

Blockseminar Empirische Forschungsmethoden in der Softwaretechnik

Statistik für EmpirikerInnen (in der Softwaretechnik)

Wissenschaftliche Methode:

1. Beobachten (Schauen, Messen, Befragen)
2. Hypothese entwickeln
3. Experiment durchführen oder Fallstudien auswerten
- 4. Statistische Bestätigung der Hypothese**
- 5. Korrelationen prüfen und Regressionsformel finden (Modellbildung)**

- Nach dem Experiment oder der Umfrage:
Was sagen mir die Daten?
 - Bloßes Draufschaun ist
 - fehleranfällig
 - nicht überzeugend
 - vage
- Statistische Auswertungen der Daten erforderlich
- Wahrscheinlichkeiten der Aussagen hinter den Daten

Ein paar Begriffe aus der W.theorie und Statistik

- Stichprobe X (*sample*)
- Wahrscheinlichkeit p (*probability*)
- Verteilung $P(X)$ (*distribution*)
- Gleichverteilung (*equal distribution*)
- Normalverteilung $N(\mu; \sigma)$ (*normal distribution*)
- (arithmetischer) Mittelwert \bar{X} (*mean*)
- Median X_{Med} (*median*)
- Varianz $\text{Var}(X)$ (*variance*)
- Standardabweichung s_X (*standard deviation*)

1. Hypothesentest

z.B. "Entwurfsmethode E führt zu einer geringeren Zahl von Entwurfsfehler als die alternative Methode A."

Frage: Wie gut unterstützen meine erhobenen Daten diese Hypothese?

2. Korrelationsanalyse

z.B. „Die Projektdauer sinkt linear mit der Anzahl der Mitarbeiter in diesem Projekt.“

Fragen: Wie eng hängen zwei Beobachtungen zusammen?
Mit welcher Formel lässt sich der Zusammenhang beschreiben?

Beispiel

- Hypothese H_1 : „Entwurfsmuster E führt zu einer geringeren Zahl von Entwurfsfehler als die alternative Methode A.“
- $H \leftrightarrow E < A$
- Stichproben (Beobachtungen):
E = (0 4 0 2 0 0 2 1 1 0 0 0 1 3 6 0 3 0 1 0)
A = (0 4 1 2 2 3 5 1 0 0 6 6 1 3 3 1 2 1 4 0)
- Nullhypothese H_0 : „E ist nicht besser als A.“

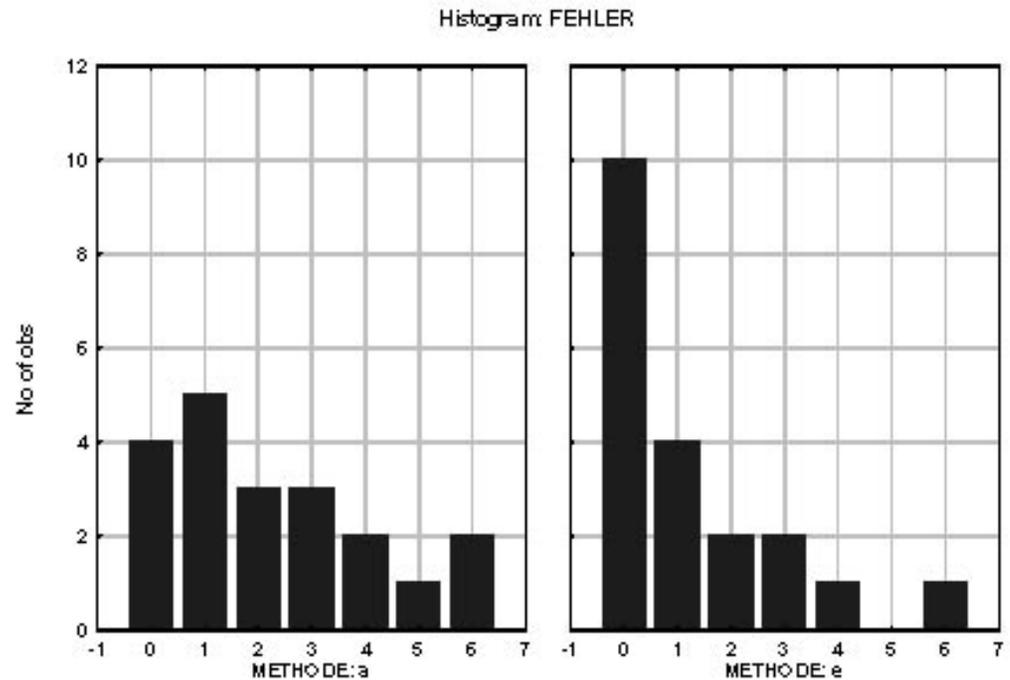
Hypothesentest / Statistischer Signifikanztest

- Berechnen von Fehler 1. Art (α -Fehler):
 H_1 wird als wahr angenommen, obwohl sie nicht gilt
- = Wie wahrscheinlich ist es, dass der beobachtete Effekt zufällig („Nullhypothese“) ist, die Annahme der (Alternativ-)Hypothese also falsch ist?
- Wenn *gering*, dann ist Hypothese unterstützt.
- gering = <50%? 5%? 1‰? (Signifikanzlevel)
- Wenn *nicht gering*, dann keine Aussage.

$$\bar{E} = 1.2 < \bar{A} = 2.25$$

$$E_{\text{Med}} = 0.5 < A_{\text{Med}} = 2.0$$

Sieht gut aus!
 Also alles paletti?



Wie wahrscheinlich ist es trotzdem, dass dies hier ein zufälliger Effekt ist, dass die Hypothese also dennoch falsch ist?

Zwei übliche Testverfahren:

- t-Test:
Prüft Hypothese, dass die Mittelwerte zweier Verteilungen unterschiedlich sind
- Mann-Whitney-U-Test
(Wilcoxon-Rangsummentest):
Prüft Hypothese, dass die Mediane zweier Verteilungen unterschiedlich sind (grob verkürzt!)
- jeweils einseitig oder zweiseitig

t-Test

- Sind die Mittelwerte unterschiedlich?
- Voraussetzungen:
 1. Beide Stichproben sind normalverteilt
 2. Beide Stichproben haben eine ähnliche Varianz (ähnlich \approx maximal um Faktor 2 verschieden)
- Wenn diese nicht gelten, dann „übersieht“ der t-Test evtl. vorhandene Unterschiede
→ wir sind auf der sicheren Seite
- Anwendung am Beispiel ergibt $p = 0.075$
- Könnten die Mittelwerte also eigentlich gleich sein?

Mann-Whitney-U-Test

- Sind die Mediane unterschiedlich?
- Voraussetzungen:
 - Ordinalskala, also: Werte sind geordnet
- *Keine* Anforderungen an die Form der Verteilung (Nicht-Parametrisch)
- Median ist robuster gegenüber Ausreißer. (Typisch in der Softwaretechnik!)
- Anwendung am Beispiel ergibt $p = 0.055$

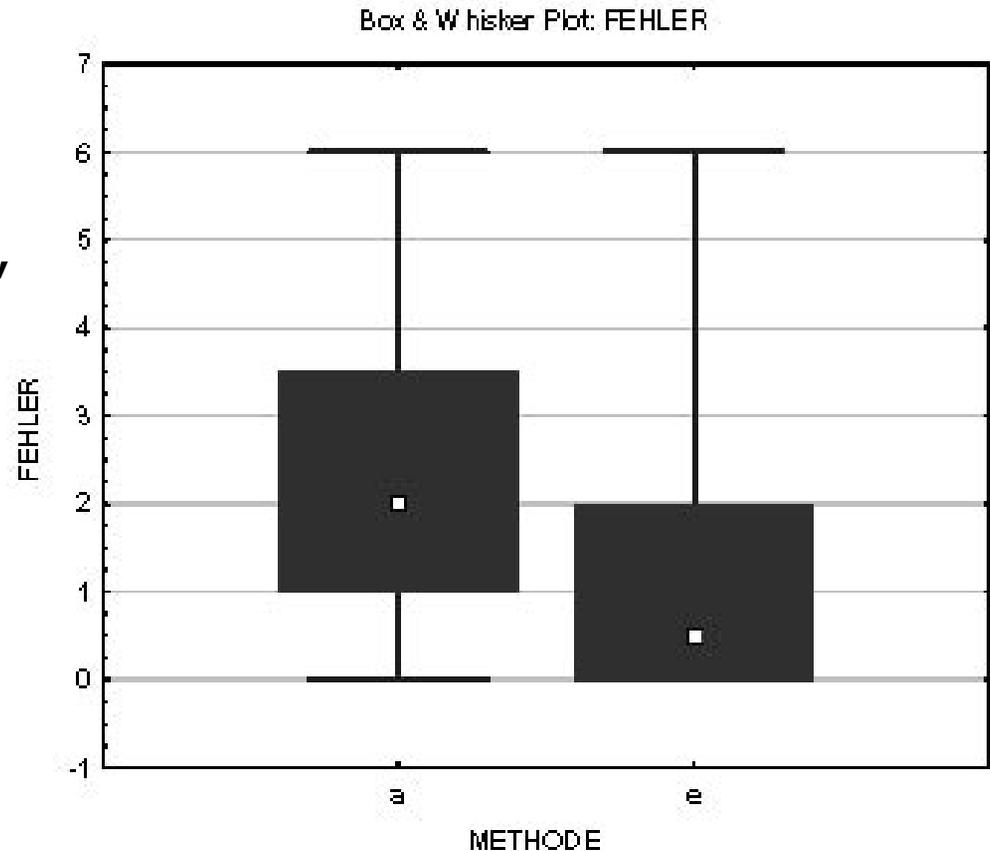
- Reicht das? Was, wenn nicht?

Boxplot

Quantile:
x % der Werte sind kleiner,
100-x % sind größer

~25%, 50% (Median),
~75%-Quantile
eingezeichnet.

Whiskers (~10%/~90%)
verdeutlichen die
Schwankung



Weitere Testmethoden:

- Bootstrap resampling test: Vergleich zweier Mittelwerte ohne Normalverteilungs-Voraussetzung
- Interquartile range test: Vergleich der Unterschiede der Boxen zweier Verteilungen
- F-Test: Vergleich der Varianzen zweier Verteilungen

1. Hypothesentest

z.B. "Entwurfsmethode E führt zu einer geringeren Zahl von Entwurfsfehler als die alternative Methode A."

Frage: Wie gut unterstützen meine erhobenen Daten diese Hypothese?

2. Korrelationsanalyse

z.B. „**Die Projektdauer sinkt linear mit der Anzahl der Mitarbeiter in diesem Projekt.**“

Fragen: Wie eng hängen zwei Beobachtungen zusammen? Mit welcher Formel lässt sich der Zusammenhang beschreiben?

Korrelationsanalyse / Regression

Ziel:

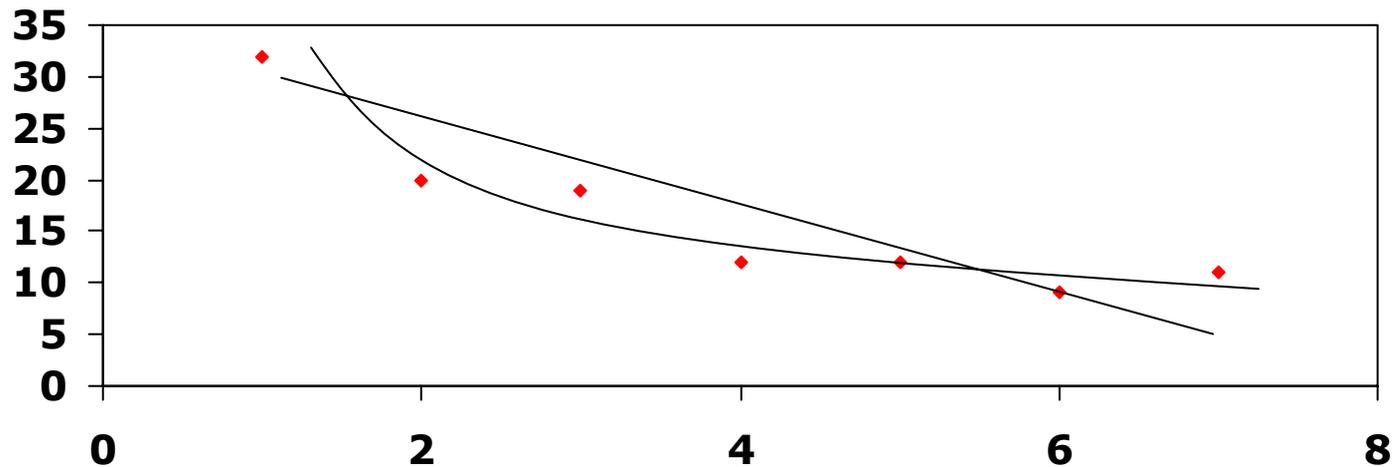
- Den Zusammenhang (Korrelation) zwischen zwei Dimensionen quantifizieren
- Aus den Daten eine Formel für den Zs.hang finden

Beispiel (zwei-dimensional):

- Zusammenhang zwischen Projektdauer und Mitarbeiterzahl
- Beobachtungen: Wochen \times Mitarbeiterzahl
(32,1) (20,2) (19,3) (12,4) (12,5) (9,6) (11,7)

Lineare Korrelation / Regression

- Korrelationskoeffizient $-1 \leq r \leq 1$
 - Pearson: normalverteilt
 - Spearman: nicht-normalverteilt
- Positive Korrelation (linear abhängig): $r > 0$
Negative Korrelation: $r < 0$
Unkorreliert: $r = 0$
- Regressionsgerade z.B. gemäß Minimierung des quadratischen Abstands (least square linear regression)



- Berechnung ist nur einfach, wenn der Zusammenhang
 - linear
 - quadratisch oder
 - logarithmisch ist
- Regressionsformel findet man immer
- Korrelation sollte aber wirklich bestehen
- Begründet auch keinen kausalen Zusammenhang

Vielen Dank!