Software Effort Estimation by Analogy and Regression Toward the Mean

In The Thirteenth International Conference on Software Engineering & Knowledge Engineering (SEKE01). -, Buenos Aires, Argentina

Abstract

Estimation by analogy is, simplified, the process of finding one or more projects that are similar to the one to be estimated and then derive the estimate from the values of these projects. If the selected projects have an unusual high or low productivity, then we should adjust the estimates toward productivity values of more average projects. The size of the adjustments depends on the expected accuracy of the estimation model. This paper evaluates one adjustment approach, based on the findings made by Sir Francis Galton in the late 1800s regarding the statistical phenomenon “regression toward the mean” (RTM). We evaluate this approach on several data sets and find indications that it improves the estimation accuracy. Surprisingly, current analogy based effort estimation models do not, as far as we know, include adjustments related to extreme analogues and inaccurate estimation models. An analysis of several industrial software development and maintenance projects indicates that the effort estimates provided by software professionals, i.e., expert estimates, to some extent are RTM-adjusted. A student experiment confirms this finding, but also indicates a rather large variance in how well the need for RTM-adjustments is understood among software developers.