

Grundlegende Begriffe



Alle Fächer haben ihre eigene Sprache (auch Jargon genannt), und die Statistik bildet da keine Ausnahme. Am besten gehen Sie damit um, indem Sie sich diese Fachsprache von Anfang an zu eigen machen. Wenn Sie dann Aufgaben bearbeiten, erkennen Sie die Hinweise in den Formulierungen und denken gleich in der richtigen Richtung. Sie können auch im Inhalts- oder Stichwortverzeichnis dieses Buches nach solchen Fachbegriffen suchen, um schnell genau die Aufgaben zu finden, mit denen Sie sich rasch vertraut machen wollen. Für die Statistik gilt dasselbe wie überall im Leben: Sobald Sie die Sprache verstehen, fühlen Sie sich gleich wohler.

Die Aufgaben, die Sie bearbeiten werden

In diesem Kapitel schauen Sie quasi aus der Vogelperspektive auf die gebräuchlichsten Begriffe der Statistik und, was vielleicht noch wichtiger ist, auf die Zusammenhänge, in denen sie verwendet werden. Ein Überblick:

- ✓ Die großen Vier: Grundgesamtheit, Stichprobe, Parameter und Statistik (Maßzahl)
- ✓ Die statistischen Ausdrücke, die Sie berechnen, wie Mittelwert, Median, Standardabweichung, Standardwert (Z-Wert) und Perzentil
- ✓ Arten von Daten, Diagrammen und Verteilungen
- ✓ Begriffe der Datenanalyse wie Konfidenzintervall, Fehlergrenze und Hypothesentests

Worauf Sie achten sollten

Achten Sie besonders auf folgende Punkte:

- ✓ Finden Sie in jeder Fragestellung die großen Vier: Sie werden Ihnen überall begegnen.
- ✓ Verstehen Sie das Konzept der Verteilung – es gehört zu den verwirrendsten Konzepten in der Statistik, wird aber immer wieder verwendet. Erledigen Sie das also gleich, bevor es später Sie erledigt.
- ✓ Achten Sie nicht nur auf die Bezeichnungen der Maßzahlen und Analysen, die Sie berechnen, sondern auch auf deren Interpretation, insbesondere im gegebenen Zusammenhang einer Aufgabe.

Erkennen von Grundgesamtheit, Parameter und Maßzahl (1–4)

Sie möchten wissen, in welchem prozentualen Anteil aller Haushalte einer Großstadt der Haushaltsvorstand eine alleinstehende Frau ist. Um diesen Prozentsatz schätzen zu können, befragen Sie 200 Haushalte und ermitteln, wie viele davon von einer alleinstehenden Frau geführt werden.

1. Was ist in diesem Beispiel die Grundgesamtheit?
2. Was ist in diesem Beispiel die Stichprobe?
3. Was ist in diesem Beispiel der Parameter?
4. Was ist in diesem Beispiel die statistische Größe (Maßzahl)?

Unterscheiden von numerischen und kategorialen Variablen (5–6)

Bearbeiten Sie die Aufgaben zu numerischen und kategorialen Variablen.

5. Welches Element der folgenden Aufzählung ist ein Beispiel für eine numerische Variable (auch metrische oder quantitative Variable genannt)?
 - A. Die Farbe eines Autos
 - B. Das Bundesland, in dem jemand wohnt
 - C. Die Postleitzahl einer Person
 - D. Die Körpergröße einer Person, gemessen in Zentimetern
 - E. Sowohl (C) als auch (D)
6. Welches Element der folgenden Aufzählung ist ein Beispiel für eine kategoriale Variable (auch qualitative Variable genannt)?
 - A. Dauer der Schulbildung in Jahren
 - B. Fachrichtung des Studiums
 - C. Abitur oder kein Abitur
 - D. Jahreseinkommen (in Euro)
 - E. Sowohl (B) als auch (C)

Verstehen von Verzerrung (Bias), Variablen und Mittelwert (7–11)

Sie möchten wissen, welcher prozentuale Anteil der Kunden eines Kaufhauses Frauen beziehungsweise Männer sind. Also stellen Sie an einem Samstagmorgen drei Stunden lang an allen vier Eingängen des Kaufhauses Datenerfasser auf und weisen sie an, aufzuzeichnen, wie viele Männer und Frauen in dieser Zeit das Geschäft betreten.

7. Warum kann das Erheben der Daten an einem einzigen Samstagmorgen über drei Stunden zu einer Verzerrung (die man auch Bias nennt) in den Daten führen?

- A. Der Ansatz geht davon aus, dass die am Samstag Einkaufenden repräsentativ für die Gesamtheit aller Personen sind, die unter der Woche in diesem Geschäft einkaufen.
 - B. Der Ansatz geht davon aus, dass der prozentuale Anteil der Frauen an den Einkaufenden am Samstagmorgen genau so groß ist wie zu anderen Zeiten oder an anderen Wochentagen.
 - C. Vielleicht kaufen Paare am Samstagmorgen eher gemeinsam ein als während der übrigen Woche. Dadurch wären die Anteile von Frauen und Männern ähnlicher als zu anderen Zeiten.
 - D. Die Teilnehmenden an dieser Studie wurden nicht zufällig ausgewählt.
 - E. Alle Antwortmöglichkeiten sind richtig.
8. Eine Variable ist ein Merkmal eines jeden Individuums, zu dem Daten erhoben werden. Welche der Elemente in der folgenden Aufzählung sind also Variablen in dieser Studie?
- A. Der Tag, den Sie für die Datenerhebung auswählen
 - B. Das Geschäft, das Sie für die Erhebung auswählen
 - C. Das Geschlecht der einzelnen Einkaufenden, die in der beobachteten Zeit in das Geschäft kommen
 - D. Die Zahl der Männer, die in diesem Zeitraum das Geschäft betreten
 - E. Sowohl (C) als auch (D)
9. In dieser Studie ist _____ eine kategoriale Variable und _____ eine quantitative Variable.
10. Mit welcher Art von Diagramm ließen sich die Anteile von Frauen und Männern an den Einkaufenden in geeigneter Weise darstellen?
- A. Balkendiagramm
 - B. Zeitdiagramm
 - C. Tortendiagramm
 - D. Sowohl (A) als auch (C)
 - E. Sowohl (A) als auch (B) und (C)
11. Wie würden Sie den Mittelwert der Zahl der Einkaufenden pro Stunde berechnen?

Verstehen verschiedener Maßzahlen und Datenanalysebegriffe (12–17)

Beantworten Sie die Fragen zu den verschiedenen Maßzahlen und Datenanalysebegriffen.

12. Bei welchem der folgenden Datensätze ist der Median 3?
- A. 3, 3, 3, 3, 3
 - B. 2, 5, 3, 1, 1
 - C. 1, 2, 3, 4, 5
 - D. 1, 2, 4, 4, 4
 - E. Sowohl (A) als auch (C)
13. Im Mathetest erreicht Susanne eine Punktzahl, die auf dem 90. Perzentil liegt. Was bedeutet das?

14. Sie haben 100 Leute befragt und herausgefunden, dass 60 % davon Schokolade mögen und 40 % nicht. Welches der Elemente der folgenden Aufzählung beschreibt die Verteilung der Variablen »Schokolade / keine Schokolade«?
- A. Eine Tabelle der Ergebnisse
 - B. Ein Tortendiagramm der Ergebnisse
 - C. Ein Balkendiagramm der Ergebnisse
 - D. Ein Satz, der die Ergebnisse beschreibt
 - E. Alle oben genannten Antworten
15. Nehmen Sie an, dass in Ihrem Ergebnis zu einem Test steht, dass Ihr Z-Wert 0,70 beträgt. Was sagt Ihnen das darüber, wie gut Sie in diesem Test abgeschnitten haben?
16. Nach einer landesweiten Befragung wird berichtet, dass 65 % der befragten Amerikaner mit dem Präsidenten zufrieden sind, wobei die Fehlergrenze bei 6 Prozentpunkten liegt. Was bedeutet das?
17. Sie möchten den prozentualen Anteil aller Deutschen schätzen, die vorhaben, in diesem Sommer zwei Wochen oder länger Urlaub zu machen. Welches statistische Verfahren sollten Sie anwenden, um einen Bereich plausibler Werte für den wahren Prozentsatz zu ermitteln?

Verwenden statistischer Verfahren (18–19)

Sie lesen in einem Bericht, dass 60 % aller Schulabsolventen während ihrer Schulzeit Sport getrieben haben.

18. Sie glauben, dass der Anteil der Sport treibenden Absolventen höher ist, als im Bericht angegeben. Welches statistische Verfahren verwenden Sie, um herauszufinden, ob Sie recht haben?
19. Sie glauben, dass der Anteil der zur Schulzeit Sport treibenden Absolventen höher ist, als der Bericht behauptet. Wenn Sie einen Hypothesentest durchführen, um den Bericht zu prüfen, über welchen der folgenden p -Werte als Ergebnis würden Sie sich am meisten freuen?
- A. $p = 0,95$
 - B. $p = 0,50$
 - C. $p = 1$
 - D. $p = 0,05$
 - E. $p = 0,001$

Arbeiten mit Standardabweichungen (20)

Beantworten Sie die Frage zur Standardabweichung.

20. Welcher Datensatz hat die höchste Standardabweichung (führen Sie keine Berechnungen durch)?
- A. 1, 2, 3, 4
 - B. 1, 1, 1, 4
 - C. 1, 1, 4, 4
 - D. 4, 4, 4, 4
 - E. 1, 2, 2, 4