ANÁLISIS DE LA VARIANZA
EN
REGRESIÓN LINEAL SIMPLE

Regresión Lineal Simple

Objetivo:

Hallar una función o un modelo matemático para predecir y estimar el valor de una variable a partir de valores de otra, ambas cuantitativas.

La variable Y: que es la dependiente (respuesta, predicha, endógena). Es la variable que se desea predecir o estimar y

la variable X: que es la independiente (predictora, explicativa, exógena). Es la variable que provee las bases para estimar.

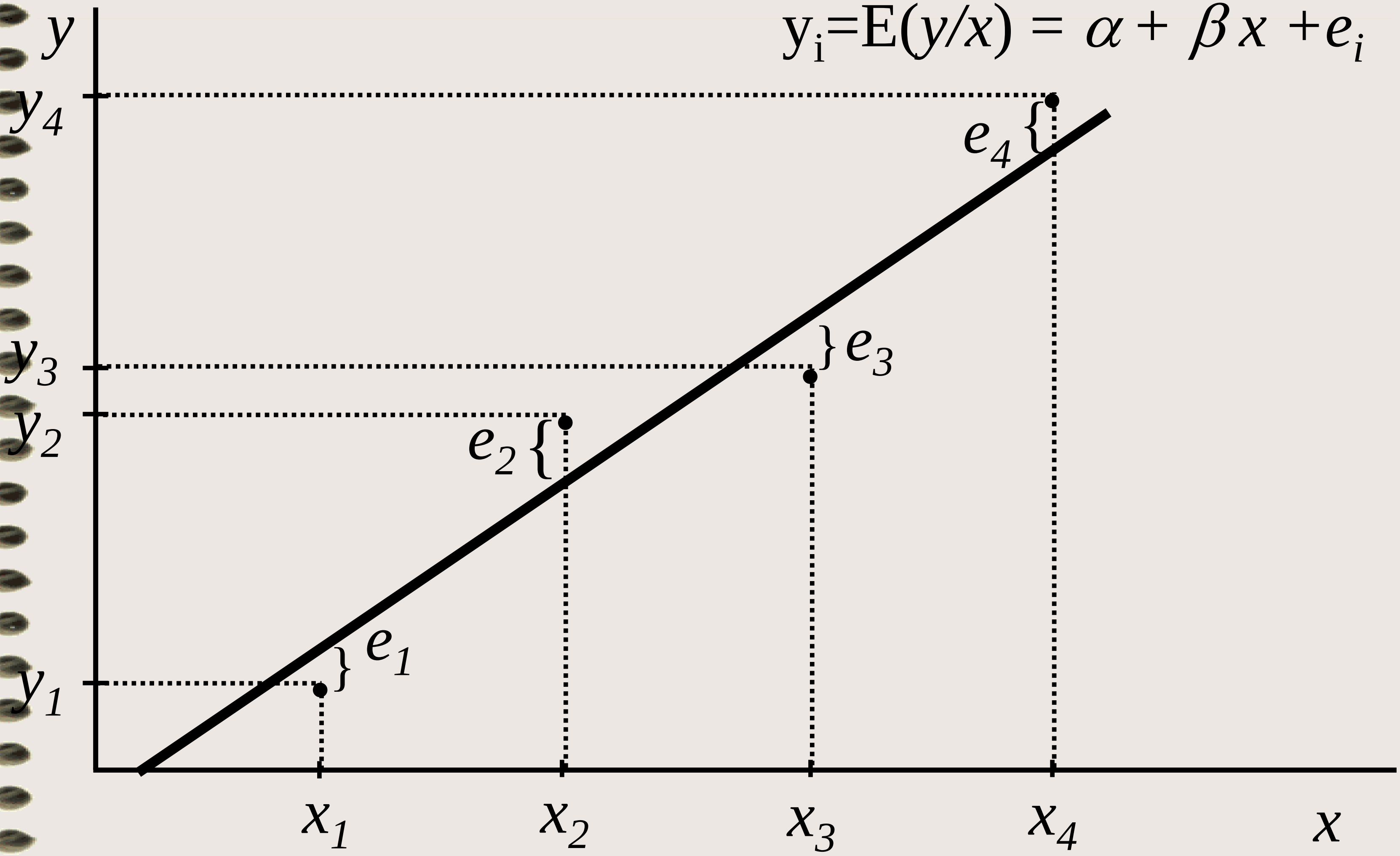
Modelo teórico:

$$\mu_{y/x} = \alpha + \beta x$$

- Este modelo implica que todas las medidas de las subpoblaciones de “y” están sobre la misma recta.
- α y β son los coeficientes de regresión de la población y geométricamente representan la ordenada al origen y la pendiente de la recta, respectivamente.

Regresión Lineal Simple

$$y_i = E(y/x) = \alpha + \beta x + e_i$$



Modelo estadístico

Modelo estimado:

$$\hat{y} = a + bx$$

Donde:

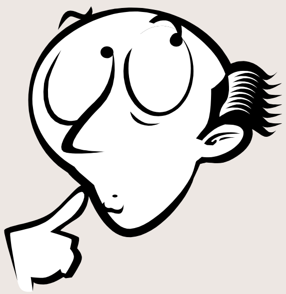
(a) es un estimador de α

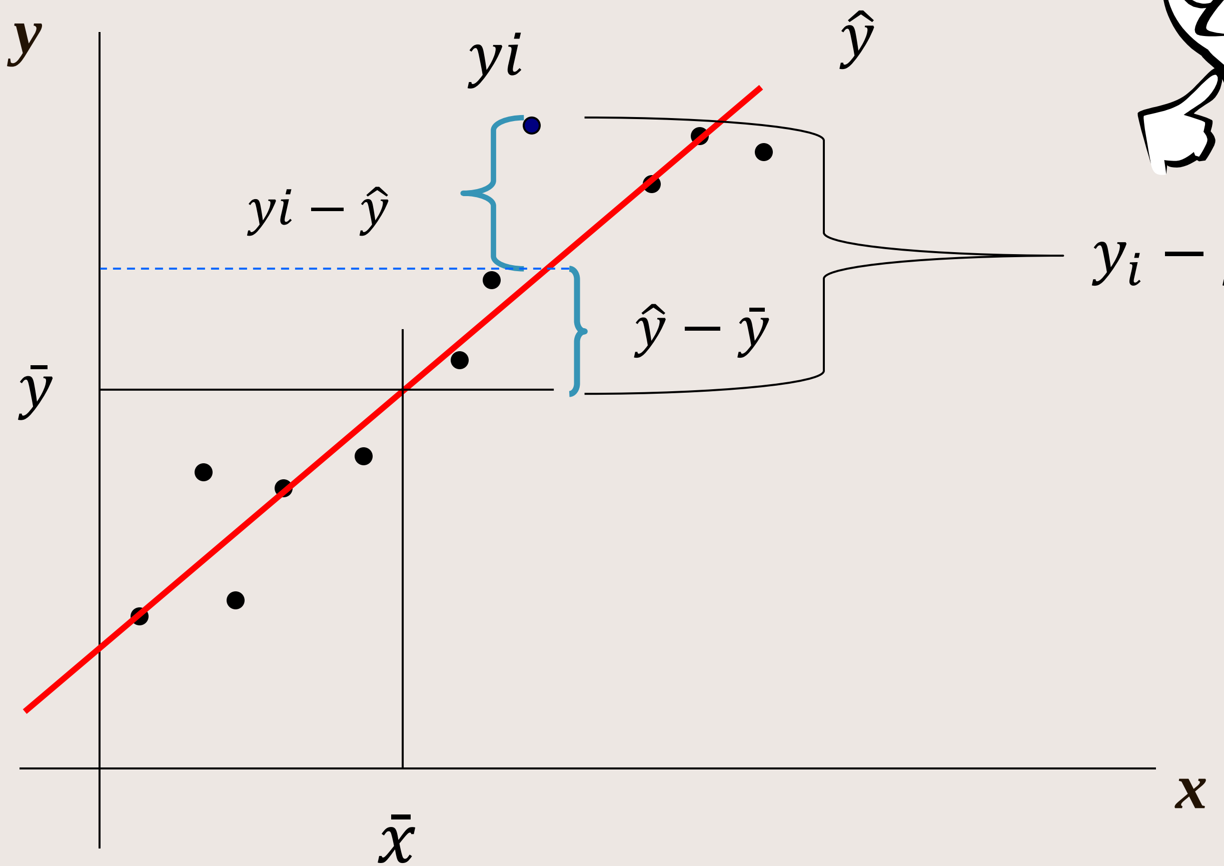
(b) es un estimador de β

Además $e \cong N(0, \sigma^2)$

ANALISIS DE LA VARIANZA EN REGRESION LINEAL SIMPLE

$$\hat{y} = a + bx$$





Análisis de Variancia en el análisis de regresión

- El enfoque desde el análisis de variancia se basa en la partición de sumas de cuadrados y grados de libertad asociados con la variable respuesta Y.
- La variación de los Y_i se mide convencionalmente en términos de las desviaciones

$$(Y_i - \bar{Y}_i)$$

- La medida de la variación total SC_{tot}, es la suma de las desviaciones al cuadrado

$$\sum (Y_i - \bar{Y}_i)^2$$

Desarrollo formal de la partición

Consideremos la desviación

$$(Y_i - \bar{Y}_i)$$

Podemos descomponerla en

$$\underset{T}{\underline{(Y_i - \bar{Y})}} = \underset{R}{\underline{(\hat{Y}_i - \bar{Y})}} + \underset{E}{\underline{(Y_i - \hat{Y}_i)}}$$

(T): desviación total

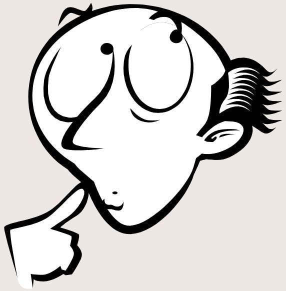
(R): es la desviación del valor ajustado por la regresión con respecto a la media general

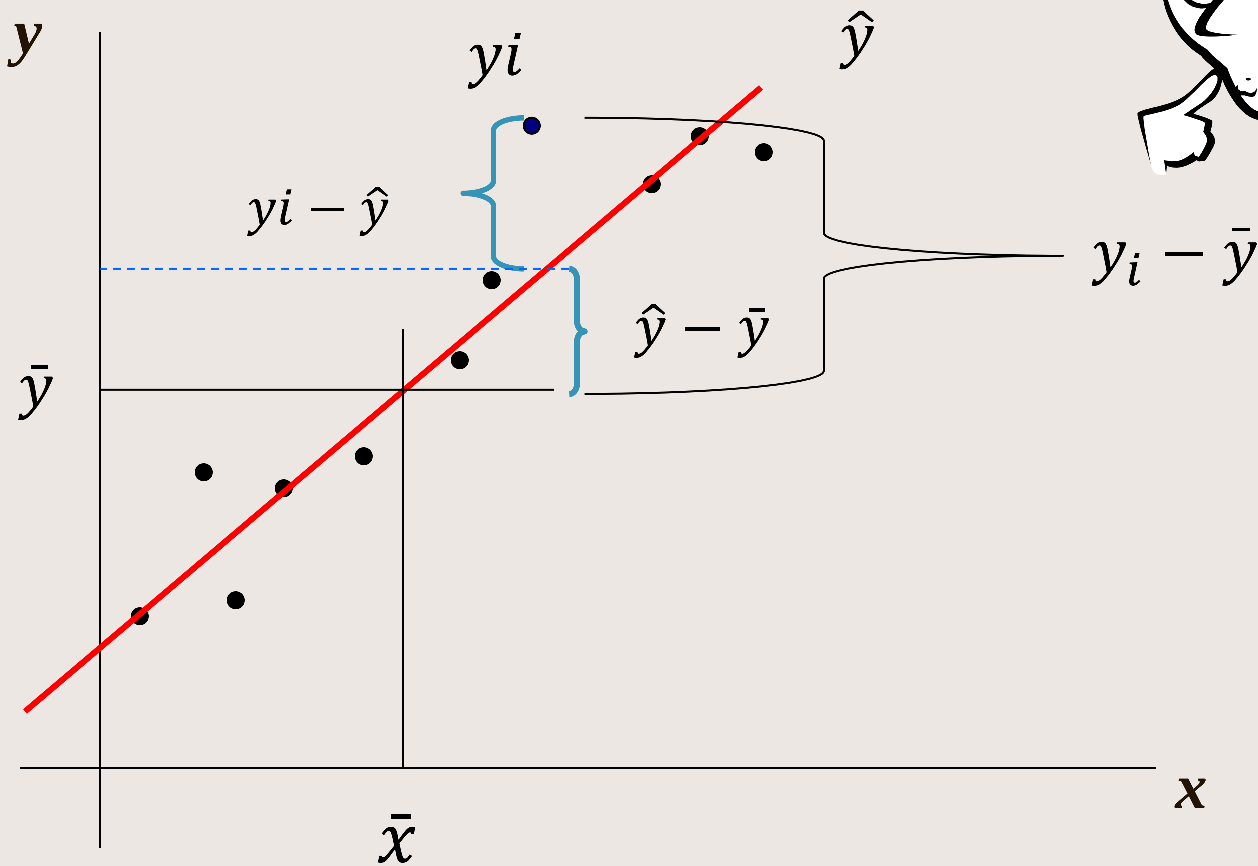
(E): es la desviación de la observación con respecto a la línea de regresión

REGRESION LINEAL SIMPLE

Estimar los valores de ***y*** (variable dependiente) a partir de los valores de ***x*** (variable independiente)

$$\hat{y} = a + bx$$





Desarrollo formal de la partición

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

Si consideramos todas las observaciones y elevamos al cuadrado para que los desvíos no se anulen:

$$\sum (Y_i - \bar{Y})^2 = \sum [\ (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2$$

Desarrollando el cuadrado del binomio obtenemos:

$$\sum (Y_i - \bar{Y})^2 = \sum (Y_i - \hat{y})^2 + 2\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum \left(\hat{Y}_i - \bar{y}\right)^2$$

Si consideramos en el segundo término: $e_i = y_i - \hat{y}_i$

$$\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum e_i(\hat{y}_i - \bar{y}) = \sum e_i(\hat{y}_i - \bar{y}) = \sum e_i(a + bx_i) - \bar{y} \sum e_i =$$

$$= a\sum e_i + b\sum e_i x_i = 0$$

e_i es independiente de x_i

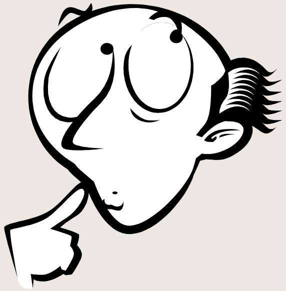
Aplicando propiedad distributiva el 2do término se anula quedando

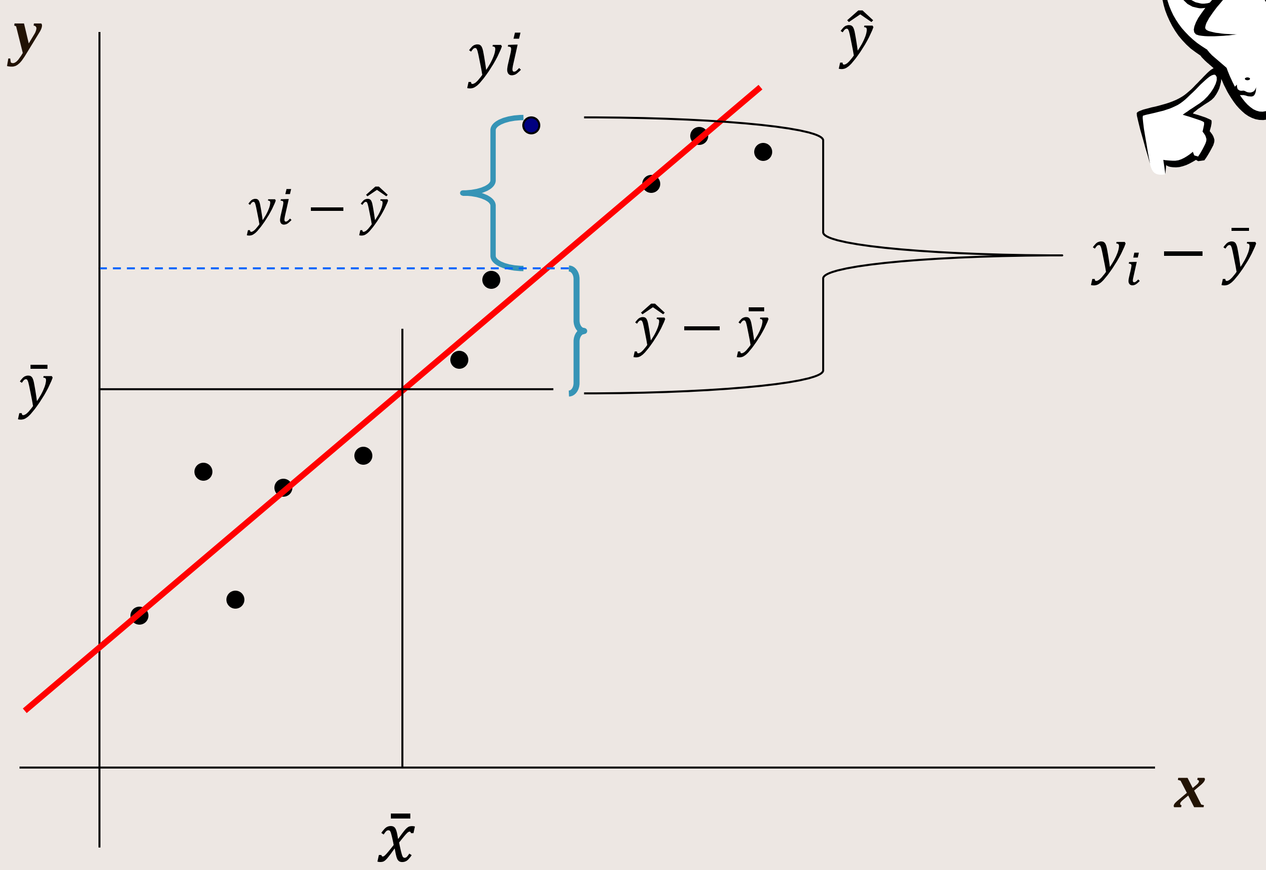
$$\sum (Y_i - \bar{Y})^2 = \sum (Y_i - \hat{y})^2 + \sum \left(\hat{Y}_i - \bar{y}\right)^2$$

REGRESION LINEAL SIMPLE

Estimar los valores de ***y*** (variable dependiente) a partir de los valores de ***x*** (variable independiente)

$$\hat{y} = a + bx$$





Desarrollo formal de la partición

Si consideremos todas las observaciones y elevamos al cuadrado para que los desvíos no se anulen

$$\underbrace{\sum (Y_i - \bar{Y})^2}_{SC_{tot}} = \underbrace{\sum (\hat{Y}_i - \bar{Y})^2}_{SC_{reg}} + \underbrace{\sum (Y_i - \hat{Y}_i)^2}_{SC_{er}}$$

(SC_{tot}): Suma de cuadrados total

(SC_{reg}): Suma de cuadrados de la regresión

(SC_{er}): Suma de cuadrados del error

Dividiendo por los grados de libertad, (n-1), (1) y

(n-2), respectivamente cada suma de cuadrados, se obtienen los cuadrados medios del análisis de variancia.

Cada un de estos cuadrados medios tiene una distribución Ji Cuadrado.

Suma de cuadrados debido a la regresión

$$SCregresión = \sum \left(\hat{Y}_i - \bar{y}\right)^2$$

Reemplazamos $\hat{y} = a + bx$

Obtenemos $SCregresión = \sum (a + bx_i - (a + b\bar{x}\ ))^2$

Cancelamos “a” $SCregresión = \sum (bx_i - b\bar{x})^2$

$$SCregresión = b^2 \sum (x_i - \bar{x})^2$$

Estimación de la variancia de los términos del error (σ^2)

$$SC_{error}=\sum(Y_i-\hat{y})^2=\sum(Y_i-a-bx_i)^2=\sum e_i{}^2$$

También se puede obtener por diferencia

$$SC_{error}=SC_{total}-SC_{regresión}$$

$$SC_{error}=\sum(Y_i-\hat{y})^2=\sum(Y_i-\bar{Y})^2-\sum\left(\hat{Y}_i-\bar{y}\right)^2$$

Reemplazando $SC_{error}=\sum(Y_i-\bar{Y})^2-b^2\sum(x_i-\bar{x})^2$

Estimación de la variancia de los términos del error (σ^2)

La suma de cuadrados del error, tiene n-2 grados de libertad asociados con ella, ya que se tuvieron que estimar dos parámetros.

Por lo tanto, las desviaciones al cuadrado dividido por los grados de libertad, se denomina cuadrados medios

$$CM_e = \frac{SC_e}{n-2} = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$$

Donde CM es el Cuadrado medio del error o cuadrado medio residual. Es un estimador insesgado de $\boldsymbol{\sigma^2}$

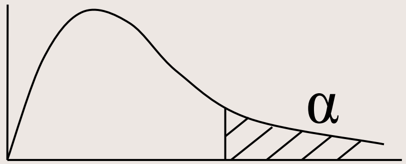
Tabla del análisis de varianza

Fuentes de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	E(CM)	F
Debido a la regresión	1	SCRegresión	CMReg= SCReg/1	$\sigma^2{}_e + \beta^2 \sum (x - \bar{x})^2$	$\frac{CMReg}{CMe}$
Debido al Error	n-2	SCerror	CMe= SCe/(n-2)	$\sigma^2{}_e$	
Total	n-1	SCTotal	-----	-------	

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

$\alpha = \ldots$

$$F_{1;(n-2)} \cong \frac{CM_{Regresión}}{CM_{Error}}$$



$F_{(1,\ n\text{-}2)}$

Regla de decisión

Rechazo H_0 si Fcal ≥ Ftabla

No rechazo H_0 si Fcal <Ftabla

Tabla del análisis de varianza

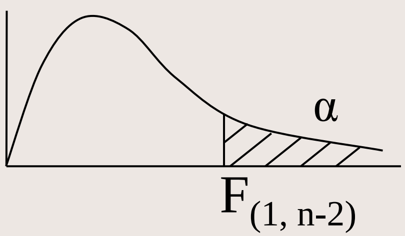
Fuentes de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	F
Debido a la regresión	1	$b^2 \sum (x_i - \bar{x})^2$	CMReg=SCReg/1	$\frac{CMReg}{CMe}$
Debido al Error	n-2	SCerror	CMe=SCe/(n-2)	
Total	n-1	$\sum (Y_i - \bar{Y})^2$	-----	

$$S^2{}_e = \frac{1}{n-2}\left(\sum (y_i - \bar{y})^2 - b^2 \sum (x_i - \bar{x})^2\right)$$

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

$\alpha = \ldots$

$$F_{1;(n-2)} \cong \frac{CM_{Regresión}}{CM_{Error}}$$



Regla de decisión

Rechazo H_0 si Fcal ≥ Ftabla

No rechazo H_0 si Fcal <Ftabla

Sea X el volumen de precipitación pluvial -lluvia (m^3) e Y el volumen de escurrimiento (m^3) en determinado lugar. Se realizaron 15 observaciones que se muestran en la tabla siguiente:

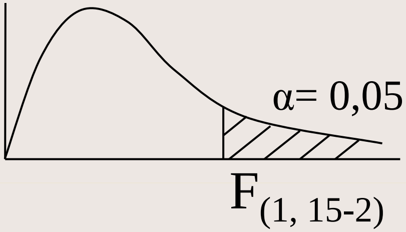
No. Observ.	X	Y
1	5	4
2	12	10
3	14	13
4	17	15
5	23	15
6	30	25
7	40	27
8	47	46
9	55	38
10	67	46
11	72	53
12	81	70
13	96	82
14	112	99
15	127	100

ANOVA en Regresión

$$\begin{cases} H_o: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$

α= 0,05

$$F_{1;(n-2)} \cong \frac{CM_{Regresión}}{CM_{Error}}$$



Fuentes de variación	Grados de libertad	Suma de cuadrados	Cuadrados Medios	F	Valor crítico de F
Regresión	1	14078,72	14078,7216	512,65	7,8961E-12
Residuos	13	357,0116	27,4624372		
Total	14	14435,73			

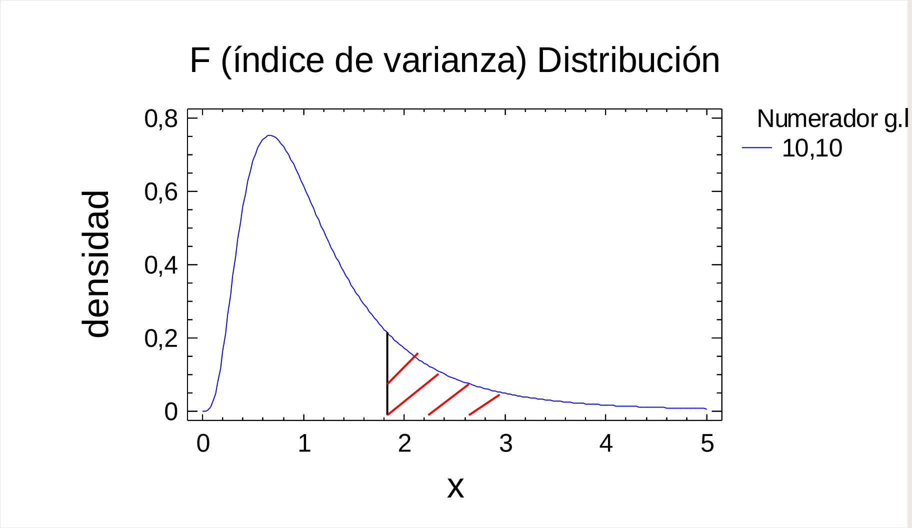
Ftabla: F1,13;0,95= 4,667. Como el valor de F calculado es mayor que 4,667 rechazo la hipótesis nula.

<u>**Conclusión**</u>: Con un nivel de significación del 5% tengo evidencias suficientes para suponer que existe una relación funcional poblacional donde, el volumen de escurrimiento está en función del volumen de precipitación pluvial, o que sea, por cada metro^3 que se incrementa la precipitación, se modifica o cambia el valor medio poblacional del volumen de escurrimiento.

F de Snedecor

La distribución F es una distribución continua que relaciona dos variables aleatorias independientes con distribuciones de chi-cuadrado, cada una dividida entre sus grados de libertad. La distribución F es asimétrica hacia la derecha y es descripta por los grados de libertad de su numerador (ν_1) y denominador (ν_2).

$$P(F_{n_1,n_2} > F_{n_1,n_2,p}) = p$$



n_1=Grados de libertad del numerador

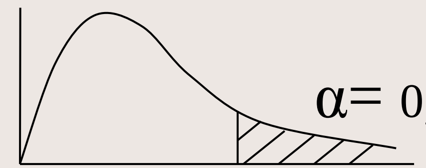
n_2=Grados de libertad del denominador

Está tabulada

$$F_{n_1,n_2,p}$$

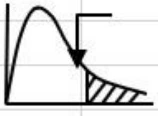
En ANOVA en regresión F=CMreg/CMerror

Búsqueda en la tabla F



Valores críticos de la distribución F (cola superior)

alfa= 0,05



	n_1 grados de libertad													
n_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**14**	**16**	**20**
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	242,98	243,91	245,36	246,46	248,01
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385	19,396	19,405	19,413	19,424	19,433	19,446
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,786	8,763	8,745	8,715	8,692	8,660
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,936	5,912	5,873	5,844	5,803
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,704	4,678	4,636	4,604	4,558
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,027	4,000	3,956	3,922	3,874
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,603	3,575	3,529	3,494	3,445
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388	3,347	3,313	3,284	3,237	3,202	3,150
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,102	3,073	3,025	2,989	2,936
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,943	2,913	2,865	2,828	2,774
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,818	2,788	2,739	2,701	2,646
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,717	2,687	2,637	2,599	2,544
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,635	2,604	2,554	2,515	2,459
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,565	2,534	2,484	2,445	2,388
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	2,507	2,475	2,424	2,385	2,328

Prueba de hipótesis para el coeficiente de regresión β

$$\begin{cases} H_o: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$

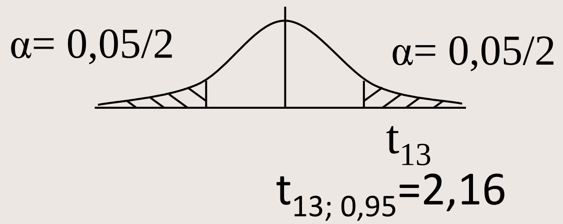
$$t = \frac{b - \beta}{S_b} \approx t_{(n-2)}$$

Regla de decisión

Rechazo H_0 si
tcal ≥ 2,16 ó
tcal ≤ -2,16

No rechazo H_0 si
-2,16< t cal < 2,16

α= 0,05



$$t_{calculado} = \frac{0,8269 - 0}{0,0365} = 22,64$$

6- Decisión: Como tcal es >2,16 Rechazo Ho.

Conclusión: Con un nivel de significación del 5% tengo evidencias suficientes para suponer que existe una relación funcional poblacional donde, el volumen de escurrimiento está en función del volumen de precipitación pluvial, o que sea, por cada metro^3 que se incrementa la precipitación se modifica o cambia el valor medio poblacional del volumen de escurrimiento.

Coeficiente de Determinación (R^2)

Esta suma de variaciones tiene una propiedad que permite escribir:

$$\sum(Y_i - \bar{Y})^2 = \sum(Y_i - \hat{y})^2 + \sum\left(\hat{Y}_i - \bar{y}\right)^2$$

$$SC_{total} = SC_{error} + SC_{regresión}$$

Dividimos a ambos miembros por SC_{total}

$$\frac{SC_{total}}{SC_{total}} = \frac{SC_{error}}{SC_{total}} + \frac{SC_{regresión}}{SC_{total}}$$

$$1 = \frac{SC_{error}}{SC_{total}} + \frac{SC_{regresión}}{SC_{total}} \leftarrow R^2$$

$$R^2 = \frac{\sum\left(\hat{Y}_i - \bar{y}\right)^2}{\sum(Y_i - \bar{Y})^2} = \frac{b^2\sum(x_i - \bar{x})^2}{\sum(Y_i - \bar{Y})^2}$$

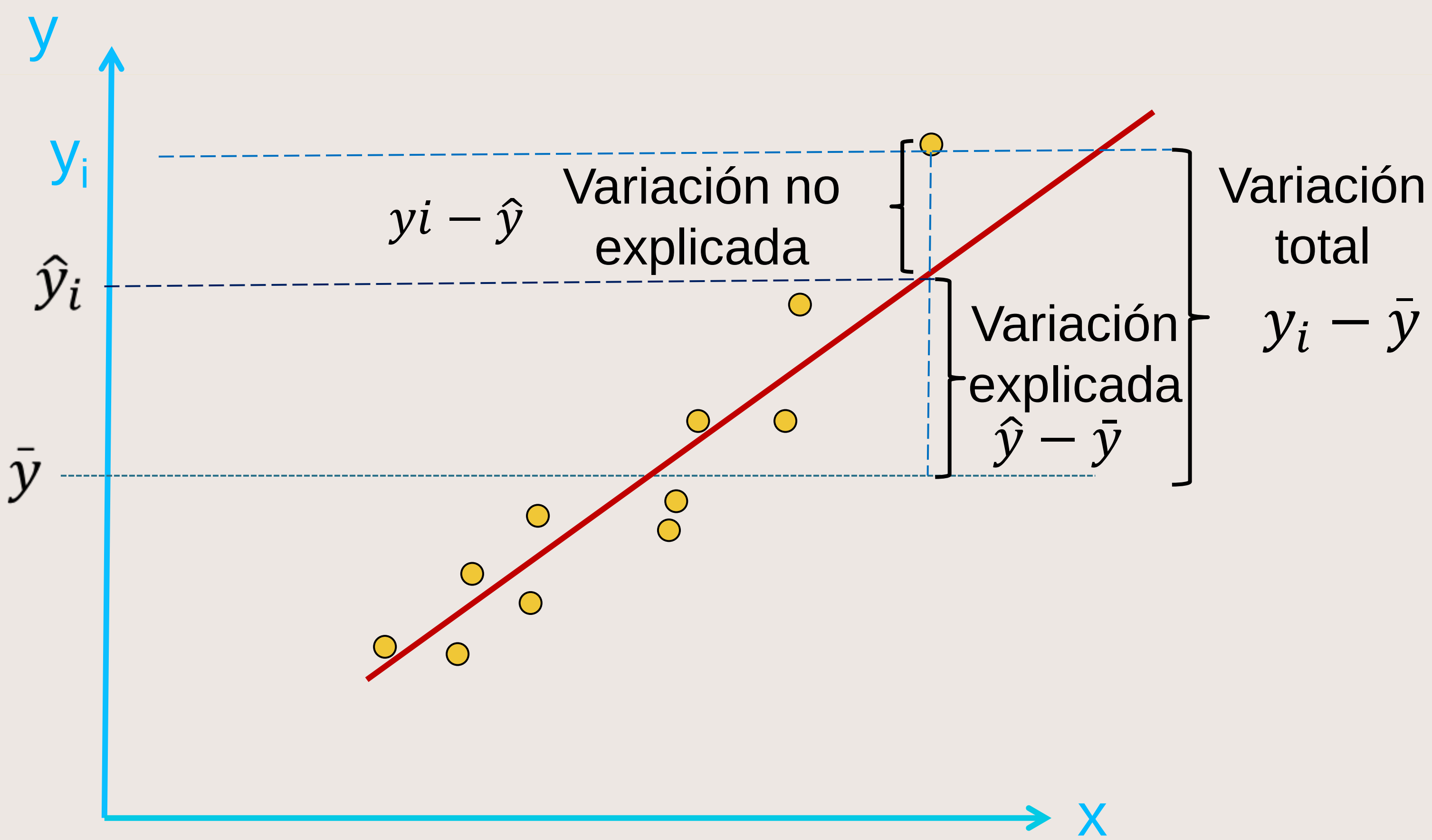
Coeficiente de Determinación (R^2)

La razón de la variación explicada a la variación total se llama Coeficiente de Determinación.

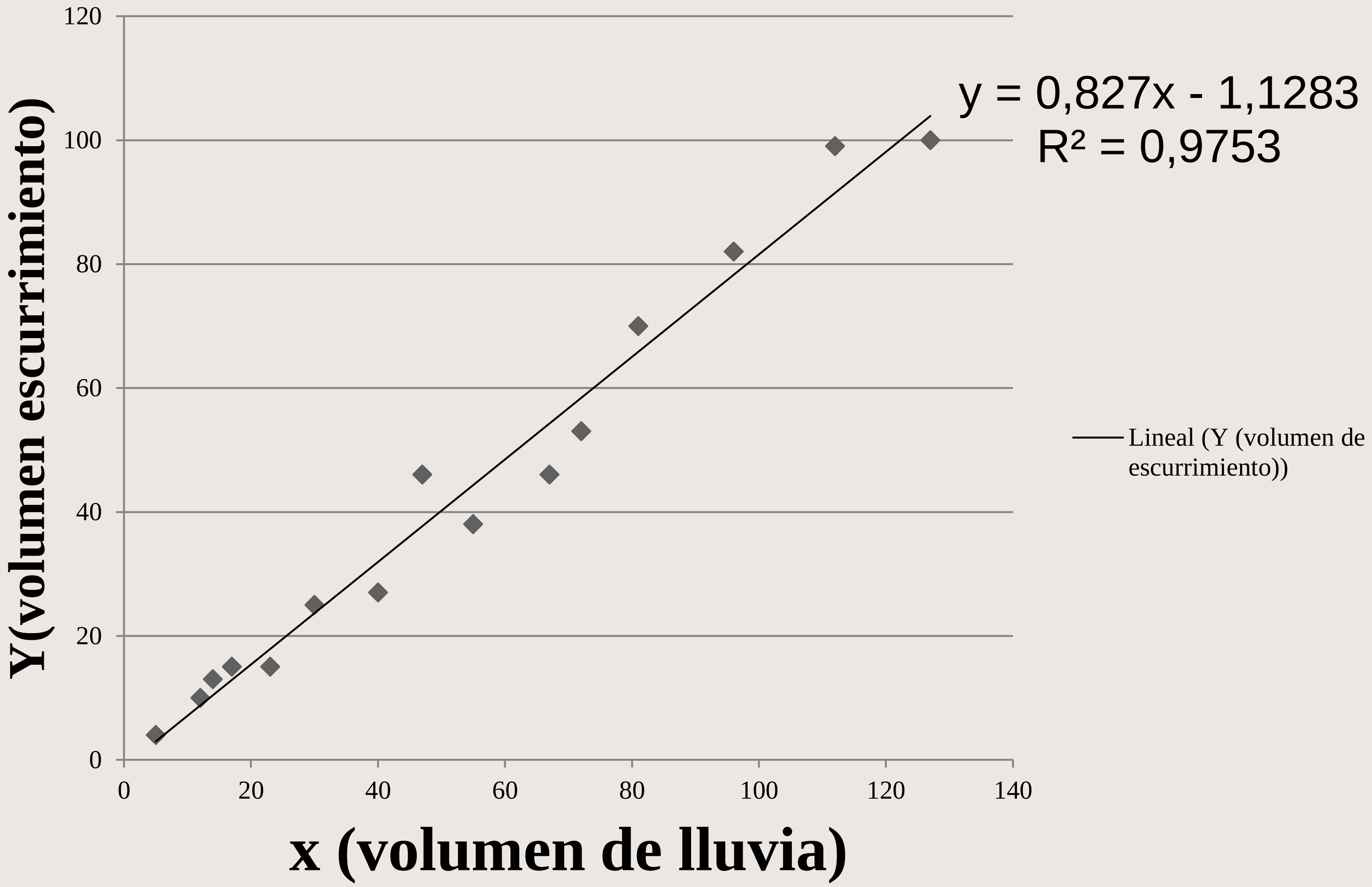
$$R^2 = \frac{\sum\left(\hat{Y}_i - \bar{y}\right)^2}{\sum(Y_i - \bar{Y})^2} = \frac{b^2 \sum(x_i - \bar{x})^2}{\sum(Y_i - \bar{Y})^2}$$

Si la variación explicada es cero, es decir, la variación total es toda no explicada, esta razón es cero.
Si la variación no explicada es cero, es decir, la variación total es toda explicada, la razón es uno.
En los demás casos **la razón se encuentra entre 0 y 1**. Puesto que la razón es siempre no negativa, se denota por R^2

Coeficiente de Determinación



$$R^2 = \frac{\sum\left(\hat{Y}_i - \bar{y}\right)^2}{\sum(Y_i - \bar{Y})^2} = \frac{b^2 \sum(x_i - \bar{x})^2}{\sum(Y_i - \bar{Y})^2}$$



120
100
80
60
40
20
0
0
20
40
60
80
100
120
140
Y(volumen escurrimiento)
x (volumen de lluvia)
y = 0,827x - 1,1283
R² = 0,9753
Lineal (Y (volumen de escurrimiento))

Coeficiente de Determinación

Es importante saber que el resultado del coeficiente de determinación oscila entre 0 y 1. Cuanto más cerca de 1 se sitúe su valor, mayor será el ajuste del modelo a la variable que estamos intentando explicar.

El coeficiente de determinación, se define como la proporción de la varianza total de la variable explicada por la regresión.

Calcule el coeficiente de determinación del ejemplo que se viene desarrollando.

Coeficiente de Determinación

$$R^2 = \frac{\sum\left(\hat{Y}_i - \bar{y}\right)^2}{\sum(Y_i - \bar{Y})^2} = \frac{b^2 \sum(x_i - \bar{x})^2}{\sum(Y_i - \bar{Y})^2}$$

Recordemos que: $\hat{y} = -1{,}1283 + 0{,}827x$

y que $\sqrt{\left(\sum x_i{}^2 - \frac{(\sum x_i)^2}{n}\right)} = \sqrt{20586{,}4}$ $\sqrt{\left(\sum y_i{}^2 - \frac{(\sum y_i)^2}{n}\right)} = \sqrt{14435{,}733}$

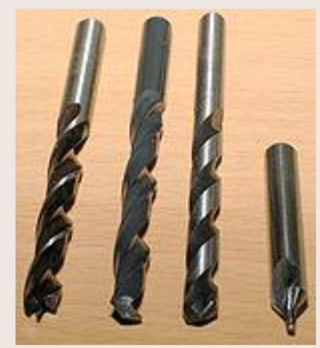
$$R^2 = \frac{b^2 \sum\left(\hat{Y}_i - \bar{y}\right)^2}{\sum(Y_i - \bar{Y})^2} = \frac{0{,}827^2 * 20586{,}4}{14435{,}733} = 0{,}9753$$

El 97,53% de la variabilidad está explicada por el modelo de regresión.

EJEMPLO

Un ingeniero encargado del área de calidad de una empresa manufacturera, desea analizar la vida útil de una herramienta de corte (el tiempo que mantiene una calidad aceptable de funcionamiento) para presentar un plan de reemplazo. Ya que sin duda, las herramientas de corte pueden determinar el éxito o fracaso de un proceso de mecanizado.

Las herramientas de corte más conocidas son: brocas, fresas, limas, sierras, herramientas de tornear, etc.



Fresa

Brocas helicoidales

Teniendo en cuenta que la vida útil se ve afectada por varios aspectos como: el ambiente operacional, las condiciones de producción o de mantenimiento y el desgaste presentado por su uso, decide comenzar a investigar la relación funcional entre la velocidad de corte (metros por minuto) y el tiempo de vida (horas de uso) de la herramienta. Para ello tomó herramientas nuevas, del mismo tipo, y a cada una (al azar) las sometió a diferentes velocidades de corte registrando en cada caso la vida útil en horas. Los datos recogidos se muestran en la tabla:

Velocidad (Metros por minuto)	Vida (Horas)
20	8,7
20	9,5
25	8,5
25	7,7
25	8,4
30	7,3
30	6,1
30	7,3
35	6,8
35	5,7
35	6,1
40	4,3
40	4,2

$\Sigma x = 390$ $\Sigma x^2 = 12250$ $\Sigma y = 90,6$ $\Sigma y^2 = 663,9$

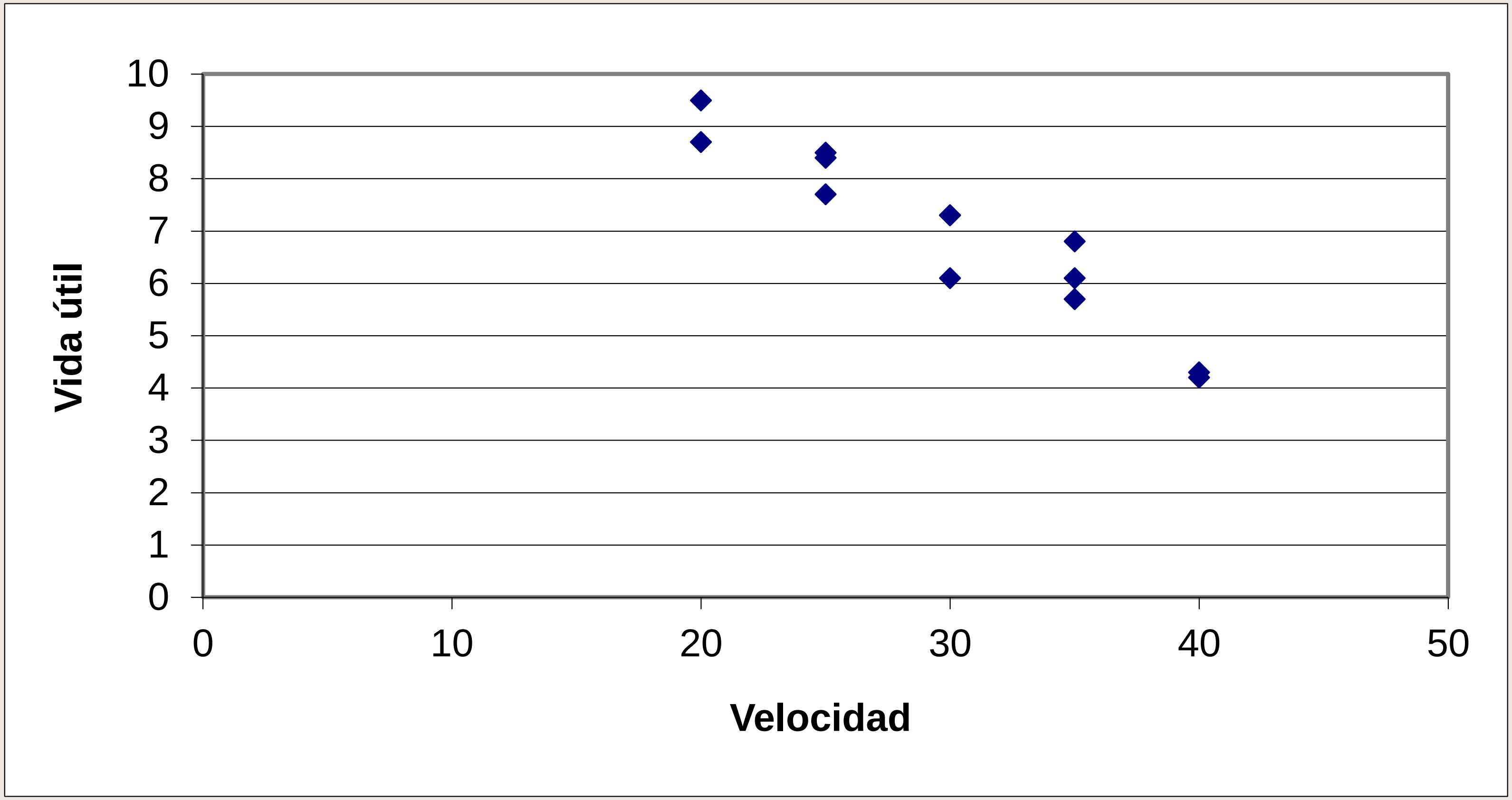
$n = 13$ $\Sigma xy = 2591$

Defina las variables. Realice el diagrama de dispersión. Analice la relación funcional entre estas dos variables.

Y: duración de la herramienta, es decir el tiempo de vida útil (en horas). Es la variable respuesta, dependiente, predicha o endógena. Es la variable que se desea predecir o estimar.

X: velocidad de corte (en m/min), que es la independiente (predictora, explicativa, exógena). Es la variable que provee las bases para estimar. En este caso la variable independiente es controlable, es decir la fija el experimentador o el operario de la máquina.

Dibujar el diagrama de dispersión.



Se observa que cuando aumenta la velocidad de corte (metros/minuto) disminuye el tiempo de vida (horas) de la herramienta. También se observa que la relación funcional podría ser estimada por una recta.

Suponiendo que la población sobre la que se ha tomado la muestra (de tiempos de vida tiene distribución normal para cada velocidad prefijada) realice las siguientes actividades:

a)Halle la recta de regresión por el método de los mínimos cuadrados luego dibújela en el diagrama de dispersión.

n=13 Σx = 390 Σx² =12250 Σy = 90,6 Σy² = 663,9 Σxy = 2591

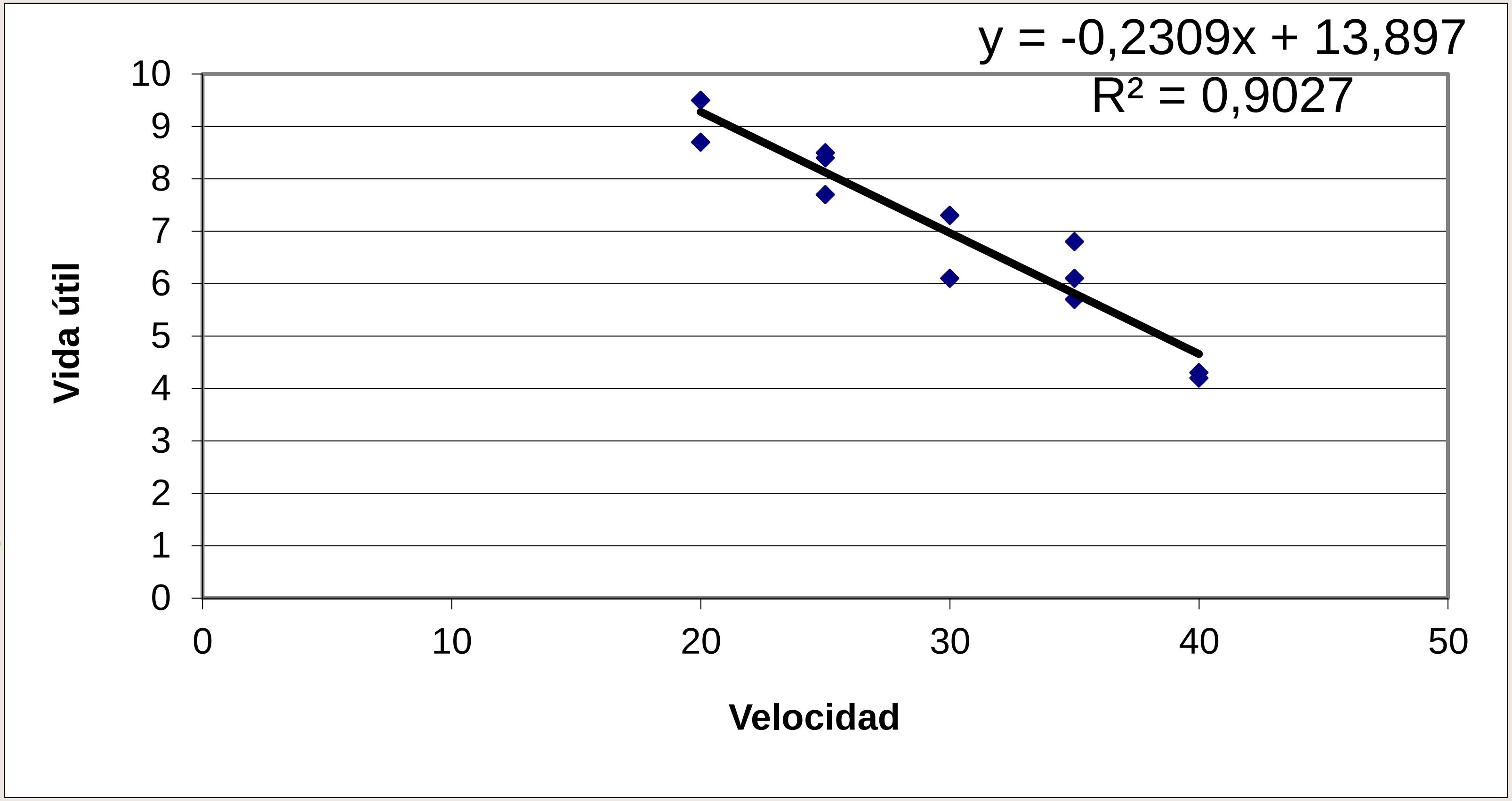
$$b = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}$$

$$b = \frac{2591 - \frac{390 * 90,6}{13}}{12250 - \frac{390^2}{13}} = \frac{-127}{550} = -0,2309$$

$$a = \bar{y} - b\bar{x}$$

$$a = \frac{90,6}{13} - 0,2309 * \frac{290}{13} = 13,897$$

$$\hat{y} = 13,897 - 0,2309x$$



y = -0,2309x + 13,897
R² = 0,9027
Vida útil
Velocidad
0
1
2
3
4
5
6
7
8
9
10
0
10
20
30
40
50

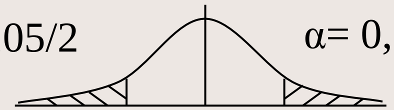
Pruebe la hipótesis más importante para decidir si continúa con el análisis del problema de regresión. Concluya e interprete. Use $\alpha = 0{,}05$.

La Prueba de hipótesis más importante es la del para el coeficiente de regresión β

$$\begin{cases} H_o: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$

$$t = \frac{b - \beta}{S_b} \approx t_{(n-2)}$$

$\alpha = 0{,}05/2$ $\alpha = 0{,}05/2$



t_{11}

$\alpha = 0{,}05$

$t_{11;\,0{,}975} = 2{,}201$

<u>Regla de decisión</u>

Rechazo H_0 si tcal $\geq 2{,}201$ ó tcal $\leq -2{,}201$

No rechazo H_0 si
$-2{,}201 < t\ cal < 2{,}201$

Cálculos

$$S^2_e = \frac{1}{n-2}\left[\sum y_i{}^2 - \frac{(\sum y_i)^2}{n} - b\left(\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}\right)\right]$$

$$S^2{}_e = \frac{1}{11}\left[663{,}9 - \frac{90{,}6^2}{13} - (-0{,}2309) * (-127)\right] = 0{,}287$$

$$S^2_b = \frac{S_e^2}{\sum(x-\bar{x}^2)^2} = \frac{S_e^2}{\sum x^2 - (\sum x)^2/n} = \frac{0{,}287}{550} = 0{,}000522$$

$S^2_e = 0{,}287$ $S_b = 0{,}02286$

$t_{cal} = (-0{,}2309 - 0)/0{,}02286$ $t_{cal} = -10{,}1006$

Como $t_{cal} < -2{,}201$ rechazo Ho

Conclusión: Con un nivel de significación del 5 % tengo evidencias suficientes para suponer que existe una relación funcional poblacional del tiempo de vida útil de la herramienta en función de la velocidad de corte, o que sea, por cada metro/minuto que se incrementa la velocidad de corte se modifica o cambia el valor medio poblacional del tiempo de vida útil de la herramienta.

Obtenga un intervalo de confianza del 95% para el coeficiente de regresión.

Intervalo de confianza para el coeficiente de Regresión

$$P\left(b - t_{n-2;1-\alpha/2} S_b < \beta < b + t_{n-2;1-\alpha/2} S_b\right) = 1-\alpha$$

$$(-0{,}281214 < \beta < -0{,}180585)$$

Con una confianza de 95 %, podría decir que el intervalo (-0,2812 ; -0,1805) horas/(metros/minuto) encerraría al verdadero valor de la pendiente de la recta de regresión. Esto es, con una confianza de 95 %, podría decir que el intervalo (-0,2812 ; -0,1805) horas/(metros/minuto) encerraría al verdadero cambio del promedio poblacional del tiempo de vida de la herramienta, para un aumento unitario en la velocidad de corte.

- A continuación se muestra la tabla de análisis de la varianza obtenida con el uso del Excel empleando el menú Herramientas, análisis de datos, regresión

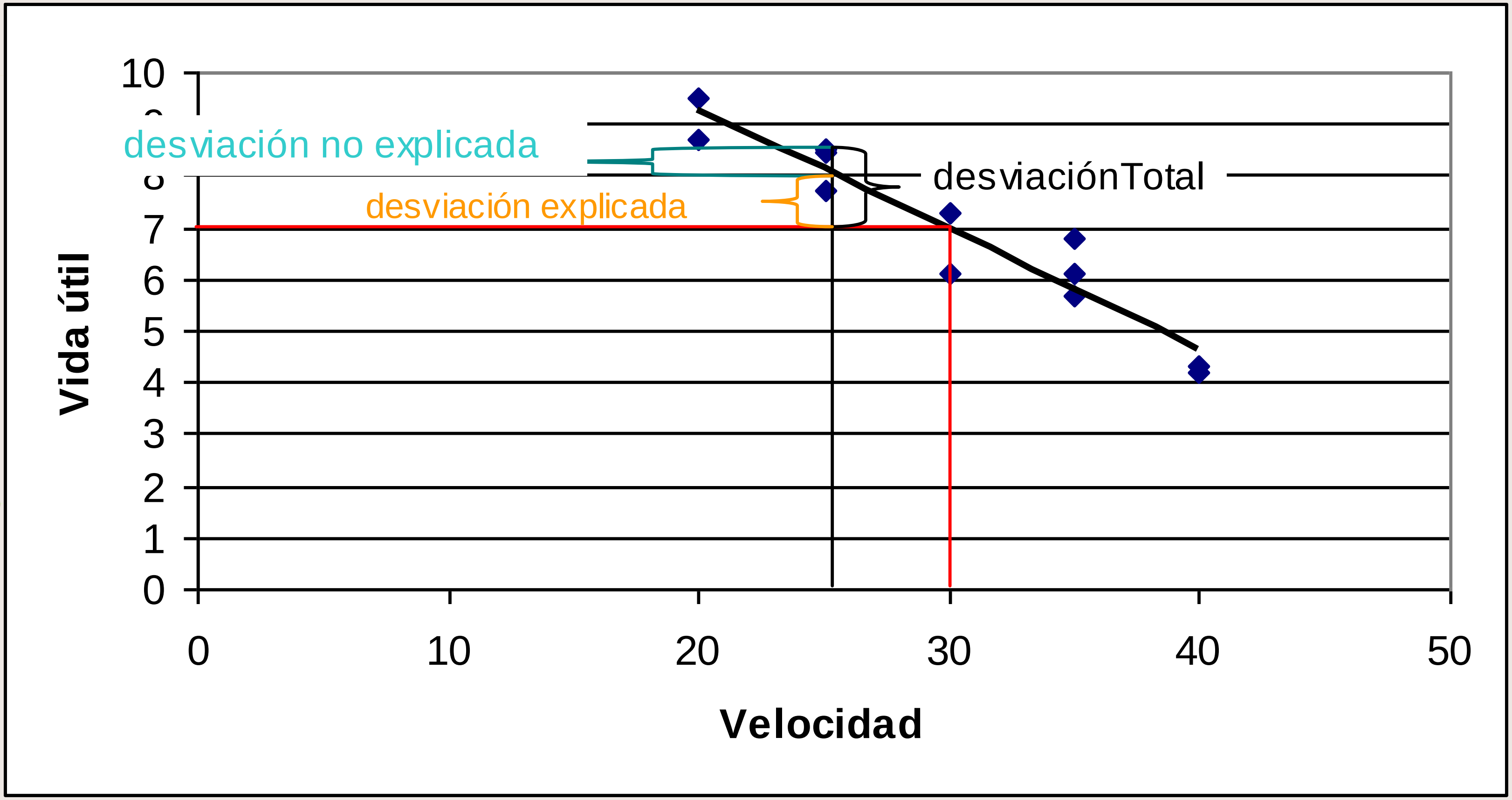
ANÁLISIS DE VARIANZA				
	Grados de libertad	Suma de cuadrados	Cuadrados Medios	F
Regresión	1	29,3254545	29,3254545	102,01004
Error	11	3,16223776	0,28747616	
Total	12	32,4876923		

n=13 $\Sigma x = 390$ $\Sigma x^2 = 12250$ $\Sigma y = 90,6$ $\Sigma y^2 = 663,9$ $\Sigma xy = 2591$

$S^2_e = 0,287$ $S^2_y = 32,487$ $S^2_x = 550$

- Seleccione un punto del diagrama de dispersión e indique la desviación sin explicar, la explicada y la total.
- Plantee las hipótesis correspondientes.
- Plantee la regla de decisión con α= 0,05
- Concluya en términos del problema empleando los valores que se observan en la tabla.

d) Compare los resultados con el item b).



desviación no explicada
desviaciónTotal
desviación explicada
Vida útil
Velocidad
10
7
6
5
4
3
2
1
0
0
10
20
30
40
50

ANOVA en Regresión

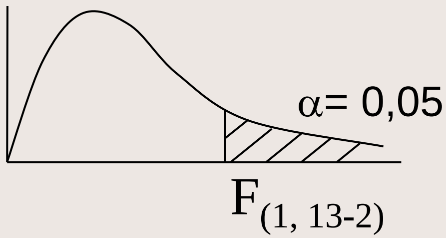
$$\begin{cases} H_o: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$

$\alpha = 0{,}05$

$$F_{1;(n-2)} \cong \frac{CM_{Regresión}}{CM_{Error}}$$

ANÁLISIS DE VARIANZA

	Grados de libertad	*Suma de cuadrados*	*Promedio de los cuadrados*	*F*	*Valor crítico de F*
Regresión	1	29,3254545	29,3254545	102,01004	6,6927E-07
Residuos	11	3,16223776	0,28747616		
Total	12	32,4876923			



Regla de decisión

F (1 ;11) y una probabilidad a la derecha de 0,05.

Rechazo Ho si $F_{cal} \geq 4{,}844$

No rechazo Ho si $F_{cal} < 4{,}844$

Conclusión: Con un nivel de significación del 5 % tengo evidencias suficientes para suponer que existe una relación funcional poblacional del tiempo de vida útil de la herramienta en función de la velocidad de corte, o que sea, por cada metro/minuto que se incrementa la velocidad de corte se modifica o cambia el valor medio poblacional del tiempo de vida útil de la herramienta.

Se obtiene la misma respuesta con ambos métodos. Esto se debe, y teóricamente se demuestra que una (distribución t con n-2 grados de libertad)^2 es equivalente a una distribución F con 1 y n-2 grados de libertad.

Si toma el valor crítico de la t= 2,2010 y lo eleva al cuadrado obtiene el valor crítico de la F= 4,844401. Si ambas pruebas se realizan con un software estadístico veremos que los p valores son iguales.

Fin

Gracias

Gracias