Statistik Klausur 1, Dr. Martin Franzen

Wintersemester 2024
04.11.24 11:00 Uhr - 13:00 Uhr
Hauptgebäude 273
Studiengänge UX, ID
Dauer 90min

Punkte

- **Aufgabe A: Arithmetisches Mittel** (4 Punkte)
- **Aufgabe B: Median** (4 Punkte)
- **Aufgabe C: Modus** (4 Punkte)
- **Aufgabe D: Varianz, Standardabweichung** (7 Punkte)
- **Aufgabe E: Skalenniveaus** (4 Punkte)
- **Aufgabe F: Pearson Korrelationskoeffizient** (10 Punkte)

Bewertung

- alle Ergebnisse, Rechenwege, Begründungen richtig → 33 Punkte

Hilfsmittel

- 1 Blatt DIN A4 Papier, Taschenrechner (kein GTR)

Abgabe

- Namen, Matrikelnummer auf jedes abgegebene Blatt schreiben

Aufgabe A: Arithmetisches Mittel (4 Punkte)

Sei $(x_1, \ldots, x_n) \in \mathbb{R}^n$, $n \in \mathbb{N}$ ein Datensatz. Dann berechnen wir das arithmetische Mittel $\overline{x}$ wie folgt

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

a) Sei der gegebene Datensatz $(1, 2, 3, 4, 5, 6, 7) \in \mathbb{R}^7$. Berechne $\overline{x}_1$ und gebe einen Rechenweg an.

b) Sei der gegebene Datensatz $(7, -7, 7, -7, 7, -7, 7) \in \mathbb{R}^7$. Berechne $\overline{x}_2$ und gebe einen Rechenweg an.

c) Sei der gegebene Datensatz $\big((7+7), (7-7), (7*7), (7:7)\big) \in \mathbb{R}^4$. Berechne $\overline{x}_3$ und gebe einen Rechenweg an.

d) Sei der gegebene Datensatz

$$3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73 \in \mathbb{R}^{20}.$$

Berechne $\overline{x}_4$ und gebe einen Rechenweg an.

Aufgabe B: Median (4 Punkte)

Sei $(x_1, \ldots, x_n) \in \mathbb{R}^n$, $n \in \mathbb{N}$ ein geordneter Datensatz, d.h. $x_1 \leq x_2 \leq \ldots \leq x_n$. Dann berechnen wir den Median $\overline{Md}$ wie folgt

- $\overline{Md} = x_{(n+1)/2}$, falls n ungerade
- $\overline{Md} = \left(x_{n/2} + x_{n/2+1}\right)/2$, falls n gerade

a) Sei der gegebene Datensatz $(1, 2, 3, 4, 5, 6, 7) \in \mathbb{R}^7$. Berechne $\overline{Md_1}$ und gebe einen Rechenweg an.

b) Sei der gegebene Datensatz $(7, -7, 7, -7, 7, -7, 7) \in \mathbb{R}^7$. Berechne $\overline{Md_2}$ und gebe einen Rechenweg an.

c) Sei der gegebene Datensatz $\left((7+7), (7-7), (7*7), (7:7)\right) \in \mathbb{R}^4$. Berechne $\overline{Md_3}$ und gebe einen Rechenweg an.

d) Sei der gegebene Datensatz

$$3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73 \in \mathbb{R}^{20}.$$

Berechne $\overline{Md_4}$ und gebe einen Rechenweg an.

Aufgabe C: Modus (4 Punkte)

Sei $(x_1, \dots, x_n) \in \mathbb{R}^n$, $n \in \mathbb{N}$ ein Datensatz. Dann berechnen wir den Modus $\overline{Mo} \in \mathbb{R}$ bzw. die Menge der Modi $\overline{\mathbf{Mo}} \subset \{x_1, \dots, x_n\}$ wie folgt - wir unterscheiden dabei drei Fälle, wobei die Funktion # die Anzahl der Elemente einer gegebenen Menge zurückgibt

- Fall $\#\overline{\mathbf{Mo}} = 0$, alle Daten kommen gleich häufig oder jedes Datum kommt genau einmal vor: es gibt keinen Modus $\overline{Mo}$ und die Menge der Modi $\overline{\mathbf{Mo}}$ besteht aus der leeren Menge $\emptyset$

- Fall $\#\overline{\mathbf{Mo}} = 1$, ein Datum x_i für ein $i \in \{1, \dots, n\}$ kommt häufiger als alle anderen Daten vor: der Modus $\overline{\mathrm{Mo}}$ ist das häufigste Datum x_i und die Menge der Modi ist die einelementige Menge $\overline{\mathbf{Mo}} = \{x_i\}$

- Fall $\#\overline{\mathbf{Mo}} > 1$, zwei oder mehr Daten kommen gleich häufig und häufiger als alle anderen Daten vor: die Menge der Modis $\overline{\mathbf{Mo}}$ besteht aus einer Teilmenge $\overline{\mathbf{Mo}} \subset \{x_1, \dots, x_n\}$

a) Sei der gegebene Datensatz $(1, 2, 3, 4, 5, 6, 7) \in \mathbb{R}^7$. Berechne die Anzahl der Elemente $\#\overline{\mathbf{Mo}}_1$. Begründe.

b) Sei der gegebene Datensatz $(7, -7, 7, -7, 7, -7, 7) \in \mathbb{R}^7$. Berechne die Anzahl der Elemente $\#\overline{\mathbf{Mo}}_2$. Begründe.

c) Sei der gegebene Datensatz $\big((7 + 7), (7 - 7), (7 * 7), (7 : 7)\big) \in \mathbb{R}^4$. Berechne die Anzahl der Elemente $\#\overline{\mathbf{Mo}}_3$. Begründe.

d) Sei der gegebene Datensatz

$$3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73 \in \mathbb{R}^{20}.$$

Berechne die Anzahl der Elemente $\#\overline{\mathbf{Mo}}_4$. Begründe.

Aufgabe D: Varianz, Standardabweichung (7 Punkte)
Sei $(x_1, \ldots, x_n) \in \mathbb{R}^n$, $n \in \mathbb{N}$ ein Datensatz. Dann berechnen wir die Varianz s^2 bzw. die Standardabweichung s wie folgt, wobei $\bar{x}$ das arithmetische Mittel der Daten ist

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{bzw.}$$
$$s = \sqrt{s^2}$$

a) Sei der gegebene Datensatz $(1, 3, 5) \in \mathbb{R}^3$. Berechne s_1^2 sowie s_1 und gebe einen Rechenweg an.

b) Sei der gegebene Datensatz $(1, 4, 7) \in \mathbb{R}^3$. Berechne s_2^2 sowie s_2 und gebe einen Rechenweg an.

c) Sei der gegebene Datensatz $(2, 4, 6) \in \mathbb{R}^3$. Berechne s_3^2 sowie s_3 und gebe einen Rechenweg an.

d) Sei der gegebene Datensatz $(3, 6, 9, 12) \in \mathbb{R}^4$. Berechne s_4^2 sowie s_4 und gebe einen Rechenweg an.

Aufgabe E: Skalenniveaus (4 Punkte)

- Nominalskala: Kategorische Daten, die keine natürliche Reihenfolge oder Abstand haben.
- Ordinalskala: Kategorische Daten, die eine Reihenfolge haben, aber bei denen die Abstände zwischen den Werten nicht gleichmäßig oder bedeutsam sind.
- Intervallskala: Numerische Daten, die eine konstante Differenz haben, jedoch keinen absoluten Nullpunkt.
- Verhältnisskala: Numerische Daten, die sowohl eine konstante Differenz als auch einen absoluten Nullpunkt haben.

a) Sei der Datensatz *Augenfarben einer Gruppe von Personen*
(braun, blau, grün, braun, grau, blau)

Welches Skalenniveau hat dieser Datensatz? Begründe.

b) Sei der Datensatz *Zufriedenheit der Kunden einer Firma auf einer Skala von Eins bis Fünf*
(1, 4, 3, 5, 2, 4, 3)

Welches Skalenniveau hat dieser Datensatz? Begründe.

c) Sei der Datensatz *Gehälter von Angestellten in Euro pro Monat*
(2500, 3100, 4000, 2750, 3200)

Welches Skalenniveau hat dieser Datensatz? Begründe.

d) Sei der Datensatz *Geburtsjahre von Personen in einer Familie*
(1980, 2001, 1975, 1993, 1965)

Welches Skalenniveau hat dieser Datensatz? Begründe.

Aufgabe F: Pearson Korrelationskoeffizient (10 Punkte)

Sei $((x_1, y_1), \dots, (x_n, y_n)) \in \mathbb{R}^{2n}$, $n \in \mathbb{N}$ ein gegebener Datensatz.
Seien $\bar{x}$ das arithemtische Mittel von $(x_1, \dots, x_n) \in \mathbb{R}^n$ und $\bar{y}$ das arithmetische Mittel von $(y_1, \dots, y_n) \in \mathbb{R}^n$.

Dann berechnen wir den Pearson Korrelationskoeffizient $r \in [-1, 1]$ wie folgt

$$
\begin{aligned}
\Delta_{xy} &:= \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \\
\Delta_{x} &:= \sum_{i=1}^{n} (x_i - \bar{x})^2 \\
\Delta_{y} &:= \sum_{i=1}^{n} (y_i - \bar{y})^2 \\
r &:= \frac{\Delta_{xy}}{\sqrt{\Delta_x \Delta_y}}
\end{aligned}
$$

Bemerkung: Für $|r| = 1$ liegen alle Datenpunkte $((x_1, y_1), \dots, (x_n, y_n)) \in \mathbb{R}^{2n}$ auf einer Geraden.

a) Sei der Datensatz gegeben durch

Alter	Einkommen
1	0
5	9
9	0

Berechnen Sie den Pearson-Korrelationskoeffizienten r_1 zwischen dem Alter und dem Einkommen (auf die zweite Nachkommastelle runden) und geben Sie einen Rechenweg an.

b) Sei der Datensatz gegeben durch

Stunden gelernt	Punkte im Test
1	1
2	-2
3	1

Berechnen Sie den Pearson-Korrelationskoeffizienten r_2 zwischen den gelernten Stunden und den Punkten im Test (auf die zweite Nachkommastelle runden) und geben Sie einen Rechenweg an.

c) Sei der Datensatz gegeben durch

Stunden gelernt	Punkte im Test
7	1
14	2
21	1

Berechnen Sie den Pearson-Korrelationskoeffizienten r_3 zwischen den gelernten Stunden und den Punkten im Test (auf die zweite Nachkommastelle runden) und geben Sie einen Rechenweg an.

d) Sei der Datensatz gegeben durch

Körpergröße (cm)	Gewicht (kg)
7	7
14	14
21	21

Berechnen Sie den Pearson-Korrelationskoeffizienten r_4 zwischen der Körpergröße und dem Gewicht (auf die zweite Nachkommastelle runden) und geben Sie einen Rechenweg an.