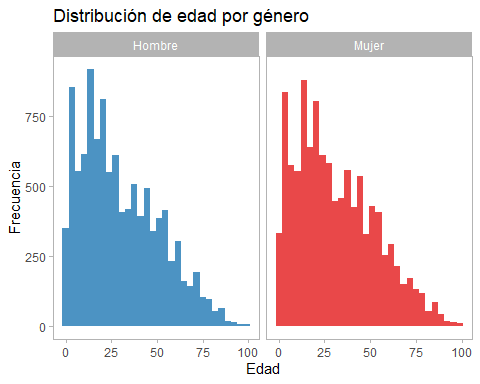
Análisis Econométrico

Práctica 1: Econometría

José Burgos

**Ejercicio 1: Considere la base de datosBase *ENCFT 20161 – 20164.xlsx* que contiene información sobre el lado de la oferta del mercado de trabajo.**

1. Importe la información a R, creando las bases de datos correspondientes.
2. Adjunte el módulo de MIEMBROS y el de OCUPACIÓN, creando una variable id que concatene la información de la vivienda, el hogar y el miembro, de modo que en una misma base de datos queden las características de los individuos y las variables de mercado laboral.
3. Cree la variable EDAD y MUJER, donde esta última sea igual a 1 si el individuo es mujer y 0 en caso contrario.
4. Usando histogramas, grafique la distribución de la **edad** por **género**.



1. Cree la variable salario por hora (W) utilizando la información de ingreso laboral en la ocupación principal. Tenga en cuenta que el ingreso reportado en la base de datos no está necesariamente en la misma escala para todos los individuos (p. ej., algunos reportan salario por hora, otros por mes, etc.).
2. Muestre en una tabla la distribución de edad por percentil de ingreso. *En particular, reporte los percentiles 5, 25, 50, 75 y 95.*

============= =========== ============ ============ ============ ============ Rango de Edad Percentil 5 Percentil 25 Percentil 50 Percentil 75 Percentil 95 ============= =========== ============ ============ ============ ============ 35-44 25.00 48.38 66.25 102.81 218.75 Menos de 25 16.12 37.50 50.00 75.00 125.00 25-34 25.00 45.00 62.50 87.50 187.50 45-54 18.75 43.75 62.50 100.00 250.00 55-64 13.44 37.50 61.88 97.81 251.87 65 o más 12.50 32.25 43.75 70.00 148.75 ============= =========== ============ ============ ============ ============

1. Para el caso del modelo de regresión lineal, muestre las propiedades de muestra finita del estimador de mínimos cuadrados.

* **Propiedades de muestra finita:**
  1. **Insesgadez**:
  + *por expectativas iteradas* . Esto demuestra que, para cualquier muestra (de tamaño ), el estimador de MCO está centrado en el valor verdadero del parámetro.
  1. **Varianza condicional**
* De aquí se obtienen las varianzas individuales y las covarainzas entre componentes de .
  1. **Insesgadez del estimados de**
* El estiamdor habitual de la varianza del error es:
* Bajo los supuestos clásicos se tiene , es decir, es insesgado para en muestras finitas.
  1. **Optimalidad (Teorema de Gauss-Markov)**

1. **Modelo de regresión lineal**
2. Muestre que las ecuaciones normales de mínimos cuadrados implican que y que .

Demostración: Para demostrar que las ecuaciones normales de mínimos cuadrados implican que y , partimos de la definición del error , donde es la predicción del modelo.

La ecuación de regresión lineal es:

1. Muestre que la solución para el término constante es:: .
2. Muestre que la solución para es
3. Los datos en el archivo *KoopandTobias2004.xls* son una extracción de 15 observaciones de una muestra de 2,178 individuos de un conjunto de variables. Sea igual a la constante, educación, experiencia y habilidad. Mientras que sea igual a la educación de la madre, la educación del padre, y el número de hermanos. Sea el salario.
4. Compute los coeficientes de MCO en la regresión de sobre X1. Reporte los coeficientes.

**Modelo:**

| **Characteristic** | **Beta** | **95% CI** | **p-value** |
| --- | --- | --- | --- |
| (Intercept) | 1.7 | 0.30, 3.0 | 0.021 |
| education | 0.01 | -0.09, 0.12 | 0.8 |
| experience | 0.07 | -0.03, 0.18 | 0.2 |
| ability | 0.03 | -0.19, 0.24 | 0.8 |
| Abbreviation: CI = Confidence Interval | | | |

1. Compute los coeficientes de MCO en la regresión de sobre X1 y X2. Reporte los coeficientes.

| **Characteristic** | **Beta** | **95% CI** | **p-value** |
| --- | --- | --- | --- |
| (Intercept) | 0.05 | -2.1, 2.2 | >0.9 |
| education | 0.03 | -0.08, 0.13 | 0.6 |
| experience | 0.10 | -0.01, 0.21 | 0.061 |
| ability | 0.03 | -0.25, 0.31 | 0.8 |
| education01 | 0.10 | -0.06, 0.26 | 0.2 |
| education02 | 0.00 | -0.10, 0.10 | >0.9 |
| siblings | 0.06 | -0.10, 0.22 | 0.4 |
| Abbreviation: CI = Confidence Interval | | | |

1. Regrese cada una de las tres variables en X2 sobre todas las variables en X1. Estas nuevas variables denótenlas como ¿Cuáles son las medias muestrales de estas variables? Explique el resultado.

==============================================================  
 Dependent variable:   
 --------------------------------  
 education01 education02 siblings  
 (1) (2) (3)   
--------------------------------------------------------------  
education -0.084 0.027 -0.048   
 (0.241) (0.355) (0.208)   
   
experience -0.286 0.001 -0.056   
 (0.236) (0.348) (0.204)   
   
ability 0.503 0.873 -0.958\*\*  
 (0.486) (0.718) (0.421)   
   
Constant 13.760\*\*\* 12.000\*\* 3.320   
 (3.035) (4.478) (2.630)   
   
--------------------------------------------------------------  
Observations 15 15 15   
R2 0.263 0.155 0.384   
Adjusted R2 0.062 -0.076 0.216   
Residual Std. Error (df = 11) 1.293 1.907 1.120   
F Statistic (df = 3; 11) 1.307 0.672 2.284   
==============================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 1: Medias muestrales

Regresiones de X2 sobre X1

|  | media |
| --- | --- |
| education01 | 12.06667 |
| education02 | 12.66667 |
| siblings | 2.20000 |

*Las medias muéstrales de las variables en* *representan el valor promedio de cada variable después de haber sido ajustadas por las variables en* *. Esto significa que estas medias reflejan el efecto de la educación, experiencia y habilidad sobre la educación de los padres y el número de hermanos, eliminando la variabilidad que podría estar asociada con estas variables. En otras palabras, las medias muéstrales indican el nivel promedio de educación de los padres y el número de hermanos, controlando por las características del individuo como la educación, experiencia y habilidad.*

1. Compute para la regresión de sobre X1 y X2. Repita el cómputo para el caso en la que el término constante es omitido de X1. ¿Qué sucede con R2?

Table 1: Cálculo de R² para los modelos

Comparación entre modelos con y sin constante

|  | R2 |
| --- | --- |
| Con constante | |
| y sobre X1 | 0.1833511 |
| y sobre X2 | 0.5161341 |
| Sin constante | |
| y sobre X1 | 0.9803976 |
| y sobre X2 | 0.9929909 |

*En esta regresión cuando se omite el término constante, el* *aumenta, lo que indica que el modelo sin constante explica una mayor proporción de la variabilidad de la variable dependiente. El problema de esto, es que al eliminar el término constante estamos omitiendo la media de la variable dependiente, en este caso la media del salario, lo que puede llevar a una interpretación errónea de los resultados. Con esto el modelo ya no explica la variación respecto a la media de* *.*

1. Compute el R2 para la regresión con todas las variables incluyendo el término constante. Interprete los resultados.

*El* *para la regresión con todas las variables incluyendo el término constante es 0.5161341 . Esto nos indica que aproximadamente el 51.6% de la variabilidad del salario puede ser explicada por las variables independientes incluidas en el modelo (educación, experiencia, habilidad, educación de los padres y número de hermanos).*

1. Considere el siguiente modelo de función de producción para un conjunto de N industrias.

Donde es el nivel de producción en la industria , es la productividad en la industria que depende de la productividad agregada común a todos los sectores y de un componente idiosincrático asociado a la industria. Asimismo, y son el factor capital y trabajo, respectivamente.

1. Tal como está formulado ¿por qué no es posible estimar este modelo por mínimos cuadrados ordinarios?

*Este modelo no es posible estimar por mínimos cuadrados ordinarios (MCO) porque violaría el primer supuesto de linealidad, ya que no es lineal en los parámetros. La función de producción es una función no lineal en los parámetros* *y* *, lo que significa que no se puede expresar como una combinación lineal de los parámetros.*

1. Aplique la transformación apropiada para que este modelo sea estimable por MCO y formule el nuevo modelo econométrico.

*Para transformar el modelo de función de producción a una forma lineal, podemos tomar el logaritmo natural de ambos lados de la ecuación. Esto nos dará una forma lineal en los parámetros:*

1. La base de datos k401k es un subconjunto de los datos analizados por Papke (1995) para estudiar la relación entre la participación en un plan de pensión y la generosidad del plan. La variable prate es el porcentaje de trabajadores que están inscritos en el plan y que tienen cuenta activa; esta es la variable que se quiere explicar. La medida de la generosidad es la tasa de contribución (de la empresa) al plan, mrate. Esta variable es la cantidad promedio con la que la empresa contribuye al plan de cada trabajador por cada peso que aporte el trabajador. Por ejemplo, si *mrate = 0.50*, entonces a una contribución de 1 del trabajador corresponde una contribución de 50 centavos de la empresa.
   1. Encuentre el promedio de la tasa de participación y el promedio de la tasa de contribución para la muestra.

| Medida | Prate | Mrate |
| --- | --- | --- |
| Promedio | 87.36 | 0.73 |

ii) Ahora, estime el modelo

y de los resultados, el tamaño de la muestra y R-cuadrada.