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.

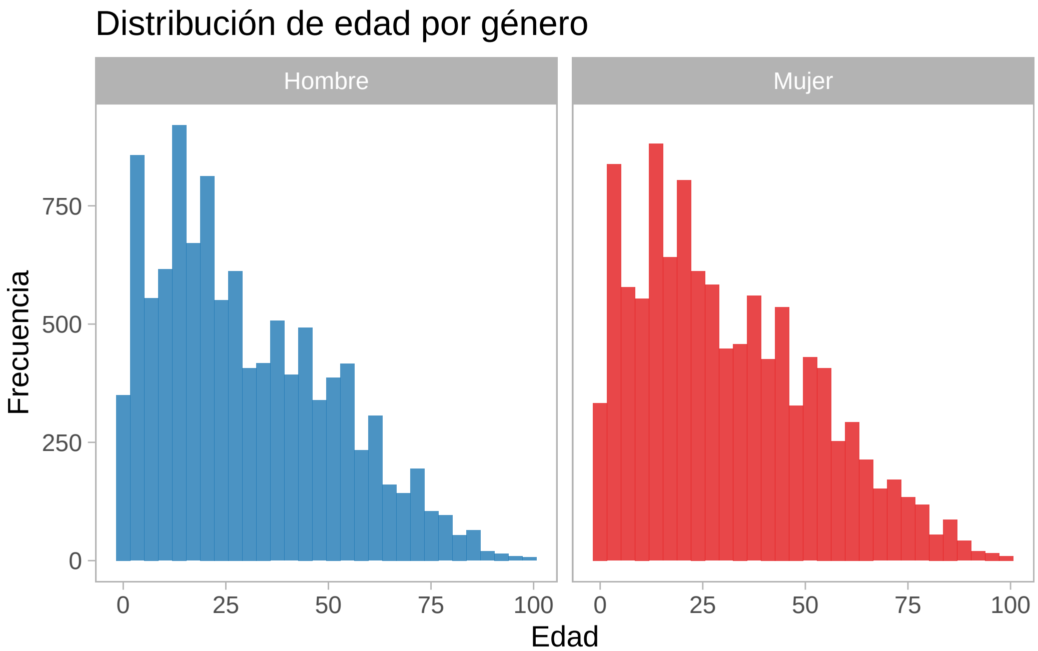
a) Importe la información a R, creando las bases de datos correspondientes.

b) 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.

c) Cree la variable `EDAD` y `MUJER`, donde esta última sea igual a 1 si el individuo es mujer y 0 en caso contrario.

```
encft <- encft |>
  mutate(
    #Cuestionario: mujer = 2 y hombre = 1
    mujer = ifelse(sexo == 2, 1, 0)
  )
```

d) Usando histogramas, grafique la distribución de la **edad** por **género**.



e) 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.).

f) Muestre en una tabla la distribución de edad por percentil de ingreso. *En particular, reporte los percentiles 5, 25, 50, 75 y 95.*

2. 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**:
   $$\hat{\beta} = (X'X)^{-1}X'y \;\rightarrow\; E[\hat{\beta}|X] = \beta$$
   por expectativas iteradas $E[\hat{\beta}] = \beta$. Esto demuestra que, para cualquier muestra (de tamaño $n >\_$), el estimador de MCO está centrado en el valor verdadero del parámetro.

   2. **Varianza condicional**

   $$Var(\hat{\beta}|X) = Var[(X'X)^{-1}X'\epsilon|X] = \sigma^2(X'X)$$

   De aquí se obtienen las varianzas individuales y las covarainzas entre componentes de $\hat{\beta}$.

3. **Insesgadez del estimados de σ^2**

El estiamdor habitual de la varianza del error es:

$$s^2 = \frac{e'e}{n-(k+1)}, \qquad e = y - X\hat{\beta}$$

Bajo los supuestos clásicos se tiene $E[s^2|X] = \sigma^2$, es decir, s^2 es insesgado para σ^2 en muestras finitas.

4. **Optimalidad (Teorema de Gauss-Markov)**

3. **Modelo de regresión lineal**

$$y = \alpha + \beta x + \varepsilon$$

a) Muestre que las ecuaciones normales de mínimos cuadrados implican que $\sum_i e_i = 0$ y que $\sum_i x_i e_i = 0$.

b) Muestre que la solución para el término constante es:: $\hat{\alpha} = \bar{y} - b\bar{x}$.

c) Muestre que la solución para b es $b = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}(x_i-\bar{x})^2}$

6. 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 $X1$ igual a la constante, educación, experiencia y habilidad. Mientras que sea $X2$ igual a la eduación de la madre, la educación del padre, y el número de hermanos. Sea y el salario.

a) Compute los coeficientes de MCO en la regresión de y sobre X1. Reporte los coeficientes.

Modelo:

$$wage = \beta_0 + \beta_1 education + \beta_2 experience + \beta_3 ability + \epsilon$$

```r
modelo1 <- lm(wage ~ education + experience + ability, data = koopand)

gtsummary::tbl_regression(modelo1, intercept = TRUE)
```

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

c) Regrese cada una de las tres variables en X2 sobre todas las variables en X1. Estas nuevas variables denótenlas como X_2^* ¿Cuáles son las medias muestrales de estas variables? Explique el resultado.

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