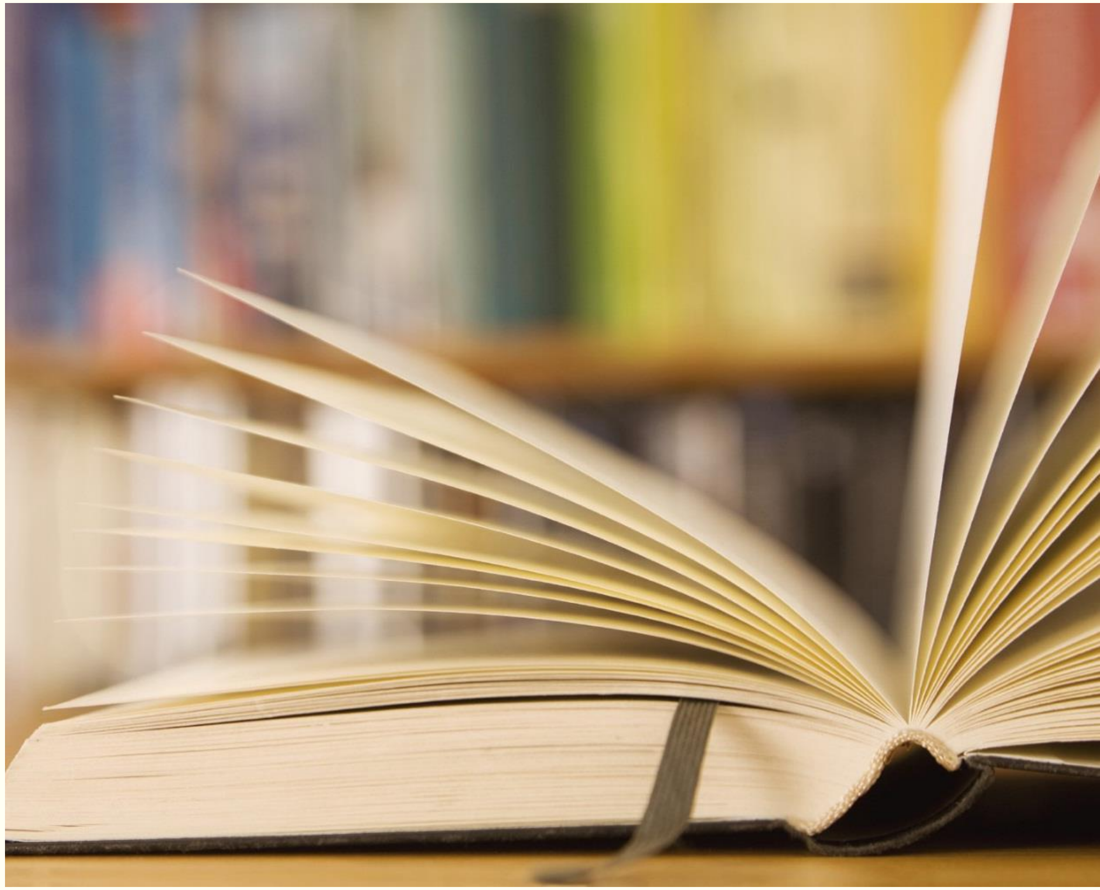
UNIVERSIDAD NACIONAL DE CHIMBORAZO

FACULTAD DE INGENIERÍA

CARRERA DE INGENIERÍA CIVIL

ASIGNATURA: Estadística

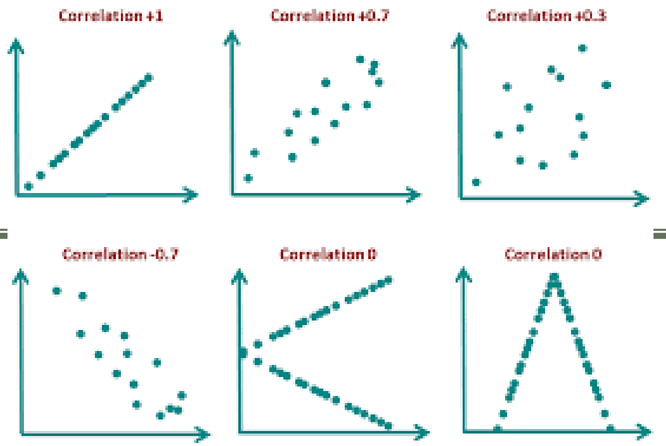
DOCENTE: Ing. Lidia Castro M.Sc



UNIDAD 3 → ESTADÍSTICA INFERENCIAL: ANÁLISIS DE REGRESIÓN Y CORRELACIÓN

OBJETIVO DE LA UNIDAD

Identificar problemas que requieren de análisis estadístico para su entendimiento, aplicando los análisis de regresión y correlación.



1. Teoría del muestreo.
2. Estimación de Parámetros.
3. Regresión y Correlación.

Introducción

La enorme ventaja de aproximar información discreta o funciones "complejas", con funciones analíticas sencillas, radica en su mayor facilidad de evaluación y manipulación, situación necesaria en el campo de la ingeniería.

DATOS ⟹

Procesados.- A partir de los datos se debe obtener un modelo matemático.

Interpretados.- Se debe estudiar la influencia entre variables independientes y dependientes.

Definiciones

Extrapolación

Consiste en obtener nuevos valores más allá de los límites de los datos observados

Interpolación

Consiste en obtener nuevos valores dentro del rango de los datos observados.

Aplicaciones

- Análisis de tendencias
- Prueba de hipótesis
- Integración numérica
- Diferenciación numérica
- Resolución de ecuaciones diferenciales

Regresión lineal simple

Cuando tan solo hay una variable independiente x, la función y que explica el análisis de regresión se escribe de la siguiente forma:

$$y = b_o + b_1 x$$

Donde

$b_o \rightarrow$ Término independiente

$b_1 \rightarrow$ Coeficiente que marca la importancia que la variable $\boldsymbol{x}$ supone en la variable $\boldsymbol{y}$

$$y_i = b_o + b_1 x_i + e_i$$

$$i = 1, 2, \dots, n$$

$e_i \rightarrow$ Es el error cometido en cada una de las observaciones

Para obtener los coeficientes de b_0 y b_1 se plantea el siguiente sistema:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}.$$

$$X^T y = X^T X b.$$

$$e = \|X\hat{b} - y\|.$$

Ejemplo 1

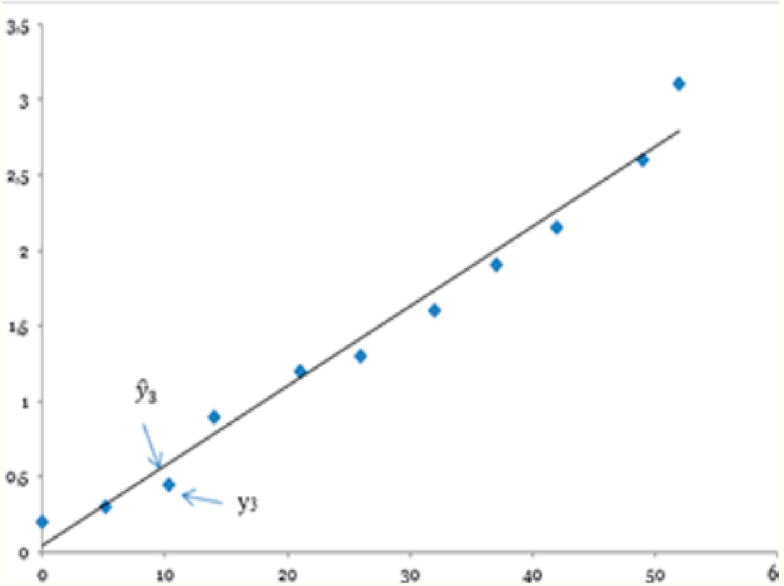
Obtener los coeficientes de la recta de regresión $\boldsymbol{y = b_o + b_1 x}$ que mejor se ajusta a los datos.

i	1	2	3	4
$\mathbf{x_i}$	2	5	7	8
$\mathbf{y_i}$	1	2	3	3

Regresión lineal por mínimos cuadrados

- Ajustar los puntos de un diagrama de dispersión -> **método de mínimos cuadrados**

- Encontrar una recta que minimice el error entre los puntos dispersado (mejor ajuste entre dos variables).



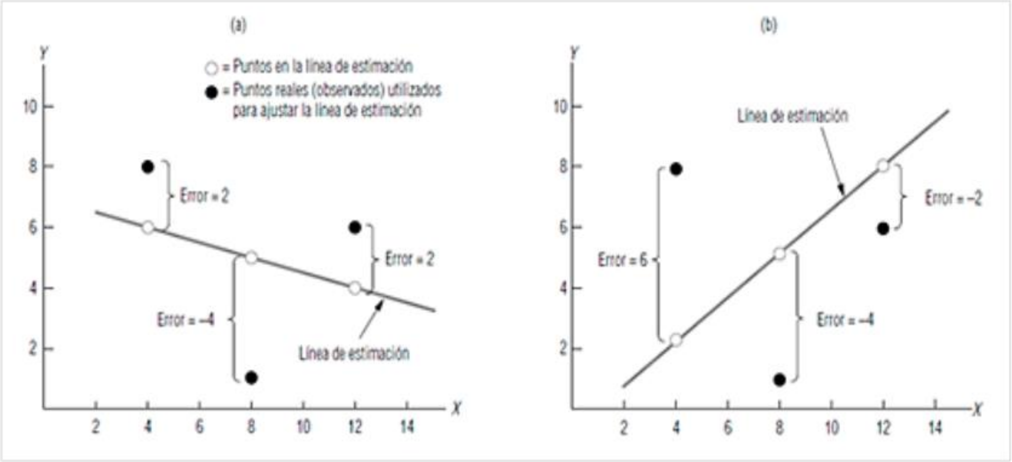
Residuos -> diferencias entre los valores observados y los estimados por la recta

El error total **no** se puede calcular así :

- Sumando los residuos porque se anulan

- Sumando los valores absolutos de los residuos porque no considera puntos muy alejados.

Solución, **elevar al cuadrado los residuos ->** método de mínimos cuadrados.



Recta que minimice los cuadrados de los residuos

La recta puede expresarse mediante la ecuación en forma explícita

$$y = a + bx$$

Donde; x e y son las variables independiente y dependiente respectivamente, **b** la pendiente **y** a la ordenada en el origen

$$\sum_1^n (y_j - \hat{y}_j)^2 \, Mínimo \rightarrow \sum_1^n (y_j - a - bx_i)^2 \, Mínimo$$

$$\sum_1^n y_j = na + b\sum_1^n x_i \quad a = \bar{y} - b\bar{x}$$

$$\sum_1^n y_j x_i = (\bar{y} - b\bar{x})\sum_1^n x_i + b\sum_1^n {x_i}^2$$

$$y - \overline{y} = \frac{S_{xy}}{S_x^2}(x - \overline{x})$$

- Donde $\bar{x}$ e $\bar{y}$ son los valores medios de x y de y respectivamente.
- S_{xy} es la covarianza y $(Sx)^2$ es la varianza de x.

Ejemplo 2

Un jugador de baloncesto ha encestado las siguientes canastas desde diferentes distancias, escribir la recta de regresión obtenida por el método de mínimos cuadrados.

Distancia	Canastas
2,5	23
3,2	21
4,7	19
6,2	15
8,7	12
13,2	8
17,0	4
18,5	1

Error Estándar de Estimación

- Ver si la estimación es **buena,** mediante el error estándar de la estimación.

- Significado similar a la desviación estándar que evalúa lo que los datos se dispersan de la recta de regresión obtenida.

- Su cálculo se lleva a cabo así:

$$S_e = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}}$$

- Cuando se eleva el error de la estimación al cuadrado se obtiene el error **cuadrático medio.**

Coeficiente de determinación/correlación

Al hacer un ajuste por regresión lineal se puede calcular de un **coeficiente que nos dé información sobre el grado de asociación** entre las variables.

$$r^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}$$

- Este valor es siempre es positivo
- Si es igual a 1 se trata de correlación perfecta entre variables.

Ejemplo 3

La siguiente tabla recoge las ganancias en dólares obtenidas por un hotel en función de los huéspedes de lunes a domingos excluyendo el miércoles.

a. Escribir la recta de regresión y el coeficiente de correlación.

b. Calcular el error estándar de la estimación.

Día de la semana	Ganancias	Número de huéspedes
Lunes	1200	12
Martes	1500	16
Jueves	3670	31
Viernes	7200	69
Sábado	10090	102
Domingo	5800	56

Ejemplo 4

La producción anual de vehículos de una empresa es determine la ecuación lineal que predice la producción y calcule para el año 2014.

AÑO	VEHÍCULOS
2000	1090
2002	1120
2004	1250
2006	1490
2008	1650

¡GRACIAS!