Research Paper

September 2021

Authors:

Kamran Sadigli
Azerbaijan Diplomatic Academy

Rauf Zeynalli
Azerbaijan Diplomatic Academy

Abstract

With the help of this analysis, we will try to understand the relationship between College GPA and other significant variables.

Econometrics I

Research Paper

Kamran Sadigli

Rauf Zeynalli

Introduction

Students and instructors are wondering about the effect or the relation between the high school GPA and the college (university) GPA through the years. For students, educators, and higher education institutions, high school course grades are crucial academic performance markers. Nonetheless, academic test results are frequently regarded as more reliable and objective indicators of academic preparation than students' grades because all students are evaluated on the same tasks under the same conditions. One fundamental assumption behind the emphasis on test scores in policy and practice is that college admission tests are reliable and consistent predictors of preparedness. However, the focus on test scores over grades in policy and practice recommendations contradicts research suggesting that high school grade point averages are stronger predictors of college outcomes than test scores. Some prior studies show that high school GPAs are more predictive than Achievement Test Score or SAT to predict college freshman GPA or graduation. However, according to College Board, "Combination of

GPA. Apart from that, we think there are some other reasons can affect positively to College GPA such as, number of skipped classes, having computers, and graduating from prestigous high school.

Thus, the purpose of our project is to understand the significance of the High school GPA and Achievement Test Score on predicting College GPA. The empirical analysis will provide all the necessary information to get the results which will make everything clear. The paper aims to analyze the cross-sectional data to get the more clear result to answer this question – "Does High school GPA, Achievement Test Scores, and other variables are directly proportional to college GPA?".

Literature Review

The relationship between college GPA and university entrance exam and high school GPA is among the most studied topics among researchers. Previous studies depict that most researchers concluded a positive relationship between high school GPA and university entrance exam on college GPA. On the other hand, several studies show that the relation between ACT score and college GPA is not that significant, and some researchers find out that there is no correlation.

A research carried by Joel. O and Kimberly (2017), using data from 2 different CSM college programs in the USA, demonstrate assessment of student's undergraduate year performance and the university admission score. Researchers applied correlation and regression analyses to find out the relation and predictivity between 2 score variables. At first, data contained 160 students (N=160) but removing some outliers and errors decreased to 155. Researchers used three years of admission data starting from 2013 to 2015. The authors took UGPA (undergraduate GPA) as the dependent variable and ACT score as a primary independent variable in the study. Apart from these two variables, researchers used some other variables such as work experience, high-school GPA. As in other previous studies on this topic, authors used ordinary least square (OLS) to explain variance in university performance. Also, in this study, researchers didn't use randomly selected data since they chose only students who are already studying in the two mentioned programs. Researchers found that ACT scores were statistically significant after rejecting the null hypothesis, making it a helpful predictor. Analyzing the linear regression model, researchers fail to reject the hypothesis that the excellent performance of students in the CSM program depends on ACT score. The authors came to the conclusion that university entrance score is a reliable and valid predictor for long term academic performance in CSM programs.

complete their degree on time.

However, most studies show a positive relationship between these variables. A study carried in the USA by Soares, 2012, between 890 students from 6 different universities found that the relationship is not that significant as mentioned in previous studies. Soares also used college GPA as the dependent variable and ACT score as the primary independent variable. He also used different independent variables such as the number of class missed during high school and student grades from math during high school. Most importantly, Soares states that more than entrance exam score, grades from math are a much better predictor of academic GPA at college.

Data Description

To analyze the relation between college GPA and high school GPA, we used "GPA1" cross-sectional dataset contains 141 observations.

In this analysis, we have used one dependent and five independent variables for our empirical analysis. For our research question, College GPA (colGPA) has been selected as the dependent variable. The main independent variable is high-school GPA (hsGPA). The below table provides detailed data of our selected variables:

Table 1: Data Description

Variable Name	Variable Description
colGPA	MSU GPA
hsGPA	High school GPA
ACT	Achievement test score
PC	=1 of pers computer at school, = 0 of pers not computer at school
skipped	avg lectures missed per week
gradMI	=1 if Michigan high school, = 0 if not Michigan high school

First of all, we did not use encode command because we do not have string variables in this dataset. In this part of the analysis, we have two dummy variables: PC and gradMI. Firstly, for creating dummy variables, we tabulated PC and gradMI.

=1 of pers | computer at | sch	Freq.	Percent	Cum.

=1 if Michigan high school	Freq.	Percent	Cum.
0	16	11.94	11.94
1	118	88.06	100.00
Total	134	100.00	

Before the analysis, we need to check possible outliers and found that ACT has some outliers. To find potential outliers first, we scattered colGPA and hsGPA, and we dropped hsGPA less than 2.7. Secondly, we scattered ACT and skipped variables and dropped for ACT variable which less than 17 and more than 32, and for skipped we need to drop average lectures skipped more than 4.5.

Table 2: Summary Statistics

Variable	Obs	Mean	Std.Dev	Min	Max
colGPA	134	3.059702	0.3709608	2.2	4
hsGPA	134	3.419403	0.3009952	2.7	4
ACT	134	24.16418	2.575752	19	31
have_pc	134	0.3955224	0.4907974	0	1
mch_grad	134	0.880597	0.3254789	0	1
skipped	134	1.020522	0.9990825	0	4

Our summary statistics table shows that our dependent variable ranges between 2.2 and 4, indicating that our dataset includes both low and high GPAs. For example, the average dependent variable (college GPA) is 3.05, an average explanatory variable (hsGPA) is 3.41. It is worth prob>z is bigger than 0.05.

H0 = variable is normally distributed

Ha = variable is not normally distributed,

we reject H0 if prob>z > 0.05

Variable	Obs.	W	V	Z	Prob>Z
ColGpa	134	0.98278	1.820	1.349	0.08867
HsGpa	134	0.99339	0.699	-0.808	0.79040
ACT	134	0.99178	0.869	-0.316	0.62406
Skipped	134	0.93990	6.351	4.166	0.00002

As we can see from the distribution, College GPA is normally distributed, high school GPA is also normally distributed, and achievement score is normally distributed. However average skipped classes per week is right skewed.

For dummy variables, we created pie charts. We can see from the pie chart only 39.55% of the students have pc and also pie chart shows us that, 88.06% of the students graduated from Michigan high school.

SLR Assumptions:

1)First SLR assumption is about linearity between the variables. As we see from the fitted values relationship between college Gpa and High School GPA is linear.

2) Second SLR assumption states that, our data should be collected randomly, however, as we know from the data description, we did not collect the data by ourselves, so therefore, we can not surely tell that our data is collected randomly. Finally, we can not say that SLR 2 is holded.

3) Third SLR assumption says that, sample outcomes should not be all the same value. As we saw from the first assumption sample outcomes are not all the same values, so SLR 3 is holded.

Empirical Analysis (Bivariate Regression)

First of all, as we mentioned before, we have used five independent variables that may influence college GPA. Our empirical analysis starts with a linear equation as below:

colGPA = $\beta_0 + \beta_1$*hsGPA+ $\beta_2$shsGPA+ ε

To start our empirical analysis, first, we did bivariate regression for our dependent, main independent and square of main independent variable

colGPA = 9.224067 - 4.132889*hsGPA+ 0.6762422shsGPA + ε

We get initial results as shown below in table 4 after our simple regression. So, the initial outcome depicts that one unit increase on hsGPA will increase colGPA by 0.4784.

Number of obs = 134

F(2, 131) = 14.93

Prob > F = 0.0000

R-squared = 0.1856

Root MSE = .33731

colGPA	Coef.	Std. Error	t	P>|t|	[95% Conf.Interval]
hsGPA	. -4.132889	. 1.947185	-2.12	0.036	-7.984886 -.2808925
shsGPA	.6762422	.2851961	2.37	0.019	.1120563 1.240428
_cons	9.224067	3.306532	2.79	0.006	2.682958 15.76518

First of all, we need to check whether there is a heteroskedasticity issue with the Breusch-Pagan test:

H0 = there is no heteroskedasticity issue

Chi2(1)	0.63
Prob>chi2	0.4262

From the table we see that our p-value is 0.4262, which is higher than 0.05. This indicates that we fail to reject H0 and conclude that we have no heteroskedasticity issue.

Decision-making rules:

There are 3 methods of decision making rule for T-test. The first one is T statistics:

H0: βi = 0 (statistically not significant)

Ha: βi : is not equal to 0 (statistically significant)

We will reject H0 if tstat > tcritical

T stat is equal to coefficient / standard error which is -2.12 and 2.37. To find t critical, we need to look at the T distribution table. Degrees of freedom in this case is 132. We need to check-in table 2 tailed test and 0.05 significance level. We will get 1.96. Our t-stat for both variable bigger than t-critical, so we reject H0 and it means hsGPA and shsGPA are statistically significant on colGPA.

The second method for decision making is p-value. We reject H0 if p value is less than alpha (alpha is a significance level. Our p value is 0.036 and 0.019 and the significance level is 0.05, so we reject H0 and get the same result).

Third method is the confidence interval. We need to reject the H0 if confidence interval does not contain the 0. Our confidence interval is -7.984886 -.2808925 As we can see from here, our confidence interval does not contain zero and we need to reject H0.

When we find partial derivative of our regression we see that our result is not logical. Our first partial derivative is . -4.132889 + 1.3524844hsGPA. The mean of hsgpa which is 3.419403. Now we will insert mean into the equation and : -4.132889 + 1.3524844*3.419403=0.4918002. Because we have coefficient of hsGPA<0 and shsGPA>0 we will have U shaped parabola. In order to find turning point we will divide 4.132889/1.3524844= 3.055775. Increase in hsGPA under 3.05 will decrease colGPA(we assume it is because of error in data). So we conclude that one point increase in hsgpa if it is higher than turning point (3.05)will rise colGPA by 0.4918002.

MLR Assumptions:

For efficiency of our research, it has to meet Gauss Markov Assumptions or best linear unbiased estimators. Now we will provide detailed information about five MLR assumptions one by one:

1) From our population sample it is evident that the relation between independent and dependent variables are linear.

2) Second MLR assumption states that we should have a random sample of *n* observations in our population model. As we discussed on the data description part, we didn't collect data by ourselves and used "GPA" dataset for our research, which can lead to our model not fully meeting MLR 2 assumption. Because we don't have detailed background about how data collected we can not fully say it doesn't meet.

3) Third assumption of MLR states that there should not appear any collinearities between explanatory variables is also meet by our dataset. On empirical analysis part we did multicollinearity test and observed that the given table shows values that less than 10. Thus result allows us to mention that we don't have perfect collinearity between independent variables in our data.

Variable	**VIF**	**1/VIF**
ACT	1.18	0.844671
hsGPA	1.18	0.848497
have_pc	1.06	0.945859
skipped	1.06	0.946372

5) Fifth assumption of MLR states about homoskedasticity which means there should not be any relationship between independent variables and variance of unobserved factors. We will easily meet this assumption by just using the Breusch-Pagan test. We did this test on empirical analysis part using test and reject H0 which means we had ***heteroskedasticity issue*** to solve that we used robust command.

Multiple Linear Regression

However, to get a clearer result of the relation between hsGPA and colGPA we need to include more explanatory variables in our regression. As we mentioned before, we have five independent

variables that can influence college GPA. Our empirical analysis will be done by cross sectional dataset by using linear estimation. As shown below is our first linear equation:

colGPA = β_0 + β_1*hsGPA+ β_2*ACT + β_3*have_pc + β_4*mch_grad + β_5* skipped + ε

Before regression we have to check if there is a heteroskedasticity issue:

H0 = there is not heteroskedasticity issue

Ha = there is a heteroskedasticity issue,

we reject H0 if p value < 0.05

We need to use **estat hettest** command to find p value:

Chi2(1)	4.03
Prob>chi2	0.0447

Using the Breusch-Pagan test we find that there is a heteroskedasticity issue. From the table we see that p value is smaller than 0.05 so we reject H0 and discover that we have heteroskedasticity issue. To solve ***heteroskedasticity issue,*** we use robust command. Also, we checked whether if our model has ***a multicollinearity issue***. For that we used **the VIF** command. We found no problem because of shallow vif values (lower than 10).

Variable **VIF** **1/VIF**

Mean VIF 1.1

reg colGPA hsGPA ACT skipped have_PC mich_grad, robust

colGPA	***Coefficient***	***Standart Deviation***	***T statistics***	***P value***	***Confidence Interval***	
hsGPA	0.3932372	0.1143496	3.44	0.001	0.1669771	0.6194974
ACT	0.183072	0.120894	1.51	0.132	-0.0056137	0.0422281
skipped	-.0883878	0.0285168	-3.10	0.002	-0.1448131	-0.0319625
have_PC	0.1063925	0.0612556	1.74	0.085	-0.014812	0.2275971
mich_grad	0.1770632	0.0879402	2.01	0.046	0.0030584	0.3510679
_cons	1.164886	0.4177595	2.79	0.006	0.3382777	1.991495

Number of observations = 134

Prob > F = 0.0000

F(6, 89) = 8.98

R squared = 0.2742

Our multiple linear regression is:

colGPA = 1.164886 + 0.3932372 *hsGPA+ 0.183072 *ACT -0.0883878* skipped + 0.1063925 *have_pc + 0.1770632*mch_grad + ε

Interpretations:

1 point increase in high school GPA, is expected to increase college GPA by 0.3932372under ceteris paribus condition.

1 score increase in achievement score, is expected to increase college GPA by 0.183072under

ceteris paribus condition.

1 more missed lecture per week in skipped, is expected to decrease college GPA by -0.0883878 under ceteris paribus.

Students with PC is expected to have 0.1063925 more college GPA than students without PC, under ceteris paribus condition.

We have seen that, only **skipped variable** has negative impact on college GPA. From our table, we can report that, high- school GPA, number of missed classes per week and graduating from Michigan (prestigious high school) have statistically significant college GPA at 5% significance level. Now, we will conduct t-test for each independent variable to find if they have individual significant impact on college GPA. So, we will have these hypotheses for each explanatory variable:

H_0: $\beta_I = 0$ (Independent variable does NOT have significant effect on college GPA)

H_1: $\beta_I \neq 0$ (Independent variable has significant effect on college GPA)

We have 3 methods to test whether it has significant impact: p value, t statistics, and confidence interval.

P value test

First of all, we are checking high-school GPA. Our decision-making rule is that, when p value is less than 0.05 (significance level), we reject null hypotheses and conclude that the independent variable will have a statistically significant effect on the dependent variable. As we see from the table p value for high school is 0.001, which is less than 0.05, we reject the H_0 and conclude that high-school GPA is statistically significant on college GPA.

Second independent variable is ACT, now we are conducting p test for ACT. As we see from the table, p value for ACT is 0.132, which is higher than 0.05, so we do not reject H_0, and conclude that, ACT has not statistical significance on college GPA.

Third independent variable is skipped, now we are conducting p test for skipped. As we see from the table, p value for skipped is 0.002, which is less than 0.05, so we reject H_0, and conclude that, skipped has statistical significance on college GPA.

Fourth independent dummy variable is have_pc, now we are conducting p test for have_pc. As we see from the table, p value for have_pc is 0.085, which is higher than 0.05, so we do not reject H_0, and conclude that, have_pc has not statistical significance on college GPA.

Our last independent dummy variable is mich_grad, now we are conducting p test for mich_grad. As we see from the table, p value for mich_grad is 0.046, which is less than 0.05, so we reject H_0, and conclude that, mich_grad has statistical significance on college GPA.

T-Test

After finishing with p test, we will start working on t test. Our decision-making rule will tailed test with 0.05 significance level at 134 - 5 – 1 = 128 degrees of freedom, which we will take 1.960 from table as our critical value. We will reject H_0 if tstat > tcritical.

We will start with the high-school GPA independent variable. From table we see that t statistics for high-school GPA is 3.44, which is higher than 1.96. In conclusion, we reject the H_0, and conclude that, high-school GPA is statistically significant on college GPA.

Recommendations Discover more

Project

How the unavoidable debt should be financed?

Kamran Sadigli · Rauf Zeynalli

View project

Article Full-text available

Effectiveness of Audit Committee Attributes on Firm Performance of CNX Nifty Companies in India

June 2021 · MUDRA Journal of Finance and Accounting

Mahesh Bendigeri · Raju L Hyderabad

This research paper has been undertaken to examine the effectiveness of audit committee attributes on the firm performance of CNX Nifty listed firms in India. A sample of 45 companies covering 7 years data has been selected. Audit committee size, independent directors, women director, frequency of audit committee meeting, director qualification are selected as independent variables. To measure ... [Show full abstract]

View full-text

Conference Paper Full-text available

Owner-Manager's Innovativeness and Innovation in Small Firms: A Study in Manufacturing Firms

March 2014

Suchi Dubey

The available literature in innovation and performance are mostly related to big firms and meager resources are available over its relationship in small and medium size firms. This research paper is an attempt to inquire owner-managers innovativeness on business innovation of small firms. Innovativeness constructs which is mostly used in consumer research studies were adopted to measure the ... [Show full abstract]

View full-text

Article Full-text available

How the unavoidable debt should be financed?

July 2022

Kamran Sadigli · Rauf Zeynalli · Elchin Babayev · [...] · Emin Mammadov

View full-text

Research Full-text available

Free trade agreements signed by Turkey and its role in Turkey foreign trade

September 2020

Rauf Zeynalli

Free trade agreements signed by Turkey and its role in Turkey foreign trade

View full-text

Rauf Zeynalli

The role of children in Ancient Rome, family structure during that time

Article

The relationship of track and field to variable components of physical fitness /

George Everett. Sinor

Typescript. Research Paper (M.A.)--Austin Peay State University, 1968. Includes bibliographical references (leaves [30]-32).

Article Full-text available

Analyzing Relationship between Title Length of Research Papers and Number of Authors

February 2010

Jatinderkumar R. Saini