

Stochastik für die Naturwissenschaften

Dr. C.J. Luchsinger

11. Planung von Experimenten, Miscellanea

Literatur Kapitel 11: Stahel: Kapitel 14

auf Englisch: **Design of Experiments** DOE

Sie studieren nicht Mathematik oder Statistik, sondern Biologie, Geographie, Erdwissenschaften oder Biomedizin, also Fächer, in denen das Erlernte aus der Analysis und Wahrscheinlichkeitstheorie und Statistik lediglich angewendet wird. Diese Fächer sind denn auch nicht besonders beliebt. Es ist aber so, dass in der modernen Naturwissenschaft und Technik die Mathematik und Statistik eine grosse Rolle spielt. Deshalb ist es sinnvoll, wenn Sie nach diesen beiden obligatorischen Vorlesungen noch

- * weitere Vorlesungen dazu an dieser Uni besuchen
- * später weitere, massgeschneiderte Kurse besuchen
- * ein Statistik-Paket kennenlernen (zum Beispiel www.r-project.org)
- * statistische Beratung aufsuchen, wenn Sie Daten analysieren müssen (an Uni, ETHZ gratis für Studis und ForscherInnen)

11.1 Ethische und rechtliche Grundlagen; Umgang mit Fehlern in der Forschung

Fast alle Staaten kennen ein **Statistikgesetz**. In der Schweiz gibt es dazu auch die kantonalen Bestimmungen. Unsere Statistikgesetze wurden sogar durch die bilateralen Verträge mit der EU berührt. Sie haben als StaatsbürgerIn oder Unternehmen in einem gewissen Rahmen die *Pflicht*, (korrekte) Daten an die Behörden abzuliefern. Dazu noch ein praktischer Tipp, den ich hier nicht publizieren darf...

Zum Nachdenken: In den USA gibt es eine sogenannte **Freedom of Information Act FOIA**. Sie gibt jedem US-Bürger das Recht, Zugang zu Dokumenten der Exekutive (der

Regierung) der Vereinigten Staaten zu *verlangen*. Bei uns läuft das unter dem Begriff **Öffentlichkeitsprinzip**. Einschränkungen betreffen Datenschutz sowie Anstrengungen und Ermittlungen von Polizei, Armee und Geheimdiensten. Dazu kann man eventuell die teils umstrittene Forderung ableiten, dass Forschungsdaten (Rohdaten) auf Anfrage offenzulegen sind. Wenn die Forschung mit öffentlichen Geldern finanziert wird, folgt auch hiervon, dass die Resultate offenzulegen sind. Was wenn die Forschung (teilweise) privat finanziert wird?

Fehlerbehebung und Fälschung in der Forschung: Der Fall Peter Chen; Forschung in Naturwissenschaft und Technik im Vergleich zu den Sozial- und Gesellschaftswissenschaften

Sie haben Papers auf der Vorlesungshomepage (u.a. Guidelines). Es wird auf die grosse Fehlerrate in den statistischen Analysen aufmerksam gemacht. Um solche Fehler zu verhindern, ist eine gute Ausbildung in Statistik und eine begleitende statistische Beratung in der Forschung notwendig.

Zwei Verhalten, welche vom ungeübten Auge auf den ersten Blick kaum als unsaubere Forschung erkannt werden:

- * das α wird nachträglich festgesetzt (damit das Resultat signifikant wird)
- * vor allem bei Zeitreihen: die Daten werden ohne *vorgängige* Hypothesenwahl betrachtet, dann werden "Muster" von Auge spontan erkannt und dann erst werden dazu Hypothesen formuliert, welche dieses Muster beschreiben. Kaum überraschend sind diese Hypothesen dann hochsignifikant (Finanzwelt und Volkswirtschaftslehre)

11.2 Versuchsplanung, ein paar mathematische Überlegungen

Einleitend sei für den Teil 11.2 und 11.3 bemerkt, dass Sie lediglich bei Laborversuchen die volle Kontrolle über die Daten und deren Erhebung haben. Überall sonst muss man die Daten nehmen, welche man gerade bekommt. Das kann heissen, dass Sie statt der Grösse, welche Sie eigentlich messen wollten, nur eine mit dieser Grösse korrelierte Grösse messen können.

Grosses n im Hinblick auf KI und Tests: $V(\bar{X})$, \sqrt{n} , feinere Unterschiede messbar bzw
sicherere Resultate

$n \gg k$ wo n Anzahl Daten und k erklärende Faktoren im Modell

Multiples Testen (Bonferroni)

Kausalität vs Korrelation; gemeinsamer, zugrundeliegender Faktor

Wie teilt man Untersuchungsobjekte 2 Gruppen zu? Das Prinzip der Randomisierung

Gruppengröße bei 2 Gruppen

Placebo-kontrollierte Doppelblindstudie: ist das wirklich *der* Goldstandard?

Mit Teil der Daten exploratorische Datenanalyse und dann mit anderem Teil konfirmatorische Datenanalyse (nach zufälliger Einteilung)

Standardabweichung und Standardfehler

Wo wählt man bei einer Regression (wenn möglich) die x -Werte?

Nochmals: P-Wert und α

Verteilung des P-Werts unter der \mathcal{H}_0 -Hypothese; Publikation nichtsignifikanter Resultate

Das Simpson-Paradoxon: <https://de.wikipedia.org/wiki/Simpson-Paradoxon>

11.3 Versuchsplanung, konkretes Vorgehen

Sukzessiver Verlust an Information:

1. Realität (Naturgesetze)
2. Was wir sehen/messen
3. Theorie dazu in natürlicher Sprache
4. Math/Stat Modell

Ich verweise hier auf Stahel Kapitel 14; dort hat es in 14.4 auch eine sehr gute Checkliste.
Den Stahel empfehle ich Ihnen für Ihr weiteres Studium und die Forschung!

11.4 Paper/Report/Arbeit schreiben, was gehört in Ihr Paper/Report/Arbeit

Beiblätter: Titelblatt — Copyright — Adresse — Danksagungen [engl Acknowledgments] (selbst bei gewaltigen Problemen mit dem/der BetreuerIn!) — Abstract 1 p, allenfalls ein 5 p Management Summary — Inhaltsverzeichnis [Table of contents] — allfällige wirtschaftliche Abhängigkeiten [Competing interests] — Liste der Anhänge [Table of appendixes] — Liste der Tabellen [List of tables] — Liste der Graphiken [List of figures] — Referenzen — Programm code — allenfalls CV

Worum geht es: Einführung [Introduction] — Problemdefinition [Statement of the problem] — Hintergrundinformationen zum Problem [Background of the problem] — Rationale Erklärung der Untersuchung [Rationale for the study] — Forschungsfrage und/oder Hypothesen [Research questions and/or Hypotheses] — Annahmen, Einschränkungen [Assumptions, Limitations] — Definition der wichtigsten Wörter

Literaturstudium: Vergangene Forschung — verwandte Forschung

Methoden: Fragebogen — Tests — Instrumente — Definition der relevanten Population — Methode der Stichprobenerhebung — eigentliche Versuchsplanung

Datenanalyse: Resultate — \mathcal{H}_0 angenommen oder verworfen — Tabellen

Zusammenfassung, Schlussfolgerungen, offene Fragen

Und noch eine hoffentlich unnötige Bemerkung: falls Sie in Ihrer Arbeit die statistische Datenanalyse von jemand anderem machen lassen und diese Person nicht mal in der Arbeit erwähnen (das ist illegal), dann schauen Sie doch wenigstens, dass Sie die dort verwendeten Methoden kennen. Konkret: sagen Sie der Person, sie soll Ihnen doch bitte die verwendeten Methoden erklären (inkl. warum diese Methode und nicht eine andere; Vor- und Nachteile).