

STATISTIK

Eine interdisziplinäre Einführung



Hans-Joachim Mittag

Inhaltsverzeichnis (Stand: 27. Mai 2010)

Vorwort zum Kurs	i
I Beschreibende Statistik	1
Lernziele zu Teil I	2
1 Statistik, Daten und statistische Methoden	3
1.1 Statistik im privaten Alltag, in Politik und Gesellschaft	3
1.2 Aufgaben und Teilbereiche der Statistik	5
1.3 Methodenkompetenz als Lernziel	6
2 Grundbegriffe der Statistik	10
2.1 Statistische Einheit, Merkmal und Grundgesamtheit	10
2.2 Merkmalsklassifikationen und Skalen	12
2.3 Operationalisierung von Merkmalen	15
3 Datengewinnung und Auswahlverfahren	18
3.1 Erhebungsarten und Studiendesigns	18
3.2 Stichprobenauswahl	25
3.3 Träger amtlicher und nicht-amtlicher Statistik	28
4 Univariate Häufigkeitsverteilungen	31
4.1 Absolute und relative Häufigkeiten	31
4.2 Die empirische Verteilungsfunktion	43
5 Kenngrößen univariater empirischer Verteilungen	47
5.1 Lagemaße	47
5.2 Streuungsmaße	53
5.3 Quantile und Boxplots	57
6 Konzentration von Merkmalswerten	62
6.1 Die Lorenzkurve	62
6.2 Konzentrationsmaße	64
7 Index- und Verhältniszahlen	71
7.1 Verhältniszahlen	71
7.2 Zusammengesetzte Indexzahlen	74

8	Bivariate Häufigkeitsverteilungen	79
8.1	Darstellung empirischer Verteilungen für diskrete Merkmale	79
8.2	Empirische Unabhängigkeit diskreter Merkmale	85
8.3	Darstellung empirischer Verteilungen für stetige Merkmale	92
9	Zusammenhangsmaße	93
9.1	Nominalskalierte Merkmale	93
9.2	Metrische Merkmale	97
9.3	Ordinalskalierte Merkmale	105
	Übungsaufgaben zu Teil I	107
II	Wahrscheinlichkeitsrechnung und schließende Statistik	115
	Lernziele zu Teil II	116
10	Zufallsvorgänge und Wahrscheinlichkeiten	117
10.1	Grundbegriffe der Wahrscheinlichkeitsrechnung	117
10.2	Zufallsstichproben und Kombinatorik	124
10.3	Bedingte Wahrscheinlichkeiten	127
11	Diskrete Zufallsvariablen	132
11.1	Wahrscheinlichkeitsfunktion und Verteilungsfunktion	132
11.2	Kenngrößen diskreter Verteilungen	138
11.3	Die Binomialverteilung	140
11.4	Die hypergeometrische Verteilung	146
12	Stetige Zufallsvariablen	153
12.1	Dichtefunktion und Verteilungsfunktion	153
12.2	Kenngrößen stetiger Verteilungen	156
12.3	Normalverteilung und Standardnormalverteilung	158
12.4	χ^2 -, t - und F -Verteilung	165
13	Bivariate Verteilungen von Zufallsvariablen	170
13.1	Unabhängigkeit von Zufallsvariablen	170
13.2	Kovarianz und Korrelation	174
14	Schätzung von Parametern	177
14.1	Punktschätzung und Eigenschaften von Schätzfunktionen	177
14.2	Schätzung von Erwartungswerten, Varianzen und Anteilswerten	180
14.3	Konfidenzintervalle für Erwartungswerte	181
15	Statistische Testverfahren	185
15.1	Arten statistischer Tests	185
15.2	Grundbegriffe und Gauß-Test für Erwartungswerte	187
15.3	t -Test für Erwartungswerte	198
15.4	χ^2 -Test für Varianzen	200

15.5	Zweistichproben-Tests für Erwartungswerte	201
16	Das lineare Regressionsmodell	204
16.1	Das einfache lineare Regressionsmodell	205
16.2	KQ-Schätzung im einfachen Regressionsmodell	207
16.3	Das Bestimmtheitsmaß	212
16.4	Das multiple lineare Regressionsmodell	214
16.5	KQ-Schätzung im multiplen Regressionsmodell	217
17	Grundzüge der Varianzanalyse	220
17.1	Das Modell der einfaktoriellen Varianzanalyse	222
17.2	Durchführung einer einfaktoriellen Varianzanalyse	223
17.3	Ausblick auf die zweifaktorielle Varianzanalyse	228
	Übungsaufgaben zu Teil II	230
III	Anhänge	239
18	Grundzüge der Matrizenrechnung	240
18.1	Grundbegriffe	240
18.2	Operationen mit Matrizen und Vektoren	242
18.3	Charakterisierung von Zufallsvektoren	247
19	Tabellenanhang	249
19.1	Verteilungsfunktion der Binomialverteilung	249
19.2	Verteilungsfunktion der Standardnormalverteilung	255
19.3	Quantile der Standardnormalverteilung	257
19.4	Quantile der χ^2 -Verteilung	258
19.5	Quantile der t-Verteilung	259
19.6	Quantile der F-Verteilung	260
20	Verzeichnisse und Internet-Ressourcen	264
20.1	Literaturverzeichnis	264
20.2	Ausgewählte Statistiklehrbücher mit Kommentaren	266
20.3	Multimedia-Software und Internet-Ressourcen	268
20.4	Symbolverzeichnis	272
20.5	Autorenregister	273
20.6	Sachregister	275

Vorwort zum Kurs

Dieses Manuskript wird an der FernUniversität Hagen in den Bachelor-Studiengängen „Politik- und Verwaltungswissenschaft“, „Soziologie“ und „Psychologie“ im Grundstudium eingesetzt. Es deckt alle Inhalte einer traditionellen Einführung in die Statistik ab, insbesondere also die beschreibende Statistik sowie Grundlagen der Wahrscheinlichkeitsrechnung und der schließenden Statistik. Die Bearbeitung des Kurses soll dazu befähigen, statistische Informationen nutzen, sachadäquat interpretieren und verständlich kommunizieren zu können. Diese als **statistische Methodenkompetenz** bezeichnete Qualifikation ist in der heutigen Wissens- und Informationsgesellschaft unverzichtbar.

Curriculare
Einbindung
dieses Kurses

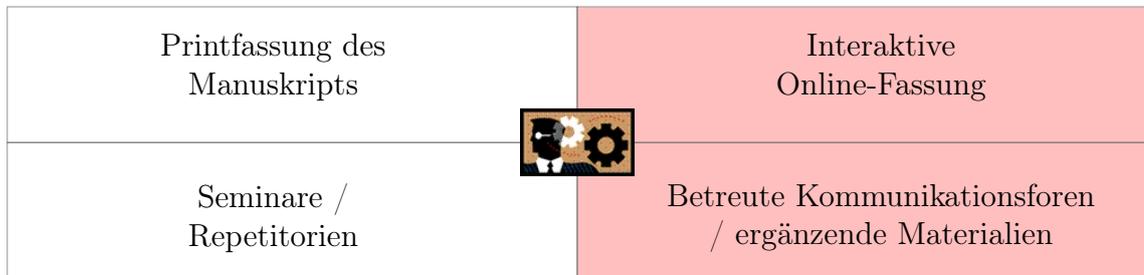
Der Kurs ist eine – zentrale – Komponente eines Medienmixes („**Blended-Learning-Ansatz**“), der „klassische“ Medien (Bücher, Präsenzlehre) und „neue“ Medien (Internet, virtuelle Kommunikationsräume) verbindet. Er wird nicht nur in Printform zur Verfügung gestellt, sondern zusätzlich auch innerhalb einer **Online-Lernplattform** (Open-Source-Lernmanagementsystem **Moodle**) als mehrfarbige pdf-Version. Die pdf-Fassung weist zahlreiche Verknüpfungen mit interessanten Web-Adressen sowie mit interaktiven oder dynamischen Elementen auf. Letztere ermöglichen es, etliche der im Kurs vorgestellten Verfahren anhand von benutzergesteuerten statistischen Experimenten unmittelbar „auszuprobieren“ oder statistische Konzepte über tongestützte Animationen nachzuvollziehen. Die Experimente und Animationen stammen zum überwiegenden Teil aus einem großen **Multimedia-Projekt „Neue Statistik“**.¹ Einige Animationen stammen aus einer **Multimedia-Software „Statistik – Beschreibende Statistik und explorative Datenanalyse“**.² Die ergänzenden Materialien sind *optionale Angebote*, die aktives Lernen unterstützen und zur besseren Verständnissicherung beitragen sollen. Der Kurs ist aber auch ohne Nutzung der statistischen Experimente und Animationen uneingeschränkt studierbar.

Realisierung
eines
Medienmixes

Der Kurs wird auf der Ebene der klassischen Medien durch Seminare oder Repetitorien ergänzt, auf der Ebene der neuen Medien durch ein in die Online-Lernplattform „Moodle“ integriertes moderiertes Kommunikationsforum. Im Forum erhalten die Kursbeleger vom Kursautor oder von virtuellen Mentoren zeitnah Antworten auf ihre Fragen. Die Lernplattform fasst auch alle im Kurs verstreuten statistischen Experimente und Animationen in virtuellen Bibliotheken zusammen. Die Lösungen zu den Übungsaufgaben sind vom Kurs abgekoppelt und in das Lernmanagementsystem „Moodle“ eingestellt.

¹Das Projekt „Neue Statistik“ wurde bis 2004 mit öffentlichen Mitteln gefördert (s. auch unter www.fernuni-hagen.de/neuestatistik – hier wird u. a. ein freier Zugang zu den statistischen Experimenten geboten – und unter www.neuestatistik.de). Im Rahmen des Projekts zeichnete die FernUniversität für statistische Experimente auf Java-Basis verantwortlich, die am Center für Digitale Systeme an der FU Berlin in Lernmodule mit Hypertextstruktur eingebettet wurden.

²Nähere Informationen zum didaktischen Konzept dieser Software und zum gegenwärtigen Entwicklungsstand sind unter www.fernuni-hagen.de/e-statistik zu finden.



Komponenten des Medienmix

Betonung von Interdisziplinarität

Der Kurs illustriert anhand zahlreicher Beispiele aus unterschiedlichen Anwendungsbereichen, dass die Statistik für viele Wissenschaften eine wichtige Servicefunktion erfüllt und alle Bereiche gesellschaftlichen Lebens durchdringt. Viele Beispiele, etwa die Darstellung von Ergebnissen der Nationalen Verzehrstudie II oder die Würdigung einer in den Medien äußerst kritisch diskutierten Statistikaufgabe vom NRW-Zentralabitur 2008, sind fast tagesaktuell.

Einbindung von Daten der amtlichen Statistik

Einige der in diesen Kurs integrierten Datensätze stammen aus der amtlichen Statistik, z. B. von Eurostat, dem Europäischen Amt für Statistik in Luxemburg, an dem der Kursautor vier Jahre als nationaler Sachverständiger tätig war. Die Daten illustrieren die Bedeutung statistischer Informationen für die Planung und das Monitoring nationaler und supranationaler Politiken. Dabei werden auch neuere Entwicklungen thematisiert und kritisch gewürdigt, etwa die zunehmende Verwendung zusammengesetzter Indizes bei internationalen Organisationen.

Bedeutung der Icons am Marginalienrand

Innerhalb des Kurses findet man am Marginalienrand Icons, die auf statistische **Experimente** oder tongestützte **Animationen** (Icon *Lautsprecher*) sowie auf besondere **Web-Links** hinweisen. Anklicken dieser Icons, die mit einem 3D-Effekt gestaltet wurden, führt bei der pdf-Version des Kurses direkt zu den genannten Elementen. Die Steuerung der statistischen Experimente erfolgt über die Schaltfläche „View“ in Verbindung mit Menüfenstern.



Icons mit Verlinkung: Statistisches Experiment, Animation mit Ton, Web-Link

Daneben findet man am Marginalienrand noch Icons, die nur der Orientierung dienen und nicht mit Links verknüpft sind. Diese Icons sind zur besseren Unterscheidung ohne 3D-Effekt ausgeführt, wirken also flacher. Zu nennen sind hier Icons, die auf **Aufgaben** und die zugehörigen **Lösungen** verweisen. In den beiden Literaturverzeichnissen im Kursanhang und auch an einigen Stellen innerhalb des Kurses wird noch ein Icon **Literatur** zur Orientierungsunterstützung verwendet.

In das Manuskript sind zahlreiche **Beispiele** integriert, die deutlich vom normalen Text abgesetzt sind. Dies gilt auch für die an einigen Stellen eingestreuten **Exkurse**, die



Weitere Icons: Aufgabe, Lösung, Literatur

sich an Leser mit Interesse an detaillierteren zusätzlichen Informationen richten. Exkurse können übersprungen werden, ohne dass der rote Faden verloren geht.

Der vorliegende Kurs ist in drei Teile gegliedert. Der erste Teil (Kurseinheit 1) thematisiert die beschreibende Statistik, während sich der zweite Teil (Kurseinheit 2) den Grundlagen der Wahrscheinlichkeitsrechnung und der schließenden Statistik widmet. Daneben gibt es noch diverse Verzeichnisse und Tabellen, die den dritten Teil dieses Manuskripts konstituieren (Kursanhang).

Struktur des Manuskripts

Dank für die Programmierung der statistischen Experimente, auf die dieser Kurs vielfach Bezug nimmt, gebührt vor allem Herrn *Dr. Ulrich Marty*, Arbeitsbereich „Interaktive Medien“ am Lehrgebiet „Praktische Informatik I“ der FernUniversität Hagen (Prof. Dr. G. *Schlageter* / Dr. E. *Heuel*). Er wurde von Frau B. *Schweiger*, Castrop-Rauxel, tatkräftig unterstützt. Herr H. *Kusenber*g, ebenfalls am Arbeitsbereich „Interaktive Medien“ tätig, hat verschiedene Grafiken mit Photoshop bearbeitet und die verwendeten Icons gestaltet. Herrn A. *Hölzle*, Bonn, gilt Dank für seine sehr engagierte und kompetente Mitwirkung bei der Gestaltung der \LaTeX -Umgebung für dieses Manuskript. Ohne die Unterstützung von Herrn *Hölzle* hätte dieses Manuskript kaum die jetzige Form bekommen. Herr Prof. Dr. K.-H. *Renner* und Herrn T. *Heydasch*, Lehrgebiet „Statistische Methodenlehre, Diagnostik und Evaluation“ der FernUniversität, Herr Dr. H.-G. *Sonnenberg*, Arbeitsbereich „Statistik und quantitative Methoden“ der FernUniversität, Frau E. *Gaus-Faltings*, Braunschweig, Herr Dr. B. *Tewes*, Katholische Universität Eichstätt, und Herr J. *Rothenstein*, Universität Wuppertal, haben wertvolle Korrekturhinweise geliefert. Frau *Gaus-Faltings* beantwortet auch Fragen von Studierenden, die in der Plattform „Moodle“ in kursbezogenen Diskussionsforen eingestellt werden. Dank gebührt ferner den Firmen und Institutionen, die kostenfrei Bildmaterial zur Verfügung gestellt haben. Die Namen dieser Unternehmen und Einrichtungen sind jeweils an der Stelle vermerkt, an der das überlassene Material eingebunden ist. Im Einzelnen sind hier zu nennen:

Danksagungen

Fa. Böhme und Weihs Systemtechnik GmbH, Sprockhövel (Herr Dr. N. *Böhme*)
Center für Digitale Systeme, Berlin (Herr Prof. Dr. N. *Apostolopoulos*, Herr A. *Schulz*)
Fa. Evonik Goldschmidt GmbH, Essen (Herr Dr. W. *Wolfes*)
Forschungsgruppe Wahlen, Mannheim (Herr B. *Weber*)
Fa. Q-DAS GmbH, Weinheim (Dr. E. *Dietrich*)
GfK Telecontrol AG, Schweiz (Frau Dr. T. *Hackenbruch*)
Hessischer Rundfunk, Frankfurt (Herr C. *Bender*)
Statistisches Amt von Malta (Herr R. *Mizzi*)
Fa. TNS Infratest, München (Herr M. *Kögel*).

Etliche Rückmeldungen zum Manuskript – etwa Fehlerhinweise zum Kurs, Anregungen

Konstruktive Rückmeldungen zum Kurs sind erwünscht!

für inhaltliche Ergänzungen oder Kommentare zur Ausgestaltung der virtuellen Lernplattform – kamen auch von Studierenden, denen ich hierfür sehr zu Dank verpflichtet bin. Anregungen zur Weiterentwicklung, die per E-Mail (joachim.mittag@fernuni-hagen.de) kommuniziert werden können, sind auch in der Zukunft willkommen.

Hans-Joachim Mittag

Hagen, im Mai 2010