



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
ESCUELA DE INGENIERÍA  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS

ICS2563 — Econometría Aplicada — 1' 2022

## Tarea 4 – Respuesta Pregunta 1

Para comparar las características de los individuos de las diferentes ciudades fijamos el tiempo en  $\text{post} = 1$  ya que así podremos comparar como eran las características de los individuos de diferentes ciudades previo a que ocurra el terremoto, el cual no afectó a todo el país sino que solo al centro.

Primero que nada creamos 3 bases de datos nuevas, a partir de la base de datos original. La primera base de datos contiene la información para  $\text{post} = 1$  en la ciudad 0, la segunda la información de  $\text{post} = 1$  en la ciudad 1 y la tercera la información que respecta a  $\text{post} = 1$  en la ciudad 3.

Lo primero que podemos notar es que la base de datos asignada a la ciudad 3 tiene casi 1000 observaciones más que la ciudad 0 y 1 por lo que se podría decir que en esta ciudad habita más gente. Las ciudades 0 y 1 tienen una diferencia muy pequeña de observaciones.

|          |               | Female    |           |           | immig      |           |           | age      |          |           |
|----------|---------------|-----------|-----------|-----------|------------|-----------|-----------|----------|----------|-----------|
|          | Observaciones | Mean      | sd        | p-value   | Mean       | sd        | p-value   | Mean     | sd       | p-value   |
| Ciudad 0 | 3019          | 0,5061297 | 0,5000453 | 0,273459  | 0,09307718 | 0,2905887 | 0,3907173 | 38,57138 | 6,126872 | 9,1exp-6  |
| Ciudad 1 | 2995          | 0,5018364 | 0,5000801 | 0,5976635 | 0,1048414  | 0,3064001 | 0,0855064 | 40,19299 | 6,463051 | 3,9exp-24 |
| Ciudad 2 | 3986          | 0,4884596 | 0,4999295 | 0,1283138 | 0,0938284  | 0,2916264 | 0,3957224 | 38,40843 | 6,357858 | 8,2exp-14 |

|          | educ     |          |            | exp      |          |            | log_income |           |            |
|----------|----------|----------|------------|----------|----------|------------|------------|-----------|------------|
|          | Mean     | sd       | p-value    | Mean     | sd       | p-value    | Mean       | sd        | p-value    |
| Ciudad 0 | 8,078503 | 2,121273 | 0,2520585  | 24,52435 | 6,531338 | 0,0001726  | 2,291271   | 0,5753521 | 0,132127   |
| Ciudad 1 | 7,573626 | 2,185496 | 4,25exp-58 | 26,64341 | 6,880582 | 4,47exp-61 | 2,42741    | 0,6067065 | 3,78exp-39 |
| Ciudad 2 | 8,552183 | 2,19213  | 1,32exp-58 | 23,88585 | 6,714935 | 1,87exp-34 | 2,223015   | 0,6333247 | 7,01exp-27 |

Sobre los salarios promedio de cada ciudad, podemos decir que este es más alto en la ciudad número 1, siendo en la ciudad 0 y 2 bastante similares.

Sobre las variables de esta base de datos, como se puede ver en la tabla, las variables de si es inmigrante o mujer presentan un valor p mayor a 0.05. No así las variables de edad, educación y experiencia laboral.



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
ESCUELA DE INGENIERÍA  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS

ICS2563 — Econometría Aplicada — 1' 2022

## Tarea 4 – Respuesta Pregunta 2

a)

Mostramos los resultados de tres modelos de regresión lineal en el periodo 1. Se eligió este periodo de tiempo por las razones de la pregunta 1.

En el primer modelo se decidió comparar el salario con solo los años de educación:

| Coefficients: |          |            |         |            |
|---------------|----------|------------|---------|------------|
|               | Estimate | Std. Error | t value | Pr(> t )   |
| (Intercept)   | 1.757828 | 0.022718   | 77.38   | <2e-16 *** |
| educ          | 0.067398 | 0.002701   | 24.95   | <2e-16 *** |

En el segundo modelo se decidió añadir los años de experiencia y los años de experiencia al cuadrado, tal como lo hizo Mincer en su función de ingresos.

|             | Estimate   | Std. Error | t value | Pr(> t )     |
|-------------|------------|------------|---------|--------------|
| (Intercept) | -2.039e-01 | 6.175e-02  | -3.302  | 0.000964 *** |
| educ        | 1.263e-01  | 2.480e-03  | 50.938  | < 2e-16 ***  |
| exp         | 6.866e-02  | 3.985e-03  | 17.228  | < 2e-16 ***  |
| exp2        | -3.395e-04 | 7.214e-05  | -4.706  | 2.55e-06 *** |

Por último se decidió añadir variables como si se es mujer, la edad, la ciudad y si se es migrante.

|             | Estimate   | Std. Error | t value | Pr(> t )     |
|-------------|------------|------------|---------|--------------|
| (Intercept) | 1.457e+00  | 1.845e-01  | 7.896   | 3.18e-15 *** |
| educ        | 4.018e-01  | 3.081e-02  | 13.040  | < 2e-16 ***  |
| exp         | 3.490e-01  | 3.172e-02  | 11.001  | < 2e-16 ***  |
| exp2        | -4.229e-04 | 7.182e-05  | -5.889  | 4.02e-09 *** |
| female      | -8.151e-02 | 9.921e-03  | -8.216  | 2.37e-16 *** |
| age         | -2.747e-01 | 3.079e-02  | -8.920  | < 2e-16 ***  |
| city        | -5.285e-02 | 6.000e-03  | -8.809  | < 2e-16 ***  |
| immig       | -1.467e-01 | 1.677e-02  | -8.751  | < 2e-16 ***  |

Podemos ver que en todas las variables los p.values son pequeños y menores a 0.05, lo que nos dice que son estadísticamente significativos. En el primer modelo se puede ver un beta de la variable educ equivalente a 0.067398. Al ser positivo podemos concluir que mientras más años de escolaridad, mayor sueldo. En el modelo 2 se aprecia que el beta de la variable educ aumenta y pasa a ser 0.1263. Esto se puede deber a que las variables de experiencia laboral hicieron aún más relevante la educación. Por último el coeficiente beta de educ en el tercer modelo aumenta aún más hasta llegar a 0.418. Podemos concluir que las variables agregadas en este modelo están dando mayor protagonismo a el factor de la educación en el aumento de salario.

b)

Ahora procedemos a mostrar los resultados de 3 modelos de regresión lineal con las mismas variables que los anteriores pero considerando todos los periodos de tiempo de la base de datos.

|             | Estimate | Std. Error | t value | Pr(> t ) |     |
|-------------|----------|------------|---------|----------|-----|
| (Intercept) | 1.599388 | 0.013318   | 120.09  | <2e-16   | *** |
| educ        | 0.070407 | 0.001582   | 44.52   | <2e-16   | *** |

|             | Estimate   | Std. Error | t value | Pr(> t ) |     |
|-------------|------------|------------|---------|----------|-----|
| (Intercept) | -2.210e-01 | 3.640e-02  | -6.072  | 1.28e-09 | *** |
| educ        | 1.266e-01  | 1.477e-03  | 85.703  | < 2e-16  | *** |
| exp         | 6.104e-02  | 2.350e-03  | 25.979  | < 2e-16  | *** |
| exp2        | -2.339e-04 | 4.258e-05  | -5.493  | 3.98e-08 | *** |

|             | Estimate   | Std. Error | t value | Pr(> t ) |     |
|-------------|------------|------------|---------|----------|-----|
| (Intercept) | 5.394e-01  | 8.263e-02  | 6.527   | 6.81e-11 | *** |
| educ        | 2.428e-01  | 1.289e-02  | 18.843  | < 2e-16  | *** |
| exp         | 1.799e-01  | 1.337e-02  | 13.456  | < 2e-16  | *** |
| exp2        | -2.655e-04 | 4.224e-05  | -6.285  | 3.32e-10 | *** |
| female      | -9.061e-02 | 5.956e-03  | -15.213 | < 2e-16  | *** |
| age         | -1.167e-01 | 1.292e-02  | -9.028  | < 2e-16  | *** |
| city        | -2.900e-02 | 3.601e-03  | -8.052  | 8.45e-16 | *** |
| immig       | -1.870e-01 | 1.007e-02  | -18.574 | < 2e-16  | *** |

Podemos ver que en todas las variables los p.values son pequeños y menores a 0.05, lo que nos dice que son estadísticamente significativos. Además se puede ver que en los coeficientes beta de la variables de los dos primeros modelos no hay grandes diferencias. Sin embargo en el modelo 3 existen diferencias mucho más notorias entre coeficientes betas. Por ejemplo el beta de educ en el modelo 3 del periodo 1 es de 0.418 y en el de todos los periodos es de 0.2428. La diferencia es significativa en la variable city también, con valores beta de -0.05285 y -0.029 respectivamente. Esto nos quiere decir que el efecto de vivir en alguna ciudad sobre el salario es mayor en el promedio de todos los periodos que en el periodo 1 de tiempo.



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
ESCUELA DE INGENIERÍA  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS

ICS2563 — Econometría Aplicada — 1' 2022

## Tarea 4 – Respuesta Pregunta 3

Para este problema crearemos un diseño en el que solo se tomarán en cuenta los habitantes de la ciudad 1 para los periodos de tiempo 0, 1 y 2. Compararemos la información de  $t=0$  y  $t=1$ , la cual será nuestro control, con la de  $t=2$ , en donde ya ocurrió el terremoto y será nuestro treatment. Creemos que este es el mejor diseño ya que las otras ciudades están en posiciones geograficas diferentes y pueden estar sujetas a otras variables que no conocemos, lo que puede dañar la comparación. Además que el número de observaciones obtenido en la pregunta 1 para la ciudad 1 (2995) nos parece suficiente para poder ver el efecto en el salario.

| Variable   | Treatment | Control  | p value for difference |
|------------|-----------|----------|------------------------|
| log_income | 1.962028  | 2.331713 | 1.01E-140              |

Se aprecia una gran diferencia entre los promedios de las variable log?income en los distintos periodos. hay una clara disminución del ingreso entre las muestras. En el tiempo en que ya habia ocurrido el terremoto el sueldo era menor. La diferencia se puede corroborar con el bajo pvalue de 1.01E-140. Debido a que pvalue en este caso es menor a 0.05, la variable no alcanza un intervalo de confianza del 95%. Se rechaza la hipótesis nula y se acoge la hipótesis alternativa, lo que nos dice que los datos son significativamente distintos.

Procedemos a analizar otras variables para ver si estas son estadísticamente significativas y afectan más que el terremoto en el salario. Para esto creamos una base de datos controlada para que sean solo de la ciudad 1 en todos los periodos de tiempo y generamos un modelo de regresión lineal que compare salario con muchas otras variables teniendo como variable dependiente si ocurrió el terremoto o no.

| Coefficients: |            |            |         |          |     |
|---------------|------------|------------|---------|----------|-----|
|               | Estimate   | Std. Error | t value | Pr(> t ) |     |
| (Intercept)   | 8.851e-01  | 1.695e-01  | 5.221   | 1.82e-07 | *** |
| earthquake    | -4.576e-01 | 1.155e-02  | -39.608 | < 2e-16  | *** |
| educ          | 2.520e-01  | 2.719e-02  | 9.268   | < 2e-16  | *** |
| exp           | 2.032e-01  | 2.812e-02  | 7.228   | 5.30e-13 | *** |
| exp2          | -2.188e-04 | 7.265e-05  | -3.011  | 0.00261  | **  |
| female        | -9.849e-02 | 1.073e-02  | -9.177  | < 2e-16  | *** |
| age           | -1.396e-01 | 2.721e-02  | -5.128  | 2.99e-07 | *** |
| immig         | -2.328e-01 | 1.741e-02  | -13.376 | < 2e-16  | *** |

Vemos que el terremoto es la variable que tiene el beta con mayor diferencia a 0, por lo que es la variable que más influye en el salario disminuyendolo. La variable que le sigue en nivel de influencia es la educación, pero la variable terremoto afecta mucho más. -0.4576 del beta del terremoto comparado con 0.252 de la educación.

Se concluye que el terremoto afecto a promedio de los salarios de la ciudad 1, bajandolos.



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
ESCUELA DE INGENIERÍA  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS

ICS2563 — Econometría Aplicada — 1' 2022

## Tarea 4 – Respuesta Pregunta 4

Mediremos los salarios logaritmicos de los migrantes y no migrantes en los tres periodos  $t=0$ ,  $t=1$  y  $t=2$  y en la ciudad 1. A la vez encontraremos los pvalues para la diferencia entre  $t=0$  con  $t=1$  y  $t=1$  con  $t=2$ . Además calcularemos pvalue para las diferencias entre salarios de migrantes y no migrantes manteniendo el periodo. Con esto veremos la significancia de las diferencias. Tomemos en cuenta que la muestra de migrantes tiene un tamaño que va de 314 a 321 y que la muestra de no migrantes tiene un tamaño de entre 2661 y 2681.

| Variable                                  | Promedio Ingreso en $t=0$ | Promedio Ingreso en $t=1$ | Promedio Ingreso en $t=2$ | p.value diferencia $t=0$ y $t=1$ | p.value diferencia $t=1$ y $t=2$ |
|---|---------------------------|---------------------------|---------------------------|----------------------------------|----------------------------------|
| Migrante                                  | 2.13803                   | 2.295228                  | 1.608146                  | 0.001049358                      | 8.33E-36                         |
| No Migrante                               | 2.247935                  | 2.442892                  | 2.004318                  | 3.41E-32                         | 3.66E-140                        |
| p.value diferencia migrante y no migrante | 0.001947157               | 5.36E-05                  | 2.41E-20                  |                                  |                                  |

Vemos que los p.value de todas las diferencias son menores a 0.05, por lo que son diferencias significativamente estadísticas. También podemos observar una gran caída en los sueldos de los migrantes y los no migrantes en  $t=2$ . la caída es de alrededor de 0.687082 en los migrantes y de 0.438574 en los Chilenos. Debemos averiguar si esta diferencia de la caída entre migrantes y no migrantes se debe mayoritariamente al subsidio o a otros factores.

Debido a que hay muchas variables, haremos un modeo de regresión lineal para cuantificar el efecto de ser migrante o no en el sueldo y poder ver la relevancia del subsidio. La variable dependiente será  $\ln(wage)$  y la base de datos serán todos los datos de la ciudad 1 en el tiempo 2, por lo que tendrán considerado el efecto del terremoto.

| Coefficients: |            |            |         |              |
|---------------|------------|------------|---------|--------------|
|               | Estimate   | Std. Error | t value | Pr(> t )     |
| (Intercept)   | 0.1939911  | 0.2145254  | 0.904   | 0.36592      |
| immig         | -0.4171092 | 0.0313568  | -13.302 | < 2e-16 ***  |
| educ          | 0.2179390  | 0.0324204  | 6.722   | 2.14e-11 *** |
| exp           | 0.1733090  | 0.0351193  | 4.935   | 8.46e-07 *** |
| exp2          | -0.0002451 | 0.0001316  | -1.863  | 0.06262 .    |
| female        | -0.1081194 | 0.0193886  | -5.576  | 2.68e-08 *** |
| age           | -0.1063968 | 0.0324981  | -3.274  | 0.00107 **   |

Se puede ver que casi todas las variables son menores a 0.05, por lo que hay diferencias significativas. También podemos apreciar que el coeficiente beta con mayor diferencia a 0 es justamente el de la variable  $\ln(wage)$  con -0.4171092. Esto lo podemos interpretar como que hay muchas variables que explican la diferencia entre el sueldo en el periodo  $t=2$  de la ciudad 2, sin embargo la que tiene mayor peso y por tanta diferencia

es la de ser migrante o no. La variable que le sigue es la del nivel de educación, y tiene alrededor de la mitad de influencia. El subsidio a Chilenos puede explicar la gran importancia que tiene la variable *immig* en este modelo.



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
ESCUELA DE INGENIERÍA  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS

ICS2563 — Econometría Aplicada — 1' 2022

## Tarea 4 – Respuesta Pregunta 5

a)

Se nos dice que habilidad es una variable fija en el tiempo, pero la variable nivel de educación, por el contrario, debiera ir aumentando en promedio con el tiempo, ya que una persona no nace educada, sino que se adquiere la educación mediante pasa el tiempo. Esto podría probocar un sesgo en el que, al no tomar en cuenta la variable habilidad, se aumenta de manera errónea el efecto de la educación. Especialmente en gente joven, ya que como en promedio tienen menos educación, se acentúa la relevancia de la variable habilidad. Mientras las personas sean más adultas en promedio tendrán más educación y predominará esa variable sobre habilidad.

b)

Se puede reducir el sesgo agregando la variable habilidad en una regresión lineal que compare salarios y educación. Con esto podemos ver los coeficientes betas específicamente de educación y de habilidad para medir cuanto esta influyendo la variable. Por la explicación de 4a) concluimos que el sesgo de habilidad en las personas con edad más avanzada es menor, por lo que una solución para atenuar el sesgo producido por habilidad puede ser considerar a personas de una cierta edad para arriba. Quizas puede ser una buena idea considerar a personas de 24 años para arriba, ya que esta es una edad donde muchas personas ingresan al mundo laboral.



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
ESCUELA DE INGENIERÍA  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS

ICS2563 — Econometría Aplicada — 1' 2022

## Tarea 4 – Respuesta Pregunta 6

Primero debemos separar a las personas que se trasladan de ciudad. Creamos 6 data frames según las ciudades de traslado con la información de las personas en el periodo  $t=1$  y  $t=2$  debido a que aquí es cuando ocurre el terremoto.

Ya con los datos procedemos a obtener una tabla con los promedios de el log\_income de las personas según el traslado y el periodo  $t$ .

| Traslado de ciudad x a ciudad y | Promedio Ingreso en $t=1$ | Promedio Ingreso en $t=2$ | p.value      |
|---------------------------------|---------------------------|---------------------------|--------------|
| 0,1                             | 2.361932                  | 1.86331                   | 2.39E-14     |
| 0,2                             | 2.176769                  | 2.191883                  | 0.8199006    |
| 1,0                             | 2.306172                  | 2.180475                  | 0.03884042   |
| 1,2                             | 2.301259                  | 2.320958                  | 0.7904732    |
| 2,0                             | 2.261711                  | 2.024945                  | 0.0004653488 |
| 2,1                             | 2.223302                  | 1.625185                  | 3.77E-16     |

Vemos que los traslados hacia la ciudad uno son los con mayor diferencia en promedio y con un p.value menor. La diferencia en (0,1) es de 0.49 y la diferencia de (2,1) es de 0.598117. En la pregunta 3 vimos que existía una diferencia de 0.369685 en la diferencia de salarios entre los dos primeros periodos y  $t=2$ . En los dos casos hay una disminución en los salarios, pero en el caso que mide a los trasladados la disminución es mayor. Esta razón se puede deber a múltiples factores. Uno de ellos puede ser el tamaño de la muestra. En el problema 3 el número de observaciones es de 2995, mientras que los datos de los trasladados tienen de 314 a 414 observaciones, lo que lo puede hacer más impreciso. El ocupar los datos de personas que se trasladan y extrapolarlos al resto de la población me parece erróneo porque es una muestra muy pequeña y además es un grupo de personas con características particulares que no reflejan sin sesgos al resto de las personas, sin embargo creemos que pueden mostrar una tendencia general.