

Guía 7: Regresión y Correlación

Estadística - Ingeniería en
Informática

Primer Cuatrimestre 2023

The background of the slide is composed of two large, overlapping geometric shapes. A teal-colored shape occupies the top-left corner, while a light gray shape occupies the bottom-left corner. The rest of the slide is white.

Ejercicio Resuelto...

Ventas y Publicidad

Un comerciante minorista realizó un estudio para determinar la relación que hay entre los gastos semanales de publicidad y las ventas. Los datos son (en miles de pesos):

Costos de publicidad (\$)	Ventas (\$)
40	385
20	400
25	395
20	365
30	475
50	440
40	490
20	420
50	560
40	525
25	480
50	510

Gráfico de Dispersión

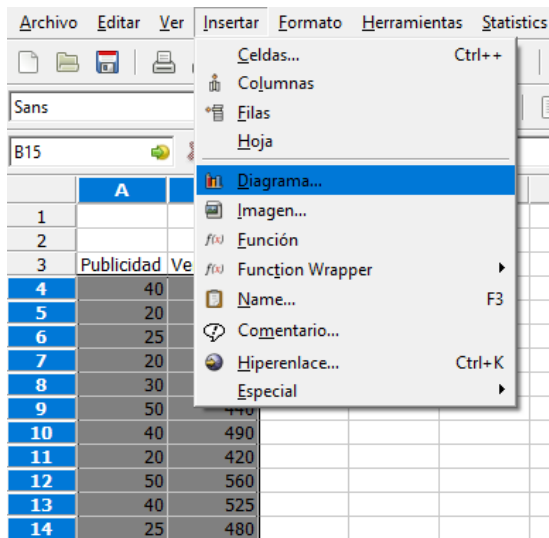


Gráfico de Dispersión

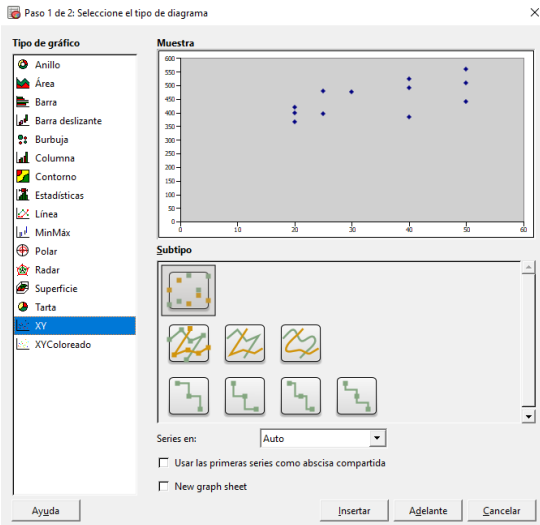
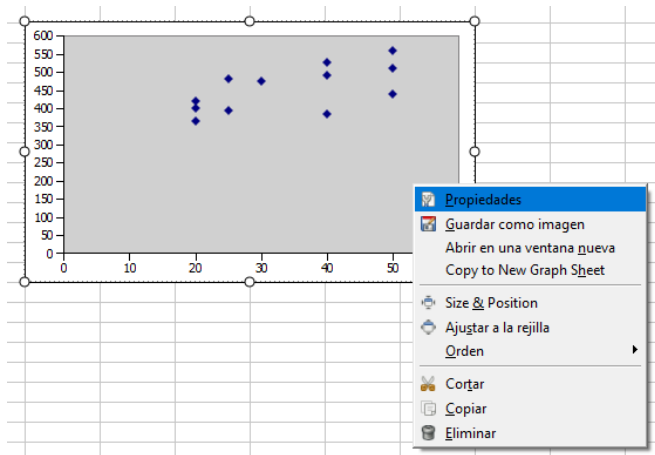
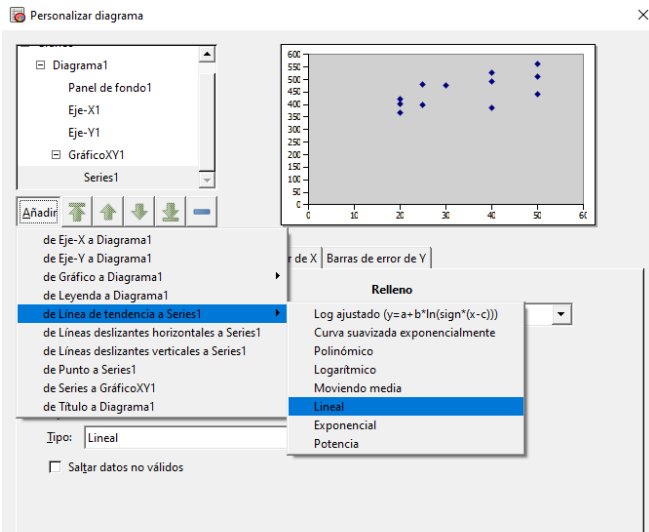


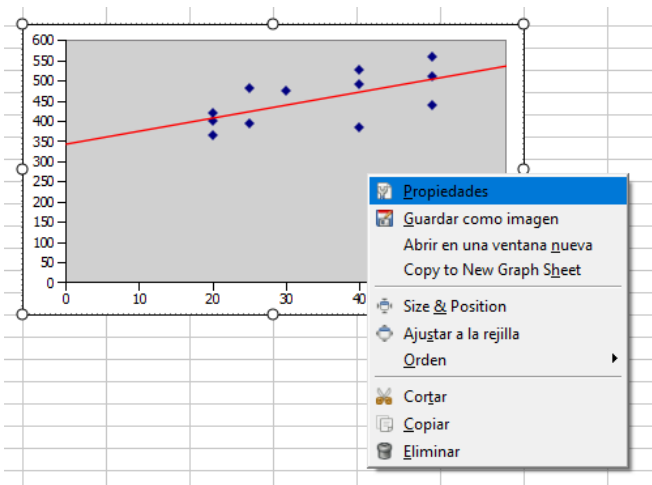
Gráfico de Dispersión



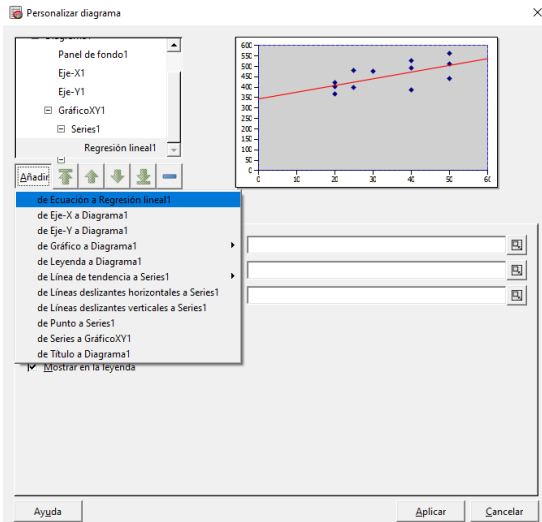
Recta de Regresión



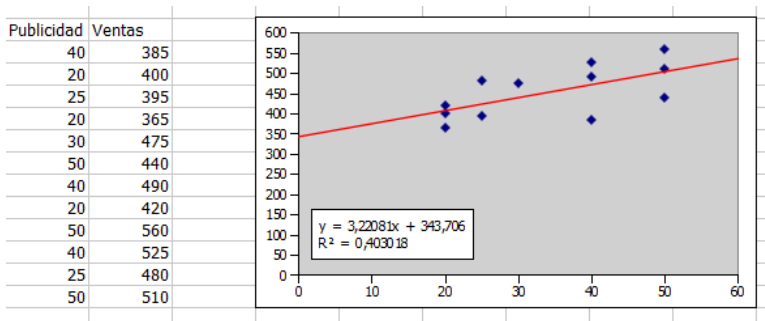
Recta de Regresión



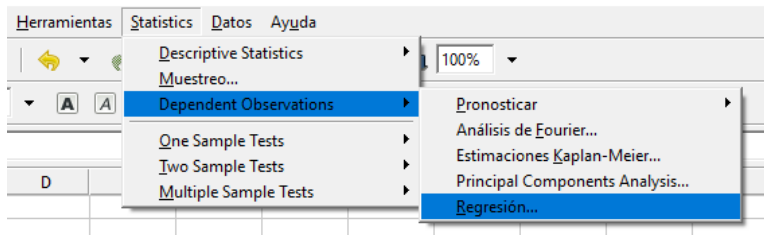
Recta de Regresión



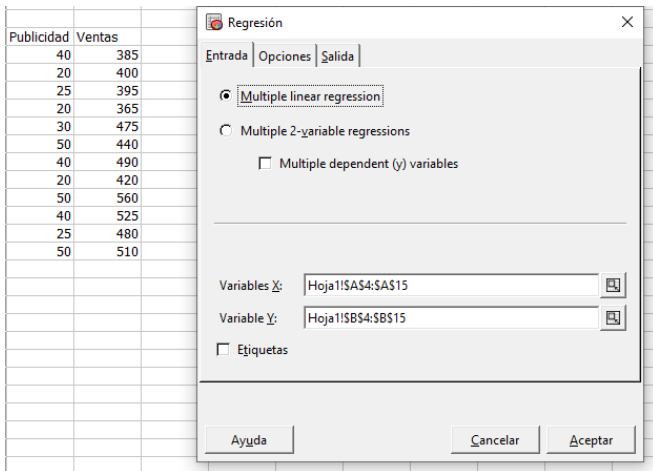
Recta de Regresión



Análisis de la Varianza de la Regresión



Análisis de la Varianza de la Regresión



Análisis de la Varianza de la Regresión

SALIDA RESUMEN		Response Variable:	Columna 2			
Estadísticas de regresión						
R múltiple	0,63483726753313					
Cuadrado R	0,40301835624893					
Error estándar	50,2256961960273					
Cuadrado R ajustado	0,34332019187382					
Observaciones	12					
Análisis de varianza						
	df	Suma de los cuadrados	Media cuadrática	F	Significancia de F	
Regresión	1	17030,0444162437	17030,0444162437	6,75093380956573	0,02657089506158	
Residual	10	25226,2055837563	2522,62055837563			
Total	11	42256,25				
	Coefficientes	Error estándar	Estadísticas-t	Valor-P	Intervalos de Confianza	
Interceptar	343,705583756345	44,7661786190334	7,67779592449311	1,6847392283314E-005	243,960321920305	443,450845592385
Columna 1	3,22081218274112	1,23960544526242	2,59825591687303	0,02657089506158	0,45879912901778	5,98282523646446

Importante: $S^2_{y|x} = CME = \frac{SCE}{n-2} = 2522,62$

Análisis de la Varianza de la Regresión

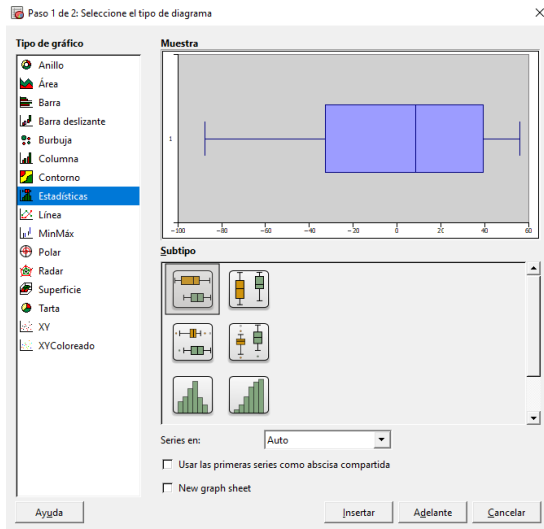
<i>Constant</i>	<i>Columna 1</i>	<i>Prediction</i>	<i>Columna 2</i>	<i>Residual</i>	<i>Leverages</i>	<i>Internally studentized</i>	<i>Externally studentized</i>	<i>p-Value</i>
1	40	472,53807106599	385	-87,5380710659898	0,10406091370563	-1,8413	-2,0935	6,58%
1	20	408,121827411167	400	-8,12182741116749	0,20558375634495	-0,1814	-0,1723	86,70%
1	25	424,225888324873	395	-29,2258883248731	0,13451776649726	-0,6255	-0,6037	56,10%
1	20	408,121827411167	365	-43,1218274111675	0,20558375634495	-0,9633	-0,9495	36,72%
1	30	440,329949238579	475	34,6700507614214	0,09390862944147	0,7252	0,7050	49,87%
1	50	504,746192893401	440	-64,746192893401	0,23604060913742	-1,4749	-1,5323	15,98%
1	40	472,53807106599	490	17,4619289340102	0,10406091370563	0,3673	0,3506	73,40%
1	20	408,121827411167	420	11,8781725888325	0,20558375634495	0,2653	0,2524	80,64%
1	50	504,746192893401	560	55,253807106599	0,23604060913742	1,2586	1,2736	23,47%
1	40	472,53807106599	525	52,4619289340102	0,10406091370563	1,1035	1,1091	29,61%
1	25	424,225888324873	480	55,7741116751269	0,13451776649726	1,1937	1,2094	25,73%
1	50	504,746192893401	510	5,25380710659897	0,23604060913742	0,1197	0,1136	91,20%

Análisis de los Residuos

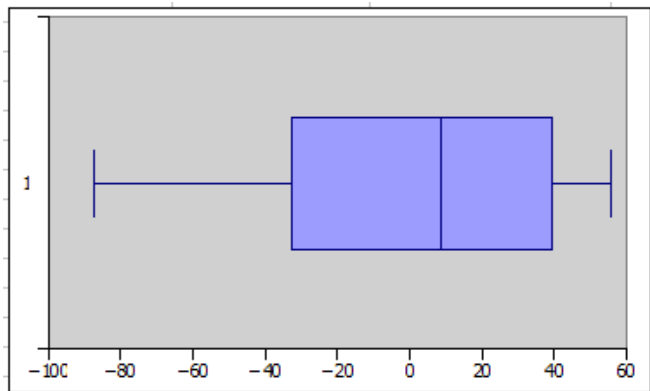
The screenshot shows a spreadsheet application with the 'Insertar' menu open. The 'Diagrama...' option is highlighted. The spreadsheet data is as follows:

	A		D	E
			Columna 2	Residual
37	Constant			
38			807106599	-87,5380710659898
39			827411167	-8,12182741116749
40			888324873	-29,2258883248731
41			827411167	-43,1218274111675
42			949238579	34,6700507614214
43			192893401	-64,746192893401
44			807106599	17,4619289340102
45			406,121827411167	11,8781725888325
46			504,746192893401	55,253807106599
47			472,53807106599	52,4619289340102
48			424,225888324873	55,7741116751269
49			504,746192893401	5,25380710659897
50				

Análisis de los Residuos



Análisis de los Residuos



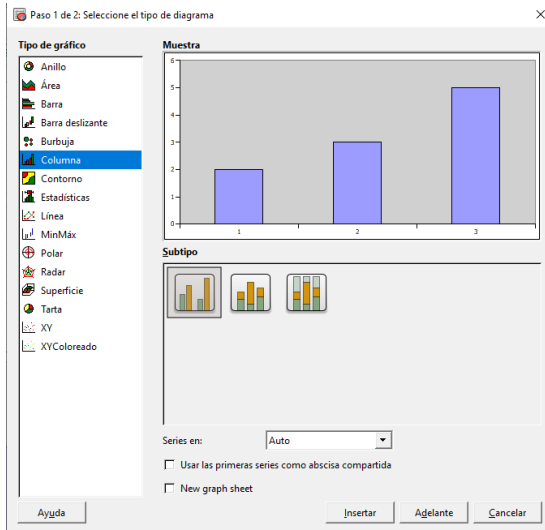
Análisis de los Residuos

Statistics Datos Ayuda

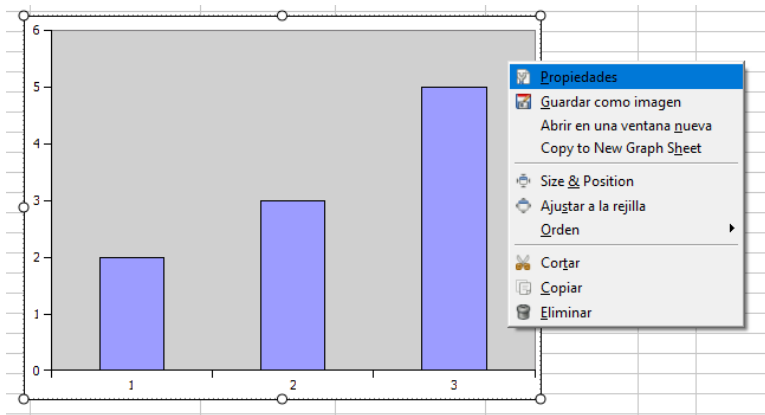
- Descriptive Statistics
 - Correlación...
 - Covarianza...
 - Estadística descriptiva...
 - Tablas de frecuencia
 - Tabla de frecuencias...
 - Histograma...
 - Rangos y percentiles...
- Muestreo...
- Dependent Observations
- One Sample Tests
- Two Sample Tests
- Multiple Sample Tests

	Columna 2	E	F	G	Ex
		Residual	Leverages	Internally studentized	
172,53807106599	385	-87,5380710659898	0,10406091370563	-1,8413	
18,121827411167	400	-8,12182741116749	0,20558375634495	-0,1814	
24,225888324873	395	-29,2258883248731	0,13451776649726	-0,6255	
18,121827411167	365	-43,1218274111675	0,20558375634495	-0,9633	
10,329949238579	475	34,6700507614214	0,09390862944147	0,7252	
14,746192893401	440	-64,746192893401	0,23604060913742	-1,4749	
172,53807106599	490	17,4619289340102	0,10406091370563	0,3673	
18,121827411167	420	11,8781725888325	0,20558375634495	0,2653	
14,746192893401	560	55,253807106599	0,23604060913742	1,2586	
172,53807106599	525	52,4619289340102	0,10406091370563	1,1035	
24,225888324873	480	55,7741116751269	0,13451776649726	1,1937	
14,746192893401	510	5,25380710659897	0,23604060913742	0,1197	

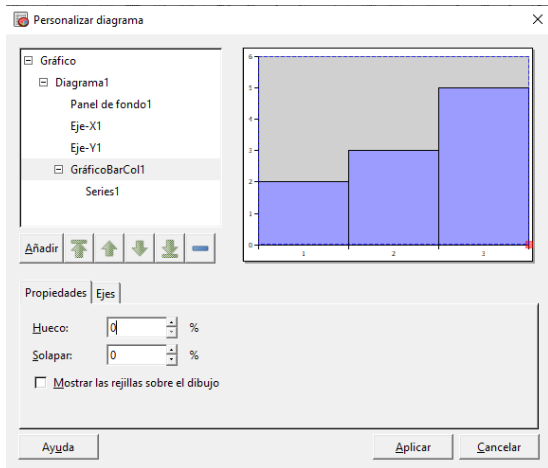
Análisis de los Residuos



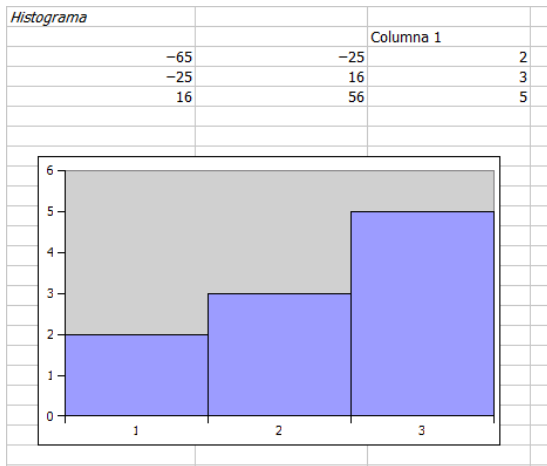
Análisis de los Residuos



Análisis de los Residuos

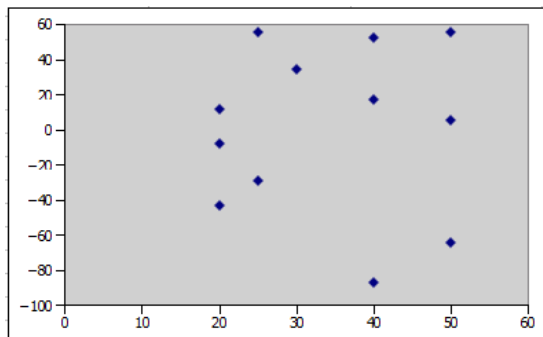


Análisis de los Residuos



Análisis de los Residuos

Gráfico de dispersión: r_i versus x_i



Predicción/Pronóstico

- a) Cuál es el valor esperado de ventas cuando se invierte 45000\$ en publicidad?

Predicción/Pronóstico

- a) Cuál es el valor esperado de ventas cuando se invierte 45000\$ en publicidad?

$$343.71 + 3.22 * 45 = 488.61$$

Predicción/Pronóstico

- a)Cuál es el valor esperado de ventas cuando se invierte 45000\$ en publicidad?

$$343.71 + 3.22 * 45 = 488.61$$

Con intervalo de predicción: (458.30, 518.98)

- b) Pronosticar las ventas para una inversión de 75000\$.

Predicción/Pronóstico

- a)Cuál es el valor esperado de ventas cuando se invierte 45000\$ en publicidad?

$$343.71 + 3.22 * 45 = 488.61$$

Con intervalo de predicción: (458.30, 518.98)

- b) Pronosticar las ventas para una inversión de 75000\$.

$$343.71 + 3.22 * 75 = 585.21 \quad \text{Ojo con este valor!}$$

Con intervalo de pronóstico: (372.69, 604.59)

Que podemos decir de las longitudes de los intervalos?

Predicción/Pronóstico

Intervalo para la Predicción	Intervalo para el Pronóstico
$\left(\hat{Y}_h \pm t_{n-2; 1-\frac{\alpha}{2}} \hat{\sigma}_{pred} \right)$ $\hat{\sigma}_{pred} = \hat{S}_{y/x} \sqrt{\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$	$\left(\hat{Y}_h \pm t_{n-2; 1-\frac{\alpha}{2}} \hat{\sigma}_{pron} \right)$ $\hat{\sigma}_{pron} = \hat{S}_{y/x} \sqrt{1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_i (X_i - \bar{X})^2}}$

The background of the slide is composed of two large, overlapping geometric shapes. A teal-colored shape occupies the top-left corner, while a light beige shape occupies the bottom-left corner. The rest of the slide is white. The text is centered in the white area.

Otro ejercicio para resolver...

Puestos vacantes

Con la finalidad de cubrir puestos vacantes la empresa *Cognisoft* de *Dacta* decide evaluar en una primera instancia a los 113 postulantes a través de un examen nivelatorio, brindándoles la bibliografía necesaria para que todos tengan los mismos conocimientos mínimos. Luego, se los evaluará mediante un examen final y serán candidatos a tomar los puestos aquellos que hayan obtenido 70 puntos o más en el examen. Los datos se encuentran en Estadilandia.

- a) Elabore un diagrama que permita ver si hay algún tipo de relación entre las variables y haga una conclusión exploratoria.
- b) Estime el modelo que crea conveniente.
- c) Es válido dicho modelo?
- d) Cuantifique el ajuste del modelo y decida si es bueno.
- e)Cuál es la nota del examen final esperada si se obtuvieron 40 puntos en el examen de nivelación?