

PRÁCTICA 7. Regresión lineal

Estudiantes:

Ejercicio 1

Una empresa fabricante de patinetes eléctricos está interesada en conocer cómo afecta a las ventas del producto, la publicidad en la televisión, en la radio y en los periódicos. Para ello, se recoge el incremento en ventas (en miles de euros) y los gastos mensuales (en miles de euros) de la publicidad en los diferentes medios correspondientes a los últimos 24 meses.

Suponiendo que los datos son una muestra aleatoria simple (m.a.s.) representativa de la población, se desea realizar un análisis de regresión.

Contamos con una muestra de datos que recoge los incrementos en las ventas (en euros) que se hicieron de patinetes en los 24 meses (en general, cuando más tiempo pasa más se incrementan las ventas de patinetes).

(a) Identifique la variable respuesta y las variables independientes consideradas en este estudio:

Variable respuesta:

Variables independientes:

Podemos leer el fichero `patinete.xlsx` disponible en Poliformat

```
library(readxl)
```

```
patinete <- read_xlsx("patinete.xlsx")
```

```
View(patinete)
```

```
attach(patinete)
```

(b) Obtener un diagrama de dispersión para las ventas y cada una de las variables de gasto en publicidad. ¿Hay alguna relación que parezca más lineal? ¿Por qué?

Por ejemplo el diagrama de dispersión para las ventas y la publicidad en televisión es `plot(Pub_tv~Ventas)`, donde `Pub_tv` es la columna del fichero de datos.

(c) Obtenga el coeficiente de correlación lineal. ¿Se confirma la mayor relación lineal entre Ventas y la variable que detectaste en el apartado b)?

Para calcular la correlación entre ventas y la publicidad en televisión es `cor(Pub_tv,Ventas)`

(d) Obtenga el modelo de regresión lineal **simple** entre la variable Ventas y la publicidad seleccionada en el apartado c) e interprete su valor:

NOTA: Utilizar el código siguiente:

```
modelo<-lm( Ventas ~publicidad)
```

```
summary(modelo)
```

Pendiente

Término independiente:

Escribe el modelo de regresión lineal simple obtenido:

Modelo:

(e) A partir del test F (ANOVA) (última línea del modelo obtenido en el apartado anterior).

Hipótesis del test:

H_0 :

H_1 :

NOTA: El resultado del test de significación global está en la última línea del test realizado anteriormente.

Valor del estadístico F :

Grados de libertad del modelo:

Grado de libertad residuales:

Valor p : (con 4 cifras decimales de precisión)

Explique sus conclusiones en el contexto del problema:

Interprete el valor de R^2 en el contexto del problema:

(f) Utiliza los residuos (R los almacena en la variable `modelo$residuals`) para validar el modelo (es decir, para comprobar el cumplimiento de las tres condiciones asumidas en realizar la regresión):

1. *Independencia* (entre las observaciones). Esta condición debe garantizar la forma en que se seleccionan los individuos (utilizando un muestreo probabilístico o un diseño aleatorizado). La muestra es aleatoria, por lo que asumimos que se cumple.

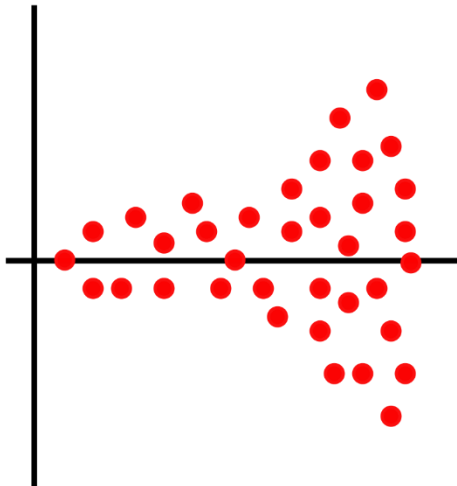
2. *Normalidad*. Compruebe si es aceptable pensar que los residuos proceden de una población normal de media igual a cero.

NOTA: Utilice el siguiente código de R:

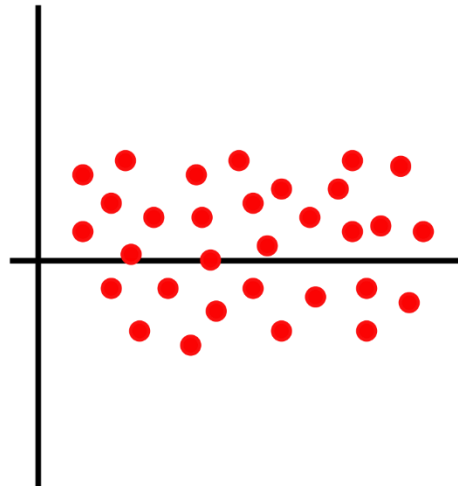
```
summary(modelo$residuals)
boxplot(modelo$residuals, horizontal=TRUE)
hist(modelo$residuals)
qqnorm(modelo$residuals)
```

3. *Homoscedasticidad* (Igualdad de varianzas). Compruebe si la dispersión de los residuos es similar en los todos los valores de la variable ajustada. Con este gráfico se comprueba que la variabilidad es más o menos similar.

```
plot(modelo$residuals~ fitted(modelo))
```



Ejemplo de NO homocedasticidad.



Ejemplo de homocedasticidad.

Después de estos análisis, ¿cree que el modelo de regresión simple es un modelo válido para predecir el incremento en ventas?

Ejercicio 2: Introducción a la regresión lineal múltiple

Obtenga el modelo de regresión lineal múltiple entre la variable Ventas y los gastos en publicidad.

NOTA: Utilizar el código siguiente:

```
modelo2<-lm( Ventas ~ Pub_tv+Pub_radio+Pub_Periodico)
summary(modelo2)
```

(a) ¿Qué variables no son significativas? ¿Por qué?

Elimina (una a una) las variables que no sean significativas.

(b) Escribe el modelo resultante:

El anterior código también te muestra el valor del estadístico F (ANOVA) en R para validar el modelo según el test de significación global.

Hipótesis del test:

H_0 :

H_1 :

Valor del estadístico F :

Grados de libertad del modelo:

Grados de libertad residuales:

Valor p :

Explique sus conclusiones en el contexto del problema:

Interprete el valor de R^2 en el contexto del problema: