

II Jornadas Internacionales de Estadística Aplicada 5 y 6 de diciembre de 2019 Regresión lineal Múltiple

Autor: Sarapura Pablo Rodolfo

Institución: Facultad de ingeniería, Universidad Nacional de Jujuy

Datos de contacto: rapid12300@gmail.com 3884171107

RESUMEN

En el presente informe se buscará ver si existe una relación del desempleo y el precio del dólar con los ingresos y gastos de la Administración Nacional (que es el ámbito del Presupuesto Nacional que vota todos los años el Congreso de la Nación), para esto se utilizara la técnica estadística de la regresión lineal múltiple.

Palabras claves: DESOCUPACION-DÓLAR-GASTO-INGRESO

INTRODUCCION

En estos tiempos observamos que se habla mucho acerca del desempleo, y del precio del dólar que se ha elevado. Lo que queremos comprobar es si estos tomados como variables tienen alguna relación.

La recopilación de datos fue hecha de diferentes bases de datos.

METODOLOGIA

La técnica estadística utilizada para hacer este informe es la regresión lineal múltiple, este método nos permite ver si las variables están relacionadas, y si existe una relación poder predecir valores futuros.

INTRODUCCION TEORICA

La regresión lineal es una técnica estadística destinada a analizar las causas de por qué pasan las cosas. A partir de los análisis de regresión lineal múltiple podemos:

- Identificar que variables independientes (causas) explican una variable dependiente (resultado)
- Comparar y comprobar modelos causales
- Predecir valores de una variable, es decir, a partir de unas características predecir de forma aproximada un comportamiento o estado

El modelo que queremos llegar es:

$$y = a + b_1X_1 + b_2X_2$$

Dónde:

- **Y**=Variable dependiente
- **a**=Intercepto de la variable Y

- X_1, X_2 = valores de las variables independientes
- b_1, b_2 = Pendientes asociadas con cada variable.

Para llegar al modelo nos apoyaremos en este sistema de ecuaciones.

$$\begin{aligned}\sum Y &= na + b_1 \sum X_1 + b_2 \sum X_2 \\ \sum Y * X_1 &= a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_2 * X_1 \\ \sum Y * X_2 &= a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2\end{aligned}$$

- **Error estándar de la estimación múltiple:** El error estándar de la regresión es el valor que muestra la diferencia entre los valores reales y los estimados de una regresión.

$$Se = \sqrt{\frac{\sum Y^2 - a \sum Y - b_1 \sum X_1 Y - b_2 \sum X_2 Y}{N - 3}}$$

- **Coeficiente de determinación:** se denomina coeficiente de determinación y es una medida de la proporción de la variabilidad explicada por el modelo ajustado

$$R^2 = \frac{a \sum Y + b_1 \sum X_1 Y + b_2 \sum X_2 Y - n \bar{Y}^2}{\sum Y^2 - n \bar{Y}^2}$$

- **Coeficiente de correlación:** La correlación determinará la calidad del ajuste. Si r es cercano o igual a 1, el ajuste será bueno y las predicciones realizadas a partir del modelo obtenido serán muy fiables

$$R = \sqrt{\frac{a \sum Y + b_1 \sum X_1 Y + b_2 \sum X_2 Y - n \bar{Y}^2}{\sum Y^2 - n \bar{Y}^2}}$$

- **Análisis de Residuos** Se define un residuo como la diferencia entre el valor observado y el valor estimado

DESARROLLO

Empezamos armando la tabla

Y=Porcentaje de desocupación

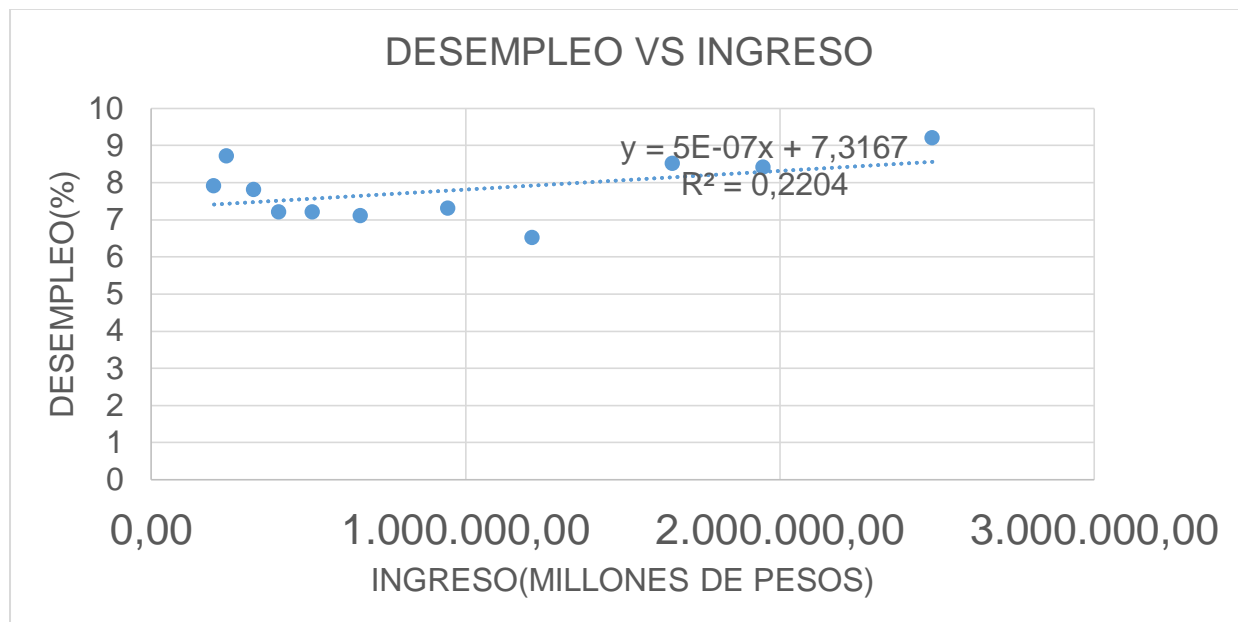
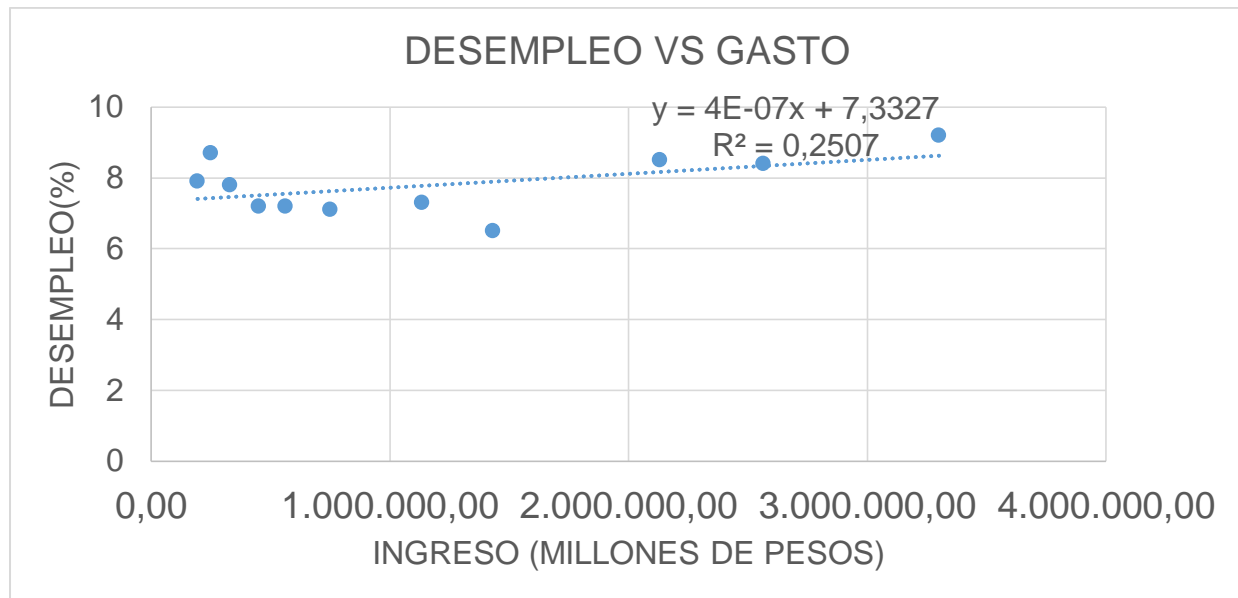
X_1 : Gasto del estado en millones de pesos

X_2 : Ingreso del estado en millones de pesos

AÑO	Y	X_1	X_2
2008	7,9	192.974,10	200.098,80
2009	8,7	249.914,10	240.485,40
2010	7,8	329.034,20	327.154
2011	7,2	450.548,20	407.334,60
2012	7,2	561.833	512.760,60
2013	7,1	750.934,90	665.876,80
2014	7,3	1.134.364	944.178

2015	6,5	1.431.662	1.212.308
2016	8,5	2.131.113,30	1.659.383,40
2017	8,4	2.564.004,60	1.947.405,40
2018	9,2	3.298.979,50	2.485.244,10

Para asegurarnos que ambas variables independientes expliquen en significancia a la variable dependiente hacemos un diagrama de dispersión y observamos el coeficiente de determinación.



Observando ambas gráficas y viendo que su coeficiente de determinación es de un 22% y un 25% bastante elevado veo que no puedo despreciar a ninguna variable independiente.

Tabla de datos necesarios para el sistema de ecuación.

ΣY	85,8	\bar{Y}	7,8
ΣX_1 GASTO	13.095.362,20	\bar{Y}^2	60,84
ΣX_2 INGRESO	10.602.229,10	ΣY^2	676,020
ΣX_1^2	26.625.980.524.365,900		
ΣX_2^2	16.160.681.398.415,900		
$\Sigma X_1 \cdot X_2$	20.712.916.172.394,300		
$\Sigma X_1 \cdot Y$	106.475.372,650		
$\Sigma X_2 \cdot Y$	85.676.926,800		
N	11		

- Formando el sistema de ecuaciones y resolviendo por el método de determinantes obtenemos el siguiente modelo:

$$Y=9,140+0,000015GASTO-0,00002INGRESO$$

Observando la siguiente ecuación puedo deducir que a más ingreso menos desempleo habrá, como así también a más gasto más desempleo.

- Calculando el error estándar de estimación múltiple:

$$Se= 0,38499098$$

Este error es la dispersión de los valores, es decir que tan dispersos están del valor real.

- Calculo del coeficiente de determinación:

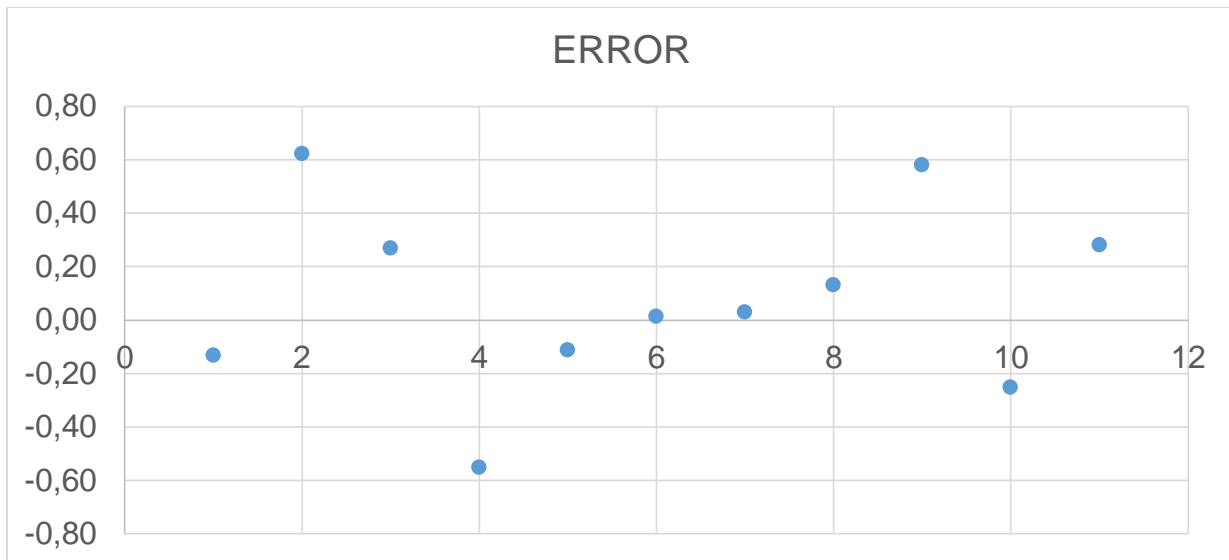
$$R^2= 0,82511144$$

El coeficiente de determinación indica el porcentaje que las variables independientes explica a la dependiente en este caso es muy elevado un 82,5%. Podemos decir que la ecuación calculada es muy útil.

- Calculo del coeficiente de correlación:

$$R= 0,90835645$$

Nos indica el grado de relación que hay entre las variables independientes e independientes. En este caso es de 0,91



Podemos observar que el modelo no es tan exacto ya que hay puntos que se van del intervalo del **Se** (Error estándar de estimación múltiple)

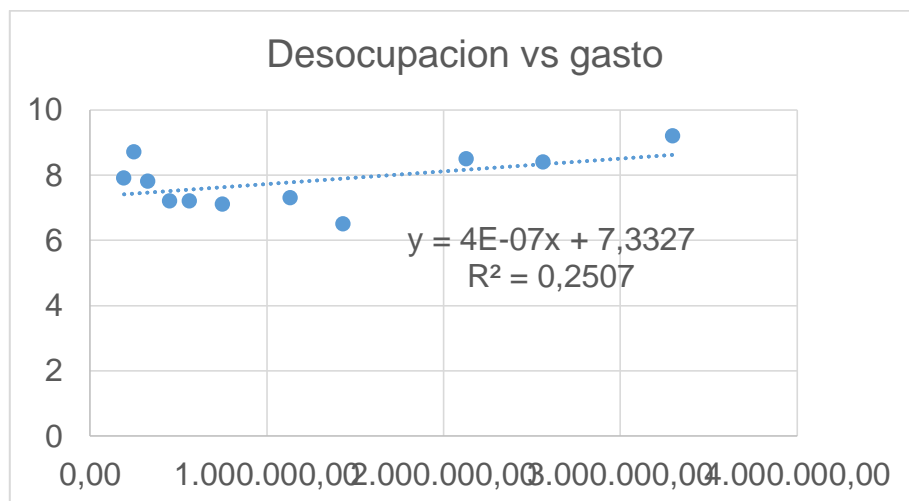
CONCLUSION

Podemos observar que, si hay una relación del ingreso y el gasto del estado con la desocupación, pero el modelo que obtuvimos no es tan explicativo, ya que hay ciertos valores que se van del error estándar de estimación por lo cual para valores futuros no podemos estar tan seguros si es viable el resultado.

Como segunda parte armando una tabla combinando el GASTO-PRECIO DEL DÓLAR-DESOCUPACION obtengo:

AÑO	Y(GASTO)	X1(DESOCUPACION)	X2(PRECIO DÓLAR)
2008	192.974,10	7,9	3,45
2009	249.914,10	8,7	3,80
2010	329.034,20	7,8	3,96
2011	450.548,20	7,2	4,25
2012	561.833	7,2	4,79
2013	750.934,90	7,1	6,20
2014	1.134.364	7,3	8,50
2015	1.431.662	6,5	13,12
2016	2.131.113,30	8,5	15,40
2017	2.564.004,60	8,4	17,16
2018	3.298.979,50	9,2	38,14

Observando las gráficas de dispersión de cada variable independiente con la dependiente podemos observar que influyen en un 25% y un 87% a la variable dependiente por lo cual no vamos a despreciar ninguna.



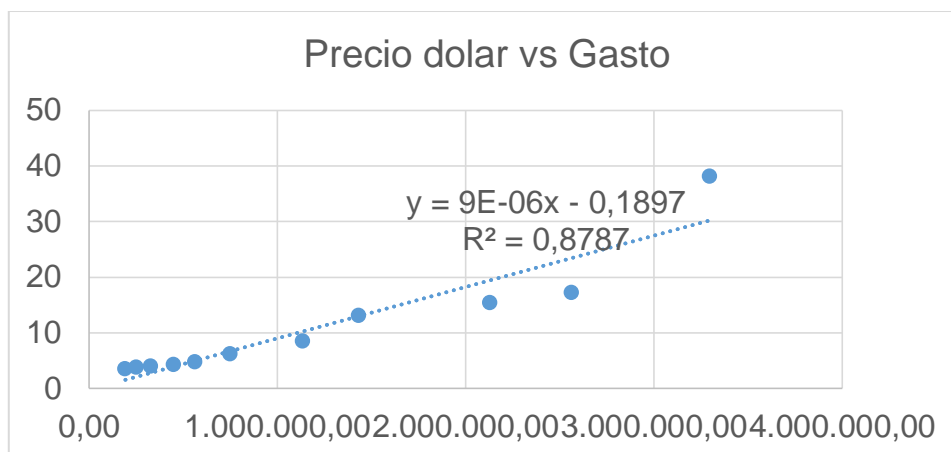


Tabla de los datos necesarios para armar el sistema de ecuación

ΣY	13.095.362,20	\bar{Y}	1190487,473
ΣX_1 DESOCUPACION	85,80	\bar{Y}^2	1,41726E+12
ΣX_2 PRECIO DOLAR	118,77	ΣY^2	26.625.980.524.365,90
ΣX_1^2	676,02		
ΣX_2^2	2.352,14		
$\Sigma X_1 \cdot X_2$	973,573		
$\Sigma X_1 \cdot Y$	106.475.372,65		
$\Sigma X_2 \cdot Y$	243.246.260,30		
N	11		

- Formando el sistema de ecuaciones y resolviendo por el método de determinantes obtenemos el siguiente modelo

$$Y = 410653,9455 - 33887X_1(\text{DESOCUPACION}) + 96705X_2(\text{DÓLAR})$$

De acá podemos decir que habiendo menos desocupación va a haber menos gasto, como así también cuando aumenta el precio del dólar significa que hay mas gasto.

- Calculando el error estándar de estimación múltiple

$$Se = 408236,4794$$

Este nos indica el error medio que puede haber acerca del valor real del gasto obtenido.

- Calculo del coeficiente de determinación

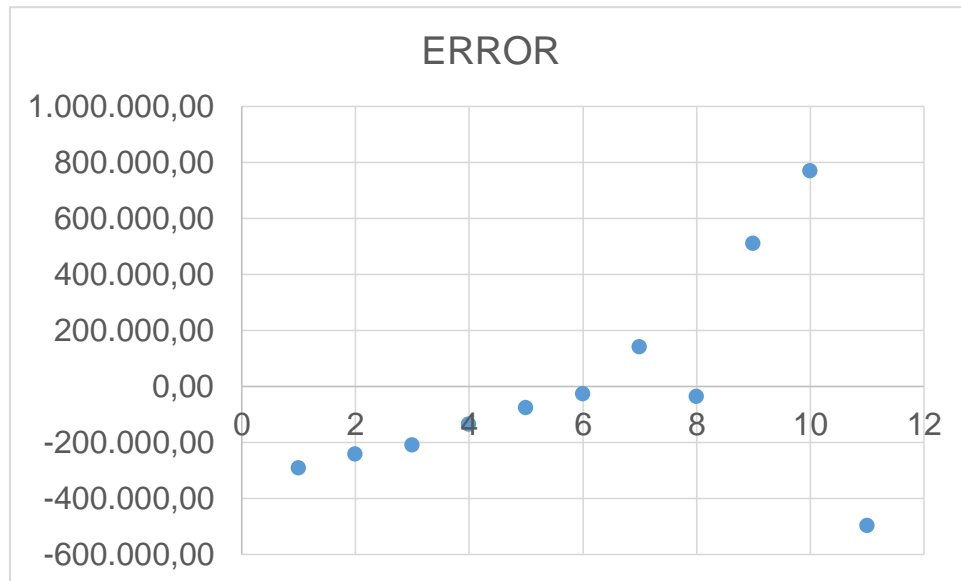
$$R^2 = 0,87919154$$

Las variables independientes explican 88% a la variable dependiente

- Cálculo del coeficiente de correlación

$$R = 0,937652141$$

El grado de relación que hay entre las variables es alto de un 0,94



Viendo los puntos del error observamos que hay 3 que se van del intervalo del Error de estimación múltiple.

CONCLUSION

Si existe relación entre el costo del dólar, la desocupación y el gasto del estado.

Como podemos ver en la gráfica de error parecería que el error se va incrementando en cada observación, por lo cual el modelo que obtuvimos no es confiable.

BIBLIOGRAFIA

- Probabilidad y estadística para ingeniería y ciencias, 9ª Edición. Ronald E. Walpole, Raymond H. Myers, Sharon I. Myers y Keying Ye. Ed. Pearson educación
- Bases de datos:

<https://www.presupuestoabierto.gob.ar/sici/destacado-PBI#>

<https://knoema.es/atlas/Argentina/Tasa-de-desempleo>

- Para el precio del dólar herramienta de google

