

Appendix To: Coverage Is Not Strongly Correlated With Test Suite Effectiveness

Laura Inozemtseva and Reid Holmes
School of Computer Science
University of Waterloo
Waterloo, ON, Canada
{linozem,rtholmes}@uwaterloo.ca

ABSTRACT

This document provides the full dataset summarized in the paper “Coverage Is Not Strongly Correlated With Test Suite Effectiveness”.

1. TABLES FOR RESEARCH QUESTION 2

The tables in this section give the correlation between coverage and effectiveness when suite size is ignored. Table 1 gives the results for the normalized kill score (i.e., number of mutants killed divided by number of non-equivalent mutants covered); Table 2 gives the results for the raw kill score (i.e., number of mutants killed divided by total number of non-equivalent mutants).

Project	Statement	Decision	Mod. Cond.
Apache POI	0.75	0.76	0.77
Closure	0.83	0.83	0.84
HSQLDB	-0.35	-0.35	-0.35
JFreeChart	0.50	0.53	0.53
Joda Time	0.80	0.80	0.80

Table 1: The Kendall τ correlation between effectiveness (normalized kill score) and different types of coverage when suite size is ignored. All entries are significant at the 99.9% level.

2. TABLES FOR RESEARCH QUESTION 3

The tables in this section give the correlation between coverage and effectiveness when suite size is controlled for. There are two tables for each subject program: one gives the results for the normalized kill score (i.e., number of mutants killed divided by number of non-equivalent mutants covered); the other gives the results for the raw kill score (i.e., number of mutants killed divided by total number of non-equivalent mutants).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

Project	Statement	Decision	Mod. Cond.
Apache POI	0.94	0.94	0.94
Closure	0.95	0.95	0.95
HSQLDB	0.81	0.80	0.79
JFreeChart	0.91	0.95	0.92
Joda Time	0.85	0.85	0.85

Table 2: The Kendall τ correlation between effectiveness (raw kill score) and different types of coverage when suite size is ignored. All entries are significant at the 99.9% level.

Apache POI (Normalized)			
Size	Statement	Decision	Mod. Cond.
3	-0.17	-0.13	-0.07
10	0.14	0.16	0.22
30	0.18	0.27	0.28
100	0.15	0.21	0.22
300	0.41	0.40	0.42
1000	0.49	0.46	0.49

Table 3: The correlation between three coverage types and effectiveness (normalized kill score) for fixed size suites for Apache POI. All entries are significant at the 99.9% level.

Apache POI (Raw)			
Size	Statement	Decision	Mod. Cond.
3	0.85	0.84	0.85
10	0.75	0.73	0.77
30	0.61	0.67	0.67
100	0.51	0.48	0.46
300	0.67	0.64	0.63
1000	0.77	0.73	0.69

Table 4: The correlation between three coverage types and effectiveness (raw kill score) for fixed size suites for Apache POI. All entries are significant at the 99.9% level.

Closure (Normalized)			
Size	Statement	Decision	Mod. Cond.
3	0.12	0.14	0.18
10	-0.14	-0.13	-0.04 [†]
30	-0.27	-0.22	-0.16
100	-0.04 [†]	-0.03 [†]	-0.01 [†]
300	0.12	0.12	0.15
1000	0.12	0.10	0.13
3000	0.13	0.17	0.19

Table 5: The correlation between three coverage types and effectiveness (normalized kill score) for fixed size suites for Closure. Entries marked with a dagger are not significant at the 99.9% level.

Closure (Raw)			
Size	Statement	Decision	Mod. Cond.
3	0.79	0.80	0.80
10	0.71	0.72	0.69
30	0.69	0.73	0.70
100	0.72	0.72	0.68
300	0.70	0.66	0.57
1000	0.65	0.62	0.56
3000	0.52	0.52	0.46

Table 6: The correlation between three coverage types and effectiveness (raw kill score) for fixed size suites for Closure. All entries are significant at the 99.9% level.

HSQLDB (Normalized)			
Size	Statement	Decision	Mod. Cond.
3	-0.76	-0.74	-0.69
10	-0.46	-0.44	-0.38
30	-0.17	-0.20	-0.20
100	0.09	0.05 [†]	0.02 [†]
300	0.21	0.19	0.07

Table 7: The correlation between three coverage types and effectiveness (normalized kill score) for fixed size suites for HSQLDB. Entries marked with a dagger are not significant at the 99.9% level.

HSQLDB (Raw)			
Size	Statement	Decision	Mod. Cond.
3	0.77	0.75	0.72
10	0.33	0.30	0.33
30	0.27	0.23	0.19
100	0.34	0.31	0.24
300	0.41	0.40	0.26

Table 8: The correlation between three coverage types and effectiveness (raw kill score) for fixed size suites for HSQLDB. All entries are significant at the 99.9% level.

JFreeChart (Normalized)			
Size	Statement	Decision	Mod. Cond.
3	-0.25	-0.08	-0.20
10	-0.42	-0.26	-0.33
30	-0.28	-0.17	-0.19
100	-0.09	-0.01 [†]	0.03 [†]
300	0.03 [†]	0.11	0.20
1000	0.06 [†]	0.13	0.20

Table 9: The correlation between three coverage types and effectiveness (normalized kill score) for fixed size suites for JFreeChart. Entries marked with a dagger are not significant at the 99.9% level.

JFreeChart (Raw)			
Size	Statement	Decision	Mod. Cond.
3	0.70	0.84	0.56
10	0.66	0.83	0.68
30	0.65	0.78	0.69
100	0.53	0.68	0.57
300	0.46	0.64	0.56
1000	0.46	0.62	0.60

Table 10: The correlation between three coverage types and effectiveness (raw kill score) for fixed size suites for JFreeChart. All entries are significant at the 99.9% level.

Joda Time (Normalized)			
Size	Statement	Decision	Mod. Cond.
3	0.00 [†]	0.00 [†]	0.00 [†]
10	-0.02 [†]	-0.01 [†]	-0.01 [†]
30	0.00 [†]	0.00 [†]	-0.01 [†]
100	0.01 [†]	0.00 [†]	-0.02 [†]
300	0.03 [†]	0.04 [†]	0.04 [†]
1000	0.00 [†]	0.00 [†]	0.00 [†]
3000	-0.04 [†]	-0.03 [†]	-0.02 [†]

Table 11: The correlation between three coverage types and effectiveness (normalized kill score) for fixed size suites for Joda Time. Entries marked with a dagger are not significant at the 99.9% level.

Joda Time (Raw)			
Size	Statement	Decision	Mod. Cond.
3	-0.04 [†]	-0.03 [†]	-0.05 [†]
10	0.01 [†]	0.00 [†]	0.00 [†]
30	0.00 [†]	0.00 [†]	0.00 [†]
100	0.00 [†]	-0.01 [†]	-0.04 [†]
300	0.03 [†]	0.04 [†]	0.04 [†]
1000	0.00 [†]	0.01 [†]	0.00 [†]
3000	-0.01 [†]	0.00 [†]	0.00 [†]

Table 12: The correlation between three coverage types and effectiveness (raw kill score) for fixed size suites for Joda Time. Entries marked with a dagger are not significant at the 99.9% level.