



GUÍA LABORATORIO ESTADÍSTICA DESCRIPTIVA

Elaborado por Ing. Wilson Castro Zapata

El laboratorio de estadística Descriptiva incluye los recursos:

- Ejercicio 35. Reporte laboratorio de estadística Descriptiva. Página en www.seactuario.com que puede acceder libremente en sin entrar a sesión en:
<https://www.seactuario.com/ContMatematicas/ProbEstadistica/CuestEstadDescriptiva.htm>
O ingresando sesión para los estudiantes inscritos, con lo cual podrá Grabar la calificación obtenida.
- Plantilla de Excel. En el mismo Ejercicio 35 lo puede descargar.
- Cuadernos de Python en archivo .ipynb sin comprimir y también comprimido. Los puede descargar del Ejercicio 35.
- Esta guía.

Videos del docente relacionados:

1. Desarrollo de la parte de Estadística descriptiva datos sin agrupar:

<https://www.youtube.com/watch?v=Nyf4-UuF7fY&t=923s>

2. Datos Agrupados:

<https://www.youtube.com/watch?v=9jJ8gPfYXqA>

<https://www.youtube.com/watch?v=gkQWHTyVl6A>

Se ha realizado el laboratorio para que el estudiante pueda llevar a cabo los cálculos para todos los estadísticos tanto mediante la definición como con la función directa de Excel y de Python o de R-RStudio con diferentes librerías.

Al abrir el Ejercicio 35, descargue la Plantilla de Excel para llevar aquí los procedimientos, así como el archivo de Python. Luego genere los datos del Ejercicio presionando el botón 'Generar Datos M.L.', que generará 50 o 52 datos de acuerdo con un modelo de regresión lineal distribuido normalmente:

Cantidad de puntos generados: 0

Datos x:

x

0

Generar Datos M.L.

Reporte de los resultados del Laboratorio sobre Estadística Descriptiva

Con los datos generados, calcule los diferentes términos que se indican, redondeando a tres decimales. Tiene hasta dos opciones para evaluarse luego de lo deberá reiniciar el cuestionario con otros datos.

$$\sum_{i=1}^n x_i^2 =$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 =$$

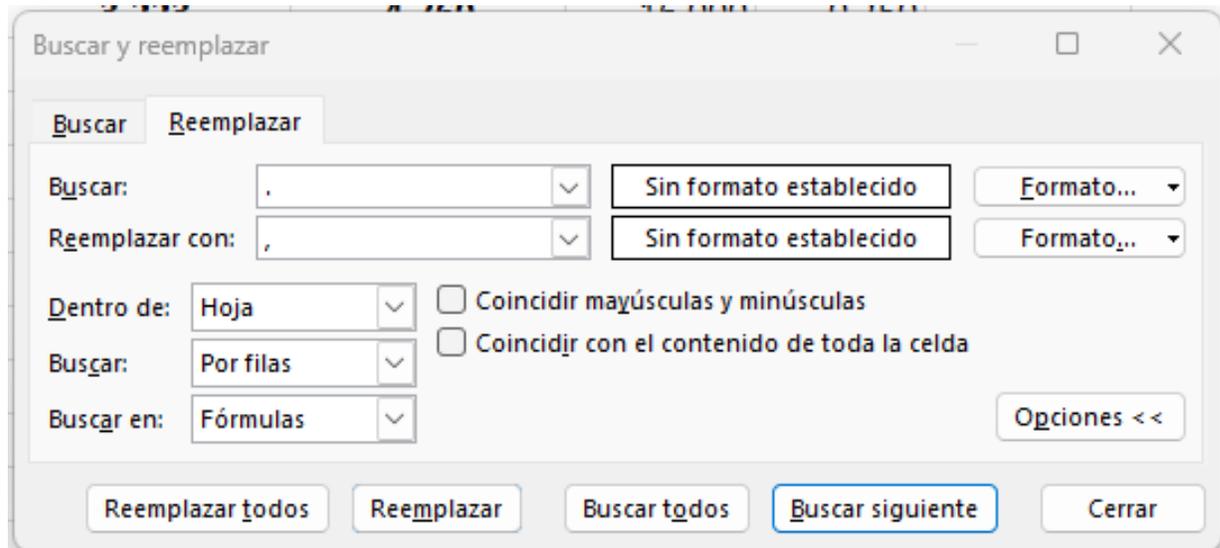
Luego seleccione sus datos y cópieles en su plantilla de Excel en la hoja Estad.Descriptiva reemplazando los datos debajo de la variable X

Datos x:	
x	
11.8	
9.7	
19.9	
16.7	
13.1	
21.4	
13.5	
17.1	
16.5	
15.9	
8.7	
17.4	
10.3	
18.1	
23.5	

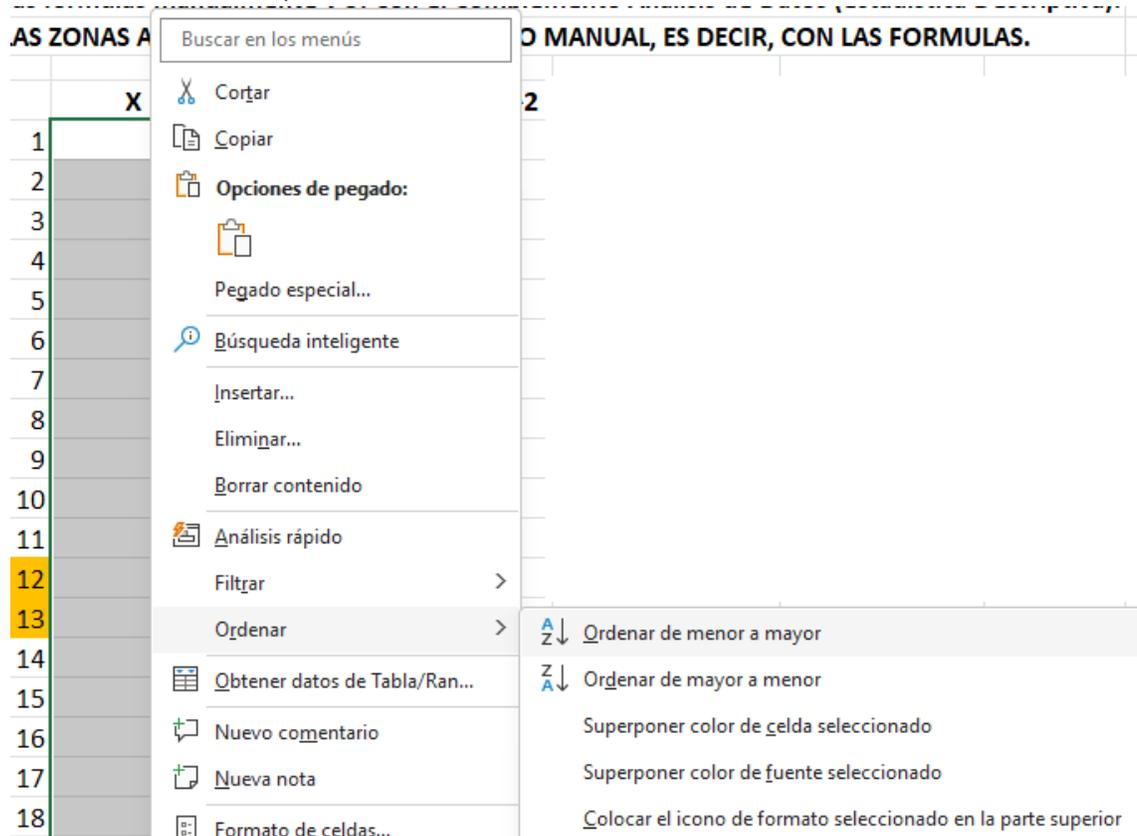
Copie sus datos y realice toda la Estadística Descriptiva		
2: Con las fórmulas manualmente y 3: Con Excel		
LAS ZONAS AZULES DEBEN TENER		
No.	X	Xi - Xm
1	4	-2,7
2	4	-2,7
3	4,1	-2,6
4	4,3	-2,4
5	4,3	-2,4
6	4,4	-2,3
7	4,4	-2,3
8	4,7	-2,0
9	4,8	-1,9
10	4,8	-1,9

A continuación, verifique si su instalación de Excel está configurado con punto o con coma para el punto decimal. Si está con coma, al copiar va a aparecer el mensaje de Error en las celdas: #!VALOR!, caso en el cual debe cambiar los puntos por comas.

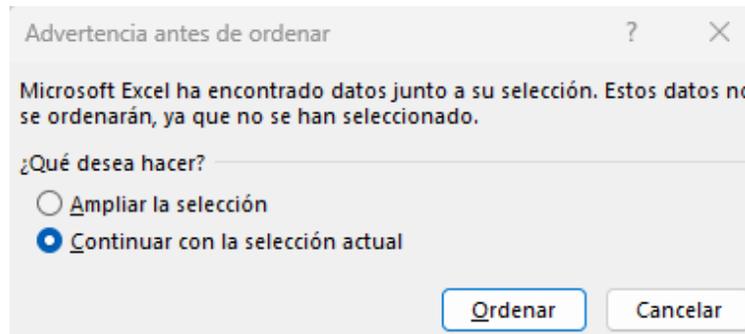
- Cambiar el punto decimal de punto a coma. Dé click en la pestaña **Inicio** y de allí a **Buscar y seleccionar** y cambie como se ilustra a continuación (Cambiar punto, Reemplazar con coma) presionando luego el botón **Reemplazar todos**.



- Ordene los datos de Menor a mayor. Teniendo los datos seleccionados, dé clic derecho y presione Ordenar y luego Ordenar de menor a mayor.



Si le aparece el cuadro de diálogo siguiente, indique que **Continuar con la selección actual**



A continuación complete las columnas que se requieren:

Copie sus datos y realice toda la Estadística Descriptiva de las tres formas: 1: Con las funciones de Excel,

2: Con las fórmulas manualmente y 3: Con el Complemento Análisis de Datos (estadística Descriptiva):

LAS ZONAS AZULES DEBEN TENER EL CÁLCULO MANUAL, ES DECIR, CON LAS FORMULAS.

No.	X	$X_i - X_m$	$(X_i - X_m)^2$	$[(X_i - X_m)/s]^3$	$[(X_i - X_m)/s]^4$	X_i^2	$1/X_i$
1	4	-2,7	7,1	-3,223	4,760	16,000	0,250
2	4	-2,7	7,1	-3,223	4,760	16,000	0,250
3	4,1	-2,6	6,6	-2,874	4,087	16,810	0,244
4	4,3	-2,4	5,6	-2,255	2,956	18,490	0,233
5	4,3	-2,4	5,6	-2,255	2,956	18,490	0,233
6	4,4	-2,3	5,2	-1,981	2,488	19,360	0,227
7	4,4	-2,3	5,2	-1,981	2,488	19,360	0,227
8	4,7	-2,0	3,9	-1,296	1,412	22,090	0,213
9	4,8	-1,9	3,5	-1,108	1,147	23,040	0,208
10	4,8	-1,9	3,5	-1,108	1,147	23,040	0,208
11	4,9	-1,8	3,1	-0,940	0,921	24,010	0,204

En que es necesario primero calcular la varianza s^2 y la desviación estándar s para poder completar las correspondientes de las desviaciones al cubo y a la cuarta potencia.

Como la Celda B60 se ha renombrado como n para facilitar los cálculos, se puede observar que puede ingresar las expresiones así como las funciones de Excel como se indica a continuación (Suponga que tiene 50 datos y se hace referencia al vector de datos como debe aparecer en la plantilla de Excel):

Estadístico	Definición	Formula en Excel	Función Excel
Media	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$	=SUMA()/CONTAR() = B58/B60	=PROMEDIO(B6:B57)
Media Geométrica	$G = \sqrt[n]{\prod_{i=1}^n x_i}$	=PRODUCTO(B6:B57)^(1/n)	=MEDIA.GEOM(B6:B57)
Media Armónica:	$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$	= n/H58	=MEDIA.ARMO(B6:B57)
Varianza s^2 :	$\frac{\sum (x_i - \bar{x})^2}{n - 1}$	= D58/(n-1)	=VAR.S(B6:B57)

Desv. Estándar s:	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$	= RAIZ(C63)	=DESVEST.M(B6:B57)
Coef. Variación C.V. :	$C.V. = \frac{s}{\bar{x}} * 100$	=100*C64/B61	N.A.

Para el **Coefficiente de Asimetría**, se utiliza la fórmula implementada en Excel para poder verificar el resultado:

Coeficiente Asimetría =	$\frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$
-------------------------	--

De tal manera que las expresiones para el cálculo manual y con la función de Excel, son respectivamente:

$$=n/((n-1)*(n-2))*E58$$

$$=COEFICIENTE.ASIMETRIA(B6:B57)$$

Para la curtosis, también se utiliza la fórmula implementada en Excel para poder verificar el resultado:

Curtosis =	$\left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$
------------	---

Como la fórmula es extensa, ya se encuentra implementada en la plantilla de Excel:

$$= (n*(n+1)/((n-1)*(n-2)*(n-3)))*F58-3*(n-1)^2/((n-2)*(n-3))$$

$$=CURTOSIS(B6:B57)$$

CUARTILES

Para los cuartiles, Excel utiliza las funciones CUARTIL.EXE() o Exclusivo y CUARTIL.INC o inclusivo.

Se realiza el cálculo de cada cuartil de forma manual por medio de la interpolación dado el dato correspondiente, así:

Cuartil 1: Q1 corresponde al dato $X_{\frac{n+1}{4}}$, de tal manera que si n=50 el dato corresponde a $X_{12.75}$

La interpolación en tal caso es: $Q_1 = X_{12.75} = X_{12} + 0.75*(X_{13}-X_{12})$

En este caso se hace la interpolación entre los datos 12 y 13, que en la plantilla corresponde a:

$$Q_1 = X_{12.75} = B17 + 0,75 * (B18 - B17)$$

Cuartil 2 o mediana: Q2 o Me corresponde al dato $X_{\frac{n+1}{2}}$, de tal manera que si n=50 el dato corresponde a $X_{25.5}$

La interpolación en tal caso es: $Q_2 = X_{25.5} = X_{25} + 0.5*(X_{26}-X_{25})$

En este caso se hace la interpolación entre los datos 25 y 26 correspondiente a 0.5 o la mitad entre uno y otro, es equivalente a realizar el promedio entre estos datos:

$$Q_2 = X_{25.5} = \text{PROMEDIO}(B30:B31)$$

Cuartil 3: Q3 corresponde al dato $X_{3(\frac{n+1}{4})}$, de tal manera que si $n=50$ el dato corresponde a $X_{38.25}$

La interpolación en tal caso es: $Q_3 = X_{38.25} = X_{38} + 0.25*(X_{39}-X_{38})$

En este caso se hace la interpolación entre los datos 38 y 39, que en la plantilla corresponde a:

$$Q_3 = X_{38.25} = B43 + 0,25 * (B44 - B43)$$

Para encontrar los Bigotes del Diagrama de Box and whisker (Caja y bigotes) o Boxplot, se requiere el rango intercuartílico R.I.Q. que se define como:

$$\text{R. I. Q.} = Q_3 - Q_1$$

Y de esta forma las longitudes teóricas máximas de los bigotes (es el caso de la igualdad), son:

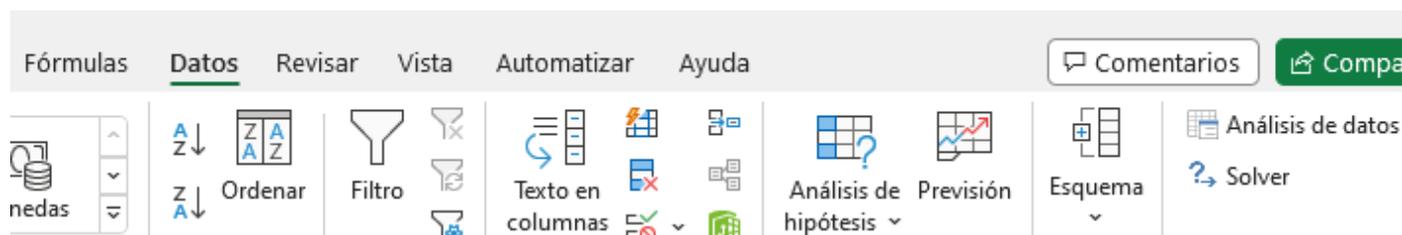
$$L_{\text{inf}} \geq Q_1 - 1.5 * \text{R. I. Q}$$

$$L_{\text{sup}} \leq Q_3 + 1.5 * \text{R. I. Q}$$

Una vez se tenga la parte de la hoja de Estadística descriptiva, se ingresan todos los valores en el cuestionario Ejercicio 35 y se puede evaluar y Grabar la nota.

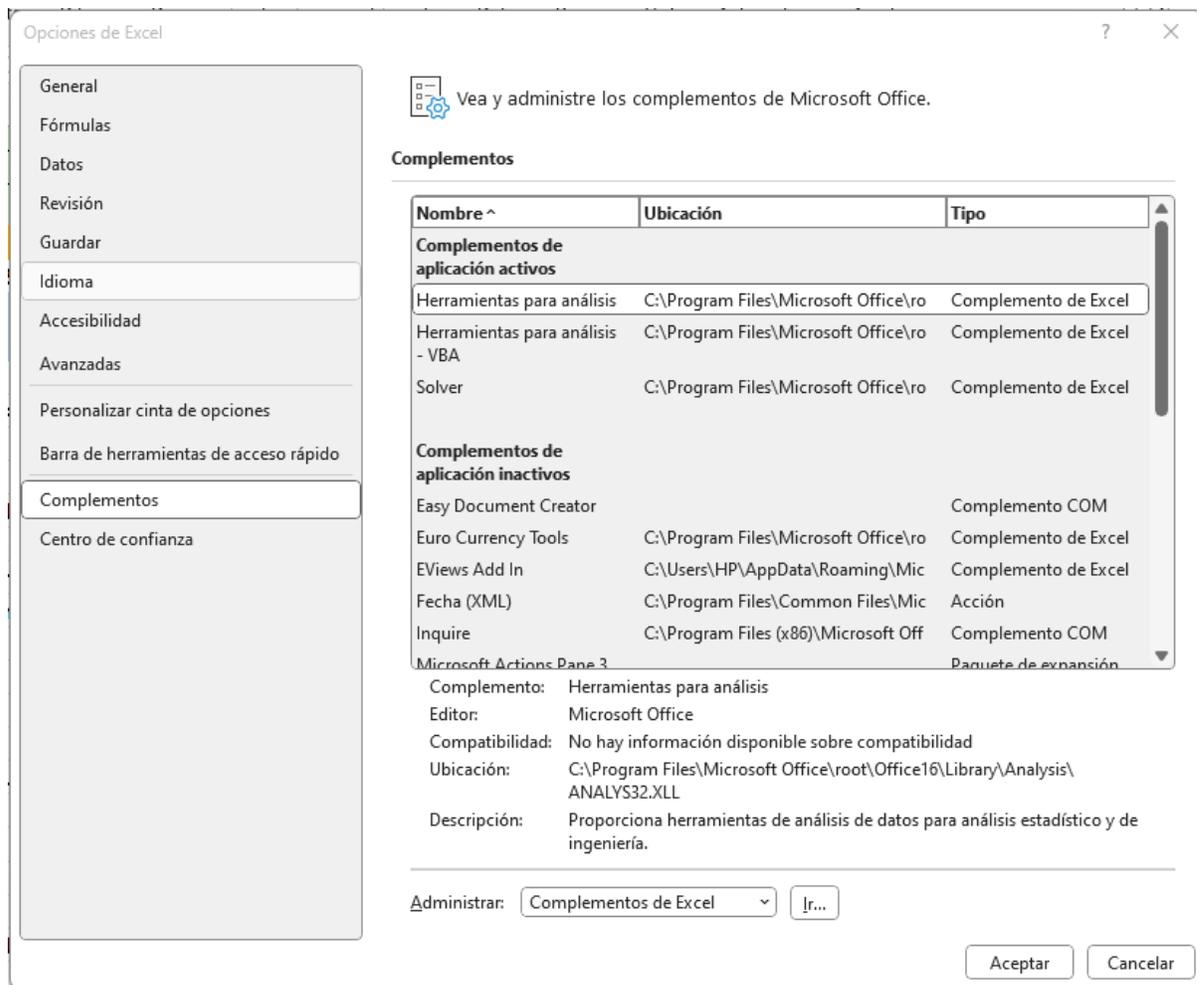
Complemento Análisis de datos Estadística descriptiva

Verifique si el complemento está habilitado haciendo click en la pestaña **Datos** y debe aparecer a la derecha superior como **Análisis de datos**.



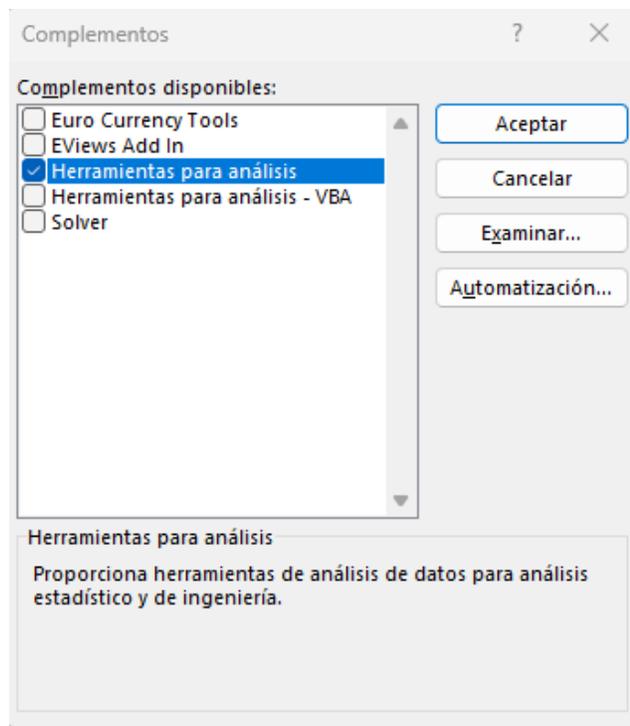
Si no aparece se debe habilitar así:

- haga clic en la pestaña **Archivo** y luego **Opciones**, se despliega el cuadro de diálogo siguiente:

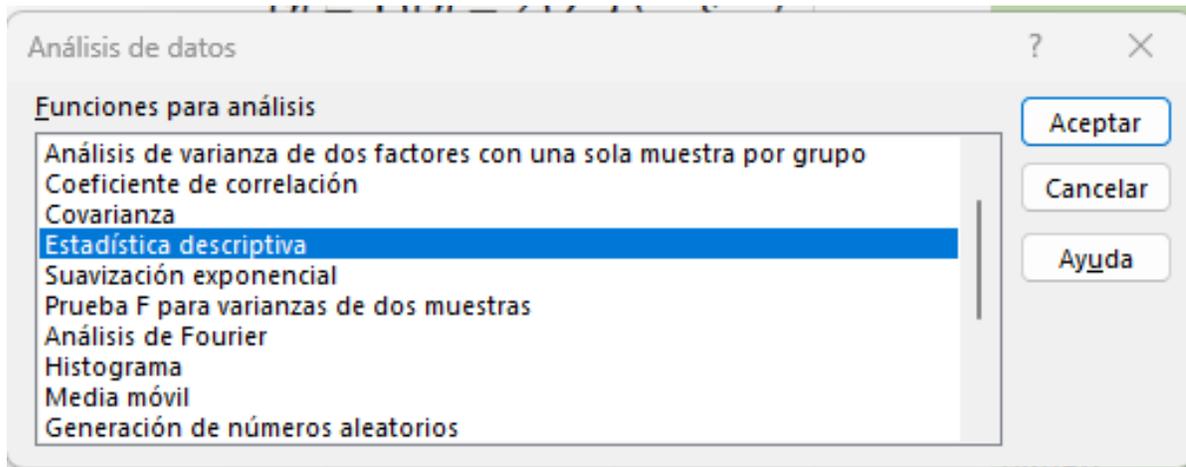


Presione el botón de la parte inferior que dice Ir...

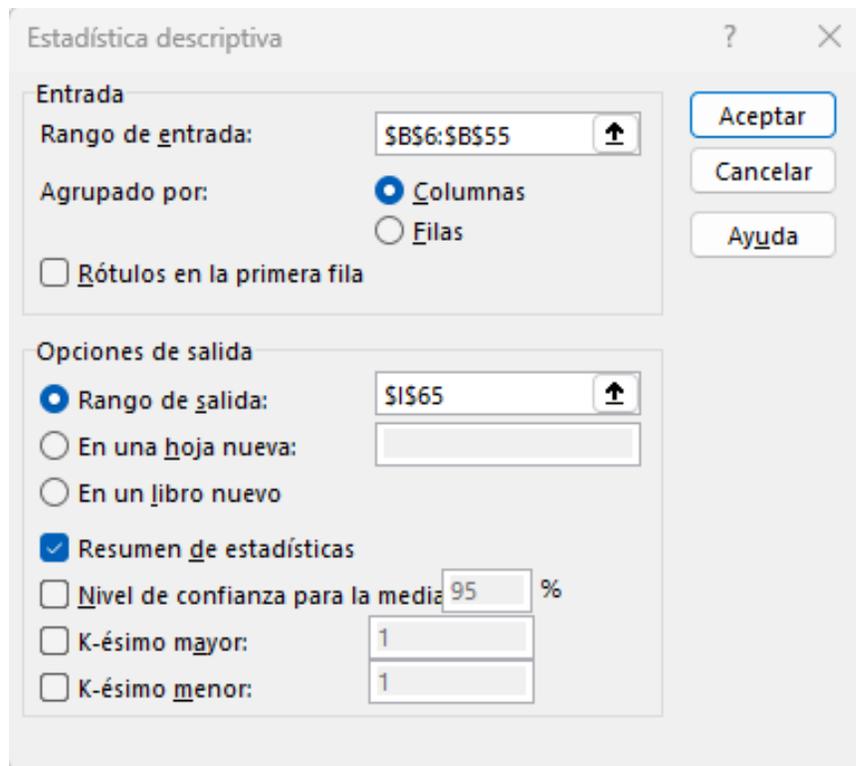
Y en la ventana de diálogo que de muestra, seleccione Herramientas para análisis y luego Aceptar:



Ahora que está habilitado, selecciona la herramienta en Datos(Análisis de datos y seleccione Estadística descriptiva:



El **Rango de entrada** es la matriz o vector de sus datos, **Rango de salida** es la celda donde desea que se despliegue el informe, y seleccione **Resumen de estadísticas** como se ilustra:



De esta forma se obtienen resultados para comparar con los ya obtenidos como:

3. COMPLEMENTO ANÁLISIS DE DATOS (ESTADÍSTICA)
COLOQUE EN ESTA ZONA LOS RESULTADOS

<i>Columna1</i>	
Media	6,672
Error típico	0,2558226
Mediana	6,15
Moda	5,7
Desviación estándar	1,8089393
Varianza de la muestra	3,2722612
Curtosis	-1,120523
Coefficiente de asimetría	0,282554
Rango	5,9
Mínimo	4
Máximo	9,9
Suma	333,6
Cuenta	50

HOJA DE DATOS AGRUPADOS

Al continuar con la Hoja de Datos Agrupados, se deben reemplazar los datos por los propios obtenidos. Se debe especificar la cantidad total de datos en la celda C15 con lo cual se obtienen las clases y el ancho de clase. Ahora solo es reemplazar el dato del limite inferior de la primera clase con el dato mínimo o un poco más pequeño y verificar que la última clase cubra a la totalidad de los datos, desde el mínimo al máximo.

Forma Intervalo:	
(]	
CLASES	
4	4 4,9
4	4,9 5,8
1	5,8 6,7
3	6,7 7,6
3	7,6 8,5
4	8,5 9,4
4	9,4 10,3
7	Totales:
8	
8	
9	n= 50
1	K= 6,61

Para encontrar las frecuencias absolutas simples (n), se puede usar la función FRECUENCIA(), como se observa a continuación:

DATOS AGRUPADOS: Copie los datos y ajuste la tabla de acuerdo a las

Datos	Forma Intervalo:		Marcas de clase	Frecuencias Abs	
	()		Simple	Acum
x	CLASES		x_i	ni	
4	4	4,9	=FRECUENCIA(A5:A54;C5:C10)		
4	4,9	5,8	FRECUENCIA(datos; grupos)	10,00	
4,1	5,8	6,7	6,25	6,00	
4,3	6,7	7,6	7,15	5,00	
4,3	7,6	8,5	8,05	7,00	
4,4	8,5	9,4	8,95	7,00	
4,4	9,4	10,3	9,85	4	
4,7	Totales:			50,00	

Es decir, los parámetros son:

- **datos:** Se selecciona toda la matriz de los datos originales (aquí de A5:A54) para 50 datos
- **grupos:** Se selecciona la matriz de los límites superiores de cada clase hasta la penúltima clase (aquí C5:C10)

En versiones más modernas de Excel, al hacer **Enter** se despliegan todas las frecuencias absolutas.

En versiones anteriores de Excel es necesario una vez se han digitado los parámetros, **seleccionar todo el rango** (diferente a hacer el copiado automático) donde se desea que salgan las frecuencias y haciendo clic en la barra de fórmulas mantener presionados los botones **Ctrl + Shift** y presionar **Enter**.

Una vez se haya realizado el procedimiento anterior, se completa la tabla de Frecuencias para Datos agrupados en clases.

Estadísticos en Datos agrupados

Para las expresiones y buscando claridad en las celdas que se seleccionan, se utiliza la plantilla como se ilustra a continuación:

	A	B	C	D	E
1	DATOS AGRUPADOS: Copie los datos y ajuste la tabla d				
2	Datos	Forma Intervalo:			Frecuencias
3		(]		Marcas de clase	Simple
4	x	CLASES		x_i	n_i
5	4	4	4,9	4,45	11,00
6	4	4,9	5,8	5,35	10,00
7	4,1	5,8	6,7	6,25	6,00
8	4,3	6,7	7,6	7,15	5,00
9	4,3	7,6	8,5	8,05	7,00
10	4,4	8,5	9,4	8,95	7,00
11	4,4	9,4	10,3	9,85	4
12	4,7	Totales:			50,00
13	4,8				
14	4,8				www.seacti
15	4,9	n=	50	F. STURGES	

Se utilizan en este caso las marcas de clase (x_i) y las frecuencias absolutas simples (n_i)

- Media: Es necesario realizar la función SUMAPRODUCTO() con la primera matriz con las x_i y la segunda matriz con las n_i :
= SUMAPRODUCTO(D5:D11;E5:E11)/C15

- Media Geométrica: Es necesario realizar el producto entre cada las x_i a la potencia n_i y luego la raíz n-ésima,

$$G = \sqrt[n]{\prod_{i=1}^n x_i^{n_i}}$$

Que en caso de 50 datos se tiene:

$$= (D5^E5 * D6^E6 * D7^E7 * D8^E8 * D9^E9 * D10^E10 * D11^E11)^{(1/C15)}$$

- Media Armónica: Como es n sobre la suma de los recíprocos de los datos, pero se tiene la frecuencia en que se repite cada dato se aplica la fórmula:

$$H = \frac{n}{\sum_{i=1}^n \frac{n_i}{x_i}}$$

que da en este caso se realiza como:

$$= C15 / (E5/D5 + E6/D6 + E7/D7 + E8/D8 + E9/D9 + E10/D10 + E11/D11)$$

Mediana y cuartiles en datos agrupados

Para la mediana y los cuartiles, se realiza el procedimiento de identificar la clase en la cual se encuentra la medida de posición y utilizando las expresiones correspondientes:

Cuartiles	Deciles	Percentiles
$Q_k = L_i + \left(\frac{\frac{kN}{4} - N_{i-1}}{n_i} \right) \cdot A$	$D_k = L_i + \left(\frac{\frac{kN}{10} - N_{i-1}}{n_i} \right) \cdot A$	$P_k = L_i + \left(\frac{\frac{kN}{100} - N_{i-1}}{n_i} \right) \cdot A$

Medidas de Variabilidad, de forma y concentración

Para encontrar la varianza, desviación estándar, coeficiente de asimetría y curtosis, se debe completar la tabla de desviaciones con un resultado para el ejemplo dado de:

	K	L	M	N
1	Varianza ()	Tome las marcas de clase Xi		
2	DESVIACIONES			
3	Desv.	Desv.2	Desv.3	Desv.4
4	(xi - xm)	ni*(xi - xm)^2	ni*[(xi - xm)/s]^3	ni*[(xi - xm)/s]^4
5	-2,232	54,800	-19,974	24,368
6	-1,332	17,742	-3,859	2,810
7	-0,432	1,120	-0,079	0,019
8	0,468	1,095	0,084	0,021
9	1,368	13,100	2,927	2,188
10	2,268	36,007	13,336	16,532
11	3,168	40,145	20,769	35,963
12	3,276	164,009	13,202	81,902

En que se multiplican las desviaciones por la frecuencia absoluta simple antes de realizar la potencia. Luego se utilizan las sumas de estas desviaciones y se deben utilizar las fórmulas correspondientes para calcular los estadísticos.