

Statistik I - Sitzung II

Bernd Schlipphak

Institut für Politikwissenschaft

Woche 2

- 1 Erste Grundbegriffe
 - Forschungsdesigns
 - Fälle, Ausprägungen und Variablen
 - Skalierungsniveaus

- 2 Quantitative Daten - Erhebung, Art und Verfügbarkeit

Erste Grundbegriffe - Übersicht

- *Wiederholung: Forschungsdesigns*
- *Wiederholung: Fälle, Ausprägungen und Variablen*
- *Neu: Skalierungsniveaus*
- *Später: Kausalmodell \Rightarrow Sitzung 6*

Deskriptive Forschung

- **Deskriptive** Forschung = Beschreibende Forschung
- Beschreibende Forschungsergebnisse oft als Startpunkt analytischer Forschung

Induktive Forschung

- **Induktive** Forschung(sdesigns)
 - = vom Kleinen aufs Große schließend
 - = von der Empirie auf die Theorie schließend
- Schritte des induktiven Forschungsdesigns
 - Fragestellung / Puzzle
 - Empirische Datenerhebung / Analyse
 - Generierung eines theoretischen Arguments / Modell

Deduktive Forschung

- **Deduktive** Forschung(sdesigns)
 - = vom Großen aufs Kleine schließend
 - = Testen der Theorie an der Empirie
- Schritte des deduktiven Forschungsdesigns
 - Fragestellung / Puzzle
 - Theoretisches Argument / Hypothesen
 - Operationalisierung der Variablen
 - Empirischer Test

Deduktiv = quantitativ?

- Deduktive und induktive Forschungsdesigns wollen beide kausale Zusammenhänge entdecken / überprüfen
- **Früher** Annahme, dass deduktive Designs nur mit quantitativen Daten und induktive Designs nur mit qualitativen Daten arbeiten
 - **Veraltete Schwarz-Weiß-Vorstellung!**
- In Statistik I und II aber nur Nutzung **deduktiver** Forschungsdesigns

Fälle, Ausprägungen und Variablen

- Was versteht man unter diesen Begriffen und wie stehen sie zueinander?
 - *Fall* = statistische Einheit / Merkmalsträger
 - *Variable* = messbar gemachte Dimension (Merkmal)
 - *Ausprägung* / *Eigenschaft* = Werte, die ein Fall auf einer Variable annehmen kann / aufweist

Sprachgebrauch

- Ein Fall (= 1 Person, 1 Haushalt, 1 Staat, 1 Parteiprogramm) 'trägt' hinsichtlich eines Merkmals (einer Variablen) eine bestimmte Ausprägung
 - *Oder:* Ein Fall nimmt auf einer Variable eine Ausprägung / einen Wert an
 - Bsp.: Eine Person (Fall) nimmt auf der Variable AfD-Wahlabsicht ja (=1) / nein (=0) den Wert / die Ausprägung 1 an (oder: die Person will die AfD wählen)
 - Bsp.: Eine Person (Fall) nimmt auf der Variable 'Haarfarbe' den Wert 'blond' an

Erste Grundbegriffe - Zwischenstand

- Wiederholung: Forschungsdesigns
- Wiederholung: Fälle, Ausprägungen und Variablen
- Neu: Skalierungsniveaus
 - Variablen sind je nach Art ihrer Ausprägungen unterschiedlich skaliert

Die Skalenniveaus von Variablen

- Nominalskala
- Ordinalskala
- Metrische Skala
 - Verhältnisskala
 - Intervallskala
 - manchmal zusätzlich: Absolutskala (Behnke/Behnke 2006)

Variablen auf Nominalskalenniveau

- Qualitativen Ausprägungen einer Variablen (V=Haarfarbe, A=blond, grün, rot, braun) werden quantitative Ausprägungen (willkürlich) zugeordnet (blond=1, grün=2, braun=5, etc.)
- Diese quantitativen Ausprägungen stehen aber in keiner Beziehung zueinander, sondern können nur als An- oder Abwesenheit einer Ausprägung interpretiert werden

Variablen auf Nominalskalenniveau

- Qualitativen Ausprägungen einer Variablen (V=Haarfarbe, A=blond, grün, rot, braun) werden quantitative Ausprägungen (willkürlich) zugeordnet (blond=1, grün=2, braun=5, etc.)

Fall	Originale Ausprägung: Haarfarbe	Quantitative Ausprägung: Haarfarbe
Fall A	blond	1
Fall B	braun	5
Fall C	grün	2
Fall D	schwarz	13
Fall E	braun	5

Variablen auf Ordinalskalenniveau

- Qualitativen Ausprägungen einer Variablen, die **hierarchisch** aufeinander aufbauen (V=Bildung, A = Abitur, Realschule, Hauptschule, Grundschule), werden quantitative Ausprägungen zugeordnet, welche diese Hierarchie spiegeln (1= Abitur bis 4= Grundschule oder umgekehrt) \Rightarrow Rangfolge!
- Bei ordinalskalierten Variablen geht es darum, ob ein Fall eine Eigenschaft in **schwächerem**, **stärkerem** oder **gleichem** Maß besitzt als bzw. wie ein anderer Fall

Variablen auf Ordinalskalenniveau

- Qualitativen Ausprägungen einer Variablen, die **hierarchisch** aufeinander aufbauen, werden quantitative Ausprägungen zugeordnet, welche diese Hierarchie spiegeln (1= Grundschule bis 4= Abitur oder umgekehrt) \Rightarrow Rangfolge!

Fall	Originale Ausprägung: Bildung	Quantitative Ausprägung: Bildung
Fall A	Abitur	4
Fall B	Realschule	3
Fall C	Hauptschule	2
Fall D	Grundschule	1
Fall E	Abitur	4

Variablen auf Ordinalskalenniveau

- Man kann damit - je nach Codierung – für **ordinal skalierte** Variablen die Aussage treffen: Ein numerischer Wert 4 steht für eine Ausprägung, die größer/kleiner/stärker/schwächer ist als die Ausprägung mit dem Wert 3

Variablen auf Ordinalskalenniveau

- Man kann damit - je nach Codierung – für **ordinal skalierte** Variablen die Aussage treffen: Ein numerischer Wert 4 steht für eine Ausprägung, die größer/kleiner/stärker/schwächer ist als die Ausprägung mit dem Wert 3
- In unserem Beispiel kennzeichnet die Ausprägung mit dem numerischen Wert 4 (Abitur) ein höheres Bildungsniveau als die Ausprägungen mit den Werten 3 (Realschule), 2 (Hauptschule) oder 1 (Grundschule). Der Wert 3 verdeutlicht eine höhere Bildungsstufe als die Werte 2 oder 1

Variablen auf Ordinalskalenniveau

- Man kann damit - je nach Codierung – für **ordinal skalierte** Variablen die Aussage treffen: Ein numerischer Wert 4 steht für eine Ausprägung, die größer/kleiner/stärker/schwächer ist als die Ausprägung mit dem Wert 3
- ABER: Für ordinal skalierte Variablen können keine mathematischen Beziehungen zwischen den Werten hergestellt werden \Rightarrow Die Ausprägung mit dem Wert 4 ist NICHT doppelt so groß wie die Ausprägung mit dem Wert 2!

Variablen auf Nominal-/Ordinalskalenniveau

- Eine Spezialform der nominalen / ordinalen Variablen ist die **dichotome** Variable (auch: **Dummy-Variable**), die nur zwei Ausprägungen besitzt
- Mit dieser Variable wird oft die (Nicht-)Anwesenheit eines Merkmals / einer Variablen für einen Fall abgebildet
 - Dummy-Variable *FDP-Wahl*: Ausprägung 1 = Fall wählt die FDP, Ausprägung 0 = Fall wählt die FDP nicht
- ABER: Dummy-Variablen werden auch für Variablen verwendet, für die es nur zwei Ausprägungen gibt
 - Dummy-Variable *Geschlecht*: Ausprägung 1 (oder 0) = weiblich, Ausprägung 0 (oder 1) = männlich

Variablen auf Nominal-/Ordinalskalenniveau

- Eine Spezialform der nominalen / ordinalen Variablen ist die dichotome Variable (auch: Dummy-Variable), die nur zwei Ausprägungen besitzt
- Manchmal sind auch Dummy-Variablen hierarchisch, d.h. die Anwesenheit (= 1) einer Variable wird als stärker / schwächer / kleiner / größer definiert als die Nicht-Anwesenheit (= 0)
 - Dummy-Variable *Hohe Bildung*: Ausprägung 1 = (mindestens) Abitur, Ausprägung 0 = alle niedrigeren Abschlüsse
 - \Rightarrow ordinale Interpretation der Dummy-Variable

Variablen auf metrischem Skalenniveau

- Ausprägungen einer Variablen sind (meistens) bereits als quantitative Ausprägungen erfasst (V = Monatl. Einkommen in EUR, A = 1000, 1001, 2000, 3000, etc.)

Variablen auf metrischem Skalenniveau

- Ausprägungen einer Variablen sind (meistens) bereits als quantitative Ausprägungen erfasst (V = Monatl. Einkommen in EUR, A = 1000, 1001, 2000, 3000, etc.)

Fall	Originale Ausprägung: Einkommen	Quantitative Ausprägung: Einkommen
Fall A	1000	1000
Fall B	1001	1001
Fall C	2000	2000
Fall D	3000	3000
Fall E	2000	2000

Variablen auf metrischem Skalenniveau

- Ausprägungen einer Variablen sind (meistens) bereits als quantitative Ausprägungen erfasst (V = Monatl. Einkommen in EUR, A = 1000, 1001, 2000, 3000, etc.)
- Mit den Ausprägungen einer metrisch skalierten Variablen (Intervallskala, Verhältnis- oder Ratioskala, selten: Absolutskala) können sinnvoll Differenzen und Summen gebildet werden
- Aussagen über Größe der Abstände zwischen zwei Ausprägungen möglich!

Zusammenfassung Skalierungsniveaus

Merksatz Skalierungsniveau

Jedes Skalenniveau weist neben seinen spezifischen Eigenschaften auch alle Eigenschaften der niedrigeren Skalenniveaus auf \Rightarrow metrisch skalierte Variablen können auch als nominal oder ordinal skaliert behandelt werden, aber nicht umgekehrt!

Zusammenfassung Skalierungsniveaus

- Wozu brauchen wir dieses Wissen über die Skalenniveaus von Variablen(-ausprägungen)?
- Sie spielen eine wichtige Rolle dafür, für welche Variablen überhaupt bestimmte Maße berechnet und sinnvoll interpretiert werden können
- Das beginnt bereits bei der univariaten Darstellung von quantitativen Daten!

Der Datensatz

- **Datensatz** = Datenmenge, welche die Ausprägungen einer Variablen für eine Menge von Objekten enthält (auch: = Urliste, Rohdaten, Primärdaten)
- In dem Datensatz wird einem Objekt (= **Fall**) der Wert (= **Ausprägung**) einer **Variablen** zugeordnet
- Der Datensatz besteht also aus einer Tabelle, in welcher die Zeilen Fällen (= Objekten) entsprechen, die Spalten Variablen und die einzelnen Zellen den Ausprägungen der Fälle auf den Variablen

Der Datensatz

Untitled2 [DataSet2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

The screenshot shows the IBM SPSS Statistics Data Editor interface. The top row contains variable names, all labeled 'var'. The left column contains case numbers from 1 to 29. A red oval highlights the top row of variables, with an arrow pointing to it from the label 'Variablen'. Another red oval highlights the left column of case numbers, with an arrow pointing to it from the label 'Fälle'. A small red rectangle highlights the cell at row 7, column 8, with an arrow pointing to it from the label 'Ausprägung des Fall 7 auf der Variable'. The text 'Data View' is written in green in the lower right area of the grid.

	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var
1																
2																
3																
4																
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																
16																
17																
18																
19																
20																
21																
22																
23																
24																
25																
26																
27																
28																
29																

Labels in the image:

- Variablen (points to the top row of variables)
- Fälle (points to the left column of case numbers)
- Ausprägung des Fall 7 auf der Variable (points to the cell at row 7, column 8)
- Data View (text in green)

Der Datensatz

DATA_Q71G_HDI_COMBINED_FINAL.sav [DataSet1] - IBM SPSS Statistics Dateneditor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Direktmarketing Grafik Extras Erweiterungen Fenster Hilfe

	PSRAID	COUNTRY	INTERNET	Q67	COMPUTER	Q148E	SMRTPHN	Q69	WOHNORT	EINKOMMEN_Z_H	BILDUNG
1	143734	2	1	1	0	2	0	2	0	0	8
2	143774	2	1	1	0	2	0	2	1	0	8
3	143776	2	1	1	0	2	0	2	1	0	12
4	143784	2	1	1	1	1	0	2	1	0	12
5	143791	2	1	1	0	2	0	2	0	0	5
6	143794	2	1	1	0	2	0	2	0	0	5
7	143820	2	1	1	0	2	0	2	0	0	15
8	143869	2	1	1	0	2	1	1	1	0	6
9	143872	2	1	1	0	2	1	1	1	0	10
10	143873	2	1	1	0	2	1	1	1	0	5
11	143879	2	1	1	0	2	1	1	1	0	10
12	143880	2	1	1	0	2	1	1	1	0	10
13	143892	2	1	1	0	2	.	.	1	0	9
14	143911	2	0	2	0	2	1	1	1	0	3
15	143925	2	1	1	0	2	0	2	1	0	12
16	143926	2	1	1	1	1	1	1	1	0	15
17	143934	2	1	1	0	2	0	2	1	0	10
18	143943	2	1	1	0	2	0	2	0	0	10
19	143946	2	1	1	0	2	0	2	0	0	12
20	143948	2	1	1	0	2	0	2	0	0	10
21	143952	2	1	1	0	2	0	2	1	0	12
22	143967	2	1	1	0	2	1	1	1	0	0
23	143975	2	1	1	1	1	1	1	1	0	0
24	143978	2	1	1	0	2	1	1	1	0	0
25	143988	2	1	1	0	2	0	2	1	0	10

Der Datensatz

- **Datensatz** - Wo kommen die Daten her?
- Quantitative Daten entstehen durch
 - Beobachtung
 - automatische oder manuelle Inhaltsanalyse von Text, Bild und Sprache
 - Befragung

Die Beobachtung

- Die **Beobachtung** kann qualitativ oder quantitativ erfolgen
- Dabei geht es darum, wissenschaftlich relevante Aspekte von Objekten – Personen, Ereignissen oder Situationen – systematisch und objektiv zu erfassen oder abzubilden
 - Vorteil: Oft kein Problem der Reaktivität der beobachteten Objekte
 - Problem der Reaktivität bei teilnehmender Beobachtung
- Beispiel 1: Konflikthäufigkeit
- Beispiel 2: Abstimmungsverhalten

Die Inhaltsanalyse

- Die **Inhaltsanalyse** kann ebenfalls qualitativ oder quantitativ erfolgen
- Dabei geht es darum, induktiv oder deduktiv visuelles oder auditives Material im Hinblick auf (politische) Ansichten oder Emotionen auszuwerten
- Grundgedanke - von Ansichten oder Emotionen im untersuchten Text, Bild oder Sprache kann auf die Ansicht oder Emotion des/der Urheberin (Partei, Institution, ...) geschlossen werden
 - Vorteil: Zu analysierendes Material oft verfügbar, wenig reaktiv
 - Probleme mit Größe des Materials (bei manueller Analyse) und Validität der Analyse (bei automatisierten Analyseverfahren)
- Beispiel: Wahlprogramme und Parteienpositionen

Die Befragung

- Die **Befragung** kann ebenfalls qualitativ oder quantitativ erfolgen
- Dabei geht es bei quantitativen, standardisierten Befragungen (Surveys) darum, politische und gesellschaftliche Einstellungen der Befragten zu erheben
 - Vorteil: direkte Messung der Einstellung, kein Rückschluss von Text auf Person nötig
 - Probleme mit Reaktivität der Daten (Interviewer Bias, Cognitive Bias, problematische Fragestellungen)

Der Datensatz

- Quantitative Daten entstehen durch
 - Beobachtung
 - automatische oder manuelle Inhaltsanalyse von Text, Bild und Sprache
 - Befragung
- Unterscheiden lassen sich dabei zwei Datenarten - die **Aggregatdaten** und die **Individualdaten**

Aggregatdaten

- **Aggregatdaten** sind Daten, die auf einer höheren Ebene Daten einer unteren Ebene zusammenfassen (aggregieren)
 - Vor allem werden dabei tatsächlich beobachtbare wirtschaftliche oder politische Strukturdaten gemessen (Wahlbeteiligung, Anzahl Ehrenamtliche / NGOs pro Gemeinde, Land, etc.) und / oder von niedrigerer (etwa kommunaler Ebene) auf höhere (regionaler, nationaler) Ebene aggregiert
- Aggregatdaten damit oft auf Grundlage objektiv nachvollziehbarer Auszählung / Beobachtung erhoben (Bsp.: Strukturdaten)

Aggregatdaten

- **Aggregatdaten** sind Daten, die auf einer höheren Ebene Daten einer unteren Ebene zusammenfassen (aggregieren)
 - Vor allem werden dabei tatsächlich beobachtbare wirtschaftliche oder politische Strukturdaten gemessen (Wahlbeteiligung, Anzahl Ehrenamtliche / NGOs pro Gemeinde, Land, etc.) und / oder von niedrigerer (etwa kommunaler Ebene) auf höhere (regionaler, nationaler) Ebene aggregiert
- Aggregatdaten damit oft auf Grundlage objektiv nachvollziehbarer Auszählung / Beobachtung erhoben (Bsp.: Strukturdaten)
 - Manchmal werden daher (Struktur-)Aggregatdaten auch als 'harte' Daten bezeichnet

Aggregatdaten - Datenbanken

- Datenquellen

- Eurostat – *Daten auf EU- und Nationalstaatsebene*
- OECD / IWF (Internat. Währungsfonds) / WTO (Welthandelsorganisation) – *Wirtschafts- und Strukturdaten von und zwischen Nationalstaaten sowie Regionalverbünden*
- Vereinte Nationen – *Globale und nationale Wirtschafts-, politische und Sozialdaten*
- CIA World Factbook – *Nationale Wirtschafts-, Politik- und Sozialdaten*
- Statistisches Bundesamt – *Nationale Wirtschafts-, Politik- und Sozialdaten*

Aggregatdaten - Datenbanken

- Eurostat – <http://ec.europa.eu/eurostat/web/main/home>
- OECD - <http://stats.oecd.org/>
- IMF - <http://data.imf.org/>
- WTO – <http://stats.wto.org/>
- Vereinte Nationen – <http://unstats.un.org/unsd/default.htm>
- CIA – <https://www.cia.gov/the-world-factbook/>
- Statist. Bundesamt – <https://www.destatis.de/DE/Startseite.html>

Individualdaten

- Individualdaten stellen Daten auf der Individualebene dar
- D.h., ein Individualdatensatz beinhaltet die Ausprägungen, die (menschliche) Individuen auf abgefragten oder erhobenen Variablen einnehmen
- Individualdaten werden vor allem über Surveys, Experimente etc. gewonnen
- Manchmal werden daher Individualdaten im Falle von Surveys auch als 'weiche' Daten bezeichnet

Aggregat- und Individualdaten

Merksatz Aggregatdaten

Auch 'weiche' Individualdaten können auf höhere Ebene aggregiert werden:
z. Bsp. der Anteil Euroskeptiker pro Land / Anteil der Bevölkerung, der zufrieden mit der Wirtschaftslage ist, etc. Dadurch werden aber Individualdaten nicht von 'weichen' zu 'harten' Daten!

Individualdaten - Datenbanken

- Politikwissenschaftlich wichtige Datenquellen
 - European Social Survey (ESS)
 - European Values Study (EVS) / World Values Survey (WVS)
 - Eurobarometer (EB)
 - Regionalbarometer (Arab Barometer, Afrobarometer, Latinobarometro)
 - Nationale / Europäische Vor- und Nachwahlstudien

Individualdaten - Datenbanken

- ESS - <http://www.europeansocialsurvey.org/>
- EVS - <http://www.europeanvaluesstudy.eu/>
- WVS - <http://www.worldvaluessurvey.org/>
- Regionalbarometer - <http://globalbarometers.org/>
- Nationale Wahlstudien, EB, ISSP – <http://www.gesis.org/>

Individualdaten - Datenbanken

- Im Bereich der Sozialwissenschaften allgemein spielen in Deutschland auch folgende Datensätze eine Rolle
 - SOEP (Sozioökonomisches Panel) – <http://www.diw.de/soep>
 - ALLBUS - <http://www.gesis.org/allbus/allbus-home/>
 - Mikrozensus - <http://tinyurl.com/czesvtr>

Grundlegend zum Problem der Datenqualität

Merksatz Datenqualität

Schlechte Daten führen zu schlechter Forschung, ungültigen Ergebnissen und problematischen normativen Ratschlägen auf Grundlage dieser Ergebnisse!

Vorbereitung Tutorium

IT - Software SPSS

← → ↻

https://www.uni-muenster.de/IT/services/arbeitsplatz/software/spss.html

Erste Schritte | Behrke_2005.pdf

9) Anmelden

WWU MÜNSTER

IT-PORTAL | SERVICES | DIE WWU IT

NEUE BESCHÄFTIGTE

DE | EN

Home > Services > Arbeitsplatz > Software

SPSS

HINWEIS: Ab sofort können die neuen SPSS Statistics- und Amos-Lizenzen bei uns erworben werden. Um die Sicherheitstische bereit zu schütten, müssen Sie SPSS auf die aktuellste Version updaten (von 27 bzw. 28 auf 27.0.1 bzw. 28.0.1) und anschließend das entsprechende FixPack installieren. Befolgen Sie dafür bitte die folgende Anleitung: [Software SPSS Update und FixPack](#)

Software: SPSS ist eine Software zur statistischen Datenanalyse. Die Campus Edition besteht aus IBM Statistics Premium (gemeinhin als "SPSS" bekannt) und IBM SPSS AMOS. IBM SPSS Statistics Premium steht für Windows und macOS zur Verfügung. AMOS ist nur unter Windows nutzbar. Im Normalfall reicht die kostenlose Netzwerklizenz für die Online-Nutzung von SPSS. Eine kostenpflichtig Einzelplatzlizenz benötigen Sie nur, wenn Sie häufig offline mit SPSS arbeiten müssen.

Zielgruppe: Studierende, Beschäftigte

Nutzungskosten:

- Netzwerklizenz (Online-Nutzung): kostenlos
- Einzelplatzlizenz (Offline-Nutzung): Statistics: 50,00 €, Amos: 10,00 €

Kontakt: [IT-Beratung](#)

INTERNET & KOMMUNIKATION

ARBEITSPLATZ

- Cloud-Speicher / sclebo
- Drucken
- Geräte-Ausleihe (Verwaltung)
- Indico
- LearnSmart
- Multifunktionsdrucker
- Netzlaufwerke & Ordner (Verwaltung)
- Notizen-Tool / HedgeDoc
- PC-Pools
- Persönlicher Speicherplatz
- Remote-Desktop
- Remotedesktopverbindung (Homeoffice)
- Software**
- Tipps
- Veranstaltungskalender