

Statistik Klausur 1, Dr. Martin Franzen

Wintersemester 2023

31.10.23 10:00 - 12:00

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 auf jedes Blatt schreiben

1. Ergebnisse → **DigiExam**
während der DigiExam-Klausur
2. Rechenwege → **Canvas**
nach Ende der DigiExam-Klausur

Aufgabe A: Arithmetisches Mittel (4 Punkte)

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

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

- a) Sei der gegebene Datensatz $(1, 2, 3, 4, 5) \in \mathbb{R}^5$. Berechne \bar{x}_1 und gib einen Rechenweg an!
- b) Sei der gegebene Datensatz $(5, -5, 5, -5, 5) \in \mathbb{R}^5$. Berechne \bar{x}_2 und gib einen Rechenweg an!
- c) Sei der gegebene Datensatz $(20, 40, 80, 160, 80) \in \mathbb{R}^5$. Berechne \bar{x}_3 und gib einen Rechenweg an!
- d) Sei der gegebene Datensatz $(10, 100, 1000, 10000, 100000) \in \mathbb{R}^5$. Berechne \bar{x}_4 und gib einen Rechenweg an!

DigiExam: $A1 = \bar{x}_1$, $A2 = \bar{x}_2$, $A3 = \bar{x}_3$, $A4 = \bar{x}_4$

Aufgabe B: Median (4 Punkte)

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

- $\overline{Md} = x_{(n+1)/2}$, falls n ungerade
 - $\overline{Md} = (x_{n/2} + x_{n/2+1}) / 2$, falls n gerade
- a) Sei der gegebene Datensatz $(1, 2, 3, 4, 5) \in \mathbb{R}^5$. Berechne \overline{Md}_1 und gib einen Rechenweg an!
- b) Sei der gegebene Datensatz $(5, -5, 5, -5, 5) \in \mathbb{R}^5$. Berechne \overline{Md}_2 und gib einen Rechenweg an!
- c) Sei der gegebene Datensatz $(20, 40, 80, 160, 80) \in \mathbb{R}^5$. Berechne \overline{Md}_3 und gib einen Rechenweg an!
- d) Sei der gegebene Datensatz $(10, 100, 1000, 10000, 100000) \in \mathbb{R}^5$. Berechne \overline{Md}_4 und gib einen Rechenweg an!

DigiExam: $B1 = \overline{Md}_1$, $B2 = \overline{Md}_2$, $B3 = \overline{Md}_3$, $B4 = \overline{Md}_4$

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 $\overline{Mo} \subset \{x_1, \dots, x_n\}$ wie folgt - wir unterscheiden dabei drei Fälle, wobei die Funktion $\#$ die Anzahl der Elemente einer gegebenen Menge zurückgibt

- Fall $\#\overline{Mo} = 0$, alle Daten kommen gleich häufig oder jedes Datum kommt genau einmal vor: es gibt keinen Modus \overline{Mo} und die Menge der Modi \overline{Mo} besteht aus der leeren Menge \emptyset
 - Fall $\#\overline{Mo} = 1$, ein Datum x_i für ein $i \in \{1, \dots, n\}$ kommt häufiger als alle anderen Daten vor: der Modus \overline{Mo} ist das häufigste Datum x_i und die Menge der Modi ist die einelementige Menge $\overline{Mo} = \{x_i\}$
 - Fall $\#\overline{Mo} > 1$, zwei oder mehr Daten kommen gleich häufig und häufiger als alle anderen Daten vor: die Menge der Modis \overline{Mo} besteht aus einer Teilmenge $\overline{Mo} \subset \{x_1, \dots, x_n\}$
- a) Sei der gegebene Datensatz $(1, 2, 3, 4, 5) \in \mathbb{R}^5$. Berechne die Anzahl der Elemente $\#\overline{Mo}_1$. Begründe!
- b) Sei der gegebene Datensatz $(5, -5, 5, -5, 5) \in \mathbb{R}^5$. Berechne die Anzahl der Elemente $\#\overline{Mo}_2$. Begründe!
- c) Sei der gegebene Datensatz $(20, 40, 80, 160, 80) \in \mathbb{R}^5$. Berechne die Anzahl der Elemente $\#\overline{Mo}_3$. Begründe!
- d) Sei der gegebene Datensatz $(10, 100, 1000, 10000, 100000) \in \mathbb{R}^5$. Berechne die Anzahl der Elemente $\#\overline{Mo}_4$. Begründe!

DigiExam: $C1 = \#\overline{Mo}_1$, $C2 = \#\overline{Mo}_2$, $C3 = \#\overline{Mo}_3$, $C4 = \#\overline{Mo}_4$

Aufgabe D: Varianz, Standardabweichung (7 Punkte)

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

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

- a) Sei der gegebene Datensatz $(1, 2, 3, 4, 5) \in \mathbb{R}^5$. Berechne s_1^2 sowie s_1 und gib einen Rechenweg an!
- b) Sei der gegebene Datensatz $(5, -5, 5, -5, 5) \in \mathbb{R}^5$. Berechne s_2^2 sowie s_2 und gib einen Rechenweg an!
- c) Sei der gegebene Datensatz $(20, 40, 80, 160, 80) \in \mathbb{R}^5$. Berechne s_3^2 sowie s_3 und gib einen Rechenweg an!
- d) Sei der gegebene Datensatz $(10, 100, 1000, 10000, 100000) \in \mathbb{R}^5$. Berechne s_4^2 sowie s_4 und gib einen Rechenweg an!

$$\text{DigiExam: } D1 = 2 s_1^2, D2 = s_2^2, D3 = s_3^2, D4 = s_4^2$$

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 *Geschlechter von Teilnehmern in einem Kurs*
(männlich, weiblich, weiblich, männlich, männlich, divers)

Welches Skalenniveau hat dieser Datensatz? Begründe!

b) Sei der Datensatz *Abschlussnoten einer Klasse*
(Sehr Gut, Gut, Befriedigend, Ausreichend, Mangelhaft, Ungenügend)

Welches Skalenniveau hat dieser Datensatz? Begründe!

c) Sei der Datensatz *Datensatz Temperaturen einer Stadt in Grad Celsius an verschiedenen Tagen*
(22,5; 19,8; 20,1; 21,9; 23,0)

Welches Skalenniveau hat dieser Datensatz? Begründe!

d) Sei der Datensatz *Gewicht von fünf verschiedenen Äpfeln in Gramm*
(120, 135, 150, 125, 145)

Welches Skalenniveau hat dieser Datensatz? Begründe!

DigiExam:

Sei Nominalskala - 1, Ordinalskala - 2, Intervallskala - 3, Verhältnisskala - 4

E1 = Antwort von a): 1, 2, 3 oder 4,

E2 = Antwort von b): 1, 2, 3 oder 4,

E3 = Antwort von c): 1, 2, 3 oder 4,

E4 = Antwort von d): 1, 2, 3 oder 4,

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 arithmetische 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
0	0
5	10
10	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	4
2	5
3	6

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 |
|-----------------|----------------|
| 1 | 6 |
| 2 | 5 |
| 3 | 4 |

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) |
|------------------|--------------|
| 1 | 1 |
| 2 | 2 |
| 3 | 1 |

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

DigiExam: $F1 = |r_1|$, $F2 = |r_2|$, $F3 = |r_3|$, $F4 = |r_4|$