UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

Quantitative Methoden

Hausaufgabe

- Validitätsarten
- Sinn der ANOVA
- Paper

Weitere Validitätsarten

- Konstruktvalidität
  - Beschreibt, wie gut das Konstrukt gemessen wird
  - Bsp: Programmverständnis, Intelligenz
- Statistical Conclusion Validity
  - Angemessenheit der statistischen Methoden

Multiples Testen-Beispiel (1)

- Faktor mit 4 Stufen, jeweils paarweise Vergleiche
- Insgesamt: $\binom{4}{2} = 6$

- Wahrscheinlichkeit, eine H_0 korrekterweise zu behalten: 0.95
- Wahrscheinlichkeit, zwei H_0 korrekterweise zu behalten: 0.95*0.95
- Wahrscheinlichkeit, sechs H_0 korrekterweise zu behalten: 0.95^6

Multiples Testen-Beispiel (2)

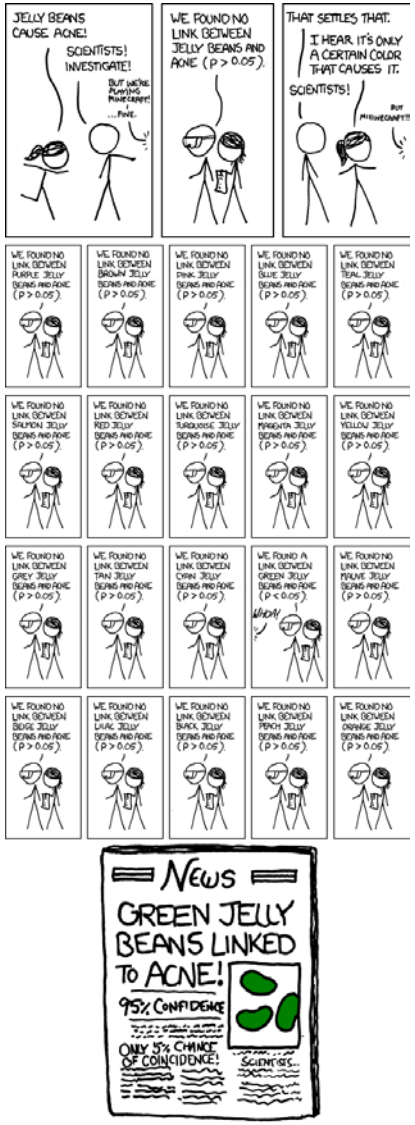
- Wahrscheinlichkeit, dass bei sechs Tests mindestens einer signifikant ist:
- $1 - 0.95^6 = 0.26$

Multiples Testen

- Bei mehreren Signifikanztests muss das Signifikanzniveau angepasst werden
- Bonferoni-Korrektur:
  - t: Anzahl Tests
  - $\alpha' = \alpha / t$
  - $\alpha / 6 = 0.0083$

- α-Fehler (Fehler 1. Art), dass grüne Jellybeans Akne verursachen: 64%
- Angepasstes Signifikanzniveau:

  0.05/20 = 0.0025



ANOVA mit R

- http://rtutorialseries.blogspot.de/
- One-Way Omnibus ANOVA:

```
> anova(lm(Values ~ Group, dataOneWay))
Analysis of Variance Table

Response: Values
          Df Sum Sq Mean Sq F value    Pr(>F)
Group      1     60  60.000  64.444 5.503e-11 ***
Residuals 58     54   0.931
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

R: Zusammenfassung

- Two-Way Omnibus ANOVA:
  - anova(lm(Values ~ Group * Gender, dataTwoWay))

```
> anova(lm(Values ~ Group * Gender, dataTwoWay))
Analysis of Variance Table

Response: Values
             Df Sum Sq Mean Sq F value  Pr(>F)
Group         1 60.000  60.000 67.9245 3.1e-11 ***
Gender        1  0.267   0.267  0.3019 0.58489
Group:Gender  1  4.267   4.267  4.8302 0.03212 *
Residuals    56 49.467   0.883
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

R: t-Test

Demo.R

```
> shapiro.test(rt)

 Shapiro-Wilk normality test

data:  rt
W = 0.9472, p-value = 0.5559


> input <- read.csv("G:/work/lehre/EMCS/rt.csv", sep=";", dec=",")
> rt <- input[,'time']
> rt1 <- subset(input,group==1)[,'time']
> rt2 <- subset(input,group==2)[,'time']
> t.test(rt1, rt2)

 Welch Two Sample t-test

data:  rt1 and rt2
t = 1.5222, df = 10.566, p-value = 0.1573
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.095727 27.583584
sample estimates:
mean of x mean of y
 50.74243  39.49850
```

R: Mann-Whitney-U Test

Demo.R

```
> wilcox.test(rt1,rt2,paired=FALSE)

 Wilcoxon rank sum test

data:  rt1 and rt2
W = 31, p-value = 0.1807
alternative hypothesis: true location shift is not equal to 0
```

R: Chi^2

- http://ww2.coastal.edu/kingw/statistics/R-tutorials/independ.html

```
> row1 = c(91,90,51)                    # or col1 = c(91,150,109)
> row2 = c(150,200,155)                 # and col2 = c(90,200,198)
> row3 = c(109,198,172)                 # and col3 = c(51,155,172)
> data.table = rbind(row1,row2,row3)    # and data.table = cbind(col1,col2,col3)
> data.table
     [,1] [,2] [,3]
row1   91   90   51
row2  150  200  155
row3  109  198  172
> chisq.test(data.table)

        Pearson's Chi-squared test

data:  data.table
X-squared = 25.086, df = 4, p-value = 4.835e-05
```

R: Korrelation

Demo.R

```
> cor.test(rt,rtTask2, method="pearson")

 Pearson's product-moment correlation

data:  rt and rtTask2
t = 4.6652, df = 11, p-value = 0.0006878
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4792664 0.9426838
sample estimates:
      cor
0.8150282
```

R: Korrelation

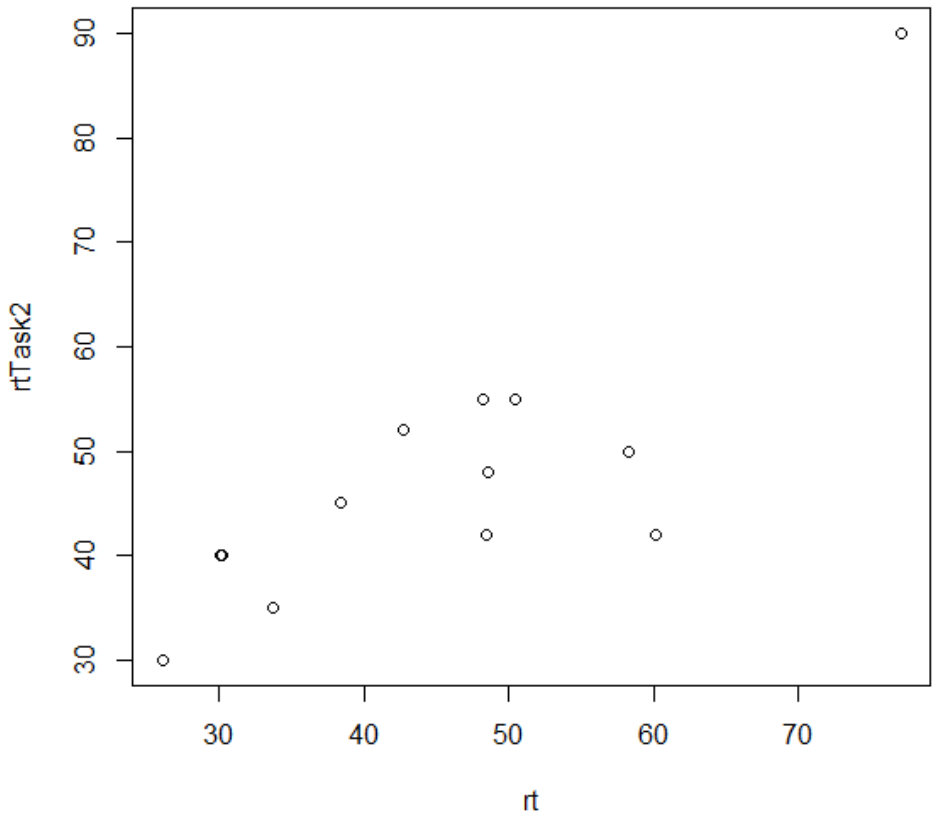
Demo.R

```
> cor.test(rt,rtTask2, method="spearman")

 Spearman's rank correlation rho

data:  rt and rtTask2
S = 102.9219, p-value = 0.005786
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.7172475

Warning message:
In cor.test.default(rt, rtTask2, method = "spearman") :
  Kann exakten p-Wert bei Bindungen nicht berechnen
> 
```

	A	B	C	D	E
1	probCode	group	time	time2	
2	ATM	1	42,744	52	
3	BQV	1	60,1	42	
4	cno	1	30,139	40	
5	ikx	1	77,047	90	
6	KQR	1	58,231	50	
7	LOF	1	48,54	48	
8	OLCAA	1	38,396	45	
9	BTM	2	48,438	42	
10	mdp	2	48,245	55	
11	RPR	2	26,052	30	
12	TZX	2	50,436	55	
13	VND	2	30,077	40	
14	vtd	2	33,743	35	
15					

Demo.R

```
plot(rt,rtTask2)
```



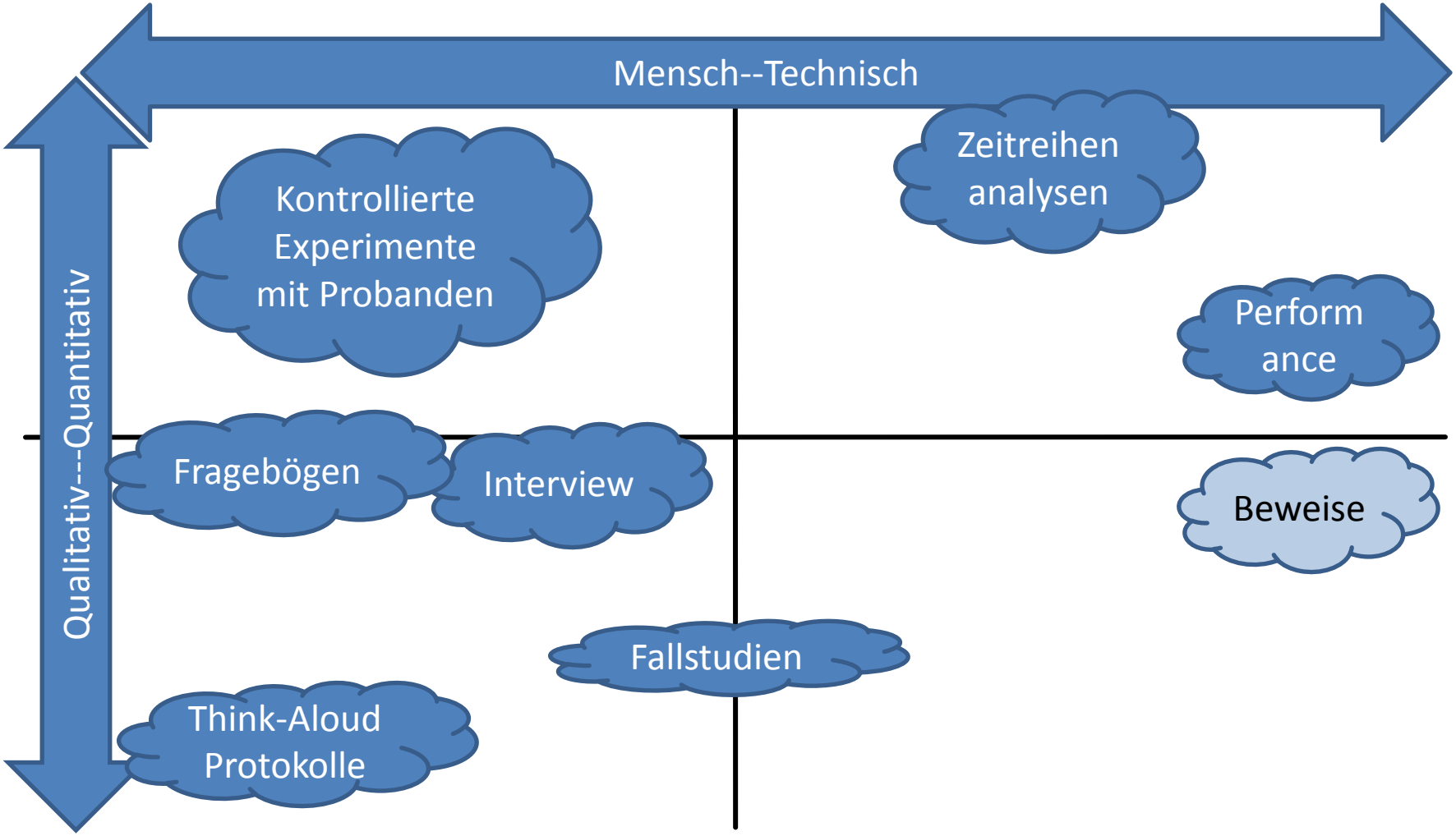
Black Story

- Hausaufgabe zum 5.6.:
  - Paper aussuchen und auf Spiel vorbereiten
  - Zusammenfassung schreiben (s. Folien „BeispielExperimente“)
  - Anregungen (Paper-Titel):
    - Using Students as Subjects: An Empirical Evaluation
    - An Empirical Study of the Effects of Personality in Pair Programming using the Five-Factor Model
    - Understanding Exception Handling: Viewpoints of Novices and Experts
    - The Relevance of Application Domain Knowledge: The Case of Computer Program Comprehension

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

Qualitative Methoden

Überblick

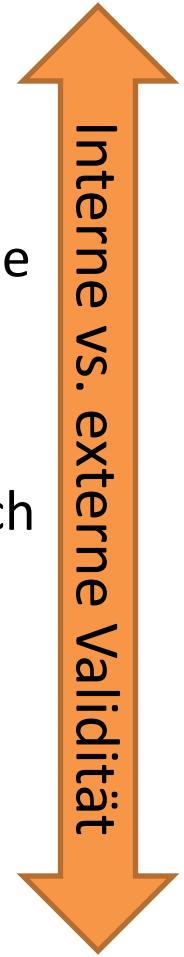

Mensch--Technisch
Qualitativ---Quantitativ
Kontrollierte Experimente mit Probanden
Zeitreihen analysen
Perform ance
Fragebögen
Interview
Beweise
Fallstudien
Think-Aloud Protokolle

Lernziele

- Einsatzmöglichkeit von qualitativen Methoden verstehen
- Wert von Fallstudien einschätzen können
- Chancen und Risiken von Fragebögen und Interviews verstehen

Laboruntersuchung vs. Feldstudie

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



Qualitative Methoden

- Interpretation von verbalem Material
- Fokus auf Erfahrung
- Offene Befragungen
- “Mehr Details als ein Messwert”
- Realismus statt Laborbedingungen

- Keine statistischen Signifikanztests
- Mehr Zeitaufwand
- Schwer vergleichbar

Oberflächliche Abgrenzung

Quantitativ

- "Naturwissenschaftlich"
- Labor
- Erklären
- "Harte Methoden"
- Messen
- Stichprobe
- Zahlen
- Abstraktion

Qualtitativ

- "Geisteswissenschaftlich"
- Feld
- Verstehen
- "Weiche Methoden"
- Beschreiben
- Einzelfall
- Texte, Bilder
- Komplexität

Qualitative und quantitative Methoden

- Kombination qualitativer und quantiativer Methoden typisch

- Programmverständnis:
  - Beobachten von Entwicklern, während sie Fehler in Software beheben
  - Lösungsstrategie von Entwicklern beobachten und abstrahieren
  - Zeit und Qualität von Fehlerbehebung
  - Zusammenhang zwischen Lösungsstrategie und Zeit/Qualität von Fehlerbehebung untersuchen

Fallstudien

Fallstudie

- Detaillierte Untersuchung eines einzigen Beispiels (oder weniger einzelner Beispiele)
- Oft im User-Interface-Bereich

- Beispiele:
  - Beobachten, wie Entwickler mit neuem Tool umgehen
  - Anwenden eines neuen Programmierparadigmas auf bestehende Implementierung

Evaluieren neuer Methoden

- Vom Autor selbst auf eigenem Beispiel
- Vom Autor selbst auf bestehendem Beispiel
- Von Drittem auf eigenem Beispiel
- Von Drittem auf bestehendem Beispiel
- Von neutralem Dritten auf bestehendem Beispiel
- Kontrolliertes Experiment

Fallstudien zur Theoriebildung

- Pilotstudie, Erkundungsexperiment
- In frühen Phasen der Untersuchung
- Zum Bilden von Theorien (die dann z.B. quantitativ untersucht werden)

Fallstudien und Quantitative Methoden

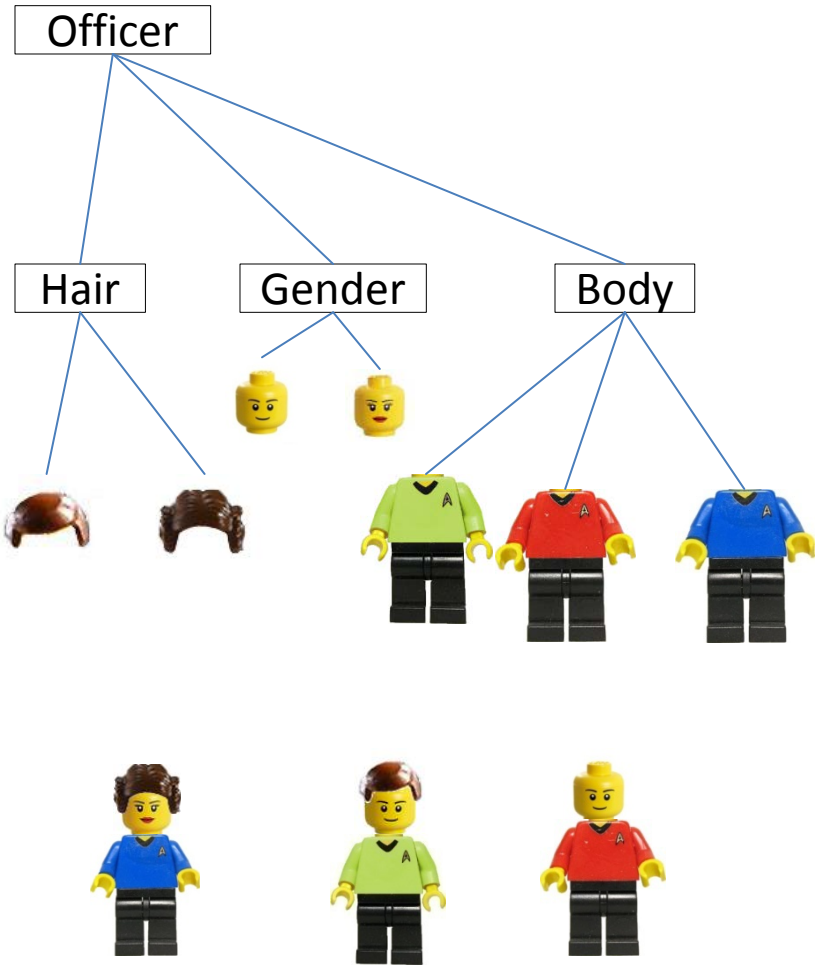
- Innerhalb einer Fallstudie Messungen möglich
  - z.B. Geschwindigkeitsvorteil durch neuen Datenbankindex
  - Inferenzstatistik für Hypothesen über diesen Fall

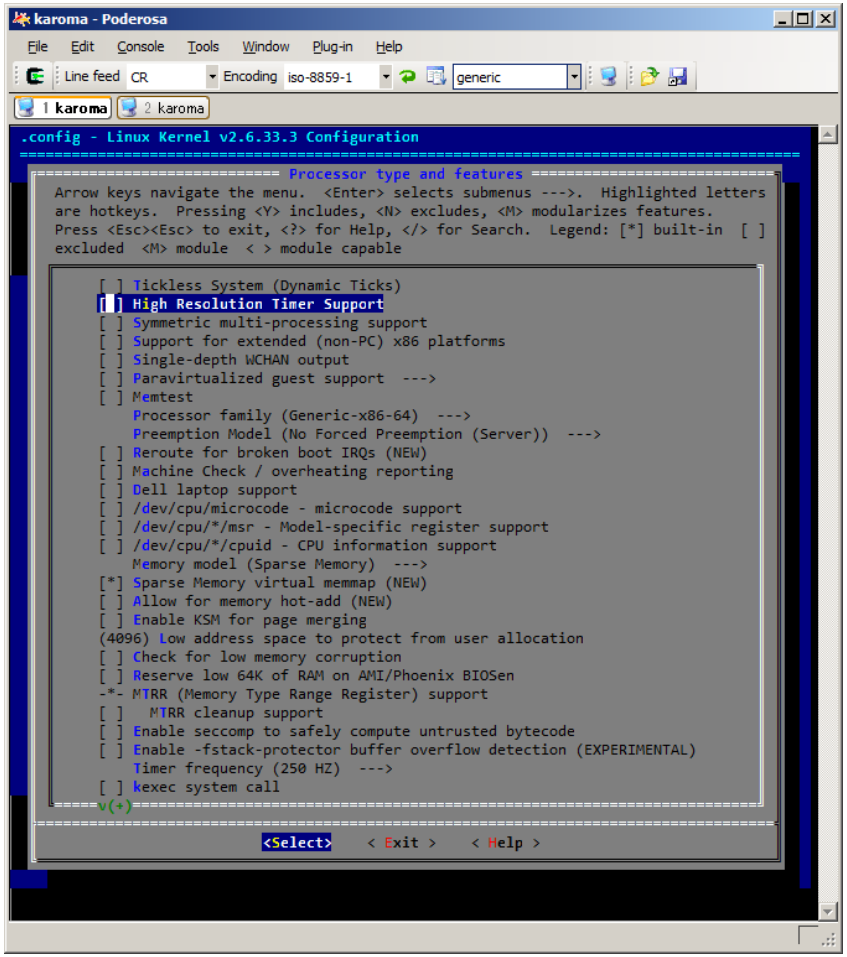
- Kein Schluss auf allgemeine Fälle (externe Validität)

Aufgabe

- Nehmen Sie zu folgenden Aussagen Stellung:
  - Theoretisches Wissen ist wertvoller als praktisches Wissen
  - Man kann nicht von einem Fall verallgemeinern; daher sind Fallstudien sinnlos für Wissenschaft
  - Fallstudien sind gut, um Hypothesen zu generieren, aber zum Überprüfen sind andere Methoden besser

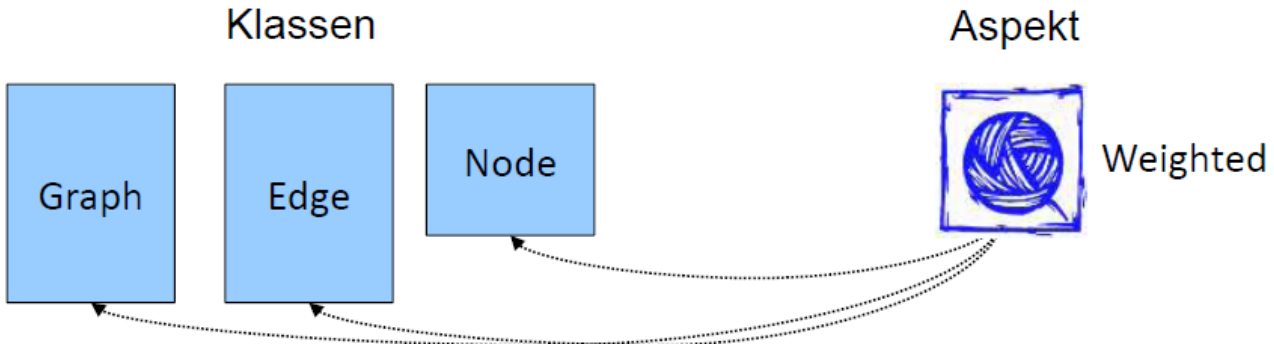
Beispiel: Aspekte für Produktlinien





Aspekt-orientierte Programmierung (AOP)

- Modularisierung von einem querschneidenen Belang in einem Aspekt
- Dieser Aspekt beschreibt die Änderungen dieses Belangs in der restlichen Software



Aspekte für Produktlinien

- Ausgangspunkt
  - Forscher schlugen AOP fuer Produktlinien vor
  - viele Publikationen, wenig Erfahrung
  - keine grossen Beispiele
- Idee
  - Umsetzen einer praktischen AOP Produktlinie
  - Zerlegung eines bestehenden Systems (statt Neuentwicklung)
  - Dadurch Realismus

Kästner, Apel, Don Batory. **A Case Study Implementing Features Using AspectJ. In SPLC,** pages 223-232. 2007.

Auswahl der Fallstudie

- Ein einziges Projekt: **Berkeley DB Java Edition**
- Eingebetette Datenbank
- Wohlbekannte Domäne
- Realistische Größe (ca. 84K Codezeilen, 300 Klassen), aber nicht zu gross
- Realistisch als Produktlinie benutzbar (eingebetette Systeme)

Beobachtungen

- Neue Sprachkonstrukte kaum verwendet
- Wenig querschneidende Belange
- Fragilität
- Lesbarkeit und Verständlichkeit
- Diverse Argumente, weitgehend subjektiv

Reflektion

- Für diesen Fall ist AOP ungeeignet
- Nur einziger Fall, aber realistisch
- Keine statistischen Tests oder Vergleiche
- Widerlegt Hypothese, dass Aspekte geeignet sind für Produktlinien
- Teils subjektiv

Aufgabe

- Diskutieren Sie, in wie weit die Ergebnisse der Fallstudie nützlich sind
- Was hätten Sie anders gemacht?

Kritik an Fallstudien

- Unkontrolliert und subjektiv -> unzuverlässig
- Tendenz zur Bestätigung bestehender Hypothesen
- Nicht verallgemeinerbar
- Viele Details, schwer zusammenfassbar

Lernen durch Fallstudien

- Betrachten eines Problems im Kontext
- Lernen aus Einzelfällen
  - Regel-Lernen für Einsteigerlevel
  - Experten durch praktische Erfahrung
  - Probleme wirklich verstehen (learning by doing)
- Realistische Details
- Nicht abstrahiert/simplifiziert auf einfache Modelle
- Verhindert “Elfenbeinturm-Forschung”

- Beweis kaum möglich, aber lernen aus Erfahrungen

Fallstudie zum Falsifizieren

- Fallstudie kann eine Hypothese falsifizieren
- Gut gewähltes Beispiel kann reichen ("Wenn schon einfache Beispiele nicht klappen…")

- Beispiel
  - Galileo Schwerkraftexperiment mit Fallbeispiel (Feder vs. Blei) statt Experimentserie
  - AOP für bekannte nichttriviale querschneidende Belange in Datenbanken

Auswahl von Fällen

Auswahl	Begründung
Zufall	Reduziert Voreingenommenheit; eher verallgemeinerbar
Extremer Fall	Ungewöhnlicher Fall; besonders problematisch oder besonders geeignet; Verdeutlicht einen Punkt sehr stark
Maximale Variation	Mehre sehr unterschiedliche Fälle (z.B. drei Fälle die sich durch Größe/Sprache/Erfahrung unterscheiden)
Kritischer Fall	Erlaubt Schlussfolgerungen wie: “Wenn es hier (nicht) klappt, klappt es in allen Fällen (nicht)” z.B. zur Plausibilitätspruefung einer Theorie
Paradigmatisch	Allgemeiner typischer Fall, der von mehreren Forschern wiederverwendet wird; Theorien basieren auf diesem Fall

Auswahl von Fallstudien

- Auswahl von guten Fallstudien erfordert Erfahrung
  - Abhängig vom Zweck
  - Machbarkeit zeigen?
  - Maximales Potential einer Methode aufzeigen?
  - Praktische Anwendbarkeit demonstrieren?
  - Bestehende Meinung widerlegen?
  - Methoden vergleichen?

- Gilt auch für Auswahl von Benchmarks

Fallstudien erfordern Selbstreflektion

- Gefahr der Verfälschung und Manipulation
  - Auswahl von sehr vorteilhaftem (trivialen) Fall
  - "Vergessen" von Problemen
  - Vereinfachende Annahmen
- Protokoll führen, eigene Arbeit kritisch überprüfen
- Erwartungen vor der Fallstudie und Hypothesen transparent machen
- In der Praxis tendieren Fallstudien zum Widerlegen von Hypothesen

Fallstudien zusammenfassen

- Fallstudienbeschreibungen oft lang, subjektiv und anekdotisch
- Oft nicht knapp zusammenfassbar, da reale Fälle komplex sind
- Erfahrungen im Kontext weitergeben
  - Aus Erfahrungen anderer lernen
  - Zusammenfassung nicht immer erwünscht
- Details in Anhang

Fragebögen

Aufgabe

- Entwerfen Sie einen Fragebogen
  1. Wie intiutiv ist die Interaktion mit dem iPad?
  2. Wie zufrieden sind Studierende an der FIM?
- Stellen Sie die Ergebnisse vor

Fragebögen

- In Informatik oft benutzt, aber meist oberflächlich
- Vor Beginn Literatur dazu lesen
- Experten befragen

- Wenn möglich, etablierten Fragebogen benutzen

Beispiel

- Geschlossene Fragen quantitativ auswerten
- Likert-Skala, z.B. 1-5
  - Wie erfahren bist du im Umgang mit folgenden Programmiersprachen?

	sehr unerfahren	unerfahren	mittel	erfahren	sehr erfahren
Java	1	2	3	4	5
C	1	2	3	4	5
Haskell	1	2	3	4	5
Prolog	1	2	3	4	5

Falsche Antworten?

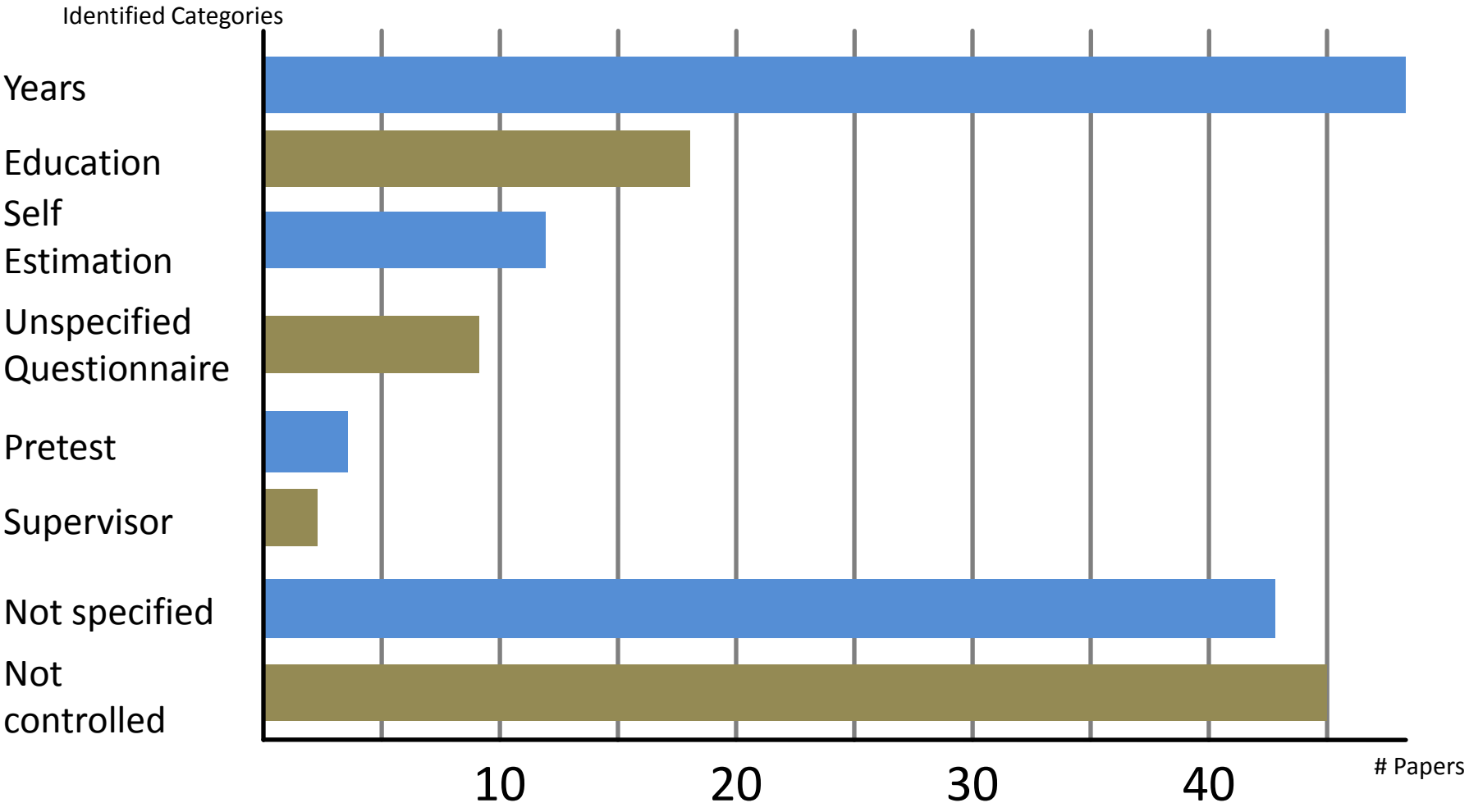
Frage	Antwort
Immatrikulation	1945
Seit wie vielen Jahren programmierst du?	99
Wie viele Programmierkurse hast du belegt	99
Java, C, Haskell, Prolog, Programmierparadigmen	5
Anzahl weiterer Programmiersprachen mit mittlerer Erfahrung	99
In welcher Domäne waren/sind diese Projekte hauptsächlich angesiedelt?	Nirgendwo. Ich habe meine unerträglichen Fähigkeiten vor der Menschheit verborgen weil sonst alle in eine tiefe Depression verfallen wären.

Vorteil von Fragebögen

- Geringe Kosten
- Große Zielgruppen
- Gut zur Ergänzung
- Online durchführbar (aber: missverständliche Fragen?)

- Tools: PROPHET, SurveyMonkey, EFSSurvey

Beispiel: Programmiererfahrung

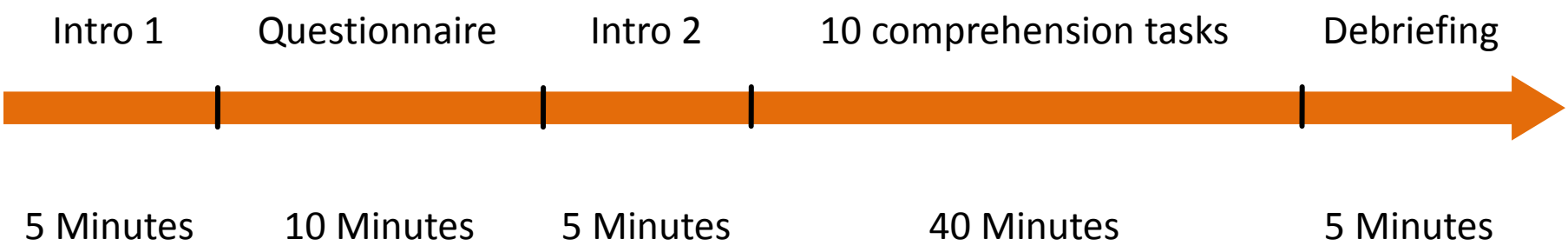


Questionnaire

Years	Self Estimation	Education	Other
y.Prog	s.PE	e.Years	o.Size
y.ProgProf	s.Experts	e.Courses	o.Age
	s.ClassMates		
	s.Java		
	s.C		
	s.Haskell		
	s.Prolog		
	s.NumLanguages		
	s.ObjectOriented		
	s.Imperative		
	s.Functional		
	s.Logical		

Evaluation

Participants: 128 students from three different German universities



Comprehension Tasks

```
1.   public static void main(String[] args) {
2.     int array[] = {14, 5, 7};
3.     for (int counter1 = 0; counter1 < array.length; counter1++)
4.     for (int counter2 = counter1; counter2 > 0; counter2--)
5.       if (array[counter2 - 1] > array[counter2]) {
6.         int variable1 = array[counter2];
7.         array[counter2] = array[counter2 - 1];
8.         array[counter2 - 1] = variable1;
9.       }
10.    for (int counter3 = 0; counter3 < array.length; counter3++)
11.      System.out.println(array[counter3]);
12.  }
```

What does executing this method print?

Ergebnis

- 2 relevante Fragen:
  - Erfahrung mit logischer Programmierung
  - Programmiererfahrung im Vergleich zu Kommillitonen

Nächster Schritt

- Experiment replizieren
- Überprüfen, ob dieselben Fragen extrahiert werden

Literatur

- Bortz & Döring. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4., überarb. Aufl., 2006. Kapitel 4 und 5.

Lernziele

- Einsatzmöglichkeit von qualitativen Methoden verstehen
- Wert von Fallstudien einschätzen können
- Chancen und Risiken von Fragebögen und Interviews verstehen

Hausaufgabe

- 5.6.: Experiment aussuchen und vorbereiten
- 12.6.: Paper lesen: „Five Misunderstandings About Case-Study Research”