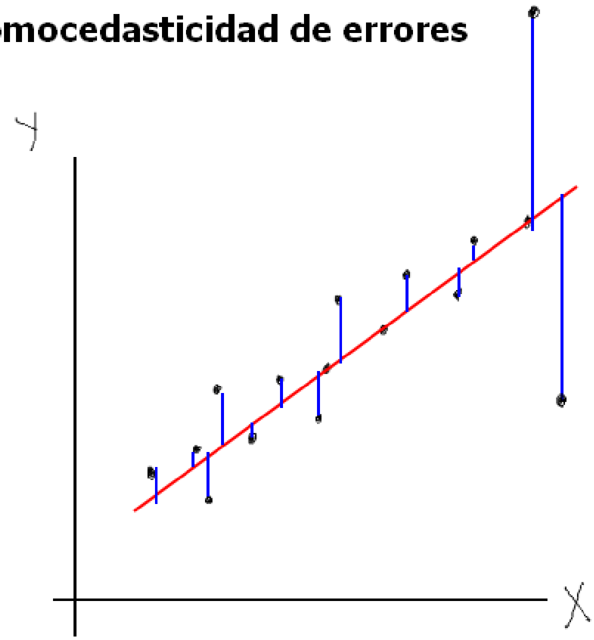
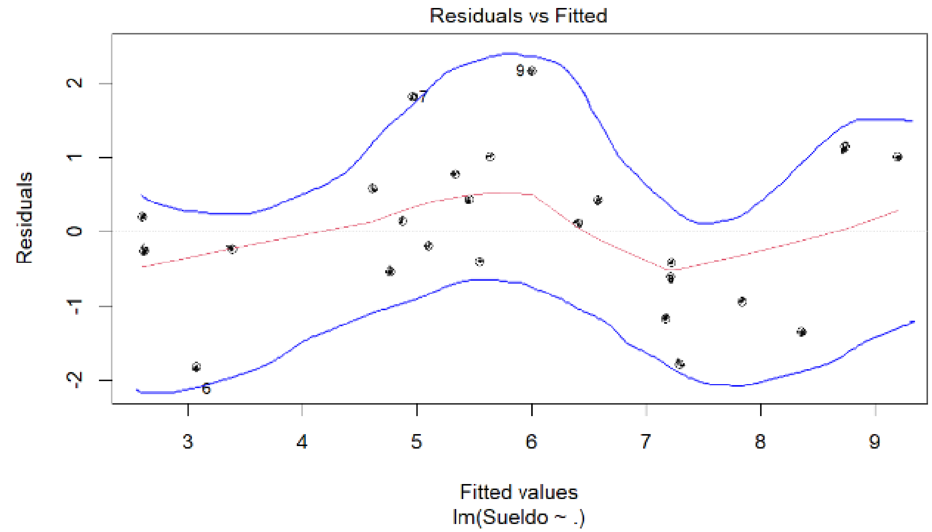
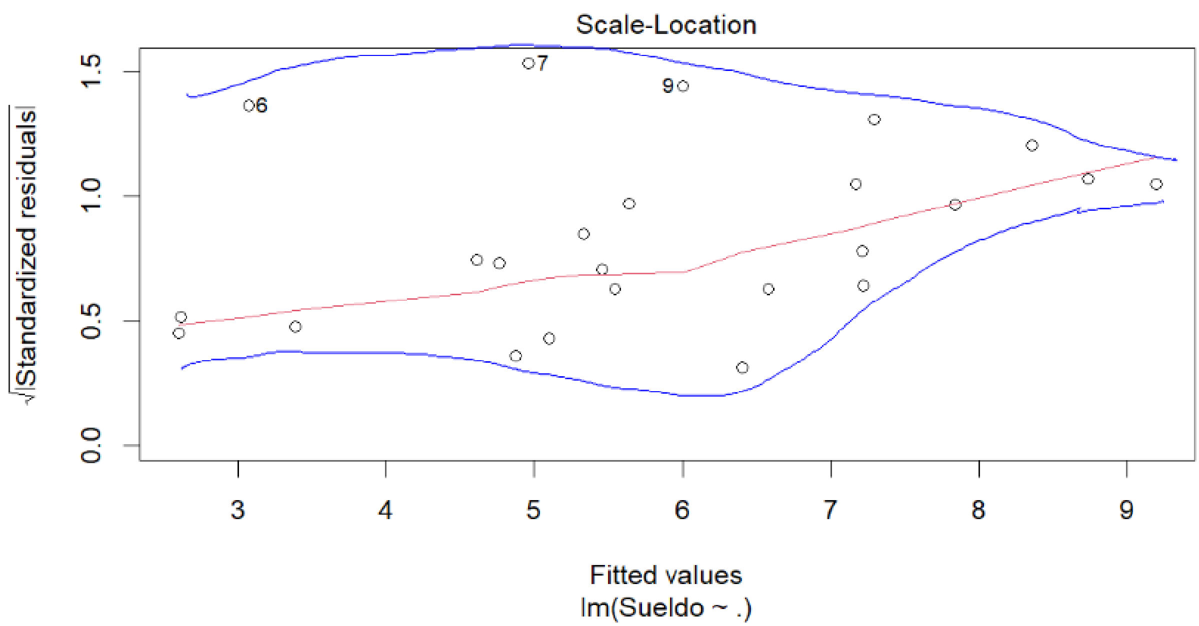
Homocedasticidad de errores





No habría $\underbrace{\text{homogeneidad de varianzas}}_{\text{homocedasticidad}}$ de los errores



Scale-Location
1.5
1.0
0.5
0.0
√|Standardized residuals|
7
9
6
3
4
5
6
7
8
9
Fitted values
lm(Sueldo ~ .)

Prueba de Breusch Pagan
H0: La varianza de los errores es constante (homocedasticidad)
H1: La varianza de los errores no es constante (heterocedasticidad)

$\alpha = 0.05$

→ car

```
> modelo |> ncvTest()
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.3867398, Df = 1, p = 0.53402
```

No se rechaza Ho → Se cumple el supuesto de homocedasticidad de errores

```
> modelo |> ols_test_breusch_pagan()
```

→ olsrr

Ordinary Least Squares → Mínimos Cuadrados Ordinarios

```
 Breusch Pagan Test for Heteroskedasticity
 -----------------------------------------
 Ho: the variance is constant
 Ha: the variance is not constant

              Data
 -----------------------------------
 Response : Sueldo
 Variables: fitted values of Sueldo

        Test Summary
 ----------------------------
 DF            =    1
 Chi2          =    0.3867398
 Prob > Chi2   =    0.5340181
```

```
> modelo |> bptest(studentize = F)

	Breusch-Pagan test

data:  modelo
BP = 0.86645, df = 4, p-value = 0.9293

> modelo |> bptest(studentize = T)

	studentized Breusch-Pagan test

data:  modelo
BP = 1.0883, df = 4, p-value = 0.8961
```

No estandarizo los residuales

Sí estandarizo los residuales

muestra pequeña

no se verifica normalidad

si hay valores influenciales

cap. 5

$e \sim N(\mu, \sigma^2)$

$e \sim Pois(\mu)$ $\longrightarrow E(e) = \mu$, $\longrightarrow V(e) = \mu$

$V(e) = f(E(e))$

Motivos

- La varianza es función de la media
- Los errores son multiplicativos $\longrightarrow$ $Y = (\beta_0 + \beta_1 X_1 + \beta_2 X_2) e$
- Falta de normalidad
- Desgaste o mejora en el proceso de toma o recolección de datos

Consecuencias

- Falta de precisión en las estimaciones intervalares.
- Pruebas de hipótesis basadas en las distribuciones, t, Chi cuadrado, F no son válidas.
- Predicciones ineficientes.

$$IC(\mu|\mathbf{x}) = \hat{\mu} \pm t_{1-\alpha/2,n-k}\sqrt{\hat{\sigma}^2\mathbf{x}'(X'X)^{-1}\mathbf{x}}$$

$$IP(y|\mathbf{x}) = \hat{y}_0 \pm t_{1-\alpha/2,n-k}\sqrt{\hat{\sigma}^2\left(1+\mathbf{x}'(X'X)^{-1}\mathbf{x}\right)}$$

FV	GL	SC	CM	Fcalc
Reg	K-1	SCReg	CMReg	CMReg/CME
Error	n-K	SCE	CME = $\hat{\sigma}^2$	
Total	n-1	SCTotal		

Acciones a tomar

Unidad 4

- Transformar la variable respuesta. Se sugiere Box Cox. ¿Qué sucede con las estimaciones y predicciones? $y \rightarrow \sqrt{y}$
- Utilizar mínimos cuadrados ponderados, de tal modo que $V(Y|X) = \sigma^2 X$ $\quad \sigma^2 f(x)$
- Considerar un modelo que contemple heterocedasticidad

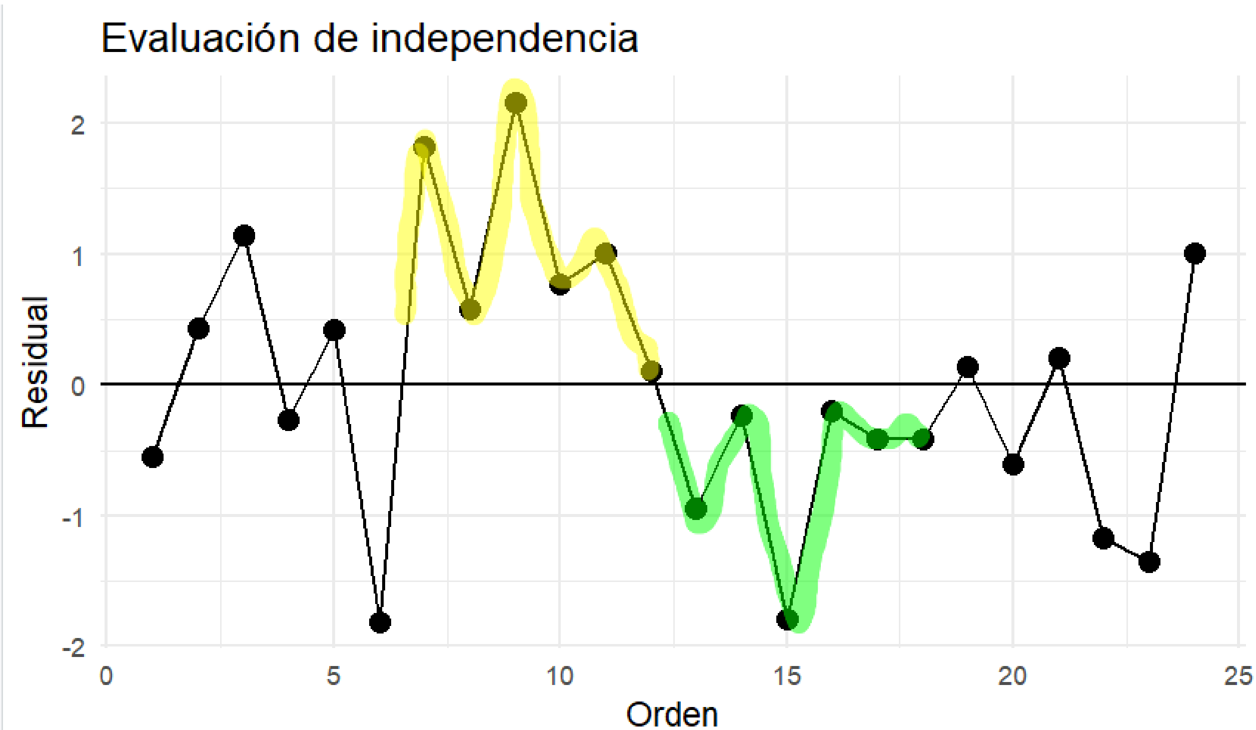
Unidad 4

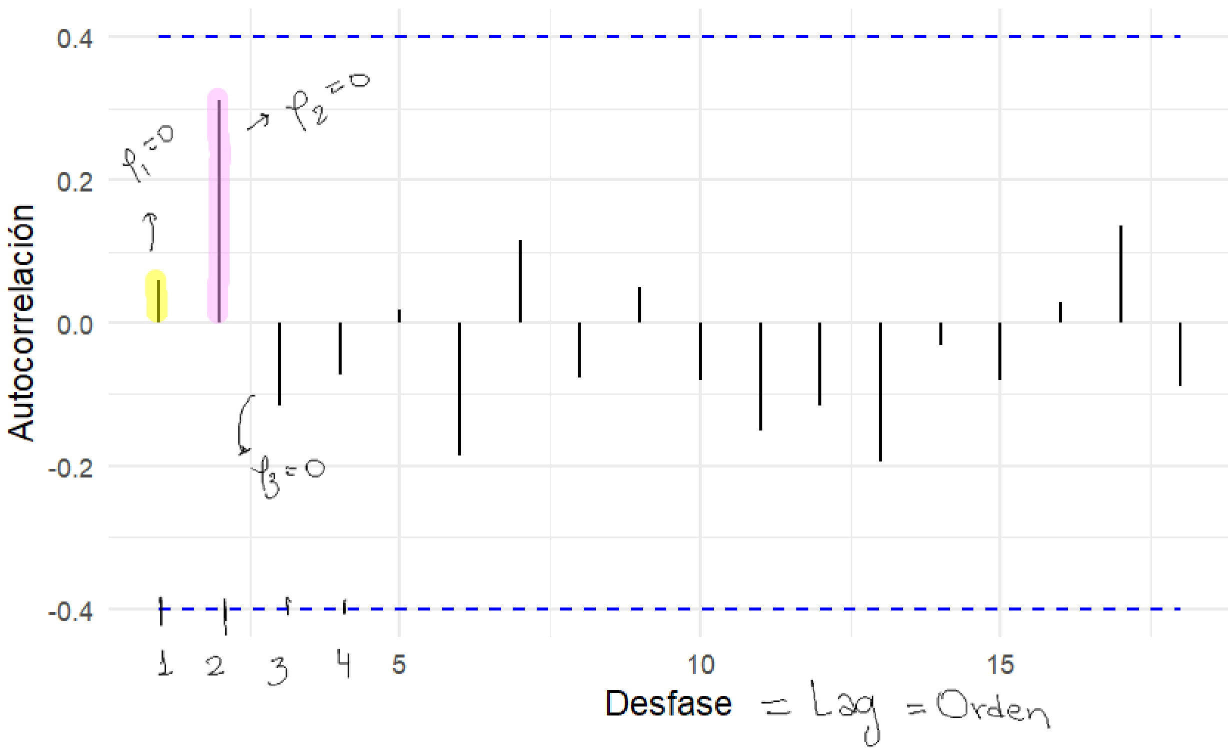
Regresión Poisson
Regresión Gamma
Regresión no lineal, etc.

Independencia de errores

Supuesto:

$\underbrace{\text{Auto}}_{\text{a sí mismo}}\underbrace{\text{correlación}}_{\text{asociación}}$ de errores → Los errores no se asocian entre sí / consigo mismos





e	e^*
e_1	e_2
e_2	e_3
e_3	e_4
e_4	e_5
e_5	e_6

$\hat{\rho}_1$

Autocorrelación de desfase 1

e	e^*
e_1	e_3
e_2	e_4
e_3	e_5
e_4	e_6
e_5	e_7

$\hat{\rho}_2$

Autocorrelación de desfase 2

$$\rho_1 = \rho_2 = \rho_3 = \dots = 0$$

Mes	Gasto en publicidad	Ganancia
1	6	10
2	4.5	11
3	8	9
4	9.2	14.1

Datos temporales o secuenciales

$\downarrow$

es probable que falle el supuesto de indep.

Prueba de Durbin Watson

$H_0: \rho_1 = 0$ ✓

$H_1: \rho_1 \neq 0$

$\alpha = 0.05$

```
> modelo |> dwtest(alternative = "two.sided")

        Durbin-Watson test

data:  modelo
DW = 1.8251, p-value = 0.6178
alternative hypothesis: true autocorrelation is not 0
```

$\left.\begin{array}{l} pv = 0.62 \\ \alpha = 0.05 \end{array}\right\}$ No se rechaza H_0

Se cumple el supuesto de independencia

```
> modelo |> durbinWatsonTest(alternative = "two.sided",
+                            max.lag = 10,
+                            reps = 1e5)
 lag Autocorrelation D-W Statistic p-value
   1      0.06083440      1.825133 0.61824
   2      0.31228900      1.239576 0.06954
   3     -0.11592874      1.986495 0.64536
   4     -0.07200947      1.894077 0.67010
   5      0.01955072      1.688519 0.87098
   6     -0.18651831      1.964531 0.32578
   7      0.11658913      1.215914 0.35800
   8     -0.07713629      1.582569 0.65528
   9      0.04915486      1.137808 0.53794
  10     -0.08010438      1.240744 0.97516
 Alternative hypothesis: rho[lag] != 0
```

$> \alpha$

$H_0: \rho_2 = 0$ ✓

$H_1: \rho_2 \neq 0$

$\alpha = 0.05$

$pv = 0.0695$

No se rechaza H_0

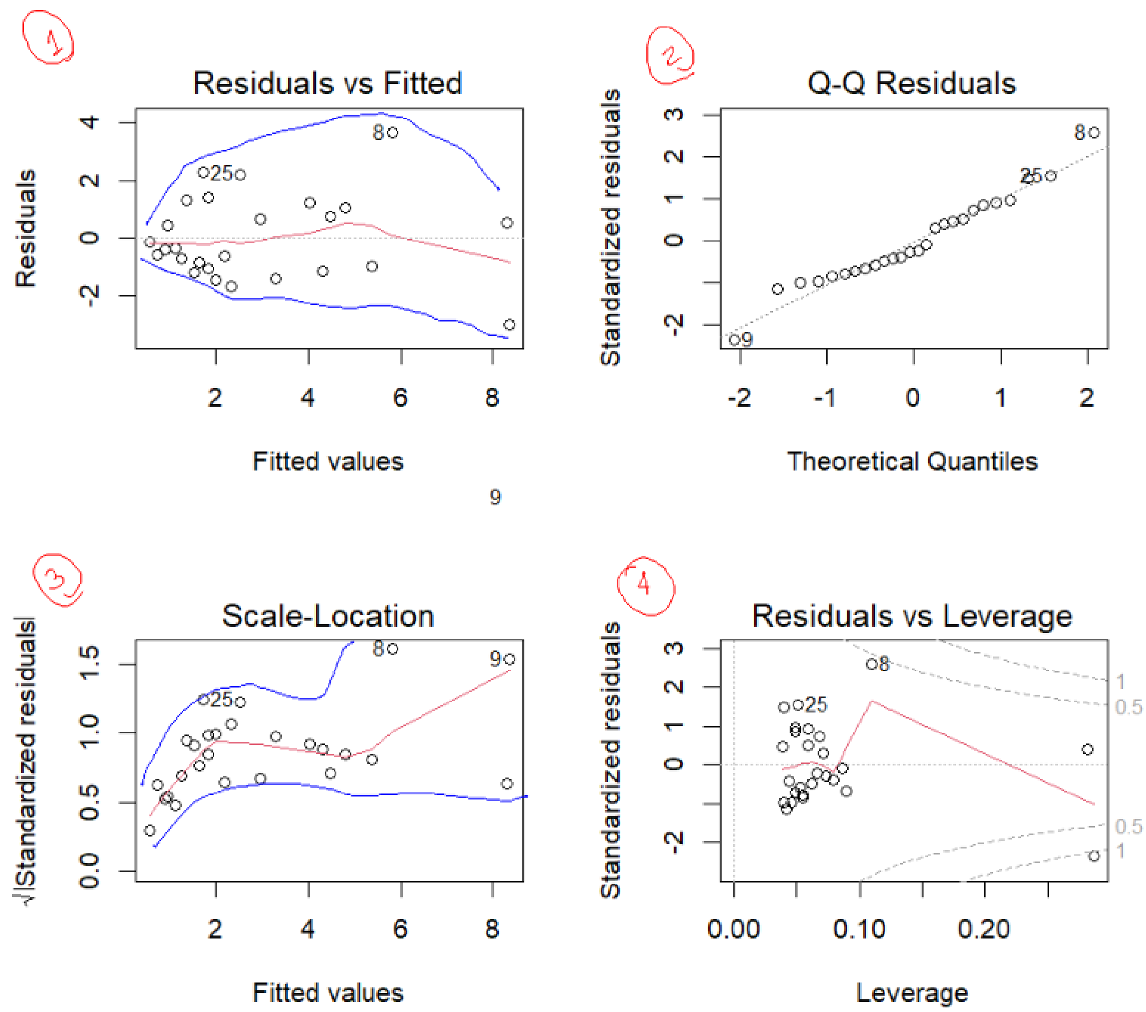
$H_0: \rho_3 = 0$ ✓

$H_1: \rho_3 \neq 0$

$\alpha = 0.05$

$pv = 0.645$

No se rechaza H_0



Normalidad de errores

(2) Al parecer sí se cumple el supuesto de normalidad de errores porque los puntos están alineados sobre la recta, sin desviación o alejamiento sistemático.

H0: Los errores siguen una dist. Normal
H1: Los errores no siguen una dist. Normal

```
> lm(y~x, datos2) |> residuals() -> res
> shapiro.test(res)

        Shapiro-Wilk normality test

data:  res
W = 0.96909, p-value = 0.5998
```

No se rechaza Ho. ✓ cumple el sup.

Homocedasticidad de errores

(1) y (3)

No se estaría cumpliendo el supuesto porque la dispersión no parece homogénea, sino que se va ampliando

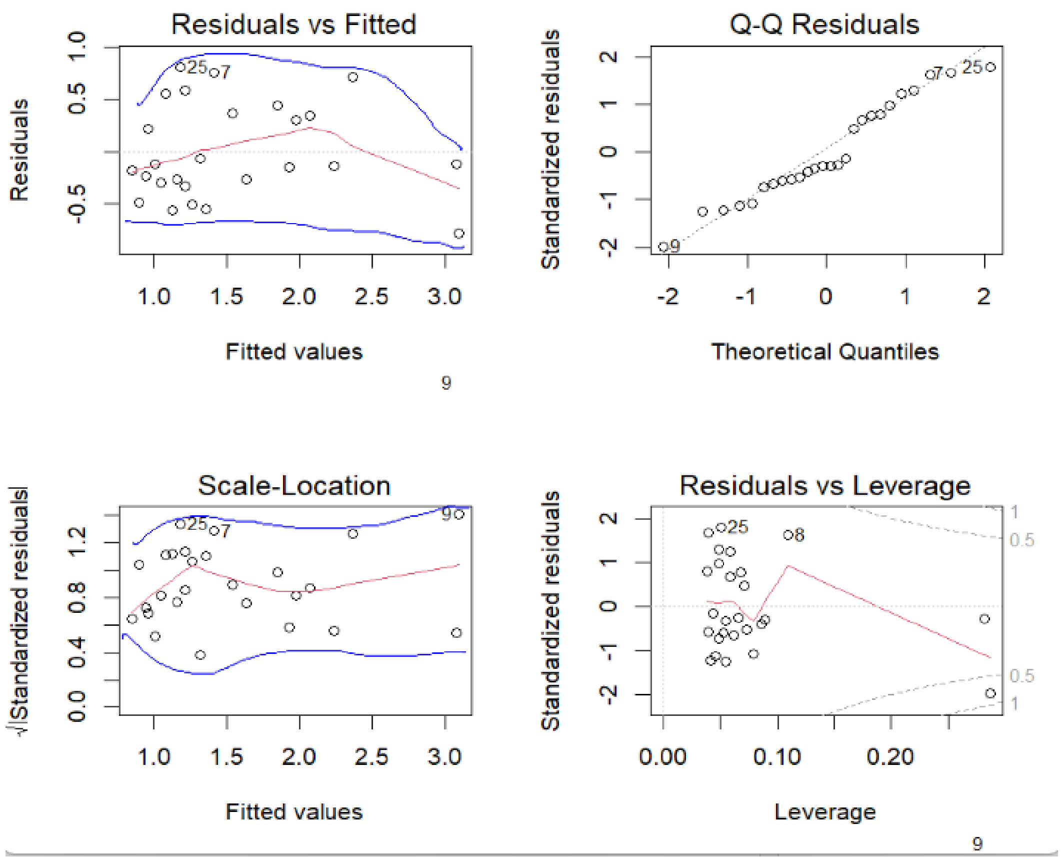
Ho: Homog.
H_1: No homog.

```
> lm(y~x, datos2) |> ncvTest()
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 5.407324, Df = 1, p = 0.020052
```

< 0,05

No se cumple el supuesto

Aplicamos $y' = \sqrt{y}$




```
> lm(y1~x, datos2) |> ncvTest()
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.2540729, Df = 1, p = 0.61422
```
$> \alpha$

Sí se cumple el supuesto de homog. de varianzas de los errores.

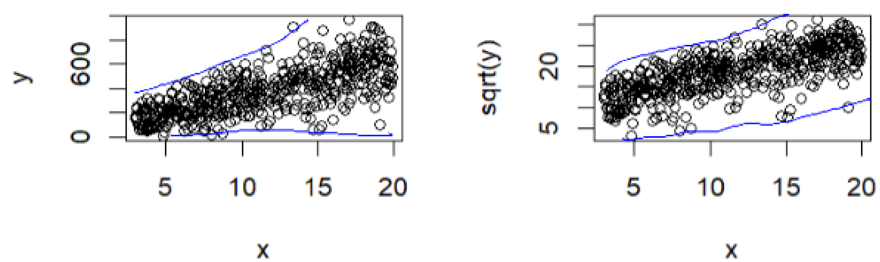
```
> lm(y1~x, datos2) |> residuals() |> shapiro.test()

	Shapiro-Wilk normality test

data:  residuals(lm(y1 ~ x, datos2))
W = 0.94187, p-value = 0.1489
```
$> \alpha$

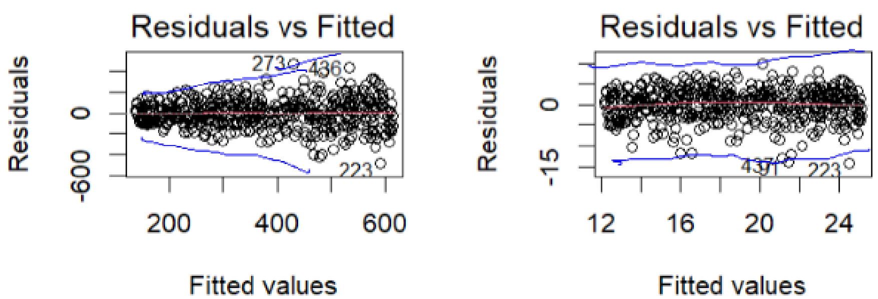
Se cumple la Normalidad de errores

Ejemplo 2

Se generaron datos con $e \sim N(0, \sigma^2 = 45\mu)$

ORIGINAL TRANSFORMADO






```
> lm(y~x) |> ncvTest() #ORIGINAL
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 45.72062, Df = 1, p = 1.3638e-11
> lm(sqrt(y)~x) |> ncvTest() #TRANSFORMADO
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 1.536166, Df = 1, p = 0.21519
```

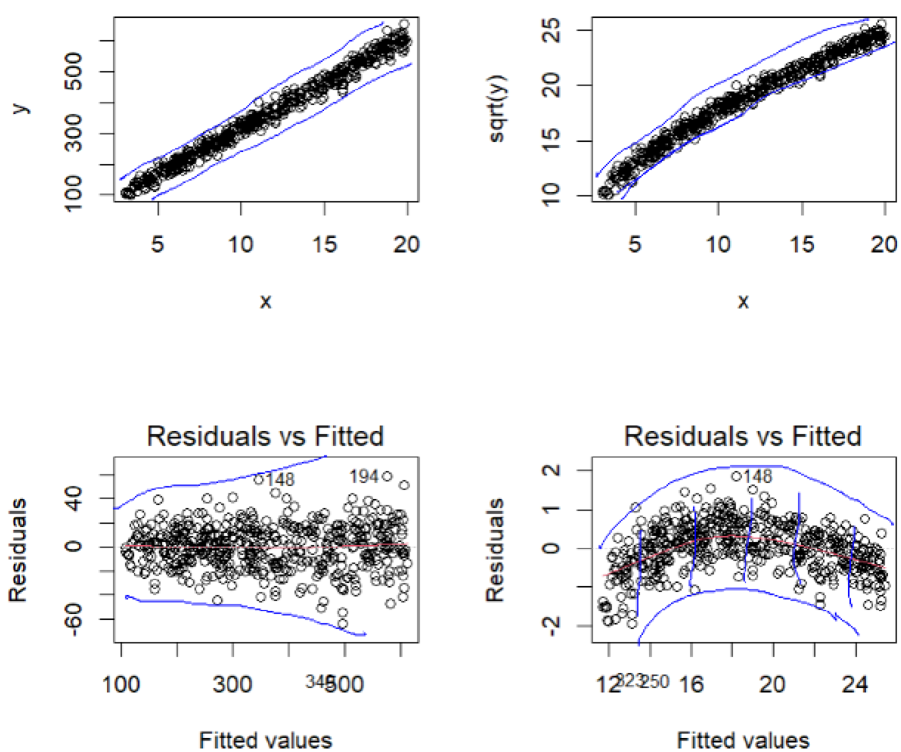
$p = 1.3638e{-11} < \alpha$ → No cumple

$p = 0.21519 > \alpha$ → Sí cumple

Ejemplo 3

Se generaron datos Poisson ($\sigma^2 = \mu$)

ORIGINAL — TRANSFORMADO $\sqrt{y}$




```
> lm(y~x) |> ncvTest()
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 21.48025, Df = 1, p = 3.5749e-06 <α
> lm(sqrt(y)~x) |> ncvTest()
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 17.5705, Df = 1, p = 2.7685e-05 <α
```

la transformación no garantiza el cumplimiento del supuesto

pvalor original: 0.00000357

pvalor transformado: 0.0000277 >