1. Aufgabe: Im Rahmen der statistischen Qualitätskontrolle wird ein Posten mit höchstens 2 % Ausschussanteil als gut angesehen. Ein Posten mit mehr als 4 % Ausschussanteil ist hingegen ein schlechter Posten. Das Risiko des Produzenten beträgt 2,5 % und das des Konsumenten 1 %.

a) Bestimmen Sie für diese Werte die Annahme- und Ablehnungsgerade eines sequentiellen Stichprobenplanes.

b) Wie groß ist der erwartete Stichprobenumfang für die Fälle, dass der Ausschussanteil gleich 2% oder 4% oder c_s ist?

c) Wieviele Stücke müssen mindestens geprüft werden, bis die Lieferung als gut angenommen wird?

Lösung:

$p_\alpha = 0,02\ , \quad \alpha = 0,025 \quad$ und $\quad p_\beta = 0,04\ , \quad \beta = 0,01.$

a)

$$\begin{aligned}
d &= \ln\left(\frac{p_\beta \cdot (1-p_\alpha)}{p_\alpha \cdot (1-p_\beta)}\right) = \ln\left(\frac{0,04 \cdot 0,98}{0,02 \cdot 0,96}\right) = 0,7137665 \\
a &= \frac{\ln\left(\frac{1-\alpha}{\beta}\right)}{d} = \frac{\ln\left(\frac{0,975}{0,01}\right)}{0,7137665} = \underline{6,42} \\
b &= \frac{\ln\left(\frac{1-\beta}{\alpha}\right)}{d} = \frac{\ln\left(\frac{0,99}{0,025}\right)}{0,7137665} = \underline{5,15} \\
c_s &= \frac{\ln\left(\frac{1-p_\alpha}{1-p_\beta}\right)}{d} = \frac{\ln\left(\frac{0,98}{0,96}\right)}{0,7137665} = \underline{0,028888}
\end{aligned}$$

Annahmegerade: $-a + c_s \cdot k = -6,42 + 0,028888 \cdot k$

Ablehnungsgerade: $b + c_s \cdot k = 5,15 + 0,028888 \cdot k$

b) • $p = p_\alpha = 0,02$ und $L(p_\alpha) = 1 - \alpha = 0,975$

$$\mathbf{E}N = \frac{b-(a+b)L(p)}{p-c_s} = \frac{5,15-(5,15+6,42)\cdot 0,975}{0,02-0,028888} = \underline{689,8}$$

• $p = p_\beta = 0,04$ und $L(p_\alpha) = \beta = 0,01$

$$\mathbf{E}N = \frac{b-(a+b)L(p)}{p-c_s} = \frac{5,15-(5,15+6,42)\cdot 0,01}{0,04-0,028888} = \underline{453,1}$$

• $p = c_s$

$$\mathbf{E}N = \frac{ab}{c_s(1-c_s)} = \frac{6,42 \cdot 5,15}{0,028888 \cdot (1-0,028888)} = \underline{1178,6}$$

c) Sind alle $X_k = 0$ für $k = 1, 2, \ldots$, so wird die Lieferung angenommen, falls die Annahmegerade größer gleich Null ist:

$$-a + c_s \cdot k \geq 0 \qquad \Longrightarrow \qquad k \geq \frac{a}{c_s} = \frac{6,42}{0,028888} = \underline{222,2}$$

Selbst wenn alle geprüften Stücke gut sind, müssen damit mindestens 223 Stück geprüft werden.

---

2. Aufgabe: Ein Zeitschriftenverlag möchte wissen, ob sich die Lesezeiten in den Segmenten Wirtschaftsmagazin und Computerzeitschriften unterscheiden.

Es liegt folgende Stichprobe der in Minuten gemessenen Lesezeit vor.

Wirtschaftsmagazin	92	105	91	59	91	113	147	
Computerzeitschriften	148	136	123	104	62	119	83	89

Die Lesezeiten sind normalverteilt mit der gleichen Varianz in beiden Sparten. Die geschätzten Standardabweichungen lauten $s_1 = 26,8$ und $s_2 = 28,9$. Testen Sie zu einem Niveau $\alpha = 0,05$, ob sich in den beiden Sparten die erwarteten Lesezeiten signifikant unterscheiden?

Lösung:

X_1 - zufällige Lesezeit beim Wirtschaftsmagazin. $\quad X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$
X_2 - zufällige Lesezeit beim Computermagazin. $\quad X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

Normalverteilt und gleiche Varianz: $\sigma_1^2 = \sigma_2^2 \quad \Longrightarrow \quad$ Doppelter-t-Test:

1. $H_0 : \mu_1 = \mu_2 \qquad$ gegen $\qquad H_A : \mu_1 \neq \mu_2$
2. $\alpha = 0,05$
3. $$T = \frac{\overline{X}_1 - \overline{X}_2}{S_g}\sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad \text{mit} \quad S_g^2 = \frac{1}{n_1 + n_2 - 2}\left((n_1 - 1)S_1^2 + (n_2 - 1)S_2^2\right)$$
4. $K = \{t \mid |t| \geq t_{n_1+n_2-2,1-\frac{\alpha}{2}}\}$
   $n_1 = 7$ und $n_2 = 8$, $\quad t_{13,0.975} = 2.16 \qquad \Longrightarrow \qquad K = \{t \mid |t| \geq 2,16\}$

5. $\overline{x}_1 = 99,71429$, $\overline{x}_2 = 108$ und $s_1 = 26,8$, $s_2 = 28,9$.

$$\begin{aligned} s_g^2 &= \frac{1}{13}(6 \cdot 26,8^2 + 7 \cdot 28,9^2) = 781,2238 \\ t &= \frac{99,71429 - 108}{\sqrt{781,2238}}\sqrt{\frac{7 \cdot 8}{15}} = -0,573 \end{aligned}$$

6. $|t| = 0,573 \not\geq 2,16 \quad \Longrightarrow \quad t \notin K \quad \Longrightarrow H_0$ wird angenommen. D.h. die erwarteten Lesezeiten in den beiden Segmenten unterscheiden sich nicht signifikant.

3. Aufgabe: 100 Leser eines Wirtschaftsmagazins wurden nach ihren durchschnittlichen Lesezeiten befragt. Es soll untersucht werden, ob die Lesezeit normalverteilt ist. Dazu liegt folgendes Testergebnis vor.

Goodness-of-Fit Tests for Lesezeit

Chi-Square Test

	Lower Limit	*Upper Limit*	*Observed Frequency*	*Expected Frequency*	*Chi-Square*
at or below		48,0	1	2,06	0,55
	48,0	60,0	3	3,00	0,00
	60,0	72,0	6	5,75	0,01
	72,0	84,0	10	9,39	0,04
	84,0	96,0	12	13,08	0,09
	96,0	108,0	19	15,52	
	108,0	120,0	15	15,71	
	120,0	132,0	12	13,55	0,18
	132,0	144,0	12	9,97	0,41
	144,0	156,0	4	6,25	0,81
	156,0	168,0	2	3,34	0,54
above	168,0		4	2,38	1,10

Chi-Square = 4,53981

a) Bestimmen Sie die 2 fehlenden Werte in der Tabelle (Spalte Chi-Square).

b) Führen Sie den Test zu Ende und treffen Sie die Testentscheidung zum Niveau $\alpha = 0,05$.

c) Mit den gleichen Daten wurde ein weiterer Test mit folgendem Ergebnis durchgeführt.

Tests for Normality for Lesezeit

Test	*Statistic*	*P-Value*
Shapiro-Wilk W	0,986744	0,859585

Welche Hypothese wird getestet und wie lautet die Testentscheidung bei $\alpha = 0,05$?

Lösung: X- zufällige durchschnittliche Lesezeit.

a)

$$i = 6: \quad \frac{(h_i - np_i)^2}{np_i} = \frac{(19 - 15,52)^2}{15,52} = \underline{0,78}$$

$$i = 7: \quad \frac{(h_i - np_i)^2}{np_i} = \frac{(15 - 15,71)^2}{15,71} = \underline{0,03}$$

b) $H_0:$ „durchschnittliche Lesezeit X ist normalverteilt“
$H_A:$ „durchschnittliche Lesezeit X ist nicht normalverteilt“

$t = 0,55 + 0,00 + 0,01 + ... + 0,54 + 1,10 = 4,54$
$K = \{t \mid t > \chi^2_{k-m-1;1-\alpha}\}$ $\quad k = 12$ und $m = 2$ $\quad \chi^2_{9,0.95} = 16,92$

$$t = 4,54 \not> 16,92 = \chi^2_{9,0.95} \quad \Longrightarrow \quad H_0 \text{ wird angenommen.}$$

Die durchschnittliche Lesezeit unterscheidet sich nicht signifikant von einer Normalverteilung.

c) H_0 : „durchschnittliche Lesezeit X ist normalverteilt“
H_A : „durchschnittliche Lesezeit X ist nicht normalverteilt“

$$p = 0,859585 > 0,05 = \alpha \quad \Longrightarrow \quad H_0 \text{ wird angenommen.}$$

Die durchschnittliche Lesezeit unterscheidet sich nicht signifikant von einer Normalverteilung.

---

4. Aufgabe: Für die 7 größten Flughäfen (nach Anzahl der abgefertigten Passagiere) wird der Zusammenhang zur Einwohnerzahl der Metropolregion der Stadt untersucht. Folgende Daten liegen vor.

Stadt	Passagiere in Millionen	Einwohner der Metropolregion in Millionen
Atlanta	95	6
Peking	82	21
London	70	15
Tokio	67	36
Chicago	66	9
Los Angeles	64	13
Paris	62	10

Bestimmen Sie die Rangkorrelation von Kendall zwischen der Anzahl der abgefertigten Passagiere und der Einwohnerzahl der Metropolregion der Stadt.

Lösung:

X - zufällige Anzahl der abgefertigten Passagiere
Y - zufällige Einwohnerzahl

j	1	2	3	4	5	6	7
Passagiere	62	64	66	67	70	82	95
Einwohnerzahl	10	13	9	36	15	21	6
q_j	2	2	1	3	1	1	0

$$\sum_{j=1}^{7} q_j = 2 + 2 + ... + 1 + 0 = 10$$

$$r_{X,Y}^{(K)} = 1 - \frac{4 \cdot \sum\limits_{j=1}^{7} q_j}{n \cdot (n-1)} = 1 - \frac{4 \cdot 10}{7 \cdot 6} = \frac{1}{21} = \underline{0,0476}$$

5. Aufgabe: Für 25 der größten Flughäfen wurde die Anzahl der abgefertigten Passagiere in den Jahren 2009 und 2012 erfasst. Aus den Daten (Anzahl der Passagiere in Millionen) erhält man das folgende Ergebnis.

Simple Regression - Passagiere 2012 vs. Passagiere 2009

Dependent variable: Passagiere 2012
Independent variable: Passagiere 2009
Linear model: Y = a + b*X

Coefficients

	Least Squares	*Standard*	*T*	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
Intercept	9,58842	4,30419	2,22769	0,0355
Slope	0,944436	0,0864419	10,9257	0,0000

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	3903,43	1	3903,43	119,37	0,0000
Residual	784,802	24	32,7001		
Total (Corr.)	4688,23	25			

Predicted Values

		95,00%		*95,00%*	
	Predicted	*Prediction*	*Limits*	*Confidence*	*Limits*
X	*Y*	*Lower*	*Upper*	*Lower*	*Upper*
30,0		25,4698	50,3733	33,9524	41,8906
80,0	85,1433	71,8357	98,451	78,995	91,2917

a) Wie lautet die Modellgleichung?

Lösung:

$$Y = a + b \cdot X + \varepsilon$$

Y - Passagiere 2012, X - Passagiere 2009, ε - zufällige Fehler.

b) Wie lautet die geschätzte Regressionsfunktion?

Lösung:

$$\hat{y}(x) = \hat{a} + \hat{b} \cdot x = 9,58842 + 0,944436 \cdot x$$

c) Schätzen Sie die Varianz des Fehlers.

Lösung:

$$\hat{\sigma}^2 = \frac{1}{n-2} SSR = \frac{1}{25-2} 784,802 = \underline{34,1}$$

d) Welche Hypothesen werden in der Tabelle Coefficients getestet und wie lauten die Testentscheidungen bei $\alpha = 0,05$?

$H_0: \ a = 0$ gegen $H_A: \ a \neq 0$ $\quad p = 0,0355 < \alpha \quad \Longrightarrow \quad H_0$ wird abgelehnt.
$H_0: \ b = 0$ gegen $H_A: \ b \neq 0$ $\quad p = 0,0000 < \alpha \quad \Longrightarrow \quad H_0$ wird abgelehnt.

Beide Parameter, a und b, sind signifikant von 0 verschieden.

e) Die Passagierzahl im Jahr 2009 war 30 Millionen. Welche Passagierzahl wird nach diesem Modell für das Jahr 2012 prognostiziert?
Lösung:

$$\hat{y}(30) = 9,58842 + 0,944436 \cdot 30 = 37,9215$$

$37,9215 \cdot 10^6$ Passagiere werden für das Jahr 2012 prognostiziert.

f) Die Passagierzahl im Jahr 2009 war 80 Millionen. Wie lautet das 95 % Prognoseintervall für die Passagierzahl im Jahr 2012?

Lösung: Prediction Limits

$$[71,8357\,;\ 98,451]$$

---

6. Aufgabe:

a) Angenommen H_0 ist richtig. Ein Signifikanztest zum Niveau $\alpha = 0,03$ wird nicht nur einmal, sondern 1000 mal mit verschiedenen Stichproben durchgeführt. Wieviele Annahmen von H_0 sind bei diesen 1000 Tests zu erwarten?

Lösung:

$$p = P(H_0 \text{ wird abgelehnt } | H_0 \text{ ist richtig}) = \alpha = 0,03$$

$\Longrightarrow$ erwartete Anzahl, dass H_0 abgelehnt wird: $np = 1000 \cdot 0,03 = 30$
$\Longrightarrow$ erwartete Anzahl, dass H_0 angenommen wird: $1000 - 30 = \underline{970} = n(1-p)$

b) Gegeben sind folgende Stichprobenpaare:

i	1	2	3
x_i	1	2	3
y_i	3	6	

Geben Sie einen Wert für y_3 so an, dass der gewöhnliche Korrelationskoeffizient kleiner 1 ($r_{X,Y} < 1$) und die Spearmansche Rangkorrelation gleich 1 ($r^{(S)}_{X,Y} = 1$) ist. Begründen Sie die Wahl des Wertes kurz!

Lösung:
$r_{X,Y} = 1$, falls die Punktepaare (x_i, y_i) alle auf einer Geraden mit positiven Anstieg liegen. Das wäre hier der Fall, falls $y_3 = 9$. Dann liegen die Punktpaare auf der Geraden $y = 3x$.

$r^S_{X,Y} = 1$, falls durch die Punktepaare (x_i, y_i) ein streng monoton wachsender Zusammenhang beschrieben wird. Das wäre hier der Fall, falls $y_3 > 6$.

Zusammen muss also gelten, $y_3 > 6$ und $y_3 \neq 9$. Also zum Beispiel $y_3 = e^\pi$ (oder auch $y_3 = 7$ oder $y_3 = 12$).

c) Es gilt für zwei Zufallsvariablen X und Y, $\mathbf{Var}(X) = 1$, $\mathbf{Var}(Y) = 2$ und $\mathbf{Corr}(X,Y) = 0,7$. Wie groß ist $\mathbf{Var}(X+Y)$?

Lösung:

$$\begin{aligned}
& \mathbf{Corr}(X,Y) &=& \frac{\mathbf{Cov}(X,Y)}{\sqrt{\mathbf{Var}X \cdot \mathbf{Var}Y}} = \frac{\mathbf{Cov}(X,Y)}{\sqrt{1 \cdot 2}} = 0,7 \\
\Longrightarrow \quad & \mathbf{Cov}(X,Y) &=& \sqrt{2} \cdot 7 \\
\Longrightarrow \quad & \mathbf{Var}(X,Y) &=& \mathbf{Var}X + \mathbf{Var}Y + 2 \cdot \mathbf{Cov}(X,Y) \\
& &=& 1 + 2 + 2 \cdot \sqrt{2} \cdot 0,7 = \underline{4,98}
\end{aligned}$$

d) Betrachten Sie den einstufigen (n,c)-Stichprobenplan. Wie verändert sich das Produzentenrisiko, falls man

i. n vergrößert und c verkleinert?

ii. n verkleinert und c verkleinert?

Lösung:

i. Wird n vergrößert und c verkleinert, so **sinkt** die Wahrscheinlichkeit dafür, dass die Lieferung angenommen wird. Das Produzentenrisiko **steigt** also.

ii. Wird n verkleinert und c verkleinert, so ist keine Aussage über die Veränderung des Produzentenrisikos machbar. Es kann sich in beide Richtungen ändern. Das hängt immer von den konkreten Werten ab.

e) Zeigen Sie, dass aus der Unabhängigkeit von X und Y die Unkorreliertheit folgt.

Lösung:

Sind X und Y unabhängig, dann gilt:

$$\mathbf{E}(X \cdot Y) = \mathbf{E}X \cdot \mathbf{E}Y.$$

Und damit:

$$\begin{aligned}
\Longrightarrow \quad & \mathbf{Cov}(X,Y) &=& \mathbf{E}(X \cdot Y) - \mathbf{E}X \cdot \mathbf{E}Y \stackrel{!}{=} 0 \\
\Longrightarrow \quad & \mathbf{Corr}(X,Y) &=& \frac{\mathbf{Cov}(X,Y)}{\sqrt{\mathbf{Var}X \cdot \mathbf{Var}Y}} = 0
\end{aligned}$$

d.h. X und Y sind unkorreliert.