

Willkommen zur Vorlesung Statistik

Prof. Dr. Wolfgang Ludwig-Mayerhofer

Universität Siegen – Philosophische Fakultät, Seminar für Sozialwissenschaften

Übersicht

- Formalien, Organisatorisches
- Was ist Statistik?
- Terminologie und Messniveaus

Vorlesung Statistik – für wen?

- BA Social Science/BA Sozialwissenschaften (Pflicht)
- Lehramt SoWi: Für einige Studiengänge Wahlmöglichkeit im Modul Didaktik und Methoden
- Studium Generale
- Sonstige: ??

Nicht-Teilnehmende bitte in LSF abmelden!!

Leistungen/Noten

- Bitte beachten Sie den Unterschied zwischen neuen (2011) und alten (alles vor 2011) Studienordnungen!
- Bitte versuchen **Sie, Ihre** Studien- und Prüfungsordnung zu verstehen! Es gibt inzwischen so viele Ordnungen, dass ich nicht in der Lage bin, einen Überblick zu behalten.
- Grundsätzlich gilt aber für BA Sozialwissenschaften
 - **Alte** Studienordnung(en): Note aufgrund von Klausur am Ende der Vorlesungszeit.
 - **Neue** Studienordnung(en): Gesamtnote für Modul durch Erstellen einer schriftlichen Arbeit nach Absolvieren des ganzen Moduls.

Studienleistung/Prüfung

- Für Fach-BA SoWi, Studienordnung 2011: Regelmäßige Bearbeitung von Tutoriumsaufgaben.
→ 3 LP: Mindestens 8 Aufgaben
- Ohne Bearbeitung der Tutoriumsaufgaben können die Leistungspunkte nicht erworben werden (dokumentierte Nachbereitung der Vorlesung)
- Die Bearbeitung der Tutoriumsaufgaben beginnt nächste Woche. Suchen Sie sich bis dahin eine/n Partner/in – die Aufgaben sollen in 2-er Teams bearbeitet werden. Bitte wählen Sie gut; spätere Partnerwechsel sind nicht möglich!
- Alle anderen: Klausur, möglichst in der vorletzten Vorlesungswoche (Teilnahme ohne Note: Klausur mit geringeren Anforderungen).

Mehr zum Tutorium

- Es gibt kein regelmäßig stattfindendes mündliches Tutorium! Die Bearbeitung der Aufgaben erfolgt mit Hilfe von Moodle. **Nur** Studierende, die beim Bearbeiten erkennen lassen, dass sie große Schwierigkeiten haben, **müssen** dann an einem mündlichen Tutorium teilnehmen (dies gilt zumindest ab 3 KP/LP).
- Sie müssen sich im Moodle-Kurs „Statistik 1“ als Teilnehmer/in anmelden. Um sich allgemein bei Moodle anzumelden, nutzen Sie die Zugangsdaten, mit denen Sie auch Ihr Uni-Mail-Konto einsehen können. Die Anmeldung zu Moodle finden Sie unter <https://moodle.uni-siegen.de/> (Login ganz klein oben rechts). Für den Kurs selbst brauchen Sie keine gesonderten Zugangsdaten.

Was ist Statistik?

„ ... im materiellen Sinn die geordnete Menge von Informationen in Form empir. Zahlen („Statistiken“); im instrumentalen Sinn (Stat. Methoden) der Inbegriff der Verfahren, nach denen empir. Zahlen gewonnen, dargestellt, verarbeitet, analysiert und für Schlußfolgerungen, Prognosen und Entscheidungen verwendet werden.“

Der große Brockhaus Kompaktausgabe, Band 21, 18. Aufl. 1984, S. 29

Beispiel für Statistik im materiellen Sinn

Amtliche Statistik: Das Statistische Jahrbuch (hier: 2010)

13 Land- und Forstwirtschaft
13.15 Anbau und Ernte von Gemüse *)

Jahr Land	Gemüse im Freiland												
	insgesamt *)	dar. Kohlgemüse							dar. Blattgemüse				
		Blumenkohl	Brokkoli	Chinakohl	Grünkohl	Kohlrabi	Ratkohl	Weißkohl	Wirsing	Eisalat	Feldsalat	Kapf Salat	Spinat
Anbaufläche in ha													
2003/2008 D	107 033	5 093	2 135	1 011	1 039	2 230	2 357	6 626	1 206	4 499	1 912	2 784	3 476
2008	132 625	4 819	2 214	1 006	1 149	2 317	2 374	6 767	1 154	4 585	2 318	2 392	3 528
2009	131 072	4 573	2 244	1 049	934	2 353	2 311	6 241	1 112	4 819	2 262	2 213	3 546
davon:													
2009 nach Ländern													
Baden-Württemberg	9 349	152	167	65	(15)	120	181	578	74	393	651	290	139
Bayern	12 802	312	77	234	(11)	191	591	943	87	282	46	393	50
Berlin und Bremen	130	-	-	-	-	-	-	-	-	26	-	0	7
Brandenburg	5 497	28	3	0	8	22	73	49	8	18	3	7	42
Hamburg	430	14	3	3	4	19	1	16	6	5	9	19	11
Hessen	6 779	153	32	37	16	129	178	517	74	32	155	78	159
Mecklenburg-Vorpommern	1 965	80	482	0	2	5	9	41	1	94	37	95	1
Niedersachsen	18 371	767	-	228	399	532	63	235	91	3 381	24	71	250
Nordrhein-Westfalen	20 203	999	294	226	430	798	681	970	469	478	(87)	574	1 265
Rheinland-Pfalz	18 020	1 199	297	206	/	313	(43)	(81)	(74)	80	1 128	659	(789)
Saarland	136	-	2	0	-	3	-	-	-	-	6	10	2
Sachsen	4 727	227	2	1	2	131	45	66	22	7	1	11	225
Sachsen-Anhalt	5 509	69	2	-	6	-	9	15	5	-	-	1	511
Schleswig-Holstein	5 793	345	54	47	37	21	399	2 530	189	2	2	4	2
Thüringen	1 361	226	42	1	0	7	33	194	8	0	0	1	94
Ertrag in dt je ha													
2003/2008 D	X	262,5	149,0	395,1	177,9	287,0	549,4	700,3	335,0	261,4	89,4	277,9	174,4
2008	X	259,8	139,7	409,7	167,6	308,0	571,1	713,4	352,3	223,3	90,9	295,1	177,1
2009	X	283,9	170,6	405,6	169,7	330,2	618,0	809,4	346,3	251,1	89,1	299,9	171,5
davon:													
2009 nach Ländern													

In dieser Vorlesung

„ . . . im materiellen Sinn die geordnete Menge von Informationen in Form empir. Zahlen („Statistiken“); **im instrumentalen Sinn (Stat. Methoden) der Inbegriff der Verfahren**, nach denen empir. Zahlen gewonnen, **dargestellt, verarbeitet, analysiert** und für Schlußfolgerungen, Prognosen und Entscheidungen verwendet werden.“

Der große Brockhaus Kompaktausgabe, Band 21, 18. Aufl. 1984, S. 29

Zahlen sind Daten!

„ ... der Inbegriff der Verfahren, nach denen empir. Zahlen gewonnen, **dargestellt, verarbeitet, analysiert** ... werden.“

- Die zu analysierenden Zahlen (oder Symbole) beziehen sich auf empirische Informationen → **Daten**
- Es geht also nicht um abstrakte Zahlenspielerei – sondern darum, die in den Daten enthaltene Information bestmöglich zu nutzen.

Statistik ist ...

- **leicht**: Außer Grundrechenarten (und Ziehen von Quadratwurzeln) keine Mathematikkenntnisse erforderlich (zumindest gilt dies für diese Vorlesung).
- **wichtig**: Statistik durchzieht unser Leben – von PISA-Studien über Armuts-/Reichtumsbericht der Bundesregierung bis zu Qualitätskontrolle (Mindesthaltbarkeit von Lebensmitteln).
- **ein Weg zum beruflichen Erfolg**: Absolventen mit guten Statistikkenntnissen (auf Master-Niveau) sind am Arbeitsmarkt sehr gefragt.

Hilfsmittel

- Lehrbücher (siehe Vorlesungsübersicht und Semesterapparat in der Bibliothek). Bitte denken Sie daran: Die Vorlesungsunterlagen sind **kein** Lehrbuch, sondern enthalten nur summarische Darstellungen.
- Übungsaufgaben sowie gegebenenfalls mündliche Tutorien
- Ein Tipp von Gero von Randow:

„Ein paar Bemerkungen zum Umgang mit Formeln. Zugegeben, Formeln sind die Geheimwaffe einer internationalen Verschwörung gegen Ihr Selbstbewußtsein. Aber am besten tun Sie so, als würde Ihnen das nichts ausmachen, das verwirrt den Gegner.“

Gero von Randow. Das Ziegenproblem. Denken in Wahrscheinlichkeiten. Rowohlt 1992, S. 17.

Was macht (man mit) Statistik? I

→ Daten anschauen; siehe ausgeteiltes Blatt!

Was macht (man mit) Statistik? II

→ Daten visualisieren (oder sonstwie zusammenfassen)

```
7* | 2
8* | 0
9* |
10* |
11* | 46
12* | 0003445778899
13* | 11236688
14* | 0179
```

Daten zu Studiendauer; Quelle: Ausgeteilte Daten

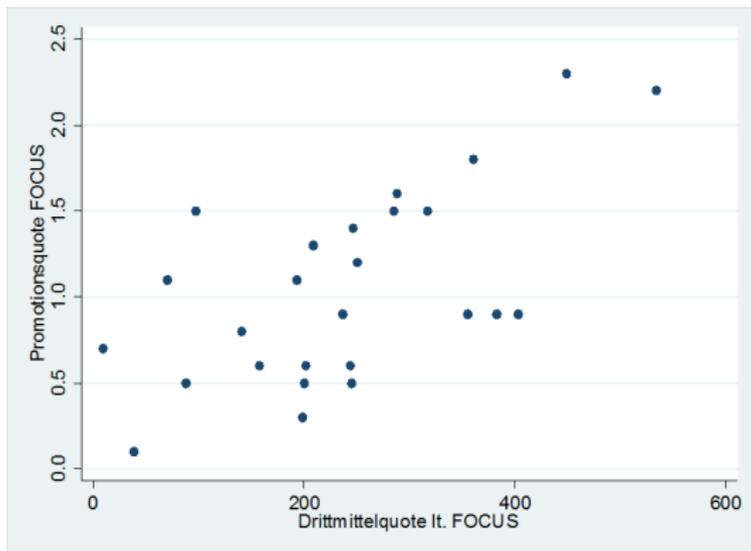
Was macht (man mit) Statistik? III

→ Kennzahlen für die „zentrale Tendenz“ und die „Streuung“ (oder Variabilität) der Datenwerte errechnen.

Ein wichtiges Maß für die zentrale Tendenz kennt jeder/r: den Mittelwert, genauer: das arithmetische Mittel („Durchschnitt“).

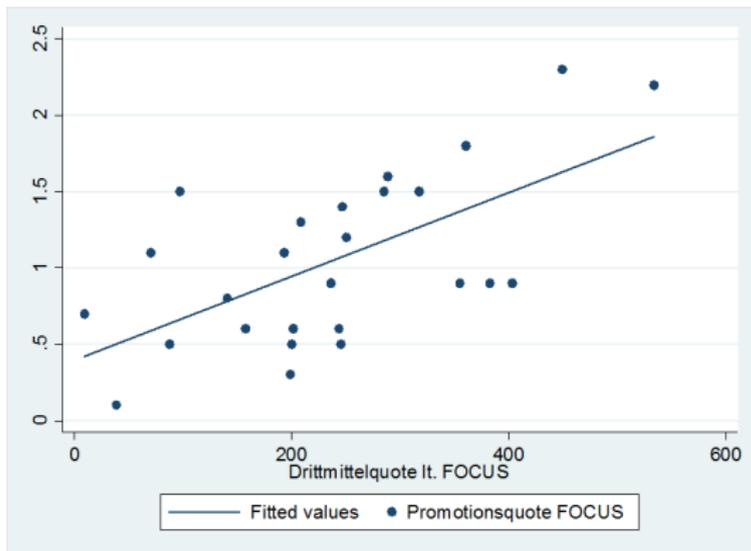
Was macht (man mit) Statistik? IV

→ Zusammenhänge untersuchen



Was macht (man mit) Statistik? V

→ Zusammenhänge modellieren: $Y = a + b \cdot X$



Eine wesentliche Unterscheidung

① **Deskriptive Statistik:**

Daten „beschreiben“ → Daten und ihre Zusammenhänge durch Kennziffern (z. B. Mittelwerte), Graphiken und statistische Modelle auf das Wesentliche reduzieren.

Hinweis: In dieser Vorlesung werden einige Elemente der sog. „explorativen Datenanalyse“ der deskriptiven Statistik zugeordnet.

② **Inferenzstatistik oder schließende Statistik:**

Wenn Stichprobe vorliegt – was sagen die Daten über die Grundgesamtheit? Mit welcher (Un-) Sicherheit ist unser Schluss auf die Grundgesamtheit behaftet?

Mit Inferenzstatistik beschäftigen wir uns in dieser Vorlesung nur kurz – aber intensiv!

Ein wenig Terminologie

- **Urliste**: Ausdruck v.a. statistischer Fachwissenschaftler für die Auflistung der Messwerte
- **Datenmatrix**: Die Anordnung der Daten vor allem (aber nicht nur) in EDV-Programmen für die weitere Analyse
- Die Messwerte (Daten) beziehen sich auf (ein oder mehrere) **Merkmale**. Beispiele: Drittmittelquote; Promotionsquote; Region.
- Die Merkmale weisen in aller Regel mehrere **Ausprägungen** auf (die Ausprägungen/Werte unterscheiden sich von einer Untersuchungseinheit zur nächsten). Dann spricht man auch von **Variablen**.
- Die Untersuchungseinheiten heißen auch **Merkmalsträger**. (Im ausgeteilten Datenblatt: Untersuchungseinheiten = Universitäten.)

Messniveaus/Skalenniveaus

- 1 **Nominalskala**: Unterschiedliche Werte bedeuten „etwas Anderes“, sonst nichts.
- 2 **Ordinalskala**: Werte lassen hinsichtlich „Mehr“ oder „Weniger“, „Besser“ oder „Schlechter“ etc. ordnen, die Abstände lassen sich aber über die Ordnung hinaus nicht numerisch interpretieren.
- 3 **Intervallskala**: Abstände sind inhaltlich interpretierbar (Unterschied zwischen 30°C und 20°C ist identisch mit Unterschied zwischen 20°C und 10°C).
- 4 **Verhältnisskala (Ratioskala)**: Skala hat echten Nullpunkt → Verhältnisse zwischen Messwerten sind interpretierbar (Temperatur in Kelvin!).

In der Forschungspraxis lässt sich die Grenze zwischen Ordinal- und Intervallskala nicht immer eindeutig ziehen.

Messniveaus/Skalenniveaus: Beispiele

- 1 **Nominalskala**: Parteipräferenz; Staatsangehörigkeit; Bundesland; Region.
- 2 **Ordinalskala**: Hiermit bezeichnet man sowohl
 - echte Rangdaten, wie etwa die Platzierung beim Campuslauf (1., 2., ... 271., ...), als auch
 - kategoriale Abstufungen, vor allem in Umfragen. Beispiel: Die Befragten werden um Angabe gebeten, wie oft sie in Büchern lesen: (Beinahe) täglich – mehrmals die Woche – ca. 1x die Woche – mehrmals monatlich – ca. 1x im Monat – seltener – nie.
- 3 **Intervallskala**: Grad Celsius; Geburtsjahr (in menschlichen Kalendern).
- 4 **Verhältnisskala (Ratioskala)**: Körpergröße; Zahl der Kinder; Zahl der Kindergartenplätze pro Einwohner.

Einige weitere Begriffe

- Intervall- und Ratioskala werden auch als **metrische** Skalen bezeichnet, entsprechend spricht man von metrischen Variablen (oder metrischen Daten). Manchmal bezeichnet man beide Skalenniveaus auch als Kardinalskala.
- Nominalskalierte und ordinale Variablen (Daten) (vor allem mit wenigen Ausprägungen) werden auch als **kategoriale**, manchmal auch als qualitative Variablen (Daten) bezeichnet.
- Wir unterscheiden ferner zwischen
 - **Diskreten** Merkmalen: Ausprägungen sind abzählbar.
 - **Stetigen Merkmalen**: Im Prinzip sind alle Werte (evtl. innerhalb eines Intervalls) möglich.

(Auch diese Unterscheidung ist relativ.)