



Wie man mit Statistik schlecht informiert: Praktiken der Ergebnisdarstellung bei Experimenten im Software Engineering

Marcus Michalsky

michalsk@inf.fu-berlin.de

Institut für Informatik

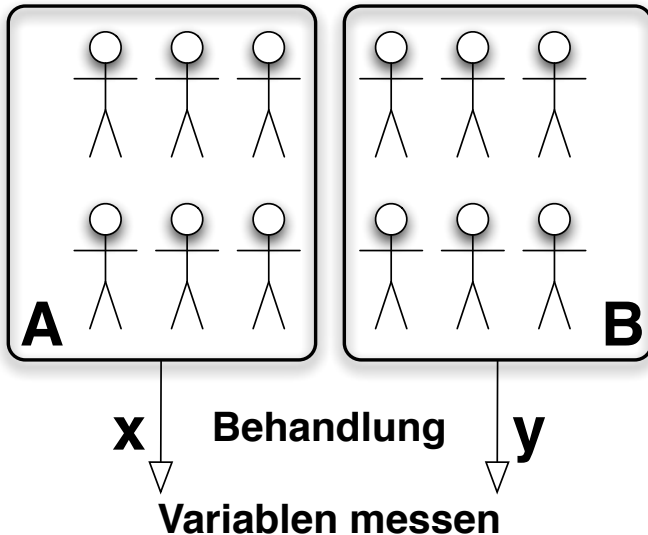
28. Januar 2010



Überblick

Methodik

Ergebnisse



Praktiken der Ergebnisdarstellung von kontrollierten Experimenten unter drei Gesichtspunkten

- ▶ Verwendung beschreibender Statistiken
- ▶ Angabe der Ergebnisse statistischer Auswertungsverfahren
- ▶ Präsenz, Verfügbarkeit und Zustand von Rohdaten

- ▶ Artikel der Zeitschriften:
 - ▶ "Empirical Software Engineering" (EMSE)
 - ▶ "IEEE Transactions on Software Engineering" (TSE)
- ▶ Artikel der Konferenz:
 - ▶ "International Conference on Software Engineering" (ICSE)

Suchverfahren: Titel, "keywords" → "abstract" → Artikel

(a) Experiment

<i>id</i>	...
EMSE09-123	...
ICSE08-456	...
TSE07-789	...
...	...

(b) abh. Variablen

<i>id</i>	<i>var</i>	...
EMSE09-123	worktime	...
EMSE09-123	defectsN	...
...

Tabelle → CSV → Statistiksoftware R

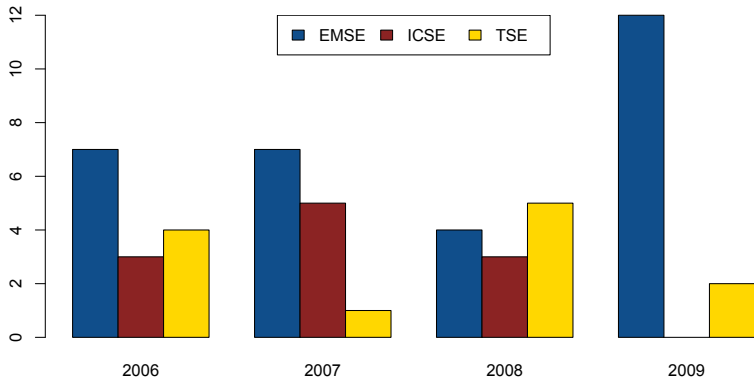


Abbildung: # Experimente je Zeitschrift/Konferenz pro Jahr

- ▶ 36 / 490 (7%) untersuchten Artikeln enthalten 53 Experimente

	%
Mittelwert	71%
Standardabweichung	56%
Tabellen	71%
Boxplots	48%

Tabelle: Angabe von beschreibenden Statistiken

Table 5 Descriptive statistics

Task/method			Time (min)			F-Measure (%)		
			Median	Mean	Std. Dev.	Median	Mean	Std. Dev.
Exp I	All	MT	120.00	125.20	27.22	53.41	52.38	16.06
		AT	75.00	83.75	23.78	65.47	61.36	17.16
	T1	MT	115.00	115.50	22.54	49.47	47.22	17.79
		AT	72.50	80.00	22.24	52.66	53.61	14.62
	T2	MT	122.50	135.00	29.06	59.91	57.52	13.01
		AT	87.50	87.50	25.85	71.93	69.10	16.58
Exp II	All	MT	142.00	138.90	14.58	75.46	71.59	14.50
		AT	87.50	85.25	13.40	75.35	75.39	13.46
	T1	MT	140.00	134.50	19.01	79.44	73.75	14.42
		AT	92.50	92.17	12.36	65.13	64.98	9.76
	T2	MT	142.00	143.30	7.71	66.76	69.43	15.60
		AT	78.50	78.33	11.33	78.07	79.25	7.91

Task: T1 (tracing use cases onto code classes) or T2 (tracing interaction diagrams onto test cases)
 Method: MT (manual tracing) or AT (ADAMS Re-Trace)

Abbildung: Beschreibende Statistiken aus EMSE09-57

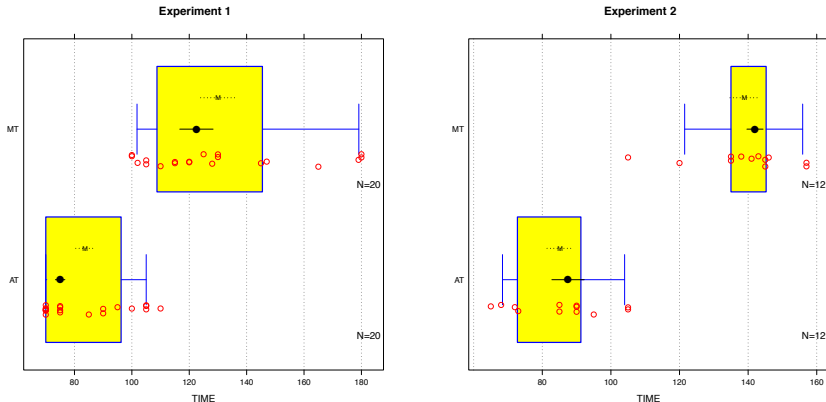


Abbildung: Boxplots für EMSE09-57

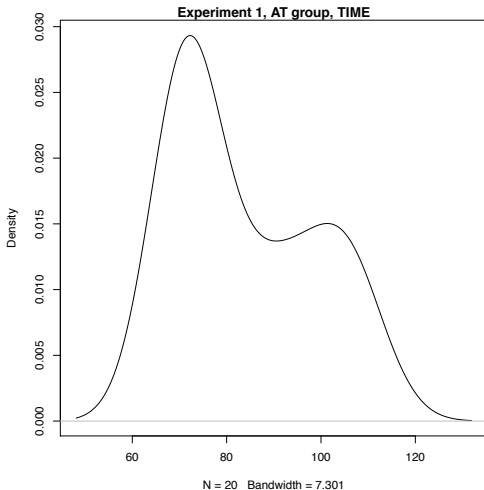
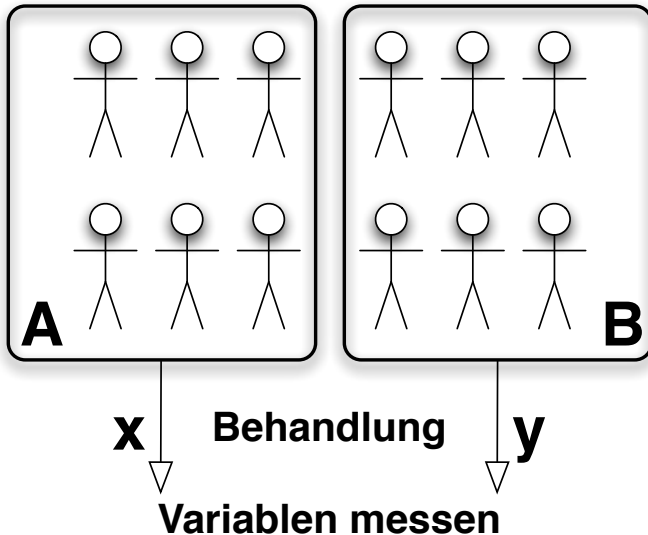


Abbildung: Dichteplot für EMSE09-57



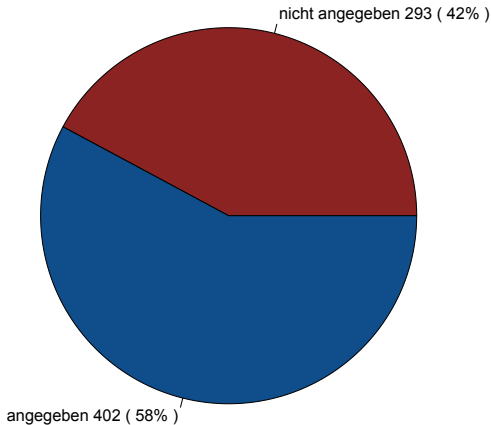


Abbildung: Anteil an berichteten p-Werten

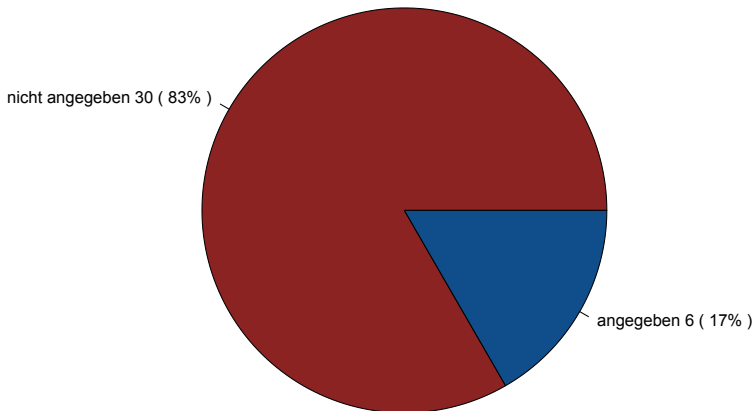


Abbildung: Anteil an berichteten Konfidenzintervallen (je Artikel)

- ▶ Experimente berichten p-Werte von $7.1 \cdot 10^{-05}$ und 0.001
- ▶ → Zeitlicher Unterschied vorhanden, aber wie groß ?
- ▶ berechnete Konfidenzintervalle ergeben

$$CI_{\text{Exp1}} = (-61, -34) \quad CI_{\text{Exp2}} = (-65, -42)$$

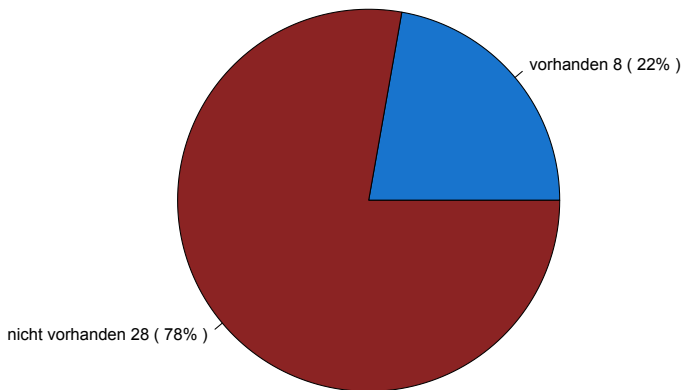


Abbildung: Verfügbarkeit von Rohdaten

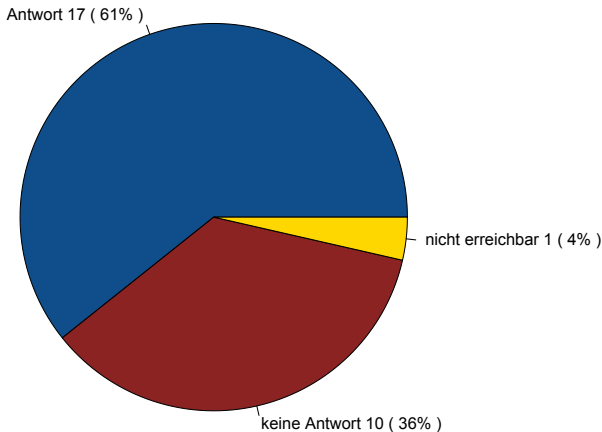


Abbildung: Ergebnis der Rohdatenanfrage

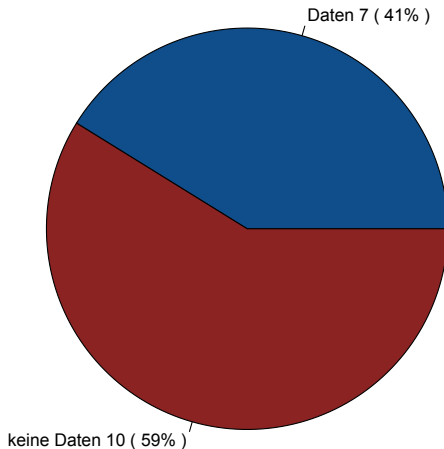
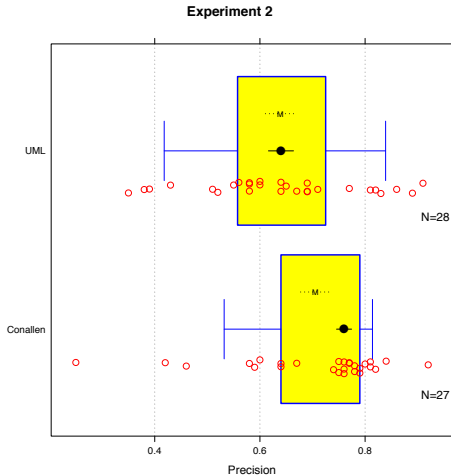


Abbildung: Anteil positiver/negativer Antworten



- ▶ $CI_{out} = (-0.02, 0.16)$
- ▶ $CI = (0.01, 0.15)$

Abbildung: Boxplot für ICSE07-375

- ▶ Artikel mit kontrollierten Experimenten erfasst
- ▶ auf drei Aspekte der Ergebnisdarstellung hin untersucht
- ▶ Ergebnisse
 - ▶ beschreibende Statistiken
 - ▶ Auswertungsverfahren
 - ▶ Rohdaten
- ▶ Anschauungsbeispiele