Available on Gale & affiliated international databases

ProQuest Start here. AsiaNet PAKISTAN

Journal of **Humanities & Social Sciences**

University of Peshawar JHSS XIX, No. 1, 2011

Modelling the Grade Point Average (G.P.A.): A Case study of the Postgraduate students of the University of AJK

Kamran Abbas [a], Muhammad Zakria [b], Syed Masroor Ahmad [a]

[a] Department of Statistics, University of Azad Jammu & Kashmir, Muzaffarabad, Pakistan
[b] Department of Mathematics & Statistics, University of Agriculture, Faisalabad, Pakistan

Abstract

Considerable research has been undertaken on the grade point average (GPA) of the students. In the present study, an attempt is made to forecast the GPA by fitting a polynomial regression model on the GPA of the Masters level students of the University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan. The data was found to be acceptable for the regression modelling after testing the assumptions. The Best subset, backward elimination and stepwise regression procedures were adopted to fit the model. Good of fit of the models is measured by the coefficient of determination, i.e. $R^2{}_p$, $R^2{}_{adj}$, MSE and Mallow's C_p etc. The model $\hat{Y} = 3.63 + 0.186X_1 - 0.124X_4 + 0.0246X_6$ with $R^2{}_p$, $R^2{}_{adj}$, MSE values 71.1%, 70.6%, and 0.033 respectively is found to be the parsimonious model. The results indicated that the three variables, i.e. study hours at home (X_1), sleeping hours (X_4) and qualification of father (X_6) significantly affect the GPA of the Masters level students and provide sufficient information to forecast the GPA of post graduate students of the said University.

Keywords: Regression model, Grade Point Average, MSE, $R^2{}_{adj}$, Mallow's C_p

Introduction

Examination is a measure that evaluates not only whether students have learnt whatever was taught to them but also their capabilities and potential. Examinations also inculcate in students certain qualities, which later stand them in good stead in practical life. In Pakistan, two examination systems are in operation. In the annual system the examination is held at the end of the session and students are evaluated only once in an academic year (though there may be other tests but they have no bearing on students' grades). Those who pass the exam are promoted to or enrolled in the next class whereas the failed students are retained in the same class. In the semester system, one academic year is divided into two semesters of 18 weeks each with a mid-semester test (in the 9^{th} week) and a final test (in the 18^{th} week). In addition to these two exams, a teacher may also assess students through quizzes, surprise test, assignments and projects. The grading system that is used in the semester system is called grade point average (GPA). The grade point average is convertible into percentage in accordance with a set formula. A grade point average is calculated as follows:

$$\frac{\text{Total Grade Points (of all courses)}}{\text{Total Credit Hours}}$$

Where a grade point = Value x course credit hours
Value is the decimal number given to a percentage mark on a scale of 4.0.

Grebennikov & Skaines (2009) identified a set of variables which significantly affected the GPA of the Students of the University of Western Sydney (UWS). Their sample consisted of 8,896 undergraduate students from the 2004 session. The study then integrated a number of characteristics associated with low probability of success in a profile of UWS students. It was concluded that such students either had relatively poor academic achievement or did not complete their studies or both. Lebcir, et al. (2008) investigated the factors affecting academic performance of international students in a project management courses. Ortiz and Dehon (2008) also discussed the factors which help in improving the GPA of students in Belgium. Fozdar, et al. (2006) conducted a survey to examine the factors responsible for student dropout from the Bachelor of Science (B.Sc.) programme at the Indira Gandhi National Open University, India. One reason was their low GPA. Bowers (2005) studied the comparison of GPA regression equations for two groups of students at the University of Illinois by using the students' High School Percentile Rank (HSPR) and their scores on

School and College Ability Test (SCAT). Results were significantly different for men and women. HSPR and SCAT verbal scores were useful predictors of GPA for all groups. However, separate regression equations for the prediction of GPA indicated significant differences in the regression coefficients of all of the independent variables among the groups. Smith, et al. (2005) reported different variables which affected the GPA of actuarial undergraduates in a mathematics course. Wolaver (2002) examined the effects of heavy drinking on grade point average, study hours as well as other college study efforts using simultaneous equation models. Approximately 1993 college alcohol drinkers were used in this study, and he concluded that GPA decreased directly with intoxication and indirectly by reducing study hours. Kelly, et al. (2001) also discussed the relationship between the length of sleep and the grade point average of the college students. Braunstein, et al. (2001) examined the impact of financial factors on college persistence, i.e., family income and financial aid on the enrolment decisions over the accepted applicants at a single institution of higher learning. He used companion analysis to show that financial factors impact students' GPA. Though a very important assessment topic, no such study, unfortunately, has so far been conducted in Pakistan. Hence, the purpose of the present study was to identify the factors/variables that significantly affect the grade point average of the students of AJK University, Muzafarabad.

Materials and Method

Approximately 510 students were enrolled in different master degree programs during the year 2005-2006 in the University of Azad Jammu and Kashmir, Muzaffarabad. These students made up the population of the present study. A random sample of 180 students was selected from the population using the simple random sampling technique. This sample comprised approximately 35% of the total population. A questionnaire was developed consisting of 12 questions and the information was collected from each student by direct investigation. Out of the twelve variables, only eight important variables were used in the analysis in which the variable GPA (Y) was treated as response variable while the other seven variables as exogenous variables, i.e. study hours at home (X_1), study hours at a library (X_2), stay during the study (X_3), sleeping hours (X_4), family size (X_5), father's qualification (X_6) and mother's qualification (X_7).

The objectives of the study were achieved by fitting the polynomial regression model. Different statistical techniques, i.e. residual plots, Durbin-Watson (DW) statistics, Variance Inflation Factor (VIF) and Mean Squared Error (MSE) were

used to test the basic regression assumptions: i.e., linearity of the model, homoscedasticity, autocorrelation as well as the multicolinearity among the different variables.

Results and Discussion

Figure 1 presents the normal probability plot of the residuals of the grade point average (Y). Most of the points are on the line or close to the line which indicate that the distribution of the residuals is approximately normal. Figure 2 is the plot of the studentized residuals verses fitted values $\hat{Y}_i$. The pattern of the residual plot is clearly random which indicates the homoscedasticity of the error variance as well as acceptability of the models. Correspondingly, Durbin Watson (DW) test was applied to examine the degree of autocorrelation. The calculated value of Durbin Watson was 1.99 which is close to 2. It indicates the absence of autocorrelation in the grade point average data. Moreover, the multicolinearity among the exogenous variable has also been examined by variance inflation factor. The Variance Inflation Factor (VIF) values presented in Table 2 are less than 10. It implies the nonexistence of multicolinearity among the variables (Draper and Smith, 2001).

The value of R_p^2 (in Figure 3 and Table 4) reveals that 71.7% of the variability in GPA has been explained by study hours at home (X_1), stay during the study (X_3), sleeping hours (X_4) and qualification of father (X_6). These results are supplemented with findings of Draper and Smith (2001) when (P-1) ≥ 4 predictors are included in the regression model reflecting a parsimonious model. The position of points also indicates that it is quite obvious that the value of R_p^2 is stables and that there would be an insignificant change in the value of R_p^2 by including more regressors in the model.

Figure 4 reveals the trend of $R^2{}_{adj}$ versus the number of parameters. It also indicates the same number of predictors to fit the data and recommends the same regression model with $R^2{}_{adj}$ = 70.6%. Figure 5 presents the graph of the mean squared error and the number of parameters. It indicates that the inclusion of only four predictors may produce the best regression model. The subset of the four predictors, i.e. study hours at home (X_1), stay during the study (X_3), sleeping hours (X_4) and father's qualification (X6), provides sufficient information to

predict the grade point average of students. Figure 6 and Table 4 depict Mallow's C_p value 4.2 which is close to the number of parameters (P), indicating a highly significant variation due to these four variables.

The summary statistics of GPA data were obtained and presented in Table 1. It is revealed that the average GPA is 3.48, ranging from 2.53 to 4. The coefficient of skewness indicates that the distribution is itself asymmetric. This condition is also reinforced by the coefficient of kurtosis which ranges from -0.703 to 4.042. Correlations among these eight variables are presented in Table 3. The most correlated variable with that of the GPA(Y) is the study hours at home (X_1), i.e. r = 0.743.

Table 4 shows the significance of different regression models and evaluation Statistics. The values of $R^2{}_p$, $R^2{}_{adj}$, MSE and Mallow's C_p criterion of the subset regression model are 71.7%, 71.1%, 0.033 and 4.2 respectively. After including the four predictor variables, the value of the mean squared error becomes stable at 0.033. The best subset of the predictors to predict the grade point average are study hours at home (X_1), stay during the study (X_3), sleeping hours(X_4) and father's qualification (X_6). The backward elimination and stepwise regression models identified the same three predictors, except the stay during the study (X_3) predictor. Different goodness of fit statistics, i.e. R^2, $R^2{}_{adj}$, MSE and P-values for these two models are 71.1%, 70.6%, 0.033 and 0.00 respectively. According to these models, study hours at home (X_1), sleeping hours (X_4) and father's qualification (X_6) significantly affect the grade point average of students.

Conclusion and Recommendations

The best subset regression model identified four predictor variables, i.e. study hours at home (X_1), stay during the study (X_3), sleeping hours (X_4) and qualification of father (X_6) as having a significant effect on the prediction of the grade point average whereas backward elimination and stepwise regression models identified three f the four predictors, i.e. study hours at home (X_1), sleeping hours (X_4) and qualification of father (X_6). The model with these three variables may be considered as parsimonious model for forecasting purposes. These results indicate that students, who have relatively higher GPA, spared more hours to study at home and less sleeping hours with respect to their father's qualification. Kelly, et al. (2001) found the relationship between the

length of sleep and the grade point average among college students. Results from Kelly, et al. (2001) are also in agreement with our findings. Ortiz and Dehon (2008) reported that the father's occupation seems to predominate whereas father's qualification is positively correlated to that of the GPA of the students. These results are also consistent with that of our parsimonious model.

It is recommended that similar studies be conducted in other universities of Pakistan so as to establish the variables that impact students' GPA. This will help teachers guide their low-achieving students to minimise the negative impact (if any) of some of the variables.

References

Bowers, J. (2005). "The comparison of GPA regression equations for regularly admitted and disadvantaged freshmen at the University of Illinois." *Journal of Educational Measurement, 7* (4), pp. 219-225.

Braunstein, A., McGrath, M., & Pescatrice, D. (2001). "Measuring the Impact of Financial Factors on College Persistence." *Journal of College Student Retention: Research, Theory and Practice*, 2 (3), pp.191-203.

Draper, N, R., & Smith, H. (2003). *Applied Regression Analysis* (3rd ed.). John Wiley & Sons (Asia) Pte. Ltd.

Fozdar, B. I., Kumar, L. S., & Kannan, S. (2006). "A survey of a study on the reasons responsible for student dropout from the bachelor of science program at India Gandhi National Open University". *International Review of Research, 7* (3).

Grebennikov, L., & Skaines, I. (2009). "University of Western Sydney Students at Risk: Profile and Opportunities." *Journal of Institutional Research, 14* (1), pp.58-70.

Kelly,W. E., Kelly, K,E., & Clanton, R.C. (2001). "The relationship between sleep length and grade-point average among college students." *College Student Journal, 35*, pp. 84-88.

Lebcir, R. M., Wells, H. & Bond, A. (2008). "Factors affecting academic performance of international students in project management course." *International Journal of Project Management, 26* (3), pp. 268-274.

Ortiz, E. A., & Dehon, C. (2008). "What are the factors of success at university? A case Control study in Belgium." *CESifo Economic Studies, 54* (2), pp. 121-148.

Smith, R. M., & Schumacher, P. A. (2005). "Predicting success for actuarial students in undergraduate mathematics courses." *College Student Journal, 39*.

Wolaver, A. M. (2002). "Effect of heavy drinking in college on study effort, GPA, and major choice." *Journal of Contemporary Economic Policy, 20* (4), pp. 415-428.

Table 1: Summary statistics of endogenous and exogenous variables

Variable	n	Min.	Max.	Mean	SE. Mean	Skew ness	Kurtosis
Y	180	2.53	4.00	3.48	0.03	-0.369	-0.703
X_1	180	1.000	5.00	3.14	0.07	-0.102	-0.033
X_2	180	0.000	3.00	0.43	0.05	1.850	3.900
X_3	180	1.000	3.00	2.46	0.06	-0.990	-0.518
X_4	180	5.000	10.00	6.89	0.10	0.381	-0.537
X_5	180	3.000	12.00	5.99	0.14	0.508	-0.027
X_6	180	0.000	7.00	5.12	0.08	-1.453	4.042
X_7	180	0.000	7.00	4.08	0.12	-0.955	0.830

Table 2: Variance inflation factor (VIF)of each exogenous variable

Predictors	X_1	X_2	X_3	X_4	X_5	X_6	X_7
VIF	1.8	1.0	1.3	1.5	1.1	1.3	1.3

Table 3: Correlation matrix among endogenous and exogenous variables

	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Y	1							
X_1	0.743	1						
X_2	-0.017	0.027	1					
X_3	-0.417	-0.478	-0.072	1				
X_4	-0.741	-0.562	-0.047	0.254	1			
X_5	-0.153	-0.150	0.038	0.118	0.143	1		
X_6	0.001	-0.098	0.001	0.053	0.072	-0.228	1	
X_7	-0.055	-0.104	0.016	0.174	0.068	-0.174	0.430	1

Table 4: Model fittings using Best Subset, Backward elimination and Stepwise regression

Best subset regression	$\hat{Y} = \theta_0 + \theta_1 X_1 + \theta_3 X_3 + \theta_4 X_4 + \theta_6 X_6$				
	θ_o	θ_1	θ_3	θ_4	θ_6
Coefficients (P-Value)	3.78 (0.00)	0.169(0.00)	-0.0403(0.059)	- 0.125(0.00)	0.0248(0.040)
R^2_p =71.7%, R^2_{adj} = 71.1% , MSE= 0.033,P- value =0.00, Mallow's C_p =4.2					
Backward Elimination	$\hat{Y} = \theta_0 + \theta_1 X_1 + \theta_4 X_4 + \theta_6 X_6$				
	θ_o	θ_1	B_4	θ_6	
Coefficients (P-Value)	3.63(0.00)	0.186(0.00)	-0.124(0.00)	0.0246 (0.042)	
R^2_p =71.1%, R^2_{adj} = 70.6% , MSE= 0.033, P value =0.00					
Stepwise Regression	$\hat{Y} = \theta_0 + \theta_1 X_1 + \theta_4 X_4 + \theta_6 X_6$				
	θ_o	θ_1	B_4	θ_6	
Coefficients (P-Value)	3.63(0.00)	0.186(0.00)	-0.124(0.00)	0.0246 (0.042)	
R^2_p =71.1%, R^2_{adj} = 70.6% , MSE= 0.033,P value =0.00					

Figure 1: Plot of R-square versus parameters

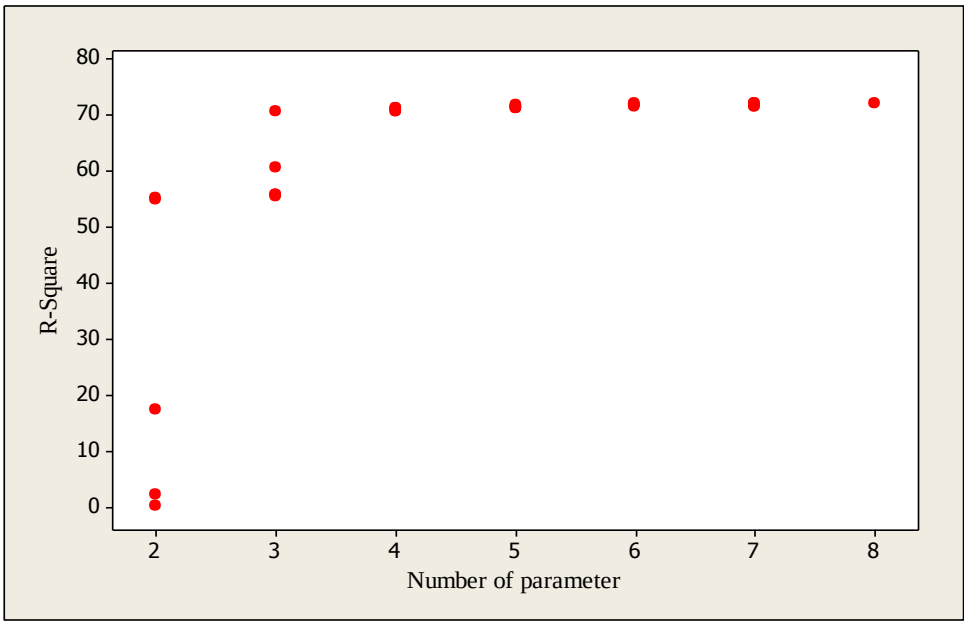

Figure 2: Plot of Adjusted R-square versus parameters

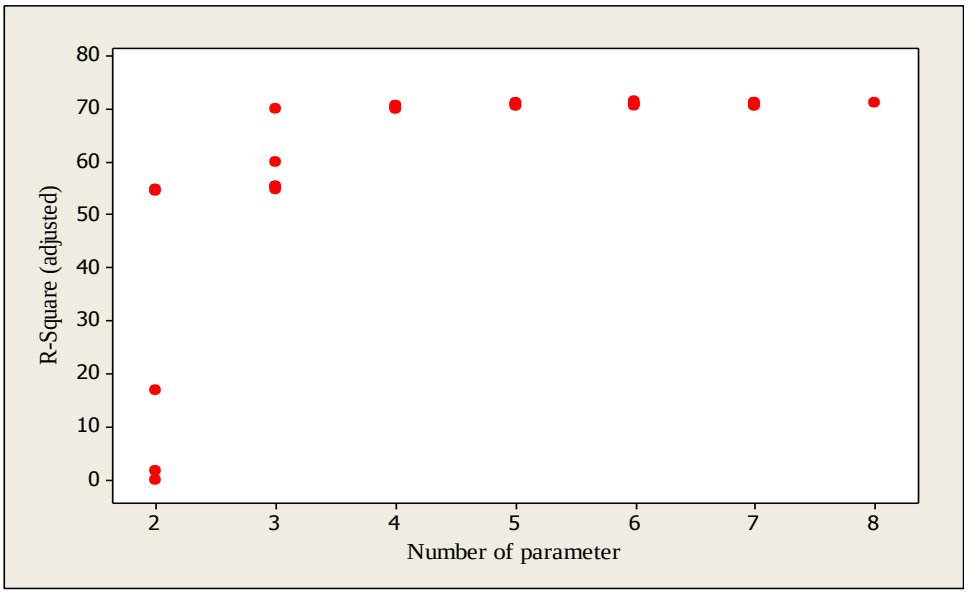

Figure 3: Plot of MSE versus number of parameters

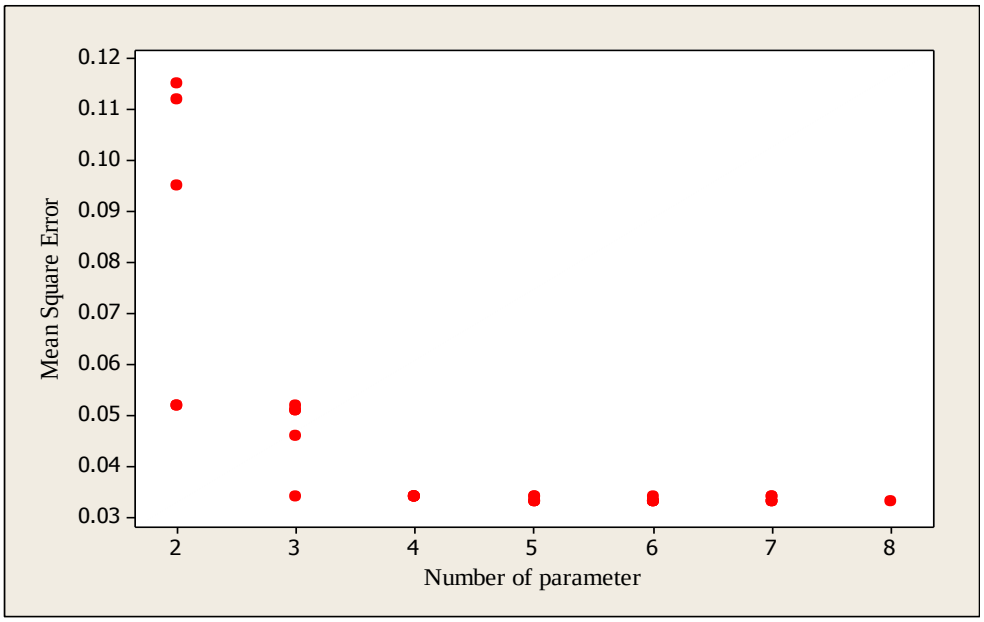

Figure 4: Plot of Adjusted R-square versus parameters