Modelos Lineales y diseño de experiementos

Sebastián Alcívar

Regresión lineal

Se realizó la regresión lineal de los valores de Utilidad-Ventas con 40 observaciones. Se obtuvo la siguiente información

Call:

lm(formula = Utilidad ~ Ventas, data = data)

Residuals:

Min 1Q Median 3Q Max

-676.35 -302.04 42.59 303.67 612.49

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 137.08270 282.69543 0.485 0.631

Ventas 0.43994 0.01859 23.663 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 367.4 on 38 degrees of freedom

Multiple R-squared: 0.9364, Adjusted R-squared: 0.9348

F-statistic: 559.9 on 1 and 38 DF, p-value: < 2.2e-16

Si obtenemos el valor del percentil 0.975 de la tabla t:

> qt(0.975 , df =38)

[1] 2.024394

Donde el valor t es menor al percentil obtenido y por lo tanto no se rechaza la hipótesis nula obteniendo que el intercepto no es significativo.

Con esta información, ya que el intercepto no es significativo, se procederá a centrar los datos mediente el siguiente código en R.

utilidad\_c<-data[,"Utilidad"]-mean(data[,"Utilidad"])

ventas\_c<-data[,"Ventas"]-mean(data[,"Ventas"])

Comprobamos sus medias.

> mean(utilidad\_c)

[1] -3.638145e-13

> mean(ventas\_c)

[1] 7.276235e-13

Los valores de las medias son “bastante” cercanos a cero, y se puede considerar que los nuevos datos obtenidos se encuentran centrados.

Luego, procedemos ha realizar nuevamente la regresión lineal pero con los datos centrados.

Obteniendo la siguiente información.

Call:

lm(formula = Utilidad ~ Ventas, data = data\_c)

Residuals:

Min 1Q Median 3Q Max

-676.35 -302.04 42.59 303.67 612.49

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -6.442e-13 5.809e+01 0.00 1

Ventas 4.399e-01 1.859e-02 23.66 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 367.4 on 38 degrees of freedom

Multiple R-squared: 0.9364, Adjusted R-squared: 0.9348

F-statistic: 559.9 on 1 and 38 DF, p-value: < 2.2e-16

Al eliminarce el intercepto nuestro modelo solo depende de beta2 y como se observa en la información obtenida, nos da que el intercepto es practicamente nulo.

De la tabla ANOVA podemos verificar si nuestra varible depende o no linealmente.

Df Sum Sq Mean Sq F value Pr(>F)

Ventas 1 75578286 75578286 559.9 <2e-16 \*\*\*

Residuals 38 5129142 134977

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Obtenemos el percentil 0.975 de la tabla F

> qf(0.95 , df1=1,df2=38)

[1] 4.098172

Y se verifica que la Utildad y las Ventas tienen una dependencia lineal, ya que se rechazó la hipótesis nula de beta2.

Para analizar los residuos, verifiquemos primero que estan centrados

> mean(res\_c)

[1] -7.771561e-16

Mediante los respectivos gŕaficos usando el siguiente código.

hist(res\_c,15)

qqline(res\_c,col="red")

Se puede observar que tienen una tendencia normal.

Del siguiente código obtemos los gráficos que nos permiten observar que la varianza de los residuos es constante y finalmente se ve que no existe un patrón en los residuos por lo que no son dependientes entre si.

plot(data\_c[,"Ventas"],data\_c[,"Utilidad"])

plot(res\_c,pred\_c)