Daria Yermolova
Econometrics 2
Final
a.

Since the treatment (vouchers) was given to families whose income was below 50 (Andalisian curency) per month (a cut-off was used), this is a clear case of regression discontinuity. Thus, this is RD regression model. In this case we need to look at the data right around the cut-off and compare families with income aroung 50 (just above and just below) who got and didn't get a voucher, see the results of their kids in exams and check for significant difference.

$$y_i = \alpha + \beta X_i + \sigma * voucher_i + \phi * familyincome_i + \epsilon_i$$

Where y_i is the studying results of children in a respective family. $X_i$is a set of control variables, $voucher_i$ is a dummy variable that is equal to 1 if family recieved a voucher (treatment) and 0 if it did not. The coefficient σ shows the impact of the treatment on studying results. $f(familyincome_i)$ is a function of representative family's income.

b.

In case the researches is interested in one particular candidate/party or if there is a two-party system like in the US or UK, one can use a probit model, estimating the probility a treated and untreated family would vote for a particular candidate/party. We still need to look at families with income close to 50, since it is still RD. Theoretically, we should expect a positive effect of the program on the voting behaviour (if the candidate/party are the ones who were in power when the program happened), since people can feel support from the previous government, but that is just an assumption. Ideally, we should include data of previous voting and political beliefs (compliant with the candidate/party or not).

$$vote_i = \alpha + \beta' X_i + \sigma * voucher_i + f(familyincome_i) + \epsilon_i$$

c.

In this case we could try to use differences in differences approach, comparing the development of voting intentions among treated and untreated families of approximately same income in the two periods. Or we could still stick to the model in b if there is enough data. However, voting behaviour tends to be quite sticky and I wouldn't expect a significant effect.

d.

I would definitely say we need more data to estimate long-run outcomes. A lot of kids might still be in their early years of school, when computers are not of that much use yet, others need to get used to using them. Additionally, a lot of children start to pay more attention to studying right before their last year to get into a good college.

e.

The data on 2007 can be of great use, since we can include the overall time trend in studying of each particular child. (Maybe some of them have been trying to pay more attention to studying in the past year already and continued to do so when they got the computer or the opposite.)

f-g.

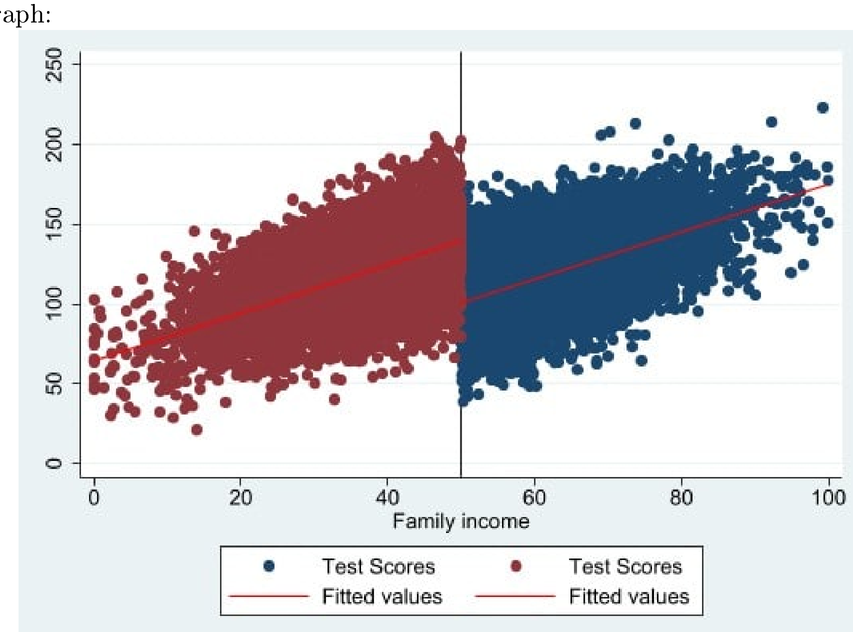
The regression without the measurement error gives the following result:

```
. reg y D z

      Source |       SS           df       MS      Number of obs   =    19,987
-------------+----------------------------------   F(2, 19984)     =   4678.44
       Model |  3761086.93         2  1880543.46   Prob > F        =    0.0000
    Residual |  8032755.64    19,984   401.95935   R-squared       =    0.3189
-------------+----------------------------------   Adj R-squared   =    0.3188
       Total |  11793842.6    19,986  590.105202   Root MSE        =    20.049

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           D |    39.7377   .4704904    84.46   0.000     38.8155     40.6599
           z |   1.504221   .0156995    95.81   0.000    1.473448    1.534993
       _cons |   24.98318   .9909948    25.21   0.000    23.04075    26.92561
------------------------------------------------------------------------------
```

The coefficient on treatment is significant at 99% confidence. On average the treatment increases the child's scrore by 39.74. We see a similar result on a graph:



h. The regression with the measurement error gives the following result:

```
. reg y D1 z1

      Source |       SS           df       MS      Number of obs   =    19,987
-------------+----------------------------------   F(2, 19984)     =   2972.84
       Model |   2704331.5         2  1352165.75   Prob > F        =    0.0000
    Residual |  9089511.08    19,984  454.839425   R-squared       =    0.2293
-------------+----------------------------------   Adj R-squared   =    0.2292
       Total |  11793842.6    19,986  590.105202   Root MSE        =    21.327

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          D1 |   31.62343   .5011446    63.10   0.000     30.64115    32.60572
          z1 |   1.284502   .0166693    77.06   0.000     1.251829    1.317175
       _cons |    39.2832   1.060614    37.04   0.000     37.20431    41.36209
------------------------------------------------------------------------------
```

The coefficient on treatment is still significant, but it is now much smaller. On average the treatment increases the child's scrore by 31.62. The effect of children from wealthier families being more successful in school is interfering with the impact of treatment and the result we get is an underestimation of the policy effect.

i. Since the measurement error is not classical (not purely random, non-zero mean) it is affecting our judgement and (in this case) decreasing the perceived effect of the treatment (attenuation bias).

j. If the mistakes end up to classical measurement error (are purely random, approximately same amount of people getting the laptop by mistake as not getting a laptop by mistake, and we have the correct data on who actually got it (D values are correct), we can still estimate the effect of the policy. However, as we saw above, if it is not the case, such a mistake can lead to drastic bias in the results.