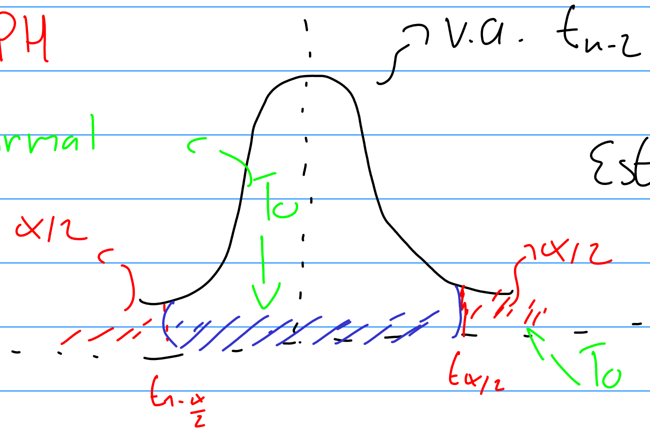
Sobre PH



Estamos condensando la info en T_0

$$\begin{cases} H_0: \beta_J = 0 \\ H_1: \beta_J \neq 0 \end{cases}$$

De entrada asumimos H_0 cierta

Rechazo H_0 si P-valor $< \alpha$ o si $|T_0| > t_{\frac{\alpha}{2}, n-2}$

Supongamos $|T_0| > t_{\alpha/2}$

$$¿P(t_{n-2} > t_{\alpha/2, n-2})? = \alpha/2$$

$$P(t_{n-2} > |T_0|) < \alpha/2 \Rightarrow \underbrace{2P(t_{n-2} > |T_0|)}_{\text{P-valor}} < \alpha$$

```
Call:
lm(formula = y ~ x, data = datos)

Residuals:
    Min      1Q  Median      3Q     Max
-26.662 -16.285  -5.674  12.833  42.896

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  21.7586     2.3494   9.262 6.03e-14 ***
x            -2.8183     0.7365  -3.827 0.000272 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19.94 on 73 degrees of freedom
Multiple R-squared:  0.1671,	Adjusted R-squared:  0.1557
F-statistic: 14.64 on 1 and 73 DF,  p-value: 0.0002716
```

ambos valores p son menores que α (0.05) $\Rightarrow$ rechazo

4) Usando mínimos cuadrados ordinarios (OLS) se optimiza

$$S(\beta_0,\beta_1) = \sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_i)^2$$

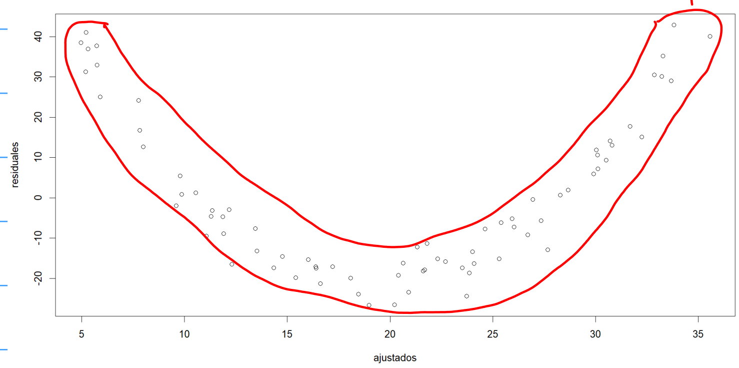
, con $\hat\beta_0$ y $\hat\beta_1$ los valores que minimizan S

$$\frac{\partial S}{\partial\beta_0} = -2\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_i)$$

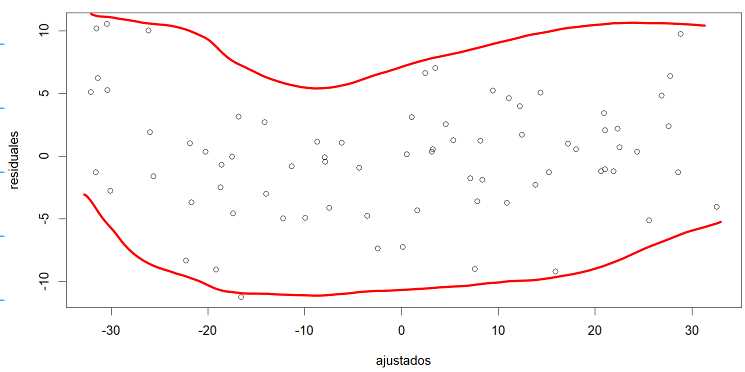
evaluemos esto ($\hat\beta_0$ y $\hat\beta_1$)

$$\frac{\partial S(\hat\beta_0,\hat\beta_1)}{\partial\beta_0} = -2\sum_{i=1}^{n}\underbrace{(y_i-\overbrace{\hat\beta_0-\hat\beta_1X_i}^{\hat y_i})}_{e_i} = 0 \Rightarrow \underbrace{\sum_{i=1}^{n}\frac{e_i}{n}}_{\text{media de los residuales es cero}} = 0$$

5)



Se ve claramente una forma de U $\Rightarrow$ no hay linealidad y no se concluye nada respecto a la varianza constante.



Varianza no constante

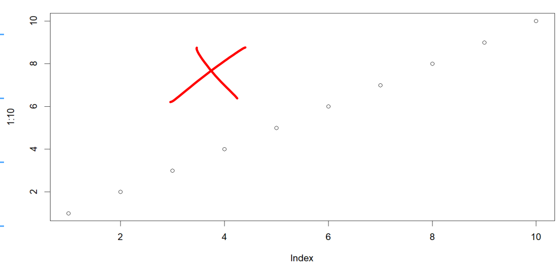
6) Pruebas de bondad de ajuste (Shapiro Wilk)

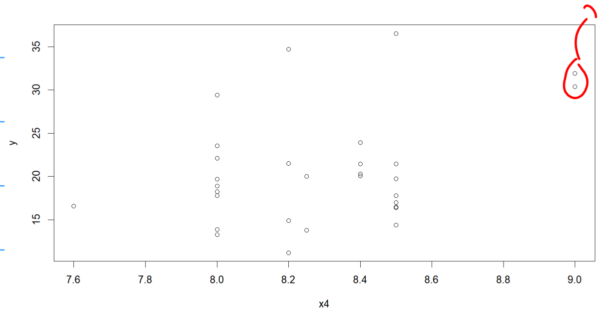
$$\begin{cases} H_0: \text{Los datos provienen de una población normal} \\ H_1: \text{Los datos NO provienen de una población normal} \end{cases}$$

Se rechaza Ho si p-valor $< \alpha$

8) Falta de ajuste

repetición





No se puede, no hay valores de x repetidos

Se puede porque hay repetición

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

Igual que en Anova

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	$SSLOF$ +	$m-2$ +	$MSLOF = \frac{SSLOF}{m-2}$	$\boxed{F_0 = \frac{MSLOF}{MSPE}}$
Error Puro	$SSPE$	$n-m$	$MSPE = \frac{SSPE}{n-m}$	
Total	SST	$n-1$		

iguales

Clave

Hipótesis
$$\begin{cases} H_0: E[Y|X_i] = \beta_0 + \beta_1 X_i \\ H_1: E[Y|X_i] \neq \beta_0 + \beta_1 X_i \end{cases}$$

```
Analysis of Variance Table

Response: y
            Df  Sum Sq Mean Sq F value  Pr(>F)
FO(x4)       1  157.32 157.325  4.3693 0.04517
Residuals   30 1080.22  36.007
Lack of fit  5  141.53  28.307  0.7539 0.59121
Pure error  25  938.69  37.547
```

P-valor de la prueba de significancia

↳ P valor de la falta de ajuste

¿Qué concluimos? Como el p-valor es enorme, no se rechaza Ho, es decir, el ajuste lineal está bien, además, el modelo es significativo a un nivel de significancia de 0.05.

9)

Modelo	Denominación	Transformación
$Y = \beta_0 e^{\beta_1 X}\varepsilon$ $\varepsilon \nsim$ Normal	Modelo exponencial multiplicativo	Se ajusta $Y^* = \beta_0^* + \beta_1 X + \varepsilon^*$ con $Y^* = \ln(Y)$ ↳ $\varepsilon^* \sim$ Normal
$Y = \beta_0 X^{\beta_1}\varepsilon$ $\nsim$ Normal	Modelo potencial multiplicativo	Se ajusta $Y^* = \beta_0^* + \beta_1 X^* + \varepsilon^*$ con $Y^* = \ln(Y)$ y $X^* = \ln(X)$ Normal
$Y = \beta_0 + \beta_1 \ln(X) + \varepsilon$	Modelo logarítmico	Se ajusta $Y = \beta_0 + \beta_1 X^* + \varepsilon$ con $X^* = \ln(X)$
$Y = \beta_0 + \beta_1 (1/X) + \varepsilon$	Modelo recíproco	Se ajusta $Y = \beta_0 + \beta_1 X^* + \varepsilon$ con $X^* = 1/X$

Nota: Verificar el residual una vez linealizado el modelo OJO

- Si $\ln(Y) \sim N(\mu, \delta^2) \Rightarrow Y \sim \text{log-normal}(\mu, \delta^2)$
- $\ln(\hat{Y}_i) = \beta_0^* + \beta_1 X_i + \varepsilon_i^* \Rightarrow Y_i = \hat{\beta}_0 e^{\hat{\beta}_1 X_i} \varepsilon_i$ NO es la recta media

porque e^{μ} es la mediana en una log-normal y la media $e^{\mu+\delta^2}$. Ahora si $\delta^2 \approx 0$ $e^{\mu} \approx e^{\mu+\delta}$

- $\hat{\beta}_0^* \neq \hat{\beta}_0$

```
Call:
lm(formula = log(sustancia) ~ t, data = datos.transform)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.12184 -0.08420  0.00987  0.06587  0.15086 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.425612   0.034957   40.78   <2e-16 ***
t           -0.801611   0.006005 -133.49   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0863 on 28 degrees of freedom
Multiple R-squared:  0.9984,	Adjusted R-squared:  0.9984 
F-statistic: 1.782e+04 on 1 and 28 DF,  p-value: < 2.2e-16
```

$\hat{\beta}_0^*$ y $\hat{\beta}_1^*$ $\Rightarrow \hat{\beta}_0 = e^{\hat{\beta}_0^*}$

- Todo se lee igual y por lo general se concluye en función del $\ln(y)$.
- El R^2 de un modelo normal y de uno transformado no son comparables.
- se analizan los residuales del modelo transformado.