Econometrics

EXTRACURRICULAR ACTIVITIES: GOOD OR BAD?

Abstract
The paper attempts to investigate a causal relationship between the cognitive abilities in mathematics of students and the extra-curricular activities a student takes part in. Taking multiple factors and control variables into account a regression has been run.
This paper systematically introduces the problem and provides previous studies conducted in this field.
It concludes the discussion with encouraging students to take part in extra-curricular activities.

Group 11
Zaid Abid (18100129)
Moomal Irfan (18110021)
Bakhtawar Soomro (18020479

Table of Contents

Introduction

Extracurricular activities at school have been seen to enhance the cognitive abilities like perception, attention, and memory, visual and spatial processing of students that in turn help Students achieve better grades in subjects like Mathematics. Since the past many years parents have had problems deciding whether they should allow their children to spend time on extra-curricular activities and spend extra money on that or they should tell their children to go with the traditional ways of studying and only focus on academics. Educationists like Stephen and Schaben are also keen to know the relationship between academic achievement and participation in extracurricular activities in order to justify that extracurricular activities do have some effect on the academic performance of students.

This paper consists of four sections. The first section is literature review which highlights the previous researches that have been conducted regarding the relationship between extra-curricular activities and cognitive ability in mathematics. Moreover, the second section contains description of the data methodology. The next part of the paper presents models and attempts to explain the trends and relationships between the dependent and independent variables.

Literature Review

The results of Beckett's study along with the findings of Fejgin (1994) and Hanson (1998) support the claim that playing school sports increases academic achievement on math assessments. The study shows that not all sports contribute towards this claim.

Pascarella & Terenzini have researched on Who the students are (sex, race or socioeconomic status), where they go to school (type of institution) and what they do while enrolled at the institution (coursework or co-curricular activities) all impact how students change during college. This change shows changes in students' cognitive skills, affective characteristics, etc.

Building on the fact that extra- curricular does impact cognitive abilities that help in mathematics, another study shows quasi-experimental data that indicates that an additional hour per day of curricular time to physical activity classes does not affect the academic performance of primary school students negatively. While the opposite implication has a small effect. Cross-sectional observations have shown a positive association between academic performance and physical activity (a form of extracurricular activity).

Harris Cooper and Jaffery Valentine conducted a study. The study consisted of Four hundred twenty-four students in Grades 6- 12 and 1 parent of each completed a questionnaire regarding the participation of student participation five following activities , homework, television viewing, extracurricular activities, other types of structured after-school groups, and jobs. And then the test scores and class grades were collected. After-school activities contributed significantly to the prediction of achievement even after the student's gender, grade level, ethnicity, free-lunch eligibility and other factors were kept as control variables. The findings showed that more time in extracurricular activities and other structured groups and less time in jobs and television viewing were related to higher test scores and class grades.

On the contrary, some studies show results from multiple regression analyses that explored the relationship between co-curricular involvement and the need for cognition. They also present the results of the conditional effects, including race, sex and institutional type, and their influence on the effect of co-curricular involvement on the need for cognition. Collectively, these results indicate that co-curricular involvement does not significantly contribute to a student's need for cognition.

Description of Data

The data that we used to test the question under observation was taken from the "Early Childhood Longitudinal Survey" which is a nationally representative data of the United States. Originally, kindergarten kids were surveyed in 1998/99 and then there were follow-up surveys in different grades. The data that we are using to test our theory is from the 5th grade follow-up. The data has various types of information relating to these children; it has their academic scores in the form of reading, math and science, their family background, their home environment as well as school environment and also detailed information regarding their extracurricular activities. We only used certain relevant variables from this dataset that we thought could have an impact of the math scores of a child. Dataset has a total of 8105 observations that are spread over the students from different regions of USA (Northeast, Midwest, South and West).

The variables that we used for our analysis and their descriptive statistics are as follows. We were able to control for many factors that affect a student's maths scores and hence, this helped us obtain more accurate estimates of the effect of extracurricular activities on math scores and so, helped us reduce omitted variable bias.

We included mother's education as a control because we believe that mothers who are educated might be more encouraging to their children in terms of studies. It could also be the case that the mother is intelligent hence the child is intelligent too through inheritance. Hence, it also serves as a control for innate ability. We have also needed to include gender in order to capture any effects that being of a certain gender can have on your education. For instance, in some households boys might be more encouraged to do well in school as compared to girls. However, we created a dummy variable for male and female and included male in the regression in order to capture the effect of being a male on math scores. We included afternoon TV as a control because we wanted to control for the impact of watching TV on test scores. It would be that watching TV reduces your thinking ability. Also, we only included TV afternoon on weekdays because we assumed that this is the study time for most students and hence, those who spend this time watching TV might have lower math scores. We also included the variable which controls for working mothers. This is because we suspect that children with stay at home mothers are able to get more attention and hence do better in studies. Apart from this, the rationale for including drug problem variable is that we believe that consumption of drugs can harm your mental ability and hence reduce math scores. Access to home computer is added as a control because children can use the computer to play educational games which can sharpen their ability to think and hence, increase math scores. We also included school types in order to control for the fact that different schools may have different ways of teaching.

We have changed certain variables from their original form into new variables in order to make our interpretation simpler. For instance, race was provided to us as a categorical variable. We turned it into individual dummy variables for each race in order to capture the effects of each race on math scores. Similarly, we generated a variable called 'momworks' which is a dummy with the value unity if mom works and zero otherwise. The data provided us the work status of mom but we believe using the dummy variable will make interpretation simpler. *Wherever we used the dummy variables, we included n-1 variables in order to avoid perfect multicollinearity*. For example, out of the four races given, we included only the dummy variables of only 3. Furthermore, we generated lincome which took the log values of income. The rationale behind this was the firstly, income is positive and continuous which fulfils the conditions for using log values. Secondly, it helped us achieve a more close to normal distribution of our error variances. The reason for using log on math scores is similar.

One thing that was missing from the data was the hours spent studying. We believe this is an important variable which will differentiate hardworking students. In order to control for this, we included the library card variable. Here we assumed that children with a card study more and hence are more hardworking.

Methodology

The aim of this paper was to see whether taking part in extra-curricular activities has a positive effect of children's math scores. If we simply ran a regression with math scores as the dependent variable and the binary variables representing several extra-curricular activities as the independent, explanatory variables then the results would have been drastically misleading. The reason for this is that the results would fail to accommodate the effects of other factors such as home environment, time spent on studying and students' background. And hence, the model would give out biased and inconsistent estimators of beta values. Also the correlation between independent variables and the error term (such as negative correlation between taking part in ECAs and time spent studying) would cause the explanatory variables to become endogenous.

As an initial regression, we ran the following equation:

$$\textbf{math\_test} = \beta_0 + \beta_1\textbf{gender} + \beta_2\textbf{race} + \beta_3\textbf{mom\_educ} + \beta_4\textbf{family\_income} + \beta_5\textbf{hhsize} + \beta_6\textbf{part\_dance} + \beta_7\textbf{part\_athletics} + \beta_8\textbf{part\_club} + \beta_9\textbf{part\_music} + \beta_{10}\textbf{part\_art} + \beta_{11}\textbf{tv\_afternoon\_mf} + \beta_{12}\textbf{both\_parents} + \beta_{13}\textbf{has\_library\_card} + \beta_{14}\textbf{has\_home\_computer} + \beta_{15}\textbf{mom\_work\_status} + \beta_{16}\textbf{school\_type} + \beta_{17}\textbf{problem\_drugs}$$

The regression output is shown in *Exhibit 2a*. We found several problems with the variables. These problems are discussed in the previous section of Data Description. Eventually we resolved these problems and came up with the following final model:

$$\textbf{lmath} = \beta_0 + \beta_1\textbf{male} + \beta_2\textbf{race1} + \beta_3\textbf{race2} + \beta_4\textbf{race3} + \beta_5\textbf{mom\_educ} + \beta_6\textbf{lincome} + \beta_7\textbf{hhsize} + \beta_8\textbf{part\_dance} + \beta_9\textbf{part\_athletics} + \beta_{10}\textbf{part\_club} + \beta_{11}\textbf{part\_music} + \beta_{12}\textbf{part\_art} + \beta_{13}\textbf{tv\_afternoon\_mf} + \beta_{14}\textbf{both\_parents} + \beta_{15}\textbf{has\_library\_card} + \beta_{16}\textbf{has\_home\_computer} + \beta_{17}\textbf{momworks} + \beta_{18}\textbf{schooltype1} + \beta_{19}\textbf{schooltype2} + \beta_{20}\textbf{problem\_drugs}$$

After running the above model in STATA we got the output shown in *Exhibit 2b*. The R-Squared value came out to be 0.2554 which means that the independent explanatory variables describe almost 25.54% variation in math scores. This value makes sense since an important factor in math scores is the amount of time spent on studying for the test which is not present in the data set. Also a student's past performance is a big factor in current test scores but since the data did not have time-series variables, this goes unaccounted for as well. The F-Stat came out to be 138.63 and Prob > F was 0 which indicated that the coefficients are jointly significant. A big surprise was the result of the Ramsey's RESET Test shown in *Exhibit 2c*. The null hypothesis that "model has no omitted variables" was not rejected at even 15% significance level as the p-value was 0.1563. This was quite surprising initially. However, after looking at the RESET test properties, this can be explained by the fact that the functional form of the model is correct. After checking for pair-wise correlation (shown in *Exhibit 2d*) between the variables, it was concluded that the problem of perfect multicollinearity does not exist between the variables. This can also be corroborated using a VIF test (also shown in *Exhibit 2d*) in which it is generally stated that if any variable's VIF is greater than equal to 4 then multicollinearity is high and troublesome. But the test shows that all of the VIFs are well below 4 reiterating the fact that there is no problem of multicollinearity. After checking the plot of fitted values against residuals, Breusch-Pagan Test and White Test (shown in *Exhibit 2e*), we came to the conclusion that the model is heteroskedastic. This problem is taken care of by using robust standard errors and all of the analysis done in this paper is done using the robust standard errors. The regression run in STATA with robust standard errors is shown in *Exhibit 2f*.

To check for the Multiple Linear Regression Assumptions, residuals from the regression were summarized and it clearly showed (*Exhibit 2g*) that the mean of residuals was 0. Similarly, after running regression of residuals on all of the independent variables one by one using a "for loop" (see do file) concluded that there was no correlation between the variables and residuals. This ensured that there is no endogeneity problem in the model. The large sample size allows us to assume asymptotic normality of the data. The kdensity of the errors however shows that the errors are not normally distributed. The discussion regarding this is done in the Limitations section. Regardless, almost all of the Multiple Linear Regression Assumptions were satisfied.

This paper will now focus on the interpretations of coefficients generated by running the above-mentioned final regression in STATA.

Being a male is related to an increase the math test score by around 2.24%. This effect is statistically significant with a p-value of 0. This could imply several things. This might be because of the fact that boys are expected to focus more on getting good grades and thus given more attention by parents. Or this might be because of the common theory that boys perform better at math compared to girls.

The coefficients of race dummy variables also give an interesting result. It is suggested that being black, represented by race2, is correlated with a negative effect on math scores. Being black is correlated with a reduction in the math scores by 6.06% and this effect is statistically significant. This might imply a lack of focus of black children towards studies or discrimination by teachers towards black children. Similarly being Hispanic is negatively correlated with math scores by around 1.80% which is statistically significant. After looking at the above relations, one would expect that being white would have a positive relationship with math scores but the regression output suggests otherwise. However, the coefficient is insignificant with a high p-value of 0.241.

The number of years of mother's education has an expected positive effect of 1.96% on math scores. This effect is statistically significant with a high t-stat of 15.75. This suggests that the more years the mother has gained education, the more they can help the children in studies and create a positive effect in the children's math scores. On the other hand, the effect of a working mom on child's performance although negative is highly insignificant which suggests that mother's working status has no effect on child's test scores. Students also living with both parents have a significant positive effect of almost 0.822% on math test scores which implies that children living in a happy family with both parents contributing and relatively less social problems tend to get more marks than students who have problems at home.

Other factors related to a child's home are discussed now. The effect of family income, although positive and significant, is very low. A 10 percent increase in family income increases the math score by just 0.116 percent. This implies that despite having more resources (as represented by higher family income), a child gains quite less in terms of maths scores. An increase of 1 in household size is associated with a fall of 0.398% in math scores. This effect is significant and the directional movement is as expected but the magnitude is much smaller than expected. One would assume that as household size increases, there would be more sources of distraction for the child and this should result in lack of study time and eventually a substantial decrease in scores. But this is not supported by the regression output. The effect of watching television in the afternoon during weekdays is significant at 5% significance level. Children who watch television score around 0.297% less than children who do not watch television.

A relatively major positive contributor to math scores is the binary variable has_home_computer. This shows that having a home computer is related with an increase in math test scores by 3.43%. The effect is very significant and implies that having home computer allows students to play games and work on their cognitive skills to increase their problem solving abilities. This goes against the common perception that having a home computer will have negative effect on child's school performance.

The proxy that this paper uses for time spent studying is has_library_card. However as it turns out that the effect is negative and significant at 5% significance level. The effect is as small as 0.722%. This shows that this is not a very good proxy as one would imagine that time spent studying would have a positive effect on test scores. But this is not the case with this model. This implies that having library card does not mean that the student spends time in library. However, one interpretation of this could be that since our model does not account for ability, it is causing the estimate on library card variable to be biased. This could be because of the fact that students who might not be very intelligent spend more time studying in library but still end up with lower scores because of lower ability. More on this is discussed in the limitations section.

The school related factors suggest that being in a Catholic school, has a negative and significant effect of 1.37% on math test scores. This implies that the common theory that Catholic schools focus more on religion than studies might be correct as students tend to do better in private and other schools. Similarly the effect of being in public school is also negative but insignificant. Another school related factor is the drug problem in schools. This turns out to be very significant that children in schools with drug problem tend to score about 1.39% less in math tests as expected.

Now coming to the basic question that this paper asks, "Whether taking part in extra-curricular activities has a positive effect on math test scores?" Individually the effects of taking part in ECAs come out to be positive and significant for athletics, club and music. And the effects come out to be positive and insignificant for dance and art. These effects have several interesting implications which are discussed in detail in the Conclusions section.

Limitations

As stated earlier, the model suffered from heteroscedasticity. This was not very problematic as robust standard errors were incorporated in the inference. The biggest problem that this model faced was that several very important variables were not recorded in the survey. Two of them being "time spent on studying" and "past performance." Due to the unavailability of these variables, these were present in the error term and the model was unable to control for them. This explains the low R-Squared value as well. Also because of these missing variables, none of the variables showed more than 10% effect on math scores. The third limitation this paper faced was OLS itself. OLS is limited in the sense that it cannot correct the non-normality of the residuals. And hence our residuals violated the normality assumption.

Conclusion

The paper comes to the conclusion that, despite the limitations mentioned previously, taking part in extracurricular activities indeed has a positive effect on students' math scores. All of the coefficients for different ECAs are positive and doing the F-test on them concludes that they are jointly significant (shown in *Exhibit 3*). Individually, taking part in dance is correlated with an increase of 0.0484% but this coefficient is insignificant. Taking part in athletics is correlated with increase in math scores of 1.25%, taking part in club activities are correlated with increase in math scores by 0.656% and taking part in music is correlated with increase in math scores of 1.95%. All these effects are highly significant. Lastly, taking part in art is correlated with increase in test scores of 0.166% which is insignificant. These results open up another area for further research which investigates the relationship different kinds of extracurricular activities have on cognitive abilities and which are most effective. This paper is very important in the sense that it contradicts the common perception of parents that spending time on activities instead of studying has a negative effect on grades.

The findings of this paper are consistent with the studies done by Harry Cooper and Jeffery Valentine who also suggest that extra-curricular activities are related to higher test scores. The results of Beckett's study and those of Pascarella & Terenzini are also similar to what our paper suggests.

Recommendations

The survey data was lacking on one major aspect and that is, the time a student spends studying. Moreover, past performance should also be given because that could be highly correlated with our y variable. Hence, this paper recommends that the survey should take into account ‘hours spent studying’ and ‘past performance’.

Apart from that, this paper recommends that more categorical variables should be included. For instance, the race variable should also include Asian because it’s hypothesized that Asians do better in maths. Similarly, data more school types should be provided.

Works Cited

http://www.academicjournals.org/article/article1379502528_Adeyemo2222.pdf

http://www.jstor.org/stable/pdf/3090254.pdf

http://ir.uiowa.edu/cgi/viewcontent.cgi?article=2733&context=etd

https://ijbnpa.biomedcentral.com/articles/10.1186/1479-5868-5-10

http://psycnet.apa.org/journals/edu/91/2/369/

http://ir.uiowa.edu/cgi/viewcontent.cgi?article=2733&context=etd

Exhibits

1

Variables	Description
Lmath	log values of scores on math
Male	dummy with value 1 for males
race1	dummy with value 1 for whites
race2	dummy with value 1 for blacks
race3	dummy with value 1 for Hispanic
mom_educ	1=1-8 years of education, 2=9-11 years, 3=HS graduate, 4=some college, 5=college graduate
Lincome	log values of annual family income
Hhsize	number of people in household
part_dance	dummy with value 1 for those who take part in dance
part_athletics	dummy with value 1 for those who take part in athletics
part_club	dummy with value 1 for those who take part in club
part_music	dummy with value 1 for those who take part in music
part_art	dummy with value 1 for those who take part in art
tv_afternoon_mf	average hours per day that student watches TV in afternoon from Monday to Friday
both_parents	dummy with value 1 for those who currently live with both parents
has_library_card	dummy with value 1 if student has a library card
has_home_computer	dummy with value 1 if student has access to home computer
Momworks	dummy with value 1 if mom works
schooltype1	dummy with value 1 if student's school is public
schooltype2	dummy with value 1 if student's school is Catholic
problem_drugs	dummy with value 1 for those whose school has a drug problem

2a

VARIABLES	(1) math_test
Gender	-2.199***
	(0.207)
Race	-0.190*
	(0.0972)
mom_educ	2.076***
	(0.110)
family_income	1.54e-05***
	(1.86e-06)
Hhsize	-0.442***
	(0.0794)
part_dance	-0.0894
	(0.299)
part_athletics	1.273***
	(0.223)
part_club	0.726***
	(0.220)
part_music	2.001***
	(0.223)
part_art	0.0669
	(0.314)
tv_afternoon_mf	-0.464***
	(0.124)
both_parents	2.024***
	(0.272)
has_library_card	-0.737**
	(0.308)
has_home_computer	3.712***
	(0.321)
mom_work_status	0.247**
	(0.122)
school_type	-0.157
	(0.177)
problem_drugs	-1.615***
	(0.246)
Constant	91.07***
	(0.750)
Observations	8,105
R-squared	0.234

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

```
Number of obs =      8105
F( 17,  8087) =    145.48
Prob > F      =    0.0000
R-squared     =    0.2342
Adj R-squared =    0.2326
Root MSE      =    8.7602
```

2b

VARIABLES	(1) lmath
Male	0.0224***
	(0.00215)
race1	-0.00406
	(0.00349)
race2	-0.0606***
	(0.00478)
race3	-0.0180***
	(0.00405)
mom_educ	0.0196***
	(0.00119)
lincome	0.0116***
	(0.00132)
hhsize	-0.00398***
	(0.000826)
part_dance	0.000484
	(0.00311)
part_athletics	0.0125***
	(0.00233)
part_club	0.00656***
	(0.00229)
part_music	0.0195***
	(0.00232)
part_art	0.00166
	(0.00327)
tv_afternoon_mf	-0.00297**
	(0.00130)
both_parents	0.00822***
	(0.00295)
has_library_card	-0.00722**
	(0.00321)
has_home_computer	0.0343***
	(0.00338)
momworks	-0.00165
	(0.00244)
schooltype1	-0.00341
	(0.00417)
schooltype2	-0.0137***
	(0.00479)
problem_drugs	-0.0139***
	(0.00257)
Constant	4.380***
	(0.0149)
Observations	8,105
R-squared	0.255

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

```
Number of obs =     8105
F( 20,  8084) =   138.63
Prob > F      =   0.0000
R-squared     =   0.2554
Adj R-squared =   0.2535
Root MSE      =   .09113
```

2c

```
. ovtest

Ramsey RESET test using powers of the fitted values of lmath
       Ho:  model has no omitted variables
                F(3, 8081) =       1.74
                  Prob > F =       0.1563
```

2d

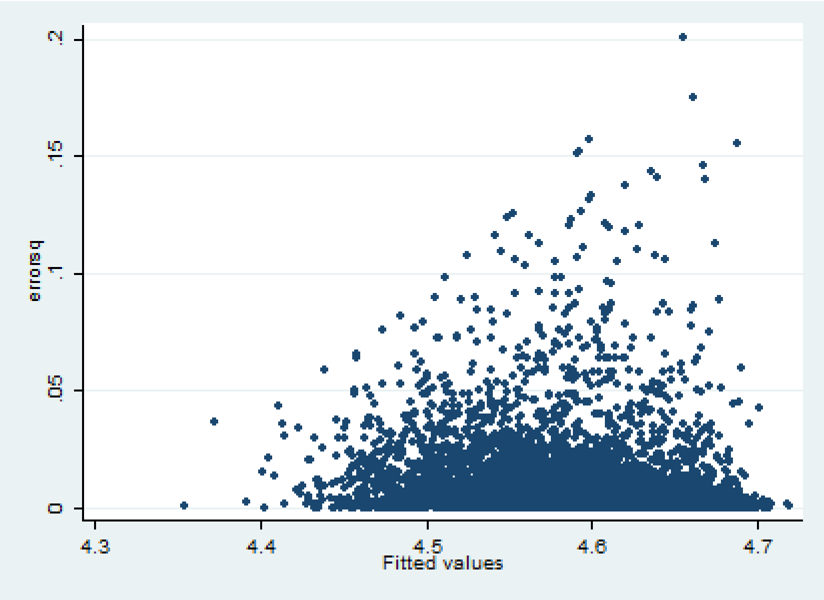
	lmath	male	race1	race2	race3	mom_educ	lincome
lmath	1.0000						
male	0.0990	1.0000					
race1	0.2623	0.0101	1.0000				
race2	-0.2372	0.0007	-0.4068	1.0000			
race3	-0.1749	0.0000	-0.6060	-0.1481	1.0000		
mom_educ	0.3850	-0.0037	0.3118	-0.0985	-0.3311	1.0000	
lincome	0.3435	0.0052	0.3307	-0.2324	-0.2362	0.4826	1.0000
hhsize	-0.0919	0.0054	-0.1522	-0.0210	0.1077	-0.1399	-0.0561
part_dance	0.0107	-0.2930	-0.0324	-0.0025	0.0140	0.0808	0.0590
part_athle~s	0.2135	0.1194	0.2191	-0.0995	-0.1534	0.2807	0.2414
part_club	0.1271	-0.0737	0.1675	-0.0599	-0.1288	0.1942	0.1285
part_music	0.1986	-0.1029	0.0825	-0.0743	-0.0960	0.2449	0.1913
part_art	0.0356	-0.0561	-0.0270	0.0166	-0.0199	0.0683	0.0417
tv_afterno~f	-0.1394	0.0163	-0.1618	0.1445	0.0937	-0.1823	-0.1773
both_parents	0.1657	0.0085	0.1619	-0.2879	-0.0144	0.1360	0.2916
has_librar~d	0.0872	0.0617	-0.0253	0.0490	-0.0164	-0.1302	-0.0997
has_home_c~r	0.2831	-0.0037	0.2681	-0.1542	-0.1982	0.3257	0.3219
momworks	0.0600	0.0177	0.0612	0.0168	-0.0858	0.1537	0.0866
schooltype1	-0.1178	0.0138	-0.1408	0.0824	0.1013	-0.2473	-0.2367
schooltype2	0.0773	-0.0120	0.1046	-0.0792	-0.0632	0.1941	0.1818
problem_dr~s	-0.1793	-0.0013	-0.1666	0.1043	0.1088	-0.2120	-0.2204

	hhsize	part_d~e	part_a~s	part_c~b	part_m~c	part_art	tv_~n_mf
hhsize	1.0000						
part_dance	-0.0092	1.0000					
part_athle~s	-0.0577	0.0027	1.0000				
part_club	-0.0332	0.0622	0.1265	1.0000			
part_music	-0.0188	0.1182	0.1222	0.1443	1.0000		
part_art	-0.0157	0.1003	0.0322	0.0742	0.1332	1.0000	
tv_afterno~f	0.0341	-0.0445	-0.1218	-0.0625	-0.1061	-0.0319	1.0000
both_parents	0.2733	0.0105	0.1112	0.0800	0.0970	0.0014	-0.0773
has_librar~d	0.0063	-0.0394	-0.0568	-0.0840	-0.0797	-0.0590	0.0710
has_home_c~r	-0.0615	0.0306	0.1545	0.1071	0.1424	0.0160	-0.1082
momworks	-0.1514	0.0135	0.0992	0.0332	0.0058	0.0061	-0.0047
schooltype1	0.0299	-0.0406	-0.1665	-0.0628	-0.1230	-0.0304	0.1221
schooltype2	-0.0216	0.0405	0.1424	0.0681	0.0823	0.0266	-0.1046
problem_dr~s	0.0734	-0.0104	-0.0855	-0.0400	-0.0781	-0.0032	0.0716

	both_p~s	has_li~d	has_ho~r	momworks	school~1	school~2	prob~ugs
both_parents	1.0000						
has_librar~d	-0.0345	1.0000					
has_home_c~r	0.2090	-0.1120	1.0000				
momworks	-0.0508	-0.0123	0.0937	1.0000			
schooltype1	-0.0818	0.0794	-0.1429	-0.0195	1.0000		
schooltype2	0.0764	-0.0613	0.1259	0.0241	-0.7897	1.0000	
problem_dr~s	-0.0853	0.0448	-0.1569	-0.0051	0.0671	-0.0360	1.0000

Variable	VIF	1/VIF
race1	2.78	0.359381
schooltype1	2.77	0.360722
schooltype2	2.67	0.374093
race3	2.37	0.421571
race2	1.84	0.544354
mom_educ	1.61	0.620668
lincome	1.55	0.646432
both_parents	1.31	0.765189
has_home_c~r	1.23	0.812197
hhsize	1.18	0.850413
part_athle~s	1.16	0.858744
part_music	1.13	0.885064
male	1.13	0.885109
part_dance	1.12	0.891938
problem_dr~s	1.08	0.922103
part_club	1.08	0.923343
tv_afterno~f	1.08	0.930177
momworks	1.06	0.946617
has_librar~d	1.04	0.958535
part_art	1.04	0.965178
Mean VIF	1.51	

2e




```
. hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of lmath

         chi2(1)      =   482.35
         Prob > chi2  =   0.0000
```

VARIABLES	(1) errorsq
yhat	0.781*
	(0.402)
yhatsq	-0.0911**
	(0.0439)
Constant	-1.656*
	(0.922)
Observations	8,105
R-squared	0.037

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

```
Number of obs =     8105
F(  2,  8102) =   157.76
Prob > F      =   0.0000
R-squared     =   0.0375
Adj R-squared =   0.0372
Root MSE      =   .01458
```

2f

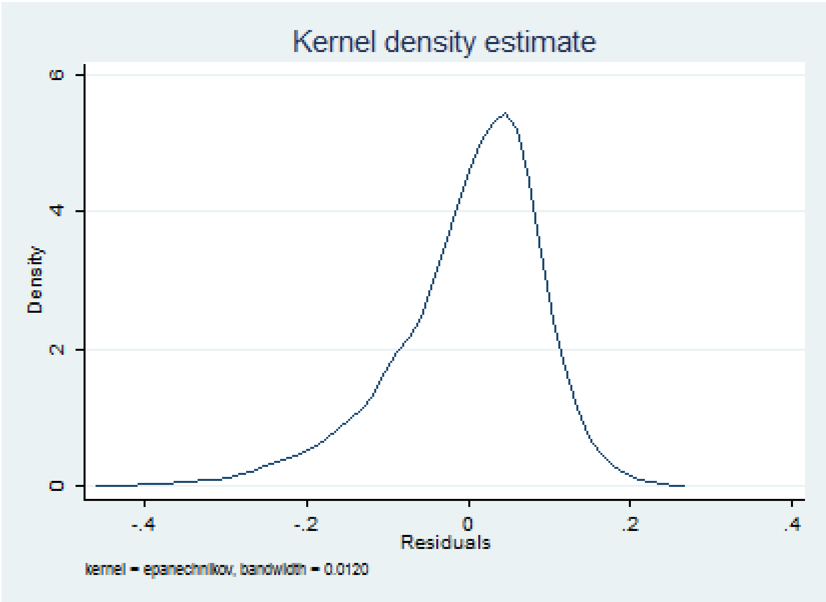
VARIABLES	(1) lmath
male	0.0224***
	(0.00216)
race1	-0.00406
	(0.00346)
race2	-0.0606***
	(0.00516)
race3	-0.0180***
	(0.00428)
mom_educ	0.0196***
	(0.00125)
lincome	0.0116***
	(0.00166)
Hhsize	-0.00398***
	(0.000875)
part_dance	0.000484
	(0.00302)
part_athletics	0.0125***
	(0.00249)
part_club	0.00656***
	(0.00220)
part_music	0.0195***
	(0.00218)
part_art	0.00166
	(0.00313)
tv_afternoon_mf	-0.00297**
	(0.00142)
both_parents	0.00822***
	(0.00313)
has_library_card	-0.00722**
	(0.00339)
has_home_computer	0.0343***
	(0.00408)
momworks	-0.00165
	(0.00252)
schooltype1	-0.00341
	(0.00346)
schooltype2	-0.0137***
	(0.00391)
problem_drugs	-0.0139***
	(0.00278)
Constant	4.380***
	(0.0180)
Observations	8,105
R-squared	0.255

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

```
Number of obs =     8105
F( 20,  8084) =   119.86
Prob > F      =   0.0000
R-squared     =   0.2554
Root MSE      =   .09113
```

2g

Variable	Obs	Mean	Std. Dev.	Min	Max
errors	8105	3.27e-11	.0910204	-.4477191	.2563214



3

```
. test part_dance part_athletics part_club part_music part_art

 ( 1)  part_dance = 0
 ( 2)  part_athletics = 0
 ( 3)  part_club = 0
 ( 4)  part_music = 0
 ( 5)  part_art = 0

       F(  5,  8084) =   25.55
            Prob > F =    0.0000
```

DO – FILE

```
use "C:\Users\zaid_\Documents\RESOURCES\ECON 330\Project\ecls_5th_grade.dta"

regress math_test gender race mom_educ family_income hhsize part_dance part_athletics part_club part_music part_art tv_afternoon_mf both_parents has_library_card has_home_computer mom_work_status school_type problem_drugs

outreg2 using "C:\Users\zaid_\Documents\RESOURCES\ECON 330\Project\YO\0.Regression 0.doc"

gen lmath = log(math_test)

gen lincome = log(family_income)

foreach x of numlist 1/4{

gen race`x' = race==`x' if !missing(race)

}

foreach x of numlist 1/3{

gen schooltype`x' = school_type==`x' if !missing(school_type)

}

gen momworks = (mom_work_status==1 | mom_work_status==2) if !missing(mom_work_status)

gen male = gender==1

regress lmath male race1 race2 race3 mom_educ lincome hhsize part_dance part_athletics part_club part_music part_art tv_afternoon_mf both_parents has_library_card has_home_computer momworks schooltype1 schooltype2 problem_drugs

outreg2 using "C:\Users\zaid_\Documents\RESOURCES\ECON 330\Project\YO\1.Regression Number 1.doc"

predict yhat

predict errors, residuals

gen errorsq = errors*errors

gen yhatsq = yhat*yhat

scatter errorsq yhat

*graph shows hetroskedasticity

regress errorsq yhat yhatsq

outreg2 using "C:\Users\zaid_\Documents\RESOURCES\ECON 330\Project\YO\4.Regression White Test.doc"

*White Test shows hetroskedasticity

regress lmath male race1 race2 race3 mom_educ lincome hhsize part_dance part_athletics part_club part_music part_art tv_afternoon_mf both_parents has_library_card has_home_computer momworks schooltype1 schooltype2 problem_drugs, robust
```

outreg2 using "C:\Users\zaid_\Documents\RESOURCES\ECON 330\Project\YO\5.Regression with Robust SE.doc"

*used robust standard errors

vif

*multicollinearity not a problem

summ errors

*see if E(u) = 0

foreach var of varlist male race1 race2 race3 mom_educ lincome hhsize part_dance part_athletics part_club part_music part_art tv_afternoon_mf both_parents has_library_card has_home_computer momworks schooltype1 schooltype2 problem_drugs{

regress errors `var'

}

pwcorr lmath male race1 race2 race3 mom_educ lincome hhsize part_dance part_athletics part_club part_music part_art tv_afternoon_mf both_parents has_library_card has_home_computer momworks schooltype1 schooltype2 problem_drugs

*check for collinerity between errors and independent variables to conclude E(u|xi's)=0

kdensity errors

*to check normality of residuals

regress lmath male race1 race2 race3 mom_educ lincome hhsize part_dance part_athletics part_club part_music part_art tv_afternoon_mf both_parents has_library_card has_home_computer momworks schooltype1 schooltype2 problem_drugs

hettest

ovtest

*B-P and RESET Test

regress lmath male race1 race2 race3 mom_educ lincome hhsize part_dance part_athletics part_club part_music part_art tv_afternoon_mf both_parents has_library_card has_home_computer momworks schooltype1 schooltype2 problem_drugs, robust

test part_dance part_athletics part_club part_music part_art

*Testing for Joint Significance of Extra-Curricular Activities