

Abschlussbericht:
eLearning Statistik

Dr. Julia Seeber
Institut für Ökologie
Fakultät für Biologie

und

a.Univ.Prof. Dr. Gilg Seeber
Institut für Politikwissenschaft
Fakultät für Politikwissenschaft und Soziologie

1 Einleitung

Der 2005 von der American Statistical Association verabschiedete Bericht *Guidelines for Assessment and Instruction in Statistics Education for College Introductory Courses* enthält im Kern sechs Empfehlungen für die Gestaltung von einführenden Statistikkursen:

- Emphasize statistical literacy and develop statistical thinking.
- Use real data.
- Stress conceptual understanding rather than mere knowledge of procedures.
- Foster active learning in the class room.
- Use technology for developing concepts and analyzing data.
- Use assessment to evaluate and improve student learning.

(siehe dazu das Vorwort zu Alan Agresti & Christine Franklin, *Statistics: The Art and Science of Learning from Data*, Upper Saddle River, NJ: Pearson 2007).

Diese Empfehlungen haben die Antragsteller versucht, in den von ihnen gestalteten Vorlesungen, Übungen, Proseminaren und Seminaren umzusetzen. Wir waren bestrebt, den Studierenden zu zeigen, dass statistische Methoden hilfreich zur Präzisierung und Analyse von (gewissen) substanzwissenschaftlichen Fragestellungen eingesetzt werden können.

2 Ablauf des Projekts und Einsatz der finanziellen Mittel

2.1 Fakultät für Biologie:

Lehrveranstaltungen „Versuchsplanung und Statistik“ (VO1, 743.203) und „Übungen zu Versuchsplanung und Statistik“ (UE1, 743.204, 8 Parallelkurse):

In der Vorlesung werden die Prinzipien der Planung und Auswertung von naturwissenschaftlichen Untersuchungen betrachtet. Ziel ist zu lernen, wie die inhaltliche Aussage und die Verlässlichkeit der Ergebnisse richtig zu interpretieren und zu bewerten sind. Die Inhalte umfassen Grundlagen der Versuchsplanung, der deskriptiven und der analytischen Statistik und sollen den StudentInnen erste Einblicke in die Datenerhebung und -auswertung geben. In der dazugehörigen Übung sollen die Vorlesungsinhalte anhand von konkreten Fallbeispielen aufgearbeitet werden.

Die im eLearning-Projekt bewilligten Mittel wurden dazu eingesetzt, anwendungsorientierte Datensätze aus Büchern und aus dem Internet zu sammeln, aufzuarbeiten und in den eCampus zu stellen. Mithilfe dieser Datensätze konnte eine eLearning-Einheit im Rahmen der Übungen durchgeführt werden, die ein Drittel der Gesamtnote darstellte. Die Datensätze wurden mit einer Dokumentation versehen, um den Daten eine „Geschichte“ zu geben und sie damit leicht verständlich und praktisch anschaulich zu machen (siehe Beispieldatensatz im Anhang). Weiters wurde eine geeignete Fragestellung erstellt, die gemeinsam mit der Dokumentation

und den Rohdaten zu einem zip-file zusammengefasst wurde, dieses wurde anschließend in einen Fragepool im eCampus hochgeladen. Aus diesem Pool wurde jeder Studentin/jedem Studenten zufällig ein Datensatz zugewiesen, dieser musste vollständig ausgearbeitet und anschließend wieder im eCampus hochgeladen werden.

Drei studentische Mitarbeiter (siehe Projektabrechnung) wurden damit beauftragt, jeweils 20 Datensätze zu suchen und aufzuarbeiten. Diese Datensätze wurden anschließend von der Projektleiterin sondiert und die am besten geeigneten für die eLearning Einheit ausgewählt. Auf diese Art und Weise konnten für die heurige Lehrveranstaltung bereits 30 verschiedene Datensätze inkl. Fragestellungen aufbereitet und den StudentInnen zur Verfügung gestellt werden. Einige der erworbenen Datensätze sind für komplexere Auswertungen besser geeignet und werden deshalb im kommenden Wintersemester im Master-Wahlmodul „Versuchsplanung und Auswerteverfahren“, das von der Projektleiterin (gemeinsam mit Mag. Andrea Bou-Vinals) durchgeführt wird, verwendet werden.

Das heurige Sommersemester wurde dazu verwendet, die Tauglichkeit einer eLearning Einheit im Rahmen der Übung „Versuchsplanung und Statistik“ im Bachelorstudium Biologie (4. Semester) zu testen. Durch eine Studienplanänderung im Bachelor Biologie werden die Lehrveranstaltungen (VO und UE) ab dem kommenden Studienjahr im Wintersemester als VU2 abgehalten werden. Im Rahmen einer kombinierten Vorlesung und Übung wird es nun besser möglich sein, eine um theoretische Inhalte erweiterte eLearning Einheit durchzuführen. Zu diesem Zweck wird zur Zeit ein Pool an theoretischen Fragen erstellt, die bis zum Beginn der eLearning-Einheit im Wintersemester in den eCampus hochgeladen werden.

2.2 Fakultät für Politikwissenschaft und Soziologie

Lehrveranstaltungen „Statistik“ (VO2, 402.055) und „PS Statistik“ (PS2, 402.006 und 402.007)

Die Studierenden der Bachelorstudien Politikwissenschaft und Soziologie besuchen in der Regel die einführenden Statistik-Lehrveranstaltungen im zweiten Semester ihres Studiums. Inhaltlich liegt der Schwerpunkt in der statistischen Datenanalyse, die besprochenen Modell und Methoden werden anhand konkreter sozialwissenschaftlicher Fragestellungen mit (überwiegend) realen Datensätzen besprochen. VO und PS sind eng koordiniert und synchronisiert, im PS lernen die TeilnehmerInnen statistische Analysen unter Anwendung der in der VO besprochenen Methoden mithilfe des Statistikprogramms SPSS selbständig durchzuführen.

In VO und PS gleichermaßen motivieren substanzwissenschaftliche Problemstellungen zur Beschäftigung mit methodischen Fragestellungen, gleichzeitig ist es auch wichtig, reale Analysen anhand realer Datensätzen mit inhaltlich „ernst zu nehmenden“ Antworten zu bieten.

Im Rahmen des Projektes wurde eine Pool von fünfzig solcher Datensätze geschaffen und als SPSS Systemdateien und begleitender Dokumentation zur Verfügung gestellt. Die Bandbreite hinsichtlich Datenstrukturen, Größe der Datensätze und Themenstellungen ist groß und reicht z.B. von Ausschnitten (mit z.T. transformierten Variablen) aus Umfragedatensätzen mit mehreren tausend RespondentInnen (z.B. aus dem Eurobarometer oder dem European Social Survey) zu kleinen Datensätzen mit wenigen Dutzend von Untersuchungseinheiten.

Für die schriftlichen Klausuren wurden den KandidatInnen kleinere Datensätze (im eCampus) zufällig zugewiesen, für die sie am Computer einige statistische Analysen durchzuführen hatten. Die Ergebnis waren auf Papier zu bringen – und in der Folge händisch zu korrigieren.

Das Bachelorstudium Soziologie sieht eine Fortsetzung „Statistik (Vertiefung)“ mit einer zweistündigen VO und einem zweistündigen SE vor. Es ist geplant, im kommenden Wintersemester auch für diese LV auf den aufgebauten Datenpool zurückzugreifen.

3 Mehrwert für die Lehre

Die Bachelorstudien Biologie, Politikwissenschaft und Soziologie platzieren die Statistik Einführungen an den Anfang des Studiums¹. Zu diesem Zeitpunkt haben die Studierenden erst wenige substanzwissenschaftliche Lehrveranstaltungen besucht oder selbst Daten erhoben/analysiert und können deshalb die Notwendigkeit des Erlernens statistischer Methoden nur dann erkennen, wenn ihnen deren „Nützlichkeit“ für die gewählte Wissenschaft demonstriert werden kann. Dazu erscheint es uns notwendig, konkrete inhaltliche Fragestellungen zu erläutern und mit statistischen Analysen anhand realer Datensätze Antworten zu präsentieren. Die Erfahrungen und der Feedback der Studierenden in diesem Semester bestärkt uns, diesen Ansatz auch weiter zu verfolgen.

Biologie

Durch die in die Übung integrierte eLearning-Einheit konnten die TeilnehmerInnen nun erste Erfahrung in der statistischen Datenanalyse (und grundlegende Kenntnisse des Experimentellen Designs) sammeln, die für spätere Praktika und Bachelor- bzw. Masterarbeiten unbedingt nötig sein wird. Hauptaugenmerk wurde hier vor allem auf die inhaltliche Interpretation von statistischen Analysen gelegt, da dies der erste Schritt zu einem besseren Verständnis von Statistik und ihrer Anwendung in der Biologie führen sollte (und geführt hat).

¹ Aus der Sicht des Statistikers (GS) ist diese Platzierung suboptimal: Es scheint sinnvoller, Methoden dann zu erlernen, wenn die Studierenden bereits Kenntnis von den grundsätzlichen Forschungsfragen ihres Faches kennengelernt haben.

Weiters stehen die heuer gesammelten Datensätze auch den Lehrveranstaltungen in den folgenden Jahren (Bachelor-Pflichtmodul „Versuchsplanung und Statistik“, Master-Wahlmodul „Auswerteverfahren und Statistik“) zur Verfügung und können jedes Jahr erweitert und ausgebaut werden. Außerdem ist es immer schwierig, neue Datensätze für Prüfungen und Beispiele in Vorlesungen zu finden, auch für diese Anwendungen konnte und kann auf diesen Datenpool zugegriffen werden.

Politikwissenschaft und Soziologie

Neben dem Aspekt der Erlernes statistischer Methoden ist das Kennenlernen „wichtiger“ (d.h. in der wissenschaftlichen und der öffentlichen Debatte eine prominente Rolle spielenden) Datensätzen von nachhaltiger Bedeutung für den weiteren Verlauf des Studiums.

4 Anhang: Beispieldatensatz

4.1 Beschreibung des Datensatzes

Datensatz Weißfische

In 68 Flüssen wurden Langnasen-Weißfische in 75m langen Probeabschnitten gezählt. Zusätzlich wurden die Größe des Einzugsgebietes, die Menge des gelösten Sauerstoffs, Nitrat und Schwefel, sowie die mittlere Wassertiefe und die Temperatur am Probenabschnitt aufgenommen. Damit soll das Habitat dieser Fischart charakterisiert werden, um daraus Rückschlüsse über die Verbreitungsmuster der Tiere ziehen zu können.

Variable:

Name:	Name des Flusses/Einzugsgebiets
#Fische:	Anzahl der Fische
Einzugsgebiet:	Größe des Einzugsgebiets
O ₂ :	Sauerstoffkonzentration
Tiefe:	Wassertiefe an der gemessenen Stelle
NO ₃ :	gemessene Nitratkonzentration
SO ₄ :	gemessene Sulfatkonzentration
Temp:	gemessene Temperatur

Source:

Maryland Biological Stream Survey

4.2 Fragestellung

Folgende Aufgaben sind zu bearbeiten:

1. Berechnen Sie Mittelwert und Standardabweichung der Variablen Einzugsgebiet, Sauerstoffkonzentration und Sulfatkonzentration und beschreiben Sie kurz die Ergebnisse.
2. Gibt es einen Zusammenhang zwischen der Anzahl an Fischen und der Größe des Einzugsgebietes bzw. zwischen der Anzahl an Fischen und der Sauerstoffkonzentration bzw. zwischen der Anzahl an Fischen und der Sulfatkonzentration? Wenn ja, wie ist der Zusammenhang? Welche Analysen führen Sie dazu durch? Berichten Sie die wichtigsten Kennwerte dieser Analysen und interpretieren Sie die Ergebnisse.
3. Erstellen Sie ein Diagramm, das den Zusammenhang zwischen der Anzahl an Fischen und der Sauerstoffkonzentration darstellt.

5.3 Auszug aus dem Datensatz

Name des Flusses	# Fische	Einzugsgebiet [km²]	O ₂ [mg/l]	Tiefe [cm]	NO ₃ [mg/l]	SO ₄ [mg/l]	Temp. [°C]
BASIN_RUN	13	10,23	9,6	80	2,28	16,75	15,3
BEAR_BR	12	13,49	8,5	83	5,34	7,74	19,4
BEAR_CR	54	79,36	8,3	96	0,99	10,92	19,5
BEAVER_DAM_CR	19	14,45	9,2	56	5,44	16,53	17
BEAVER_RUN	37	6,97	8,1	43	5,66	5,91	19,3
BENNETT_CR	2	2,36	9,2	51	2,26	8,81	12,9
BIG_BR	72	19,38	9,4	91	4,1	5,65	16,7
BIG_ELK_CR	164	145,57	10,2	81	3,2	17,53	13,8
BIG_PIPE_CR	18	102,95	7,5	120	3,53	8,2	13,7
BLUE_LICK_RUN	1	8,97	8,5	46	1,2	10,85	14,3
BROAD_RUN	53	7,98	11,9	56	3,25	11,12	22,2
BUFFALO_RUN	16	51,07	8,3	37	0,61	18,87	16,8
BUSH_CR	32	77,08	8,3	120	2,93	11,31	18
CABIN_JOHN_CR	21	34,85	8,2	103	1,57	16,09	15
CARROLL_BR	23	15,77	10,4	105	2,77	12,79	18,4
COLLIER_RUN	18	25,49	8,6	42	0,26	17,63	18,2
CONOWINGO_CR	112	110,68	8,5	65	6,95	14,94	24,1
DEAD_RUN	25	16,77	8,7	51	0,34	44,93	23
DEEP_RUN	5	4,76	7,7	57	1,3	21,68	21,8
DEER_CR	26	33,58	9,9	60	5,26	6,36	19,1
DORSEY_RUN	8	31,62	6,8	160	0,44	20,24	22,6
FALLS_RUN	15	7,06	9,4	48	2,19	10,27	14,3
FISHING_CR	11	20,42	7,6	109	0,73	7,1	19
FLINTSTONE_CR	11	76,66	9,2	50	0,25	14,21	18,5
GREAT_SENECA_CR	87	34,90	8,6	78	3,37	7,51	21,3
GREENE_BR	33	9,00	9,1	41	2,3	9,72	20,5
GUNPOWDER_FALLS	22	51,23	9,7	65	3,3	5,98	18
HAINES_BR	98	7,96	8,6	50	7,71	26,44	16,8
HAWLINGS_R	1	4,74	8,3	73	2,62	4,64	20,5
HAY_MEADOW_BR	5	2,59	9,5	26	3,53	4,46	20,1
HERRINGTON_RUN	1	28,55	6,4	60	0,25	9,82	24,5
HOLLANDS_BR	38	7,83	10,5	85	2,34	11,44	12
ISRAEL_CR	30	25,33	9,5	133	2,41	13,77	21
LIBERTY_RES	12	1,72	8,3	62	3,49	5,82	20,2
LITTLE_ANTIETAM_CR	24	14,12	9,3	44	2,11	13,37	24