

Studienbrief 2

Grundlagen quantitativer Datenanalyse

Wiederholungsfragen Studienführer 1

Aufgabe 2.0:

Bearbeiten Sie die ersten 3 Wiederholungsfragen unter folgenden Link:

https://fomshinyapps.shinyapps.io/Lernfortschrittskontrolle_01/



Konnten Sie die Fragen gut beantworten? Toll!

Grundlagen quantitativer Datenanalyse (Kapitel 3 im Skript)

Dieser Studienführer behandelt die Folien 33 bis 71 im Skript. Was sollen Sie sich in dieser Einheit erarbeiten:

Lernergebnisse

Nach erfolgreichem Abschluss des ersten Kapitels sind Sie in der Lage,

- eine Messung zu definieren und verschiedene Typen von Skalenniveaus sicher zu unterscheiden.
- Grundbegriffe der quantitativen Datenanalyse (wie *Unabhängige Variable*, *Abhängige Variable*) und Probleme erklären, die aufgrund der Nichtberücksichtigung von Drittvariablen auftreten können.
- zentrale Begriffe um Stichproben(arten) und Populationen zu erläutern.
- die Unterschiede hinsichtlich Zufalls und Kausalität bei Beobachtungs- vs. Experimentalstudien zu erkennen.
- die Phasen des PPDAC-Zyklus zu nennen und anhand von Beispielen zu illustrieren.



Fangen Sie mit den Folien 35 bis 42 an, hier geht es um die Variablen (also das, wie Werte gespeichert werden) und wie diese gemessen werden. Auch hier tauchen Gütekriterien auf, jetzt die der Messung.

Aufgabe 2.1:

Vergleichen Sie die Gütekriterien der Messung mit denen der Forschung. Wo gibt es Gemeinsamkeiten, wo Unterschiede?



Folie 41 zeigt exemplarisch, was bei einer Messung „schief“ gehen kann. Entweder streut das Messergebnis (Varianz) oder die Messung hat eine Abweichung (Verzerrung oder Bias), beispielsweise eine Waage, die verstellt ist.

Auf den Folien 43 bis 47 geht es um das Skalenniveau von Daten, d. h. um den Informationsgehalt. Können Sie eine Eigenschaft nur zuordnen oder können Sie die Eigenschaften ordnen? Können nur die Abstände inhaltlich sinnvoll interpretiert werden oder auch die Quotienten (z. B. 2 m ist doppelt so viel wie 1 m, aber 20 °C ist nicht „doppelt“ so warm wie 10 °C – inhaltlich ist die zweite Aussage nicht sinnvoll).

Tip: Eigenschaften werden in Kategorien angegeben, daher **kategoriale** Daten. Wenn Sie etwas messen können (quasi mit dem Maßband oder dem Zollstock), sind es **metrische** Daten.



Aufgabe 2.2:

Hören Sie zur Entspannung den Song auf Folie 47.



Tip: Im Skript finden Sie an vielen Stellen Verweise auf Songs und Videos. Wenn Sie Spaß daran haben, hören oder schauen Sie sich diese an. Es lockert den Ablauf etwas auf und erreicht eine zusätzliche Verfestigung des Lernstoffes.



Schauen Sie sich die Folie 48 an: Ein Datensatz ist meist wie ein Tabellenkalkulationsblatt aufgebaut. Daten werden häufig auch in Formaten abgespeichert, die von Tabellenkalkulationsprogrammen (wie Excel, Numbers o. ä.) gelesen werden können.

Aufgabe 2.3:

Lesen Sie Kapitel 1.2 aus dem Buch *Introductory Statistics*. Sie können es hier herunterladen: https://www.openintro.org/stat/textbook.php?stat_book=isrs.



Puh, das ist ja auf Englisch! Ja, Sie haben Recht, aber es liest sich gut (viele englischsprachige Bücher lassen sich eigentlich besser lesen als vergleichbare deutschsprachige) und passt inhaltlich sehr gut zu unserer Veranstaltung.

Und nebenbei können Sie so Ihre Kompetenzen erweitern. 😊

Folie 49 greift wieder die Variation auf (das hatten wir schon bei der Messung der Daten). Aber Daten können auch aus anderen Gründen streuen.

Ziel in der quantitativen Datenanalyse ist häufig, beobachtete Daten zu modellieren, d. h., ein Modell zu finden, was einen Teil der Variation erklärt.

Aufgabe 2.4:

Überlegen Sie sich, warum Punkte einer Klausur variieren.



Wir versuchen oft, den Zusammenhang zwischen verschiedenen Variablen herzustellen (z. B. Punkte in einer Klausur hängen ab von dem Zeitaufwand, mit dem gelernt wurde). Das zeigt Folie 50.

Dass Sie bei der Auswahl der Variablen aufpassen müssen, zeigen die Folien 51 und 52. Das ist ein Beispiel des sogenannten Simpson-Paradox

Aufgabe 2.5:

Recherchieren Sie, wer Simpson war und warum das Paradox nach ihm benannt wurde.



Die Folien 54 bis 68 beschäftigen sich mit der Datenerhebung. Sie lernen Begriffe wie Stichprobe und Population kennen und wie Sie Stichproben ziehen können.

Warum ist es ideal, eine Zufallsstichprobe zu haben? Wenn ich Ihren Kurs als Stichprobe für die Studierenden der FOM in Frankfurt nehmen würde, könnten meine Ergebnisse verzerrt sein (z. B. wenn ich mich für das Monatseinkommen interessiere). Ich habe nur Bachelorstudierende und nur aus dem Tagesstudium. Besser wäre es hier z. B. eine geschichtete Stichprobe (nach Studiengängen und Zeitmodellen) zu ziehen.

In anderen Fragestellungen mag die Gelegenheitsstichprobe *Ihr Kurs* nicht verzerrt sein. z. B. wenn ich mich für die Körpergröße interessiere. Es hängt also immer auch von der Fragestellung ab, ob eine Gelegenheitsstichprobe geeignet ist.

Ein Experiment erlaubt direkt einen Kausalschluss, was in der Beobachtungsstudie nicht so ohne weiteres möglich ist. Eigentlich wollen wir aber in der Regel wissen, warum etwas passiert (also z. B. eine längere Lernzeit führt zu mehr Klausurpunkten). Deshalb greifen wir das hinten im Kapitel *Einführung in die kausale Inferenz* auf und lernen, wie auch in Beobachtungsstudien Kausalschlüsse gezogen werden können.

Aufgabe 2.6:

Zur Vertiefung können Sie die Kapitel 1.3 bis 1.5 des Buches *Introductory Statistics* lesen.



Die Folie 69 zeigt Ihnen ein paar griechische Buchstaben, die in der Statistik häufig verwendet werden. Wir werden diese später einsetzen, wenn formale Hypothesen

zu unseren Forschungsfragen formuliert werden sollen – aber das hat noch ein bisschen Zeit.

Tipp: Schon mal merken: **Griechische** Buchstaben werden in **Hypothesen** verwendet, da diese sich immer auf die **Population** beziehen. Lateinische Buchstaben werden für Aussagen zur Stichprobe verwendet.



Als letzten Punkt zeigt Folie 70 den PPDAC. Welches Problem soll untersucht werden (P), welche Variablen sind dafür notwendig (P), wie kommen Sie an die Daten (D). Über diese drei Punkte können Sie sich in der Übung 17 schon einmal Gedanken machen, da die Praxistransferaufgabe der Klausur im Wesentlichen auf dem PPDAC basieren wird.

Zusammenfassung

Halten wir fest, was Sie im Kapitel Grundlagen quantitativer Datenanalyse gelernt haben:

- Sie können die Skalenniveaus von Daten zuordnen.
- Sie verstehen, worauf Sie bei einer Messung achten müssen.
- Sie können zwischen unabhängigen und abhängigen Variablen sowie Kovariablen unterscheiden.
- Sie wissen, warum Sie eine Stichprobe benötigen und was Sie bei dem Ziehen einer Stichprobe beachten müssen.
- Sie haben gelernt, dass Kausalschlüsse (A verursacht B) ohne weiteres nur bei Experimenten möglich sind.
- Sie können den PPDAC (zumindest PPD) auf eine Fragestellung aus Ihrer beruflichen Praxis anwenden.

Abschluss

Denken Sie bitte daran, dass Sie bei der nächsten Sitzung R und RStudio entweder lokal auf Ihrem Rechner benötigen oder der Zugang zur RStudio-Cloud sichergestellt sein muss. Bitte installieren Sie bei der lokalen Variante auch das im OC hinterlegte RStudio-Projekt.

Lösungshinweise

Aufgabe 2.1

Sowohl bei der Messung als auch bei der Forschung gibt es Transparenz bzw. Reproduzierbarkeit (bei der Messung bezeichnen wir das als Reliabilität), Objektivität und Validität. Letztere wird bei der Forschung noch unterschieden in interne und externe Validität. Zusätzlich gibt es bei der Messung noch die Genauigkeit einer Messung und bei der Forschung spielen auch die ethischen Aspekte eine wichtige Rolle.

Aufgabe 2.4

Unterschiedliche Lernzeiten der Studierenden, unterschiedliche Verfassung am Tag der Klausur, unterschiedliche Fachkompetenz, ...

Aufgabe 2.5

Edward Hugh Simpson war ein britischer Statistiker, der das Phänomen 1951 beschrieb. Ihm fiel auf, dass die Bewertung verschiedener Gruppen unterschiedlich ausfällt in Abhängigkeit davon, ob die Gruppen kombiniert werden oder nicht.

Hinweise

Dieser Studienbrief wurde von Matthias Gehrke auf Basis des Konzepts von Roswitha Grassl (beide FOM, <https://www.fom.de/>) entwickelt und stehen unter einer Creative Commons Namensnennung – Nicht-kommerziell – Weitergabe unter gleichen Bedingungen 3.0 Lizenz (<https://creativecommons.org/licenses/by-nc-sa/3.0/legalcode.de>).