# Why do girls enter engineering schools?

Jerome Duguet Maryne Floch–Le Goff

May 24, 2013

## 1 Introduction

There are few girls studying in engineering school, therefore we found it interesting to study why these girls have decided to enter these schools. The question we wondered is which parameters appeal girls to French engineering schools and how much are they interested in their field of study?

## 2 Our model

### 2.1 Data description

Our data was collecting thanks to a poll we sent to seven French engineering schools [1]. We have collected 230 available observations in 10 days. Our dependent variable is the interest of girls in their field of study, on a scale from 0 (not interested at all) to 5(passionate). To find our explanatory variables, we wondered which variables could influence the interest of girls in their field of study. However our data is rather biased: 1/3 come from ENAC and 1/3 from Chimie Montpellier. The data [2] includes information relative to girls themselves, such as nationality, age, race, family (parents position, if they are separated or together, the number of brothers and sisters). It also contains information relative to girls and their school: field of study, target job, when she decided to do the job, the number of students in the promotion, the percentage of girls, the way of knowing their school, if she has a friend in her school, distance from school to home. Finally it contains some variables explaining why she has entered the school: by challenge, by ambition, by passion, for the openings, by default or to fight against stereotypes. We drop a dummy variable which was military: is the school a military school or not? Only 2 answers out of 230 were yes so it was not concluant enough. Furthermore French military schools are in minority, so this factor could not bring anything relevant to our model. We have a lot of dummy variables: nationality, race, parents job (CEO or professor), marital situation of parents, field of study (girls field: chemistry, agro or men field: mechanics, buildings), target job(known or not yet), the way of knowing the school (publicity of the school of advice from friends), friend in the school, challenge, ambition, passion, openings, default and stereotypes. Actually, we only have the variables distance, age, number of students, percentage of girls and brothers and sisters which are not dummy; consequently only these variables will have a marginal effect.

[1] ENAC Toulouse, Chimie Montpellier, Agro Montpellier, Ecole de leau Strasbourg, Centrale Marseille, and two other schools no identified.

[2] Please find some details in appendix

### 2.2 Summary statistics

Thanks to Google Doc we obtained data translated in percentage. Here are some of the most relevant data: - 65 per cent have known their schools thanks to the publicity of the school, 23 per cent by friend; relatives seems to be important for girls for taking their decision. - 26 per cent do not know yet the job their want to do; it does not seem to be as much linked as we thought with the interest of girls in their field of study. - 78 per cent of people who answered have married or en concubinage parents; it may be a factor in favor of entering an engineering schools. - 44 per cent have a father CEO or manager, 8 per cent professor; the other half per cent could represent a lot of different positions, however it seemed more important for us to know if their parents are well-positioned, e.g. being CEO or professor. - 59 per cent have a mother CEO or manager, 15 per cent professor. - For 29 per cent, they entered the school for the openings, for 20 per cent it is a passion and for 19 per cent it was the best school they could have after their "classe prepa"; as we would expect, passion and ambitious seem to have a real importance for girls in terms of interest.

| | Coeff | Std. Error | t-stat | Prob |
|---|---|---|---|---|
| C | 2.567 | 0.779 | 3.296 | 0.001 |
| French | 0.397 | 0.338 | 1.174 | 0.241 |
| Age | 0.047 | 0.026 | 1.769 | 0.079 |
| Race | -0.109 | 0.206 | -0.528 | 0.597 |
| Field | -0.0002 | 0.16 | -0.001 | 0.999 |
| Job targeted | -0.094 | 0.154 | -0.609 | 0.543 |
| Step | -0.017 | 0.115 | -0.145 | 0.885 |
| Promotion | -0.0007 | 0.0002 | -2.718 | 0.007 |
| Perc of Girls | -0.002 | 0.004 | -0.589 | 0.557 |
| Distance | 0.0003 | 0.0002 | 1.978 | 0.049 |
| Knowledge | 0.293 | 0.129 | 2.268 | 0.024 |
| Friend | 0.024 | 0.109 | 0.223 | 0.823 |
| CEO Mother | 0.133 | 0.0.138 | 0.970 | 0.333 |
| Prof Mother | -0.116 | 0.161 | -0.719 | 0.473 |
| CEO Father | -0.085 | 0.127 | -0.667 | 0.505 |
| Prof Father | -0.237 | 0.221 | -1.073 | 0.285 |
| Together | 0.003 | 0.141 | 0.023 | 0.982 |
| Brother | -0.210 | 0.066 | -3.176 | 0.0017 |
| Challenge | 0.230 | 0.185 | 1.250 | 0.213 |
| Ambition | 0.064 | 0.119 | 0.533 | 0.595 |
| Openings | 0.028 | 0.113 | 0.249 | 0.803 |
| Passion | 0.596 | 0.118 | 5.058 | 0.000 |
| Default | -0.234 | 0.116 | -2.026 | 0.044 |
| Stereotypes | 0.0008 | 0.191 | 0.004 | 0.997 |

Table 1: First regression results.

## 2.3 First model

Our first equation[3] is therefore:

$$INTEREST_i = Xi, j'\beta + U$$

$$i = 1, 2, \ldots, n, j = 1, 2, \ldots, k$$

Where n=230 is the total number of observations, and k=23 is the total number of criterium

In table 1, please find our first model, made with every variable.

Note that n=230 and k=23. The estimated marginal effects of Age, Promotion, Distance, Knowledge, Brother, Passion and Default are all significant at 90per-cent level. They are positive for Age, Distance, Knowledge, Passion (and for the constant, which is less interesting). $\beta$ are negative for Promotion, Brother and Default. As we would expect, choosing the school by default decrease the interest of girls, such as having a bigger number of students in the promotion or having a lot of brothers and sisters.But, surprisingly, the distance from school to home have a positive effect on the interest, which means that girls are more likely to be passionate if they live far: we can explain that by more motivation. The magnitude of significant coefficients is Passion (0.596) $\geq$ Default (0.234) $\geq$ Knowledge (0.293)$\geq$ Age(0.047) $\geq$ Promotion (0.0007) $\geq$ Distance (0.0003). Even if the other parameters are not significant at usual levels, we could say a word about them, because perhaps we are not going to drop them all. We can say that having a friend in the school, having a mother who is CEO, choosing this job by challenge, for the openings and to fight against stereotypes have a positive effect on the interest that girls have. However, studying in a women field have a negative effect on the interest; maybe the fact that we have a lot of girls who answered who came from Chimie Montpellier could explain that result, so that variable may not be relevant. Finally, it is very astonishing that having a mother or a father who is professor, or a father who is CEO have a negative effect on interest. We can explain that saying that pupils coming from more modest middles are more motivated to get a better job.

[3]Please find the complete equation in Appendix

## 2.4 Test of our model

We find an adjusted R-squared equal to 21 per cent, which is quite weak[4]. Therefore, our idea is to drop useless indicators. Using the correlation matrix, we find that field and job, field and girl and job and girl are highly correlated. [5]. Looking to the p-values, we drop the variables Field and Girls which do not seem to be significant. But we still may have too many explanatory variables, thats why, looking at the variables with the biggest p-value, we make a joint test (Wald test) testing if $\beta_6 = \beta_{11} = \beta_{16} = \beta_{20} = \beta_{23} = 0$. These estimated marginal coefficients correspond to: step, friend, together, openings and stereotypes. The results are such that we drop them. We can see that some variables still have p-value superior to 10 percent but they are not so high, thats why we can keep them.

## 2.5 New model

After dropping these variables, thanks to EViews we find a new equation [6].

In table 2, please find our new model, made with every saved variable:

We find the same trends as for the first equation, so it is reassuring. In addition the adjusted R-

[4]Actually we would need to compare this value to another model; furthermore we know that a "good" R-squared depends on the context.

[5] Correlation values : field/job : 0.54 ; field/girl : 0.70 ; job/girl : 0.57

[6]Please find the complete equation in Appendix

| | Coeff | Std. Error | t-stat | Prob |
|---|---|---|---|---|
| C | 2.491 | 0.729 | 3.41 | 0.0008 |
| French | 0.385 | 0.329 | 1.17 | 0.243 |
| Age | 0.0473 | 0.0257 | 1.84 | 0.0672 |
| Race | -0.109 | 0.198 | -0.551 | 0.582 |
| Job targeted | -0.0369 | 0.122 | -0.300 | 0.764 |
| Promotion | -0.0006 | 0.00024 | -2.75 | 0.0064 |
| Distance | -0.003 | 0.00015 | 2.03 | 0.0431 |
| Knowledge | 0.295 | 0.123 | 2.38 | 0.0180 |
| CEO Mother | 0.139 | 0.132 | 1.05 | 0.295 |
| Prof Mother | -0.119 | 0.155 | -0.771 | 0.441 |
| CEO Father | -0.0723 | 0.122 | -0.591 | 0.554 |
| Prof Father | -0.224 | 0.213 | -1.05 | 0.292 |
| Brother | -0.212 | 0.0641 | -3.31 | 0.0011 |
| Challenge | 0.247 | 0.167 | 1.47 | 0.1410 |
| Ambition | 0.0629 | 0.114 | 0.551 | 0.5822 |
| Passion | 0.0578 | 0.109 | 5.28 | 0.000 |
| Default | -0.228 | 0.111 | -2.04 | 0.0419 |

Table 2: First regression results.

square is equal to 23 percent, which is better than the previous model; our new model is more precise. However, we still have variables which are insignificant at usual levels: French, White, Job, CM, PM, CF, PF, Challenge, Ambition. We could have tested them jointly and see if their estimated marginal effect are null but we decide here to keep them because their p-value is not so high. [7].

# 3 Validation of the model

## 3.1 Linearity and non multicolienarity

It is obvious that our model respect the linearity and the non multicolinearity, as we assume a multilinear regression and our k (number of variables) is inferior to our sample (230). So these points are clearly verified.

## 3.2 Homoscedasticity

Then, we can easily see that there is no heteroscedasticity with a White Test. See table 3.

## 3.3 Normality of errors

To test the normality, we have first used a Jarque Bera (see figure 1) which shows us that the model is

[7]Please find the complete numerical equation in Appendix

| | | | |
|---|---|---|---|
| F-statistic | 0.830770 | Prob.F(127,100) | 0.8386 |
| Obs*R-squared | 117.0553 | Prob.Chi-square(127) | 0.7254 |
| Scaled explained | 101.3967 | Prob.Chi-Square(127) | 0.9542 |

Table 3: Results of the test of homoscedasticity.

exogeneous[8], but there are some extreme values which disturb us. Here, the hypothesis of normality is reject at 95 percent of significance. Probably if we remove them we would see that the residuals are normal.

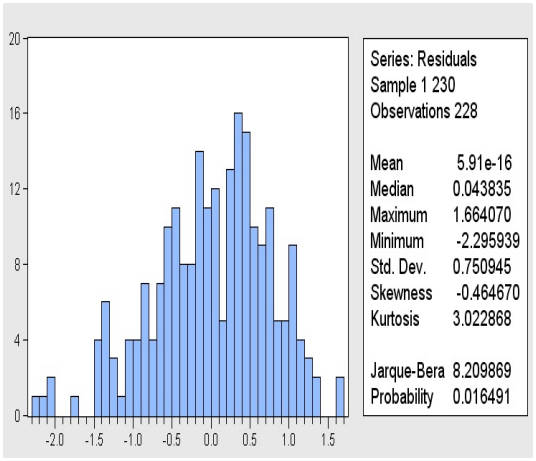


Figure 1: Jarque Bera Test.

We also have built the quantile to quantile plot, and we can say that we have the same phenomenon, if we remove the extreme values, our model would be correct because the residuals would be normal. See the figure 2.

# 4 Concluding comments

Our paper has examined the interest of girls to enter engineering schools. Our model is not very strong (adjusted $R^2 = 0.23$). This is certainly due to the fact that we have in majority dummy variables, not enough observations and extreme values disturbing. Moreover we have some variables, due to our low number of observations, that are not relevant: the age has an importance on interest according to our model, but all

[8]E(u)=0

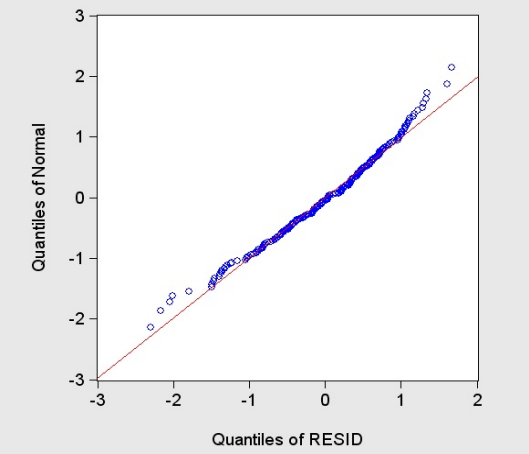


Figure 2: Quantile quantile plot.

our values were between 20 and 25 years old so it does not mean a lot of things. We would have create a new variable AGESQUARED which would measure more precisely the influence of that factor. It is the same for the parameter French: most of girls who answered were French; and idem for the parameter white. We can not conclude on a so small sample. The variables of parents job also seem to play a role in the interest ($\beta$ quite high) but their p-value are also higher than what we should have so maybe we should have drop these variables. It could also have been interesting to integrate some other variables, such as financial ones for example. We could also have dropped the insignificant variables found in our second model (French, White, Job, CM, PM, CF, PF, Challenge, Ambition) but it seemed to us that our model would have been weak with so few variables remaining. However, it would certainly have improved our model.

What we can say to conclude our study about our results is that girls seem to be interested in their subject, and so come to engineering schools mainly because they are passionate, for the challenge of entering a men world, because they heard from the school from a friend or someone from their belongings. But the more they have brothers and sisters for example, the less they are interested.

# 5 Apendix

## 5.1 Details about the model

We have selected several variables which seem to us relevant to describe which parameters appeal girls to enter engineering schools.

- Nationality: 1 if French, 0 if everything else.
- Age: has a range of 20 to 25years. This little difference of age between the observations is due that we have sent our poll to girls currently in engineering schools. In fact, it was difficult to find a way to diffuse our poll to current engineer.
- Race: 1 if white, 0 if else.
- Field of study: 1 if it is a "women" field ie chemistry, agronomy; 0 if it is a "men" field: mechanics, buildings. With this dummy we want to check if the stereotypes are verified: does girls prefer "girls" field?
- Job targeted: 1 if the girls knows what job she wants, 0 if not. We answered on our poll the question "which job do you want to do?" and we had a lot of different answers which were difficult to classify. 20per-cent wanted to be searchers, 4per-cent Chief Executive Officers, 24per-cent manager, 26per-cent still do not know and 19per-cent of "other" answers. That's why, we 23per-cent of "I don't know", we decided to make it our dummy variable.
- Step: it is the step of her life when she was when she decided to do the job: 1 if it was before classe prepa (a choice thought); 0 if after. If she does not know, and so did not answer to the question, we consider that it is 0.
- Promotion: number of girls in the promotion. It has a range of 100 to 1500 students, with a median of 300students, quite common in engineering schools. The biggest values (around 1500) corresponds to people who have answered the number of students in their entire school.
- Girls: the percentage of girls in the promotion. As we have targeted several different schools, we have approwimately: 1/3 of [0;25per-cent]; 1/3 of [25;50per-cent] and 1/3 of [50;75per-cent]. We can make the hypothesis that it would not be very relevant.

- Distance from home to school: It has a range of 0km (from the town when her school is) to 10 000 (a ceiling we have fixed for far destinations: New Caledonia, Tahiti, Madagascar etc).
- Knowledge: represents how she had knew the school. We put 1 if it was from advice from friends or family, 0 if it was from the school publicity (internet, forums, etc).
- Friend: 1 if she has a friend in the school before entering it; 0 if else.
- Mother CEO (or manager): 1 if yes; 0 else. we have put this variable to see if the fact that having a well-payed mother with a high responsibilities job is relevant.
- Mother professor: 1 if yes, 0 if else.
- Father CEO: 1 if yes; 0 else.
- Father professor: 1 if yes; 0 else.
- Together: it is a dummy variable which quantifies the effect of parents married or "en concubinage". 1 if yes; 0 if no.
- Brothers: This variables quantifies the number of brothers and sisters of the girl (has a range from 0 to 3).
- Challenge: this variable corresponds to the question: has the girl entered her school to prove that she can enter a "men" world? 1 if yes, 0 if no.
- Ambition: does the girl want to be well-payed? 1 if yes.
- Openings: does the girl enter the school for the professional openings? 1 if yes.
- Passion: does the girl enter the school because she is passionate by the subject? 1 if yes.
- Default: it is a variable that quantifies the impact on the interest of having chosen the school because it was the best school she could have after her classe prepa.
- Stereotypes: has the girl enter the school to fight against stereotypes, to prove that it is not a "men" world?

### 5.2 First regression equation

$$\begin{aligned} INTEREST_i &= \beta_0 + \beta_1 FRENCH_i + \beta_2 AGE_i \\ &+\beta_3 WHITE_i + \beta_4 FIELD_i + \beta_5 JOB_i \\ &+\beta_6 STEP_i + \beta_7 PROMOTION_i + \beta_8 GIRLS_i \\ &+\beta_9 DISTANCE_i + \beta_{10} KNOWLEDGE_i + \beta_{11} FRIEND_i \\ &+\beta_{12} CEOM_i + \beta_{13} PROFESSORM_i + \beta_{14} CEOF_i \\ &+\beta_{15} PROFESSORF_i + \beta_{16} TOGETHER_i + \beta_{17} BROTHERS_i \\ &+\beta_{18} CHALLENGE_i + \beta_{19} AMBITION_i + \beta_{20} OPENINGS_i \\ &+\beta_{21} PASSION_i + \beta_{22} DEFAULT_i + \beta_{23} STEREOTYPES_i \\ &+u_i,\ i = 1, 2, \ldots, 230 \end{aligned}$$

### 5.3 Second regression equation

$$\begin{aligned} INTEREST_i &= \beta_0 + \beta_1 FRENCH_i + \beta_2 AGE_i \\ &+\beta_3 WHITE_i + \beta_4 JOB_i + \beta_5 PROMOTION_i \\ &+\beta_6 DISTANCE_i + \beta_7 KNOWLEDGE_i + \beta_8 CEOM_i \\ &+\beta_9 PROFESSORM_i + \beta_{10} CEOF_i + \beta_{11} PROFESSORF_i \\ &+\beta_{12} BROTHERS_i + \beta_{13} CHALLENGE_i + \beta_{14} AMBITION_i \\ &+\beta_{15} PASSION_i + \beta_{16} DEFAULT_i + u_i, \\ &i = 1, 2, \ldots, 230 \end{aligned}$$

Numerical equation:

$$\begin{aligned} INTEREST_i &= 2.49 + 0.385 FRENCH_i + 0.047 AGE_i \\ &-0.109 WHITE_i - 0.0369 JOB_i - 0.000685 PROMOTION_i \\ &-0.000310 DISTANCE_i + 0.295 KNOWLEDGE_i \\ &+0.139 CEOM_i - 0.119 PROFESSORM_i - 0.0723 CEOF_i \\ &-0.224 PROFESSORF_i - 0.212 BROTHERS_i + \\ &0.247 CHALLENGE_i + 0.063 AMBITION_i + 0.578 PASSION_i \\ &-0.228 DEFAULT_i + u_i,\ i = 1, 2, \ldots, 230 \end{aligned}$$