International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

DEVELOPMENT OF A DECISION TOOL FOR COST JUSTIFICATION OF SOFTWARE USABILITY

Burchan Aydin, Ph.D.[1] and Dr. Mario G. Beruvides, Ph.D., P. E.[2]
Texas Tech University, Department of Industrial Engineering[1,2]
Email: burchan.aydin@ttu.edu [1] , mario.beruvides@ttu.edu [2]

Abstract:

Cost justification of software usability has been criticized to be subjective and in some cases biased. This study is aimed at providing scientific basis to be used in usability cost justification models. Two decision tools were developed that help usability practitioners and/or engineer managers iteratively determine expected benefits of usability. Pareto analysis, interaction modeling, distribution fitting, bootstrap resampling, quantile regression, goal programming optimization modeling, and sensitivity analysis were used to develop these decision tools. The benefits of usability incorporated into the decision tools were increased sales, increased traffic of the web sites, reduced error rate of users, and decreasing task time.

Keywords : Usability, Cost Justification of Usability

INTRODUCTION

Usability is often described as 'ease of use'. However, it is definitely well beyond this short definition [22]. It is defined in ISO 9241-11 guidelines as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (1998, p. 2) [13]. Thus, according to the ISO standards, the three main factors that should be considered in order to design for usability are: effectiveness, efficiency, and user satisfaction. Effectiveness is "the accuracy and completeness with which users achieve specified goals", whereas efficiency is "the use of resources in relation to the accuracy and completeness with which users achieve goals" (ISO 9241-11, 1998, p. 2). Common effectiveness metrics are task completion rates and error rates. On the other hand, efficiency is usually measured by time on task and task flow deviation. Task flow deviation is "the ratio of the optimal number of steps to complete a task to the average number of steps to complete the task" [24]. A definition in terms of efficiency factor is provided by Porat and Tractinsky (2012) as: "Usability work is a focus on the design and evaluation of activities in order to aid users accomplish the intended tasks efficiently" [21]. User satisfaction is "the freedom from discomfort, and positive attitudes towards the use of the product" [13].

In the software industry, usability is a measure of how well a software program facilitates user learning, helps users remember what they have learned, reduces error rates, increases efficiency and effectiveness, and how much users are satisfied with the software [7]. Usability not only benefits the end users of the software, but also the developers, the companies they work for, and the vendors [8]. Marcus (2005) listed the return on investment of usability for software internally used as: (1) increased user productivity, (2) decreased user errors, (3) decreased training costs, (4) decreased development costs, and (5) decreased user support costs.

ISSN 2304-0777 www.jitbm.com

Moreover, returns on investment of usability for externally sold software are listed as: (1) increased sales, (2) decreased customer support costs, (3) savings gained from making changes earlier in the development life cycle, and (4) reduced cost of training offered through vendor company [18].

Investment costs are inevitable. The key point is to validate the investment with the profits gained. According to survey results [11], 62% of people are convinced of the value of user centered design by testimonials, case stories, and demonstrated results, 23% by support from customers, management, and outside companies, and only 15% by cost-benefit analysis. Even though survey results showed that cost-benefit analysis was not a primary method to influence people of the value of user centered design, 54% of the participants in the same survey believed that cost-benefit analysis would be convincing if in-depth information about how the model worked and the assumptions and data used would be provided (Harrison et. al., 1994, p. 233) [11]. Even though it is not a primary convincing factor, cost-benefit analysis is the main tool used by usability researchers to cost-justify usability in the literature. Management looks for quantified cost and benefit figures for all investment alternatives. “The cost-benefit analysis is a method of analyzing projects for investment justification [14]”. The method has the following steps [20]:

- Identification of relevant costs and benefits,
- Measurement of costs and benefits,
- Comparison of cost and benefit streams accruing during the lifetime of the project,
- Project selection

Donahue (2001) listed the steps of cost-benefit analysis for usability as (p. 1): “Selecting a usability technique, determining the proper unit of measurement, making a rational assumption about the benefit’s magnitude, and translating the estimated benefit into a monetary figure. Considering these steps, it can be stated that cost-benefit analysis is performed to cost justify usability by estimating the costs and benefits of usability activities and comparing against the costs of not performing any usability activities” [8].

Justification of usability investment has been a major part of the usability research domain during dot.com era, yet there has been a declining trend in usability justification model development lately [1]. Majority of the models in cost justification of usability domain were gathered in a book by Bias and Mayhew in 1994 [3]. After that there were a few primary theories, yet most of the establishments were modification of those primary models to different business cases such as organization type. This research study is a much needed re-focus on the cost justification of usability. Existing cost justification models are based on many assumptions. Currently, there is no study that provides reliable numerical ranges for these assumptions. However, it is crucial that usability practitioners show reliable numbers to decision makers to promote usability investments. How much increase or decrease may happen in any performance measure after usability investments is not apparent. Therefore, it is assigned to the usability practitioner to make assumptions on the magnitude of the usability benefits. The accuracy of these usability benefit expectations depend mainly on the consistency of usability practitioner or researcher. Consistency in usability testing endeavors is a major problem as reported in Vermeeren et. al., 2008 [29]. For instance, a variation of 37% to 72% has been detected on within-team consistency, which indicates considerable subjectivity [29].This paper provides two decision tools for usability practitioners to determine expected changes in sales, traffic rate, error rates of users and task times. These expected changes shall further be converted into monetary figures for justifying usability investments by using cost justification theories.

METHODOLOGY

This paper involves three stages of methods to develop decision tools for cost justification of usability. The first stage provides target values and constraints for the goal programming models that are developed in third stage, while the second stage is to test validity of these constraints. Figure-1 demonstrates the basis of the three-stage methodology of this paper.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

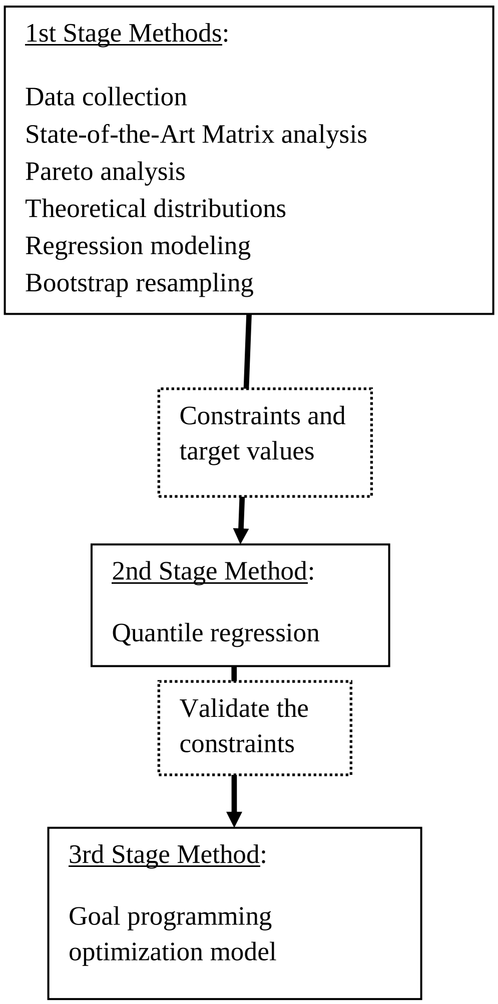

Figure 1- Methodology

Data Collection

In this research, a quantitative research methodology was used and data collection was held through a complete literature review. Types of data are percentages. The steps of the methodology used for data collection can be listed as follows.

1. A fairly extensive literature search for relevant studies was conducted through professional journals, using online databases to identify studies addressing the topic
2. Each study's results are converted to a common statistical index in order to ensure data comparability between studies.

The first point in the list has been established by a literature review via online databases, and electronic journals. A State-of-the-Art Matrix Analysis technique for literature review was conducted to assist in the data collection of the research articles surveyed [2]. See also [28]; and [27] for further examples of this literature review technique. The second point was fulfilled by selecting the studies that includes a numeric data as a result of any usability improvement, and a common index was picked as 'percentage changes in usability benefits' as a result of usability improvement.

A total of 135 studies including hypothetical and real world cases were explored with a total of 385 data points. The total percentage data available is 156. Any data point, which is out of the range of 3 standard deviations below or above the mean, was accounted as an outlier. There were two outliers detected for increased sales category. The results with and without outliers are both provided in results section. Table-1 shows the number of hypothetical versus real data gathered for usability benefits.

Table 1- Number of Hypothetical and Real Data

Benefit	Hypothetical	Real	Total
Increased sales	4	44	48
Decreased task time	2	39	41
Decreased error rate	4	30	34
Increased website traffic	0	19	19
Decreased training time	7	7	14

First of all, usability benefit types mentioned in the literature were collected. The mostly mentioned benefit was found as increased sales. See Figure 2 for the Pareto chart of usability benefits mentioned in the literature.

The results indicated that increased sales, decreased task times, reduced maintenance and

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

support costs, increased user satisfaction and loyalty, decreased errors, reduced training costs, reduced development costs, increase in website traffic, and increased completion rates were the usability benefits that are mostly mentioned in the literature.

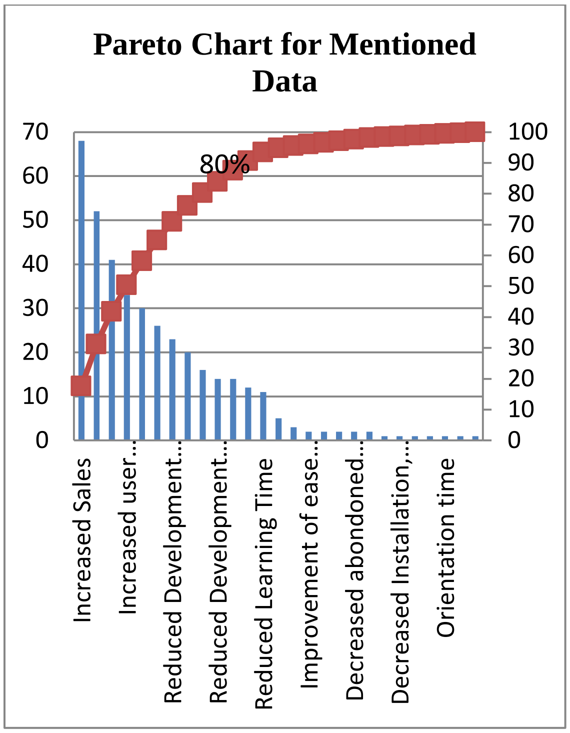

Figure 2 - Pareto Analysis for Usability Benefits Mentioned in the Literature

The next step was gathering the numerical data in these studies. Increased sales category was found to be the benefit with most numeric data availability. See Figure 3 for Pareto Analysis for numerical data availability of usability benefits.

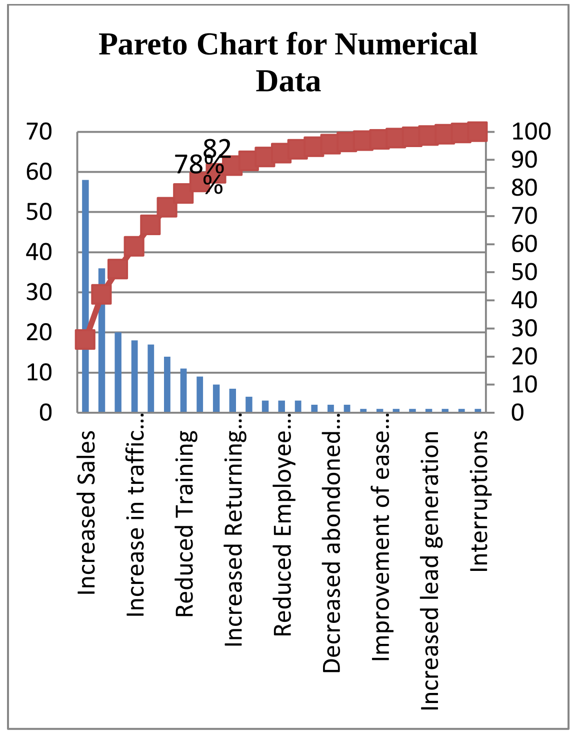

Figure 3- Pareto Chart for Numeric Data Availability

According to numeric data availability Pareto, increased sales, decreased task time, completion rate, increase in traffic, reduced maintenance and support costs, decreased error rates, and reduced training are the significant benefits.

The problem at this point was that there were differences in terms of unit of measurements in between studies for the same benefit types. For instance, in some cases the values of a benefit before and after usability improvements were not provided. Only data available was how many more units sold after usability improvement, or how many less errors were made after usability improvement (for example, 1 error per day is eliminated with usability improvement, [17]). On the other hand, majority of the studies provided percentage changes in a benefit type before and after usability improvement (for example, error rates decreased by 2.5%, [19]). In order to have comparable data, the studies providing percentage change data were selected, and another Pareto Analysis was made to determine benefit types with sufficient data for statistical purposes. Another key reason for selecting percentages as the common index point was that it helped address the issue of different cost basis in different studies. The costs in

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

each study were most likely different, but by taking benefit percentages; costs were accounted relatively equal. Furthermore, there were many benefit categories that could be grouped together due to similarity. First, decreased error rate, decreased failure rate, and increased completion rates were grouped. The values for increased completion rates were reversed to count as decreased error rate. Next, user satisfaction and loyalty, reduced employee turnover, and purchasing experience were grouped since user satisfaction is the key point for employee turnover, and purchasing experience. Considering these grouped components, a final Pareto analysis was conducted to conclude the variables to be evaluated. Figure 4 shows the Pareto Analysis with the final usability benefits list.

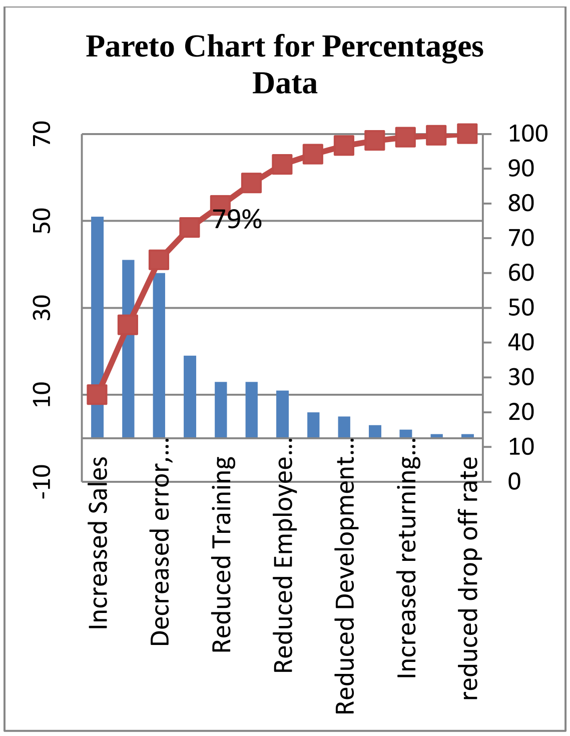

Figure 4- Pareto Chart for Percentages Data

According to the final Pareto chart, the usability benefits to be focused in this research are decided to be increased sales, decreased task time, decreased error rate, increase in traffic and decrease in training time. The rest of the usability benefits are saved for future research endeavor due to current data unavailability. There are very important benefit categories such as reduced maintenance and support cost and reduced development time etc, that will be investigated in future studies.

It should be noted here that training time reduction data do not involve reduced learning time data. Training is producing a motivated user who has the essential skills needed to apply what has been learned and then to continue to learn on the job [5]. It is about the training that companies offer for their employees for ease of usage of companywide software. On the other hand, a general definition of learning is: "the process by which an activity originates or is changed through reacting to an encountered situation, provided that the characteristics of the change in activity cannot be explained on the basis of native response tendencies, maturation, or temporary states of the organism (e.g., fatigue, drugs, etc) [12]". Reduced learning time is another usability benefit that is out of scope of this dissertation as a result of insufficient data as can be seen in Figure 4.

Sample Distributions

Hypothetical data were eliminated from statistical analysis. Distributions of the usability benefit data samples (with only real data) were found by ARENA input analyzer module. The best fitting distribution is picked as the sample distribution, which is an estimate of the population distribution function. Arena performs the Chi-squared and Kolmogorov-Smirnov goodness-of-fit tests for distribution fitting. The null hypothesis for the test is that the data follows the chosen distribution. Therefore, failing to reject the hypothesis means accepting the fitted distribution. It is not suggested to use the Chi-Square test for ratios and percentages. Moreover, the Kolmogorov test is preferred over Chi-Square test if the sample size is small. Thus, only the result of the Kolmogorov-Smirnov test will be taken into consideration. Arena also provides p-values which are the type-1 error levels. A large p value indicates a lack of evidence to reject the null hypothesis.

Two-way Interactions

Selected usability benefits were checked for any correlation by Spearman's Rho non-parametric test. The correlated variables were further analyzed by using MS Excel Data Analysis module and R statistical software. Two way interaction models were found for those variables with significant interaction.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Three-way and four-way interactions were omitted due to data limitations.

Bootstrap

The bootstrap method was introduced by Efron (1979) [9]. It is a method to generalize a sample statistic to the parent population in a scientific approach [25]. It helps gathering information cheaply and in a timely fashion. The sample data is used as a surrogate population, and a large number of samples of same size are resampled from the sample data. These large numbers of samples are named bootstrap samples. The sample statistic is then determined for each bootstrap sample, and histogram of this set of computed statistics is the bootstrap distribution of the statistic [25]. The bootstrap distributions for lower quartiles of selected usability benefits were found in this study. The details of the method will be illustrated in the results section.

Quantile Regression

First ordinary least square regression was used to find the correlation models. The next step is conducting quantile regression to see if these models are valid for different quantiles of the dependent variables.

In this research study, quantile regression modeling is utilized to see:

- The effect of website traffic on sales for usability investments with low, medium, and high percentage changes in sales,
- The effect of error rate on task time for usability investments with low, medium, and high percentage changes in task time.

Thus, the main difference is exploring the impacts of independent variables on dependent variables for different quantiles of the dependent variable.

The first quantile regression analysis was conducted for sales and traffic rate. Sales is assumed to be the dependent variable, and traffic rate is counted as independent variable. Secondly, quantile regression was conducted for error rate and task time. Decreased task time is counted as the dependent variable, whereas decreased error rate is processed as the independent variable.

In order to justify the use of quantile regression modeling, following assumptions should be met ([15]; [16]; [4]):

- The coefficients determined by the quantile regression must be significantly different than zero. In other words, zero should be out of the confidence interval for the quantile regression coefficient. Confidence intervals are found by using the output of the quantile regression in R software. The significance level is selected as 5%.
- The coefficients determined by the quantile regression should be significantly different than the coefficients estimated by the ordinary least squares regression, namely linear regression. In other words, the confidence intervals found via the output of the R software for each quantile regression coefficient should exclude the linear regression coefficients.
- The coefficients determined by quantile regression for different quantiles should be significantly different than each other. So as to check if each coefficient is significantly different than each other, quantile regression plots are analyzed via R software.
- Heterosdecasticity should be checked. The Breusch-Pagan test statistics must be significantly different than zero. In R software, the command '> bptest' applies the Breusch-Pagan test. The null hypothesis in Breusch-Pagan test is that error variances are equal. On the other hand, the alternative hypothesis is that the error variances are a multiplicative function of one or more variables. A large chi-square indicates that heteroscedasticity is present.

If all the assumptions are met, the quantile regression results will indicate reliably the marginal relationships between usability impact factors. This will help software vendors and developers determine

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

how much to invest on usability, since they will know the level of website traffic increase and error rate decrease that no further provides any value to the company. Table 2 provides the number of data points for interactions of variables.

Table 2-The Number of Interaction Data Points in the Data Table

Usability Benefit	Sales	Traffic	Error Rate	Task Time
Sales		5	3	3
Traffic	5		2	0
Error Rate	3	2		4
Task Time	3	0	4	
Total Sample Size in Data Table	44	19	30	39

The first step on the quantile regression was to use only these data points to analyze sales and traffic interaction (5 data points), and task time and error rate interaction (4 data points). These interaction data were analyzed in R-Studio software. The second step was to create random samples from the original samples to increase the number of interaction data points. For sales, 1,000 random samples were created with a sample size of 19, in order to match the sample size of traffic sample in the data table. These 1,000 samples were ranked from smallest to largest (See Table 3).

Table 3-Random samples

Random Samples (1 to 1000)						
		Sample 1	Sample 2	Sample 3	....	Sample 1000
Ranks (1 to 19)	1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	....	$X_{1,1000}$
	2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	....	$X_{2,1000}$
	⋮	⋮	⋮	⋮		⋮
	19	$X_{19,1}$	$X_{19,2}$	$X_{19,3}$	....	$X_{19,1000}$

The point of interest is the row averages in order to determine the expected value of the ranks 1 trough 19.

$$E[Rank_i]=\frac{\sum_{j=1}^{1000} xij}{1000} \quad \text{for } i=1,..,19 \qquad \text{Eq 1}$$

Once the expected values of the ranks were found, they were matched with the traffic data ranked smallest to largest. The same random sampling procedure was used for task time data in order to match the 30 error rate data points.

$$E[Rank_i]=\frac{\sum_{j=1}^{1000} xij}{1000} \quad \text{for } i=1,..,30 \qquad \text{Eq 2}$$

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

The last step was conducting the quantile regression in R-Studio and checking the quantile regression assumptions.

RESULTS

Distribution Results

Weibull distribution fits the 'percentage increase in sales' sample without any outliers data. However, exponential distribution also fits the data very good, which is expected as exponential distribution is a subset of Weibull. A decision was made here to assume exponential distribution. The square error for Weibull fit is 0.00156, whereas 0.00225 for exponential fit. See Table 4 for the ARENA input analyzer solution for increased sales sample without outliers in data table.

Table 4 Increased Sales Distribution Fitting, Outliers Excluded

Distribution Summary	Distribution	Exponential
	Expression	"-0.001 + EXPO(1.38)"
	Square Error	0.002253
Kolmogorov-Smirnov Test	Test Statistic	0.12
	Corresponding p-value	> 0.15
Data Summary	Number of Data Points	44
	Min Data Value	0
	Max Data Value	9
	Sample Mean	1.38
	Sample Std Dev	1.86

See Table 5 for the ARENA input analyzer solution for increased sales sample with outliers in data table.

Table 5 Increased Sales Distribution Fitting with outliers

Distribution Summary	Distribution	Weibull
	Expression	"-0.001 + WEIB(1.71, 0.596)"
	Square Error	0.002920
Kolmogorov-Smirnov Test	Test Statistic	0.144
	Corresponding p-value	> 0.15
Data Summary	Number of Data Points	46
	Min Data Value	0
	Max Data Value	46.2
	Sample Mean	3.04
	Sample Std Dev	8.2

The next best fitting distribution is exponential distribution however Kolmogorov-Smirnov test p-value is less than 0.01, therefore we reject the null hypothesis that exponential model fits data. Thus, when the outliers exist in data, increased sales followed a Weibull distribution as provided in Table 7.

Next, beta distribution fits the 'percentage decrease of task time' sample data. See Table 6 for ARENA input analyzer results.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Table 6 Decreased Task Time Distribution Fitting

Distribution Summary	Distribution	Beta
	Expression	BETA(1.17, 1.4094)
	Square Error	0.015179
Kolmogorov-Smirnov Test	Test Statistic	0.077
	Corresponding p-value	> 0.15
Data Summary	Number of Data Points	39
	Min Data Value	0.04
	Max Data Value	0.98
	Sample Mean	0.453
	Sample Std Dev	0.263

Beta distribution fits the 'percentage decrease of error rate' sample data. See Table 7 for ARENA input analyzer results.

Table 7 Decreased Error Rate Distribution Fitting

Distribution Summary	Distribution	Beta
	Expression	1 * BETA(1.58, 0.913)
	Square Error	0.013359
Kolmogorov-Smirnov Test	Test Statistic	0.171
	Corresponding p-value	> 0.15
Data Summary	Number of Data Points	30
	Min Data Value	0.09
	Max Data Value	1
	Sample Mean	0.645
	Sample Std Dev	0.266

On the other hand, gamma distribution fits the 'increased traffic time' sample data. See Table 8 for distribution summary retrieved from ARENA input analyzer.

Table 8 Increased Traffic Rate Distribution Fitting

Distribution Summary	Distribution	Gamma
	Expression	GAMM(1.18, 1.04)
	Square Error	0.009704
Kolmogorov-Smirnov Test	Test Statistic	0.104
	Corresponding p-value	> 0.15
Data Summary	Number of Data Points	19
	Min Data Value	0.06
	Max Data Value	4.54
	Sample Mean	1.23
	Sample Std Dev	1.27

Finally, triangular distribution fits the 'decreased training time' sample data as summarized in Table 9.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Table 9-Decreased Training Time Distribution Fitting

Distribution Summary	Distribution	Triangular
	Expression	TRIA(0.17, 0.918, 1)
	Square Error	0.167297
Kolmogorov-Smirnov Test	Test Statistic	0.558
	Corresponding p-value	0.0164
Data Summary	Number of Data Points	7
	Min Data Value	0.25
	Max Data Value	1
	Sample Mean	0.787
	Sample Std Dev	0.302

It should be noted that even though the best fit is triangular distribution, Kolmogorov-Smirnov test result indicates that it is not a good fit. The null hypothesis is rejected. .

Two-way Interactions Analysis Results

Spearman's rho test was used to determine existence of any two way interactions. Increased sales and traffic was found to have a dependency. Moreover, decreased error rate and decreased task time was detected to have dependency. Rest of the two way interactions were detected to be insignificant. Training time is not correlated with any other variables, thus is removed from further analysis.

- Test Statistics for Sales and Traffic: Spearman's Rho=1 (Positive Correlation)

Hypothesis:

H01: *S and Tf are mutually independent*

H11: *There is either a tendency for larger values of Tf to be paired with larger values of S, or there is a tendency for smaller values of Tf to be paired with larger values of S.*

Two-tailed test

Decision Criteria: Reject Ho at the level of alpha equals 0.05, if absolute value of Sperman's Rho is greater than its 1-alpha/2 quantile. This value is 0.8 as given in Quantiles of Sperman's rho table in Conover, 2007 [6].

Spearman's Rho=1>0.8

Thus, we reject the null hypothesis at a significance level of 0.95. There is a significant positive correlation between increased sales and traffic rate.

- Test Statistics for Error rate and task time: Spearman's Rho=0.95

Hypothesis:

H01: *ER and TaT are mutually independent*

H11: *There is either a tendency for larger values of TaT to be paired with larger values of ER, or there is a tendency for smaller values of TaT to be paired with larger values of ER.*

Two-tailed test

Decision Criteria: Reject Ho at the level of alpha equals 0.05, if absolute value of Sperman's Rho is greater than its 1-alpha/2 quantile. This value is again 0.8 as given in Quantiles of Sperman's rho table in Conover, 2007 [6].

Spearman's Rho=0.95>0.8

Thus, we reject the null hypothesis at a significance level of 0.95. There is a positive correlation between decreased error rate and decreased task time.

For these significant two-way interactions, parametric non-linear regression was used to determine the relationship models. First, increased

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

sales and traffic rate interaction were analyzed. See Table 10 for the sales versus traffic rate regression output.

Table 10-Increased Sales x Increased Traffic Rate Regression Summary Output

R Square	0.880279813				
Adjusted R Square	0.840373084				
Standard Error	1.194773806				
Observations	5				
ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	31.48807	31.48807	22.05843	0.018256
Residual	3	4.282453	1.427484		
Total	4	35.77052			
		Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept		-0.259746244	0.881994	-0.2945	0.787579
Increased Traffic Rate		5.070378837	1.079576	4.69664	0.018256

R-square is almost 0.9 and Significance F value is 0.01, thus increased sales and increased traffic rate regression model is reliable.

Increased sales = -0.2597462 +5.0703788*Increased Traffic Rate Eq 3

Next two-way interaction to be shown is between task time and error rate. See Table 11 for regression output summary.

Table 11-Decreased Task Time x Decreased Error Rate Regression Summary Output

R Square	0.970907039				
Adjusted R Square	0.956360559				
Standard Error	0.030535707				
Observations	4				
ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.062235141	0.062235141	66.74515202	0.014653848
Residual	2	0.001864859	0.000932429		
Total	3	0.0641			
		Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept		0.222958075	0.038870687	5.735892378	0.029075612
Decreased Error Rate		0.447573831	0.054784137	8.169770622	0.014653848

R square is almost 1, and Significance F is 0.01, thus regression equation is reliable.

Decreased task time= 0.222958+0.4475738*Decreased error rate Eq 4

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Bootstrap Results

The bootstrap method is used for the usability benefits with significant correlation; increased sales, increased traffic, decreased error rate, and decreased task time. The bootstrap sample size is denoted as n (which equals to the sample size), whereas the number of bootstrap samples is denoted as N. There are many different recommendations on how to determine the number of bootstrap samples in the literature. Efron and Tibshirani (1986) suggested usage of at least 250 bootstrap samples for finding confidence intervals [10]. Conover (2007) stated that for estimating simply the mean and the standard error of an estimator, 100 or 200 replications are needed unless confidence intervals are needed [6]. In this study, the point estimate for the lower quantile of the population is the key interest, thus the confidence intervals are not found. The approach used in this study in determining the number of bootstrap samples is taking the second power of the sample size ($N=n^2$) as recommended in Singh and Xie (2008) [25]. The difference from their recommendation is that after finding n^2, the closest upper thousand value was selected. Table 12 shows the selected N for each benefit type.

Table 12-Bootstrap Sample Sizes and the Number of Samples

	n	n^2	N
Increased Sales	44	1936	2000
Increased Traffic	19	361	1000
Decreased Error Rate	30	900	1000
Decreased Task Time	39	1521	2000

In order to create the bootstrap samples, INDEX function was used in Microsoft Excel. The formula is provided below.

=INDEX($A$2:$A$45, ROWS($A$2:$A$45)*RAND()+1) Eq 5

Rows function returns the number of rows in the array, and index function returns a value from the referred position of the single column. Thus, Equation 5 enables random selection of a number from the sample.

The sample size (n) for increased sales is 44. Therefore, the sample size for each bootstrap samples is 44 as well. The statistic investigated is the lower quantile of increased sales. Table 13 shows the descriptive results.

The values of the overall population parameters are unknown. The data analysis results lack the usability investment data that is kept confidential by companies. It is just a representative random sample from a whole population of companies that have invested on usability. Since, the population parameters are unknown, standard error is taken into consideration to estimate the variation of the statistic rather than standard deviation. Standard error indicates how precisely you know the statistic of the population [26]. Standard error in Table 13 is calculated by dividing the standard deviation by the square root of the number of bootstrap samples. Very small standard error indicates that 0.28844 is a very precise estimation for the increased sales population's lower quantile. Figure 5 shows the bootstrap distribution of increased sales lower quantile.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Table 13-Descriptive Statistics of the Lower Quantile Statistic of the Bootstrap samples of Sales

Mean	0.28844
Standard Error	0.001588663
Median	0.25
Mode	0.24
Standard Deviation	0.071047187
Sample Variance	0.005047703
Kurtosis	3.015876451
Skewness	1.576551555
Range	0.54
Minimum	0.15
Maximum	0.69
Sum	576.88
Count	2000
Confidence Level(95.0%)	0.003115609

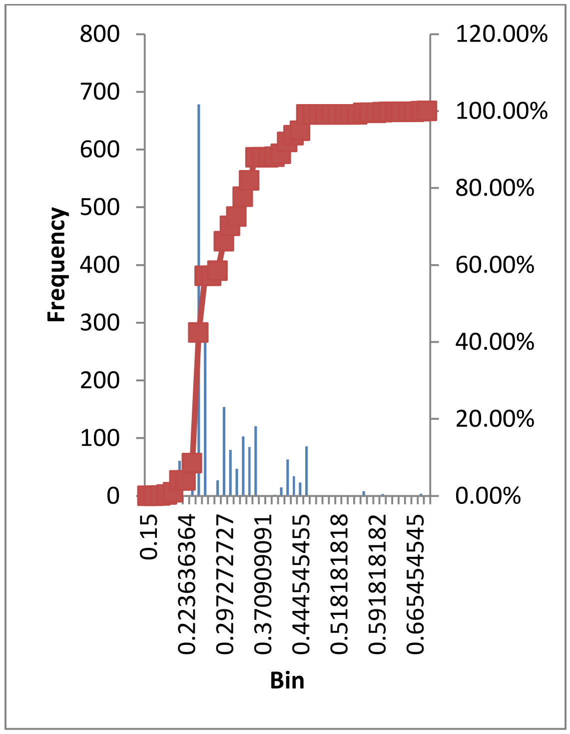

Figure 5 - Increased Sales Lower Quantile Bootstrap Distribution

The next usability benefit sample investigated is the web site traffic. The traffic bootstrap resampling yielded the following descriptive statistics.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Table 14-Descriptive Statistics of the Lower Quantile Statistic of the Bootstrap samples of Web Site Traffic

Mean	0.41168
Standard Error	0.00540605
Median	0.425
Mode	0.2
Standard Deviation	0.170954315
Sample Variance	0.029225378
Kurtosis	0.943345886
Skewness	0.598495433
Range	1.1025
Minimum	0.0675
Maximum	1.17
Sum	411.68
Count	1000
Confidence Level(95.0%)	0.010608516

The standard error is less than 0.01, which indicates that 041168 is a precise estimation for the lower quantile of the traffic population. Figure 6 is the histogram of the bootstrap distribution for traffic rate.

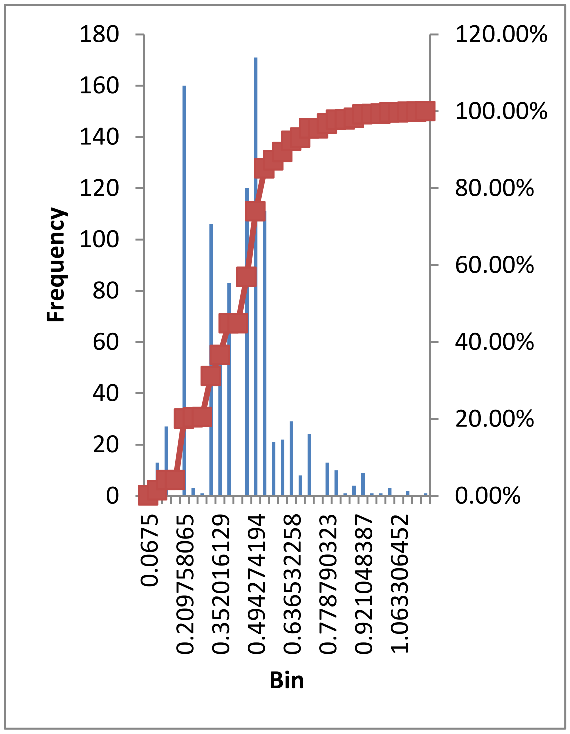

Figure 6 - Increased Traffic Lower Quantile Bootstrap Distribution

The bootstrap statistics for the error rate are provided in Table 15.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Table 15-Descriptive Statistics of the Lower Quantile Statistic of the Bootstrap Samples of Error Rate

Mean	0.48438
Standard Error	0.003991428
Median	0.4575
Mode	0.45
Standard Deviation	0.126220034
Sample Variance	0.015931497
Kurtosis	-0.46639272
Skewness	-0.052572036
Range	0.6825
Minimum	0.1175
Maximum	0.8
Sum	484.38
Count	1000
Confidence Level(95.0%)	0.007832544

The standard error is very low, which indicates very narrow confidence interval. The estimate of 048438 for the lower quantile of the error rate is precise. Figure 7 shows the histogram.

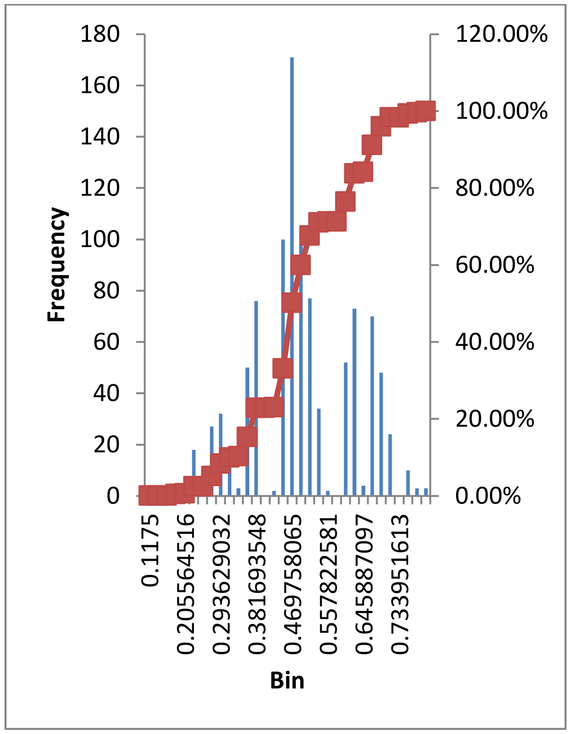

Figure 7 - Decreased Error Rate Lower Quantile Bootstrap Distribution

Finally, the results of the bootstrap analysis for the decreased task time are provided below.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Table 16-Descriptive Statistics of the Lower Quantile Statistic of the Bootstrap samples of Task Time

Mean	0.25781
Standard Error	0.000960638
Median	0.265
Mode	0.25
Standard Deviation	0.042961049
Sample Variance	0.001845652
Kurtosis	1.386176074
Skewness	-0.769764818
Range	0.37
Minimum	0.1
Maximum	0.47
Sum	515.62
Count	2000
Confidence Level(95.0%)	0.001883957

The standard error is less than 0.001, indicating that the estimation of 025781 for the lower quantile of decreased task time population is very precise. The histogram is demonstrated below.

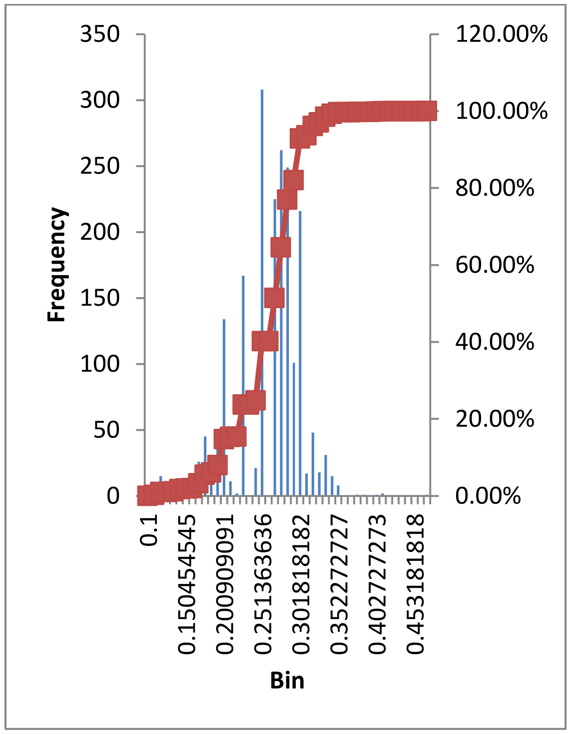

Figure 8 - Decreased Task Time Lower Quantile Bootstrap Distribution

Quantile Regression Results

The quantiles investigated are the lower quantile (25th percentile), the median (50th percentile), and the upper quantile (75th percentile).

Sales and Traffic Quantile Regression Results

Table 17 shows the coefficients and the confidence intervals for sales and traffic quantile regression.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Table 17-Quantile Regression Coefficients and Confidence Intervals for Sales and Traffic with Original Data Points

25th Percentile	**Coefficients**	**Lower Bound**	**Upper Bound**
Intercept	-0.05345000	-inf	0.57343
Increased Traffic Rate	5.043170	-inf	inf
50th Percentile	**Coefficients**	**Lower Bound**	**Upper Bound**
Intercept	-0.05345000	-inf	inf
Increased Traffic Rate	5.043170	-inf	inf
75th Percentile	**Coefficients**	**Lower Bound**	**Upper Bound**
Intercept	-0.15667	-0.32186	inf
Increased Traffic Rate	6.33333	-inf	inf

The infinitely wide confidence intervals are caused by the insufficient number of data. The coefficients are not significantly different than zero, as the confidence intervals include zero. Moreover, quantile regression coefficients are not significantly different from the ordinary least squares regression coefficients. See Table 18 for a comparison.

Table 18- Comparison of Ordinary Least Square regression and Quantile Regression

Increase in sales	OLS Regression	Quantile Regression at 25th quantile	Quantile Regression at 50th quantile	Quantile Regression at 75th quantile
Intercept	-0.2597	-0.05345000	-0.05345000	-0.15667
Increased website traffic	5.0704	5.043170	5.043170	6.33333

- *: Significantly different quantile regression coefficients than zero at 5%
- +: Significantly different quantile regression coefficients from OLS coefficients at 5%

As can be seen, the assumptions of the quantile regression are not satisfied with these 5 original interaction data points. Next, the results for the expected value of ranks data set are shown in Table 19.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Table 19-Quantile Regression Coefficients and Confidence Intervals for Sales and Traffic with Random Sampled Ranked Data Set

25th Percentile	**Coefficients**	**Lower Bound**	**Upper Bound**
Intercept	-0.10845	-0.25244	-0.08848
Increased Traffic Rate	1.08102	0.88635	1.27745
50th Percentile	**Coefficients**	**Lower Bound**	**Upper Bound**
Intercept	-0.07960	-0.24485	-0.00343
Increased Traffic Rate	1.10562	1.05796	1.47636
75th Percentile	**Coefficients**	**Lower Bound**	**Upper Bound**
Intercept	-0.05281	-0.29013	0.02167
Increased Traffic Rate	1.24361	1.08203	1.46424

As can be seen, the infinite confidence intervals were eliminated by increasing the sample size. Table 20 shows the comparisons with the ordinary least squares regression.

Table 20- Comparison of Ordinary Least Square regression and Quantile Regression

Increase in sales	OLS Regression	Quantile Regression at 25th quantile	Quantile Regression at 50th quantile	Quantile Regression at 75th quantile
Intercept	-0.19047	-0.10845 *	-0.07960 *	-0.05281
Increased website traffic	1.27180	1.08102 *	1.10562 *	1.24361 *

- *: Significantly different quantile regression coefficients than zero at 5%
- +: Significantly different quantile regression coefficients from OLS coefficients at 5%

The coefficients are mostly significantly different than zero, but they are not significantly different than the ordinary least square regression as indicated by the confidence intervals. Thus, the assumptions of the quantile regression were not satisfied. Although the quantile regression assumptions were not satisfied, the results showed an increase on the impact of website traffic on increased sales for the upper quantile of sales as the coefficient increases from 1.08 to 1.24. However, this finding cannot be concluded as the assumptions of the quantile regression were not satisfied. Figure 9 also shows that quantile regression results are not significantly different than the ordinary least squares results. The dotted lines show the ordinary least square regression limits, intercepting the confidence limits of the quantile regression.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

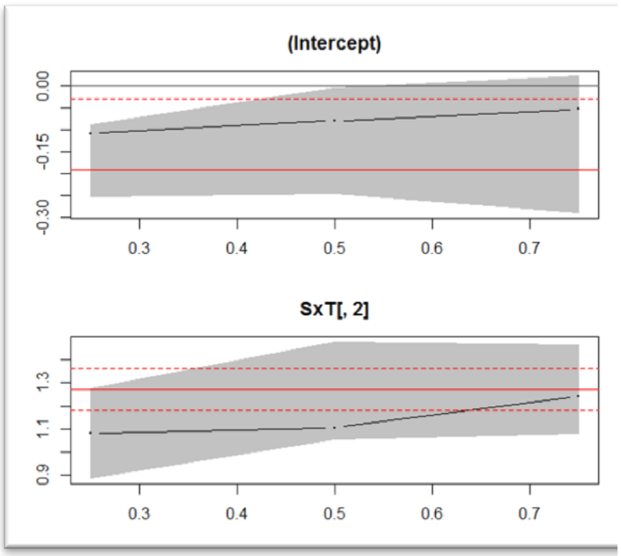

Figure 9 Comparison of the Quantile Regression and Ordinary Least Square Regression Coefficients for Sales and Traffic Rate

Task Time and Error Rate Quantile Regression Results

Table 21 shows the coefficients and the confidence intervals for task time and error rate quantile regression.

Table 21-Quantile Regression Coefficients and Confidence Intervals for Task Time and Error rate with Original Data Points

25th Percentile	**Coefficients**	**Lower Bound**	**Upper Bound**
Intercept	-0.23254	-inf	inf
Increased Error Rate	0.3968300	-inf	inf
50th Percentile	**Coefficients**	**Lower Bound**	**Upper Bound**
Intercept	0.22415	-inf	inf
Increased Error Rate	0.44615	-inf	inf
75th Percentile	**Coefficients**	**Lower Bound**	**Upper Bound**
Intercept	0.21631	-inf	inf
Increased Error Rate	0.49231	-inf	inf

The confidence intervals are infinitely wide as a result of insufficient number of data. The coefficients are not significantly different than zero, as the confidence intervals include zero. Moreover, quantile regression coefficients are not significantly different from the ordinary least squares regression coefficients. See Table 22 for a comparison.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Table 22- Comparison of Ordinary Least Square regression and Quantile Regression

Decrease in Task Time	OLS Regression	Quantile Regression at 25th quantile	Quantile Regression at 50th quantile	Quantile Regression at 75th quantile
Intercept	-0.2597	0.2325397	0.2241538	0.2163077
Decreased Error rate	5.0704	0.3968254	0.4461538	0.4923077

- *: Significantly different quantile regression coefficients than zero at 5%
- +: Significantly different quantile regression coefficients from OLS coefficients at 5%

As can be seen, the assumptions of the quantile regression are not satisfied with these 4 original interaction data points. Next, the results for the expected value of ranks data set are shown in Table 23.

Table 23-Quantile Regression Coefficients and Confidence Intervals for Task Time and Error Rate with Random Sampled Ranked Data Set

25th Percentile	**Coefficients**	**Lower Bound**	**Upper Bound**
Intercept	-0.01322	-1.08961	0.02667
Increased Error Rate	0.59428	0.49440	1.68603
50th Percentile	**Coefficients**	**Lower Bound**	**Upper Bound**
Intercept	-0.01306	-0.20349	-0.00453
Increased Error Rate	0.70995	0.62035	1.04829
75th Percentile	**Coefficients**	**Lower Bound**	**Upper Bound**
Intercept	-0.05303	-0.07802	-0.01486
Increased Error Rate	0.97791	0.85146	1.01547

The infinite confidence intervals were eliminated by increasing the sample size to 30. Table 24 shows the comparisons with the ordinary least squares regression.

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Table 24- Comparison of Ordinary Least Square regression and Quantile Regression

Decrease in Task Time	OLS Regression	Quantile Regression at 25th quantile	Quantile Regression at 50th quantile	Quantile Regression at 75th quantile
Intercept	-0.19047	-0.01322	-0.01306 *	-0.05303 *
Decreased Error rate	1.27180	0.59428 *	0.70995 *	0.97791 *

- *: Significantly different quantile regression coefficients than zero at 5%
- +: Significantly different quantile regression coefficients from OLS coefficients at 5%

The coefficients are significantly different than zero, except for the intercept at 25th quantile. However, they are not different than the ordinary least square regression significantly since the confidence intervals of the quantile regression coefficients includes the coefficient for the ordinary least square regression coefficient. Thus, the assumptions of the quantile regression were not satisfied. Figure 10 also demonstrates that there is no significant difference as the limits of the ordinary least square regression intersects the limits of the quantile regression.

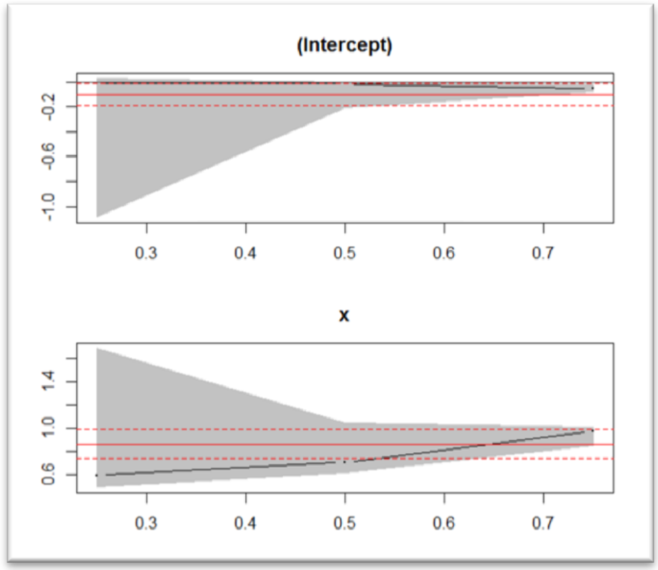

Figure 10- Comparison of the Quantile Regression and Ordinary Least Square Regression Coefficients for Task Time and Error Rate

Since there are no significant differences based on the quantiles, the ordinary least square interaction results were used as the constraints in the goal programming models explained in the next section.

Goal Programming Model

There are two models developed: (1) Sales and traffic rate, and (2) error rate and task time. Training time has been eliminated as no significant correlation was detected with any other variable.

Increased Sales and Increased Traffic Rate Model

The objective is to minimize the maximum deviation from target values for the means of increasing sales, and increasing traffic rate. The variables of the model are shown in Table 25.

The target values for the main decision variables were found as follows. Data collection yielded a table for percentages changes in usability benefits. Bootstrap samples created by resamling these samples in the data table. The lower quantiles of each bootstrap samples were found. The mean of the lower quantiles of the bootstrap samples were picked as the target values for the goal programming models in this study (Population Means: MS, MTf).

ISSN 2304-0777 www.jitbm.com

Moreover, the interaction constraints of the two goal programming models were provided in the interactions section above.

Table 25- Variables of the Sales x Traffic Rate Model

Variable Type	Notation
Main Decision Variables	S: Percentage increase in sales due to usability improvement Tf: Percentage increase in traffic rates due to usability improvement
Deviational Variables	d_i^- : underachievement from target value M_i; d_i^+ : overachievement from target value M_i; Where; **i**: S, Tf
Weights	w_S^-: weight of the negative deviation from M_s w_S^+: weight of the positive deviation from M_s w_{Tf}^-: weight of the negative deviation from M_{Tf} w_{Tf}^+: weight of the positive deviation from M_{Tf}
Maximum Deviation from Means	Q

The weights listed in Table 2 shall be assigned by usability practitioners depending on the significance of the benefit types based on company type and company goals.

Variable Q is the maximum of the weighted percentage deviations from the target means. Weighted percentage deviations are calculated as follows:

$w_i^- \times d_i^- / M_i$ for underachievements Eq. 6

$w_i^+ \times d_i^+ / M_i$ for overachievements Eq. 7

Where; **i**: S, Tf

The objective of the model is minimizing the maximum of the weighted percentage deviations from target means; thus minimizing the predefined variable Q.

MIN Q Eq. 8

Subject to

$w_S^- * d_S^- / M_s \leq Q$ Sale Goal Eq.9

$W_{Tf}^- * d_{Tf}^- / M_{Tf} \leq Q$ Traffic Goal Eq. 10

$S + d_S^- - d_S^+ = M_s$ (actual versus target sale) Eq. 11

$Tf + d_{Tf}^- - d_{Tf}^+ = M_{Tf}$ (actual versus target traffic rate) Eq. 12

$S = -0.259746244 + 5.070378837*Tf$ Eq 13

$d_i^-, d_i^+ \geq 0$ Eq. 14

Where; i: S, Tf

$w_S^-, w_{Tf}^- \geq 0$ Eq. 15

First set of constraints are to set the model as a MINMAX optimization. As the objective function is minimizing Q, these constraints make sure that Q is the maximum weighted deviation from target means. It should be noted that the objective function will ignore the positive deviations from targets, because of the fact that more percentage increase in sales or traffic are favorable. However, less increase in sales or traffic than target value should be penalized. This will help usability practitioners perform a conservative justification of usability. The value of these penalty weights are to be

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

decided by the users of the model. Any biased manipulation of the weights will not make any significant overall impact, since an enormous weight assignment for any under or over achievement for a decision variable will cause serious underachievement in the other decision variable. Thus, users of the model are encouraged to assign realistic weights depending on their company type. Second set of constraints are to set the actual versus target functions. The third set of constraints is the two-way interaction between sales and traffic rate. Sales is set as the Y variable, whereas traffic rate is the X variable. The last set of constraints is to set all decision variables as non negative.

Decreased Task Time and Decreased Error Rate Model

The objective is to minimize the maximum deviation from target values for the means of decreasing task time, and decreasing error rate. The variables of the model are shown in Table 26.

Table 26- Variables of the Error Rate and Task Time Model

Variable Type	Notation
Main Decision Variables	TaT: Absolute percentage decrease in task time due to usability improvement ER: Absolute percentage decrease in error rate due to usability improvement
Deviational Variables	d_i^-: underachievement from target value M_i; d_i^+: overachievement from target value M_i; Where; **i**: TaT, ER
Weights	w_{ER}^-: weight of the negative deviation from M_{ER} w_{ER}^+: weight of the positive deviation from M_{ER} w_{TaT}^-: weight of the negative deviation from M_{TaT} w_{TaT}^+: weight of the positive deviation from M_{TaT}
Maximum Deviation from Means	Z

The weights listed in Table 26 should be assigned by usability practitioners depending on the significance of the benefit types based on company type and company key expectations.

Variable Z is the maximum of the weighted percentage deviations from the target means. Weighted percentage deviations are calculated as follows:

$w_i^- \times d_i^- / M_i$ for underachievements Eq. 16

$w_i^+ \times d_i^+ / M_i$ for overachievements Eq. 17

Where; **i**: ER, TaT

The objective of the model is minimizing the maximum of the weighted percentage deviations from target means; thus minimizing the predefined variable Z.

MIN Z Eq. 18

Subject to

$W_{TaT}^- * d_{TaT}^- / M_{TaT} \leq Z$ Task Time Goal Eq 19

$W_{ER}^- * d_{ER}^- / M_{ER} \leq Z$ Error Rate Goal Eq. 20

$TaT + d_{TaT}^- - d_{TaT}^+ = M_{TaT}$ (actual versus target task time) Eq. 21

$ER + d_{ER}^- - d_{ER}^+ = M_{ER}$ (actual versus target error rate) Eq. 22

Decreased task time= 0.222958075+0.44757381*Decreased error rate Eq.23

$d_i^-, d_i^+ \geq 0$ Eq. 24

Where; i: ER, TaT

$w_{ER}^-, w_{TaT}^- \geq 0$ Eq. 25

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

First set of constraints are to set the model as a MINMAX optimization. As the objective function is minimizing Z, these constraints make sure that Z is the maximum weighted deviation from target means. It should be noted that the objective function will ignore the positive deviations from targets, because of the fact that more percentage decrease in error rate and task time are favorable. However, less decrease in error rate or task time than target values should be penalized. This will help usability practitioners perform a conservative justification of usability. Second set of constraints are to set the actual versus target functions. The third set of constraints is the two-way interaction between task time and error rate. Task time is set as the Y variable, whereas error rate is the X variable. The last set of constraints is to set all decision variables as non negative.

Sensitivity Analysis

The two goal programming models developed are non-linear as multiplications of the variables are involved. The significant points to take into consideration in the sensitivity analysis of non-linear optimization in MS Excel solver module are the reduced gradients for the variables, and Lagrange multipliers for the constraints [23].

Sensitivity Analysis for Sales and Traffic Model

Assuming all the weights are assigned a value of 1, which means the company's perspective on under and over deviations for increased sales and increased traffic rates are the same, the sensitivity analysis results are provided in Table 27.

Table 27- Sales Traffic Model Sensitivity Report with equal weights

Adjustable Cells			**Final**	**Reduced**
	Cell	**Name**	**Value**	**Gradient**
	$C$4	Actual Percent Change Sales	0.475	0
	$D$4	Actual Percent Change Traffic	0.145	0
	$C$5	Under Sales	0	0.479
	$D$5	Under Traffic	0.267	0
	$C$6	Over Sales	0.187	0
	$D$6	Over Traffic	0	2.429
Constraints			**Final**	**Lagrange**
	Cell	**Name**	**Value**	**Multiplier**
	$C$11	SalesxTraffic	-0.259	0.479
	$C$7	Goal Sales	0.288	-0.479
	$D$7	Goal Traffic	0.412	2.429

The same results are attained if the weights assigned are 10,000, or 100,000 for the under and over achievements for sales and traffic. This indicates that the model is very robust; as long as the four weights are assigned equal by the usability researcher and/or engineering manager who uses the model, the above results will be achieved.

In regular nonlinear models, reduced gradient shows the change in the objective function

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

in case that decision variable changes one unit. For instance, if under achievement from target sales increases 1 unit, the objective function increases by 0.479. Similarly, if over achievement from target traffic increases by 1 unit, the objective function increases by 2.429. On the other hand, the Lagrange multiplier shows that if the right hand side of the interaction constraint of sales and traffic increases by 1 unit, then the objective function increases by 0.479. It should be noted clearly at this point that the objective function is not the point of interest in this goal programming model. The key interest is the actual sales and traffic changes depending on the weights assigned to under and over achievements. The possible outcomes when the penalty weights are different can be seen in Figure 11.

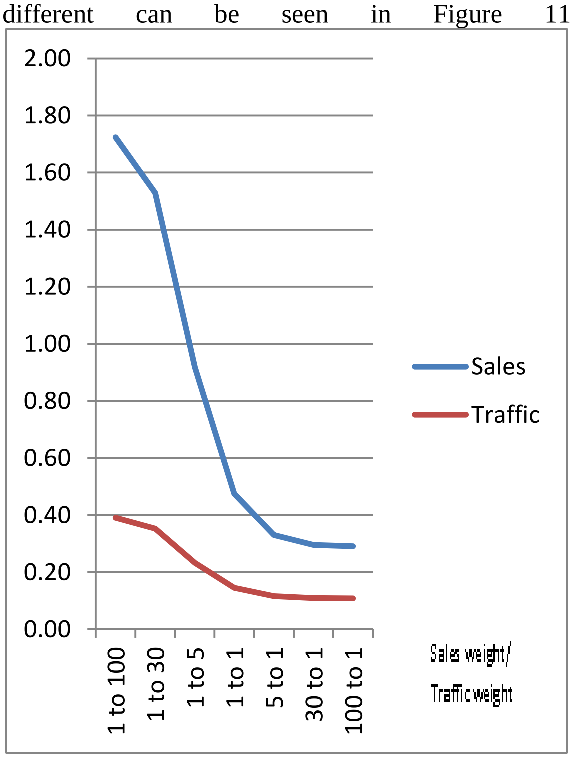

Figure 11 Sales Traffic Model Sensitivity Analysis

The x-axis on Figure 4.1 shows the rate of 'penalty weight assigned to deviations from sales target value' to 'penalty weight assigned to deviations from traffic target value'. The plot starts with a 1 to 100 rate. Thus, an iteration of the model with this weight ratio results in traffic rates very close to its target value. At this point the sales are very high. An initial perspective by looking at the plot could be using these weights since both sales and traffic rate are high, however there is the cost issue that needs to be addressed. Achieving these levels of increase in sales and traffic requires more usability investment, which causes more initial costs. Usability practitioner or engineer manager using the model must conduct a cost benefit analysis. In the region where weights are 1 to 30, and 1 to 100, (the very left of x-axis), there is an indication that the benefit expectation results for sales and traffic rate is starting to get insensitive to further changes in the x-axis. Figure 12 better shows this insensitivity.

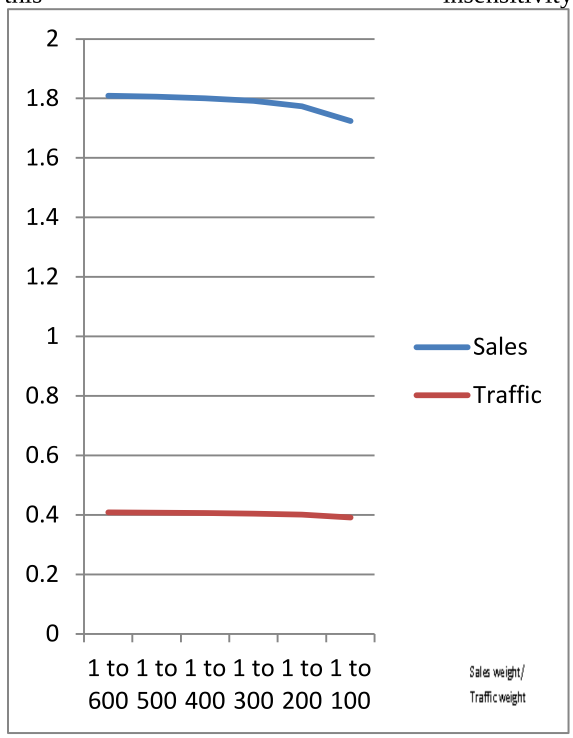

Figure 12 Sensitivity Analysis for Higher Weights for Traffic

In Figure 12, the rate of 1 to 200 is the cutoff point after which the insensitivity starts. Users of the model should not spend time on the region smaller than 1 to 200 as it is practically useless.

Moreover the very right end of the x-axis in Figure 11 shows an indication of insensitivity. Figure 13 proves this insensitivity. The usability benefit expectations are insensitive to changes in the rates

ISSN 2304-0777 www.jitbm.com

higher than 200

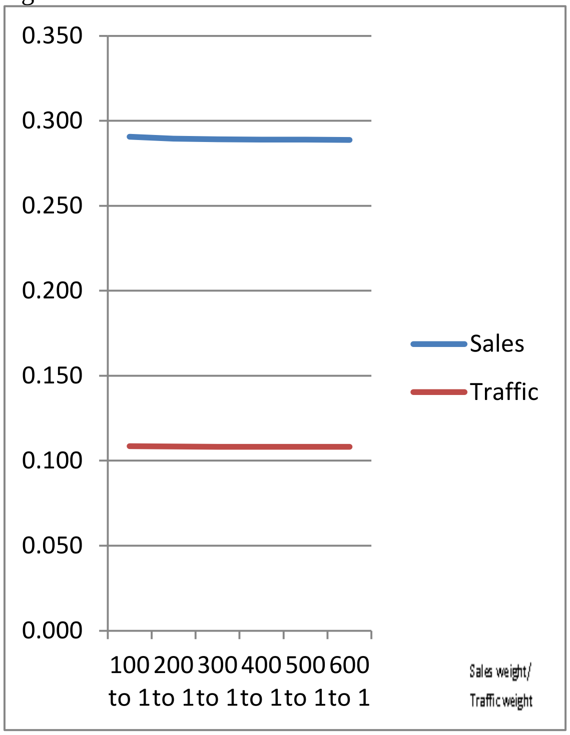

Figure 13 Sensitivity Analysis for Higher Weights for Sales

To sum up the sensitivity findings for this model, users should focus on the region in between the ratios 1 to 200 and 200 to 1.

Sensitivity Analysis for Task time and Error Rate Model

The non-linear optimization sensitivity result of MS Excel solver module is given in Table 28.

Table 28- Task Time Error Rate Model Sensitivity Report with equal weights

Adjustable Cells		Final	Reduced
Cell	**Name**	**Value**	**Gradient**
$C$4	Actual Percent Change Task Time	0.357	0
$D$4	Actual Percent Change Error Rate	0.298	0
$C$5	Under Task Time	0	3.879
$D$5	Under Error Rate	0.186	0
$C$6	Over Task Time	0.098	0
$D$6	Over Error Rate	0	1.736
Constraints		**Final**	**Lagrange**
Cell	**Name**	**Value**	**Multiplier**
$C$11	ERxTaT Task Time	0.223	3.879
$C$7	Goal Task Time	0.258	-3.879
$D$7	Goal Error Rate	0.484	1.736

Similar to the discussions for the sales traffic model, the interest is not the objective function. Therefore, there is no need to explore the reduced gradients and the Lagrange multipliers. See Figure 14 for possible outcomes of the model with

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

different weights assigned.

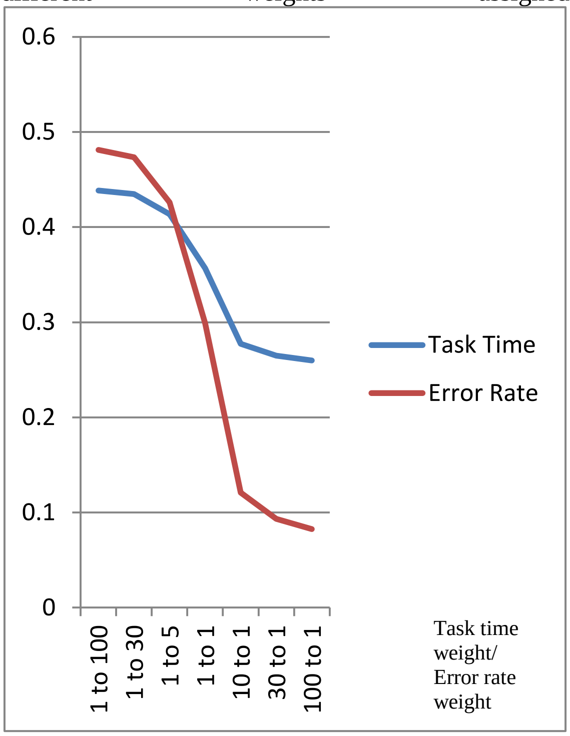

Figure 14 Task Time and Error Rate Model Sensitivity Analysis

The region to the very left of the plot seems to be the optimum region since both task time and error rate are high. However, the costs associated with achieving those usability benefit levels should be considered. A cost benefit analysis is needed. Penalty weight ratios less than 1 to 100 and more than 100 to 1 were analyzed in order to explore sensitivity. See Figure 15 for small weight ratios.

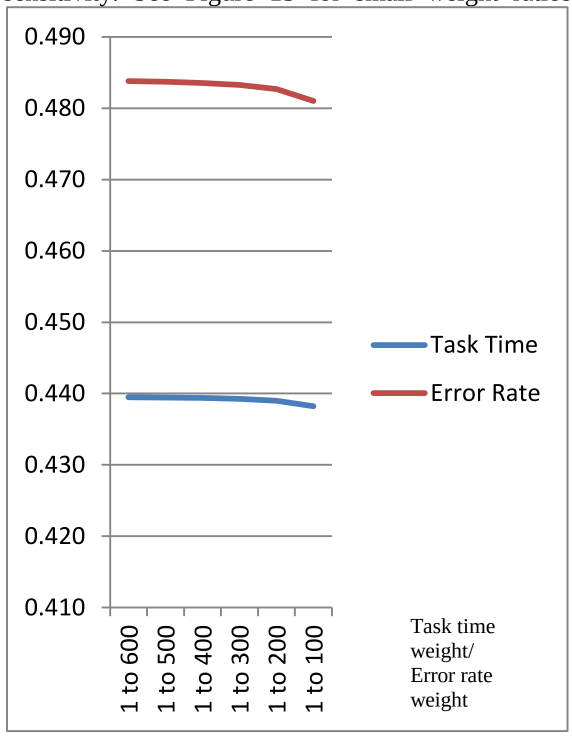

Figure 15 Sensitivity Analysis for Higher Weights for Error Rate

The cutoff point is 1 to 200 penalty weight ratio. Usability benefit expectations are insensitive to further decreasing the penalty weight ratio. See

International Journal of Information Technology and Business Management
29th August 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

Figure 16 for high weight ratios.

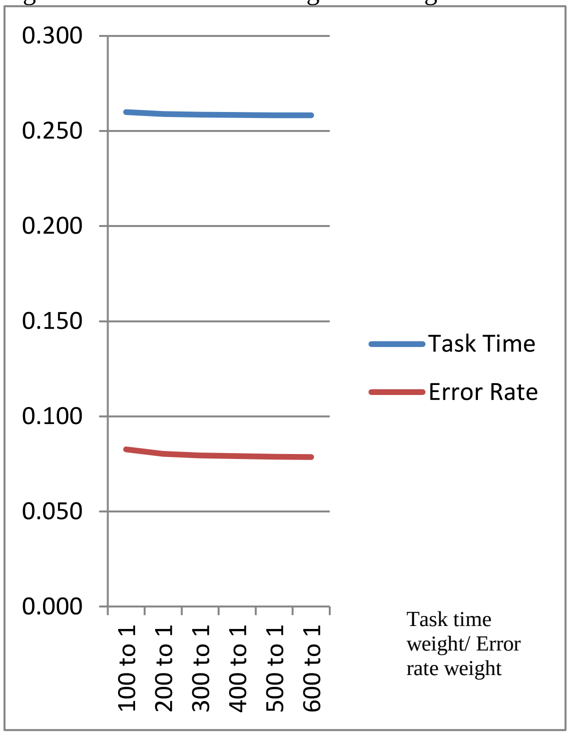

Figure 16 Sensitivity Analyses for Higher Weights for Task Time

The cutoff point is 200 to 1 penalty weight ratio. Usability benefit expectations for task time and error rate are insensitive to weight ratio changes after this point.

CONCLUSION

This study provided two decision tools for usability practitioners. The first tool includes sales and traffic benefits, while the second model covers error rate and task time benefits. The key findings from this study are:

- The interaction between sales and web site traffic in big scale companies and small scale companies that invest on usability are the same.
- The interaction between error rate and task time in big scale companies and small scale companies that invest on usability are the same.
- No matter how much focus is given on increasing usability to increase sales, as long as traffic rate cannot be increased over 40%, the highest increase on sales will be around 180%.
- No matter how much focus is given on increasing usability to decrease task time, as long as error rates cannot be decreased more than 49%, the highest decrease on task time will be around 44%.

It should be noted that these values are going to be used in cost justification of usability models to find overall benefit. Then, benefits will be compared against investment costs with recommended financial ratio in a specific cost justification model. Thus, the key contribution of the decision tools developed in this dissertation is that they provide the usability benefit expectations needed to estimate the benefits part of the cost justification of usability models. The subjectivity in estimations of these assumptions in existing models was addressed by using scientific methods in this study.

REFERENCES

1. Aydin, B., Millet, B., Beruvides, M. G., (2011). *The State-Of-The-Art Matrix Analysis for Cost-Justification of Usability Research*. Paper presented at the Proceedings of the American Society for Engineering Management: Winds of Change, Staking paths to explore new horizons, Lubbock, Texas (pp. 221-229). Red Hook, NY: Curran Associates, Inc.
2. Beruvides, M. G., & Omachonu, V. (2001). A systemic-statistical approach for manageing research information: The State-of-the-Art-Matrix analysis. Proceedings of the 2001 Industrial Engineering Research Conference.
3. Bias, R. G., & Mayhew, D. J. (1994). *Cost-justifying usability*. San Diego, CA: Academic Press.
4. Buchinsky, M. (1998). Recent Advances in Quantile Regression Models: A Practical Guideline for

ISSN 2304-0777 www.jitbm.com

Empirical Research. *Journal of Human Resources*, *33*(1).

5. Compeau, D., Olfman, L., Sei, M., & Webster, J. (1995). End-user training and learning. *Communications of the ACM, 38*(7), 24-26.
6. Conover, W. J. (2007). *Practical Nonparametric Statistics* (3rd Ed.). John Wiley & Sons Inc.
7. Donahue, G. M., Weinschenk, S., & Nowicki, J. (1999). Usability is Good Business. *IEEE Software*, *18*(1), 31-37.
8. Donahue, G. M. (2001). Usability and the bottom line. *Software, IEEE, 18*(1), 31-37.
9. Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annuals of Statistics, 1*-26.
10. Efron, B., Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other methods of statistical accuracy. Statistical Science, 1(1), 54-77.
11. Harrison, M., Henneman, R., & Blatt, L. (1994). Design of a human factors cost-justification tool. In R. G. Bias & D.J. Mayhew (Eds.), *Cost-justifying usability* (pp. 203-239). San Diego, CA: Academic Press.
12. Hilgard, E. R., & Bower, G. H. (1966). *Theories of Learning*. New York: Appleton-Century-Crofts.
13. ISO/IEC. 9241-14 Ergonomic requirements for office work with visual display terminals (VDT)s - Part 14 Menu dialogues, ISO/IEC 9241-14: 1998 (E), 1998.
14. Karat, C. M. (1994). A business case approach to usability cost justification. In R. G. Bias, & D. J. Mayhew (Eds.), *Cost-justifying usability* (pp. 45-70). San Diego, CA: Academic Press.
15. Katchova, Ani [econometricsacademy]. (2013, February 24). *Quantile Regression* [Video file]. Retrieved from http://www.youtube.com/watch?v=P9lMmEkXuBw.
16. Koenker, R. (2005). *Quantile regression*. Cambridge University press.
17. Mantei, M. M., & Teorey, T. J. (1988). Cost/benefit analysis for incorporating human factors in the software lifecycle. *Communications of the ACM, 31*(4), 428-439.
18. Marcus, A. (2005). User interface design's return on investment: Examples and statistics. In R. G. Bias & D. J. Mayhew (Eds.), *Cost justifying usability: An update for the Internet age* (pp. 17-40). San Francisco, CA: Morgan Kaufmann.
19. Mayhew, D. J., & Tremaine, M. M. (2005). A basic framework. In R. G. Bias & D. J. Mayhew (Eds.), Cost justifying usability: An update for the Internet age (pp. 41-102). San Francisco, CA: Morgan Kaufmann.
20. Nas, T. (1996). *Cost-benefit analysis: Theory and application*. Thousand Oaks, CA: Sage Publications, Inc.
21. Porat, T., & Tractinsky, N. (2012) It's a pleasure buying here: the effects of web-store design on consumers' emotions and attitudes. Human–Computer Interaction, 27:3, 235-276.
22. Qusenbery, W. (2001). What does usability mean: Looking beyond 'ease of use'. Retrieved from http://wqusability.com/articles/more-than-ease-of-use.html. Accessed on April 2012.
23. Ragsdale, C. T. (1998). Spreadsheet Modeling and Decision Analysis. (2nd Ed.) Cincinati, OH. South-Western College Publishing.
24. Sherman, P. (2006). *Usability success stories : How organizations improve by making easier-to-use software and web sites*. Burlington, VT: Ashgate Publishing Company.
25. Singh, K., & Xie, M. (2008). Bootstrap: a statistical method. *Unpublished Working Paper. Rutgers University. Retrieved from http://www.stat.rutgers.edu/home/mxie/stat586/handout/Bootstrap1.pdf.*
26. Streiner, D. L. (1996). Maintaining standards: differences between the standard deviation and standard error, and when to use each. *Can J Psychiatry, 41:* 498-502.
27. Sumanth, D. J., Omachonu, V. K., & Beruvides, M. G. (1990). A review of the state-of-the-art research on white-collar/knowledge worker productivity. International Journal of Technology Management, 5, 337-355.
28. Tolk, J., Cantu, J., & Beruvides, M. G. (2013). High reliability organization research: Transforming theory into practice, a literature review for healthcare. ASEM Annual Conference on Engineering Management.
29. Vermeeren, A. P., Attema, J., Akar, E., de Ridder, H., Von Doorn, A. J., Erbuğ, Ç., Berkman, A. E., & Maguire, M. C. (2008). Usability problem reports for comparative studies: consistency and inspectability. *Human–Computer Interaction, 23*(4), 329-380.