

SE Geographie und Ökonomie

Einheit 8: Datensätze und Datenquellen

Bernhard Schmidpeter

bernhard.schmidpeter@jku.at

Institut für Volkswirtschaftslehre

SS 2022

Klassifikation von Datensätzen (1)

- **Datengewinnung:**
 - ▶ **Primärdaten:** (neue) Daten werden z.B. mittels Befragung erhoben.
 - ▶ **Sekundärdaten:** bereits veröffentlichtes Daten werden verwendet.
- **Umfang:**
 - ▶ **Vollerhebung:** Daten werden zu allen Elementen der Grundgesamtheit erhoben.
 - ▶ **Teilerhebung (Stichprobe):** Daten werden nur zu einem Teil der Elementen der Grundgesamtheit erhoben.
- **“Dimension”** der Daten:
 - ▶ **Querschnittsdaten:** Mehrere Beobachtungseinheiten werden zu einem Zeitpunkt erhoben.
 - ▶ **Zeitreihendaten (Längsschnittdaten):** Eine Beobachtungseinheit wird zu mehreren Zeitpunkten (oder Zeitperioden) erhoben.
 - ▶ **Paneldaten:** Zweidimensionale Daten, wo Daten über mehrere Beobachtungseinheiten zu mehreren Zeitpunkten erhoben werden.

Klassifikation von Datensätzen (2)

- **Erhebungseinheit:**

- ▶ **Mikrodaten** (Einzeldaten): beziehen sich auf Erhebungseinheiten (Personen, Haushalte, Firmen).
 - ★ Umfragedaten: systematische Erhebung von Informationen für eine Stichprobe von einzelnen Erhebungseinheiten.
 - ★ Verwaltungsdaten: werden primär erhoben, um wirtschaftliche Aktivität zu messen (und darauf aufbauend Markteingriffe vorzunehmen, wie Subventionen zu vergeben oder Steuern einzuheben); z.B. Daten aus dem Steuerregister, Sozialversicherungsdaten ...
 - ★ Abgrenzung nicht immer eindeutig; Volkszählung passt etwa in beide Kategorien.
- ▶ **Makrodaten** (Gruppendaten): Daten werden auf Ebene von Gruppen (Sektoren, Gemeinden, Länder, Altersklassen) aggregiert.

Beispiele für Mikro-Datensätze

- Beispiele für **Haushaltspaneldaten** (Stichproben)
 - ▶ Deutsches Sozioökonomisches Panel (German Socio-Economic Panel; SOEP, seit 1989)
 - ▶ British Household Panel Survey (BHPS, seit 1991)
 - ▶ Panel Study of Income Dynamics (PSID, seit 1968)
- Beispiele für **Querschnittserhebungen** (Stichproben)
 - ▶ Konsumerhebung der Statistik Austria (seit 1999/2000, alle 5 Jahre): Sehr detaillierte Daten zu Konsumausgaben von Haushalten in Österreich.
 - ▶ European Social Survey (ESS, seit 2002, alle 2 Jahre): Erhebt Meinungen zu sozialen und politischen Themen aus über 30 europäischen Ländern.
 - ▶ Eurobarometer der Europäischen Kommission (seit 1973): Erhebt Daten zu sozialen und politischen Einstellungen in den Ländern der EU.

Beispiele für Mikro-Datensätze

Beispiele für **Vollerhebungen**

- Volkszählung der Statistik Austria: Von 1951 bis 2001 in Abständen von 10 Jahren durchgeführt.
 - ▶ Im Vergleich zu den Haushaltspaneldata: Sehr große Zahl an Erhebungseinheiten, aber (vergleichsweise) geringe Zahl an Merkmalen.
- In den 2000er Jahren: Übergang zu Registerdaten. Bei Registerdaten werden Datensätze erstellt, die auf administrativen Daten basieren, und nicht mehr extra erhoben werden müssen:
 - ▶ Zentrales Melderegister (ZMR) für Wohnbevölkerung
 - ▶ Gebäude- und Wohnungsregister
 - ▶ Arbeitsstättenzählung
 - ▶ Abgestimmte Erwerbsstatistik

Mögliche Datenquellen für Österreich

- Die **Statistik Austria** sammelt eine große Zahl an Daten in verschiedenen Bereichen.
 - ▶ Viele Datensätze sind **kostenpflichtig** oder (aus Gründen der Geheimhaltung) gar nicht oder nur anonymisiert zugänglich.
 - ▶ Üblicherweise werden **ausgewählte Makrodaten gratis** veröffentlicht. Die Mikrodaten werden dafür auf (zum Beispiel) Bundesländer oder Sektoren aggregiert.
 - ▶ Für Forschungs- und Lehrzwecke werden **Standardisierte (Mikro-)Datensätze** (SDS) unentgeltlich zu Verfügung gestellt, siehe dazu folgenden **Link**. Man muss dazu aber meist eine Datenanfrage an die Statistik Austria stellen.
- Die Seite **data.gv.at** bietet einen Katalog **offener Datensätze** und Dienste aus der öffentlichen Verwaltung, die auf den sogenannten **Open Data-Prinzipien** basieren. Diese Daten sind frei zugänglich und können von dieser Seite heruntergeladen werden. Der Nachteil dieser Seite ist, dass die (vielen!) Datensätze **schlecht strukturiert** sind.

Seminarteil

Übersicht:

- Projekt (in Teams zu 2 bis 3 Personen), indem eine ökonomisch relevante Forschungsfrage gestellt wird, die mit einem geeigneten Datensatz mit quantitativen Methoden untersucht und beantwortet wird.
- Begleitung (Coaching) durch LV-Leiter.
- kurzer Seminarbericht (ca. 10 Seiten) und Präsentation der Ergebnisse in letzter Einheit; Einarbeitung von Feedback.

Mögliche Datensätze:

- Konsumerhebung: ich habe eine Anfrage an die Statistik Austria gestellt, und würde den Datensatz über Moodle bereit stellen.
- European Social Survey (ESS): frei zugänglich; ich werde Daten zu Österreich von der 9. Welle (Jahr 2018) in Moodle bereit stellen.
- Bitte Dokumentation der Variablen beachten!
- Sie können auch gerne jeden anderen Datensatz verwenden!

Zeitplan

- Bis 29.05.2022: Abgabe von ersten Schätzungen und deskriptiven Auswertungen. Wenn schon (einzelne) Folien gemacht wurden, bitte auch abgeben.
 - ▶ Feedback zur Präsentation am 31.05.2022.
- Bis 12.06.2022: Abgabe der Präsentation. Zu diesem Zeitpunkt sollte die Präsentation schon weit gediehen sein.
 - ▶ Feedback zur Präsentation am 14.06.2022.
- Bis 19.06.2022: Abgabe der fertigen Präsentation.
 - ▶ Präsentation am 21.06.2022
- Bis 12.07.2022: Abgabe der Seminararbeit.
 - ▶ Am 28.06 besteht die Möglichkeit zusätzliches Feedback zur Seminararbeit zu bekommen.