Estimação do número de erros de um *Software*: Erros fáceis e difíceis de detecção.

George Lucas Moraes Pezzott
Universidade Federal de São Carlos
Rodovia Washington Luís, km 235 - SP-310 São Carlos - São Paulo - Brasil
george.pezzott@gmail.com

José Galvão Leite
Instituto de Matemática e Estatística da Usp.
Rua do Matão, 1010, Cidade Universitária, Butantã São Paulo, SP – Brasil
leite@ufscar.br

Luis Ernesto Bueno Salasar
Universidade Federal de São Carlos
Rodovia Washington Luís, km 235 - SP-310 São Carlos - São Paulo - Brasil
luis.salasar@gmail.com

RESUMO

O processo de estimação do número de erros de um *software* é baseado no método de captura-recaptura múltipla, que consiste em examinar um *software*, em paralelo, por certo número de revisores. O modelo probabilístico considerado supõe revisores independentes e homogêneos e erros independentes e heterogêneos, e a probabilidade de detecção depende exclusivamente se o erro é de fácil ou difícil detecção. Sanathanan (1972) propõe um método via máxima verossimilhança em duas etapas: primeiramente estimam-se os parâmetros *nuisances* e, em seguida, o parâmetro de interesse *N* é estimado pela função de verossimilhança condicional considerando as estimativas dos parâmetros *nuisances* obtida na primeira etapa. Realizaremos a inferência sobre o parâmetro de interesse pela função de verossimilhança completa e pelo método proposto por Sanathanan. Os métodos de estimação serão aplicados a um conjunto de dados reais da literatura e comparados a partir de um estudo de simulação feito com auxílio do algoritmo EM.

PALAVARAS CHAVE: Método de captura-recaptura, Erros de software, algoritmo EM.

Área principal : Estimação de população.

ABSTRACT

The estimating process of the number of errors is a software is based on the method for multiple capture-recapture, comprehend by examine a software, in parallel, by the number of reviewers. The probabilistic model assumes that the reviewers are independent and homogeneous, the errors are heterogeneous and independent and probability of detection depends only if the error is easy or difficult to be detected. Sanathanan (1972) proposeed a method via maximum likelihood in two steps: first we estimate the parameters nuisances and then the parameter of interest N is estimated by considering the conditional likelihood function using the nuisances parameters estimates obtained in the first step. We will make the inference about the parameter of interest by the full likelihood function and by the method proposed by Sanathanan. The estimation methods will be applied in a real set of data from the literature and compared from a simulation study using the EM algorithm.

KEYWORDS: Capture-recapture method, Software Errors, EM algorithm

Main area: Population estimate.