

Lab05 - PRUEBAS DE HIPÓTESIS PARA MODELO LINEAL

Aranda Huerta, Milene - Escriba Flores, Daniel

2025-04-16

Contents

Pregunta A: Regresion y diagrama de dispersion	2
Lectura de datos	2
Regresion Lineal	2
Grafico de dispersion	2
Pregunta B: Linealidad e independencias	3
Supuesto de Linealidad entre la variable dependiente y la independiente	3
Interpretacion:	4
Supuesto de Independencia de los residuos	4
Interpretacion:	4
Pregunta C: Normalidad y Homocedasticidad	4
Supuesto de Normalidad de los residuos	4
Interpretación:	5
Grafico de ayuda	5
Supuesto de Varianza constante (homocedasticidad) de los residuos	5
Interpretacion:	6
Grafico de ayuda	6
Conclusiones finales	6

Contexto:

El director del Zoológico de Tampa estudia la relación entre el número de visitantes, en miles, y la temperatura alta, en grados Fahrenheit. Selecciona una muestra de 15 días y la información muestral recopilada se tabula a continuación:

Visitantes (miles)	Temperatura (°F)	Visitantes (miles)	Temperatura (°F)
2.0	86	2.2	84
0.6	71	2.5	66
2.0	89	1.3	76
2.1	73	3.6	84
2.2	76	1.0	75
2.1	75	1.8	72
0.5	68	2.1	76
0.3	72		

De acuerdo a lo anterior, haga lo siguiente:

Pregunta A: Regresion y diagrama de dispersion

- a. Lea la base de datos, haga un análisis de regresión que prediga el número de visitantes mediante la temperatura y muestre el diagrama de dispersión junto con su recta de regresión

Lectura de datos

```
library(readxl) #Llamamos a la libreria

## Warning: package 'readxl' was built under R version 4.4.3

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

##      visitantes temperatura
## 1           2.0           86
## 2           0.6           71
## 3           2.0           89
## 4           2.1           73
## 5           2.2           76
## 6           2.1           75
## 7           0.5           68
## 8           0.3           72
## 9           2.2           84
## 10          2.5           66
## 11          1.3           76
## 12          3.6           84
## 13          1.0           75
## 14          1.8           72
## 15          2.1           76
```

Regresion Lineal

- variable independiente : Temperatura en (°F)
- Variable dependiente : Visitantes en miles

```
x = data$temperatura
y = data$visitantes

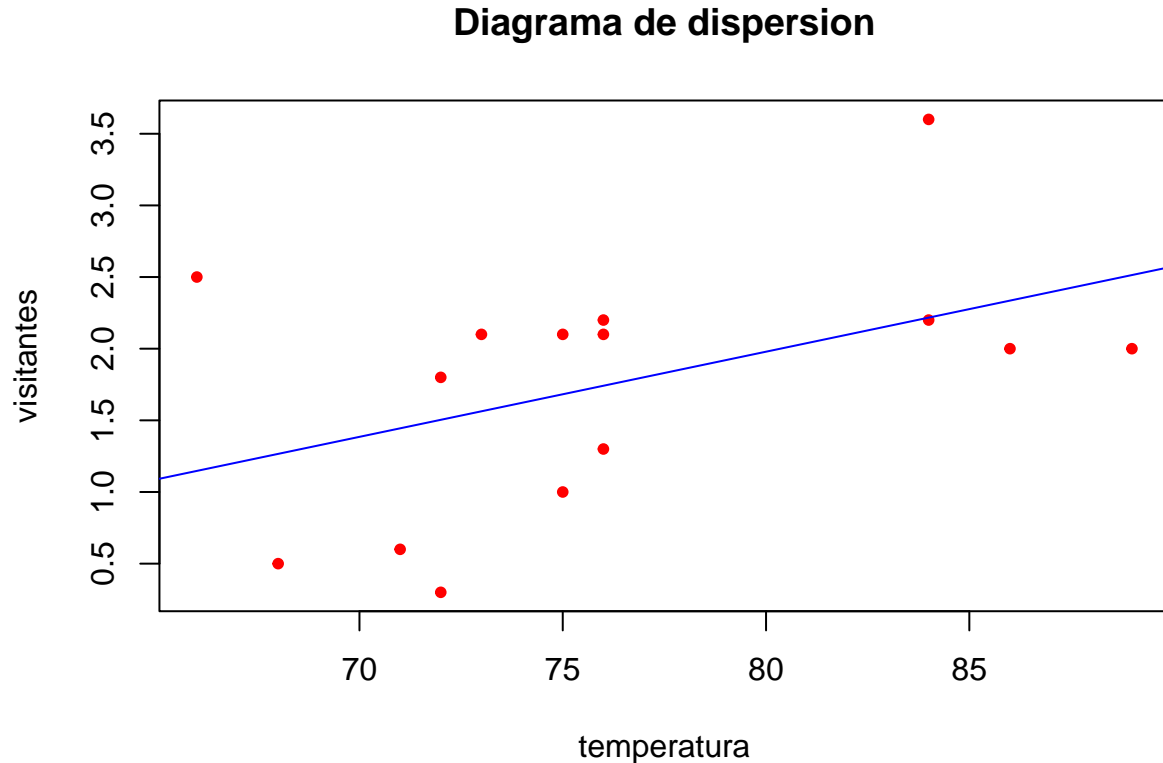
modelo = lm(y ~ x)

modelo$coefficients

## (Intercept)          x
## -2.77452790  0.05942075
```

Grafico de dispersion

```
plot(y ~ x, pch=20, col="red", xlab="temperatura", ylab="visitantes", main="Diagrama de dispersion")
abline(modelo,col="blue")
```



Pregunta B: Linealidad e independencias

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

Supuesto de Linealidad entre la variable dependiente y la independiente

Haremos uso de la función `anova()` que se basa en el estadístico F. Si el p-valor es menor a 0.05, se debe rechazar la hipótesis nula H_0 , de que $B_1 = 0$ (no existe una relación lineal entre X e Y), y en caso contrario se debe aceptar H_0 .

```
anova(modelo) # analysis of variance
```

```
## Analysis of Variance Table
##
## Response: y
##          Df Sum Sq Mean Sq F value Pr(>F)
## x          1  2.2188  2.21877   3.4262 0.08702 .
## Residuals 13  8.4186  0.64758
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretacion:

El p-valor (0.08702) es mayor a 0.05 => Se debe aceptar la hipótesis nula H_0 de que $B_1 = 0$. O sea. Es decir, no existe una relacion lineal entre X e Y.

Supuesto de Independencia de los residuos

Si el p-valor ('p-value') es menor a 0.05, se debe rechazar la hipótesis nula H_0 , de que NO existe autocorrelación entre los errores, lo que sugiere que hay evidencia de autocorrelación. y en caso contrario se debe aceptar H_0 .

```
library(lmtest)
```

```
## Cargando paquete requerido: zoo
```

```
##
```

```
## Adjuntando el paquete: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```
dwtest(modelo)
```

```
##
```

```
## Durbin-Watson test
```

```
##
```

```
## data:  modelo
```

```
## DW = 2.1539, p-value = 0.6579
```

```
## alternative hypothesis: true autocorrelation is greater than 0
```

Interpretacion:

El p-valor (0.6579) es mayor que 0.05, lo que indica que no hay evidencia estadística suficiente para rechazar la hipótesis nula (H_0) de que no hay autocorrelación entre los residuos. Por lo tanto, se sugiere que los residuos no están autocorrelacionados

Pregunta C: Normalidad y Homocedasticidad

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

Supuesto de Normalidad de los residuos

Hacemos uso de la función `shapiro.test()` que se basa en el estadístico Shapiro-Wilk sobre los residuos. Si el p-valor ('p-value') es menor a 0.05, se debe rechazar la hipótesis nula H_0 de que los residuos se distribuyen normalmente, y en caso contrario se debe aceptar H_0 .

```
residuos = resid(modelo)
```

```
shapiro.test(x = residuos)
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data:  residuos
## W = 0.94609, p-value = 0.4652
```

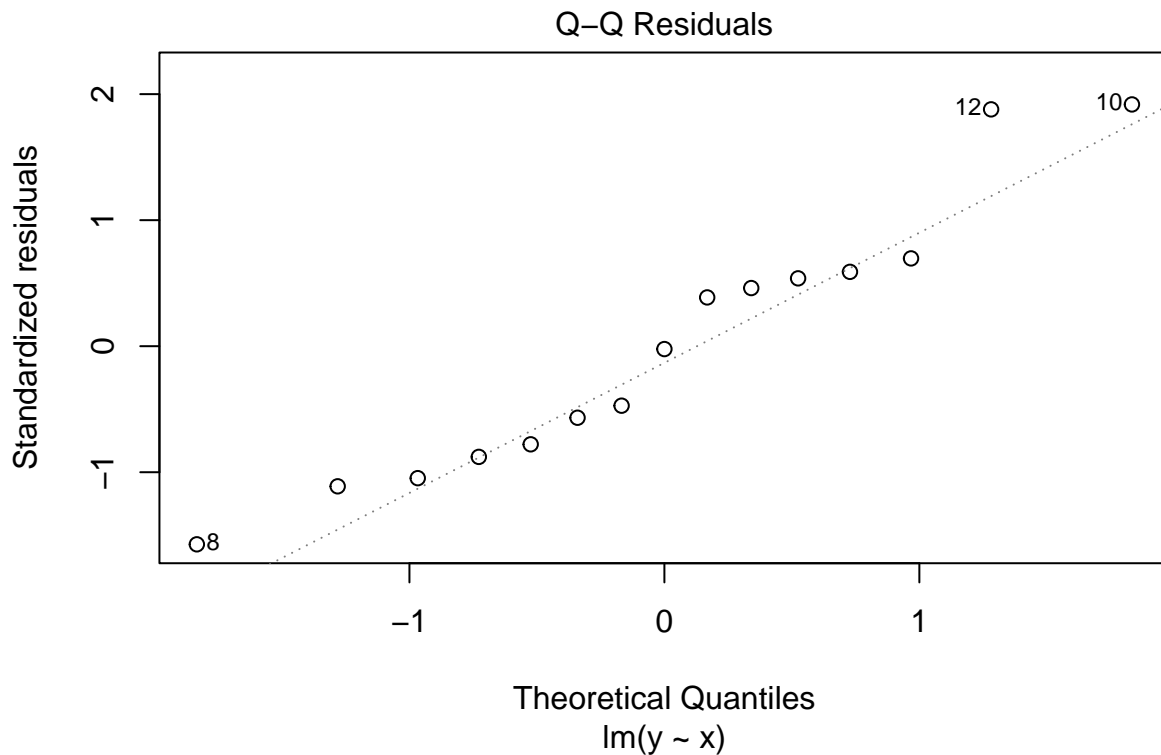
Interpretación:

El p-value = 0.4652. Ya que el p-value es mayor que 0.05 entonces se acepta la H_0 de que los residuos se distribuyen normalmente (lo que es bueno...)

Grafico de ayuda

Tambien podemos verificar la normalidad de los residuos por medio de un gráfico Q-Q, en el que se deben encontrar alrededor de la recta.

```
plot(modelo, 2)
```



Supuesto de Varianza constante (homocedasticidad) de los residuos

Por ultimo, para probar la varianza constante (homocedasticidad) de los residuos hacemos uso de la función `bptest()` que se basa en el estadístico Breusch-Pagan sobre los residuos. Previo a ellos debemos instalar el paquete “`lmtest`” en caso de no tenerlo.

Si el p-valor (‘p-value’) es menor a 0.05, se debe rechazar la hipótesis nula H_0 , de que los residuos tienen varianza constante, y en caso contrario aceptar H_0 .

```
library(lmtest)
```

```
bptest(modelo)
```

```
##
## studentized Breusch-Pagan test
##
## data: modelo
## BP = 1.0652, df = 1, p-value = 0.302
```

Ho: Los residuos tienen varianza constante H1: Los residuo no tienen varianza constante.

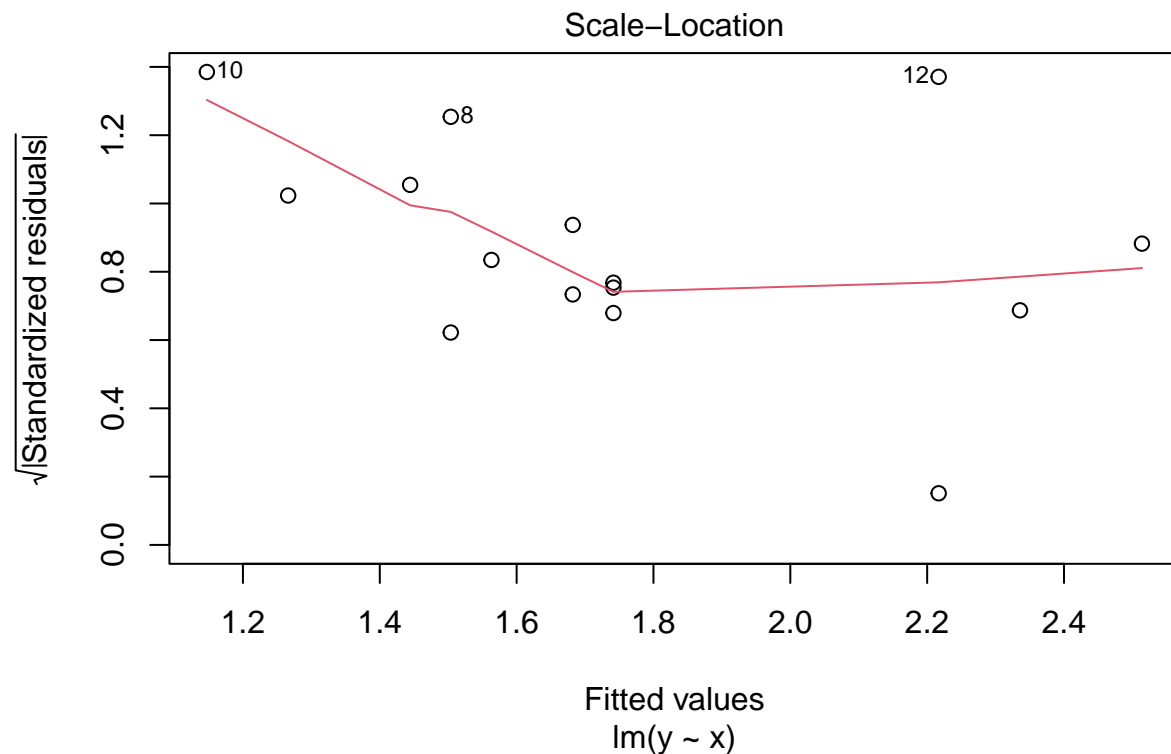
Interpretacion:

El p-value = 0.302. Ya que es mayor que 0.05 entonces se acepta la Ho de que los residuos tienen varianza constante.

Grafico de ayuda

Podemos hacer un gráfico de las raices cuadradas de los residuales estandarizados.

```
plot(x = modelo, which = 3)
```



Conclusiones finales

- Linealidad: No existe una relación lineal significativa entre la temperatura y el número de visitantes, ya que el p-valor (0.08702) es mayor que 0.05, lo que lleva a aceptar la hipótesis nula de que $B1 = 0$.
- Independencia de los residuos: Los residuos no están autocorrelacionados, ya que el p-valor (0.6579) es mayor que 0.05, lo que indica que no hay evidencia estadística suficiente para rechazar la hipótesis nula de que no hay autocorrelación entre los residuos.

- Normalidad de los residuos: Los residuos se distribuyen normalmente, ya que el p-valor (0.4652) es mayor que 0.05, lo que lleva a aceptar la hipótesis nula de que los residuos se distribuyen normalmente.
- Homocedasticidad: Los residuos tienen varianza constante, ya que el p-valor (0.302) es mayor que 0.05, lo que lleva a aceptar la hipótesis nula de que los residuos tienen varianza constante.

En resumen, aunque no se encontró una relación lineal significativa entre la temperatura y el número de visitantes, los residuos del modelo cumplen con los supuestos de independencia, normalidad y homocedasticidad.
