Vorlesung Statistik II

Florian Wickelmaier

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Sommersemester 2026

Literatur

Lehrbücher

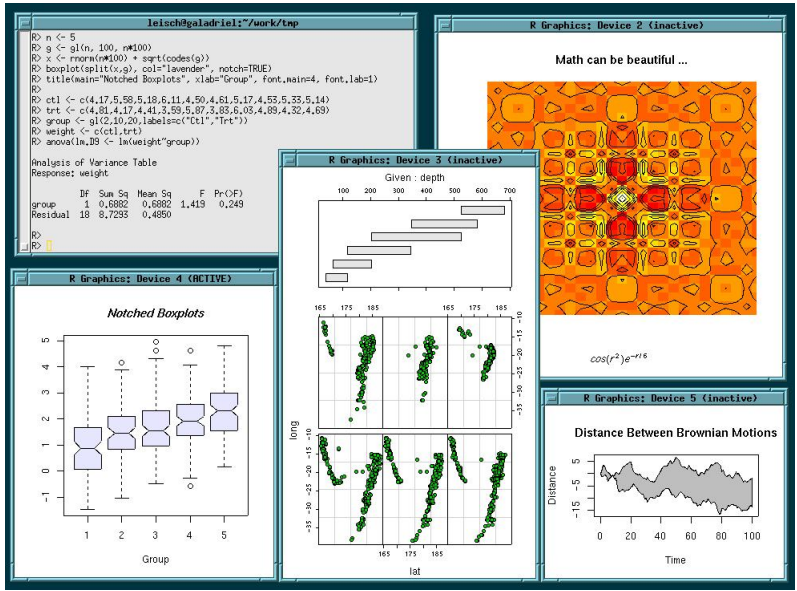
- Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I. & Tutz, G. (2016). *Statistik*. Berlin: Springer. doi:10.1007/978-3-662-50372-0
- Bosch, K. (2011). *Elementare Einführung in die Wahrscheinlichkeitsrechnung* (11. Aufl.). Braunschweig: Vieweg.
- Bosch, K. (2010). *Elementare Einführung in die angewandte Statistik* (9. Aufl.). Braunschweig: Vieweg.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer.
- Hays, W.L. (1994). *Statistics* (5th ed.). New York: Holt, Rinehart and Winston.
- Heller, J. (2012). Experimentelle Psychologie. München: Oldenbourg. (Kap. 4)

Online-Lehrtexte

- Siehe Homepage

Open-Source-Software: R

https://www.r-project.org/



Einführung

Statistik I

- Deskriptive Statistik, explorative Datenanalyse

  Untersuchungsergebnisse (Stichprobe) werden in geeigneter Weise beschrieben und zusammengefasst (Tabellen, grafische Darstellungen)

  Berechnung von Kenngrößen (Statistiken), die möglichst viel Information über die vorliegenden Daten liefern

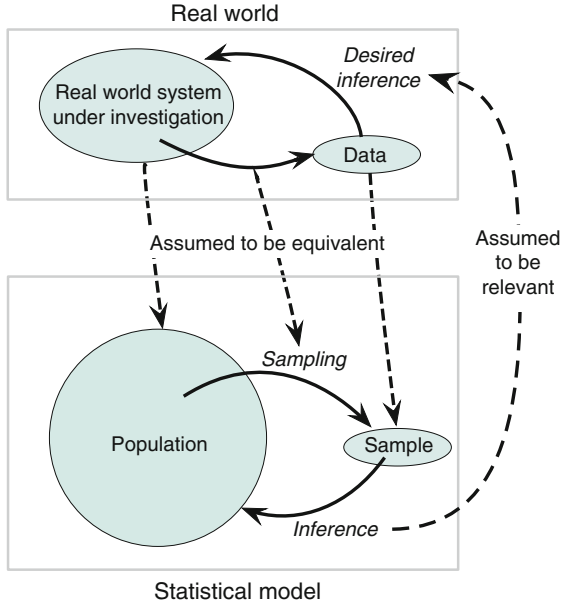
- Wahrscheinlichkeitstheoretische Grundlagen

  Erfassung des zufälligen Geschehens in Zufallsexperimenten

  Wahrscheinlichkeitsraum als Grundlage der Charakterisierung zufälliger Ereignisse durch Zufallsvariablen und ihre Verteilungen und Parameter

Statistische Inferenz

Lew, M. J. (2020). A reckless guide to p-values: local evidence, global errors.



Statistische Inferenz

Zufallsexperimente und Wahrscheinlichkeitsraum

Das durchgeführte Zufallsexperiment bestimmt den Wahrscheinlichkeitsraum und damit

- den Wertebereich
- die Verteilung

der betrachteten Zufallsvariablen, also das *daten-erzeugende Modell*

Der Wahrscheinlichkeitsraum wird

- unmittelbar festgelegt durch die Art des Zufallsexperiments (z. B. Multiple-Choice-Aufgaben, Ja/Nein-Antworten)
- ergibt sich approximativ oder im Rahmen einer asymptotischen Betrachtung (z. B. Merkmale wie Körpergröße, IQ?)

Erinnerung: Zentraler Grenzwertsatz

Für jedes n seien die Zufallsvariablen $X_1, X_2, \ldots, X_n$ stochastisch unabhängig mit Erwartungswerten $E(X_i)$ und Varianzen $Var(X_i)$

Unter sehr allgemeinen Bedingungen (die im Wesentlichen besagen, dass jede Zufallsvariable bei der Summenbildung nur einen kleinen Einfluss hat) ist dann die Summe

$$S_n = X_1 + X_2 + \ldots + X_n$$

asymptotisch normalverteilt (d. h. für $n \to \infty$ normalverteilt) mit

$$\mu = E(S_n) = \sum_{i=1}^{n} E(X_i) \text{ und } \sigma^2 = Var(S_n) = \sum_{i=1}^{n} Var(X_i)$$

Die Bedingungen sind z. B. erfüllt, wenn die Zufallsvariablen identisch verteilt sind

Beispiel: Konstanzmethode

Aufgabe der Versuchsperson im Rahmen der Konstanzmethode

- Bei Darbietung eines Reizpaares (x, s) bekommt die Versuchsperson die Aufgabe, mit "Ja" zu antworten, falls der Vergleichsreiz x bezüglich der betrachteten Empfindungsgröße als intensiver erscheint als der Standardreiz s, und mit "Nein" sonst

Die experimentelle Kontrolle der Randbedingungen soll gewährleisten, dass das Teilexperiment der wiederholten Darbietung der Reizsituation (x, s) als Bernoulli-Experiment betrachtet werden kann

Die Anzahl der Fälle, bei denen eine "Ja"-Antwort erfolgt, ist daher eine binomialverteilte Zufallsvariable

Daten-erzeugendes Modell in R

```
x <- rbinom(1, size = 30, prob = 0.6)  # n = 30, p = 0.6
```

Parameterschätzung

Grundsätzlicher Unterschied zur Urnenziehung

- Die Wahrscheinlichkeit einer "Ja"-Antwort als Parameter p der Bernoulli- bzw. Binomialverteilung ist nicht bekannt

Aufgabe der Statistik

- Bereitstellung geeigneter Verfahren zur Schätzung der Parameter der unterliegenden Verteilung aus erhobenen Daten

Eine vergleichbare Situation haben wir bereits im Kontext der linearen Regression kennengelernt, wo die Methode der kleinsten Quadrate zur Parameterschätzung verwendet wurde

Statistisches Schließen

Häufig ist man nicht primär an den im Rahmen einer *Stichprobe* erhobenen Daten interessiert, sondern an den Schlüssen hinsichtlich der betrachteten *Population* (Grundgesamtheit), die diese zulassen

Beispiel: Konstanzmethode

- Ziel ist nicht die Angabe der Häufigkeit einer "Ja"-Antwort in der Stichprobe, sondern die allgemeine Charakterisierung von deren Auftretenswahrscheinlichkeit
- Frage: Welche Wahrscheinlichkeiten sind im Lichte der Information aus der Stichprobe noch plausibel?

Beispiel: Elektrodermale Aktivität

Messung des Hautwiderstands in kOhm bei Tag (X) und bei Nacht (Y) ergibt bei einer Stichprobe von 5 Personen folgende Werte

X_i	24	28	21	27	23
Y_i	20	25	15	22	18

Frage

- (Wie stark) ist der Hautwiderstand bei Nacht vermindert?

Offensichtlich gilt für die Mittelwerte in der Stichprobe $\bar{x} > \bar{y}$

Zur Beantwortung der Frage muss man aber über die gegebene Stichprobe hinaus gehen und auf die Verhältnisse der betrachteten Population schließen

Beispiel: Elektrodermale Aktivität

Die Inferenz auf die Verhältnisse in der Population erfordert eine wahrscheinlichkeitstheoretische Modellierung der Situation

- Die Merkmale X und Y werden beispielsweise beide als normalverteilt angenommen, also $X \sim N(\mu_x, \sigma_x^2)$ und $Y \sim N(\mu_y, \sigma_y^2)$
- Die gestellte Frage bezieht sich dann auf die Beziehung $\mu_x > \mu_y$

Diese Modellierung ergibt sich häufig nicht zwangsläufig aus Situation und Art der Datenerhebung, sondern ist gegebenenfalls empirisch zu begründen

- Es ist beispielsweise zu prüfen, ob die erhobenen Stichprobenwerte als Realisierungen einer normalverteilten Zufallsvariablen aufgefasst werden können

Statistischer Test

Statistische Tests sind statistische Entscheidungsverfahren, die Aufschluss über die Verhältnisse in der Population geben sollen

Im Rahmen einer wahrscheinlichkeitstheoretischen Modellierung der betrachteten Situation bilden die erhaltenen Stichprobenergebnisse die Grundlage für diese Inferenz

- Die klassische Inferenzstatistik gründet sich auf die Annahme normalverteilter Merkmale (*parametrische Tests*)
- Ist die Annahme normalverteilter Merkmale nicht gerechtfertigt, so werden *nonparametrische Tests* angeboten
- Wir werden auch eine weitere häufige Klasse statistischer Tests besprechen (*Randomisierungstests* bzw. *Resamplingverfahren*), die ohne explizite Verteilungsannahmen auskommen und wesentlich stärkere Aussagen erlauben als die nonparametrischen Tests

Population und Stichprobe

Klärung der Begriffe auf zwei verschiedenen Ebenen

- Formale Ebene

  Die in Statistik I erarbeiteten wahrscheinlichkeitstheoretischen Konzepte bilden die Grundlage der Modellierung der Population, sowie der Definition des Begriffs der Stichprobe

- Inhaltliche bzw. anwendungsbezogene Ebene

  Wie kann man die betrachtete Population inhaltlich charakterisieren?

  Welche Möglichkeiten der Erhebung einer Stichprobe existieren?

Population und Stichprobe

Formal betrachtet beziehen sich die statistischen Annahmen zur Population auf die Verteilung des interessierenden Merkmals in der Grundgesamtheit von Untersuchungseinheiten (also auf das Model und dessen Parameter)

Statistische Inferenzen hinsichtlich der Population erfolgen über die Ziehung einer (einfachen) Zufallsstichprobe (random sampling)

- Jedes Element der Grundgesamtheit hat dabei unabhängig von den übrigen die gleiche Wahrscheinlichkeit in die Stichprobe aufgenommen zu werden

Ausgehend von Stichprobenkennwerten sollen dann Inferenzen bezüglich der Populationsparameter gezogen werden

Hierzu wird aus der Verteilung in der Population die Verteilung der Stichprobenkennwerte abgeleitet (sampling distribution)

Population und Stichprobe

Die Inferenz von Stichprobenkennwerten auf Populationsparameter, beruht entscheidend auf der Ziehung einer Zufallsstichprobe

> *Häufig sind bei sozialwissenschaftlichen Forschungsfragen nicht alle Untersuchungseinheiten, die zu einer Population gehören, bekannt, so dass die Ziehung einer "echten" Zufallsstichprobe unmöglich oder doch zumindest mit einem unzumutbaren Aufwand verbunden ist. Man begnügt sich deshalb gelegentlich mit sog. "anfallenden" oder "ad hoc"-Stichproben (...) in der Hoffnung, auch so zu aussagefähigen Resultaten zu gelangen. Vor dieser Vorgehensweise sei nachdrücklich gewarnt.* *(Bortz, 1999, S. 87)*

Redefinition der entsprechenden Population als Ausweg?

> *...letztlich lässt sich für jede "Stichprobe" eine fiktive Population konstruieren, für die diese "Stichprobe" repräsentativ erscheinen mag. Die Schlüsse, die aus derartigen Untersuchungen gezogen werden, beziehen sich jedoch nicht auf real existierende Populationen und können deshalb wertlos sein.* *(Bortz, 1999, S. 87)*

Verfahren der Stichprobenziehung

Während eine formale Definition der Begriffe Population und Stichprobe problemlos möglich ist, offenbart deren inhaltliche Interpretation grundlegende Probleme

Insbesondere die Festlegung der als Population betrachteten Grundgesamtheit ist nicht ohne weiteres möglich

- Was ist beispielsweise die Grundgesamtheit in der oben betrachteten Untersuchung der elektrodermalen Aktivität?

Nicht nur deswegen stellt sich die Frage nach der grundsätzlichen Realisierbarkeit einer (einfachen) Zufallsstichprobe, deren Vorliegen bei den Verfahren der klassischen Inferenzstatistik vorausgesetzt wird

Häufig werden in der Praxis leichter umsetzbare Verfahren zur Stichprobenziehung verwendet

Geschichtete Zufallsstichprobe

Zerlegung der Grundgesamtheit in nicht überlappende Schichten (z. B. anhand der Werte einer Schichtungsvariable)

Anschließende Ziehung einer Zufallsstichprobe in jeder der Schichten

Geschichtete Zufallsstichproben sind oftmals informativer als einfache Zufallsstichproben (d. h. sie führen zu genaueren Parameterschätzungen), was als *Schichtungseffekt* bezeichnet wird

Der Schichtungseffekt hängt von vielen Faktoren ab, u. a. auch von der Schichtungsvariable (sollte mit der interessierenden Variable hoch korrelieren und zu homogenen Schichten führen)

Klumpenstichprobe

Im Gegensatz zur "künstlich" erzeugten Zerlegung bei geschichteten Zufallsstichproben liegt oftmals eine natürliche Schichtung vor

- Diese natürlichen Gruppierungen werden auch als *Klumpen* bezeichnet (z. B. Gemeinden)

Es erfolgt eine Zufallsauswahl der Klumpen aus deren Grundgesamtheit

Innerhalb der Klumpen wird dann eine Vollerhebung durchgeführt

Keine Verfälschung der Ergebnisse, wenn die Binnenstruktur der Klumpen sehr heterogen bezüglich des interessierenden Merkmals ist und die Klumpen untereinander sehr homogen sind

Mehrstufiges Auswahlverfahren

Eine direkte Ziehung der Stichprobenelemente ist häufig nur schwer (oder gar nicht) umsetzbar

In diesen Fällen wird auf ein mehrstufiges Auswahlverfahren zurückgegriffen, bei der auf verschiedenen Ebenen eine wiederholte Zufallsauswahl stattfindet

Beispiel: Waldschadensforschung

- 1. Stufe: Auswahl einzelner Forstbezirke
- 2. Stufe: Auswahl einzelner Waldstücke in den gewählten Forstbezirken
- 3. Stufe: Auswahl einzelner Bäume in den gewählten Waldstücken

Quotenauswahl

Eine bei Meinungsumfragen häufig eingesetzte Form der Erhebung, die aber *nicht* mehr als zufällige Auswahl angesehen werden kann

Zur Erhöhung der Repräsentativität der Stichprobe, erfolgt die Auswahl anhand von *Quoten*, die eine Verteilung bestimmter Merkmale in der Stichprobe festlegen

- Beispiel: Für eine Untersuchung unter Psychologiestudierenden sollte die Stichprobe die Geschlechterverteilung der Grundgesamtheit widerspiegeln

Quotenpläne beschreiben die Auswahl hinsichtlich verschiedener relevanter Merkmale

Unter Einhaltung der Quoten kann der Interviewer die Stichprobenelemente frei auswählen (Problem der Subjektivität)

Statistik II: Themen

- Stichprobentheorie
- Parameterschätzung (Punktschätzung)
- Konfidenzintervalle (Intervallschätzung)
- Statistisches Testen von Hypothesen
- Regressionsanalyse
- Varianzanalyse
- Kovarianzanalyse
- Allgemeines lineares Modell