Probeklausur

Statistik I für Wirtschaftswissenschaften

Universität: Technische Universität Berlin
Kurs/Modul: Statistik I für Wirtschaftswissenschaften
Bearbeitungszeit: 120 Minuten
Erstellungsdatum: September 19, 2025

Bearbeitungszeit: 120 Minuten.

Aufgabe 1.

(a) Gegeben sei eine Stichprobe von Renditen

$$12.3,\ 9.8,\ 11.5,\ 14.2,\ 10.7,\ 13.4,\ 8.9,\ 9.3,\ 15.1,\ 12.0,\ 11.2,\ 13.7,\ 10.1,\ 9.9,\ 8.4$$

(alle Werte in Prozent). Berechnen Sie die folgenden Kennzahlen der Stichprobe: $\bar{x}$ (Mittelwert) und den Median.

(b) Bestimmen Sie die Standardabweichung s der Renditen.

(c) Bestimmen Sie das 1. Quartil Q_1 und das 3. Quartil Q_3 sowie den Interquartilsabstand $\mathrm{IQR} = Q_3 - Q_1$.

(d) Geben Sie eine kurze Häufigkeitsverteilung der Werte an, gruppiert in Klassenbreiten von 2 Einheiten. Skizzieren Sie anschließend grob eine empirische Verteilungsform.

(e) Ziehen Sie eine kurze wirtschaftliche Interpretation der Verteilung und eventuelle Ausreißer.

Aufgabe 2.

(a) Eine Urne enthält 3 rote und 2 blaue Kugeln. Es werden zwei Kugeln ohne Zurücklegen gezogen. Bestimmen Sie die Wahrscheinlichkeiten:

$$P(\text{beide rot}), \quad P(\text{genau eine rot}), \quad P(\text{kein rot}).$$

(b) Sei X die Anzahl roter Kugeln, die gezogen werden. Bestimmen Sie die Verteilung von X (d.h. $P(X = 0), P(X = 1), P(X = 2)$) und den Erwartungswert $E(X)$.

(c) Berechnen Sie $P(\text{Rot bei der zweiten Ziehung} \mid \text{Rot bei der ersten Ziehung})$.

(d) Interpretieren Sie kurz, wie sich die Abhängigkeit der Ziehungen auf Wahrscheinlichkeiten auswirkt.

Aufgabe 3.

(a) Gegeben seien die Datensätze

$$\{(1,2),(2,4),(3,5),(4,4),(5,5),(6,7)\}.$$

Berechnen Sie das Pearson-Korrelationskoeffizient r zwischen X und Y.

(b) Bestimmen Sie die einfache lineare Regression von Y auf X: $Y = a + bX$. Geben Sie die Werte von a und b an.

(c) Bestimmen Sie R^2 (Anteil der Varianz von Y, der durch das Modell erklärt wird).

(d) Interpretieren Sie die Ergebnisse in einem wirtschaftswissenschaftlichen Kontext.

Aufgabe 4.

(a) Ein Datensatz enthält 20 Beobachtungen der monatlichen Haushaltsausgaben (in Tausend Euro):

$$0.8,\ 1.2,\ 1.0,\ 0.9,\ 1.4,\ 0.6,\ 1.8,\ 2.0,\ 1.3,\ 0.7,\ 1.1,\ 1.9,\ 2.2,\ 1.5,\ 1.0,\ 0.8,\ 1.7,\ 1.4,\ 1.3,\ 2.1.$$

Bilden Sie eine Häufigkeitsverteilung mit Klassen 0.5–1.0, 1.0–1.5, 1.5–2.0, 2.0–2.5. Geben Sie die relativen Häufigkeiten an.

(b) Bestimmen Sie Mittelwert, Median, Modus und Varianz der Daten.

(c) Interpretieren Sie die Verteilung im Hinblick auf typische monatliche Ausgaben.

(d) Diskutieren Sie, wie Ausreißer das Ergebnis beeinflussen könnten.

Lösungen

Bearbeitungszeit: 120 Minuten.

Aufgabe 1.

(a) Zunächst sei $n = 15$ und die Werte seien geordnet:

$$12.3,\ 9.8,\ 11.5,\ 14.2,\ 10.7,\ 13.4,\ 8.9,\ 9.3,\ 15.1,\ 12.0,\ 11.2,\ 13.7,\ 10.1,\ 9.9,\ 8.4$$

Es gilt

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{170.5}{15} \approx 11.37\%, \qquad \text{Median} = 11.2\%.$$

Begründung/Schritte: - Summe der Renditen: 170.5. - Mittelwert: $170.5/15 \approx 11.3667$. - Die sortierten Werte haben den mittleren Wert (8. von 15) als Median: 11.2.

(b) Die Stichproben-Standardabweichung s sei die Wurzel aus der Stichprobenvarianz $(n-1)^{-1}\sum(x_i - \bar{x})^2$. Die quadrierten Abweichungen summieren sich zu

$$\sum(x_i - \bar{x})^2 \approx 58.6733.$$

Daraus folgt

$$s^2 = \frac{58.6733}{14} \approx 4.191 \quad \Rightarrow \quad s \approx 2.05\%.$$

(c) Quartile: Nach der Sortierung der 15 Werte gilt

$$Q_1 = 9.8, \quad Q_3 = 13.4, \quad \text{IQR} = Q_3 - Q_1 = 3.6.$$

(d) Häufigkeitsverteilung mit Klassenbreite 2 (als [links geschlossen, rechts offen]):

Klasse	absolute Häufigkeit	relative Häufigkeit
8.0–10.0	5	$\frac{5}{15} = 0.333$
10.0–12.0	4	$\frac{4}{15} \approx 0.267$
12.0–14.0	4	$\frac{4}{15} \approx 0.267$
14.0–16.0	2	$\frac{2}{15} \approx 0.133$

Skizze der groben empirischen Verteilung (Häufigkeiten): - 8–10: 5 Beobachtungen - 10–12: 4 Beobachtungen - 12–14: 4 Beobachtungen - 14–16: 2 Beobachtungen Grobe grafische Darstellung (Textform): 8–10 10–12 12–14 14–16

(e) Wirtschaftliche Interpretation: Die Renditewerte liegen zentriert um ca. 11–12%, mit einem moderaten array von Abweichungen. Der obere Tail reicht bis ca. 15

Aufgabe 2.

(a) Eine Urne enthält 3 rote und 2 blaue Kugeln. Es werden zwei Kugeln ohne Zurücklegen gezogen.

$$P(\text{beide rot}) = \frac{\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10} = 0.30,$$

$$P(\text{genau eine rot}) = \frac{\binom{3}{1}\binom{2}{1}}{\binom{5}{2}} = \frac{6}{10} = 0.60,$$

$$P(\text{kein rot}) = \frac{\binom{2}{2}}{\binom{5}{2}} = \frac{1}{10} = 0.10.$$

(b) Sei X die Anzahl roter Kugeln, die gezogen werden. Dann gilt

$$P(X = 0) = 0.10, \quad P(X = 1) = 0.60, \quad P(X = 2) = 0.30.$$

Der Erwartungswert ist

$$E(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) = 0.60 + 0.60 = 1.2.$$

(c) Gesucht ist $P(\text{Rot bei der zweiten Ziehung} \mid \text{Rot bei der ersten Ziehung})$. Nach dem ersten roten Zug bleiben noch 2 rote und 2 blaue Kugeln von insgesamt 4. Also

$$P(\text{Rot}_2 \mid \text{Rot}_1) = \frac{2}{4} = \frac{1}{2} = 0.50.$$

(d) Interpretation: Die Ziehungen sind abhängig: Das Ereignis der ersten roten Kugel beeinflusst die Wahrscheinlichkeiten der zweiten Ziehung. Konkret sinkt der direkte Anteil rot von $\frac{3}{5} = 0.60$ auf $\frac{2}{4} = 0.50$, da eine rote Kugel entfernt wurde. Dies illustriert die Konsequenz von Abhängigkeiten bei Ziehungen ohne Zurücklegen.

Aufgabe 3.

(a) Gegeben seien die Datensätze

$$\{(1,2),(2,4),(3,5),(4,4),(5,5),(6,7)\}.\}$$

Berechnen Sie das Pearson-Korrelationskoeffizient r zwischen X und Y.
Zunächst

$$\bar{X} = \frac{1+2+3+4+5+6}{6} = 3.5, \qquad \bar{Y} = \frac{2+4+5+4+5+7}{6} = \frac{27}{6} = 4.5.$$

Dann

$$\sum(x_i - \bar{X})(y_i - \bar{Y}) = 13.5, \quad \sum(x_i - \bar{X})^2 = 17.5, \quad \sum(y_i - \bar{Y})^2 = 13.5.$$

Daraus folgt

$$r = \frac{\sum(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum(x_i - \bar{X})^2 \sum(y_i - \bar{Y})^2}} = \frac{13.5}{\sqrt{17.5 \cdot 13.5}} \approx \frac{13.5}{15.373} \approx 0.878.$$

(b) Bestimmen Sie die einfache lineare Regression von Y auf X: $Y = a + bX$. Von der Theorie gilt

$$b = \frac{\mathrm{Cov}(X,Y)}{s_X^2} = r\,\frac{S_Y}{S_X}, \quad a = \bar{Y} - b\,\bar{X}.$$

Mit $S_X = \sqrt{17.5} \approx 4.1833$, $S_Y = \sqrt{13.5} \approx 3.6742$ und $r \approx 0.878$ ergibt sich

$$b \approx 0.878\,\frac{3.6742}{4.1833} \approx 0.771, \qquad a \approx 4.5 - 0.771 \cdot 3.5 \approx 1.80.$$

Also

$$Y \approx 1.80 + 0.77\,X.$$

(c) Bestimmen Sie R^2 (Anteil der Varianz von Y, der durch das Modell erklärt wird).

$$R^2 = r^2 \approx (0.878)^2 \approx 0.770 \ (\approx 77.0\%).$$

(d) Interpretation: Die lineare Beziehung zwischen X und Y ist positiv und moderat stark. Ein Anstieg von X um 1 Einheit führt im Mittel zu einer Zunahme von Y um ca. 0.77 Einheiten. Rund 77% der Varianz von Y lassen sich durch das Modell erklären; dies deutet auf eine recht informative lineare Zusammenhangsstruktur hin, wobei jedoch weitere Einflussgrößen bestehen, die hier nicht modelliert sind.

Aufgabe 4.

(a) Eine Häufigkeitsverteilung mit Klassen 0.5–1.0, 1.0–1.5, 1.5–2.0, 2.0–2.5 (links geschlossen, rechts offen):

Klasse	absolute Häufigkeit	relative Häufigkeit
0.5–1.0	5	0.25
1.0–1.5	8	0.40
1.5–2.0	4	0.20
2.0–2.5	3	0.15

Zusammenfassung: - Minimum = 0.6, Maximum = 2.2, Mittelwert 1.335.

(b) Mittelwert, Median, Modus und Varianz der Daten.
- Mittelwert

$$\bar{x} = \frac{26.7}{20} = 1.335 \text{ (Tsd. Euro)}.$$

- Median: Die geordneten Werte lauten 0.6, 0.7, 0.8, 0.8, 0.9, 1.0, 1.0, 1.1, 1.2, 1.3, 1.3, 1.4, 1.4, 1.5, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2. Bei $n = 20$ ist der Median der Durchschnitt der 10. und 11. Beobachtung:

$$\text{Median} = \frac{1.3 + 1.3}{2} = 1.30 \text{ (Tsd. Euro)}.$$

- Modus: Mehrfachmodus möglich. Werte mit Häufigkeit 2 sind: 0.8, 1.0, 1.3, 1.4. Also multimodal mit den Modi 0.8, 1.0, 1.3, 1.4 (jeweils in Tausend Euro).
- Varianz (Stichprobenvarianz): Normale Berechnung liefert

$$S^2 \approx 0.236 \quad \Rightarrow \quad s \approx 0.486 \text{ (Tsd. Euro)}.$$

Hinweis: Bei der Berechnung wurde der Mittelwert als 1.335 verwendet; die Einheiten sind Tausend Euro.

(c) Interpretation der Verteilung: Der Durchschnitt liegt bei ca. 1.34 Tsd. Euro, Median ca. 1.30 Tsd. Euro. Die Verteilung ist relativ breit gestreut (Streuung ca. 0.49 Tsd. Euro). Typische monatliche Ausgaben liegen also um ca. 1.2–1.5 Tsd. Euro, mit Ausreißern nach oben bis ca. 2.2 Tsd. Euro.

(d) Diskussion, wie Ausreißer das Ergebnis beeinflussen könnten: - Der Mittelwert reagiert empfindlich auf Ausreißer (z. B. 2.2 oder 0.6), während der Median robuster ist. - Die Varianz (und damit die Standardabweichung) wird ebenfalls durch Extremwerte erhöht. - Optionen zur Robustheit: Verwendung des Medians, des interquartile Bereichs (IQR) oder eines trimmed mean; alternative Maße wie MAD (Median absolute deviation) könnten sinnvoll sein.