

# *SE Geographie und Ökonomie*

## **Einheit 8: Datensätze und Datenquellen**

**Dieter Pennerstorfer**  
dieter.pennerstorfer@jku.at

Institut für Volkswirtschaftslehre  
Johannes Kepler Universität Linz



# 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: systematischen 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 **Haushaltspaneldata**n (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 4 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.
- **Präsentation der Ergebnisse** in den letzten beiden Einheiten und **kurzer Seminarbericht** (ca. 10 Seiten)
- Siehe den **Leitfaden für Seminararbeit.pdf** auf Moodle als Anleitung für die Erstellung des Seminarberichts, aber auch für die Ausarbeitung der Präsentation.
- Nehmen Sie Feedback als Verbesserungsvorschläge auf und berücksichtigen Sie diese Rückmeldungen bei der Überarbeitung von Präsentation und Seminarbericht.

# Seminarteil: Mögliche Datensätze

- **Konsumerhebung:**

- ▶ Testfile mit ca. 5 % der Haushalte, die bei der Konsumerhebung 2014/15 teilgenommen haben.
- ▶ Die Beobachtungseinheit ist ein Haushalt. Der Datensatz besteht aus 3 Teil-Datensätzen über die Ausgaben des Haushalts innerhalb einer 14-tägigen *Buchführungszeitraum* (TESTK14\_AUSGAGG), sowie Informationen über den Haushalt (TESTK14\_HHVAR) und die Haushaltsmitglieder (TESTK14\_PVAR).
- ▶ Die verschiedenen Datensätze können über eine eindeutige Haushaltskennzahl (*OB*) miteinander verknüpft werden.

- **European Social Survey (ESS):**

- ▶ Daten zu Österreich von der 9. Welle (Jahr 2018) sind auf Moodle bereit gestellt (generell sind Daten des ESS frei zugänglich).

- Bitte beachten Sie die Dokumentation der Variablen auf Moodle.

- Sie können auch gerne **jeden anderen Datensatz** verwenden!



# Seminarteil: Zeitplan

Zeitplan für die **Gruppen A und B** bzw. für die **Gruppen c und D**.

- Bis **08.01.2024** bzw. **15.01.2024**: Abgabe der Präsentation. Zu diesem Zeitpunkt sollte die Präsentation schon weit gediehen sein.
  - ▶ Feedback zur Präsentation am **09.01.2024** bzw. **16.01.2024**.
  - ▶ Termine finden um 18:15 Uhr (Gruppe **A** und **C**) bzw. 19:00 Uhr (Gruppe **B** und **D**)
- Bis **22.01.2024** bzw. **29.01.2024**: Abgabe der fertigen Präsentation.
  - ▶ Präsentation am **23.01.2024** bzw. am **30.01.2024**. Kurzes Feedback im Anschluss.
- Bis 11.02.2024: Abgabe der fertigen Seminararbeit.
- Schreiben Sie mir bis 20.12.2023 eine mail (= eine mail pro Gruppe), ob die Termine passen und ob Sie wollen, dass Feedbackgespräche Online oder in Präsenz abgehalten werden.

Feedbackgespräche dauern etwa 30 Minuten und finden (an den entsprechenden Tagen) immer Dienstag Abend statt (siehe Moodle für den genauen Zeitplan). Sollten Sie zu einem Besprechungstermin keine Zeit haben, versuchen Sie, selbständig mit einer andere Gruppe zu tauschen (und informieren Sie mich kurz) bzw. sagen Sie mir möglichst zeitnah Bescheid, falls das nicht möglich ist.