

Analysis Report

This report is structured as follows.

Contents

Sample Characterisation	2
Reliability Tests	2
Normality	5
Correlation Analysis	6
Multiple Linear Regression	7
Normality, linearity and homoscedasticity of residuals (errors)	8
Conclusion	12
References	13

Sample Characterisation

This section is dedicated to present the general descriptive statistics of the variables of interest to this study. The table below shows the characteristics of the sample.

Variables		Count	%
Gender	Male	97	45.54%
	Female	116	54.46%
	Non-Binary	0	0.00%
Age	18-24	97	45.54%
	25-40	31	14.55%
	41-56	42	19.72%
	57+	43	20.19%
Service Usage	All of the Time	12	5.63%
	Most of the Time	56	26.29%
	Sometimes	80	37.56%
	Not Often	65	30.52%
	I Have Never Used It	0	0.00%

54.4% of the sample is female and almost half (45.54%) is under 25 years old. The totality of respondents has used the system at least sometimes.

Reliability Tests

The first step before the analysis takes place was to execute reliability analysis. Reliability is an assessment of the degree of consistency between multiple measurements of a variable. One form of reliability is test–retest, by which consistency is measured between the responses for an individual at two points in time. The objective is to ensure that responses are not too varied across time periods so that a measurement taken at any point in time is reliable. A second and more commonly used measure of reliability is internal consistency, which applies to the consistency among the variables in a summated scale. The rationale for internal consistency is that the individual items or indicators of the scale should all be measuring the same construct and thus be highly intercorrelated (Hair et al., 2014). The analysis in this study was done using Cronbach's Alpha. The table below shows the descriptive statistics and Alpha coefficients for the scales under study.

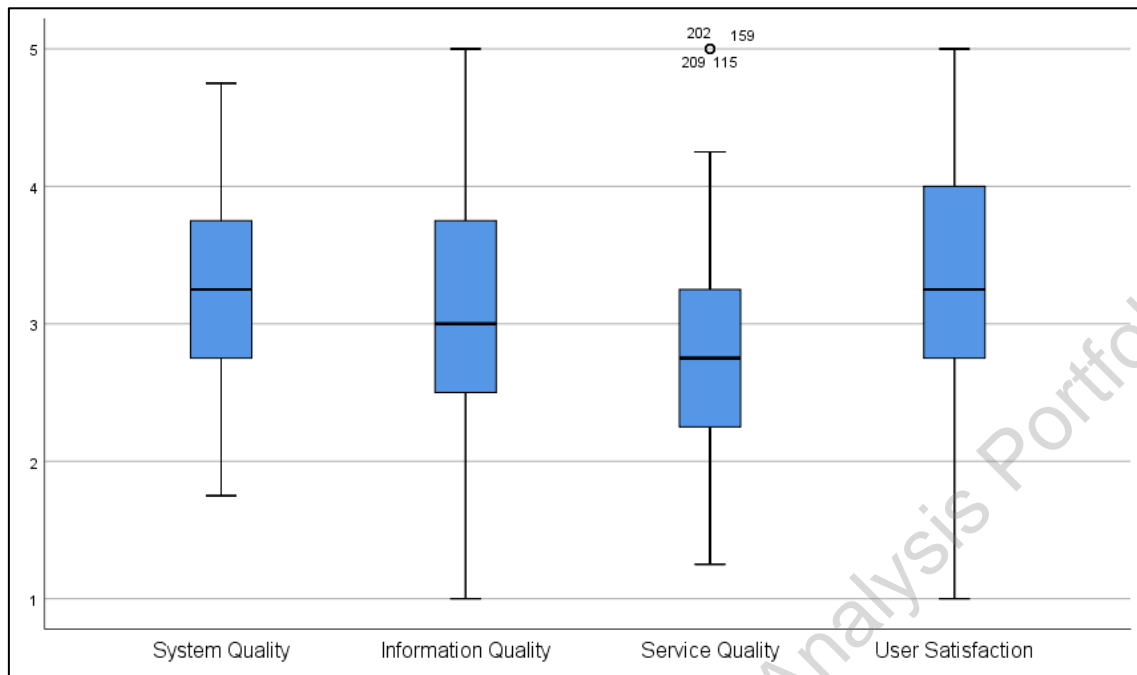
Descriptive Statistics

		Mean	Std. Deviation	λ
System Quality	Q5_1_SystemQuality	3.84	0.925	0.658
	Q5_2_SystemQuality	2.91	0.989	
	Q5_3_SystemQuality	3.83	0.931	
	Q5_4_SystemQuality	2.76	1.066	
Information Quality	Q6_1_InformationQuality	2.85	1.097	0.729
	Q6_2_InformationQuality	3.03	1.000	
	Q6_3_InformationQuality	3.15	1.162	
	Q6_4_InformationQuality	3.54	0.978	
Service Quality	Q7_1_ServiceQuality	2.45	0.973	0.770
	Q7_2_ServiceQuality	3.05	1.015	
	Q7_3_ServiceQuality	2.83	1.191	
	Q7_4_ServiceQuality	2.87	0.967	
Satisfaction	Q8_Satisfaction	3.35	0.892	0.879
	Q9_Satisfaction	3.03	0.931	
	Q10_Satisfaction	3.26	0.973	
	Q11_Satisfaction	3.31	1.013	

All of the tested scalars reached a minimum Alpha of 0.65, which can be considered acceptable. The item scores were averaged to form the overall scale scores. The descriptive statistics of the resulting scales are shown below (by sociodemographic classes and in total). In terms of User Satisfaction, men are more satisfied than women ($M = 3.31$) and the most satisfied age group is 18-24 ($M = 3.39$). Those who use the system all of the time ($M = 3.35$) or most of the time ($M = 3.48$) are also the most satisfied.

Variables		System Quality	Information Quality	Service Quality	User Satisfaction
		Mean	Mean	Mean	Mean
Gender	Male	3.284	3.134	2.773	3.307
	Female	3.377	3.149	2.821	3.179
	Non-Binary				
	Total	3.335	3.142	2.799	3.237
Age	18-24	3.562	3.466	3.008	3.392
	25-40	3.258	2.653	2.734	3.113
	41-56	3.494	3.137	2.536	3.179
	57+	2.721	2.767	2.634	3.035
	Total	3.335	3.142	2.799	3.237
Service Usage	All of the Time	3.896	3.146	2.875	3.354
	Most of the Time	3.371	3.321	3.027	3.482
	Sometimes	3.291	3.184	2.716	3.206
	Not Often	3.254	2.935	2.692	3.042
	I Have Never Used It	-	-	-	-
	Total	3.335	3.142	2.799	3.237

The figure below shows the distribution of scores using boxplots. The graph literally box in observations that are around the median (horizontal line in the middle of the box). The box edges represent the interquartile range of values. That is, the 25th percentile (lowest edge) and the 75th percentile (highest edge). 50% of values lie inside the box. The whiskers (lines protruding from the box), represent the minimum and maximum values observed among the cases. Outliers (participants of which values surpass 1.5 times the interquartile range) are presented as dots outside the whiskers, extreme outliers (more than 3 times this range) are represented as stars. No outliers are present in any variable.



Normality

The scales were also examined with respect to normality. One of the methods to examine if scales follow a normal distribution is to look at values of skewness and kurtosis. Both values should remain between -1 and 1 to indicate normality. As can be seen in the table below, no values surpass this threshold, which indicates no substantial departs from normality.

Descriptive Statistics

	N	Mean	Std. Deviation	Skewness	Kurtosis
System Quality	213	3.335	0.688	0.136	-0.649
Information Quality	213	3.142	0.789	0.169	-0.277
Service Quality	213	2.799	0.801	0.227	0.041
User Satisfaction	213	3.237	0.817	-0.208	-0.114
Valid N (listwise)	213				

Correlation Analysis

Correlation coefficients are indicators of associations between variables (Pallant, 2010). Values between 0.10 and 0.29 indicate a small degree of association, while values between 0.30 and 0.49 are considered medium and values higher than 0.50 represent a high degree of association (Cohen, 1988). This test was used to measure the association between the four scales under study.

There are several different statistics available, depending on the level of measurement and the nature of your data. Pearson's coefficient 'r' is designed for interval level (continuous) variables, whereas Spearman's 'rho' is designed for use with ordinal level or ranked data and is particularly useful when the data does not meet the criteria for Pearson correlation (Pallant, 2010). As the variables under study are measured in a continuous level, Pearson's test was used. The table below shows the correlation matrix.

Correlations

	System Quality	Information Quality	Service Quality	User Satisfaction
System Quality	1	0.725	0.602	0.704
Information Quality	0.725	1	0.675	0.795
Service Quality	0.602	0.675	1	0.704
User Satisfaction	0.704	0.795	0.704	1

All pairs of variables are strongly and positively correlated ($r > 0.500$). This means that when the value of one scale increases, the value for the other scales are expected to increase as well. The strongest association was observed between user satisfaction and information quality ($r = .795$). All associations were statistically significant ($p < .001$).

In order to test how each of the measures of quality influences user's satisfaction, a multivariate regression analysis was executed and is presented next.

Multiple Linear Regression

Multiple Regression analysis is a technique used to explore the relationships between a continuous dependent variable and two or more independent (or predictor) variables (Pallant, 2010). The objective of multiple regression analysis is to use the independent variables whose values are known to predict the single dependent value selected by the researcher. Each independent variable is weighted by the regression analysis procedure to ensure maximal prediction from the set of independent variables. The weights denote the relative contribution of the independent variables to the overall prediction and facilitate interpretation as to the influence of each variable in making the prediction, although correlation among the independent variables complicates the interpretative process. The set of weighted independent variables forms the regression variate, a linear combination of the independent variables that best predicts the dependent variable (Hair et al., 2014).

The model was significant, $F(3, 213) = 164.808$, $p < .001$, $R^2 = .703$. The R-square value means the total variance in the variable 'User Satisfaction' that is explained by all the variables that were inserted in the model. In other words, all variables explain 70.3% of the variance in User Satisfaction.

Next, we look at beta coefficients and their significance (columns 'B' and 'Sig.', below). A p-value (Sig.) of less than 0.05 indicated the model is significant at the 95% confidence level (or 5% significance level). This means that the regression variate (set of independent variables) has a significant effect on the outcome variable.

While the t-statistics is not directly interpretable (same as the F-statistic), the p-value represents the results of the significance test of the variable. Values below 0.05 represent statistical significance of the predictor. That is, the variable significantly influences the outcome variable. The beta coefficient refers to the absolute change in the outcome variable that happens for a one-unit change on the independent variable. So, negative beta coefficients indicate an inverse relationship between the variables, whereas positive values represent the opposite.

Coefficients^a

Model	Unstandardized Coefficients		t	Sig.	Collinearity Statistics	
	B	Std. Error			Tolerance	VIF
(Constant)	0.143	0.155	0.928	0.355		
System Quality	0.246	0.067	3.684	0.000	0.451	2.215
1 Information Quality	0.484	0.063	7.691	0.000	0.385	2.595
Service Quality	0.270	0.053	5.042	0.000	0.518	1.930

a. Dependent Variable: User Satisfaction

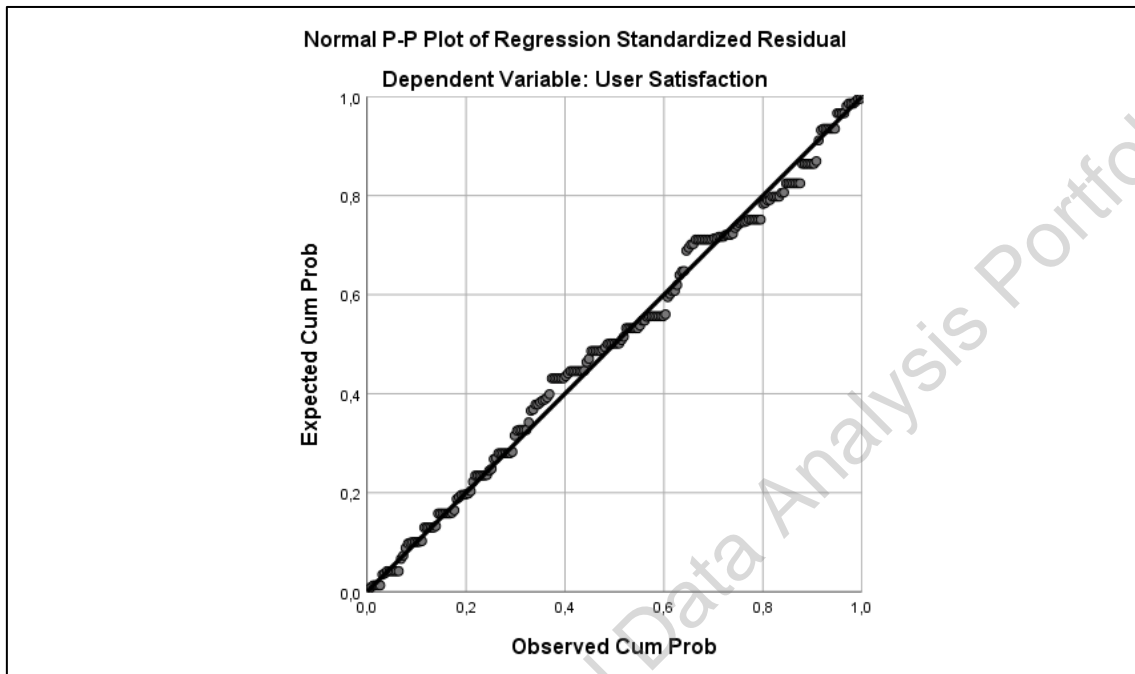
System Quality was a significant predictor of User Satisfaction ($\beta = .246$, $p < 0.001$), as well as Information Quality ($\beta = .484$, $p < 0.001$) and Service Quality ($\beta = .270$, $p < 0.001$). The highest beta coefficient was observed for Information Quality, which suggests that this indicator has the highest impact on user's satisfaction, followed by service quality and system quality. For each one-unit increase on information quality (assuming the indicator was scaled from 1 to 5), user's satisfaction is expected to increase by 0.484 (again considering the agreement scale from 1 to 5).

The existence of multicollinearity was assessed for the model. It refers to an excessive correlation between independent variables, which is not desirable. It can be tested using correlation coefficients, which should not exceed 0.8 (Pallant, 2010), since this would create bias in the model. The correlation matrix for all possible pairs of independent variables was shown earlier and none of the values exceeded 0.800, which clearly does not pose a risk to violating the multicollinearity assumption. An alternative method to test for multicollinearity is looking at Variance Inflation Factors (VIFs) and tolerances (shown in the table above). When VIF values are below 10.0, multicollinearity can be assumed to be absent (Hair et al., 2014).

Normality, linearity and homoscedasticity of residuals (errors).

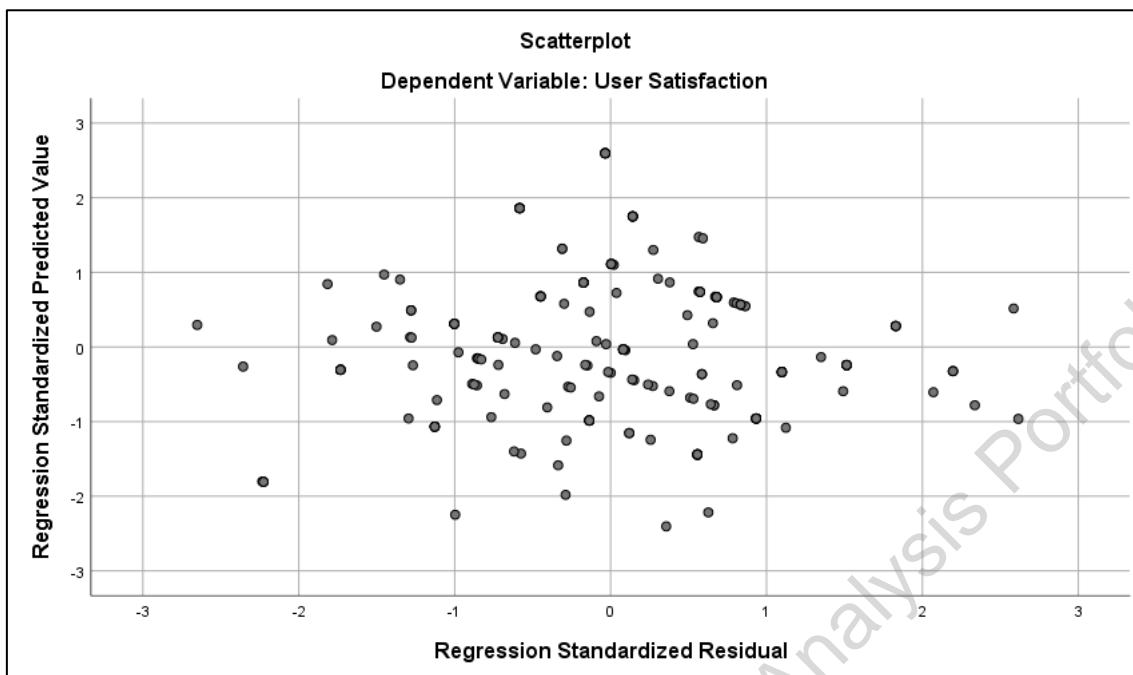
Lastly, violations of the assumptions of normality, linearity and homoscedasticity of residuals (errors) were examined for the regression model. The next figure shows a P-P

plot, which is used to assess the normality of residuals. The observations should follow a diagonal pattern to suggest normality of residuals (Tabachnick and Fidell, 2014).



The graph suggests that no substantial violations of normality are present.

The next figure shows a scatterplot of standardized residuals and standardized predicted values of the dependent variable. If points are well distributed along the X and Y axes, this would suggest homoscedasticity and linearity. Nonlinearity is indicated when most of the residuals are above the zero line on the plot at some predicted values and below the zero line at other predicted values. Lack of homoscedasticity is indicated if values are more dispersed for a given predicted values than at other values (Tabachnick and Fidell, 2014).



The effect of quality indicators on satisfaction, controlling for age groups

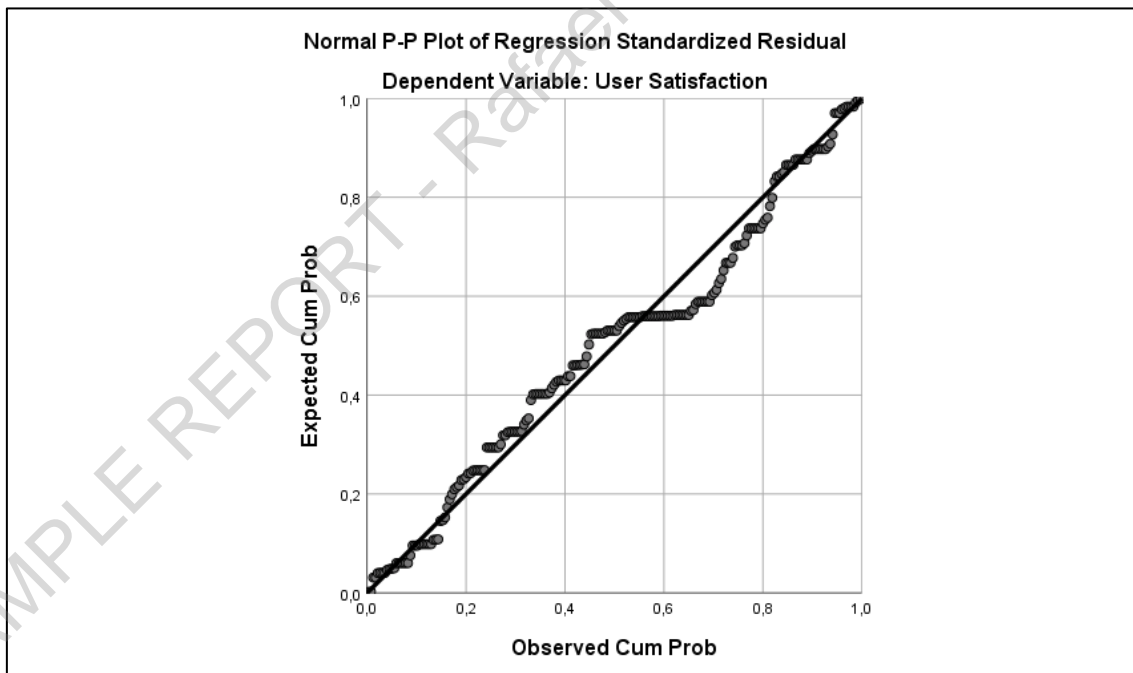
Age groups were included as controlling variables, to check if the significant effects of the quality indicators on user satisfaction change. The results are shown below (Model 2). Age group from 18 to 24 years old is omitted since it was selected to be the reference group and to avoid the redundancy of variables. Thus, the interpretation of the beta coefficients needs to be done against this age group (18-24). The introduction of the age groups in the model significantly increase the explanatory power of the model, $F(3, 210) = 7.927$, $p < .001$. The final R-square was .734. The effects of the quality indicators remained positive and significant while controlling for age. Information quality, again, showed the highest impact, $\beta = .557$, $p < 0.001$. Belonging to the '25-40' age group has a positive impact on user satisfaction compared to the younger age group '18-24', $\beta = .332$, $p = 0.001$. Being 57 years old or older also makes a positive impact on satisfaction in comparison to being 18 to 24, $\beta = .389$, $p < 0.001$. Being between 41 and 56 years old does not have a significant effect on satisfaction, compared to belonging to the youngest age group (18-24).

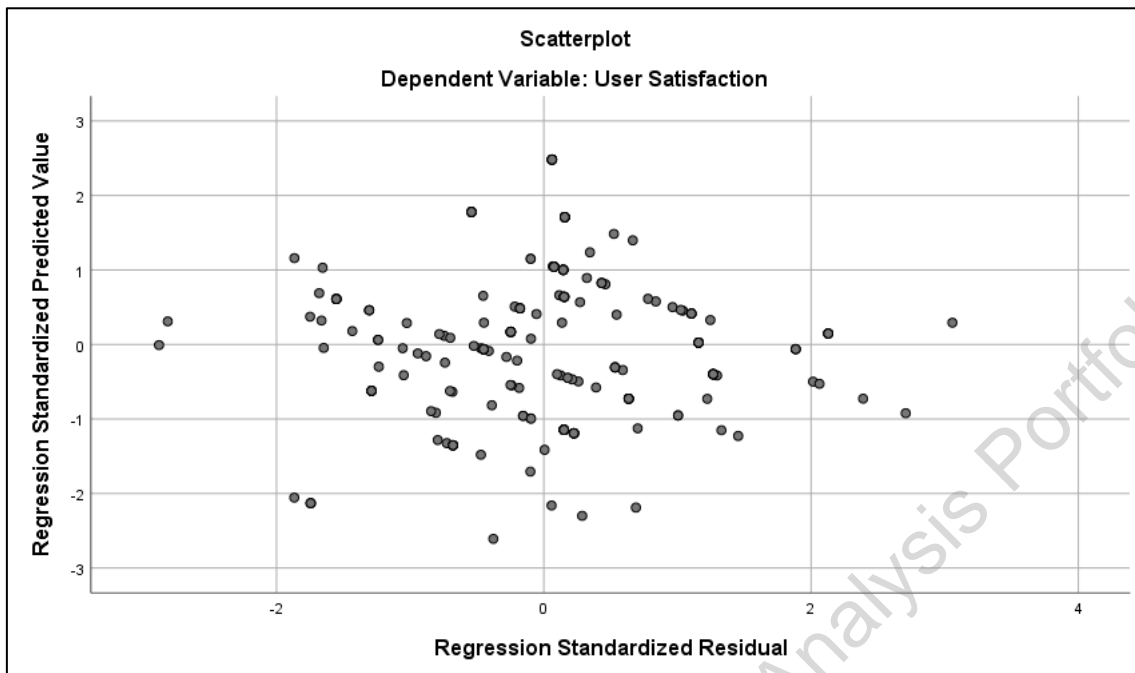
Coefficients^a

Model	Unstandardized Coefficients		t	Sig.	Collinearity Statistics	
	B	Std. Error			Tolerance	VIF
1	(Constant)	0.143	0.155	0.928	0.355	
	System Quality	0.246	0.067	3.684	0.000	0.451
	Information Quality	0.484	0.063	7.691	0.000	0.385
	Service Quality	0.270	0.053	5.042	0.000	0.518
2	(Constant)	-0.347	0.180	-1.926	0.055	
	System Quality	0.331	0.072	4.605	0.000	0.353
	Information Quality	0.557	0.065	8.614	0.000	0.332
	Service Quality	0.210	0.054	3.886	0.000	0.461
	Q3_Age=25-40	0.332	0.096	3.447	0.001	0.746
	Q3_Age=41-56	0.092	0.082	1.119	0.265	0.806
	Q3_Age=57+	0.389	0.090	4.329	0.000	0.661

a. Dependent Variable: User Satisfaction

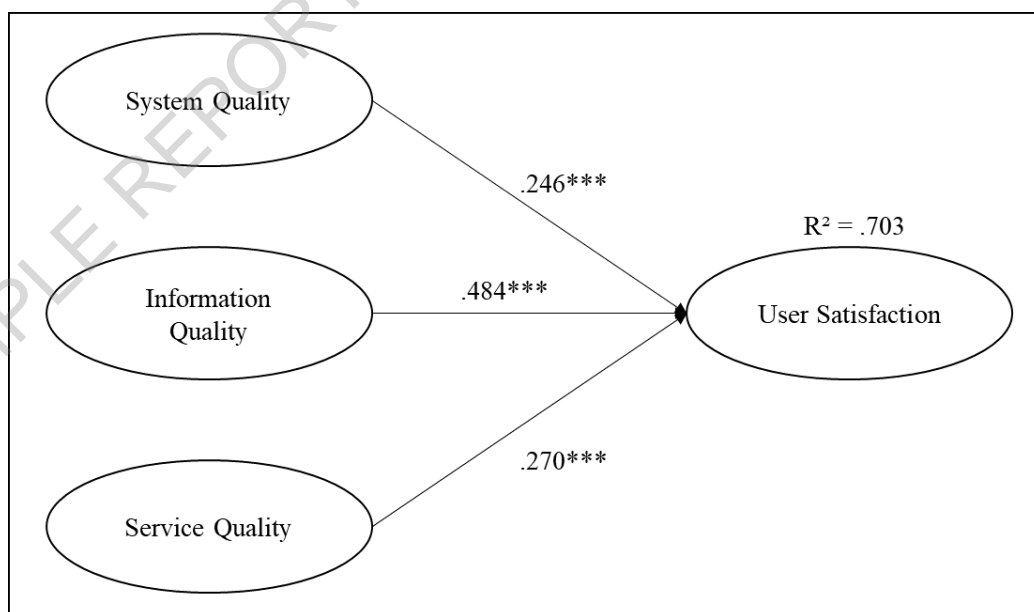
The images below show the scatterplot and P-P plot for residuals, with no substantial departs from normality.





Conclusion

The figure below shows the conceptual model tested here and the corresponding path coefficients.



References

Cohen, J., 1988. Statistical power analysis for the behavioral sciences, 2nd ed. Erlbaum, Hillsdale, NJ.

Hair, J.F., Black, W., Babin, B., Anderson, R., 2014. Multivariate data analysis, Seventh. ed. Pearson Education, Inc., Edinburgh.

Pallant, J., 2010. SPSS Survival Manual, 4th ed. McGraw-Hill, Berkshire, England.

Tabachnick, B.G., Fidell, L.S., 2014. Using multivariate statistics / Barbara G. Tabachnick, Linda S. Fidell.

SAMPLE REPORT - Rafael Data Analysis Portfolio