

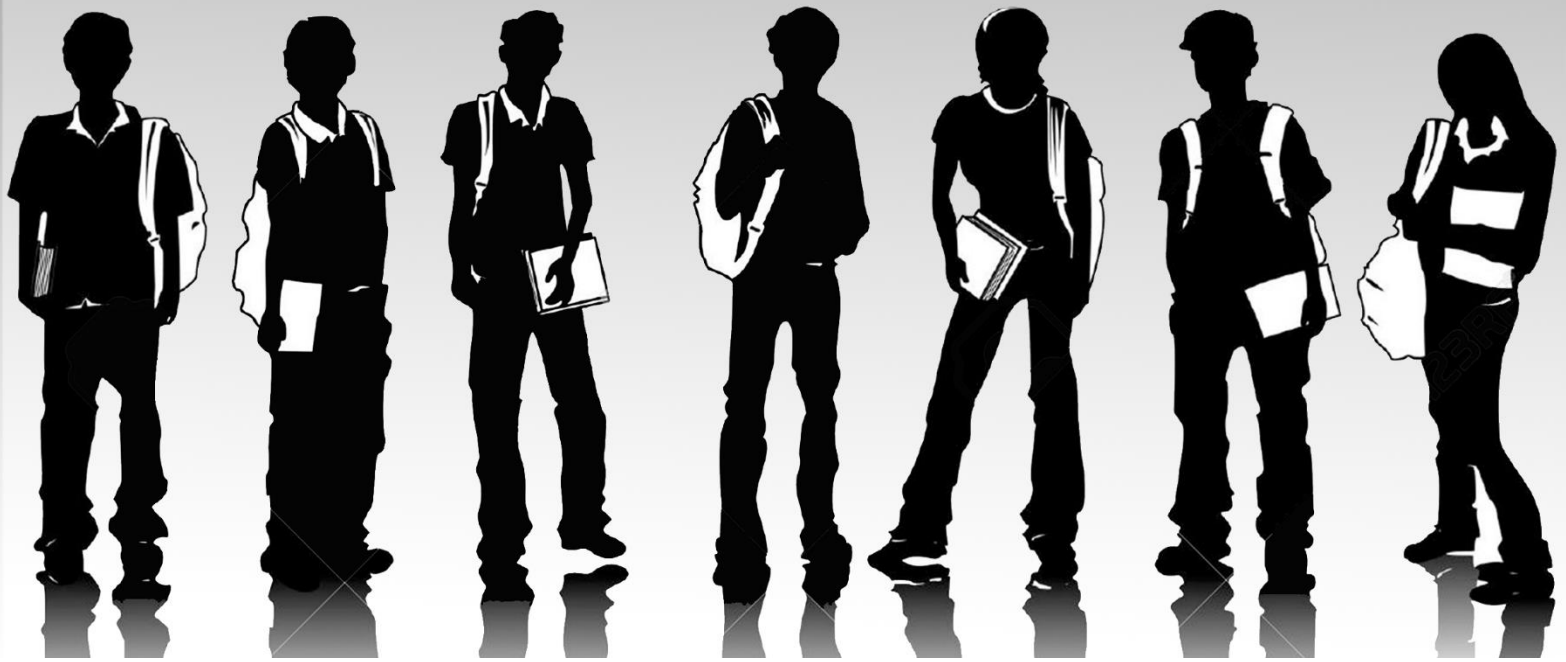
An ornate, symmetrical decorative border in a dark gray color frames the central text. It features intricate scrollwork, floral motifs, and a central crest-like element at the top.

# VARIOUS FACTORS AFFECTING COLLEGE GPA

an econometric study



Antonio Jurlina Joe Romano Jacob Wood



## PART I

### 1. Data

Project data consists of a random sample of student records (and a self-reported survey) for 141 University of Maine students from 2014. The data set was revised and cleaned, to attain maximum usability and incur minimal loss in valuable observations. Table 1. details all the relevant data corrections. Variables regarding the average number of lectures missed per week and the average number of alcoholic drinks consumed per week were assumed to have a biased pattern of missing values (skewed towards the higher end). Therefore, as Table 1. notes, these variables were corrected with conditional means. We didn't feel comfortable speculating about the reasons behind some missing values in variables describing numbers of work hours per week. This, in turn, led to us dropping observations (15 in total) which had values missing for both work-hour-related variables.

### 2. Methodology

With careful consideration and extensive literature review, we settled on a linear model (similar to that of Balsa, et al. 2011 and Lindo, et al. 2011) for OLS regression. The linear relationship between college GPA and various factors that affect it is useful in revealing a unit-by-unit effect of multiple variables, since we feel that each additional unit under consideration (for all relevant variables) will have a linearly increasing (or decreasing) effect. In case of binary variables, our coefficients will provide intercept shifts (premiums) that we might expect from certain variables. The following equation captures some of the relationships of interest:

$$colGPA = \beta_0 + \beta_1 alcohol + \beta_2 greek + x\delta + u,$$

where  $x\delta$  represents all additional factors being controlled for (see Table 3.) and  $\beta$  parameters represent the coefficients to be estimated. We assumed that our set of explanatory variables (detailed summary in Table 2.) will result in a model with heteroskedastic errors. This is mostly due to intuitive reasoning that students involved with various organizations, clubs, activities and groups will also be the ones with smaller GPA variance (toward the higher end) compared to students with more free time that have less (assumed) structure and discipline and will have their GPA vary greatly around the mean, subject to random chance and unforeseen events. Keeping this intuition in mind, we performed White's General Test and found that our errors were, indeed, heteroskedastic. To correct for this, we used White's robust standard errors in our linear regression model.

## **PART II**

### **1. Results**

#### *1.1. Alcohol consumption and greek life*

Students tend to exist within a socially rich lifestyle consisting of multiple outlets in direct relationship with individual preferences. Therefore, any effect that a particular activity might have on college GPA, must be described relative to another appropriate activity and/or attribute (or a multitude of them). This is necessary for direct comparison, as well as the isolation of effects an activity might have on its own (keeping all other effects in check). With that in mind, our results pertaining to alcohol consumption and greek life participation will be presented in several simplified ways. Further details can be found in Table 3.

First, our results for average weekly consumption of alcohol (with all studied effects accounted for) were not significant enough to allow for any substantial claims to be made. We

compared alcohol consumption between male and female students to determine whether there was a certain effect associated with either sex. The results seem to indicate no significant effect on average GPA from alcohol consumption, for either male or female students. However, we do not feel confident enough to promulgate such claims. We encountered the same problem with results on greek life participation. Being a member of a sorority or a fraternity showed no significant impact on average GPA (holding everything else constant) and even if it did, we would not feel confident enough in our results to substantiate such claims. Finally, we observed no clear difference in effects alcohol consumption might have between members and non-members of greek organizations.

### *1.2. Athletics*

Although we are not confident enough to generalize and promote this claim, we have observed a potential effect athletics might have on average GPA. We noticed that athletes attain a 0.15 average premium on GPA, holding everything else constant. However, due to the limitations of our results, we are unable to say whether this effect is positive or negative.

### *1.3. Other academic and non-academic activities*

The most significant predictor of college GPA, based on our results, was high school GPA. We observed, while holding everything else constant, that a unit increase in high school GPA corresponds to almost 0.5 points increase in college GPA, on average. This observation comes with a high degree of confidence associated with it (see Table 3.). Before interpreting these results further, we wish to assess an additional variable of interest – whether a student comes from a Maine high school or not. Our results indicate a clear positive premium of 0.2 on GPA, holding

everything else constant, that students coming to University of Maine as graduates of Maine high schools can expect, on average.

Literature review indicates that high school GPA is a decent predictor of college GPA, but mostly on the higher end of grading spectrum (Noble and Sawyer, 2002). Additionally, high school GPA is a better predictor of college GPA than ACT scores are (Sawyer, 2013). With this in mind, our results surpass usual literature expectations and we can speculate as to the reasons why. In our data set, 111 students were Maine high school graduates, while only 15 were not. Besides a small sample of out-of-state students, we believe that another factor might contribute towards our overly explanatory results. Given a fairly unimpressive standing University of Maine holds in the minds of high achieving, out-of-state students, it is reasonable to assume that the University receives applications from applicants with poorer academic standing. Contrasting that, many higher achieving Maine residents will gravitate towards lower costs of in-state tuition. Summing it up, the average GPA of Maine residents is expected to be higher on average and, therefore, our results might overestimate sets of applicants with higher numbers of out-of-state students.

Furthermore, our focus shifts towards study groups and the impact that belonging to at least one might have on average GPA (see Table 3.). We found that students belonging to at least one study group (holding everything else constant), can expect a 0.15 positive premium on their GPA, on average. The reasoning seems straightforward – those participating in study groups will have a stronger academic drive and will be exposed to ways other students might solve same problems. In addition, the informal setting of study groups is more conducive to idea exchange and it side-steps the free-rider issue which is inherent to mandatory projects (Jain and Kapoor, 2015).

Finally, we observed a significant effect the average number of classes skipped per week has on GPA (see Table 3.). Our results indicate that for every additional increase in average classes

per week missed (holding everything else constant), students suffer a 0.08 reduction in GPA, on average. Common reasons include sleeping-in, lack of interest in class material and using the time to work on material for another class (Dobkin, et al. 2007). Literature speculates that missing class introduces a source discrepancy between the nuance-rich material covered in class and dryer material introduced in textbooks. We would add a caveat that large variance between instructors might introduce a lack of clarity in results. Additionally, we must point out that our sample contains 99 students majoring in business and only 27 students majoring in something other than business.

## **2. Limitations**

We started out with a sample consisting of 141 observations. This rather small sample was further reduced after we excluded 15 more observations that had missing values. Size of our sample implies small variability and originality in our data, making our conclusions less significant, overall. Further adding to the reduced variability, we chose to fill in some missing values with average values for that category. Moreover, our estimates of relationships between variables of interest and college GPA were artificially improved when we filled in some missing values with category averages contingent on other observations. Also, as noted in Part 1 of the report, the differences between observed and predicted and expected and predicted values for variables of interest were not uniform across our data.\* This means that we are more likely to falsely perceive an observation as insignificant and reject its validity and explanatory power. Finally, we wish to point out a possibility that any variable not considered stands a chance of being correlated with variables we did consider (e.g. inherent ability of an individual might be a predictor of GPA). By omitting any such variable, we introduce the likelihood of overstating the effects considered variables have on GPA.†

***Word Count: 1450***

\* heteroskedasticity

† endogeneity

### **Works Cited:**

- Balsa, A. I., Giuliano, L. M., & French, M. T. (2011). *The effects of alcohol use on academic achievement in high school*. *Economics of Education Review*, 30(1), 1-15. doi:10.1016/j.econedurev.2010.06.015
- Dobkin, C., Gil, R., & Marion, J. (2007). *Causes and Consequences of Skipping Class in College*. Economics Department, UC Santa Cruz. Retrieved April 09, 2018, from [https://www.researchgate.net/publication/245622925\\_Causes\\_and\\_Consequences\\_of\\_Skiping\\_Class\\_in\\_College](https://www.researchgate.net/publication/245622925_Causes_and_Consequences_of_Skiping_Class_in_College)
- Jain, T., & Kapoor, M. (2015). *The Impact of Study Groups and Roommates on Academic Performance*. *Review of Economics and Statistics*, 97(1), 44-54. doi:10.1162/rest\_a\_00454
- Lindo, J., Swensen, I., & Waddell, G. (2011). *Alcohol and Student Performance: Estimating the Effect of Legal Access*. *Journal of Economic Literature*, 18(21), 32nd ser. doi:10.3386/w17637
- Noble, J., & Sawyer, R. (2002). *Predicting Different Levels of Academic Success in College Using High School GPA and ACT Composite Score*. ACT Research Report Series.
- Richard Sawyer (2013) *Beyond Correlations: Usefulness of High School GPA and Test Scores in Making College Admissions Decisions*, *Applied Measurement in Education*, 26:2, 89-112, DOI: 10.1080/08957347.2013.765433

**Table 1: Data clean-up summary**

Variable	Cleaning performed	Method
colGPA	No	N/A
age	No	N/A
soph	No	N/A
junior	No	N/A
senior	No	N/A
senior5	No	N/A
male	No	N/A
campus	No	N/A
major	No	N/A
hsGPA	No	N/A
ACT	No	N/A
job19	Yes	drop & unconditional mean
job20	Yes	drop & unconditional mean
athlete	No	N/A
study	No	N/A
greek	No	N/A
car	No	N/A
romantic	Yes	unconditional mean
clubs	No	N/A
skipped	Yes	conditional mean
alcohol	Yes	conditional mean
gradME	No	N/A
fathcoll	Yes	unconditional mean
mothcoll	Yes	unconditional mean

- Summary of methods used to clean our data



**Table 2: General summary**

Variable	Obs	Mean	Std. Dev.	Min	Max
colGPA	126	3.05873	0.3651345	2.2	4
age	126	20.94444	1.316392	19	30
soph	126	0.015873	0.1254832	0	1
junior	126	0.3650794	0.4833744	0	1
senior	126	0.515873	0.501743	0	1
senior5	126	0.1031746	0.3054011	0	1
male	126	0.531746	0.5009832	0	1
campus	126	0.1746032	0.3811428	0	1
major	126	0.7857143	0.4119639	0	1
hsGPA	126	3.389683	0.302742	2.5	4
ACT	126	24	2.814249	16	31
job19_clean	126	0.4152542	0.4787696	0	1
job20_clean	126	0.1587302	0.3668831	0	1
athlete	126	0.2142857	0.4119639	0	1
study	126	0.4285714	0.4968472	0	1
greek	126	0.3015873	0.4607792	0	1
car	126	0.7936508	0.4062996	0	1
romantic	126	0.4799887	0.4995999	0	1
clubs	126	0.6111111	0.4894441	0	1
skipped_clean	126	1.051778	1.090061	0	5
alcohol_clean	126	1.913866	1.317166	0	7
gradME	126	0.8809524	0.3251373	0	1
fathcoll_clean	126	0.622807	0.4628667	0	1
mothcoll_clean	126	0.5603448	0.478143	0	1

- General summary of all variables used for regression (with mean, standard deviation, minimum and maximum values listed)

**Table 3: Regression results**

Number of obs	126
F(25, 100)	6.44
Prob > F	0
R-squared	0.4297
Root MSE	0.30829

colGPA	Coef.	Robust Std. Err.	t	P> t	95 % Confidence	Interval
age	0.021251	0.0320725	0.66	0.509	-0.0423796	0.0848822
soph	0.057362	0.4153544	0.14	0.89	-0.7666895	0.8814132
junior	-0.08574	0.1133371	-0.76	0.451	-0.3106001	0.1391152
senior	-0.06369	0.0913556	-0.7	0.487	-0.2449368	0.1175569
senior5	0	(omitted)				
greek	-0.21134	0.1187554	-1.78	0.078	-0.4469436	0.0242711
male	-0.01464	0.1266708	-0.12	0.908	-0.2659501	0.2366723
greek x male	0.017986	0.1443904	0.12	0.901	-0.2684807	0.3044523
campus	-0.13956	0.0826665	-1.69	0.094	-0.3035691	0.0244467
major	0.112604	0.0729317	1.54	0.126	-0.0320905	0.2572983
hsGPA	0.465165	0.1115209	4.17	0	0.2439105	0.6864193
ACT	0.006732	0.0121163	0.56	0.58	-0.0173065	0.0307704
job19_clean	-0.00053	0.0685132	-0.01	0.994	-0.136461	0.1353956
job20_clean	-0.04827	0.100433	-0.48	0.632	-0.2475293	0.1509832
athlete	-0.15225	0.0833446	-1.83	0.071	-0.3176	0.0131066
study	0.1539	0.0594459	2.59	0.011	0.0359611	0.2718391
car	-0.04157	0.0724444	-0.57	0.567	-0.185296	0.1021594
romantic	0.066255	0.0626001	1.06	0.292	-0.0579417	0.1904519
clubs	0.061457	0.0595272	1.03	0.304	-0.0566436	0.1795571
skipped_clean	-0.07794	0.0288525	-2.7	0.008	-.1351834	0.0206985
alcohol_clean	-0.07649	0.0538082	-1.42	0.158	-0.1832405	0.0302673
male x alcohol_clean	0.044754	0.0593827	0.75	0.453	-0.0730601	0.1625672
greek x alcohol_clean	0.131238	0.0460115	2.85	0.005	0.0399522	0.2225232
gradME	0.214187	0.0732499	2.92	0.004	0.068861	0.3595124
fathcoll_clean	0.067925	0.0694952	0.98	0.331	-0.0699519	0.2058011
mothcoll_clean	-0.10811	0.0658437	-1.64	0.104	-0.2387411	0.0225229
_cons	0.793883	0.9209647	0.86	0.391	-1.033285	2.62105

- Summary of OLS estimates, their standard deviations and significance at the 5% level (several interaction terms were generated for the purpose of answering project-specific questions)

**Table 4: Variance inflation factor**

Variable	VIF	1/VIF	$\sqrt{VIF}$
age	2.02	0.495944	1.421267
soph	1.62	0.617838	1.272792
junior	4.66	0.214775	2.158703
senior	3.55	0.281868	1.884144
greek	4.78	0.209043	2.186321
male	4.87	0.205372	2.206808
greek x male	3.78	0.264866	1.944222
campus	1.37	0.731928	1.17047
major	1.35	0.742253	1.161895
hsGPA	1.59	0.628599	1.260952
ACT	1.42	0.701952	1.191638
job19_clean	1.41	0.707642	1.187434
job20_clean	1.59	0.630067	1.260952
athlete	1.38	0.724699	1.174734
study	1.21	0.828686	1.1
car	1.26	0.791349	1.122497
romantic	1.26	0.795813	1.122497
clubs	1.2	0.829877	1.095445
skipped_clean	1.4	0.713442	1.183216
alcohol_clean	6.5	0.153892	2.54951
male x alcohol	11	0.090894	3.316625
greek x alcohol	6.15	0.162695	2.479919
gradME	1.16	0.860434	1.077033
fathcoll_clean	1.44	0.692915	1.2
mothcoll_clean	1.46	0.68314	1.208305

<b>Mean VIF</b>	<b>2.78</b>
-----------------	-------------

- Summary of variance inflation factors (due to heteroskedasticity) for each variable used in our regression model

## Appendix A: Stata log-file

```

-----
name: <unnamed>
log: D:\UMaine\Spring (2018)\ECO 485\Stata\Empirical_Project_2.log
log type: text
opened on: 11 Apr 2018, 08:13:26

.
. use UMaine_GPA.dta //choose data

.
. // following commands represent our data clean-up process
. ///////////////////////////////////////////////////////////////////
> drop u1 u2 u3 u4 u5 u6 u7 u8

. drop drive bike walk sibling

. misstable summarize

```

Variable	Obs=.	Obs>.	Obs<.	Unique values	Min	Max
job19	23		118	2	0	1
job20	15		126	2	0	1
romantic	1		140	2	0	1
skipped	14		127	8	0	5
alcohol	22		119	13	0	7
fathcoll	13		128	2	0	1
mothcoll	10		131	2	0	1

```

-----

.
. summarize romantic

```

Variable	Obs	Mean	Std. Dev.	Min	Max
romantic	140	.4785714	.5013343	0	1

```

. replace romantic = r(mean) if romantic == .
variable romantic was byte now float
(1 real change made)

.
. generate job19_clean = job19
(23 missing values generated)

. generate job20_clean = job20
(15 missing values generated)

. drop if job19 == . & job20 == .
(15 observations deleted)

. summarize job19_clean

```

Variable	Obs	Mean	Std. Dev.	Min	Max
job19_clean	118	.4152542	.4948672	0	1

```

. replace job19_clean = r(mean) if job19_clean == .
(8 real changes made)

.
. generate mothcoll_clean = mothcoll
(10 missing values generated)

. summarize mothcoll_clean

```

Variable	Obs	Mean	Std. Dev.	Min	Max
mothcoll_c~n	116	.5603448	.4984985	0	1

```
. replace mothcoll_clean = r(mean) if mothcoll_clean == .
(10 real changes made)
```

```
. generate fathcoll_clean = fathcoll
(12 missing values generated)
```

```
. summarize fathcoll_clean
```

Variable	Obs	Mean	Std. Dev.	Min	Max
fathcoll_c~n	114	.622807	.4868237	0	1

```
. replace fathcoll_clean = r(mean) if fathcoll_clean == .
(12 real changes made)
```

```
. generate majory = "business"
```

```
. replace majory = "other" if business == 0
(27 real changes made)
```

```
. label define major 1 "business" 0 "other"
```

```
. encode majory, generate(major)
```

```
. drop majory
```

```
. generate alcohol_clean = alcohol
(17 missing values generated)
```

```
. regress alcohol age soph junior senior senior5 male campus major hsGPA ACT job19_clean
job20_clean athlete study car sk
> iped romantic clubs greek gradME fathcoll_clean mothcoll_clean, robust
note: senior5 omitted because of collinearity
```

Linear regression	Number of obs	=	100
	F(21, 78)	=	4.55
	Prob > F	=	0.0000
	R-squared	=	0.3798
	Root MSE	=	1.2174

	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
alcohol						
age	-.1133956	.1380473	-0.82	0.414	-.3882267	.1614354
soph	-.438	.6789179	-0.65	0.521	-1.789622	.9136216
junior	-.4757598	.5101044	-0.93	0.354	-1.4913	.53978
senior	.3356907	.4087754	0.82	0.414	-.4781186	1.1495
senior5	0	(omitted)				
male	.7565491	.3317633	2.28	0.025	.096059	1.417039
campus	-.8865903	.2659068	-3.33	0.001	-1.41597	-.3572106
major	.1201771	.3510906	0.34	0.733	-.5787906	.8191447
hsGPA	.3709321	.5121592	0.72	0.471	-.6486986	1.390563
ACT	.0116831	.0460052	0.25	0.800	-.0799062	.1032724
job19_clean	-.1743867	.3033803	-0.57	0.567	-.7783705	.4295971
job20_clean	-.3482278	.3242186	-1.07	0.286	-.9936976	.2972419
athlete	.0849272	.3032069	0.28	0.780	-.5187114	.6885658
study	-.0542044	.2873907	-0.19	0.851	-.6263554	.5179466
car	.3059174	.2867401	1.07	0.289	-.2649383	.876773
skipped	.3599371	.1221327	2.95	0.004	.1167896	.6030846
romantic	-.1508504	.2998294	-0.50	0.616	-.7477647	.446064
clubs	.2672687	.275573	0.97	0.335	-.2813549	.8158923
greek	.031575	.328727	0.10	0.924	-.6228701	.6860201
gradME	.1618562	.3055574	0.53	0.598	-.4464618	.7701742

```

fathcoll_clean | .3040568 .278426 1.09 0.278 -.2502469 .8583605
mothcoll_clean | .3031311 .2681988 1.13 0.262 -.2308116 .8370738
_cons | 1.333749 4.310867 0.31 0.758 -7.248529 9.916027
-----

```

```

. predict alcoholhat
(option xb assumed; fitted values)
(11 missing values generated)

```

```

. summarize alcohol

```

```

Variable | Obs Mean Std. Dev. Min Max
-----+-----
alcohol | 109 1.915596 1.417037 0 7

```

```

. summarize alcoholhat

```

```

Variable | Obs Mean Std. Dev. Min Max
-----+-----
alcoholhat | 115 1.902775 .8387463 .1706877 4.109898

```

```

. replace alcohol_clean = r(mean) if alcohol_clean == .
(17 real changes made)

```

```

. drop alcoholhat

```

```

. generate skipped_clean = skipped
(11 missing values generated)

```

```

. regress skipped age soph junior senior senior5 male campus major hsGPA ACT job19_clean
job20_clean athlete alcohol stud
> y car romantic clubs greek gradME fathcoll_clean mothcoll_clean, robust
note: senior5 omitted because of collinearity

```

```

Linear regression                               Number of obs   =       100
                                                F(21, 78)         =       1.52
                                                Prob > F          =       0.0953
                                                R-squared         =       0.2991
                                                Root MSE         =       1.0947

```

```

-----+-----
skipped | Coef. Robust Std. Err. t P>|t| [95% Conf. Interval]
-----+-----
age | .0143772 .1256326 0.11 0.909 -.2357382 .2644926
soph | .4125938 1.478162 0.28 0.781 -2.5302 3.355388
junior | .0916726 .4646303 0.20 0.844 -.8333352 1.01668
senior | -.2290638 .4339745 -0.53 0.599 -1.093041 .6349131
senior5 | 0 (omitted)
male | .0840323 .2491577 0.34 0.737 -.4120025 .5800671
campus | -.1957296 .2542043 -0.77 0.444 -.7018114 .3103523
major | -.2207643 .3020908 -0.73 0.467 -.8221809 .3806523
hsGPA | -.4897715 .4153194 -1.18 0.242 -1.316609 .3370659
ACT | .037708 .0491215 0.77 0.445 -.0600854 .1355015
job19_clean | -.0652601 .2889345 -0.23 0.822 -.6404846 .5099643
job20_clean | -.0221759 .3121515 -0.07 0.944 -.6436217 .5992699
athlete | -.4404475 .2104252 -2.09 0.040 -.8593719 -.0215231
alcohol | .2910622 .1281268 2.27 0.026 .0359813 .5461431
study | -.1796578 .2265819 -0.79 0.430 -.6307478 .2714321
car | .0249061 .3023714 0.08 0.935 -.577069 .6268813
romantic | .0504093 .2422552 0.21 0.836 -.4318837 .5327023
clubs | -.1018344 .2979852 -0.34 0.733 -.6950773 .4914085
greek | .3001129 .3102684 0.97 0.336 -.3175841 .9178099
gradME | -.3389273 .3477308 -0.97 0.333 -1.031206 .3533515
fathcoll_clean | -.2443219 .2701136 -0.90 0.369 -.7820768 .293433
mothcoll_clean | -.4291269 .2445282 -1.75 0.083 -.9159452 .0576913
_cons | 2.026876 3.462074 0.59 0.560 -4.865583 8.919336
-----

```

```

. predict skippedhat

```

```
(option xb assumed; fitted values)
(17 missing values generated)
```

```
. summarize skipped
```

Variable	Obs	Mean	Std. Dev.	Min	Max
skipped	115	1.05	1.141425	0	5

```
. summarize skippedhat
```

Variable	Obs	Mean	Std. Dev.	Min	Max
skippedhat	109	1.07037	.676979	-.4186085	2.536748

```
. replace skipped_clean = r(mean) if skipped_clean == .
(11 real changes made)
```

```
. drop skippedhat
```

```
. ////////////////////////////////////////////////////
> // end of data clean-up process
```

```
.
. summarize colGPA age soph junior senior senior5 male campus major hsGPA ACT job19_clean
job20_clean athlete study greek
> car romantic clubs skipped_clean alcohol_clean gradME fathcoll_clean mothcoll_clean // basic
data summary
```

Variable	Obs	Mean	Std. Dev.	Min	Max
colGPA	126	3.05873	.3651345	2.2	4
age	126	20.94444	1.316392	19	30
soph	126	.015873	.1254832	0	1
junior	126	.3650794	.4833744	0	1
senior	126	.515873	.501743	0	1
senior5	126	.1031746	.3054011	0	1
male	126	.531746	.5009832	0	1
campus	126	.1746032	.3811428	0	1
major	126	.7857143	.4119639	0	1
hsGPA	126	3.389683	.302742	2.5	4
ACT	126	24	2.814249	16	31
job19_clean	126	.4152542	.4787696	0	1
job20_clean	126	.1587302	.3668831	0	1
athlete	126	.2142857	.4119639	0	1
study	126	.4285714	.4968472	0	1
greek	126	.3015873	.4607792	0	1
car	126	.7936508	.4062996	0	1
romantic	126	.4799887	.4995999	0	1
clubs	126	.6111111	.4894441	0	1
skipped_cl~n	126	1.051778	1.090061	0	5
alcohol_cl~n	126	1.913866	1.317166	0	7
gradME	126	.8809524	.3251373	0	1
fathcoll_c~n	126	.622807	.4628667	0	1
mothcoll_c~n	126	.5603448	.478143	0	1

```
.
. regress colGPA age soph junior senior senior5 i.greek##i.male campus major hsGPA ACT
job19_clean job20_clean athlete st
> udy car romantic clubs skipped_clean c.alcohol_clean##i.male c.alcohol_clean#i.greek gradME
fathcoll_clean mothcoll_cle
> an
note: senior5 omitted because of collinearity
```

Source	SS	df	MS	Number of obs	=	126
Model	7.16099866	25	.286439947	F(25, 100)	=	3.01
Residual	9.50439804	100	.09504398	Prob > F	=	0.0001
				R-squared	=	0.4297

```
-----+-----
Total | 16.6653967      125  .133323174  Adj R-squared = 0.2871
Root MSE = .30829
```

colGPA	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.0212513	.0297444	0.71	0.477	-.0377608	.0802634
soph	.0573619	.279566	0.21	0.838	-.4972891	.6120128
junior	-.0857425	.1230926	-0.70	0.488	-.3299546	.1584697
senior	-.0636899	.1035149	-0.62	0.540	-.2690605	.1416807
senior5	0 (omitted)					
1.greek	-.2113363	.1308869	-1.61	0.110	-.4710122	.0483397
1.male	-.0146389	.1214545	-0.12	0.904	-.2556011	.2263233
greek#male						
1 1	.0179858	.1431958	0.13	0.900	-.2661105	.3020821
campus	-.1395612	.0845639	-1.65	0.102	-.3073337	.0282113
major	.1126039	.0776912	1.45	0.150	-.0415333	.2667411
hsGPA	.4651649	.1148809	4.05	0.000	.2372445	.6930853
ACT	.0067319	.0116947	0.58	0.566	-.0164701	.029934
job19_clean	-.0005327	.0684658	-0.01	0.994	-.1363669	.1353015
job20_clean	-.048273	.0946861	-0.51	0.611	-.2361275	.1395815
athlete	-.1522467	.0786265	-1.94	0.056	-.3082395	.0037461
study	.1539001	.0609663	2.52	0.013	.0329447	.2748554
car	-.0415683	.0762916	-0.54	0.587	-.1929287	.1097921
romantic	.0662551	.0618699	1.07	0.287	-.056493	.1890033
clubs	.0614568	.061844	0.99	0.323	-.0612399	.1841534
skipped_clean	-.0779409	.0299486	-2.60	0.011	-.1373582	-.0185236
alcohol_clean	-.0764866	.0533652	-1.43	0.155	-.1823617	.0293884
male#c.alcohol_clean						
1	.0447535	.0575008	0.78	0.438	-.0693263	.1588334
greek#c.alcohol_clean						
1	.1312377	.0531613	2.47	0.015	.0257672	.2367083
gradME	.2141867	.0914285	2.34	0.021	.0327951	.3955783
fathcoll_clean	.0679246	.0715668	0.95	0.345	-.0740618	.209911
mothcoll_clean	-.1081091	.0697742	-1.55	0.124	-.2465391	.0303208
_cons	.7938827	.8623142	0.92	0.359	-.9169241	2.504689

```
. estat imtest, white
```

```
White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity
```

```
chi2(125) = 126.00
Prob > chi2 = 0.4581
```

```
Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	126.00	125	0.4581
Skewness	18.53	25	0.8192
Kurtosis	4.60	1	0.0319
Total	149.13	151	0.5277

```
. estat vif
```

Variable	VIF	1/VIF
age	2.02	0.495944
soph	1.62	0.617838
junior	4.66	0.214775
senior	3.55	0.281868



```

1.greek |      4.78    0.209043
1.male  |      4.87    0.205372
greek#male |
1 1    |      3.78    0.264866
campus  |      1.37    0.731928
major   |      1.35    0.742253
hsGPA   |      1.59    0.628599
ACT     |      1.42    0.701952
job19_clean |      1.41    0.707642
job20_clean |      1.59    0.630067
athlete |      1.38    0.724699
study   |      1.21    0.828686
car     |      1.26    0.791349
romantic |      1.26    0.795813
clubs   |      1.20    0.829877
skipped_clean |      1.40    0.713442
alcohol_clean |      6.50    0.153892
male# |
c. |
alcohol_clean |
1 |      11.00    0.090894
greek# |
c. |
alcohol_clean |
1 |      6.15    0.162695
gradME |      1.16    0.860434
fathcoll_clean |      1.44    0.692915
mothcoll_clean |      1.46    0.683140
-----
Mean VIF |      2.78

```

```

. regress colGPA age soph junior senior senior5 i.greek##i.male campus major hsGPA ACT
job19_clean job20_clean athlete st
> udy car romantic clubs skipped_clean c.alcohol_clean##i.male c.alcohol_clean#i.greek gradME
fathcoll_clean mothcoll_clean
> an, robust
note: senior5 omitted because of collinearity

```

```

Linear regression          Number of obs   =      126
                          F(25, 100)       =       6.44
                          Prob > F          =      0.0000
                          R-squared         =      0.4297
                          Root MSE       =      .30829

```

```

-----
               |
               |      Coef.   Robust
colGPA         |      Std. Err.   t   P>|t|   [95% Conf. Interval]
-----+-----
age            |      .0212513   .0320725   0.66   0.509   -.0423796   .0848822
soph           |      .0573619   .4153544   0.14   0.890   -.7666895   .8814132
junior         |     -.0857425   .1133371  -0.76   0.451   -.3106001   .1391152
senior         |     -.0636899   .0913556  -0.70   0.487   -.2449368   .1175569
senior5        |      0 (omitted)
1.greek        |     -.2113363   .1187554  -1.78   0.078   -.4469436   .0242711
1.male         |     -.0146389   .1266708  -0.12   0.908   -.2659501   .2366723
greek#male     |
1 1            |      .0179858   .1443904   0.12   0.901   -.2684807   .3044523
campus         |     -.1395612   .0826665  -1.69   0.094   -.3035691   .0244467
major          |      .1126039   .0729317   1.54   0.126   -.0320905   .2572983
hsGPA          |      .4651649   .1115209   4.17   0.000   .2439105   .6864193
ACT            |      .0067319   .0121163   0.56   0.580   -.0173065   .0307704
job19_clean    |     -.0005327   .0685132  -0.01   0.994   -.136461   .1353956
job20_clean    |     -.048273    .100433   -0.48   0.632   -.2475293   .1509832
athlete        |     -.1522467   .0833446  -1.83   0.071   -.3176   .0131066
study          |      .1539001   .0594459   2.59   0.011   .0359611   .2718391
car            |     -.0415683   .0724444  -0.57   0.567   -.185296   .1021594
romantic       |      .0662551   .0626001   1.06   0.292   -.0579417   .1904519
clubs          |      .0614568   .0595272   1.03   0.304   -.0566436   .1795571

```

skipped_clean		-.0779409	.0288525	-2.70	0.008	-.1351834	-.0206985
alcohol_clean		-.0764866	.0538082	-1.42	0.158	-.1832405	.0302673
male#c.alcohol_clean							
1		.0447535	.0593827	0.75	0.453	-.0730601	.1625672
greek#c.alcohol_clean							
1		.1312377	.0460115	2.85	0.005	.0399522	.2225232
gradME		.2141867	.0732499	2.92	0.004	.068861	.3595124
fathcoll_clean		.0679246	.0694952	0.98	0.331	-.0699519	.2058011
mothcoll_clean		-.1081091	.0658437	-1.64	0.104	-.2387411	.0225229
_cons		.7938827	.9209647	0.86	0.391	-1.033285	2.62105

```

.
. save UMaine_GPA_clean.dta, replace
file UMaine_GPA_clean.dta saved

. log close
  name: <unnamed>
  log: D:\UMaine\Spring (2018)\ECO 485\Stata\Empirical_Project_2.log
  log type: text
  closed on: 11 Apr 2018, 08:13:30

```

<pre> //***** * Authors: Antonio J. Joe R. Jacob W * Date: 4/9/2019 * File: EMPIRICAL_PROJECT_2.do * *****/  clear //clear previous data  cd "D:\Online\Spring (2018)\ECO 485\Stata" //setting my directory  // Store results in a log file (diary) log using "D:\Online\Spring (2018)\ECO 485\Stata\Empirical_Project_2", replace text  use Online_CPA.dta //choose data  // Following commands represent our data clean-up process //////////////////// drop u1 u2 u3 u4 u5 u6 u7 u8 drop drive bike walk stb1ng misstable summarize  summarize romantic replace romantic = r(mean) if romantic == .  generate job19_clean = job19 generate job20_clean = job20 drop if job19 == 1 &amp; job20 == . summarize job19_clean job20_clean replace job19_clean = r(mean) if job19_clean == .  generate mothcoll_clean = mothcoll summarize mothcoll_clean replace mothcoll_clean = r(mean) if mothcoll_clean == . generate fathcoll_clean = fathcoll summarize fathcoll_clean replace fathcoll_clean = r(mean) if fathcoll_clean == .  generate majory = "Business" replace majory = "other" if business == 0 label define major 1 "Business" 0 "other" encode majory, generate(major) drop majory  generate alcohol_clean = alcohol regress alcohol age soph junior senior seniors male campus major hactra ACT job19_clean job20_clean athlete study car skipped romantic clubs greek grade fathcoll_clean mothcoll_clean, robust predict alcoholhat summarize alcoholhat replace alcohol_clean = r(mean) if alcohol_clean == . drop alcoholhat  generate skipped_clean = skipped regress skipped age soph junior senior seniors male campus major hactra ACT job19_clean job20_clean athlete study car skipped romantic clubs greek grade fathcoll_clean mothcoll_clean, robust predict skippedhat summarize skippedhat replace skipped_clean = r(mean) if skipped_clean == . drop skippedhat //////////////////// // end of data clean-up process  summarize colGPA age soph junior senior seniors male campus major hactra ACT job19_clean job20_clean athlete study greek car romantic clubs skipped_clean alcohol_clean grade fathcoll_clean mothcoll_clean // basic data summary  regress colGPA age soph junior senior seniors 1.greek###1.male campus major hactra ACT job19_clean job20_clean athlete study car romantic clubs skipped_clean c.alcohol_clean###1.male c.alcohol_clean###1.greek grade fathcoll_clean mothcoll_clean estat lmtest, white  estat vif  regress colGPA age soph junior senior seniors 1.greek###1.male campus major hactra ACT job19_clean job20_clean athlete study car romantic clubs skipped_clean c.alcohol_clean###1.male c.alcohol_clean###1.greek grade fathcoll_clean mothcoll_clean, robust save Online_CPA_clean.dta, replace log close </pre>	
--	--

Appendix B: Stata Do-file