



Im Bereich der datenbasierten Entscheidungen spielen bestimmte zentrale Begriffe eine entscheidende Rolle. Der erste Schritt in jeder statistischen Analyse besteht darin, zu klären, ob die Daten aus einer Population oder einer Stichprobe stammen.

Population bezeichnet dabei die Gesamtheit aller relevanten Einheiten für eine Untersuchung. Sie wird mit dem Großbuchstaben **I** symbolisiert. Alle aus dieser Population berechneten Werte nennt man **Parameter**.

Eine **Stichprobe**, dargestellt durch den kleinen Buchstaben **i**, ist hingegen eine Teilmenge einer Population. Sie dient als Grundlage für die Berechnung von **Statistiken**.

Ein gutes Beispiel für diese Begriffe ist eine Umfrage zu den Karrierezielen von Studenten der Universität Wien. Die Frage, welche Studenten zur Population gehören, erscheint zunächst einfach: Alle Studenten der Universität Wien. Doch bei genauerer Betrachtung wird schnell klar, dass diese Definition nicht ausreichend ist. Es müssten auch Studenten berücksichtigt werden, die derzeit im Ausland sind, sich im Fernstudium befinden oder bereits eingeschrieben, aber noch nicht angetreten sind.

In der Praxis sind Populationen oft schwer zu bestimmen und zu erfassen. Dies steht im Gegensatz zu Stichproben, die relativ einfach und kostengünstig zu erheben sind. Aus diesem Grund wird in der statistischen Analyse in der Regel auf Stichproben zurückgegriffen, statt die gesamte Population zu untersuchen.

Nehmen wir an, wir möchten eine Stichprobe aus der Studentenschaft der Universität Wien ziehen. Ein möglicher Ansatz wäre, eine Umfrage unter den Studenten in der

Kantine durchzuführen. Wenn wir 50 Studenten befragen, haben wir eine Stichprobe – allerdings eine, die nicht zufällig gewählt wurde, da nur die Studenten, die sich gerade in der Kantine aufhalten, berücksichtigt werden. Diese Stichprobe ist daher weder zufällig noch repräsentativ für die gesamte Studentenpopulation der Universität.

Eine **zufällige Stichprobe** bedeutet, dass jedes Element der Population mit der gleichen Wahrscheinlichkeit ausgewählt wird. Eine **repräsentative Stichprobe** spiegelt die Merkmale der gesamten Population wider. In unserem Beispiel wären weder Zufälligkeit noch Repräsentativität gegeben, da nur eine spezifische Gruppe von Studenten (die Kantinenbesucher) befragt wurde.

Für eine zufällig und repräsentativ ausgewählte Stichprobe müsste die Auswahl der Studenten nach dem Zufallsprinzip erfolgen und die gesamte Bandbreite an Studierenden umfassen. Eine solche Umfrage könnte theoretisch durch den Zugriff auf die Datenbank der Universität erfolgen, um eine zufällige Auswahl der Studenten zu treffen. Doch in der Praxis wäre dies ohne die Mithilfe der Universität kaum durchführbar.

<https://datatab.de/tutorial/stichprobe>

