

LAB 9

ARANDA

2025-05-16

#SEMANA 9: Prueba de hipótesis para validar un modelo multilíneal.

###1 Situación presentada: ###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 información 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:
```

```
# a. Lea la base de datos y haga un análisis de regresión múltiple. Además, escriba la ecuación de regresión.
# Leyendo la base 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")
```

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

Suscriptores	Población	Presupuesto	Ingreso
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

```
#Ajuste del Modelo de regresión múltiple
```

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

```
##
## 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
```

#Extraemos los coeficientes

```
model$coefficients
```

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

#Escribiendo la ecuación de regresión

~~# $\beta_0\beta_1\beta_2\beta_3$~~ ~~$=-5.732813186=0.007536582=0.050883394=1.097380777$~~

~~# $Suscriptores=(-5.732813186)+0.007536582 \cdot Población+0.050883394 \cdot Presupuesto+(1.097380777) \cdot Ingreso$~~

#b.Verifique los supuestos de linealidad e independencia y diga si se aceptan o rechazan las hipótesis nula H_0 .

```
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
```

#El p-value = 0.0005271 < 0.05. Entonces, al 5% de significancia estadística se rechaza la H_0 de que No existe correlación entre Suscriptores y Población. Es decir, se acepta la hipótesis alternativa H_1 de que existe correlación.

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

```
##
## Pearson's product-moment correlation
##
## data: data$Suscriptores and data$Presupuesto
## t = 4.8247, df = 23, p-value = 7.208e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.4363907 0.8626202
## sample estimates:
## cor
## 0.7092283
```

#El p-value = 7.208e-05 < 0.05. Entonces, al 5% de significancia estadística se rechaza la H_0 de que No existe correlación entre Suscriptores y Presupuesto. Es decir, se acepta la hipótesis alternativa H_1 de que existe correlación.

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

```
##
## Pearson's product-moment correlation
##
## data: data$Suscriptores and data$Ingreso
## t = 4.8356, df = 23, p-value = 7.016e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.4376789 0.8630271
## sample estimates:
## cor
## 0.7100188
```

#El p -value = $7.016e-05 < 0.05$ Entonces, al 5% de significancia estadística se rechaza la H_0 de que No existe correlación entre Suscriptores e Ingreso. Es decir, se acepta la hipótesis alternativa H_1 de que existe correlación.

```
library(GGally)
```

```
## Loading required package: ggplot2
```

```
## Registered S3 method overwritten by 'GGally':  
##   method from  
##   +.gg      ggplot2
```

```
##  
## Attaching package: 'GGally'
```

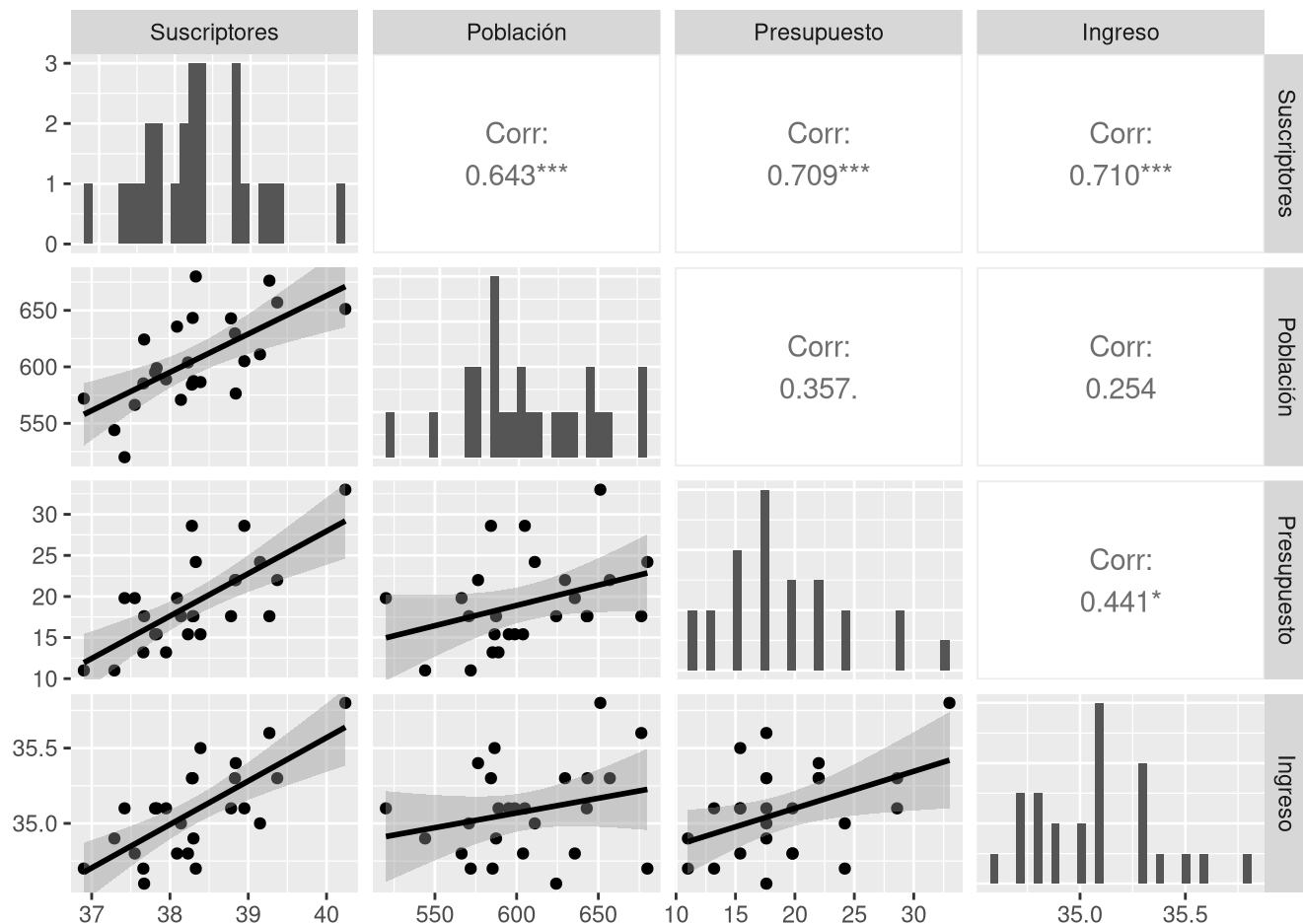
```
## The following object is masked from 'package:pander':  
##  
##   wrap
```

```
ggpairs(data, lower=list (continuous="smooth"),  
diag=list (continuous="barDiag"), axisLabels="none")
```

```
## Warning in warn_if_args_exist(list(...)): Extra arguments: 'axisLabels' are  
## being ignored. If these are meant to be aesthetics, submit them using the  
## 'mapping' variable within ggpairs with ggplot2::aes or ggplot2::aes_string.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
#Independencia
```

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
## as.Date, as.Date.numeric
```

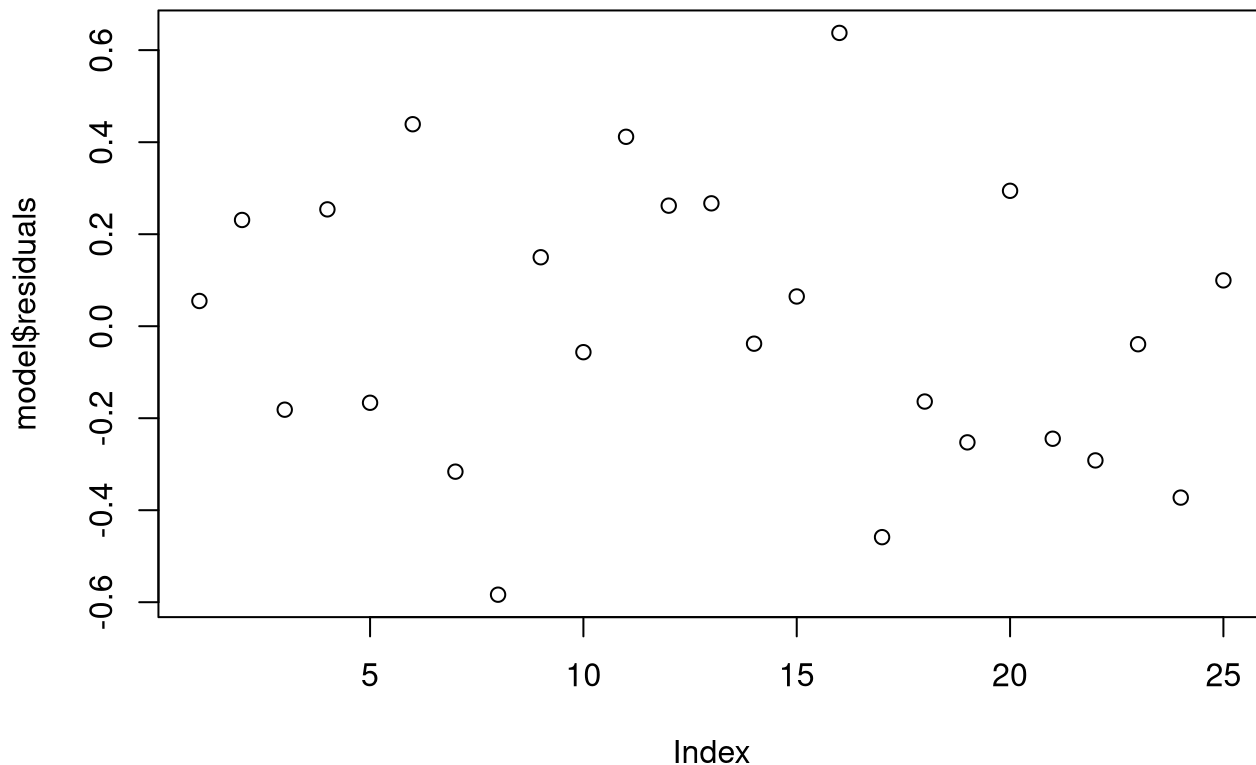
```
dwtest(model)
```

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

#El $p\text{-value} = 0.7468 > 0.05$. Entonces, al 5% de significancia estadística, NO se rechaza la hipótesis nula H_0 , de que No existe autocorrelación. Es decir, NO existe autocorrelación entre los residuos.

#Gráfico de los residuos

```
plot(model$residuals)
```



#Interpretación: El gráfico de residuos muestra que están bien distribuidos y no presentan patrones, lo que indica que los errores son independientes. No hay tendencias ni agrupamientos que sugieran problemas, aunque algunos puntos extremos podrían ser valores atípicos. En general, la aleatoriedad de los residuos sugiere que el modelo es adecuado en cuanto a la independencia de los errores.

#c. Verifique los supuestos de normalidad y homocedasticidad y diga si se aceptan o rechazan las hipótesis nula H_0 .

#Normalidad

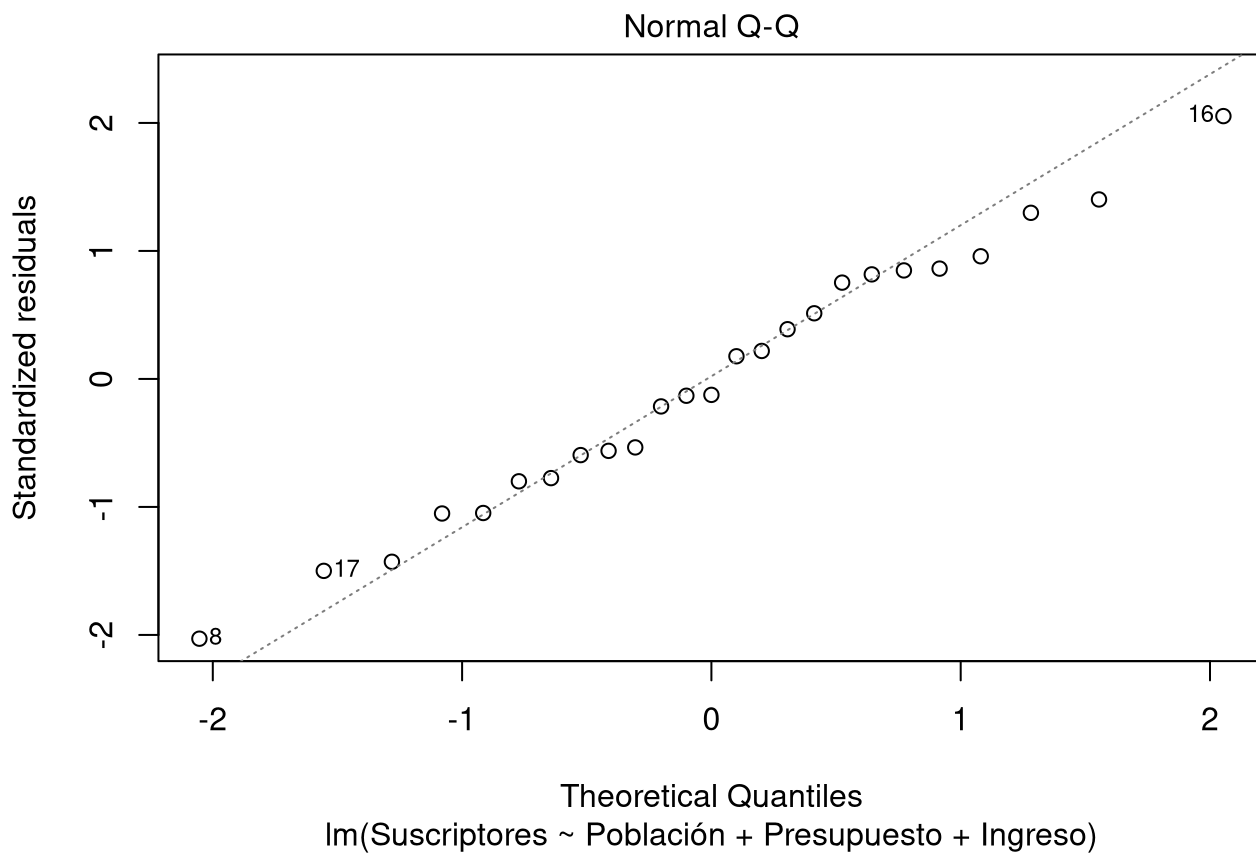
```
shapiro.test(model$residuals)
```

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

#El $p\text{-value} = 0.964 > 0.05$. Entonces, al 5% de significancia estadística, NO se rechaza la hipótesis nula H_0 , de que los residuos se distribuyen normalmente. Es decir, los residuos siguen una distribución normal.

#Gráfico Q-Q

```
plot(model, 2)
```



#Interpretacion:El gráfico Q-Q de residuos muestra que la mayoría se alinean con la línea diagonal, lo que sugiere una distribución normal. Sin embargo, hay algunas observaciones (8, 17 y 16) que se desvían, indicando posibles valores atípicos. En general, la normalidad de los residuos es aceptable, aunque hay ligeras desviaciones en los extremos que merecen ser investigadas.

#Varianza constante(homocedasticidad)


```
library(lmtest)
```

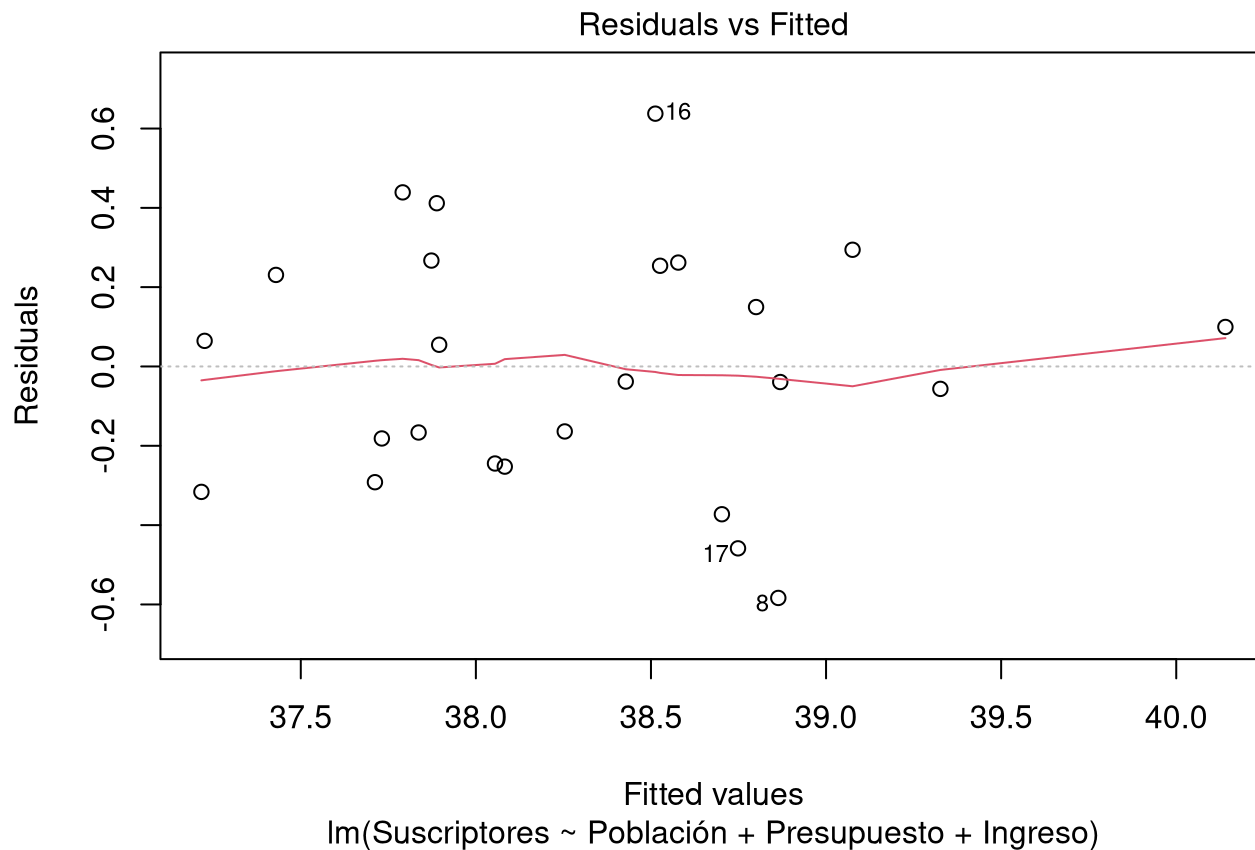
```
bptest(model)
```

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

#El p-value = 0.3327 > 0.05. Entonces, al 5% de significancia estadística, NO se rechaza la hipótesis nula H_0 , de que los residuos tienen varianza constante. Es decir, los residuos poseen una varianza constante.

```
#Grafico de los residuos
```

```
plot(model, 1)
```



#Interpretacion:El gráfico de residuos frente a valores ajustados muestra una distribución aleatoria de los residuos alrededor de cero, lo que indica que no hay patrones sistemáticos y sugiere que el modelo lineal es adecuado. Sin embargo, hay algunos puntos (8, 16 y 17) que se desvían, lo que puede señalar valores atípicos que necesitan más análisis. En general, la línea de suavizado (en rojo) se mantiene bastante plana, apoyando la homocedasticidad de los residuos.