Statistik für Jedermann

Tamara Cook

Miles Tallon

# Einleitung: Wozu überhaupt Statistik?

Wie hängen Statistik und empirische Wissenschaft zusammen? Falls du annimmst, Statistik sei nur etwas für Rechnungsprüfer und Mathenerds, oder falls dir die empirische Vorgehensweise nicht sehr vertraut ist, hilft dir dieses wissenschaftstheoretische Miniframework wahrscheinlich weiter. Du kannst aber auch den ungeduldigen Weg gehen und gleich bei den Inhalten einsteigen.

# Deskriptive Statistik: Daten beschreiben und zusammenfassen

Ein Wust an frisch gemessenen Daten liegt vor dir. Auf den ersten Blick sagen sie uns nicht viel. Hier greift aber die deskriptive Statistik, mit deren Hilfe wir Messdaten auf ihre zentralen Eigenschaften reduzieren können. Es geht hier darum, sich einen Überblick über die Daten zu verschaffen. Falls du das bereits gut kannst, wirst du mit den anderen Lerneinheiten wahrscheinlich wenig Probleme haben, auch wenn du diese hier nicht bearbeitest. Deskriptive Statistik ist nicht allzu schwierig, legt aber die Basis für die übrige Statistik. Falls dir das Thema nicht so recht geläufig ist, beginne am besten mit dieser Einheit.

## Zentrale Tendenz:

## Variablen und Summenzeichen:

## Streuung:

## Boxplot:

# Häufigkeitsverteilungen: Grafische Datenaufbereitung und ihr Nutzen

Eine Häufigkeitsverteilung beschreibt, wie oft ein bestimmter Wert innerhalb eines Merkmals, einer gemessenen Eigenschaft wie Masse oder Neugier, auftritt. Die Werte können sich um den Mittelwert herum häufen, aber auch z.B. um das Minimum. Viele Eigenschaften nähern sich einer Verteilungsform an, die sich als mathematische Funktion ausdrücken lässt, ähnlich wie die Geradengleichung eine Gerade beschreibt. In der Statistik machen wir uns diese Gesetzmäßigkeiten zunutze um mehr über die Eigenschaft zu erfahren. Es geht in dieser Lerneinheit vor allem darum, die Verteilung an der grafischen Aufbereitung zu erkennen und daraus weitere Schlüsse über das Merkmal zu ziehen.

# Skalenniveaus: Was kann man mit welchen Daten machen?

Wenn man etwas misst, beobachtet man etwas in der physischen Welt und drückt diese Beobachtung als Messwert aus. Die Messung ist eine abstrakte Repräsentation dessen, wie sich die Beobachtungen voneinander unterscheiden, in welcher Beziehung sie also zueinander stehen. Sind zwei Messwerte gleich/ungleich, oder ist einer größer als der andere, kann man den Größenunterschied als Zahl ausdrücken? Es gibt insgesamt fünf solcher Beziehungsarten, die Skalenniveaus. In der Statistik hängt es vom Skalenniveau ab, welche statistischen Verfahren man auf Messdaten anwenden kann. Wenn man versucht, etwas zu errechnen, was die Daten von ihrem Niveau aus eigentlich nicht hergeben, entsteht Unsinn; wenn man bei der grafischen Aufbereitung nicht die ganze Information des Niveaus nutzt, kann das ggf. zu einem täuschenden Eindruck führen. Dieses Thema ist der Schlüssel zu den geschönten Statistiken und suggestiven Grafiken. Es ist wohl etwas theoretisch, aber sehr fundamental. Was macht man z.B., wenn die Daten keine Zahlen sind, sondern Kategorien?

## Nominalskala: Gleich-Ungleich-Beziehungen

## Ordinalskala: Größer-Kleiner-Beziehungen

## Intervallskala:

## Verhältnisskala:

## Absolutskala:

# Wahrscheinlichkeitstheorie: Den Zufall berechnen

Der Zufall ist den meisten Menschen ein widerstrebendes Konzept. Klare Tatsachen, Sicherheit, Kontrolle scheinen da deutlich sympathischer. Zufall wird gern als völlige Unsicherheit betrachtet, dabei gibt es ein mathematisches Framework um das Ausmaß dieser Unsicherheit zu quantifizieren: die Wahrscheinlichkeit. Ohne Wahrscheinlichkeit verzichten wir auf Informationen und Kontrolle; wir überlassen es dem Zufall. Meistens können wir anhand der Wahrscheinlichkeit genauer einschätzen, ob ein Ereignis eintreten wird. Viele Forschungsfragen lassen sich nur auf der Wahrscheinlichkeitsebene beantworten und auch Statistik hängt eng mit der Wahrscheinlichkeitstheorie zusammen.

# Korrelation und Regression: Zusammenhänge

Der Zusammenhang zwischen zwei metrischen Merkmalen (nicht kategorial) kann statistisch untersucht werden. Bei einer Stichprobe von Personen misst man nicht mehr einen Wert pro Person, sondern ein Wertepaar (beide Merkmale für jede Person). Ist der eine Messwert umso größer, je größer der andere Messwert ist? Wie genau kann man den einen Messwert aufgrund des anderen bestimmen? Solche Fragestellungen ergänzen sich mit jenen nach dem Gruppenunterschied. Sie können außerdem auch gut grafisch untersucht werden.

# Inferenzstatistik: von der Messung auf die Gesamtheit schließen

Es gibt Stichproben, konkrete Messungen, und Populationen, die Gesamtheit, aus der die Stichprobe gezogen wird. Stichproben sollen ihre Gesamtheit repräsentieren. Oft ist die Analyse der Stichprobe nur ein Zwischenziel statistischer Bemühungen, während das Hauptziel darin besteht, durch die Stichprobe mehr über die Population zu erfahren. Diese Art der Statistik fällt in den Bereich der Inferenzstatistik. Sie eröffnet neue Möglichkeiten, bringt aber auch einige Fallstricke mit sich. Wenn du regelmäßig Forschungsergebnisse interpretieren oder verwenden musst, ist die praktische Relevanz des Themas für dich besonders hoch, weil Forschung vor allem an Erkenntnissen über die Gesamtheit interessiert ist und daher immer mit Methoden aus der Inferenzstatistik arbeitet.

# Signifikanztests: Gruppenunterschiede

In den Sozialwissenschaften sehr geläufig sind Fragestellungen nach Gruppenunterschieden. Bei der statistischen Signifikanz geht es darum, ob man gefundene Gruppenunterschiede in Stichproben auf die Population übertragen kann. Die Gruppenunterschiede müssen signifikant (bedeutsam) genug sein, damit wir sie verallgemeinern können. Signifikanz wird mit einem Signifikanztest geprüft, einem Verfahren aus der Inferenzstatistik. Es gibt für alle möglichen Szenarien und Fragestellungen passende Signifikanztests. Du lernst hier einige von ihnen kennen und anzuwenden.