

### Frage:

Aufgabe 3a und b) auf dem Aufgabenblatt 5

Da es sich ja bei der Aufgabe um gruppierte Daten handelt, rechnen wir mit  $\tilde{v}_i$ . Bei der Berechnung sind wir so vorgegangen, als ob wir  $v_i$  berechnen würden. Kann man einfach immer bei gruppierten Daten mit  $v_i$  rechnen und so tun als ob die vier Gruppen vier Personen sind, würde man sich zb die Berechnung des arithmetischen Mittels sparen? Ich habe  $\tilde{v}$  auch selber berechnet und kam da auf das selbe Ergebnis wie für  $v_i$ . Kann ich auch bei gruppierten Daten einfach mit  $v_i$  rechnen weil  $v_i$  und  $\tilde{v}_i$  das selbe ergeben? Oder ist das hier nur Zufall?

### Antwort:

Für individuelle Daten, wie z.B. die Konzentration des Taschengeldes pro Schüler wird, die Unterteilung der X-Achse in gleich große Teile, entsprechend der Gesamtanzahl an Individuen über

$$u_i = \frac{i}{n}, \quad (1)$$

mit  $i = 1, \dots, n$ , als Personenindex und  $n$  der Gesamtzahl der Personen, berechnet.

Für die Anteile der Merkmalsausprägung an der Gesamtsumme aller Untersuchungseinheiten wird,

$$v_i = \frac{\sum_{j=1}^i x(j)}{\sum_{j=1}^n x(j)} \quad (2)$$

berechnet. Beispielsweise der Anteil am gesamten Taschengeld die ein einzelner Schüler besitzt.

Im Falle von gruppierten Daten mit Klassenmitten  $a_1 < a_2 < \dots < a_k$ , beispielsweise der Anteil an Taschengeld einer Klasse an einer Schule werden  $\tilde{u}_i$  und  $\tilde{v}_i$  verwendet.

Hierfür wird als X-Achsenabschnitt

$$\tilde{u}_i = \sum_{j=1}^i f_j \quad (3)$$

die kumulierte relative Häufigkeit der jeweiligen Gruppe berechnet.

Für die Anteile der Merkmalsausprägung an der Gesamtsumme aller Untersuchungseinheiten wird

$$\tilde{v} = \frac{\sum_{j=1}^i f_j a_j}{\sum_{j=1}^k f_j a_j} = \frac{\sum_{j=1}^i n_j a_j}{n \bar{x}} \quad (4)$$

berechnet. Hierbei repräsentieren die Merkmalsausprägungen  $a_j$  die Klassenmitten, das arithmetische Mittel der jeweiligen Gruppe. Also das durchschnittliche Taschengeld eines Schülers in der Klasse.

Wenn man sich die beiden Formeln genau ansieht, merkt man dass sie im Prinzip natürlich das gleiche beinhalten.

$u_i$  Die  $u_i$  bzw  $\tilde{u}_i$  bezeichnen den Anteil an betrachteten Subjekten/Personen.  
Wenn man bei Einzeldaten  $u_i = \frac{i}{n}$  berechnet, ist das ja nichts anderes als das was  $\tilde{u}_i = \sum_{j=1}^i f_j$  bedeutet. Jede Person hat die rel Häufigkeit  $f_j = \frac{1}{n}$  und damit ist  $\tilde{u}_i = \sum_{j=1}^i f_j = \frac{i}{n}$

$v_i$  Ähnliches kann man sich für  $v_i$  überlegen, hier gilt ebenfalls für Einzeldaten dass  $f_j = \frac{1}{n}$  bzw  $n_j = 1$  gilt und sich dadurch die Formeln für gruppierte Daten vereinfachen zu denen für Einzeldaten.

In der angesprochenen Aufgabe Blatt5/3 kann man die Aufgabe lösen, ohne explizit exakt alle Bestandteile der Formel zu benutzen. Wichtig ist einfach, verstanden zu haben was gesucht ist.

**Beispiel:**  $u_i$  und  $v_i$  bis Gruppe 2

Gruppe 1 hat  $n_1 = 10$  Mitglieder und insgesamt 40 Mio investiert, im Mittel also  $a_1 = 4$ . Mitglieder: 10 von insgesamt 30 bedeutet  $f_1 = 0,33$ .

Gruppe 2 hat  $n_2 = 8$  Mitglieder und insgesamt 60 Mio investiert, im Mittel also  $a_2 = 7,5$ . Mitglieder: 8 von insgesamt 30 bedeutet  $f_2 = 0,27$ .

per Formel lautet die Lösung nun:

$$\tilde{u}_2 = \sum_{j=1}^2 f_j = 0,33 + 0,27 = 0,60$$

Oder man überlegt sich, dass für den betreffenden Punkt gilt dass 8 + 10 von 30 Personen betrachtet wurden, der Anteil liegt also bei  $u_2 = \frac{18}{30} = 0,6$

Gleiches gilt für  $\tilde{v}_2$ : die kumulierte Menge des Guts wird durch die Gesamtmenge geteilt. Dies kann man per Formel erreichen. Oder man überlegt sich, dass die beiden ersten Gruppen zusammen  $40 + 60 = 100$  Mio der insgesamt 250 Mio investiert haben, also  $\tilde{v}_2 = \frac{100}{250} = 0,4$  gelten muss.