

Empirische Softwaretechnik

Feldanalyse der Prozentsatzmethode zur Aufwandsschätzung

Dr. Victor Pankratius · Andreas Höfer
Wintersemester 2009/2010

IPD Tichy, Fakultät für Informatik



Feldanalyse zur Prozentsatzmethode

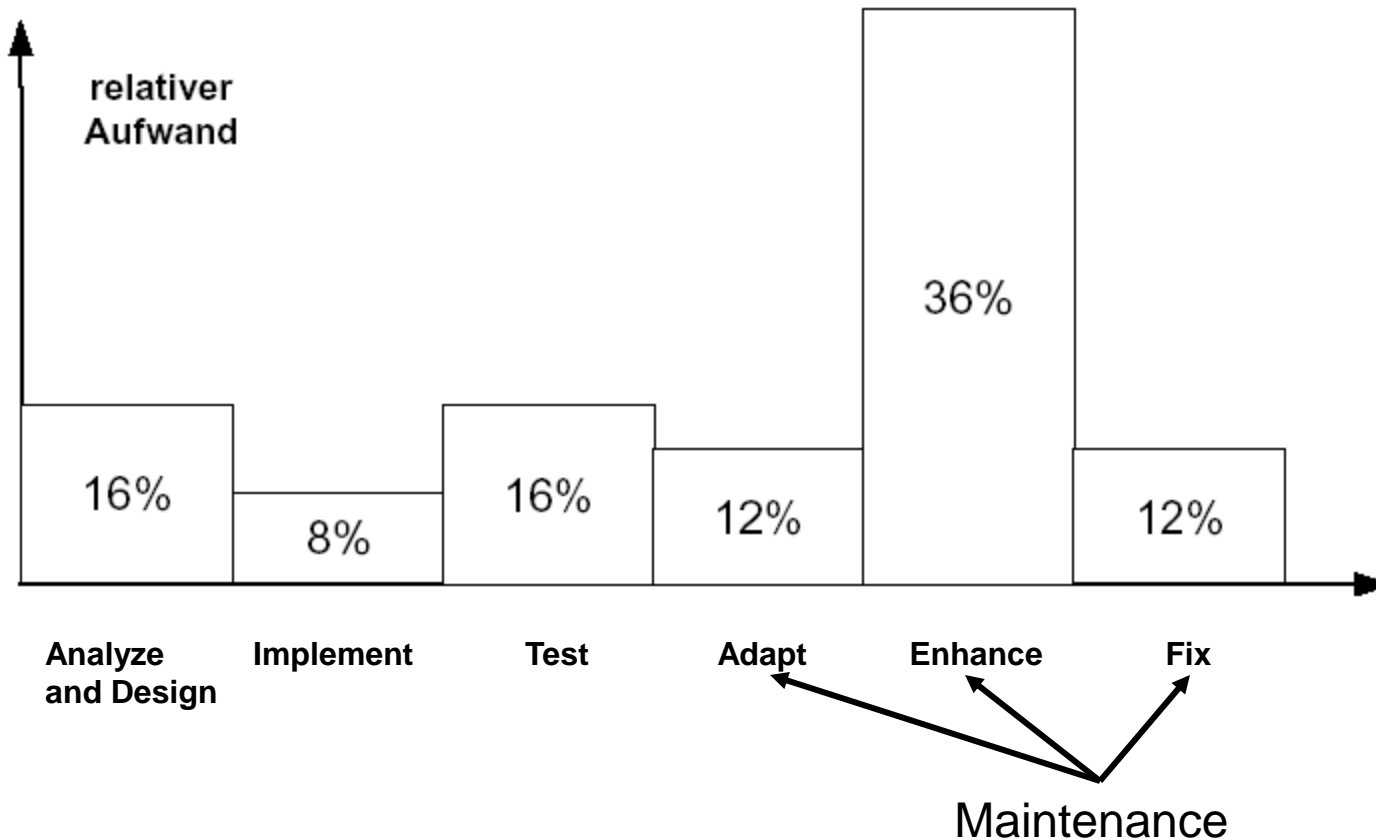
- Was wir hier lernen:
- Softwaretechnik:
 - die einfache Prozentsatzmethode zur Schätzung von Softwareentwicklungsaufwand wird widerlegt
 - Ein Regressionsansatz, um aus dem Aufwand einer Phase den Aufwand für die nächste Phase zu schätzen, wird gezeigt
- Statistik:
 - Regressionsdiagnostik
 - Anwendbar auf ähnliche Fragen bei Aufwandsschätzungen

Prozentsatzmethode

- aus früheren Projekten wird ermittelt, wie sich der Aufwand auf die einzelnen Entwicklungsphasen verteilt hat
- beim neuen Projekt schließt man eine Phase zunächst vollständig ab und ermittelt dann aus dem Ist-Aufwand für diese Phase den Soll-Aufwand für die restlichen Phasen

Prozentsatzmethode (Forts.)

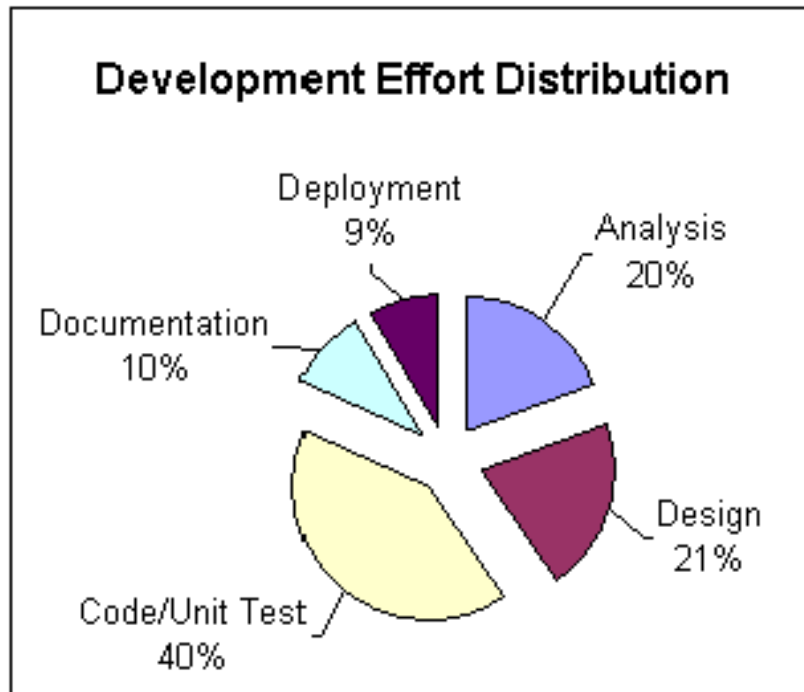
- Beispiel für Prozentsätze [Boehm 1973]:



nach B. Boehm [Datamation 1973]
 entnommen: Fairley (S.10)

Prozentsatzmethode (Forts.)

- Beispiel für Prozentsätze [Rubin 1995]:



nach H. Rubin [1995]

entnommen: Nageswaran [2001]

Prozentsatzmethode (Forts.)

- Beispiel für Schätzung:
 - Gesamtaufwand sei g
 - Entwurf war 320 Personentage (PT) Aufwand
 - $320 \text{ PT} \approx g \times 20 \%$ laut Diagramm
 - Gesamtaufwand $g \approx 320 \div 0,2 = 1600 \text{ PT}$
 - Implementierung mit Testen $\approx g \times 40 \% = 640 \text{ PT}$

Prozentsatzmethode (Forts.)

- die Methode wird immer wieder versucht...
 - Berechnung aus Schätzkurven für Aufwand und Personalverteilung in anderen Modellen
 - SLIM [Putnam 1992 Kap.17 S. 285]
 - CoCoMo [Boehm 1981 Kap.7 S. 98]
 - [Royce 2000]
 - [Jones 2002]
- ...und in Büchern gelehrt (Balzert S. 82)

Wie gut ist diese Methode?

Feldanalyse von MacDonell und Shepperd (2003)

- Using Prior-Phase Effort Records for Re-Estimation During Software Projects
- Ergebnis: die Prozentsatzmethode ist nicht korrekt, aber ein Regressionsansatz kann Verbesserungen bringen
- Int. Symposium on Software Metrics METRICS 9 (2003)
73-86
- Lesen!

Projekte

- innerhalb einer Firma
- 16 vergleichbare Einzelentwicklungen von Hardware-Test-Systemen
- C und C++
- Unix-basiert oder in spezieller Laufzeitumgebung eingebettet (Ausführungsumgebung)

Projekte (Forts.)

- über einen Zeitraum von 18 Monaten in 2000 und 2001
- stabiles Management
- klassische Wasserfall-Entwicklung

Aufwandschätzung

- durch die Manager („expert estimation“)
- vorab für jede Phase
(OE = original effort estimate)
- während der Projekte aktualisiert
(CE = current effort estimate)

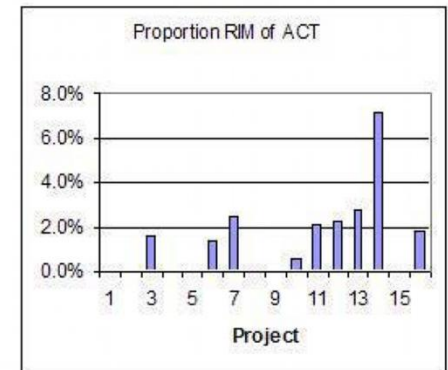
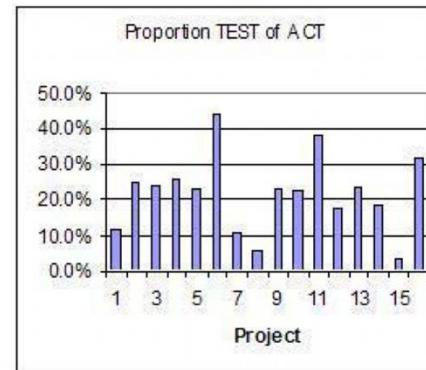
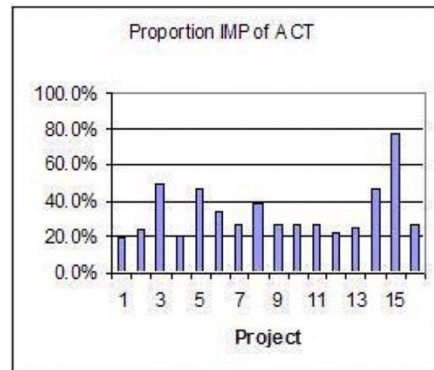
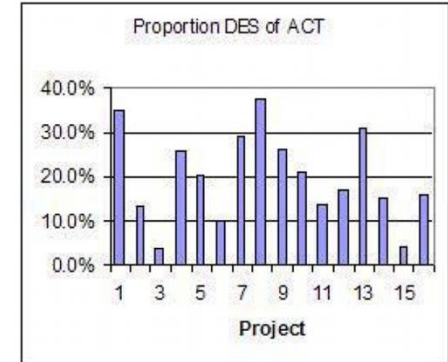
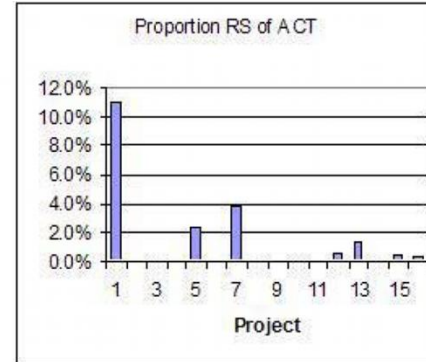
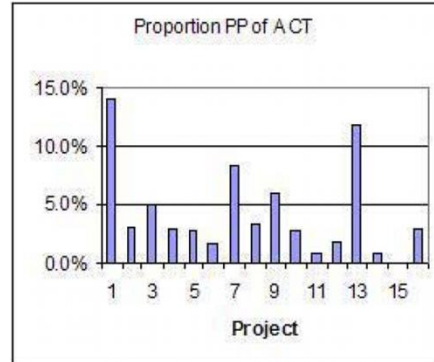
Aufwanderfassung

- laufende Erfassung des Ist-Aufwands (ACT = actual effort)
- zwischen 500 und 7800 Personenstunden tatsächlicher Gesamtaufwand
- systematische Vorgehensweise bei der Datensammlung (CMM-orientiert)

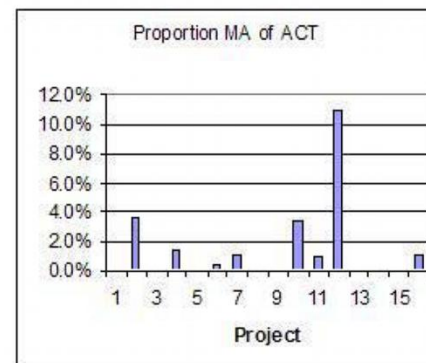
Erster Versuch...

- feste Prozentsatzverteilung, orientiert am Median oder Mittelwert der gesammelten Aufwanddaten

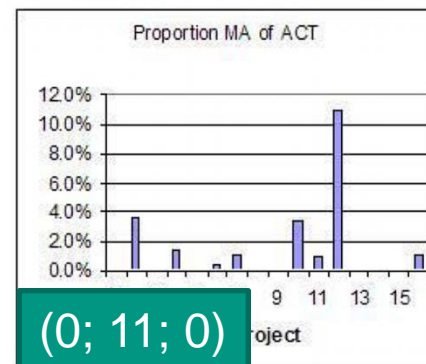
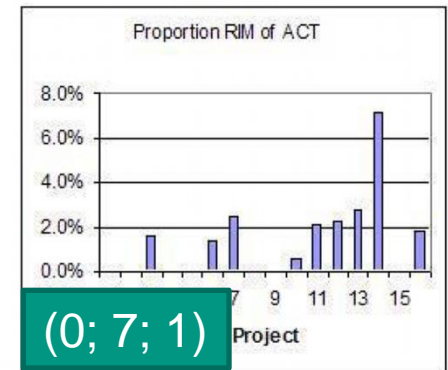
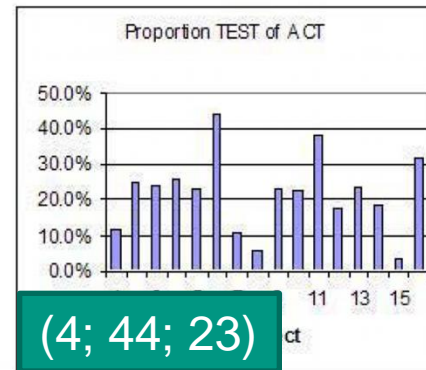
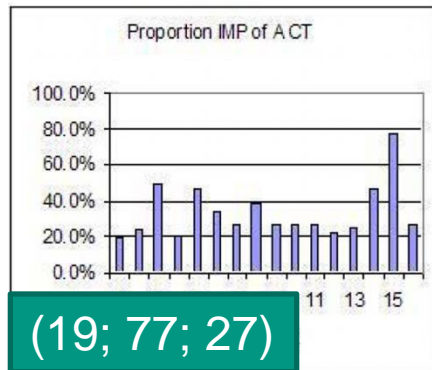
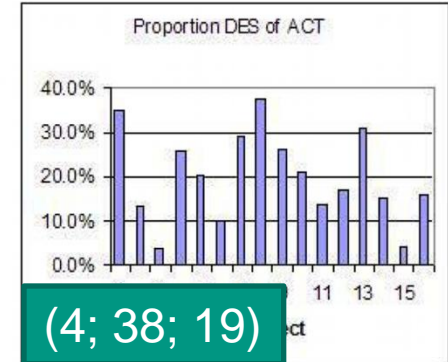
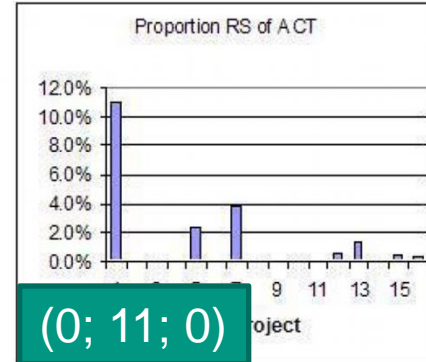
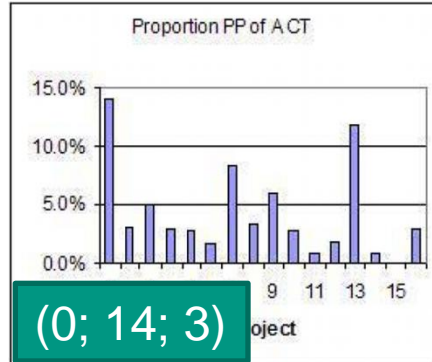
Anteile der Phasen



PP = Projektplanung
 RS = Anforderungsspezifikation
 DES = Entwurf
 IMP = Implementierung
 TEST = Testen
 RIM = Auslieferung und Dokumentation
 MA = Wartung



Anteile der Phasen



(Min.; Max.; Median) [%]

Zwischenergebnis

- Anteile der einzelnen Phasen am Gesamtaufwand schwanken stark
- verlässliche Schätzung des Aufwands mit der Prozentsatz-Methode ist nicht in Sicht!

Zweiter Versuch...

- allgemeine „Einsicht“:
gründliche Vorbereitung von Spezifikation
und Entwurf führt zu geringerem Aufwand für
Implementieren und Testen
- Ansatz:
Korrelationsanalyse der Proportionen der einzelnen
Entwicklungsphasen

Einschränkung der Studie

- RS-Aufwand oft anteilig in PP oder DES enthalten
- RIM-Aufwand oft anteilig in IMP oder TEST enthalten
- unbekannt, wann und wie sich diese Aufwände im einzelnen auf die anderen Phasen verteilt haben

Einschränkung (Forts.)

- Konsequenz: Studie beschränkt sich auf die Verhältnisse zwischen PP, DES, IMP und TEST, ohne RS und RIM (enthält damit die aufwändigsten Phasen)
- außerdem nicht in der Auswertung berücksichtigte Aufwände:
 - training and learning (TL)
 - project management (PM)
- Zahlen erst re-skalieren (ohne RS und RIM Anteile)...
- (Die folgende Analyse basiert auf den Originaldaten, ist nicht so ausführlich im Artikel ausgeführt.)

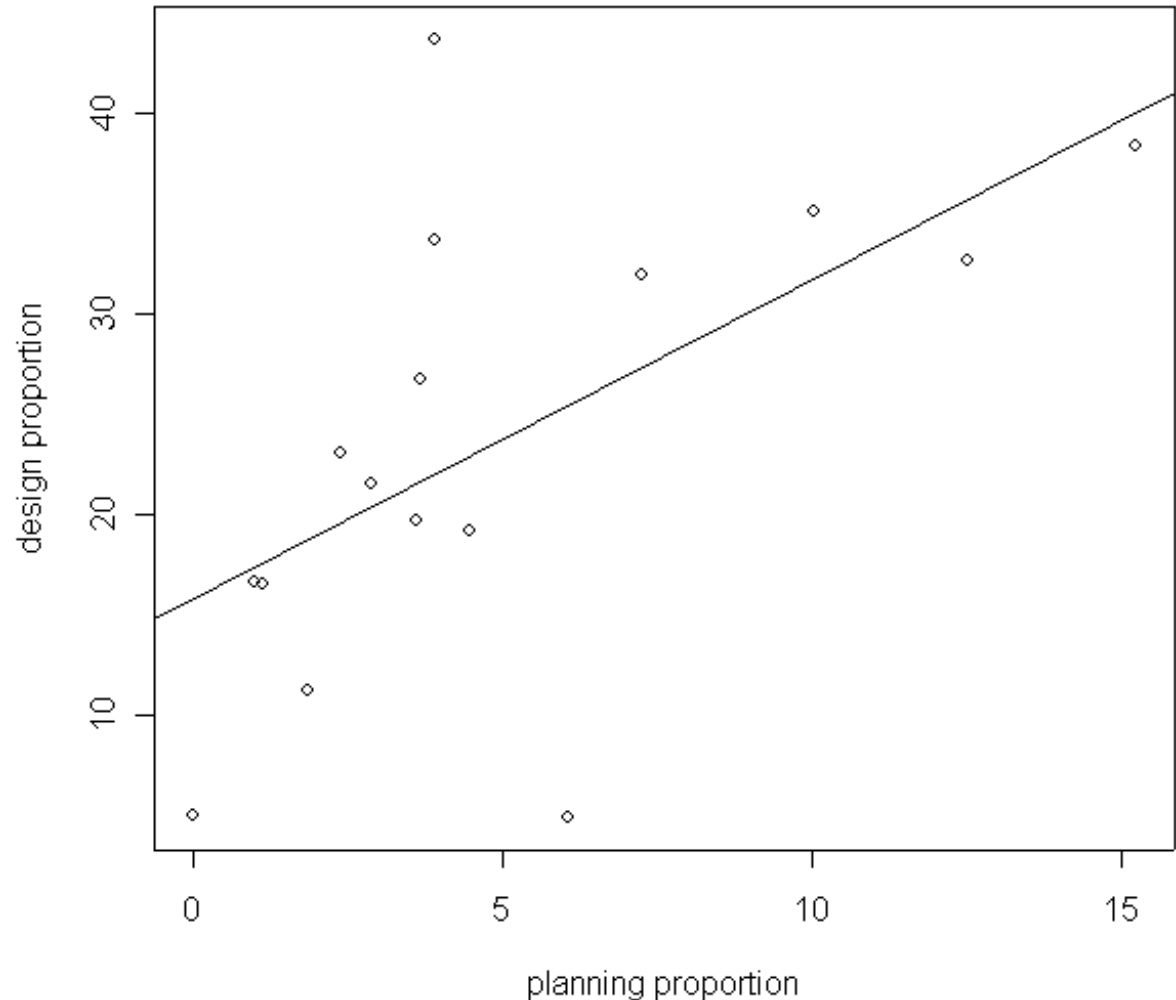
Entwurfsanteil vs. Planungsanteil

τ (Tau) Kendalls Korrelationskoeffizient für Daten, die nicht normalverteilt sind ($-1 \leq \tau \leq 1$).

■ $\tau = 0,5$

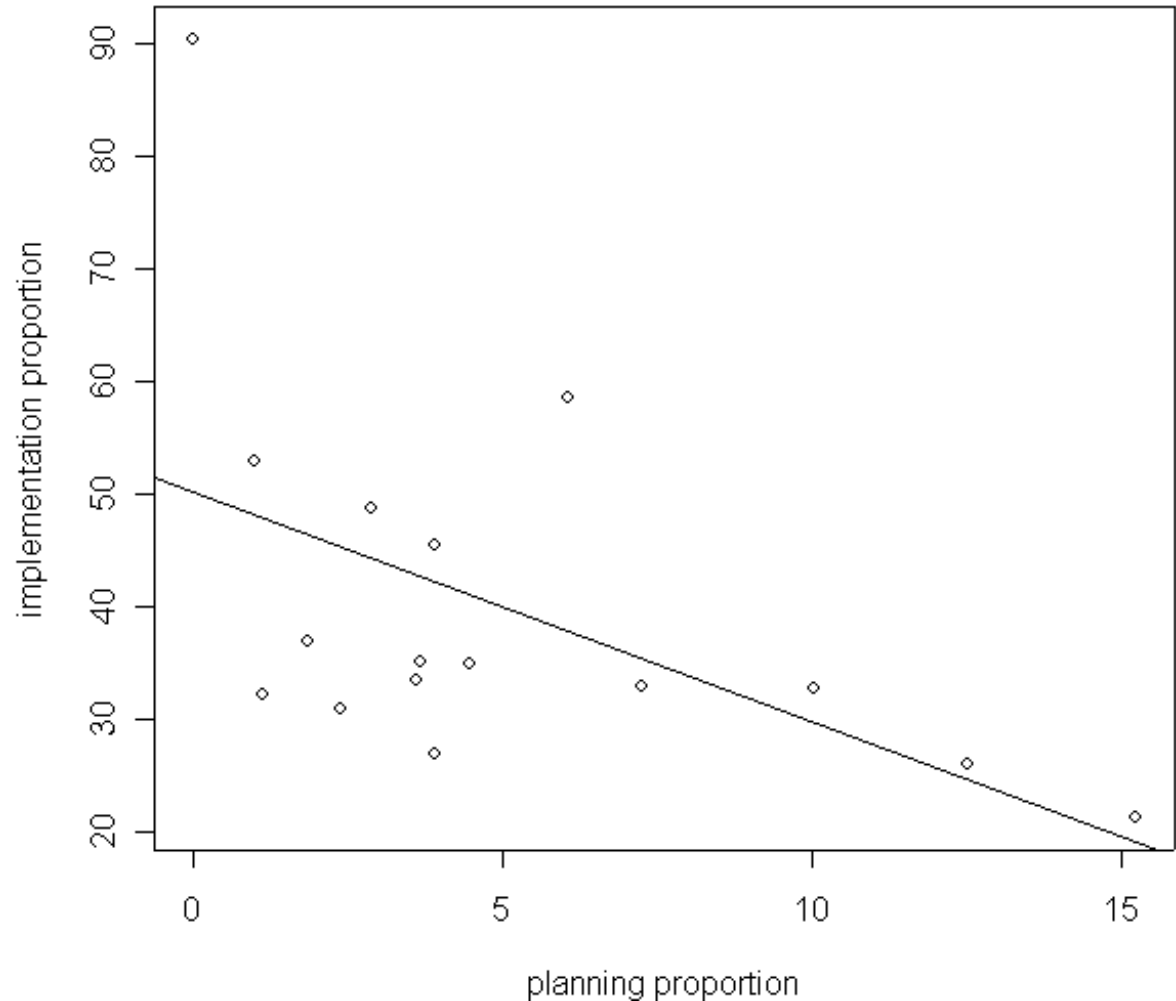
■ $p < 1\%$

p: die Wahrscheinlichkeit, dass X- und Y-Werte der Stichprobe unabhängig sind.



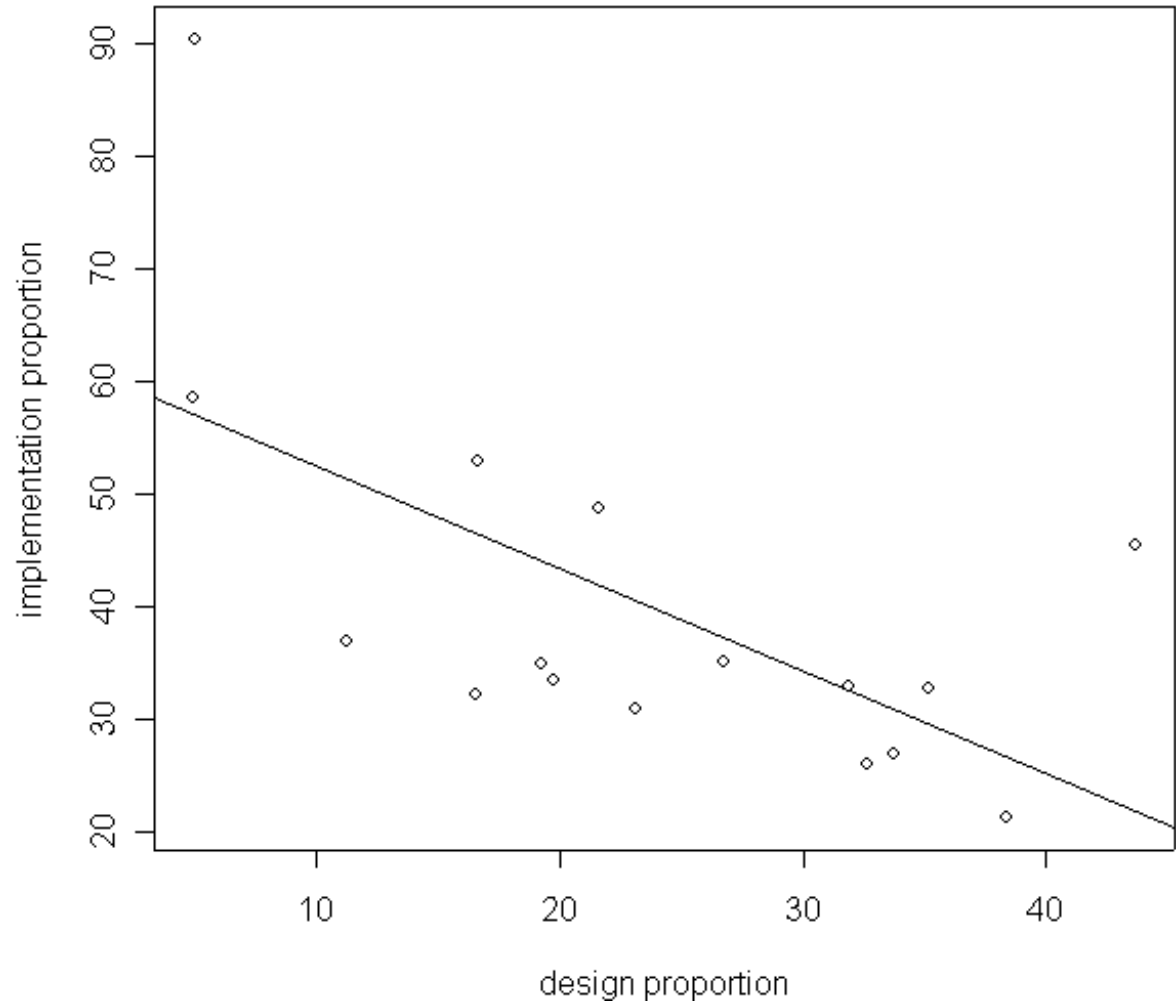
Implementierungsanteil vs. Planungsanteil

- $\tau = -0,42$
- $p = 2 \%$



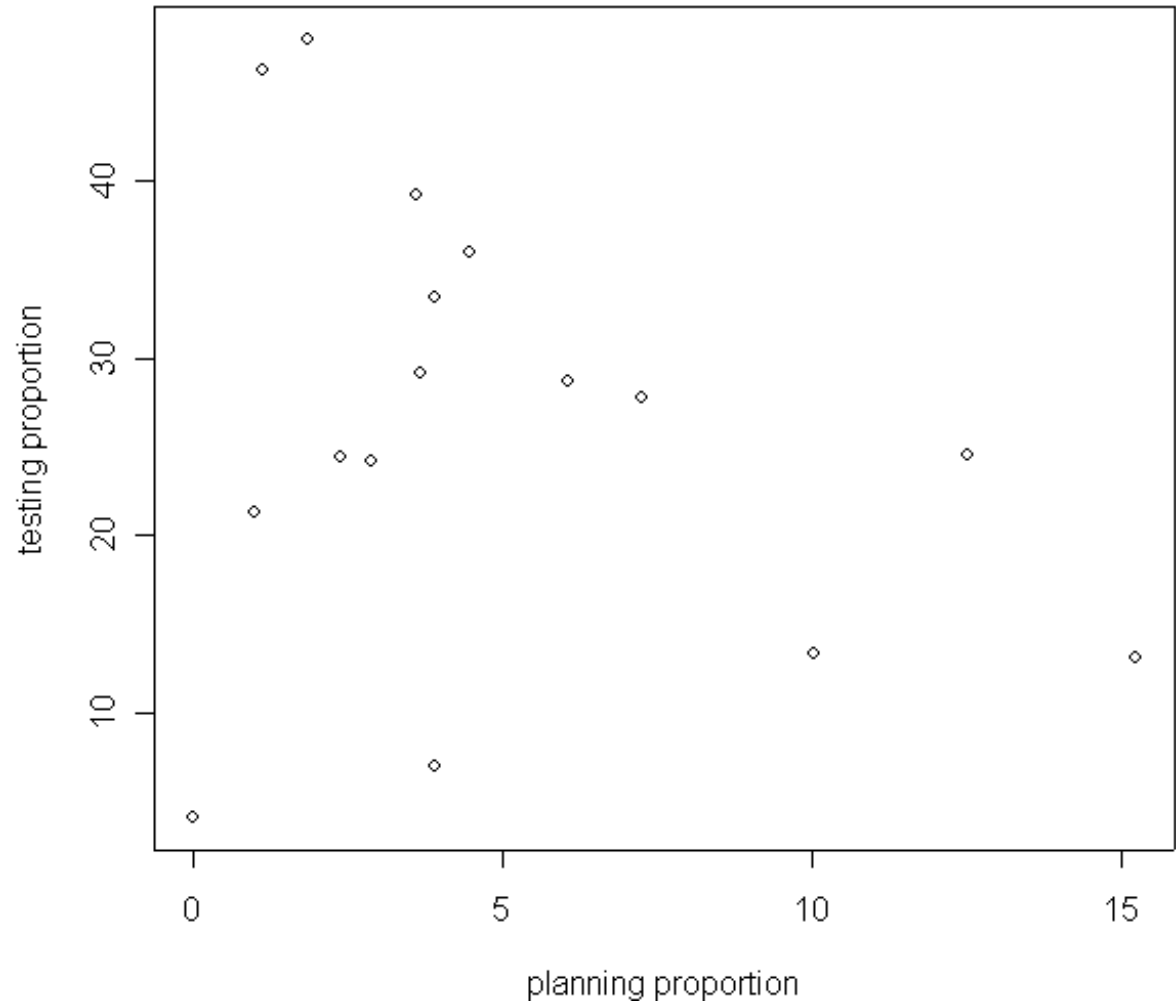
Implementierungsanteil vs. Entwurfsanteil

- $\tau = -0,48$
- $p < 1 \%$



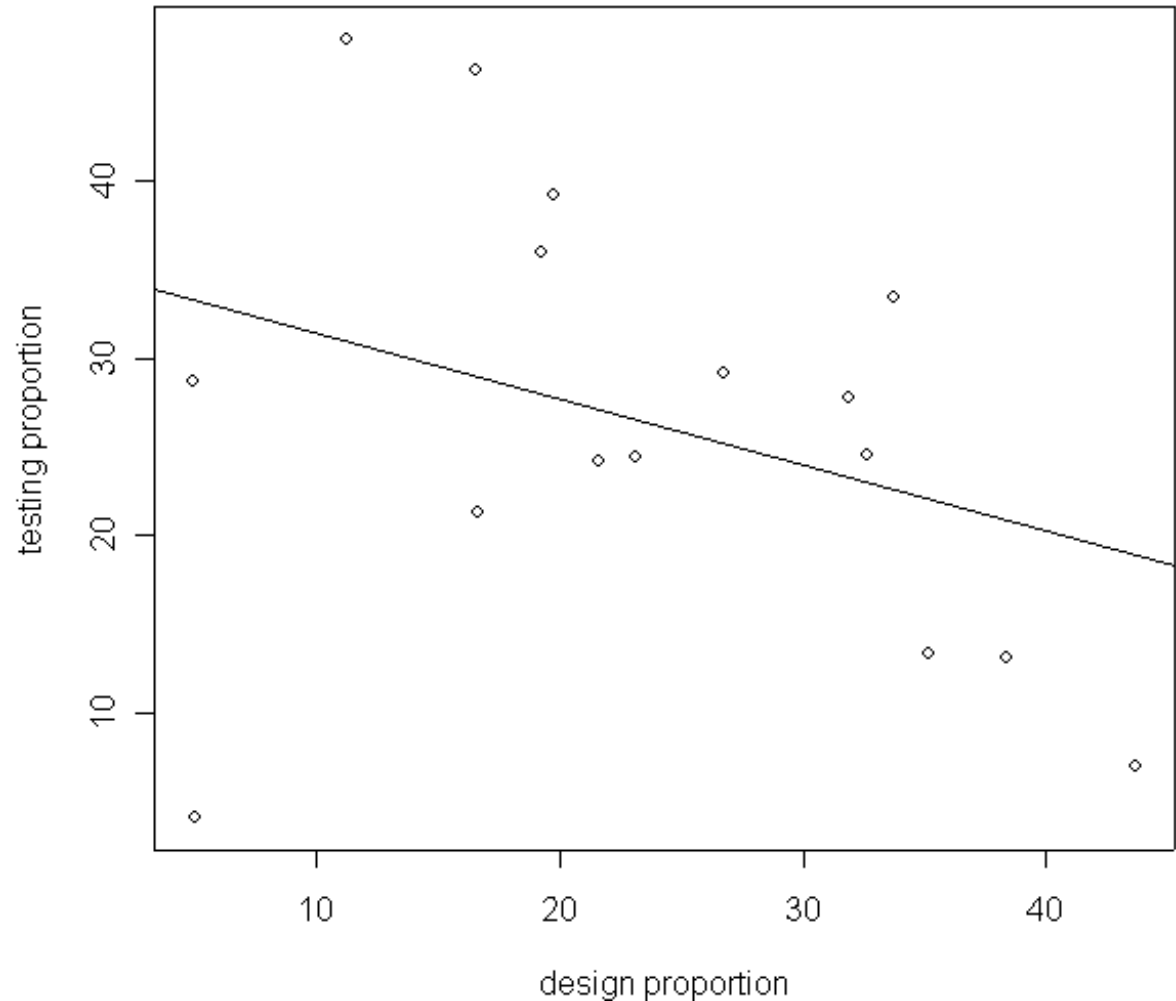
Testanteil vs. Planungsanteil

- $\tau = -0,15$
- $p = 45 \%$



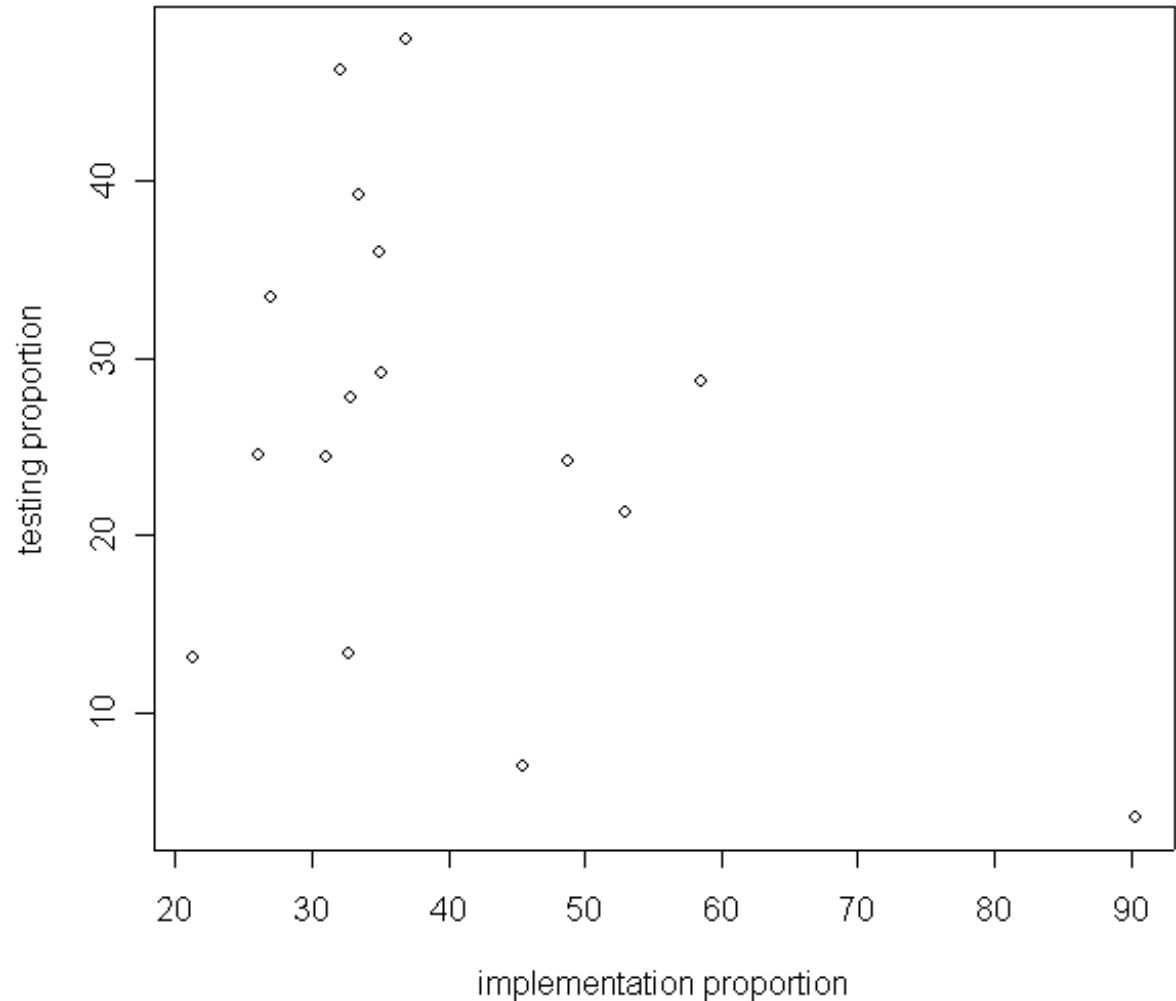
Testanteil vs. Entwurfsanteil

- $\tau = -0,32$
- $p < 10 \%$



Testanteil vs. Implementierungsanteil

- $\tau = -0,1$
- $p = 63 \%$



Interpretation

- Anteil der Entwurfsarbeit steigt mit Anteil der Planungsarbeit
- Anteil der Implementierungsarbeit sinkt, wenn Anteil der Planungsarbeit und Anteil der Entwurfsarbeit zunehmen
- Anteil der Testarbeit sinkt, wenn Anteil der Entwurfsarbeit zunimmt

Bemerkung

- klar ist: wenn der Anteil von PP + DES steigt, dann muss der Anteil von IMP + TEST sinken
- denke an ein Kuchendiagramm mit vier Teilen: PP, DES, IMP und TEST!

Zwischenergebnis (2)

- Mit Korrelationen haben wir untermauert, dass die einfache Prozentsatzmethode (mit festen Anteilen der Phasen) nicht funktioniert, denn die Phasenanteile sind offenbar nicht konstant.
- aber Vorsicht: Korrelation alleine reicht für diese Schlussfolgerung nicht aus (brauchen Schwankung der Anteile über größeren Wertebereich)

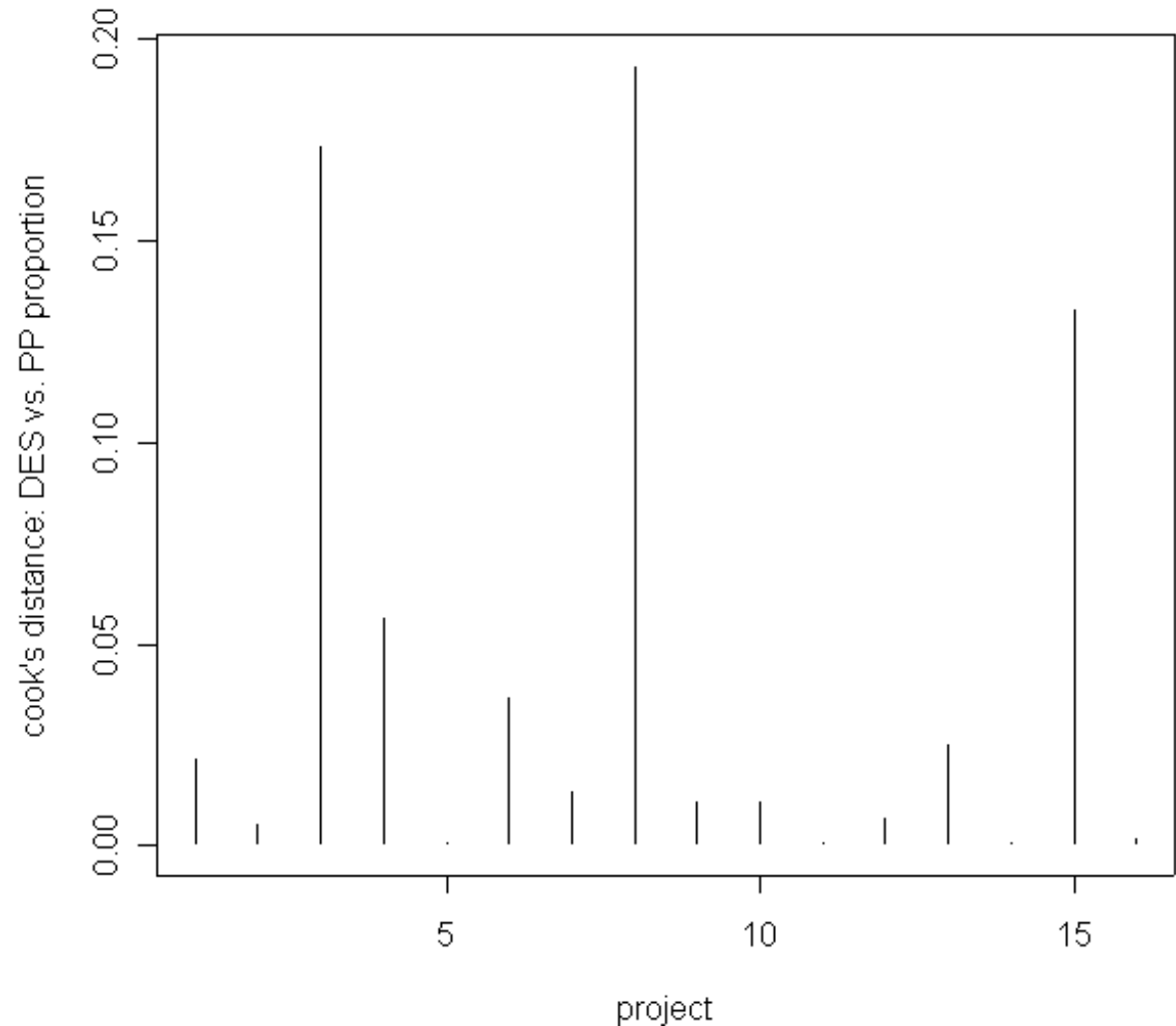
EXKURS IN DIE STATISTIK: REGRESSIONSDIAGNOSTIK

Cook-Abstand

- misst den Einfluss eines Datenpunktes auf die Stärke der Korrelation und die Lage der Regressionsgerade
- Anhaltspunkte für einflussreiche Punkte:
 - Cook-Abstand > 1
 - oder Cook-Abstand deutlich größer als bei den anderen Datenpunkten
- Diese Punkte muss man genauer betrachten, da sie das Ergebnis verfälschen können.

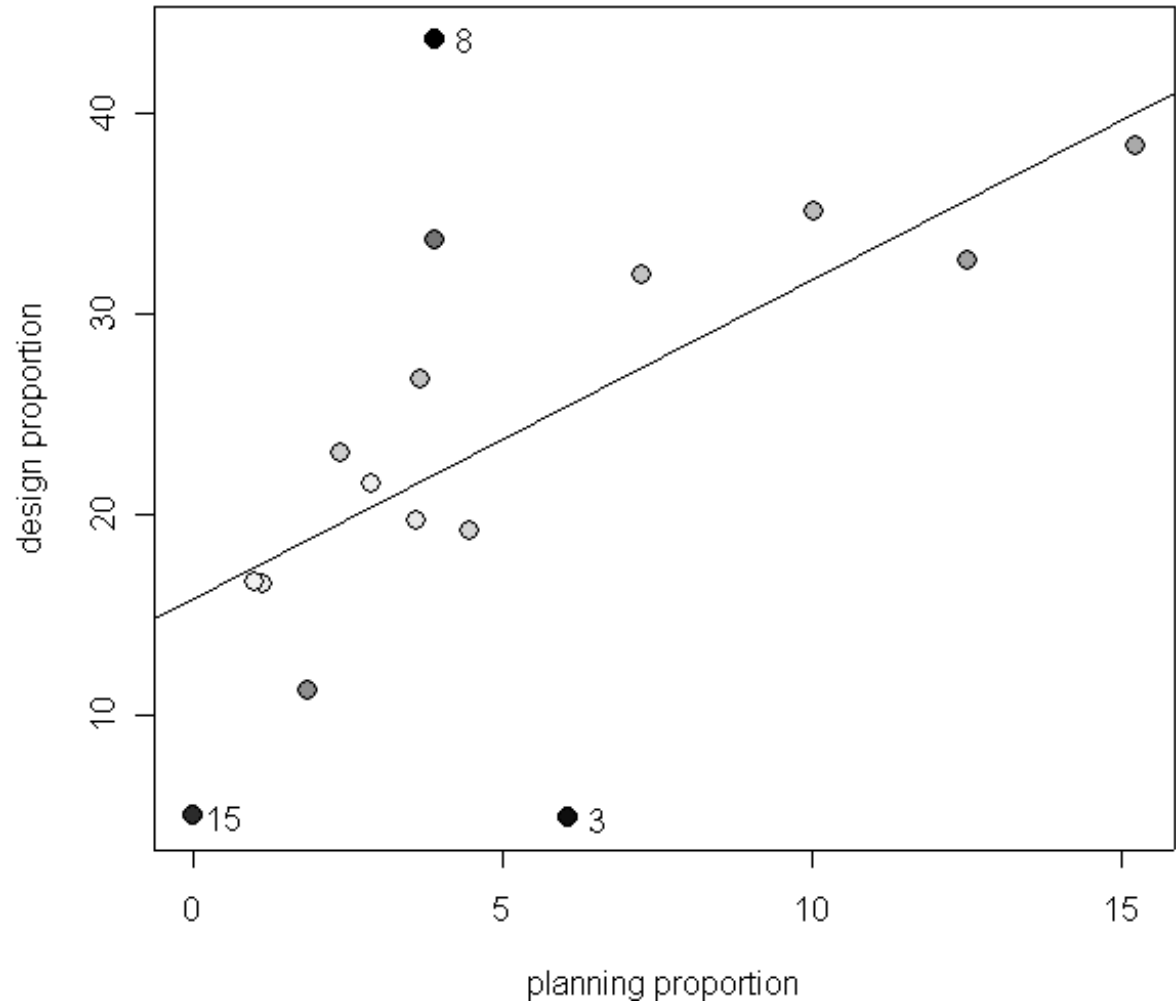
Entwurfsanteil vs. Planungsanteil (CookD)

- cook > 0,1:
3,8,15



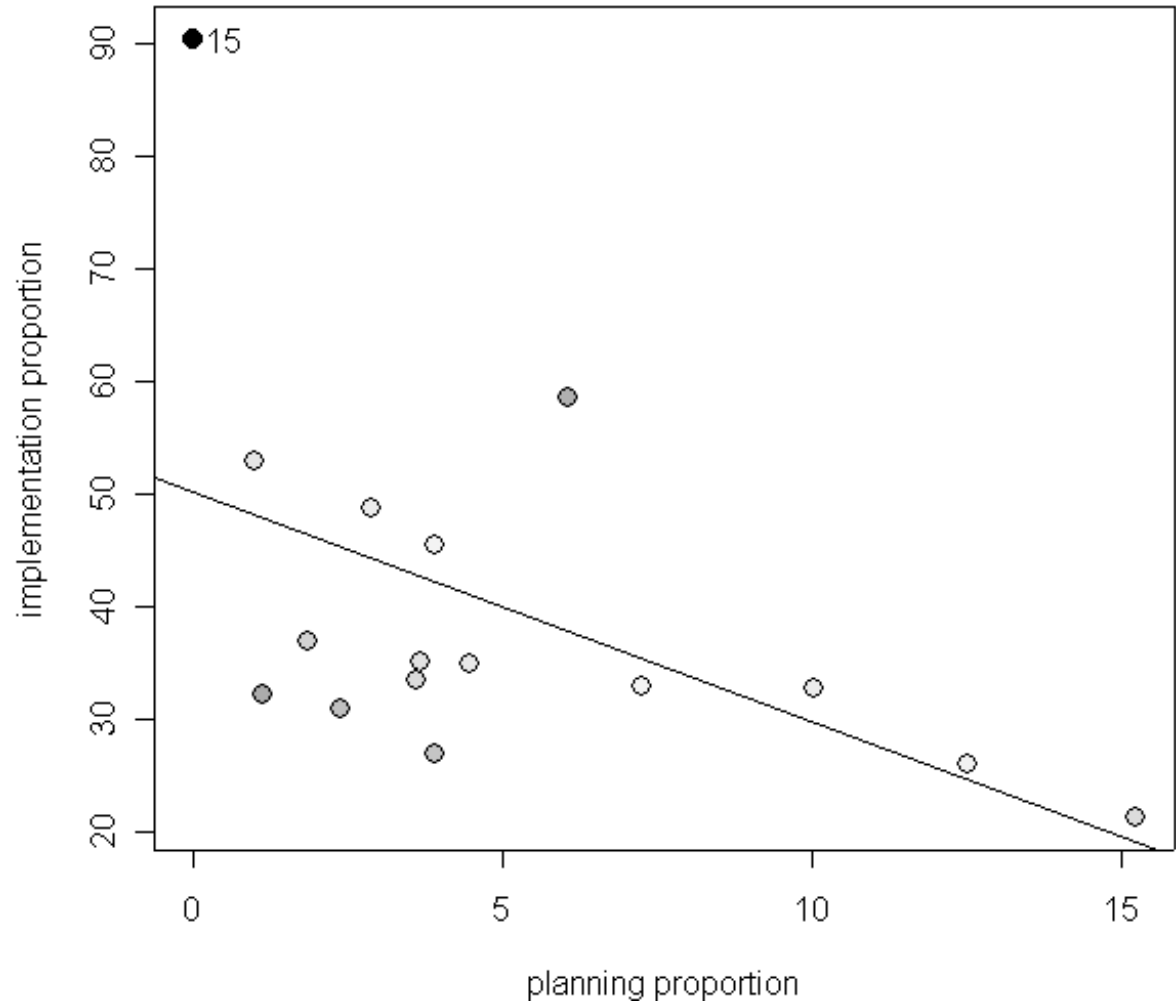
Entwurfsanteil vs. Planungsanteil (nochmal, mit Färbung des Cook-Abstands)

- $\tau = 0,5$
- $p < 1 \%$



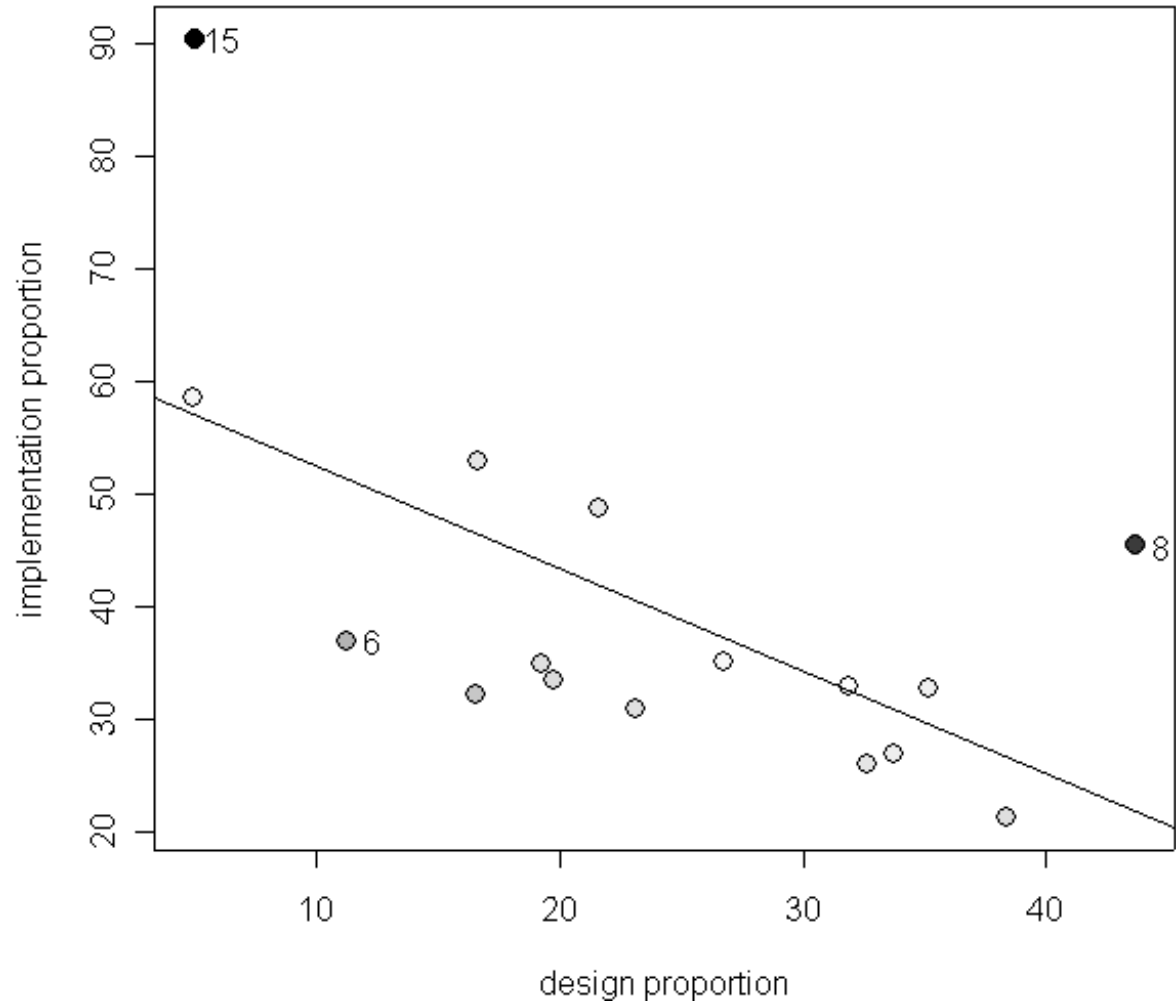
Implementierungsanteil vs. Planungsanteil

- $\tau = -0,42$
- $p = 2 \%$



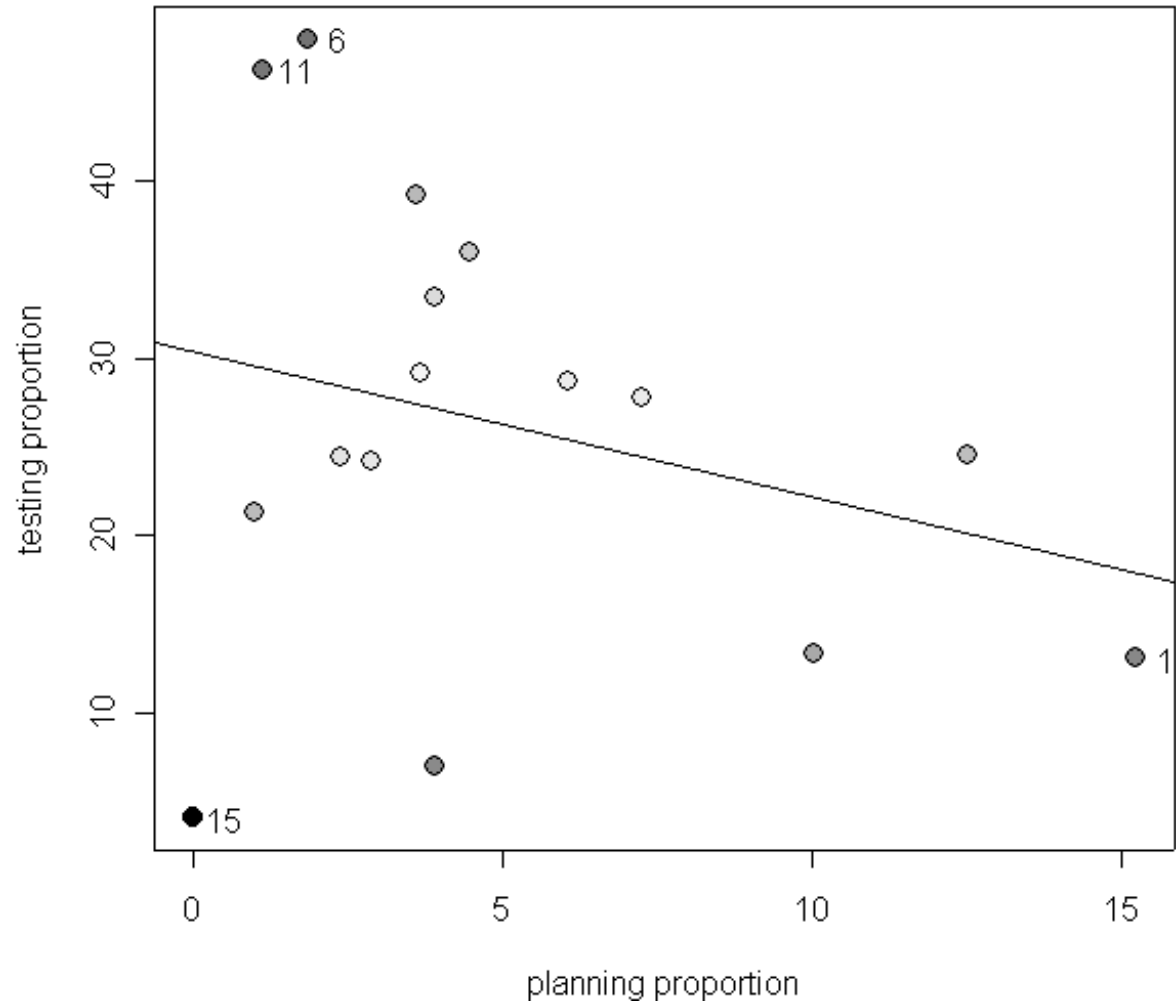
Implementierungsanteil vs. Entwurfsanteil

- $\tau = -0,48$
- $p < 1 \%$



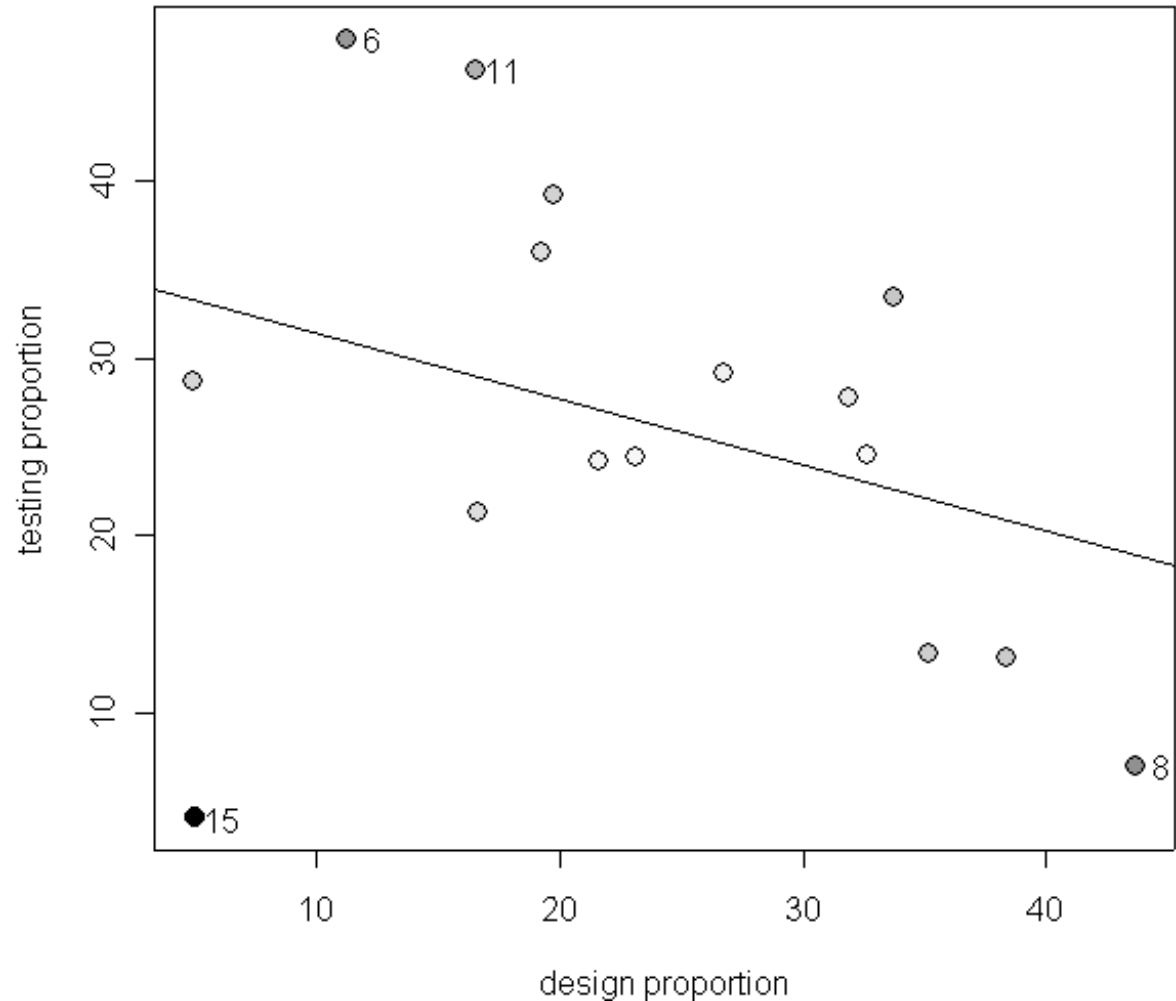
Testanteil vs. Planungsanteil

- $\tau = -0,15$
- $p = 45\%$



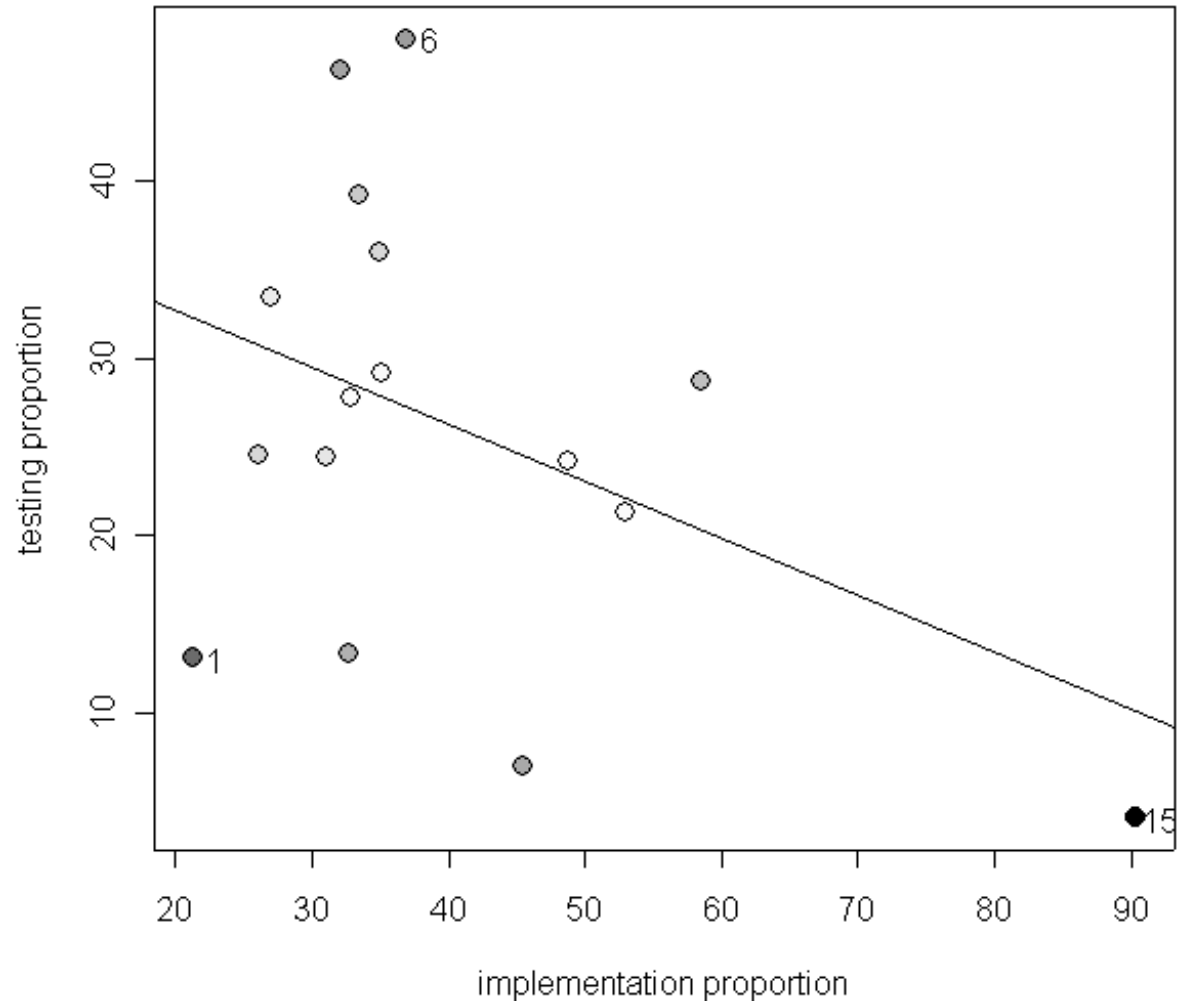
Testanteil vs. Entwurfsanteil

- $\tau = -0,32$
- $p < 10 \%$



Testanteil vs. Implementierungsanteil

- $\tau = -0,1$
- $p = 63 \%$



Einflussreiche Datenpunkte

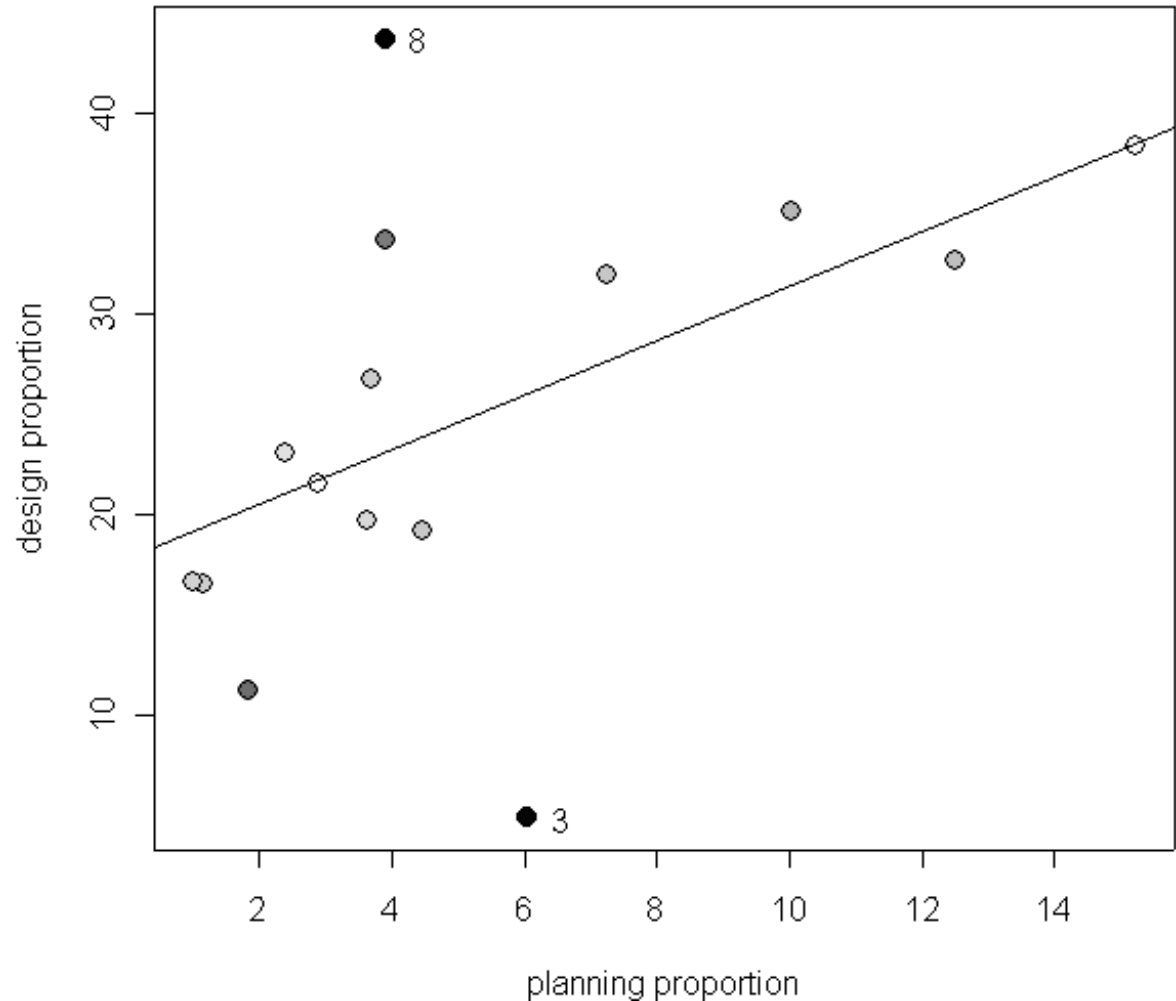
- Projekt 15: IMP ~ 90 %
- Projekt 8: DES + IMP ~ 90 %
- Projekt 6: IMP + TEST ~ 85 %
- Projekt 3: IMP + TEST ~ 85 %

Sensitivitäts-Analyse

- Datenpunkt 15 hat oft einen auffälligen Cook-Abstand
- extrem hoher Anteil der Implementierung am Gesamtaufwand bei diesem Projekt (IMP ~ 90 %)
- unklar, ob schlechte Aufwanderfassung oder merkwürdiges Projekt
- wie ändern sich die Regressionen, wenn man diesen Datenpunkt weglässt?

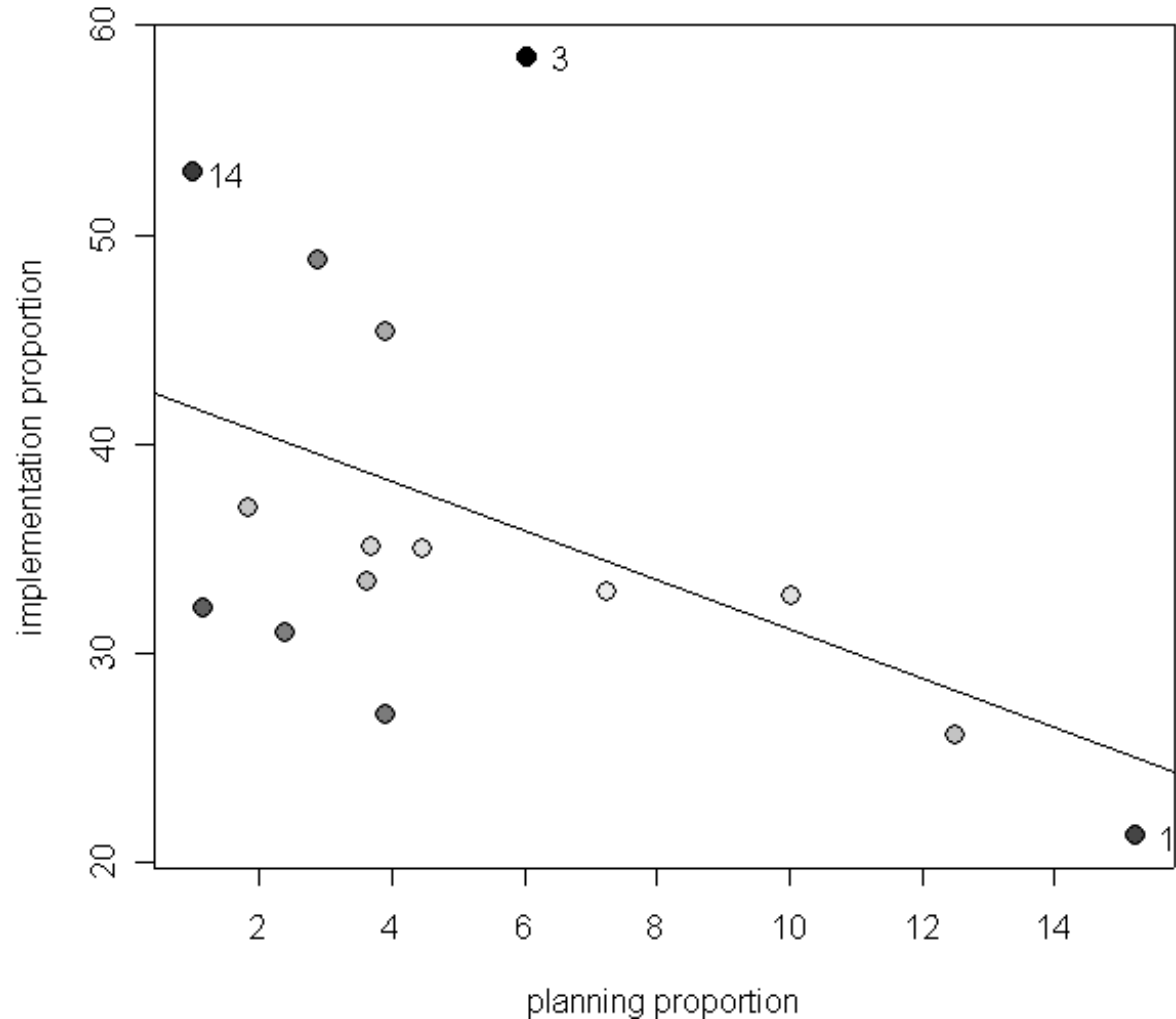
Entwurfsanteil vs. Planungsanteil (ohne 15)

- $\tau = 0,45$
- $p = 2\%$
(alt $p < 1\%$)



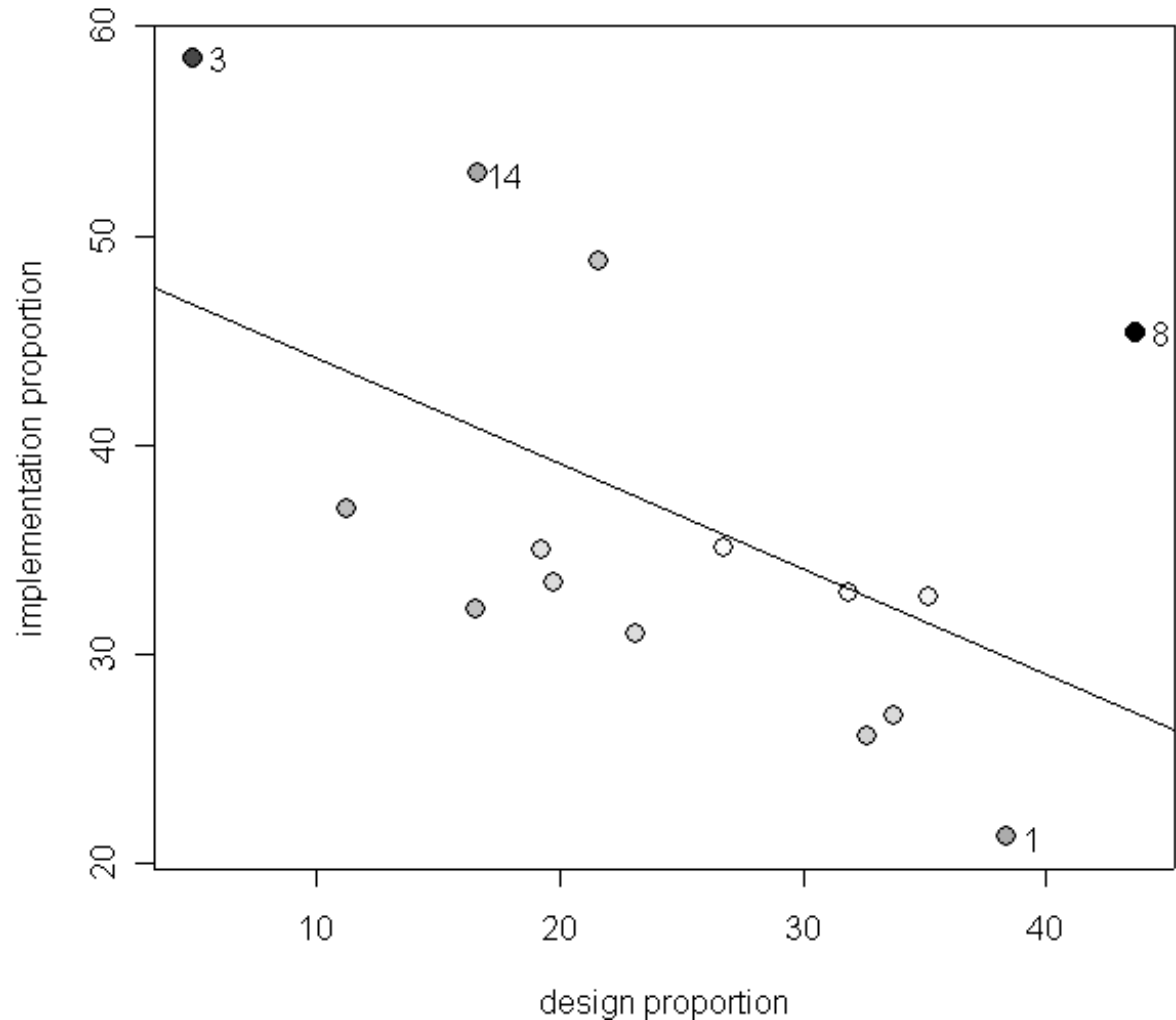
Implementierungsanteil vs. Planungsanteil (ohne 15)

- $\tau = -0,33$
- $p = 9\%$
(alt 2%)



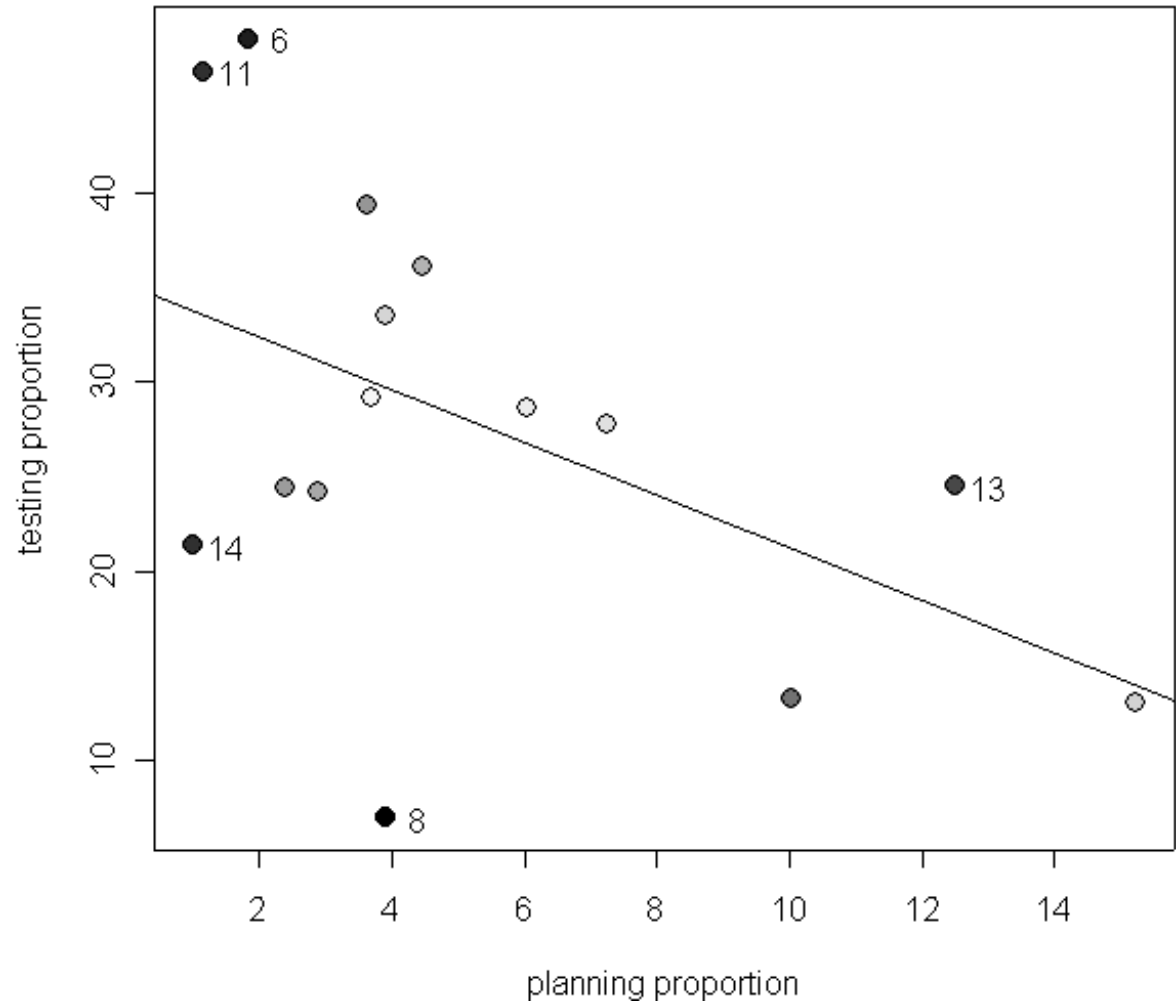
Implementierungsanteil vs. Entwurfsanteil (ohne 15)

- $\tau = -0,43$
- $p = 3 \%$
(alt $p < 1 \%$)



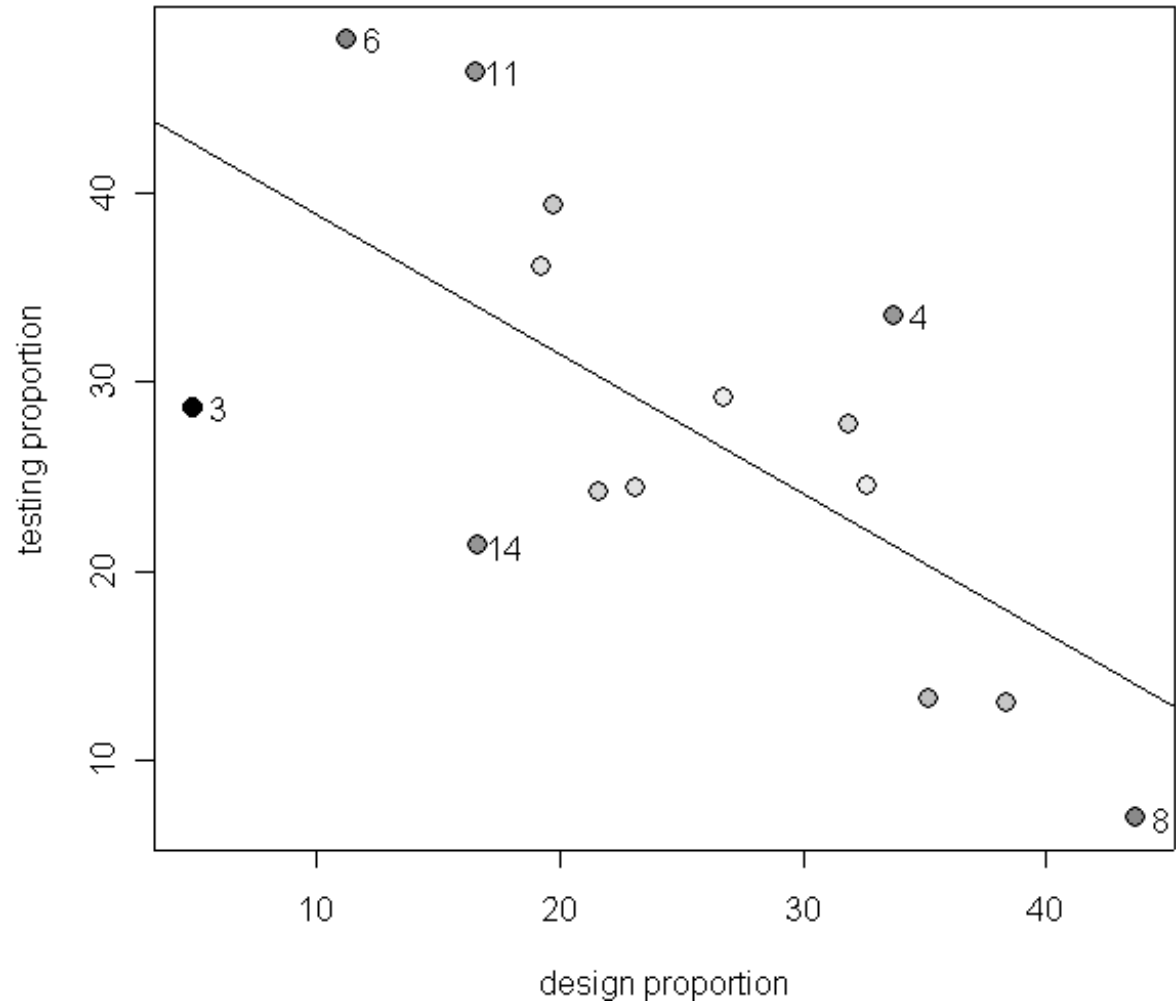
Testanteil vs. Planungsanteil (ohne 15)

- $\tau = -0,31$
- $p = 11 \%$
(alt $p = 45 \%$)



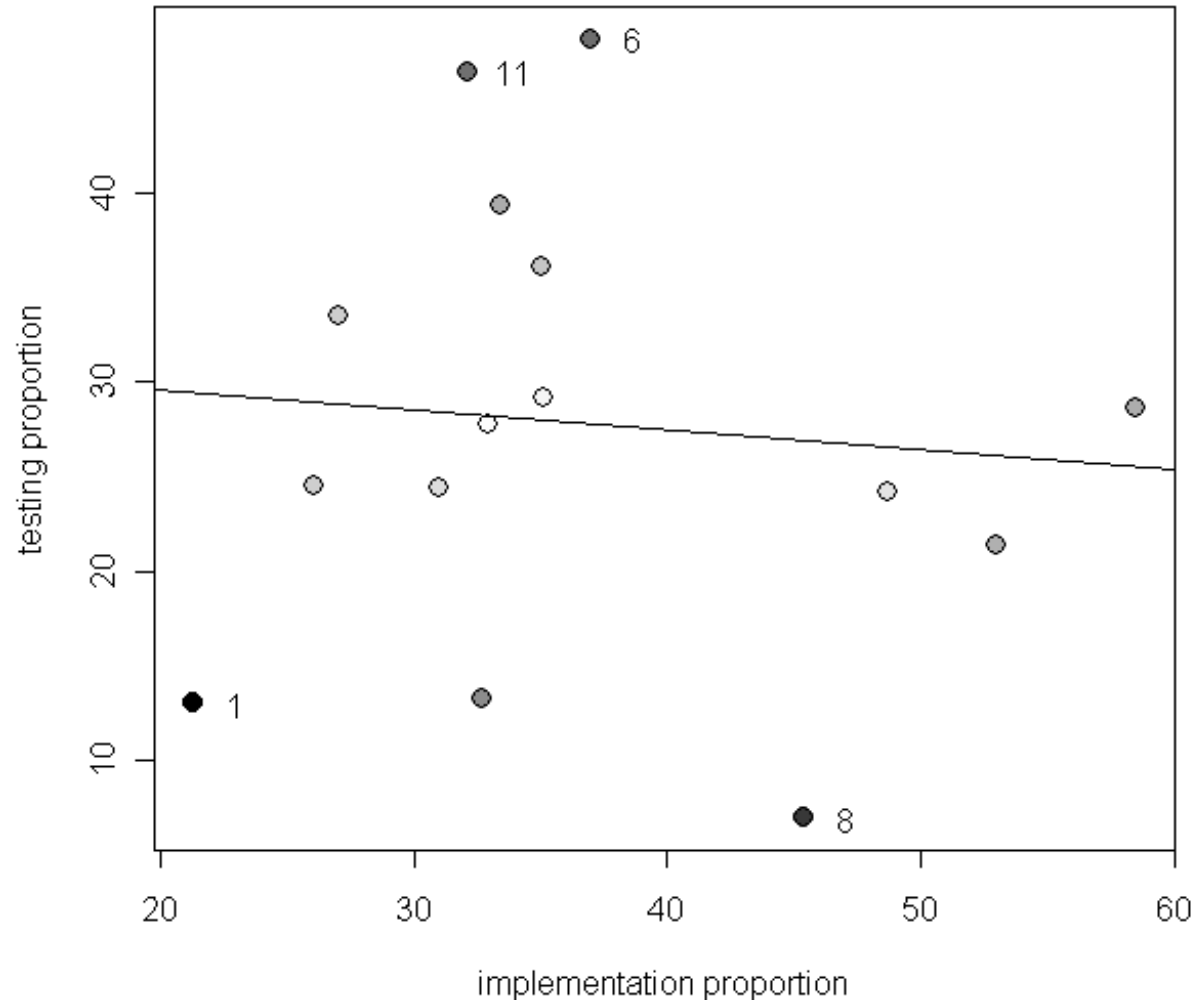
Testanteil vs. Entwurfsanteil (ohne 15)

- $\tau = -0,49$
- $p < 1 \%$
(alt $p < 10 \%$)



Testanteil vs. Imp-Anteil (ohne 15)

- $\tau = 0,03$
- $p = 100\%$
(alt $p = 63\%$)



Ergebnis der Korrelationsdiagnostik

- die meisten Trends sind bei Weglassen des Projekts 15 unverändert
- zusätzlicher Trend: Testanteil sinkt, wenn Planungsanteil zunimmt
- verstärkter Trend: Testanteil sinkt, wenn Entwurfsanteil zunimmt
- Testanteil und Implementierungsanteil sind noch deutlicher unkorreliert

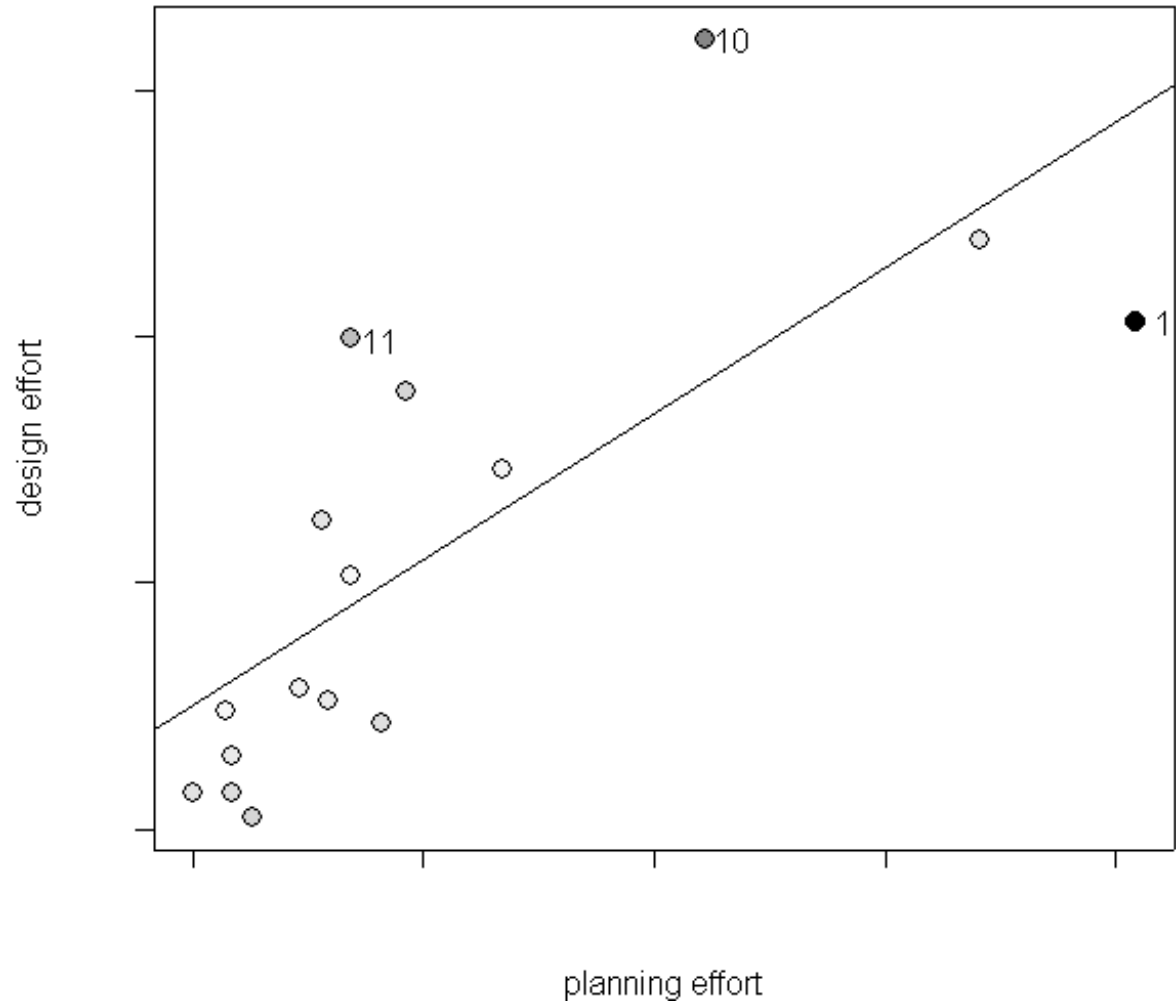
ZURÜCK ZUR ANALYSE...

Dritter Versuch...

- Korrelationsanalyse der absoluten Aufwände für die einzelnen Entwicklungsphasen (nicht mehr die Anteile)

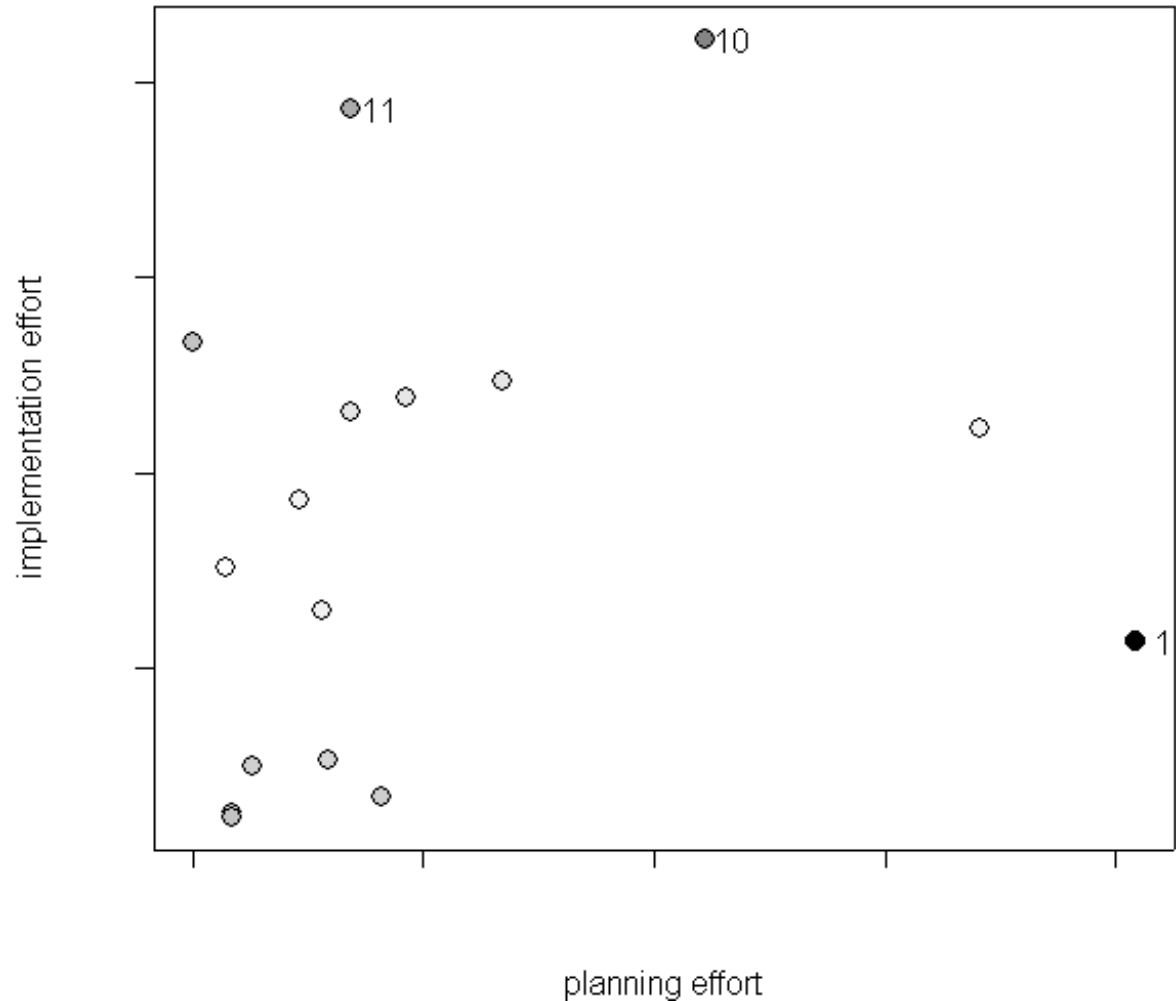
Entwurf vs. Planung

- $\tau = 0,62$
- $p < 1 \%$



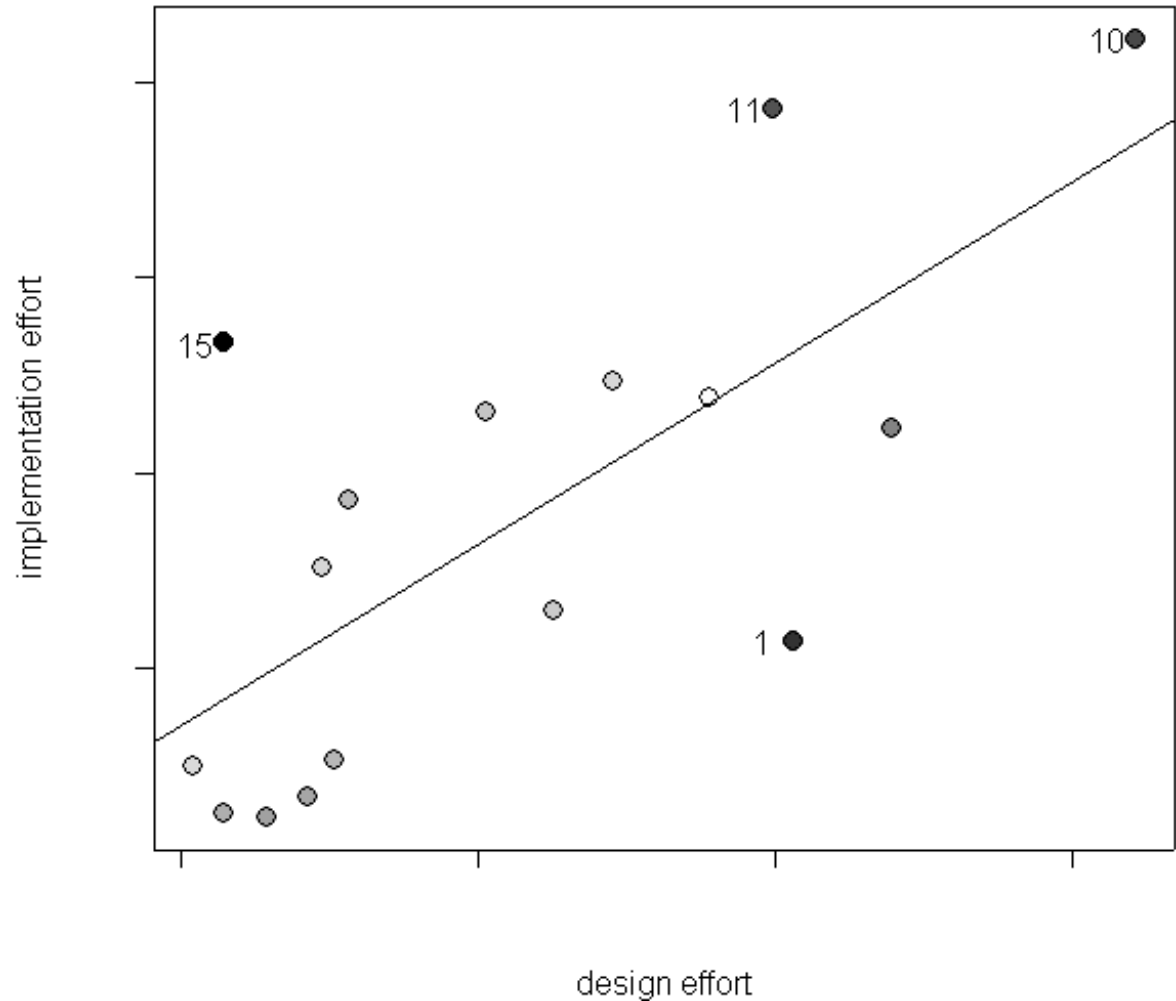
Implementieren vs. Planung

- $\tau = 0,25$
- $p = 17\%$



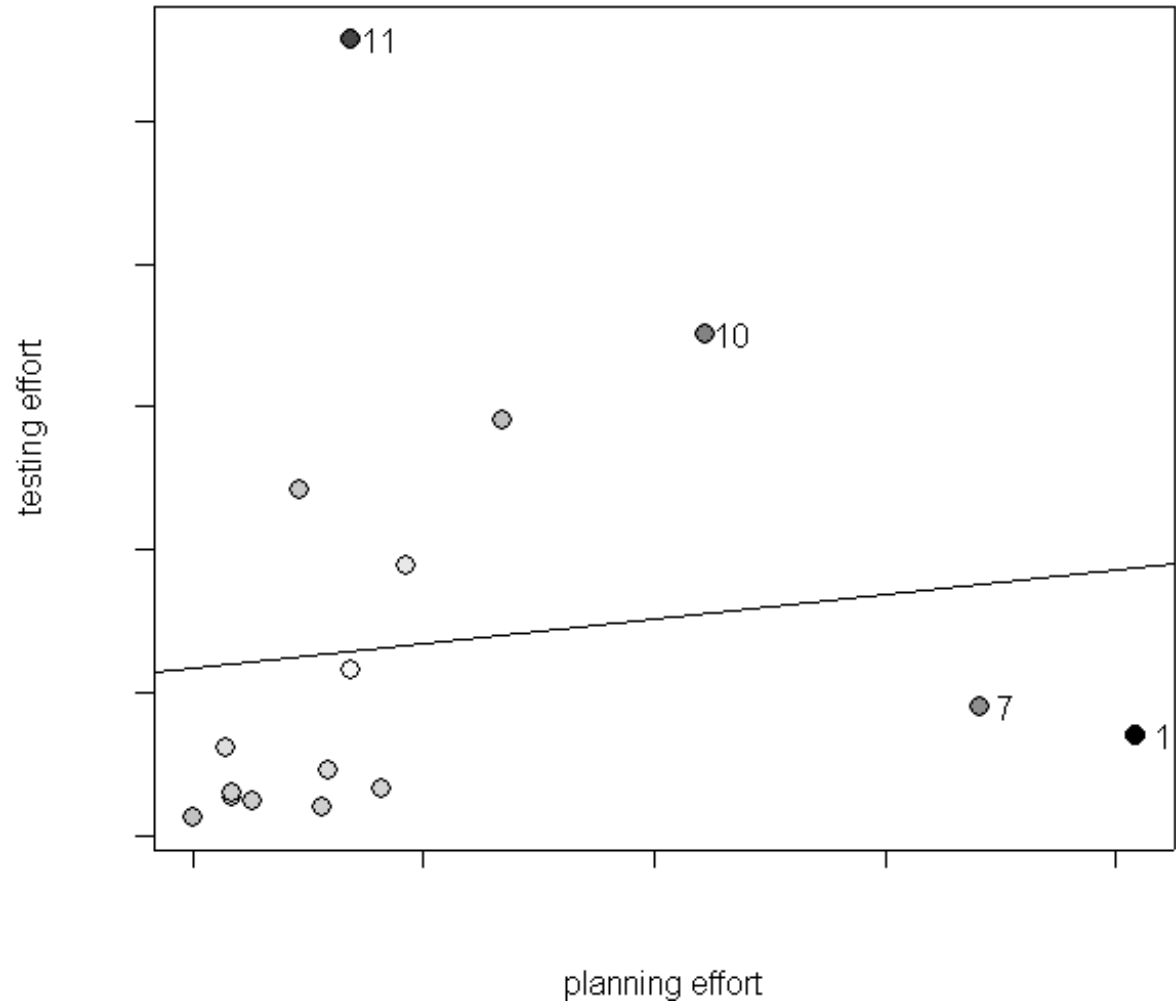
Implementieren vs. Entwurf

- $\tau = 0,48$
- $p = 1 \%$



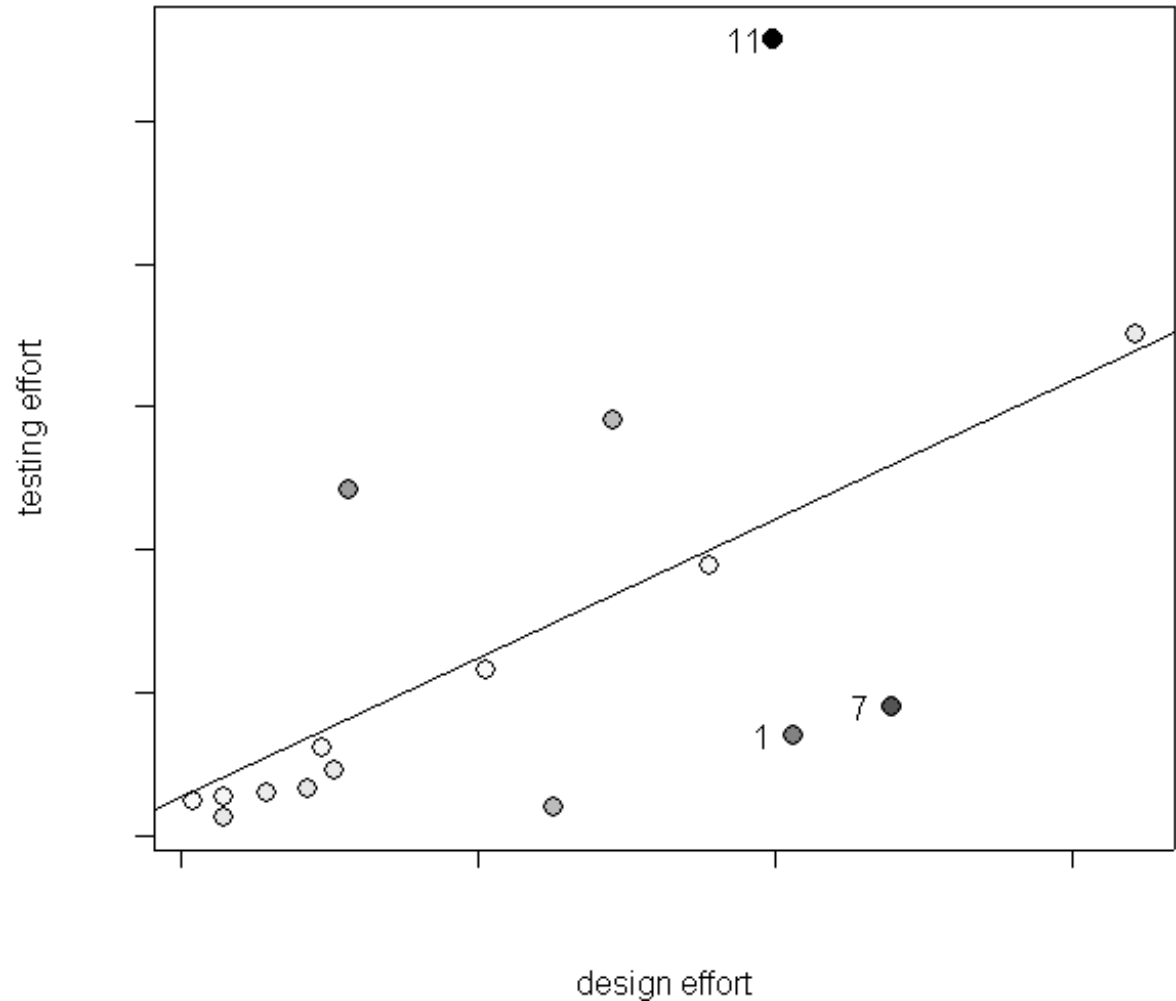
Testen vs. Planung

- $\tau = 0,4$
- $p = 3 \%$



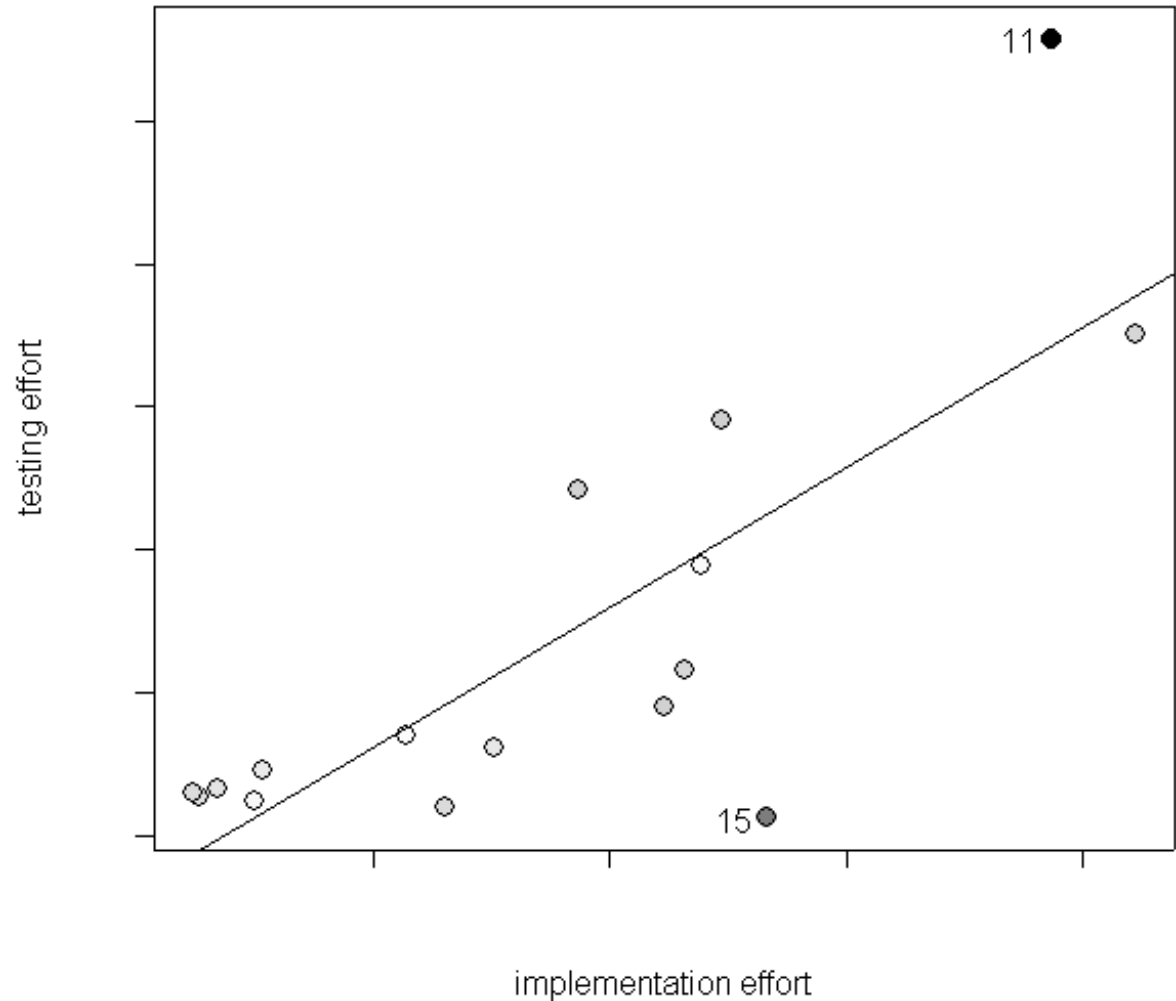
Testen vs. Entwurf

- $\tau = 0,58$
- $p < 1 \%$



Testen vs. Implementieren

- $\tau = 0,53$
- $p < 1 \%$



Korrelationskoeffizienten

	PP	DES	IMP
DES	0,62*		
IMP	0,25	0,48*	
TEST	0,40"	0,58*	0,53*

* bedeutet: signifikant auf 1 % Niveau

" bedeutet: signifikant auf 5 % Niveau

Schätzungen

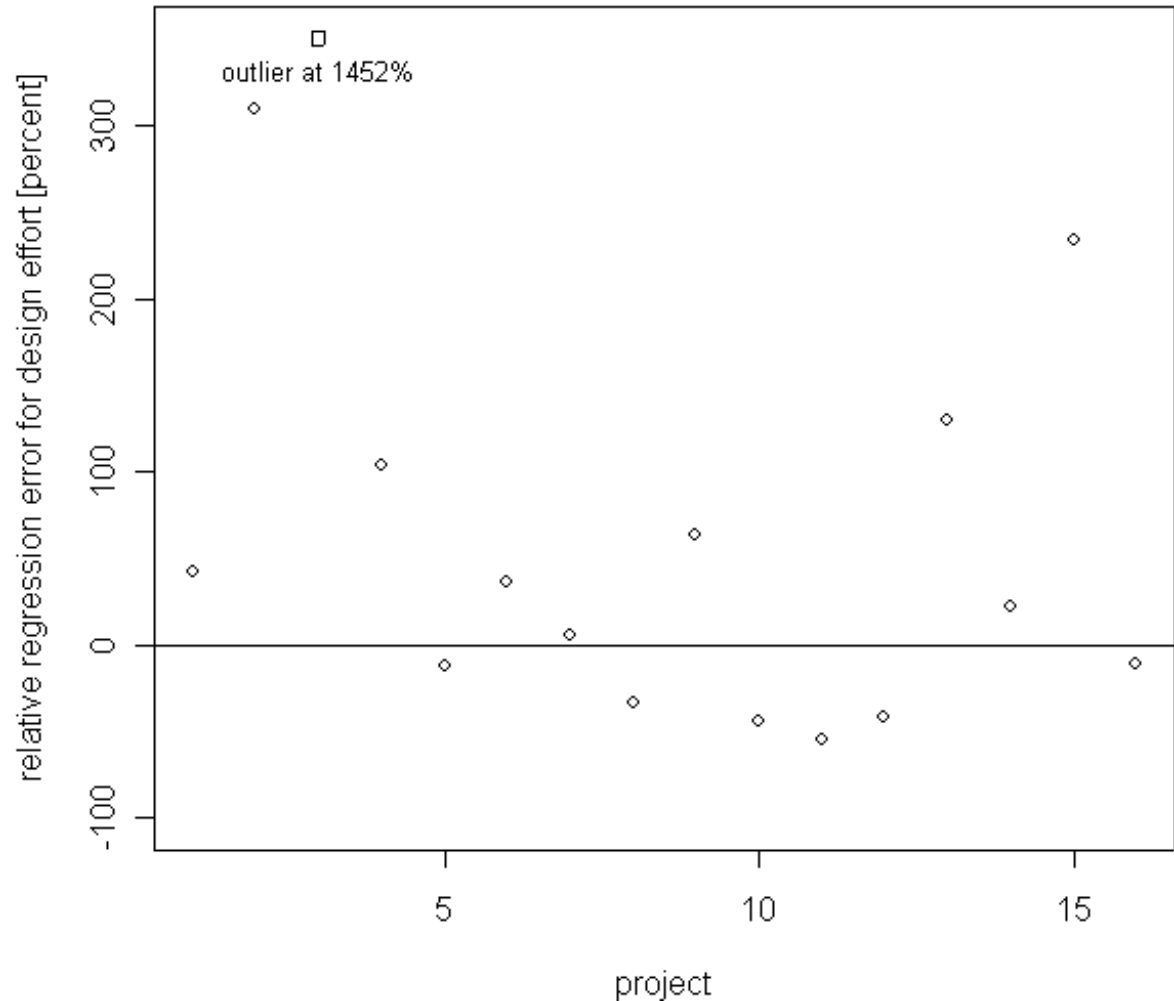
- wie groß ist der Fehler, wenn man die jeweilige Regressionsgerade zum Schätzen des Aufwands verwendet?
- eigentlich nötig: Jackknife-Prozedur!
- schnelle, ungefähre Antwort:
betrachte die Residuen, also den Abstand der Datenpunkte von der Regressionsgerade

Jackknife-Prozedur

- einen Datenwert herausnehmen
- die restlichen zur Schätzung des herausgenommenen Werts benutzen
- den Schätzfehler bestimmen (relativer Fehler in Prozent des aktuellen Wertes)
- für alle Datenpunkte wiederholen und den mittleren relativen Fehler berechnen

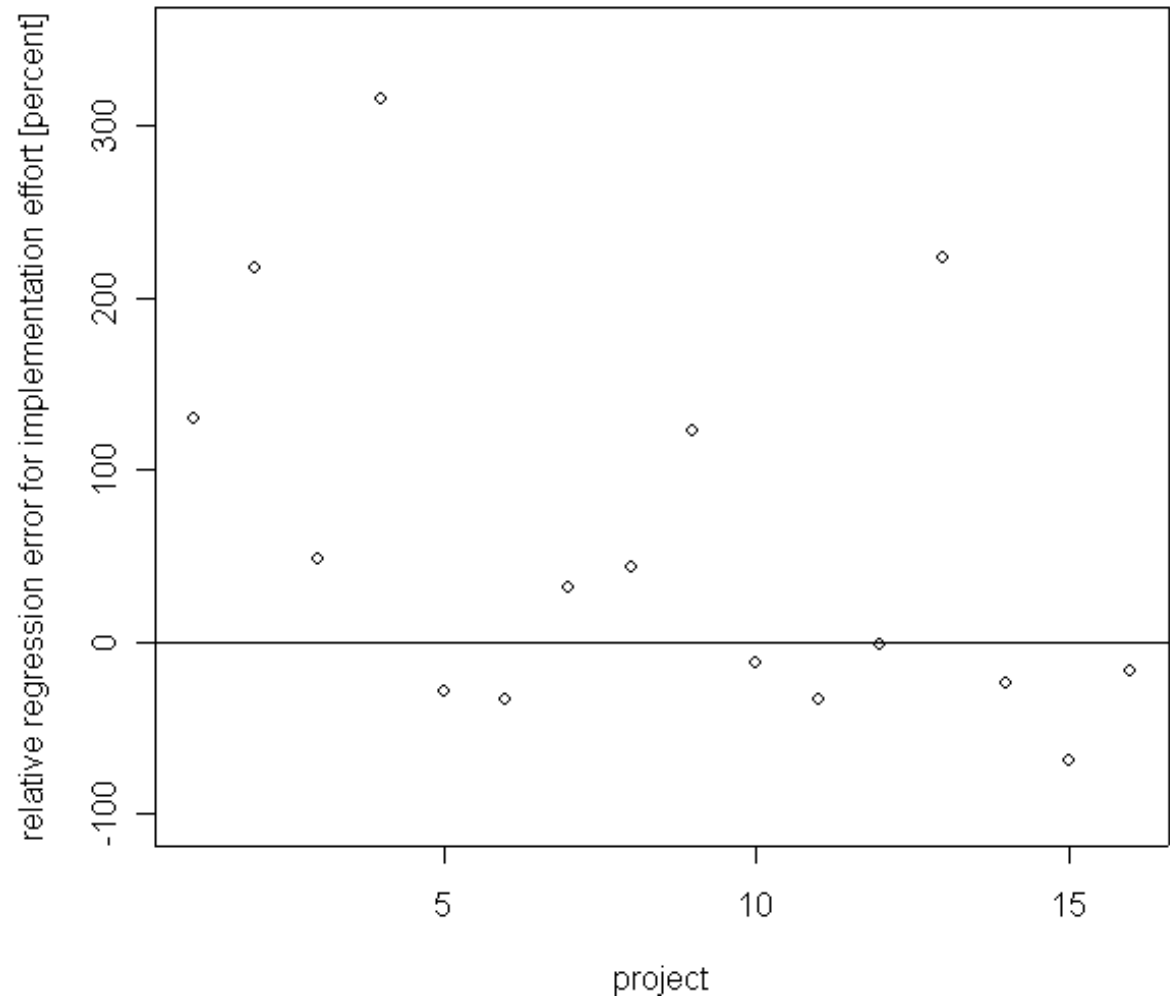
Relativer Regressionsfehler: Entwurf vs. Planung

- $\frac{1}{n} \sum |\cdot| = 162 \%$
- ohne Ausreißer:
76 %



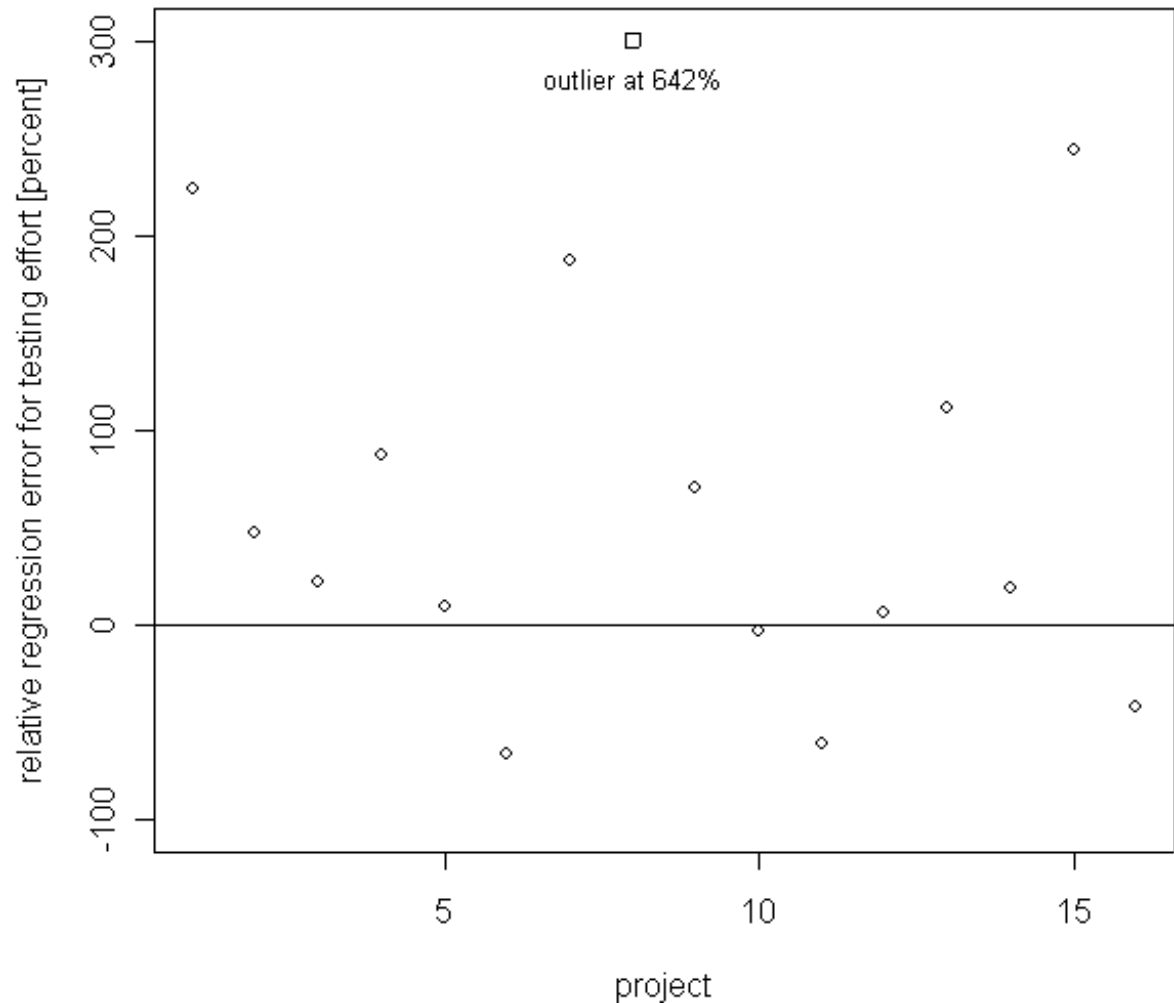
Relativer Regressionsfehler: Implementieren vs. Entwurf

■ $\frac{1}{n} \sum |\cdot| = 84 \%$



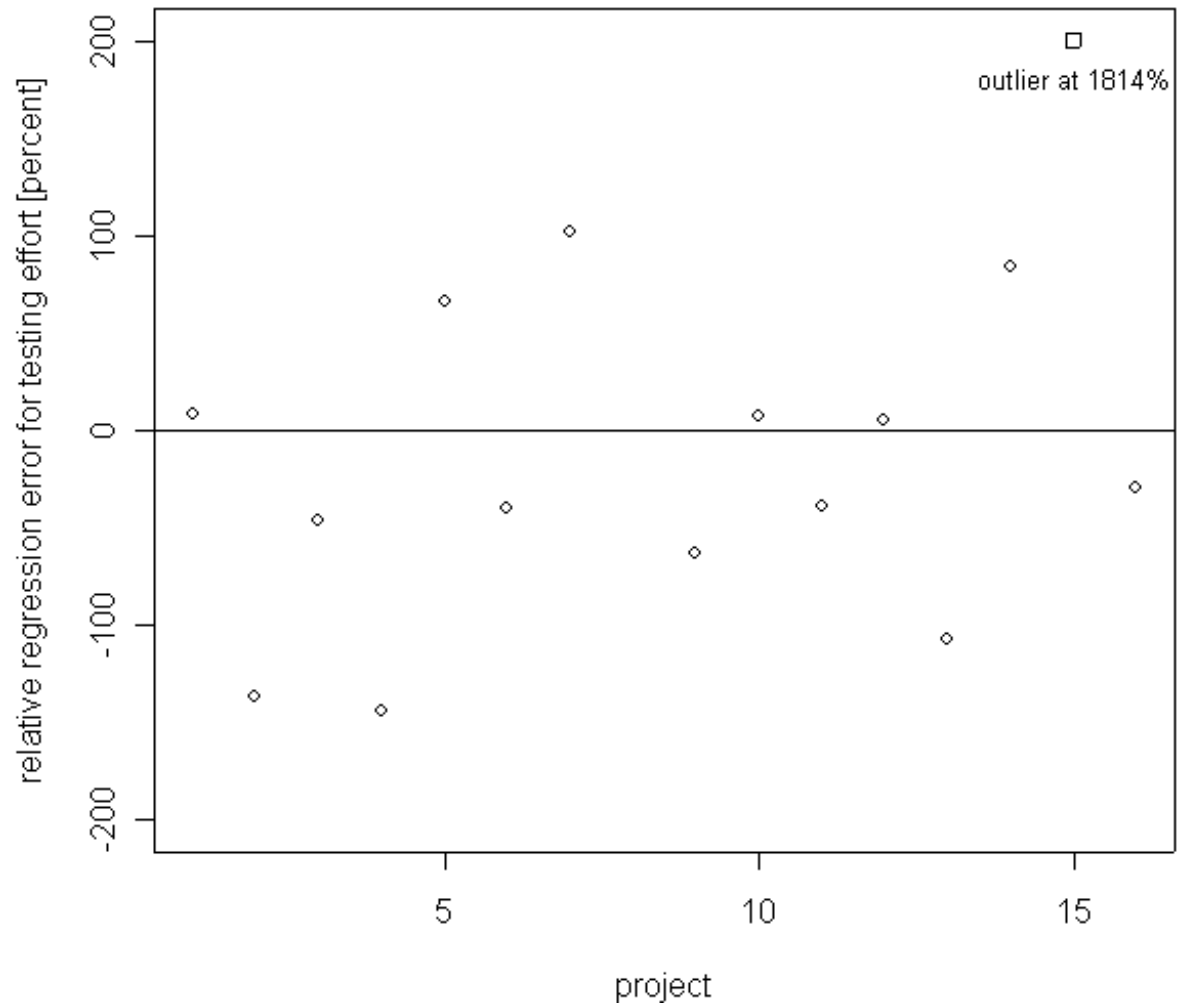
Relativer Regressionsfehler: Testen vs. Entwurf

- $\frac{1}{n} \sum |\cdot| = 115 \%$
- ohne Ausreißer:
80 %



Relativer Regressionsfehler: Testen vs. Implementieren

- $\frac{1}{n} \sum |\cdot| = 190 \%$
- ohne Ausreißer:
82 %



Maße für die Schätzqualität

- Mean Magnitude of Relative Error
 - manchmal auch „average error“ genannt
 - Frage: in welcher Größenordnung liegt der prozentuale Schätzfehler?
 - $MMRE = 1/n \sum |estimated - actual|$
 - soll: $MMRE < 25 \%$
 - bedeutet: im Mittel sind die Schätzwerte auf 25 Prozent genau

Schätzqualität (Forts.)

- Prediction Quality
 - Frage: wie viele Schätzwerte haben eine brauchbare Genauigkeit?
 - soll: $\text{PRED}(0,25) > 0,75$
 - bedeutet: mehr als 75 Prozent der Schätzwerte sind auf 25 Prozent genau
- für unsere Regressionsmodelle: der Fehler ist bei weitem zu hoch!
(MMRE \approx 80 % ohne die Ausreißer)

Zwischenergebnis (3)

- deutliche Korrelation der absoluten Aufwände von:
 - Entwurf und Planung
 - Implementierung und Entwurf
 - Testen und Entwurf
 - Testen und Implementierung
- ...aber auch massive Abweichungen von den Schätzgeraden und zu hohe prozentuale Fehler

Vierter Versuch...

- Einbeziehen weiterer Kenngrößen in die Regression, z.B.:
 - Ausführungsumgebung (Unix vs. „runtime“)
 - Programmiersprache (C vs. C++)
- kombinierte Modelle aus Regressions-Schätzwert und Manager-Schätzwert

Weitere Regressionsvariable

- Programmiersprache und Ausführungsumgebung sind eng korreliert:
 - C++ bei Unix
 - C bei „runtime“
- 12 von 16 Projekten hatten eine dieser beiden Kombinationen
- Ausführungsumgebung als die zweite Variable in der Regression gewählt;
- Ein kombiniertes Modell aus Regression und Manager-Schätzung wurde auch aufgestellt.
- In der weiteren Untersuchung wird immer das beste der Modelle verglichen.

Bemerkung

- Indikator als zweite Regressionsvariable ist nicht dasselbe wie Aufteilen des Datensatzes vor der Regression:
 - $y = ax + b + cI$ mit $I = 0$ oder 1
 - Aufteilung des Datensatzes nach Ausführungsumgebung ergäbe zwei Geraden:
 $y = sx + t$ und $y = ux + v$

Zwischenergebnis (4)

- Summe der abs. Schätzfehler in Personenstunden:
(siehe Tabelle 6 im Artikel)

	OE	Bestes Modell	Lin. Reg.
DES~PP	3243	3160	3915
IMP~DES	7888	3688	5741
TEST~DES	6351	4594	6297
TEST~IMP	6351	4245	5476

Eines der Regressionsmodelle
oder „kombinierten“ Modelle mit
Manager-Schätzungen –
siehe den Artikel

OE = original effort estimate

Ergebnisse

- kombinierte Modelle (Regression plus Expertenschätzung) hatten meist den kleinsten absoluten Fehler
- eine Schätzung des Implementierungs- und Testaufwands aus dem tatsächlichen Entwurfsaufwand scheint lohnenswert
- der Entwurfsaufwand (und nur dieser) wird von den Experten nach Abschluss der Planungsphase (CE-Wert) etwa gleich gut geschätzt wie vom besten Modell
- betrachtet man die Projekte einer Firma als „Portfolio“, dann gleicht sich der Schätzfehler der Regressionsmodelle über das Portfolio aus (positive Überschätzung und negative Unterschätzung ergeben in Summe einen kleinen absoluten Fehler).

Diskussion

- die Analyse hat glaubwürdig gezeigt:
 - die einfache Prozentsatzmethode wird in der Praxis kaum funktionieren (nicht einmal bei relativ homogenen Projekten)
 - es kann durchaus Zusammenhänge geben zwischen den Anteilen der einzelnen Phasen am Gesamtaufwand
 - die lokal gesammelten Daten entscheiden über die Qualität einer Schätzmethode

Diskussion (Forts.)

- was den absoluten Fehler in Personenstunden betrifft, bringen die Regressionsmodelle eine Verbesserung gegenüber den Expertenschätzungen
- was den prozentualen Fehler betrifft, haben die Regressionsmodelle aber eine viel zu hohe Schwankung (leider)

Diskussion (Forts.)

- es fehlt eine genaue Ursachenforschung:
 - man muss Daten sammeln über Besonderheiten der Projekte und ihren Verlauf, um Abweichungen von einer Regressionskurve erklären zu können
 - dabei kann eine Liste von Kostentreibern („cost drivers“) helfen; siehe z.B. die Liste von B. Boehm in seinem CoCoMo-Modell

Fazit

- es ist ein großer Unterschied, ob man
- mittels Korrelationsanalyse Trends herausfinden will, oder
- mittels der zugehörigen Regressionsgeraden tatsächlich Schätzungen, also Vorhersagen machen will

