

# Econometría 2502 - Taller 2

Rafael Marulanda, Carlos Puerto y Lorena Toro

27 de setiembre de 2025

## 1 Ejercicio sobre la base de datos wage1

Descargue la base de datos `wage1`.

- a) Calcule estadísticas descriptivas de `wage`, `educ`, `exper`, `tenure` y `numdep`. Muestre sus resultados. Calcule el histograma de frecuencias de cada una de las variables. Interprete sus resultados.
- b) Tabule y calcule las estadísticas descriptivas de las variables `nonwhite`, `female` y `married`. Interprete sus resultados.
- c) Calcule el número de mujeres no blancas en la muestra.
- d) Calcule el número de hombres blancos y casados de la muestra.
- e) Calcule el número de mujeres blancas y solteras en la muestra.
- f) Calcule la covarianza y la correlación entre `wage`, `educ`, `exper`, `tenure` y `numdep`.

### a) Estadísticas Descriptivas

#### Resumen de las Distribuciones de Variables

La Tabla ?? proporciona las estadísticas resumen para las variables continuas: salarios por hora, años de educación, experiencia en el mercado laboral, antigüedad en el empleo y número de dependientes económicos.

#### Análisis de la Distribución Salarial

La variable de salario, medida en dólares por hora, presenta una media de \$5.90 con una desviación estándar de \$3.69, como se muestra en la Tabla 1. La distribución muestra un sesgo positivo pronunciado, evidenciado por la diferencia entre la media (\$5.90) y la mediana (\$4.65). Este sesgo hacia la derecha indica que mientras la mayoría de los trabajadores perciben salarios modestos, una pequeña proporción de trabajadores con altos ingresos influye significativamente en el promedio hacia arriba.

Table 1: Descriptive Statistics

	(1)				
	mean	sd	min	max	count
average hourly earnings	5.90	3.69	0.53	24.98	526.00
years of education	12.56	2.77	0.00	18.00	526.00
years potential experience	17.02	13.57	1.00	51.00	526.00
years with current employer	5.10	7.22	0.00	44.00	526.00
number of dependents	1.04	1.26	0.00	6.00	526.00
Observations	526				

La Figura 1 ilustra la distribución de frecuencias de los salarios, revelando una concentración de observaciones en el rango de \$2-\$6 con una cola larga hacia la derecha que se extiende hasta el valor máximo de \$24.98. Este patrón se alinea con los fenómenos de desigualdad salarial bien documentados en la literatura de economía laboral, donde las diferencias en capital humano y la segmentación del mercado laboral contribuyen a la dispersión de ingresos.

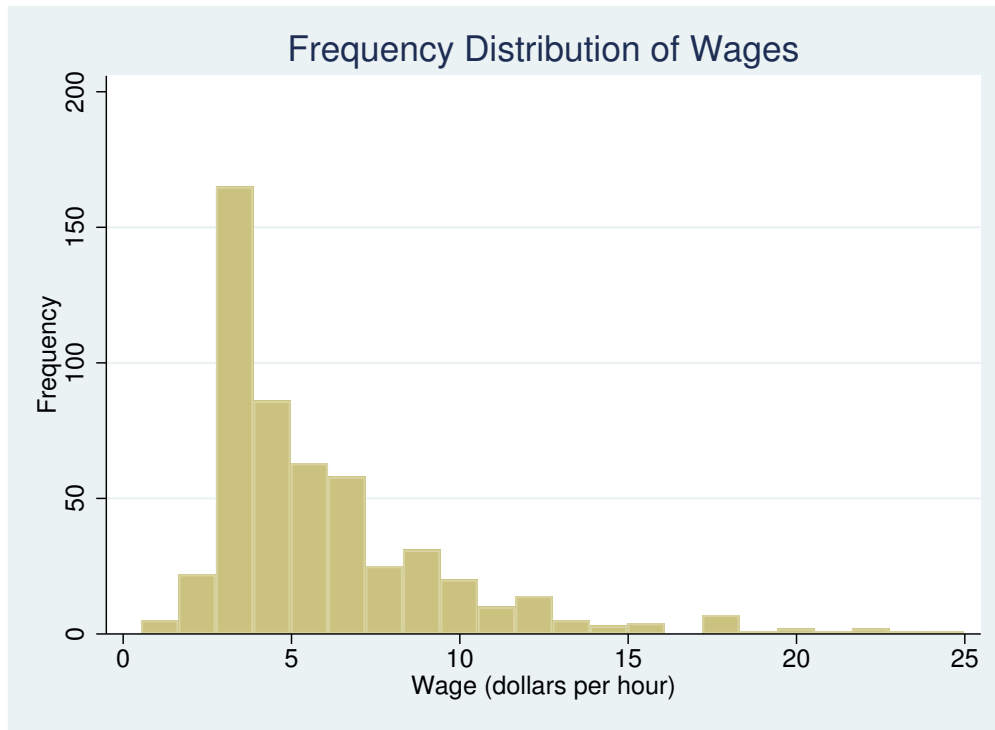


Figure 1: Distribución de Frecuencias de Salarios

## Patrones de Logro Educativo

La variable de educación presenta una media de 12.56 años y una desviación estándar de 2.77 años. La Figura 2 revela una concentración notable exactamente en 12 años de educación, correspondiente a la finalización de la educación secundaria. Este pico modal, que abarca aproximadamente 200 observaciones, domina la distribución y sugiere que la fuerza laboral consiste principalmente en graduados de secundaria.

La desviación estándar relativamente modesta indica heterogeneidad educativa limitada dentro de la muestra. El rango se extiende desde 0 hasta 18 años, capturando individuos sin educación formal hasta aquellos con títulos de posgrado. Sin embargo, la concentración alrededor de la finalización de la educación secundaria sugiere que esta muestra puede representar un segmento particular del mercado laboral, posiblemente trabajadores del sector manufacturero o de servicios.

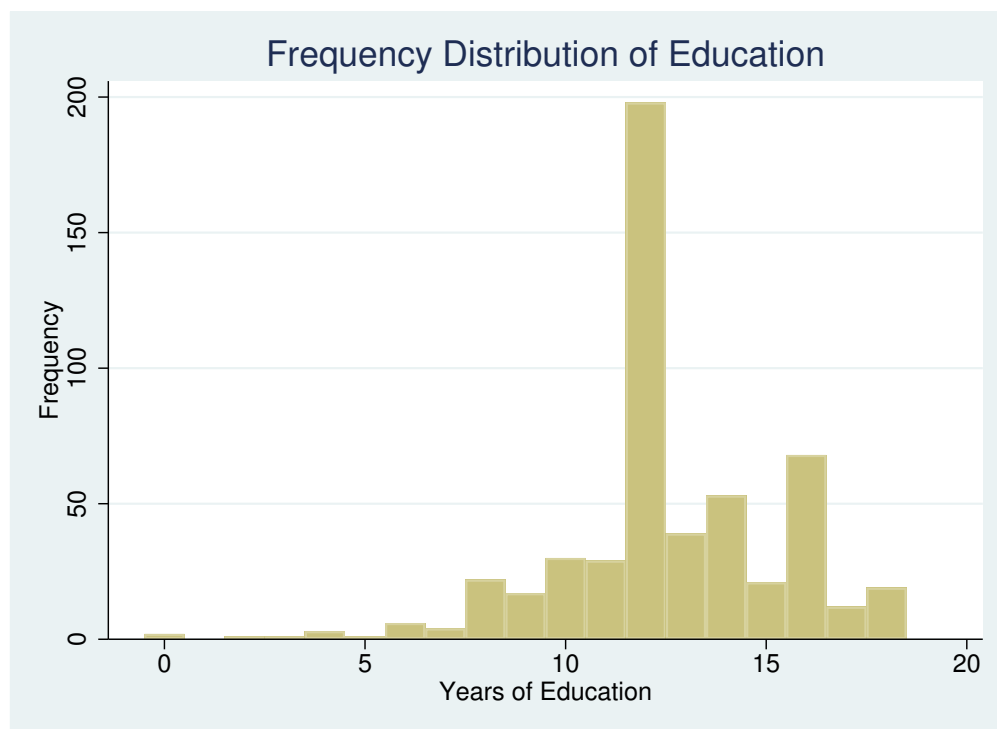


Figure 2: Distribución de Frecuencias de Años de Educación

## Distribución de Experiencia y Dinámica Profesional

La experiencia en el mercado laboral, con una media de 17.02 años y una desviación estándar de 13.57 años, muestra una variabilidad considerable a través de la muestra. La Figura 3 muestra una distribución con sesgo positivo, con la frecuencia más alta ocurriendo entre trabajadores con niveles de experiencia relativamente bajos (1-10 años), seguida de una disminución gradual en frecuencia a medida que aumenta la experiencia.

La mediana de experiencia de 13.5 años cae por debajo de la media, confirmando el sesgo positivo de la distribución. El valor máximo de 51 años indica la presencia de trabajadores acercándose a la edad de jubilación, mientras que el mínimo de 1 año sugiere recién ingresados al mercado laboral. Este amplio rango refleja una fuerza laboral que abarca múltiples etapas profesionales, desde trabajadores en inicio de carrera hasta aquellos con participación extensa en el mercado laboral.

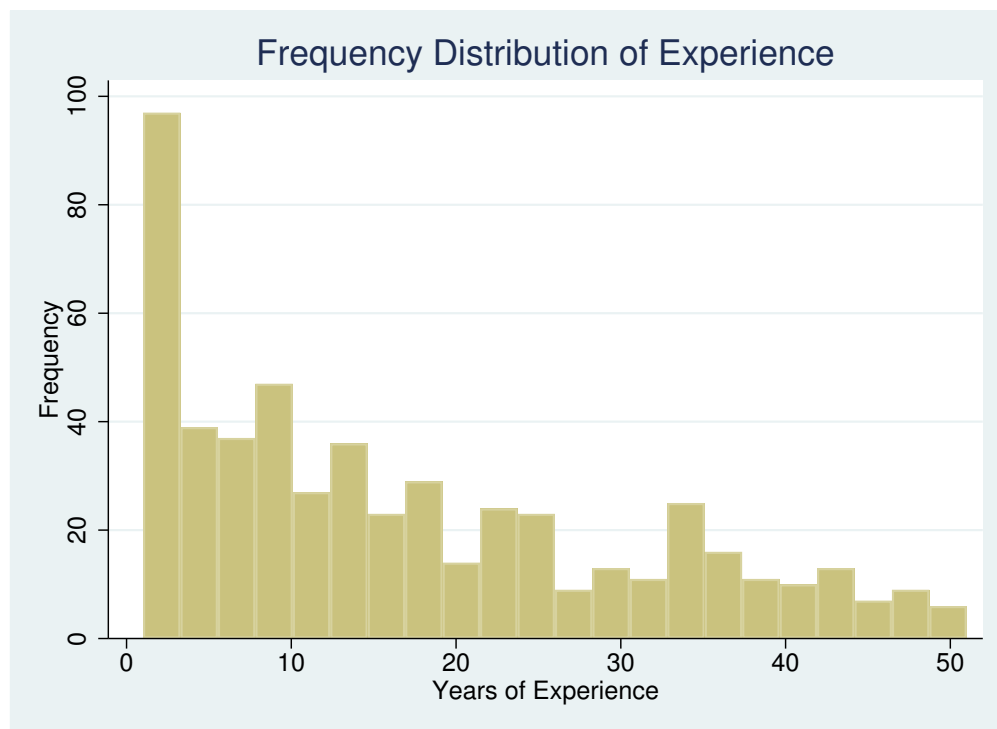


Figure 3: Distribución de Frecuencias de Años de Experiencia

### Antigüedad Laboral y Estabilidad en el Empleo

La variable de antigüedad presenta las características distribucionales más extremas entre todas las variables examinadas. Con una media de 5.10 años pero una mediana de solo 2 años, la distribución exhibe un sesgo positivo severo. La Figura 4 revela que aproximadamente 220 trabajadores ( 42% de la muestra) reportan cero antigüedad con su empleador actual, sugiriendo tasas altas de rotación laboral.

La desviación estándar de 7.22 años excede la media, indicando variabilidad extrema en los patrones de vinculación laboral. Mientras que la mayoría de los trabajadores demuestran antigüedad limitada con su empleador actual, el valor máximo de 44 años revela la presencia de trabajadores con estabilidad laboral excepcional.

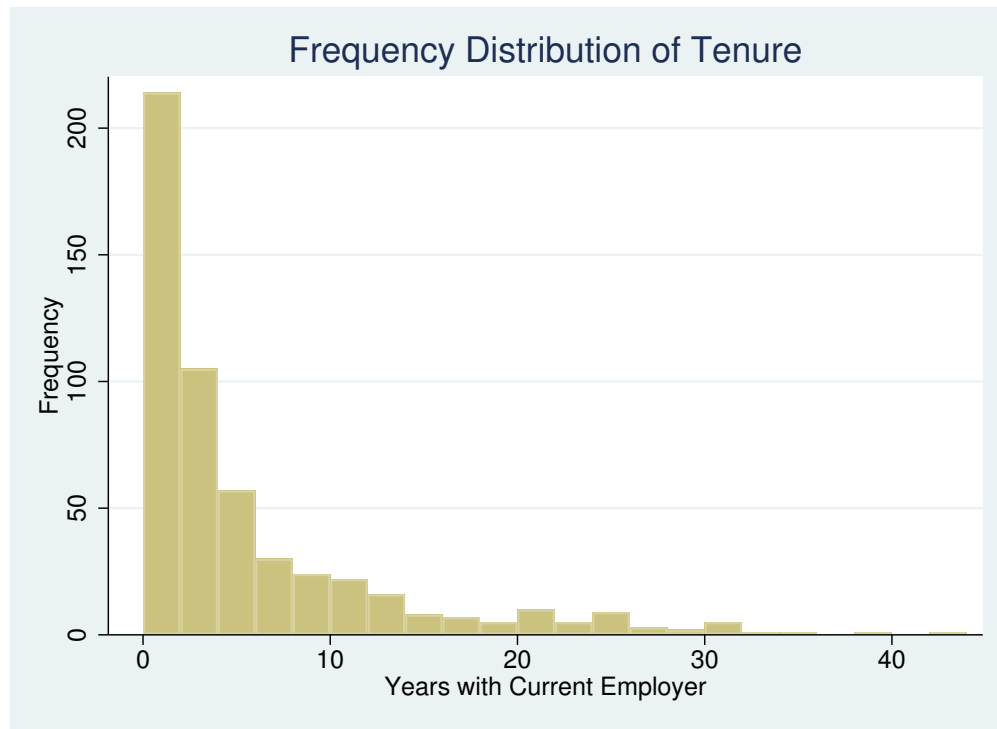


Figure 4: Distribución de Frecuencias de Años de Antigüedad

### Composición del Hogar y Dependientes

La variable del número de dependientes muestra una media de 1.04 con una desviación estándar de 1.26. La distribución se concentra fuertemente en valores bajos, con la mediana en 1 dependiente. El máximo de 6 dependientes ocurre raramente en la muestra, indicando que las familias grandes son poco comunes entre esta fuerza laboral.

### Interpretación general

El análisis descriptivo revela varios patrones importantes que caracterizan esta muestra del mercado laboral. Primero, la pronunciada desigualdad salarial, evidenciada por la distribución sesgada de salarios. Segundo, la concentración del logro educativo en el nivel de secundaria sugiere una fuerza laboral posicionada en el medio de la distribución de habilidades, no dominada ni por graduados universitarios ni por trabajadores sin educación básica.

La desconexión entre la experiencia promedio (17 años) y la antigüedad promedio (5 años) indica movilidad laboral sustancial a lo largo de las carreras de los trabajadores. Tal movilidad puede reflejar tanto cambios de trabajo voluntarios para el avance profesional como separaciones involuntarias debido a condiciones económicas.

El número promedio relativamente bajo de dependientes, junto con la proporción sustancial de trabajadores sin dependientes, puede indicar una demografía de fuerza laboral más joven. Estas características distribucionales muestran colectivamente un cuadro de un segmento

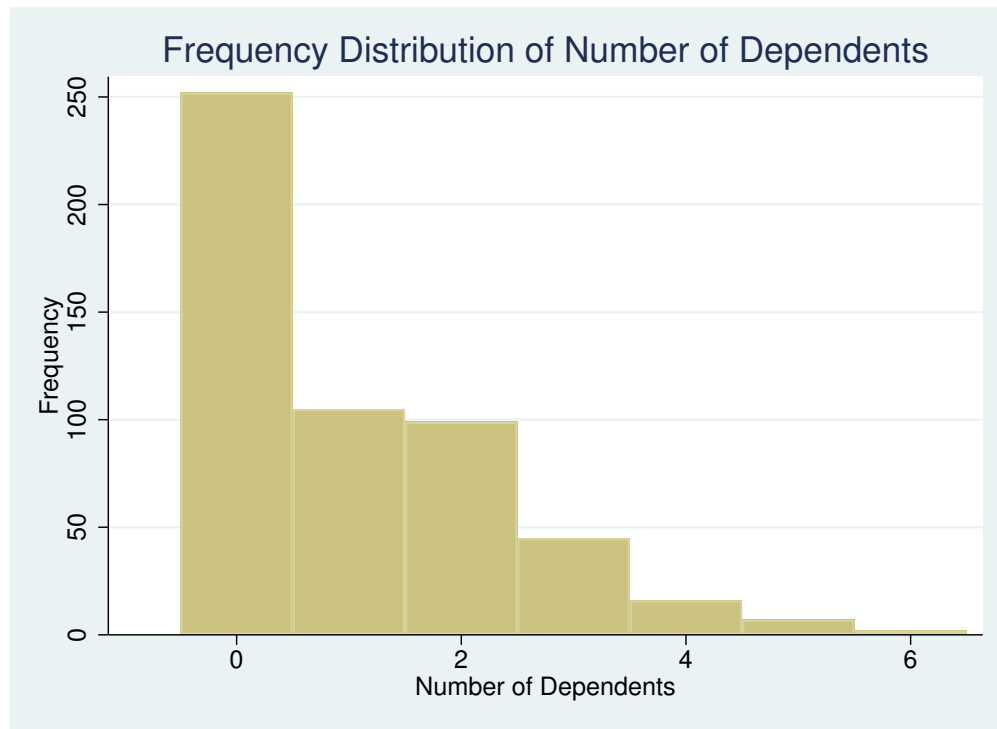


Figure 5: Distribución de Frecuencias de Número de Dependientes

del mercado laboral caracterizado por niveles de habilidad moderados, dispersión salarial significativa, alta movilidad laboral y estructuras familiares variadas.

Salida de Stata de la parte a

## b) Análisis de variables categóricas

### Distribución y Características de las Variables Binarias

El análisis de las variables categóricas **nonwhite**, **female** y **married** proporciona información fundamental sobre la composición demográfica de la muestra. La Tabla ?? presenta un resumen de estas variables binarias, mientras que las Tablas 2, ?? y ?? detallan sus distribuciones individuales.

### Composición Racial

La variable **nonwhite** revela una muestra predominantemente blanca, con solo el 10.27% (54 trabajadores) identificados como no blancos y el 89.73% (472 trabajadores) como blancos. Esta distribución refleja la composición racial del mercado laboral estadounidense durante el período de recolección de datos, aunque muestra una subrepresentación de minorías étnicas comparada con los promedios nacionales actuales.

La Tabla ?? presenta diferencias notables en las características económicas por grupo racial. Los trabajadores blancos perciben salarios promedio de \$5.94 por hora, mientras que los no

Table 2: Distribution by Race

	(1) nonwhite b/pct
0	472 89.7
1	54 10.3
Total	526 100.0
<i>N</i>	526

Frequency and percentage shown

blancos reciben \$5.48, una diferencia de \$0.47 o aproximadamente 8% menos. Esta brecha salarial persiste a pesar de que los trabajadores no blancos tienen ligeramente más experiencia laboral (17.59 vs 16.95 años). Los trabajadores no blancos presentan menor logro educativo promedio (11.87 vs 12.64 años), equivalente a casi un año menos de escolaridad. Esta diferencia educativa podría explicar parcialmente la disparidad salarial observada, aunque no puede descartarse la presencia de discriminación laboral.

### Distribución por Género

La variable **female** muestra una distribución relativamente equilibrada con 52.09% hombres (274 trabajadores) y 47.91% mujeres (252 trabajadoras). Esta proporción cercana a la paridad sugiere una muestra representativa en términos de participación laboral por género para el período analizado.

Sin embargo, la Tabla ?? revela disparidades económicas sustanciales entre géneros. Los hombres ganan en promedio \$7.10 por hora comparado con \$4.59 para las mujeres, representando una brecha salarial del 54.7%. Esta diferencia dramática constituye la disparidad más pronunciada entre todas las variables categóricas analizadas. Las mujeres también presentan menor educación promedio (12.32 vs 12.79 años), menor experiencia laboral (16.43 vs 17.56 años), y notablemente menor antigüedad laboral (3.62 vs 6.47 años). La diferencia en antigüedad sugiere mayor rotación laboral femenina, posiblemente relacionada con interrupciones de carrera por responsabilidades familiares o discriminación en retención laboral.

### Estado Civil

La variable **married** indica que el 60.84% de la muestra (320 trabajadores) está casada, mientras que el 39.16% (206 trabajadores) permanece soltera. Esta mayoría de trabajadores casados es consistente con patrones demográficos típicos de la fuerza laboral adulta.

La Tabla ?? revela que los trabajadores casados perciben salarios significativamente superiores (\$6.57 vs \$4.84), una prima matrimonial del 35.7%. Los casados también exhiben mayor

experiencia laboral (20.47 vs 11.66 años) y antigüedad (6.49 vs 2.95 años). Estas diferencias sugieren que el matrimonio puede estar correlacionado con mayor estabilidad laboral y acumulación de capital humano, aunque también podría reflejar efectos de selección donde individuos con mejores perspectivas económicas tienen mayor probabilidad de casarse.

## **Raza y Género**

La distribución por género es notablemente similar entre grupos raciales: 51.91% de los trabajadores blancos son hombres comparado con 53.70% entre no blancos. Esta similitud sugiere que no existe segregación ocupacional significativa por género dentro de grupos raciales en esta muestra. Sin embargo, la combinación de ser mujer y no blanca (25 trabajadoras, 4.75% de la muestra) representa el grupo más vulnerable a discriminación múltiple. Tablas ??

## **Raza y Estado Civil**

Los trabajadores blancos muestran mayor tasa de matrimonio (61.86%) comparado con los no blancos (51.85%). Esta diferencia de 10 puntos porcentuales podría reflejar factores socioeconómicos que afectan las decisiones matrimoniales, incluyendo estabilidad económica y normas culturales diferenciadas. ??

## **Género y Estado Civil**

La Tabla ?? revela patrones matrimoniales marcadamente diferentes por género. El 68.61% de los hombres están casados versus solo 52.38% de las mujeres. Esta disparidad sugiere que las mujeres solteras tienen mayor participación laboral que los hombres solteros, consistente con teorías de especialización familiar donde las mujeres casadas enfrentan mayores barreras para la participación laboral. ??

## **Interpretación General**

Los resultados revelan segmentación significativa del mercado laboral según características demográficas. Las brechas salariales observadas—8% por raza, 54.7% por género, y 35.7% por estado civil—sugieren que factores más allá del capital humano influyen en la determinación salarial. La intersección de estas características probablemente amplifica las desventajas: una mujer no blanca soltera enfrenta múltiples penalizaciones salariales acumulativas.

La menor antigüedad laboral de mujeres y trabajadores no blancos indica mayor vulnerabilidad a interrupciones laborales, menor acumulación de capital humano específico a la empresa, y potencialmente menor acceso a promociones internas. Estos patrones sugieren la necesidad de políticas laborales que aborden no solo las brechas salariales directas, sino también las barreras estructurales que impiden la acumulación de experiencia y antigüedad en grupos desfavorecidos.

El análisis también sugiere que el matrimonio funciona como un marcador de estabilidad económica y laboral, aunque la causalidad permanece ambigua. Los trabajadores casados



podrían recibir primas salariales debido a mayor productividad percibida, discriminación estadística favorable, o simplemente porque individuos económicamente exitosos tienen mayor probabilidad de casarse y permanecer casados.

Salida de Stata de la parte b)

## 2 Regresión lineal simple: predicción de ventas

Una pequeña empresa contrata a un consultor para predecir el valor de las ventas semanales de su producto si el gasto en publicidad semanal aumenta a \$750. El consultor lleva un registro de cuánto gastó la empresa en publicidad por semana y las ventas semanales correspondientes durante los últimos seis meses.

El consultor escribe:

“Durante los últimos seis meses, el gasto semanal promedio en publicidad ha sido de \$500 y las ventas semanales promedio han sido de \$10,000. Según los resultados de una regresión lineal simple, predigo que las ventas serán de \$12,000 si se gastan \$750 por semana en publicidad”.

- a) ¿Cuál es la regresión simple estimada utilizada por el consultor para hacer esta predicción?
- b) Dibuje un gráfico de la línea de regresión estimada. Busque los valores semanales promedio en el gráfico.

### a) Regresión Simple Estimada

El modelo de regresión lineal simple poblacional se especifica como:

$$\text{Ventas} = \beta_0 + \beta_1 \cdot \text{Publicidad} + u$$

Donde:

- *Ventas* es la variable dependiente ( $y$ ).
- *Publicidad* es la variable independiente ( $x$ ).
- $\beta_0$  es el intercepto.
- $\beta_1$  es la pendiente.
- $u$  es el término de error.

El consultor utiliza un modelo de regresión lineal. La ecuación de esta línea estimada es:

$$\text{Ventas} = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{Publicidad}$$

Para encontrar los valores de los estimadores, se usa la información dada:

- El gasto semanal promedio en publicidad ( $\bar{x}$ ) es de \$500.
- Las ventas semanales promedio ( $\bar{y}$ ) son de \$10,000.
- Una predicción específica: cuando el gasto en publicidad ( $x_p$ ) es 750, las ventas predichas ( $\hat{y}_p$ ) son \$12,000.

Una propiedad de la regresión lineal es que la línea de regresión siempre pasa por el punto de las medias ( $\bar{x}, \bar{y}$ ). Por lo tanto, este punto debe satisfacer la ecuación de regresión:

$$\text{Ecuación 1: } \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \implies 10,000 = \hat{\beta}_0 + \hat{\beta}_1(500)$$

Además, se sabe que el punto de la predicción (750, 12,000) también se encuentra sobre la línea de regresión estimada:

$$\text{Ecuación 2: } 12,000 = \hat{\beta}_0 + \hat{\beta}_1(750)$$

Ahora se tiene un sistema de dos ecuaciones lineales con dos incógnitas ( $\hat{\beta}_0$  y  $\hat{\beta}_1$ ). Se resuelve restando la Ecuación 1 de la Ecuación 2:

$$\begin{aligned} (12,000 - 10,000) &= (\hat{\beta}_0 + 750\hat{\beta}_1) - (\hat{\beta}_0 + 500\hat{\beta}_1) \\ 2,000 &= 250\hat{\beta}_1 \end{aligned}$$

Despejando  $\hat{\beta}_1$ :

$$\hat{\beta}_1 = \frac{2,000}{250} = 8$$

Esta pendiente ( $\hat{\beta}_1 = 8$ ) indica el efecto marginal. Significa que, por cada dólar adicional gastado en publicidad, las ventas semanales aumentarán en \$8.

Se substituye el valor de  $\hat{\beta}_1$  en la Ecuación 1 para encontrar el intercepto ( $\hat{\beta}_0$ ):

$$\begin{aligned} 10,000 &= \hat{\beta}_0 + 8(500) \\ 10,000 &= \hat{\beta}_0 + 4,000 \\ \hat{\beta}_0 &= 10,000 - 4,000 = 6,000 \end{aligned}$$

El intercepto ( $\hat{\beta}_0 = 6,000$ ) es el valor predicho de las ventas semanales si el gasto en publicidad es cero.

Por lo tanto, la regresión simple estimada utilizada por el consultor es:

$$\text{Ventas} = 6,000 + 8 \cdot \text{Publicidad}$$

## b) Gráfico de la línea de Regresión Estimada

Una línea se define con dos puntos, entonces usando los puntos dados se define esta. El eje horizontal ( $X$ ) representa el *Gasto en Publicidad* y el eje vertical ( $Y$ ) representa las *Ventas Semanales*.

Los puntos clave en esta línea son:

- El intercepto en  $Y$ :  $(0, 6000)$
- El punto de las medias:  $(\bar{x}, \bar{y}) = (500, 10000)$
- El punto de predicción:  $(750, 12000)$

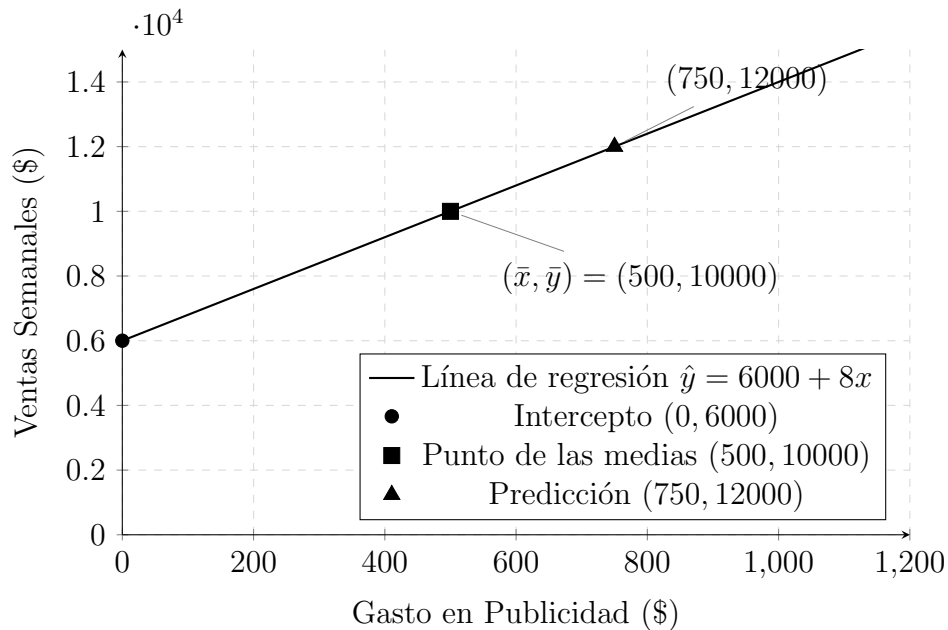


Figure 6: Línea de regresión estimada y puntos clave (intercepto, medias y predicción).

El punto de las medias actúa como el “centro de gravedad” para la nube de datos a partir de la cual se estimó la regresión.

### 3 Relación entre temperatura y ventas de refrescos

Un vendedor de gaseosas en los juegos de baseball de la Universidad de Costa Rica observa que se venden más refrescos cuanto más cálida es la temperatura en el momento del juego. Basado en 32 juegos en casa que cubren cinco años, el proveedor estima que la relación entre las ventas de refrescos y la temperatura es

$$\hat{y} = -240 + 8x,$$

donde  $y$  es el número de refrescos que vende y  $x$  la temperatura en grados Fahrenheit.

- a) Interprete los parámetros estimados. ¿Tienen sentido las estimaciones? ¿Por qué o por qué no?
- b) En un día en el que se pronostique que la temperatura a la hora del juego será 80°F, prediga cuántos refrescos venderá el vendedor.
- c) ¿Por debajo de qué temperatura son cero las ventas previstas?
- d) Dibuje un gráfico de la línea de regresión estimada.

#### a) Interpretación de los Parámetros Estimados

La función de regresión muestral estimada es:

$$\hat{y} = -240 + 8x$$

Donde  $\hat{y}$  es el número predicho de refrescos vendidos y  $x$  es la temperatura en grados Fahrenheit.

##### 1. Interpretación del Intercepto ( $\hat{\beta}_0 = -240$ ):

El intercepto,  $\hat{\beta}_0$ , representa el valor predicho de la variable dependiente ( $y$ ) cuando la variable independiente ( $x$ ) es igual a cero.

Entonces el intercepto de -240 sugiere que si la temperatura fuera de 0°F, se "venderían" -240 refrescos. Evidentemente, esto es un sinsentido, ya que no se puede vender una cantidad negativa de un producto.

Esta falta de sentido se debe a la extrapolación fuera del rango muestral. Los datos de temperatura utilizados para estimar el modelo probablemente no incluyen temperaturas de 0°F o cercanas a cero. Luego el intercepto no tiene una interpretación económica válida en este caso.

##### 2. Interpretación del Coeficiente de Pendiente ( $\hat{\beta}_1 = 8$ ):

El coeficiente de pendiente,  $\hat{\beta}_1$ , mide el *efecto marginal* de la variable independiente sobre la variable dependiente. Es el cambio predicho en  $y$  ante un cambio de una unidad en  $x$ :

$$\hat{\beta}_1 = \frac{\Delta \hat{y}}{\Delta x}.$$

Un valor de  $\hat{\beta}_1 = 8$  indica que, dado *ceteris paribus* (manteniendo todo lo demás constante), por cada aumento de un grado Fahrenheit en la temperatura, se predice que las ventas de refrescos aumentarán en 8 unidades.

Esta estimación sí tiene sentido. Es esperado que en días más cálidos aumente la demanda de bebidas frías como los refrescos. El signo positivo del coeficiente es consistente con la intuición económica.

## b) Predicción de Ventas a 80°F

Para predecir las ventas en un día con una temperatura de 80°F, sustituimos  $x = 80$  en la ecuación de regresión estimada:

$$\hat{y} = -240 + 8(80)$$

$$\hat{y} = -240 + 640$$

$$\hat{y} = 400$$

Se predice que el vendedor venderá 400 refrescos cuando la temperatura sea de 80°F.

## c) Temperatura para Ventas Nulas

Para encontrar la temperatura por debajo de la cual las ventas previstas son cero, se establece  $\hat{y} = 0$  en la ecuación y se despeja  $x$ :

$$0 = -240 + 8x$$

$$240 = 8x$$

$$x = \frac{240}{8}$$

$$x = 30$$

Las ventas previstas son cero a una temperatura de **30°F**. Por debajo de esta temperatura, el modelo predeciría ventas negativas, lo que refuerza la idea de que el modelo solo es válido dentro de un rango relevante de temperaturas.

#### d) Gráfico de la línea de regresión estimada

Se traza la línea de regresión estimada

$$\hat{y} = -240 + 8x,$$

con sus puntos clave: intercepto, predicción en  $x = 80$  y temperatura de ventas nulas en  $x = 30$ .

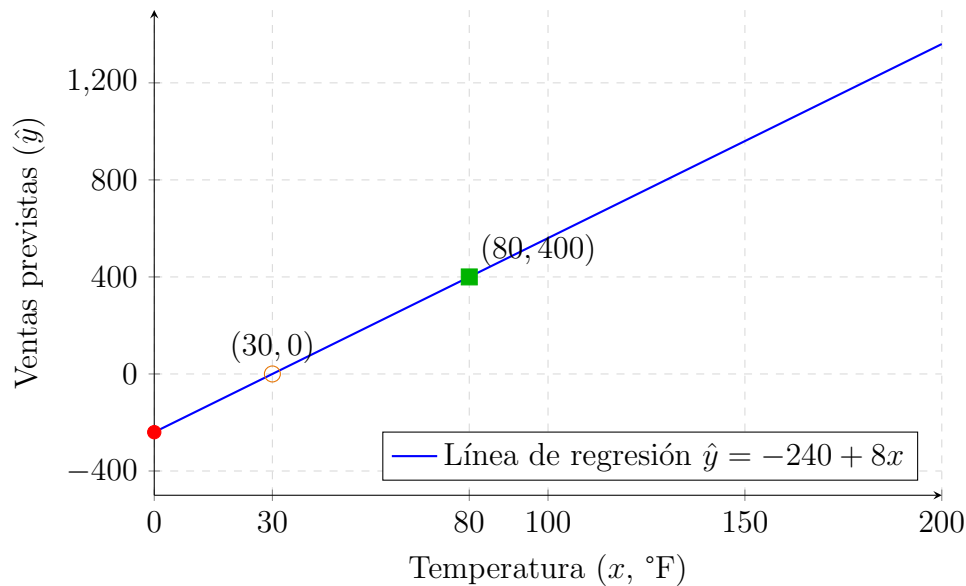


Figure 7: Línea de regresión estimada y puntos relevantes.

## 4 Modelo CAPM

El modelo de valoración de activos de capital (CAPM) es un modelo importante en el campo de las finanzas. Explica las variaciones en la tasa de rendimiento de un activo en función de la tasa de rendimiento de un portafolio que consta de todas las acciones que cotizan en bolsa, lo que se denomina cartera de mercado. Generalmente, la tasa de rendimiento de cualquier inversión se mide en relación con su costo de oportunidad, que es el rendimiento de un activo libre de riesgo. La diferencia resultante se denomina prima de riesgo, ya que es la recompensa o castigo por realizar una inversión arriesgada.

El CAPM dice que la prima de riesgo del activo  $j$  es proporcional a la prima de riesgo de la cartera de mercado. Es decir,

$$r_j - r_f = \beta_j(r_m - r_f)$$

Donde  $r_j$  y  $r_f$  son los rendimientos del activo  $j$  y la tasa libre de riesgo, respectivamente,  $r_m$  es el rendimiento del portafolio de mercado y  $\beta_j$  es el valor “beta” del activo  $j$ -ésimo. La beta de una acción es importante para los inversores, ya que revela la volatilidad de la acción.

Mide la sensibilidad del retorno del activo  $j$  a la variación en todo el mercado de valores. Como tal, los valores de beta inferiores a 1 indican que la acción es “defensiva” ya que su variación es menor que el del mercado. Una beta mayor que 1 indica una “acción agresiva”. Los inversores normalmente quieren conocer una estimación de la beta de una acción antes de comprarla. El modelo CAPM que se muestra arriba es el “modelo económico” en este caso.

El “modelo econométrico” se obtiene al incluir una constante en el modelo (aunque la teoría lo diga debe ser cero) y un término de error. Esto es:

$$r_j - r_f = \beta_j(r_m - r_f) + u$$

- a) Explique por qué el modelo econométrico anterior es un modelo de regresión simple como los discutidos.
- b) En el archivo de datos `capm4.dat` hay datos sobre los rendimientos mensuales de seis empresas (Microsoft, GE, GM, IBM, Disney y Mobil-Exxon), la tasa de rendimiento de la cartera de mercado (MKT) y la tasa de rendimiento sobre el activo libre de riesgo (RISKFREE). Las 132 observaciones abarcan desde enero de 1998 hasta diciembre de 2008. Estime el modelo CAPM para cada empresa y comente sus valores beta estimados. ¿Qué firma parece más agresiva? ¿Qué empresa parece estar más a la defensiva?
- c) Estime el modelo para cada empresa bajo el supuesto de que  $\alpha_j = 0$  ¿Cambian mucho las estimaciones de los valores beta?

## a) El CAPM como un Modelo de Regresión Lineal Simple

El modelo econométrico CAPM es un ejemplo de un modelo de regresión lineal simple porque su estructura matemática es idéntica a la de una regresión simple. Para mostrar lo anterior se debe identificar la variable dependiente, la variable independiente y los parámetros de la regresión.

Un modelo de regresión lineal simple general se define como:

$$Y = \beta_0 + \beta_1 X + u$$

El modelo econométrico CAPM es:

$$(r_j - r_f) = \alpha + \beta_j(r_m - r_f) + u$$

Se puede ver la correspondencia directa entre los dos modelos si se definen las variables de la siguiente manera:

- **Variable Dependiente ( $Y$ ) (exceso de rendimiento del activo  $j$ ):**

$$Y = (r_j - r_f)$$

- **Variable Independiente ( $X$ ) (exceso de rendimiento de la cartera de mercado):**

$$X = (r_m - r_f)$$

- **Intercepto ( $\beta_0$ ) (parámetro alfa ( $\alpha$ )):**

$$\beta_0 = \alpha$$

- **Coefficiente de Pendiente ( $\beta_1$ ) (beta del activo ( $\beta_j$ )):**

$$\beta_1 = \beta_j$$

- **Término de Error ( $u$ ):** Es el término de error estocástico, que captura todos los demás factores y riesgos (riesgo no sistemático) que afectan el rendimiento del activo  $j$  y que no están correlacionados con el rendimiento del mercado.

El modelo es una regresión lineal simple porque muestra una relación en línea recta entre una única variable dependiente (el exceso de rendimiento del activo) y una única variable independiente (el exceso de rendimiento del mercado).

## b) Estimación de los rendimientos en exceso y análisis de betas bajo el CAPM

- **Estimación del modelo CAPM y comentarios sobre los valores  $\beta$ :**

Se estimó el modelo general

$$r_{i,t}^{ex} = \alpha_i + \beta_i (r_{m,t} - r_{f,t}) + \varepsilon_{i,t},$$

para cada empresa, donde  $r_{i,t}^{ex}$  es el rendimiento en exceso de la acción  $i$ . Los alfas y betas extraídos de Stata para cada modelo se muestran en la Tabla 3.

Table 3: Estimaciones de  $\alpha$  y  $\beta$  del modelo CAPM para cada empresa

<b>Empresa</b>	$\hat{\alpha}$	$\hat{\beta}$
Microsoft (MSFT)	0.0012	1.3231
General Electric (GE)	0.0004	0.9027
General Motors (GM)	-0.0006	1.2583
IBM	0.0009	1.1864
Disney (DIS)	-0.0002	0.9015
Exxon-Mobil (XOM)	0.0007	0.4098

Los modelos por empresa serían:



- Microsoft (MSFT):

$$r_{MSFT,t}^{ex} = 0.0012 + 1.32 (r_{m,t} - r_{f,t}) + \varepsilon_t$$

Beta mayor a 1. Acción agresiva, amplifica los movimientos del mercado.

- General Electric (GE):

$$r_{GE,t}^{ex} = 0.0004 + 0.90 (r_{m,t} - r_{f,t}) + \varepsilon_t$$

Beta inferior a 1. Comportamiento ligeramente defensivo.

- General Motors (GM):

$$r_{GM,t}^{ex} = -0.0006 + 1.26 (r_{m,t} - r_{f,t}) + \varepsilon_t$$

Beta mayor a 1. Agresiva.

- IBM:

$$r_{IBM,t}^{ex} = 0.0009 + 1.19 (r_{m,t} - r_{f,t}) + \varepsilon_t$$

Beta mayor a 1. Agresiva.

- Disney (DIS):

$$r_{DIS,t}^{ex} = -0.0002 + 0.90 (r_{m,t} - r_{f,t}) + \varepsilon_t$$

Igual a GE, cercano al mercado.

- Exxon-Mobil (XOM):

$$r_{XOM,t}^{ex} = 0.0007 + 0.41 (r_{m,t} - r_{f,t}) + \varepsilon_t$$

Beta bastante inferior a 1. Altamente defensiva, reacciona menos a los movimientos del mercado.

En todos los casos, los interceptos  $\hat{\alpha}$  no son estadísticamente significativos (teóricamente ceros), lo cual sugiere ausencia de rendimientos anormales.

- **Firma más agresiva:**

La empresa más agresiva es **Microsoft**, con un  $\hat{\beta} \approx 1.32$ , lo cual indica una sensibilidad superior a la unidad frente a los rendimientos del mercado.

- **Firma más defensiva:**

La empresa más defensiva es **Exxon-Mobil**, con un  $\hat{\beta} \approx 0.41$ , lo que refleja un perfil más estable.

## 5 Modelos de Regresión para Precios de Vivienda

El archivo `stockton4.dat` contiene datos sobre 15,009 casas vendidas en Stockton, California, durante el período 1996–1998.

- a) Grafique el precio de venta de las casas en función del área habitable para todas las casas en la muestra.
- b) Estime el modelo de regresión:

$$sprice = \beta_0 + \beta_1 livarea + u$$

para todas las casas en la muestra. Interprete las estimaciones y dibuje un boceto de la línea ajustada.

- c) Estime el modelo cuadrático:

$$sprice = \beta_0 + \beta_1 livarea^2 + u$$

para todas las casas en la muestra. ¿Cuál es el efecto marginal de 100 pies cuadrados adicionales de área habitable en una casa con 1500 pies cuadrados de área habitable?

- d) En el mismo gráfico, trace las líneas ajustadas de los modelos lineal y cuadrático. ¿Cuál parece ajustarse mejor a los datos? Compare la suma de los residuos al cuadrado (SSE) de ambos modelos. ¿Cuál es menor?
- e) Estime el modelo de regresión en (c) utilizando solo casas en lotes grandes. Repita la estimación para casas que no están en lotes grandes. Interprete las estimaciones. ¿Cómo se comparan los resultados?
- f) Grafique el precio de venta de las casas en función de la edad de la casa (AGE). Estime el modelo lineal:

$$sprice = \beta_0 + \beta_1 age + u$$

e interprete los coeficientes estimados. Luego, repita el ejercicio con el modelo log-lineal:

$$\ln(sprice) = \beta_0 + \beta_1 age + u$$

Basándose en los gráficos y el ajuste visual de las líneas de regresión estimadas, ¿cuál de estos dos modelos preferiría? Explique.

# Anexos

## A Código del problema 1

## B Código del problema 4

```
/* *****  
* STATA DO-FILE for CAPM Analysis  
*  
* Description: This script estimates the Capital Asset Pricing Model (CAPM)  
* for six different companies to determine their market risk (beta).  
* Data file:   capm4.dat  
* Date:        September 17, 2025  
* *****  
  
* ——— 1. SETUP ———  
  
clear all  
  
cd "C:\repos\ujav-2502-econometrics\Data"  
  
* ——— 2. LOAD AND PREPARE DATA ———  
  
use "ecnm-2502-hwrk-0204.dta", clear /// equivalent to "cpmn4.dta"  
  
* See the data and variable names after loading  
describe  
  
* a) Market excess return (market return – risk-free rate)  
gen mktrf = mkt – riskfree  
label var mktrf "Market Excess Return (MKT – RF)"  
  
* b) Company excess returns (company return – risk-free rate)  
gen msft_ex = msft – riskfree  
gen ge_ex   = ge – riskfree  
gen gm_ex   = gm – riskfree  
gen ibm_ex  = ibm – riskfree  
gen dis_ex  = dis – riskfree  
gen xom_ex  = xom – riskfree  
  
* Label the new variables for clarity  
label var msft_ex "Microsoft Excess Return"  
label var ge_ex   "GE Excess Return"  
label var gm_ex   "GM Excess Return"  
label var ibm_ex  "IBM Excess Return"  
label var dis_ex  "Disney Excess Return"  
label var xom_ex  "Mobil-Exxon Excess Return"
```

```

* — 3. ESTIMATE CAPM FOR EACH COMPANY —
* The CAPM model is: company_excess_return = alpha + beta * market_excess_ret
* A simple linear regression for each company is runned. The estimated coeffi
* on the ‘mktrf’ variable is the company’s Beta.

```

```

di ""
di "-----"
di "1. Estimating CAPM for Microsoft"
di "-----"
regress msft_ex mktrf

```

```

di ""
di "-----"
di "2. Estimating CAPM for General Electric (GE)"
di "-----"
regress ge_ex mktrf

```

```

di ""
di "-----"
di "3. Estimating CAPM for General Motors (GM)"
di "-----"
regress gm_ex mktrf

```

```

di ""
di "-----"
di "4. Estimating CAPM for IBM"
di "-----"
regress ibm_ex mktrf

```

```

di ""
di "-----"
di "5. Estimating CAPM for Disney"
di "-----"
regress dis_ex mktrf

```

```

di ""
di "-----"
di "6. Estimating CAPM for Mobil-Exxon"
di "-----"
regress xom_ex mktrf

```

## C Código del problema 5