

# Empirische Bewertung der Informatik – Klausurvorbereitung

ausgearbeitet von Hannes Restel

zum Sommersemester 2008 (Juli 2008)

## Quellen:

- [Vorlesungsfolien von Prof. Lutz Prechelt \(FU Berlin\)](#)
- tabellarische Übersicht Empirischer Methoden von Christian Kopf

## **Empirie und Empirische Studien**

- Woran man eine gute empirische Studie erkennt → Klare Fragestellung am Beginn? + Klares Ergebnis der Studie am Ende formuliert? + Kann Ergebnis leicht zur Datenanalyse zurückverfolgt werden? + Existiert Verbindung von dem Ergebnis der Datenanalyse zu den Ergebnissen?
- Gültigkeit und Arten von Validitäten
  - Konstruktgültigkeit (*construct validity*)
    - ist die Studie unseren Zielen entsprechend angelegt?
  - Interne Gültigkeit (*internal validity*)
    - „Messe ich genau das, was ich messen möchte?“
    - Zuverlässigkeit (*reliability*) → Ergebnisgleichheit bei Wiederholung gegeben?
  - Externe Gültigkeit (*external validity*)
    - Generalisierbarkeit

## **Umfragen (Surveys)**

- Ansatz: „Frage viele Leute was Du wissen möchtest“
- Schritte: Ziele bestimmen → Zielpopulation auswählen → Umfrageinstrument erstellen → Umfrage durchführen (überwacht, unüberwacht, halbüberwacht) → Daten zusammentragen, prüfen, analysieren → Forschungsfrage beantworten
- Vorteile → kostengünstig, sehr flexibel
- Nachteile → subjektiv, Validierung der Richtigkeit der Antworten ist schwierig, Glaubwürdigkeit schwer zu beweisen
- Gegenstand der Fragen → Querschnittsfragen („snapshot“), Veränderungen über die Zeit („longitudinal“), Retrospektive Fragen (historische Gründe für aktuellen Status)

- Auswahl der Zielpopulation → Wie viele? Welcher Art? Wie erreiche ich die Leute? Wie motiviere ich die Leute zur Teilnahme? Erwartete Anzahl an irrelevanten/ Falschantworten erwarte ich?
- Entwurf des Fragebogens/Umfrageinstruments (*questionnaire*) → (teilweise) Wiederverwendung und/oder Anpassung vorhandener Fragebögen ist hilfreich, minimiere Anzahl der Fragen, Demographische Fragen ans Ende, Freitextfelder ans Ende, für jede Frage bewerten: Zweck/Sinn/Verständlichkeit/Eindeutigkeit dieser Frage?
  - Arten von Fragen → offene Frage (Freitext), geschlossene Fragen (multiple choice, Kategorien, Zahlenwerte)
- Beispiel der Vorlesung → Relevanzbewertung von Aspekten der Informatik-Ausbildung (Lethbridge, 2000)
  - Web-basierte Umfrage → 200 Teilnehmer
  - Ergebnisse → einige Dinge werden gelehrt, welche im Beruf später nicht benötigt werden (z.B. stetige Mathematik) und andere werden im Beruf benötigt und nicht gelehrt (z.B. Software Management, Soft skills, Qualitätssicherung)
    - einige Probleme der internen und externen Gültigkeit

### **Kontrollierte Experimente**

- Beschreibung: „Ändere nur eine Sache und halte alle anderen Sachen konstant und schaue was passiert“
  - Vorteile → einzige Methode zum Beweis kausaler Zusammenhänge, bei guter Durchführung hohe Glaubwürdigkeit, quantitative statistische Datenanalyse anwendbar, Ergebnisse einfach zu interpretieren
  - Nachteile → meist sehr kostspielig, nur für Detailfragen anwendbar, Generalisierung schwierig weshalb Relevanz meist schwierig ist
  - die Kontrolle bezieht sich darauf, dass alle unabhängigen Variablen bekannt und steuerbar sind plus Wiederholbarkeit ist gegeben → Zweck der Kontrolle ist es *Konstanz* herzustellen
    - Bedrohung der Konstanz durch → individuelle Unterschiede, Einfluss äußerer Faktoren bei Langzeitexperimenten, Reifung und Sophistikation der Probanden, Messinstrumente, Ausscheiden von Probanden, Einfluss des Experimentators, Sequenzeffekte (=Reihenfolge-Abhängigkeit der Telexperimente)
    - Herstellung der Kontrolle der unabhängigen Variablen durch Durchschnittsbildung der der Gruppen mittels
      - *Randomisierung* (und im Extremfall *double blind testing*)
      - *Gegenbalancierung* (dieselbe Anzahl an Probanden löst Aufgaben in allen möglichen Reihenfolgen → reduziere Sequenzeffekte)
      - *Matching* (gruppiere Probanden nach Eigenschaften und weise Probanden mit gleichen Eigenschaften zufällig unterschiedlichen Gruppen zu)
- Unterschiede gleichen sich (statistisch!) aus

- Beispiele der Vorlesung:
  - Flussdiagramme vs. Pseudocode
    - sehr hohe interne Gültigkeit (*internal validity*) aber äußerst geringe externe Gültigkeit, Glaubwürdigkeit und insbesondere Relevanz (=Generalisierung)
  - Nützlichkeit von Entwurfsmuster-Dokumentation zum Codeverständnis
    - genutzt wurde Gegenbalancierung mit Randomisierung → trotz ungleicher Gruppen dadurch gute interne Gültigkeit gegeben

## **Quasi-Experimente**

- „Quasi-Experimente sind wie kontrollierte Experimente, jedoch ohne vollständige Kontrolle und meist ohne Randomisierung der Teilnehmergruppen“
- Durch das Aufweichen der Kontrolle werden sehr interessante Studien ermöglicht, welche möglicherweise hohe Verallgemeinerbarkeit (*relevance*) besitzen
- Vorteile → können sehr gutes Kosten-Nutzen-Verhältnis bieten, bei guter Durchführung hohe Glaubwürdigkeit erreichbar
- Nachteile → opportunistisch (Durchführung determiniert von Umwelt), manchmal schwer die gute Glaubwürdigkeit zu beweisen, oftmals schwer die Ergebnisse zu generalisieren
- Beispiele der Vorlesung:
  - „Vergleich von Programmiersprachen“ (Lutz Prechelt)
  - „Wirkung von Arbeitsplatzbedingungen“ (Tom DeMarco & Tim Lister)
    - Zweier-Teams aus 35 Organisationen müssen Aufgabe lösen → Beide Entwickler eines Teams lösen dieselbe Aufgabe für sich → erstaunliche Korrelation der benötigten Arbeitszeit zwischen den beiden Entwicklern des Teams → Die liegt an der Organisation und ist zurückzuführen auf die Arbeitsbedingungen → Hohe Indizien dafür, dass die Arbeitsbedingungen die Arbeitsleistung beeinflussen
- Besonderheit von Prechelts Boxplots: die Whiskers („Schnurrhaare“) umfassen die linken und rechten 10% der Datenmenge
- Generische Methode der Quasi-Experimente:
  - nur eine Variable (*independent variable*) variiert, der Rest ist konstant (wie bei kontrolliertem Experiment)
  - einige Variablen werden beobachtet, woraus sich der Einfluss der *independent variable* ableiten lässt (angegeben in Konfidenz)
  - aber: einige bzw. viele Attribute lassen sich nicht kontrollieren und variieren → z.B. die Zusammensetzung der Kontrollgruppen: die Teilnehmer weisen sich selbst Gruppen zu oder Teilnehmer werden von außen zugewiesen oder Gruppen bestehen bereits → Gruppen können verschieden eingewiesen, verschieden überwacht verschieden ausgestattet usw sein → das Messsystem (*measurement*) ist weniger zuverlässig als beim kontrollierten Experiment
- Generische Auswertung von Quasi-Experimenten:

- siehe: kontrollierte Experimente

## **Benchmarking (Vergleichstests)**

- Ziel: Standardisiertes Messen der Performanz eines Systems oder einer Methode
- Benchmark = Leistungsmaß + Aufgabe(n) + direkter Vergleich mehrerer Ergebnisse
- ein guter Benchmark ist ein mächtiges und kosteneffektives Evaluationswerkzeug
- Vorteile: hohe Glaubwürdigkeit, Relevanz, Zugänglichkeit, Portierbarkeit, Ergebnisse leicht verständlich, Skalierbarkeit, hohe Kontrolle (= "kontrolliertes Experiment in der Dose"), Wiederholbarkeit, Aufwand der Durchführung vergleichsweise gering und evtl automatisierbar
- Nachteile: meist hohe Kosten der Benchmarkaufstellung, Aufgabenfindung und -Zusammenstellung meist schwierig, Überanpassung der Testobjekte an Benchmark möglich → Verfälschung der Ergebnisse
- Ausarbeiten des Benchmarks ist komplex und teuer, das Durchführen und auswerten jedoch vergleichsweise günstig
- Alles kann gemessen und verglichen werden: Hardware, Software, Methoden → deshalb können selbst Messungen eine subjektive Komponente besitzen und deshalb die Glaubwürdigkeit verringern (ins Besondere beim Messen menschlicher Eigenschaften)
- Bausteine des Benchmark
  - Leistungsmaß (*performance measure*) → oftmals quantitativ und automatisiert, kann aber auch qualitativ und manuell sein
  - Aufgabe(n) (*task(s)*) → ein oder mehrere Aufgaben, welche vom Versuchsobjekt durchgeführt werden müssen; Aufgabe(n) sollte(n) relevant und repräsentativ sein
  - Vergleich (*comparison*) → sammelt und misst die Ergebnisse; entweder die Ergebnisse eines Testobjekts anhand eines Referenzwerts oder mehrere Testobjekte untereinander
- Generisches Vorgehen bei Erstellung eines Benchmarks (von Sim, Easterbrook, Holt):
  - Einigung auf Leistungsmaß → Einigung auf Benchmark-Ansatz → Benchmark-Inhalte definieren → Benchmark-Vorgehen definieren → Ergebnisformat definieren → Zusammenstellen und Verteilen des Benchmarks → Ergebnisse sammeln, auswerten und veröffentlichen
- Qualitätsattribute: Ein Benchmark sollte zugänglich, preisgünstig, eindeutig, portierbar, skalierbar, lösbar und relevant und glaubwürdig sein.
- Beispiele der Vorlesung:
  - SPEC CPU2000
    - SPEC ([www.spec.org](http://www.spec.org)) ist ein non-profit Konsortium welche diverse Hardware- und Software-Benchmarks entwirft (z.B. CPU, Java, Mail Server, Graphics, Web Services)

- CPU2000 ist Zusammenstellung (= *Package*) mehrerer existierender „real-world“ Programme verschiedener Couleur, welche auf alle Unix- und Windows-Systeme portierbar sind. Programme umfassen viele Programmiersprachen, Anwendungsgebiete und balancieren diverse Aspekte (wie Pipelineing, Cacheing, Memory Performance, Integer Schwerpunkte, FloatingPoint Schwerpunkte)
- die *Performance*, d.h. Das Ergebnis von SPEC CPU2000 wird an einem Referenzwert gemessen, was die Vergleichbarkeit herstellt
  - Performance wird gemessen mit geometrischem Mittel über alle Einzelbenchmarks
- TREC → „Text Retrieval Conference“
  - Vergleich von Text-Retrieval-Werkzeugen
  - Leistungsmaß sind *precision* und *recall*

### **Fallstudien (case studies)**

- Ziel → Beobachte (= erforsche, charakterisiere, validiere) etwas Spezifisches sehr detailliert während es gerade geschieht und beziehe so viele Informationsquellen wie möglich ein, wobei das Interessensobjekt direkt in Kontakt mit dem Fallstudienleiter treten kann. Fallstudien sind Mix aus qualitativer und quantitativer Forschung
- Eine Fallstudie ist eine entlang der Zeitdimension fortwährende detaillierte Beobachtung eines Interessensphänomens in seiner natürlichen Umgebung
- Achtung: in der Informatik wird der Begriff „Fallstudie“ sehr locker verwendet und ist selten eine „formale Fallstudie“
  - wenn z.B. ein kontrolliertes Experiment keine statistisch signifikanten Ergebnisse erzielt wird es einfach Fallstudie genannt
- Vorteile → reichhaltige Ergebnisse, sehr hohe Glaubwürdigkeit möglich
- Nachteile → schwierige Methode da viele Fertigkeiten notwendig, Generalisierung schwierig, Beurteilung der Relevanz schwierig, nur ein Interessensobjekt pro Fallstudie (der „Fall“), geeignete Präsentation des Fallstudienreports schwierig
  - wenig Kontrolle wird vom Empiriker ausgeübt, da mehr meist nicht möglich ist
- keine fixes Ende → meist währt die Fallstudie so lange, bis man mit Beobachtungen zufrieden ist → Beobachtungen deshalb sehr breit und detailliert
- nutze alle möglichen Datenquellen des Interessensobjekts → Dokumentation, Archive, Interviews, direkte Beobachtung, Physische Artefakte (selten bei Informatik)
- nutze Triangulation → beobachte einen Sachverhalt mehrfach aus unterschiedlichen Blickwinkeln (mehrere Beobachter, mehrere Datensätze, ..)
- Bausteine einer Fallstudie → Forschungsfrage, Behauptungen (können fehlen), Forschungsobjekt(e) (teilweise schwierig zu identifizieren), Analysemethode
- im Anschluss an Beobachtung werden Erklärungen (Warum-Fragen) und/oder

Beschreibungen (Wie-Fragen) produziert und daraus Schlüsse gezogen (→ Beantwortung der Forschungsfrage)

- Fa Fallstudien stets Einzelbeobachtungen sind, ist Generalisierung schwer → nutze *Analytische Generalisierung* → z.B. erhebe mehrere Fallstudien, deckt sich Ergebnis der Fallstudie mit vorhandener Theorie?
- Tipps und Hinweise → halte Daten in einer wohlgeordneten Datenbank, erzeuge Beweiskette von Fragen zu Beobachtungen zu Schlussfolgerungen
- Schwierigkeit des Aufbaus der Ausarbeitung → top-down Fallbeschreibung oder bottom-up Analysebeschreibung?, bei multi-Fall Fallstudien ein Kapitel pro Fallstudie oder funktionale Zusammenhänge pro Kapitel?, chronologisch?, Frage-und-Antwort-Format?, Spannungsbogen?
- Beispiele der Vorlesung:
  - Einarbeitung eines neuen „Immigranten“ in eine Softwareteam
    - Ansatz: Explorative multi-Fall Fallstudie → d.h. jeder Immigrant wurde mehrmals zu unterschiedlichen Zeitpunkten befragt und versucht Übereinstimmungen zu finden
  - Anforderungs-Inspektionen bei DaimlerChrysler
    - Unkonventioneller Ansatz, da Inspektoren aktiv mit Spezifikation arbeiten mussten: z.B. UML-Diagramme erzeugen
    - im Vorfeld aufgestellte Hypothesen sollten durch Fallstudie bestätigt/widerlegt werden
    - Ergebnisse u.a. → Vorbereitungszeit und Anzahl ermittelter Defekte korrelieren stark
    - Kritik → keine Beobachtung entlang der Zeitdimension, nur quantitativ, kein Fokus auf Warum- und Wie-Fragen

## **Andere Methoden**

- Arten von Methoden
  - beobachtend
    - Fallstudie (ein Beobachtungsobjekt), Feldstudie (Menge von Beobachtungsobjekten), Umfrage(?)
  - historisch
    - Literaturstudie, Altdatenanalyse, Umfrage
  - kontrolliert
    - Kontrolliertes Experiment (eher Prozess-basiert), Quasi-Experiment (eher Prozess-basiert), Benchmarks (eher Produkt-basiert), Simulation (sowohl Prozess- als auch Produkt-basiert)
- Simulation

- Ansatz: Modellbildung eines Prozesses oder Systems → Implementierung des Modells → Parameter setzen und Modell ausführen; Parameter variieren → Änderungen der Variablen beobachten
- Mit Simulation können Systeme untersucht werden, welche noch gar nicht existieren oder deren Manipulation unmöglich ist (z.B. Ausfall des Internets eines ganzen Landes) → „was-wäre-wenn-Spiele“ möglich
- Vorteile: Simulation ermöglicht volle Kontrolle; Komplexe Systeme können zu geringen Kosten „beobachtet“ werden
- Nachteile: Validierung des Modells kann sehr schwierig sein (z.B. Validierung von Klimaentwicklung-Modelle sogar unmöglich)
- Beispiel: Skalierung von Peer2Peer-Netzen
  - Messe Erfolgsrate, hop counts und Verzögerung in P2P-Netzen, welche mit verschiedenen Techniken und Optimierungen arbeiten, wobei jeweils unterschiedliche Nutzerzahlen und Dateiverarbeitungen zu Grunde liegen
- Altdaten-Analyse (*legacy data*)
  - Ansatz: Analysieren existierender Datensätze → untersuche neue Fragestellungen oder aggregiere vorhandene Daten
  - Vorteile → große Datenmengen können mit geringem Aufwand genutzt werden, nachträglich gefundene Fragestellungen können beantwortet werden
  - Nachteile → ggf. sind zusätzliche Daten nicht zu beschaffen, da sie nicht existieren (z.B. Alter der Teilnehmer einer Umfrage, wenn das Alter jedoch nicht erfasst wurde)
  - Beispiel: Studie über Codeverfall bei einem komplexen Software-System
    - Loess-Kurven, Nadeldiagramme und SeeSoft-Software wurden in Datenanalyse verwendet
- Literaturstudie
  - Ansatz: Durchsicht und Analyse bereits veröffentlichter empirischer Studien eines ähnlichen Themas → Schlussfolgerungen auf Grund der Vereinigung der Daten oder Ergebnisse dieser Studien
  - Vorteile → verhältnismäßig geringer Aufwand, kostengünstig, vereinigte Sicht mehrerer Studien auf ein Forschungsobjekt kann Glaubwürdigkeit und/oder Relevanz erhöhen
  - Nachteile und Probleme → Studien enthalten oftmals benötigte Teilinformationen nicht, oft existieren nur sehr wenige Studien zum selben Thema, „uninteressante“ oder ergebnisschwache Studien werden oft nicht veröffentlicht → Bild der empirischen Bewertung eines Themas wird verzerrt
  - Beispiel: Effizienz von Reviews

## **Datenanalyse (data analysis)**

- Was ist Datenanalyse → Prozess der Umwandlung von Rohdaten in Daten welche direkte Schlussfolgerungen ermöglichen
- Ziele der Datenanalyse → Treffe gültige Aussagen über die Grundgesamtheit (=Population) auf Basis einer Stichprobe (*sample*), entweder auf Grundlage der Stichprobe allein (*frequentist aproach*) oder durch Hinzunahme von apropi-Annahmen (*Bayes Ansatz*)
- Schritte der Datenanalyse: Daten verfügbar machen (sammeln, umformatieren, pre processing) → Daten validieren (finden und korrigieren von Lücken, Ausreißern) → Daten erkunden → Daten analysieren (messen, modellieren, vergleichen)
- Datenanalyse dient dem Erreichen der wichtigsten Qualitätsziele Glaubwürdigkeit und Relevanz, kann sich aber nur auf die Glaubwürdigkeit beziehen
- Die Datenanalyse muss → korrekt, illustrativ (d.h. einfach verständlich: Grafik, Grafik, Grafik!) und informativ (d.h. Datenanalyse muss neutral der Beantwortung der Forschungsfrage dienen) sein, um Glaubwürdigkeit zu erzielen.
- Probleme der Datenanalyse:
  - nur ein Aspekt der Daten zu betrachten ist nicht ausreichend, jedoch die Gesamtheit zu betrachten ist schwierig
  - häufig sind Tendenzen beobachtbar, diesen jedoch einen semantischen Inhalt zu geben uns schwierig
- Mögliche Aufgaben der Datenanalyse: Erkunden, Messen, Modellieren zur Vorhersage, Modellieren zum Verständnis, Vergleichen
- Erkunden
  - wenn man nicht weiss, was man von den Daten erwartet
  - Überblick über die Daten bekommen und Strukturen entdecken
  - Ziel: Hypothesen für die spätere Erforschung aufstellen; Besonderheiten, Artefakte, Probleme in den Datensätzen finden (z.B. Ausreisser)
- Messen
  - z.B. „Zähle die Anzahl der Defekte in einem Entwurfsdokument“
  - auch dies kann schwierig und uneindeutig sein, wegen → Stochastischen Fehlern (Messfehlern, „verschiedene Menschen messen Länge eines Tisches unterschiedlich“), systematischen Messfehlern, Korruption einzelner Datenpunkte
  - Ziel: Werte einer Messung ermitteln; Messfehler und ihre Größe bestimmen
- Modellieren zur Erklärung
  - Ziel: Beschreibe den Mechanismus, welcher die Daten generiert hat → Aufstellen von Regeln, um diese Daten auch künstlich produzieren zu können; dient zur Aufstellung von Theorien
- Modellieren zur Vorhersage
  - Ziel: Aufstellen von Vorhersagemodellen → Maschinelles Lernen (z.B. Neuronale

Netze)

- Vergleichen
  - Ziel: Vergleiche Datenpunkte oder Datensätze untereinander; Stichpunkte hier sind *Signifikanztests* und *Konfidenzintervalle*
- Tipps zur Datenvalidierung + Datenerkundung + Datenanalyse + Schlussfolgerungen:
  - immer vorsichtig, skeptisch sein!
  - Redundanzen sind gut
  - Manuell eingegebene Daten immer doppelt kontrollieren
  - Mißtraue Daten, welcher in direkter Umgebung von Fehlerhaften Daten liegen
  - zweifle erwartete/unerwartete Regelmäßigkeiten stets an
  - Grafik, Grafik, Grafik!
  - Überblick über den Datensatz bekommen (`nrow`, `names`, `sapply`)
  - einzelne Variablen beobachten (`summary`, `sort`, `table`, `plot`, `boxplot`, `barplot`)
  - Paare von Variablen beobachten (`ScatterPlotMatrix`, `plot(df$columnA ~ df$columnB)`)
  - apriori erstellte Erwartungen überprüfen (vergleichende Boxplots)
  - ziehe Außenstehende hinzu und lasse z.B. deine Schlüsse von diesen überprüfen
  - Verwende nur diejenigen Techniken, welche Du selbst verstehst und beherrscht
  - Glaubwürdigkeit, Gültigkeit, Illustration sind allesamt wichtiger als Präzision!
  - Hole dir professionelle Hilfe, soweit möglich
  - erstelle diverse Checklisten und arbeite diese ab!
  - Sind alle Schlussfolgerungen durch die Daten eindeutig abgedeckt?
  - Können alle Schlussfolgerungen bis hin zu den Daten einfach rückverfolgbar?
  - Traust Du selbst Deinen Schlussfolgerungen?

### **Techniken der Datenanalyse und Statistik**

- Zur Datenanalyse werden verschiedene Schätzungen der Variable(n) benötigt:
  - geschätzter erwarteter Wert der Variablen → (beschnittener) Mittelwert, Median, Modus (=häufigster Wert), ..
  - geschätzte erwartete Abweichung der Variablen → Varianz, Standardabweichung (`sqrt(var(x))`), mittlere Absolute Abweichung (`mad(x, constant=1)`), Quartile, Quantile-Bereiche, ..
    - Normalverteilung (Gauß) → 68%/95%/99.7% der Datenpunkte liegen innerhalb der 1/2/3fachen Standardabweichung
  - geschätzte Fehlergröße der Variablen → Standardfehler der Abweichung

$(\sqrt{\text{var}(x) / \text{length}(x)})$ ), Konfidenzintervalle, alpha/beta-Fehler, ...

- Bootstrapping für Fehlerschätzer → mehrfache Ziehungen mit Zurücklegen aus der Stichprobe incl. statistischer Auswertungen dieser Meta-Stichproben, um die Grundgesamtheit zu simulieren und somit exaktere Aussagen über die Fehler in der Stichprobe treffen zu können
- Vergleich mehrerer Stichproben bzgl. einer Variablen untereinander → wie macht man dies?
  - Ziel → Ermittle Konfidenzintervalle für den Unterschied statistischer Maße der mehreren Proben (z.B. finde Konfidenzintervall für die Aussagekraft des Unterschieds des Mittelwerts zweier Experimentgruppen)
  - Möglichkeit 1 → Unter der Annahme der Normalverteilung der verschiedenen Stichproben nutze den *T-Test* → ermittelt die Wahrscheinlichkeit dafür, dass der beobachtete Unterschied Zufall ist (← Ohne Garantie! Kann auch genau andersherum sein!)
  - Möglichkeit 2 → mit Bootstrapping
- Ermittle Beziehungen zwischen Variablen
  - relevante Fragen → Existiert eine Beziehung? Wie groß ist diese Beziehung?
  - Lässt sich leicht mit der Korrelation herausfinden. Aber Achtung: Annahme hierbei ist, dass wieder Normalverteilungen vorliegen! →  $\text{cor}(x, y)$  oder  $\text{cor}(\log(x, 2), \log(y, 2))$  (Achtung: Logarithmus beeinflusst Korrelationskoeffizienten!) oder  $\text{cor}(\text{rank}(x), \text{rank}(y))$  zum Ignorieren der Skala
    - Achtung Achtung: Vorhandene Korrelation impliziert nicht, dass kausaler Zusammenhang vorliegen muss!!
  - Hinweis: Wenn keine geeignete Funktion zur Interpolation der Datenpunkte gefunden werden kann, so sollte *lokale Regression* mit Hilfe von *LOESS-Kurven* angewandt werden.