

Willkommen zur Vorlesung Empirische Methoden I

12. Sitzung: (Quantitative) Datenauswertung

Prof. Dr. Wolfgang Ludwig-Mayerhofer

Universität Siegen – Philosophische Fakultät, Seminar für Sozialwissenschaften

Themen der Vorlesung

- Ursache-Wirkungs-Beziehungen von den Ursachen her betrachten!
- Statistische Unsicherheit berücksichtigen
- Keine Angst vor Regressionsmodellen

Vorbemerkungen

Wissen über und Fertigkeiten in Verfahren zur Auswertung standardisierter Daten gehören in den Bereich der sozialwissenschaftlichen Statistik.

Diese kann selbstredend im Rahmen einer einzelnen Vorlesungsstunde auch nicht annähernd vermittelt werden. Aus diesem Grund werden hier nur einige Hinweise für den Umgang mit oder die Interpretation von wissenschaftlichen (oder eben nicht-so-wissenschaftlichen) Ergebnissen gegeben.

Mögliche Fehlschlüsse I

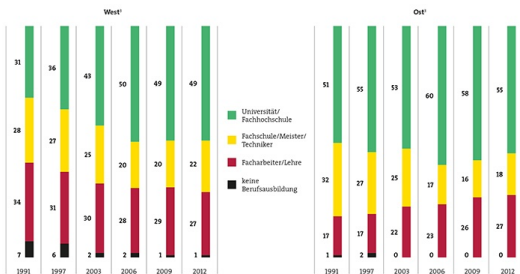
„Untersucht man . . . klinisch-stationär behandelte Drogenabhängige, meist Heroinsüchtige oder Polytoxikomane, so stellt man fest, daß sie ihre Drogenkarriere zu 98 bis 100 % mit Haschisch begonnen hatten.“
(Karl-Ludwig Täschner, Das Cannabis-Problem 1979, Seite 169).

→ Führt Haschisch also zur Heroinabhängigkeit?

Mögliche Fehlschlüsse II

Immer weniger Kinder von Facharbeitern an den (westdeutschen) Universitäten → Verschlechterung der Chancen von Arbeiterkindern?

Höchster beruflicher Abschluss der Eltern¹ von Studierenden² in West und Ost
(1991-2012, in Prozent)



¹ Zuordnung nach dem beruflichen Abschluss des Elternteils mit dem höchsten Abschluss

² Studierende an Universitäten und Fachhochschulen

³ ohne Berlin

Datenquelle: IHS - Institut für Hochschulforschung Hannover

Quelle: <http://www.bpb.de/izpb/198031/bildungsexpansion-und-bildungschancen?p=1>

Mögliche Fehlschlüsse III

In beiden (und vielen anderen vergleichbaren) Fällen kann es zu Fehlinterpretationen kommen:

- Drogenabhängigkeit: Die Annahme bricht zusammen, wenn man feststellt, dass sich auch unter Nicht-Abhängigen sehr viele (ehemalige) Haschischkonsumenten befinden.
Der richtige Ansatz lautet, Haschischkonsumenten und Nicht-Konsumenten zu vergleichen. Nur wenn die Haschischkonsumenten einen größeren Anteil von späteren Drogenabhängigen aufweisen, kann man von einem möglichen Ursache-Wirkungs-Zusammenhang sprechen.
- Arbeiterkinder: Der Rückgang der Arbeiterkinder kann auch auf einen Rückgang der Arbeiter an der Gesamtbevölkerung zurückgehen (und tut dies wahrscheinlich auch). Aussagen über die Chancen von Arbeiterkindern müssen Arbeiterkinder und andere hinsichtlich der Frage vergleichen, wie groß jeweils der Anteil derer ist, die es auf die Hochschule schaffen.

Mögliche Fehlschlüsse IV

Ein fiktives Beispiel (Tabelle zeigt absolute Zahlen):

	1990		2010	
	Arbeiter	Andere	Arbeiter	Andere
Berufliche Bildung	400	1000	200	1100
Studium	100	500	100	600
n	500	1500	300	1700

Anteil der Arbeiterkinder unter den Studierenden:

1990: 17 Prozent, 2010: 14 Prozent

Anteil der Studierenden unter den Arbeiterkindern:

1990: 20 Prozent, 2010: 33 Prozent

Stichproben als Unsicherheitsfaktor

Sozialwissenschaftliche Untersuchungen sind (wie alle wissenschaftliche Tätigkeit!) aus verschiedenen Gründen mit Unsicherheit behaftet:

- Allgemein: Wie weit reicht die Aussagekraft von Untersuchungen (andere Bevölkerungsgruppen? Andere Kulturen? Andere Zeiträume?)?
- Wie gut sind die Erhebungsinstrumente?
- Wie gut ist Umsetzung des Forschungsdesigns (auch wenn es auf dem Papier ‚perfekt‘ ist)?
- Der größte Unsicherheitsfaktor resultiert jedoch aus der Tatsache, dass meist mit Stichproben gearbeitet wird. Gleichzeitig ist dieser Unsicherheitsfaktor (bei guter Stichprobenqualität) am besten abzuschätzen bzw. in seinen Auswirkungen berechenbar.

Schlüsse von Stichproben auf die Grundgesamtheit

Stichproben sind grundsätzlich mit Unsicherheit behaftet – dass eine Stichprobe genau mit der Grundgesamtheit übereinstimmt, ist sehr unwahrscheinlich. Allerdings ist es auch unwahrscheinlich, dass sie sehr stark von der Grundgesamtheit abweicht.

Diese Erkenntnis (die sich mit Hilfe der mathematischen Stochastik beweisen und genau quantifizieren lässt) ist Grundlage der sog. Inferenzstatistik (von lateinisch *inferre* = [u. a.] schließen; daher im Deutschen auch oft: schließende Statistik). Der Schluss von der Stichprobe auf die Grundgesamtheit ist vor allem in zwei Formen wichtig:

- Berechnung sog. Konfidenzintervalle
- Berechnung der statistischen Signifikanz von Ergebnissen

Konfidenzintervalle I

Konfidenzintervalle beziehen sich auf einzelne Größen, die aus Stichproben berechnet werden, beispielsweise auf Anteilswerte oder Mittelwerte.

In Qualitätsmedien werden sie gelegentlich unter dem Begriff „Fehlermarge“ oder „Fehlertoleranz“ erwähnt:

Untersuchungsanlage

Grundgesamtheit: Wahlberechtigte Bevölkerung ab 18 Jahren

Stichprobe: Repräsentative Zufallsauswahl / Dual Frame (Festnetz- und Mobilfunkstichprobe)

Erhebungsverfahren: Computergestützte Telefoninterviews (CATI)

Fallzahl: 1001 Befragte

Erhebungszeitraum: 15. bis 17. Juni 2015

Fehlertoleranz: 1,4* bis 3,1** Prozentpunkte

*bei einem Anteilswert von 5 Prozent **bei einem Anteilswert von 50 Prozent

Konfidenzintervalle II

Wird ein Konfidenzintervall/eine ‚Fehlertoleranz‘ angegeben, so heißt das in der Regel:

„Der Bereich von ... bis ... enthält **mit 95-prozentiger Wahrscheinlichkeit** den wahren Wert der Grundgesamtheit“.

Am **Beispiel** der vorherigen Folie (‚Fehlertoleranz‘ von 3,1 Prozentpunkten bei einem in einer Stichprobe ermittelten Anteilswert von 50 Prozent):

„Eine sichere Aussage über den wahren Anteilswert (den Wert der Grundgesamtheit) ist nicht möglich. Der Wert von 50 Prozent ist nur der Wert, den wir bei den zufällig befragten Personen ermittelt haben. Dass dies exakt dem wahren Wert entspricht, ist möglich, aber unwahrscheinlich.

Wollen wir eine Aussage mit größerer Wahrscheinlichkeit machen, so können wir dies nur erreichen, indem wir statt eines einzelnen Wertes einen Bereich angeben, in dem der wahre Wert mit einiger Wahrscheinlichkeit liegt. Im vorliegenden Fall gilt:

Mit 95-prozentiger Wahrscheinlichkeit enthält der Bereich von 46,9 bis 53,1 den wahren Wert der Grundgesamtheit.“

Konfidenzintervalle: FAQ I

- Was heißt „95-prozentige Wahrscheinlichkeit“?

Es heißt: In 95 Prozent aller Fälle, in denen wir eine Stichprobe dieses Umfangs (und auch sonst mit gleichen Eigenschaften) gezogen haben, enthält das berechnete Intervall den wahren Wert, in 5 Prozent der Fälle liegt der wahre Wert außerhalb des Intervalls. Ob der Wert nun gerade bei dieser Stichprobe in dem Intervall liegt oder nicht, können wir nicht wissen.

- Warum gerade „95-prozentige Wahrscheinlichkeit“?

95 Prozent ist eine Konvention, die sich in Sozialwissenschaft und Meinungsforschung herausgebildet hat. Man könnte auch Intervalle für andere Wahrscheinlichkeiten berechnen.

(Wir können auch nicht sicher wissen, ob es bei der ARD-Umfrage um 95-prozentige Wahrscheinlichkeit geht; eigentlich muss das angegeben werden. Die ARD weiß aber auch, dass die meisten Menschen nicht über Statistik nachdenken wollen.)

Konfidenzintervalle: FAQ II

- Sind Konfidenzintervalle immer so groß?

Wie breit ein Konfidenzintervall ist, hängt von mehreren Faktoren ab:

- Dem (mutmaßlich) wahren Wert in der Grundgesamtheit (das KI ist bei Anteilswerten von 50 Prozent größer als bei 10 Prozent; bei Mittelwerten hängt es von der Streuung der Merkmals in der Grundgesamtheit ab);
- dem Umfang der Stichprobe (große Stichproben haben kleinere Konfidenzintervalle, denn sie erlauben eine genauere Schätzung);
- der Art der Stichprobe (mehrstufige Stichproben haben meistens größere Konfidenzintervalle als einstufige);
- und natürlich von der gewählten Wahrscheinlichkeit – eine größere Wahrscheinlichkeit für das Konfidenzintervall führt zu einem breiteren Intervall.

Konfidenzintervalle: FAQ III

- **Und woher weiß man das alles?**
Intelligente Menschen ohne Angst vor Mathematik haben darüber nachgedacht.
- **Und stimmt das auch alles wirklich?**
In der Theorie: ja. Ob es in der Praxis der Sozialforschung immer gilt, ist fraglich, da die Menschen sich leider nicht immer so verhalten, wie es gemäß den Methoden der Sozialforschung wünschenswert wäre (ein Beispiel bei Schnell & Noack 2014).
- **Kann ich das alles auch verstehen?**
Siehe oben. Sie brauchen allerdings ein Statistik-Lehrbuch oder eine Lehrveranstaltung (oder beides).

Statistische Signifikanz I

In Berichten über empirische Untersuchungen tauchen häufig Sätze wie der folgende auf:

„Der Einfluss von X auf Y (oder: der Zusammenhang zwischen X und Y, oder: der Unterschied zwischen X und Y) ist (statistisch) signifikant (oder auch: nicht signifikant).“

- Häufig sind solche Sätze noch etwas ausführlicher i. S. v. genauer; es heißt „... ist auf 5-prozentigem Niveau (nicht) signifikant“ oder „ist (statistisch [nicht]) signifikant ($p < 0,05$) [manchmal auch: $\alpha < 0,05$]“.
- Die korrekte Ausdrucksweise ist „statistisch signifikant“, da der Begriff „signifikant“ auch in anderen Kontexten verwendet wird (signifikant = bedeutend, bedeutsam); die korrekte Ausdrucksweise wird aber nicht immer gewählt. In Verbindung mit Ausdrücken wie „5-prozentiges Signifikanzniveau“ oder „ $p < 0,05$ “ oder dgl. ist aber immer „statistisch signifikant“ gemeint.

Statistische Signifikanz II

Aber was heißt das nun?

Der Begriff der statistischen Signifikanz bezieht sich ebenfalls auf die Unsicherheit, die mit der Verwendung von Stichproben (statt Totalerhebungen) verbunden ist. Er wird verwendet, wenn wir Hypothesen über Zusammenhänge zwischen Merkmalen untersuchen wollen (oder: über Einflüsse von einem Merkmal auf ein anderes; oder: über Unterschiede zwischen zwei oder mehreren Gruppen).

Ein „statistisch [nicht] signifikantes“ Ergebnis besagt:

„Wir haben eine Annahme über einen Zusammenhang (Unterschied, Einfluss) in einer Stichprobe geprüft, und es ist einigermaßen plausibel, dass diese Annahme in ‚Wirklichkeit‘, das heißt: in der Grundgesamtheit, aus der die Stichprobe gezogen wurde, [nicht] zutrifft“.

Statistische Signifikanz: FAQ I

- Kann man das denn etwas genauer sagen?

Jein. Forscher/-innen verwenden meist ein „Signifikanzniveau“ von 5 Prozent. Das bedeutet: In nur 5 Prozent aller Fälle, in denen in Wahrheit gar kein Zusammenhang (Unterschied, Einfluss) besteht, erhalte ich rein zufallsbedingt trotzdem ein statistisch signifikantes Ergebnis. Das heißt, das Risiko, fälschlicherweise das Bestehen eines Zusammenhanges anzunehmen, ist ziemlich klein.

Diese Wahrscheinlichkeitsangabe gilt aber nur für den Fall, dass in der Grundgesamtheit kein Zusammenhang besteht. Ich kann daher nicht im Umkehrschluss sagen: In (nur) 5 Prozent aller Fälle, in denen ich Hypothesen anhand von Stichproben prüfe, begehe ich einen Irrtum. Das Ausmaß des Irrtums hängt vielmehr davon ab, wie viele Hypothesen ich prüfe, bei denen in Wahrheit ein (oder kein) Zusammenhang besteht.

Statistische Signifikanz: FAQ II

- Uff. Das klingt aber kompliziert!
Korrekt. Es ist tatsächlich kompliziert! Wenn man will, kann man es aber durchaus verstehen. Man muss nur seinen Kopf gebrauchen. Und ja, ein wenig lernen. (Aber deshalb studieren Sie ja.)

Statistische Signifikanz: FAQ II

- Uff. Das klingt aber kompliziert!

Korrekt. Es ist tatsächlich kompliziert! Wenn man will, kann man es aber durchaus verstehen. Man muss nur seinen Kopf gebrauchen. Und ja, ein wenig lernen. (Aber deshalb studieren Sie ja.)

- Oh je. Muss ich das jetzt gleich lernen?

Was heißt „müssen“? Es heißt: „Kann ich das jetzt gleich lernen“? Die Antwort lautet aber: Leider nein. Man braucht dafür paar (einfache, aber eben doch erst zu erwerbende) Grundlagen. Dafür haben wir im Bachelor-Studium keine Zeit. Aber im Master! (Gilt aber nicht für Lehramt, nur für Fachstudiengänge.)

Statistische Signifikanz: FAQ III

- Aber die Sozialwissenschaftler, die Untersuchungen durchführen, verstehen das schon alle?
Äh, hmm, . . . ja doch, also jedenfalls viele. Alle wohl nicht. Dafür spricht jedenfalls die Tatsache, dass man immer wieder Äußerungen finden kann, die von beschränktem Sachverstand sprechen.
- Was für Äußerungen meinen Sie?
Der wichtigste Fehler ist es, einen statistisch signifikanten Zusammenhang mit einem ‚starken‘, ‚großen‘, ‚wichtigen‘ Zusammenhang zu verwechseln. Dass man das nicht darf, lernt jede/-r Anfänger/-in, es wird aber immer wieder vergessen.
- Und was heißt grade nochmal „statistisch signifikant“, wenn es nicht „groß“ oder „wichtig“ heißt?
Es heißt nur: Es ist einigermaßen plausibel, dass der Zusammenhang auch in der Grundgesamtheit besteht.

Regressionsmodelle

Regressionsmodelle sind statistische Verfahren, mit denen simultan die Einflüsse vieler Merkmale (sog. unabhängige Variablen; auch: Prädiktoren, Kovariaten) auf ein anderes Merkmal (abhängige Variable; auch Zielvariable) statistisch untersucht werden können.

Beispiele:

- Das Erwerbseinkommen von Menschen hängt ab von der Bildung, der Berufserfahrung, der Betriebsgröße, dem Geschlecht und einigem anderen mehr.
- Ob jemand sein Studium abbricht, hängt ab von der Motivation, der Befähigung, Belastungen in anderen Lebensbereichen (Betreuung von Kindern oder Pflegebedürftigen), der finanziellen Situation (z. B. Notwendigkeit, nebenher Geld zu verdienen), u. a. m.

Ein Beispiel: Armut (Giesselmann & Goebel 2013, S. 295)

schon Regressionsmodelle überprüft.

	OLS	FE	
Bruttolohn pro Stunde	-0,0012** (0,0001)	-0,0005** (0,0001)	
ISEI	-0,0009** (0,0001)	-0,0002 (0,0001)	
Alter (Referenz: 18 - 29)			
30 - 49	-0,011* (0,005)	-0,011** (0,003)	
50 - 64	-0,011* (0,006)	-0,005 (0,004)	
Bildungsniveau (Referenz: kein Abschluss)			
Noch in der Schule	-0,166** (0,057)	-0,033 (0,031)	
Hauptschulabschluss ohne ber. Ausbildung	-0,031 (0,021)	0,017 (0,013)	
Hauptschulabschluss und ber. Ausbildung	-0,059** (0,020)	0,028* (0,013)	
Mittlere Reife ohne ber. Ausbildung	-0,074** (0,023)	0,011 (0,015)	
Mittlere Reife und ber. Ausbildung	-0,075** (0,020)	0,019 (0,014)	
Abitur ohne ber. Ausbildung	-0,072** (0,024)	0,030 (0,016)	
Abitur und ber. Ausbildung	-0,078** (0,020)	0,032* (0,016)	
Fachhochschulabschluss/ Bachelor	-0,058** (0,021)	0,052** (0,016)	
Hochschulabschluss	-0,064** (0,020)	0,053** (0,017)	
Erwerbsstatus (Referenz: Voll erwerbstätig)			
Teilzeitbeschäftigt	0,085** (0,012)	0,075** (0,006)	
In Ausbildung	0,126** (0,013)	0,120** (0,006)	
Geringfügig erwerbstätig	0,244** (0,023)	0,171** (0,007)	
Haushaltstyp (Referenz: Einpersonenhaushalt)			
Paar ohne Kinder	0,004 (0,005)	0,006 (0,004)	
Alleinerziehende	0,047** (0,011)	0,036** (0,007)	
Paar mit Kind(ern)	0,036** (0,005)	0,018** (0,004)	
Mehrgenerationenhaushalt	0,094** (0,019)	0,074** (0,010)	
Sonstige Kombination	-0,001 (0,011)	-0,0002 (0,010)	
Anzahl Erwerbstätigen im HH (Ref.: eine Person)			
mehr als ein Erwerbstätiger	-0,041** (0,004)	-0,026** (0,002)	
Region (Referenz: Westdeutschland)			
Ostdeutschland	0,036** (0,004)	0,047** (0,010)	
Konstante	0,171** (0,021)	0,033* (0,014)	
N	87278	87278	

Standardfehler in Klammern

* p<0,05 ** p<0,01

Quelle: SOEP v28, eigene gewichtete Berechnungen

Tabelle 3: Lineare Regressionen auf die Armutswahrscheinlichkeit erwerbstätiger Männer in Deutschland (1995-2011)

Die Zahlen (sog. Regressionskoeffizienten) geben an, wie sich die abhängige Variable ändert, wenn das jeweilige Merkmal um eine Einheit zunimmt bzw. die jeweilige Ausprägung annimmt.

Abhängige Variable = Armutrisiko (als Anteilswert).

Lesebeispiel (linke Spalte): Die Eigenschaft „alleinerziehend“ steigert das Armutrisiko um 4,7 Prozentpunkte.

Regressionsmodelle interpretieren

- Die elementare Interpretation: Positiver Effekt (Pluszeichen) oder negativer Effekt (Minuszeichen)?
(Aufgepasst: Bei einzelnen Modellen stehen Koeffizienten > 1 für einen positiven und Koeffizienten zwischen 0 und 1 für einen negativen Effekt. Achten Sie auf Begriffe wie „Odds Ratio“ oder „Effektkoeffizienten“.)
- Bessere Interpretation: *Um wieviel* ändert sich die abhängige Variable, wenn die unabhängige Variable um eine Einheit zunimmt?
(Wird in jüngster Zeit gelegentlich visualisiert, vor allem bei Modellen, die nicht einfach zu interpretieren sind.)
- Achten Sie darauf, ob Effekte überhaupt statistisch signifikant sind – in der Regel erkennbar an den hochgestellten Sternchen.

Alles Weitere müssen Sie dem guten Glauben überlassen ... oder sich vertiefte Statistik-Kenntnisse aneignen!

Literatur

Giesselmann, Marco/Goebel, Jan (2013): Soziale Ungleichheit in Deutschland in der Längsschnittperspektive. Befunde zur Armutsproblematik auf Basis des Sozio-oekonomischen Panels (SOEP), in: Analyse & Kritik 35/2, 277-302.

Schnell, Rainer/Noack, Marcel (2014): The Accuracy of Pre-Election Polling of German General Elections, in: methods, data, analyses 8/1, 5-24.