

Deskriptive Statistik - Aufgabe 8

Der Chef eines Unternehmens möchte sich über die täglich anfallenden Arbeitswege seiner Mitarbeiter informieren. Vom Personalbüro wird ihm auf seine Anfrage hin nachfolgende Tabelle übersandt:

Tabelle: Tägliche Anfahrtswege von 100 Beschäftigten zum Arbeitsplatz

km (von ... bis unter ...)	Anzahl der Beschäftigten
0 - 1	7
1 - 5	24
5 - 15	35
15 - 30	18
30 - 50	16

Aufgabe:

Bestimmen Sie den Median.

Achtung: Median bei klassifizierten Daten verwenden!!

Lösung – Kurzfassung:

Median: 10,43 km

Überlegungen zur Lösung:

Hauptziel ist erst einmal, die Klasse zu finden, in der der mittlere Fall (Median) sicher zu finden ist. Das ist einfach festzustellen. Er liegt zwischen Fall 50 und 51 und liegt hier damit offenbar in der Klasse, die die Fälle 32-66 beinhaltet, hat also einen Wert zwischen 5 und 15 km.

Er nimmt jedoch nicht so einfach den mittleren Wert in dieser Klasse an (das wären hier 10 km). Das Gewicht der Fälle unterhalb und oberhalb der betrachteten Klasse verschiebt die Wahrscheinlichkeit, wo der Median zu finden ist¹. Wenn in den Klassen unterhalb bspw. schon 49 Fälle zu finden wären (und nicht nur 31 wie hier), wäre es wahrscheinlich, dass der Median nahe der unteren Grenze, also dem Wert 5 km liegt. Das nimmt der Rechenweg auf.

¹ Die Werte unterhalb, innerhalb und oberhalb der identifizierten Klasse sind komplementär und ergeben die Gesamtzahl n. Es reicht also für die Rechnung, mit den Fällen unterhalb und innerhalb der Medianklasse sowie der Gesamtzahl n zu rechnen.

Lösung – Erläuterung:

- 1) Es bietet sich bei dieser Aufgabenart an, kumulierte Häufigkeiten zu berechnen, d.h. die Anzahl der Fälle in den Klassen nach und nach aufzuaddieren bis die Gesamtsumme n erreicht ist (hier = 100), s. rechte Spalte.

Klasse	km (von ... bis unter ...)	Anzahl der Beschäftigten	Kumulierte Häufigkeiten
A	0 - 1	7	7
B	1 - 5	24	31
C	5 - 15	35	66
D	15 - 30	18	84
E	30 - 50	16	100

- 2) Es muss die Kategorie bestimmt werden in dem der Median liegt. Der Median ist der Wert in der Mitte $\rightarrow (100+1):2 = 50,5$ (Formel für ungerade n)

In welcher Kategorie der 50,5 Wert liegt, kann jetzt in der Spalte der kumulierten Häufigkeiten abgelesen werden. Hier ist es die dritte Kategorie (5-15 km). Sie enthält die ‚Fälle‘ 32-66, wie man sie in einer geordneten Urliste, geordnet nach Länge der Anfahrtswege, ablesen könnte. ‚Fall 50,5‘ findet sich also hier.

(Zufälligerweise liegt diese Kategorie auch in der Mitte. Das muss jedoch nicht so sein.)

Nun ist Fall 50,5 nicht zu identifizieren. Entweder wurden die Originaldaten mit den exakten Entfernungsdaten nachklassifiziert und z.B. aus Datenschutzgründen vernichtet, oder es wurden nur ‚grobe Daten‘ (Entfernungsklassen) bei den Beschäftigten abgefragt. Deshalb ist der rechnerische Weg gefragt, um den Median annäherungsweise zu rekonstruieren.

- 3) Die gefundene richtige Klasse C, 5-15 km ist jetzt der sog. ‚Eingriffsspielraum‘ m .
- 4) x_{mu} stellt die untere Grenze dieses Eingriffsspielraums dar. Hier 5 km.
- 5) f_m stellt die Anzahl der Fälle der Eingriffsklasse dar. Hier 35
- 6) h stellt die ‚Höhe‘ des Eingriffsspielraums dar. Hier: $15-5 = 10$ km
- 7) $\sum_{i=1}^{m-1} f_i$ stellt die Anzahl der Fälle unterhalb der Eingriffsklasse, d.h. der Fälle der ersten Klasse bis zur letzten Klasse darunter dar. Hier: 31 (re. Spalte, Wert hier bei Klasse B)

$$\begin{aligned} \text{Median klassifizierter Daten : } Z &= x_{mu} + \frac{\frac{n}{2} - \sum_{i=1}^{m-1} f_i}{f_m} \cdot h = 5 + \frac{100}{2} - 31}{35} \cdot 10 \\ &= 5 + \frac{19}{35} \cdot 10 = 5 + 5,43 = 10,43 \end{aligned}$$