# Estadística Descriptiva Bivariada

#### Dr. Cristian Carvajal Muquillaza

Universidad de Playa Ancha



DEPARTAMENTO DISCIPLINARIO DE MATEMÁTICA Y ESTADÍSTICA



### Contenido

- 1 Preliminares
- 2 Medidas de Dependencia
- 3 Medidas de Asociación
- 4 Modelo de Regresión
- 5 Inferencia en el Modelo de Regresión



DEPARTAMENTO DISCIPLINARIO DE MATEMÁTICA VESTADÍSTICA



# Estadística Descriptiva Bivariada

#### Estadística Descriptiva Bivariada

La **Estadística Descriptiva Bivariada** aborda el estudio de los sucesos en los que intervienen dos variables simultáneamente, por ejemplo:

- Edad y Género,
- Escolaridad e Ingreso,
- Peso y Estatura,
- Estado civil y Nacionalidad,



DEPARTAMENTO DISCIPLINARIO DE MATEMÁTICA



#### Tabla de Contingencia

Medidas de Dependencia

Sean las variables A y B, tal que,  $A = a_1, ..., a_I$  y  $B = b_1, ..., a_I$ , entonces:

		$b_1$	 $b_j$	 $b_J$	
	$a_1$	$n_{11}$	 $n_{1j}$	 $n_{1J}$	$n_{1.}$
Variable A	•••				
	$a_i$	$n_{i1}$	 $n_{ij}$	 $n_{iJ}$	$n_{i.}$
Φ <b>&gt;</b>					
	$a_I$	$n_{I1}$	 $n_{Ij}$	 $n_{IJ}$	$n_{I.}$
		n <sub>.1</sub>	$n_{.j}$	$n_{J}$	n

**AMENTO** MÁTICA **ÍSTICA** 



# Estadística Descriptiva Bivariada

### Tabla de Contingencia: **Distribuciones Marginales**

Sean las variables A y B, tal que,  $A = a_1, ..., a_I$  y  $B = b_1, ..., a_I$ , entonces:

■ Variable A:

$$n_{i.} = \sum_{j=1}^{J} n_{ij}$$

Variable B:

$$n_{.j} = \sum_{i=1}^{I} n_{ij}$$





# Estadística Descriptiva Bivariada

#### Tabla de Contingencia: **Distribuciones Condicionales**

Sean las variables A y B, tal que,  $A = a_1, ..., a_I$  y  $B = b_1, ..., a_I$ , entonces:

1 Variable A condicionada a  $B = B_I$ :

$$n_{i|j} = n_{ij} | n_{.j}$$

2 Variable B condicionada a  $A = A_I$ :

$$n_{j|i} = n_{ij}|n_{i.}$$



Facultad de Ciencias Naturales y Exactas





# Medidas de Dependencia: Variables Cualitativas

#### Contraste de Independencia Chi-Cuadrado:

Suponga una tabla de contingencia  $I \times J$ , y suponga que se quiere contrastar la Hipótesis Nula de Independencia, es decir:

Modelo de Regresión

$$H_0: p_{ij} = p_{i.} \cdot p_{.j}$$
  $V/S$   $H_1: p_{ij} \neq p_{i.} \cdot p_{.j}$ 

Para contrastar la hipótesis utilizamos la cantidad pivotal de Pearson, que se define:

$$X^{2} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \hat{m_{ij}})^{2}}{\hat{m_{ij}}}$$

Donde

$$\hat{m_{ij}} = \frac{n_{i.} \times n_{.j}}{n}$$

MÁTICA STICA

# Medidas de Dependencia: Variables Cualitativas

#### Contraste de Independencia Chi-Cuadrado:

Bajo la Hipótesis Nula el estadístico  $X^2$  se distribuye asintóticamente Chi-Cuadrado con  $(I-1) \times (J-1)$  grados de libertad, es decir:

Modelo de Regresión

$$X^{2} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \hat{m}_{ij})^{2}}{\hat{m}_{ij}} \sim \chi^{2}_{(I-1)\times(J-1)}$$

Entonces se rechaza la Hipótesis Nula de Independencia al nivel  $\alpha$ , cuando se verifica

Se Rechaza  $H_0$ 

$$X_{obs}^2 \ge \chi_{\alpha;(I-1)\times(J-1)}^2$$



# Medidas de Dependencia: Variables Cualitativas

Medidas de Asociación

### Ejemplo: Contraste de Independencia Chi-Cuadrado

Se realiza un estudio descriptivo en un colegio, sobre la nacionalidad de los estudiantes desglosado por genero, los resultados se pueden observar en la tabla de contingencia adjunta:

		NACIONALIDAD					
		Chileno	Chileno Venezolano Haitiano Peruc		Peruano	Total	
GENERO	Hombre	150	85	47	15		
RO	Mujer	130	96	39	28		
/							

**AMENTO** Determine si las variables son Independientes, con un nivel de Con-NARIO fianza del 95 %.

MÁTICA STICA

#### Covarianza:

Sean  $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$  una muestra de n datos bivariados, entonces la covarianza es un estadístico que permite medir la fuerza de asociación entre las variables X e Y, y se define:

$$Cov(X,Y) = S_{XY} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

- Si  $S_{XY} > 0 \Rightarrow$  Asociación directa
- Si  $S_{XY} < 0 \Rightarrow$  Asociación inversa
- Si  $S_{XY} = 0 \Rightarrow$  No existe Asociación

Naturales y Exactas | YESTADÍSTICA



#### Covarianza: Propiedades

Sean X, Y variables aleatorias y a, b constantes, entonces:

- $\mathbf{II}$  Cov(X,a)=0
- $\mathbf{Cov}(X,X) = Var(X)$
- Cov(X,Y) = Cov(Y,X)
- 4 Cov(aX, bY) = abCov(X, Y)
- $5 \quad Cov(X+a,Y+b) = Cov(X,Y)$





#### Coeficiente de Correlación de Pearson:

Sean  $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$  una muestra de n datos bivariados, entonces el Coeficiente de Correlación de Pearson  $r_{xy}$  es un estadístico que mide el grado de asocioción lineal entre dos variables cuantitativas, y se define:

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

**Obs:**  $-1 \le r_{xy} \le 1$ 



AMENTO SCIPLINARIO MATEMÁTICA STADÍSTICA



Medidas de Asociación 00000000

#### Coeficiente de Correlación de Pearson:

Si  $r_{xy} = 1$  entonces existe una correlación lineal directa perfecta entre X e Y.



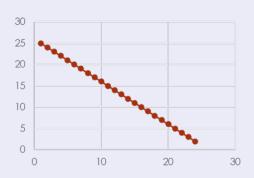
MÁTICA STICA



Medidas de Asociación 00000000

#### Coeficiente de Correlación de Pearson:

Si  $r_{xy} = -1$  entonces existe una correlación lineal inversa perfecta entre X e Y.



MÁTICA STICA



00000000

#### Coeficiente de Correlación de Pearson:

Si  $r_{xy} = 0$  entonces No existe una correlación lineal entre **X** e **Y**.









#### Coeficiente de Correlación de Pearson:

Es poco probable que  $r_{xy}$  tome exáctamente uno de estos tres valores -1, 0, 1, lo que se da en la práctica es que tome un valor que pertenezca al intervalo [-1, 1], entonces:

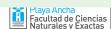


### Ejemplo:

Se quiere estudiar la relación existente entre el grado de percepción de los estudiantes del clima escolar y el rendimiento en la asignatura de matemática, los resultados de la encuesta se ve en la tabla adjunta:

Percepción	10	8	4	12	7	8	15	3
Nota	5.2	5.0	3.4	6.5	5.0	6.0	6.8	4.5

- 1 Construir una gráfica de dispersión de los datos,
- 2 Calcular e interpretar  $r_{xy}$ .





#### Modelo de Regresión Lineal Simple

El objetivo de un **modelo de regresión lineal simple** es tratar de explicar la relación causal que existe entre una variable X y Y, las cuales pueden tener las siguientes denominaciones:

Y	X
Dependiente	Independiente
Explicada	Explicativa
Predicha	Predictora
Respuesta	Regresora

Téngase presente que la dirección de causalidad es que X explica a Y. MÁTICA

Naturales y Exactas | Y ESTADÍSTICA

#### Modelo de Regresión Lineal Simple

El modelo de regresión lineal simple tiene la siguiente expresión:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

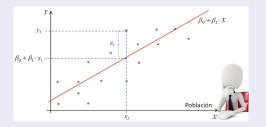
En donde  $\beta_0$  es la ordenada en el origen,  $\beta_1$  es la pendiente de la recta y  $\varepsilon$  es una variable que incluye un conjunto grande de factores, cada uno de los cuales influye en la respuesta sólo en pequeña magnitud, a la que llamaremos error.





#### Modelo de Regresión Lineal Simple

Gráficamente, un modelo de regresión lineal se representa como una recta, tal como se observa en la figura. Asumiremos que se trata de una representación á nivel poblacional.



MÁTICA

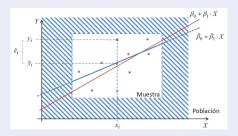




#### Modelo de Regresión Lineal Simple

En la práctica no se puede observar el modelo poblacional, sino que sólo una parte de éste a través de una muestra, lo cual llamaremos modelo estimado.

Medidas de Asociación



MÁTICA



#### Estimación del Modelo de Regresión Lineal Simple

Para estimar el modelo, debemos estimar los coeficientes  $\beta_0$  y  $\beta_1$ , para esto utilizaremos el método de estimación denominado "Mínimos Cuadrados Ordinarios" (MCO), el cual intenta minimizar la suma de los cuadrados de los errores  $(\varepsilon_i^2)$ .

Asumiendo que los  $\varepsilon_i$  son v.a.i.i.d con  $\mathbb{E}[\varepsilon_i] = 0$  v  $Var[\varepsilon_i] = \sigma^2$ , para todo i. Entonces, considerando i = 1, ..., n:

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$$





### Estimación del Modelo de Regresión Lineal Simple

Se plantea el siguiente proceso de optimización:

$$\begin{split} \hat{\beta}_0, \hat{\beta}_1 &= \arg\min\left(\sum_{i=1}^n \varepsilon_i^2\right) \\ &= \arg\min\left(\sum_{i=1}^n (Y_i - \hat{Y}_i)^2\right) \\ &= \arg\min\left(\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2\right) \\ &= \arg\min\Psi \end{split}$$

AMENTO NARIO MÁTICA ÍSTICA



#### Estimación del Modelo de Regresión Lineal Simple

El propósito será resolver el siguiente sistema de ecuaciones:

$$\frac{\partial \Psi}{\partial \hat{\beta}_0} = 0$$

$$\frac{\partial \Psi}{\partial \hat{\beta}_1} = 0$$





#### Estimación del Modelo de Regresión Lineal Simple

Condición de primer orden (CPO)

$$\frac{\partial \Psi}{\partial \hat{\beta}_0} = \sum_{i=1}^n 2(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) = 0$$
 
$$\frac{\partial \Psi}{\partial \hat{\beta}_1} = \sum_{i=1}^n 2(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i) = 0$$



DE MATEMÁTICA



$$\sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) = 0$$

$$\sum_{i=1}^{n} y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_1 x_i = 0$$

$$n\bar{y} - n\hat{\beta}_0 - n\hat{\beta}_1 \bar{x} = 0$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
Fully a Ancha Facultad de Ciencias Naturales y Exactas

$$\sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i) = 0$$

$$\sum_{i=1}^{n} x_i Y_i - x_i \hat{\beta}_0 - \hat{\beta}_1 x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_1 x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$
Universidad de Playa Ancha Flaya Ancha Flaya

000000000000

$$\begin{split} \sum_{i=1}^{n} x_{i} y_{i} - \sum_{i=1}^{n} x_{i} (\bar{y} - \hat{\beta}_{1} \bar{x}) - \hat{\beta}_{1} \sum_{i=1}^{n} x_{i}^{2} &= 0 \\ \sum_{i=1}^{n} x_{i} y_{i} - n \bar{x} (\bar{y} - \hat{\beta}_{1} \bar{x}) - \hat{\beta}_{1} \sum_{i=1}^{n} x_{i}^{2} &= 0 \\ \sum_{i=1}^{n} x_{i} y_{i} - n \bar{x} \bar{y} + n \hat{\beta}_{1} \bar{x}^{2} - \hat{\beta}_{1} \sum_{i=1}^{n} x_{i}^{2} &= 0 \\ n \hat{\beta}_{1} \bar{x}^{2} - \hat{\beta}_{1} \sum_{i=1}^{n} x_{i}^{2} &= n \bar{x} \bar{y} - \sum_{i=1}^{n} x_{i} y_{i} \\ & \hat{\beta}_{1} \bar{x}^{2} - \hat{\beta}_{1} \sum_{i=1}^{n} x_{i}^{2} &= n \bar{x} \bar{y} - \sum_{i=1}^{n} x_{i} y_{i} \\ & \hat{\beta}_{1} \bar{x}^{2} - \hat{\beta}_{1} \sum_{i=1}^{n} x_{i}^{2} &= n \bar{x} \bar{y} - \sum_{i=1}^{n} x_{i} y_{i} \\ & \hat{\beta}_{1} \bar{x}^{2} - \hat{\beta}_{1} \sum_{i=1}^{n} x_{i}^{2} &= n \bar{x} \bar{y} - \sum_{i=1}^{n} x_{i} y_{i} \\ & \hat{\beta}_{1} \bar{x}^{2} - \hat{\beta}_{1} \sum_{i=1}^{n} x_{i}^{2} &= n \bar{x} \bar{y} - \sum_{i=1}^{n} x_{i} y_{i} \\ & \hat{\beta}_{1} \bar{x}^{2} - \hat{\beta}_{1} \sum_{i=1}^{n} x_{i}^{2} &= n \bar{x} \bar{y} - \sum_{i=1}^{n} x_{i} y_{i} \\ & \hat{\beta}_{2} \bar{y}_{1} \bar{y}_{2} \bar{y}_{1} \bar{y}_{2} \bar{y}_{2} \bar{y}_{1} \bar{y}_{2} \bar{y}_{2} \\ & \hat{\beta}_{2} \bar{y}_{2} \bar{y}_{1} \bar{y}_{2} \bar{y}_$$

0000000000000

# Modelo de Regresión

$$n\hat{\beta}_{1}\bar{x}^{2} - \hat{\beta}_{1}\sum_{i=1}^{n}x_{i}^{2} = n\bar{x}\bar{y} - \sum_{i=1}^{n}x_{i}y_{i}$$

$$\hat{\beta}_{1}\left(n\bar{x}^{2} - \sum_{i=1}^{n}x_{i}^{2}\right) = n\bar{x}\bar{y} - \sum_{i=1}^{n}x_{i}y_{i}$$

$$\hat{\beta}_{1} = \frac{n\bar{x}\bar{y} - \sum_{i=1}^{n}x_{i}y_{i}}{n\bar{x}^{2} - \sum_{i=1}^{n}x_{i}^{2}}$$



DE MATEMÁTICA



#### Resumen Estimación del Modelo de Regresión Lineal Simple

#### Dado el modelo

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\varepsilon}_i$$

Pendiente 
$$\hat{\beta}_1 = \frac{n\bar{x}\bar{y} - \sum_{i=1}^n x_i y_i}{n\bar{x}^2 - \sum_{i=1}^n x_i^2}$$
Constante 
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$



DEPARTAMENTO
DISCIPLINARIO
DE MATEMÁTICA
VESTADÍSTICA



#### Estimación de la varianza residual

Un estimador insesgado de la varianza residual está dado por la expresión:

$$\hat{\sigma^2} = \frac{\sum_{i=1}^n \varepsilon^2}{n-2} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}$$

Donde

$$\frac{(n-2)\hat{\sigma^2}}{\sigma^2} \sim \chi^2_{(n-2)}$$





# <u>Inferencia</u> en el Modelo de Regresión

## Intervalos de confianza para $\beta_0$ , $\beta_1$ y $\sigma^2$

Queremos ahora obtener intervalos de confianza para  $\beta_0$ ,  $\beta_1$  y  $\sigma^2$ , todos a un nivel de confianza de  $(1 - \alpha)100\%$  . Considerando las distribuciones de las siguientes cantidades pivotales:

$$\frac{\widehat{\beta_0} - \beta_0}{\sqrt{\widehat{\sigma^2} \left[ \frac{1}{n} + \frac{\bar{x}^2}{\sum (x - \bar{x})^2} \right]}} \sim t_{(n-2)} \qquad \frac{\widehat{\beta_1} - \beta_1}{\sqrt{\frac{\widehat{\sigma^2}}{\sum (x - \bar{x})^2}}} \sim t_{(n-2)} \qquad \frac{(n-2)\widehat{\sigma^2}}{\sigma^2} \sim \chi_{(n-2)}^2$$





#### Intervalo de confianza para $\beta_0$

A partir del pivote, se plantea

$$P\left(-t_{\alpha/2} \le \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x - \bar{x})^2}\right]}} \le t_{\alpha/2}\right) = 1 - \alpha$$



DE MATEMÁTICA



# Inferencia en el Modelo de Regresión

#### Intervalo de confianza para $\beta_0$

Despejando  $\beta_0$ , tenemos

$$I.C_{(1-\alpha)100}(\beta_0) = \hat{\beta}_0 \pm t_{(\alpha/2;n-2)} \sqrt{\hat{\sigma}^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{\sum (x - \bar{x})^2} \right]}$$

es decir, el  $I.C_{(1-\alpha)100}(\beta_0)$  está dado por:

$$\left[\hat{\beta}_{0} - t_{\left(\frac{\alpha}{2}; n-2\right)} \sqrt{\hat{\sigma}^{2} \left[\frac{1}{n} + \frac{\bar{x}^{2}}{\sum (x - \bar{x})^{2}}\right]}; \hat{\beta}_{0} + t_{\left(\frac{\alpha}{2}; n-2\right)} \sqrt{\hat{\sigma}^{2} \left[\frac{1}{n} + \frac{\bar{x}^{2}}{\sum (x - \bar{x})^{2}}\right]}\right]_{\text{MENTO}}$$

MÁTIC

MÁTIC

MÁTIC

Naturales y Exactas | YESTADÍSTICA

# Inferencia en el Modelo de Regresión

#### Intervalo de confianza para $\beta_1$

A partir del pivote, se plantea

$$P\left(-t_{\alpha/2} \le \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum (x - \bar{x})^2}}} \le t_{\alpha/2}\right) = 1 - \alpha$$



DEPARTAMENTO
DISCIPLINARIO
DE MATEMÁTICA



# Inferencia en el Modelo de Regresión

#### Intervalo de confianza para $\beta_0$

Despejando  $\beta_1$ , tenemos

$$I.C_{(1-\alpha)100}(\beta_1) = \hat{\beta}_1 \pm t_{(\alpha/2;n-2)} \sqrt{\frac{\hat{\sigma}^2}{\sum (x-\bar{x})^2}}$$

es decir, el  $I.C_{(1-\alpha)100}(\beta_1)$  está dado por:

$$\left[\hat{\beta}_1 - t \left(\frac{\alpha}{2}; n-2\right) \sqrt{\frac{\hat{\sigma}^2}{\sum (x-\bar{x})^2}}; \hat{\beta}_1 + t \left(\frac{\alpha}{2}; n-2\right) \sqrt{\frac{\hat{\sigma}^2}{\sum (x-\bar{x})^2}}\right]$$

AMENTO NARIO MÁTICA

INATURATES Y EXACTAS | YESTADÍSTICA

#### Intervalo de confianza para $\sigma^2$

A partir del pivote, se plantea

$$P\left(\chi_{\alpha/2}^2 \le \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \le \chi_{1-\alpha/2}^2\right) = 1 - \alpha$$

$$P\left(\chi_{\alpha/2}^2 \leq \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sigma^2} \leq \chi_{1-\alpha/2}^2\right) = 1 - \alpha$$





# Inferencia en el Modelo de Regresión

### Intervalo de confianza para $\sigma^2$

Despejando  $\sigma^2$ , tenemos

$$I.C_{(1-\alpha)100}(\sigma^2) = \left[ \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\chi^2_{(\alpha/2:n-2)}}; \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\chi^2_{(1-\alpha/2:n-2)}} \right]$$







### Análisis de varianza del MRLS

Consideremos el test

$$H_0: \beta_1 = 0 \quad v/s \quad H_1: \beta_1 \neq 0$$

Estas hipótesis prueban la contribución de la variable regresora X para explicar la respuesta Y, una vez que  $H_0$  sea verdadera, esto es  $\beta_1 = 0$ , esa contribución no es significativa. En ese caso o el MRLS no es adecuado o la variable regresora escogida no contribuye a la explicación de la variable Y.





### Análisis de varianza del MRLS

Así para un test de nivel  $\alpha$  rechazamos  $H_0$ , si

$$\frac{\widehat{\beta}_1^2 \sum_{i=1}^n (x_i - \overline{x})^2}{\widehat{\sigma}^2} >_1 F_{(n-2)}(\alpha),$$

del cuantil  $(1-\alpha)$  de la distribución F con 1 y (n-2) grados de libertad.





#### **ANOVA**

#### **Definiciones**

- $\sum_{i=1}^{\infty} (y_i \overline{y})^2$ : Suma de cuadrados totales (SQT), Variación total de Y en torno a su media.
- $\sum_{i=1}^{n} (y_i \widehat{y})^2 :$ Suma de cuadrados del error (SQE), Variación total de Y en torno a la recta.
- $\sum_{i=1}^{n} (\widehat{y}_i \overline{y})^2$ : Suma de cuadrados de la regresión (SQR), Variación de las esperanzas específicas de Y, dado x, en torno a su media.

**AMENTO** MÁTICA INGLUI GLES Y LAGULAS I TESTADÍSTICA

### TABLA ANOVA

Fuente	GL	SQ	QM	$F_0$
(Fuente de	(Grados de	(Suma de	(Cuadrado	
Variación)	Libertad)	Cuadrados	Medio)	
Regresión	1	SQR	SQR	$\frac{SQR}{SQE/(n-2)}$
Error	n-2	SQE	SQE/(n-2)	
Total	n – 1	SQT		

Medidas de Asociación

SOT = SOE + SOR.



DISCIPLINARIO DE MATEMÁTICA Y ESTADÍSTICA



- **1** Primero calculamos *SOT*.
- 2 Calculamos *SQR*.
- Obtenemos la SQE por sustracción: SOT SOR.

Medidas de Asociación

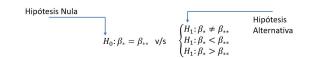
- En la cuarta columna hacemos la división de la suma de los cuadrados por los grados de libertad.
- En la quinta columna obtenemos el valor de la estadística de prueba, .

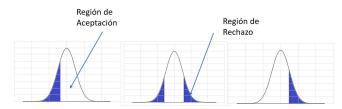
Así rechazamos  $H_0: \beta_1 = 0$  versus  $H_1: \beta_1 \neq 0$ , si  $F_0 > 1$   $F_{n-2:1-\alpha}$ , de la distribución F.





#### Test de Hipótesis





DEPARTAMENTO
DISCIPLINARIO
DE MATEMÁTICA



# Inferencia en el Modelo de Regresión

### Test de Hipótesis para $\beta_0$

$$H_1: \beta_0 < \beta_{00}$$



$$H_0$$
:  $\beta_0 = \beta_{00}$  V/S  $H_1$ :  $\beta_0 \neq \beta_{00}$ 



$$T = \frac{\widehat{\beta_0} - \beta_{00}}{\sqrt{\widehat{\sigma^2} \left[ \frac{1}{n} + \frac{\bar{x}^2}{\sum (x - \bar{x})^2} \right]}}$$

$$H_1: \beta_0 > \beta_{00}$$





# Inferencia en el Modelo de Regresión

### Test de Hipótesis para $\beta_1$





$$H_0$$
:  $\beta_1 = \beta_{10}$   
V/S  
 $H_1$ :  $\beta_1 \neq \beta_{10}$ 



$$T = \frac{\widehat{\beta_1} - \beta_{10}}{\sqrt{\frac{\widehat{\sigma^2}}{\sum (x - \bar{x})^2}}}$$







# Modelo de Regresión

#### Ejemplo: Modelo de Regresión Lineal simple

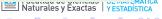
Se quiere realizar un estudio sobre la relación entre el tiempo diario dedicado al estudio y el rendimiento en la asignatura de matemática, los resultados de la encueta son los siguientes

Tiempo	2,5	3,1	4,7	1,5	3,6	5,3
Rendimiento	3,5	4,1	5,3	2,8	3,8	5,4

Construir una gráfica de dispersión de los datos,

Medidas de Asociación

- Calcular e interpretar  $r_{xy}$ ,
- Determine el modelo de regresión lineal,



## Modelo de Regresión

#### Ejemplo: Modelo de Regresión Lineal simple

Tiempo	2,5	3,1	4,7	1,5	3,6	5,3
Rendimiento	3,5	4,1	5,3	2,8	3,8	5,4

- Encuentre  $I.C_{80\%}(\beta_0)$ ;  $I.C_{95\%}(\beta_1)$  y  $I.C_{90\%}(\sigma^2)$
- Dado los siguientes test, determine si se rechazan o no la Hipótesis Nula.

**1** 
$$H_0: \beta_0 = 1, 5 \quad v/s \quad H_1: \beta_0 > 1, 5; \quad \alpha = 5\%$$

2 
$$H_0: \beta_1 = 0$$
  $v/s$   $H_1: \beta_1 \neq 0;$   $\alpha = 10\%$ 



