Regresión lineal simple

Laboratorio FINTRADE

7/7/2020

## Regresiones lineales simples

Las regresiones lineales comunmente son usadas para variables dependientes continuas. Aquí la función es lineal, es decir, requiere la determinación de dos parámetros: la pendiente y la ordenada en el origen.

## Aplicación en RStudio

La base de datos con la que se trabajará será “Boston” esta base recoge la mediana del valor de viviendasen 506 áreas residenciales de Boston, teniendo un total de 13 variables, para visualizar esta base de datos una vez instalada la librería se debe escribir en la consola View(Boston), y para obtener más información acerca de la misma, en la ventana de **help** escribir el nombre de la base.

Es necesario instalar el paquete “MASS” que contiene a la librería “Boston”

library(MASS)  
data("Boston")  
attach(Boston)

Con esta regresión se pretende predecir el valor de la vivienda en función de la tasa impositiva de las propiedas en Boston. Para esto, se emplea la función lm() para generar un modelo lineal.

reg1<-lm(medv~tax)

La función lm() genera un objeto que almacena toda la información del modelo, y para visualizar los principales parámetros del modelo generado se utiliza summary().

summary(reg1)

##   
## Call:  
## lm(formula = medv ~ tax)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14.091 -5.173 -2.085 3.158 34.058   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 32.970654 0.948296 34.77 <2e-16 \*\*\*  
## tax -0.025568 0.002147 -11.91 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.133 on 504 degrees of freedom  
## Multiple R-squared: 0.2195, Adjusted R-squared: 0.218   
## F-statistic: 141.8 on 1 and 504 DF, p-value: < 2.2e-16

Se observa que la variable “tax” está significativamnte relacionada con la variable respuesta. La evaluación del modelo en conjunto puede hacerse a partir de los valores RSE o del valor arrojado por el summary.

* Residual standar error (RSE): En promedio, cualquier predicción del modelo se aleja 8.133 unidades del verdadero valor.
* : El predictor **tax** empleado en el modelo es capaz de explicar el 54.44% de la variabilidad observada en el precio de las viviendas.

Los dos coeficientes de regresión ( y ) estimados por el modelo son significativos y se pueden interpretar como:

* Intercept(): El valor promedio del precio de la vivienda cuando el tax es 0, es de 32.97 unidades.
* Predictor tax (): por cada unidad que se incrementa el predictor tax el precio de la vivienda disminuye en promedio 0.025 unidades.

La estimación de todo coeficiente de regresión tiene asociada un error estándar, por lo tanto todo coeficiente de regresión tiene su correspondiente intervalo de confianza.

confint(reg1, level = 0.95)

## 2.5 % 97.5 %  
## (Intercept) 31.10755524 34.83375374  
## tax -0.02978712 -0.02134908

Dado que el p-value del predictor **tax** ha resultado significativo a un 0.05%, su intervalo de confianza del 95% no contiene el valor de 0.