

Software Adaptability Metrics Model Using Ordinary Logistic Regression

Udo, E. N.^{1*}, Akwukwuma, V. V. N.²

¹ Department of Computer Science, University of Uyo, Uyo, Akwa Ibom State, Nigeria.

² Department of Computer Science, University of Benin, Benin City, Nigeria.

*Corresponding Author; Tel: +2348023339501; Email: edwardudo@yahoo.com

Manuscript submitted December 10, 2018; accepted January 18, 2019

doi: 10.17706/jsw.14.3.116-128

ABSTRACT: Adaptability is an important software quality characteristic and a major non-functional requirement in software and should therefore be given adequate attention during software quality measurement and predictions especially now that the environment in which software products operate becomes highly unpredictable due to rapid changes in hardware platform as well as changes in the operating system requirements. In this work the results from the analysis of object-oriented software source code using our previously developed software analyser to measure the values of some internal properties using object-oriented software metrics based on formulation of decision rules in conjunction with binary logic combination of the possible internal software properties which aided in the prediction of adaptability level of a given software, is used as the dataset for ordinal logistic regression analysis. This analysis was used to formulate the adaptability model based on proportional odds assumption. The results showed that software with low coupling, inheritance and complexity were more likely to be adaptable than those with high values. Conversely, software with low cohesion were less likely to be adaptable than those with high cohesion. High cohesion was associated with adaptability since its odds ratio (low/high) was 7.97 (>1) while low coupling, inheritance and complexity were associated with adaptability since their odds ratios (low/high) were 0.15, 0.22 and 0.05 (<1) respectively. The resulted model fitted the data well and the estimated cumulative odds were the same across all the ordinal categories, thus the proportional odds assumption held.

Key words: Metrics, model, ordinal logistic regression, proportional odds, software adaptability.

1. Introduction

Adaptability is the degree of ease a system shows in being modified to meet a changed or deleted requirement. It therefore characterizes the ability of a system to still function properly in an event of change to new specifications. Adaptability is an important software quality characteristic and a major non-functional requirement in software [1]. This requirement should therefore be given adequate attention during software quality measurement and predictions especially now that the environment in which software products operate becomes highly unpredictable due to rapid changes in hardware platform as well as changes in the operating system requirements. Adaptable software is software that will still perform as expected in the face of any technological or operating environment changes. That is software that is evolving and scalable.

As time elapses, the original environment (CPU, operating system, business rules, external product characteristics, etc) for which the software was developed is likely to change [2]. This environmental change is mostly due to changes in technology, organizational structure, human perception and needs [3]. Other changes may be as a result of changing customer requirements, need for faster development of new software, adding of new software features and fixing software defects during the maintenance phase of software life cycle.

If a software is designed without bearing adaptability in mind, the software will degrade and malfunction in the face of continuous software requirements changes and this will result in losses as most organizations and establishments are relying heavily on software for the day to day running of their enterprises and discharging of their corporate functions. Therefore, measurement of software adaptability should be a major concern to software developers and other software engineering community.

One of the main objectives of research in metrics is to formulate a relation that will establish a relationship between the internal and the external attributes. This relation is called a model or Link.

One of the approaches to formulate a model is the use of statistical techniques such as regression, which is a statistical procedure that attempts to predict values of a given variable (dependent variable) based on the values of one or more other variables (independent variables). The result of a regression is usually an equation (or model) which summarizes the relationship between dependent variable and independent variable(s).

In this work, adaptability metrics model is formulated using ordinal logistic regression: analyses of the relationship between multiple independent variables and ordinal dependent variable, to show the relationship between internal properties of object-oriented software and adaptability. The attributes and the relationship involved in the metrics model can be adjusted and improved by software developers and users.

As early as 1977, [4] brought up a quality model that has three quality of software product: product transition (adaptability to new environment), product revision (ability to undergo changes), and product operations (its operation characteristics). Product revision includes Maintainability, Flexibility, and Testability. Product Transition includes Portability, Reusability, and Interoperability. Product operation includes Reliability, Efficiency, Integrity and Usability.

Boehm quality model also identified portability as a key quality factor [5]; FURPS model by Grady [6] also identified portability as a key software quality factor; Dromey quality model in [7] also identified portability as a criteria for software quality model; Reference [8] described software quality in terms of six quality attributes and identified adaptability as a sub-characteristic of portability; [9] also included adaptability as a key quality factor.

Reference [10] introduced software adaptability as a quality factor into McCall quality model of 1977. They stated that adaptability makes software platform independent and increases the capability of software to adapt to changes.

Reference [11] reviewed the existing quality models and showed that all the models contain portability or adaptability as either a quality factor or criteria. This informs our decision to develop a model to predict adaptability in object-oriented software.

There are a lot of software metrics in literature that measures several quality factors like maintainability, complexity, reliability etc using different methods including statistical methods at different phases of the software lifecycle, *yet there are no metrics to measure adaptability at the coding phase of the software lifecycle using statistical methods.*

The aim of this work therefore is to:

formulate adaptability model using statistical techniques.

investigate the relationship between internal object-oriented software properties and adaptability using the values from the statistical analysis.

evaluate the formulated model.

2. Methodology

- 1) Analysis of source codes written in java using a software analyzer to extract the values of coupling, cohesion, inheritance and complexity using object oriented design metrics suite (CBO, RFC, LCOM1,DIT, NOC and WMC) of [12] and NOM/LCOM2 metrics by [13]. Values of these object-oriented design metrics were matched against the estimated threshold values proposed by [14]–[17]. Metrics values above these threshold values were considered as high (1) while those below were considered as low(0).
- 2) Rules were formulated using the threshold values range and the desirable level of the internal software properties: coupling (low), cohesion (high), inheritance (low) and complexity (low) to determine the attributes level of given software. Low were depicted as **0** while high is depicted as **1**. Attributes levels were computed for 200 different open source object oriented codes written in java. Binary logic combinations of these attributes level were performed to determine the adaptability level.
- 3) Statistical analysis using SPSS 17.0 package was carried out on the dataset as follows:
 - (a) **Ordinal Logistic Regression (OLR) Analysis** - The output from the software analyser served as input into the statistical package for the analysis of the 40 object-oriented software source codes. Adaptability was the dependent variable with 3 ordered categories: Poorly Adaptable, Fairly Adaptable and Adaptable coded as 1, 2 and 3 respectively. The independent variables were coupling, cohesion, inheritance and complexity each coded as Low (0) or High (1). OLR was used to analyse the dataset since the outcome variable was ordinal in nature and the explanatory variables were categorical. The ordinal logistic regression analysis was based on the proportional odds assumption [18]–[20]:

$$\ln \left[\frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J} \right] = \alpha_j + \beta_1 x_1 + \dots + \beta_p, \quad j = 1, \dots, J - 1 \quad (1)$$

where π_j = the probability of being at or below category j given a set of predictors; α_j = the intercept equivalent term; β_1, β_p = the logit coefficients (the log-odds ratio of the event) and x_1, x_p = the explanatory variables.

This model estimates the relationship between a set of predictor variables and an outcome variable using a logit link function. It also estimates the cumulative odds of being at or below a particular level of the response variable.

The frequency distribution of adaptability levels was carried out on the dataset. Bivariate crosstabs analysis was also conducted on the dataset to show the pattern that existed on the data and the level of association between software properties and adaptability. Values of several statistics for measuring the strength of association between the dependent variable and the predictor variables were also inspected. Multicollinearity check was conducted to examine the interrelationship among the predictor variables.

- (b) **Calculation of Odds Ratios (OR)** - Odds ratios for the four explanatory variables (coupling, cohesion, inheritance and complexity) were calculated using estimated response probabilities tables, one of the outputs from ordinal logistic regression analysis.
- (c) **Formulation of Adaptability Model** - Based on the proportional odds assumption, adaptability model was formulated, **which** estimated the cumulative odds of being at or below a particular level of the response variable. For each predictor variable, the estimated cumulative odds were the same

across all the ordinal categories.

The authors carried out the first two aspects of the objectives and validated the results which were published in [21], [22]. This work models software adaptability based on the proportional odds assumption using ordinary logistic regression.

3. Ordinal Logistic Regression Analysis

In order to effectively carry out ordinal logistic regression analysis, data to be used must pass these four assumptions [23]:

- 1) Dependent variable should be measured at ordinal level.
- 2) One or more independent variables should either be continuous, ordinal or categorical (including dichotomous variables)
- 3) There should be no multicollinearity, which is a state in which two or more predictor variables are highly correlated with each other. If multicollinearity is present in data, the statistical inference made about the data may not be reliable as the predictors are explaining almost the same variability in the outcome.
- 4) There should be proportional odds, which mean that each independent variable has an identical effect at each cumulative split of the ordinal dependent variable.

The data used in this research work has passed the four assumptions, which are highlighted accordingly hereunder:

- (a) The **dependent** variable was Adaptability with the following ordered categories: Poorly Adaptable – 1; Fairly Adaptable – 2; Adaptable – 3.
- (b) There were four (4) independent variables with each having 2 categories:
 - i. Coupling – Low (0), High (1)
 - ii. Cohesion – Low (0), High (1)
 - iii. Inheritance – Low (0), High (1)
 - iv. Complexity – Low (0), High (1)
- (c) To test for multicollinearity among the predictor variables, a tolerance statistics was computed for the independent variables. This was done by running a linear regression on the independent variables and checked the results of the “collinearity statistics” in SPSS. Values of Tolerance and Variance Inflator Factor (VIF) from the resulting output were used to test multicollinearity. A tolerance of less than 0.20 and a VIF of more than 5 indicate a multicollinearity problem.

The independent variables in this work were tested for multicollinearity as indicated in (c) above and obtained the result displayed in Table 1.

Table 1. Multicollinearity Checks

| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
|--------------|-----------------------------|------------|---------------------------|---------|------|-------------------------|-------|
| | B | Std. Error | Beta | | | Tolerance | VIF |
| 1 (Constant) | 2.743 | .042 | | 64.638 | .000 | | |
| Coupling | -.723 | .038 | -.434 | -19.086 | .000 | .991 | 1.009 |
| Cohesion | .725 | .039 | .434 | 18.665 | .000 | .945 | 1.058 |
| Inheritance | -.678 | .038 | -.408 | -17.749 | .000 | .967 | 1.034 |
| Complexity | -.820 | .039 | -.493 | -21.173 | .000 | .942 | 1.061 |

The results showed that there was no multicollinearity among the predictor variables as the value of Tolerance for all the predictor variables was greater than 0.2 and that of VIF was less than 5. For instance,

the tolerance and VIF for coupling was 0.991 and 1.009 respectively.

(d) The model developed in this study is based on the proportional odds assumption.

4. Fitting Ordinal Logistic Model

Reference [18]) stated that ordinal regression models are specialized cases of the general linear model:

$$Y = Xb + e \tag{2}$$

Logistic Regression is part of a family of models called the generalized linear model. The main feature of this model is that instead of using y directly, it is modeled through what is called a “Link” function:

$$G(y) = Xb + e \tag{3}$$

To fit a binary logistic regression model, one estimates a set of regression coefficients that predict the probability of the outcome of interest:

$$\ln \left[\frac{\text{prob}(\text{even})}{1-\text{prob}(\text{event})} \right] + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \tag{4}$$

The quantity on the left hand side is called the logit (log of the odds that an event occurs). The coefficients in the logistic regression model tell how much the logit changes based on the values of the predictor variables.

Binary logistic regression can be modified to incorporate the ordinal nature of a dependent variable by defining the probabilities differently instead of considering the probability of an individual event, the probability of that event and all events that are ordered before it are considered.

For dependent variable with three categories, one can model the following odds:

$$\theta_1 = \text{prob}(\text{score of } 1) / \text{prob}(\text{score} > 1)$$

$$\theta_2 = \text{prob}(\text{score of } 1 \text{ or } 2) / \text{prob}(\text{score} > 2)$$

The odds are in the form:

$$\theta_j = \text{prob}(\text{score} \leq j) / \text{prob}(\text{score} > j) \tag{5}$$

“Equation (5)” can be written as:

$$\theta_j = \text{prob}(\text{score} \leq j) / (1 - \text{prob}(\text{score} \leq j)) \tag{6}$$

Since the probability of a score greater than j is 1 minus the probability of a score less than or equals to j . Suppose that response Y has J categories and the probability for category i is given by [20]:

$$P(Y = i) = \pi_i \text{ for } i = 1, \dots, J \tag{7}$$

Consider explanatory variables x_1, \dots, x_p . Sometimes, there may be a latent continuous variable (a variable which is not directly observable and is assumed to affect the response variable) Y for which the cutpoints C_1, \dots, C_{J-1} define J ordinal categories with associated probabilities π_1, \dots, π_j . A cumulative probability for Y is

the probability that Y falls at or below a particular point. For outcome category j, the cumulative probability is:

$$P(Y \leq j) = \pi_1 + \dots + \pi_j, j = 1, \dots, J \quad (8)$$

where $P(Y \leq 1) \leq P(Y \leq 2) \leq \dots \leq (P(Y \leq j)) = 1$

The logit of the cumulative probabilities, called cumulative logit is:

$$\begin{aligned} \text{logit}[P(Y \leq j)] &= \ln \left[\frac{P(Y \leq j)}{1 - P(Y \leq j)} \right] \\ &= \ln \left[\frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J} \right], j = 1, \dots, J - 1 \end{aligned} \quad (9)$$

The cumulative logit model is given by:

$$\text{logit}(P(Y \leq j)) = \ln \left[\frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J} \right] = \alpha_j + \beta_{j1}x_1 + \dots + \beta_{jp}x_p, j = 1, \dots, J - 1 \quad (10)$$

If the intercepts α_j depends on the category j, but the other regression coefficients for explanatory variables do not depend on j, then the model is:

$$\ln \left[\frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J} \right] = \alpha_j + \beta_1x_1 + \dots + \beta_px_p, j = 1, \dots, J - 1 \quad (11)$$

“Equation (11) is called the proportional odds model”, which is widely used as the default for ordinal regression analysis in SPSS. This model estimates the relationship between a set of predictor variables and an outcome variable using a logit link function. It is also known as the cumulative logit model because it estimates the commulative odds of being at or below a particular level of the response variable. In addition, for each predictor variable, the estimated cumulative odds are assumed to be the same across all the ordinal categories. This assumption is called proportional odds assumption.

According to [19], the ordinal regression model can be expressed on the logit scale as follows:

$$\ln(Y'_j = \text{logit} \left[\frac{\pi_j(x)}{1 - \pi_j(x)} \right] = \alpha_j + (-\beta_1x_1 - \beta_2x_2, -\dots, -\beta_px_p) \quad (12)$$

where $\pi_j(x) = \pi(Y \leq j | x_1, x_2, \dots, x_p)$ is the probability of being at low or below category j, given a set of predictors $j = 1, 2, \dots, j-1$. α_j are the cutpoints and $\beta_1, \beta_2, \dots, \beta_p$ are the logit coefficients. This proportional odds model estimates different cut points, but the effect of any predictor is assumed to be the same across these cut points therefore for each predictor, only one logit coefficient is estimated.

5. Dataset Description

Frequency distribution of adaptability level of the analysed software was carried out. The result showed that 60.0% of the software were adaptable, 22.5% were fairly adaptable and 17.5% were poorly adaptable. A simple bivariate cross tabulations analysis was also done on the datasets, which shows that the data

showed a strong association between coupling, cohesion, inheritance and complexity with adaptability.

To show variables interrelationship, ordinal logistic regression analysis was conducted for adaptability. The pseudo-R² values were inspected to see the predictive power of the model. Fig. 1 shows the values of several pseudo-R² that can be used to measure strength of association between the dependent variable and the predictor variables. The values of these pseudo-R² statistics were high, indicating that the model predicted the outcome for individual cases.

| | |
|---------------|-------|
| Cox and Snell | .850 |
| Nagelkerke | 1.000 |
| McFadden | 1.000 |

Link function: Logit.

Fig. 1. Pseudo R-square.

One of the outputs used in building a model in SPSS is the parameter estimates table which contains the estimated coefficients for the model. Table 2 shows the parameter estimates for the dataset. Thresholds are the intercepts equivalent terms (the α_j 's) which are used in the calculations of predicted probabilities. The estimates labeled locations are the coefficients for the predictor variables.

Table 2. Parameters Estimates for 200 Datasets

| | | Estimate | Std. Error | Wald | df | Sig. | 95% Confidence Interval | |
|-----------|--------------------|----------------|------------|--------|----|------|-------------------------|-------------|
| | | | | | | | Lower Bound | Upper Bound |
| Threshold | [Adaptability = 1] | 7.973 | 3.716 | 4.604 | 1 | .032 | .690 | 15.257 |
| | [Adaptability = 2] | 21.214 | 6.406 | 10.965 | 1 | .001 | 8.658 | 33.770 |
| Location | [Coupling=0] | 12.798 | 4.129 | 9.609 | 1 | .002 | 4.706 | 20.890 |
| | [Coupling=1] | 0 ^a | . | . | 0 | . | . | . |
| | [Cohesion=0] | -16.893 | 5.503 | 9.423 | 1 | .002 | -27.679 | -6.107 |
| | [Cohesion=1] | 0 ^a | . | . | 0 | . | . | . |
| | [Inheritance=0] | 13.358 | 4.440 | 9.053 | 1 | .003 | 4.657 | 22.060 |
| | [Inheritance=1] | 0 ^a | . | . | 0 | . | . | . |
| | [Complexity=0] | 16.781 | 5.298 | 10.032 | 1 | .002 | 6.397 | 27.165 |
| | [Complexity=1] | 0 ^a | . | . | 0 | . | . | . |

Link function: Logit.

a. This parameter is set to zero because it is redundant.

By default, the inbuilt model for Ordinal Logistic Regression (OLR) analysis is the proportional odds model with a null hypothesis that the model without predictors is as good as a model with predictors. By examining the adequacy of the values and the proportional odds assumption, it was found from the model fitting information that the difference between the two log-likelihoods, which is the chi-square, has an observed significance level of less than 0.005. This means that the null hypothesis that the model without predictors is as good as the model with predictors can be rejected.

However, the goodness-of-fit statistics in Table 3 suggests that the model with predictors fits the data well since the observed and the expected cell counts are similar; the value of each statistics is small and the observed significance level is large.

Table 3. Goodness of Fit for Model

| | Chi-Square | Df | Sig. |
|----------|------------|----|-------|
| Pearson | .019 | 10 | 1.000 |
| Deviance | .039 | 10 | 1.000 |

Link function: Logit.

There is no sufficient evidence to reject the proportional odds assumption that the regression coefficients are the same for all three categories of adaptability, since the observed significance level in the test of parallel lines is large (> .005).

6. Calculation Of Odds Ratios For Predictor Variables

Coupling – Analysis is carried out with coupling alone as a factor. The parameter estimate table for the analysis is shown in Table 4.

Table 4. Parameter Estimates Using Only Coupling

| | Estimate | Std. Error | Wald | df | Sig. | 95% Confidence Interval | |
|------------------------------|----------------|------------|--------|----|------|-------------------------|-------------|
| | | | | | | Lower Bound | Upper Bound |
| Threshold [Adaptability = 1] | .055 | .207 | .070 | 1 | .792 | -.351 | .460 |
| [Adaptability = 2] | 1.612 | .245 | 43.374 | 1 | .000 | 1.133 | 2.092 |
| Location [Coupling=0] | 1.890 | .293 | 41.700 | 1 | .000 | 1.316 | 2.463 |
| [Coupling=1] | 0 ^a | . | . | 0 | . | . | . |

Link function: Logit.

a. This parameter is set to zero because it is redundant.

The ratio of odds for lower to higher scores for software having low coupling and those having high coupling is $e^{-\beta}$ [Exp (-1.890)] which is 0.15. This ratio is the same over all the adaptability levels. To find out whether the model fits with coupling, a standard statistical method is to compare the observed values with the expected values.

In calculating the expected values, the values of the coefficients, from the estimate column in Table 4 are used to calculate the cumulative predicted probabilities from the logistic model for each case. The logistic model as given by [24] is written out in “(13)”:

$$Prob(event j) = 1/(1 + e^{-(\alpha_j - \beta x)}) \tag{13}$$

where j is the ordinal category of the dependent variable, α is the intercept equivalent term and β is the coefficient of the independent variable x.

The logistic model parameters (α and β) are inputted into the software written to calculate the predicted probabilities, expected probabilities and odds ratios. Fig. 2 shows the user interface for inputting the logistic parameters

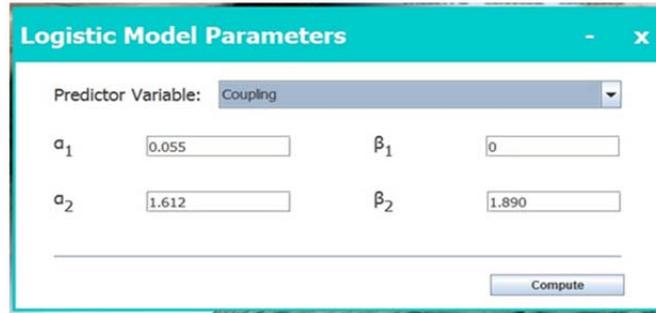


Fig. 2. User interface to input logistic parameters.

For software with high coupling, $\alpha = 0.055$ and 1.612 , $\beta = 0$ (values taken from estimate column of Table 4), the predicted probabilities are computed as:

| Score | Probability |
|---------------|--------------------|
| Prob(score 1) | 0.513746534902355 |
| Prob(score 2) | 0.8336888749911537 |
| Prob(score 3) | 1.0 |

The values of the cumulative predicted probabilities are used to calculate the estimated probability of the individual scores for software with high coupling. The probability for the individual scores is calculated using the formular:

$$\text{Prob}(\text{score} = j) = \text{prob}(\text{score less than or equal to } j) - \text{prob}(\text{score less than } j)$$

The result is computed from the software thus:

| Score | Probability |
|-----------------|---------------------|
| Prob(score = 1) | 0.513746534902355 |
| Prob(score = 2) | 0.31994234008879874 |
| Prob(score = 3) | 0.16631112500884626 |

Probability of software with low coupling is calculated using the values from estimate column of Table 4, where $\alpha = 0.055$ and 1.612 ; $\beta = 1.890$.

The result is displayed thus:

| Score | Probability |
|-------------------------|---------------------|
| Prob(score 1) | 0.13764370760149627 |
| Prob(score 1 or 2) | 0.4309441707529759 |
| Prob(score 1 or 2 or 3) | 1.0 |

Using the estimated response probabilities shown in Table 5, the odds ratios for the adaptability levels are calculated.

Table 5. Estimated Response Probabilities for Coupling

| Coupling | | Estimated Cell Probability for Response Category: 1 | Estimated Cell Probability for Response Category: 2 | Estimated Cell Probability for Response Category: 3 |
|----------|----------------|---|---|---|
| Low | Mean | .1377 | .2935 | .5688 |
| | N | 111 | 111 | 111 |
| | Std. Deviation | .00000 | .00000 | .00000 |
| High | Mean | .5136 | .3201 | .1662 |
| | N | 89 | 89 | 89 |
| | Std. Deviation | .00000 | .00000 | .00000 |
| Total | Mean | .3050 | .3054 | .3897 |
| | N | 200 | 200 | 200 |
| | Std. Deviation | .18732 | .01325 | .20057 |

The user interface to input the values from the estimated response probabilities table for the three categories of adaptability for both low and high levels for the calculation of the odds ratios is shown in Fig. 3.

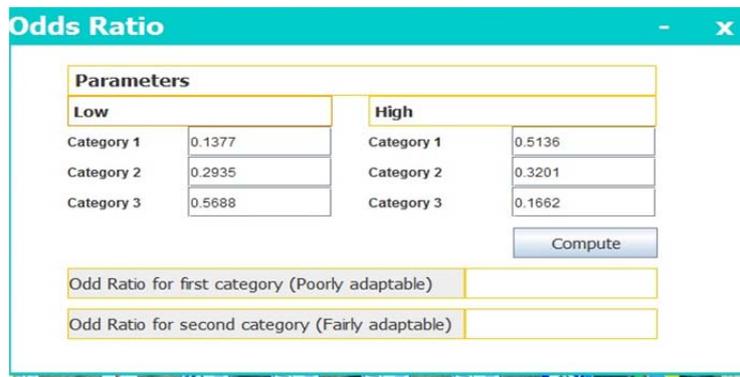
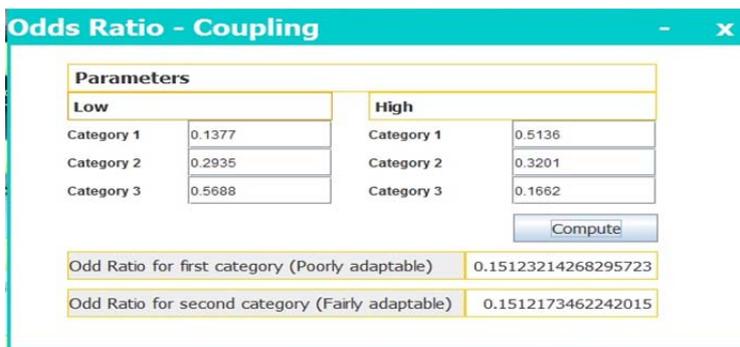


Fig. 3 – User Interface to input Parameters to Calculate Odds Ratios. The odds ratios for the first and the second categories is computed and displayed thus:



The same steps are followed to calculate odds ratio the remaining three (3) predictor variables: Cohesion, Inheritance and Complexity.

The results of all the calculations are summarized in Table 6 showing the values of $e^{-\beta}$, cumulative expected probabilities for high level, estimated probabilities for low level, probability of software with low level and the odds ratios for each category.

Table 6. Summary of Results

| | | COUPLING | COHESION | INHERITANCE | COMPLEXITY |
|--|------------------|----------|----------|-------------|------------|
| | $e^{-\beta}$ | 0.15 | 7.97 | 0.22 | 0.05 |
| CUMULATIVE PREDICTED PROBABILITY FOR HIGH LEVEL | P(S=1) | 0.51 | 0.12 | 0.44 | 0.58 |
| | P(S=2) | 0.83 | 0.41 | 0.77 | 0.91 |
| | P(S=3) | 1 | 1 | 1 | 1 |
| ESTIMATED PROBABILITY FOR LOW LEVEL | P(S=1) | 0.14 | 0.53 | 0.15 | 0.07 |
| | P(S=2) | 0.29 | 0.32 | 0.27 | 0.27 |
| | P(S=3) | 0.57 | 0.15 | 0.58 | 0.66 |
| ESTIMATED PROBABILITY FOR HIGH LEVEL | P(S=1) | 0.51 | 0.12 | 0.44 | 0.58 |
| | P(S=2) | 0.32 | 0.29 | 0.33 | 0.34 |
| | P(S=3) | 0.17 | 0.59 | 0.23 | 0.09 |
| PROBABILITY OF SOFTWARE WITH LOW VALUE | P(S=1) | 0.14 | 0.52 | 0.14 | 0.06 |
| | P(S=1 or 2) | 0.43 | 0.85 | 0.42 | 0.34 |
| | P(S=1 or 2 or 3) | 1 | 1 | 1 | 1 |
| ODDS RATIO (LOW/HIGH) | CATEGORY 1 | 0.15 | 7.97 | 0.22 | 0.05 |
| | CATEGORY 2 | 0.15 | 7.97 | 0.22 | 0.05 |
| | CATEGORY 3 | 1 | 1 | 1 | 1 |
| ODDS RATIO (HIGH/LOW) | CATEGORY 1 | 6.6 | 0.13 | 4.59 | 19.4 |
| | CATEGORY 2 | 6.6 | 0.13 | 4.59 | 19.4 |
| | CATEGORY 3 | 1 | 1 | 1 | 1 |

7. Findings

The sum of the expected (estimated) cumulative probabilities for all the independent variables for both the low level and high level are 1. The odds ratios for both low/high and high/low are constant across the possible cut points of the outcome. High cohesion is associated with adaptability because the odd ratio (low/high) is greater than 1.

Low coupling, inheritance and complexity are associated with adaptability because their odd ratios (low/high) are less than 1. All the odd ratios are the same with $e^{-\beta}$ across the different adaptability levels which is the [Exp(-1.890)], [Exp(2.076)], [Exp(-1.536)] and [Exp(2.991)] respectively.

The adaptability model is formulated using the proportional odds model in "(11)":

$$\ln \left[\frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J} \right] = \alpha_j + \beta_1 x_1 + \dots + \beta_p x_p, \quad j = 1, \dots, J - 1$$

where β 's are the odds ratios; the coefficient of the predictors variables, x_1, x_p . The predictor variables (independent variables) are coupling, cohesion, inheritance and complexity with odds ratios of 0.15, 0.13, 0.22 and 0.05 respectively. The odds ratios of low/high are used for coupling, inheritance and complexity since low values of these properties are desirable for adaptability. The odds ratio of high/low is used for cohesion since high value of cohesion is desirable for adaptability.

The proposed adaptability model is therefore given by:

$$P = \alpha + 0.15\text{COU} + 0.13\text{COH} + 0.22\text{INH} + 0.05\text{COM} \quad (14)$$

where P is the cumulative estimated probability since ordinal regression does not model the probability of an individual event but the probability of that event and all others above it in the ordinal ranking. α is the intercept equivalent term.

Coupling, Cohesion, Inheritance and Complexity are the predictors which are respectively shortened to COU, COH, INH and COM and can take values of 0 or 1.

The model based on the proportional odds assumption is valid since the effects of the explanatory variables are the same across the different threshold and the significance level in test of parallel lines was $> .005$

References

- [1] Subramanian, N., & Chung, L. (2001). Software architecture adaptability: An NFR approach. *Proceedings of the 4th International Workshop on Principles of Software Evolution, Vienna, Austria* (pp. 52 – 61).
- [2] Pressman, R. S. (2005). *Software engineering – A practitioner's Approach*. 4th Edition, McGraw-Hill Companies Inc, New York.
- [3] Chung, L., & Subramanian, N. (2001). Process-oriented metrics for software architecture adaptability. *Proceedings of 5th IEEE International Symposium Requirements Engineering* (pp. 310 -311).
- [4] McCall, J. A., Richards, P. K., & Walters, G. F. (1977). Factors in software quality. RADC TR-77-369, Rome Air Development Center, Rome
- [5] Boehm, B. W., Brow, J. R., Lipow, M., McLeod, G., & Merritt, M. (1978). Characteristics of software quality. North Holland Publishing. Amsterdam, the Netherlands.
- [6] Grady, R. B. (1992). *Practical software metrics for project management and process improvement*. Prentice Hall, New Jersey, USA.
- [7] Dromey R. G. (1995). A model for software product quality. *IEEE Transaction on Software Engineering*. 21, 146-162.
- [8] ISO/IEC (2001). *Software Engineering Product Quality – Part 1: Quality Model*. International Standard Organization/International Electrotechnical Commission, Technical Report.
- [9] ISO/IEC 25010 (2011): *Software engineering– system and software quality requirement and evaluation*

(SQuaRE). *System and Software Quality Model*.

- [10] Chatterjee, D., Orisa, B., & Malu, R. (2013). A quality model for adaptability. *International Journal of Computer Science and Engineering Technology*, 4(12); 1466 – 1469
- [11] Lee, M. (2014). Software quality factors and software quality metrics to enhance software quality assurance. *British Journal of Applied Science and Technology*, 4(21). 3069 – 3095
- [12] Chidamber, S., & Kemerer, C. (1994). A metrics suite for object-oriented design. *IEEE Transactions on Software Engineering*, 20(6), 476 – 493.
- [13] Bansiya, J., & Davis, C. G. (2002). A hierarchical model for object-oriented design quality assessment. *IEEE Transactions on Software Engineering*, 28(1), 4 - 17.
- [14] Rosenberg, L., Hammer, T., & Shaw, J. (1998). Software metrics and reliability. *Proceedings of the 9th International Symposium on Software Reliability Engineering*.
- [15] Benlarbi, S., El Emam, K., Goel, N., & Rai, S. (2000). Threshold for object-oriented measures. *Proceeding of the 11th International Symposium on Software Reliability Engineering*.
- [16] Chandra, E., & Linda, P. (2010). Class break point determination using CK metrics thresholds. *Global Journal of Computer Science and Technology*, 10(14), 73 – 77.
- [17] Herbold, S., Grabowski, J., & Waack, S. (2010): Calculation of optimization of thresholds for sets of software metrics.
- [18] Norusis, M. (2008). *The SPSS 16.0 Guide to Data Analysis*. Prentice Hall Press, New Jersey, USA.
- [19] Liu, X. (2009). Ordinal regression analysis: Fitting the proportional odds model using Stata, SAS and SPSS. *Journal of Modern Applied Statistical Method*, 8(2), 632 – 645.
- [20] Citko, D., Milewska, A., Wasilewska, J., & Kaczmarek, M. (2012). Ordinal logistic regression for the analysis of skin test reactivity to common aeroallergen. *Studies in Logic, Grammar and Rhetoric, Medical University of Bialystok, Poland*, 29(42), 87 – 98.
- [21] Udo, E, N., & Akwukwuma, V. V. N. (2015). Object-oriented software quality assessment using metrics threshold values. *World Journal of Applied Science and Technology*, 7(2), 191 – 199.
- [22] Akwukwuma, V. V. N., & Udo, E, N. (2015). Predicting adaptability level of object-oriented software using metrics and threshold values. *Pacific Journal of Science and Technology*, 16(2), 124-134.
- [23] Ordinal Logistic Regression Analysis – URL: https://statistics.laerd.com/spss_tutorial (last visited, December 2018)
- [24] Park, H. (2013). An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*, 43(2), 154 – 164.

Edward N. Udo is a lecturer in the Department of Computer Science, University of Uyo, Uyo, Nigeria. His main research interests include are among others, object-oriented software engineering, software metrics, and measurement. He is a member of NCS, CPN, and IEEE.