

---

## Übungsserie 2: 29.08.2024

Thema «Statistik mit Python» - Grundlagen: Lagemasse, Streumasse und Visualisierungen

---

### Teil 1: Lagemasse

#### Aufgabe 1: Mittelwert berechnen

**Theorie:** Der Mittelwert (arithmetisches Mittel) ist das am häufigsten verwendete Mass für die zentrale Tendenz. Er wird berechnet, indem man die Summe aller Werte durch die Anzahl der Werte teilt.

**Datensatz:** [12, 15, 23, 10, 18, 20, 14]

#### Aufgabe 2: Median berechnen

**Theorie:** Der Median ist der mittlere Wert eines geordneten Datensatzes und teilt die Daten in zwei gleich grosse Hälften. Er ist robuster gegenüber Ausreissern als der Mittelwert.

**Formel:**

- Für ungerade Anzahl an Werten: Der Median ist der mittlere Wert.
- Für gerade Anzahl an Werten: Der Median ist der Durchschnitt der beiden mittleren Werte.

**Datensatz:** [7, 9, 13, 12, 8, 11, 10]

#### Aufgabe 3: Modus berechnen

**Theorie:** Der Modus ist der Wert, der in einem Datensatz am häufigsten vorkommt. Es kann mehr als einen Modus geben, wenn mehrere Werte die gleiche Häufigkeit haben.

**Formel:**

- Der Modus ist der Wert mit der höchsten Häufigkeit im Datensatz.

**Datensatz:** [5, 8, 9, 8, 10, 5, 7]

#### Aufgabe 4: Gewichteten Mittelwert berechnen

**Theorie:** Der gewichtete Mittelwert ist eine Variante des Mittelwerts, bei der einige Datenpunkte mehr Einfluss (Gewicht) haben als andere. Er wird häufig verwendet, wenn bestimmte Werte wichtiger sind als andere.

**Datensatz:** Werte: [3, 7, 9, 15], Gewichte: [2, 1, 3, 4]

### Aufgabe 5: Mittelwert und Median vergleichen

**Theorie:** Der Mittelwert und der Median sind beides Lagemasse, aber sie reagieren unterschiedlich auf Ausreisser. Der Mittelwert wird durch extreme Werte stärker beeinflusst, während der Median robuster gegenüber Ausreissern ist.

**Datensatz:** [21, 19, 15, 22, 20, 24, 18]

## Teil 2: Streuungsmasse

### Aufgabe 6: Spannweite berechnen

**Theorie:** Die Spannweite ist das einfachste Streuungsmass und wird berechnet, indem man den kleinsten Wert vom grössten Wert des Datensatzes subtrahiert. Sie gibt die Gesamtstreuung der Daten an.

**Datensatz:** [45, 22, 36, 48, 33, 25, 50]

### Aufgabe 7: Varianz berechnen

**Theorie:** Die Varianz ist ein Mass dafür, wie weit die einzelnen Werte eines Datensatzes im Durchschnitt vom Mittelwert entfernt sind. Sie wird berechnet, indem man die Abweichungen der einzelnen Werte vom Mittelwert quadriert und den Durchschnitt dieser Quadrate bildet.

**Datensatz:** [3, 7, 4, 6, 8, 5, 9]

### Aufgabe 8a: Standardabweichung berechnen

**Theorie:** Die Standardabweichung ist die Quadratwurzel der Varianz und gibt die durchschnittliche Abweichung der Datenwerte vom Mittelwert an. Sie ist ein gängiges Mass für die Streuung der Daten.

**Datensatz:** [12, 15, 14, 16, 11, 13, 17]

### Aufgabe 8b: Standardabweichung für Stichprobendaten berechnen

**Theorie:** Die Standardabweichung einer Stichprobe wird ähnlich wie die Standardabweichung der gesamten Population berechnet, jedoch mit einer kleinen Anpassung im Nenner. Diese Anpassung, bei der man im Nenner 1 vom Stichprobenumfang subtrahiert, wird als Bessel-Korrektur bezeichnet. Diese Korrektur wird durchgeführt, um eine Verzerrung in der Schätzung der Varianz und der Standardabweichung zu vermeiden.

Warum wird 1 im Nenner subtrahiert?

Der Grund für die Subtraktion von 1 im Nenner (also die Verwendung von  $n-1$  anstelle von  $n$ ) liegt darin, dass eine Stichprobe eine unvollständige Abbildung der Population ist. Wenn man den Mittelwert einer Stichprobe berechnet, nähert er sich tendenziell dem tatsächlichen Populationsmittelwert an, aber er ist nicht exakt. Diese Korrektur sorgt dafür, dass die Berechnung der Varianz (und damit auch der Standardabweichung) nicht systematisch unterschätzt wird. Durch das

Dividieren durch  $n-1$  wird diese Schätzung „entzerrt“, um genauer die Populationsparameter widerzuspiegeln.

**Datensatz:** [12, 15, 14, 16, 11, 13, 17]

#### **Aufgabe 9: Quartilsabstand berechnen**

**Theorie:** Der Quartilsabstand (Interquartilsabstand, IQR) ist ein Streuungsmass, das die Spannweite der mittleren 50% der Daten angibt. Er wird berechnet, indem man das erste Quartil vom dritten Quartil subtrahiert.

**Datensatz:** [5, 9, 14, 7, 11, 8, 10]

#### **Aufgabe 10: Mittlere absolute Abweichung berechnen**

**Theorie:** Die mittlere absolute Abweichung (MAD) ist das durchschnittliche absolute Mass, wie weit die Werte eines Datensatzes von ihrem Mittelwert entfernt sind. Sie ist ein alternatives Streuungsmass zur Standardabweichung.

**Datensatz:** [22, 27, 24, 19, 25, 21, 23]