Escuela de Economía Universidad de Carabobo
Profesor: Exaú Navarro Pérez. **Econometría**

Regresión Múltiple:

Municipio	Ocupados	Población Analfabeta Mayor de 10 años	Total de Viviendas
Bejuma	18.874	1.835	12.221
Carlos Arvelo	49.602	5.996	38.703
Diego Ibarra	39.807	2.951	26.114
Guacara	67.956	3.100	47.273
Juan José Mora	23.705	2.405	16.760
Libertador	59.787	5.619	42.998
Los Guayos	59.052	2.598	37.471
Miranda	11.337	1.352	7.594
Montalbán	9.857	796	6.264
Naguanagua	67.452	2.594	44.186
Puerto Cabello	69.892	4.027	46.074
San Diego	42.620	975	24.684
San Joaquín	25.210	1.348	16.817
Valencia	334.491	17.091	221.464

Fuente: INE, Datos Censo 2011

$$X'X = \begin{bmatrix} 14 & 52.687 & 588.623 \\ 52.687 & 422.021.527 & 5.004.549.811 \\ 588.623 & 5.004.549.811 & 62.208.291.265 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 0,1435698333062770000 & -0,0000394403732687499 & 0,0000018144334831000 \\ -0,0000394403732687500 & 0,0000000623436490702 & -0,0000000046422491354 \\ 0,0000018144334831000 & -0,0000000046422491354 & 0,0000000003723675620 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 879.642 \\ 7.478.474.842 \\ 93.485.343.044 \end{bmatrix}$$

$$(X'X)^{-1}\,(X'Y) = b = \begin{bmatrix} 959,152657885 \\ -2,4402506071 \\ 1,6900177984 \end{bmatrix}$$

Regresión Múltiple:

Tabla 1: ***Ocupados, Población Analfabeta y Total de Viviendas en el Estado Carabobo (Año 2011).***

Municipio	Ocupados	Población Analfabeta Mayor de 10 años	Total de Viviendas
Bejuma	18.874	1.835	12.221
Carlos Arvelo	49.602	5.996	38.703
Diego Ibarra	39.807	2.951	26.114
Guacara	67.956	3.100	47.273
Juan José Mora	23.705	2.405	16.760
Libertador	59.787	5.619	42.998
Los Guayos	59.052	2.598	37.471
Miranda	11.337	1.352	7.594
Montalbán	9.857	796	6.264
Naguanagua	67.452	2.594	44.186
Puerto Cabello	69.892	4.027	46.074
San Diego	42.620	975	24.684
San Joaquín	25.210	1.348	16.817
Valencia	334.491	17.091	221.464

Fuente: INE, Datos Censo 2011

Modelo a Estimar:

Para los catorce municipios del estado Carabobo se formulará un modelo que explique el empleo municipal en función del nivel educativo de cada municipio y el stock o inventario de capital en cada uno de ellos.

$$y = b_0 + b_1 X_1 + b_2 X_2 + e \qquad (1)$$

Donde y es la variable endogena y representa el empleo en cada municipio, X_1 es la variable para explicar el factor educación a nivel municipal y X_2 la variable explicativa del stock de capital.

Se estimará un modelo de empleo con dos variables independientes o exogenas y una variable dependiente o endogena. El modelo tiene la siguiente forma:

$$\hat{y} = b_0 + b_1 X_1 + b_2 X_2 \qquad (2)$$

Con y: número de ocupados para cada municipio.
X_1: Población analfabeta mayor de 10 años en cada municipio.
X_2: Total de Viviendas en los catorces municipios del estado Carabobo.

Estas variables representan el empleo municipal, el nivel educativo y el stock de capital respectivamente. Se toma el total de viviendas en cada municipio como una variable proxi del stock de capital, al no tener directamente esta variable, es decir no hay un registro a nivel municipal para la variable capital.

Solución Matricial del Modelo:

Para resolver el modelo y estimar los coeficiente b_0, b_1 y b_2 debemos plantear las diferentes matrices que nos permitan realizar las operaciones matriciales y obtener los diferentes estadísticos del modelo.

La matriz X tiene 3 columnas, la primera con unos (1) para el término independiente, la segunda columna contiene los valores la variable que representa la población analfabeta y la tercera columna tiene los valores para la variable total de viviendas:

$$X = \begin{pmatrix} 1 & 1.835 & 12.221 \\ 1 & 5.996 & 38.703 \\ 1 & 2.951 & 26.114 \\ 1 & 3.100 & 47.273 \\ 1 & 2.405 & 16.760 \\ 1 & 5.619 & 42.998 \\ 1 & 2.598 & 37.471 \\ 1 & 1.352 & 7.594 \\ 1 & 796 & 6.264 \\ 1 & 2.594 & 44.186 \\ 1 & 4.027 & 46.074 \\ 1 & 975 & 24.684 \\ 1 & 1.348 & 16.817 \\ 1 & 17.091 & 221.464 \end{pmatrix}_{(14x3)}$$

La matriz X tiene orden o tamaño (14x3), es decir 14 filas y tres columnas.

Luego transponemos la matriz X para obtener la matriz X', donde las columnas de X pasan a ser las filas de X' y las filas de X las columnas de X'. Esta matriz X' tiene orden (3x14)

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1.835 & 5.996 & 2.951 & 3.100 & 2.405 & 5.619 & 2.598 & 1.352 & 796 & 2.594 & 4.027 & 975 & 1.348 & 17.091 \\ 12.221 & 38.703 & 26.114 & 47.273 & 16.760 & 42.998 & 37.471 & 7.594 & 6.264 & 44.186 & 46.074 & 24.684 & 16.817 & 221.464 \end{pmatrix}_{(3x14)}$$

Al multiplicar la matriz X por la matriz X' se obtiene la matriz X'X, que es una matriz cuadrada de orden (3x3)

$$X'X = \begin{bmatrix} 14 & 52.687 & 588.623 \\ 52.687 & 422.021.527 & 5.004.549.811 \\ 588.623 & 5.004.549.811 & 62.208.291.265 \end{bmatrix}_{(3x3)}$$

Al ser la matriz X'X una matriz invertible, podemos obtener su matriz inversa $(X'X)^{-1}$, que es también una matriz cuadrada de orden (3x3):

$$(X'X)^{-1} = \begin{bmatrix} 0{,}1435698333062770000 & -0{,}0000394403732687499 & 0{,}0000018144334831000 \\ -0{,}0000394403732687500 & 0{,}0000000623436490702 & -0{,}0000000046422491354 \\ 0{,}0000018144334831000 & -0{,}0000000046422491354 & 0{,}0000000003723675620 \end{bmatrix}_{(3x3)}$$

Ahora planteamos el vector columna o matriz **y**, que contiene los valores u observaciones de la variable endogena o dependiente, en este caso los datos correspondientes al número de ocupados por cada municipio del estado Carabobo, este vector es de orden (14x1):

$$y = \begin{pmatrix} 18.874 \\ 49.602 \\ 39.807 \\ 67.956 \\ 23.705 \\ 59.787 \\ 59.052 \\ 11.337 \\ 9.857 \\ 67.452 \\ 69.892 \\ 42.620 \\ 25.210 \\ 334.491 \end{pmatrix}_{(14x1)}$$

Luego multiplicamos la matriz X' de orden (3x14) por el vector y de orden (14x1) para obtener el vector X'y de orden (3x1):

$$X'y = \begin{bmatrix} 879.642 \\ 7.478.474.842 \\ 93.485.343.044 \end{bmatrix}_{(3x1)}$$

Finalmente multiplicamos la matriz inversa $(X'X)^{-1}$ de orden (3x3) por el vector X'y de orden (3x1) para obtener el vector de coeficientes **b**, que es un vector de orden (3x1) que contiene, para este ejemplo, los tres coeficiente estimados mediante el método de los mínimos cuadrados ordinarios:

$$(X'X)^{-1}(X'y) = b = \begin{bmatrix} 959,152657885 \\ -2,4402506071 \\ 1,6900177984 \end{bmatrix}_{(3x1)}$$

Validación Económica del Modelo Estimado:

El modelo estimado se expresa de acuerdo a la siguiente función de regresión:

$$\hat{y} = 959,1526 - 2,44025\,X_1 + 1,690017\,X_2 \qquad (3)$$

La interpretación de los distintos coeficientes la podemos hacer de la siguiente forma:
b_0 = 959,1526 es el término independiente o el valor de la variable endógena cuando las variables explicativas del modelo toman valor cero en forma simultanea o a la vez. Para este modelo en particular, sería la cantidad de ocupados (en promedio) en cada municipio que no depende ni de el número de analfabetos ni del total de viviendas, el signo esperado para este coeficiente es positivo.

b_1 = -2,44025 es el coeficiente de la variable X_1: representa esta variable al número de analfabetos mayor a 10 años en cada municipio, la interpretación del valor estimado para este coeficiente b_1 es que si el número de viviendas permanece constante o no cambia, un aumento o disminución en una unidad en el número de personas analfabetas, hará disminuir o aumentar el promedio de ocupados por municipio en 2,44 personas. Es decir que una disminución en la población analfabeta, hará aumentar el número de ocupados en promedio en cada municipio, por lo que el signo esperado para esta variable es negativo. En este caso el signo que tiene el valor estimado, está acorde con el signo esperado.

b_2 = 1,690017 es el coeficiente estimado para la variable X_2 que representa al total de viviendas existente en cada municipio del estado Carabobo, el valor estimado para este coeficiente b_2 se puede interpretar en que si el número de personas analfabetas mayor de 10 años en cada municipio permanece constante o no cambia, un aumento o disminución en una unidad en el total de viviendas, hará aumentar o disminuir el promedio de ocupados por municipio en 1,69 personas. Es decir que un aumento en el total de viviendas en cada uno de los municipios del estado Carabobo, hará aumentar el número de ocupados en promedio en cada municipio, por lo que se espera un signo positivo para esta variable. En este caso el signo que tiene el valor estimado (positivo), está acorde con el signo esperado.

Validación Estadística:

Para la validación estadística del modelo estimado, se realizarán las pruebas de hipótesis que permitan determinar si los coeficientes estimados del modelo son o no son significativos, es decir si los valores t calculados para cada coeficiente caen en la región crítica o de rechazo de la hipótesis nula de que los verdaderos valores poblacionales son cero: $\beta_i = 0$. En esta etapa también se determinará la bondad de ajuste del modelo, mediante el calculo del coeficiente de determinación R^2.

Como primer paso se procede al calculo de la matriz de varianzas y covarianzas, pues esta matriz contiene en su diagonal principal las varianzas de las estimaciones hechas mediante el método de los mínimos cuadrados ordinarios que luego nos permitirán obtener los correspondientes errores estándar de los coeficientes estimados.

Para calcular la matriz de varianzas y covarianzas obtenemos primero la varianza de los errores, que en este caso se estima, al no poder conocer el verdadero valor de σ^2, un estimador insesgado de la varianza es:

$$\hat{\sigma}^2 = S_e^2 = \frac{e'e}{n-k} = \frac{y'y - b'X'y}{n-k}$$

Es decir, se calcula e'e y se divide entre los grados de libertad del modelo. Para obtener e'e realizamos primero el cálculo de y'y, luego calculamos b'X'y para realizar la resta y'y – b'X'y.

Calculo de y'y:

$$y'y = \begin{pmatrix} 18.874 & 49.602 & 39.807 & 67.956 & 23.705 & 59.787 & 59.052 & 11.337 & 9.857 & 67.452 & 69.892 & 42.620 & 25.210 & 334.491 \end{pmatrix}_{(1x14)} \begin{pmatrix} 18.874 \\ 49.602 \\ 39.807 \\ 67.956 \\ 23.705 \\ 59.787 \\ 59.052 \\ 11.337 \\ 9.857 \\ 67.452 \\ 69.892 \\ 42.620 \\ 25.210 \\ 334.491 \end{pmatrix}_{(14x1)}$$

y'y = 140.639.342.130

Calculo de b'X'y:

$$b'X'y = \begin{pmatrix} 959{,}152657885 & -2{,}4402506071 & 1{,}6900177984 \end{pmatrix}_{(1x3)} \begin{pmatrix} 879{,}642 \\ 7.478.474.842 \\ 93.485.343.044 \end{pmatrix}_{(3x1)}$$

b' X'y = 140.586.251.819,37

Ahora procedemos a calcular e'e:

e'e = 140.639.342.130 - 140.586.251.819,37

e'e = 53.090.310,63434

e'e es también la suma de los cuadrados de los residuos SCR, es decir:

SCR = 53.090.310,63434

Como n es el tamaño de la muestra o el número de observaciones y el estado Carabobo tiene 14 municipios, entonces:

n = 14.

El número de coeficientes a estimar en el modelo es igual a 3, entonces:

k = 3.

Y los grados de libertad:

n – k = 14 – 3 = 11.

Procedemos ahora a calcular la varianza estimada de los errores:

$$S_e^2 = \frac{e'e}{n-k} = \frac{53.090.310,63434}{11}$$

$$\boldsymbol{S_e^2 = 4.826.391,87585}$$

Una vez calculada la varianza, realizamos la multiplicación de un escalar por una matriz, es decir la varianza por la matriz inversa:

$$\boldsymbol{Var-Cov} = 4.826.391,87585 \begin{pmatrix} 0,14356983330627700 & -0,00003944037326874990 & 0,000001814433483100 \\ -0,000039440373268750 & 0,0000000623436490702 & -0,0000000046422491354 \\ 0,0000018144334831 & -0,00000000464224913540 & -0,0000000003723675620 \end{pmatrix}_{(3x3)}$$

Al realizar la multiplicación obtenemos la matriz de varianzas y covarianzas:

$$\boldsymbol{Var-Cov} = \begin{pmatrix} \mathbf{692.924,27709} & -190,3546971 & 8,75716702 \\ -190,3546971 & \mathbf{0,3008949} & -0,02240531 \\ 8,75716702 & -0,02240531 & \mathbf{0,00179719} \end{pmatrix}_{(3x3)}$$

En la diagonal principal de la matriz de varianzas y covarianzas tenemos las varianzas de las estimaciones de cada uno de los coeficientes de la función de regresión, la raíz cuadrada de estas varianzas son los errores estándar de los coeficientes:

$$S(b_0) = \sqrt{692.927,27709} = 832,4207332151$$

$$S(b_1) = \sqrt{0,3008949} = 0,5485388604$$

$$S(b_2) = \sqrt{0,00179719} = 0,0423932987$$

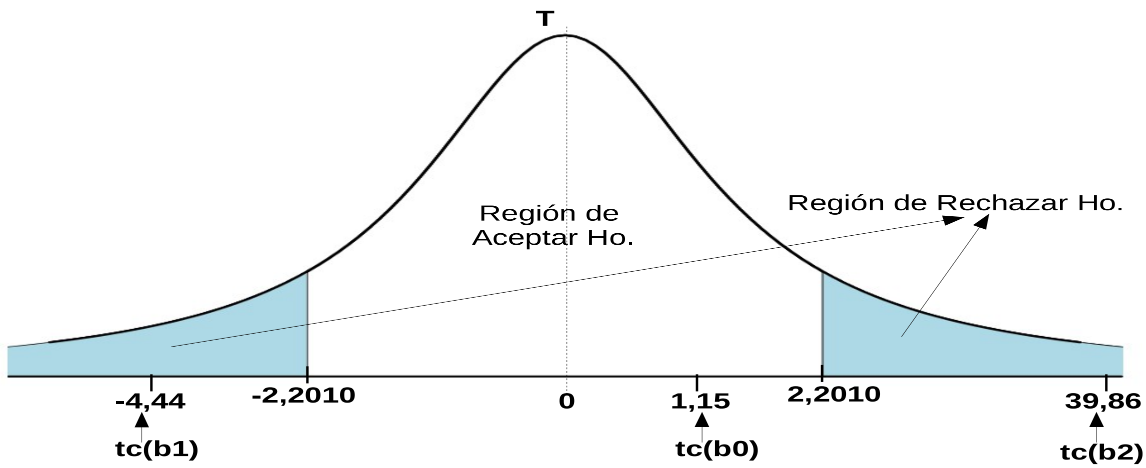
El siguiente paso es encontrar los valores t calculados, y luego realizar las correspondientes pruebas de hipótesis de significancia de las estimaciones realizadas para los parámetros del modelo.

$$t_c(b_0) = \frac{959{,}1526}{832{,}4207332151} = 1{,}1522$$

$$t_c(b_1) = \frac{-2{,}44025}{0{,}5485388604} = -4{,}4486$$

$$t_c(b_2) = \frac{1{,}690017}{0{,}0423932987} = 39{,}8652$$

El valor crítico en la tabla t, para 11 grados de libertad y con una significancia del 5% (dos colas) es: 2,201



En el gráfico se observa que tanto b1 como b2 son significativos, es decir la muestra nos da una confianza del 95% para aceptar que los coeficientes son distintos de cero. No así para el término independiente o coeficiente b0.

Bondad de Ajuste:

Para la bondad de ajuste calculamos R^2 y el R^2 modificado o ajustado.

Para el cálculo de R^2 debemos calcular SCT, para esto utilizamos la siguiente formula:

$$SCT = y'y - n(\bar{y})^2$$

Calculamos la media de y:

$$\bar{y} = \frac{\sum y}{n} = \frac{879.642}{14} = 62.831{,}57$$

Elevamos la media de y al cuadrado:

$$(\bar{y})^2 = (62.831{,}57)^2 = 3.947.806.368{,}18$$

Multiplicamos por n = 14:

$$n(\bar{y})^2 = 14*(3.947.806.368,18) = 55.269.289.154,57$$

Como y'y = 140.639.342.130 podemos aplicar la formula y calcular SCT:

$$SCT = 140.639.342.130 - 55.269.289.154,57 = 85.370.052.975,42$$

De SCT = SCE + SCR despejamos SCE:

SCE = SCT – SCR

Si SCR = 53.090.310,63 entonces:

$$SCE = 85.370.052.975,42 - 53.090.310,63 = 85.316.962.664,79$$

Ahora procedemos a calcular el R^2

$$R^2 = \frac{85.316.962.664,79}{85.370.052.975,42} = 0,9993 = 99,93$$

Que se interpreta como que el 99,93% de la variación en la variable endógena y, es explicada por la regresión en un 99,93%.

El R^2 ajustado se calcula con la siguiente fórmula:

$$\bar{R}^2 = 1 - (1 - R^2)\left(\frac{n - 1}{n - k}\right) = 0,9992 = 99,92$$

Predicción o Estimación:

Para predicción o estimación usamos la siguiente fórmula:

$$\hat{y}_0 / x_0 = x'_0 b$$

Donde x_0 es el vector columna con los valores de las variables explicativas para un municipio y b es el vector columna con los coeficientes estimados:

Ejemplo: Calcule el número de empleados para el municipio San Diego si las personas analfabetas de este municipio disminuyeron a 970 personas, y el numero total de viviendas aumentó a 24.952 viviendas.

$$\hat{y}_0 = \begin{pmatrix} 1 & 970 & 24.952 \end{pmatrix}_{(1x3)} \begin{pmatrix} 959,152657885 \\ -2,4402506071 \\ 1,6900177984 \end{pmatrix}_{(3x1)}$$

$$\hat{y}_0 = 40.761,43367$$

Intervalo de confianza del 95% para la predicción puntual o individual: