

Unidad 8: Modelamiento Estadístico1: Regresión Lineal

Modelamiento Lineal

Usted está realizando una consultoría para un organismo internacional para explicar los ingresos económicos que cada país recibe por cooperación internacional. Para ello usted, mediante criterios teóricos, selecciona algunas variables que buscan esclarecer las diferencias entre los países, una primera variable que reconoce es la cantidad de exportaciones, una segunda variable que toma en cuenta son la cantidades de importaciones, una tercera variable es el ingreso PBI per cápita y, finalmente, usted considera como tercer factor una variable dicotómica que es el régimen del gobierno, y sus dos categorías son autoritarismo y democracia.

En estos casos es necesario realizar un modelo de regresión lineal, que sería usado para explicar la relación entre una variable dependiente (cooperación internacional) y las variables independientes (Importación, exportación, PBI per cápita y régimen). El modelo de Regresión Lineal nos permite determinar el efecto de las variables independientes sobre el modelo y predecir valores de respuesta a partir de estos resultados. El *Modelo de Regresión Lineal* tiene como requisito que la variable dependiente sea numérica continua; mientras que las independientes pueden ser numéricas, ordinal o categórica. A continuación, veamos cómo funciona este modelo con la base de datos UE, para lo cual armaremos nuestra ecuación con las siguientes variables: *GDP Deflator*, como variable dependiente y *Government Bonds*, *Consumer Prices*, *Total Reserve Minus Gold*, como variables independientes.

Abriendo base de datos de indicadores financieros de la Unión Europea. Se utiliza el paquete *foreign* para exportar data de SPSS.

```
library(foreign)
```

```
data2 <- read.spss("UE.sav", to.data.frame=TRUE, use.value.labels = TRUE)
```

Existen dos formas de realizar el modelo de regresión lineal, podemos trabajar con el comando *glm* o *lm*. Cada uno de estos comandos tienen características distintas pero los resultados serán los mismos.

Probemos con la primera forma de solicitar el modelo. La familia nos indica el tipo de modelo a realizarse, en este caso solicitamos el gaussiano que hace referencia a un modelo con una variable dependiente numérica continua. Algunas particularidades de este comando es que nos brinda el valor de *AIC* (*Akaike Information Criterion*), que es un valor que se utiliza para comparar modelos. En ambas formas tenemos que observar los coeficientes e interceptos para generar la ecuación de la regresión.

```
modelo<-glm(GDPDeflator ~ Governmentbonds + ConsumerPrices + TotalReservesMinusGold,  
family = gaussian, data = data2) #creamos el objeto modelo, para lo cual planteamos la  
fórmula de la ecuación de la regresión lineal
```

```
summary(modelo) #pedimos un resumen
```

Call:

```
glm(formula = GDPDeflator ~ Governmentbonds + ConsumerPrices +  
TotalReservesMinusGold, family = gaussian, data = data2)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.83403	-0.67586	-0.05295	0.58537	1.83679

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.131e+01	2.482e+01	-1.262	0.2232
Governmentbonds	-2.403e-01	1.116e-01	-2.153	0.0451 *
ConsumerPrices	1.306e+00	2.427e-01	5.381	4.1e-05 ***
TotalReservesMinusGold	-2.243e-05	1.234e-05	-1.817	0.0859 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.020216)

Null deviance: 48.297 on 21 degrees of freedom

Residual deviance: 18.364 on 18 degrees of freedom

(8 observations deleted due to missingness)

AIC : 68.459

Number of Fisher Scoring iterations: 2

Probemos con la segunda forma de solicitar el modelo. La particularidad de esta forma, radica en que nos brinda un resumen del modelo con R², R²ajustado y el Error Estándar. Por eso, observamos el R² para ver en qué porcentaje las independientes están explicando la dependiente

```
modelo2<-lm(GDPDeflator ~ Governmentbonds + ConsumerPrices + TotalReservesMinusGold, data2)
```

```
summary(modelo2)
```

Call:

```
lm(formula = GDPDeflator ~ Governmentbonds + ConsumerPrices +  
    TotalReservesMinusGold, data = data2)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.83403	-0.67586	-0.05295	0.58537	1.83679

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.131e+01	2.482e+01	-1.262	0.2232
Governmentbonds	-2.403e-01	1.116e-01	-2.153	0.0451 *
ConsumerPrices	1.306e+00	2.427e-01	5.381	4.1e-05 ***
TotalReservesMinusGold	-2.243e-05	1.234e-05	-1.817	0.0859 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.01 on 18 degrees of freedom

(8 observations deleted due to missingness)

Multiple R-squared: 0.6198, Adjusted R-squared: 0.5564

F-statistic: 9.78 on 3 and 18 DF, p-value: 0.0004746

Ahora es momento de interpretar los resultados para conocer si el modelo es significativo. Se debe observar la significancia de cada una de las variables independientes; es necesario que al menos una de las variables haya pasado esta prueba. Para esto observamos la última columna: $Pr(>|t|)$ -> *p-value for the t-test*, tenemos que esperar que el valor de cada variable independiente no supere el 0,05 para que se acepte la Hipótesis de que si tiene efecto causal sobre la dependiente. Asimismo, observamos que es posible llegar a estas conclusiones con los asteriscos (*) que se encuentran al costado de dichos números. Se espera que la variable independiente tenga al menos un asterisco. Para ello observamos la columna del t value de cada variable, y observamos las cifras, independientemente del signo. En ese sentido, podemos ver que la variable *ConsumerPrices* tiene más efecto que *Governmentbonds* sobre la variable dependiente. Nótese que se omitió *TotalReservesMinusGold*, ya que no pasó la prueba.

En este caso, observamos que hay dos variables que han pasado la prueba y tiene efecto sobre la dependiente *GDP deflator*: *Governmentbonds GovBo* y *Consumer Prices ConsPric*. Ahora

tenemos que plantear la ecuación. Para esto se observa la columna de *Estimate*, donde se encuentran todos los coeficientes de las dependientes.

$$\text{GDP} = -3.131\text{e}+01 -2.403\text{e}-01 (\text{GovBo}) + 1.306\text{e}+00 (\text{ConsPric})$$

Desagreguemos el decimal y la ecuación final sería la siguiente:

$$\text{GDP} = -31,307 - 0,2403(\text{GovBo}) + 1.306(\text{ConsPric})$$

En esta ecuación, también podemos observar qué sentido tiene la relación de cada variable significativa con la dependiente. La variable *GovBo* tiene un signo negativo, por lo tanto la relación es inversa; mientras que con la variable (*ConsPric*) la relación es directa, signo positivo.

Para determinar la capacidad explicativa del modelo, como lo mencionamos líneas arriba, tenemos que fijarnos en el R² (medida que nos brinda el modelo²) para ver cuánto está explicando mi modelo. En este caso el R² calculado es 0,61 esto significa que las independientes explican un 61%. También se debe observar la intensidad de cada una de las relaciones significativas.