Tercer taller

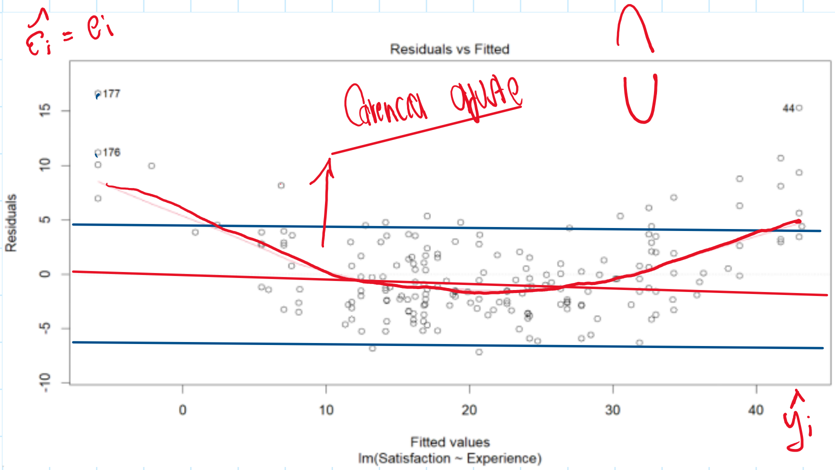
miércoles, 20 de noviembre de 2024 1:59 p. m.

Observación: Al interpretar los parámetros, es esencial fijarse en su signo. Por ejemplo, B1 negativo, con un valor de -8. Su interpretación práctica sería: Por un cambio unitario en X, Y disminuye en promedio 8 unidades. (No sería Y disminuye en promedio -8 unidades, ya que cambia el sentido de la interpretación.)

$\boxed{Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \; ; \; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)}$: $\boxed{Y: \text{Satisfacción}; \; X: \text{Años trabajados}}$

• $\boxed{\hat{\beta}_0 \to \beta_0}$; $\boxed{\hat{\varepsilon}_i \to \varepsilon_i}$

$\hat{\varepsilon}_i = e_i$



• El supuesto de media cero siempre se cumple (Errores)

$\boxed{\sum_{i=1}^{n} \hat{\varepsilon}_i = 0}$

- A partir de la línea roja, se podría decir que el modelo está capturando un efecto no considerado por el modelo lineal.
- Hay patrón de carencia de ajuste. Y no decimos nada sobre la varianza.

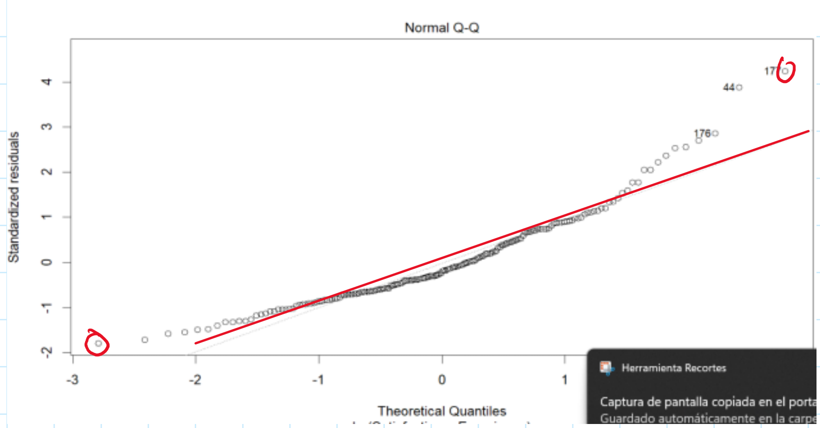
- Existe una jerarquía: Es idóneo primero aplicar la prueba de carencia de ajuste y para luego considerar otras pruebas, errores y hacer inferencia sobre los parámetros (Prueba de significancia de la regresión y de los parámetros).

Probar criterio normalidad: (1). Gráfica

```
Call:
lm(formula = Satisfaction ~ Experience, data = datos)

Residuals:
    Min      1Q  Median      3Q     Max
-7.1882 -2.7119 -0.8805  2.5452 16.5908

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -8.0242     0.8563  -9.371   <2e-16 ***
Experience    5.4675     0.1516  36.066   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.998 on 188 degrees of freedom
Multiple R-squared:  0.8737,	Adjusted R-squared:  0.8731
F-statistic:  1301 on 1 and 188 DF,  p-value: < 2.2e-16
```

gl

R^2 $\sqrt{MSE}$



(2). Analítica: $\boxed{\text{Shapiro-wilk}}$ $\boxed{H_0: \varepsilon_i \sim \text{normal} \quad \boxed{i=1,\dots,n} \\ H_1: \varepsilon_i \nsim \text{normal}}$: $\boxed{H_0: \varepsilon_i \sim N(0, \sigma^2) \\ H_1: \varepsilon_i \nsim N(0, \sigma^2)}$

(2). Analítica: Shapiro-Wilk

$H_0: \varepsilon_i \sim normal$ $i=1,\dots,n$: $H_0: \varepsilon_i \sim N(0,\sigma^2)$

$H_1: \varepsilon_i \nsim normal$ $\qquad$ $H_1: \varepsilon_i \nsim N(0,\sigma^2)$

(Esto no)

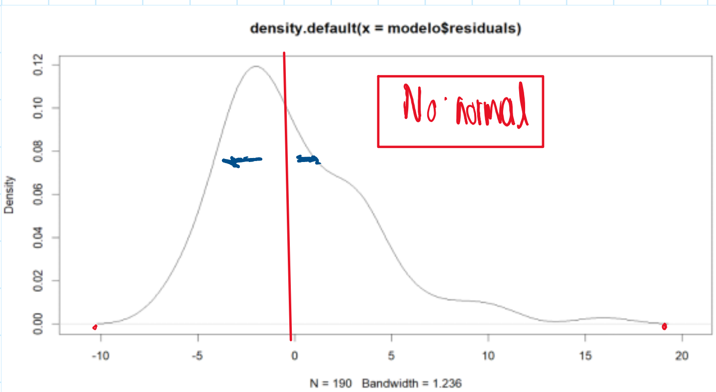
Pval < α $\quad$ α = 0,05 $\quad$ α = 0,1

```
        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.93066, p-value = 7.217e-08
```

p < α

- En función del valor P obtenido, existe evidencia suficiente para rechazar la H0. -> En términos del problema, los errores no tienen una distribución normal.

(Pval ≈ 0,05 $\quad$ α = 0,05)



- NO se cumplen los supuestos del modelo.

(Probar significancia de parámetros y regresión global)

$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$

$T^2 = F$ $\quad$ Si y solo si $F \sim f_{1,\dots}$ ~ Una única covariable regresora

$S.E(\hat{\beta}_j) \neq \widehat{S.E(\beta_j)}$

$\sigma^2 \qquad$ MSE

$H_0: \beta_j = 0 \quad j = 0,1$

$H_1: \beta_j \neq 0$

~ $H_0: \beta_0 = 0$: $H_1: \beta_1 = 0$

$H_1: \beta_0 \neq 0 \qquad H: \beta_1 = 0$

$T = \dfrac{\hat{\beta}_j - 0}{S.E(\hat{\beta}_j)} \sim t_{n-2}$

Bajo Ho Cierto

T_{cal}: Evalúo

$P(|t_{n-2}| > |T_{calc}|) < \alpha$

$R_c: |T_{cal}| > t_{\alpha/2,\,n-2}$

$T_{calc}(\hat{\beta}_0)$

Pval $(\beta_0) < \alpha$

Pval $(\beta_1) < \alpha$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -8.0242     0.8563  -9.371   <2e-16 ***
Experience    5.4675     0.1516  36.066   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.998 on 188 degrees of freedom
Multiple R-squared:  0.8737,    Adjusted R-squared:  0.8731
F-statistic:  1301 on 1 and 188 DF,  p-value: < 2.2e-16
```

$T_{cal}(\hat{\beta}_1)$

- Hay evidencia suficiente para rechazar la hipótesis nula, en el caso de j= 0, 1.
- Suponiendo que el modelo funciona, no hay carencia de ajuste, los supuestos sobre los errores, se podría concluir en términos prácticos.

$X_{min} = 0$ $\quad$ > 0 (No se puede interpretar)

Crear Intervalos de confianza:

```
                2.5 %    97.5 %
(Intercept) -9.713459 -6.335017
Experience   5.168465  5.766559
```

- Por un cambio unitario en los años trabajados por los empleados del centro universitario, la satisfacción promedio cambia 5.46 unidades.
- Con una confianza del 95% se puede determinar que el efecto de la experiencia está comprendido entre 5,1 y 5,7 unidades.

$T_{cal} = 36,066 \qquad F_{cal} = 1301$

$T^2 = (36,066)^2 = F$

Falo.

$P(|t_{n-2}| > |T_{cal}|) = P(f_{1,188} > F_{cal})$

```
Analysis of Variance Table
```

5,1 y 5,7 unidades.

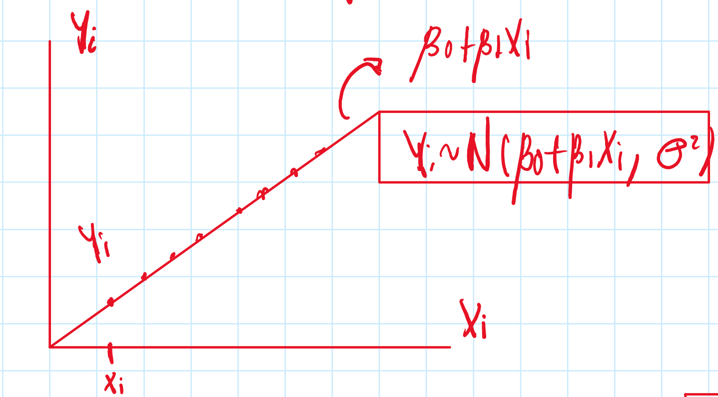
```
Analysis of Variance Table

Response: Satisfaction
            Df  Sum Sq Mean Sq F value    Pr(>F)
Experience   1 20790.7   20791  1300.8 < 2.2e-16 ***
Residuals  188  3004.8      16
---
```

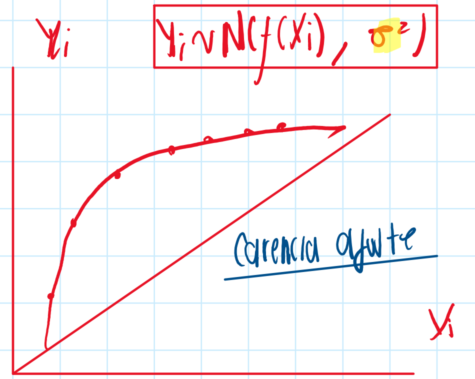
Fcalc. SSE SSR MSR MSE Pval

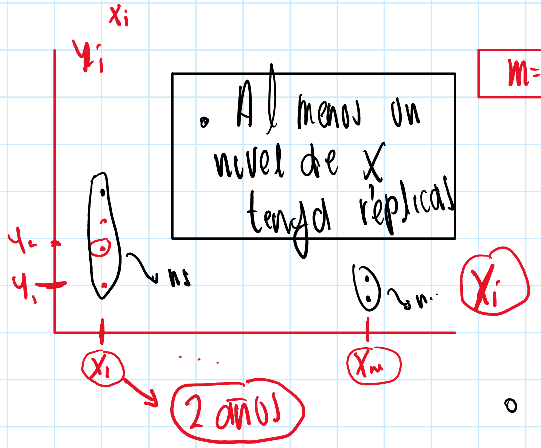
$\boxed{P(|t_{n-2}| > |T_{cal}|) = P(f_{1,188} > F_{cal})}$

$F = \dfrac{SSR/1}{SSE/n-2} = \dfrac{MSR}{MSE} \sim f_{1,n-2}$

Prueba de carencia de ajuste:







$\boxed{m = \text{niveles}}$

$H_0: E[Y|X_i] = \beta_0+\beta_1 X_i$

$H_1: E[Y|X_i] \neq \beta_0+\beta_1 X_i$

$\boxed{P_{val} < \alpha}$

$\boxed{F_{LOF} \neq F_{sig}}$

$F_{LOF} = \dfrac{SSLOF/(m-2)}{SSPE/(n-m)} = \dfrac{MSLOF}{MSPE}$

Bajo H_0 cierta $\sim f_{m-2,n-m}$

• $SSPE = \sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-\bar{Y}_i)^2$ $(n-m)$ $\quad SSLOF = \sum_{i=1}^{m} n_i(\bar{y}_i - \hat{y}_i)$ $m-2$

+

m = 100

$\boxed{\begin{matrix} SSE = SSPE + SSLOF \\ (n-2) = m-2 + n-m \end{matrix}}$

Análisis de varianza que incorpora la prueba de falta de ajuste en el modelo de RLS

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F Calculado
Regresión	SSR	1	$MSR = \frac{SSR}{1} = SSR$	$F_0 = \frac{MSR}{MSE}$
Error	SSE	$n-2$	$MSE = \frac{SSE}{n-2}$	
Falta de Ajuste	SSLOF	$m-2$	$MSLOF = \frac{SSLOF}{m-2}$	$F_0 = \frac{MSLOF}{MSPE}$
Error Puro	SSPE	$n-m$	$MSPE = \frac{SSPE}{n-m}$	
Total	SST	$n-1$		

```
Lack of fit test - Anova Table
             Sum Sq  Df Mean Sq   F value    Pr(>F)
Regression  20790.7   1 20790.7 1300.7898 < 2.2e-16 ***
Residuals    3004.8 188    16.0
Lack of fit  2130.9  98    21.7    2.2392  6.69e-05 ***
Pure error    873.9  90     9.7
Total       23795.5 189
```

?0 niveles con 3 réplicas

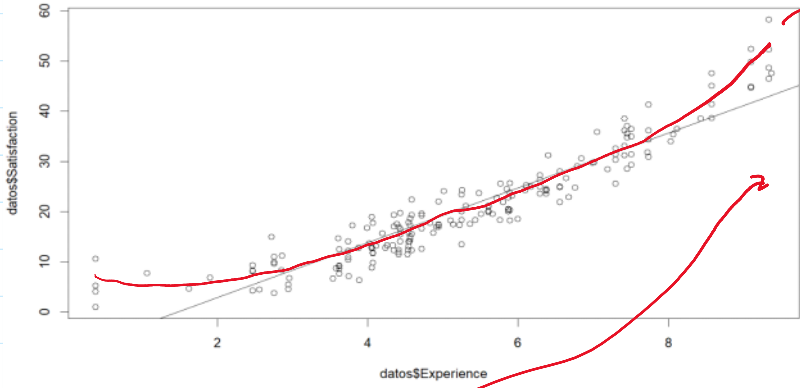
SSLOF SSPE $m-2 = 98$ MSLOF MSPE F_{LOF} cal. Pval $< \alpha$

$\boxed{\text{• Hay carencia de ajuste}}$

- Generalmente, el problema de normalidad, está asociado a una heterocedasticidad. $\boxed{\text{Varianza no constante}}$

- El problema de carencia de ajuste, podría solucionarse con una transformación.

$\boxed{Y_i = \beta_0 + \beta_1\sqrt{X_i} + \varepsilon_i;\ \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)}$



Exponencial multiplicativo:

$$\boxed{Y_i = \beta_0 e^{\beta_1 X} \cdot \varepsilon_i}\ ; \quad \boxed{\varepsilon_i \sim lognorm(\mu = 0, \sigma)}$$

$$\boxed{Y_i^* = \ln(Y_i)}$$

Linealizar

$$\ln(Y_i) = \ln(\beta_0 e^{\beta_1 X_i} \cdot \varepsilon_i)$$

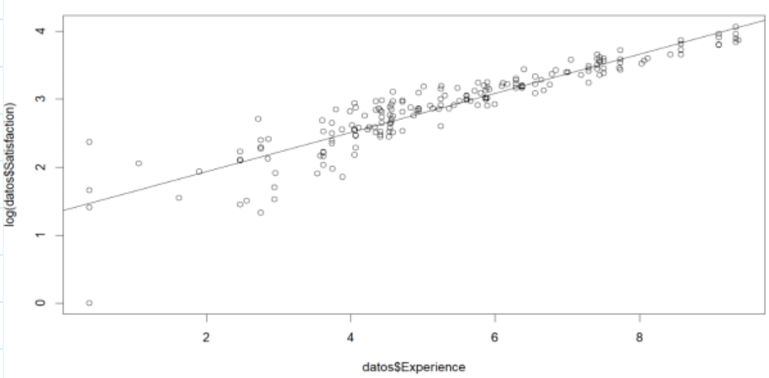
$$Y_i^* = \ln(\beta_0) + \beta_1 X_{1i} + \ln(\varepsilon_i)$$

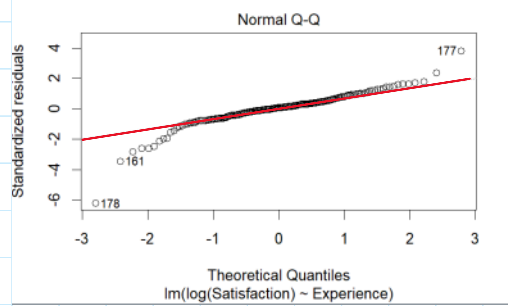
$$\boxed{Y_i^* = \beta_0^* + \beta_1 X_{1i} + \varepsilon_i^*\ ; \quad \varepsilon_i^* \sim N(0, \sigma^2)}$$

$\sim \text{Log}(Y)$.

- Verificar supuestos modelo!

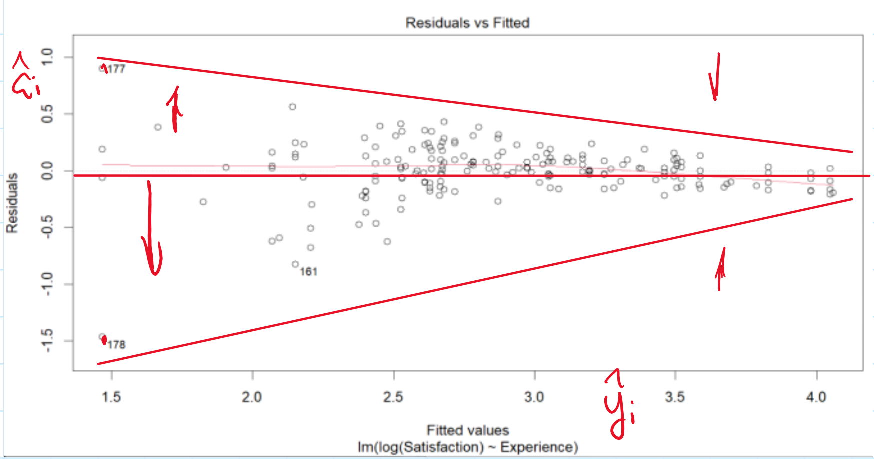
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.357843   0.051531   26.35   <2e-16 ***
Experience  0.288230   0.009123   31.59   <2e-16 ***
---
```






```
        Shapiro-Wilk normality test

data:  modelo2$residuals
W = 0.89831, p-value = 4.504e-10
```

- En general, se busca no rechazar la prueba de normalidad.
- En general, se busca no rechazar la prueba de carencia de ajuste.



$$\boxed{Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i}$$

Prueba T

Significancia regresión: F

$$\boxed{H_0: \beta_1 = \cancel{\beta_2} = 0}$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$$

$$\boxed{H_0: \beta_1 = 0}$$