Aussagekraft von Erwartungswerten (2)

Zwei Personengruppen zu je 9 Personen werden nach ihren Einkommen befragt. Die Antworten der Angehörigen der beiden Gruppen liegen als Tabellen vor. Die Befragten der ersten Gruppe machen folgende Angaben:

(1) Erwerbstätige von X-Dorf

i	**Einkommen (€)**
1	2730
2	3450
3	3100
4	3090
5	2990
6	3340
7	3190
8	3030
9	3430

(2) Erwerbstätige von Y-Dorf

Die Befragten der zweiten Gruppe machen folgende Angaben:

j	**Einkommen (€)**
1	8400
2	430
3	2990
4	770
5	1280
6	6400
7	2520
8	1060
9	4500

Lösungsskizze

Wir berechnen zunächst den Erwartungswert. Da hier Gleichverteilung der Umfragewerte vorliegt, gilt

$$E(X) = {}^{1}/_{N} \sum_{i=1}^{N} x_i$$

$$= 3150 \text{ [€]}$$

Entsprechend gilt für

$$E(Y) = {}^{1}/_{N} \sum_{i=1}^{N} y_i$$

$$= 3150 \text{ [€]}$$

Offensichtlich ist E(X) für X-Dorf eine repräsentative Maßzahl. E(Y) dagegen ist keine Maßzahl die die Einkommensverhältnisse von Y-Dorf aussagekräftig repräsentiert. Dieser Befund soll nun präzisiert werden, indem wir die Abweichungen der Umfragewerte von den jeweiligen Mittelwerten quantifizieren.

Wir berechnen zunächst die Abweichungen der x_i vom Mittelwert E(X). Da die Summe der Abweichungen vom Mittelwert gleich 0 ist, gehen wir zu quadratischen Abweichungen über. Die Summe der quadratischen Abweichungen ist

$$\sum_{i=1}^{N} (x_i - E(X))^2 = 428600{,}0 \text{ [€}^2\text{]}$$

Division durch n (hier gilt: n= 9) ergibt die durchschnittliche quadratische Abweichung, die sog. ***Varianz von X***. Diese wird von V(X) bezeichnet. Hier gilt

$$V(X) = 47622{,}2 \text{ [€}^2\text{]}$$

Die Varianz ist eine wenig anschauliche Größe und besitzt auch nicht die gleiche Dimension wie die Grundgesamtheit. Daher ziehen wir die Wurzel aus der Varianz und erhalten die Standardabweichung

$$s(X) = 218{,}2\ [€]$$

Die Standardabweichung besitzt nun dieselbe Dimension wie die Grundgesamtheit und ist ein geeigneter Parameter, um die Abweichung vom Mittelwert zu charakterisieren. Es gilt (bei *hinreichend großem* N), daß (*mindestens*) ca. 2/3 der Umfragewerte in einer Umgebung von [μ -s, μ +s] um den Mittelwert μ verteilt sind. Hier gilt, daß ca. 2/3 der Umfragewerte im Bereich [3150 € +/- 218 €] liegen. Um die Standardabweichung von X mit der Standardabweichung anderer Verteilungen vergleichbar zu machen, wird nun die relative Standardabweichung von X, d.h.

$$s_{rel}(X) = s(X) / E(X)$$

berechnet. Hier gilt

$$s_{rel}(X) = 0{,}069$$

Die relative Standarrabweichung ist gering. Sie beträgt nur 6,9% vom Erwartungswert E(X). Die Standardabweichung $s_{rel}(X)$ wird auch als ***Variationskoeffizient*** bezeichnet.

Wir berechnen nun den Erwartungswert von Y. Da hier Gleichverteilung der Umfragewerte vorliegt, gilt

$$E(Y) = 3150\ [€]$$

Offensichtlich ist E(Y) keine Maßzahl, die die Einkommensverhältnisse von Y-Dorf aussagekräftig repräsentiert. Dieser Befund soll nun präzisiert werden. Wir berechnen zunächst die Abweichungen der y_i vom Mittelwert E(Y). Die Summe der quadratischen Abweichungen ist

$$\sum_{i=1}^{N} (y_i - E(Y))^2 = 61297800 \, [€^2]$$

Division durch n (hier gilt: n= 9) ergibt die durchschnittliche quadratische Abweichung, die sog. Varianz von Y. Hier gilt

$$V(Y) = 6810866{,}7 \, [€^2]$$

Wir ziehen die Wurzel aus der Varianz und erhalten die Standardabweichung

$$s(Y) = 2609{,}8 \, [€]$$

Es gilt (bei *hinreichend großem* N), daß ca. 2/3 der Umfragewerte in einer Umgebung von $[\mu\text{-}s, \mu + s\}$ um den Mittelwert μ verteilt sind. Hier gilt, daß ca. 2/3 der Meßwerte im Bereich [3150 € +/- 2882 €] liegen. Um die Standardabweichung mit Standardabweichung von Y anderen Verteilungen vergleichbar zu machen, wird nun die Standardabweichung

$$s_{rel}(Y) = s(Y) / E(Y)$$

berechnet. Hier gilt

$$s_{rel}(Y) = 0{,}828$$

Die Standardabweichung beträgt also 82,8% vom Erwartungswert E(Y). Die Standardabweichung $s_{rel}(Y)$ erreicht beinahe die Größenordnung des Mittelwerts selbst. Daher ist E(Y) als Parameter nicht aussagekräftig.

Der Mittelwert stellt also nur dann eine aussagekräftige Größe dar, wenn die relative Standardabweichung klein gegenüber dem Mittelwert ist.