

# Empirische Methoden für Informatiker

## Teil 3: Vorgehen bei Empirische Forschung

Christian Kästner

# Kontext

---

- ▶ Bisher: Performancemessungen
- ▶ Relativ einfach
  - ▶ rel. geringer Aufwand, automatisierbar, oft wiederholbar
  - ▶ klare objektive Messverfahren (Zeit, Speicher)
  - ▶ kaum menschlicher Einfluss (eher technische Störvariablen)
- ▶ Aber:
  - ▶ Lesbarkeit und Verständlichkeit von Quelltext?
  - ▶ Korrektheit von Quelltext?
  - ▶ Qualität von Vorschlägen für Autovervollständigung?
  - ▶ Softwareentwicklungskosten und Wartungskosten?
  - ▶ Entwicklerproduktivität?

# Probleme (Auszug)

---

- ▶ Teure Experimente, wenige Probanden
- ▶ Menschlicher Einfluss
  - ▶ Frageboegen
  - ▶ subjektive Einschaeztungen
  - ▶ Erwartungshaltung
- ▶ Schwierige Messung (Korrektheit, Lesbarkeit, ...)
  - ▶ X ist 20% lesbarer als Y?
  - ▶ Entwickler mit 50% mehr Erfahrung produzieren 90% weniger Fehler?

# Agenda: Vorgehen bei empirischen Evaluierungen

---

- ▶ Variablen
- ▶ Abwaegungen zur internen/externen Validitaet
- ▶ Fishing for Results
- ▶ Quantitative vs. qualitative Untersuchungen
- ▶ Ausblick



# Variablen

# Unabhaengige Variablen

---

- ▶ Bewusst kontrolliert variiert
- ▶ Ergeben sich i.d.R. aus der Forschungshypothese
- ▶ Abstufungen muessen klar erkennbar/messbar sein
  
- ▶ Variablen und ihre Stufen, z.B:
  - ▶ Sortierverfahren (Quicksort vs. Mergesort)
  - ▶ Sprache (Java vs. Scala vs. Haskell)
  - ▶ Datenbanksystem (Oracle vs. MySQL)
  - ▶ Speichergroesse (...)

# Abhaengige Variablen

---

- ▶ In Abhaengigkeit der unabhaengigen Variablen gemessen
- ▶ Muss messbar sein
- ▶ z.B.
  - ▶ Geschwindigkeit von Quicksort vs. Mergesort
  - ▶ Speicherverbrauch von Scala vs. Java vs. Haskell
  - ▶ Verstaendlichkeit der Implementierung

# Drittvariablen / Stoervariablen

---

- ▶ Beeinflussen abhaengige Variablen
  - ▶ (zusätzlich zur unabhängigen Variablen)
- ▶ z.B. Prozessorgeschwindigkeit, Speichergroesse, Hintergrundprozesse, Erfahrung des Entwicklers
- ▶ Umgang z.B.:
  - ▶ Miterheben, um spaeter den Einfluss zu ermitteln
  - ▶ Konstant halten, um Einfluss zu eliminieren
  - ▶ Zufaellig verteilen (Randomisierung), um Einfluss im Mittel aufzuheben



# Laboruntersuchung vs. Felduntersuchung

---

- ▶ Konstanthalten von Drittvariablen im Labor
  - ▶ “Quicksort ist schneller als Mergesort bei den Daten X auf Computer Y wenn implementiert mit Z von V’.”
  - ▶ Zuverlaessige Messung der abhaengigen Variablen (hohe interne Validitaet)
  - ▶ Nicht verallgemeinerbar auf andere Belegungen der Drittvariable (geringe externe Validitaet)
  - ▶ Aus praktischen und ethischen Gruenden nicht immer moeglich
- ▶ Untersuchung im Feld, Drittvariablen nicht immer kontrollierbar
  - ▶ Hohe externe Validitaet
  - ▶ Geringe interne Validitaet



=> Kompromiss

# Phasen der empirischen Forschung

nach Bortz und Schuster (2011)

# Phasen der empirischen Forschung

---

- ▶ Erkundungsphase
- ▶ Theoretische Phase
- ▶ Planungsphase
- ▶ Untersuchungsphase
- ▶ Auswertungsphase
- ▶ Entscheidungsphase

# Erkundungsphase

---

- ▶ Literatur sichten
- ▶ Informationsgespräche mit Praktikern
- ▶ Forschungshypothesen formulieren

# Theorie

---

“Theorien haben die Funktion, Sachverhalte zu beschreiben, zu erklären und vorherzusagen. Im Kern bestehen sozialwissenschaftliche Theorien aus einer Vernetzung von gut bewährten Hypothesen bzw. anerkannten empirischen ‘Gesetzmäßigkeiten’”

(Bortz und Doering)

# Theoretische Phase

---

- ▶ Ist die Forschungshypothese praezise formuliert? (Begriffe klar definiert?)
- ▶ Informationsgehalt der Forschungshypothese? Sinnvoll fuer die Anwendung?
  - ▶ “Kommentare im Quelltext beeinflussen Entwickler”
  - ▶ “Je mehr Kommentare im Quelltext, desto einfacher die Wartung”
  - ▶ Wenn-Dann, Je-Desto
- ▶ Logisch konsistent? (Tautologie, Widerspruch?)
- ▶ Empirisch ueberpruefbar?
  - ▶ Messbar? Falsifizierbar?

# Planungsphase

---

- ▶ Auswahl der Variablen: *Was* messen?
  - ▶ Unabhaengige Variablen – werden kontrolliert variiert
  - ▶ Abhaengige Variablen – werden gemessen
  - ▶ Kontrollvariablen, Stoervariablen
- ▶ Operationalisierung: *Wie* messen?
  - ▶ Wie misst man zuverlaessig Speicherauslastung?
  - ▶ Insb. bei komplexen Variablen wie “Verstaendlichkeit” schwierig
- ▶ Stichprobenumfang: Wieviel messen?
- ▶ Planung der statistischen Auswertung
  - ▶ Welche Tests, welche Signifikanzniveaus

# Untersuchungsphase

---

- ▶ Durchfuehrung von Untersuchung/Messung/Experiment
- ▶ Stark abhaengig von Versuchsaufbau
- ▶ Ausblick: Versuchsleitereffekte bei kontrollierten Experimenten



# Auswertungsphase

---

- ▶ Ergebnisse objektiv bewerten (bevorzugt standardisierte Testverfahren)
- ▶ Daten darstellen/aufbereiten
- ▶ Hypothesen ueberpruefen (Inferenzstatistik, z.B. Signifikanztests)

# Hypothesen

---

- ▶ Hypothesen koennen nur widerlegt, nicht bewiesen werden
- ▶ Drei Arten von Hypothesen
  - ▶ Forschungshypothese: Beschreibt Untersuchungsgegenstand, Forschungsfragen
  - ▶ Nullhypothese  $h_0$ : Aus Forschungshypothese abgeleitet, zu widerlegende Aussagen
  - ▶ Alternativhypothese  $h_1$ : Negation der Nulltypothese
- ▶ Beispiel
  - ▶ Forschungshypothese  $h_1$ : “Verfahren X verbessert die Lesbarkeit”
  - ▶ Nullhypothese  $h_0$ : “Die Lesbarkeit unterscheidet sich nicht”

Statistisches  
Hypothesen-  
paar

# Entscheidungsphase

---

- ▶ Statistisches Test widerlegen die Nullhypothese ueberzeugend
  - ▶ Interpretation als Bestaetigung fuer die Alternativhypothese
  - ▶ “kein Grund, um an der Theorie zu zweifeln”
  - ▶ Dennoch: Kein absoluter Beweis fuer Korrektheit!
  - ▶ Bei wiederholter Bestaetigung zunehmende Akzeptanz, Aufnahme in die Praxis
- ▶ Statistischer Test widerlegt die Nullhypothese nicht
  - ▶ Untersuchung war nicht geeignet, die Nullhypothese zu entkraefen
  - ▶ Fehler in der Untersuchung?
  - ▶ Nullhypothese gilt?
  - ▶ Zweifel an der Allgemeingultigkeit der urspr. Hypothese
  - ▶ Ueberarbeiten der Hypothese

# Exhaustion / Ausschöpfung

---

- ▶ Modifikation einer Theorie durch falsifizierte Hypothese
  - ▶ “X verbessert die Lesbarkeit bei Probanden im 4. Semester in Programmiersprachen ausser Java”
- ▶ Durch wieviele Exhaustionen kann eine Theorie belastet werden?
  - ▶ Informationsgehalt der Theorie geht gegen Null
  - ▶ Interesse an Theorie laesst bei mehrfacher Falsifikation/Exhaustion nach, Alternativen werden gesucht

- ▶ “So ist die empirische Basis der objektiven Wissenschaft nichts ‘Absolutes’; die Wissenschaft baut nicht auf Felsengrund. Es ist eher ein Sumpfland, über dem sich die kühne Konstruktion ihrer Theorien erhebt; sie ist ein Pfeilerbau, dessen Pfeiler sich von oben her in den Sumpf senken – aber nicht bis zu einem natürlichen ‘gegebenen’ Grund. Denn nicht deshalb hört man auf, die Pfeiler tiefer hineinzutreiben, weil man auf eine feste Schicht gestoßen ist: Wenn man hofft dass sie das Gebäude tragen werden, beschließt man, sich vorläufig mit der Festigkeit der Pfeiler zu begnügen.”

(Popper, 1966)

## **An exploratory study of the effect of aspect-oriented programming on maintainability**

**Marc Bartsch · Rachel Harrison**

### 3.7 Experimental design

A between-subjects design was used to test the hypotheses. The dependent variable was the system used (object-oriented or aspect-oriented) and the independent variables were software understanding and modifiability. Learning effects were avoided since each subject only answered questions with one system. The data were collected online. After a participant had sent in solutions to all five exercises from the aspect-oriented tutorial, he or she was sent an email as described in the previous section. The subject was asked to mail back the answers to the questionnaire. Neither during the preparation phase nor during the experimental phase were the names of the participants disclosed within the group. All participants were locally separated. Also, there were no deadlines set for the solution to the exercises and the questionnaire. This design was established for the



## Fishing for Results

# Erkundungsexperiment vs. Inferenzstatistik

---

- ▶ Erkundungsexperiment zur Bildung von Theorien
  - ▶ Daten erheben, um Hypothesen zu formulieren
  - ▶ In erhobenen Daten nach Zusammenhaengen suchen
- ▶ Inferenzstatistik
  - ▶ Ueberpruefung von Hypothesen
- ▶ **Wichtig:** Bildung der Hypothese und Ueberpruefen der Hypothese auf unterschiedlichen Daten
  - ▶ Kann sonst leicht zufaellige Zusammenhaenge belegen
  - ▶ Suchen und belegen von Zusammenhaengen in den gleichen Daten ist unwissenschaftlich



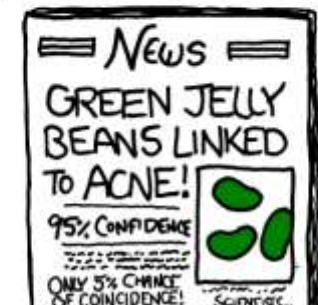
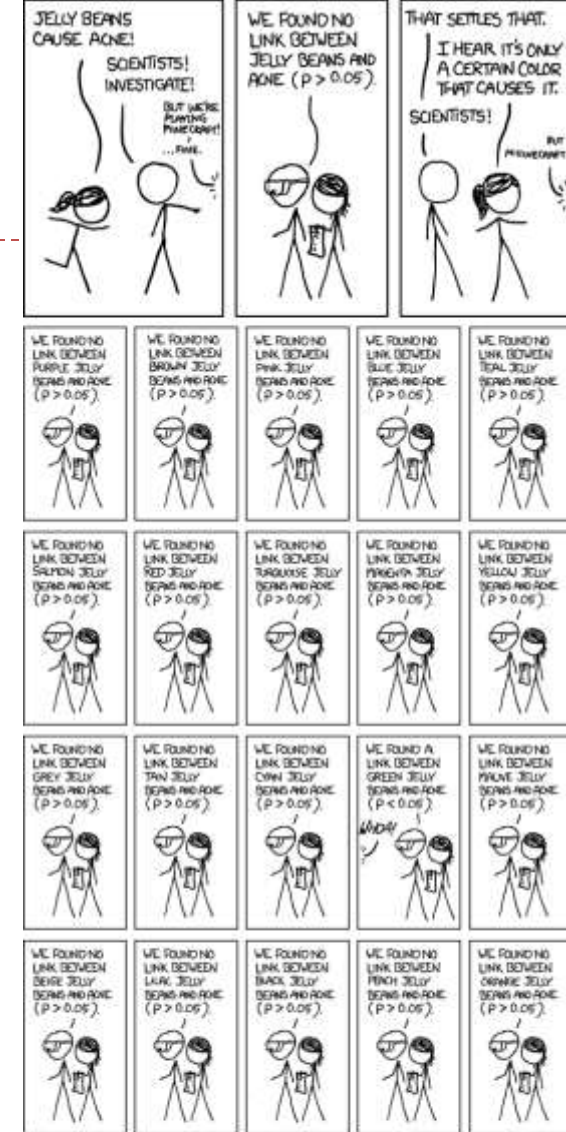
# Fishing for Results

- ▶ Irgendwelche statistisch signifikanten Zusammenhaenge finden sich immer

Deshalb:

- ▶ Hypothese vor Beginn des Experiments festlegen und begruenden
- ▶ Hypothesen, die sich beim Auswerten ergeben, als solche klar markieren
  - ▶ Erkundungsexperiment / Explorative Studie

“Fuer den sinnvollen Einsatz der Inferenzstatistik ist es erforderlich, dass vor Untersuchungsbeginn eine theoretisch gut begruendete Hypothese formuliert wurde.”





Ausblick

# Qualitative vs. Quantitative Methoden

---

## ▶ Qualitative Methoden

- ▶ Verbale Daten, Beobachtungsprotokolle, Interviewtexte, Fotos
- ▶ Mehr Zeit, weniger Personen
- ▶ “reichhaltige Daten”
- ▶ Interpretation

## ▶ Quantitative Methoden

- ▶ Numerische Daten, Messwerte
- ▶ Skalen
- ▶ “genormte Daten”
- ▶ Statistik

## Aufgaben bis 21.5 (wg. Sport Dies)

---

- ▶ Lesen Sie eine der drei Publikationen (Paper 2, 3, oder 4) und bereiten Sie sich auf eine Präsentation und Diskussion vor.
- ▶ Lesen Sie Abschnitt 5.5 von Paper 5. Identifizieren sie einzelne Threats to Validity. Bereiten Sie sich auf eine kurze Präsentation und Diskussion vor.
- ▶ Recherchieren Sie die Hintergründe zu denen im Comic <http://xkcd.com/882/> thematisierten Problem. Warum ist es ein Problem? Wie kann man es vermeiden?

# Literatur

---

- ▶ Bortz und Döring. **Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler**. Kapitel 1, Springer, 2006.
  - ▶ aus Uninetz als PDF verfügbare:  
<http://www.springerlink.com/content/r35544/>
- ▶ Bortz und Schuster. **Statistik für Human- und Sozialwissenschaftler**. Kapitel 1, Springer, 2010.
  - ▶ <http://www.springerlink.com/content/w68677/>