1 Planung einer Fachschaftsfeier – ein Einführungsbeispiel

Um die Grundbegriffe der statistischen Denkweise kennenzulernen, betrachten wir zunächst ein einfaches Beispiel: Der Hauptorganisator der diesjährigen Fachschaftsfeier ist besorgt, dass der dafür gebuchte Raum dieses Jahr nicht ausreichen könnte, denn anders als in den vergangenen Jahren haben ihn dieses Mal bereits viele Studierende auf die Fachschaftsfeier angesprochen. Die Feier war schon immer sehr beliebt, etwa 40 % aller Studierenden des Faches nahmen im letzten Jahr daran teil. Um Vorbereitungen treffen zu können, falls dieser Anteil wesentlich höher ausfallen sollte, befragt der Organisator auf dem Campus willkürlich eine *Stichprobe* von $n = 60$ Studierenden seines Fachs, ob sie beabsichtigen, die Party zu besuchen. Da die Befragung nur einen Tag vor der Feier stattfindet, gehen wir der Einfachheit halber davon aus, dass alle, die zusagen, auch wirklich erscheinen werden.

Er macht also *Beobachtungen* $x_1, \ldots, x_n$. Diese Beobachtungen sind *kategoriell.* Sie nehmen Werte in zwei Kategorien an, $x_i \in \{ja, nein\}$. Nehmen wir an, der *Beobachtungsvektor* $\mathbf{x} := (x_1, \ldots, x_n)^t$ sei

$$\mathbf{x} = (ja, nein, nein, nein, nein, ja, nein, ja, ja, ja, ja, ja, \ldots, nein)^t.$$

Da $\mathbf{x}$ sehr ‚unübersichtlich' ist, betrachtet er nur noch den Anteil an positiven Antworten in der Stichprobe: Von den befragten 60 Studierenden möchten tatsächlich 35, also ein Anteil von etwa $35/60 \approx 0.58$ zur Party kommen! Wenn der wahre Anteil Partybesucher ebenso hoch wäre, würde der alte Raum niemals ausreichen!

Der Organisator hat hier ein Problem und verschiedene Lösungsmöglichkeiten. Zum einen kann er einen größeren Raum buchen, wenn das einfach möglich ist. Vielleicht ist das aber nicht so leicht, und zudem wäre es auch sehr ungemütlich, wenn in einem größeren Raum schließlich doch nur weniger Besucher feiern. Daher stellt sich die Frage, ob seine Beobachtungen tatsächlich darauf hindeuten, dass in diesem Jahr der Andrang wesentlich größer sein wird, oder ob es möglich ist, dass in seiner Stichprobe der Anteil an Partybesuchern wesentlich höher ist als in der *Grundgesamtheit,* oder *Population* aller Studierenden.

M. Messer und G. Schneider, *Statistik,* https://doi.org/10.1007/978-3-662-59339-4_1

Zur Behandlung dieses Problems gibt es eine Reihe statistischer Handwerkszeuge, die wir im Folgenden anhand dieses einfachen Beispiels zusammenfassen.

Zunächst werden die Beobachtungen geeignet dargestellt und übersichtlich mit wenigen Kennzahlen zusammengefasst (beschreibende bzw. deskriptive Statistik). Zweitens möchte man die Beobachtungen in Beziehung setzen zu Aussagen über eine größere Population, z. B. über alle Studierenden dieses Faches an dieser Hochschule (Inferenzstatistik).

Deskriptive Statistik In unserem Beispiel war eine einfache und intuitive *Statistik* – d. h. Funktion der Beobachtungen – zur sinnvollen Beschreibung der Beobachtungen die *relative Häufigkeit* $\hat{p}$ der Partybesucher

$$\hat{p}(\mathbf{x}) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{ja\}}(x_i),$$

wobei die Indikatorfunktion $\mathbb{1}_{\{ja\}}$ gegeben ist durch

$$\mathbb{1}_{\{ja\}}(x_i) := \begin{cases} 1, & \text{falls } x_i = ja, \\ 0, & \text{sonst.} \end{cases}$$

Der Anteil der Partymuffel ergibt sich dann als $1 - \hat{p}(\mathbf{x})$.

Anhand dieser Statistik werden die Daten ‚zusammengefasst'. Der unübersichtliche Datenvektor $\mathbf{x}$ wird auf gewisse Art und Weise verständlich. Dabei haben wir nicht zu viel Information verloren, denn der Organisator ist hier nicht daran interessiert, ob nun gerade die k-te Person zur Party kommt oder nicht. Hätte er andererseits nur an besagter k-ter Person ein Interesse, so wäre die Betrachtung der relativen Häufigkeit nicht sinnvoll, denn die Entscheidung von Person k können wir i. Allg. nicht aus $\hat{p}(\mathbf{x})$ ableiten. In diesem Fall würden wir wohl besser die Statistik $\pi_k(\mathbf{x}) := x_k$ anschauen. Wir müssen also kontextabhängig entscheiden, welche Statistik wir zur Beschreibung der Daten heranziehen.

Das Verständnis für die Daten wird typischerweise durch eine geeignete *grafische Darstellung* gefördert. In Abb. 1.1 sind die Häufigkeiten als Balken dargestellt. Die Länge eines Balkens gibt direkt die absolute (untere Achse) bzw. relative (obere Achse) Häufigkeit der entsprechenden Antworten an. Grafische Darstellungen sollten klar und unmissverständlich sein. Die Botschaften sollten direkt ersichtlich sein. Letzteres wird insbesondere durch

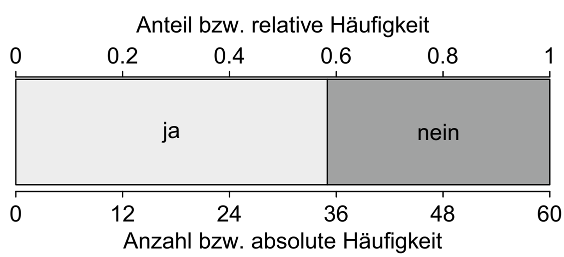

Abb. 1.1 Ergebnis der Befragung nach der Teilnahme an der Fachschaftsparty

eine einfache und optisch ansprechende Darstellung unterstützt. Ganz wichtig: Die Achsenbeschriftungen sollten hinreichend groß sein, und gegebenenfalls sollte eine Legende angefertigt werden.

Schließende Statistik Nun möchten wir die Beobachtungen in Beziehung setzen zu Aussagen über die größere, nicht beobachtete Population aller Studierenden dieses Faches. Wir interpretieren dabei die auf dem Beobachtungsvektor $\mathbf{x}$ der Stichprobe basierende Auswertung der Statistik $\hat{p}(\mathbf{x}) \approx 0.58$ als *Schätzung* des wahren Anteils p und stellen fest, dass die Schätzung $\hat{p}(\mathbf{x})$ von der Behauptung $p^{(0)} = 0.4$ abweicht. Ziel der *schließenden* Statistik ist es, die Größe einer solche Abweichung von Beobachtung und Behauptung, hier $d(\mathbf{x}) := \hat{p}(\mathbf{x}) - p^{(0)} \approx 0.18$, zu beurteilen.

Wir fragen daher: Ist es eine sinnvolle Annahme, dass in dieser Population der *wahre* Anteil Partybesucher wie im letzten Jahr trotzdem $p^{(0)} = 0.4$ ist, sodass doch kein größerer Raum gebraucht wird? Beziehungsweise: Gibt uns die Abweichung $d(\mathbf{x}) = \hat{p}(\mathbf{x}) - p^{(0)} \approx 0.18$ Grund, daran zu zweifeln, dass der wahre Anteil trotzdem $p^{(0)} = 0.4$ ist?

Dazu müssen wir entscheiden, ob $|d(\mathbf{x})|$ ‚groß' ist. Aber was soll das bedeuten? Dieses $|d(\mathbf{x})|$ kann für verschiedene Beobachtungen $\mathbf{x}$ prinzipiell alle Werte im Intervall $[0, 0.6]$ annehmen. Was ist ein großer Wert? Wir machen ein Gedankenexperiment: Angenommen, ein Kollege des Organisators hätte zusätzlich an einem anderen Tag eine Stichprobe $\mathbf{x}^{(a)}$ erhoben, so würde diese vermutlich etwas anders ausfallen als $\mathbf{x}$, und auch $\hat{p}(\mathbf{x}^{(a)})$ würde sich typischerweise von $\hat{p}(\mathbf{x})$ unterscheiden – die Stichprobe $\mathbf{x}^{(a)}$ wird i. Allg. nicht einmal denselben Umfang n haben. Wiederholten wir dieses Experiment immer wieder, so würden wir in den Daten eine gewisse *Variabilität* beobachten. Die Grundidee ist nun, diese Variabilität durch den Zufall zu beschreiben: Wir interpretieren jede Beobachtung x_i des Datenvektors $\mathbf{x}$ als Realisierung einer Zufallsvariable X_i. Dies liefert uns den Zufallsvektor $\mathfrak{X} = (X_1, \ldots, X_n)^t$. Diesen Vektor $\mathfrak{X}$ zusammen mit einer Annahme an mögliche Verteilungen von $\mathfrak{X}$ bezeichnen wir als *statistisches Modell.* Es ist ein einfaches Konstrukt, das uns helfen soll, die beobachtete Variabilität in den Daten zu beschreiben.

Im vorliegenden Fall ist ein einfaches Modell gegeben durch die Annahme, dass die Komponenten von $\mathfrak{X}$ unabhängig und identisch Bernoulli-verteilt sind, mit einem Erfolgsparameter $p \in (0.1)$ (wobei ‚Erfolg' hier Antwort *Ja* bedeutet). Das Tolle an der Sache ist nun, dass wir im Rahmen des Modells einen Begriff für die Größe der Abweichung $d(\mathbf{x})$ erhalten, indem wir sie mit der Verteilung der zufälligen Abweichung $d(\mathfrak{X})$ vergleichen.

Die Anzahl der Erfolge beim n-fachen Münzwurf mit Erfolgsparameter p ist binomialverteilt mit Parametern n und p, wir schreiben $b(n, p)$. Abb. 1.2 zeigt die Verteilung von $\hat{p}(\mathfrak{X})$ für $p = p^{(0)} = 0.4$ und $n = 60$. Wir sehen, dass, wenn die Behauptung $p^{(0)} = 0.4$ zutrifft, die Wahrscheinlichkeit für eine mindestens so große Abweichung von $p^{(0)}$ sehr klein ist, nämlich kleiner als 0.01. Dies ergibt sich als die Summe der Kreise in Abb. 1.2.

Wenn die Behauptung stimmt, ist das, was wir beobachtet haben, also ziemlich unwahrscheinlich, d. h., es käme durch Zufall fast nie vor. In diesem Sinne geben uns die Daten Anlass, daran zu zweifeln, dass der Anteil der Partybesucher unter allen Studierenden dieses Jahr 0.4 ist.

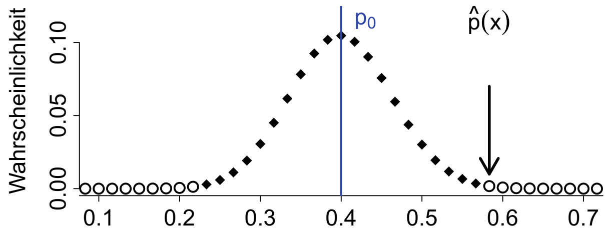

Abb. 1.2 Verteilung von $\hat{p}(\mathfrak{X})$ für $p = p^{(0)} = 0.4$ und $n = 60$

Interpretation der Ergebnisse Die obige ‚Bewertung der Unwahrscheinlichkeit der Beobachtung' war ein (vereinfachtes) Beispiel eines *statistischen Tests.* Von einem mathematischen Standpunkt aus ist gar nicht so viel passiert. Erfahrungsgemäß ist vor allem die *Interpretation* des Ergebnisses beim Erlernen der Statistik einer der anspruchsvollsten Teile der statistischen Schlussweise.

Eine der häufigsten Fehlinterpretationen ist zum Beispiel, dass wir mit dem Test eine Aussage über die Gültigkeit der Behauptung, dass $p^{(0)} = 0.4$, oder des Modells selbst machen können. Das würden wir natürlich liebend gerne, aber das können wir nicht! Wir können damit keine Aussagen über die Population machen. Denn wir haben lediglich eine (Wahrscheinlichkeits-)Aussage über die Beobachtungen(!) gemacht – unter der Annahme, dass die Behauptung stimmt. Wir haben also festgestellt, dass die Beobachtungen nicht gut zur Behauptung passen – nicht mehr und nicht weniger.

Zudem sollten wir uns bewusst machen, dass wir – ungeachtet der Frage, ob die Behauptung (dass $p^{(0)} = 0.4$ ist) stimmt – vorab die Modellannahme getroffen haben, dass unser statistisches Modell der unabhängigen und Bernoulli-verteilten Zufallsvariablen die Realität vernünftig beschreibt. Das ist leider grundsätzlich fragwürdig. Vielleicht hat der Organisator zwar einigermaßen zufällig Studierende auf dem Campus befragt, aber er kann in diesem Befragungsaufbau schlecht andere verzerrende Effekte korrigieren. Zum Beispiel könnten Studierende, die häufiger am Campus anzutreffen sind, auch eher zu den Partybesuchern zählen. Seine nette Art bei der Befragung animierte womöglich auch einige, die zur Party gar nicht erscheinen werden, trotzdem eine Zusage zu machen. Viele dieser Aspekte sind bei der Versuchsplanung zu beachten, wobei es praktisch nie vollständig möglich ist, alle Punkte zu berücksichtigen. So gesehen ist ein statistisches Modell praktisch immer eine Vereinfachung, und damit genau genommen falsch. Dennoch bietet es eine Möglichkeit, die komplexe ‚wahre' Welt in Hinblick auf eine Fragestellung in einem einfachen theoretischen Rahmen zu approximieren.