

Was genau beschreibt die beschreibende Statistik?

In diesem Kapitel ...

- ▶ Erfahren, worum es überhaupt geht in der beschreibenden Statistik
- ▶ Die guten Daten ins Töpfchen, die schlechten ins Kröpfchen
- ▶ Quantitativ von qualitativ unterscheiden

Das Arbeitsgebiet der beschreibenden Statistik lässt sich so zusammenfassen: Mit der beschreibenden Statistik lassen sich Daten übersichtlich darstellen. Dazu werden nicht nur Tabellen oder grafische Methoden verwendet, sondern darüber hinaus Kennzahlen definiert und berechnet, die dieser Übersichtlichkeit dienen.

Erste Daten werden erhoben, erste Stichproben genommen

Das Wichtigste gleich vorweg: Eine statistische Auswertung ist nur sinnvoll, wenn genügend Daten vorliegen und wenn Sie den Daten trauen können. Taugen die Daten nichts oder sind es (viel) zu wenige, so ist jede weitere Rechnung für die Mülltonne. So toll Ihr Statistikprogramm auch sein mag, es kann von sich aus nicht beurteilen, ob die Daten etwas taugen.

Als Mathematiker wird man häufig gefragt, ob man bei dieser oder jener statistischen Auswertung nicht helfen könne. Ich mache das immer gerne und habe dabei festgestellt, dass es meistens nicht die statistischen Verfahren sind, die Probleme machen, sondern die Daten und dabei vor allem die folgenden Tücken:

- ✓ Die Datenmenge ist so gering, dass keine vernünftigen Schlüsse gezogen werden können. Beispiel: Sie untersuchen die Wirksamkeit einer Therapiemethode in einem Krankenhaus. Sie haben aber nur vier Patienten.
- ✓ Es gibt zu viele Daten, denen Sie nicht trauen können, weil nicht klar ist, woher sie kommen oder ob in der Übermittlung der Daten etwas schiefgelaufen ist. Beispiel: Sie erhalten Daten von verschiedenen Institutionen.

- ✓ Es gibt Ausreißer, die das Ergebnis verfälschen. Beispiel: Ihnen liegen 100 Messdaten aus einer Messung vor. 98 davon liegen zwischen 11 und 12. Die übrigen beiden aber lauten 1000 und 1200. Was tun? Jede Mittelwertberechnung und sonstige statistische Auswertung wird durch die beiden Ausreißer extrem beeinflusst, aber können Sie sie einfach weglassen?



Da es meistens nicht möglich ist, eine Totalerhebung (die komplette Grundgesamtheit liegt als Datenmaterial vor) zu machen, beschränkt man sich auf eine Teilerhebung. Diese wird *Stichprobe* genannt.

Einteilung der Merkmale

Grundsätzlich lassen sich Merkmale in *quantitative* und *qualitative* Merkmale aufteilen. Die quantitativen Merkmale haben in der Statistik die bei Weitem größere Bedeutung.

Quantitative Merkmale – zählen und messen

Die Ausprägungen quantitativer Merkmale sind Zahlen aus Messungen oder aus Zählungen. Man unterscheidet dabei

- ✓ **Quantitativ-diskrete Merkmale**, bei denen die Ausprägungen einzelne Punkte sind und die Daten vorzugsweise aus Zählungen stammen. Beispiele hierfür sind Ergebnisse von Verkehrszählungen, Anzahl von Defektstücken, Alter (in Jahren) und so weiter.
- ✓ **Quantitativ-stetige Merkmale**, bei denen die Ausprägungen typischerweise aus einem bestimmten Intervall sind und die Daten vorzugsweise aus Messungen stammen. Beispiele hierfür sind Längen, Größen, Gewicht, Temperaturen und so weiter.

Diese Unterscheidung zwischen *stetig* und *diskret* zieht sich wie ein roter Faden durch die Statistik und somit auch durch dieses Buch.



Die Ausprägungen quantitativ-diskreter Merkmale sind beispielsweise die natürlichen Zahlen inklusive der Null, also 0, 1, 2, 3 ...

Die Ausprägungen quantitativ-stetiger Merkmale hingegen sind prinzipiell die reellen Zahlen $\mathbb{R}$. Da aber keine Messung so exakt durchgeführt werden kann, werden die Ergebnisse in Intervallen angegeben. Dazu ein Beispiel:

Sie messen von mehreren Personen die Körpergröße. Bei der ersten Messung notieren Sie 173.8 cm. Klar ist, dass diese Person nicht exakt 173.8 cm groß ist, Sie messen es eben nur nicht genauer. Dennoch handelt es sich um eine quantitativ-stetige Größe. Mit der Angabe 173.8 cm meinen Sie letztlich das Intervall zwischen 173.75 cm und 173.84 cm.

Qualitative Merkmale – beschreiben und bestaunen

Qualitative Merkmale sind beschreibende Eigenschaften.



Achtung: Beschreibende Eigenschaften können mit Zahlen codiert sein. Vor allem in Datenbanken taucht das häufig auf, um Speicherplatz zu sparen. Dann steht beispielsweise die Zahl 1 für die Farbe Grün, die Zahl 2 für die Farbe Rot und so weiter. Dennoch handelt es sich aber natürlich um ein qualitatives Merkmal, denn die Zahlen stellen ja nur Platzhalter dar und haben nicht die gleiche Bedeutung wie bei den quantitativ-diskreten Merkmalen.

Man unterscheidet

- ✓ **Ordinale Merkmale**, denen eine natürliche Reihenfolge zugrunde liegt. Ein Beispiel: Ich stelle Ihnen die Frage »Wie gerne lesen Sie dieses Buch?« und biete Ihnen die Antwortmöglichkeiten: »sehr gerne«, »gerne«, »mäßig gerne«, »nicht so gerne«, »ich hasse es!«. Die Antworten haben eine natürliche Reihenfolge, aber sie repräsentieren nicht ein Ergebnis wie aus einer Messung.
- ✓ **Nominale Merkmale**, bei denen keine natürliche Reihenfolge zugrunde liegt. Beispiele hierfür sind Farbe, Wohnort, Geschlecht und so weiter.



Ein recht nützliches Vorgehen, quantitative von qualitativen Merkmalen zu unterscheiden, ist auch das Folgende:

Sie überlegen sich, ob die Differenz zwischen zwei Ausprägungen eine Bedeutung hat oder nicht. Hat sie eine, so ist das Merkmal quantitativ, ansonsten qualitativ.

Zum Abschluss des Kapitels noch zu jeder Kategorie ein Beispiel:

- ✓ **Quantitativ-diskret:** Gestern fuhren 150 Autos über die Kreuzung. Heute waren es 162. Also fuhren heute 12 Autos mehr über die Kreuzung. Die Differenz hat eine Bedeutung, also ist das Merkmal quantitativ. Genauer gesagt ist es *quantitativ-diskret*.
- ✓ **Quantitativ-stetig:** In Eisdorf war die durchschnittliche Temperatur im letzten Jahr 6.4 °C. In Ofenhausen war sie 12.2 °C. Also war es in Ofenhausen im Schnitt um 5.8 °C wärmer. Die Differenz hat eine Bedeutung, also ist das Merkmal quantitativ. Genauer gesagt ist es *quantitativ-stetig*.
- ✓ **Ordinal:** Franz kreuzte im Hotel bei der Frage, wie sehr er mit der Sauberkeit des Hotels zufrieden war, die Antwort »sehr« an. Heike, seine Frau, hingegen »mittelmäßig«. Es kann keine Differenz ermittelt werden. Das Merkmal ist qualitativ, genauer gesagt *ordinal*.
- ✓ **Nominal:** Studenten werden nach ihren Studiengängen befragt. Diese werden mit Kennziffern codiert. Pascal studiert Maschinenbau (Kennziffer 3), Nadine studiert Mathematik (Kennziffer 5). Die Differenz zwischen den beiden Kennziffern beträgt 2 ($5 - 3 = 2$). Die Kennziffer 2 stünde für Jura, aber das hat gar keine Bedeutung für die Auswertung. Das Merkmal ist *nominal*.

Das ist nicht so schwierig, nicht wahr? Dann kommt jetzt was fürs Auge, und zwar die verschiedenen Darstellungsmöglichkeiten dieser Merkmale.