Herbst 2018

Konzept der Veranstaltung „Methoden der Datenanalyse I“

Dr. Manuela Pötschke
UNIVERSITÄT KASSEL; ARBEITSGEBIET ANGEWANDTE STATISTIK

Vorwort

Viele Studierende, manche Kolleginnen und Kollegen, aber vor allem Menschen außerhalb der Universitäten denken, dass eine immer wiederkehrende Veranstaltung für die Dozentin keinerlei oder nur wenig Vorbereitungsarbeit mit sich bringt, weil die Inhalte ja von Jahr zu Jahr gleichblieben. Für die Konzeption der Lehrveranstaltungen zu den Methoden der Datenanalyse am Fachbereich 05 der Universität Kassel trifft das definitiv nicht zu. Vielmehr werden die Inhalte und vor allem die Didaktik der Veranstaltungen von Semester zu Semester weiterentwickelt. In die inhaltlichen Weiterentwicklungen fließen aktuelle Methodendiskussionen ein. Die didaktische Weiterentwicklung orientiert sich an Erkenntnissen der Lernforschung und an den Ergebnissen der kontinuierlich durchgeführten Forschung im Rahmen des Projektes „Statistik lernen" (vgl. www.uni-kassel.de/go/statistik_lernen). Die Lernforschung zeigt z.B., dass das exemplarische Lernen an aktuellen Beispielen besser geeignet ist, auch abstrakte Konzepte zu verstehen, als der generelle Bezug auf formale statistische Modelle. Deshalb besteht eines der grundlegenden Prinzipien der Datenanalyseveranstaltungen im Zugang über Beispiele aus relevanten sozialwissenschaftlichen Surveys. Das erfordert dann für jede Veranstaltung die neue Aufbereitung passender Beispiele. Auch zeigt die neuere Lernforschung, dass der Einsatz multimedialer Mittel in der Lehre hilfreich sein kann. Dabei sind die Facetten der Wirksamkeit spezifischer Anwendungen aber noch nicht vollständig ausgeleuchtet. Hier hat die Veranstaltung auch experimentellen Charakter. Neuerungen werden ausprobiert und in den Kanon der zu verwendenden Methoden aufgenommen, wenn sie sich bewehren. Im anderen Fall werden sie im nächsten Durchgang wieder verworfen.

Insgesamt bedeutet das, dass ausgehend von einem Grundstock an konzeptionellen Überlegungen jede Veranstaltung wieder neu überdacht und angepasst wird. Das vorliegende Paper ist dem Grundstock gewidmet und stellt die basalen Überlegungen zur Veranstaltung „Methoden der Datenanalyse I" vor.

Inhaltsverzeichnis

Formale Einordnung der Veranstaltung „Methoden der Datenanalyse I" 1

Grundlegende didaktische Prinzipien der Veranstaltung 1

Inhaltliches Konzept der Veranstaltung 3

Formale Einordnung der Veranstaltung „Methoden der Datenanalyse I“

Die Veranstaltung richtet sich an Studierende in den Hauptfächern Soziologie und Politikwissenschaft und an Studierende anderer Hauptfächer mit dem Nebenfach Soziologie. Die empfehlenden Studienpläne sehen die Veranstaltung im 2. (Hauptfach Soz.) bzw. 4. Semester (Hauptfach Powi und Nebenfach Soz.) vor.

Die Veranstaltung umfasst eine Plenumssitzung á 4 SWS und ein Tutorium á 2 SWS. Die Plenumssitzung beinhaltet Inputsequenzen und Anwendungssequenzen.

Die zentralen Lernziele der Veranstaltung sind:

- Studierenden können die deskriptiven Ergebnisse aus empirischen Studien anderer nachvollziehen und verstehen und die verwendeten Maße und Darstellungsformate kritisch beurteilen.
- Studierende können für eigene empirische Fragen angemessene deskriptive Darstellungsformen auswählen und korrekt interpretieren.
- Studierende können das Grundprinzip des statistischen Testens beschreiben und spezifische Hypothesentests angemessen interpretieren.

Um die Ziele zu erreichen, liegen der Veranstaltung grundlegende didaktische Prinzipien zugrunde.

Grundlegende didaktische Prinzipien der Veranstaltung

Komplexe Methoden der Datenanalyse werden auf der Basis von Beispielanwendungen eingeführt.

Der Bezug auf praktische Beispiele erfolgt in verschiedener Hinsicht. Ausgangspunkt der Einführungen sind Anwendungen der Verfahren, die in Fachartikeln verwendet wurden. Dadurch ist der Zugang konkret und inhaltlich nachvollziehbar, da die Ergebnisse der statistischen Analysen auch verbalisiert vorliegen. Danach schließen sich weitere Beispiele zur Einführung an, die dann den Softwareoutput stärker in den Fokus nehmen. Diese Beispiele basieren auf einem zentralen Survey für die Sozialforschung, dem ALLBUS. Der ALLBUS hat mehrere Vorteile für Studierende. Der Datensatz selbst und zugehörige Dokumentationen sind leicht zugängig (jeder Studierende kann sich die Daten, den Fragebogen und Methodenmaterial nach einer Anmeldung auf www.gesis.org herunterladen). Der Datensatz umfasst Informationen zu aktuellen Themen von allgemeinem Interesse und er ist dabei nicht zu komplex. In den Tutorien können die Studierenden dann eigenen Fragestellungen auf der Basis des ALLBUS oder weiterer Surveys nachgehen und damit stärker ihre eigenen Interessen einbringen. Das wird möglich, weil die Tutorien systematisch in die Nutzung von Softwareanwendungen zur Datenanalyse einführen. Im Mittelpunkt stehen dabei (in den Semestern in unterschiedlicher Gewichtung) die Programme SPSS, R und STATA.

Der Zugang zu spezifischen Verfahren erfolgt über eine Herausforderung oder Aufgabe, vor der die Datenanalyse steht.

Die inhaltliche Gliederung der Veranstaltung orientiert sich an den Abschnitten eines empirischen Berichts. Dadurch gewinnen die Studierenden nicht nur Sicherheit in der Anwendung der einzelnen Verfahren, sondern auch in der Strukturierung eines empirischen Forschungsvorhabens. Für die Abfassung empirischer Berichte, Hausarbeiten und Abschlussarbeiten ist das eine essentielle Kompe-

tenz (vgl. das Paper „Hinweise für Forschungsberichte und empirische Abschlussarbeiten“ in diesem Material).

Die Einführung ausgewählter Themen erfolgt problemorientiert und selbst entdeckend.

Oft werden in der universitären Einführungslehre (vermeintliche) fachliche Probleme vom Dozierenden vorgetragen, um direkt im Anschluss die Lösung für das Problem darzustellen. Für zwei Themen – Einführung in Standardabweichung /Varianz und Auswahl angemessener statistischer Maße - wird dieses Vorgehen aufgegeben. Die Studierenden werden vielmehr mit der Aufgabe konfrontiert, für einige Variablen und Verteilungen Maße zur Beschreibung von Heterogenität selber zu entwickeln und Prinzipien abzuleiten, welche statistischen Maße bei welchen Variablen hilfreich und angemessen sind.

Neuere Erkenntnisse der Lernforschung werden in die Konzeption aufgenommen und die Lehre entsprechend gestaltet.

Der Schwerpunkt liegt dabei auf „wünschenswerten Erschwernissen“ im Lernprozess, weil empirische Studien deren Wirkungen auf einen nachhaltigen Lernerfolg zeigen konnten.[1] Zu den vorgesehenen Erschwernissen zählen Aufgaben zur Generierung eigener Lösungen zu statistischen Problemen, der Zugang zu den Gegenständen über verschachteltes Lernen und die Konstruktion von Testgelegenheiten. Das verschachtelte Lernen schlägt sich dabei auch in der Präsentation der Themen nieder. Diese erfolgt aufgaben- und problemorientiert, wobei in einem ersten Schritt mögliche statistischen Konzepte gesammelt und sie erst danach systematisiert werden.

Testgelegenheiten bieten sich durch die Bereitstellung eines umfangreichen Onlinetests, der mehrfach und parallel zur Veranstaltung eigenständig bearbeitet werden kann, durch die gemeinsame Besprechung der Lösungen einer Probeklausur zu Beginn des letzten Drittels der Veranstaltung und durch Aufgabenblätter, die zum Teil im Tutorium bearbeitet werden können.

Darüber hinaus war in die Veranstaltung bisher die begleitende Übung von Veranstaltungsinhalten nach dem Prinzip des verteilten Lernens über eine Kooperation mit der Kollegin Barzagar-Nazari eingebunden. Die Übungen werden solange angeboten, wie die Kooperation besteht.

Der Heterogenität in den Voraussetzungen wird mit einem Set von Lernunterstützungen begegnet.

Dazu zählen E-Learning-Angebote, Literaturempfehlungen, Handlungsanleitungen und Testgelegenheiten. Die Bereitstellung der Unterstützungen erfolgt über die Homepage „von Studis für Studis“ (www.uni-kassel.de/go/statistik_lernen) und über die Lernplattform Moodle.

Den E-Learning-Angeboten kommt hier eine besondere Funktion zu. Entwickelt wurden multimediale Szenarien, Videofilme und Web Based Trainings, die jeweils spezifische Ziele verfolgen. Die multimedialen Szenarien greifen thematisch einzelne Überlegungen im Zusammenhang mit statistischen Themen auf und stellen diese in aller Ausführlichkeit dar. Bei der Konzeption dieser Anwendungen haben wir uns von typischen Fehlern oder Problemen bei der Aneignung statistischen Grundlagenwissens durch Studierende leiten lassen.[2] Die Filme stellen statistische Konzepte mit dem Verweis auf Alltagssituationen der Studierenden dar oder weisen einen eher motivationsfördernden Charakter auf. Die aktuellsten Entwicklungen sind WBTs. Diese sind als eigenständige Lehreinheiten zu verste-

[1] Vgl. Lipowsky, F., Richter, T., Borromeo Ferri, R., Ebersbach, M. & Hänze. M. (2015). Wünschenswerte Erschwernisse beim Lernen. Schulpädagogik Heute, 11.

[2] Z.B. stellt für einige Studierende der Übergang aus der Datenmatrix zur Häufigkeitstabelle ein Problem dar. Ihm ist deshalb ein eigenes Szenario gewidmet.

hen, die jeweils ein Thema in den Blick nehmen. Neben der interaktiven Erarbeitung spielen die Übungen hier eine besondere Rolle.

Eine besondere Verbindung stellen Quiz über die digitale Anwendung ARSnova (vgl. https://arsnova.thm.de/mobile) dar. Zu Beginn einer Veranstaltung eingesetzt, können Fragen zu den Inhalten vergangener Sitzungen an das Bisherige anknüpfen und zur Wissensaktivierung beitragen. In der Veranstaltung kann die Anwendung genutzt werden, um den Lernfortschritt zu dokumentieren und auf fehlende Links in den kognitiven Verknüpfungen der Studierenden hinzuweisen. Zum Ende wird die Anwendung vor allem als Instrument der Ergebnissicherung eingesetzt.[3]

Ebenfalls zur Lernunterstützung wird begleitend zu allen Sitzungen auf Beiträge aus zwei neueren Handbüchern zu Methoden und auf angewandte Beiträge aus der Kölner Zeitschrift für Soziologie und Sozialpsychologie verwiesen. Die Handbuchbeiträge thematisieren den Gegenstand der jeweiligen Sitzung und eignen sich sowohl für die Nachbereitung der Sitzung als auch für deren Vorbereitung. Zu den Anwendungsbeiträgen werden Lesefragen bereitgestellt, die die Erfassung der wesentlichen Überlegungen der Beiträge unterstützen sollen und gleichzeitig einen anderen Zugang zu den besprochenen Inhalten erlauben.

Die Lernfortschritte werden durch eine begleitende Lernforschung dokumentiert und analysiert.

Im Rahmen der Veranstaltung werden die Studierenden zweimal befragt: zu Beginn des Semesters nach ihren Vorkenntnissen, Einstellungen und Erwartungen bezogen auf das Themenfeld der Datenanalyse und die Veranstaltung. Am Ende des Semesters werden sie gebeten, den Lernprozess und ihre Klausurvorbereitung zu reflektieren und einzuschätzen, welche Unterstützungsmaßnahmen ihnen wie geholfen haben. Neben diesem Kranz an regelmäßigen Indikatoren werden besondere Fragekomplexe aufgenommen, wenn in den Veranstaltungen didaktische Neuerungen probiert wurden. Ergänzt werden die Befragungsdaten durch spezifische und allgemeine Ergebnisse aus der Klausur.

Inhaltliches Konzept der Veranstaltung

Das inhaltliche Konzept orientiert sich an den Strukturen empirischer Aufsätze (vgl. das Paper „Hinweise für Forschungsberichte und empirische Abschlussarbeiten“ in diesem Material) und an den Schritten, die im Rahmen einer eigenen Datenanalyse bis zur Interpretation von Erklärungsmodellen zu gehen sind. Diese Schritte umfassen die Operationalisierung von latenten Konstrukten, die theoretisch abgeleitete Hypothesen generieren. Ohne hier explizit ausgewählte Theorie ausführlich besprechen zu können, ist der Link von theoretischen Überlegungen hin zu Daten und Datenbeschreibung herzustellen. Dieser Link wird über den Prozess der Operationalisierung sichergestellt. Dieses Thema wird dabei in enger Anknüpfung an die Überlegungen in der Veranstaltung zu „Methoden der Datenerhebung“ besprochen. Ein besonderes Augenmerk legen wir in diesem Zusammenhang darauf, die Studierenden mit ausgewählten wichtigen Surveys für die empirische Sozialforschung bekannt zu machen. Die Ziele sind dabei einerseits, den Studierenden die Bandbreite möglicher nutzbarer empirischer Informationen zu vergegenwärtigen und andererseits die Komplexität der Daten zu verdeutlichen, damit die Studierenden in eigenen Analysen den Auf für das Datenhandling besser einschätzen können.

[3] ARSnova bietet weitere Anwendungen, die in der Veranstaltung bisher jedoch nicht genutzt werden. Die Nutzung von Feedbackfragen und Einschätzungen der Studierenden wird im nächsten Semester erprobt und bei positiver Evaluation in das allgemeine Konzept aufgenommen. Der Grad der Interaktivität wird dadurch deutlich erhöht. Die Anforderungen an die Flexibilität der Dozentin steigt enorm an.

Anschließend werden drei prinzipielle Möglichkeiten der angemessenen univariaten Beschreibung (Deskription) empirischer Verteilungen eingeführt. Empirische Verteilungen geben Auskunft darüber, wie viele Fälle auf die einzelnen Kategorien einer Variable entfallen. Zur Beschreibung können Tabellen, Grafiken und statistische Lage- und Streuungsmaße herangezogen werden. Welche Form gewählt wird hängt davon ab, ob die Verteilung vollständig berichtet werden soll (Grafiken und Tabelle) oder ob typische Werte und Heterogenitätsinformationen gegeben werden sollen (Lage- und Streuungsmaße). Welche konkreten Formen gewählt werden hängt dann vom Skalenniveau der Variablen ab. Da die Skalenniveaus auch für weitere Entscheidungen im Prozess statistischer Analysen eine herausragende Bedeutung haben, wird ihrer Einführung breiter Raum gegeben.

Die univariate Einführung wird in zwei Hinsichten erweitert: zum einen werden bivariate Verteilungen betrachtet und nach Zusammenhängen gefragt und zum anderen wird thematisiert, was uns beobachtete Stichprobenergebnisse über Sacherhalte in der nicht beobachteten Grundgesamtheit sagen können.

Aus diesen Überlegungen ergibt sich die folgende thematische Gliederung:

- Datenanalyse I im Forschungsprozess
- Hypothesen und Operationalisierung
- Univariate Beschreibung von Verteilungen
- Skalenniveaus
- Bivariate Beschreibung von Verteilungen
- Grundlagen des Schätzens und Testens
- Anwendung spezifischer Hypothesentests