

STATISTICS FOR THE BEHAVIORAL SCIENCES

by GRAVETTER AND WALLNAU · Language: DE

Summary by booksummaryai.com —

<https://booksummaryai.com/country/de/statistics-for-the-behavioral-sciences-zusammenfassung>

Summary

In 60 Sekunden

Gravetter und Wallnau bauen die Statistik als ein einziges durchgehendes Argument auf: Du beschreibst eine Stichprobe und schließt dann von ihr auf die dahinterliegende Grundgesamtheit. Deskriptive Werkzeuge – Mittelwerte, Mediane, Standardabweichungen – fassen Daten ehrlich zusammen, und z-Werte machen verschiedene Verteilungen vergleichbar. Die Verteilung der Stichprobenmittelwerte und der zentrale Grenzwertsatz erklären, warum Stichproben die Population nie perfekt abbilden – und genau deshalb existieren Hypothesentests. Von t-Tests über die Varianzanalyse bis zu Korrelation und Chi-Quadrat kommt in jedem Kapitel eine weitere Methode hinzu, mit der du entscheidest, ob ein Ergebnis echt ist oder nur Stichprobenfehler.

Warum es sich lohnt

Wenn du Forschung in Psychologie, Pädagogik oder den Gesundheitswissenschaften liest, entschlüsselt dir dieses Buch die Zahlen hinter jeder Behauptung, der du begegnest. Es ist für Studierende geschrieben, die Angst vor Mathe haben: Jede Formel kommt mit einem klaren Grund in Alltagssprache, warum es sie überhaupt gibt – vom Wippen-Mittelwert bis zur Gerichtsverhandlungs-Logik des Hypothesentests. Am Ende kannst du den Ergebnisteil eines Fachartikels lesen und weißt, wofür jede einzelne Statistik bürgt.

Die Kernideen

Populationen, Stichproben und die ehrliche Lücke zwischen ihnen

Eine Forscherin will etwas über alle Interessierten wissen – sagen wir, alle Studierenden –, kann aber nur eine Handvoll davon messen. Die volle Gruppe ist die Population; die Handvoll ist die Stichprobe, ausgewählt, um sie zu vertreten. Aus der Stichprobe kommen Statistiken, Werte, die die tatsächlich gemessenen Personen beschreiben; aus der Population kommen Parameter, Werte, die alle beschreiben. Deskriptive Statistik fasst die Daten zusammen und vereinfacht sie, die du hast. Dann kommt der unbequeme Teil, den das Buch ins Zentrum der Inferenz stellt: Die Stichprobenstatistiken sind von Natur aus unvollkommene und variable Repräsentationen der Populationsparameter. Diese eingebaute Diskrepanz – der Stichprobenfehler – ist der Grund, warum der Rest des Buches existiert.

Die gesamte Disziplin ruht auf einem Geständnis: Deine Stichprobe wird die Population nie exakt treffen, und keine noch so große Sorgfalt beseitigt diese Lücke. Was die Statistik stattdessen bietet, ist eine Möglichkeit, die Lücke zu messen – zu sagen, wie falsch eine Stichprobenstatistik wahrscheinlich ist. Wenn du akzeptierst, dass jede Verallgemeinerung diesen natürlichen Fehler trägt, hört die Maschinerie der späteren Kapitel auf, wie Ritual auszusehen, und beginnt wie Schadensbegrenzung auszusehen.

Probier es aus: Nimm eine Statistik aus einer Schlagzeile dieser Woche und frage: Wurde sie an einer Stichprobe oder an der gesamten Population gemessen, und wie groß war die Stichprobe?

Der Mittelwert ist eine Wippe

Stell dir eine Wippe vor, mit den Werten als Gewichten, die entlang eines Bretts platziert sind. Der Mittelwert ist der Gleichgewichtspunkt der Verteilung – die Stelle, an der das Brett ins Lot kommt, berechnet als Summe der Werte geteilt durch die Anzahl der Werte ($\mu = \Sigma X/N$ für eine Population, $M = \Sigma X/n$ für eine Stichprobe). Schiebe nun ein Gewicht weit nach außen ans Ende: Ein einzelner extremer Wert zieht den Gleichgewichtspunkt zu sich. Deshalb empfiehlt das Buch den Median für schiefe Verteilungen, denn der Median wird durch extreme Werte nicht beeinflusst – er ist einfach der Mittelpunkt der geordneten Werte, der Punkt, unter den 50 % fallen. Für Kategorien ohne jede Ordnung, wie Augenfarbe, überlebt nur der Modus: der Wert mit der größten Häufigkeit und das einzige Lagemaß, das mit Nominaldaten verwendbar ist.

Die Wahl eines Lagemaßes ist in Wirklichkeit die Wahl, mit welcher Art von Verzerrung du leben kannst. Der Mittelwert antwortet auf jeden Wert, was ihn präzise und zerbrechlich macht; der Median ignoriert die Größe, was ihn unter Schiefe robust macht; der Modus ignoriert alles außer Beliebtheit. Das Skalenniveau entscheidet für dich – Nominaldaten lassen gar keine Wahl.

Probier es aus: Nimm deine letzten fünf Einkäufe, berechne Mittelwert und Median und sieh, welches Maß einen 'typischen' Betrag besser repräsentiert.

Standardabweichung: die durchschnittliche Entfernung vom Mittelwert

Zwei Klassen können denselben Mittelwert haben, während die eine voller fast identischer Werte ist und die andere überall verstreut. Um das zu erfassen, misst das Buch Entfernung: Ziehe den Mittelwert von jedem Wert ab, quadriere die Abweichungen und bilde den Durchschnitt – das ist die Varianz, das Mittel der quadrierten Abweichungen, $\sigma^2 = \Sigma (X - \mu)^2 / N$ für eine Population. Die Summe der quadrierten Abweichungen, SS, lässt sich entweder mit der definitorischen oder der rechnerischen Formel finden. Ziehe aus der Varianz die Quadratwurzel und du erhältst die Standardabweichung, $\sigma = [\Sigma (X - \mu)^2 / N]^{1/2}$, ein Maß für die Standard- oder durchschnittliche Entfernung vom Mittelwert. Bei Stichproben gibt es eine Korrektur: Teile SS durch $n-1$, nicht durch n , weil die Stichprobenvariabilität systematisch nach unten verzerrt ist, und die $n-1$ -Korrektur macht die Stichprobenvarianz zu einer unverzerrten Schätzung der Populationsvarianz.

Variabilität ist kein Nachgedanke zum Mittelwert; sie ist das, was den Mittelwert vertrauenswürdig oder bedeutungslos macht. Ein Mittelwert von 50 bedeutet etwas anderes, wenn die Werte innerhalb von 2 Punkten klustern, als wenn sie sich über 20 erstrecken. Die $n-1$ -Korrektur ist der stille Meisterzug des Buches: Sie gibt zu, dass eine Stichprobe die Streuung ihrer Population unterschätzt, und zahlt einen kleinen Preis im Nenner, um eine ehrliche Schätzung zu kaufen.

Probier es aus: Berechne die Standardabweichung deiner täglichen Bildschirmzeit für eine Woche und sieh, wie weit ein 'typischer' Tag von deinem Durchschnitt abweicht.

z-Werte: alle Verteilungen auf dasselbe Lineal bringen

Du schreibst in einer Prüfung 70 und in einer anderen 85. Welche Leistung ist besser? Die rohen Zahlen können es nicht sagen, denn die beiden Prüfungen können unterschiedliche Mittelwerte und unterschiedliche Streuungen haben. Die Lösung des Buches ist der z-Wert: den Mittelwert abziehen und durch die Standardabweichung teilen, $z = (X - \mu) / \sigma$. Das Vorzeichen sagt dir, ob der Wert über oder unter dem Mittelwert liegt; die Zahl sagt dir, wie viele Standardabweichungen er davon entfernt ist. Der angegebene Zweck besteht darin, X-Werte in z-Werte umzurechnen, sodass die entstehende Verteilung einen Mittelwert von 0 und eine Standardabweichung von 1 hat und verschiedene Verteilungen damit vergleichbar werden. Deine 70 könnte zwei Standardabweichungen über dem Schnitt liegen, während deine 85 kaum eine halbe Standardabweichung darüber liegt – und jetzt ist der Vergleich ehrlich.

Ein z-Wert ist eine Übersetzung, keine Verwandlung der Person. Er verlagert jeden Wert auf eine universelle Skala, auf der null für Durchschnitt steht und jede Einheit für eine Standardabweichung. Die Überraschung ist, wie viel Information Vorzeichen und Größe tragen: Die Lage innerhalb der Verteilung, nicht die rohe Zahl, macht Werte über Tests, Kurse oder Messungen hinweg vergleichbar.

Probier es aus: Rechne einen deiner Testwerte mit dem Klassendurchschnitt und der Standardabweichung in einen z-Wert um und sieh, wo du tatsächlich stehst.

Wahrscheinlichkeit als Anteil

Bevor ein Hypothesentest überhaupt funktionieren kann, braucht das Buch eine Definition von Wahrscheinlichkeit, mit der die Arithmetik etwas anfangen kann. Seine Antwort ist entwaffnend schlicht: Wahrscheinlichkeit ist ein Bruchteil oder Anteil aller möglichen Ausgänge. Die Wahrscheinlichkeit von A gleich der Anzahl der als A eingestuften Ausgänge geteilt durch die Gesamtzahl der möglichen Ausgänge. Keine Metaphysik, keine Langfristphilosophie – nur Zählen. Mit dieser Definition in der Hand werden die Anteile unter einer Verteilung von Werten zu Wahrscheinlichkeiten, bestimmte Werte zu ziehen, und die Brücke von beschreibenden Kurven zu inferenzstatistischen Entscheidungen ist geschlagen.

Wahrscheinlichkeit als Anteil zu definieren ist das, was eine Kurve von Werten zum Doppeldienst als Aussage über Zufall befähigt. Wenn du den Anteil der Ausgänge in einem Bereich kennst,

kennst du die Wahrscheinlichkeit, dort zu landen. Jeder p-Wert, den du je lesen wirst, stammt von dieser schlichten Division ab: günstige Ausgänge durch alle Ausgänge.

Probier es aus: Nimm dir in dieser Woche eine Ja/Nein-Entscheidung vor, schreibe die möglichen Ausgänge als Bruch auf und prüfe, ob dein Bauchgefühl mit der Rechnung übereinstimmt.

Die Verteilung der Stichprobenmittelwerte und der Zentrale Grenzwertsatz

Ziehe jede mögliche Zufallsstichprobe des Umfangs n aus einer Grundgesamtheit, berechne den Mittelwert jeder Stichprobe und sammle all diese Mittelwerte: Diese Sammlung ist die Verteilung der Stichprobenmittelwerte, die das Buch ein zentrales Konzept der Inferenzstatistik nennt. Der Zentrale Grenzwertsatz beschreibt dann ihre Form: Für jede Grundgesamtheit mit Mittelwert μ und Standardabweichung σ hat die Verteilung der Stichprobenmittelwerte den Mittelwert μ , die Standardabweichung $\sigma/\sqrt{n}$ und nähert sich mit n gegen unendlich der Normalverteilung an. Sie ist normalverteilt, wenn die Grundgesamtheit selbst normalverteilt ist oder der Stichprobenumfang relativ groß ist – mindestens $n \geq 30$. Der Mittelwert dieser Verteilung entspricht dem Grundgesamtheitsmittelwert, und ihre Standardabweichung – der Standardfehler, $\sigma_M = \sigma/\sqrt{n}$ – ist ein Maß für die durchschnittliche Entfernung zwischen einem Stichprobenmittelwert und dem Grundgesamtheitsmittelwert.

Dieses Kapitel erklärt, warum der Stichprobenfehler aus Kapitel 1 kein Chaos ist, sondern Struktur. Die Stichprobenmittelwerte türken sich um den Grundgesamtheitsmittelwert in einem vorhersagbaren, meist normalen Muster auf, und größere Stichproben pressen diesen Haufen enger zusammen – teile σ durch $\sqrt{n}$ und sieh zu, wie der Fehler schrumpft. Der Standardfehler macht aus „Stichproben sind unvollkommen“ eine Zahl, die du berechnen kannst – genau das, was Hypothesentests später ausgeben werden.

Probier es aus: Wirf 10 Münzen, mittlere die Köpfe pro 10 Würfe, wiederhole das fünfmal und beobachte, wie sich deine fünf Mittelwerte um 0,5 scharen.

Der Prozess vor der Jury: Hypothesentesten

Das Buch inszeniert das Hypothesentesten als Gerichtsverhandlung. Der Angeklagte gilt als unschuldig, so wie die Nullhypothese als wahr gilt, bis die Daten etwas anderes nahelegen, und die Jury muss zwischen zwei möglichen Urteilen entscheiden. Zwei Fehler sind möglich: einen Unschuldigen verurteilen – das ist ein Fehler 1. Art, also eine Nullhypothese zurückweisen, die tatsächlich wahr ist – oder einen Schuldigen freisprechen, ein Fehler 2. Art, also eine Nullhypothese beibehalten, die in Wirklichkeit falsch ist. Das Signifikanzniveau α ist die Wahrscheinlichkeit, dass der Test zu einem Fehler 1. Art führt; das Symbol β steht für die Wahrscheinlichkeit des Fehlers 2. Art. Jedes 'signifikante' Ergebnis in einer Fachzeitschrift ist ein Urteil, das genau unter diesen Risiken gefällt wurde.

Das Alpha-Niveau ist keine magische Schwelle, sondern eine vorab festgelegte Toleranz dafür, eine unschuldige Nullhypothese zu verurteilen. Senkst du α , verurteilst du weniger unschuldige Hypothesen – sprichst dafür mehr schuldige frei, weil die beiden Fehler gegeneinander wirken.

Wer Forschung mit dieser Analogie liest, stellt nicht mehr die Frage 'Ist es signifikant?', sondern 'Welchen Fehler hat diese Studie bewusst riskiert – und zu welchem Preis?'

Probier es aus: Notiere vor dem Lesen der Ergebnisse einer Studie ihr Alpha-Niveau und frage dich, welchen Fehler – Fehllalarm oder übersehener Effekt – die Forschenden zu minimieren beschlossen haben.

Die t-Statistik: Mittelwerte testen, wenn σ unbekannt ist

Der z-Test des vorigen Kapitels verlangt etwas, das echte Forschung selten liefert: die Standardabweichung der Population. In der Praxis hast du eine Stichprobe und kein σ . Die t-Statistik schließt diese Lücke: Sie prüft Hypothesen über einen unbekannten Populationsmittelwert μ , wenn σ unbekannt ist, mit der Formel $t = (M - \mu) / s_M$, wobei s_M der geschätzte Standardfehler $s/\sqrt{n}$ ist. Die Standardabweichung der Stichprobe selbst vertritt die der Population, und der geschätzte Standardfehler vertritt den wahren. Die Logik des Tests bleibt unverändert – du vergleichst deinen Stichprobenmittelwert mit dem von der Nullhypothese behaupteten Mittelwert – nur das Maß ist nun aus den Daten geschätzt, die du tatsächlich hast.

Die t-Statistik ist die Arbeitsversion des z-Tests, gebaut für den häufigen Fall, dass die Population ein Unbekannter ist. s statt σ einzusetzen kostet Gewissheit – deshalb haben t-Verteilungen dickere Enden als die Normalverteilung –, bringt aber die Anwendbarkeit auf nahezu jedes echte Experiment. Eine Stichprobe, ein Mittelwert, eine ehrliche Fehlerschätzung: Das ist die ganze Maschinerie.

Probier es aus: Suche in einer Arbeit, die du gerade liest, einen t-Wert für eine Stichprobe und prüfe, ob sein Standardfehler aus der Stichprobe stammt und nicht aus einer bekannten Population.

Zwei Stichproben, zwei Designs: unabhängig oder wiederholt

Es gibt zwei Wege, zwei Behandlungen zu vergleichen, und die Wahl verändert den Test. Der t-Test für unabhängige Stichproben nutzt zwei getrennte Stichproben – eine eigene Gruppe pro Behandlungsbedingung –, um die Mittelwertdifferenz zwischen zwei Populationen zu bewerten, mit einer Nullhypothese der Form $\mu_1 - \mu_2 = 0$. Der t-Test für abhängige Stichproben setzt stattdessen jede Person beiden Bedingungen aus oder beobachtet jeden Menschen zu zwei verschiedenen Zeitpunkten und arbeitet mit Differenzwerten $D = X_2 - X_1$, wobei der Mittelwert dieser Differenzen bewertet wird. Dieselbe Frage – unterscheiden sich die Behandlungen? –, aber die Daten kommen unterschiedlich geformt an: zwei unabhängige Haufen von Werten gegenüber einer Spalte intraindividueller Veränderungen.

Das Design bestimmt die Rechenart. Getrennte Gruppen zwingen den Test, zwei Stichprobenmittelwerte gegeneinander zu vergleichen; dieselben Personen in beiden Bedingungen erlauben dem Test, die Veränderung jeder Person direkt zu verfolgen. Der Weg über abhängige Stichproben entfernt still die Unterschiede zwischen den Personen aus dem Vergleich – deshalb ziehen Forschende, die ihre Teilnehmer zweimal testen können, ihn oft vor.

Probier es aus: Entscheide bei deiner nächsten Vergleichsfrage zuerst, ob eine Gruppe beide Bedingungen durchlaufen kann – das kann das Rauschen halbieren.

ANOVA und die F-Quotient: viele Behandlungen auf einmal vergleichen

Ein Forscher mit drei oder vier Behandlungen könnte t-Test für t-Test durchführen—aber jeder zusätzliche Test würfelt erneut, erhöht das experimentweise Alpha-Niveau und steigert das Risiko eines Typ-I-Fehlers. Die Varianzanalyse löst dies mit einem einzigen Test: ANOVA bewertet Mittelwertsunterschiede zwischen zwei oder mehr Behandlungen, indem sie einen F-Quotienten berechnet, $F = \text{Varianz zwischen Behandlungen} / \text{Varianz innerhalb der Behandlungen}$, oder $MS_{\text{zwischen}} / MS_{\text{innerhalb}}$. Wenn sich die Behandlungen wirklich unterscheiden, schwillt die Varianz zwischen ihnen über die Hintergrundvarianz innerhalb von ihnen an, und der Quotient wächst. Eine Zahl, eine Entscheidung, kein kumulierender Fehler.

Der F-Quotient ist ein Signal-Rausch-Argument in einer einzigen Division: Unterschiede zwischen Behandlungen oben, gewöhnliche Variabilität unter den Teilnehmern unten. Die Einsicht hinter ANOVA ist, dass das Stapeln paarweiser Tests dein Typ-I-Fehlerrisiko multipliziert, also schützt ein Omnibus-Test das Alpha-Niveau, das du festlegst. Erst nachdem F gesprochen hat, verdienen paarweise Vergleiche ihren Auftritt.

Probier es aus: Wenn du eine Studie siehst, die drei oder mehr Gruppen vergleicht, prüfe, ob sie einen einzigen Gesamt-F berichtet, bevor du irgendwelchen paarweisen Behauptungen vertraust.

Messwiederholungs- und zweifaktorielle ANOVA: Personen entfernen, Interaktionen hinzufügen

Wenn dieselben Teilnehmer in jeder Behandlungsbedingung dienen, kann die Messwiederholungs-ANOVA—dargestellt als Erweiterung des t-Tests für Messwiederholungen—individuelle Unterschiede vollständig aus dem Fehlerterm entfernen, weil jede Person als ihre eigene Baseline fungiert. Ihr F-Quotient wird zu $MS_{\text{zwischen}} / MS_{\text{fehler}}$, wobei die Person-zu-Person-Varianz vom Nenner subtrahiert wird. Dann weitet das Buch den Blick: Eine zweifaktorielle ANOVA bewertet Mittelwertsunterschiede aus zwei unabhängigen Variablen gleichzeitig und berechnet drei separate F-Quotienten—einen für den Haupteffekt von Faktor A, einen für den Haupteffekt von Faktor B und einen für die $A \times B$ -Interaktion. Eine Interaktion tritt auf, wenn der Effekt eines Faktors vom Niveau eines anderen Faktors abhängt.

Diese Kapitel zeigen ANOVA als flexible Architektur statt als einen einzelnen Test. Messwiederholungen kaufen Kraft, indem sie die Varianz löschen, die daher kommt, dass Menschen einfach unterschiedlich sind. Zweifaktorielle Designs kaufen Realismus, indem sie prüfen, ob Effekte über Bedingungen hinweg Bestand haben—oder sich biegen, was eine Interaktion offenlegt und was ein Haupteffekt allein nie zeigen kann.

Probier es aus: Frage bei deinem nächsten Datenvergleich, ob sich der Effekt einer Variablen über die Niveaus einer anderen ändert—das ist die Interaktion, die es zu plotten lohnt.

Korrelation und Chi-Quadrat: Beziehungen ohne Mittelwerte messen

Nicht jede Frage dreht sich um Mittelwertsunterschiede. Die Pearson-Korrelation misst Grad und Richtung der linearen Beziehung zwischen zwei Variablen, berechnet als $r = SP/(SSx \cdot SSy)$ aus der Summe der Produkte der Abweichungen. Ihre Werte laufen von $-1,00$, einer perfekten negativen Beziehung, bis $+1,00$, einer perfekten positiven, wobei 0 überhaupt keine lineare Beziehung bedeutet. Und wenn deine Daten keine Scores, sondern Zählungen sind—wie viele Personen in jede Kategorie fallen—vergleicht der Chi-Quadrat-Anpassungstest die Häufigkeitsverteilung aus einer Stichprobe mit einer hypothetischen Verteilung, die von der Nullhypothese angegeben wird, unter Verwendung von $\chi^2 = \sum (f_o - f_e)^2 / f_e$, wobei f_o die beobachtete Häufigkeit und f_e die erwartete Häufigkeit ist.

Korrelation beantwortet 'Bewegen sich diese beiden Variablen zusammen, und wie eng?', während Chi-Quadrat beantwortet 'Passen meine beobachteten Zählungen zu dem, was die Nullhypothese vorhersagte?' Beide verzichten vollständig auf Mittelwerte—eine arbeitet mit Kovariation, die andere mit Häufigkeiten.

Probier es aus: Berechne oder plote für einen gepaarten Zahlensatz in deinem Leben—Schlafstunden gegenüber Stimmungsbewertung—die Korrelation und prüfe ihr Vorzeichen und ihre Stärke.

Schätzen: von Ja-oder-Nein zu einem Wertebereich

Hypothesentests liefern ein Urteil; Schätzen liefert eine Messung. Statt nur zu fragen, ob eine Behandlung wirkt, kannst du abschätzen, wo sich ein Anteil der Populationsparameter befindet. Eine Punktschätzung ist eine einzelne Zahl, etwa ein Stichprobenmittelwert; eine Intervallschätzung ist ein Wertebereich – ein Konfidenzintervall. Für einen unbekannten Populationsmittelwert μ wird das Intervall berechnet als $M \pm t(sM)$: der Stichprobenmittelwert im Zentrum, der geschätzte Standardfehler, skaliert durch t , bestimmt die Breite. Das Ergebnis ist ein Bereich, der mit gewählter Zuverlässigkeit angibt, wo der Populationswert plausibel liegt.

Ein Konfidenzintervall verwandelt eine binäre Entscheidung in einen Bereich mit Breite, und die Breite selbst ist Information: weite Intervalle gestehen Unsicherheit ein, schmale beanspruchen Präzision. Weil das Intervall aus demselben geschätzten Standardfehler aufgebaut ist, der auch die t -Statistik antreibt, sind Schätzen und Hypothesentest zwei Lesarten einer einzigen Rechnung – verwerfen oder nicht, und um wie viel.

Probier es aus: Wenn du das nächste Mal einen berichteten Mittelwert liest, suche nach seinem Konfidenzintervall und beurteile die Studie anhand der Breite des Intervalls, nicht nur anhand des Mittelwerts.

Die besten Zitate

- "die Stichprobenkennwerte sind naturgemäß unvollkommene und variable Abbilder der Populationsparameter" - Gravetter und Wallnau

- "das arithmetische Mittel ist der Gleichgewichtspunkt der Verteilung" - Gravetter und Wallnau
- "der Median wird durch extreme Werte nicht beeinflusst" - Gravetter und Wallnau
- "der Standardfehler ist ein Maß für die durchschnittliche Distanz zwischen einem Stichprobenmittelwert und dem Populationsmittelwert" - Gravetter und Wallnau

Kritik

Die Milde des Buchs gegenüber der Mathematik hat ihren Preis: Herleitungen bleiben knapp gehalten, wer Beweise sucht oder einen matrizenalgebraischen Weg zu Varianzanalyse und Regression will, braucht eine dichtere Begleitlektüre. Die Beispiele stützen sich auf lehrbuchhafte Szenarien statt auf unordentliche echte Datensätze, und Effektstärken, Power sowie moderne Resampling-Methoden werden im Vergleich zur klassischen Testpalette vergleichsweise knapp behandelt. Auch die formelzentrierte Reihenfolge kann die frühen Kapitel mechanisch wirken lassen, bis sich der inferentielle Ertrag in den Kapiteln 7 und 8 einstellt.

In einem Satz

Jede Statistik in diesem Buch ist eine Methode, um zu messen, wie weit die ehrliche Unvollkommenheit einer Stichprobe von der Wahrheit der Grundgesamtheit abweicht, die dich eigentlich interessiert.

Für wen sich das Buch eignet

Studierende der Psychologie, Pädagogik, Pflege oder eines jeden verhaltenswissenschaftlichen Kurses, der mit dem Lehrbuch von Gravetter und Wallnau arbeitet, dazu Selbstlerner, die den klassischen Statistik-Werkzeugkasten ohne Beweise erklärt bekommen wollen. Forschende, die zu den Grundlagen zurückkehren, werden die Gerichts- und Wippen-Analogien als Auffrischung zu schätzen wissen.