

Lab09: PRUEBAS DE HIPÓTESIS PARA MODELO MULTILINEAL

Daniel Escriba Flores

Contents

Contexto:	1
Pregunta A: Datos, Regresion Multiple y Ecuacion	2
Lectura de datos	2
Regresion Multiple	3
Ecuacion de Regresion	3
Pregunta B: Linealidad e independencia	4
Linealidad	4
Interpretacion	4
Independencia	4
Interpretacion	5
Pregunta C: Normalidad y Homocedasticidad	5
Normalidad	5
Interpretacion	5
Homocedasticidad	5
Interpretacion	5

Contexto:

El Times-Observer es un periódico en la ciudad Metro. Al igual que muchos periódicos en la ciudad, el Times-Observer pasa por dificultades financieras. La gerente de circulación estudia otros periódicos en ciudades similares en Estados Unidos y Canadá, con interés particular en las variables que se relacionan con el número de suscriptores. Esta reúne la informacion muestral de 25 periódicos en ciudades similares empleando la siguiente notación:

• Suscriptores : Número de suscriptores (en miles) • Población : Población metropolitana (en miles) • Presupuesto : Presupuesto en publicidad del periódico (en miles de dólares) • Ingreso : Ingreso familiar medio en el área metropolitana (en miles de dólares)

Deseando predecir el número de suscriptores con base a las otras variables, haga lo siguiente:

Pregunta A: Datos, Regresion Multiple y Ecuacion

- a. Lea la base de datos y haga un análisis de regresión múltiple. Además, escriba la ecuación de regresión.

Lectura de datos

```
library(readxl)
library(pander)
library(equatiomatic)

data = read_excel("periódicos.xlsx", sheet="Hoja1")
data = data[-1]
data = as.data.frame(data)

# Usar pander para formatear la tabla
pander(data, caption = "Tabla de datos de suscripciones a periodicos de USA y Canada")
```

Table 1: Tabla de datos de suscripciones a periodicos de USA y Canada

Suscriptores	Población	Presupuesto	Ingreso
37.95	588.9	13.2	35.1
37.66	585.3	13.2	34.7
37.55	566.3	19.8	34.8
38.78	642.9	17.6	35.1
37.67	624.2	17.6	34.6
38.23	603.9	15.4	34.8
36.9	571.9	11	34.7
38.28	584.3	28.6	35.3
38.95	605	28.6	35.1
39.27	676.3	17.6	35.6
38.3	587.4	17.6	34.9
38.84	576.4	22	35.4
38.14	570.8	17.6	35
38.39	586.5	15.4	35.5
37.29	544	11	34.9
39.15	611.1	24.2	35
38.29	643.3	17.6	35.3
38.09	635.6	19.8	34.8
37.83	598.9	15.4	35.1
39.37	657	22	35.3
37.81	595.2	15.4	35.1
37.42	520	19.8	35.1
38.83	629.6	22	35.3
38.33	680	24.2	34.7
40.24	651.2	33	35.8

Regresion Multiple

```
modelo = lm(formula = Suscriptores ~ Población +
            Presupuesto +
            Ingreso,
            data= data)

summary(modelo)
```

```
##
## Call:
## lm(formula = Suscriptores ~ Población + Presupuesto + Ingreso,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58362 -0.24463 -0.03801  0.25393  0.63750
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.732813   8.427349  -0.680 0.503769
## Población    0.007537   0.001813   4.157 0.000446 ***
## Presupuesto  0.050883   0.014139   3.599 0.001689 **
## Ingreso      1.097381   0.245029   4.479 0.000207 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3269 on 21 degrees of freedom
## Multiple R-squared:  0.8348, Adjusted R-squared:  0.8112
## F-statistic: 35.38 on 3 and 21 DF,  p-value: 2.144e-08
```

```
modelo$coefficients
```

```
## (Intercept)    Población  Presupuesto    Ingreso
## -5.732813186  0.007536582  0.050883394  1.097380777
```

Ecuacion de Regresion

```
## Ecuación de regresión:
```

```
## Suscriptores = -5.732813 + 0.007536582 * Población + 0.05088339 * Presupuesto + 1.097381 * Ingreso
```

```
# Versión básica
extract_eq(modelo)
```

$$\text{Suscriptores} = \alpha + \beta_1(\text{Población}) + \beta_2(\text{Presupuesto}) + \beta_3(\text{Ingreso}) + \epsilon \quad (1)$$

```
# Versión con valores de coeficientes
extract_eq(modelo, use_coefs = TRUE)
```

$$\text{Suscriptores} = -5.73 + 0.01(\text{Población}) + 0.05(\text{Presupuesto}) + 1.1(\text{Ingreso}) \quad (2)$$

Pregunta B: Linealidad e independencia

- b. Verifique los supuestos de linealidad e independencia y diga si se aceptan o rechazan las hipótesis nula 0.

Linealidad

```
cor.test(data$Suscriptores, data$Población)

##
## Pearson's product-moment correlation
##
## data: data$Suscriptores and data$Población
## t = 4.0262, df = 23, p-value = 0.0005271
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3322627 0.8277985
## sample estimates:
##      cor
## 0.6429787
```

Interpretacion

- El valor de p-valor = 0.0005271 es menor a 0.05 , por lo tanto, se rechaza la hipótesis nula (H_0) . Esto indica que existe una correlación lineal estadísticamente significativa entre la variable Suscriptores y Población .

El coeficiente de correlación es 0.643 , lo que sugiere una relación positiva moderada a fuerte entre ambas variables.

Independencia

```
library(lmtest)

dwtest(modelo)

##
## Durbin-Watson test
##
## data: modelo
## DW = 2.279, p-value = 0.7468
## alternative hypothesis: true autocorrelation is greater than 0
```

Interpretacion

El p-valor = 0.7468 es mayor a 0.05 , por lo tanto, no se rechaza la hipótesis nula (H_0) . Esto indica que no existe autocorrelación significativa entre los residuos del modelo. Los errores son independientes, lo cual cumple con el supuesto de independencia en regresión lineal.

Pregunta C: Normalidad y Homocedasticidad

- c. Verifique los supuestos de normalidad y homocedasticidad y diga si se aceptan o rechazan las hipótesis nula 0.

Normalidad

```
shapiro.test(modelo$residuals)

##
##  Shapiro-Wilk normality test
##
## data:  modelo$residuals
## W = 0.98508, p-value = 0.964
```

Interpretacion

El p-valor = 0.964 es mayor a 0.05 , por lo tanto, no se rechaza la hipótesis nula (H_0) . Esto indica que los residuos siguen una distribución normal , cumpliendo así con el supuesto de normalidad en el modelo de regresión.

Homocedasticidad

```
library(lmtest)

bptest(modelo)

##
##  studentized Breusch-Pagan test
##
## data:  modelo
## BP = 3.4097, df = 3, p-value = 0.3327
```

Interpretacion

El p-valor = 0.3327 es mayor a 0.05 , por lo tanto, no se rechaza la hipótesis nula (H_0) . Esto indica que los residuos tienen varianza constante , es decir, se cumple el supuesto de homocedasticidad en el modelo de regresión.