

Introduction
Variables
First Model
Second Model

Is ENAC for the elite ?

LAVRILLOUX Martin, SILVE Lucas, VIDAL Pablo

ENAC, Econometrics

April 07, 2017



Table of content

- 1 Introduction
 - Objectives
 - Online Study
 - The datas
- 2 Variables
- 3 First Model
- 4 Second Model

Objectives

1.

Identifying the parameters witch could show that engineering schools are elitist

2.

Compare three different type of schools : ENAC, POLYTECHNIQUE and CENTRALE LYON

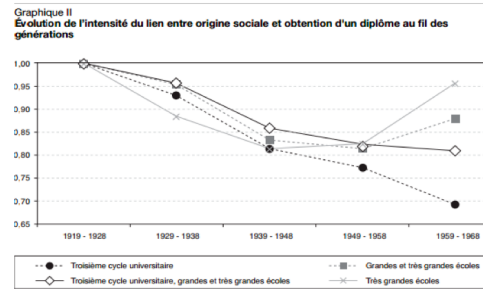
3.

Compare our results with an online study from 1940 to 1980

Online study

Observations

- Graph from a report of Insee about "grandes écoles"
- Strong link with social origin



Database

- Use of a Google Form
- $n = 634$
- 3 different schools
- total of $k = 6$ variables

Explanatory variables	Details	Expectations	Reality
Income	Euros	?	?
Large town	dummy	+	+
Proximity of a large town	dummy	+	?
Rural environment	dummy	-	-
Polytechnique	dummy	+	+
Centrale Lyon	dummy	+	-
ENAC	dummy	-	+
Private employee	dummy	+	+
Civil servant	dummy	+	+
Manual labour	dummy	-	?
Executive	dummy	+	+
Unemployed	dummy	-	-
Profession	dummy	+	+
Siblings	number	+	+

Introduction
Variables
First Model
Second Model

Descriptive Variables
Possible Dependent Variables
Choice of the variables
Analysis of the correlation matrix

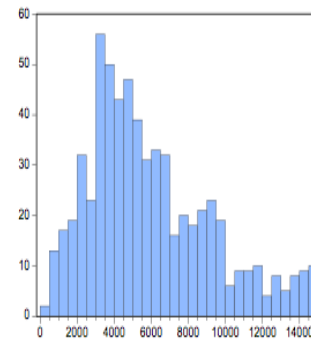
Table of content

- 1 Introduction
- 2 **Variables**
 - Descriptive Variables
 - Possible Dependent Variables
 - Choice of the variables
 - Analysis of the correlation matrix
- 3 First Model
- 4 Second Model



Income

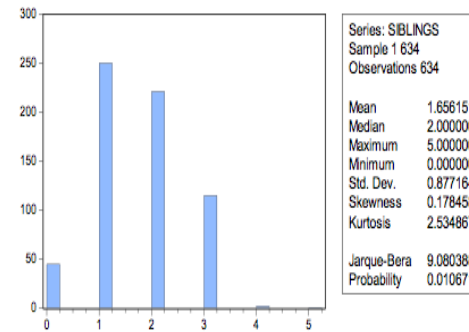
- SK superior to 0 so high concentration of "low" income
- mean of 5964 euros



Series: INCOME	
Sample 1 634	
Observations 632	
Mean	5963.701
Median	5075.500
Maximum	14997.00
Minimum	5.000000
Std. Dev.	3379.648
Skewness	0.795598
Kurtosis	3.025204
Jarque-Bera	66.69019
Probability	0.000000

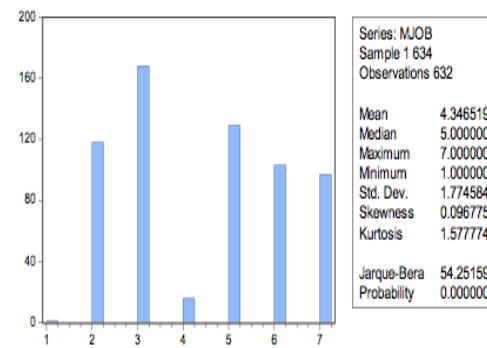
Siblings

- 39% for 1
- 35% for 2
- 18% for 3



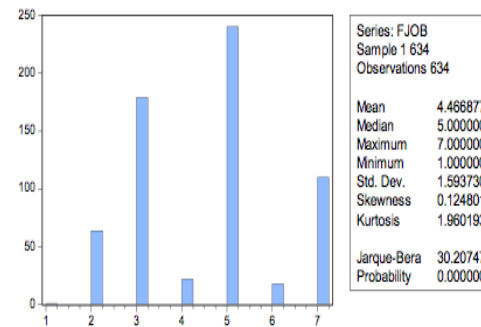
MJob

- Good repartition, around 20% for each



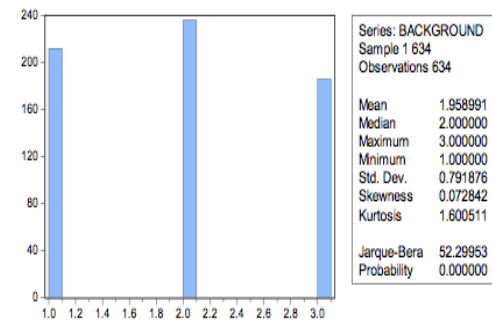
FJob

- Executive and Civil servant are the most represented with 66%
- Profession and private employee around 30%
- Manual and Unemployed less than 5%



Background

- 33% from Big cities
- 37% near Big cities
- 30% from rural environnement



Possible Dependent variables

- For our first model, the aim is to show the impact of variables such as BACKGROUND, JOB, SCHOOL, SIBLINGS on INCOME
- For others models it will be interesting to choose ENAC and POLYTECHNIQUE and CENTRALE LYON as Dependent variables in order to compare the linear regressions

Choice of variables

How do we choose our variables ?

- Analysis of the correlation matrix
- Suppression of some variables to avoid multicollinearity
- Non significant variables
- Try to increase the Adjusted R squared

Introduction
Variables
First Model
Second Model

Descriptive Variables
Possible Dependent Variables
Choice of the variables
Analysis of the correlation matrix

Analysis of the correlation matrix

	BACKGROUND=1	BACKGROUND=2	BACKGROUND=3	MIDB=1	MIDB=2	MIDB=3	MIDB=4	MIDB=5	MIDB=6	MIDB=7	FIDB=1	FIDB=2	FIDB=3	FIDB=4	FIDB=5	FIDB=6	FIDB=7	SRLINGS	INCOME	SCHOOL=1	SCHOOL=2	SCHOOL=3
BACKGROUND=1	1.00000	-0.54984	-0.45814	-0.02851	0.04471	-0.071205	-0.029240	0.039579	-0.023219	0.041593	-0.02851	0.017489	-0.020163	-0.080276	0.075108	-0.101750	0.036345	0.004540	0.157430	0.128965	-0.072256	-0.064952
BACKGROUND=2	0.54984	1.00000	-0.495136	-0.030609	-0.008232	0.005714	-0.081207	0.062024	-0.013172	-0.026724	-0.030609	-0.051131	0.029699	-0.056436	0.111691	-0.013185	-0.010808	0.006592	0.076625	-0.030355	-0.032547	0.054952
BACKGROUND=3	-0.45814	-0.495136	1.00000	0.061820	-0.037045	0.067729	0.117321	-0.103941	0.025514	-0.014776	0.061820	0.036065	0.020504	0.142998	-0.196205	0.119410	0.070241	-0.011469	-0.223619	-0.101464	0.196887	0.009693
MIDB=1	-0.02851	-0.030609	0.061820	1.00000	-0.018975	-0.024089	-0.006447	-0.020170	-0.017658	-0.017639	1.00000	-0.013077	-0.029209	-0.007597	0.031238	-0.006849	-0.018371	0.061356	-0.064590	0.054984	-0.019774	-0.036672
MIDB=2	0.04471	-0.008232	-0.037045	-0.018975	1.00000	-0.287061	-0.076825	-0.240557	-0.210424	-0.203049	-0.018975	0.189476	-0.053026	0.043349	0.024462	-0.057031	-0.111003	-0.059897	0.070194	0.051506	0.043576	0.015527
MIDB=3	-0.071205	0.005714	0.067729	-0.024089	-0.287061	1.00000	-0.097529	-0.305134	-0.267134	-0.257771	-0.024089	-0.099979	0.266898	0.002425	0.146838	0.047250	-0.117108	0.017338	-0.064946	0.007785	-0.046237	-0.021075
MIDB=4	-0.029240	-0.081207	0.117321	-0.006447	-0.076825	-0.097529	1.00000	0.081661	-0.071492	-0.068986	-0.006447	0.049418	0.033001	0.124126	-0.064063	0.027790	-0.047796	-0.062899	-0.113591	0.053526	0.087699	-0.026872
MIDB=5	0.039579	0.062024	-0.103941	-0.020170	-0.240557	-0.305134	-0.081661	1.00000	-0.223672	-0.213832	-0.020170	-0.085380	-0.028251	-0.031746	0.149405	0.007793	-0.076768	-0.070539	0.155354	-0.096628	0.047310	-0.028952
MIDB=6	-0.023219	-0.013172	0.025514	-0.017658	-0.210424	-0.267134	-0.071492	-0.223672	1.00000	-0.188954	-0.017658	0.029197	-0.134937	0.009294	0.060749	0.052888	0.033786	-0.015355	-0.103557	0.004210	-0.024954	0.015862
MIDB=7	0.041593	-0.026724	-0.014776	-0.017639	-0.203049	-0.257771	-0.068986	-0.213832	-0.188954	1.00000	-0.017639	-0.110182	-0.111896	-0.081292	-0.044049	-0.046887	0.336492	0.038294	0.098988	0.004961	-0.056680	0.040256
FIDB=1	-0.02851	-0.030609	0.061820	1.00000	-0.018975	-0.024089	-0.006447	-0.020170	-0.017658	-0.017639	1.00000	-0.013077	-0.029209	-0.007597	0.031238	-0.006849	-0.018371	0.061356	-0.064590	0.054984	-0.019774	-0.036672
FIDB=2	0.017489	-0.051131	0.036065	-0.013077	-0.080476	-0.099979	0.049418	-0.085380	0.029197	-0.110182	-0.013077	1.00000	-0.202879	-0.062389	-0.236541	-0.056248	-0.153070	-0.127952	0.123276	0.045709	0.013160	0.033103
FIDB=3	-0.050163	0.029699	0.020504	-0.029699	-0.053026	0.266898	0.033001	-0.028251	-0.134937	-0.111896	-0.029699	-0.202879	1.00000	-0.196602	0.491800	-0.078239	-0.289224	-0.041543	-0.187330	0.019240	0.023810	-0.038968
FIDB=4	-0.080276	-0.056436	-0.142998	-0.007597	0.043349	0.002425	0.124126	-0.031746	0.009294	-0.081292	-0.007597	-0.062389	-0.196602	1.00000	-0.148033	0.032676	-0.087646	-0.048231	-0.116030	0.025671	-0.029083	-0.001270
FIDB=5	0.075108	0.111691	-0.196205	-0.031238	-0.024692	-0.146838	-0.044063	0.149405	0.060749	-0.044049	-0.031238	-0.236541	-0.491800	-0.149033	1.00000	0.134364	-0.360395	0.04095	-0.061118	0.003768	-0.030354	0.036113
FIDB=6	-0.101750	-0.013185	0.119410	-0.006849	-0.057031	0.047250	0.027730	0.007983	0.052888	-0.046887	-0.006849	-0.056248	-0.107829	0.032676	-0.134364	1.00000	-0.079018	0.023484	-0.173961	0.018950	0.036759	-0.020242
FIDB=7	0.036345	-0.010808	0.070241	-0.018371	-0.111003	-0.117108	-0.047796	-0.076768	0.033786	0.336492	-0.018371	-0.152070	-0.289224	-0.087946	-0.360395	-0.079018	1.00000	0.024487	0.095039	-0.017161	0.013831	0.005239
SRLINGS	0.004540	0.006592	-0.011469	0.061356	-0.059897	0.017338	0.062899	0.101535	0.038294	0.061356	-0.127952	-0.041543	-0.048231	0.104095	0.013484	0.024487	1.00000	0.109295	0.100308	-0.178053	0.046695	
INCOME	0.157430	0.076625	-0.223619	-0.064590	-0.070194	-0.046436	-0.113591	-0.053526	-0.035527	-0.039888	-0.064590	-0.123276	-0.187330	-0.116030	-0.073061	-0.095039	0.109295	1.00000	0.097929	-0.004919	-0.089998	
SCHOOL=1	0.128965	-0.030355	-0.101464	0.054984	-0.051906	0.067865	-0.035525	-0.029628	0.004210	0.004961	0.054984	-0.045709	0.019240	0.029471	0.003768	0.015862	-0.017161	-0.010083	0.097929	1.00000	-0.358622	-0.666961
SCHOOL=2	-0.072256	-0.032547	0.101464	-0.054984	-0.046436	-0.067865	-0.035525	-0.029628	0.004210	0.004961	0.054984	-0.045709	0.019240	0.029471	0.003768	0.015862	-0.017161	-0.010083	0.097929	-0.358622	1.00000	-0.666961
SCHOOL=3	-0.064952	0.054952	0.009693	-0.019774	-0.036672	-0.021075	-0.026872	-0.028952	-0.015862	-0.040256	-0.033103	-0.038968	-0.036113	-0.030354	-0.036113	-0.030354	-0.036113	-0.030354	-0.036113	-0.030354	-0.036113	1.00000

LAVRILLOUX Martin, SILVE Lucas, VIDAL Pablo

Is ENAC for the elite ?

Introduction
Variables
First Model
Second Model

Linear Regression
Positive marginal effects
Negative marginal effects
Residuals
White Test
Improvement of the model
1st Model 2

Table of content

- 1 Introduction
- 2 Variables
- 3 **First Model**
 - Linear Regression
 - Positive marginal effects
 - Negative marginal effects
 - Residuals
 - White Test
 - Improvement of the model
 - 1st Model 2



Introduction
Variables
First Model
Second Model

Linear Regression
Positive marginal effects
Negative marginal effects
Residuals
White Test
Improvement of the model
1rst Model 2

The first linear regression

Observations

- 5 insignificant variables
- F statistic of 12.56 superior to $F(15,618)$ so significant
- Adjusted R squared of 0.22

Dependent Variable: INCOME
Method: Least Squares
Date: 04/06/17 Time: 18:18
Sample: 1 634
Included observations: 634-

	Coefficient	Std. Error	t-Statistic	Prob.
C	1810.057	982.5331	1.842235	0.0659
BACKGROUND=1	127.1952	284.8959	0.446462	0.6554
BACKGROUND=3	-917.2507	302.8584	-3.028645	0.0026
FJOB=2	1385.124	748.8746	1.851860	0.0641
FJOB=3	1072.551	685.6797	1.564216	0.1153
FJOB=5	3216.834	683.2023	4.708465	0.0000
FJOB=6	-1078.775	956.4840	-1.127855	0.2598
FJOB=7	2626.548	722.3733	3.912863	0.0001
MJOB=2	867.7535	804.2925	1.078903	0.2811
MJOB=3	1389.534	793.4258	1.751310	0.0804
MJOB=5	2323.105	807.5182	2.876845	0.0042
MJOB=6	329.8287	818.2045	0.402866	0.6928
MJOB=7	1594.359	834.8254	1.909812	0.0566
SCHOOL=1	758.9281	269.7612	2.813333	0.0051
SCHOOL=2	719.0473	325.9201	2.206207	0.0277
SIBLINGS	350.3317	140.2352	2.498172	0.0127
R-squared	0.234641	Mean dependent var.		5999.480
Adjusted R-squared	0.215974	S.D. dependent var.		3352.375
S.E. of regression	2968.368	Akaike info criterion		18.85444
Sum squared resid	5.42E+09	Schwarz criterion		18.96721
Log likelihood	-5932.576	Hannan-Quinn criter.		18.89524
F-statistic	12.56966	Durbin-Watson stat		2.057731
Prob(F-statistic)	0.000000			

Navigation icons: back, forward, search, etc.

Positive marginal effects

- marginal effect of having a father or a mother who is a executive is positive
- wage of executive is usually higher than other jobs.
- FJOB=5 is 3216.834, so when a student has a father who is executive, his household income is supposed to increase its monthly income by 3216 euros
- Regarding the school of the students, we remarks that the estimated income of the household of Polytechnique student is supposed to be higher than the household of ENAC student

Introduction
Variables
First Model
Second Model

Linear Regression
Positive marginal effects
Negative marginal effects
Residuals
White Test
Improvement of the model
1st Model 2

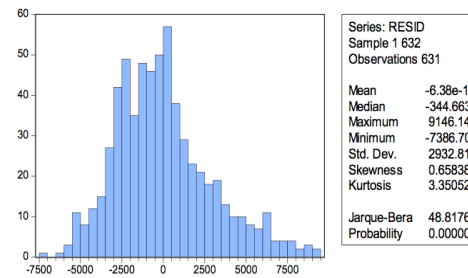
Negative marginal effects

- marginal effect of living in a rural environment is negative
- coherent because you need less money to live in countryside than in city.

Residuals

Observations

- $JB = 48.8$
- Superior to 4.61
- Reject the assumption of normality



White Test

Observations

- nRsquare = 107.7
- ProbChisquared > 0.05
- Reject Heteroskedasticity

Heteroskedasticity Test: White

F-statistic	1.131685	Prob. F(97,533)	0.2008
Obs*R-squared	107.7628	Prob. Chi-Square(97)	0.2138
Scaled explained SS	120.3079	Prob. Chi-Square(97)	0.0545

Improvement of the model

Observations

- $q = 5$
- $qF = 13.68$
greater than
11.07
- Reject the null
of
simplification

Dependent Variable: INCOME
Method: Least Squares
Date: 04/06/17 Time: 18:46
Sample: 1 632
Included observations: 631

	Coefficient	Std. Error	t-Statistic	Prob.
C	3284.101	397.6797	8.258155	0.0000
BACKGROUND=3	-1143.894	269.1713	-4.249688	0.0000
FJOB=2	667.4949	442.0393	1.510035	0.1315
FJOB=5	2420.973	292.8950	8.265669	0.0000
FJOB=7	2025.656	373.1115	5.429092	0.0000
MJOB=3	883.7290	311.3312	2.838549	0.0047
MJOB=5	1765.366	333.3386	5.296015	0.0000
MJOB=7	1091.123	381.7209	2.858431	0.0044
SCHOOL=1	732.4491	269.9073	2.712706	0.0068
SCHOOL=2	693.8485	326.1961	2.127090	0.0338
SIBLINGS	336.8570	140.4246	2.398847	0.0167
R-squared	0.218191	Mean dependent var	5999.480	
Adjusted R-squared	0.205581	S.D. dependent var	3352.375	
S.E. of regression	2987.978	Akaike info criterion	18.95086	
Sum squared resid	5.54E+09	Schwarz criterion	18.93739	
Log likelihood	-5939.286	Hannan-Quinn criter.	18.88997	
F-statistic	17.30323	Durbin-Watson stat	2.038974	
Prob(F-statistic)	0.000000			

Improvement of the model

Observations

- R-squared is nearly equal to the one of the last model
- the marginal effect of Centrale Lyon is negative
- leads to an approximate 744 decrease in income of household.

Dependent Variable: INCOME
Method: Least Squares
Date: 04/06/17 Time: 19:56
Sample: 1 632
Included observations: 631

	Coefficient	Std. Error	t-Statistic	Prob.
C	2548.312	972.7412	2.619722	0.0090
BACKGROUND=1	129.0885	283.8476	0.456895	0.6479
BACKGROUND=3	-920.0040	301.6749	-3.049653	0.0024
FJOB=2	1379.584	744.7326	1.852456	0.0644
FJOB=3	1066.889	683.3731	1.561210	0.1190
FJOB=5	3211.418	681.0415	4.715451	0.0000
FJOB=6	-1086.322	953.4799	-1.139324	0.2550
FJOB=7	2819.761	719.3977	3.919614	0.0001
MJOB=2	873.1976	802.2639	1.088417	0.2768
MJOB=3	1398.540	788.9444	1.772672	0.0768
MJOB=5	2329.494	804.5719	2.893892	0.0039
MJOB=6	337.2462	814.8831	0.413858	0.6791
MJOB=7	1603.298	830.5568	1.930389	0.0540
SCHOOL=3	-744.5710	239.1553	-3.113337	0.0019
SIBLINGS	353.0298	138.1610	2.555205	0.0109
R-squared	0.234625	Mean dependent var	5999.480	
Adjusted R-squared	0.217230	S.D. dependent var	3352.375	
S.E. of regression	2965.990	Akaike info criterion	18.95129	
Sum squared resid	5.42E+09	Schwarz criterion	18.95701	
Log likelihood	-5932.583	Hannan-Quinn criter.	18.89236	
F-statistic	13.48815	Durbin-Watson stat	2.056570	
Prob(F-statistic)	0.000000			

Introduction
Variables
First Model
Second Model

Linear Regression1
Analysis
Linear Regression2
Analysis
Linear Regression3
Analysis
Conclusion

Table of content

- 1 Introduction
- 2 Variables
- 3 First Model
- 4 **Second Model**
 - Linear Regression1
 - Analysis
 - Linear Regression2
 - Analysis
 - Linear Regression3



Linear Regression1

Dependent Variable: SCHOOL=1
Method: Least Squares
Date: 04/06/17 Time: 19:43
Sample: 1 632
Included observations: 631

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.211885	0.153953	1.376295	0.1692
INCOME	1.39E-05	6.33E-06	2.189340	0.0289
BACKGROUND=1	0.116478	0.044808	2.599483	0.0096
BACKGROUND=3	-0.047099	0.048079	-0.979620	0.3277
FJOB=2	-0.191613	0.118134	-1.621993	0.1053
FJOB=3	-0.136161	0.108351	-1.256670	0.2093
FJOB=5	-0.191353	0.109649	-1.745136	0.0815
FJOB=6	-0.046878	0.150802	-0.310857	0.7560
FJOB=7	-0.200972	0.115236	-1.744008	0.0817
MJOB=2	0.072005	0.127016	0.566896	0.5710
MJOB=3	0.161345	0.125104	1.289689	0.1976
MJOB=5	0.093730	0.128183	0.731219	0.4649
MJOB=6	0.128957	0.128913	1.000341	0.3175
MJOB=7	0.118209	0.131676	0.897730	0.3697
SIBLINGS	0.046572	0.021943	2.122420	0.0342
R-squared	0.049001	Mean dependent var	0.345483	
Adjusted R-squared	0.027387	S.D. dependent var	0.475903	
S.E. of regression	0.469341	Akaike info criterion	1.348510	
Sum squared resid	135.6930	Schwarz criterion	1.454230	
Log likelihood	-410.4548	Hannan-Quinn criter.	1.389571	
F-statistic	2.267126	Durbin-Watson stat	0.522681	

Analysis

- very small R-squared, F-statistic=2.267126, reject the null, relevant model
- SIBLINGS, BACKGROUND=1, INCOME, FJOB=5 and FJOB=7 are significant
- father self-employed or executive has a strong impact on being a student of Polytechnique and the numbers of brothers and sisters a small impact
- surprisingly the income has really poor impact when we look at previous models

Linear Regression2

Dependent Variable: SCHOOL=2
Method: Least Squares
Date: 04/06/17 Time: 19:44
Sample: 1 632
Included observations: 631

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.368558	0.127738	2.885258	0.0040
INCOME	6.96E-06	5.25E-06	1.325178	0.1856
BACKGROUND=1	-0.032002	0.037178	-0.860764	0.3897
BACKGROUND=3	0.080620	0.039893	2.020922	0.0437
FJOB=2	0.110372	0.098019	1.126026	0.2606
FJOB=3	0.146172	0.089901	1.625925	0.1045
FJOB=5	0.107245	0.090979	1.178789	0.2389
FJOB=6	0.268560	0.125124	2.146358	0.0322
FJOB=7	0.155208	0.095614	1.623276	0.1050
MJOB=2	-0.172117	0.105388	-1.633175	0.1029
MJOB=3	-0.260953	0.103801	-2.513967	0.0122
MJOB=5	-0.195681	0.106357	-1.839858	0.0663
MJOB=6	-0.225684	0.106962	-2.109944	0.0353
MJOB=7	-0.282466	0.109255	-2.585391	0.0100
SIBLINGS	-0.079232	0.018207	-4.351835	0.0000
R-squared	0.068050	Mean dependent var	0.198098	
Adjusted R-squared	0.046869	S.D. dependent var	0.398883	
S.E. of regression	0.389423	Akaike info criterion	0.975183	
Sum squared resid	93.41656	Schwarz criterion	1.080904	
Log likelihood	-292.6704	Hannan-Quinn criter.	1.016245	
F-statistic	3.212824	Durbin-Watson stat	1.268750	
<small>Prob(F-statistic)</small>	<small>0.000000</small>			



Analysis

- BACKGROUND=3, FJOB=6(Unemployed), MJOB=3, MJOB=5, MJOB=6, MJOB=7 and SIBLINGS are significant
- strong impact of the mother employment on the fact of being a student at the ENAC

Linear Regression3

Dependent Variable: SCHOOL=3
Method: Least Squares
Date: 04/06/17 Time: 19:45
Sample: 1 632
Included observations: 631

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.419557	0.162633	2.579773	0.0101
INCOME	-2.08E-05	6.68E-06	-3.113337	0.0019
BACKGROUND=1	-0.084476	0.047335	-1.784665	0.0748
BACKGROUND=3	-0.033520	0.050790	-0.659975	0.5095
FJOB=2	0.081241	0.124795	0.650998	0.5153
FJOB=3	-0.010011	0.114460	-0.087466	0.9303
FJOB=5	0.084109	0.115832	0.726127	0.4680
FJOB=6	-0.221682	0.159304	-1.391568	0.1646
FJOB=7	0.045764	0.121733	0.375942	0.7071
MJOB=2	0.100112	0.134177	0.746120	0.4559
MJOB=3	0.099608	0.132157	0.753712	0.4513
MJOB=5	0.101951	0.135411	0.752904	0.4518
MJOB=6	0.096727	0.136181	0.710281	0.4778
MJOB=7	0.164257	0.139100	1.180851	0.2381
SIBLINGS	0.032660	0.023180	1.408960	0.1594

R-squared	0.032743	Mean dependent var	0.456418
Adjusted R-squared	0.010760	S.D. dependent var	0.498492
S.E. of regression	0.495803	Akaike info criterion	1.458208
Sum squared resid	151.4255	Schwarz criterion	1.563929
Log likelihood	-445.0647	Hannan-Quinn criter.	1.499270
F-statistic	1.489481	Durbin-Watson stat	1.028598
Prob(F-statistic)	0.109393		



Introduction
Variables
First Model
Second Model

Linear Regression1
Analysis
Linear Regression2
Analysis
Linear Regression3
Analysis
Conclusion

Analysis

- F-statistic=1.489481
- the model is not significant

Conclusion

- Mean of income greater than the average household income
- First Model : Shows a strong link with social origin, the job of fathers counts more than the one of mothers, people from Polytechnique come from a well-off background better than ENAC or Centrale Lyon.
- Second Model : Problem of the dummy variables, the impact of jobs vary depending on the school. People from Polytechnique comes from large town and income has a better influence. People from ENAC comes from a rural environnement and income has a positive impact conversely to Centrale Lyon.