

SE Geographie und Ökonomie

Einheit 1: Einführung und Grundbegriffe

Dieter Pennerstorfer
dieter.pennerstorfer@jku.at

Institut für Volkswirtschaftslehre
Johannes Kepler Universität Linz Linz



Einordnung

Empirische Forschung ist die Konfrontation des (theoretischen) Modells mit (empirischen) Beobachtungen.

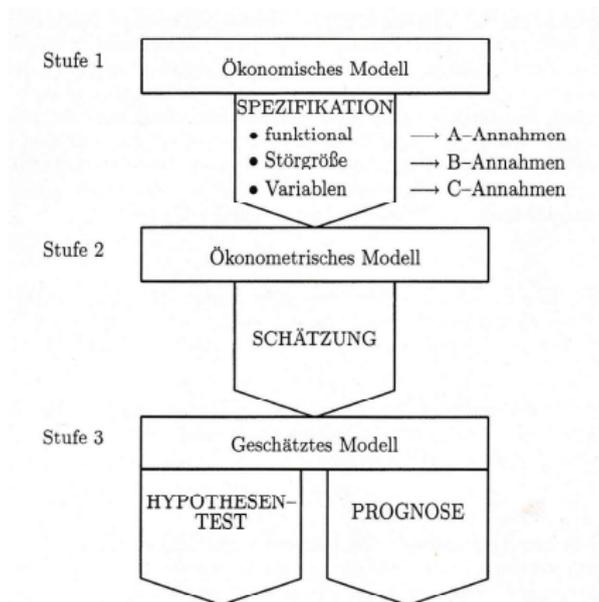


Abbildung 1.1: Die vier Aufgaben ökonometrischer Analyse.

Ziele der Lehrveranstaltung

Am Ende der Lehrveranstaltung können Sie:

- 1 eine **ökonomisch relevante Fragestellung** (oder Hypothese) formulieren, die einen vermuteten Zusammenhang zwischen zwei (oder mehreren) Variablen herstellt.
- 2 einen **geeigneten Datensatz** suchen und aufbereiten.
- 3 den Datensatz **deskriptiv auswerten**, um einen ersten Überblick über die Daten und den vermuteten Zusammenhang zu bekommen.
- 4 eine einfache **statistische Auswertung** durchführen (eine sogenannte **lineare Regression**), um den Zusammenhang fundierter zu untersuchen.
- 5 die **Ergebnisse** (in Bezug auf die Größe der Koeffizienten und die statistische Signifikanz, aber auch inhaltlich) **richtig interpretieren**. Sie können daraus auch etwaige Politikempfehlungen (Prognose) ableiten.

Ziele der Lehrveranstaltung: Beispiel (1)

1 Fragestellung:

Wohnen Single-Haushalte mit höherem Einkommen in größeren Wohnungen als Single-Haushalte mit niedrigem Einkommen? Gibt es Unterschiede zwischen Männern und Frauen?

Hypothese: Single-Haushalte mit höheren Einkommen wohnen in größeren Wohnungen, während es keine Unterschied zwischen den Geschlechtern gibt.

2 Datensatz:

Auszug aus dem *Deutschen Sozio-oekonomischen Panel* (GSOEP) des Jahres 2009. Das GSOEP ist eine Befragung einer Stichprobe der deutschen Bevölkerung. Der Datensatz wird in MS-Excel eingelesen.

3 Deskriptive Auswertung der Daten:

4 Schätzung einer linearen Regressionsgleichung:

$$WGr_i = \alpha + \beta_1 Eink_i + \beta_2 Frau_i + \epsilon_i$$

Auswertung erfolgt mit dem Excel-Add-Inn *Analyse-Funktion*.

Ziele der Lehrveranstaltung: Beispiel (2)

	Koef.	Std. Err.	t-Wert	p-Wert	95 % Konfidenzintervall	
Eink	0.81	0.11	7.46	0.00	0.60	1.02
Frau	5.10	2.35	2.17	0.03	0.49	9.71
Interzept	54.87	2.79	19.66	0.00	49.39	60.34

Ergebnisse und Interpretation:

- Ergebnisse: Single-Haushalte mit höherem Einkommen wohnen (im Durchschnitt) in größeren Wohnungen. Dieser Zusammenhang ist statistisch gut abgesichert. Das bedeutet, dass dieser positive Zusammenhang mit einer hohen Wahrscheinlichkeit auch in der Grundgesamtheit (alle Single-Haushalte in Deutschland) vermutet werden kann. Wenn das Einkommen eines Haushalts um 1,000 Euro höher ist, ist die Wohnungsfläche um durchschnittlich $0,8 \text{ m}^2$ größer. Single-Frauen wohnen in durchschnittlich 5 m^2 größeren Wohnungen als Single-Männer.

Ziele der Lehrveranstaltung: Beispiel (2)

	Koef.	Std. Err.	t-Wert	p-Wert	95 % Konfidenzintervall	
Eink	0.81	0.11	7.46	0.00	0.60	1.02
Frau	5.10	2.35	2.17	0.03	0.49	9.71
Interzept	54.87	2.79	19.66	0.00	49.39	60.34

Ergebnisse und Interpretation:

- Ergebnisse: Single-Haushalte mit höherem Einkommen wohnen (im Durchschnitt) in größeren Wohnungen. Dieser Zusammenhang ist statistisch gut abgesichert. Das bedeutet, dass dieser positive Zusammenhang mit einer hohen Wahrscheinlichkeit auch in der Grundgesamtheit (alle Single-Haushalte in Deutschland) vermutet werden kann. Wenn das Einkommen eines Haushalts um 1,000 Euro höher ist, ist die Wohnungsfläche um durchschnittlich $0,8 \text{ m}^2$ größer. Single-Frauen wohnen in durchschnittlich 5 m^2 größeren Wohnungen als Single-Männer.
- Interpretation (selektiv): Die Interpretation, dass Single-Frauen bei gleichem Einkommen in größeren Wohnungen leben, weil sie Frauen sind (und Frauen vielleicht eine höhere Präferenz für mehr Wohnraum haben) ist nicht zulässig, da andere Merkmale nicht berücksichtigt wurden. So sind etwa über 70% der Frauen, aber weniger als 50% der Männer in Single-Haushalten vor 1960 geboren. Unter den Single-Haushalten wohnen ältere Personen durchschnittlich in größeren Wohnungen. Wenn man das Alter der Haushaltspersonen berücksichtigt, spielt das Geschlecht keine Rolle mehr.

Ziele der Lehrveranstaltung: Beispiel (3)

Politikempfehlung (selektiv):

- Wenn es ein Politikziel ist, die Wohnsituation von Single-Haushalten zu verbessern, dann kann die Erhöhung der Haushaltseinkommen von Single-Haushalten (etwa durch Transferzahlungen) zu einer Verbesserung der Wohnsituation (i.s.v. größeren Wohnungen) führen. Zwei Aspekte sind allerdings zu beachten: Erstens, die Ergebnisse könnten durch andere Eigenschaften bestimmt sein (etwa dem Vermögen der Haushalte), die die Schätzergebnisse möglicherweise verzerren. Selbst wenn eine kausale Interpretation zulässig ist (also höhere Einkommen dazu führen, dass Haushalte in größeren Wohnungen leben), muss bedacht werden, dass Transferzahlungen von 10.000 Euro notwendig sind, damit die Wohnungsgröße um durchschnittlich 8 m^2 ansteigt. Andere Eingriffe, wie Mietzuschüsse oder öffentlicher Wohnbau, können daher effizientere (d.h. kostengünstigere) Maßnahmen darstellen, um das Politikziel zu erreichen.

Aufbau der Lehrveranstaltung (1)

1. Vorlesungsteil

- Grundbegriffe
- univariate, bivariate und multivariate Datenanalyse
- deskriptive (beschreibende) und schließende Statistik
- lineare Einfachregression und multiple lineare Regression (Kleinstquadratschätzer)
- Hypothesentests
- mögliche Datenquellen
- Übungsaufgaben zwischen den Einheiten
 - ▶ sollen in fixen Teams zu 2 bis 3 Personen gelöst werden
 - ▶ Abgabe bis Sonntag (23:59 Uhr) vor der nächsten Einheit als Word / pdf (Fragen und Antworten) + Excel-file (Berechnungen)
 - ▶ Diskussion über Schwierigkeiten in nächster Einheit (bitte Laptop mitnehmen)
 - ▶ Vorlesungsteil dauert ca. 9×2 SWS und schließt nicht mit Klausur ab
 - ▶ Übungen werden nur auf Vollständigkeit überprüft
 - ▶ Vorlesungsteil macht 30 % der Gesamtnote aus (abgegebene Übungen + Mitarbeit in der LV)

2. Seminarteil

Aufbau der Lehrveranstaltung (2)

1. Vorlesungsteil

2. Seminarteil

- 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 bzw. den letzten beiden Einheiten; Einarbeitung von Feedback
- Vortrag macht 25 % und Seminarbericht 45 % der Gesamtnote aus

Fachwissenschaftliche Seminar hat 2 SWS und **6 ECTS** (= 150 Echtstunden zu 60 Minuten)

- **Literaturgrundlage**

- ▶ Christine Duller, 2019: Einführung in die Statistik mit EXCEL und SPSS (1. Teil der Vorlesung; online über Moodle verfügbar)
- ▶ Ludwig von Auer, 2003: Ökonometrie: Eine Einführung (2. Teil der Vorlesung; in Bibliothek verfügbar)

- **MS-Excel + Add-In “Analyse-Funktionen”**

Es wird eine grundlegende Vertrautheit mit Excel vorausgesetzt (etwa: Menüband, Bearbeitungsleiste, absoluter bzw. relativer Zellbezug, Matrixformel). Sie sollten auch einfache Befehle (Summe, Mittelwert, Wenn-Dann-Befehl...) kennen oder sich selbständig aneignen können (z. B. Duller, 2019, Kapitel 3; Google/Youtube).

- Auszug aus dem **Deutschen Sozio-oekonomischen Panel** (GSOEP) des Jahres 2009 als Übungsdatensatz (wird über Moodle bereitgestellt).
- Nutzen Sie Unterstützung von:
 - ▶ Folien, Literatur und sonstige Lehrmaterialien
 - ▶ Ihren KollegInnen und mir
 - ▶ Excel-Hilfe, Internet, Youtube, ...

Grundbegriffe (1)

- **Grundgesamtheit:** Das ist die Menge aller Objekte (Personen, Betriebe oder Ähnliches), über die man Informationen gewinnen will. Dabei ist eine exakte räumliche, zeitliche und sachliche Abgrenzung notwendig.
- **Erhebungseinheit:** Ein einzelnes Element der Grundgesamtheit. Die Anzahl der Erhebungseinheiten bildet den Umfang der Grundgesamtheit ($= N$).
- **Stichprobe:** Eine Teilmenge der Grundgesamtheit. Beinhaltet $n \leq N$ Erhebungseinheiten. (Der Unterschied wird später wieder aufgegriffen.)
- **Repräsentative Stichprobe:** Die Stichprobe zeichnet ein möglichst genaues Abbild der Grundgesamtheit.
- **Einfache Zufallsstichprobe:** Jedes Element der Grundgesamtheit hat die gleiche Chance, in die Stichprobe zu gelangen.
- **Datengewinnung:**
 - ▶ Primärstatistik: (neue) Daten werden z.B. mittels Befragung erhoben
 - ▶ Sekundärstatistik: bereits veröffentlichtes Datenmaterial wird verwendet
- **Merkmal (Variable):** Die interessierende Eigenschaft der Erhebungseinheiten. Jedes Merkmal besitzt verschiedene **Ausprägungen**.
- **Wertebereich:** Alle Ausprägungen eines Merkmals bilden den Wertebereich.

Grundbegriffe (2)

- **Deskriptive (beschreibende) Statistik:**

Beschreibung und Darstellung der Daten, etwa die Aufbereitung in Form von Tabellen und Grafiken und die Berechnung einfacher statistischer Kennzahlen. Erster Überblick über die Daten.

- **Induktive (schließenden) Statistik:**

Bei repräsentativer Stichprobe (etwa: Zufallsstichprobe) erlauben die Methoden der induktiven Statistik Rückschlüsse von der Stichprobe auf die Grundgesamtheit. Rückschlüsse von der (oft kleinen) Stichprobe auf die (unbekannte) Grundgesamtheit sind möglich, aber mit einer Unsicherheit behaftet.

- **Univariate, bivariate und multivariate Statistik:**

Üblicherweise werden an den Erhebungseinheiten mehrere Merkmale erhoben. Greift man zur Analyse nur ein einziges Merkmal heraus, so spricht man von univariater Statistik. Betrachtet man zwei (mehrere) Merkmale, spricht man von bivariaten (multivariaten) Verfahren.

Merkmalsausprägungen (1)

- **Nominal skaliert:** (qualitative oder kategoriale Merkmale)

Ausprägungen können unterschieden, aber nicht in eindeutiger Weise geordnet werden (z.B. Geschlecht, Familienstand, Beruf). Abstand zweier Merkmalsausprägungen ist nicht definiert.

- **Ordinal skaliert:** (Rangmerkmal)

Ausprägungen können in eine Ordnungsbeziehung wie größer, kleiner, besser oder schlechter zueinander stehen (z. B. Schulnoten, Güteklassen). Abstand zweier Merkmalsausprägungen allerdings nicht klar definiert und nicht interpretierbar.

- **Metrisch skaliert:** (kardinalskaliert)

Wenn Ausprägungen Vielfache einer Einheit sind (z.B. Länge, Einkommen). Die Ausprägungen sind voneinander verschieden, haben eine eindeutige Anordnung und einen eindeutig definierten Abstand. Bei **verhältnisskalierten Merkmalen** gibt es einen natürlichen Nullpunkt (z.B. Preis) und das Verhältnis zweier Ausprägungen lässt sich sinnvoll interpretieren (Produkt A ist doppelt so teuer wie Produkt B). **Intervallskalierte Merkmale** haben keinen natürlichen Nullpunkt, daher können auch Verhältnisse nicht sinnvoll interpretiert werden (z.B. Temperatur in Grad Celsius).

Merkmalsausprägungen (2)

Merkmalsausprägungen	Unterscheiden	Ordnen	Summen / Differenzen	Quotienten
Nominal	Ja	Nein	Nein	Nein
Ordinal	Ja	Ja	Nein	Nein
Metrisch				
Intervallskaliert	Ja	Ja	Ja	Nein
Verhältnisskaliert	Ja	Ja	Ja	Ja

Stetige und diskrete Merkmale

- **Stetige Merkmale:** Die Ausprägungen eines Merkmals können beliebige Zahlenwerte aus einem Intervall annehmen können (z. B. Länge, Gewicht).
- **Diskrete Merkmale:** Die Ausprägungen eines können (bei geeigneter Skalierung) nur ganzzahlige Werte annehmen können (z. B. Fehlerzahlen, Schulnoten, Geschlecht). Diskrete Merkmale haben abzählbar viele Ausprägungen.
- **Dichotome Merkmale:** Sonderform von diskreten Merkmalen mit nur zwei Ausprägungen (z. B. Geschlecht, Klausur bestanden).
- **“Zwischenstufen:”**
 - ▶ *quasistetige Merkmale:* Aufgrund der Definition diskret, verfügen aber über eine so feine Abstufung, dass man sie als stetige Merkmale behandeln kann (z. B. Preis, Miete).
 - ▶ *diskretisierte Merkmale:* Stetige Merkmale werden nur in diskreter Form erfasst (z. B. Alter in ganzen Jahren, Einkommen in 1.000-Euro-Intervallen).