
Auswertung statistischer Daten 2

Dr. Elke Warmuth

Sommersemester 2018

Kenngößen von Häufigkeitsverteilungen

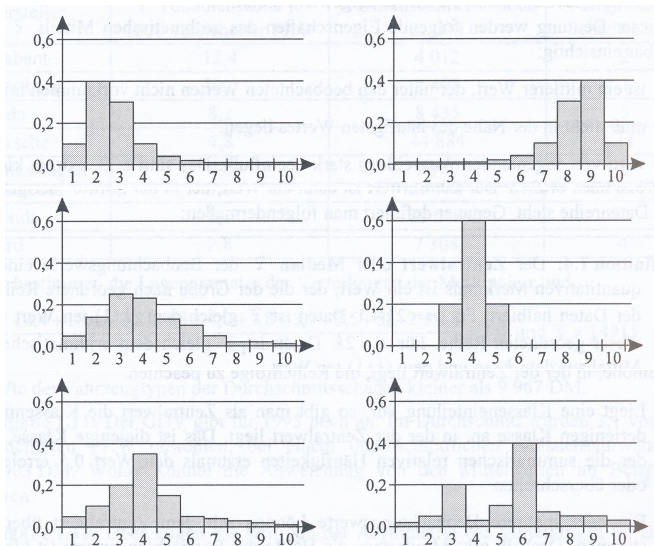
Mittelwerte

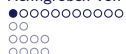
Streuungsmaße 1

Boxplot

Streuungsmaße 2

Glockenförmige Verteilungen





Mittelwerte

- ▶ Ziel: Lage der Verteilung auf der Achse beschreiben
- ▶ nur bei quantitativen Merkmalen möglich
- ▶ hier:
 - ▶ arithmetisches Mittel (Durchschnitt) \bar{x}
 - ▶ Median (Zentralwert) $x_{0,5} = x_{1/2} = \tilde{x} = z$
 - ▶ Modalwerte
 - ▶ Viertelwerte $x_{0,25} = x_{1/4}$, $x_{0,75} = x_{3/4}$

arithmetisches Mittel \bar{x} n Beobachtungen

| | | | | |
|--------------------|------------|------------|---------|------------|
| Merkmalsausprägung | x_1 | x_2 | \dots | x_s |
| Häufigkeit | $H_n(x_1)$ | $H_n(x_2)$ | \dots | $H_n(x_s)$ |
| rel. Häufigkeit | $h_n(x_1)$ | $h_n(x_2)$ | \dots | $h_n(x_s)$ |

$$\begin{aligned} \bar{x} &= \frac{x_1 \cdot H_n(x_1) + x_2 \cdot H_n(x_2) + \dots + x_s \cdot H_n(x_s)}{n} \\ &= x_1 \cdot h_n(x_1) + x_2 \cdot h_n(x_2) + \dots + x_s \cdot h_n(x_s) \end{aligned}$$

bei Klasseneinteilung: x_i – Repräsentant der Klasse

Deutungen und Eigenschaften von \bar{x}

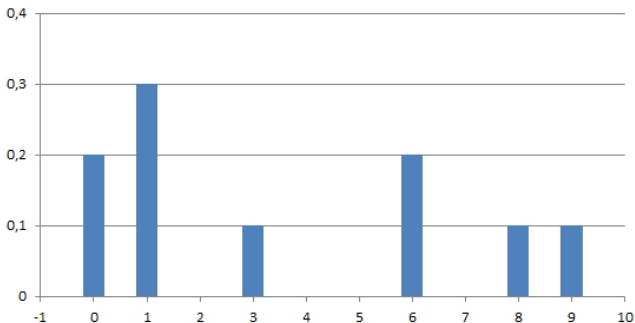
- ▶ gleichmäßige Aufteilung der Gesamtmessung auf die Messpunkte:
„Wenn alle gleich groß wären, gleich viel bekommen würden, ..., dann ... “
- ▶ markiert den Schwerpunkt der Verteilung
- ▶ ist empfindlich gegen Ausreißer
- ▶ ist i. Allg. nicht der häufigste Wert
- ▶ kommt i. Allg. unter den Beobachtungswerten gar nicht vor
- ▶ ist bei asymmetrischer Verteilung stark verschieden vom Median



Beispiel mit TR erzeugte Pseudozufallszahlen:

| | | | | | | | | | | |
|--------------------|-----|-----|---|-----|---|---|-----|---|-----|-----|
| Merkmalsausprägung | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Häufigkeit | 2 | 3 | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 1 |
| rel. Häufigkeit | 0,2 | 0,3 | 0 | 0,1 | 0 | 0 | 0,2 | 0 | 0,1 | 0,1 |

10 Pseudozufallszahlen



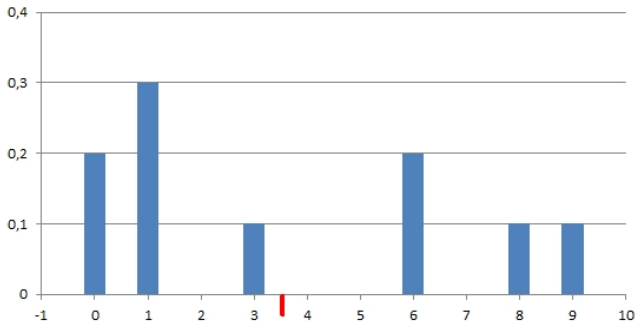
Beispiel mit TR erzeugte Pseudozufallszahlen:

| | | | | | | | | | | |
|--------------------|-----|-----|---|-----|---|---|-----|---|-----|-----|
| Merkmalsausprägung | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Häufigkeit | 2 | 3 | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 1 |
| rel. Häufigkeit | 0,2 | 0,3 | 0 | 0,1 | 0 | 0 | 0,2 | 0 | 0,1 | 0,1 |

$$\begin{aligned}
 \bar{x} &= \frac{0 \cdot 2 + 1 \cdot 3 + 2 \cdot 0 + 3 \cdot 1 + 4 \cdot 0 + 5 \cdot 0 + 6 \cdot 2 + 7 \cdot 0 + 8 \cdot 1 + 9 \cdot 1}{10} \\
 &= 1 \cdot 0,3 + 3 \cdot 0,1 + 6 \cdot 0,2 + 8 \cdot 0,1 + 9 \cdot 0,1 \\
 &= 3,5
 \end{aligned}$$



10 Pseudozufallszahlen



Modalwert

ist der Beobachtungswert mit der größten relativen Häufigkeit. Es kann mehrere Modalwerte geben.

○○○○○●○○○○

○○

○○○○

○○○○

Mittelwerte

Median $x_{1/2}$

ist der Wert, der die der Größe nach geordnete Reihe der Daten halbiert.

Für $n = 2k + 1$ Daten ist $x_{1/2}$ gleich dem $(k + 1)$ -ten Wert in dieser geordneten Reihe.

Für $n = 2k$ Daten ist $x_{1/2}$ gleich dem arithmetischen Mittel aus dem k -ten und dem $(k + 1)$ -ten Wert.

Um $x_{1/2}$ zu bestimmen, muss im Wesentlichen nur zählen können.

Deutungen und Eigenschaften von $x_{1/2}$

- ▶ markiert die Mitte der Verteilung
- ▶ mindestens die Hälfte der Beobachtungswert ist kleiner oder gleich dem Median, mindestens die Hälfte ist größer oder gleich dem Median („mindestens“ darf man nicht weglassen!)
- ▶ ist unempfindlich gegen Ausreißer
- ▶ ist i. Allg. nicht der häufigste Wert
- ▶ ist bei symmetrischer Verteilung nahe dem arithmetischen Mittel

am besten \bar{x} und \tilde{x} angeben

Bildungsstandards: Begründen die Wahl eines Mittelwertes – schwierig!

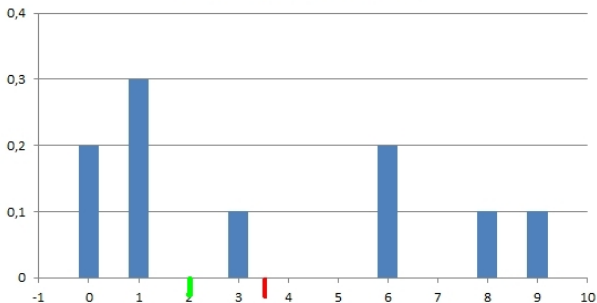


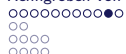
Beispiel mit TR erzeugte Pseudozufallszahlen:

| | | | | | | | | | | |
|--------------------|-----|-----|---|-----|---|---|-----|---|-----|-----|
| Merkmalsausprägung | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Häufigkeit | 2 | 3 | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 1 |
| rel. Häufigkeit | 0,2 | 0,3 | 0 | 0,1 | 0 | 0 | 0,2 | 0 | 0,1 | 0,1 |

$$x_{1/2} = \frac{1 + 3}{2} = 2$$

10 Pseudozufallszahlen





Viertelwerte $x_{1/4}$ und $x_{3/4}$

Durch $x_{1/2}$ wird die geordnete Datenreihe in eine „untere“ und eine „obere“ Hälfte geteilt.

Der untere Viertelwert $x_{1/4}$ ist der Median der „unteren“ Hälfte, der obere Viertelwert $x_{3/4}$ ist der Median der „oberen“ Hälfte.

Beispiel mit TR erzeugte Pseudozufallszahlen:

| | | | | | | | | | | |
|--------------------|-----|-----|---|-----|---|---|-----|---|-----|-----|
| Merkmalsausprägung | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Häufigkeit | 2 | 3 | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 1 |
| rel. Häufigkeit | 0,2 | 0,3 | 0 | 0,1 | 0 | 0 | 0,2 | 0 | 0,1 | 0,1 |

$$x_{1/4} = 1, x_{3/4} = 6$$



$$x_{1/4} = 1, x_{1/2} = 2, x_{3/4} = 6$$

10 Pseudozufallszahlen

