19th International Symposium on Software Reliability Engineering

STUDENT PAPER: The Effect of the Number of Defects on Estimates Produced by Capture-Recapture Models

Gursimran Singh Walia
Mississippi State University
gw86@cse.msstate.edu

Jeffrey C. Carver
University of Alabama
carver@cs.ua.edu

Abstract

Project managers use inspection data as input to capture-recapture (CR) models to estimate the total number of faults present in a software artifact. The CR models use the number of faults found during an inspection and the overlap of faults among inspectors to calculate the estimate. A common belief is that CR models underestimate the number of faults but their performance can be improved with more input data. This paper investigates the minimum number of faults that has to be present in an artifact before the CR method can be used. The result shows that the minimum number of faults varies from ten faults to twenty-three faults for different CR estimators.

1. Introduction

Software inspections help developers detect faults early and avoid rework. However, evidence indicates that the effectiveness of inspections varies widely. Project manager can use a reliable estimate of the number of faults to determine whether to reinspect. The capture-recapture (CR) method can be used during software inspections to estimate the number of faults in an artifact. During an inspection, each inspector finds (captures) some faults. If the same fault is captured by multiple inspectors then it has been re-captured. The total number of faults is estimated based on the unique faults found by each inspector and the overlap of the faults found by different inspectors [2].

While previous research has found this approach to be useful, the consensus is that the CR models generally underestimate [2]. Little research has focused on the effects of the two main inputs to the CR method, namely the *number of inspectors* and the *number of faults*. Therefore, it is important to understand the direct impact of each of these factors. In a previous paper, we analyzed the effect of the number of inspectors on the accuracy of the CR estimates independently [6]. In this paper, we examine the effect of the number of faults independently. The CR method uses four models (M_o, M_t, M_h, and M_{th}), each with a different set of assumptions. Each model has a set of mathematical estimators to perform the calculations. [2]. Estimators used in this study are: UMLE, CMLE, EE for the M_o and M_t models; SC and EE for the M_h and M_{th} models; and JK for the M_h model [2]. Due to space restrictions, detailed descriptions of models and estimators are omitted. For all estimators, the input data is a matrix with rows representing faults and columns representing inspectors. An entry in a matrix is 1 if the fault was found by the inspector and 0 otherwise.

2. The Study

To address the impact of the number of faults, the question driving this research is: *What is the minimum number of faults that have to be present in a document before the CR estimators can be used?* An answer to this question provides project managers with insight into the type of artifact for which CR models will be useful. The data was drawn from an inspection study conducted at Microsoft Research in which 73 participants individually inspected a generic requirements document describing a financial system. For use in previous studies, the document was seeded, by other researchers, with 30 realistic faults [1]. The resulting data was organized into a matrix with 73 columns and 30 rows.

To understand the impact of independently varying the number of faults, the number of inspectors had to be held constant. Based on our previous results [3], we chose the number of inspectors to be 18. Ten virtual inspection teams of size 18 were created by randomly choosing ten sets of 18 inspectors from the pool of 73, thereby creating 10 matrices with 18 columns. In each matrix, there were some rows that were zero because none of the 18 inspectors found that fault. Then, for each fault count (ranging from 1 to 30), ten virtual inspections were created by randomly selecting the

1071-9458 2008

DOI 10.1109/ISSRE.2008.61

number of rows (equal to the fault count) from each of the ten matrices just created. Therefore, 300 total virtual inspections were created, 10 for each fault count from 1 to 30.

Using these estimates, the estimators are evaluated using three metrics: accuracy, precision, and failure rate. The **accuracy** is measured as the relative error (RE) of an estimate. It is calculated as: *(Estimated number of faults – Actual number of faults)/Actual number of faults*. In this study, the *actual number of faults* was set equal to be the *fault count* being evaluated (1-30). For each fault count, the accuracy of the estimator was calculated as the median RE for the 10 samples. The accuracy is satisfactory when the RE is within 20% of the actual value [2, 3]. The **precision** of an estimator at each fault count is determined by the variability of the 10 RE estimates. The **failure rate** of an estimator was calculated as the number of times the estimator failed to produce an estimate.

3. Analysis and Results

As an overview of the results, Figure 1 shows the median RE (representing accuracy) for each CR estimator (represented by a line) for all fault counts. All CR estimators severely overestimate when the fault count is less than seven. Also, the accuracy of all the estimators improves as the fault count increases; with the improvement being more for some estimators than others. In practice, however, it is important for the estimate to be both precise and accurate.

An approach for augmenting the accuracy analysis with precision is to calculate three values for each fault count and estimator combination: a) the median estimate, b) the 75th percentile, and c) the 25th percentile. Together b) and c) define the interquartile range. Figure 2 shows this analysis for the M_0-UMLE estimator, with the median RE (solid line) appearing between the upper and lower bounds on the estimates (dotted lines). When analyzing the accuracy and precision together, a similar criterion was used as was used for accuracy alone. The only difference is that the RE at all three points (median, 75th percentile, and 25th percentile) is used. The results of the example in Figure 2 show that the estimator severely overestimates when the fault count is less than 10. Overall, the results showed that the EE estimator for M_t, M_h requires a minimum of 7 faults; the UMLE, CMLE, SC, and M_o-EE estimator require 10 or more faults; and the JK estimator requires 23 or more faults. Also, the EE and UMLE estimator for M_0 and M_t failed to produce an estimate with as many as 13 to 24 faults.

Based on these results, most of the estimators can produce satisfactory estimates on artifacts that contain 10 or more faults. However, the JK estimator can only be used for artifacts containing at least 23 faults. The EE estimator for M_h and M_t is not recommended because of its high failure rates. An important caveat of these results is that a team size of eighteen inspectors was used for the analysis, which is larger than normally used in software organizations. The next step is to study the interaction between the *inspectors* and *faults* by varying both of them simultaneously.

4. Acknowledgements

We thank the employees of Microsoft who participated in the study that provided the data.

5. References

[1] Carver, J., Nachiappan, N., and Page, A. "The Impact of Educational Background on Effectiveness of Requirements Inspections: An Empirical Study." To appear in *IEEE Transactions on Software Engineering*.

[2] Petersson, H., Thelin, T., Runeson, P., Wohlin, C. "Capture-Recapture in Software Inspections after 10 Years Research-Theory, Evaluation, and Application." *The Journal of Systems and Software*, 72(2):249-264.

[3] Walia, G. S., Carver. J., Nagappan, N. "The Effect of the Number of Inspectors on the Defect Estimates Produced by Capture-Recapture Estimators". In *Proceedings of the 30th International Conference on Software Engineering*, May, 10-18 2008. Leipzig, Germany. p. 331-340.

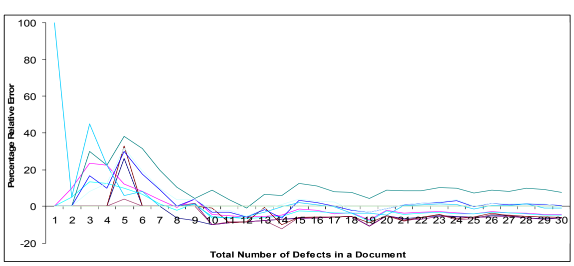

Figure 1. Median Relative Error as a function of actual fault count

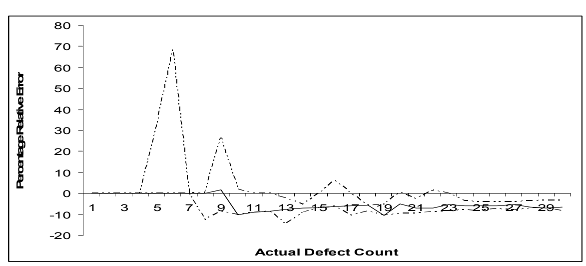

Figure 2. Variation of Mt-UMLE estimates as a function of the actual fault count