

## Vorlesung "Empirische Bewertung in der Informatik"

Freie Universität Berlin, Institut für Informatik, Arbeitsgruppe Software Engineering  
Prof. Dr. Lutz Prechelt, Stephan Salinger  
Übungsblatt 8 SS 2007 zu den Übungen 2007-06-25/2007-07-02

### Aufgabe 8: (Auswertung der Befragung)

Dieses Übungsblatt geht über zwei Wochen.

Erfassen Sie die Antwortdaten Ihrer Umfrage und werten Sie sie statistisch aus.

- Sie sollten für eventuelle spätere Nachprüfungen jede eingegangene Antwort mit einer Nummer versehen und diese in die Daten mit aufnehmen.
- Legen Sie fest, wie Sie unklare bzw. mehrdeutige Antworten handhaben wollen und *schreiben sie dies auf*.
- Erfassen Sie die Daten aller Fragebögen.
  - Die quantitativen Daten als Textdatei mit einem Datensatz pro Zeile und Semikolon oder TAB als Feldtrenner. (Das Einlesen in R muss mit der Funktion `read.table` möglich sein.)
  - Die qualitativen Daten als Textdatei in einem geeigneten und selbsterklärenden Format.
- Validation: Prüfen Sie die Daten durch Plausibilitätsprüfungen sowie mindestens stichprobenhafte Kontrolle auf Erfassungsfehler.
  - Die ganz gründliche Methode wäre, sämtliche Daten von beiden Teammitgliedern separat (also doppelt) erfassen zu lassen, und dann die Ergebnisse zu vergleichen. Der Vergleich lässt sich (zumindest bei allen schematischen Antworten) natürlich leicht automatisieren

8-1: Schicken Sie die Datei(en) mit Ihren *validierten* Rohdaten **bis spätestens Montag, 9:00 Uhr, der ersten Woche an salinger[klam\*mer\*af\*fe]inf.fu-berlin.de** (Betreff: [UE Empir07-8.1][Gruppe<Gruppennummer>]).

- Fertigen Sie unter Einsatz von R summarische Übersichten an:
  - Demographisches Profil der Teilnehmer
  - Verteilung der Antworten zu jeder Frage
  - Gibt es Fragen mit auffallend hohen oder niedrigen Antworten oder auffallend starken oder geringen Schwankungen der Antworten?
    - Falls ja: Sind die Ursachen vermutlich inhaltlicher Art (die Beobachtungen also gültig) oder lag eine Schwäche im Fragebogen zu Grunde?
  - Gibt es Fragen, die auffallend häufig gar nicht beantwortet wurden oder bei denen die Antworten das Schema des Fragebogens verletzen (z.B. mehrere Antworten, wo nur eine erlaubt war)?
    - Falls ja: Woran liegt das vermutlich?
- Untersuchen Sie unter Einsatz von R Zusammenhänge:
  - Gibt es bei bestimmten Fragen erkennbare Unterschiede zwischen unterschiedlich charakterisierbaren Teilnehmern/Teilnehmergruppen? Die Charakterisierung der Teilnehmer hängt hierbei von Ihrem Fragebogen ab. Es können z.B. verschiedene Fähigkeitsstufen oder verschiedenen Anwendungsdomänen sein.

- Falls ja: Fügen sich mehrere solche Unterschiede zu einem erkennbaren Gesamtbild über Fähigkeitsstufen, Anwendungsdomänen etc.?
  - Sind die einzelnen Unterschiede statistisch signifikant?
  - Lassen sich irgendwelche Zusammenhänge auf einleuchtende Weise visualisieren?
- Nehmen Sie abschließend auch noch nicht-quantitative Information in die Analyse auf (falls Sie so etwas erhoben haben, z.B. Kommentarfelder):
  - Gibt es in den qualitativen Antworten deutliche Trends und Wiederholungen?
  - Falls ja: Unterstützen diese Ergebnisse die aus den verwandten quantitativen Fragen? Oder gibt es Widersprüche?
  - Falls nein: Erklärt das, warum zugehörige quantitative Fragen kein eindeutiges Bild ergeben haben?
- Diskutieren Sie Ihre Ergebnisse und/oder offene Fragen bei der Analyse informell mit dem anderen Team, das bei der Fragebogendurchsicht Ihren Fragebogen bearbeitet hatte.
- Beantworten Sie folgende Fragen:
  - Welches sind die eindeutigsten Aussagen, die sich aus der Analyse der Daten gewinnen lassen?
  - Welche erheblichen Mängel an Ihrem Fragebogen haben Sie angesichts der Daten bemerkt?
  - Welche Erfassungsfehler haben Sie in Ihren Daten trotz vorheriger Validierung im Laufe der Analyse noch entdeckt? Wodurch?
  - Wozu lassen sich *keine* klaren Aussagen ableiten, obwohl Sie das erwartet hätten?
  - Welche genaueren Analysen hätten Sie gern durchgeführt, wozu Ihnen aber die Statistik-Kenntnisse fehlen?
- **Behalten Sie bei Ihrer Analyse Ihre ursprüngliche Forschungsfrage im Auge und versuchen Sie aus den Daten so viel wie möglich zu deren Beantwortung herauszuholen.**
- Wenn Sie irgendeine Auswertung nicht wie gewünscht mit R hinbekommen, fragen Sie auf der se\_v\_empir-Mailingliste um Hilfe nach.

8-2: Schicken Sie die Antworten auf diese Fragen **bis spätestens Montag, 9:00 Uhr, der zweiten Woche an salinger[klam\*mer\*af\*fe]inf.fu-berlin.de** (Betreff: [UE Empir07-8.2][Gruppe<Gruppennummer>]) und senden Sie als Anhang eine Textdatei mit, die die R-Kommandos enthält, die Sie für Ihre Auswertungen benutzt haben (Name: Gruppe<Gruppennummer>.R)