $g(11) = 0,25$.

i) Wie groß ist die erwartete Lauflänge, bis diese Abweichung vom Sollwert erkannt wird?

Lösung:

$$\mathbf{E}N = \frac{1}{p} = \frac{1}{g(\mu)} = \frac{1}{g(11)} = \frac{1}{0,25} = 4$$

---

ii) Wie groß ist die Wahrscheinlichkeit für den Fehler 2. Art, d.h. für unterlassenen Alarm?

Lösung:

$$1 - g(11) = 1 - 0,25 = \underline{0,75}$$

---

iii) Wie verändert sich $g(11)$, falls der Stichprobenumfang für jede Prüfung erhöht wird?

Lösung:
Durch Erhöhung des Stichprobenumfanges wird die Güte verbessert, d.h. $g(11)$ wird größer und die Wahrscheinlichkeit für unterlassenen Alarm $1 - g(11)$ wird kleiner.

---

b) Beschreiben Sie kurz die Funktionsweise eines 2-stufigen Stichprobenplanes mit: $n_1 = 50$, $n_2 = 50$, $c_1 = 0$, $c_2 = 4$ und $c_3 = 5$.

Lösung:
Man zieht aus der Lieferung eine erste Stichprobe vom Unfang $n_1 = 50$.

X_1 ist die zufällige Anzahl der fehlerhaften Einheiten in dieser Stichprobe.

Ist $x_1 = 0 = c_1$ dann wird die Lieferung angenommen und ist $x_1 > 4 = c_2$ so wird die Lieferung abgelehnt.

Ist $c_0 = 0 < x_1 \leq 4 = c_2$ so zieht man aus der Lieferung eine zweite Stichprobe vom Umfang $n_2 = 50$. X_2 ist die zufällige Anzahl der fehlerhaften Einheiten in dieser zweiten Stichprobe.

Ist $x_1+x_2 \leq 5 = c_3$, so wird die Lieferung angenommen und ist $x_1+x_2 > 5 = c_3$ dann wird die Lieferung abgelehnt.

---

c) Sei $F_{(X,Y)}(x, y)$ die gemeinsame Verteilungsfunktion des Zufallsvektors (X, Y) an der Stelle (x, y) mit $x, y \in \mathbb{R}$. Geben Sie den richtigenen Ausdruck für

i. $\lim_{x \to \infty} F_{(X,Y)}(x, y)$

ii. $\lim_{x \to -\infty} F_{(X,Y)}(x, y)$

an.

Zur Auswahl stehen dabei dabei folgende Ausdrücke:
-1, 0, 1, $-\infty$, ∞, $F_Y(y)$ und $F_X(x)$.

Lösung:

$$\lim_{x \to \infty} F_{(X,Y)}(x, y) = F_Y(y)$$

$$\lim_{x \to -\infty} F_{(X,Y)}(x, y) = 0$$
