INTRODUCTORY ECONOMETRICS
Dpt of Econometrics & Statistics (EA3)
University of the Basque Country UPV/EHU
OCW Self Evaluation
Time: 21/2 hours

SURNAME:____
NAME:__________________________ **ID#:**__________________________

Specific competences to be evaluated in this exercise:

1. To analyse critically the basic elements of Econometrics in order to understand the logic of econometric modelling and be able to specify causal relationships among economic variables.
2. To identify the relevant statistical sources in order to be able to search for, organise and systematically arrange available economic data.
3. To use with confidence appropriate statistical methods and available computing tools in order to correctly estimate and validate econometric models.
4. To handle econometric prediction tools in order to estimate unknown or future values of an economic variable.
5. To interpret adequately the results obtained in order to be able to write meaningful reports about the behaviour of economic data.

EXERCISE

In order to carry out a study on employees' wages, a company collects information from its 500 employees[1] as follows:

W_i: **present Wage** of the *i*-th employee (in thousands of euros),

In_i: **Initial wage** of the *i*-th employee (in thousands of euros),

T_i: **service Time** in the company of the *i*-th employee (in months),

E_i: **previous Experience** of the *i*-th employee (in months),

i	W_i	In_i	T_i	E_i	i	W_i	In_i	T_i	E_i
1	46.24	31.09	35	71	...				
2	127.28	53.19	233	139	...				
3	36.91	30.5	29	47	486	57.13	33.57	99	69
4	69.85	38.37	78	98	487	62.11	33.37	94	71
5	33.86	30.5	16	53	488	58.57	30.5	113	59
6	36.43	30.5	26	27	489	50.64	32.38	77	61
7	90.69	42.46	167	98	490	38.37	30.5	26	56
8	34.97	30.5	21	41	491	20.5	20.5	25	0
9	52.51	31.79	78	52	492	71.16	44	45	103
10	84.41	35.84	196	74	493	35.73	30.5	10	27
11	48.28	31.72	27	77	494	37.92	30.5	28	24
12	61.38	35.81	103	89	495	56.56	30.5	96	60
13	39.62	30.5	21	49	496	24.53	20.5	21	36
14	79.05	46.72	38	133	497	38.84	30.5	15	47
15	100.69	59.37	42	180	498	38.52	30.5	27	36
...					499	64.33	41.76	28	110
...					500	94.69	41.94	214	101

Please do not write under this line.

[1]Source: own computer generated data.

FIRST PART

The expert hired to do the study begins by specifying a SLRM for the **present Wage** in terms of the **Initial wage**:

$$W_i = \beta_0 + \beta_1 In_i + u_i, \qquad i = 1, \dots, N \tag{1}$$

Using the following sample information,

$$\sum_{i=1}^{500} W_i = 27771.40 \qquad \sum_{i=1}^{500} (W_i - \overline{W})^2 = 234530.53$$

$$\sum_{i=1}^{500} In_i = 17012.09 \qquad \sum_{i=1}^{500} (In_i - \overline{In})^2 = 21200.65$$

$$\sum_{i=1}^{500} W_i In_i = 1005208.51 \qquad \sum_{i=1}^{500} (W_i - \overline{W})(In_i - \overline{In}) = 60309.40$$

1. (1 point) Write down the **sample regression function** for the proposed model.
   - A. $\widehat{W}_i = \beta_0 + \beta_1 In_i$
   - B. $\widehat{W}_i = -41.25 + 2.84 In_i$
   - C. $SRF = 73.1\%$
   - D. $\widehat{W}_i = -21.84 + 4.25 In_i$
   - E. $E(Y_i) = \beta_0 + \beta_1 X_i$
   - F. $\overline{W}_i = \widehat{\beta}_0 + \widehat{\beta}_1 \overline{In}_i$

   1. ________________

2. (1 point) Give an interpretation for the **estimated** coefficient of the **Initial wage** (In).
   - A. It is equal to the estimated coefficient of the dependent variable *ceteris paribus*.
   - B. It is the estimated proportion of dependent variable explained by the regression *ceteris paribus*.
   - C. The present Wage increases 4.25 euros when the Initial wage increases 1 euro.
   - D. The present Wage increases 2840 euros when the Initial wage increases 1000 euros.
   - E. The present Wage increases 2.84% when the Initial wage increases 1% *ceteris paribus*.
   - F. It is the estimated increase when the regression increases one measurement unit *ceteris paribus*.

   2. ________________

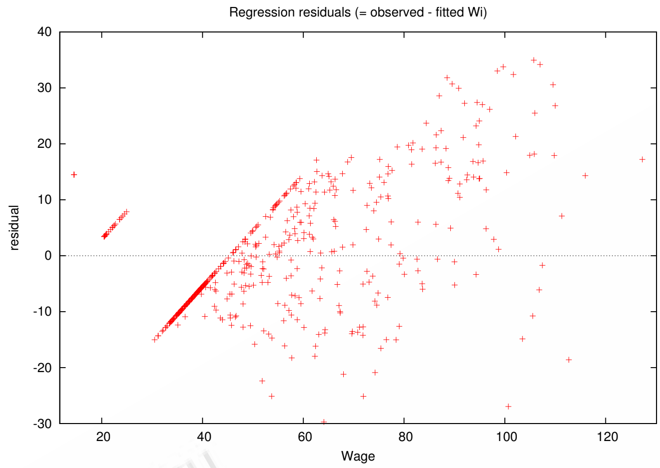
3. (a) (1 point) What assumptions **about the error term** are necessary for the OLS estimator to be the **BLU Estimator**?

   ____
   ____
   ____
   ____

   (b) (1 point) and for **hypothesis testing**?
   - A. The same.
   - B. None of them.
   - C. We need assumptions about disturbances not about errors.
   - D. Hypothesis testing needs an accurate model therefore there can be no errors in our regression.
   - E. We also need the errors to be normal.
   - F. All of the above.

   (b) ________________

4. (2 points) The following graph shows the residuals for Model 1. Comment on the validity of the assumptions for OLS to be BLUE.



5. (1 point) Write down the sample regression function if the W and In variables were measured in **thousands of dollars** (assume an exchange rate of \$1.3 per euro).

A. $\widehat{W}_i = 1.3\beta_0 + 1.3\beta_1 In_i$

B. $\overline{W}_i = \widehat{\beta}_0 + 1.3\widehat{\beta}_1\overline{In}_i$

C. $E(Y_i) = 1.3\beta_0 + \beta_1 X_i$

D. $\widehat{W}_i = -21.84 + 5.23 In_i$

E. $\widehat{W}_i = -53.62 + 2.84 In_i$

F. $SRF = 1.3 \times 73.1\%$

5. ____________

Please do not write under this line.

SECOND PART

The expert decides to make use of all the available information, that is, he now also includes the **service Time** and **previous Experience** explanatory variables into the model. The following results are thus obtained:

Model 2: OLS estimates using the 500 observations 1–500
Dependent variable: W

	Coefficient	Std. Error	t-ratio	p-value
const	−12.326	1.06981	−11.5223	0.0000
In	1.29671	0.0491288	26.3940	0.0000
T	0.202709	0.00287339	70.5468	0.0000
E	0.134346	0.0106343	12.6334	0.0000

Sum of squared residuals	4760.34
Standard error of the regression ($\widehat{\sigma}$)	3.09798
Unadjusted R^2	0.979703
$F(3,496)$	7980.25

6. (1 point) Write down the **proposed GLRM**.
   A. $E(Y_t) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}$
   B. $Y_t = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$
   C. $W_i = \gamma_0 + \gamma_1 In_i + \gamma_2 T_i + \gamma_3 E_i$
   D. $W_i = \gamma_0 + \gamma_1 In_i + \gamma_2 T_i + \gamma_3 E_i + u_i$
   E. $\widehat{W}_i = -12.3267 + 1.29671 In_i + 0.202709 T_i + 0.134346 E_i$
   F. $\overline{W}_i = -12.3267 + 1.29671 \overline{In}_i + 0.202709 \overline{T}_i + 0.134346 \overline{E}_i + \overline{u}_i$

6. ________________

7. (1 point) Write down the **sample regression function**.
   A. $E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}$
   B. $W_i = \beta_0 + \beta_1 In_i + \beta_2 T_i + \beta_3 E_i$
   C. $\widehat{W}_i = -11.5223 + 26.3940 In_i + 70.5468 T_i + 12.6334 E_i$
   D. $\widehat{W}_i = -12.3267 + 1.29671 In_i + 0.202709 T_i + 0.134346 E_i$
   E. $\overline{W}_i = -12.3267 + 1.29671 \overline{In}_i + 0.202709 \overline{T}_i + 0.134346 \overline{E}_i + \overline{u}_i$
   F. $W_i = \gamma_0 + \gamma_1 In_i + \gamma_2 T_i + \gamma_3 E_i + u_i$

7. ________________

8. (1 point) Write down the **first two values** of vector $X'Y$.
   A. 27721 & 1005208
   B. 234530 & 21200
   C. -11.5223 & 26.3940
   D. -12.3267 & 1.29671
   E. $X_1'Y$ & $X_2'Y$
   F. $\sum W In_t$ & $\sum W T_t$

8. ________________

9. (a) (1 point) Write down the expression used to calculate the **goodness-of-fit** of the estimated model.
   A. $R^2 = 1 - RSS/ESS$
   B. $R^2 = TSS/ESS$
   C. $R^2 = \sum \widehat{W}_i^2 - T\overline{W}^2 / \sum W_i^2 - T\overline{W}^2$
   D. $R^2 = (\sum \overline{W}_i^2 - T\widehat{W}^2)/(\sum W_i^2 - T\widehat{W}^2)$
   E. $R^2 = (\sum \widehat{W}_i^2 - T\overline{W}^2)/(\sum W_i^2 - T\overline{W}^2)$
   F. None of the above.

(a) ________________

(b) (1 point) Interpret the result.

A. Proportion of regression explained by the variables.
B. Expected increase of dependent variable when the explanatory variable increases one measurement unit *ceteris paribus*.
C. Proportion of TSS explained by the regression.
D. Proportion of dependent variable variance explained by the regression.
E. Proportion of ESS explained by the RSS/TSS.
F. It is positive when the fit is good and viceversa.

(b) ________________

10. (1 point) According to Model 2, what is the **estimated present Wage** for the first employee?

A. 46.24 euros.
B. 44.62 euros.
C. 44620 euros.
D. -12.326 euros.
E. $\beta_0 + \beta_1 In_i$
F. $\beta_0 + \beta_1 In_i + \beta_2 T_i + \beta_3 E_i$

10. ________________

11. (1 point) Angela A. has a **service Time** in the company of 60 months **more** than Brandon B., however, they both started with the same **previous Experience** and **Initial wage**. What is the **estimated difference** between the **present Wages** of Alice and Brandon?

A. $\beta_0 + \beta_1 In_i + \beta_2 \times 60 + \beta_3 E_i$
B. Since Angela is a woman she will probably earn less than Brandon.
C. 12.16 euros *ceteris paribus*.
D. $60 \times \beta_2$
E. 12160 euros.
F. $W_{\text{Angela}} - W_{\text{Brandon}}$

11. ________________

12. (2 points) Test the **individual significance** of the variable "employee's **service Time**".

A. $0.202709 < 1.96 \Rightarrow$ Don't reject H_0: employee's service Time is a relevant variable at 5% significance.
B. $70.55 > 1.96 \Rightarrow$ Don't reject H_0: employee's service Time is a relevant variable at 5% significance.
C. $70.55 > 1.65 \Rightarrow$ Reject H_0: $\beta_2 = 0$ at 5% significance.
D. $0.2027/0.00287 > 1.65 \Rightarrow$ Reject H_0: employee's service Time is a relevant variable at 5% significance.
E. $0.2027/0.00287 > 1.96 \Rightarrow$ Reject H_0: $\beta_2 = 0$ at 5% significance.
F. $1.96 < |0.2027/0.00287| \Rightarrow$ Don't Reject H_0: $\beta_2 = 0$ at 5% significance.

12. ________________

13. (1 point) Test the **overall significance** of the explanatory variables in Model 2.

A. R-squared $= 0.97970 > 0.85 \Rightarrow$ Reject $H_0 \rightarrow In, T, E$ jointly relevant at 5% significance.
B. p-value(F-statistic) $< 0.00001 \Rightarrow$ Reject $H_0 \rightarrow In, T, E$ jointly relevant at 5% significance.
C. p-value(F-statistic) $< 0.00001 \Rightarrow$ Don't Reject $H_0 \rightarrow In, T, E$ NOT jointly relevant at 5% significance.
D. $(R^2/k)/(1-R^2)/(T-k) > 0.00001 \Rightarrow$ Reject $H_0 \rightarrow In, T, E$ jointly relevant at 5% significance.
E. F-statistic $> \mathscr{F}(4,495) \Rightarrow$ Reject $H_0 \rightarrow In, T, E$ jointly relevant at 5% significance.
F. F-statistic $< \mathscr{F}(3,496) \Rightarrow$ Don't Reject $H_0 \rightarrow In, T, E$ NOT jointly relevant at 5% significance.

13. ________________

14. The expert thinks that **previous Experience** and **service Time** have the **same effect** on **present Wage** and decides to incorporate this information into the model in order to obtain estimators with a smaller variance.

(a) (1 point) Write down the **null** (H_0) and **alternative** (H_a) hypotheses (in terms of the corresponding β coefficients).

A. H_0: $E_i = T_i$ vs. H_a: $E_i \neq T_i$
B. H_0: $0.2027 > 0.1343$ vs. H_a: $0.2027 <= 0.1343$
C. H_0: $\beta_2 - \beta_3 = 0$ vs. H_a: $\beta_3 - \beta_2 = 0$

D. H_0: $\beta_2 = \beta_3$ vs. H_a: $\beta_2 \neq \beta_3$

E. H_0: $\widehat{\beta}_2 = \widehat{\beta}_3$ vs. H_a: $\widehat{\beta}_2 \neq \widehat{\beta}_3$

F. H_0: $\beta_2 > \beta_3$ vs. $H_a \Rightarrow$ Reject H_0 at 5% significance.

(a) ______________

(b) (1 point) Write down the **restricted** model.

A. $W_i = \beta_0 + \beta_1 In_i + \beta_2 T_i + \beta_3 E_i + u_i$

B. $W_i = \beta_0 + \beta_1 In_i + \beta_2 T_i + (1-\beta_2)E_i + u_i$

C. $W_i = \widehat{\beta}_0 + \widehat{\beta}_1 In_i + (1-\widehat{\beta}_2)E_i + \widehat{u}_i$

D. $\widehat{W}_i = \widehat{\beta}_0 + \widehat{\beta}_1 In_i + \widehat{\beta}_2(T_i + E_i)$

E. $W_i = \beta_0 + \beta_1 In_i + \beta_2(T_i + E_i) + u_i$

F. $\widehat{W}_i = -12.3267 + 1.29671 In_i + 0.202709(T_i + E_i)$

(b) ______________

(c) (2 points) Given the following information:

$$\widehat{W}_i = \underset{(0.87655)}{-8.53682} + \underset{(0.032185)}{1.07565}\, In_i + \underset{(0.0025862)}{0.194511}\,(T_i + E_i) \quad ; \quad \sum_{i=1}^{500} \widehat{u}_i^2 = 5085.58$$

can you tell whether the expert's reasoning is correct? Why?

A. CORRECT because $R^2 = 1 - 5085.58/TSS$ is high.

B. CORRECT because T_i and E_i are relevant variables at 5% significance.

C. CORRECT because $0.194511/0.00258618 = 75.21 > t(496) \Rightarrow Z_i = T_i + E_i$ is significant at 5% significance.

D. NOT CORRECT because of the presence of multicolinearity in $Z_i = T_i + E_i$.

E. NOT CORRECT because $5.82 > t(496) \Rightarrow$ Reject expert's reasoning at 5% significance.

F. NOT CORRECT because $5.82 > \mathcal{F}(496) \Rightarrow$ Reject expert's reasoning at 5% significance.

(c) ______________

Please do not write under this line.

THIRD PART

In the past, it has been alleged that women workers may be suffering from some sort of wage discrimination in the company. In order to take appropriate actions, the company's advisory board asks the expert to find evidence to proof or disproof this. After estimating the appropriate model, he obtains the following results:

Model 3: OLS estimates using the 500 observations 1–500
Dependent variable: W

	Coefficient	Std. Error	t-ratio	p-value
const	−10.016	0.376698	−26.5889	0.0000
In	1.22923	0.0161561	76.0842	0.0000
T	0.200350	0.000943168	212.4222	0.0000
E	0.0885570	0.00402965	21.9764	0.0000
D	−0.346670	0.234570	−1.4779	0.1401
D×E	0.0824304	0.00306608	26.8846	0.0000

$$\sum_{i=1}^{500} In_i = 17012, \quad \sum_{i=1}^{500} T_i = 34879, \quad \sum_{i=1}^{500} E_i = 35764, \quad \sum_{i=1}^{500} D_i = 299, \quad \sum_{i=1}^{500} D_i \times E_i = 22033.$$

$$(X'X)^{-1} = \begin{bmatrix} 0.13747 & -0.00514 & 0.00001 & 0.00054 & -0.03025 & 0.00040 \\ -0.00514 & 0.00025 & 0.00000 & -0.00005 & -0.00012 & 0.00000 \\ 0.00001 & 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00054 & -0.00005 & 0.00000 & 0.00002 & 0.00043 & -0.00001 \\ -0.03025 & -0.00012 & 0.00000 & 0.00043 & 0.05331 & -0.00064 \\ 0.00040 & 0.00000 & 0.00000 & -0.00001 & -0.00064 & 0.00001 \end{bmatrix}$$

Sum of squared residuals	509.924
Standard error of the regression ($\widehat{\sigma}$)	1.01599
Unadjusted R^2	0.997826
$F(5, 494)$	45342.5

15. (a) (1 point) Write down the proposed model.

A. $E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}$
B. $W_i = \gamma_0 + \gamma_1 In_i + \gamma_2 T_i + \gamma_3 E_i + D_i(1 + E_i) + u_i$
C. $W_i = \beta_0 + \beta_1 In_i + \beta_2 T_i + \beta_3 E_i + \beta_4 D_i + \beta_5 D_i E_i + u_i$
D. $\widehat{W}_i = \beta_0 + \beta_1 In_i + \beta_2 T_i + \beta_3 E_i + \gamma D_i + \delta D_i E_i$
E. $\overline{W}_i = -10.016 + 1.229\overline{In}_i + 0.2000\overline{T}_i + 0.088\overline{E}_i - 0.347\overline{D}_i + 0.082\overline{D_i E_i} + \overline{u}_i$
F. $\widehat{W}_i = -10.016 + 1.229\widehat{In}_i + 0.2000\widehat{T}_i + 0.088\widehat{E}_i - 0.347\widehat{D}_i + 0.082\widehat{D_i E_i}$

(a) ________________

(b) (2 points) What is the new variable **D**? What value does it take for Angela? And for Brandon?

A. D_i is the Discrimination variable that takes a value of 1 if women are discriminated and 0 otherwise.
B. $D_i = \mathscr{I}$(men > women), where $\mathscr{I}$ is the indicator function.
C. $D_i = \mathscr{I}$(Company = discrimination subsample), where $\mathscr{I}$ is the indicator function.
D. $D_i = \mathscr{I}$(i = man), where $\mathscr{I}$ is the indicator function.
E. $D_i = 1$ if men earn more than women and 0 otherwise.
F. $D_i = 1$ or $D_i = 0$ depending on salary.

(b) ________________

(c) (2 points) What is the **average previous Experience** for men? And for women?

A. 35764/500 & 22033/500 months.
B. 22033/500 & 35764/500 months.
C. 37 & 52 months.
D. 52 & 37 months.
E. 3.8 & 6.14 years.
F. 6.14 & 3.8 years.

(c) ________________

(d) (1 point) Give an interpretation for the **estimated** coefficient of the **D×E** variable.

A. The estimated increase in wages is 0.0824304 when the variable Di×Ei increases one measurement unit *ceteris paribus*.
B. It is the expected increase in wages when the variable Di×Ei increases one measurement unit *ceteris paribus*.
C. It is the estimated increase in wages when the variable Di×Ei increases one measurement unit *ceteris paribus*.
D. Any worker (irrespective of sex) will experience a salary increase of 82.4 euros when there is an extra month of previous experience *ceteris paribus*.
E. A man's salary increases 82.4 euros more than a woman when they have an extra month of previous experience *ceteris paribus*.
F. A woman's salary increases 82.4 euros more than a woman when they have an extra month of previous experience *ceteris paribus*.

(d) ________________

(e) (2 points) Give an interpretation for the **estimated** coefficient of the **E** variable.

A. The estimated increase in wages is 0.0885570 when the variable Di×Ei increases one measurement unit *ceteris paribus*.
B. It is the average salary increase when the variable Ei increases one measurement unit *ceteris paribus*.
C. It is the expected salary increase when the variable Ei increases one measurement unit *ceteris paribus*.
D. Any worker (irrespective of sex) will experience a salary increase of 88.6 euros when there is an extra month of previous experience *ceteris paribus*.
E. A woman's salary increases 88.6 euros when she has an extra month of previous experience *ceteris paribus*.
F. A man's salary increases 88.6 euros when she has an extra month of previous experience *ceteris paribus*.

(e) ________________

(f) (1 point) Do you agree with the claim of salary discrimination? Why?

A. NO, I DON'T AGREE, because the coefficient for the dummy variable D is not significant at th5 5% significance level.
B. YES, I AGREE, but it is men who are suffering from wage discrimination because the coefficient of D is negative.
C. YES, I AGREE, women suffer from wage discrimination because F=45342.5>F(5,494).
D. YES, I AGREE, women suffer from wage discrimination because F=2958.84>F(2,494).
E. IT DEPENDS, because D×E is a relevant variable while D is not.
F. IT DEPENDS on the salary level, because the coefficients of the dummy variables are of different sign.

(f) ________________

16. (a) (1 point) From the results obtained in Model 3, the expert says that a **female** worker that after **5 years of previous Experience** entered the company **10 years ago** with an **Initial wage of 30000 euros** would have **doubled** her salary by now. Do you agree with the expert?

A. NO, because a 95% confidence interval for the Initial wage coefficient is $\beta_1 \in [1.19748; 1.26097]$ and 2 is not inside.
B. NO, because the last estimated coefficient is significant at the 5% significance level.
C. NO, because 2×30 falls outside a Wage interval of 95% confidence $= [54.21; 58.22]$.
D. YES, because 2×30 falls within a Wage interval of 95% confidence $= [56.21; 62.22]$.
E. YES, because $\widehat{W}_p = -10.016 + 1.229 \times 30000 + 0.2000 \times 10 + 0.088 \times 5 - 0.347 \times 1 + 0.082 \times 5$ falls within a Wage interval of 95% confidence.
F. That depends on the post and experience of the particular woman versus the average male worker.

(a) ________________

(b) (1 point) And if the worker **was a man**?

A. NO, because 2×30 falls outside a Wage interval of 95% confidence $= [48.81; 52.82]$.
B. NO, because a 95% confidence interval for the Initial wage coefficient is $\beta_1 \in [1.19748; 1.26097]$ and 2 is not inside.
C. YES, because $\widehat{W}_p = -10.016 + 1.229 \times 30000 + 0.2000 \times 10 + 0.088 \times 5$ falls within a Wage interval of 95% confidence.
D. YES, because 2 (that is, double) falls within a Wage interval of 95% confidence.
E. YES, because 2×30 falls within a Wage interval of 95% confidence $= [58.81; 62.82]$.
F. YES, I agree because men tend on average to earn more than women.

(b) ________________

Please do not write under this line.