Empirical Project - Data Analysis (12 pages without the do.file)

Alexia Ge, Bianca Lewis, Nick Araya, Chun Nok Larry Lo, Miles Ma

I. Title

The effect of caffeine consumption on students' school success measured by grade point average.

II. Research Question & Motivation

The research question that we want to investigate is the relationship between student caffeine consumption (intensive and extensive consumption) and their educational success, measured by grade point average.

This question is interesting for us to investigate because we know that a large majority of students consume caffeine on a daily basis, which is why we want to understand the relationship that may exist (if any) between caffeine and educational success.

III. Regression Model 1

Please see the end of data analysis for the appendix with models.

$$GPA = \beta_0 + \beta_1 \times D^{caffeine\ consumption} + \beta_2 \times i.income + \beta_3 \times D^{caffeine\ consumption} \times i.income$$
$$+ \beta_4 \times i.videogamehours + \beta_5 \times D^{caffeine\ consumption} \times i.videogamehours + \beta_6 \times BMI$$
$$+ \beta_7 \times D^{gender} + \beta_8 \times i.marijuanafrequency + \beta_9 * i.alcoholfrequency$$
$$+ \beta_{10} \times i.otherdrugfrequency + \beta_{11} \times i.growth + \varepsilon$$

The dependent variable in the regression is *GPA* and the main explanatory variable is a dummy variable for caffeine consumption (extensive measure) with 1 = at least one cup of caffeine consumption, and 0 = no caffeine. The other explanatory variables were included to try to account for confounding variables because the other explanatories could be correlated with the dependent variable and the main explanatory variable of caffeine consumption. For example, income may affect caffeine consumption and also affect GPA. We did not identify any intervening variables that were included because there are no variables that affect GPA and are affected by caffeine consumption for the sample of high school students and the level of caffeine consumption. The model was made assuming that the explanatory variables would affect caffeine consumption, rather than the other way around because the sample was done only over a short period of time before really formative behaviors could occur as a result of caffeine consumption.

The model also included interaction variables between caffeine dummy variable and income level and video game hours played, respectively. The positive sign and magnitude on caf_exten show that the effect of consuming at least one cup of caffeine per day tends to increase GPA by .104 points. The coefficients on the interaction between *caf_exten* and income show that for caffeine consumption and increasing levels of income the effect on GPA tends to decrease until the magnitudes become negative. The coefficients on the interaction between *caf_exten* and *vidgam* show that for caffeine consumption paired with increasing levels of video game hours the magnitudes are large, positive, and range from .45 to 1.3 point effect increase on GPA. Finally, some notable coefficient trends for the other variables show approximately increasing negative effects on GPA for increasing marijuana, alcohol, and other drug use.

The results have varying statistical significance, with the majority of the coefficients not being less than .05. Therefore, the results overall are not comprehensively statistically significant.

IV. Regression Model 2

Please see the end of data analysis for the appendix with models

$$GPA = \beta_0 + \beta_1 \times i.caffeineconsumption + \beta_2 \times i.income + \beta_3 \times i.caffeineconsumption \times i.income + \beta_4 \times i.videogamehours + \beta_5 \times i.caffeineconsumption \times i.videogamehours + \beta_6 \times BMI + \beta_7 \times D^{gender} + \beta_8 \times i.marijuanafrequency + \beta_9 * i.alcoholfrequency + \beta_{10} \times i.otherdrugfrequency + \beta_{11} \times i.growth + \varepsilon$$

The dependent variable in Model 2 is *GPA* and the main explanatory variable is a categorical variable for caffeine consumption (intensive measure) with No drinks = 1, One to two caffeinated drinks per day = 2, three to four =3, five to six = 4, and more than six drinks per day is represented by 5. Several other variables that we thought could be confounding variables we also included as explanatory variables. We do not think that there are any intervening variables included.

Similar to Model 1, Model 2 also includes 2 interaction variables between caffeine consumption and video games & income. We can see, due to the significance and magnitude that drinking at least 5-6 cups of caffeine a day results in a 2.029 increase in GPA. Based on the results of the interaction variable, it seems like the effect of the interaction variable comes mostly from the income part. The higher income is, the higher GPA will be. The coefficients of the caffeine x video game variable seem to vary. And just like Model 1, it seems like higher substances (alcohol, other drugs, marijuana) use results in lower GPAs.

Overall, the coefficients are not very statistically significant.

V. Regression Model 3

Please see the end of data analysis for the appendix with models.

The dependent variable in the regression is *pass*, with 1=(GPA>2), and 0=(GPA<2), measuring whether the student obtains passing grades, which we defined as above a C. The main explanatory variable was caffeine consumption (extensive measure). For caffeine consumption, we used a dummy variable with 1 defined as consuming at least one serving of caffeine drinks, and 0 defined as consuming no servings of caffeinated drinks. We included other explanatory variables to try to account for confounding variables because the other explanatory variables could be correlated with the dependent variable and the main explanatory variable of caffeine consumption, similar to model 1. We did not identify any intervening variables that were included. The model, just like model 1, also included interaction variables between the caffeine dummy variable, income level, and video game hours played, respectively.

The positive sign and magnitude on *caf_exten* show that the marginal effect of consuming at least one cup of caffeine per day (as opposed to no cups of caffeine) tends to increase Z value by 1.56, and the probability of having passing grades by about 50%, and it differs with the number of Z (holding other variables constant). (e.g. $\Phi(1.56)-\Phi(0)=0.44$, $\Phi(0.78)-\Phi(-0.78)=0.56$).

The coefficients on the interactions, between caf_exten and income and between *caf_exten* and *vidgam* do not show a significant effect. Most of the coefficients are negative or omitted.

As most of the other coefficients are negative, they were not very statistically significant, but *caf_exten* seems to be a significant coefficient that has a notable impact on *pass*.

VI. Regression Model 4

Please see the end of data analysis for the appendix with models.

1. Reason for choosing IV

As discovered in basic models 1 and 2, the frequency of alcohol consumption at high levels has a statistically significant negative effect on GPA. Given this result, we decide to investigate the key factors leading to such an impact. One hypothesis that we make is that students consuming alcohol at a high frequency are less risk-averse than others, and their risk-loving characteristics may lead to lower grades at school. Without taking into account this risk factor, our model may suffer from omitted variable bias. Hence, for the purpose of this hypothesis, we plan to use another variable in the dataset, *helmetfreq*, to account for the frequency of alcohol consumption and construct a new model with an instrumental variable. The *helmetfreq* variable records how often the sample wears a helmet when riding a bicycle, scooter, or other motorized bikes in the past year, on an always, nearly always, sometimes, seldom, and never scale. We believe that the answers to this part of the questionnaire accurately capture how students respond to risks because risk-loving individuals tend to ignore the helmet requirements and expose themselves to biking hazards.

To set up this model, we first generate a new variable *alc,* which is an indicator variable that equals 1 if the student has at least one whole drink of alcohol for more than 10 days in the past 30 days, and equals 0 if otherwise. This threshold of 10 days is based on the fact that the GPAs of students drinking alcohol for more than 10 days in the past month are shown to be significantly affected by the frequency of alcohol consumption in model 1. This indicator variable is set up because the original alcohol consumption frequency variable (*alcfreq*) is a categorical variable that cannot be used in the *ivregress* command in Stata for exogeneity reasons.

2. Good & Strong instrument check

To ensure that *helmetfreq* is a good instrument of *alc*, we first examine if *helmetfreq* is correlated with this endogenous variable. As shown in the output below, these two variables are positively correlated with a correlation coefficient of 0.25. Secondly, we investigate whether *helmetfreq*, or the risk factor, only affect GPA through its effect on *alc* and other independent variables. In our basic model, we included several risk-relevant variables including *marjfreq* and *othdrug*. We believe that students' risk-loving characteristics are most directly represented by their participation in risky activities such as drug and alcohol consumption, and these risky activities cause their grades to decrease by distracting them from focusing on school works. Other than these explanatory variables, we have not identified any other way risk-loving students can have substantially lower GPAs.

corr alc helmetfreq		
(obs=2,157)		
	alc	helmet~q
alc	1	
helmetfreq	0.2501	1

To check if *helmetfreq* is a strong instrument, we perform an F-test on the null hypothesis that the coefficient on the instruments in the first stage equals 0. The results are shown in the following table. This output yields an F-statistic of 1.87, which is significantly lower than 10, the standard for a strong instrument. This outcome implies that, even though the instrument and the endogenous variable are correlated, they are not correlated strong enough for it to be a valid instrument for the model. In addition, even with the instrumental variable, the coefficient on *caf_inten*, our main independent variable, is not statistically significant. Therefore, we conclude that, even though *helmetfreq* is a good instrument, it is a weak instrument and this instrumental variable model is not our preferred model.

test helmetfreq=0

helmetfreq = 0	
F(1, 152) =	1.87
Prob > F =	0.1738

VII. Discussion/ Conclusion

A. Main Takeaway Section

According to our preferred model of regression, Model 3, caffeine has a fairly (almost at 90%) significant coefficient. Thus while we cannot be absolutely certain that caffeine improves the GPA/academic performance of students, our results would lead us to believe that it is the case. On top of that, looking at Model 3 and our other models, income, and frequency of alcohol consumption does seem to contribute to a significant change in the passing rate of students. In most of our models, alcfreq=5 or above (consuming alcohol 10-19 days a month) results in a considerably large GPA decrease, or in model 3, a -0.2 probability of the student passing, both being significant. And at the higher levels of income, the student has a greatly higher chance of passing/getting a higher GPA.

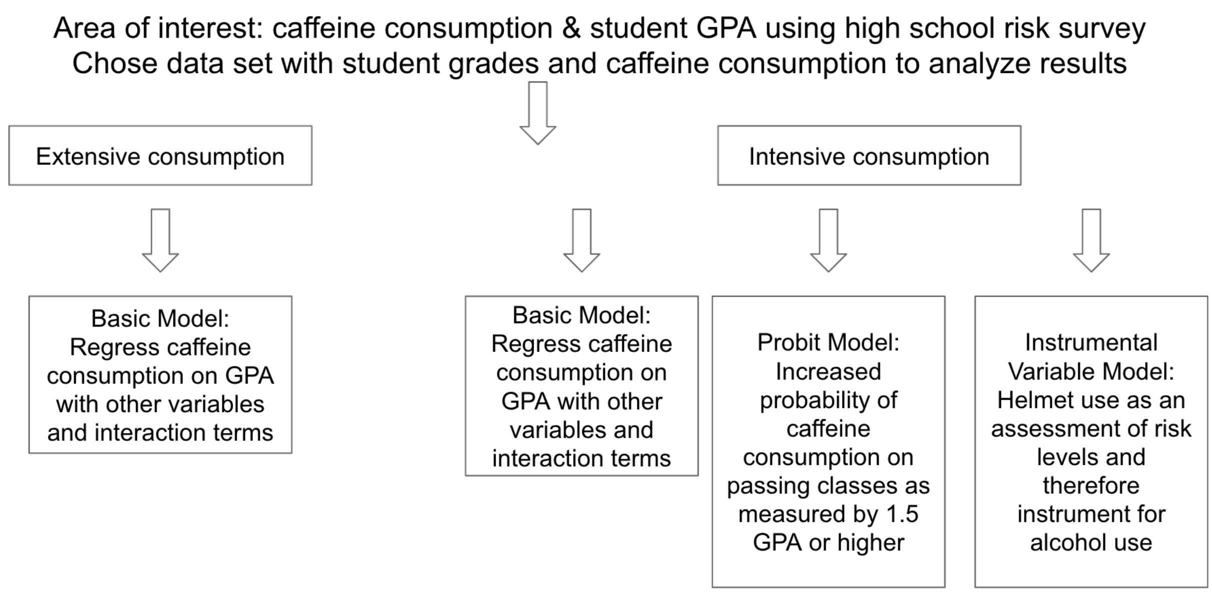
B. Limitations of Approach

The models have overall low statistical significance, which could be due to the low number of observations. Also, the models rest on the assumption that the sample was collected over a short time period and therefore there are no intervening effects of caffeine consumption on any of the other variables. This assumption holds true due to the sample age being high school students where we assume that their caffeine consumption would not have significant behavioral impacts on the other variables, because the time period was not long enough for the student to develop an extreme caffeine addiction, especially for the high school age. If the study were looking at adults of all ages over a longer time period, however, this assumption may not hold true.

C. Policy Relevance

Our study into the possible effects of caffeine consumption on student GPA is quite important, as the results could advise on future changes in school policy or other regulations. Should caffeine consumption have a positive effect on student GPA/passing rate, schools should perhaps look into offering free coffee to students in the morning, reducing the effect on GPA that disparity in income has; furthermore, in this situation, it would also become important to conduct additional research on the impact that increased caffeine consumption might have on students' sleep and overall health. However, should caffeine consumption have a negative effect on student performance, changes in school start times would need to be made, allowing students to get more sleep every night and reduce the amount of caffeine they consume. A workload reduction could serve a similar purpose. While our results are not significant and we were unable to conclude anything about the effects of caffeine consumption on GPA/passing rate, we do think that a big part of this can be contributed to the fact that the number observation was too low, and ideally, we would have access to more specific information from a larger demographic. However, through our other models, we discovered that frequent alcohol consumption has a massive negative effect on students' GPA and passing rate. Perhaps schools should implement more restrictions on alcohol or implement better alcohol education to reduce dangerous alcohol consumption and thus increase academic success.

VIII. FlowChart for Decision Process



Appendix: Regression Models & do.file

Model 1:

reg gpa caf_exten i.income caf_exten#i.income i.vidgam caf_exten#i.vidgam bmi genderS i.marjfreq i.alcfreq i.othdrug i.growth

Source	SS	df	MS			
				Number of obs	=	211
				F(36, 174)	=	3.10
Model	40.1920325	36	1.11644535	Prob > F	=	0.0000
Residual	62.7463562	174	0.360611242	R-squared	=	0.3904
Total	102.938389	210	0.490182803	Adj R-squared	=	0.2643
				Root MSE	=	0.60051

gpa	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
caf_exten	.1036037	.5134212	0.20	0.840	-.9097312	1.116939
income						
2	.5307415	.7622789	0.70	0.487	-.9737619	2.035245
3	1.5941390	.6716471	2.37	0.019	.2685146	2.919763
4	1.6252530	.6923203	2.35	0.020	.2588263	2.991679
5	.8143164	.6995120	1.16	0.246	-.5663046	2.194937
caf_exten#income						
1 1	.3182075	.7415096	0.43	0.668	-1.145304	1.781719
1 2	.1740642	.4919482	0.35	0.724	-.796890	1.145018
1 3	-.6877297	.3134272	-2.19	0.030	-1.306338	-.069121
1 4	-.6623143	.3915318	-1.69	0.093	-1.435077	.110449
1 5	0	(omitted)				
vidgam						
2	.5478035	.2463946	2.22	0.027	.0614966	1.03411
3	-.5595524	.656127	-0.85	0.395	-1.854545	.7354398
4	.3101481	.468244	0.66	0.509	-.6140211	1.234317
5	.5081727	.4836759	1.05	0.295	-.4464544	1.462800
caf_exten#vidgam						
1 1	.6822013	.5154609	1.32	0.187	-.3351595	1.699562
1 2	.2307498	.5198553	0.44	0.658	-.7952841	1.256784
1 3	1.257057	.8409014	1.49	0.137	-.4026231	2.916737
1 4	.4467698	.6577434	0.68	0.498	-.8514127	1.744952
1 5	0	(omitted)				
bmi	-.0139075	.0101583	-1.37	0.173	-.0339568	.0061419
genderS	.0285279	.1065838	0.27	0.789	-.1818356	.2388914
marjfreq						
2	.0053334	.1816639	0.03	0.977	-.353151	.3638819
3	-.0766267	.1724210	-0.44	0.657	-.416933	.2636792
4	-.0332366	.1536161	-0.22	0.829	-.336427	.2699543
5	-.0480504	.1470131	-0.33	0.744	-.338209	.2421080
alcfreq						
2	-.2005081	.1558627	-1.29	0.200	-.5081330	.1071167
3	-.2697088	.1775080	-1.52	0.130	-.6200548	.0806373
4	-.1545792	.1956428	-0.79	0.431	-.5407178	.2315594
5	-.3619347	.1704620	-2.12	0.035	-.6983741	-.0254954
6	-.8215891	.2206627	-3.72	0.000	-1.2571090	-.3860689
7	-.6720597	.2509251	-2.68	0.008	-1.1673080	-.1768111
othdrug						
2	-.1138811	.1986946	-0.57	0.567	-.5060430	.2782807
3	-.0241101	.1523615	-0.16	0.874	-.3248246	.2766044
4	-.5796301	.1825768	-3.17	0.002	-.9399803	-.2192798
5	-.1768475	.1611040	-1.10	0.274	-.4948170	.1411220
growth						
2	-.0961088	.2257486	-0.43	0.671	-.5416669	.3494492
3	-.1692570	.1968751	-0.86	0.391	-.5578277	.2193137
4	.0935274	.1790966	0.52	0.602	-.2599539	.4470088
_cons	1.341866	.7434071	1.81	0.073	-.1253899	2.809123

Model 2:

reg gpa i.caf_inten i.income i.caf_inten#income i.vidgam i.caf_inten#vidgam bmi genderS i.marjfreq i.alcfreq i.othdrug i.growth

Source	SS	df	MS			
				Number of obs	=	209
				F(61, 147)	=	3.95
Model	63.2395808	61	1.03671444	Prob > F	=	0.0000
Residual	38.5714239	147	0.262390639	R-squared	=	0.6211
				Adj R-squared	=	0.4639
Total	101.811005	208	0.489475985	Root MSE	=	0.51224

gpa	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
cafe_inten						
2	1.162035	0.6276841	1.85	0.066	-0.0784152	2.402485
3	1.240074	0.8419955	1.47	0.143	-0.4239053	2.904054
4	2.028895	0.70222	2.89	0.004	0.6411442	3.416645
5	1.052426	0.9230543	1.14	0.256	-0.7717442	2.876597
income						
2	0.6018957	0.6570959	0.92	0.361	-0.696679	1.90047
3	1.636876	0.5792067	2.83	0.005	0.4922282	2.781523
4	1.625626	0.5975821	2.72	0.007	0.4446644	2.806588
5	0.8863773	0.6030051	1.47	0.144	-0.3053014	2.078056
caf_inten#income						
2 2	-.4043321	0.7501203	-0.54	0.591	-1.8867450	1.0780810
2 3	-1.0349210	.6479299	-1.60	0.112	-2.3153810	.2455402
2 4	-1.0898740	.6799681	-1.60	0.111	-2.4336500	.2539016
2 5	-.4693723	.6740125	-0.70	0.487	-1.8013780	.8626337
3 2	.5429028	.9433085	0.58	0.566	-1.3212950	2.4071000
3 3	-1.0283240	.8423783	-1.22	0.224	-2.6930600	.6364116
3 4	-1.2943050	.8905494	-1.45	0.148	-3.0542380	.4656287
3 5	-.4596040	.8582582	-0.54	0.593	-2.1557230	1.2365140
4 2	-.7697909	.8590514	-0.90	0.372	-2.4674770	.9278951
4 3	-2.1876060	.7348486	-2.98	0.003	-3.6398390	-.7353740
4 4	-1.6266140	.8163394	-1.99	0.048	-3.2398920	-.0133373
4 5	-2.7126480	.8216659	-3.30	0.001	-4.3364520	-1.0888450
5 2	-.9553861	1.0172050	-0.94	0.349	-2.9656200	1.0548480
5 3	-1.2095660	.9410116	-1.29	0.201	-3.0692240	.6500926
5 4	-.0803531	.9322278	-0.09	0.931	-1.9226530	1.7619470
5 5	.2351641	.9333929	0.25	0.801	-1.6094380	2.0797660
vidgam						
2	.5001201	.2123420	2.36	0.020	.0804827	.9197574
3	-.4173534	.5652992	-0.74	0.462	-1.5345170	.6998098
4	.4238023	.4033237	1.05	0.295	-.3732595	1.2208640
5	.4105372	.4146536	0.99	0.324	-.4089151	1.2299890
caf_inten#vidgam						
2 2	-.2640624	.2428763	-1.09	0.279	-.7440427	.2159179
2 3	.5843115	.5950893	0.98	0.328	-.5917237	1.7603470
2 4	.2740782	.7162182	0.38	0.703	-1.1413360	1.6894920
2 5	-1.4167220	.5279725	-2.68	0.008	-2.4601190	-.3733253
3 2	-.6426147	.3052522	-2.11	0.037	-1.2458640	-.0393652
3 3	-.0152627	.6394844	-0.02	0.981	-1.2790330	1.2485080
3 4	-.3207260	.5063059	-0.63	0.527	-1.3213050	.6798526
3 5	-.1174690	.5482871	-0.21	0.831	-1.2010120	.9660742
4 2	-.3078083	.3790586	-0.81	0.418	-1.0569170	.4412999
4 3	2.8599090	.8115581	3.52	0.001	1.2560810	4.4637380
4 4	-1.4058780	.8074037	-1.74	0.084	-3.0014960	.1897400
4 5		(empty)				
5 2	-1.7118520	.4912810	-3.48	0.001	-2.6827380	-.7409658
5 3	-.8732562	.7313549	-1.19	0.234	-2.3185840	.5720718
5 4		(empty)				
5 5	-.5720010	.4873003	-1.17	0.242	-1.5350200	.3910181
bmi	-.0116959	.0092813	-1.26	0.210	-.0300379	.0066461
genderS	.0461531	.0983285	0.47	0.639	-.1481670	.2404731
marjfreq						
2	-0.049787	0.1658327	-0.3	0.764	-.377511	.2779372
3	-0.216217	0.1585641	-1.36	0.175	-.529577	.0971426
4	-0.1328299	0.1411354	-0.94	0.348	-.411746	.1460865
5	-0.2678851	0.1368262	-1.96	0.052	-.538286	.0025154
alcfreq						
2	-.1199575	.1429191	-0.84	0.403	-.4023989	.1624839
3	-.1626925	.1594333	-1.02	0.309	-.4777698	.1523849
4	-.0327523	.1773177	-0.18	0.854	-.3831734	.3176687
5	-.2750541	.1566863	-1.76	0.081	-.5847029	.0345947
6	-.8112649	.2066409	-3.93	0.000	-1.2196360	-.4028942
7	-.3689018	.2728366	-1.35	0.178	-.9080905	.1702868
othdrug						
2	-.1406205	.1915818	-0.73	0.464	-.5192309	.2379898
3	-.1730623	.1421561	-1.22	0.225	-.4539959	.1078713
4	-.6633659	.1674541	-3.96	0.000	-.9942942	-.3324375
5	-.2468043	.1552696	-1.59	0.114	-.5536532	.0600447
growth						
2	-.0809969	.2066937	-0.39	0.696	-.4894718	.3274780
3	-.0968113	.1820047	-0.53	0.596	-.4564951	.2628724
4	.2105931	.1673274	1.26	0.210	-.1200849	.5412711
_cons	1.2267530	.6587157	1.86	0.065	-.0750225	2.5285290

Model 3:

probit pass caf_exten caf_exten#income caf_exten#vidgam bmi genderS i.marjfreq i.alcfreq i.othdrug i.growth

Probit regression	Number of obs = 257
	LR chi2(34) = 68.32
	Prob > chi2 = 0.0004
Log likelihood = -104.30028	Pseudo R2 = 0.2467

pass	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
caf_exten	1.5635	1.123848	1.39	0.164	-0.6392024	3.766202
income						
2	-0.7457273	1.268908	-0.59	0.557	-3.232742	1.741287
3	1.103725	0.958906	1.15	0.25	-0.7756959	2.983146
4	1.175726	1.007138	1.17	0.243	-0.7982285	3.14968
5	0.3467682	0.5401782	0.64	0.521	-0.7119616	1.405498
caf_exten#income						
1 1	.0000000	(omitted)				
1 2	.7662339	1.2448300	0.62	0.538	-1.674	3.2060560
1 3	-.6103000	.8461579	-0.72	0.471	-2.269	1.0481390
1 4	-.3007085	.9481813	-0.32	0.751	-2.159	1.5576930
1 5	.0000000	(omitted)				
vidgam						
2	1.6262860	.8994087	1.81	0.071	-0.137	3.3890950
3	.2580207	1.1371810	0.23	0.821	-1.971	2.4868550
4	-.6814820	1.1378210	-0.60	0.549	-2.912	1.5486060
5	-.8062673	1.2946900	-0.62	0.533	-3.344	1.7312780
caf_exten#vidgam						
1 1	-.5120029	1.2546690	-0.41	0.683	-2.971	1.9471030
1 2	-1.6958870	1.4174370	-1.2	0.232	-4.474	1.0822380
1 3	-.9080438	1.6423720	-0.55	0.58	-4.127	2.3109470
1 4	.0000000	(omitted)				
1 5	.0000000	(omitted)				
bmi	-.0122786	.0258572	-0.47	0.635	-0.063	.0384005
genderS	-.0571071	.2528939	-0.23	0.821	-0.553	.4385558
marjfreq						
2	-0.1350628	0.4752535	-0.28	0.776	-1.066543	.796417
3	-0.127689	0.4254685	-0.3	0.764	-0.9615919	.706214
4	-0.2345446	0.3699303	-0.63	0.526	-0.9595947	.490506
5	-0.105165	0.3505427	-0.3	0.764	-0.7922161	.581886
alcfreq						
2	-.7963775	.4388546	-1.81	0.07	-1.657	.0637617
3	-.7006405	.4772932	-1.47	0.142	-1.636	.2348370
4	-.1855674	.5966593	-0.31	0.756	-1.355	.9838633
5	-1.1360180	.4658316	-2.44	0.015	-2.049	-.2230048
6	-1.8177450	.5424488	-3.35	0.001	-2.881	-.7545644
7	-1.0910780	.5851497	-1.86	0.062	-2.238	.0557948
othdrug						
2	-.0793407	.4451432	-0.18	0.859	-0.952	.7931239
3	-.1375802	.3492443	-0.39	0.694	-0.822	.5469260
4	-1.0958440	.3761772	-2.91	0.004	-1.833	-.3585499
5	-.3164407	.3503976	-0.90	0.366	-1.003	.3703259
growth						
2	-.4696205	.6025607	-0.78	0.436	-1.651	.7113769
3	-1.2498280	.4905763	-2.55	0.011	-2.211	-.2883160
4	-.6632905	.4761211	-1.39	0.164	-1.596	.2698897
_cons	1.4490710	1.1535730	1.26	0.209	-0.812	3.7100320

Model 4:

1. First-stage regressions

ivregress 2sls gpa i.caf_inten i.income i.caf_inten#income i.vidgam i.caf_inten#vidgam bmi gender5 i.marjfreq i.othdrug i.growth (alc = helmetfreq), first

First-stage regressions

Number of obs	=	144
F(36, 174)	=	2.06
Prob > F	=	0.0012
R-squared	=	0.5481
Adj R-squared	=	0.282
Root MSE	=	0.4029

alc	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
caf_inten						
2	0.3069108	0.6909727	0.44	0.658	-1.065827	1.679649
3	0.3533963	0.3435082	1.03	0.306	-0.3290428	1.035835
4	0.2683576	0.7035992	0.38	0.704	-1.129465	1.66618
5	0.2228842	0.4488116	0.5	0.621	-0.6687584	1.114527
income						
2	0.3631485	0.6758728	0.54	0.592	-0.9795908	1.705888
3	0.4055923	0.6301318	0.64	0.521	-0.8462745	1.657459
4	0.7110018	0.6231143	1.14	0.257	-0.5269236	1.948927
5	0.5769407	0.5787198	1	0.321	-0.5727872	1.726669
caf_inten#income						
1 1	0	(empty)				
2 2	-0.3960876	0.7753237	-0.51	0.611	-1.936403	1.144228
2 3	-0.4117287	0.714084	-0.58	0.566	-1.830381	1.006924
2 4	-0.6238817	0.6987473	-0.89	0.374	-2.012065	0.7643017
2 5	-0.6311318	0.6795468	-0.93	0.356	-1.98117	0.7189065
3 1	0	(empty)				
3 2	-0.1959074	0.4605598	-0.43	0.672	-1.11089	0.719075
3 3	-0.2379722	0.3438083	-0.69	0.491	-0.9210075	0.445063
3 4	-0.2781545	0.4196104	-0.66	0.509	-1.111784	0.5554747
3 5	0	(omitted)				
4 2	-1.027456	0.8160195	-1.26	0.211	-2.648621	0.5937095
4 3	-0.2979542	0.7316417	-0.41	0.685	-1.751488	1.15558
4 4	0.7692967	0.9441038	0.81	0.417	-1.10633	2.644924
4 5	0.2641845	0.7735555	0.34	0.734	-1.272618	1.800987
5 2	-0.1954295	0.5527753	-0.35	0.725	-1.293614	0.9027552
5 3	0.0968227	0.3998413	0.24	0.809	-0.6975318	0.8911772
5 4	-0.2465349	0.4345331	-0.57	0.572	-1.109811	0.6167409
5 5	0	(omitted)				
vidgam						
2	-0.2459578	0.1974213	-1.25	0.216	-0.6381698	0.1462541
3	-0.4996138	0.5223534	-0.96	0.341	-1.53736	0.5381325
4	-0.3764175	0.3601604	-1.05	0.299	-1.091939	0.3391041
5	-0.2314916	0.3378178	-0.69	0.495	-0.9026256	0.4396425
caf_inten#vidgam						
2 2	0.4052383	0.2382402	1.7	0.092	-0.0680674	0.878544
2 3	0.8232491	0.5890299	1.4	0.166	-0.3469615	1.99346
2 4	1.020252	0.5773184	1.77	0.081	-0.1266921	2.167195
2 5	0.2483551	0.4438508	0.56	0.577	-0.6334319	1.130142
3 2	0.2872015	0.2932721	0.98	0.33	-0.2954346	0.8698376
3 3	-0.0656161	0.6085834	-0.11	0.914	-1.274673	1.143441
3 4	0.2319866	0.490972	0.47	0.638	-0.743415	1.207388
3 5	-0.1130347	0.4625934	-0.24	0.808	-1.032057	0.8059878
4 2	0.748447	0.3556474	2.1	0.038	0.0418914	1.455003
4 3	0.5360323	0.7545344	0.71	0.479	-0.9629821	2.035047
4 4	0	(omitted)				
4 5	0	(empty)				
5 2	0.2773123	0.4626008	0.6	0.55	-0.6417249	1.196349
5 3	-0.223508	0.7482493	-0.3	0.766	-1.710036	1.26302
5 4	0	(empty)				
5 5	0.2358898	0.4315537	0.55	0.586	-0.621467	1.093247
bmi	0.0128543	0.0094228	1.36	0.176	-0.0058657	0.0315744
gender5	0.2376275	0.1100403	-2.16	0.033	0.4562418	0.0190132
marjfreq						
2	-0.0133318	0.1554356	-0.09	0.932	-0.3221319	0.2954682
3	0.2854989	0.1664021	1.72	0.09	-0.045088	0.6160858
4	0.2630372	0.1336101	1.97	0.052	-0.0024027	0.528477
5	0.3197323	0.1147263	2.79	0.006	0.0918085	0.5476562
othdrug						
2	0.0060487	0.195232	0.03	0.975	-0.3818138	0.3939111
3	-0.1051447	0.1385014	-0.76	0.45	-0.3803018	0.1700124
4	0.237294	0.2567284	0.92	0.358	-0.2727418	0.7473297
5	0.2169569	0.1401842	1.55	0.125	-0.0615435	0.4954572
growth						
2	0.1525795	0.183048	0.83	0.407	-0.2110774	0.5162363
3	0.0129489	0.1542168	0.08	0.933	-0.2934296	0.3193275
4	0.1567197	0.1404861	1.12	0.268	-0.1223804	0.4358198
helmetfreq	0.1300199	0.0607964	2.14	0.035	0.0092371	0.2508026
_cons	-0.7826208	0.6924838	-1.13	0.261	-2.158361	0.5931191

2. Instrumental variables 2SLS Regressions

Instrumental varaiables 2SLS Regression

Number of obs	=	144
F(36, 174)	=	182.54
Prob > F	=	0.0000
R-squared	=	0.4938
Root MSE	=	0.51496

gpa	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
alc	-0.9517894	0.5976945	-1.59	0.111	-2.123249	0.2196702
caf_inten						
2	2.083611	0.920186	2.26	0.024	0.2800799	3.887143
3	0.6535718	0.4798227	1.36	0.173	-0.2868634	1.594007
4	2.869104	0.918529	3.12	0.002	1.06882	4.669387
5	1.472305	0.583649	2.52	0.012	0.3283737	2.616236
income						
2	1.363034	0.904096	1.51	0.132	-0.4089617	3.135029
3	2.537515	0.8444731	3	0.003	0.882378	4.192652
4	2.365728	0.8946758	2.64	0.008	0.6121955	4.11926
5	1.770156	0.8093578	2.19	0.029	0.1838434	3.356468
caf_inten#income						
1 1	0	(empty)				
2 2	-1.526091	1.058806	-1.44	0.149	-3.601312	0.5491305
2 3	-2.285002	0.9806218	-2.33	0.02	-4.206986	-0.3630188
2 4	-2.194882	1.010662	-2.17	0.03	-4.175743	-0.2140215
2 5	-1.621406	0.9899228	-1.64	0.101	-3.561619	0.3188072
3 1	0	(empty)				
3 2	1.5286	0.6081975	2.51	0.012	0.3365545	2.720645
3 3	-0.1021672	0.4699636	-0.22	0.828	-1.023279	0.8189446
3 4	0.1425417	0.5825547	0.24	0.807	-0.9992446	1.284328
3 5	0	(omitted)				
4 2	-2.229654	1.288613	-1.73	0.084	-4.755288	0.2959808
4 3	-3.306231	0.9705617	-3.41	0.001	-5.208497	-1.403965
4 4	-3.169013	1.28785	-2.46	0.014	-5.693153	-0.6448736
4 5	-3.639991	0.9911948	-3.67	0	-5.582697	-1.697285
5 2	-1.305161	0.7128702	-1.83	0.067	-2.702361	0.0920386
5 3	-1.563773	0.5157264	-3.03	0.002	-2.574578	-0.552968
5 4	-0.4017419	0.5733456	-0.7	0.483	-1.525479	0.7219948
5 5	0	(omitted)				
vidgam						
2	0.4422744	0.3047529	1.45	0.147	-0.1550302	1.039579
3	1.100293	0.735282	1.5	0.135	-0.3408332	2.541419
4	0.370066	0.5315177	0.7	0.486	-0.6716895	1.411822
5	0.3356157	0.4610176	0.73	0.467	-0.5679621	1.239194
caf_inten#vidgam						
2 2	-0.0805616	0.4061149	-0.2	0.843	-0.8765323	0.715409
2 3	-0.6865042	0.9091486	-0.76	0.45	-2.468403	1.095394
2 4	0.8028836	1.026293	0.78	0.434	-1.208613	2.81438
2 5	-1.307235	0.5937122	-2.2	0.028	-2.47089	-0.1435808
3 2	-0.925033	0.4310706	-2.15	0.032	-1.769916	-0.0801502
3 3	-2.274872	0.7727348	-2.94	0.003	-3.789404	-0.7603395
3 4	-0.8339981	0.6695395	-1.25	0.213	-2.146271	0.4782752
3 5	-0.1324655	0.5884345	-0.23	0.822	-1.285776	1.020845
4 2	0.1421604	0.6791539	0.21	0.834	-1.188957	1.473278
4 3	1.664515	1.024098	1.63	0.104	-0.3426795	3.671709
4 4	0	(omitted)				
4 5	0	(empty)				
5 2	-1.483275	0.6270881	-2.37	0.018	-2.712345	-0.2542046
5 3	-3.315622	0.9584779	-3.46	0.001	-5.194204	-1.43704
5 4	0	(empty)				
5 5	-0.577393	0.5809757	-0.99	0.32	-1.716084	0.5612985
bmi	-0.0036163	0.013218	-0.27	0.784	-0.0295232	0.0222905
genderS	-0.0040332	0.1889918	-0.02	0.983	-0.3744504	0.366384
marjfreq						
2	-0.0381432	0.198597	-0.19	0.848	-0.4273862	0.3510999
3	-0.3810837	0.2728631	-1.4	0.163	-0.9158857	0.1537182
4	-0.0891031	0.2363522	-0.38	0.706	-0.5523449	0.3741387
5	-0.0486347	0.2583478	-0.19	0.851	-0.5549871	0.4577177
othdrug						
2	-0.181321	0.2496811	-0.73	0.468	-0.6706869	0.3080449
3	-0.4280426	0.1908429	-2.24	0.025	-0.8020879	-0.0539974
4	-0.5222679	0.3559452	-1.47	0.142	-1.219908	0.1753719
5	-0.0542311	0.2301802	-0.24	0.814	-0.5053761	0.3969138
growth						
2	-0.074356	0.2472634	-0.3	0.764	-0.5589835	0.4102714
3	-0.0868663	0.1959015	-0.44	0.657	-0.4708262	0.2970936
4	0.3268058	0.1862723	1.75	0.079	-0.0382812	0.6918927
_cons	0.2639913	0.9180439	0.29	0.774	-1.535342	2.063324

Do.file

```
// categorical
// q1 - age
gen ageyrs_less14 = (q1==1)
gen ageyrs_14 = (q1==2)
gen ageyrs_15 = (q1==3)
gen ageyrs_16 = (q1==4)
gen ageyrs_17 = (q1==5)
gen ageyrs_18 = (q1==6)
gen ageyrs_more19 = (q1==7)
gen ageyrs = q1
// q91 - caffeine consumption  (intensive)
gen caf_inten = q91
// q10 - income level
gen dincome_welf = 1 if q10==1
replace dincome_welf  = 0 if q10!=1
gen dincome_low = 1 if q10==2
replace dincome_low = 0 if q10!=2
gen dincome_middle = 1 if q10==3
replace dincome_middle = 0 if  q10!=3
gen dincome_high = 1 if q10==4
replace dincome_high = 0 if q10!=4
gen incomelvl=q10
// q144 - completion of growth for height
gen dgrowth_no = 1 if q144==1
replace dgrowth_no = 0 if q144!=1
gen dgrowth_bstart = 1 if q144==2
replace dgrowth_bstart = 0 if q144!=2
gen dgrowth_udwy = 1 if q144==3
replace dgrowth_udwy = 0 if q144!=3
gen dgrowth_compl = 1 if q144==4
replace dgrowth_compl = 0 if q144!=4
gen growth=q144
// q130 - number of hours video games
gen dvidgam_no = 1 if q130==1
replace dvidgam_no = 0 if q130!=1
gen dvidgam_less7 = 1 if q130==2
replace dvidgam_less7 = 0 if q130!=2
gen dvidgam_7to14 = 1 if q130==3
```

```
replace dvidgam_7to14 = 0 if q130!=3
gen dvidgam_15to20 = 1 if q130==4
replace dvidgam_15to20 = 0 if q130!=4
gen dvidgam_more21 = 1 if q130==5
replace dvidgam_more21 = 0 if q130!=5
gen vidgam=q130
// q66 - marijuana consumption
gen dmarjfreq_no = 1 if q66==1
replace dmarjfreq_no= 0 if q66!=1
gen dmarjfreq_once = 1 if q66==2
replace dmarjfreq_once = 0 if q66!=2
gen dmarjfreq_twice = 1 if q66==3
replace dmarjfreq_twice = 0 if q66!=3
gen dmarjfreq_3to5 = 1 if q66==4
replace dmarjfreq_3to5 = 0 if q66!=4
gen dmarjfreq_evd = 1 if q66==5
replace dmarjfreq_evd = 0 if q66!=5
gen marjfreq = q66
// q74 - alcohol frequency
gen dalcfreq_0 = 1 if q74==1
replace dalcfreq_0= 0 if q74!=1
gen dalcfreq_1to2 = 1 if q74==2
replace dalcfreq_1to2 = 0 if q74!=2
gen dalcfreq_3to5 = 1 if q74==3
replace dalcfreq_3to5 = 0 if q74!=3
gen dalcfreq_6to9 = 1 if q74==4
replace dalcfreq_6to9 = 0 if q74!=4
gen dalcfreq_10to19 = 1 if q74==5
replace dalcfreq_10to19 = 0 if q74!=5
gen dalcfreq_20to29 = 1 if q74==6
replace dalcfreq_20to29 = 0 if q74!=6
gen dalcfreq_evd = 1 if q74==7
replace dalcfreq_evd = 0 if q74!=7
gen alcfreq = q74
// q75 - alcohol intensity
gen dalcinten_0 = 1 if q75==1
replace dalcinten_0= 0 if q75!=1
gen dalcinten_1to2 = 1 if q75==2
replace dalcinten_1to2 = 0 if q75!=2
gen dalcinten_3to5 = 1 if q75==3
```

```
replace dalcinten_3to5 = 0 if q75!=3
gen dalcinten_6to9 = 1 if q75==4
replace dalcinten_6to9 = 0 if q75!=4
gen dalcinten_10to19 = 1 if q75==5
replace dalcinten_10to19 = 0 if q75!=5
gen dalcinten_20to29 = 1 if q75==6
replace dalcinten_20to29 = 0 if q75!=6
gen dalcinten_evd = 1 if q75==7
replace dalcinten_evd = 0 if q75!=7
gen alcinten=q75
// q82 - other drugs
gen dothdrug_no = 1 if q82==1
replace dothdrug_no= 0 if q82!=1
gen dothdrug_once = 1 if q82==2
replace dothdrug_once = 0 if q82!=2
gen dothdrug_twice = 1 if q82==3
replace dothdrug_twice = 0 if q82!=3
gen dothdrug_3to5 = 1 if q82==4
replace dothdrug_3to5 = 0 if q82!=4
gen dothdrug_evd = 1 if q82==5
replace dothdrug_evd = 0 if q82!=5
gen othdrug=q82
// indicator
// q2 - gender
// q91 - caffeine consumption (extensive)
// q8 - part time job
gen genderS = 1 if q2==1
replace genderS = 0 if q2==2
gen caf_exten = 0 if q91==1
replace caf_exten = 1 if q91!=1
gen ptj = 1 if q8==1
replace ptj = 0 if q8==2
// continuous
// q11 - height
gen heightinch = q11a*12+q11b if q11a!=0
// q12 - weight
gen weightpds = q12
// q6 - grade average
// recode q6 (1=4) (2=3) (3=2.5) (4=1.5) (5=0), gen(gpa)
gen gpa = 4 if q6==11
```

```
replace gpa = 3 if q6==2
replace gpa = 2.5 if q6==3
replace gpa = 1.5 if q6==4
replace gpa = 0.5 if q6==5
// regression models
// model 1
reg gpa caf_exten i.income caf_exten#income i.vidgam caf_exten#vidgam bmi genderS
i.marjfreq i.alcfreq i.othdrug i.growth
// model 2
reg gpa i.caf_inten i.income i.caf_inten#income i.vidgam i.caf_inten#vidgam bmi genderS
i.marjfreq i.alcfreq i.othdrug i.growth
// model 3
gen pass=1 if gpa>2
replace pass=0 if gpa<2 & gpa>0
probit pass caf_exten i.income caf_exten#income i.vidgam caf_exten#vidgam bmi genderS
i.marjfreq i.alcfreq i.othdrug i.growth
// model 4
gen alc = 1 if alcfreq == 5|alcfreq == 6|alcfreq == 7
replace alc = 0 if alcfreq !=5 & alcfreq != 6 & alcfreq != 7
gen helmetfreq = q16
corr alc helmetfreq
ivregress 2sls gpa i.caf_inten i.income i.caf_inten#income i.vidgam i.caf_inten#vidgam bmi
genderS i.marjfreq i.othdrug i.growth (alc = helmetfreq), first
//regress i.caf_inten i.income i.caf_inten#income i.vidgam i.caf_inten#vidgam bmi genderS
i.marjfreq i.othdrug i.growth
//test helmetfreq=0
```