Aufgabe 1: (2+4+4=10 Punkte)

Die Personalabteilung analysiert den Krankenstand im Unternehmen im ersten Halbjahr.

Monat	Jan	Feb	Mär	Apr	Mai	Jun
Fehltage	12	8	12	16	14	10

a.) Geben Sie das Skalenniveau der Analyse an.

b.) Berechnen Sie die Lagemaße der Anzahl Fehltage im Monat

- arithmetischen Mittelwert.
- Modus.

c.) Bestimmen Sie die Standardabweichung der Anzahl Fehltage im Monat.

Aufgabe 2: (3+3+4=10 Punkte)

Innerhalb eines Produktionsprozesses können fehlerhafte Bauteile produziert werden. In der Qualitätskontrolle wird ein fehlerhaftes Bauteil mit einer Wahrscheinlichkeit von $0,95$ erkannt. 1% der heilen Bauteile werden dabei fälschlicherweise als fehlerhaft eingeordnet. Unter 10.000 Bauteilen befinden sich erfahrungsgemäß 10 fehlerhafte.

a.) Bei 10.000 Bauteilen, wie oft meldet die Qualitätskontrolle ein fehlerhaftes Bauteil?

b.) Wie groß ist die Wahrscheinlichkeit, dass ein als fehlerhaft eingestuftes Bauteil wirklich fehlerhaft ist?

c.) Wie groß ist die Wahrscheinlichkeit, dass ein Bauteil nach der Qualitätskontrolle noch fehlerhaft ist?

Aufgabe 3: (4+7+3=14 Punkte)

Eine Fondsmanagerin möchte zeigen, dass die wöchentliche Rendite Ihres Fonds größer als 2% ist. Im letzen Jahr (52 Wochen) betrug die durchschnittliche Rendite 3% bei einer Standardabweichung (Volatilität) von 6%. Sie geht dabei von einer Normalverteilung der Renditen aus.

a.) Formulieren Sie das statistische Testproblem, d.h. H_0 und H_A.

b.) Testen Sie Ihre Hypothese aus a.) zum Niveau 5% und interpretieren Sie Ihr Ergebnis.

c.) Berechnen Sie die Wahrscheinlichkeit für einen Wochenverlust.

Aufgabe 4: (14+2+2+2=20 Punkte)

Ein Schuhfilialist untersucht in 5 Filialen den Zusammenhang zwischen der Anzahl Verkäufer und den Umsatz.

Filiale	Anzahl Verkäufer	Umsatz (in 1.000 Euro)
1	2	8
2	4	10
3	8	15
4	5	7
5	6	10

a.) Berechnen Sie den Korrelationskoeffizienten.

b.) Wie würden Sie die Stärke des Zusammenhanges einordnen?

c.) Zeigt diese Analyse, dass mehr Verkäufer zu einem höheren Umsatz führen?

d.) Mit welcher statistischen Methode könnten Sie den Einfluss der Anzahl Verkäufer auf den Umsatz modellieren?

Aufgabe 5: (2+2+2=6 Punkte)
Welche der folgenden Aussagen sind richtig?

a.) Der Wert einer empirischen Verteilungsfunktion an der Stelle $x = 6$ ist $0,8$, also $F(6) = 0,8$.

   i. Der Median ≤ 6.
   ii. Der Median ≥ 6.
   iii. Es ist keine Aussage über die Lage des Medians möglich.

b.) Ein $0,95$ Konfidenzintervall für den Lageparameter μ einer Normalverteilung gehe von -2 bis $+2$.

   i. 95% der Beobachtungen liegen im Intervall $[-2, +2]$.
   ii. In 95% der Fälle wird ein Konfidenzintervall den wahren Wert μ beinhalten.
   iii. Die Nullhypothese $H_0 : \mu = 0$ wird zum Signifikanzniveau 5% verworfen.

c.) Linearer Zusammenhang.

   i. Bei negativen Kontingenzkoeffizienten liegt ein negativer, linearer Zusammenhang vor.
   ii. Bei positiven Kontingenzkoeffizienten liegt ein positiver, linearer Zusammenhang vor.
   iii. Erst ab metrischen Skalenniveau können Linearzusammenhänge untersucht werden.

Lösung Aufgabe 1: (2+4+4=10 Punkte)

a.) Die Anzahl Fehltage ist metrisch diskret skaliert.

b.)
- Arithmetischer Mittelwert:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{12+8+12+16+14+10}{6} = 12.$$

- Modus: $x_{\text{mod}} = 12$ (die Merkmalsausprägung die am Häufigsten vorkommt).

c.) Zunächst Varianz berechnen:

$$\begin{aligned} s^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ &= \frac{(12-12)^2 + (8-12)^2 + (12-12)^2 + (16-12)^2 + (14-12)^2 + (10-12)^2}{6} \\ &= \frac{0+16+0+16+4+4}{6} = \frac{40}{6} = 6,67 \end{aligned}$$

und damit gilt für die Standardabweichung

$$s = \sqrt{s^2} = \sqrt{\frac{40}{6}} = 2,58.$$

Lösung Aufgabe 2: (3+3+4=10 Punkte)

Definiere zunächst folgende Ereignisse:

- BF: Bauteil fehlerhaft
- BO: Bauteil ok (nicht BF)
- QF: Bauteil in Qualitätskontrolle fehlerhaft
- QO: Bauteil in Qualitätskontrolle ok (nicht QF)

a.) Von den 10.000 Bauteilen sind 10 fehlerhaft ($P(BF) = \frac{10}{10.000}$). Von diesen werden 95% erkannt ($P(QF|BF) = 0,95$), also hier $9,5$ Stück. $10.000-10 = 9.990$ Stück sind ok ($P(BO) = 1 - P(BF)$). Davon wird 1% irrtümlich als fehlerhaft erkannt $P(QF|BO) = 0,01$, also $9.990 \cdot 0,01 = 99,9$. Insgesamt werden also $9,5 + 99,9 = 109,4$ Bauteile als fehlerhaft von der Qualitätssicherung eingestuft. (Satz von der totalen Wahrscheinlichkeit.)

$$P(QF) = P(QF|BF) \cdot P(BF) + P(QF|BO) \cdot P(BO)$$

b.) Von den $109,4$ als fehlerhaft eingestuften Bauteilen sind $9,5$ wirklich fehlerhaft, also ist die gesuchte Wahrscheinlichkeit $\frac{9,5}{109,4} = 0,0868$, also 8,7 %. (Satz von Bayes.)

$$P(BF|QF) = \frac{P(QF|BF) \cdot P(BF)}{P(QF|BF) \cdot P(BF) + P(QF|BO) \cdot P(BO)}$$

c.) Es werden $10-9,5 = 0,5$ fehlerhafte Teile nicht erkannt. Die Qualitätskontrolle wird von $9.900 - 99,9 = 9890,1$ heilen Bauteilen passiert. Die Wahrscheinlichkeit für das fehlerhafte liegt also bei $\frac{0,5}{0,5+9890,1} = 0,00005 = 5 \cdot 10^{-5}$.

$$P(BF|QO) = \frac{P(QO|BF) \cdot P(BF)}{P(QO|BF) \cdot P(BF) + P(QO|BO) \cdot P(BO)}$$

Lösung Aufgabe 3: (4+7+3=14 Punkte)

a.) Die Argumentationskette beim statistischen Hypothesentest ist folgende: Sind die Daten (Teststatistik) unter der Nullhypothese sehr unwahrscheinlich wird diese verworfen. Um zu zeigen, dass die Rendite $> 0,02$ ist, wird überprüft, ob die Daten unter der Nullhyopthese $H_0 : \mu \leq 0,02$ unwahrscheinlich sind, also

$$H_0 : \mu \leq 0,02 \quad vs. H_A : \mu > 0,02.$$

b.)

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{0,03 - 0,02}{\frac{0,06}{\sqrt{52}}} = 1,20185$$

Der kritische Wert für den einseitigen Test zum Niveau $\alpha = 5\%$ liegt bei

$$z_{1-\alpha} = z_{1-0,05} = z_{0,95} = 1,6449$$

Da aber $T = 1,20185 \not> 1,6449 = z_{0,95}$ ist, kann die Nullhypothese nicht verworfen werden.

c.)

$$\begin{aligned} P(X < 0) &= P(X \leq 0) = F(0) = \Phi(\frac{0 - 0,03}{0,06}) \\ &= \Phi(-0,5) = 1 - \Phi(0,5) = 1 - 0,6915 = 0,3085 \end{aligned}$$

Die Wahrscheinlichkeit eines Wochenverlustes liegt also bei ca. 31%.

Lösung Aufgabe 4: (14+2+2+2=20 Punkte)

a.) Es ergibt sich folgende Hilfstabelle:

i	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	2	8	-3	-2	9	4	6
2	4	10	-1	0	1	0	0
3	8	15	3	5	9	25	15
4	5	7	0	-3	0	9	0
5	6	10	1	0	1	0	0
Summe	25	50	0	0	20	38	21
Mittelwert	5	10	0	0	4	7,6	4,2

Und damit

$$r = \frac{s_{xy}}{\sqrt{s_x^2}\sqrt{s_y^2}} = \frac{4,2}{\sqrt{4}\sqrt{7,6}} = 0,76$$

b.) Da $r > 0,7$ gibt es einen hohe, positive Korrelation (=linearen Zusammenhang).

c.) Nein, nicht direkt. Es könnte auch sein, dass ein hoher Umsatz dazu führt, dass mehr Verkäufer eingestellt werden, oder es liegt ein indirekter Zusammenhang z. B. über die Verkaufsfläche vor.

d.) Mit einer linearen Regression: $y = a + b \cdot x + \epsilon$.

Lösung Aufgabe 5: (2+2+2=6 Punkte)

Folgenden Aussagen sind richtig:

a.) Der Wert einer empirischen Verteilungsfunktion an der Stelle $x = 6$ ist $0,8$, also $F(6) = 0,8$.

   i. Der Median ≤ 6.

b.) Ein $0,95$ Konfidenzintervall für den Lageparameter μ einer Normalverteilung gehe von -2 bis $+2$.

   ii. In 95% der Fälle wird ein Konfidenzintervall den wahren Wert μ beinhalten.

c.) Linearer Zusammenhang.

   iii. Erst ab metrischen Skalenniveau können Linearzusammenhänge untersucht werden.