Econometrics - Homework Assignment

"Maïté Lamothe - Florentine Oliveira - Tom Verrier"

10/11/2019

Contents

1] Description of the sample

Question 1

- interprération écart type EDUC
- comparaison sal et salbegin : gros écart d'écart type, différence éducation, salaire d'efficience (mean)
- interprétations sur gender et minotity: sample composed of more women ou jsp + regarder interprétation dummies gender et tout : much diversity
- ajouter tableau stats jobcat + commenter

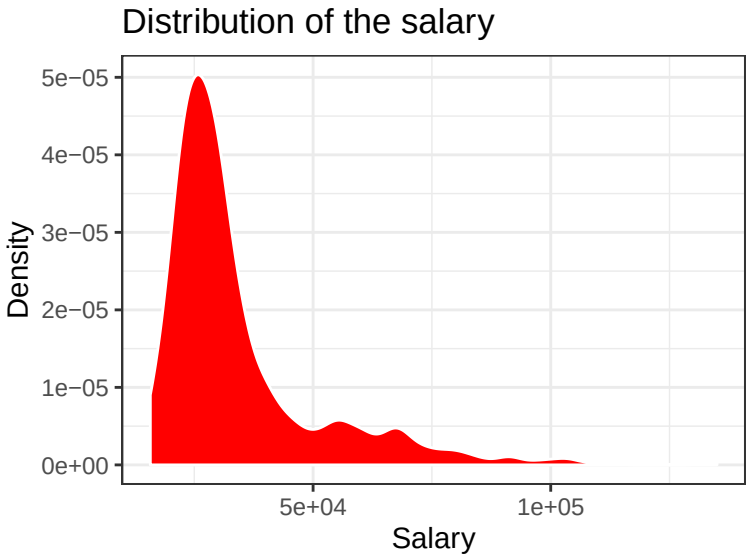
Table 1: Some statistics

variable	Mean	Sd	Min	Max
EDUC	13.49156	2.884846	8	21
SALARY	34419.56751	17075.661465	15750	135000
SALBEGIN	17016.08650	7870.638155	9000	79980

Table 2: Some other statistics

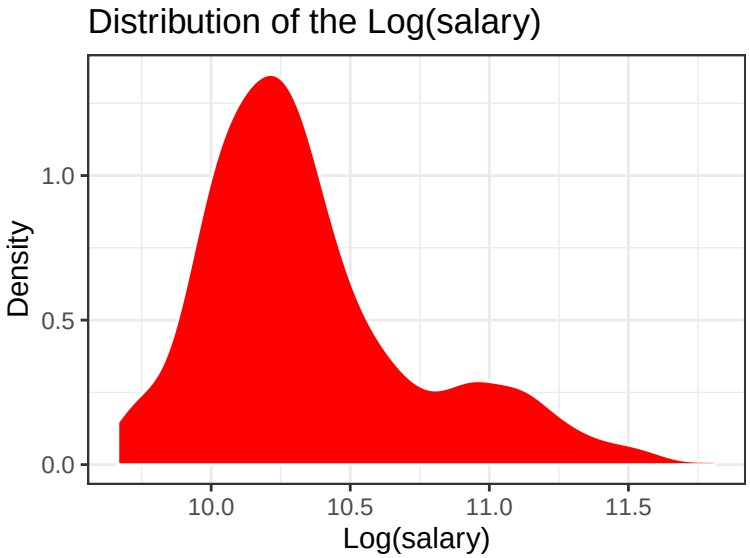
variable	Mean
GENDER	0.5443038
MINORITY	0.2194093

Question 2



It could be a good idea to use the logarithm of the variable Salary because the logarithm linearizes and smoothes the variable. In facts it decreases the expanse of the values that the variable takes (max-min is lower). Moreover, the interest to use the logarithm of this variable is that we can easily interpret the coefficient as an elasticity in a log-log model.

Question 3



We see that the variable LogSal is much more readable. The distribution is less extensive.

2] Linear Regression

Question 1

a)

We estimate the model: $LogSal = \alpha + \beta Education + \epsilon$ (**R1**)

Table 3:

	Dependent variable:
	LOGSAL
EDUC	0.096^{***}
	(0.005)
Constant	9.062^{***}
	(0.063)
Observations	474
R^2	0.485
Adjusted R^2	0.484
Residual Std. Error	0.285 (df = 472)
F Statistic	445.300^{***} (df = 1; 472)
Note:	*p<0.1; **p<0.05; ***p<0.01

We can see on the *Table 2* that the variable Education is statistically significant at the threshold of 1%, as well as the intercept.

b)

This is a log-level linear model. Then, an increase in one year of education lead to an increase of $100.\hat{\beta} = 100 * 0.096 = 9,6$ in the Salary.

c)

Question 2

We now estimate the model: $LogSal = \alpha + \beta_1 Education + \beta_2 LogSalBegin + \epsilon$ (**R2**). The results are shown in *Table 4*.

a)

The impact of education on LogSal is different from the first model **R1** because we have added an explanatory variable in the model.
Mathematically, the matrix X of the explanatory variable is now different. Therefore the vector of estimated coefficients, which is equal to $(X'X)^{-1}X'Y$, differs. There was an omitted variable bias.

b)

Theoretically: total effect is the effect shown with model R1 direct effect is the one of model R2 the indirect effect is the one captures by the regression : logsalbegin on a constant and Education We are supposed to

Table 4:

	Dependent variable:
	LOGSAL
EDUC	0.023***
	(0.004)
LOGSALBEGIN	0.869***
	(0.032)
Constant	1.647***
	(0.275)
Observations	474
R^2	0.801
Adjusted R^2	0.800
Residual Std. Error	0.178 (df = 471)
F Statistic	945.421*** (df = 2; 471)
Note:	*p<0.1; **p<0.05; ***p<0.01

find that R1=R2+

c)

We regress *LogSalBegin* on a constant and *Education*. Results are shown in *Table 5*.

Table 5:

	Dependent variable:
	LOGSALBEGIN
EDUC	0.084***
	(0.004)
Constant	8.538***
	(0.057)
Observations	474
R^2	0.470
Adjusted R^2	0.469
Residual Std. Error	0.257 (df = 472)
F Statistic	418.920*** (df = 1; 472)
Note:	*p<0.1; **p<0.05; ***p<0.01

d)

Question 3

Results of the regression of the model $\overline{LogSal} = \beta_1 \overline{education} + \beta_2 \overline{LogSalBegin} + \epsilon$ are shown in *Table 5*.

Results of the regresion of the model $DMLogSal = \beta_1 DMeducation + \beta_2 DMLogSalBegin + \epsilon$ are shown in *Table 6*.

Table 6:

	Dependent variable:
	resid_LOGSAL
resid_EDUC	0.023***
	(0.004)
resid_LOGSALBEGIN	0.869***
	(0.032)
Observations	474
R^2	0.801
Adjusted R^2	0.800
Residual Std. Error	0.178 (df = 472)
F Statistic	947.428*** (df = 2; 472)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7:

	Dependent variable:
	DMLogSal
DMeducation	0.023***
	(0.004)
DMLogSalBegin	0.869***
	(0.032)
Observations	474
R^2	0.801
Adjusted R^2	0.800
Residual Std. Error	0.178 (df = 472)
F Statistic	947.428*** (df = 2; 472)
Note:	*p<0.1; **p<0.05; ***p<0.01

a)

As we can see in *Table 5* and *Table 6*, the estimated coefficients are the same. Proove that residuals is the same as demean.

b)

Proove that demin without constant ad demean with constant leads to the same estimates.