

BASE DE DATOS REGRESIÓN

$$U.E. \text{ final value } y_2 - k \cdot y_3 \times x_1 \times x_2 \text{ and } x_1 \text{ and } x_2$$

$\Gamma = \{x_1, x_2, \dots, x_n\}$ $\Gamma \vdash_{\text{Hilbert}} \varphi$

$Y_{21} \times Y_{22} \dots Y_{29} \circ X_{21} \times X_{22} \dots X_{29}$

metabolic cycle

$$Y_{n1} Y_{n2} \cdots Y_{np} X_{n1} X_{n2} \cdots X_{np}$$

Dependientes Explicatorias

1. OBJETIVO: Estudiar la relación entre dos o más variables. Dicha relación se da entre una (o varias) variable(s) dependiente(s) cuantitativa(s) y una (o varias) variables explicatorias X que pueden ser cualitativas o cuantitativas.

2. TIPOS DE RELACION:

i) ASOCIACIÓN: los valores de X y de Y parecen estar vinculados de alguna manera en la población

iii) CAUSALIDAD: se presenta

- La relación entre X y Y es consistente en dirección y/o en magnitud
- Si se interviene o cambia X entonces responde acorde.
- Existe un mecanismo que vincula la causa (X) con el efecto (Y).

iii) INDEPENDENCIA (DEPENDENCIA)

3. RESULTADO: la relación entre X y Y se representa mediante una ecuación que constituye un modelo matemático y se estructura mediante el análisis de varianza (ANOVA).

4. NOTACIÓN

Variable	Significado
Y	Dependiente / Explicada / Predicha / Regresada / Respuesta / Endógena
X	Independiente / Predictora / Regresora / Estímulo / Exógena

5. TIPOS DE MODELOS: los que se verán en el curso

- Una sola Y (continua), una sola X (continua):

REGRESIÓN SIMPLE

- Una sola Y (continua), varias X (continuas):

REGRESIÓN MÚLTIPLE

- Una o varias Y (continuas), varias X (continuas, discretas y/o categóricas): **MODELO LINEAL GENERAL**

6. ORIGEN DEL TÉRMINO REGRESIÓN:

CALTON (Regresión a la mediocridad) + PEARSON (+1000 reg.)

NOTA: En el análisis de regresión interesa una dependencia ESTADÍSTICA, más no una dependencia DETERMINÍSTICA propia de las ciencias básicas. "Una relación estadística no puede aducir, por sí misma causalidad, para ello se debe acudir a consideraciones a priori o teóricas". A. Volta

2. MODELOS DE REGRESIÓN LINEAL.

UTILIDAD DE PREDICCIÓN DE REGRESIÓN

- DESCRIPCIÓN: Se construye una ecuación que permite describir la relación de asociación entre X y Y .

- PREDICCIÓN: Predecir o estimar el valor promedio o media de la variable dependiente, Y , ante cambios en los valores conocidos o fijos de las explicatorias, esto es: $E[Y/x]$

- CONTROL: Controlar el comportamiento de la variable respuesta Y mediante cambios en X .

REGRESIÓN LINEAL SIMPLE: REGRESIÓN CON DOS VARIABLES

La (variable) dependiente Y es analizada en términos de una sola variable explicatoria X .

El modelo de regresión simple sería:

FENÓMENO = MODELO + RESIDUAL

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

que se conoce como la Función del Regresión Poblacional FRP. Donde:

- β_0 : es el intercepto.
- β_1 : es la pendiente.
- Y : es la variable dependiente.
- X : es la variable explicatoria.
- ϵ : es el término de error aleatorio.

INTERPRETACIÓN:

β_0 : se interpreta como el valor que toma Y siempre que X valga cero. Siempre y cuando el cero esté incluido en el recorrido de X . Sino, es simplemente el intercepto y no se interpreta.

β_1 : se interpreta como el cambio en Y ante cambios unitarios de X . Es decir, Y aumenta (Si β_1 es +) o disminuye (Si β_1 es -) una cantidad β_1 por cada unidad que cambie X .

ϵ : es el término estocástico o error aleatorio, contiene la información de TODAS las variables que no se incluyeron de manera individual en el modelo, pero que, en conjunto, tienen influencia en Y .

X : es la variable explicatoria, las cuales, en principio, son variables matemáticas fijas con error de medición cero.

Y : es la variable respuesta que es aleatoria pues es función de ϵ .

Importancia de β : $\beta + \beta_1 X_{11} + \beta_2 X_{12} + \dots + \beta_N X_{NN}$

1. Vaguedad de la teoría
2. Falta de disponibilidad de datos
3. Variables centrales y variables periféricas
4. Aleatoriedad intrínseca en el comportamiento humano
5. Variables inadecuadas
6. Principio de parsimonia
7. Forma funcional incorrecta

Significado del término lineal:

- Linealidad en las variables: Los x están elevados a un exponente $\alpha = 1$. $Beta = 1$.
- Linealidad en los parámetros: β^0 .

En el modelo de Regresión Lineal:

λ puede ser o no 1.

θ tiene que ser siempre 1

Notación Matricial:

Para cada unidad experimental, la FRP puede escribirse como:

$$Y_1 = \beta_0 + \beta_1 X_{11} + \epsilon_1$$

$$Y_2 = \beta_0 + \beta_1 X_{12} + \epsilon_2$$

$$\vdots$$

$$Y_N = \beta_0 + \beta_1 X_{1N} + \epsilon_N$$

que es equivalente a:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} \\ 1 & X_{21} \\ \vdots & \vdots \\ 1 & X_{n1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

En notación matricial:

$$Y_{n \times 1} = X_{N \times p} \beta_{p \times 1} + \epsilon_{n \times 1}$$

CLASE 5:

ESTIMACIÓN DE LA FRP: FUNCIONES DE REGRESIÓN MUESTRAL FRM

El objetivo principal del análisis de regresión es estimar esta función FRP, lo que se realiza a partir de una muestra de n valores de Y (y_i), obtenidos en condiciones de observación de los variables regresoras, no necesariamente idénticas y garantizando la aleatoriedad en el orden de observación.

El procedimiento de estimación permitirá encontrar la Función de Regresión Muestral, que será una aproximación a la FRP y que se representa por:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_p X_p + e$$

Donde:

$\hat{\beta}_j$: es el estimador para β_j

\hat{Y} : es el estimador de $E[Y]$

e : es el residual, que tiene interpretación análoga

al error aleatorio.

Para las n observaciones de la muestra:

$$Y_1 = \hat{\beta}_0 + \hat{\beta}_1 X_{11} + e_1$$

$$Y_2 = \hat{\beta}_0 + \hat{\beta}_1 X_{12} + e_2$$

$$Y_n = \hat{\beta}_0 + \hat{\beta}_1 X_{1n} + e_n$$

que es equivalente a:

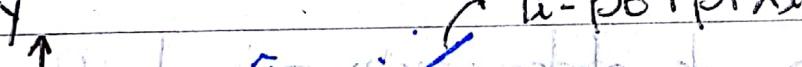
$$\begin{array}{|c|c|c|c|c|} \hline Y_1 & 1 & X_{11} & \hat{\beta}_0 & e_1 \\ \hline Y_2 & 1 & X_{12} & \hat{\beta}_1 & e_2 \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline Y_n & 1 & X_{1n} & \end{array}$$

En notación matricial:

$$Y_{nx1} = X_{nx2} \hat{\beta}_{2x1} + E_{nx1}$$

Gráfica. la estimación:

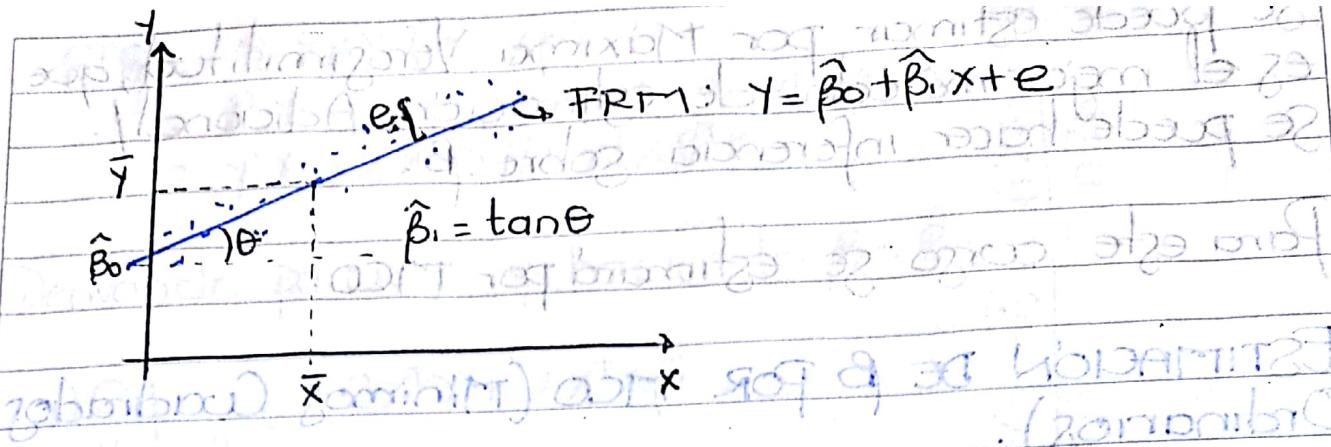
$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$



$$E[Y_i | X_i] = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_1 X_2 + \dots + \hat{\beta}_1 X_n + \hat{\beta}_0 = P$$

Gráfica. la FRT:



Propiedades de la FRM

1. La recta pasa por los puntos \bar{x}, \bar{y}
2. $\bar{Y}_i = \hat{Y}_i$
3. $\bar{e}_i = 0$
4. Los residuales e_i no están correlacionados con \hat{Y}_i : $\text{Corr}(e_i, \hat{Y}_i) = 0$
5. Los residuales e_i no están correlacionados con x_i : $\text{Corr}(e_i, x_i) = 0$

Supuestos sobre el vector E

- i) $E(E_p) = E[E_{pi}] = 0$: El valor promedio del error aleatorio es cero. Se interpreta como que la información contenida en E_p no afecta de manera sistemática a \bar{Y} . (unos $E_{pi} + y$ otros - se anulan).
- ii) $\text{Var}(E_p) = \sigma^2 I_{N \times N} = E[E_p E_p^T]$: debe cumplirse que la varianza de los E_{pi} permanezca constante, es decir que debe cumplirse el supuesto de homoscedasticidad. Además deben estar incorelacionados para estimar por MCO.
- iii) $E_1, E_2, \dots, E_N \stackrel{iid}{\sim} N(0, \sigma^2 I)$: Si esto se cumple

se puede estimar por Máxima Verosimilitud, que es el mejor método de estimación. Adicionalmente se puede hacer inferencia sobre β .

Para este curso se estimará por MCO.

ESTIMACIÓN DE β POR MCO (Mínimos Cuadrados Ordinarios).

El criterio de MCO busca determinar el valor de los coeficientes de regresión β , $\hat{\beta}$, tal que:

$$Q = \sum \epsilon_i^2$$

Sea mínima, y debe satisfacer la condición:

$$\frac{\partial Q}{\partial \beta} \Big|_{\beta=\hat{\beta}} = 0$$

Para estimar a β se parte de:

$$Y = X\beta + \epsilon$$

$$Q = \sum \epsilon_i^2 = \epsilon^T \epsilon = \sum (\epsilon_1, \epsilon_2, \dots, \epsilon_n) \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} = (\epsilon)$$

Por tanto:

$$\begin{aligned} \mathbf{e}_1^T \mathbf{e}_1 &= (\mathbf{y} - \mathbf{x}\beta)^T (\mathbf{y} - \mathbf{x}\beta) \\ &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{x} \beta^T - \mathbf{x}^T \beta^T \mathbf{y} + \mathbf{x}^T \mathbf{x} \beta^T \beta \\ &= \mathbf{y}^T \mathbf{y} - 2 \mathbf{x}^T \beta^T \mathbf{y} + \mathbf{x}^T \mathbf{x} \beta^T \beta \end{aligned}$$

Dervando parcial / y haciendo $\beta = \hat{\beta}$

$$\frac{\partial Q}{\partial \beta} \Big|_{\beta=\hat{\beta}} = -2\mathbf{x}^T \mathbf{y} + 2\mathbf{x}^T \mathbf{x} \hat{\beta} = 0$$

$$2\mathbf{x}^T \mathbf{x} \hat{\beta} = 2\mathbf{x}^T \mathbf{y}$$

$$\mathbf{x}^T \mathbf{x} \hat{\beta} = \mathbf{x}^T \mathbf{y}$$

$$\hat{\beta} = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

CLASE 6

EJEMPLO 1. los siguientes datos corresponden a los rayos cósmicos medidos en varias altitudes:

i	Altura (pies)	Rayos Cómicos	Tasa (mrem/año)
1	50	28	
2	450	30	
3	780	32	
4	1200	36	
5	4400	51	
6	4800	58	
7	5300	69	

Ajuste un modelo que describa la relación entre estas obs variables usando el método de mco.

X: Altura

Y: Rayos

$$X = \begin{bmatrix} 1 & 50 \\ 1 & 450 \\ 1 & 780 \\ 1 & 1200 \\ 1 & 4400 \\ 1 & 4800 \\ 1 & 5300 \end{bmatrix} \quad X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 50 & 450 & 780 & 1200 & 4400 & 4800 & 5300 \end{bmatrix}$$

$$Y = \begin{bmatrix} 28 \\ 30 \\ 32 \\ 36 \\ 51 \\ 58 \\ 69 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = (X^T X)^{-1} X^T Y$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 50 & 450 & 780 & 1200 & 4400 & 4800 & 5300 \end{bmatrix} \quad \begin{bmatrix} 1 & 50 \\ 1 & 450 \\ 1 & 780 \\ 1 & 1200 \\ 1 & 4400 \\ 1 & 4800 \\ 1 & 5300 \end{bmatrix}$$

$$= \begin{bmatrix} 7 & 16980 \\ 16980 & 72743400 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{16980} \text{Adj} = 4.53 \times 10^{-9} \begin{bmatrix} 7.2743400 & -16980 \\ -16980 & 72743400 \end{bmatrix}$$

$$= \begin{bmatrix} 0.3293 & -7.6873 \times 10^{-5} \\ -7.6873 \times 10^{-5} & 3.169 \times 10^{-8} \end{bmatrix}$$

$$X^T Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 50 & 450 & 780 & 1200 & 4400 & 4800 & 5300 \end{bmatrix} \begin{bmatrix} 28 \\ 30 \\ 32 \\ 36 \\ 51 \\ 58 \\ 69 \end{bmatrix}$$

$$= \begin{bmatrix} 304 \\ 951560 \end{bmatrix}$$

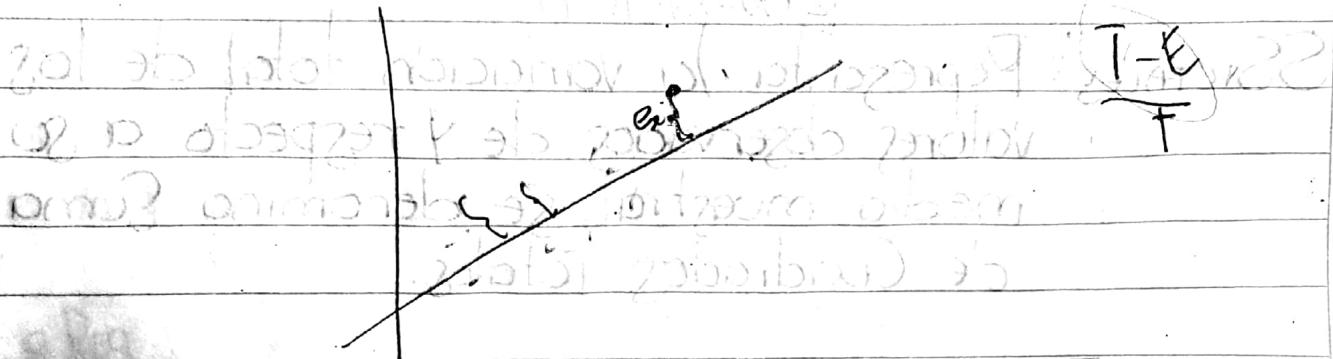
$$(X^T X)^{-1} (X^T Y) = \begin{bmatrix} 0.3293 & -7.6873 \times 10^{-5} \\ -7.6873 \times 10^{-5} & 3.169 \times 10^{-8} \end{bmatrix} \begin{bmatrix} 304 \\ 951560 \end{bmatrix}$$

$\hat{\beta}_0$
Obs
Pred

$$= \begin{bmatrix} 26.96 \\ 0.006786 \end{bmatrix}$$

$$e = Y - X\hat{\beta}$$

	28	1.50	26.96	0.693943
1	30	1.450	0.006786	-0.0266118
2	32	1.780		-0.260119
3	36	1.1200		0.889890
4	51	1.4400		-5.82684
5	58	1.4800		-1.54139
6	69	1.5300		6.06542



CLASE 7

ANÁLISIS DE VARIANZA: ANOVA

El análisis de Varianza (ANOVA) en la regresión, se hace con el fin de descomponer la varianza total en sus fuentes de variación:

$$\text{FENÓMENO} = \text{MODELO} + \text{RESIDUAL}$$
$$Y = X\hat{\beta} + e$$

Sabemos que: $y_i = \hat{y}_i + e_i \quad ①$

y haciendo: $x_i = x_i - \bar{x}$

$$y_i = y_i - \bar{y}$$

Entonces De ①

$$y_i = \hat{y}_i + e_i$$

$$y_i^2 = \hat{y}_i^2 + 2\hat{y}_i e_i + e_i^2$$

$$\sum y_i^2 = \sum \hat{y}_i^2 + 2 \sum \hat{y}_i e_i + \sum e_i^2$$

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2$$

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \bar{y})^2 + \sum e_i^2$$

$$SS_{\text{TOTALES}} = SS_{\text{Regresión}} + SS_{\text{Error}}$$

$$[\sum (x_i - \bar{x})(y_i - \bar{y})]^2$$

SS_{TOTALES} : Representa la variación total de los valores observados de Y respecto a su medio muestral, se denomina Suma de Cuadrados Totales.

$SS_{\text{REGRESIÓN}}$: Es la variación debida a las variables explicativas y se denomina Suma de

Cuadrados de la Regresión

SSError: Es la variancia no explicada, también conocida como la variancia de los residuos de los valores de y respecto de la linea de Regresión, denominado Cuadrado del Error.

Si se dividen las Sumas de Cuadrados por sus grados de libertad, se obtienen los cuadrados medios:

Cuadrado Medio de la Regresión: $MSR = \frac{SSR}{P}$

Cuadrado Medio del Error: $MSE = \frac{SSE}{n-p-1}$

Cuadrado Medio Total: $MS_T = \frac{SST}{n-1}$

El cociente de dos Cuadrados Medios independientes, tiene distribución F:

$$\frac{(SST - SSR)}{MS_E} \sim F_{P, n-p-1}$$

Tabla ANOVA

FdeV.	SS	Grados L.	MS	Fr.
Debido a Reg	SSR	P	$MSR = SSR/P$	MSR
Debido al error	SSE	n-p-1	$MSE = SSE/(n-p-1)$	MSE
Total	SST	n-1		

Con esta tabla se prueba hipótesis

$H_0: \beta = 0$: No existe modelo

$H_1: \beta \neq 0$: Si existe modelo

Que se conoce como la prueba global. Se Rechaza la hipótesis nula si el valor p de la prueba F tiende a cero ($\text{Valor } p < \alpha$). Es decir:

$F_0 \geq F_{\alpha, p, n-p-1}$ entonces Rechaza H_0 .

En caso de Rechazar H_0 se afirma que si existe modelo.

Pruebas Individuales:

En estas pruebas, basadas en el estadístico t, se prueba si:

$H_0: \beta_j = 0$

$H_1: \beta_j \neq 0$

El estadístico t se construye como:

$$t_j = \frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}} ; \quad \hat{\sigma}_{\hat{\beta}_j}^2 = \frac{\sigma^2}{\sum x_j^2} = \frac{\sigma^2}{\sum (x_j - \bar{x})^2} = S^2$$

El caso en el que se rechace H_0 , indica que x_j si aporta a la variación de y .

Para β_0

$$H_0: \beta_0 = 0$$

$$H_1: \beta_0 \neq 0$$

$$t_0 = \hat{\beta}_0 ; \quad \text{error estandar de } \hat{\beta}_0 = \sigma_{\hat{\beta}_0} = \sigma \sqrt{\frac{\sum x_i^2}{n}} = S \sqrt{\frac{2}{n-2}}$$

Intervalos de Confianza

Para β_j

$$\beta_j \in (\hat{\beta}_j \pm t_{\alpha/2} \hat{\sigma}_{\hat{\beta}_j})$$

Si el I.C. contiene a cero, x_j no explica a y .

Análoga.: $\beta_0 \in (\hat{\beta}_0 \pm t_{\alpha/2} \hat{\sigma}_{\hat{\beta}_0})$

Si el I.C. contiene a cero, no existe el intercepto.

El coeficiente de determinación R^2

Es una medida de qué tan "bueno" es el ajuste del modelo de Regresión a los datos. En porcentaje, mide la bondad del ajuste porcentual del modelo, se define como:

$$R^2 = \frac{\hat{\beta}^T X^T Y - n\bar{y}^2}{Y^T Y - n\bar{y}^2} = \frac{SSR}{SST}, \quad R^{2\%} = R^2 \times 100\%$$

Se debe tener especial cuidado con esta medida de ajuste porque, en principio, cuando R^2 está más cerca de 1, se tiene mejor modelo.

Sin embargo, esta medida crece a medida que se incorporan nuevas X al modelo.

R^2 ajustado: \hat{R}^2 se calcula como

$$\hat{R}^2 = \frac{SS_{\text{E}} / (n-p-1)}{SS_{\text{T}} / (n-1)}$$

Que permite comparar el ajuste de varios modelos de regresión, siempre y cuando n y p sigan siendo los mismos y el modelo incluya el intercepto.

En el modelo de Regresión Simple $R = \sqrt{R^2}$ es el coeficiente de Correlación entre X y Y .

En el ejemplo:

Y_i	X_i	$\hat{Y}_i = Y_i - e_i$	$\hat{Y}_i - \bar{Y}$	e_i^2	$Y_i - \bar{Y}$
28	50	$28 - 0.69 = 27.31$	-16.1185	0.4761	-15.4285
30	450	$30 + 0.02 = 30.02$	-13.4085	0.0004	-13.4285
32	780	$32 + 0.26 = 32.26$	-11.1685	0.0676	-11.4285
36	1200	$36 - 0.89 = 35.11$	-8.3185	0.7921	-7.4285
51	4400	$51 + 5.82 = 56.82$	13.3915	33.87	7.5715
58	4800	$58 + 1.54 = 59.54$	16.1115	2.3716	14.5715
69	5300	$69 - 6.06 = 62.94$	19.5115	36.7236	25.5715

$$\bar{Y} = 43.4285$$

$$\bar{X} = 2425.714$$

$$SS_T = \sum (Y_i - \bar{Y})^2 = 1527.71$$

$$SS_R = \sum (\hat{Y}_i - \bar{Y})^2 = 1453.13$$

$$SS_E = \sum e_i^2 = SS_T - SS_R = 74.58$$

3) Se realiza la prueba global:

$H_0: \beta = 0$: No existe modelo

$H_1: \beta \neq 0$: Si existe modelo

Anova

Fuente	V	SS	Grado L	MS	Fo
Regresión	1	1453.13		1453.13	97.42
Error	5	74.581.0		14.916 = S^2	
Total	6	1527.71			

$S = 3.862$

$$F_{0.05, 1, 5} = 6.608$$

Como $97.42 > 6.608$: Si existe modelo

3) las Pruebas Individuales y los I.G.

$$\hat{\beta}_0 = 26.96$$

$$\hat{J}_{\hat{\beta}_0} = \frac{\sum X_i^2}{S^2} = \frac{14.916}{72743400} (72743400) = 31554771.43$$

$$H_0: \beta_0 = 0$$

$$H_1: \beta_0 \neq 0$$

$X_i - \bar{x}$	
-2375.714	
-1975.714	
-1645.714	
-1225.714	
1974.286	
2374.286	
2874.286	

$$\hat{J}_{\hat{\beta}_0} = 2.21$$

$$\sum (X_i^2) = 72743400$$

$$\sum (X_i - \bar{x})^2 = 31554771.43$$

$$t_0 = \frac{26.96 - 12.164}{2.21} = 2.09000.0$$

$$t_{0.025, 5} = 2.571$$

Como $12.164 > 2.571$: Si existe Intercepto

I.C. para β_0

$$26.96 - 2.571(2.21) \leq \beta_0 \leq 26.96 + 2.571(2.21)$$

$$21.27809 \leq \beta_0 \leq 32.64191$$

Para β_1

$$\hat{\sigma}_{\beta_1}^2 = \frac{s^2}{\sum(x_i - \bar{x})^2} = \frac{14.916}{31554771.43} = 0.000004727$$

$$\hat{\sigma}_{\beta_1} = 0.000687$$

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

$$t_1 = \frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}} = \frac{0.006786}{0.000687} = 9.87$$

$$t_{0.025, 5} = 2.571$$

Como $9.87 > 2.571$ entonces X si explica a Y

I.C. para β_1

$$0.006786 - 2.571(0.000687) \leq \beta_1 \leq 0.006786 + 2.571(0.000687)$$

$$0.005019723 \leq \beta_1 \leq 0.008552277$$

$$4) R^2 = \frac{SSR}{SS_T} = \frac{1453.13}{1527.71} = 0.9512$$

Las alturas explican los rayos cósmicos en un 95,12%

$$\hat{R}^2 = 1 - \frac{74.58}{5} = 0.9414$$

1527.71/6

CLASE 8

MODELOS DE REGRESIÓN LINEAL MÚLTIPLE

La mayoría de las veces resulta poco viable modelar el comportamiento de una variable y solamente en términos de una variable X .

$$\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon = Y$$

Para los casos en los cuales el comportamiento de una variable respuesta requiere ser explicado por más de una variable explicatoria, existe el análisis de regresión múltiple.

Un modelo de regresión múltiple sería:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

Donde:

$\beta_1, \beta_2, \dots, \beta_p$: Son los parámetros del modelo a estimar y son los llamados coeficientes del modelo de regresión.

ϵ : Es el término estocástico o error aleatorio que, al igual que en el modelo de regresión simple, contiene la información de todas las variables que no fueron incluidas de manera individual en el modelo, pero que, en conjunto, inciden en Y .

X_1, X_2, \dots, X_p : son las variables explicatorias, las cuales en principio, son variables matemáticas fijas con error de medición cero.

Y : Es la variable respuesta, ~~que~~ ~~también~~ variable aleatoria.

La Función de Regresión ~~y~~ muestra la relación entre ~~la~~

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_p X_p + \epsilon$$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_p X_p$$

Donde:

$\hat{\beta}_j$: Es el estimador para β_j .

$$\hat{Y} = E[Y]$$

ϵ : Es el residual que tiene interpretación análoga a ϵ .

Las ecuaciones para las N unidades experimentales de la población serían:

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \dots + \beta_p X_{p1} + \epsilon_1$$

$$Y_2 = \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \dots + \beta_p X_{p2} + \epsilon_2$$

$$\vdots$$

$$Y_N = \beta_0 + \beta_1 X_{1N} + \beta_2 X_{2N} + \dots + \beta_p X_{pN} + \epsilon_N$$

Que se pueden representar de forma matricial:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{p1} \\ 1 & X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1N} & X_{2N} & \dots & X_{pN} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}$$

Empleando Notación Matricial:

$$Y = X\beta + \epsilon : \text{FRP}$$

Nx1 Rxp+1) (p+1)x1 + Nx1

$$Nx1 = N \times (p+1) (p+1)^{-1} + Nx1$$

$$I^T D = [A] \text{ inv } (A)$$

Para las n observaciones de la muestra, las ecuaciones serían:

$$Y_1 = \hat{\beta}_0 + \hat{\beta}_1 X_{11} + \hat{\beta}_2 X_{21} + \dots + \hat{\beta}_p X_{p1} + \epsilon_1$$

$$Y_2 = \hat{\beta}_0 + \hat{\beta}_1 X_{12} + \hat{\beta}_2 X_{22} + \dots + \hat{\beta}_p X_{p2} + \epsilon_2$$

$$\vdots$$

$$Y_N = \hat{\beta}_0 + \hat{\beta}_1 X_{1N} + \hat{\beta}_2 X_{2N} + \dots + \hat{\beta}_p X_{pN} + \epsilon_N$$

En forma matricial:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{p1} \\ 1 & X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1N} & X_{2N} & \dots & X_{pN} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}$$

$$Y = X\hat{\beta} + \epsilon : \text{FRM}$$

$$\text{ESTIMACIÓN DE } \beta: 1x_1\beta_1 + \dots + ux_1\beta_u + u\epsilon = Y$$

$$s_1^2x_1^2 + s_2^2x_2^2 + \dots + s_{ux}^2x_{ux}^2 + u\epsilon^2 = S^2$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Si $X^T X$ es de rango completo, es decir que $|X^T X| \neq 0$, entonces hay una solución única para $\hat{\beta}$ y como:

$$E[Y] = X\beta$$

$$\text{Var}(Y) = \sigma^2 I$$

Entonces: Propiedades de $\hat{\beta}$

i) $E[\hat{\beta}] = \beta$

$$\beta + \delta X = Y$$

ii) $\text{Var}[\hat{\beta}] = \sigma^2 C$

C : Matriz de Constantes.

iii) $\text{Cov}[\beta_i, \beta_j] = \sigma^2 C_{ij}$

iv) $\hat{\beta}$ es el mejor estimador lineal insesgado de β , en la familia de estimadores insesgados que son función lineal de Y . Si, además, $\epsilon \sim \text{Normal}_n(0, \sigma^2 I)$, entonces $\hat{\beta}$ es el estimador por máxima verosimilitud y de mínima varianza.

INTERPRETACIÓN DE LOS COEFICIENTES DE REGRESIÓN

Un coeficiente de regresión $\hat{\beta}_j$ se interpreta como

la estimación del cambio promedio en la vble respuesta Y ante un cambio unitario en X_{ij} . Siempre y cuando las demás variables permanezcan constantes.

La interpretación de β_0 es análoga al modelo de regresión simple.

CLASE 9

ANÁLISIS DE VARIANZA (ANOVA)

En la regresión, el análisis de Varianza ANOVA, se hace con el fin de descomponer la varianza Total en sus fuentes de Variación. Así:

$$\text{Fenómeno} = \text{Modelo} + \text{Residual}$$

Si se parte de:

$$Y = X\beta + e \quad Y_i = \hat{Y}_i + e_i \quad \text{si se eleva al cuadrado}$$
$$Y_i^2 = \hat{Y}_i^2 + 2\hat{Y}_i e_i + e_i^2 \quad \text{y se suman } \hat{Y}_i$$
$$\sum Y_i^2 = \sum \hat{Y}_i^2 + 2 \sum \hat{Y}_i e_i + \sum e_i^2$$

entonces

$$\sum Y_i^2 = \sum \hat{Y}_i^2 + \sum e_i^2$$

que se denominan las sumas de cuadrados, donde

$$\sum Y_i^2 = \sum (Y_i - \bar{Y})^2: \text{Es la Suma de cuadrados totales}$$

$$\sum \hat{Y}_i^2 = \sum (\hat{Y}_i - \bar{Y})^2 \quad \text{Es la Suma de cuadrados del modelo}$$

$$\sum e_i^2: \text{Es la suma de Cuadrados del error.}$$

$$(50)$$

que matricial se calculan como

$$SS_T = \mathbf{y}^T \mathbf{y} - n\bar{y}^2$$

$$SS_R = \hat{\beta}^T \mathbf{X}^T \mathbf{y} - n\bar{y}^2$$

$$SS_E = \mathbf{y}^T \mathbf{y} - \hat{\beta}^T \mathbf{X}^T \mathbf{y}$$

$$MS_T = SS_T / (n-1)$$

$$S_R = SS_R / p$$

$$MS_E = SS_E / (n-p-1) = S^2$$

INFERNERIA ESTADISTICA

La inferencia estadistica en el modelo de Regresion parte de:

$$\epsilon \sim \text{Normal}(0, \sigma^2 I)$$

$$\mathbf{y} \sim \text{Normal}(\mathbf{X}\beta, \sigma^2 I)$$

Para las pruebas de hipotesis:

$$\beta \sim \text{Normal}(\beta, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$$

• Prueba global

Se prueba si las \mathbf{X} explican de manera global a \mathbf{y} , es decir:

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

Que se compara con el estadistico $F_{p, n-p-1}$. Si $F_0 > F_{p, n-p-1}$, se afirma que las \mathbf{X} si explican de manera global a \mathbf{y} , por tanto, si existe model.

Esto se prueba usando la tabla ANOVA

F de V	SS	gdeL	TTS	To
Modelo	SSR	P	$TTS = \frac{SSR}{P}$	TTS / TTS_E
Error	SSE	$n - P - 1$	$TTS_E = SSE / (n - P - 1)$	
Total	ST	$n - 1$		

• Pruebas individuales

De manera complementaria, se pueden plantear hipótesis acerca de los coeficientes de la regresión, así:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

Se usa el estadístico $t_j = \frac{\hat{\beta}_j}{\text{Op}_{\hat{\beta}_j}} ; \text{Op}_{\hat{\beta}_j} = \sqrt{\frac{S^2}{\sum(X_j - \bar{x}_j)^2}}$

En caso de rechazar H_0 , indica que la variable X_j sí contribuye a la variación de Y . En caso contrario, una vez validados los supuestos, X_j es una candidata a salir del modelo.

COEFICIENTE DE DETERMINACIÓN R^2

Es una medida del porcentaje en el que las X explican a Y . Se calcula como:

$$R^2 \% = \frac{\hat{B}^T \times T Y - n \bar{Y}^2}{T Y - n \bar{Y}^2} = \frac{SSR}{SSE} \times 100\%$$

Al igual que en la regresión simple debe tenerse cuidado con esta medida. Debe complementarse con

d) R^2 (R cuadrado ajustado)

$$\hat{R}^2 = 1 - \frac{MSE}{MS_{\text{Total}}}$$

Suponga que se incorpora una variable explicatoria X_{p+1} y se calcula su R cuadrado ajustado \hat{R}^2_{p+1} . Si la nueva variable explicatoria no aporta información suficiente al modelo, entonces $\hat{R}^2_{p+1} < \hat{R}^2_p$.

EJEMPLO: los siguientes datos corresponden a la edad, los ingresos y los años de universidad cursados por 5 ejecutivos. Realice la estimación del modelo, los procedimientos de inferencia estadística y el ajuste.

EDAD	ANOS DE UNIVERSIDAD	INGRESO Y
37	4	31200
45	0	26800
38	5	38000
42	2	30300
31	4	25400

a) $\hat{\beta}_0 = -16278.7$

$$\hat{\beta}_1 = 960.925$$
$$\hat{\beta}_2 = 2975.66$$

b) $t_0 = \frac{-16278.7}{1594.26} = -10.2108$ Rct 161 (-23138.2, -9419.13)

$$t_1 = \frac{960.925}{35.2946} = 27.2258 \quad Rct 161 (209.068, 1112.79)$$

$$t_2 = \frac{2975.66}{93.8797} = 31.6965 \quad R_c \text{ Ho: } (2571.73, 3379.81)$$

ANOVA

Fde Y	SS	g.l.	MS	F0
Modelo	5.7479×10^7	2	2.8739×10^7	151.13
Error	112456	2	56228.2	
Total	8.7592×10^7			

$$R^2 = 99.8047\% \quad \text{porcentaje de explicación del modelo}$$

$$\hat{R}^2 = 99.6095\%$$

CLASE 10

VALIDACIÓN DE SUPUESTOS EN EL MODELO DE REGRESIÓN.

Sean X_1, X_2, \dots, X_p un conjunto de variables no aleatorias, con base en las cuales se pretende obtener información acerca del valor promedio de una variable aleatoria Y , a partir de la observación de n unidades experimentales a través del modelo:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + \varepsilon_i ; \quad i=1, 2, \dots, n$$

Donde $\{X_{1i}, X_{2i}, \dots, X_{pi}, Y_i\}$ son características asociadas a la i -ésima unidad experimental. ε_i es una variable aleatoria que se conoce como error. $\beta_0, \beta_1, \dots, \beta_p$ son constantes. Los supuestos que acompañan el modelo son:

Los supuestos que acompañan el modelo son:

i) El modelo propuesto está correcto/. especificado.
Esto significa que:

$$E[\epsilon_i] = 0 ; \text{ para } i=1, 2, \dots, N$$

lo cual es equivalente con:

$$E[Y_i] = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}, \quad i=1, 2, \dots, N$$

ii) Homogeneidad de Varianza de los errores. Es decir:

$$\text{Var}[\epsilon_i] = \sigma^2, \text{ para todo } i=1, 2, \dots, N$$

la dispersión de los errores es igual en todos los puntos X_i involucrados.

iii) No correlación de los errores. Es decir:

$$\text{Cov}(\epsilon_i, \epsilon_j) = 0, \quad i \neq j$$

iv) Distribución normal de los errores.

$$i, \dots, n, \quad \epsilon_i \sim \text{Normal}(0, \sigma^2 I), \quad i=1, 2, \dots, N$$

los supuestos iii) y iv) implican independencia de los errores.

En forma matricial, el modelo y los supuestos quedan resumirse como:

$$Y = X\beta + \epsilon; \quad \epsilon \sim \text{Normal}(0, \sigma^2 I)$$

IMPORTANCIA DE LA VALIDACIÓN DE LOS SUPUESTOS

"Validar un supuesto" es no encontrar suficiente evidencia para rechazarlo; estrictamente se prueba que el supuesto no se cumple.

La importancia de validar los supuestos, radica fundamentalmente en que ellos inciden en las cualidades de los estimadores de mínimos cuadrados que son los más comúnmente usados, los más difundidos y para los cuales existe la mayor cantidad de software computacional.

Cuando se cumplen los supuestos i), ii) y iii) los estimadores de mínimos cuadrados son los mejores estimadores lineales insesgados, si además se cumple el supuesto iv) de normalidad, le abona eficiencia a dichos estimadores, proporcionándoles propiedades distribucionales que garantizan la disponibilidad de una gran cantidad de pruebas para validar hipótesis sobre los sus parámetros, construir I.C., predicciones, etc.

Antes de iniciar la validación de los supuestos se debe cumplir la NO EXISTENCIA DE MULTICOLINEALIDAD. Si existe, se debe solucionar y proceder a validarlos.

MULTICOLINEALIDAD

El análisis de regresión se realiza con el fin de establecer relaciones cuantitativas entre una variable respuesta y uno o más variables explicatorias. Tales relaciones permiten:

- Identificar los efectos relativos de las variables explicatorias
- Estimación y/o predicción.
- Selección de un conjunto apropiado de variables regresoras para el modelo.

Tales inferencias parten de:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \text{ que se deriva de}$$

$$(X^T X) \hat{\beta} = X^T Y \star$$

* que son las denominadas ecuaciones normales. Por lo que la matriz $X^T X$ es responsable de la estimación.

La COLINEALIDAD o MULTICOLINEALIDAD indica una relación lineal perfecta (exacta) o imperfecta (no exacta) entre las variables explicatorias (los X s).

Se da una relación lineal exacta, multicolinealidad exacta, si:

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_p x_p = 0$$

Donde los α_i son constantes que, no todas son iguales a cero.

Se presenta multicolinealidad imperfecta, relación lineal no exacta, cuando:

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_p x_p + v_i \approx 0$$

Donde v_i es un término de error estocástico.

A medida que crece la colinealidad, el coeficiente de Correlación tiende a 1, siendo 1 cuando la colinealidad es perfecta.

Consecuencias de la multicolinealidad

I. Si $R_j^2 = R^2$ en la regresión (de x_j sobre las $(p-1)$ restantes variables explicatorias), entonces a medida que R_j se acerca a 1, también aumentan las varianzas de los β_j 's y, si $R_j^2 = 1$, las varianzas son infinitas. De la misma manera, aumentan las covarianzas en Valor absoluto, así:

$$\text{Var}(\hat{\beta}_j) \rightarrow \infty$$

$$|\text{Cov}(\hat{\beta}_i, \hat{\beta}_j)| \rightarrow \infty$$

La velocidad con la que aumentan las varianzas y las covarianzas de los estimadores se mide con el

Valor de Inflación de la Varianza VIF:

$$VIF = \frac{1}{1 - R_j^2}$$

El VIF es un indicador de la manera en la que la varianza de un estimador se infla por efectos de la multicolinealidad.

II. De acuerdo a I, el estadístico de prueba:

$$t_j = \frac{\hat{\beta}_j}{\text{S.E.}(\hat{\beta}_j)} \rightarrow 0$$

entonces $H_0: \beta_j = 0$ no se rechazaría.

III. Por I y II, es posible que el rechazo de H_0 solo ocurra sobre unos pocos coeficientes de regresión y, sin embargo, se tengan un R^2 alto. Es decir, que puede tenerse un F alto para el rechazo de

$H_0: \beta = 0$ (Tabla ANOVA), aun así, las pruebas individuales para β_j no lleven al rechazo de

$H_0: \beta_j = 0$. "Prueba global se Rechaza, Pruebas individuales se aceptan".

IV. los signos de los coeficientes de regresión estimados pueden contradecir el efecto real que presentan las variables explicatorias en la verdadera regresión poblacional.

V. los estimadores son muy sensibles a conjuntos

particulares de datos.

Causas de la multicolinealidad

1. El método de recolección de la información.
2. Restricciones en el modelo o la población objeto de muestreo.
3. Especificación del modelo: sobreespecificación subespecificación.
4. Modelo sobre determinado

Detección de la multicolinealidad

1. Las pruebas individuales son no significativas (Valor $p > \alpha$) y la prueba global ANOVA muestra significancia, adicionalmente R^2 es elevado.
2. Las correlaciones entre pares de variables x_i, x_j son mayores en valor absoluto, es decir:
 $|f_{ij}| > 0.5$
3. $VIF > 10$. Entre mayor sea este valor, mayor colinealidad existe.
4. Análisis de las raíces características de $X^T X$. Entre mayor sea la colinealidad, una o más raíces características serán pequeñas. Además:

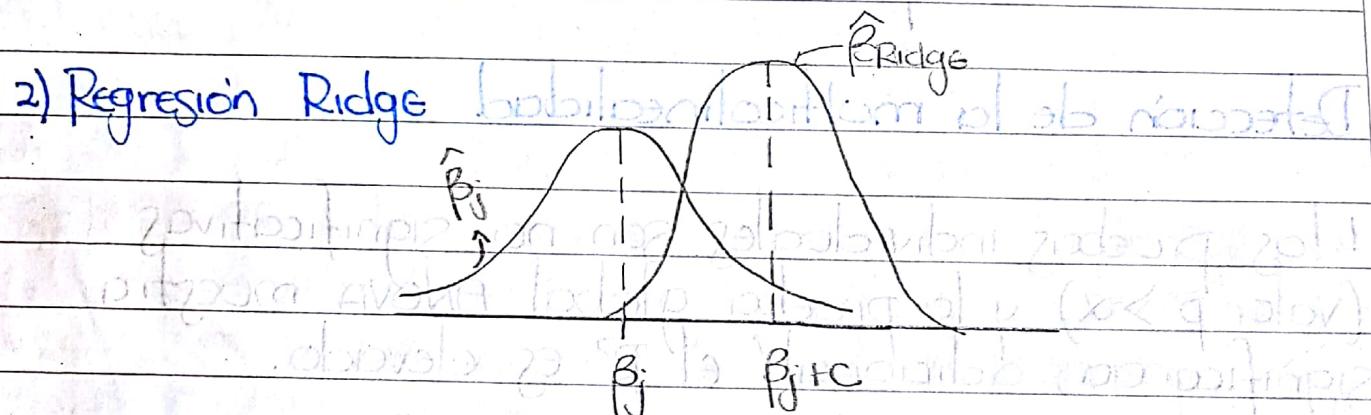
$$k = \frac{\lambda_{\max}}{\lambda_{\min}}$$

- Si $k < 100$, no hay multicolinealidad
- Si $100 \leq k \leq 1000$, hay multicolinealidad de moderada a fuerte.
- Si $k > 1000$, hay multicolinealidad severa.

CLASE 11.

Solución de la multicolinealidad

1) Convivir con la multicolinealidad



Si se parte del modelo:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \epsilon_i$$

Si se estandarizan los datos:

$$y_i^* = \frac{Y_i - \bar{Y}}{S_{Y_i}}$$

$$x_j^* = \frac{X_{ji} - \bar{X}_j}{S_{X_{ji}}}$$

y se aplica TCO, entonces

$$\beta_j = \beta_j^* \frac{y_i}{x_j} \Rightarrow \text{SESCADOS}$$

Con

$$\beta_j^* = (X^T X + C I)^{-1} X^T y$$

Con

$$0 \leq C \leq 1$$

Se hace la traza Ridge y se elige el valor de C que da la traza más estable.

3) Componentes Principales (CP)

El análisis de Componentes Principales (CP) es un procedimiento matemático que transforma un conjunto de variables CORRELACIONADAS en un conjunto menor de variables NO CORRELACIONADAS, llamadas COMPONENTES PRINCIPALES, con el propósito de reducir la dimensión de los datos y facilitar su interpretación y análisis estadístico.

En el análisis de Componentes Principales se busca MAXIMIZAR la variación de una Combinación lineal de Variables.

Suponga que X_{pp} es un vector aleatorio con matriz de covarianzas $\Sigma_{pp} = \sum_{pp} x_i x_i^T$ cuyos valores propios son $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, entonces:

$$Y_1 = a_1^T X = a_{11} x_1 + a_{12} x_2 + \dots + a_{1p} x_p$$

$$Y_2 = a_2^T X = a_{21} x_1 + a_{22} x_2 + \dots + a_{2p} x_p$$

$$\vdots$$

$$Y_p = a_p^T X = a_{p1} x_1 + a_{p2} x_2 + \dots + a_{pp} x_p$$

donde

$$A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{vmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

es una matriz ortogonal:

Entonces $A^T A = I$ y a_{ij} tienen el signo?

la i -ésima componente de x es igual al i -ésimo?

Supongamos también las parejas de valores y vectores propios de Σ .

$$(\lambda_1, c_1); (\lambda_2, c_2); \dots; (\lambda_p, c_p)$$

↓ Vector propio

Valor propio

donde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$

Entonces, la i -ésima componente principal está dada por la combinación lineal:

$$y_i = c_i^T x = c_{i1} x_1 + c_{i2} x_2 + \dots + c_{ip} x_p$$

donde los signos están en c_i .

$$\text{Var}(y_i) = c_i^T \Sigma c_i = \lambda_i$$

$$\text{Cov}(y_i, y_k) = c_i^T \Sigma c_k = 0 \quad i \neq k$$

$$x_1^2 p + \dots + x_p^2 p = \lambda_i p = k$$

$$\text{Varianza Total} = \lambda_1 + \lambda_2 + \dots + \lambda_p \quad \lambda_i p = k$$

Por ejemplo, suponga las variables x_1, x_2 y x_3 , con matriz de var-cov:

$$\Sigma = \begin{bmatrix} \text{Cov}(x_1, x_1) & \text{Cov}(x_1, x_2) & \text{Cov}(x_1, x_3) \\ -2 & \text{Cov}(x_2, x_2) & \text{Cov}(x_2, x_3) \\ 0 & 0 & \text{Cov}(x_3, x_3) \end{bmatrix}$$

Cuyos eigenvalores son:

$$\lambda_1 = 5.83 \quad \lambda_2 = 2 \quad \lambda_3 = 0.17$$

y sus correspondientes eigenvectores

$$C_1^T = [0.383 \quad -0.924 \quad 0]$$

$$C_2^T = [0 \quad 0 \quad 1]$$

$$C_3^T = [0.924 \quad 0.383 \quad 0]$$

Encuentre los CP de $X^T [x_1 \ x_2 \ x_3]$

Partimos de $Y_i = C_i^T X$

$$Y_1 = C_1^T X = [0.383 \quad -0.924 \quad 0] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= 0.383 x_1 - 0.924 x_2$$

$$Y_2 = C_2^T X = [0 \quad 0 \quad 1] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= x_3$$

$$Y_3 = C_3^T X = \{0.924, 0.383, 0\} \begin{vmatrix} X_1 \\ X_2 \\ X_3 \end{vmatrix}$$

$$= 0.924X_1 + 0.383X_2$$

Quedaria así:

$$Y_1 = 0.383X_1 - 0.942X_2$$

$$Y_2 = X_3$$

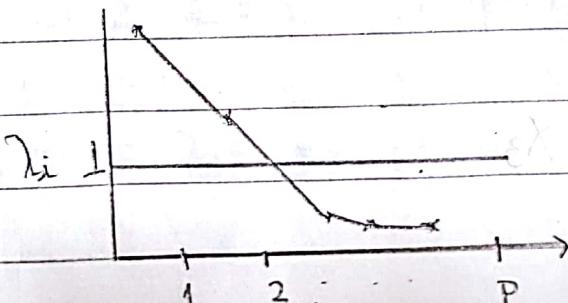
$$Y_3 = 0.924X_1 + 0.383X_2$$

X_3 está en una sola componente, esto se debe a que no está correlacionada con las demás.

Nota: la estandarización de las variables afecta los resultados en el análisis de Componentes Principales; así que debe tenerse cuidado al hacerlo solo cuando las variables originales difieren mucho en escala.

Elección del número de componentes

Se determina a través del gráfico sobre λ_i vs. i , ordenados de acuerdo a su magnitud, entonces cuando los valores propios tiendan a nivelarse "formen un codo", estos valores propios se consideran cercanos a cero y pueden ignorarse.



VIOLACIÓN DE LOS SUPUESTOS

1. Incorrecta especificación del modelo

Ocurre cuando se incumple el modelo, en el sentido en que

$$E[\epsilon] \neq 0, \text{ es decir}$$

$$E[Y] \neq X\beta$$

Un modelo puede estar incorrectamente especificado por:

- Excluir variables relevantes
- Incluir variables irrelevantes
- Planteamiento equivocado de la relación entre la variable respuesta Y , el conjunto de variables predictoras X y el error ϵ ; por ejemplo ajustar un modelo lineal que no lo sea.
- Excluir Variables relevantes

Cuando hay falta de ajuste, las esperanzas de los residuales $E(\epsilon) \neq 0$, lo cual se reflejaría a través de puntos "outliers", es decir, los residuales que se encuentran lejos de cero, lo cual será más evidente cuanto más importantes sean las variables excluidas.

b) Incluir variables irrelevantes

La varianza de los coeficientes del modelo $\text{Var}(\hat{\beta}) = (X^T X)^{-1} \sigma^2$ es inflada al considerar variables adicionales (alas necesarias) $(n - p)$

Consecuencias de una mala especificación del modelo

La consecuencia más relevante en la estimación es el subajuste o sobreajuste del modelo. Es de anotar que una incorrecta especificación del modelo

puede causar heteroscedasticidad. Por lo tanto, cuando un modelo se encuentre mal especificado y, a la vez, presente heteroscedasticidad, es recomendable solucionar el problema de especificación.

Detección

Se realiza la prueba R RESET de Ramsey donde se prueba la hipótesis:

$$H_0: \text{El modelo está correcto (específico)}$$

$$H_1: \text{No es correcto (incorrecto)}$$

los pasos para realizar la prueba son:

a) Realice la regresión por MCO y obtenga el \hat{Y} y el R^2

b) Realice una regresión auxiliar incluyendo \hat{Y} entre las variables explicativas y obtenga R_{aux}^2

c) Construya el estadístico de prueba:

$$F_0 = \frac{R^2 - R_{aux}^2}{1 - R_{aux}^2}$$

$(n - \# \text{ de parámetros de la reg. auxiliar})$

Si $F_0 > F_{\alpha, p-1, n}$, entonces Rechaza H_0 , es decir, si el valor p tiende a cero, se afirma que el modelo se encuentra incorrecto (específico).

Solución

Se recomienda el uso de transformaciones de potencia entre -2 y 2 para estimar por MCG, recuerde que $\lambda=0 \Rightarrow \ln$.

2. No homogeneidad de la varianza: Heteroscedasticidad

$$Y = X\beta + \epsilon \quad E[Y] = E[X\beta + \epsilon] = \beta$$

la heteroscedasticidad se presenta cuando:

$$\text{Var}[\epsilon] \neq \sigma^2 I$$

Entonces el estimador de Mínimos Cuadrados Ordinarios (MCO):

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Continua siendo insesgado:

$$\begin{aligned} E[\hat{\beta}] &= E[(X^T X)^{-1} X^T Y] \\ &= (X^T X)^{-1} X^T E[Y] = \beta \end{aligned}$$

Consecuencias de la heteroscedasticidad.

Sin embargo, pierde su condición de optimidad en el conjunto de estimadores lineales insesgados, es decir, no es el Mejor Estimador Lineal Insesgado (no es MELI), porque no tienen Varianza mínima.

La matriz de Varianzas-Covarianzas del vector de errores puede ser diferente de $\sigma^2 I$ bien sea porque los errores son correlacionados o porque las varianzas son distintas (o ambas). Para el primer caso: Se conoce como AUTOCORRELACIÓN, el segundo como HETEROSCEDASTICIDAD.

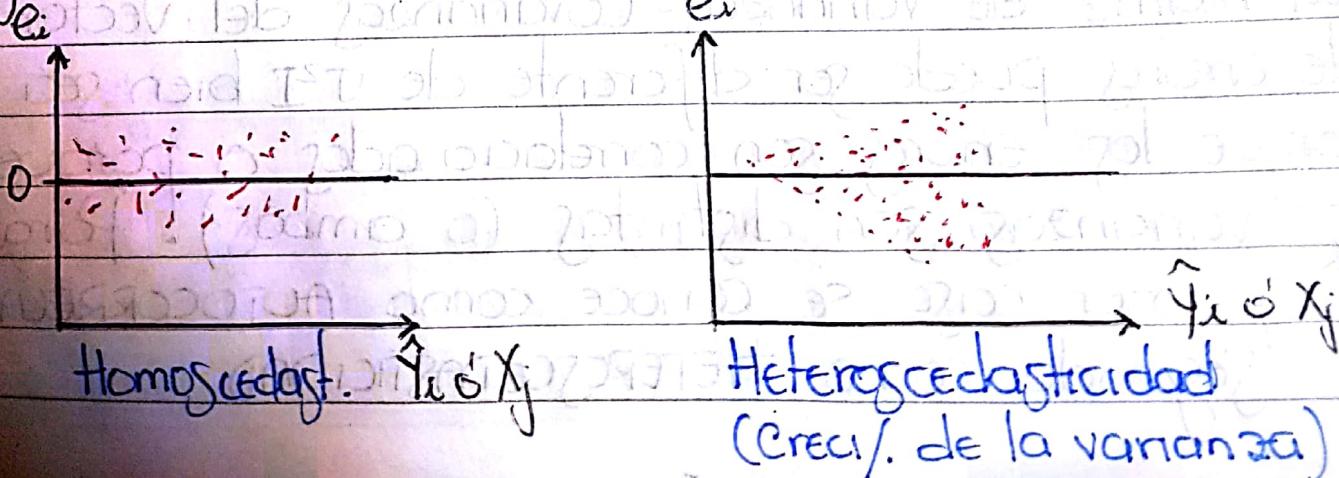
Causas de la heteroscedasticidad

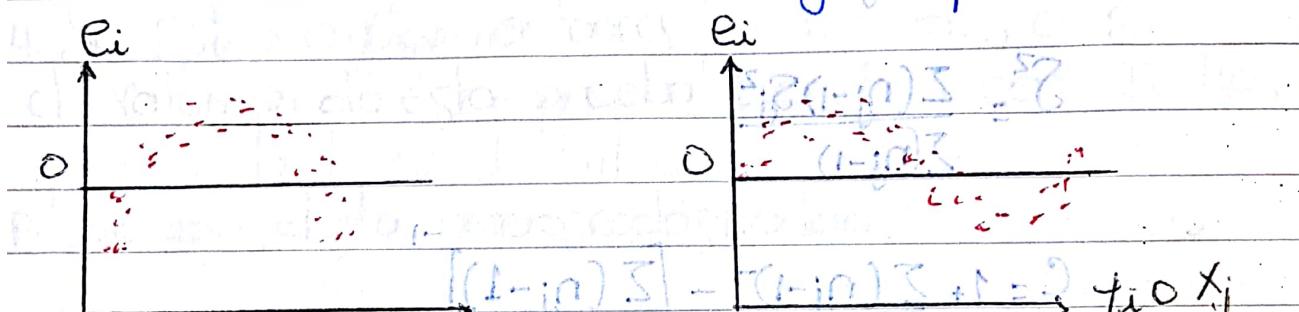
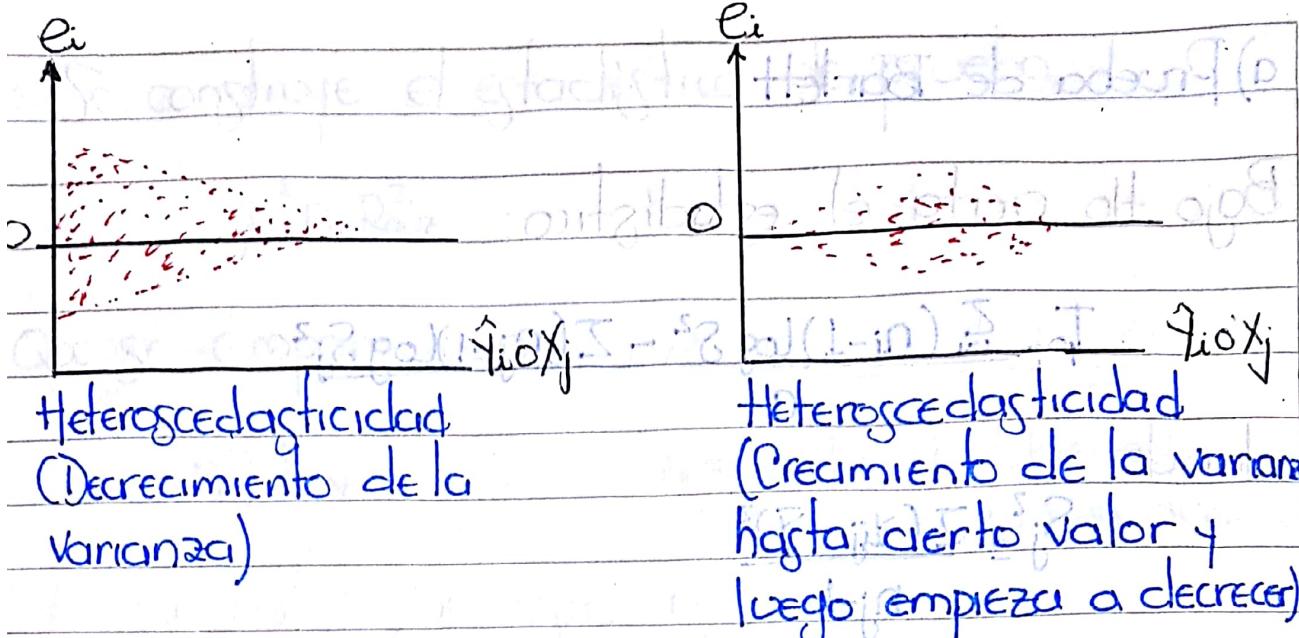
- Problemas en la recolección de la información.
- Presencia de valores atípicos
- Incorrecta especificación del modelo
- Asimetría (cola pegada) en una o más variables explicatorias.
- Incorrecta transformación de los datos y forma funcional.

Detección de la heteroscedasticidad

Existen varios métodos de detección, entre los que se encuentran los métodos gráficos y las pruebas.

Método gráfico: los gráficos más usados para la detección de heteroscedasticidad son los que se construyen con los residuales en el eje y y los predichos (\hat{y}) o alguna variable predictora (x) en el eje x. Algunos gráficos típicos pueden ser los siguientes:





Heteroscedasticidad

Heteroscedasticidad

Heterogeneidad

- Pruebas para detectar heteroscedasticidad: Para realizar estas pruebas se supone que se tienen agrupados los errores aleatorios en K grupos para los que se supone que dentro de cada grupo los errores tienen igual varianza. Así la varianza de los errores del grupo j es σ_j^2 . ($j = 1, 2, \dots, K$), se desea probar la hipótesis:

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma^2$: Homoscedasticidad

H1: al menos un par $\sigma_i^2 \neq \sigma_j^2$: heteroscedasticidad

a) Prueba de Bartlett

Bajo H_0 cierta, el estadístico:

$$T = \sum_{j=1}^k (n_j - 1) \log S^2 - \sum (n_j - 1) \log S_j^2$$

donde $S^2 = \frac{1}{C} \sum (y_{ij} - \bar{y})^2$

$$S_j^2 = \frac{1}{n_j - 1} \sum (y_{ij} - \bar{y}_j)^2$$

$$S^2 = \frac{\sum (n_j - 1) S_j^2}{\sum (n_j - 1)}$$

$$C = 1 + \frac{\sum (n_j - 1)^{-1}}{3(k-1)} - \left[\frac{\sum (n_j - 1)}{3(k-1)} \right]^2$$

Si $T_1 > \chi^2_{\alpha, k-1}$ Rechaza H_0 .

Esta prueba es muy sensible a la no normalidad de las x_i , se recomienda que, antes de usarla, se realice un diagnóstico de normalidad.

b) Prueba de White

Se construye de la sgte manera:

1. Se estiman los residuales del modelo original

2. Se realizó una regresión auxiliar de los residuales al cuadrado (e_i^2) como variable explicada y las x_i como explicadoras junto con cada x_j^2 y sus productos cruzados. Tomar el R^2 .

3. Se construye el estadístico de prueba

Como en $\chi^2 = n \cdot R_{aux}$

Que se compara con

$n \chi^2_{\alpha, v}$ ($v = \text{grados de libertad}$)
de SST en la Reg. aux.

4. Si $\chi^2 > n \chi^2_{\alpha, v}$ entonces Rechazación, o si no el valor p de esta prueba tiende a cero. Pcto.

Solución de la heteroscedasticidad

Se recomiendan soluciones como el uso de Minimos Cuadrados Ponderados (MCPP) y el estimador de White. Sin embargo, una tendencia generalizada es realizar transformaciones de potencia sobre las observaciones y_i con el propósito de estabilizar varianza. Algunas transformaciones sugeridas son:

\sqrt{y} : Para conteos de Poisson

$\sqrt{y} + \sqrt{y+1}$: Para cuando algunos y_i son muy pequeños

$\log y$: El rango de y es muy amplio

$\log(y+1)$: Si hay valores de $y_i = 0$

$1/y$: Cuando las x_j están cercanas a cero

$1/(y+1)$: Si hay $y_i = 0$

$\operatorname{sen}(\sqrt{y})$: Para proporciones binomiales

CLASE 14.

3. Correlación de los errores

En los estudios transversales, a menudo los datos se recopilan para unidades transversales (unidades experimentales) como familias, empresas, casas, universidades, autos, etc., de modo que no existe razón previa para creer que el término de error que corresponde a un caso, por ejemplo, esté correlacionado con el término de error de otro caso. Si por casualidad, se observa dicha correlación entre unidades experimentales, se conoce como AUTOCORRELACIÓN ESPACIAL.

Sin embargo, cuando las observaciones se forman como una secuencia en el tiempo, puede presentarse autocorrelación de los errores tal como ocurre en los fenómenos que son objeto de estudio en SERIES DE TIEMPO, éste tipo de correlación entre los errores se conoce como AUTOCORRELACIÓN SERIAL.

El supuesto de no correlación de los errores, no autocorrelación, simbólica/mental:

$$\text{Cov}(e_i, e_j / X_i, X_j) = E(e_i e_j) = 0 \quad i \neq j$$

Cuando existe autocorrelación

$$E(e_i e_j) \neq 0$$

Consecuencias de la autocorrelación

Como en la heteroscedasticidad, en presencia de autocorrelación los estimadores continúan siendo lineales e insesgados, al igual que consistentes, y están distribuidos de forma asintótica normal, pero dejan de ser eficientes, pero dejan de ser eficientes (es decir, no tienen varianza mínima).

Por consiguiente, las pruebas de significancia t y F usuales dejan de ser válidas y, de aplicarse, es probable que conduzcan a conclusiones erróneas sobre la significancia estadística de los coeficientes de regresión estimados.

Causas de la autocorrelación

1. Inercia: aplica para las series de tiempo (correlación serial) e indica que las observaciones sucesivas son interdependientes. Datos en serie de tiempo.
2. Sesgo en la especificación por omisión de variables relevantes.
3. Manipulación de los datos.
4. Transformaciones incorrectas de los datos.

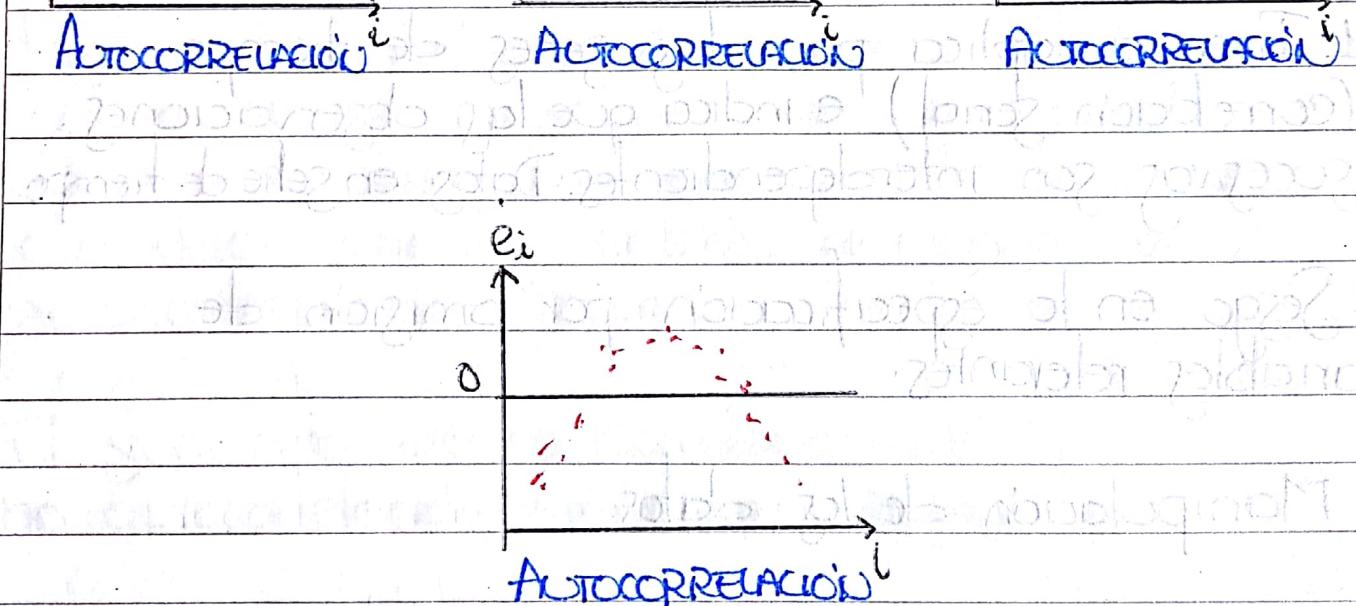
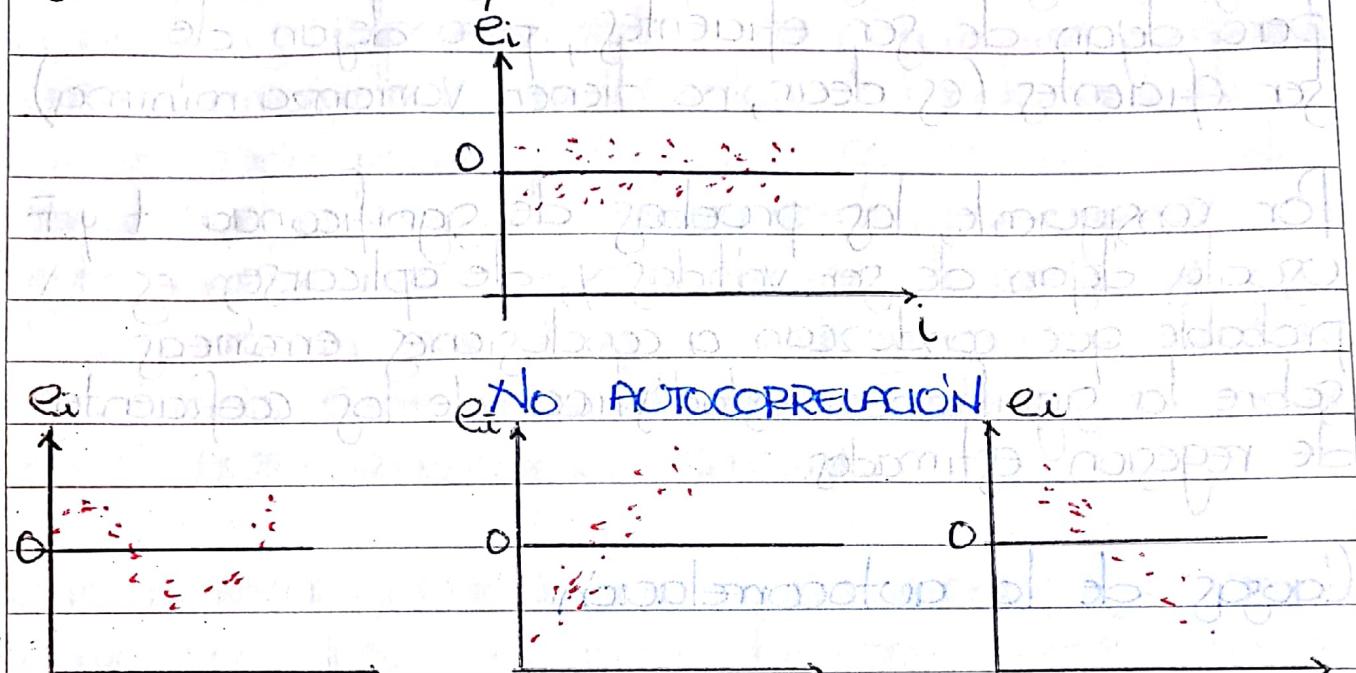
Detección de la autocorrelación

Se usa el diagnóstico gráfico y, si los datos corresponden a una serie de tiempo, la prueba de rachas.

DURBIN-WATSON

- **Método gráfico**

Se realiza analizando el gráfico de los residuos vs. el número de fila. Así:



- **Prueba de Rachas**

Dada una sucesión ordenada de dos tipos de eventos, una racha es una sucesión de uno o más eventos del mismo tipo que están seguidos y precedidos

por eventos del mismo tipo o por ninguno. Por ejemplo, si se listara la secuencia de los signos de 10 residuales y resultara así: + + + + - - - - - -

Se evidencian dos rachas.

Si la situación fuera:

- + - + - + - + - + - Se tendría 10 rachas.

Ambos casos son sospechosos de no aleatoriedad.

Ahora, sea:

N = Número total de observaciones

N_1 = número de símbolos + (es decir residuos positivos)

N_2 = " - (es decir residuos negativos)

R = número de rachas

Para probar las hipótesis:

H_0 : los residuos son independientes (no autocorrelación)

H_1 : (H_0) no es cierto (si: H_1 es cierto)

Se calcula; siempre que $N_1 > 10$ y $N_2 > 10$

R = el número de rachas $N = N_1 + N_2$

$$E(R) = \frac{2N_1N_2}{N} + 1 \quad \sigma_R^2 = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N-1)}$$

y el intervalo de Confianza al 95%.

$$E(R) - 1.96\sigma_R \leq R \leq E(R) + 1.96\sigma_R$$

Rechace H_0 si R se encuentra fuera del I.C.

• Prueba Durbin Watson

La prueba más conocida para detectar correlación serial es la del estadístico d de Durbin Watson, que se define como:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \approx 2(1 - \frac{\sum e_t e_{t-1}}{\sum e_t^2}) \approx 2(1 - \rho)$$

Por tanto como $0 \leq \rho \leq 1$, entonces

$$0 \leq d \leq 4$$

H_0 : No hay autocorrelación

H_1 : Si

(Si $d \approx 2$: no) hay autocorrelación ($\rho=0$)

$d \approx 0$: Si hay autocorrelación (ρ positiva ($\rho=1$))

$d \approx 4$: Si hay autocorrelación (ρ negativa ($\rho=-1$))

Este procedimiento se puede usar siempre y

cuando se cumpla que: $\hat{e}_t \perp \hat{e}_s$ para $t \neq s$

1. El modelo incluya el intercepto

2. Las X_s sean variables fijas

3. e_t sean un esquema autoregresivo de orden 1 (AR1)

4. $e_t \sim \text{Normal}(0, \sigma^2)$

5. y_t no esté resarcida.

6. No hay observaciones faltantes.

Solución de la autocorrelación

1. Transformaciones de potencia

$$Y^{\alpha} \quad \text{con } -2 \leq \alpha \leq 2$$

$$X_j^{\lambda} \quad \text{con } -2 \leq \lambda \leq 2$$

2. Si es una serie de tiempo, incorporar la estructura autoregresiva al modelo.

CLASE 15

4. No normalidad de los errores

El supuesto de normalidad exigido a los errores en el modelo de regresión lineal, permite la estimación por I.C. no solo para los coeficientes de regresión, sino también para la predicción. Permite el planteamiento de pruebas de hipótesis sobre los parámetros del modelo, las que a su vez facilitan procedimientos de selección de variables. Además permite la construcción de pruebas de bondad de ajuste, de homogeneidad de varianzas, de correlación de errores, etc.

Como los errores que no son observables, las pruebas se realizan sobre los residuales, el supuesto de normalidad de los errores es:

$$e_i \stackrel{iid}{\sim} \text{Normal}(0, \sigma^2)$$

Consecuencias de la no normalidad de los errores

Cuando los errores no son normales, las pruebas mencionadas y los intervalos de Confianza no son exactos, pudiendo llegar al extremo de ser inválidas.

Detección de la no normalidad

Existen diversas pruebas de diagnóstico entre las que se encuentran los métodos gráficos y las pruebas de bondad de ajuste.

→ **Métodos gráficos:** gráfico de Probabilidad normal.

→ **Pruebas:** Shapiro Wilks, Smirnov - Kolmogorov

Para estas pruebas se establecen las hipótesis:

$$H_0: e_i \sim \text{Normal}(0, \sigma^2)$$

$$H_1: e_i \text{No} \sim \text{Normal}(0, \sigma^2)$$

Donde el criterio de rechazo es:

Rechazo si valor p < λ

Solución de la no normalidad

Transformaciones de Potencia

$$y^\alpha \quad \alpha \in [-2, 2] \quad \alpha = 0 \Rightarrow \ln(y)$$

$$x_j^\lambda \quad \lambda \in [-2, 2] \quad \lambda = 0 \Rightarrow \ln(x_j)$$