Asian Journal of Economics, Business and Accounting

21(14): 30-40, 2021; Article no.AJEBA.73872
ISSN: 2456-639X

Forecasting Students' Final Exam: Results Using Multiple Regression Analysis in an Undergraduate Business Statistics Course

Gunawardena Egodawatte[1*]

[1]Department of Global Management Studies, Ted Rogers School of Business Management, Ryerson University, 350, Victoria Street, Toronto, ON, Canada.

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/AJEBA/2021/v21i1430469
Editor(s):
(1) Dr. María-Dolores Guillamón, University of Murcia, Spain.
Reviewers:
(1) Christopher O. Arimie, University of Port Harcourt Teaching Hospital, Nigeria.
(2) Lawrence Dhliwayo, University of Zimbabwe, Zimbawe.
Complete Peer review History: https://www.sdiarticle4.com/review-history/73872

Original Research Article

Received 10 July 2021
Accepted 20 September 2021
Published 27 September 2021

ABSTRACT

This paper discusses the development of a multiple regression model to predict the final examination marks of students in an undergraduate business statistics course. The marks of a sample of 366 students in the Winter 2017 semester were used to fit the regression model. The final model contained three predictor variables namely two test marks and the homework assignment mark. The marks of another 194 students from Winter 2018 were used to validate the model. The model validation showed that it can be used for future cohorts of students for prediction. The two main objectives of the study were to use the model as a teaching tool in class and to use the model to predict final examination marks of future students.

Keywords: Multiple regression; prediction; multicollinearity; forecasting; performance; statistics education; evaluation in higher education.

1. INTRODUCTION

Generally, there is a growing anxiety among students before their final examination to know how they will perform overall in the course. "Am I going to pass or fail the course?" is a commonly asked question. If we can answer this question in an effective way, it would lead some poor-

**Corresponding author: E-mail: gunawardena911@yahoo.ca, egoguna@hotmail.com;*

performing students to drop the course without academic penalty and some better-performing students to get some extra motivation to perform well in the final examination. Some students expect help form their instructor in this scenario and the instructors can use several tools available to them to comment on the students' performance. One method is that the instructor can calculate the student's current course average by averaging their marks so far and provide them with information on how they should perform in the final examination to reach a respectable grade. This is an estimated guess based on the student's past performance. This estimate would not provide them with any future predictions. Using a reliable and valid regression model to predict their final examination score from their current scors is a much better way to satisfy and motivate students further.

This study has developed a multiple regression model to predict the final examination marks of students in an undergraduate Business Statistics course in a mid-sized urban university in Ontario, Canada. A sample of 366 students' marks from two tests and an online homework assignment in the Winter 2017 semester was used to fit the regression model. Another set of 194 student marks from Winter 2018 was used to validate the model. The steps followed in the model building process were: fitting the regression model, testing the assumptions, interpreting the coefficients, testing the significance of the model and individual coefficients, and, finally, validating the model. Initially, six predictor variables were used and after fitting the model using the forward selection method in SPSS, three variables were eliminated by the procedure as insignificant to the final model. The final model contained three predictor variables namely test 1, test 2, and the homework assignment marks. Coefficients of the model were interpreted, and the model was assessed for validity to predict the final examination marks of a similar course. The model validation also showed that it can be used for future cohorts of students.

The study has two main objectives: 1) To teach students as part of their course work to fit a multiple regression model from their own data and predict their individual final examination marks. 2) Use the model to predict final examination marks for students in future semesters [1].

One advantage of this study is that the students will get a real chance to utilize the data collected from their own class rather than using the data collected by others for different purposes. The process will allow students to go through the entire model building process with guided practice before tackling a large data set on their own in later courses involving various forms of regression. The exercise also will increase students' conceptual understanding and practical problems of substantive issues related to regression.

2. THEORETICAL FRAMEWORK

2.1 Fitting a Regression Model

Multiple regression is a frequently used statistical method for analyzing data when there are multiple independent variables. The approach is used most commonly in associational relationships and it can be used in place of analysis of variance [2]. According to Kuiper [3], there are three main uses of multiple regression namely, description, prediction, and confirmation. Description is to describe the relationship between the explanatory variables and the response variable. Prediction is to use the model to make predictions and confirmation is to develop theories using the model.

Multiple regression involves the use of more than one independent variable to predict a dependent variable [4]. In a multiple linear regression model, the response, *Y*, is a random variable that is related to the independent (predictor) *k* variables X_1 , X_2 , …, X_k by the population regression equation,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \qquad (1)$$

where for the *i*th observation, $Y = Y_i$ and X_1, X_2, X_3,…. X_k, are set at values X_{i1}, X_{i2}, X_{i3},…. X_{ik}. The ε s are error components that are the deviations of the response from the true relation. They are unobservable random variables accounting for the effect of other factors on the response. The errors are assumed to be independent and normally distributed with mean 0 and unknown standard deviation, σ. Given the sample data, the regression coefficients are determined by using the method of lease squares. The sample linear regression equation can be written as

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k. \qquad (2)$$

There are several variable selection methods that can be used when deciding which variables are to be included in a regression model. In SPSS, there are methods such as forward selection, backward elimination, and stepwise. Forward selection is a step-by-step procedure, that enters variables into the model each time having the largest absolute t- value until there are no more variables having this criterion. In this method, the variable that correlates highly with the dependent variable gets entered to the equation first, and stays in the equation and this process is repeated until all significant variables are added to the model [5].

2.2 Adequacy of the Model

After fitting a regression model, it is necessary to examine whether the fitted model is adequate for prediction and estimation purposes. A simple measure to explain model adequacy is using the R^2 (adjusted). This measure explains the proportion of variability in the response, Y, explained by the predictors after adjusting for the sample size and the number of independent variables in the model. It is a measure which can be used to compare several models to select the best model. Hypothesis testing is a better way of assessing the overall significance of the model. An F-test will test the significance of all the predictors collectively to the model. T-tests are used to test the significance of the individual predictor variables to the model. The confidence intervals for each regression coefficient explains the range of values that the mean of Y can take given a unit change in the corresponding predictor variable. Finally, the standard error of the regression model measures the spread of the observed y-values about the fitted regression line. We can expect that approximately 95% of the observed y-values lie within 2 standard errors of the fitted line if the errors are normally distributed.

2.3 Testing Model Assumptions

There are four assumptions in regression analysis namely, linearity, independence, normality, and equal variance that can be tested using a variety of methods. The linearity assumption states that there is a linear relationship between the dependent variable and each independent variable. This can be checked using scatter plots between the dependent variable and each independent variable. If the plots show a linear pattern, then it is assumed to be a linear relationship between the dependent and those predictor variables. Otherwise, higher order terms for predictor variables or data transformations are necessary to come up with a reasonable model. The second assumption is that the errors (residuals) are independent of one another. Errors are the differences between the observed Y values from the data and the predicted Y values obtained from the fitted regression equation. The independence assumption can be tested using plots between the residuals and each independent variable. When the assumption is satisfied, the plots show no linear or other curvilinear patterns.

The third, normality assumption indicates that the errors are normally distributed. This can be tested using a normal probability plot, which compares the cumulative distribution of the observed residuals with the expected values derived from the normal distribution, which forms a straight line. If the observed data fall close to the expected line, this assumption is satisfied [6]. Fourth, the equal variance assumption is that the errors have constant variance for all levels of the independent variables. A residual plot can diagnose this assumption. If the graph of standardized residuals and standardized predicted values shows no pattern and the points are scattered randomly in a uniform manner on the graph, this assumption is satisfied. In situations where this assumption is violated, data transformations are necessary before fitting a regression line.

2.4 Multicollinearity

When the data for regression are routinely recorded rather than obtained from preselected settings in controlled experiments, the independent variables often are linearly related. This condition is called multicollinearity. "If two independent variables are highly related to each other, they will explain the same variation, and the addition of the second variable will not improve the forecast. In the fields such as econometrics and applied statistics, there is a great deal of concern with this problem of intercorrelation among independent variables, often referred to as multicollinearity" [4]. There are several methods of tackling multicollinearity. A covariance matrix shows the correlation coefficients between any two independent variables and if the coefficients are high and close to -1 or 1, multicollinearity exists.

Another method is to examine the Variance Inflation Factors (VIF) for individual β parameters. When VIFs are greater than 10, a serious multicollinearity problem exists [7]. VIFs less than 10 for predictors indicate that independent variables are not correlated. Another procedure is to use stepwise regression method to fit the regression line since this procedure tests the parameter associated with each variable in the presence of all the other variables already in the model. Another advanced method is to use ridge regression instead of the lease squares method to fit the regression line. In ridge regression, the β coefficients are more stable than in the least squares method.

Another method is to examine the condition index (CI) for each independent variable in the model. Ho [5] commented on how to detect a multicollinearity problem using the CI. "The condition index summarizes the findings, and a common rule of thumb is that a condition index over 15 indicates a possible multicollinearity problem and a condition index over 30 suggests a serious multicollinearity problem" [5]. One or more above methods can be used to examine multicollinearity in predictor variables. Both VIFs and CIs were used in the current study to detect multicollinearity.

2.5 Validation of the Model

One of the objectives of fitting a regression model to the current data is to use the model for future predictions. Therefore, it is important to assess the validity of the model for this purpose. Some methods of model validation are: examining the predicted values, examining the estimated model parameters, applying the model to a new data set, data splitting, and Jackknifing.

Mendenhall. & Sincich [8] provided the following formula to calculate the Mean Square Error (MSE) of the prediction.

$$MSE_{Prediction} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-(k+1)}, \quad (3)$$

where *n* is the number of new sample observations and *k* is the number of independent variables in the fitted model. y_i represents the values of the dependent variable in the new sample data and $\hat{y}_i$ represents the predicted values for the new data using the previously fitted regression model. Montgomery, Peck, & Vining [7] stated that the number of observations in the new data set should be large enough to reliably assess the model's prediction capability. They recommended to use 15 to 20 new observations at a minimum for validation. This method was used in the current study to validate the model.

2.6 The Educational Value

It is challenging when teaching multiple regression concepts without interesting real-life datasets and to put together all the concepts in one large example. For example, concepts like Prediction, validation, and testing assumptions are explained in textbooks quite often illustrated through small-scale data sets. This is challenging for students to perceive how these concepts might be applied in more realistic multi-variable problems. This article attempts to address this problem by describing a complete multiple linear regression analysis procedure by analyzing a large set of student data from a previous cohort of students. It is a comprehensive analysis in the sense that all the important steps in regression analysis namely formulating the model, testing assumptions of regression, validating the model, and prediction using the model are discussed. The analysis also contains useful practical advice on model building. Secondly, instructors can use the method for their own student data thereby serving the procudure as a teaching tool. As said earlier, a comprehesive analysis is given with real data as a module if teachers wish to adapt it. Therefore, the content of the article serves as a learning exercise for students and a teaching tool for teachers.

3. DATA ANALYSIS

3.1 Data Description

The data file contained the marks of 366 students from an Introductory Business Statistics course offered by the business school of an Ontario university during Winter 2017. This statistics course was a mandatory requirement for all Bachelor of Commerce (B. Com.) students in the business school (…… University, 2018).

As admission requirements to the B. Com. Degree, enrolling students need to have Ontario Secondary School Diploma (OSSD) or equivalent with a minimum of six Grade 12U or M courses. English (ENG4U/EAE4U) and one of Advanced Functions (MHF4U), Calculus and Vectors (MCV4U) or Mathematics of Data Management (MDM4U) are preferred courses. The minimum grades required in the subject prerequisites normally are in the 70% range and this cut-off mark differs subject to competition. In addition, a few international students enroll in each year with other equivalent international qualifications. Considering these entry requirements, it is reasonable to assume that the cohort of students enrolling to the program each year have the same baseline qualifications.

After enrolling into the Business Management program, students choose one of the seven majors namely, Economics and Management Science, Entrepreneurship, Global Management Studies, Human Recourses Management and Organizational Behaviour, Law and Business, Marketing Management, and Real Estate Management.

The evaluation framework of the Business Statistics course contains two tests (test 1 and test 2), weekly online homework (12 assignments), and a group project to be completed using the statistical software package, SPSS. In this project, students are expected to analyze some business data using SPSS software and write a report. Normally, the project will cover topics in the last 4 to 5 weeks of the course. There are 12 weekly modules in the online homework assignment spreading over all the topics in the course. The materials in test 1 and test 2 do not overlap as they are based on the materials of the first and second halves of the course respectively. The final examination covers all the topics in the course. All the questions in tests, the online homework, and the final examination have multiple choice and short answer question formats.

The data set contained one response variable and six predictor variables as follows:

- Test1 – Test 1 marks (%), 18% of the course weight
- Test2 – Test 2 marks (%), 19% of the course weight
- Mystatlab – Weekly online homework assignment marks (%), 10% of the course weight
- Project – The group project marks (%), 3% of the course weight
- Program – Major (coded as 1, 2, 3, 4, 5, 6, or 7)
- Year – Year in the program (coded as 1, 2, 3, or 4)

The response variable, final examination (%) worth 50% of the course weight.

3.2 Fitting the Model

As there was no severe multicollinearity problem among predictors (section 3.4), the forward selection method in SPSS was used to fit a regression line. All the six predictor variables were entered initially to the model. The partial plots [6] (Appendices B.1, B.2, B.3), revealed that there exists a positive linear relationship between FinaExam and Test1 and FinalExam and Test2. A weak positive linear relationship indicated between FinaExam and Mystatlab. No attempt, therefore, was made to include quadratic or higher order terms into the model. The forward selection procedure eliminated the three variables, Project, Program, and Year from the final model. The final model retained three predictor variables: Test1, Test2, and Mystatlab as significant and it read as:

FinalExam. = 20.825 + 0.221*Test1 + 0.407*Test2 + 0.060*Mystatlab (Appendix A.3). (4)

3.3 Testing Model Adequacy

The model explained 50.7% ($R^2_{adj.}$= 50.7) of the total variation in the FinalExam. marks using Test 1, Test 2, and Mystatlab marks as predictors (Appendix A.1). The overall model was significant (F (3, 362) = 126.136, p<.001) (Appendix A.2). Individual t-tests indicated that all the three variables namely, Test1 (t=5.78, p<.001), Test2 (t=11.78, p<.001), and Mystatlab (t=2.41, p<.05) are useful in predicting the FinalExam. mark (Appendix A.3). The confidence interval for the coefficient of Test1 marks indicated that if Test1 mark was increased by 1%, the FinalExam. mark would increase on average between 0.15% and 0.30%. Similarly, if Test2 mark was increased by 1%, the FinalExam. mark would increase on average between 0.34% and 0.48%. If Mystatlab mark was increased by 1%, the FinalExam. mark would increase on average between 0.01% and 0.11% (Appendix A.3). The model expected to predict FinalExam. marks within about ± 2(9.78) = ±19.56 marks (MSE = 95.662) (Appendix A.2).

3.4 Testing Regression Assumptions and Detecting Lack of Fit

Multicollinearity has not affected the parameter estimates of the current model since VIFs for all predictor variables were less than 10. (Appendix A.3). This was further established by the fact that the condition indices for model 3 coefficients were all less than 15 (Appendix A.5). The normality assumption of residuals which indicates that regression residuals are normally distributed was satisfied (Fig. 1,) since all the observed data points were close to the expected line. The equal variance assumption was also satisfied since the residuals against predicted values showed no pattern and very few points (less than 5%) were outside the ±2 standard deviations limit (Fig. 2,).

Residuals plotted against each predictor variable shows that the points are scattered randomly in the diagram with no apparent pattern or trend in the graphs (Appendix B.4, B.5, and B.6), thereby confirming no lack of fit. To detect influential observations and outliers, Cook's Distances were calculated for each observation and compared the values with the values of an F-distribution with 4 and 362 degrees of freedom ($F(4, 362) = 0.841$). None of the Cook's Distances were on or above 0.841 indicating that there were no influential observations or outliers. To detect residual correlation [9], the Durbin-Watson d statistic was calculated (d=2.237) (Appendix A.1). Since d≈2, residuals were uncorrelated thereby satisfying the independence assumption.

3.5 External Model Validation

Next, the validity of the fitted regression model was assessed using a sample of 198 students from the Winter 2018 semester who followed the same course with the same evaluation scheme. Using the fitted regression model in section 3.2 with n = 198 and k = 3, the calculated $MSE_{Prediction}$ was 94.9705. Comparing with the MSE of the fitted model (95.662), it can be concluded that the fitted model was adequate and valid for any future predictions.

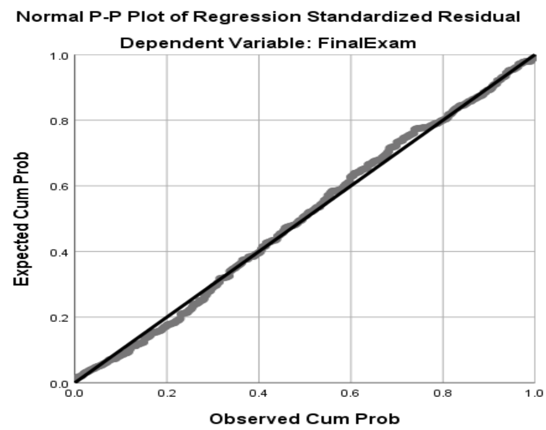


Fig. 1. The Normal probability plot

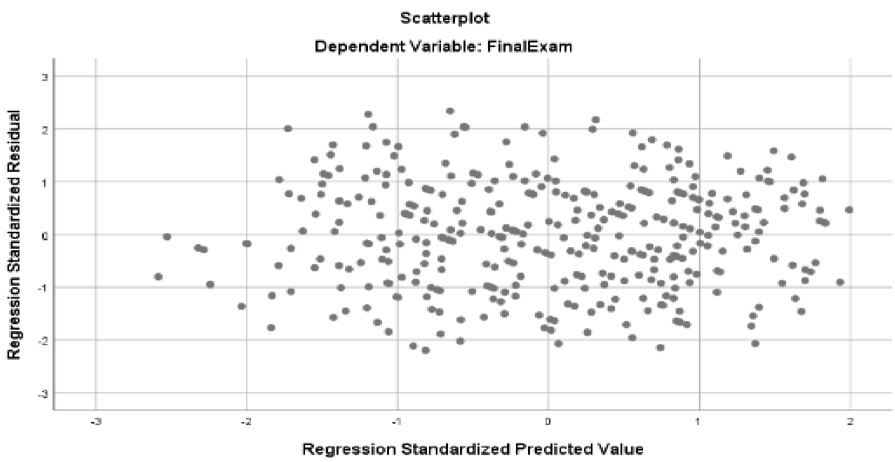


Fig. 2. Residuals vs. predicted values

4. INTERPRETATION OF RESULTS

The model explains 51% of the total variation in the final examination marks and with this reasonable amount, the model is moderately powerful in predicting final exam. marks. This is further confirmed by the significance of the overall F-test and individual t -tests. The interpretations of confidence intervals indicate that test 2 has comparatively high influence on the final examination marks. One reason for this may be that since test 2 is closer to the final examination, students would do more preparation work during this time. The expected prediction range for the final examination mark (±19.56) using the model was high possibly due to the high variations in Test1, Test2, and Mystatlab marks (Appendix A.6). All the model assumptions were completely satisfied making the fitted model valid and without lack of fit. No influential observations indicate that the performances of students were homogeneous. The external model validation with a considerably large number of students made the model powerful enough to be used with future cohorts of students.

5. CONCLUSION

As mentioned in the introduction section, this study was supposed to serve two main purposes. First, it can be used as an in-class exercise. Since multiple regression is taught towards the end of the course, the students have a chance to use their own data and can follow the step-by-step process discussed in this study to fit a multiple regression model to a moderately large data set. This serves the purpose of using the data as a teaching tool for the instructor. Second, the fitted model can be used for future predictions of final examination marks of similar cohorts of students.

On the one hand, as mentioned earlier, the predicted score will help the students to get motivated and to perform well in their final examination or they can decide on whether to continue or not with the course. Before the final examination, any student can find their course average by averaging their current marks. However, this would not provide them with any future predictions. It will only show their current performance. Hence, the advantage of using the fitted model and giving them a predicted final examination score is much better for student motivation and satisfaction.

On the other hand, the process of regression model building will show the uncertainty of predictions and various other problems involved in regression such as fitting a model with a limited number of predictor variables. The need for data on additional variables that could improve the model can also be emphasized using this example. When using the model with future cohorts, it is also possible to fit another regression model using the data of that cohort. As a further step, students can compare and contrast the two models and discuss the differences and possible reasons for the discrepancies.

6. FUTURE RESEARCH

It would be possible to improve the final model by considering additional variables such as the number of hours spent for studying, student motivation, and the number of pervious statistics courses taken to be included in the model. Thus, it may reduce the prediction range of the final examination marks (±19.56) with an improved R^2 value. Quadratic or higher order terms were not included in the current model based on the results of the scatter plots. However, inclusion of those higher order terms especially for the variables that are not included in the model (SPSS project mark, year, program), which are anticipated to be important to predict the final examination marks, would improve the prediction capability of the model. This can be observed from Appendix B.3, that showed a high concentration of Mystatlab marks and final examination marks towards the high end of the marks range.

Data transformations on the excluded variables and inputting them into the model is another option to improve the model. This suggestion was made on the observation that there was a high concentration of data points towards the right end of the diagram in the residual plot of Mystatlab marks which could have caused a slight violation of the regression assumptions. (Appendix B.6). All the suggestions above may or may not be applied to other data sets. There could well be variations when using other data sets.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Brase CH, Brase CP. Understandable statistics: Concepts and methods. Eleventh edition, Cengage Learning, CT, USA; 2015.
2. Leech NL, Gliner JA, Morgan GA. Use and interpretation of multiple regression. Journal of the American Academy of Child & Adolescent Psychiatry. 2003;42(6):738–740.
3. Kuiper S Introduction to multiple regression: How much is your car worth?. Journal of Statistics Education (online). 2008;16(3).
   Available:www.amstat.org/publications/jse/v16n3/kuiper.html
4. Hanke JE, Wichern DW. Business Forecasting. Ninth edition, Pearson Prentice Hall, NY; 2009.
5. Ho Robert. Handbook of univariate and multivariate data analysis with IBM SPSS. CRC Press, Taylor & Francis Group, London, NY; 2014.
6. Levine DM, Szabat KA, Stephan DF. Business statistics. Seventh edition, Pearson Education Inc, NY; 2016.
7. Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis, Fifth edition, Wiley, Hoboken, NJ; 2012.
8. Mendenhall W, Sincich T. A second course in statistics: Regression analysis. Seventh edition, Prentice Hall, Boston, MA; 2012.
9. Keller G. Statistics for management and economics. 9th edition. Cengage Learning, OH, USA; 2012.

APPENDIX A. REGRESSION RESULTS

Appendix A.1. Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.670[a]	.449	.448	10.35384	
2	.709[b]	.503	.500	9.84534	
3	.715[c]	.511	.507	9.78069	2.237

a. Predictors: (Constant), Test2
b. Predictors: (Constant), Test2, Test1
c. Predictors: (Constant), Test2, Test1, Mystatlab
d. Dependent Variable: FinalExam

Appendix A.2. ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	31807.134	1	31807.134	296.703	.000[b]
	Residual	39021.546	364	107.202		
	Total	70828.680	365			
2	Regression	35642.805	2	17821.402	183.857	.000[c]
	Residual	35185.876	363	96.931		
	Total	70828.680	365			
3	Regression	36199.103	3	12066.368	126.136	.000[d]
	Residual	34629.577	362	95.662		
	Total	70828.680	365			

a. Dependent Variable: FinalExam
b. Predictors: (Constant), Test2
c. Predictors: (Constant), Test2, Test1
d. Predictors: (Constant), Test2, Test1, Mystatlab

Appendix A.3. Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	32.941	2.084		15.803	.000	28.842	37.040		
	Test2	.532	.031	.670	17.225	.000	.471	.593	1.000	1.000
2	(Constant)	23.515	2.485		9.464	.000	18.629	28.402		
	Test2	.425	.034	.535	12.487	.000	.358	.491	.747	1.339
	Test1	.238	.038	.269	6.291	.000	.163	.312	.747	1.339
3	(Constant)	20.825	2.709		7.688	.000	15.498	26.152		
	Test2	.407	.035	.512	11.775	.000	.339	.475	.713	1.402
	Test1	.221	.038	.250	5.784	.000	.146	.296	.722	1.386
	Mystatlab	.060	.025	.096	2.411	.016	.011	.109	.859	1.163

a. Dependent Variable: FinalExam

Appendix A.4. Excluded Variables

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	Test1	.269[b]	6.291	.000	.314	.747	1.339	.747
	SPSS	.045[b]	1.129	.260	.059	.954	1.049	.954
	Mystatlab	.138[b]	3.382	.001	.175	.889	1.124	.889
	Program	-.006[b]	-.157	.875	-.008	.992	1.008	.992
	Year	-.051[b]	-1.324	.186	-.069	.999	1.001	.999
2	SPSS	.041[c]	1.093	.275	.057	.954	1.049	.723
	Mystatlab	.096[c]	2.411	.016	.126	.859	1.163	.713
	Program	-.017[c]	-.445	.656	-.023	.990	1.010	.739
	Year	-.038[c]	-1.022	.307	-.054	.996	1.004	.744

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
3	SPSS	.037[d]	.970	.333	.051	.951	1.052	.694
	Program	-.012[d]	-.328	.743	-.017	.988	1.012	.708
	Year	-.031[d]	-.852	.395	-.045	.990	1.010	.713

a. Dependent Variable: FinalExam
b. Predictors in the Model: (Constant), Test2
c. Predictors in the Model: (Constant), Test2, Test1
d. Predictors in the Model: (Constant), Test2, Test1, Mystatlab

Appendix A.5. Collinearity Diagnostics

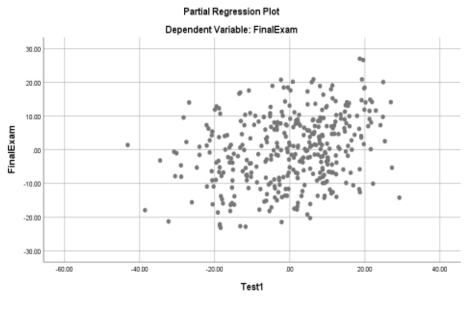
Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	Test2	Test1	Mystatlab
1	1	1.966	1.000	.02	.02		
	2	.034	7.571	.98	.98		
2	1	2.941	1.000	.00	.01	.00	
	2	.035	9.176	.42	.89	.04	
	3	.024	11.026	.57	.10	.95	
3	1	3.897	1.000	.00	.00	.00	.00
	2	.047	9.126	.00	.30	.07	.75
	3	.033	10.895	.34	.61	.18	.18
	4	.024	12.867	.66	.09	.75	.06

a. Dependent Variable: FinalExam

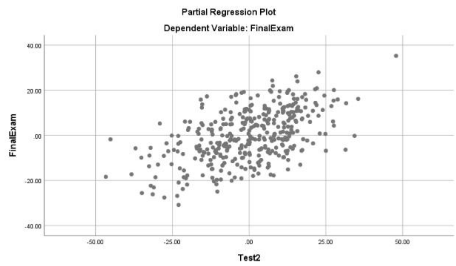
Appendix A.6. Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Test1	366	21.21	100.00	69.1597	15.78263
Test2	366	18.18	100.00	65.1457	17.53902
Mystatlab	366	1.00	100.00	83.4963	22.21835
FinalExam	366	32.00	98.00	67.6148	13.93023
Valid N (listwise)	366				

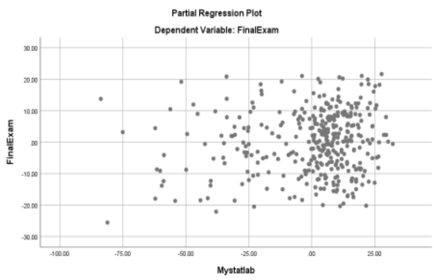
APPENDIX B. GRAPHS AND SUMMARY TABLES



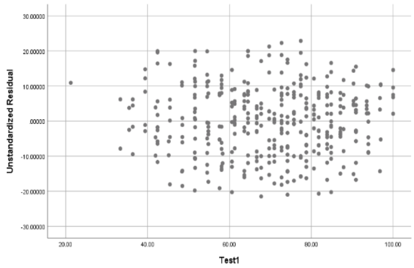
Appendix B.1. Final Exam. vs. Test 1 Marks



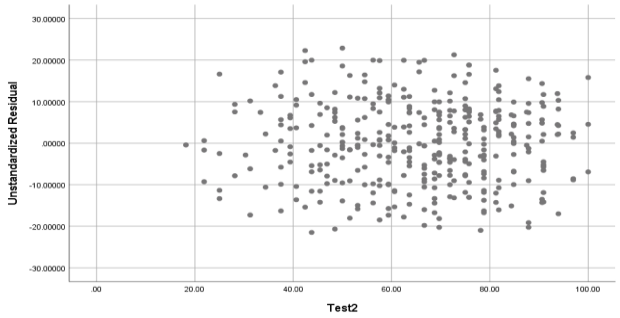
Appendix B.2. Final Exam. vs. Test 2 Marks



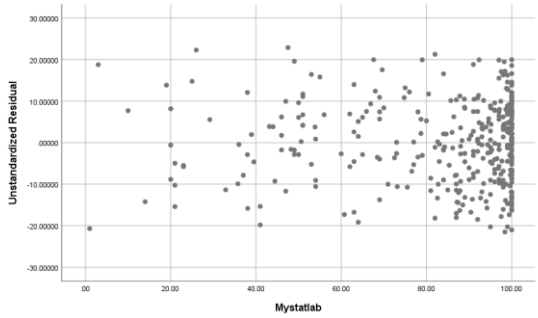
Appendix B.3. Final Exam. vs. Mystatlab Marks



Appendix B.4. Residuals vs. Test 1 Marks



Appendix B.5. Residuals vs. Test 2 Marks



Appendix B.6. Residuals vs. Mystatlab Marks

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle4.com/review-history/73872