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.

Gegeben sei eine Stichprobe von $n = 12$ Unternehmens-Umsätzen (in Tausend Euro):

$$102,\ 118,\ 95,\ 110,\ 132,\ 125,\ 98,\ 105,\ 120,\ 115,\ 107,\ 128.$$

(a) Berechnen Sie den arithmetischen Mittelwert dieser Daten.

(b) Bestimmen Sie die Standardabweichung der Umsätze.

(c) Bestimmen Sie den Median der Daten.

(d) Geben Sie die Spannweite sowie den Interquartilsabstand (IQR) der Daten an und interpretieren Sie grob die Verteilung bezüglich Symmetrie.

(e) Diskutieren Sie kurz, ob die Verteilung eher links- oder rechtsschief ist und begründen Sie Ihre Einschätzung anhand Mittelwert, Median und Verteilungsform.

Aufgabe 2.

Eine Urne A enthält drei rote Kugeln und zwei blaue Kugeln. Es werden zwei Kugeln ohne Zurücklegen gezogen.

(a) Bestimmen Sie die Wahrscheinlichkeiten

$$P(\text{rot,rot}), \quad P(\text{rot,blau}), \quad P(\text{blau,rot}), \quad P(\text{blau,blau}).$$

(b) Berechnen Sie die bedingte Wahrscheinlichkeit P(second ist rot | first ist rot).

(c) Bestimmen Sie die erwartete Anzahl roter Kugeln in den zwei Zügen.

(d) Stellen Sie die Varianz der Anzahl roter Kugeln in den zwei Zügen aus.

Aufgabe 3.

Gegeben seien die Paare (x_i, y_i) mit

$$(1,2),\ (2,4),\ (3,5),\ (4,4),\ (5,5).$$

(a) Berechnen Sie den Pearson-Korrelationskoeffizienten r zwischen x und y.

(b) Bestimmen Sie die lineare Regression der Form $y = a + bx$. Geben Sie die Parameterwerte an.

(c) Interpretieren Sie die Steigung b und beantworten Sie, ob ein steigender Zusammenhang besteht.

(d) Geben Sie den Bestimmtheitskoeffizienten R^2 an und interpretieren Sie ihn.

Aufgabe 4.

Eine Zufallsvariable X sei die Augenzahl eines fairen Würfels. Es seien n = 30 unabhängige Ziehungen zu betrachten.
(a) Verwenden Sie den zentralen Grenzwertsatz, um ungefähr die Wahrscheinlichkeit zu berechnen, dass der Stichprobenmittelwert $\bar{X}$ die Schwelle $\bar{X} > 4.2$ überschreitet.

(b) Sei nun Y die Anzahl der Erfolge einer Bernoulli-Variante mit $p = 0.6$ in einer Stichprobe der Größe $n = 12$. Berechnen Sie die Varianz von Y.

(c) Fassen Sie kurz zusammen, warum die Normalapproximation bei Teil (a) sinnvoll ist und nennen Sie eine Voraussetzung, die erfüllt sein muss.

Lösungen

Bearbeitungszeit: 120 Minuten.

Aufgabe 1. Lösung.

(a) Berechnen Sie den arithmetischen Mittelwert der Daten.

$$\bar{x} \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i \;=\; \frac{1}{12}\,(102+118+95+110+132+125+98+105+120+115+107+128) = \frac{1355}{12} \approx 112{,}92\;.$$

(b) Bestimmen Sie die Standardabweichung der Umsätze.
Mit $\sum x_i^2 = 154\,589$ und $\mu = \bar{x} = 1355/12$ gilt

$$S = \sum_{i=1}^{n}(x_i-\mu)^2 \;=\; \sum x_i^2 - n\mu^2 \;=\; 154\,589 - \frac{1355^2}{12} \;=\; \frac{19{,}043}{12},$$

und

$$s^2 = \frac{S}{n-1} = \frac{19043}{132} \approx 144{,}27, \qquad s = \sqrt{s^2} \;=\; \sqrt{\frac{19043}{132}} \approx 12{,}01.$$

(c) Bestimmen Sie den Median der Daten.
Sortiert: 95, 98, 102, 105, 107, 110, 115, 118, 120, 125, 128, 132.
Bei $n = 12$ liegt der Median als Mittelwert der beiden zentralen Werte

$$\text{Median} \;=\; \frac{110+115}{2} \;=\; 112{,}5.$$

(d) Geben Sie die Spannweite sowie den Interquartilsabstand (IQR) der Daten an und interpretieren Sie grob die Verteilung bezüglich Symmetrie.
Spannweite: $\max - \min = 132 - 95 = 37$.
Q1 und Q3 (Definition als Median der unteren bzw. oberen Halbgruppe): - Untere Hälfte: 95, 98, 102, 105, 107, 110 $\Rightarrow$ Q1 = $\frac{102+105}{2}$ = 103,5. - Obere Hälfte: 115, 118, 120, 125, 128, 132 $\Rightarrow$ Q3 = $\frac{120+125}{2}$ = 122,5.
IQR = Q3 - Q1 = 122,5 - 103,5 = 19,0.
Interpretation: Die Verteilung weist eine moderate Streuung (IQR 19) auf; der Median liegt nahe am Mittelwert (siehe unten). Die Höhenspannweite durch wenige größere Werte (z. B. 132) deutet auf eine leichte Rechts-Schiefe hin.

(e) Diskutieren Sie kurz, ob die Verteilung eher links- oder rechtsschief ist und begründen Sie Ihre Einschätzung anhand Mittelwert, Median und Verteilungsform.
Der Mittelwert $\bar{x} \approx 112{,}92$ liegt leicht über dem Median 112,5. Hinzu kommen einige größere Werte (z. B. 128, 132) in der oberen Hälfte, wodurch sich eine leichte Rechts-Schiefe ergibt. Die Verteilung ist demnach tendenziell rechtsschief.

Aufgabe 2. Lösung.

Eine Urne A enthält drei rote Kugeln und zwei blaue Kugeln. Es werden zwei Kugeln ohne Zurücklegen gezogen.

(a) Bestimmen Sie die Wahrscheinlichkeiten P(rot,rot), P(rot,blau), P(blau,rot), P(blau,blau).
- $P(\text{rot,rot}) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}$. - $P(\text{rot,blau}) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}$. - $P(\text{blau,rot}) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10}$. - $P(\text{blau,blau}) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}$.

(b) Berechnen Sie die bedingte Wahrscheinlichkeit P(second ist rot | first ist rot).

$$P(\text{second rot} \mid \text{first rot}) = \frac{P(\text{rot,rot})}{P(\text{rot})} = \frac{3/10}{3/5} = \frac{1}{2}.$$

(c) Bestimmen Sie die erwartete Anzahl roter Kugeln in den zwei Zügen.
Seien X_1, X_2 Indikatorvariablen für rote Kugeln der ersten bzw. zweiten Ziehung. Dann

$$E[X_1] = P(\text{rot}) = \frac{3}{5}, \quad E[X_2] = P(\text{rot}) = \frac{3}{5}, \quad E[X] = E[X_1 + X_2] = E[X_1] + E[X_2] = \frac{6}{5} = 1{,}2.$$

(d) Stellen Sie die Varianz der Anzahl roter Kugeln in den zwei Zügen aus.
Variablen X_1, X_2 mit $p = \frac{3}{5}$. Dann

$$\operatorname{Var}(X_1) = p(1-p) = \frac{3}{5} \cdot \frac{2}{5} = \frac{6}{25}, \quad \operatorname{Cov}(X_1, X_2) = P(\text{beide rot}) - P(\text{rot})^2 = \frac{3}{10} - \left(\frac{3}{5}\right)^2 = -\frac{3}{50},$$

und

$$\operatorname{Var}(X) = \operatorname{Var}(X_1) + \operatorname{Var}(X_2) + 2\operatorname{Cov}(X_1, X_2) = \frac{6}{25} + \frac{6}{25} + 2\left(-\frac{3}{50}\right) = \frac{9}{25} = 0{,}36.$$

Aufgabe 3. Lösung.

Gegeben seien die Paare (x_i, y_i) mit

$$(1,2),\ (2,4),\ (3,5),\ (4,4),\ (5,5).$$

(a) Berechnen Sie den Pearson-Korrelationskoeffizienten r zwischen x und y.
Mit x_i als $1, 2, 3, 4, 5$ und y_i als $2, 4, 5, 4, 5$ gilt

$$\bar{x} = 3, \quad \bar{y} = 4, \quad \sum (x_i - \bar{x})(y_i - \bar{y}) = 6, \quad \sum (x_i - \bar{x})^2 = 10, \quad \sum (y_i - \bar{y})^2 = 6.$$

Daraus

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} = \frac{6}{\sqrt{10 \cdot 6}} = \frac{6}{\sqrt{60}} = \sqrt{\frac{3}{5}} \approx 0{,}7746.$$

(b) Bestimmen Sie die lineare Regression der Form $y = a + bx$. Geben Sie die Parameterwerte an.
Mit $b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{6}{10} = \frac{3}{5}$ und $a = \bar{y} - b\,\bar{x} = 4 - \frac{3}{5} \cdot 3 = \frac{11}{5}$, erhält man

$$y = a + bx = \frac{11}{5} + \frac{3}{5}x \ = \ 2{,}2 + 0{,}6\,x.$$

(c) Interpretieren Sie die Steigung b und beantworten Sie, ob ein steigender Zusammenhang besteht.
Die Steigung $b = 3/5 = 0{,}6$ bedeutet: Bei einer Zunahme von x um 1 Einheiten steigt der vorhergesagte y um 0,6 Einheiten. Damit besteht ein positiver, steigender Zusammenhang.

(d) Geben Sie den Bestimmtheitskoeffizienten R^2 an und interpretieren Sie ihn.

$$R^2 = r^2 = \left(\sqrt{\frac{3}{5}}\right)^2 = \frac{3}{5} = 0{,}60.$$

Interpretation: Etwa 60

Aufgabe 4. Lösung.

Eine Zufallsvariable X sei die Augenzahl eines fairen Würfels. Es seien $n = 30$ unabhängige Ziehungen zu betrachten.

(a) Verwenden Sie den zentralen Grenzwertsatz, um ungefähr die Wahrscheinlichkeit zu berechnen, dass der Stichprobenmittelwert $\bar{X}$ die Schwelle $\bar{X} > 4{,}2$ überschreitet.

Für den Würfel gilt $\mathbb{E}[X] = \mu = 3{,}5$ und $\mathrm{Var}(X) = \frac{6^2-1}{12} = \frac{35}{12}$. Die Stichprobenmitte $\bar{X}$ hat dann $\mathbb{E}[\bar{X}] = \mu = 3{,}5$ und

$$\mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(X)}{n} = \frac{35}{12 \cdot 30} = \frac{35}{360}.$$

Damit

$$Z \ = \ \frac{\bar{X} - \mu}{\sqrt{\mathrm{Var}(\bar{X})}} \ = \ \frac{4{,}2 - 3{,}5}{\sqrt{35/360}} \ = \ \frac{0{,}7}{\sqrt{0{,}09722\ldots}} \approx 2{,}25.$$

Daraus folgt

$$P(\bar{X} > 4{,}2) = P(Z > 2{,}25) \approx 0{,}012 \text{ (ca. 1.2\%)}.$$

(b) Sei nun Y die Anzahl der Erfolge einer Bernoulli-Variante mit $p = 0{,}6$ in einer Stichprobe der Größe $n = 12$. Berechnen Sie die Varianz von Y.

$Y \sim \mathrm{Binomial}(n = 12, p = 0{,}6)$, daher

$$\mathrm{Var}(Y) = np(1-p) = 12 \cdot 0{,}6 \cdot 0{,}4 = 2{,}88.$$

(c) Fassen Sie kurz zusammen, warum die Normalapproximation bei Teil (a) sinnvoll ist und nennen Sie eine Voraussetzung, die erfüllt sein muss.

Die Verteilung des Stichprobenmittelwertes $\bar{X}$ konvergiert gemäß dem Zentralen Grenzwertsatz gegen eine Normalverteilung, sobald die Unabhängigkeit der Ziehungen gegeben ist und die Varianz der Grundverteilung endlich ist. Für $n = 30$ ist die Normalapproximation in der Praxis geeignet. Eine zentrale Voraussetzung ist insbesondere die Unabhängigkeit der Beobachtungen (bzw. der Ziehungen) und eine endliche Varianz der Grundverteilung.