

Grundlegende Methoden der Sozialstatistik

Als Sozialstatistik wird die systematische und regelmäßige Beobachtung und Analyse von Lebensbedingungen bestimmter Bevölkerungsgruppen oder der gesamten Bevölkerung verstanden. Datenbasis für Dauerbeobachtungen sind in der Regel zielgerichtete Erhebungen und administrative Daten. Nur in seltenen Fällen können randomisierte Studien zur Informationsgewinnung herangezogen werden. In dieser Hinsicht unterscheidet sich die Sozialstatistik grundlegend von anderen Bereichen der Statistik. Selektive Teilnahme, Antwortverweigerung, Messfehler und variierende Wahrscheinlichkeiten für eine Erhebung ausgewählt zu werden erschweren eine Auswertung der erhobenen Daten. Die beiden Veranstaltungen Grundlegende Methoden der Sozialstatistik A und B haben das Ziel, eine Einführung in die speziellen Problemstellungen der Sozialstatistik zu geben und den Studierenden das statistische Instrumentarium zu vermitteln, das eine valide Analyse der erhobenen Daten ermöglicht.

Die Veranstaltung **Grundlegende Methoden der Sozialstatistik B** setzt sich mit den verschiedenen Fehlerquellen auseinander, die die Validität der Analyseergebnisse auf Basis von Befragungsdaten unterminieren können. Neben dem Stichprobenfehler, dem in allen Teildisziplinen der Statistik eine große Bedeutung beigemessen wird, sind hier insbesondere Messfehler und Fehler, die durch Antwortausfall entstehen, zu berücksichtigen. Im Rahmen des Kurses werden die einzelnen Fehlerquellen vorgestellt und verschiedene Maßnahmen diskutiert, um deren Ausmaß möglichst gering zu halten. Zudem werden Verfahren präsentiert, die es ermöglichen, die Fehler bei der Analyse zu berücksichtigen und somit valide Ergebnisse zu erzielen. Ein Schwerpunkt der Veranstaltung liegt auf dem korrekten Umgang mit Antwortausfall. Hier wird in der Literatur zwischen einer kompletten Teilnahmeverweigerung an der Befragung (*unit nonresponse*) und der Auskunftsverweigerung bezüglich einzelner Fragestellungen (*item nonresponse*) unterschieden. Beide Formen des Antwortausfalls können zu Verzerrungen in den Analyseergebnissen führen, wenn sie nicht adäquat berücksichtigt werden. Der Kurs stellt die beiden Verfahren zur Korrektur dieser Verzerrungen vor – Gewichtung für *unit nonresponse* und Imputation für *item nonresponse* – und diskutiert die dabei getroffenen Annahmen kritisch.

Neben Strategien zur Vermeidung und der adäquaten Berücksichtigung der Fehler, die im Rahmen des Erhebungsprozesses entstehen können, ist die Einhaltung des Datenschutzes ein wichtiger Aspekt in der Sozialstatistik. Wie können die Daten an Wissenschaftler und die interessierte Öffentlichkeit weitergegeben werden, ohne die bei der Befragung zugesicherte Anonymität zu gefährden? Um dies zu ermöglichen, werden bei der Datenbereitstellung seit vielen Jahren informationsreduzierende und Daten verändernde Verfahren eingesetzt. Diese Verfahren stoßen aber im Zeitalter von Big Data zunehmend an ihre Grenzen. In der Veranstaltung werden die verschiedenen gängigen Verfahren besprochen und neuste Entwicklungen vorgestellt, die auch in Zukunft eine breite Verfügbarkeit der Daten bei gleichzeitiger Gewährleistung des Datenschutzes sicherstellen sollen.