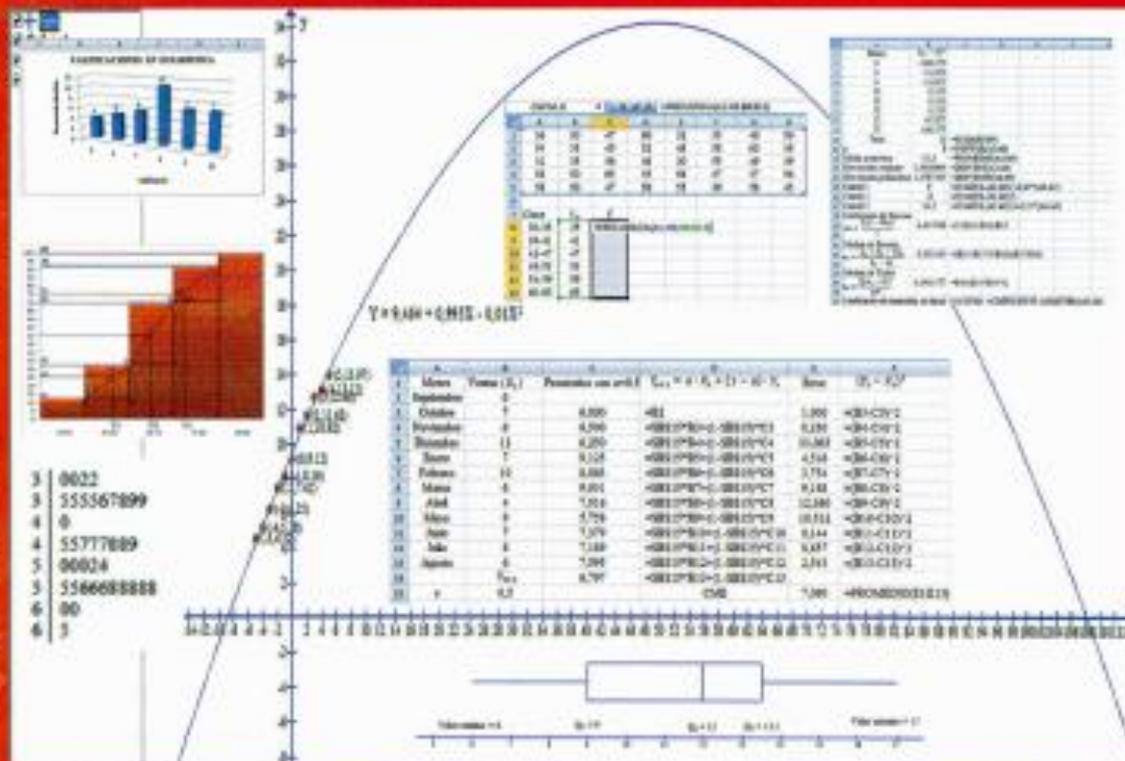




UNIVERSIDAD TÉCNICA DEL NORTE
Facultad de Ciencias Administrativas y Económicas

INTERAPRENDIZAJE DE ESTADÍSTICA BÁSICA



AUTORES:

Mgs. Mario O. Suárez I.

Msc. Fausto A. Tapia Z.

www.utn.edu.ec

IBARRA - ECUADOR

2012

MARIO ORLANDO SUÁREZ IBUJES



Nació el 24 de marzo de 1978 en la ciudad de Ibarra, Imbabura-Ecuador. Sus primeros estudios los realizó en la Escuela Fiscal Mixta "Alejandro Pasquel Monge", del Barrio "La Florida" de la ciudad de Ibarra, en la cual fue Abanderado del Estandarte de la Escuela, Abanderado del Pabellón Nacional y Mejor Egresado.

Sus estudios secundarios los realizó en el centenario Colegio "Teodoro Gómez de la Torre" de la ciudad de Ibarra, en el cual fue el Mejor Alumno en Matemática durante los tres últimos años, Abanderado del Estandarte del Colegio y Mejor Egresado.

Sus estudios de tercer nivel los realizó en la Universidad Técnica del Norte (UTN), en la cual siendo el Mejor Egresado, obtiene el título de Licenciado en Física y Matemática. En la ciudad de Quito se hace acreedor a una Mención Especial en Matemática, Premio Nacional otorgado por la VI Feria de Ciencia y Tecnología por haber triunfado con su proyecto Multiprisma (Un rompecabezas tridimensional bicolor integrado por partes prismáticas).

Sus estudios de cuarto nivel los realizó en la UTN en convenio con la Asociación de Facultades Ecuatorianas de Filosofía y Ciencias de la Educación (AFEFC), en la cual obtiene el título de Magister en Gerencia de Proyectos Educativos y Sociales a los 26 años de edad.

Sus publicaciones: Unidades para Producir Medios Instruccionales en Educación (coautor a los 24 años), Interaprendizaje Holístico de Matemática (autor), Hacia un Interaprendizaje Holístico de Álgebra y Geometría (autor), Matemática Recreativa (coautor), Interaprendizaje de Probabilidades y Estadística Inferencial con Excel, Winstats y Graph (autor) e Interaprendizaje de Estadística Básica (coautor). Sus tesis de licenciatura y maestría, así como más de 60 artículos sobre Aritmética, Álgebra, Geometría, Probabilidades, Estadística Descriptiva y Estadística Inferencial se encuentran publicados en www.monografias.com, www.scribd.com y en el repositorio digital del uniportal www.utn.edu.ec

Su experiencia profesional: Maestro de Matemática en la Escuela "Alejandro Pasquel Monge" (1998-2001), Docente e Inspector de Curso del Colegio Nacional "Ibarra" (2001-2003), Docente del Colegio Universitario "UTN" (2003-2004), Docente de la Academia Militar "San Diego" (2004-2008), Docente a los 20 años de edad en la Unidad Educativa Experimental "Teodoro Gómez de la Torre" (1998-2011). Actual Docente del Colegio Nacional Técnico "Mariano Suárez Veintimilla" y de la Universidad Técnica del Norte en la Facultad de Ciencias Administrativas y Económicas.

FAUSTO AMILCAR TAPIA ZAMBRANO



Nació el 18 de mayo de 1954 en la ciudad de Tulcán, Carchi-Ecuador. Sus primeros estudios los realizó en las Escuelas Rurales "Héctor Lara Zambrano", y "Belisario Quevedo" de Julio Andrade.

Sus estudios secundarios del ciclo básico los realizó en el Colegio Experimental "Bolívar" de Tulcán, y ciclo diversificado en el Normal Experimental "Juan Montalvo" de la ciudad de Quito.

Sus estudios de tercer nivel los realizó en la Universidad Central del Ecuador obteniendo el título de Licenciado en Psicología Educativa y Orientación Vocacional.

Sus estudios de cuarto nivel los realizó en la Universidad Central del Ecuador, obteniendo el título de Doctor en Ciencias de la Educación, y en la Universidad Técnica del Norte obteniendo el título de Magister en Docencia Universitaria e Investigación.

Sus publicaciones: Las Dificultades Específicas del Aprendizaje: Prevención, Diagnóstico y Tratamiento (autor) e Interaprendizaje de Estadística Básica (coautor).

Su experiencia profesional: Inspector General del Colegio Nacional "Pimampiro", Inspector General del Colegio Nacional Técnico "Mariano Suárez Veintimilla", Docente del Instituto de Educación Especial de Ibarra, actual Jefe del Departamento de Educación Especial de la Dirección Provincial de Educación de Imbabura y Catedrático Fundador de la Universidad Técnica del Norte en la Facultad de Ciencias Administrativas y Económicas.

Dignidades: Subdecano de la Facultad de Ciencias Administrativas y Económicas (período 2000-2003).

ISBN 978-9942-11-239-2

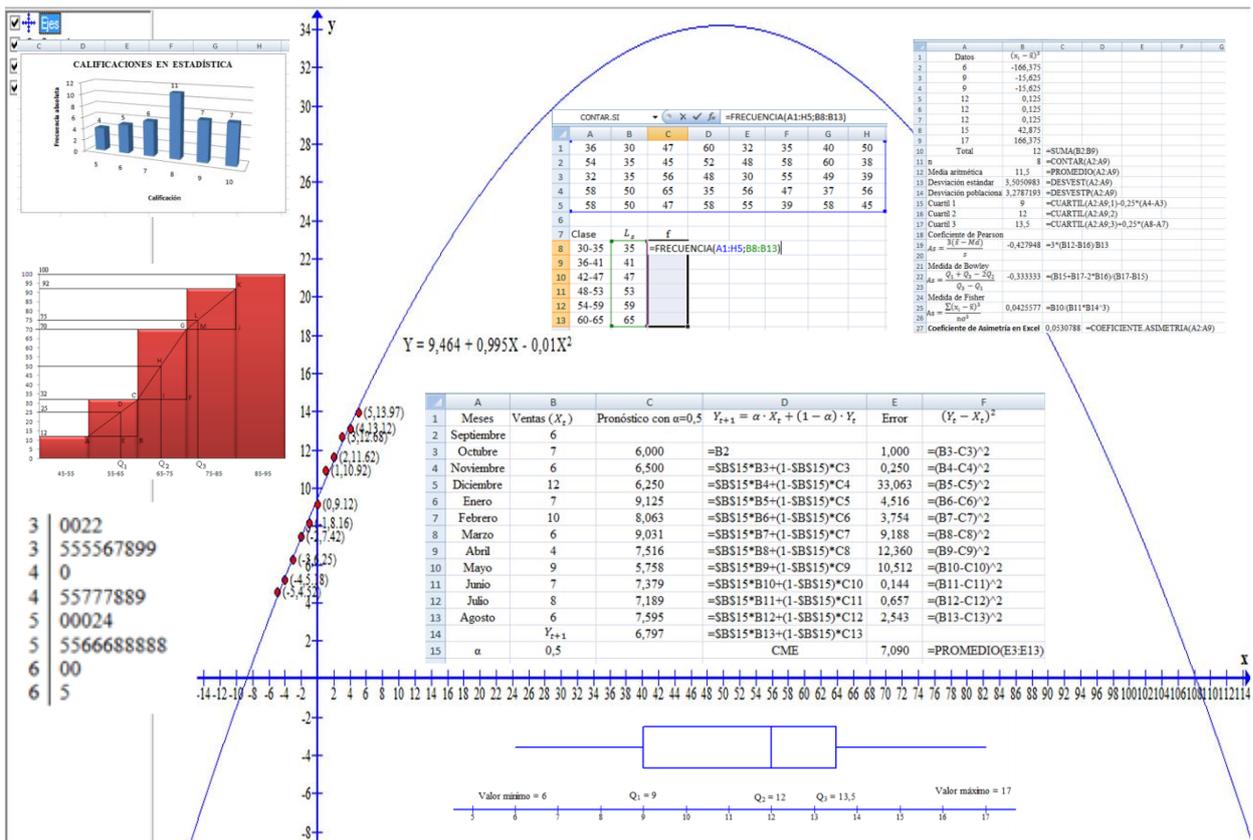


9 789942 112392



UNIVERSIDAD TÉCNICA DEL NORTE
FACULTAD DE CIENCIAS ADMINISTRATIVAS Y ECONÓMICAS

INTERAPRENDIZAJE DE ESTADÍSTICA BÁSICA



AUTORES:

Mgs. Mario O. Suárez I.
Msc. Fausto A. Tapia Z.

IBARRA - ECUADOR
2012

AUTORIDADES DE LA UNIVERSIDAD TÉCNICA DEL NORTE

Dr. Antonio Posso
RECTOR

Dr. Miguel Naranjo
VICERRECTOR ACADÉMICO

Dr. Marco Cerda
VICERRECTOR ADMINISTRATIVO

Ing. Msc. Ney Mora
DECANO FACAE

Msc. Soraya Rhea
SUBDECANA FACAE

DERECHOS RESERVADOS DEL AUTOR:

Instituto Ecuatoriano de la Propiedad Intelectual
Dirección Nacional de Derecho de Autor y Derechos Conexos
Derecho de Autor N° 038383
ISBN: 978-9942-11-239-2

Coordinación de publicación:
Universidad Técnica del Norte

Primera Edición

Esta obra no puede ser reproducida total ni parcialmente por ningún medio sin expreso consentimiento previo y por escrito de los autores.

Pedidos a: Mario: 062632166; 0985619601 mosuarez@utn.edu.ec; mgsmariosuarez@gmail.com
Fausto: 093754382, fatapia@utn.edu.ec; tapiafausto@gmail.com

DEDICATORIA

Mario

Con infinito amor en expansión
a mi esposa Dyanita Rivera y a mi hijo Mathías Suárez
por ser mi fuente de inspiración y mi más anhelado sueño hecho realidad.

Fausto

Con especial afecto a mi familia
por la paciencia y apoyo brindado para la realización de este trabajo.

AGRADECIMIENTO

Nuestra gratitud y reconocimiento a las Autoridades de la Universidad Técnica del Norte y de la Facultad de Ciencias Administrativas y Económicas por el incondicional apoyo brindado para la realización de la presente obra.

CONTENIDOS

	Pág.
CONTRAPORTADA	1
DEDICATORIA	3
AGRADECIMIENTO	4
CONTENIDOS	5
PRESENTACIÓN	9
EVALUACIÓN DIAGNÓSTICA	10
CAPÍTULO I	
DESCRIPCIÓN Y PRESENTACIÓN DE LA INFORMACIÓN ESTADÍSTICA	11
1.1.- ¿Qué es la Estadística?	12
1.1.1.- Historia	
1.1.2.- Definición	
1.1.3.- Aplicaciones	13
1.1.4.- Fines	
1.1.5.- Objetivos	
1.1.6.- Métodos	14
1.1.7.- Clasificación de la Estadística	
1.1.7.1.- Estadística Descriptiva o Deductiva	
1.1.7.2.- Estadística Inferencial o Inductiva	
1.2.- Conceptos y Definiciones Básicas	
1.2.1.- Población	
1.2.2.- Muestra	15
1.2.3.- Elemento o Individuo	16
1.2.4.- Datos Estadísticos	
1.2.5.- Censo	
1.2.6.- Encuesta	
1.2.7.- Variable	18
1.2.7.1.- Clasificación	
1.2.7.2.- Niveles de Medición	
1.3.- Tablas o Cuadros Estadísticos	21
1.4.- Distribución de Frecuencias	23
1.4.1.- Para datos sin Agrupar	
1.4.2.- Para datos Agrupados en Clases o Intervalos	28
1.5.- Gráficos Estadísticos Básicos	34
1.5.1.- Diagramas de Barras	
1.5.2.- Histogramas	42
1.5.3.- Polígono de Frecuencias	48
1.5.3.1.- Polígono de Frecuencias Acumuladas u Ojiva	54
1.5.3.2.- Polígono de Frecuencias Relativas Acumuladas Porcentuales	55
1.5.4.- Diagrama de Tallo y Hojas	56
1.5.5.- Diagrama de Sectores	57
1.5.6.- Pictogramas	60
CAPÍTULO II	
MEDIDAS DE TENDENCIA CENTRAL	65
2.1.- Media Aritmética	66
2.1.1.- Media Aritmética Simple	
2.1.1.1.- Definición	
2.1.1.2.- Métodos de Cálculo	
2.1.2.- Media Aritmética Ponderada	
2.2.- Media Geométrica	

2.2.1.- Propiedades	
2.2.2.- Métodos de Cálculo	
2.2.2.1.- Para Datos No Agrupados	
2.2.2.2.- Para Datos Agrupados en Tablas de Frecuencias	76
2.2.2.3.- Para Datos Agrupados en Intervalos	77
2.3.- Media Armónica	78
2.3.1.- Propiedades	
2.3.2.- Métodos de Cálculo	
2.3.2.1.- Para Datos No Agrupados	
2.3.2.2.- Para Datos Agrupados en Tablas de Frecuencias	80
2.3.2.3.- Para Datos Agrupados en Intervalos	81
2.4.- La Mediana	83
2.4.1.- Propiedades	
2.4.2.- Métodos de Cálculo	84
2.4.2.1.- Para Datos No Agrupados	
2.4.2.2.- Para Datos Agrupados en Tablas de Frecuencias	87
2.4.2.3.- Para Datos Agrupados en Intervalos	88
2.5.- Medidas de Posición	91
2.5.1.- Cuartiles	92
2.5.1.1.- Propiedades	
2.5.1.2.- Métodos de Cálculo	
2.5.2.- Deciles	108
2.5.2.1.- Definición	
2.5.2.2.- Métodos de Cálculo	
2.5.3.- Percentiles o Centiles	109
2.5.3.1.- Definición	
2.5.3.2.- Métodos de Cálculo	
2.6.- Moda	113
2.6.1.- Propiedades	
2.6.2.- Métodos de Cálculo	
2.6.2.1.- Para Datos No Agrupados	
2.6.2.2.- Para Datos Agrupados en Tablas de Frecuencias	116
2.6.2.3.- Para Datos Agrupados en Intervalos	
CAPÍTULO III	
MEDIDAS DE DISPERSIÓN	119
3.1.- Desviación Media o Desviación Promedio	120
3.1.1.- Propiedades	
3.1.2.- Métodos de Cálculo	
3.1.2.1.- Para Datos No Agrupados	
3.1.2.2.- Para Datos Agrupados en Tablas de Frecuencia	123
3.1.2.3.- Para Datos Agrupados en Intervalos	99
3.2.- Varianza y Desviación Estándar	126
3.2.1.- Propiedades	
3.2.2.- Métodos de Cálculo	127
3.2.2.1.- Para Datos No agrupados	
3.2.2.2.- Para Datos Agrupados en Tablas de Frecuencia	132
3.2.2.3.- Para Datos Agrupados en Intervalos	134
3.3.- Otras Medidas de Dispersión	137
3.3.1.- Rango, Amplitud Total o Recorrido	
3.3.2.- Amplitud Intercuartílica	138
3.3.3.- Rango Semi-intercuartil o Desviación Cuartílica	
3.3.4.- Rango Percentil o Amplitud Cuartílica	

3.4.-	Dispersión Relativa o Coeficiente de Variación	139
3.4.1.-	Propiedades	
3.4.2.-	Métodos de Cálculo	
CAPÍTULO IV		
MEDIDAS DE FORMA		
4.1.-	Asimetría	145
4.1.1.-	Tipos de Asimetría	146
4.1.1.1.-	Asimetría Negativa o a la Izquierda	
4.1.1.2.-	Simétrica	
4.1.1.3.-	Asimetría Positiva o a la Derecha	
4.1.2.-	Medidas de Asimetría	147
4.1.2.1.-	Coeficiente de Pearson	
4.1.2.2.-	Medida de Yule Bowley o Medida Cuartílica	
4.1.2.3.-	Medida de Fisher	
4.2.-	Curtosis o Apuntamiento	151
4.2.1.-	Tipos de Curtosis	
4.2.1.2.-	Leptocúrtica	
4.2.1.3.-	Mesocúrtica	
4.2.1.4.-	Platicúrtica	
4.2.2.-	Medidas de Curtosis	
4.2.2.1.-	Medida de Fisher	
4.2.2.2.-	Medida basada en Cuartiles y Percentiles	152
CAPÍTULO V		
CORRELACIÓN Y REGRESIÓN		
5.1.-	Análisis de Correlación	155
5.1.1.-	Diagrama de Dispersión	156
5.1.2.-	Clasificación de la Correlación	
5.1.2.1.-	Según la relación entre variables	
5.1.2.2.-	Según el número de variables	
5.1.2.3.-	Según el valor cuantitativo	
5.1.3.-	Coeficientes de Correlación	157
5.1.3.1.-	Coeficiente de Correlación de Karl Pearson	158
5.1.3.2.-	Coeficiente de Correlación por Rangos de Spearman	175
5.1.4.-	Coeficiente de Determinación	181
5.2.-	Análisis de Regresión	183
5.2.1.-	Principio de los Mínimos Cuadrados	184
5.2.1.1.-	La Recta de los Mínimos Cuadrados	
5.2.1.2.-	La Parábola de los Mínimos Cuadrados	194
5.2.1.3.-	Regresión Exponencial	200
5.2.1.4.-	Regresión Potencial	205
5.2.2.-	Error Estándar de Estimación	212
CAPÍTULO VI		
SERIES CRONOLÓGICAS		
6.1.-	Definición	219
6.2.-	Movimientos o Componentes	220
6.2.1.-	Tendencia Secular	
6.2.2.-	Movimientos Estacionales	225
6.2.3.-	Movimientos Cíclicos	226
6.2.4.-	Movimientos Irregulares o Aleatorios	
6.3.-	Modelos de Series de Tiempo	
6.3.1.-	Modelo Multiplicativo	227

6.3.2.- Modelo Aditivo	
6.4.- Métodos de Suavizamiento y Pronóstico	228
6.4.1.- Método de los Promedios Móviles	
6.4.2.- Suavización Exponencial	232
6.5.- Análisis de Tendencia	238
6.5.1.- Método de los Mínimos Cuadrados	
6.5.2.- Método de los Semipromedios	243
6.6.- Análisis de Movimientos Estacionales	250
6.6.1.- Cálculo del Índice Estacional por el Método del Porcentaje Medio	
6.6.2.- Desestacionalización de los Datos	
6.7.- Análisis de Movimientos Cíclicos e Irregulares	254
REFERENCIAS BIBLIOGRÁFICAS	259

PRESENTACIÓN

La Estadística es tan antigua como la humanidad y desde su origen ha tomado un carácter importante y estratégico, aportando al desarrollo socio-económico y político, por eso algunos estudiosos la llaman la ciencia del Estado. Los historiadores afirman que las primeras formas de la Estadística fueron los censos de población o bienes, organizados por el poder político con fines militares o fiscales. La Estadística hoy en día se la considera como una disciplina esencial en todos los campos del saber humano. Su utilización es decisiva en la planeación y programación de las actividades de cualquier índole, ayuda a manejar información para resolver problemas, predecir o pronosticar hacia el futuro, y en definitiva, para obtener conclusiones y tomar las decisiones más adecuadas.

Conocedores que muchas personas por una u otra razón, tienen cierto miedo a las asignaturas con base matemática como es la Estadística, aun cuando estamos conscientes que los cálculos matemáticos juegan un papel importante en nuestras vidas, se pone a disposición del público el presente texto con ejemplos ilustrativos que han sido cuidadosamente seleccionados y resueltos didácticamente empleando un lenguaje matemático sencillo de manera manual y recurriendo al uso de los programas de fácil comprensión como son el Excel y el Graph. En cada capítulo constan las competencias que se espera que el alumno sea capaz de alcanzar, los contenidos a tratar y las tareas de interaprendizaje. Los contenidos y las tareas de interaprendizaje se han organizado de manera secuencial e interrelacionadas entre sí para afianzar y concatenar los conocimientos y competencias que se van logrando en el desarrollo de cada capítulo del presente texto. En general, los lectores, dispondrán de los elementos básicos sobre esta fascinante disciplina, que les permitirá aclarar juicios y ordenar ideas orientadas al trabajo autónomo, reflexivo y creador durante el proceso de interaprendizaje de la misma.

Los contenidos y procesos didácticos de interaprendizaje de la presente obra ya fueron puestos en práctica en con las y los estudiantes durante algunos años, obteniéndose resultados óptimos, por lo que estamos seguros que el presente libro de Estadística Básica tendrá la acogida por parte de la comunidad académica y contribuirá a mejorar significativamente la comprensión de esta hermosa ciencia.

Seguros de que ninguna obra humana es perfecta, serán ustedes estimados lectores los que con sus sugerencias nos ayudarán a mejorar la presente propuesta.

¡Bienvenidos a incursionar en el fascinante estudio de la Estadística!

Los Autores

EVALUACIÓN DIAGNÓSTICA

Lea cuidadosamente cada una de las siguientes interrogantes y conteste según sus conocimientos, no importa si se equivoca. La presente evaluación puede ser resuelta de manera grupal o individual. Siempre trabaje con disciplina, honradez y buena voluntad. Recuerde que el éxito se refleja en nuestro trabajo y hay que lograrlo, tarea tras tarea, y merecer ese logro. Los Autores

Según la naturaleza de los siguientes enunciados, escriba en el paréntesis la letra V si es verdadero o la F si es falso. Si su respuesta es F escriba el ¿por qué? de su respuesta.

- 1) La Estadística se encarga del estudio de las características cualitativas del fenómeno. ()
- 2) A la Estadística le interesan los fenómenos de tipo cuantitativo. ()
- 3) A la Estadística solamente le interesa la recopilación de datos. ()
- 4) Los fines de la estadística son conocer las características de los fenómenos, analizarlos y predecir lo que sucederá en el futuro. ()
- 5) Los objetivos de la Estadística son recopilar, organizar, tabular y presentar gráficamente los datos, proporcionando una visión cuantitativa de los fenómenos observados. ()
- 6) Los métodos de la Estadística son recopilar, clasificar, tabular y presentar datos para la toma de decisiones y solución de problemas. ()
- 7) La estadística descriptiva busca obtener información sobre la población basándose en el estudio de los datos de una muestra tomada a partir de ella. ()
- 8) La estadística inferencial se preocupa de llegar a conclusiones basados en la muestra y luego hacerlos válidos para toda la población. ()
- 9) La muestra es el conjunto de todos los elementos que tienen una característica común ()
- 10) Las partes de una tabla o cuadro estadístico son: título, conceptos o columna Matriz y cuerpo del cuadro. ()

Conteste a las siguientes preguntas

- 11) Sugiera 5 referentes de información que usted suponga son de tipo estadístico.
- 12) ¿Qué piensa usted que es la Estadística?
- 13) ¿Para qué sirven los censos poblacionales o de alguna otra índole?
- 14) Redacte un pensamiento que indique la importancia de la Estadística.
- 15) ¿Para qué sirven los gráficos estadísticos?. Enumere los que usted conoce.
- 16) ¿Qué son las medidas de tendencia central?. Enumere las que usted conoce.
- 17) Defina con sus propias palabras lo que entiende por medidas de dispersión. Enumere las que usted conoce.
- 18) ¿Qué entiende por medidas de forma?
- 19) ¿En qué se diferencian la correlación y la regresión?
- 20) ¿Cuál es la aplicación principal de las series cronológicas?

CAPÍTULO I

DESCRIPCIÓN Y PRESENTACIÓN DE LA INFORMACIÓN ESTADÍSTICA

COMPETENCIAS:

- ✓ Describe adecuadamente con sus propias palabras conceptos básicos de Estadística y su aplicación.
- ✓ Recopila correctamente de manera práctica información estadística utilizando encuestas.
- ✓ Organiza y presenta eficientemente la información estadística en tablas y gráficos de manera manual y empleando Excel.

CONTENIDOS:

- ✓ ¿Qué es la Estadística?
- ✓ Conceptos y Definiciones Básicas.
- ✓ Tablas o Cuadros Estadísticos.
- ✓ Distribución de Frecuencias.
- ✓ Gráficos Estadísticos.

1.1) ¿QUÉ ES LA ESTADÍSTICA?

1.1.1) HISTORIA

Establecer con absoluta claridad y precisión el proceso de desarrollo de esta ciencia que actualmente se llama Estadística, es una tarea difícil ya que la información que se dispone es fragmentada, parcial y aislada.

Es seguro que desde la antigüedad se realizaron inventarios de habitantes, bienes, productos, etc. Estos inventarios o censos (palabra derivada del latín *cencere* que significa valorar o tasar) se realizaron con fines catastrales, tributarios y militares.

En Egipto ya en el año 3050 a. c se tiene noticias de estadísticas destinadas a fines semejantes a los señalados y especialmente en la construcción de las pirámides.

En China en el año 2000 a. c. se conocen estudios similares. El nacimiento de Cristo coincide con la realización de un censo poblacional en gran escala en el Imperio Romano. Durante mucho tiempo se entendía por “estadística” la información relacionada con el gobierno, la palabra misma se deriva del latín *statisticus* o *estatus* que significa “del estado”.

Ya en nuestra era, en el año 727, los árabes realizaron estadísticas similares en lo que hoy es España. En Inglaterra en el año 1083 y 1662 y en Alemania en 1741, se llevaron a cabo censos referentes a defunciones, nacimientos, enfermedades, posesión de bienes, migraciones y otros problemas y los datos obtenidos se utilizaron en la previsión y planificación. En América se realizaron encuestas mediante el sistema de “quipus”.

El desarrollo científico de la estadística comienza recién en el siglo XVII, con la introducción en el pensum de estudio de las universidades en Alemania.

A comienzos del siglo XX, una nueva aportación de la escuela inglesa, preocupada por problemas de índole agropecuaria y biométrica coloca a la estadística en el tramo final de su establecimiento como ciencia.

En general las primeras aplicaciones de la estadística tuvieron que ver directamente con las actividades del estado. Se cree que la primera persona que hizo uso de la palabra estadística fue Godofredo Achenwall (1719-1772), profesor y economista alemán, escribió sobre el descubrimiento de una nueva ciencia que llamó estadística (palabra derivada de *Staat* que significa gobierno) y que definió como “el conocimiento profundo de la situación respectiva y comparativa de cada estado”.

1.1.2) DEFINICIÓN

Existen muchas definiciones de Estadística, pero en síntesis la podemos definir como la ciencia rama de la Matemática que se ocupa de recolectar, organizar, presentar, analizar e interpretar información cuantitativa para obtener conclusiones válidas, solucionar problemas, predecir fenómenos y ayudar a una toma de decisiones más efectivas.

1.1.3) APLICACIONES

La Estadística anteriormente sólo se aplicaba a los asuntos del Estado, pero en la actualidad la utilizan las compañías de seguros, empresarios, comerciantes, educadores, etc. No hay campo de la actividad humana que no requiera del auxilio de esta ciencia, así por ejemplo:

- El educador mediante la estadística podrá conocer si un estudiante lee muy bien o regular, si la asistencia es normal o irregular, si la estatura está en relación con la edad, media aritmética de rendimiento escolar en un período determinado, etc.
- El hombre de negocios realiza encuestas estadísticas para determinar la reacción de los consumidores frente a los actuales productos de la empresa y en el lanzamiento de los nuevos.
- El economista emplea una amplia gama de estadísticas para estudiar los planes de los consumidores y efectuar pronósticos sobre las tendencias de las actividades económicas
- El gerente de una empresa eléctrica proporciona un buen servicio a la comunidad mediante la variación estacional de las necesidades de carga
- El sociólogo trata de auscultar la opinión pública mediante encuestas, para determinar su preferencia por un candidato presidencial, o su posición frente a determinados problemas económicos, políticos o sociales
- El geólogo utiliza métodos estadísticos para determinar las edades de las rocas
- El Genetista determina las semejanzas entre los resultados observados y esperados en una experiencia genética se determina estadísticamente

1.1.4) FINES

- **Conocer** las características de un grupo de casos de estudio.
- **Comparar** entre los resultados actuales y los obtenidos en experiencias pasadas para determinar las causas que han influenciado en los cambios.
- **Predecir** lo que puede ocurrir en el futuro de un fenómeno.

1.1.5) OBJETIVOS

- **Describir** numéricamente las características de los conjuntos de observaciones. Esta etapa consiste en recopilar, organizar, tabular y presentar gráficamente los datos, proporcionando una visión cuantitativa de los fenómenos observados.
- **Analizar** los datos de manera objetiva con el fin de disponer de un concepto claro de universo o población y adoptar decisiones basadas en la información proporcionada por los datos de la muestra.
- **Estimar** o predecir lo que sucederá en el futuro con un fenómeno de una manera relativamente aceptable, así por ejemplo, podemos estimar cuál será la población del país dentro de un determinado número de años conociendo la actual.

1.1.6) MÉTODOS

- **Recopilación.-** Consiste en la obtención de datos relacionados con el problema motivo de estudio, utilizando instrumentos, tales como: cuestionarios, entrevistas, informes, memorias, etc.
- **Organización.-** Consiste en realizar una crítica, corrección, clasificación y tabulación de los datos obtenidos en el paso anterior.
- **Presentación.-** Consiste en mostrar datos de manera significativa y descriptiva. Los datos deben colocarse en un orden lógico que revele rápida y fácilmente el mensaje que contienen. La presentación se la puede hacer a través de gráficos estadísticos.
- **Análisis.-** Consiste en descomponer el fenómeno en partes y luego examinar cada una de ellas con el objetivo de lograr una explicación, haciendo uso, en su mayoría, de los cálculos matemáticos.
- **Interpretación.-** Consiste en un proceso mental, mediante el cual se encuentra un significado más amplio de los datos estadísticos con el objetivo de llegar a conclusiones para la toma de decisiones y solución de problemas.

1.1.7) CLASIFICACIÓN DE LA ESTADÍSTICA

1.1.7.1) Estadística Descriptiva o Deductiva

Es un proceso mediante el cual se recopila, organiza, presenta, analiza e interpreta datos de manera tal que describa fácil y rápidamente las características esenciales de dichos datos mediante el empleo de métodos gráficos, tabulares o numéricos, así por ejemplo:

Supóngase que un docente de Matemática calcula la calificación promedio de uno de sus cursos a su cargo. Como solo se está describiendo el desempeño del curso pero no hace ninguna generalización acerca de los diferentes cursos, en este caso el maestro está haciendo uso de la Estadística Descriptiva.

1.1.7.2) Estadística Inferencial o Inductiva

Llamada también inferencia estadística, la cual consiste en llegar a obtener conclusiones o generalizaciones que sobrepasan los límites de los conocimientos aportados por un conjunto de datos. Busca obtener información sobre la población basándose en el estudio de los datos de una muestra tomada a partir de ella, así por ejemplo:

Supóngase ahora que el docente de Matemática utiliza el promedio de calificaciones obtenidas por uno de sus cursos para estimar la calificación promedio de los 5 cursos a su cargo. Como se está realizando una generalización acerca los diferentes cursos, en este caso el maestro usa la Estadística Inferencial.

1.2) CONCEPTOS Y DEFINICIONES BÁSICAS

1.2.1) POBLACIÓN

Llamado también universo o colectivo es el conjunto de todos los elementos que tienen una característica común.

Una población puede ser finita o infinita. Es **población finita** cuando está delimitada y conocemos el número que la integran, así por ejemplo: Estudiantes de la Universidad UTN. Es **población infinita** cuando a pesar de estar delimitada en el espacio, no se conoce el número de elementos que la integran, así por ejemplo: Todos los profesionales universitarios que están ejerciendo su carrera.

1.2.2) MUESTRA

Es un subconjunto de la población. Ejemplo: Estudiantes de 2do Semestre de la Universidad UTN.

Sus principales características son:

Representativa.- Se refiere a que todos y cada uno de los elementos de la población tengan la misma oportunidad de ser tomados en cuenta para formar dicha muestra.

Adecuada y válida.- Se refiere a que la muestra debe ser obtenida de tal manera que permita establecer un mínimo de error posible respecto de la población.

Para que una muestra sea fiable, es necesario que su tamaño sea obtenido mediante procesos matemáticos que eliminen la incidencia del error.

Para calcular el tamaño de la muestra suele utilizarse la siguiente fórmula:

$$n = \frac{N\sigma^2 Z^2}{(N - 1)e^2 + \sigma^2 Z^2}$$

Donde:

n = el tamaño de la muestra.

N = tamaño de la población.

σ = Desviación estándar de la población que, generalmente cuando no se tiene su valor, suele utilizarse un valor constante de 0,5.

Z = Valor obtenido mediante niveles de confianza. Es un valor constante que, si no se tiene su valor, se lo toma en relación al 95% de confianza equivale a 1,96 (como más usual) o en relación al 99% de confianza equivale 2,58, valor que queda a criterio del encuestador.

e = Límite aceptable de error muestral que, generalmente cuando no se tiene su valor, suele utilizarse un valor que varía entre el 1% (0,01) y 9% (0,09), valor que queda a criterio del encuestador.

Ejemplo ilustrativo: Calcular el tamaño de la muestra de una población de 1000 elementos.

Solución:

Se tiene N=1000, y como no se tiene los demás valores se tomará $\sigma = 0,5$, Z = 1,96 y e = 0,05.

Reemplazando valores en la fórmula se obtiene:

$$n = \frac{N\sigma^2 Z^2}{(N - 1)e^2 + \sigma^2 Z^2} = \frac{1000 \cdot 0,5^2 \cdot 1,96^2}{(1000 - 1) \cdot 0,05^2 + 0,5^2 \cdot 1,96^2} = \frac{1000 \cdot 0,25 \cdot 3,8416}{(999) \cdot 0,0025 + 0,25 \cdot 3,8416}$$
$$n = \frac{960,4}{2,4975 + 0,9604} = \frac{960,4}{3,4579} = 277,74 = 278$$

Estos cálculos realizados en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G
1	N	1000					
2	σ	0,5					
3	Z	1,96					
4	e	0,05					
5							
6	$n = \frac{N\sigma^2 Z^2}{(N-1)e^2 + \sigma^2 Z^2}$			277,74083			
7							
8				=(B1*B2^2*B3^2)/((B1-1)*B4^2+B2^2*B3^2)			

1.2.3) ELEMENTO O INDIVIDUO

Unidad mínima que compone una población. El elemento puede ser una entidad simple (una persona) o una entidad compleja (una familia), y se denomina unidad investigativa.

1.2.4) DATOS ESTADÍSTICOS

Son medidas, valores o características susceptibles de ser observados y contados. Como por ejemplo, la edad de los estudiantes de la Universidad UTN.

Los datos estadísticos pueden ser clasificados en *cuantitativos* (la diferencia entre ellos es de clase y no de cantidad), *cuantitativos* (representan magnitudes), *cronológicos* (difieren en instantes o períodos de tiempo) y *geográficos* (referidos a una localidad).

Los datos estadísticos se obtienen de *fuentes primarias* (obtenidos directamente sin intermediarios valiéndose de observaciones, encuestas, entrevistas y sondeos de opinión) y *fuentes secundarias* (obtenidos a través de intermediarios valiéndose de textos, revistas, documentos, publicaciones de prensa, y demás trabajos hechos por personas o entidades).

1.2.5) CENSO

Es una técnica de recolección de datos estadísticos que se realiza a toda la población

1.2.6) ENCUESTA

Es la técnica que nos permite recolectar datos estadísticos que se realiza una muestra de la población.

Se clasifica en:

- *Descriptiva.*- Cuando registra datos referentes a las características de los elementos o individuos.
- *Explicativa.*- Cuando averigua las causas o razones que originan los fenómenos.
- *Mixtas.*- Cuando es descriptiva y explicativa.

- **Por muestreo.**- Cuando recolecta información de grupos representativos de la población.

Su estructura es:

- Nombre de la institución que auspicia la encuesta.
- Tema de la encuesta.
- Objetivos de la encuesta.
- Datos informativos: Lugar, fecha, y otros datos que se considere necesario según la naturaleza de la información estadística a encuestarse.
- Instrucciones para el encuestado para que sepa la forma de llenar la encuesta.
- Cuestionario o listado de preguntas (cerradas, abiertas, o ambas a la vez) sobre los diferentes aspectos motivo de estudio.
- Frase de agradecimiento al encuestado, como por ejemplo, ¡Gracias por su colaboración!

Las diferentes tipos de preguntas pueden ser:

- **Abiertas.**- Son aquellas en la cual el encuestado construye la respuesta de manera libre según su opinión y de la manera que él desea. Ejemplo: ¿Qué piensa usted sobre la política educativa del actual gobierno?

- **Cerradas o dicotómicas.**- Sólo pueden ser contestadas por un “sí” o por un “no”. Ejemplo: ¿Está usted de acuerdo con la política educativa del actual gobierno?

Si ()
No ()

Como es obvio, la respuesta será forzosamente una de las alternativas planteadas: Las preguntas cerradas son fáciles de tabular y facilitan la cuantificación mediante la asignación de puntuaciones.

- **Preguntas de elección múltiple o categorizada:** Se trata en cierto modo de preguntas cerradas que, dentro de los extremos de una escala permiten una serie de alternativas de respuestas cuyos matices son fijados de antemano. Presentan dos formas: En abanico y de estimación

- **Preguntas con respuesta en abanico:** Estas preguntas permiten contestar señalando una o varias respuestas presentadas junto con la pregunta. Por ejemplo: Indique otras alternativas que considere importantes para mejorar la educación en nuestro país.

- **Preguntas de Estimación:** Son preguntas cuantitativas que introducen diversos grados de intensidad creciente o decreciente para un mismo ítem. Ejemplos:

-¿Cómo calificaría la política educativa del gobierno actual?

Excelente ()

Muy Buena ()

Regular ()

Deficiente ()

-¿En qué porcentaje está de acuerdo con la política educativa del gobierno actual?

100% () 75% () 50% () 25% () 0% ()

- ¿Le interesa conocer el modelo educativo vigente?

Nada () Poco () Algo () Mucho ()

¿Piensa culminar sus estudios superiores?

Sí () Probablemente Sí () No () Aún no decido ()

1.2.7) VARIABLE

Son caracteres susceptibles a cambio y pueden tener diferentes valores en cada elemento o individuo.

1.2.7.1) Clasificación

- *Variable Cualitativa*

Son atributos que se expresan mediante palabras no numéricas. Como por ejemplo, profesión, religión, marca de automóvil, estado civil, sexo, raza, etc.

- *Variable Cuantitativa*

Es toda magnitud representada por números. Como por ejemplo, peso, estatura, número de habitantes, etc.

- *Variable Discreta*

Es una característica cuantitativa representada por números enteros o exactos, que generalmente resultan del proceso de conteo, como por ejemplo: número de estudiantes de la promoción del año anterior.

- *Variable Continua*

Es una característica cuantitativa que puede tomar cualquier valor representado por un número racional, que generalmente resultan del proceso de medición, como por ejemplo, tiempo destinado a estudiar Estadística

1.2.7.2) Niveles de medición

- *Nivel Nominal*

Cuando los datos sólo pueden contarse y clasificados en categorías, no existe un orden específico entre las clases. Como por ejemplo, se cuentan cuántos hombres y cuántas mujeres asisten a determinado evento.

- Nivel Ordinal

Cuando se ordenan los datos por jerarquías, una categoría es mayor que otra. Como por ejemplo, excelente es mejor que bueno o bueno es mejor que regular. Otro ejemplo: Una persona puede tener mucho o poco dinero.

- Nivel de Intervalos

Cuando se incluye todas las características del nivel ordinal, pero la diferencia entre los valores tiene un significado medido en unidades iguales que son comunes y constantes, que permiten asignar números reales a todos los miembros de la clase ordenada, facilitando el establecimiento de diferencias en grados de propiedad y entre objetos sobre la base de una medida. Como por ejemplo: La diferencia entre 70 kilogramos y 60 kilogramos, es de 10 kilogramos. Otro ejemplo: Si la temperatura de hoy es de 20 grados centígrados y la de ayer fue de 25 grados centígrados, se sabe que la de hoy es 5 grados centígrados más baja que la de ayer.

- Nivel de Razón o Cociente

Este es el nivel de medición “más alto”, tiene todas las características del nivel de intervalos y además en este nivel de medición el cero tiene significado (así si se tiene 0 dólares, entonces no se poseen fondos), y la razón (o cociente) entre dos números también es significativa (Un estudiante obtiene una calificación de 3/10 y otro 6/10, el segundo estudiante obtiene el doble que el primero).

TAREA DE INTERAPRENDIZAJE N° 1

- 1) Realice un organizador gráfico (cuadro sinóptico, mapa conceptual, mentefacto, etc.) sobre la historia de la Estadística.
- 2) Etimológicamente resuma qué significa la Estadística
- 3) Defina con sus propias palabras lo que entiende por Estadística.
- 4) Proponga 3 aplicaciones de la Estadística en su vida cotidiana.
- 5) Realice un organizador gráfico sobre los fines de la Estadística.
- 6) Realice un organizador gráfico sobre los objetivos de la Estadística.
- 7) Realice un organizador gráfico sobre los métodos de la Estadística.
- 8) Defina con sus propias palabras lo que entiende por Estadística Descriptiva y Estadística Inferencial. Ilustre con un ejemplo cada definición.
- 9) Proponga 3 ejemplos de población, muestra y elemento.
- 10) Calcule el tamaño de la muestra para una población de 5000 con un error de muestreo del 5% y nivel de confianza del 95%. Realice los cálculos de manera manual y empleando Excel.
- 11) Cree y resuelva un ejercicio sobre el cálculo del tamaño de una muestra
- 12) Realice un organizador gráfico sobre los datos estadísticos.

357

13) Escriba una semejanza y una diferencia entre censo y encuesta.

14) Elabore una encuesta mixta con 10 preguntas sobre cualquier tema de su preferencia. Y aplique la encuesta empleando los conocimientos del tamaño de la muestra.

15) Según el tipo de variable, escriba en el paréntesis la letra D si es Discreta o la C si es Continua. Escriba el ¿por qué? de su respuesta.

- 15.1) Número de aulas de una universidad. ()
- 15.2) Número de goles recibidos por un equipo de fútbol. ()
- 15.3) Estatura de los compañeros del curso de Estadística. ()
- 15.4) Peso del contenido de una caja de cereal. ()
- 15.5) Número de libros leídos el año pasado. ()
- 15.6) Diámetro de un cojín. ()
- 15.7) Número de artículos defectuosos producidos por una máquina. ()
- 15.8) Volumen de aire en un aula. ()
- 15.9) Número de individuos de un sector que reciben el bono solidario. ()
- 15.10) Temperatura ambiente en una ciudad. ()

D, D, C, C, D, C, D, C, D, C

16) Según el nivel de medición de las variables, escriba en el paréntesis la letra N si es Nominal, La O si es Ordinal, la I si es de Intervalos o la R si es de Razón. Escriba el ¿por qué? de su respuesta.

- 16.1) Nivel de riesgo sobre catástrofes naturales en una ciudad. ()
- 16.2) Salarios de los miembros de una familia. ()
- 16.3) El número de años que cada empleado ha trabajado. ()
- 16.4) Un sistema para medir las preferencias de los clientes respecto a los vehículos con base en su estilo. ()
- 16.5) Un sistema para identificar las ciudades de nacimiento de los alumnos. ()
- 16.6) Un sistema para evaluar a los empleados con base en el número de días que faltan al trabajo. ()
- 16.7) Un sistema para medir las edades de los alumnos. ()
- 16.8) Un sistema para medir la calificación de un examen especial de ingreso a la Universidad. ()
- 16.9) Determinar la cantidad de dinero gastado en ropa en el mes pasado. ()

- 16.10) Determinar el número de abrigos de mujer para el invierno. ()
- 16.11) Un sistema para medir volúmenes de líquidos. ()
- 16.12) Indicar el estado civil de los compañeros de trabajo. ()
- 16.13) Determinar el tiempo utilizado para estudiar en el mes pasado. ()
- 16.14) Promedio de calificación de los estudiantes. ()
- 16.15) Calificaciones de los estudiantes en la primera prueba de Estadística. ()
- 16.16) Determinar el número de pantalones utilizados la semana pasada. ()
- 16.17) Ponderación de las calificaciones obtenidas por un estudiante en un colegio. ()
- 16.18) Número de trabajos terminados, en el último año, de los empleados de una fábrica. ()
- 16.19) Determinar si a un grupo de personas les gustan o no un determinado producto. ()
- 16.20) Las distancias entre las casas de un sector. ()

O, R, N, N, N, O, I, O, O, N, I, N, O, I, I, N, N, O, N, N, R

17) Proponga 3 ejemplos de cada una de los tipos de variables y 3 ejemplos de cada uno de los niveles de medición.

1.3) TABLAS O CUADROS ESTADÍSTICOS

Son representaciones tabulares que sirven para ordenar la información estadística, las cuales están formadas de filas (horizontales) y columnas (verticales).

Sus partes son:

- **Número.**- Es necesario sobre todo cuando existen varios cuadros
- **Título.**- Consiste en la descripción del contenido en forma entendible. Responde a las interrogantes: ¿qué?, ¿cómo?, ¿dónde?, ¿cuándo? y ¿cuánto?
- **Encabezado.**- Son los títulos de la parte superior de las columnas
- **Conceptos.**- Son descripciones que van en las filas del cuadro y son clasificaciones de los encabezados.
- **Columna Matriz.**- Se conforma de los diferentes conceptos.
- **Cuerpo.**- Constituye el contenido mismo del cuadro.
- **Fuente.**- Se pone cuando los datos han sido sacados de documentos o fuentes secundarias. Se ubica debajo del cuadro.

- **Nota de Encabezado.**- Sirven para clarificar partes del cuadro que no han sido incluidas en el título. Se ubica después del título entre paréntesis. No siempre está presente en un cuadro.

- **Nota de Pie.**- Sirven para clarificar algunas partes del cuadro que no son explicadas en ninguna parte. Se ubica después de la fuente. No siempre está presente en un cuadro.

A continuación se presenta un ejemplo ilustrativo de cuadro identificando sus partes:

Cuadro N°...

Título



INSCRIPCIONES EN INSTITUCIONES DE EDUCACIÓN SUPERIOR, POR TIPO Y POR GÉNERO 2010-2011

Nota de encabezado (en miles)

TIPO DE INSTITUCIÓN	2010			2011		
	Hombres	Mujeres	TOTAL	Hombres	Mujeres	TOTAL
Universidades
Escuelas Politécnicas
Institutos Tecnológicos
Institutos Militares

Columna matriz (pointing to the first column)
Encabezados (pointing to the top row)
Conceptos (pointing to the first column)
Cuerpo (pointing to the data cells)

Fuente: Departamento de Estadística del SENECYT

Nota de pie: Inscripciones de estudiantes regulares sólo del primer semestre.

TAREA DE INTERAPRENDIZAJE N° 2

- 1) Elabore un cuadro sobre cualquier información de su preferencia e identifique sus partes.
- 2) El Colegio D & M en el año lectivo 2010-2011 tiene 800 alumnos y para su funcionamiento dispone del siguiente personal por estamento: directivo: un rector y un vicerrector; de control: un inspector general y 3 inspectores; docente: 80 maestros; psicólogos: 2; de administración: 4, y mantenimiento: 10. Llene la siguiente tabla o cuadro:

Cuadro N° 1

PERSONAL DEL COLEGIO D & M POR ESTAMENTO Y ALUMNOS POR CADA UNO EN EL AÑO LECTIVO 2010-2011

Estamento	N° de personas	Alumnos por cada uno
Directivo	2	400
Totales		

Fuente: Departamento Administrativo del Colegio D & M

3) Investigue el consumo de energía eléctrica de su domicilio durante cada uno de los 6 últimos meses. Elabore un cuadro o tabla con los datos investigados, calculando el porcentaje de consumo de cada mes sobre el total.

4) Investigue sobre cualquier tema de su interés y elabore una tabla.

1.4) DISTRIBUCIÓN DE FRECUENCIAS

Son tablas que resumen los datos originales en frecuencias.

1.4.1) PARA DATOS SIN AGRUPAR

Los tipos de frecuencia pueden ser:

- *Frecuencia Absoluta (f)*

Es el número de veces que se repite el valor de cada variable. La suma de frecuencias absolutas es siempre al total de datos observados.

- *Frecuencia Relativa (fr)*

Indica la proporción con que se repite un valor. Es el cociente entre la frecuencia absoluta y el número total de datos. La suma de las frecuencias relativas es siempre 1

$$fr = \frac{f}{n}$$

- *Frecuencia Acumulada (fa)*

Indica el número de valores que son menores o iguales que el valor dado. Es la suma de la frecuencia absoluta primera con la segunda, este valor con la tercera, y así sucesivamente.

- Frecuencia Porcentual ($f\%$)

Llamada también frecuencia relativa porcentual. Se obtiene multiplicando la frecuencia relativa por 100. La suma de las frecuencias porcentuales es siempre 100%. Se calcula así:

$$f\% = fr \cdot 100$$

- Frecuencia Relativa Acumulada (fra)

Es la suma de la frecuencia relativa primera con la segunda, este valor con la tercera, y así sucesivamente.

- Frecuencia Relativa Acumulada Porcentual ($fra\%$)

Indica el número de valores que son menores o iguales que el valor dado. Se obtiene multiplicando la frecuencia relativa acumulada por 100. Se calcula así:

$$fra\% = fra \cdot 100$$

Ejemplo ilustrativo:

Calcular las diferentes frecuencias de las siguientes calificaciones evaluadas sobre 10 obtenidas de 40 estudiantes en la asignatura de Estadística sin agrupar en clases:

10	8	9	8	7	8	9	10
6	7	10	9	8	8	10	8
6	5	6	8	10	5	9	9
8	10	9	7	6	7	7	6
8	10	7	8	5	9	8	5

Solución:

El ejercicio resuelto se muestra en la tabla:

Calificación	f	fr	fa	$f\%$	fra	$fra\%$
5	4	$4/40 = 0,1$	4	$0,1 \cdot 100 = 10$	0,1	$0,1 \cdot 100 = 10$
6	5	$5/40 = 0,125$	$4+5 = 9$	$0,125 \cdot 100 = 12,5$	$0,1+0,125 = 0,225$	$0,225 \cdot 100 = 22,5$
7	6	$6/40 = 0,15$	$9+6 = 15$	$0,15 \cdot 100 = 15$	$0,225+0,15 = 0,375$	$0,375 \cdot 100 = 37,5$
8	11	$11/40 = 0,275$	$15+11 = 26$	$0,275 \cdot 100 = 27,5$	$0,375+0,275 = 0,65$	$0,65 \cdot 100 = 65$
9	7	$7/40 = 0,175$	$26+7 = 33$	$0,175 \cdot 100 = 17,5$	$0,65+0,175 = 0,825$	$0,825 \cdot 100 = 82,5$
10	7	$7/40 = 0,175$	$33+7 = 40$	$0,175 \cdot 100 = 17,5$	$0,825+0,175 = 1$	$1 \cdot 100 = 100$
Total	40	1		100		

En Excel se calcula de la siguiente manera:

a) Digite los datos en una hoja de cálculo. Pulse en B8 y seleccione insertar función. Clic en insertar función (fx). En el área seleccionar una categoría, seleccione Estadísticas. En el área seleccionar una función elija la función CONTAR.SI.

The screenshot shows an Excel spreadsheet with data in columns A through H and rows 1 through 5. The formula bar shows the active cell B8. The 'Insertar función' dialog box is open, with the 'Estadísticas' category selected and 'CONTAR.SI' chosen from the list of functions. The dialog box also displays the function's syntax: **CONTAR.SI(rango; criterio)** and its description: 'Cuenta las celdas en el rango que coinciden con la condición dada.'

	A	B	C	D	E	F	G	H
1	10	8	9	8	7	8	9	10
2	6	7	10	9	8	8	10	8
3	6	5	6	8	10	5	9	9
4	8	10	9	7	6	7	7	6
5	8	10	7	8	5	9	8	5

b) Pulse en Aceptar para que se abra el cuadro de Argumentos de la función. Coloque el cursor en el recuadro Rango y arrastre el ratón por la hoja seleccionando el rango A1:H5

The screenshot shows the Excel spreadsheet with the range A1:H5 selected. The formula bar displays the function: **=CONTAR.SI(A1:H5)**. The 'Argumentos de función' dialog box is open, showing the 'Rango' field set to 'A1:H5' and the 'Criterio' field empty. The dialog box also displays the function's description and the result of the formula: 'Resultado de la fórmula ='. The 'Rango' field is highlighted with a dashed border, indicating it is the active selection.

	A	B	C	D	E	F	G	H
1	10	8	9	8	7	8	9	10
2	6	7	10	9	8	8	10	8
3	6	5	6	8	10	5	9	9
4	8	10	9	7	6	7	7	6
5	8	10	7	8	5	9	8	5

c) Dentro del recuadro Rango sobre el rango allí escrito, pulse la tecla F4 con la finalidad de que el rango permanezca fijo.

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I
1	10	8	9	8	7	8	9	10	
2	6	7	10	9	8	8	10	8	
3	6	5	6	8	10	5	9	9	
4	8	10	9	7	6	7	7	6	
5	8	10	7	8	5	9	8	5	
6									
7	Calificación	f							
8	5	(\$1:\$H\$5)							
9	6								
10	7								
11	8								
12	9								
13	10								
14									
15									
16									
17									
18									
19									

The dialog box 'Argumentos de función' for the COUNTIF function is open. It shows the following settings:

- Función: CONTAR.SI
- Rango: \$A\$1:\$H\$5
- Criterio: (empty)
- Resultado de la fórmula: =

d) Coloque el cursor en el recuadro Criterio y pulse la celda A8.

The screenshot shows the same Excel spreadsheet as above. The dialog box 'Argumentos de función' for the COUNTIF function is open with the following settings:

- Función: CONTAR.SI
- Rango: \$A\$1:\$H\$5
- Criterio: A8
- Resultado de la fórmula: = 4

e) Clic en Aceptar para obtener la frecuencia absoluta de la calificación de 5

	A	B	C	D	E	F	G	H
1	10	8	9	8	7	8	9	10
2	6	7	10	9	8	8	10	8
3	6	5	6	8	10	5	9	9
4	8	10	9	7	6	7	7	6
5	8	10	7	8	5	9	8	5
6								
7	Calificación	f						
8	5	4						
9	6							
10	7							
11	8							
12	9							
13	10							

f) Para calcular las frecuencias absolutas de las otras calificaciones, pulse la esquina inferior derecha de la celda B8, hasta que aparezca una cruz, luego arrastre el curso hacia abajo.

	A	B	C	D	E	F	G	H
1	10	8	9	8	7	8	9	10
2	6	7	10	9	8	8	10	8
3	6	5	6	8	10	5	9	9
4	8	10	9	7	6	7	7	6
5	8	10	7	8	5	9	8	5
6								
7	Calificación	f						
8	5	4						
9	6	5						
10	7	6						
11	8	11						
12	9	7						
13	10	7						

g) Los cálculos de las demás frecuencias se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Calificación	f	fr		fa		f%		fra		fra%	
2	5	4	0,1	=B2/\$B\$8	4	=B2	10	=C2*100	0,1	=C2	10	=I2*100
3	6	5	0,125	=B3/\$B\$8	9	=E2+B3	12,5	=C3*100	0,225	=I2+C3	22,5	=I3*100
4	7	6	0,15	=B4/\$B\$8	15	=E3+B4	15	=C4*100	0,375	=I3+C4	37,5	=I4*100
5	8	11	0,275	=B5/\$B\$8	26	=E4+B5	27,5	=C5*100	0,65	=I4+C5	65	=I5*100
6	9	7	0,175	=B6/\$B\$8	33	=E5+B6	17,5	=C6*100	0,825	=I5+C6	82,5	=I6*100
7	10	7	0,175	=B7/\$B\$8	40	=E6+B7	17,5	=C7*100	1	=I6+C7	100	=I7*100
8	Total	40	1				100					
9		=SUMA(B2:B7)	=SUMA(C2:C7)				=SUMA(G2:G7)					

1.4.2) PARA DATOS AGRUPADOS EN CLASES O INTERVALOS

Cuando los datos contienen una gran cantidad de elementos, para facilitar los cálculos es necesario agruparlos, a estos grupos se los llama intervalos o clases. Un intervalo es una serie de números incluidos entre dos extremos, así por ejemplo, el intervalo 40 – 45 está formado por 40, 41, 42, 43, 44 y 45, siendo 40 el límite inferior, 45 el límite superior, 39,5 límite real inferior (límite inferior disminuido en 5 décimas) y 40,5 el límite real superior (límite superior aumentado en 5 décimas).

Las reglas generales para formas distribuciones de frecuencias para datos agrupados en clases son:

- **Calcule el Rango (R).**- También se llama recorrido o amplitud total. Es la diferencia entre el valor mayor y el menor de los datos.

$$R = x_{\text{máx}} - x_{\text{mín}}$$

- **Seleccione el Número de Intervalos de Clase (n_i).**- No debe ser menor de 5 y mayor de 12, ya que un número mayor o menor de clases podría oscurecer el comportamiento de los datos. Para calcular el número de intervalos se aplica la regla de Sturges, propuesta por Herberth Sturges en 1926:

$$n_i = 1 + 3,32 \cdot \log(n)$$

Siendo n el tamaño de la muestra.

- **Calcule el Ancho del Intervalo (i).**- Se obtiene dividiendo el Rango para el número de intervalos

$$i = \frac{R}{n_i}$$

Cuando el valor de i no es exacto, se debe redondear al valor superior más cercano. Esto altera el valor de rango por lo que es necesario efectuar un ajuste así:

$$\text{Nuevo } R = n_i \cdot i$$

Por ejemplo:

Si una distribución de 40 datos el valor mayor es 41 y el menor es 20 se tiene:

Calculando el Rango se obtiene:

$$R = x_{\text{máx}} - x_{\text{mín}} = 41 - 20 = 21$$

Calculando el número de intervalos se obtiene:

$$n_i = 1 + 3,32 \cdot \log(n) = 1 + 3,32 \cdot \log 40 = 6,32 = 6$$

Calculando el ancho se obtiene:

$$i = \frac{R}{n_i} = \frac{21}{6} = 3,5$$

Redondeando se obtiene: $i = 4$

Calculando el nuevo rango se obtiene:

$$\text{Nuevo } R = n_i \cdot i = 6 \cdot 4 = 24$$

El exceso de 3 que se tiene en este caso se distribuye entre $x_{\text{máx}}$ y $x_{\text{mín}}$. Por lo general se agrega al mayor y se quita al menor. Como por ejemplo, se podría agregar 2 al valor mayor y quitar 1 al valor menor, obteniéndose los siguientes nuevos valores:

$$x_{\text{máx}} = 41 + 2 = 43$$

$$x_{\text{mín}} = 20 - 1 = 19$$

O también se podría agregar 1 al valor mayor y quitar 2 al valor menor, obteniéndose los siguientes nuevos valores:

$$x_{m\acute{a}x} = 41 + 1 = 42$$

$$x_{m\acute{i}n} = 20 - 2 = 18$$

- **Forme los Intervalos de Clase agregando $i-1$** al límite inferior de cada clase, comenzando por el $x_{m\acute{i}n}$ del rango.

- **Se realiza el Conteo de Datos que cae dentro de cada clase (frecuencia absoluta)**

- **Calcule la Marca de Clase (x_m).**- Es el valor medio de cada clase, se obtiene sumando los límites superior (L_s) e inferior (L_i) del intervalo y dividiendo ésta suma entre 2

$$x_m = \frac{L_s + L_i}{2}$$

- **Calcule las Frecuencias**

Ejemplo ilustrativo:

A 40 estudiantes se les pidió que estimen el número de horas que habrían dedicado a estudiar la semana pasada (tanto en clase como fuera de ella), obteniéndose los siguientes resultados:

36	30	47	60	32	35	40	50
54	35	45	52	48	58	60	38
32	35	56	48	30	55	49	39
58	50	65	35	56	47	37	56
58	50	47	58	55	39	58	45

Solución:

1) Calculando el Rango se obtiene:

$$R = x_{m\acute{a}x} - x_{m\acute{i}n} = 65 - 30 = 35$$

2) Calculando el número de intervalos se obtiene:

$$n_i = 1 + 3,32 \cdot \log(n) = 1 + 3,32 \cdot \log 40 = 6,32 = 6$$

3) Calculando el ancho se obtiene:

$$i = \frac{R}{n_i} = \frac{35}{6} = 5,83$$

Redondeando se obtiene: $i = 6$, por lo que es necesario realizar un ajuste al rango.

Los cálculos realizados en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H
1	36	30	47	60	32	35	40	50
2	54	35	45	52	48	58	60	38
3	32	35	56	48	30	55	49	39
4	58	50	65	35	56	47	37	56
5	58	50	47	58	55	39	58	45
6								
7	n	40	=CONTAR(A1:H5)					
8	$x_{m\acute{a}x}$	65	=MAX(A1:H5)					
9	$x_{m\acute{i}n}$	30	=MIN(A1:H5)					
10	R	35	=B8-B9	=MAX(A1:H5)-MIN(A1:H5)				
11	n_i	6	=ENTERO(1+3,32*LOG10(B7))					
12	i	6	=B10/B11					

4) Calculando el nuevo rango se obtiene:

$$\text{Nuevo } R = n_i \cdot i = 6 \cdot 6 = 36$$

El exceso de 1 que se tiene en este caso se distribuye entre $x_{m\acute{a}x}$ y $x_{m\acute{i}n}$. En este ejemplo, se podría agregar 1 al valor mayor y no quitar nada al valor menor, o no agregar nada al mayor y quitar 1 al menor. Al elegir la primera opción se obtiene:

$$x_{m\acute{a}x} = 65 + 1 = 66$$

$$x_{m\acute{i}n} = 30 - 0 = 30$$

5) Formando los intervalos de clase agregando $i-1$ ($6-1=5$) al límite inferior de cada clase, comenzando por el $x_{m\acute{i}n}$ del rango se obtiene:

$$30+5 = 35; 36+5 = 41; 42+5 = 47; 48+5 = 53; 54+5 = 59; 60+5 = 65$$

6) Realizando el conteo de datos que cae dentro de cada clase, calculando la marca de clase y las frecuencias se obtiene:

Clases	f	x_m	fr	fa	$f\%$	fra	$fra\%$
30-35	8	$(30+35)/2 = 32,5$	0,2	8	20	0,2	20
36-41	6	$(36+41)/2 = 38,5$	0,15	14	15	0,35	35
42-47	5	$(42+47)/2 = 44,5$	0,125	19	12,5	0,475	47,5
48-53	7	$(48+53)/2 = 50,5$	0,175	26	17,5	0,65	65
54-59	11	$(54+59)/2 = 56,5$	0,275	37	27,5	0,925	92,5
60-65	3	$(60+65)/2 = 62,5$	0,075	40	7,5	1	100
Total	40		1		100		

A continuación se presenta algunas interpretaciones de la tabla:

El valor de $f=8$: Significa que 8 estudiantes dedicaron a estudiar la semana pasada entre 30 y 35 horas.

El valor de $xm = 50,5$: Significa que 7 estudiantes dedicaron en promedio a estudiar la semana pasada 50,5 horas.

El valor de $fr = 0,15$ y $f\% = 15\%$: Significa que el 0,15 o el 15% de los estudiantes dedicaron a estudiar la semana pasada entre 36 y 41 horas.

El valor de $fa = 26$: Significa que 26 estudiantes dedicaron a estudiar la semana pasada entre 30 y 53 horas.

El valor de $fra = 0,65$ y $fra\% = 65\%$: Significa que el 0,65 o el 65% de los estudiantes dedicaron a estudiar la semana pasado entre 30 y 53 horas.

Para realizar los cálculos de la frecuencia absoluta empleando Excel se procede de la siguiente manera:

a) Digite los datos, las clases y límites superiores de las clases.

	A	B	C	D	E	F	G	H
1	36	30	47	60	32	35	40	50
2	54	35	45	52	48	58	60	38
3	32	35	56	48	30	55	49	39
4	58	50	65	35	56	47	37	56
5	58	50	47	58	55	39	58	45
6								
7	Clase	L_s	f					
8	30-35	35						
9	36-41	41						
10	42-47	47						
11	48-53	53						
12	54-59	59						
13	60-65	65						

b) Seleccione C8:C13 donde las frecuencias absolutas deben ser calculadas.

	A	B	C	D	E	F	G	H
1	36	30	47	60	32	35	40	50
2	54	35	45	52	48	58	60	38
3	32	35	56	48	30	55	49	39
4	58	50	65	35	56	47	37	56
5	58	50	47	58	55	39	58	45
6								
7	Clase	L_s	f					
8	30-35	35						
9	36-41	41						
10	42-47	47						
11	48-53	53						
12	54-59	59						
13	60-65	65						

c) Escriba la fórmula: =FRECUENCIA (A1:H5; B8:B13)

CONTAR.SI		=FRECUENCIA(A1:H5;B8:B13)						
	A	B	C	D	E	F	G	H
1	36	30	47	60	32	35	40	50
2	54	35	45	52	48	58	60	38
3	32	35	56	48	30	55	49	39
4	58	50	65	35	56	47	37	56
5	58	50	47	58	55	39	58	45
6								
7	Clase	L_s	f					
8	30-35	35	=FRECUENCIA(A1:H5;B8:B13)					
9	36-41	41						
10	42-47	47						
11	48-53	53						
12	54-59	59						
13	60-65	65						

d) Presione CTRL+SHIFT+ENTER

J19		={FRECUENCIA(A1:H5;B8:B13)}						
	A	B	C	D	E	F	G	H
1	36	30	47	60	32	35	40	50
2	54	35	45	52	48	58	60	38
3	32	35	56	48	30	55	49	39
4	58	50	65	35	56	47	37	56
5	58	50	47	58	55	39	58	45
6								
7	Clase	L_s	f					
8	30-35	35	8					
9	36-41	41	6					
10	42-47	47	5					
11	48-53	53	7					
12	54-59	59	11					
13	60-65	65	3					

e) Los cálculos de la marca de clase y de las otras frecuencias empleando Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Clases	f	xm	fr	fa	f%	fra	fra%							
2	30	35	8	32,5	=(A2+B2)/2	0,2	=C2/SC\$8	8	=C2	20	=F2*100	0,2	=F2	20	=L2*100
3	36	41	6	38,5	=(A3+B3)/2	0,15	=C3/SC\$8	14	=H2+C3	15	=F3*100	0,35	=L2+F3	35	=L3*100
4	42	47	5	44,5	=(A4+B4)/2	0,125	=C4/SC\$8	19	=H3+C4	12,5	=F4*100	0,475	=L3+F4	47,5	=L4*100
5	48	53	7	50,5	=(A5+B5)/2	0,175	=C5/SC\$8	26	=H4+C5	17,5	=F5*100	0,65	=L4+F5	65	=L5*100
6	54	59	11	56,5	=(A6+B6)/2	0,275	=C6/SC\$8	37	=H5+C6	27,5	=F6*100	0,925	=L5+F6	92,5	=L6*100
7	60	65	3	62,5	=(A7+B7)/2	0,075	=C7/SC\$8	40	=H6+C7	7,5	=F7*100	1	=L6+F7	100	=L7*100
8	Total		40			1				100					

TAREA DE INTERAPRENDIZAJE N° 3

- 1) Defina con sus propias palabras lo que entiende por distribución de frecuencias.
- 2) Realice un organizador gráfico sobre los tipos de frecuencias
- 3) Dadas las siguientes calificaciones evaluadas sobre 10 obtenidas de 40 estudiantes en la asignatura de Estadística:

9	7	8	7	6	9	7	8
8	9	7	8	8	9	8	7
7	10	6	9	9	9	6	8
6	5	10	5	5	10	9	8
5	5	8	8	7	8	9	7

- 3.1) Terminar de llenar de manera manual la siguiente tabla:

Calificación	f	fr	fa	$f\%$	fra	$fra\%$
5		0,125		12,5		12,5
	4		9		0,225	
7		0,2		20		42,5
	11		28		0,7	70
9		0,225		22,5		
	3		40		1	100
Total	40	1		100		

- 3.2) Realice la interpretación de un valor cualquiera de f , fr , fa , $f\%$, fra y $fra\%$ de la tabla anterior.
- 3.3) Calcular la frecuencia absoluta empleando Excel.
- 3.4) Calcular fr , fa , $f\%$, fra y $fra\%$ empleando Excel.
- 4) Crear y resolver un ejercicio similar al anterior sobre cualquier tema de su preferencia.
- 5) Consulte sobre biografía de Herberth Arthur Sturges y elabore un organizador gráfico de la misma.
- 6) A 40 docentes que laboran en la Universidad UTN se les preguntó su edad, obteniéndose los siguientes resultados:

32	50	52	40	45	38	58	58
54	44	48	38	49	55	58	48
42	55	46	38	54	44	47	43
48	40	57	55	46	57	47	46
48	54	57	48	51	59	54	55

- 6.1) Calcule el rango, número de intervalos y el ancho de la clase de manera manual y empleando Excel.

$$R = 27; n_i = 6; i = 5$$

- 6.2) Calcule el nuevo rango

30

- 6.3) Calcule los nuevos $x_{m\acute{a}x}$ y $x_{m\acute{i}n}$

61 y 31 ó 60 y 30

6.4) Forme los intervalos de clase comenzando por $x_{min} = 31$. Luego realice el conteo de datos que cae dentro de cada clase de manera manual y empleando Excel, indicando cada uno los procesos seguidos.

Clases	f
31-35	1
36-40	5
41-45	5
46-50	12
51-55	10
56-60	7
Total	40

6.5) Calcule la marca de clase y las demás frecuencias de manera manual y empleando Excel, indicando cada uno los procesos seguidos.

Clases	f	xm	fr	fa	$f\%$	fra	$fra\%$
31-35	1	33	0,025	1	2,5	0,025	2,50
36-40	5	38	0,125	6	12,5	0,150	15,0
41-45	5	43	0,125	11	12,5	0,275	27,5
46-50	12	48	0,300	23	30,0	0,575	57,5
51-55	10	53	0,250	33	25,0	0,825	82,5
56-60	7	58	0,175	40	17,5	1	100
Total	40		1		100		

6.6) Realice la interpretación de un valor cualquiera de f , xm , fr , fa , $f\%$, fra y $fra\%$ de la tabla anterior.
 7) Cree y resuelva un ejercicio similar al anterior sobre cualquier tema de su preferencia.

1.5) GRÁFICOS ESTADÍSTICOS BÁSICOS

Las empresas, industrias, instituciones, etc. emplean diversos gráficos estadísticas para presentar informaciones sobre diversos asuntos relativos a ellas.

Las representaciones gráficas deben conseguir que un simple análisis visual ofrezca la mayor información posible. Según el tipo del carácter que estemos estudiando, usaremos una representación gráfica u otra.

A continuación se presenta los diagramas más empleados:

1.5.1) DIAGRAMAS DE BARRAS

Es un gráfico bidimensional en el que los objetos gráficos elementales son rectángulos de igual base cuya altura sea proporcional a sus frecuencias. Si en el eje horizontal se ubican las etiquetas con los nombres de las categorías, y en el eje vertical la frecuencia absoluta, la relativa o la frecuencia porcentual, toma el nombre de diagrama de barras vertical, y si se intercambian las ubicaciones de las categorías y las frecuencias, toma el nombre de diagrama de barras horizontal.

Ejemplo ilustrativo:

Empleando los datos de la siguiente tabla sobre las siguientes calificaciones obtenidas en una evaluación por 40 estudiantes en la asignatura de Estadística:

Calificación	f
5	4
6	5
7	6
8	11
9	7
10	7
Total	40

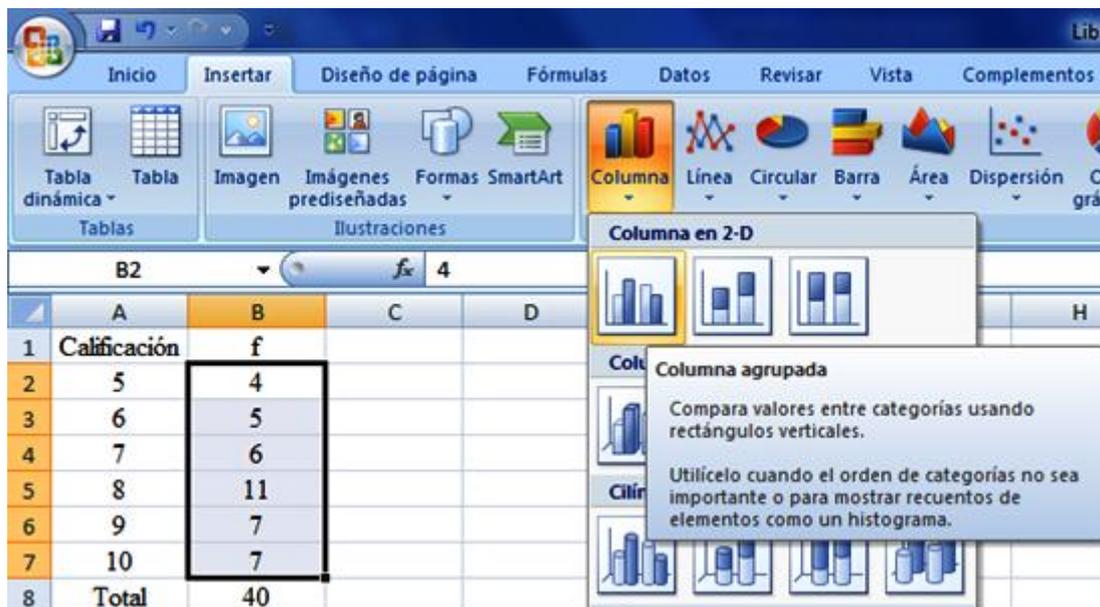
- 1) Elaborar un diagrama de barras verticales en 2 dimensiones (2D) y 3 dimensiones (3D).
- 2) Elaborar un diagrama de barras horizontales en 2 dimensiones (2D) y 3 dimensiones (3D).

Solución:

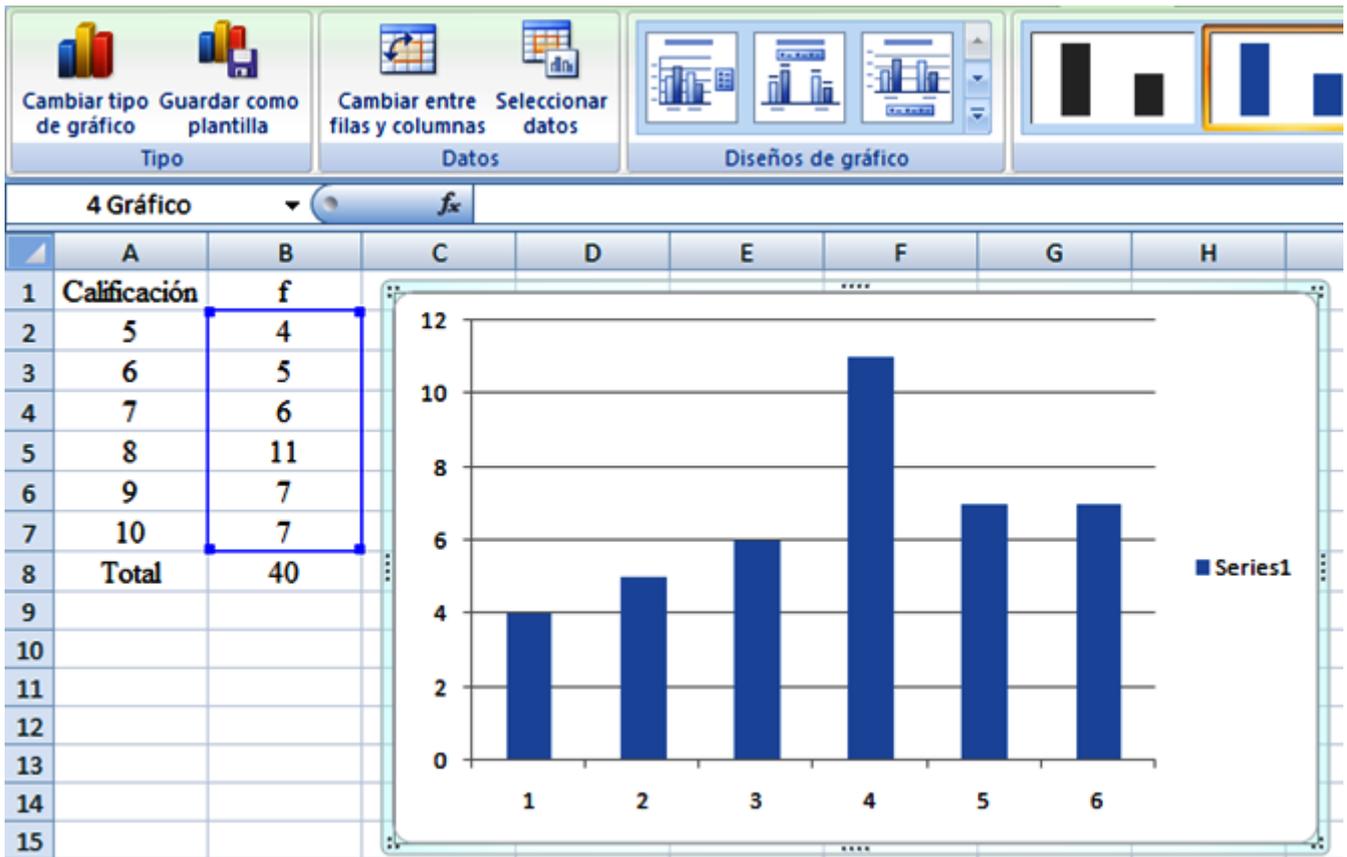
En Excel se elabora de la siguiente manera:

- 1) Barras verticales

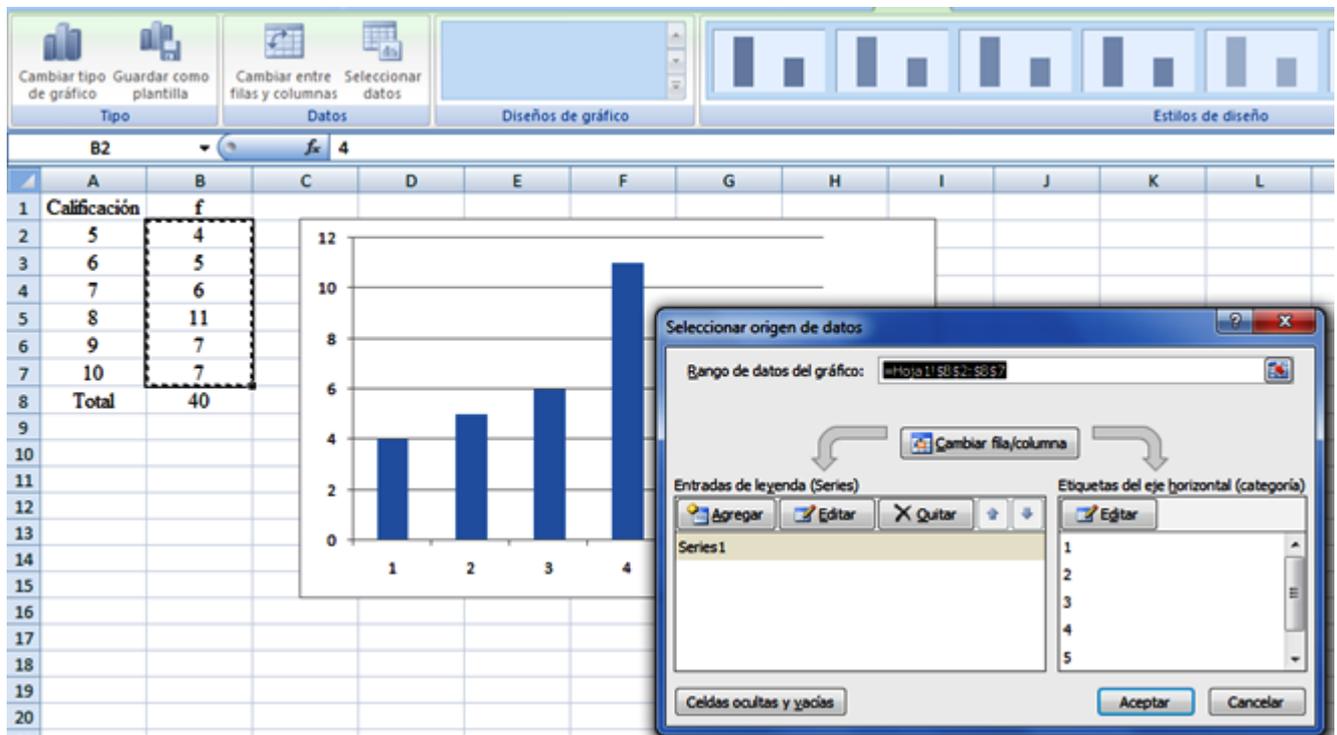
a) Seleccionar las celdas B2:B7 y luego clic en Insertar Columna.



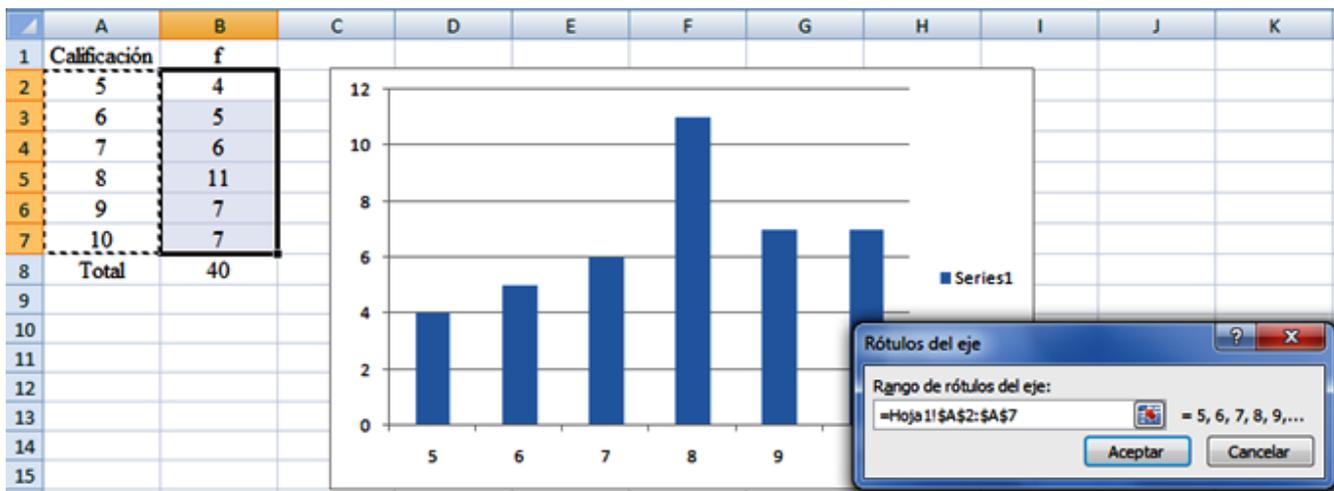
- b) Hacer clic en la primera opción de Columna en 2-D



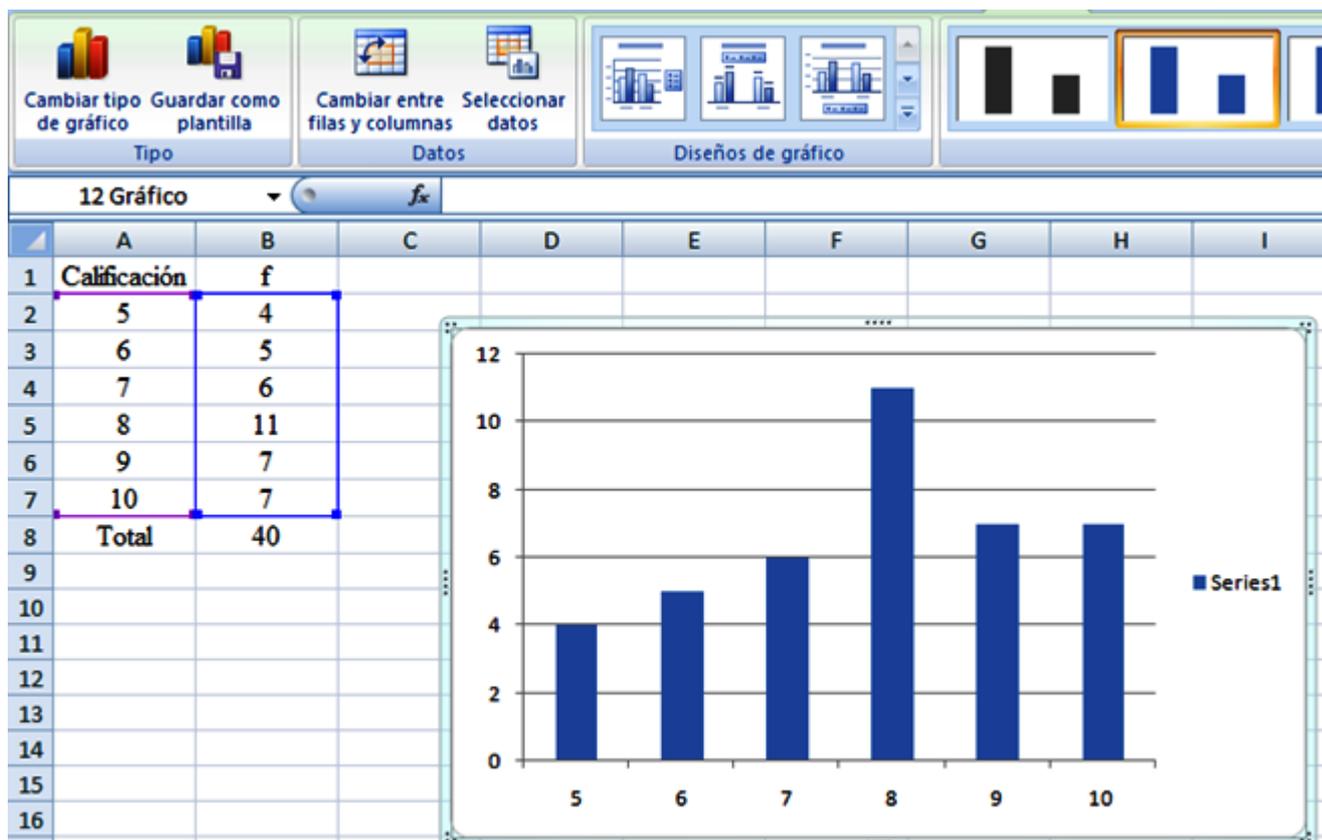
c) Clic en Seleccionar datos para que aparezca la ventana origen de datos.



d) En Etiquetas de eje horizontal (categoría) seleccionar Editar para que aparezca la ventana Rótulos del eje. En rango de rótulos del eje seleccionar A2:A7. Clic en Aceptar



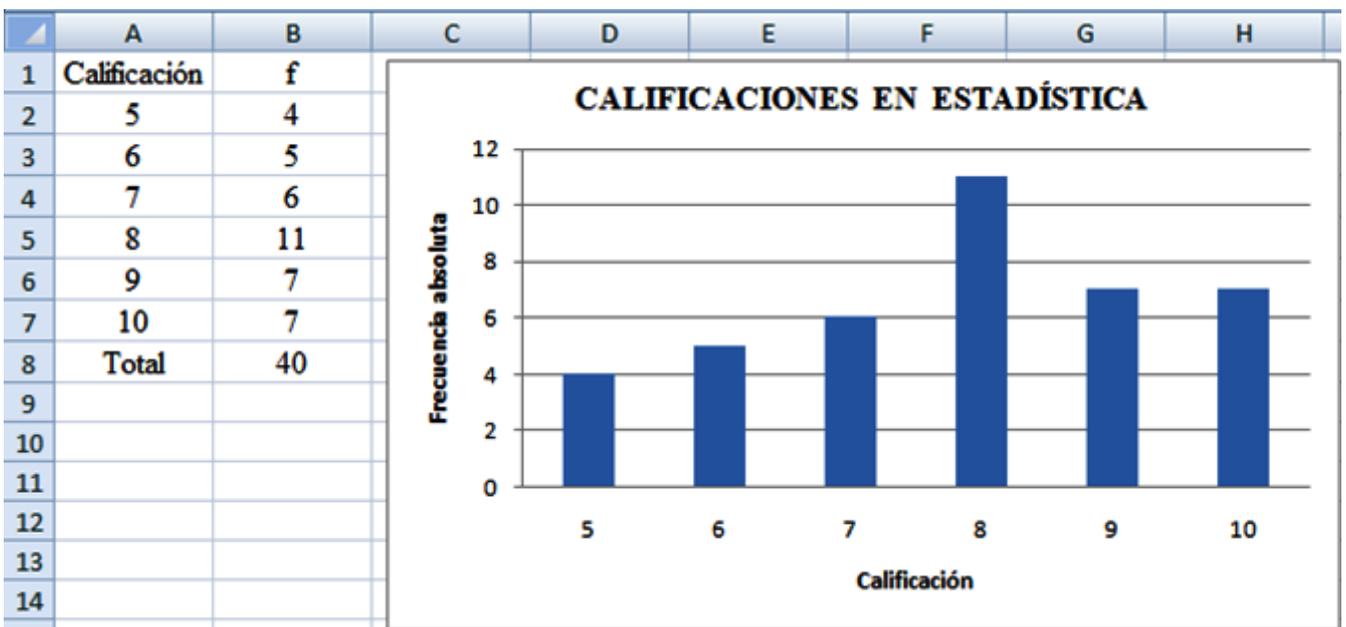
e) Clic en Aceptar



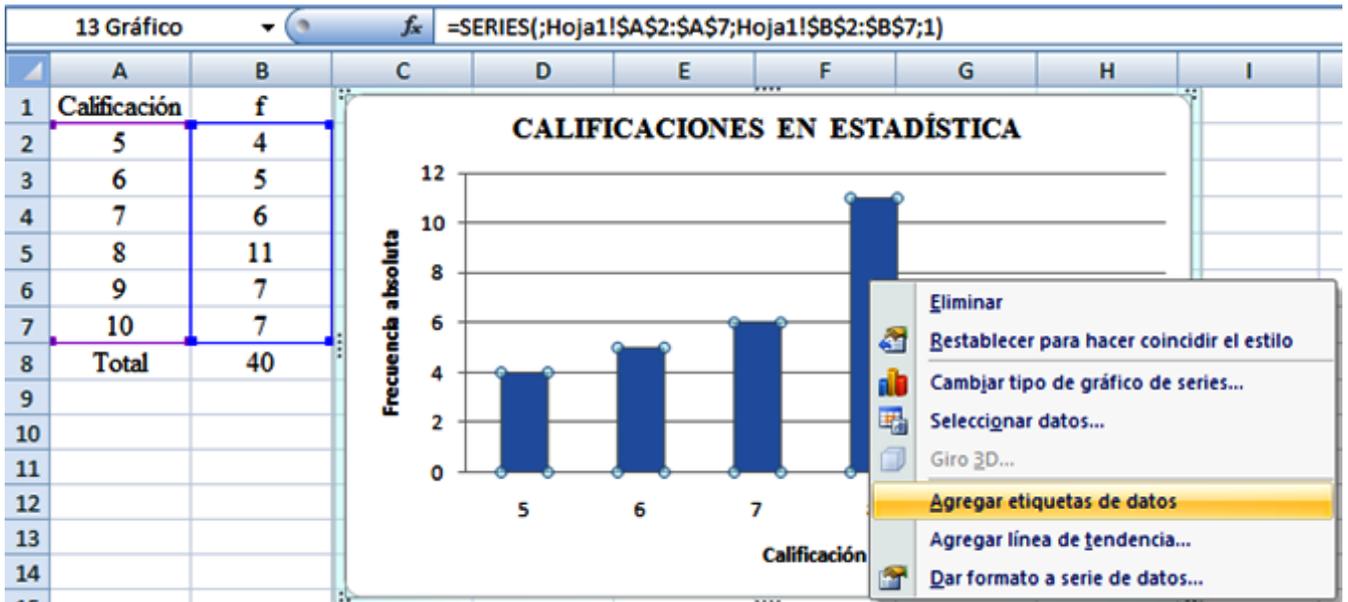
f) En Diseños de gráfico seleccionar Diseño 9.



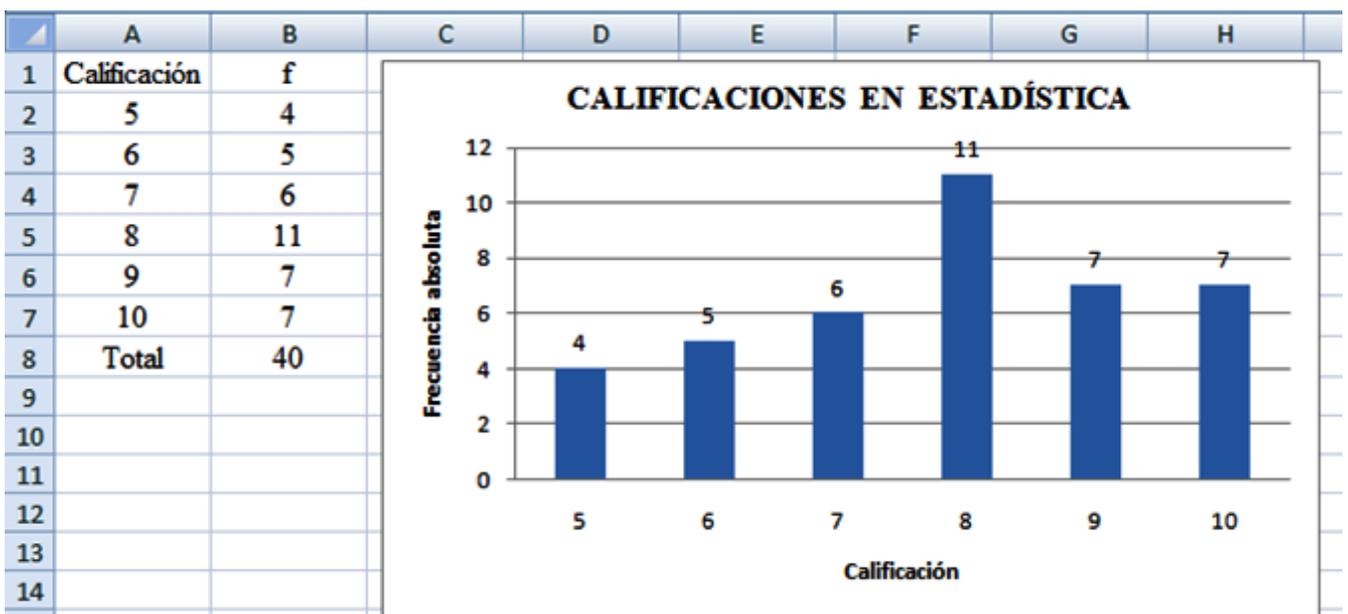
g) Borrar Series 1. En título del gráfico escribir Calificaciones en Estadística. En título del eje vertical escribir frecuencia absoluta. En título del eje horizontal escribir calificación.



h) Clic derecho en el gráfico.



i) Clic en Agregar etiquetas de datos y queda elaborado el diagrama de barras verticales en 2D

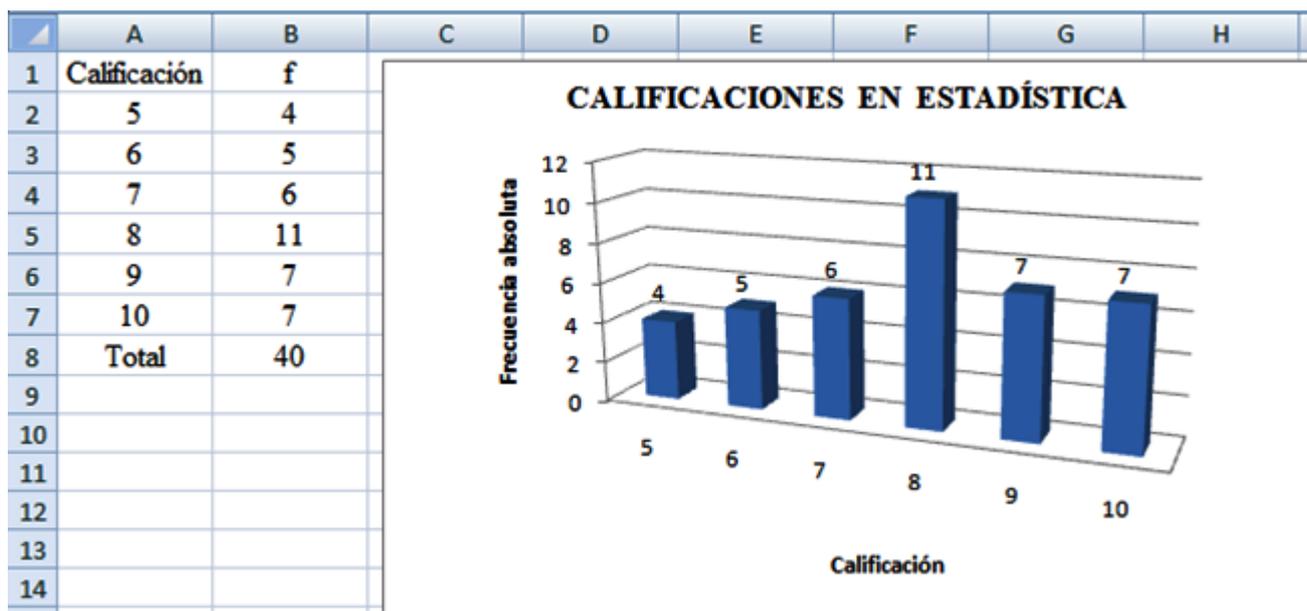


j) Para elaborar el diagrama de barras verticales en 3D, hacer clic en Columna.

The screenshot shows the Microsoft Excel interface with the 'Columna' menu open. The menu options include 'Columna en 2-D', 'Columna en 3-D', 'Cilindro', 'Cónico', and 'Pirámide'. The 'Columna en 3-D' option is highlighted, and a tooltip for 'Columna agrupada 3D' is visible, stating: 'Compara valores entre categorías y muestra columnas agrupadas en formato 3D.' The background shows a 2D bar chart with the following data:

Calificación	f
5	4
6	5
7	6
8	11
9	7
10	7
Total	40

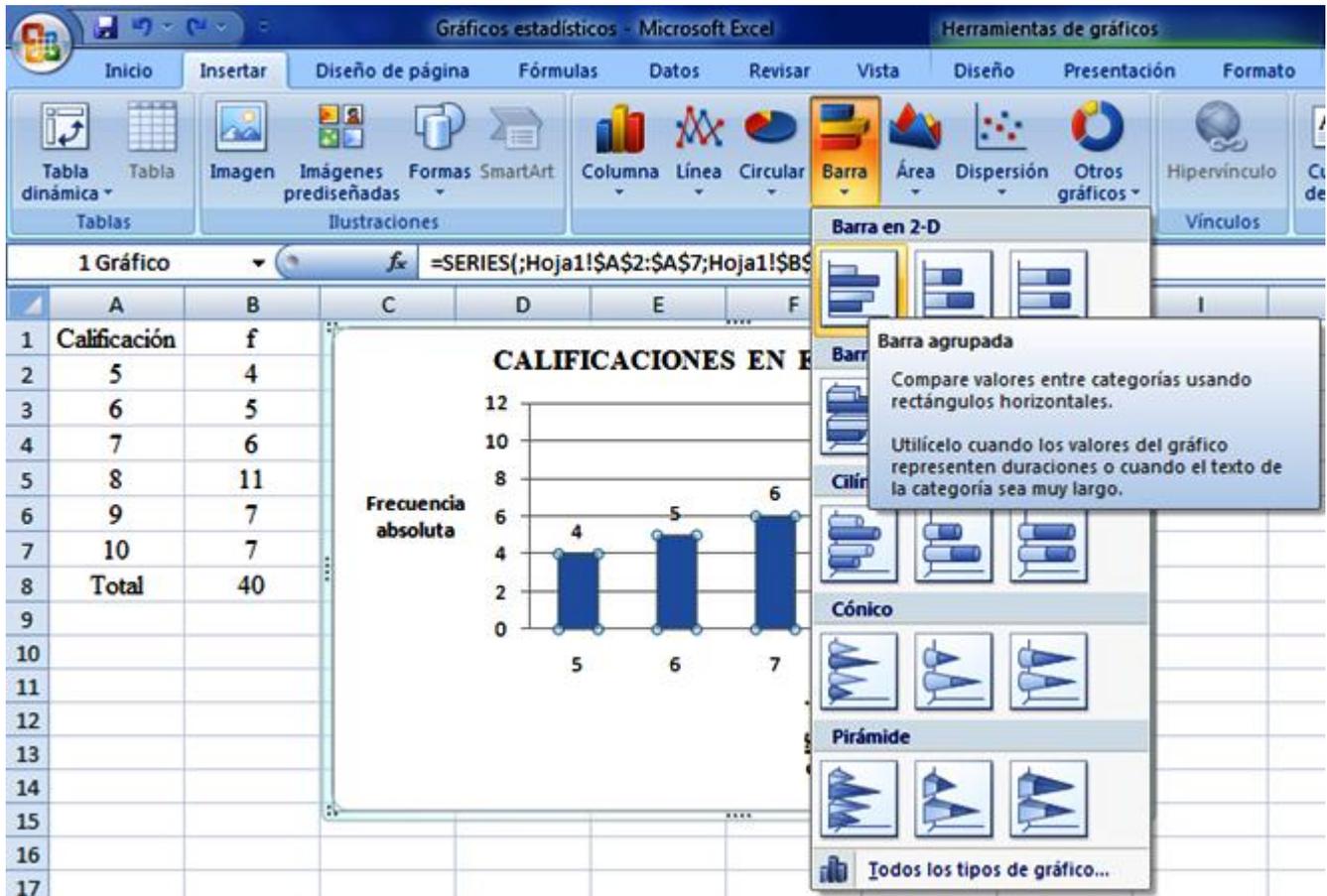
k) Clic en la primera opción Columna en 3-D



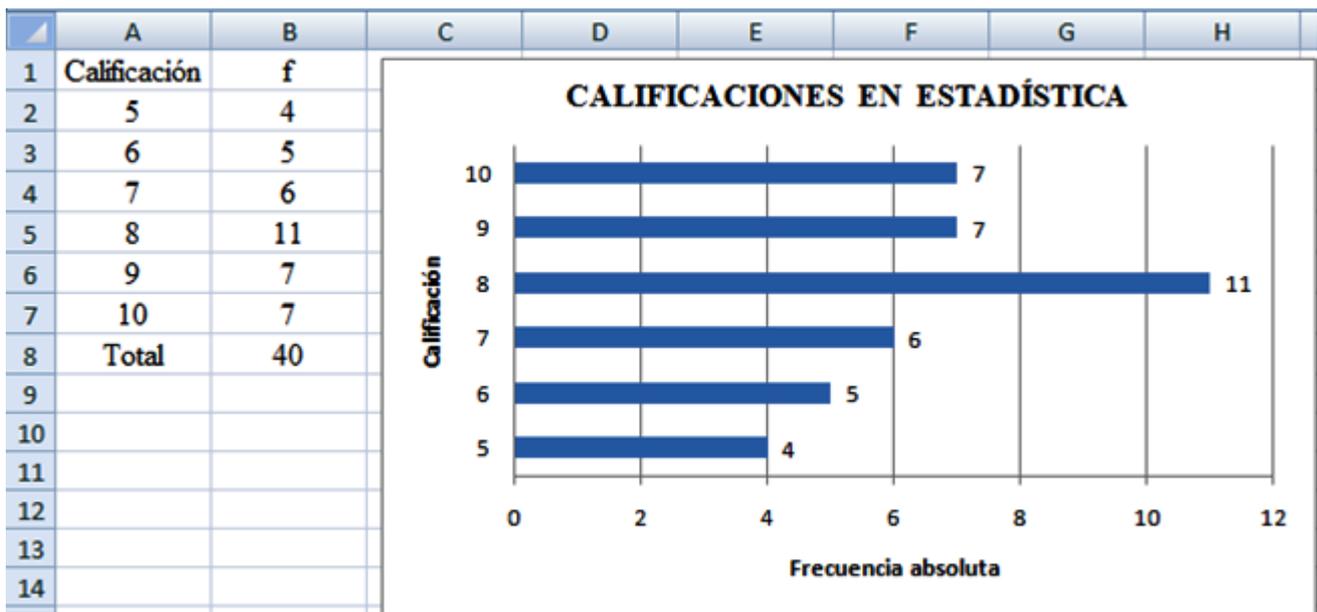
2) Barras horizontales

Utilizando el gráfico de barras verticales en 2D

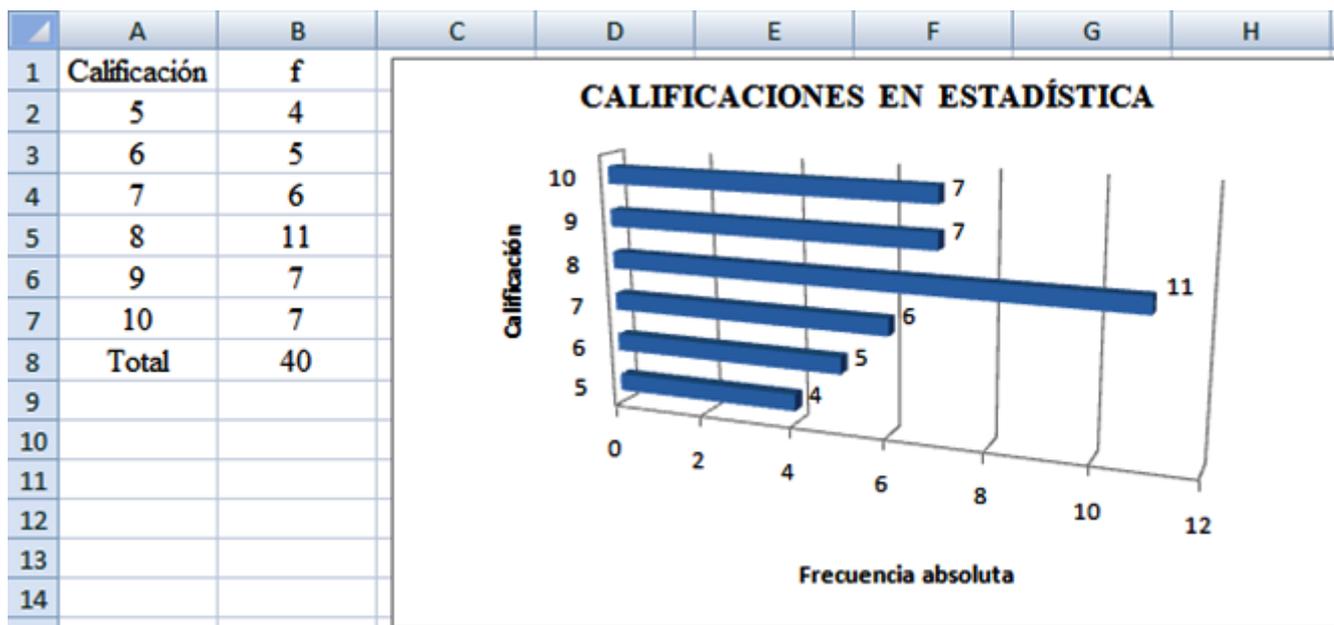
a) Clic en Barra



b) Clic en la primera opción y queda elaborado el diagrama de barras horizontales en 2D



c) A partir del gráfico anterior, para elaborar el diagrama de barras horizontales en 3D, hacer clic en Barras. Escoger la primera opción de Barra en 3-D



1.5.2) HISTOGRAMAS

Se utiliza para datos agrupados en intervalos de clase, representando en el eje horizontal los intervalos de clase o la marca de clase, y en el eje vertical se elabora rectángulos contiguos de base el ancho del intervalo y de altura proporcional a las frecuencias representadas.

Ejemplo ilustrativo

A 40 docentes que laboran en la Universidad UTN se les preguntó su edad, obteniéndose los siguientes resultados:

Clases	f	xm	fr	fa	$f\%$	fra	$fra\%$
31-35	1	33	0,025	1	2,5	0,025	2,50
36-40	5	38	0,125	6	12,5	0,150	15,0
41-45	5	43	0,125	11	12,5	0,275	27,5
46-50	12	48	0,300	23	30,0	0,575	57,5
51-55	10	53	0,250	33	25,0	0,825	82,5
56-60	7	58	0,175	40	17,5	1	100
Total	40		1		100		

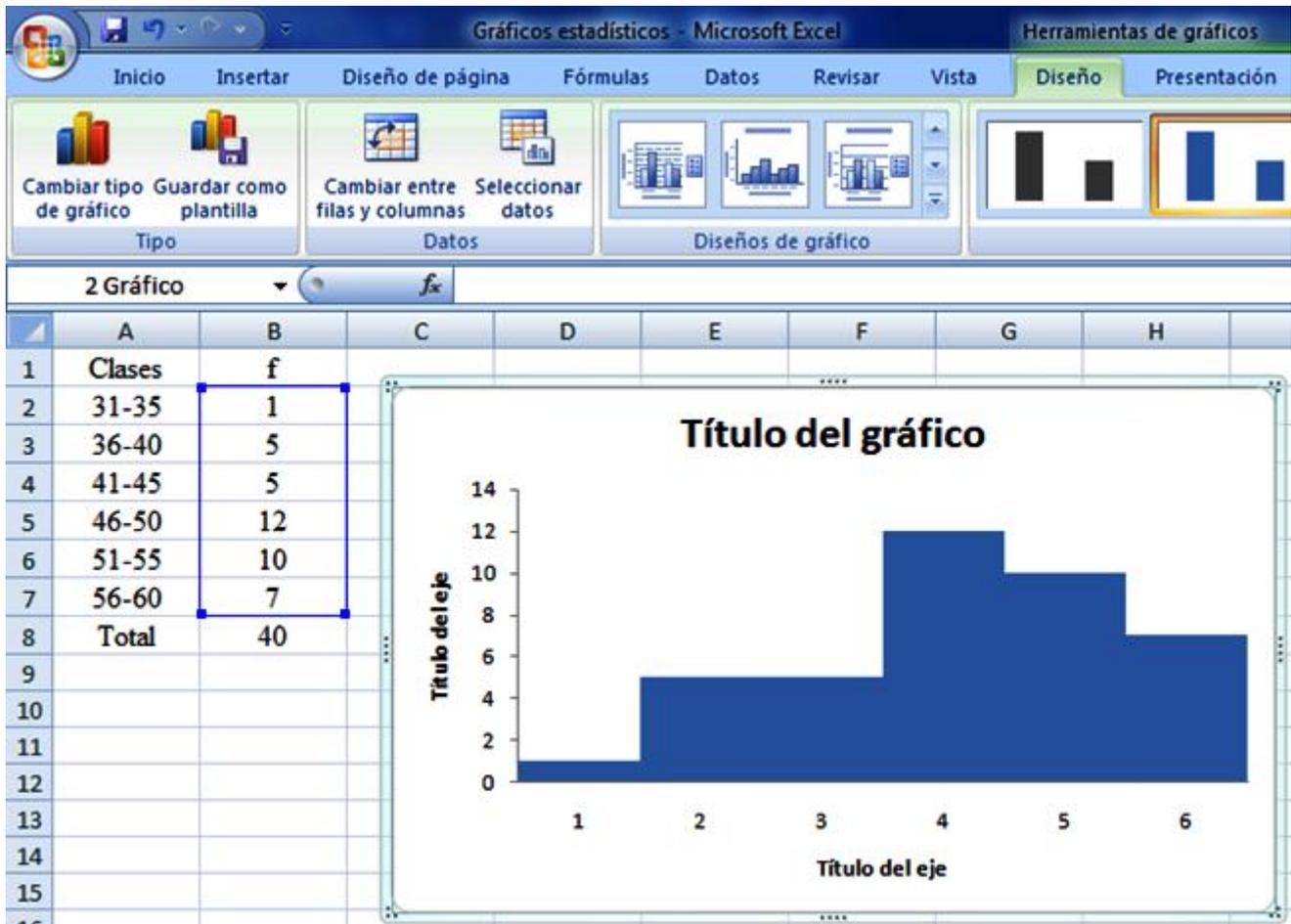
- 1) Elaborar un histograma para f
- 2) Elaborar un histograma para $f\%$
- 3) Elaborar un histograma para $fra\%$

Solución:

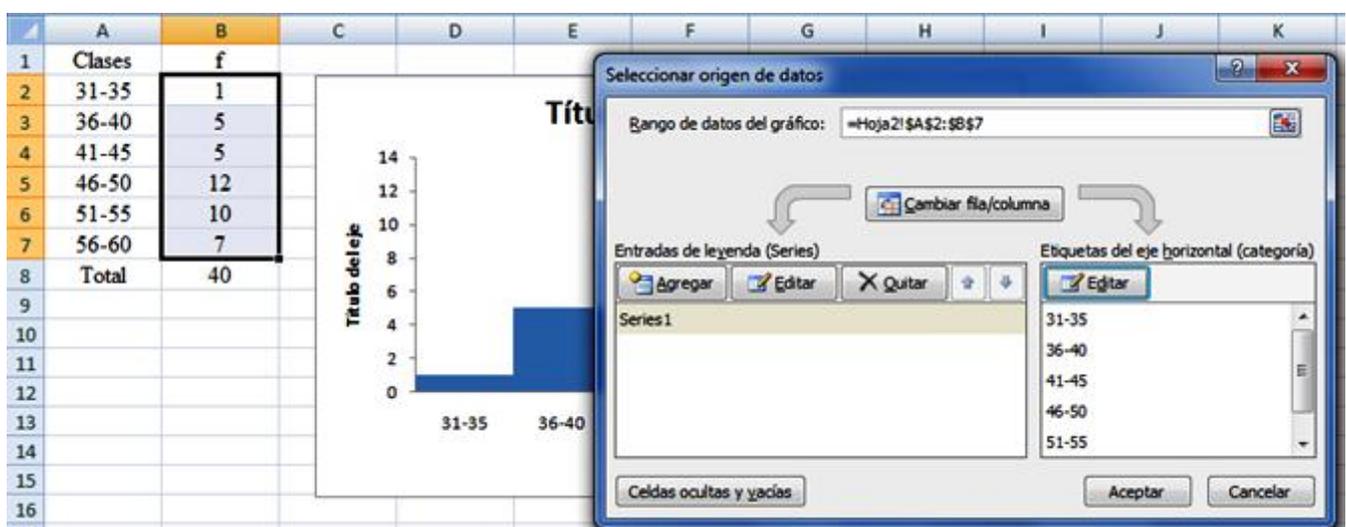
En Excel se realiza de la siguiente manera:

1) Histograma para f

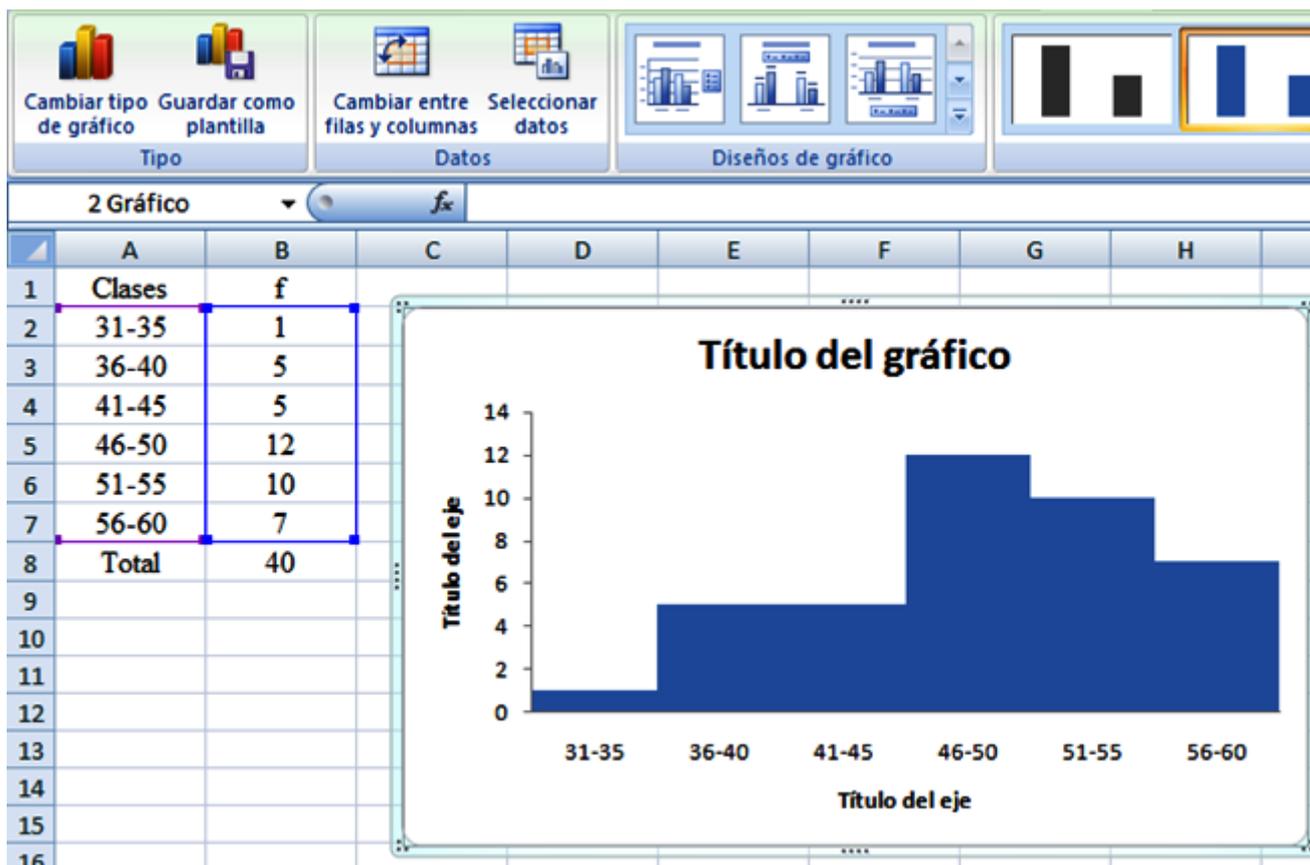
a) Seleccionar B2:B7. Insertar Columna. En diseños de gráfico seleccionar Diseño 8.



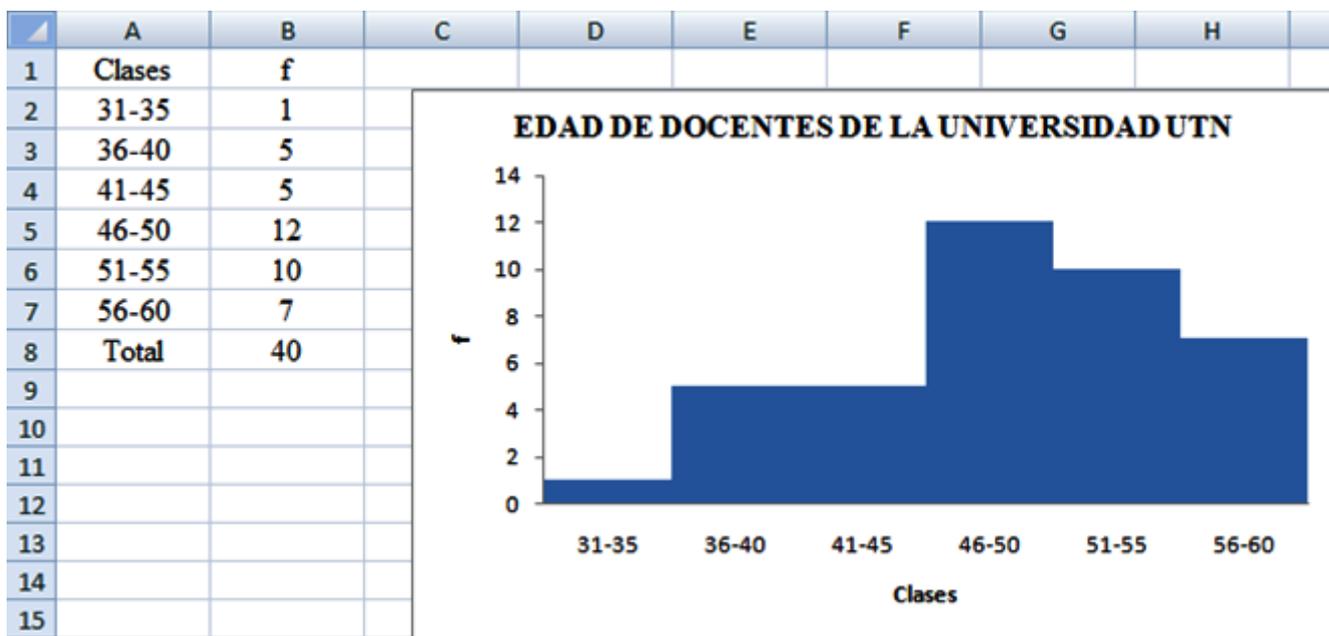
b) Clic en Seleccionar datos para que aparezca la ventana origen de datos. En Etiquetas de eje horizontal (categoría) seleccionar Editar para que aparezca la ventana Rótulos del eje. En rango de rótulos del eje seleccionar A2:A7. Clic en Aceptar.



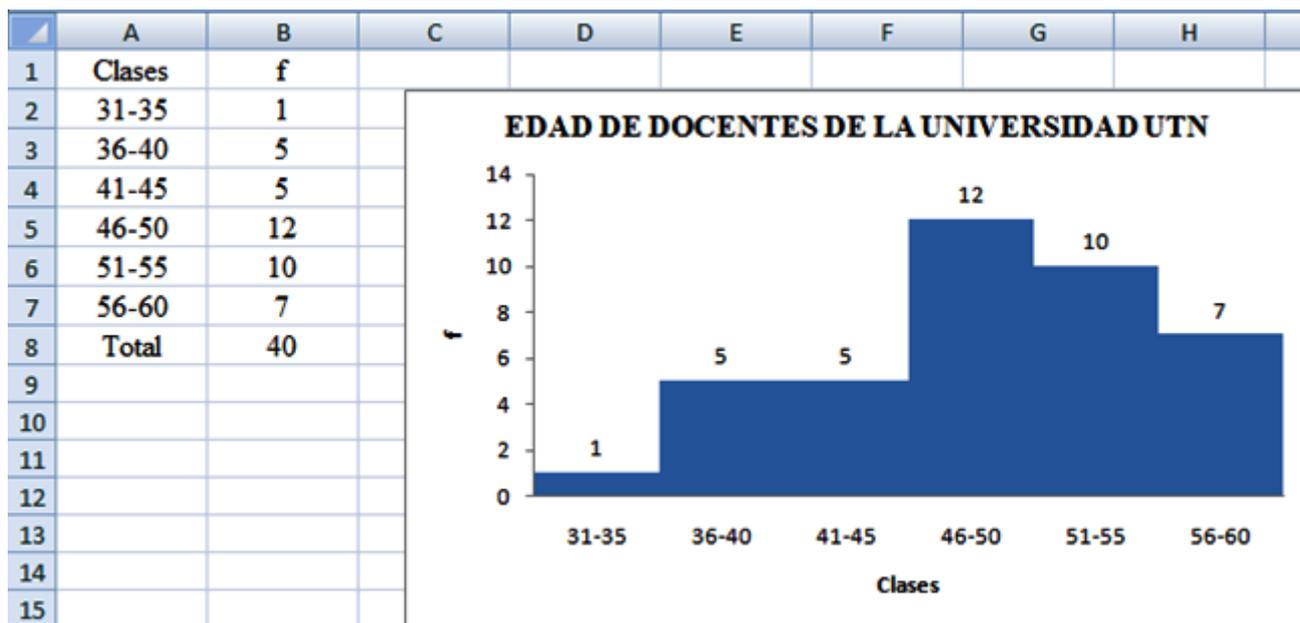
c) Clic en Aceptar



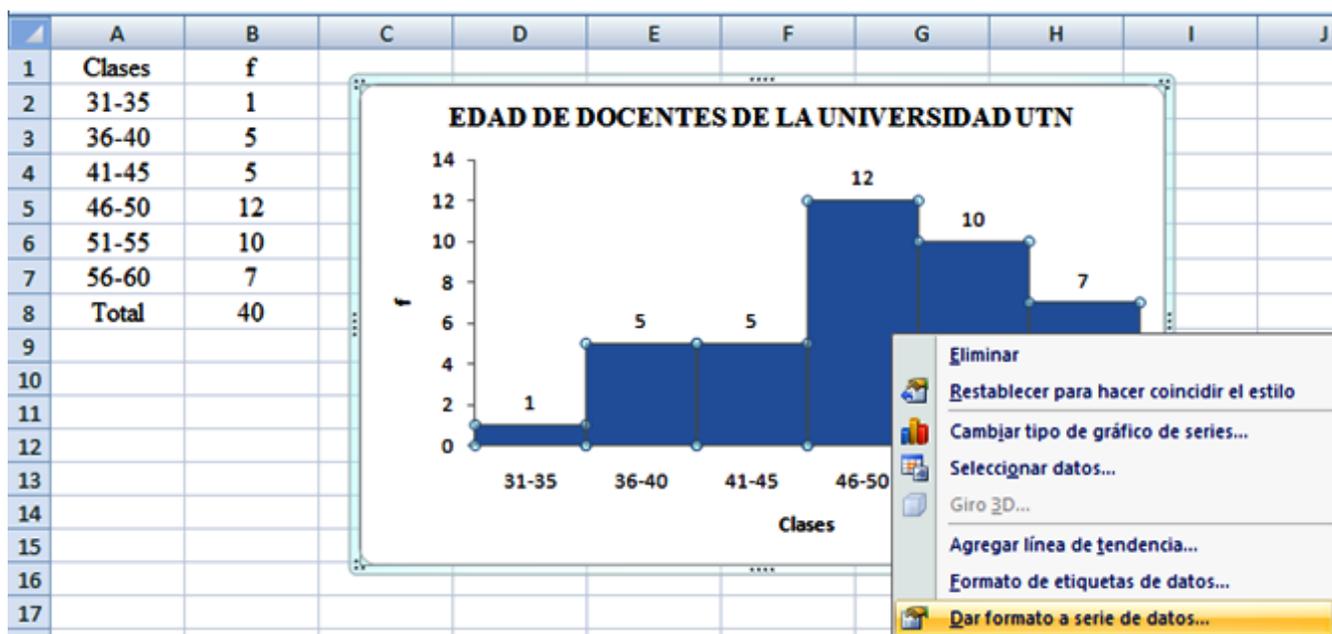
d) Escribir Edad de Docentes de la Universidad UTN en título del gráfico, Clases en Título del eje horizontal y f en Título del eje vertical.



e) Clic derecho en el gráfico. Clic en Agregar etiquetas de datos.



f) Clic derecho en el gráfico.



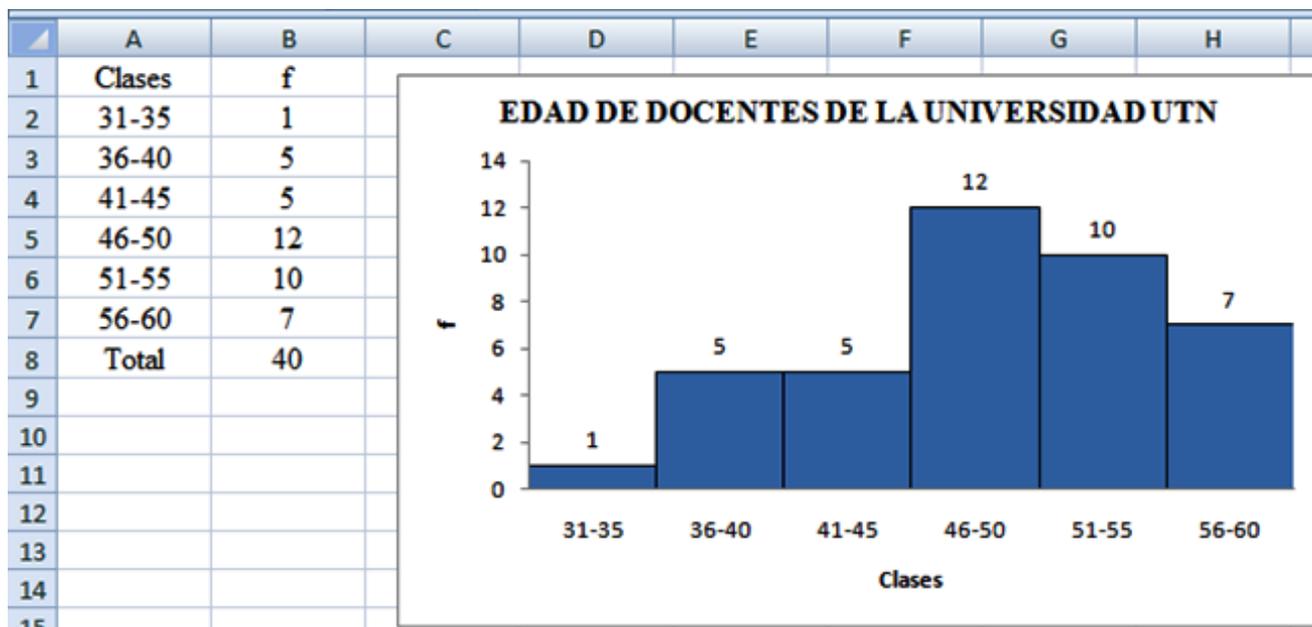
g) Clic en dar formato a serie de datos. Seleccionar línea sólida en Color del borde. Escoger un color diferente al de las barras.

The screenshot shows an Excel spreadsheet with a bar chart titled "EDAD DE DOCENTES DE LA U...". The chart displays the frequency of teachers in different age groups. The data is as follows:

Clases	f
31-35	1
36-40	5
41-45	5
46-50	12
51-55	10
56-60	7
Total	40

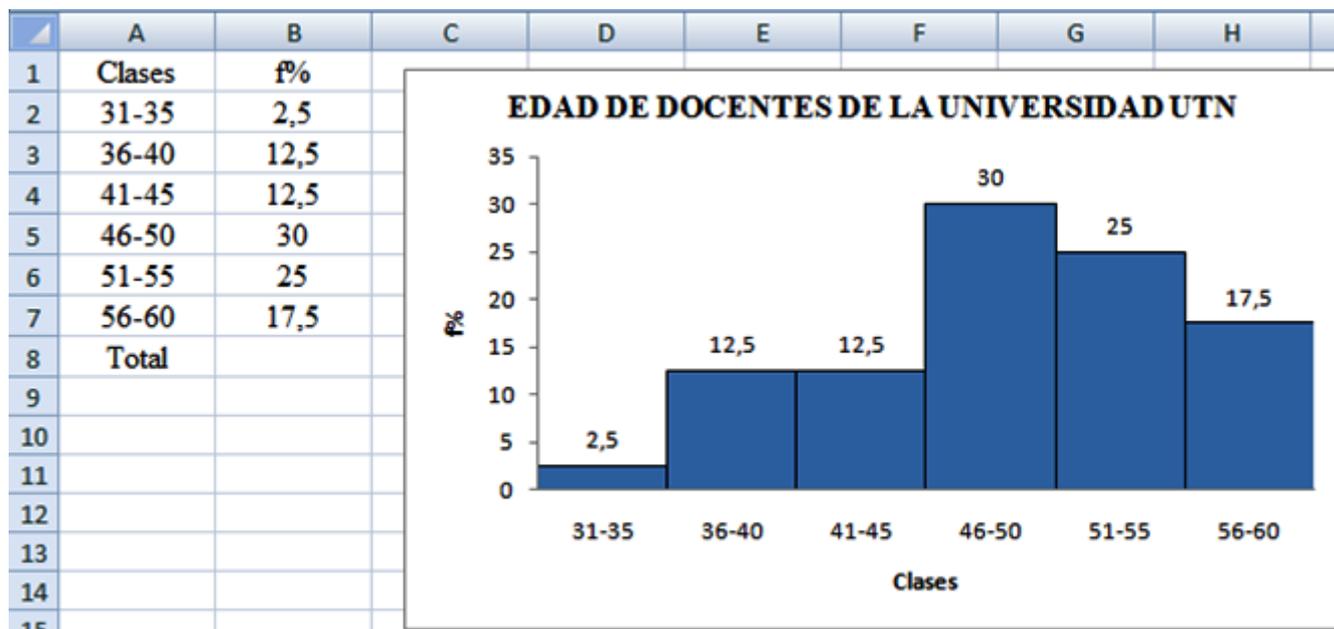
The 'Formato de serie de datos' dialog box is open, showing the 'Color del borde' section. The 'Línea sólida' option is selected. A color palette is visible, showing various theme and standard colors.

h) Clic en Cerrar



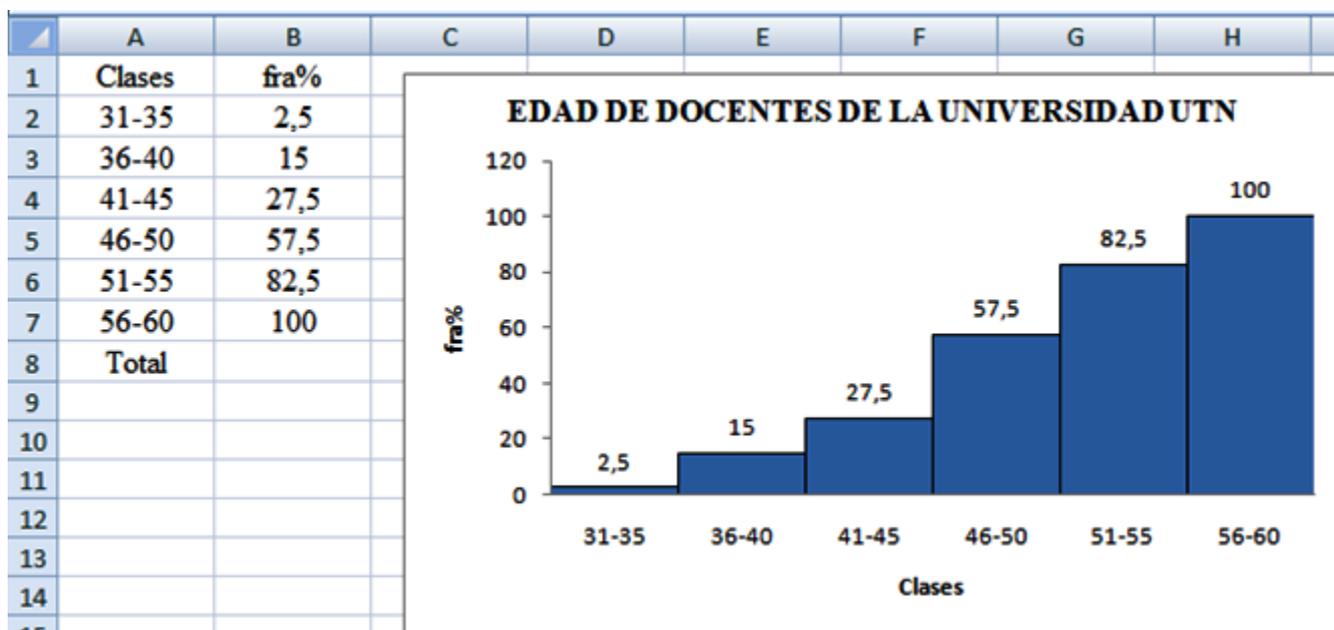
2) Histograma para f%

Utilizando el gráfico anterior, borrar la columna de fa y escribir la columna de f%. En eje vertical escribir f%.



3) Histograma para fra%

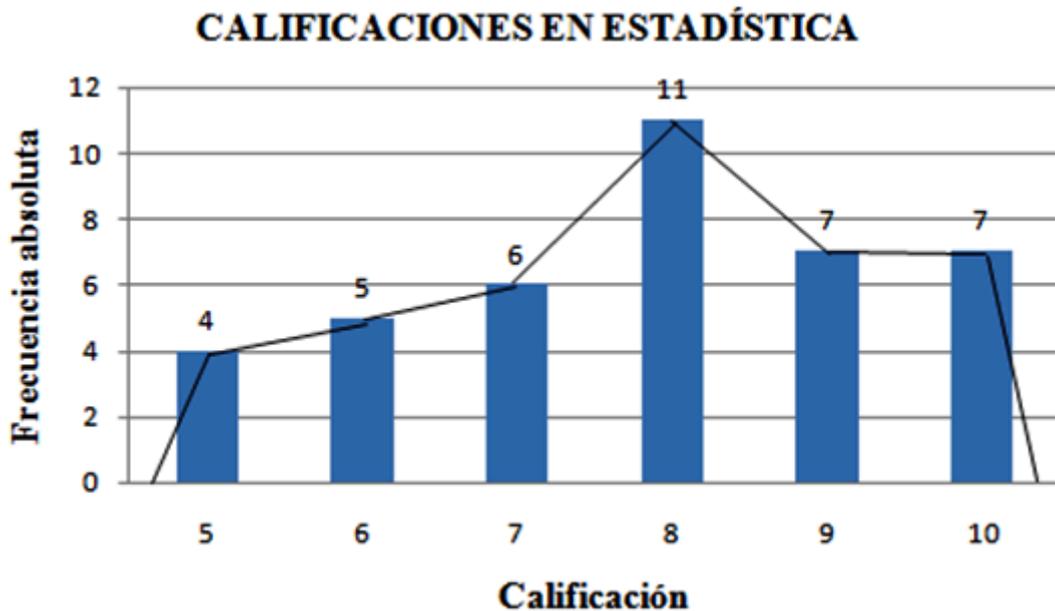
Utilizando el gráfico anterior, borrar la columna de f% y escribir la columna de fra%. En eje vertical escribir fra%.



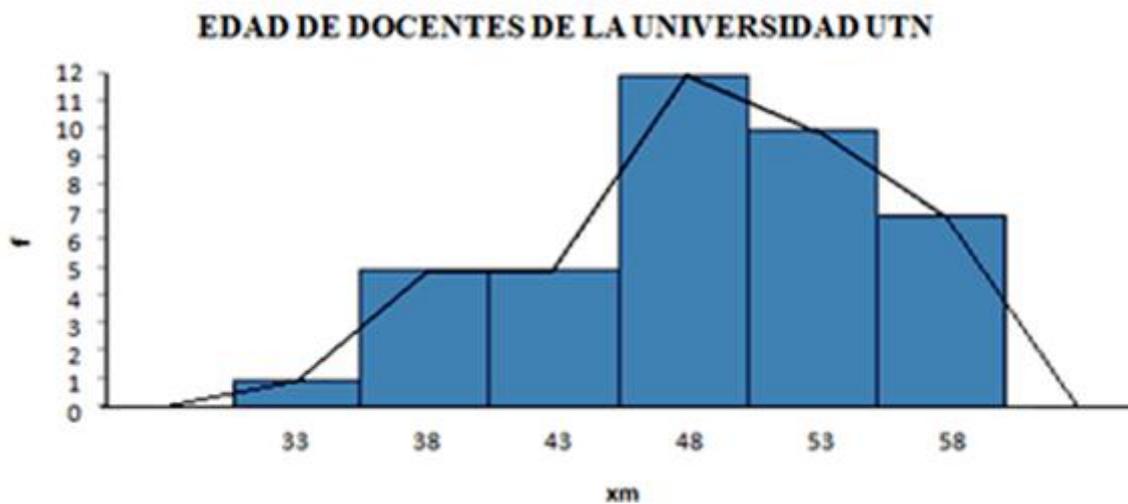
1.5.3) POLÍGONO DE FRECUENCIAS

Son gráficos lineales que se realizan uniendo:

a) Los puntos medios de las bases superiores de los rectángulos en un diagrama de barras.

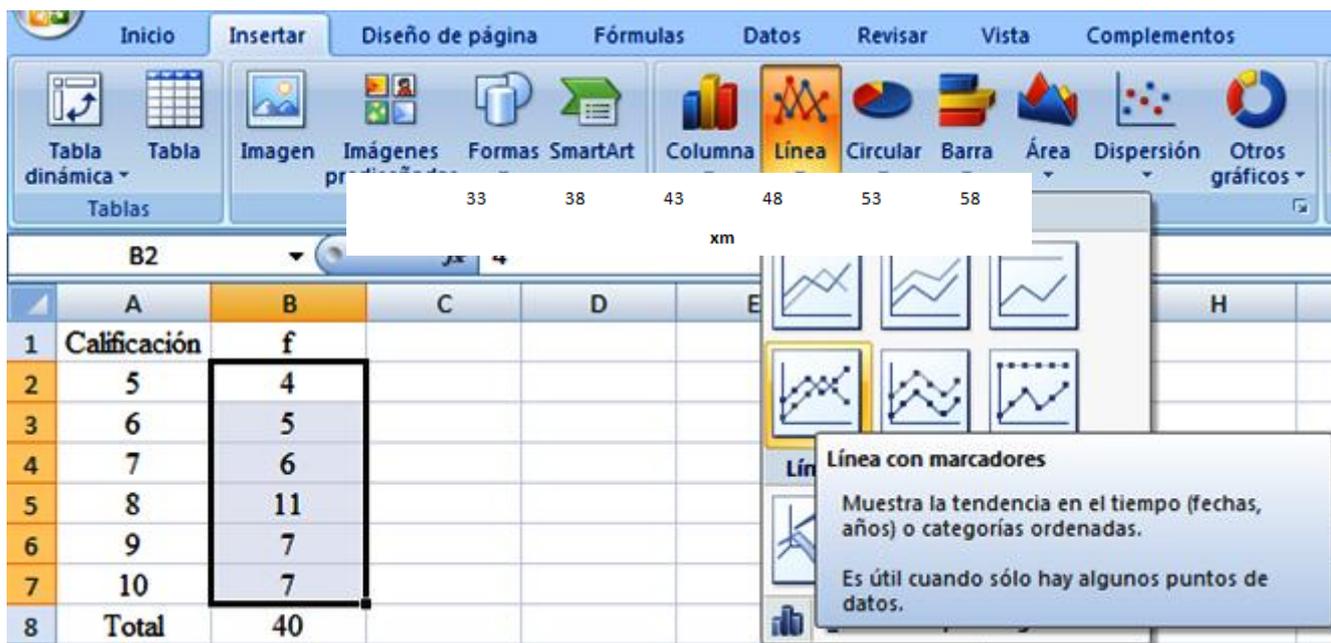


b) Los puntos medios (marcas de clase) de las bases superiores en el histograma.

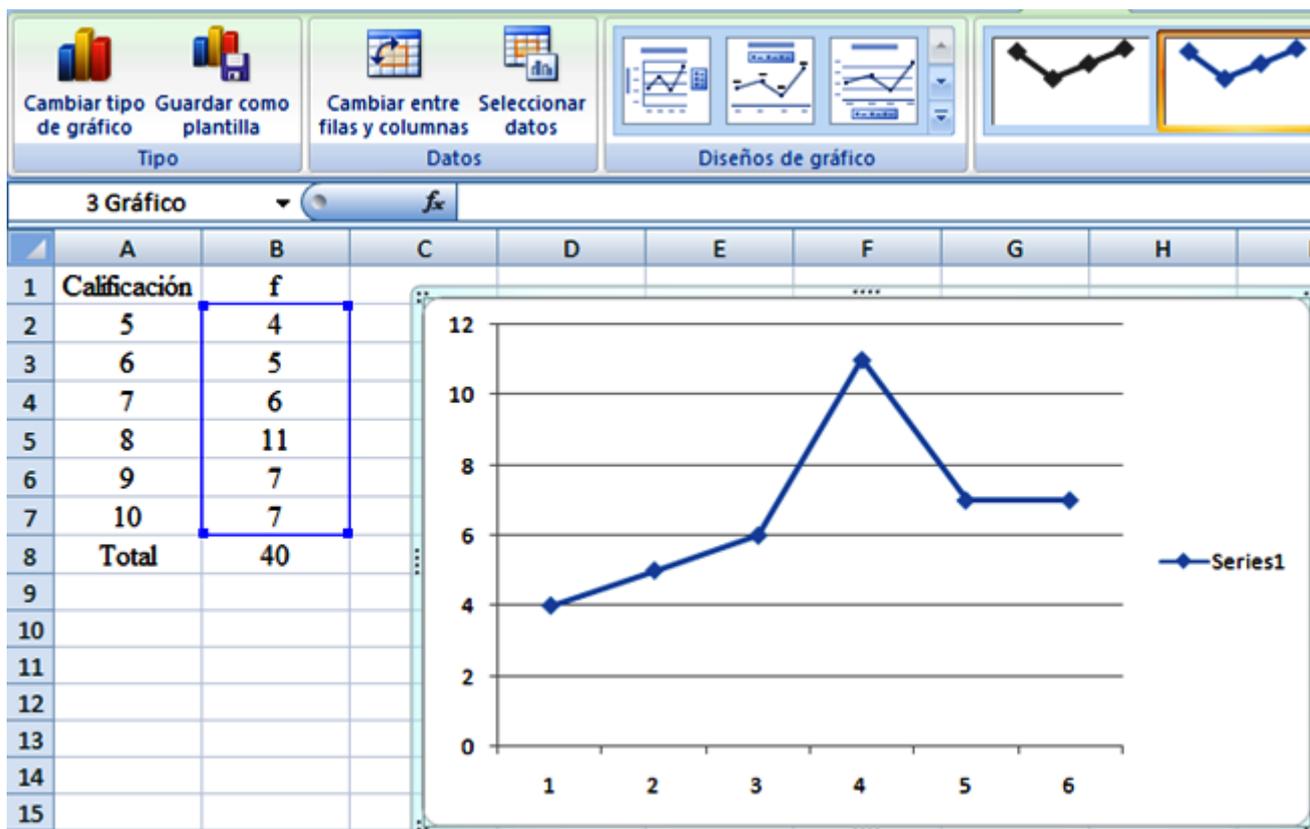


Sin las barras, empleando Excel se elabora de la siguiente manera:

a) Seleccionar A2:A7. Clic en Insertar Línea.



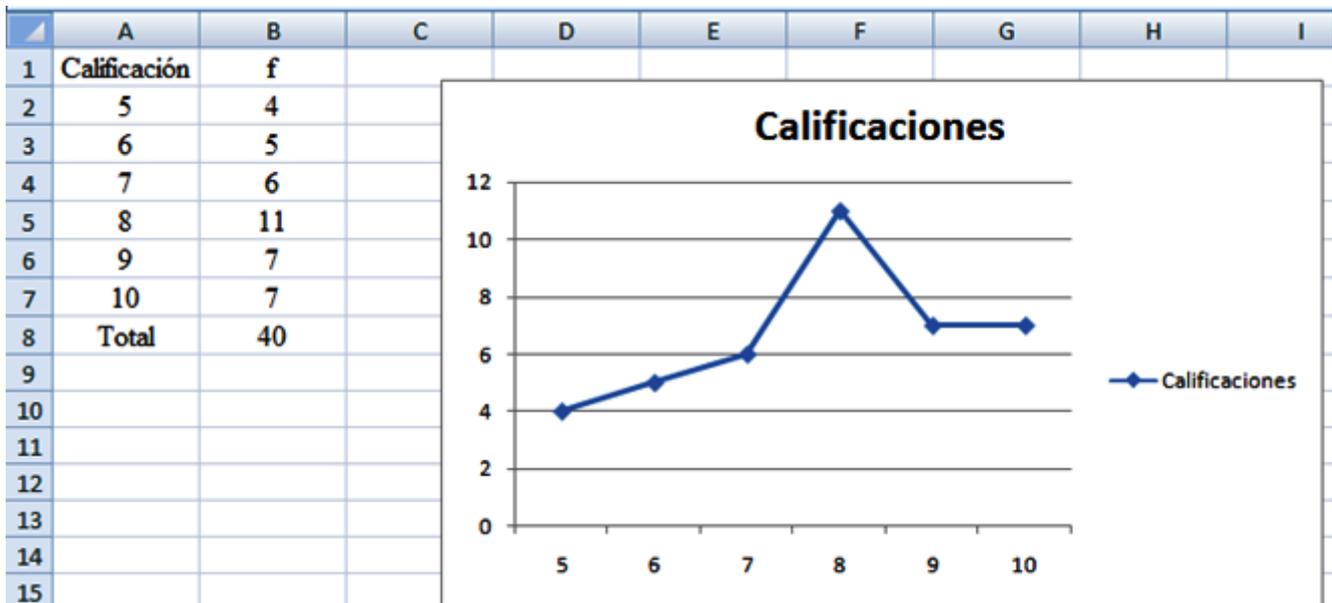
b) Escoger la tercera opción de Línea 2D



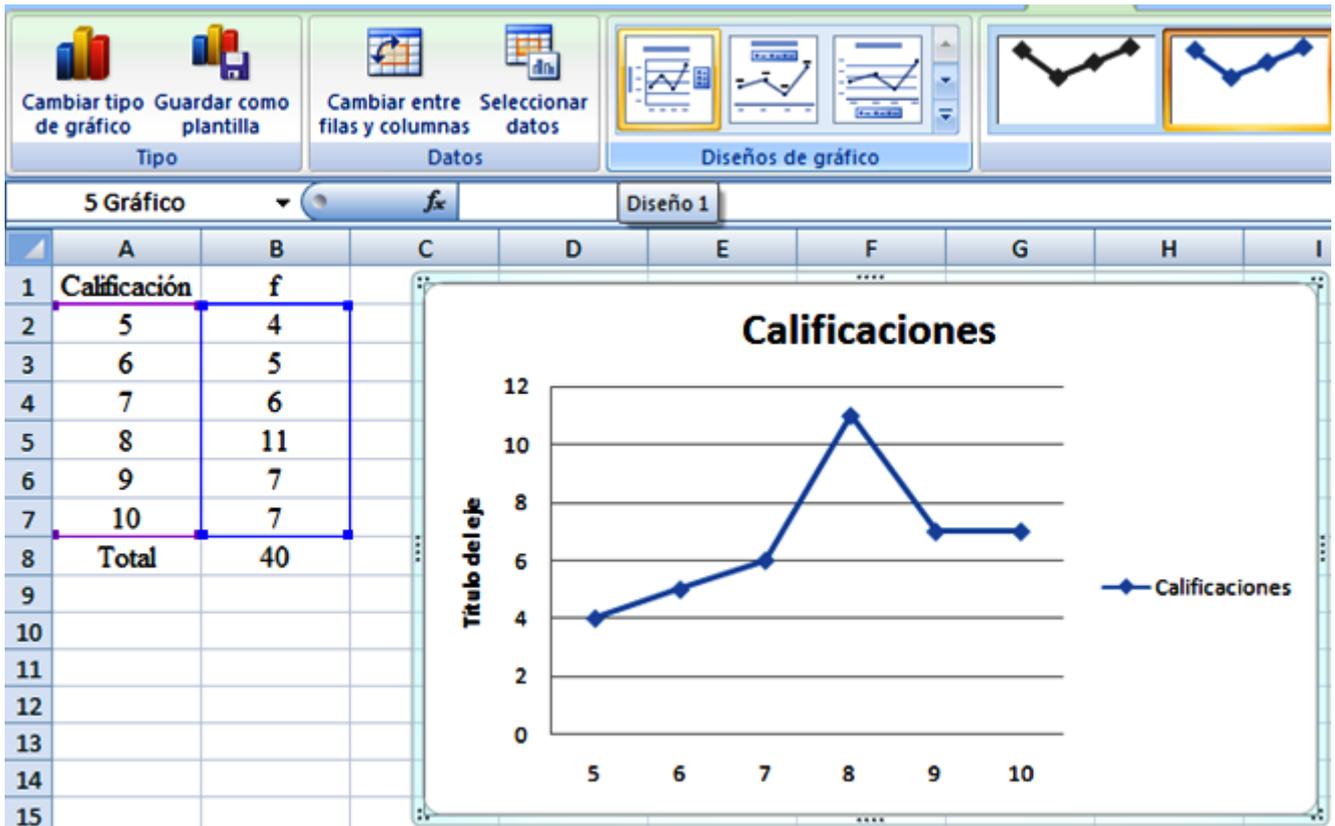
c) Clic en Seleccionar datos para que aparezca la ventana Seleccionar origen de datos. En entradas de leyenda (Series), clic en Editar, luego en nombre de serie escribir Calificaciones y clic en Aceptar. En Etiquetas de eje horizontal (categoría), clic en Editar, luego en rango de rótulos de eje seleccionar A2:A7 y clic en Aceptar.

Calificación	f
5	4
6	5
7	6
8	11
9	7
10	7
Total	40

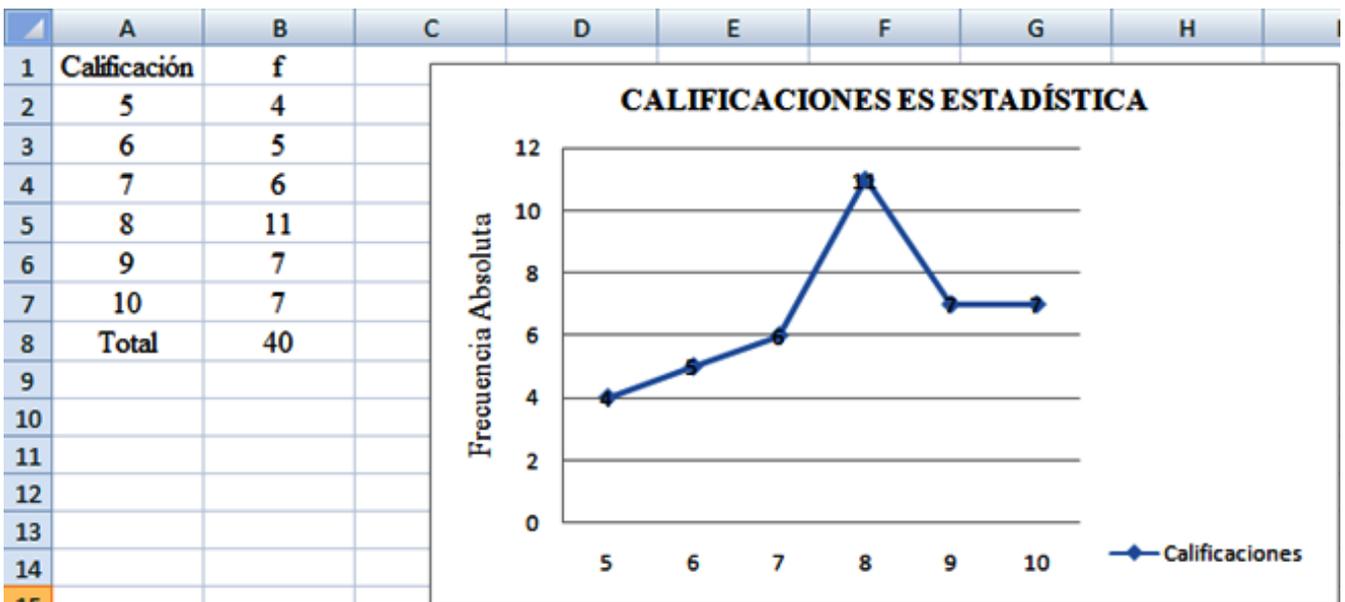
d) Clic en Aceptar



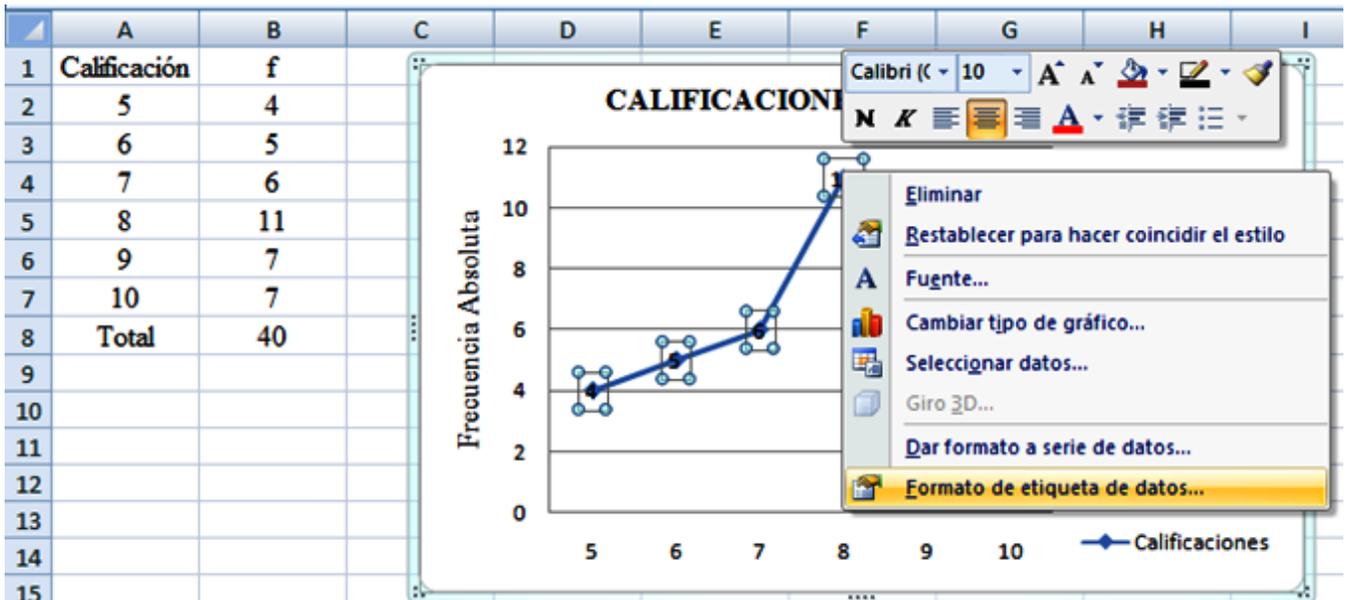
e) Seleccionar Diseño 1 en diseño de gráfico



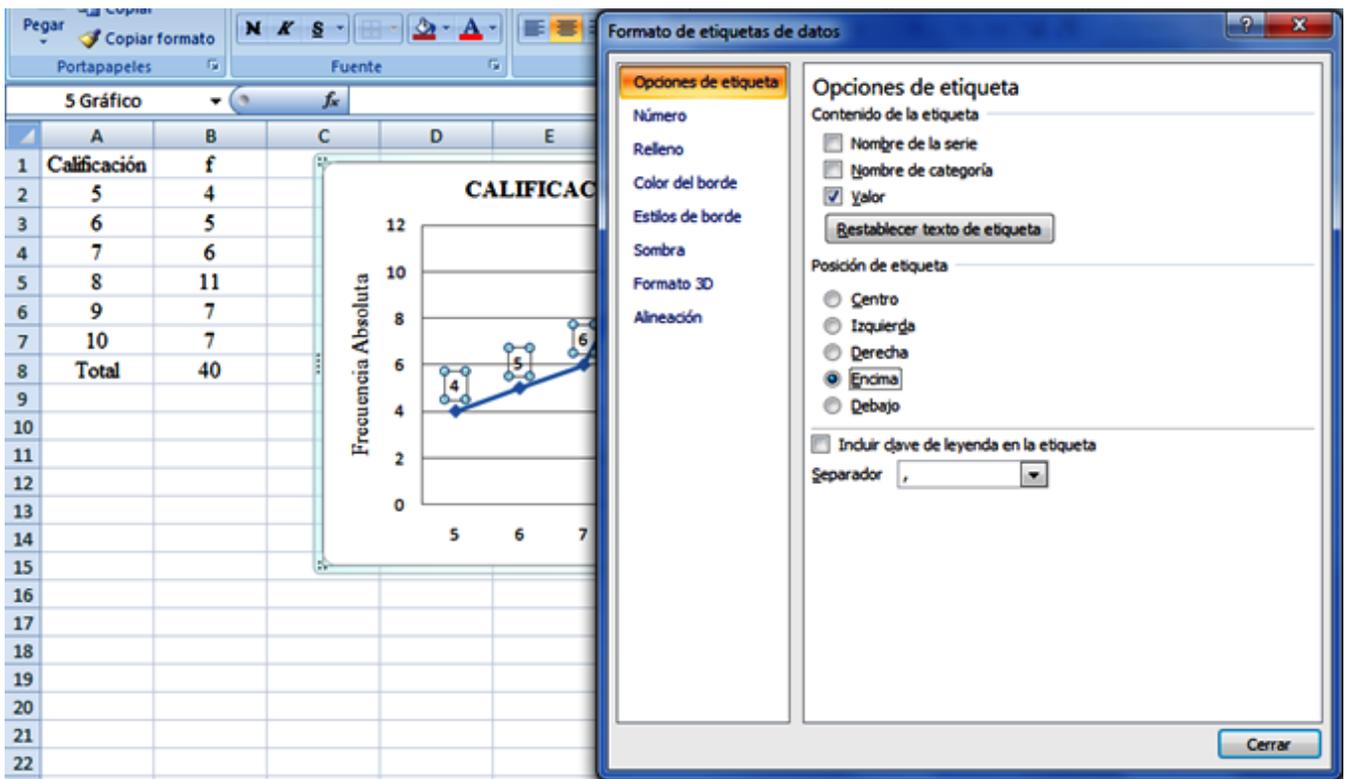
f) La serie Calificaciones arrastrar hacia más abajo. En título del gráfico escribir Calificaciones en Estadística. En título del eje vertical escribir frecuencia absoluta. Clic derecho en el gráfico y luego agregar etiqueta de datos.



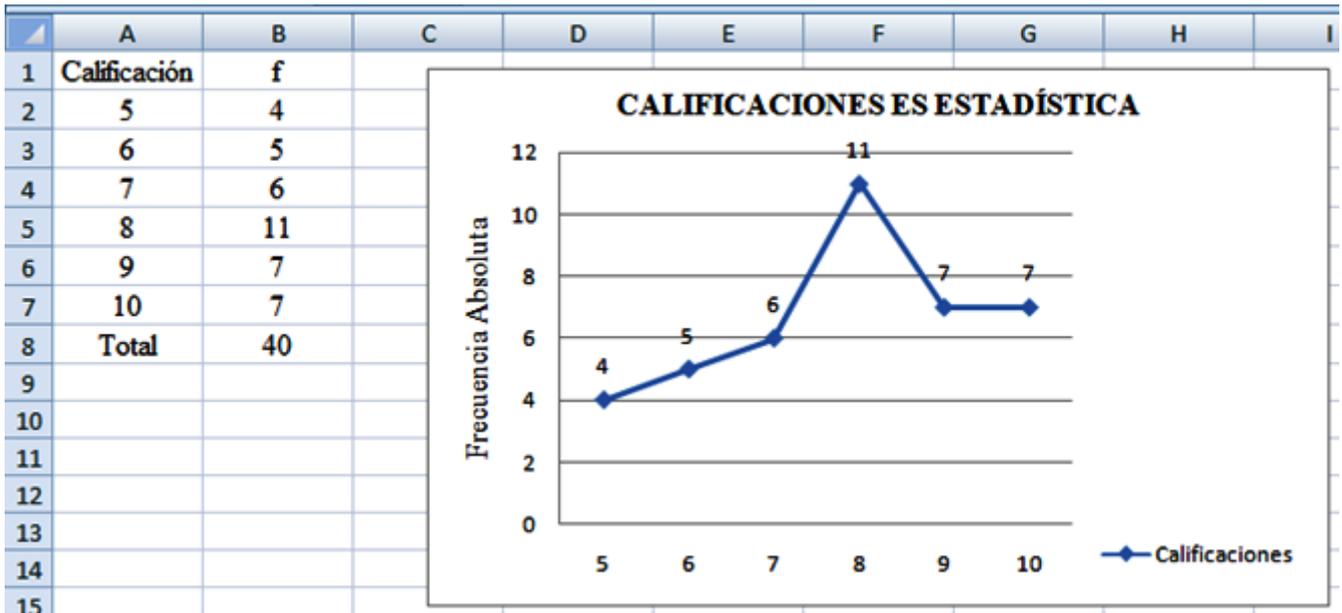
g) Clic derecho en etiqueta de datos



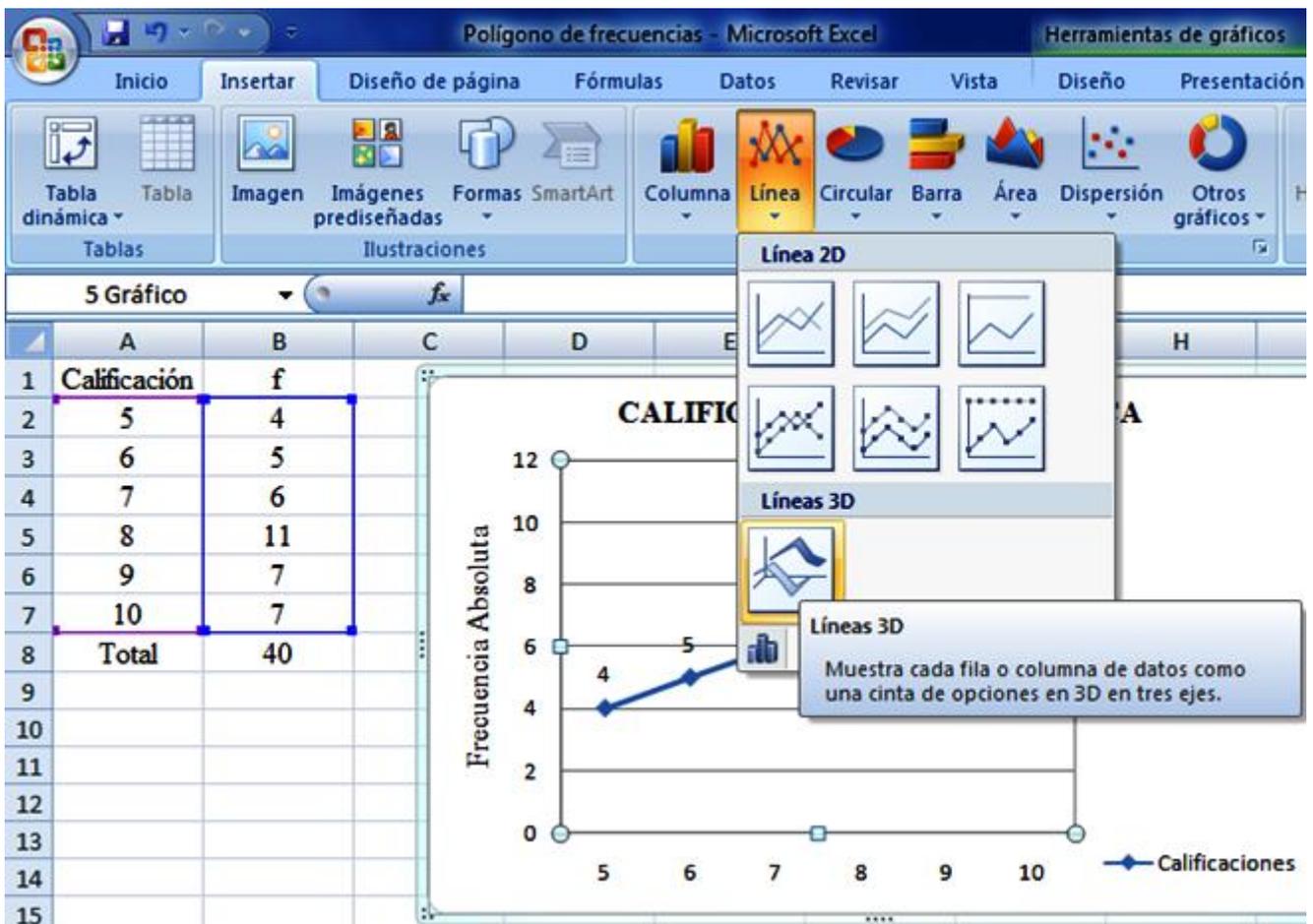
h) Seleccionar Formato de etiqueta de datos. En Posición de etiqueta seleccionar Encima.



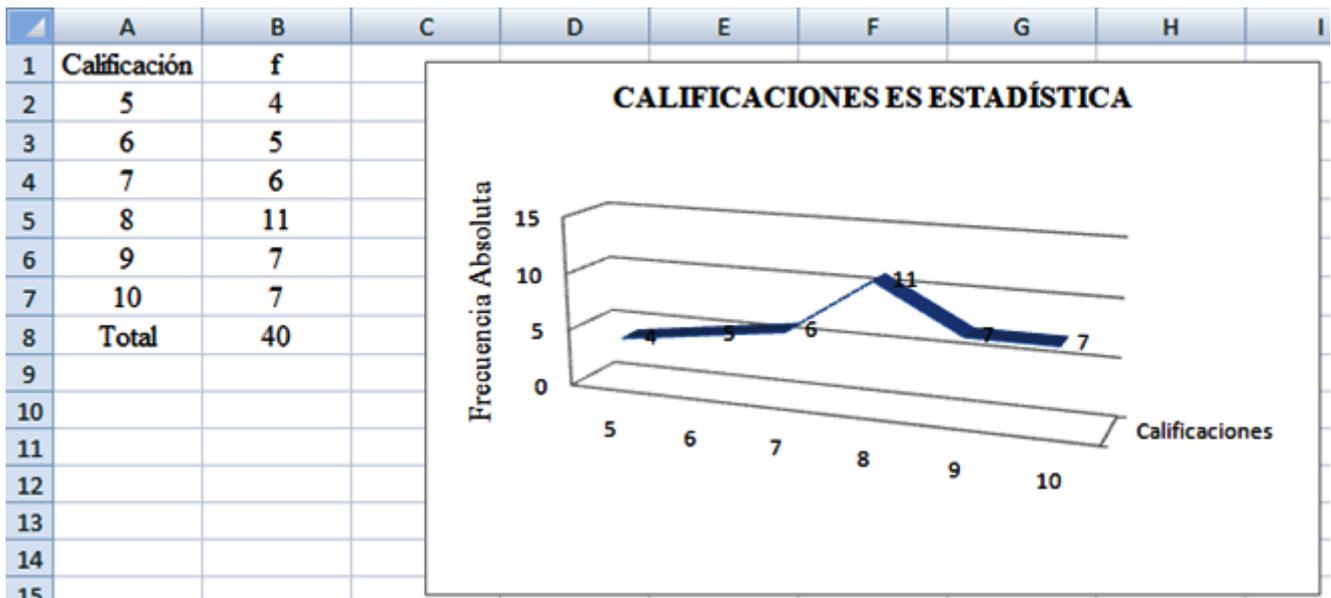
i) Clic en Cerrar y queda elaborado el polígono de frecuencias en 2D.



j) Para elaborar un polígono de frecuencias en tres dimensiones (3D), empleando el gráfico anterior hacer clic en Línea.

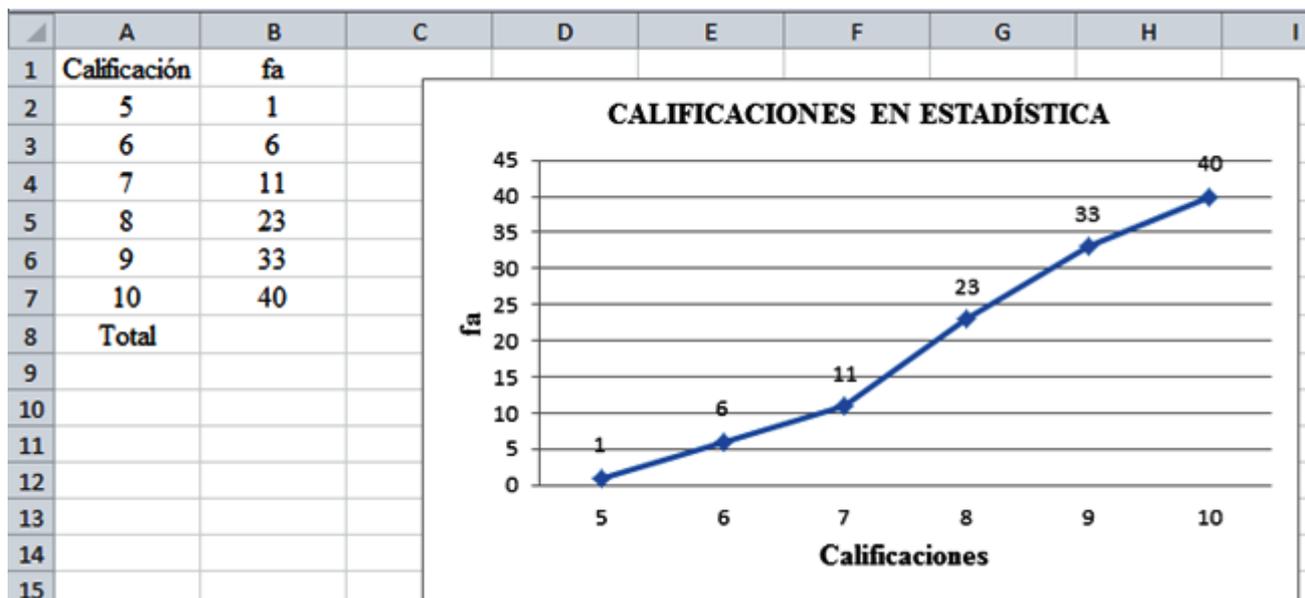


k) Clic Líneas 3D

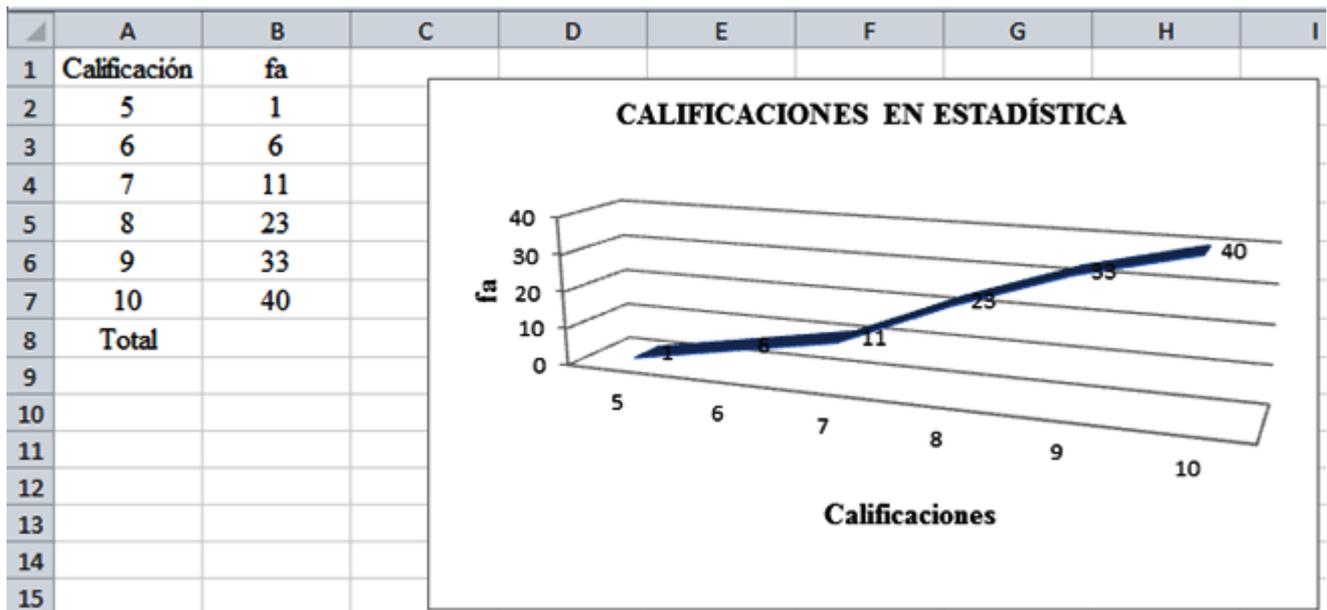


1.5.3.1) Polígono de Frecuencias Acumuladas u Ojiva.- Un gráfico que recoja las frecuencias acumuladas por debajo de cualquiera de las fronteras de clase superiores respecto de dicha frontera se llama un polígono de frecuencias acumuladas u ojiva.

Empleando polígono de frecuencias en 2D anterior, borrando la columna de la frecuencia absoluta y escribiendo la columna de la frecuencia acumulada del ejemplo del cálculo de las frecuencias sobre las siguientes calificaciones obtenidas por 40 estudiantes en una evaluación de la asignatura de Estadística se obtiene la siguiente figura que representa a una Ojiva:

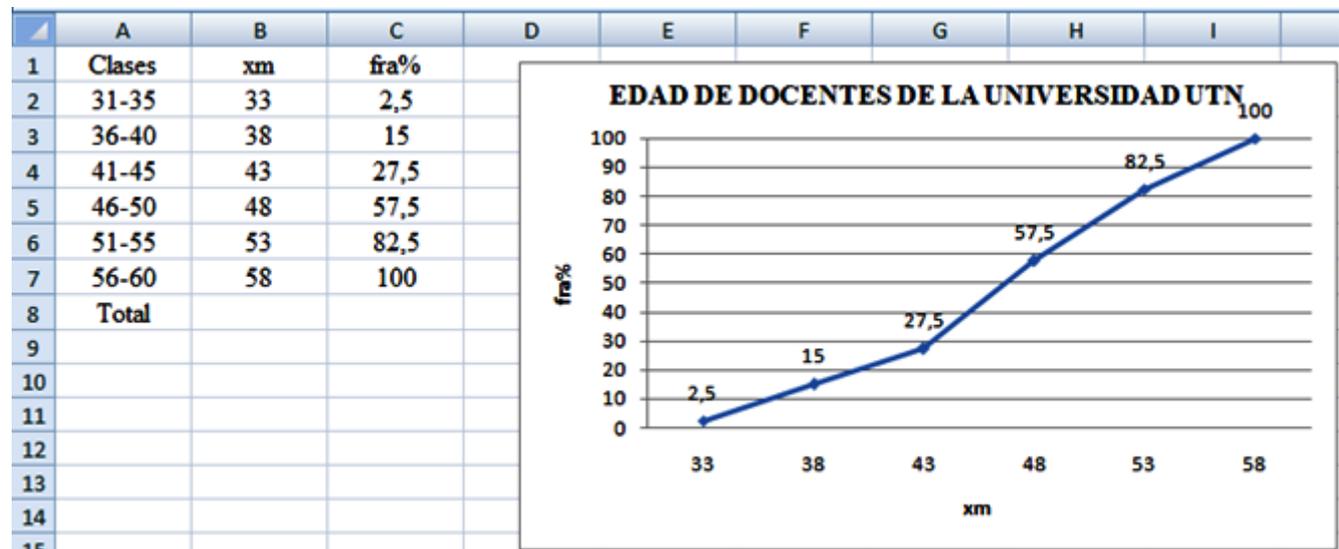


A partir del gráfico anterior, haciendo clic en Línea y luego en Líneas 3D se obtiene una Ojiva en 3D como muestra la siguiente figura:

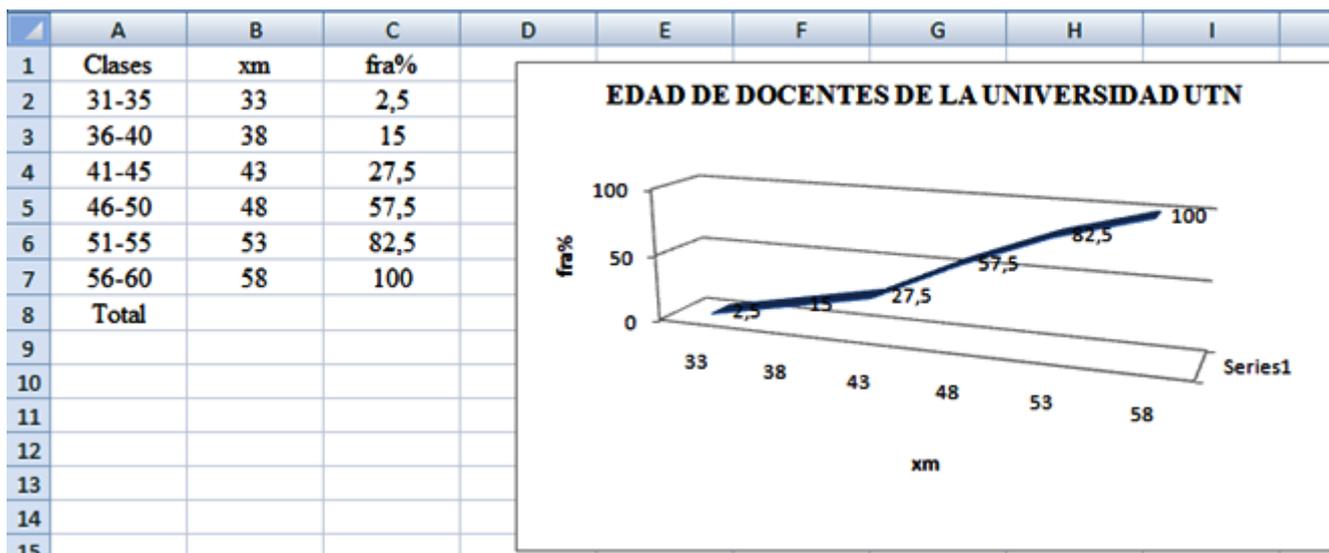


1.5.3.2) Polígono de Frecuencias Relativas Acumuladas Porcentuales.- Si se usan frecuencias $fra\%$ para realizar un polígono de frecuencias, este recibe el nombre de polígono de frecuencias relativas acumuladas porcentuales, o también llamado *ojiva de porcentajes*.

A continuación se presenta una ojiva de porcentajes elaborada en Excel empleando los datos del ejemplo de la Edad de 40 Docentes de la Universidad UTN:



La ojiva de porcentajes anterior elaborada en 3D se muestra en la siguiente figura:



1.5.4) DIAGRAMA DE TALLO Y HOJAS

En el diagrama de tallo y hojas cada dato representa su valor y, a la vez, ocupa un espacio de forma que se obtiene simultáneamente la presentación de los datos y distribución gráfica.

En este diagrama cada valor se descompone en 2 partes: el primero o primeros dígitos (el tallo) y el dígito que sigue a los utilizados en el tallo (las hojas). Por ejemplo, el valor 32 puede descomponerse en un tallo de 3 y una hoja de 2; el valor 325 puede descomponerse en un tallo de 32 y una hoja de 5; el valor 3256 puede descomponerse en un tallo de 325 y una hoja de 6. Cada tallo puede ocupar una o más filas. Si un tallo ocupa una sola fila, sus hojas contendrán dígitos del 0 al 9; si ocupa dos filas, la primera fila contendrá dígitos del 0 al 4 y la segunda fila del 5 al 9.

La ventaja de este diagrama es que refleja a primera vista las mismas impresiones gráficas que el histograma sin necesidad de elaborar el gráfico. También tiene la ventaja de conservar los valores originales de los datos.

Ejemplo ilustrativo:

A 40 estudiantes se les pidió que estimen el número de horas que habrían dedicado a estudiar la semana pasada (tanto en clase como fuera de ella), obteniéndose los siguientes resultados:

30	30	32	32	35	35	35	35
36	37	38	39	39	40	45	45
47	47	47	48	48	49	50	50
50	52	54	55	55	56	56	56
58	58	58	58	58	60	60	65

Elaborar un diagrama de tallo y hojas.

Solución:

A fin de elaborar el diagrama de tallo y hojas se ordena los datos con los dígitos iniciales de cada uno, las decenas (tallos) a la izquierda de una línea vertical, y a la derecha de esa recta el último dígito de cada dato, en este caso la unidad, conforme recorren los datos en el orden en que fueron anotados.

3		0022
3		555567899
4		0
4		55777889
5		00024
5		5566688888
6		00
6		5

Interpretaciones: Hay 4 estudiantes que dedican entre 30 y 32 horas semanales a estudiar, 10 estudiantes que dedican entre 55 y 58 horas semanales a estudiar, existe un solo estudiante que dedica 65 horas semanales a estudiar.

1.5.5) DIAGRAMA DE SECTORES

Llamado también diagrama circular o de pastel. Es un gráfico en el que a cada valor o modalidad se asigna un sector circular de área proporcional a la frecuencia que representan.

Ejemplo ilustrativo: Con los datos de la siguiente tabla sobre las calificaciones obtenidas por 40 estudiantes en una evaluación de Estadística, presentar la información a través de un diagrama de sectores:

Calificación	f
5	4
6	5
7	6
8	11
9	7
10	7
Total	40

Solución:

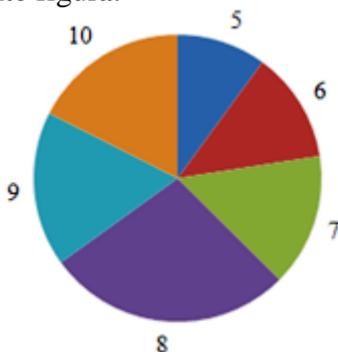
a) Se calcula la frecuencia relativa y el número de grados que representa cada calificación. El número de grados se calcula multiplicando la frecuencia relativa con 360^0 , así:

$$\text{número de grados} = fr \cdot 360^0$$

Estos cálculos se muestran en la siguiente tabla:

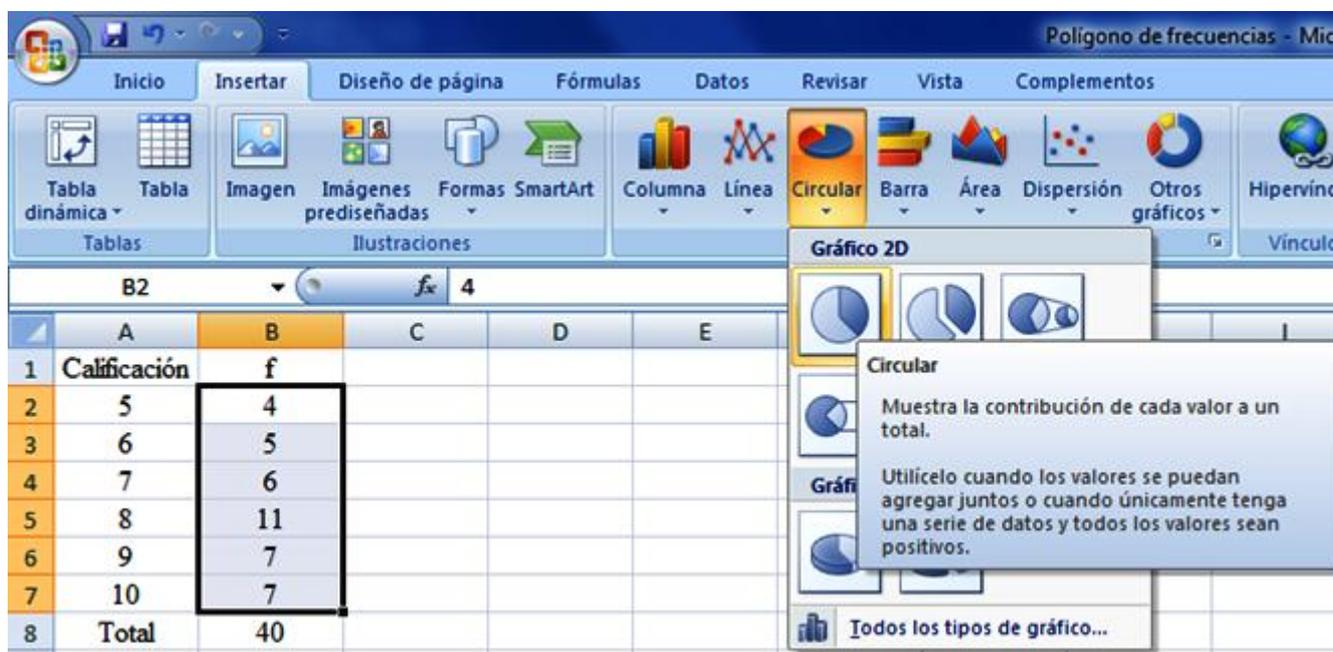
Calificación	f	fr	Grados
5	4	0,100	36
6	5	0,125	45
7	6	0,150	54
8	11	0,275	99
9	7	0,175	63
10	7	0,175	63
Total	40	1	360

b) Se dibuja una circunferencia tomando para cada calificación tantos grados como indica la tabla anterior como se muestra en la siguiente figura:

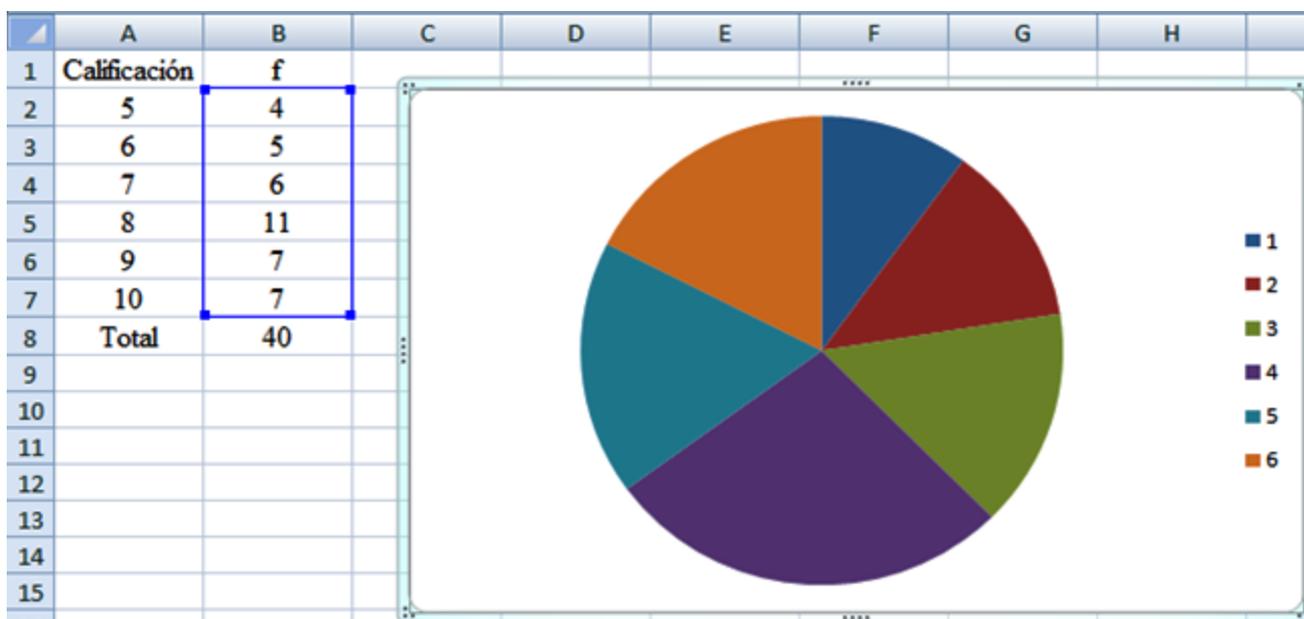


En Excel se elabora de la siguiente manera:

a) Digitar los datos. Seleccionar A2:A7. Clic en Insertar Gráfico Circular



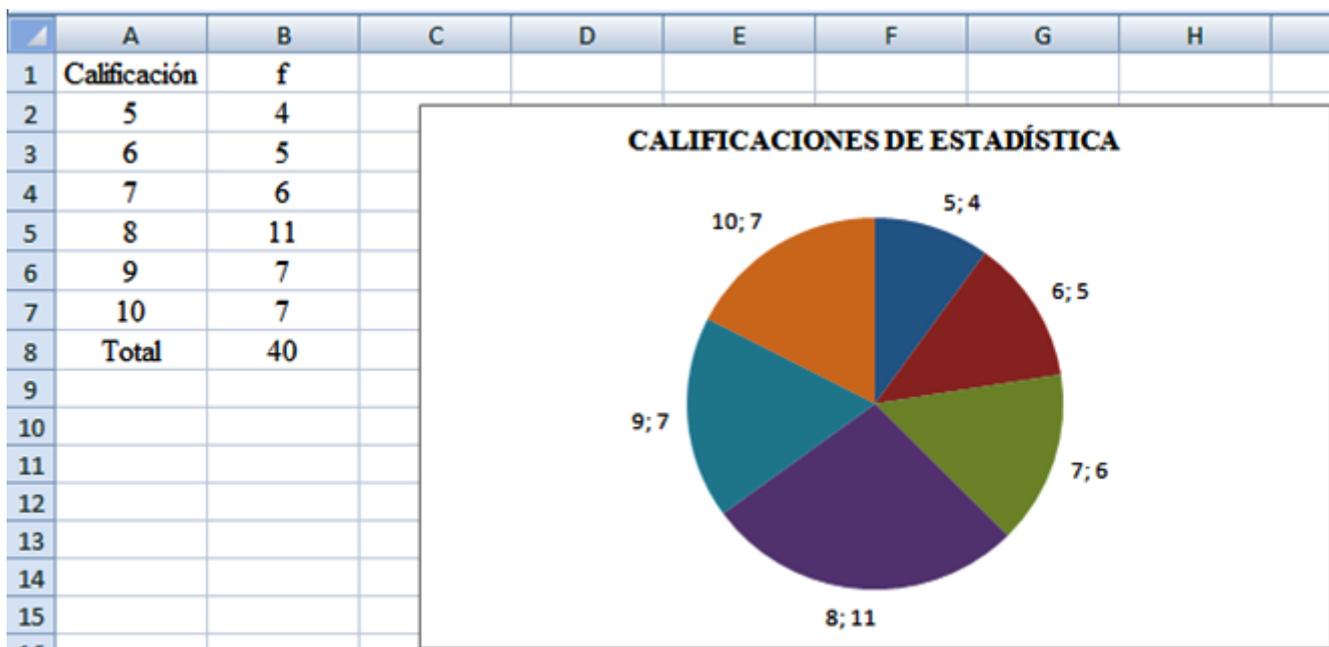
b) Clic en la primera opción de gráfico Circular 2D



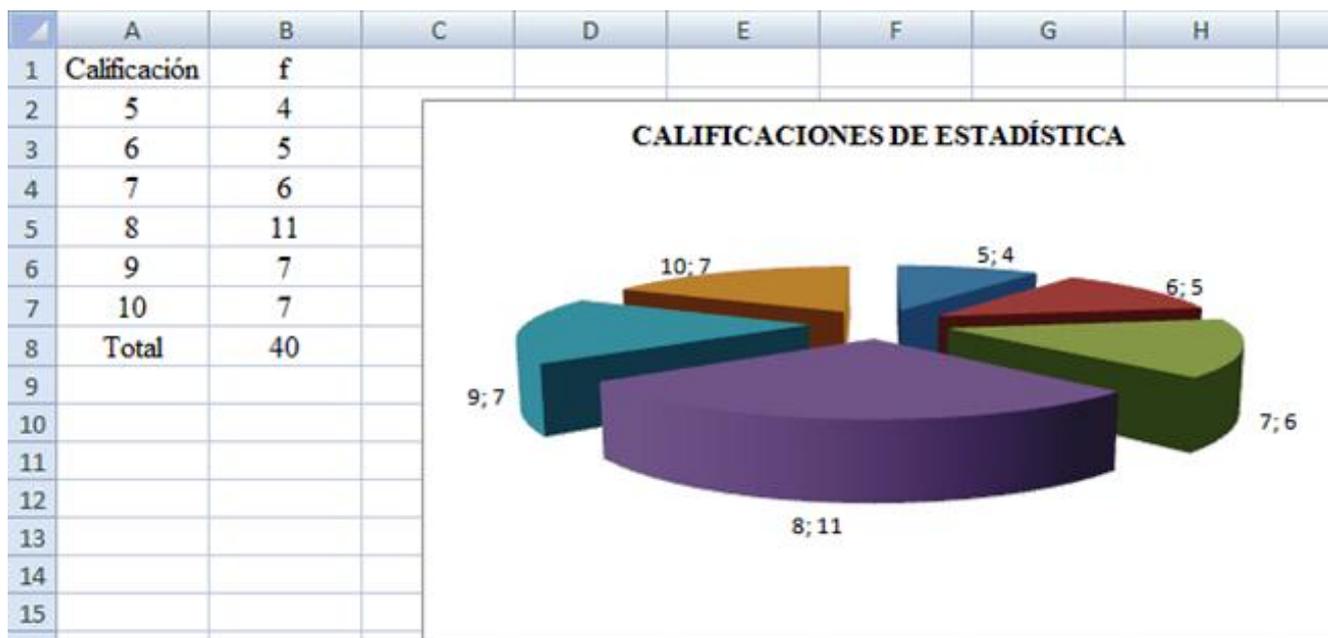
c) Clic en Diseño 5 de Diseños de gráfico.



d) Clic en Seleccionar datos para que aparezca la ventana Seleccionar origen de datos. En Etiquetas de eje horizontal (categoría), clic en Editar, luego en rango de rótulos de eje seleccionar A2:A7 y clic en Aceptar. En título del gráfico escribir Calificaciones en Estadística. Seleccionar Formato de etiqueta de datos, en contenido de etiqueta seleccionar Valor, en Posición de etiqueta seleccionar Extremo externo, en Separador seleccionar punto y como, y luego clic en Cerrar.



e) Para elaborar un diagrama de sectores en 3 dimensiones se procede de la siguiente manera: Clic en el gráfico anterior. Seleccionar Insertar Gráfico Circular. Escoger la opción 2 de Gráfico circular 3D. Se obtiene la siguiente figura:



1.5.6) PICTOGRAMAS

Son dibujos, figuras o signos llamativos alusivos al carácter que se está estudiando cuyo tamaño es proporcional a la frecuencia que representa los datos.

Ejemplo ilustrativo: Un equipo de fútbol en su trayectoria tiene 120 partidos ganados, 60 perdidos y 30 empatados. Al representar estos datos mediante pictogramas se obtiene:



Otra forma de representar los datos mediante pictogramas se muestra en la siguiente figura:



Partidos Ganados

120



Partidos Perdidos

60



Partidos Empatados

30

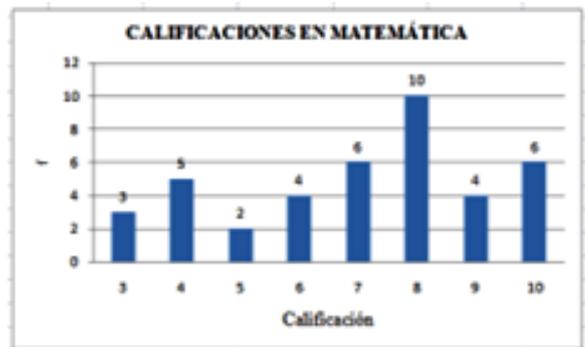
TAREA DE INTERAPRENDIZAJE N° 4

1) Presentar la información obtenida en la encuesta de la tarea de interaprendizaje N° 1 mediante tablas y gráficos estadísticos. Una tabla y un gráfico (de su preferencia) por cada pregunta.

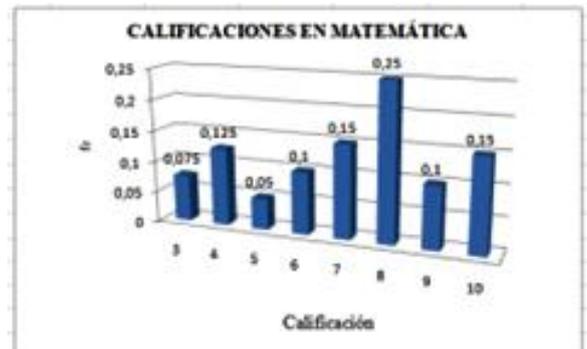
2) Las calificaciones obtenidas por 40 estudiantes en una evaluación de Matemática son:

4	6	6	8	10	10	6	8
8	8	7	7	9	8	4	8
4	7	9	7	9	10	8	9
10	5	10	6	8	4	5	7
10	4	3	3	3	8	7	8

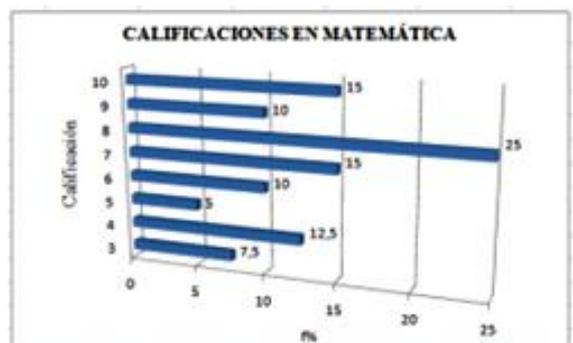
2.1) Elaborar un diagrama de barras verticales en 2D con la frecuencia absoluta de manera manual y empleando Excel.



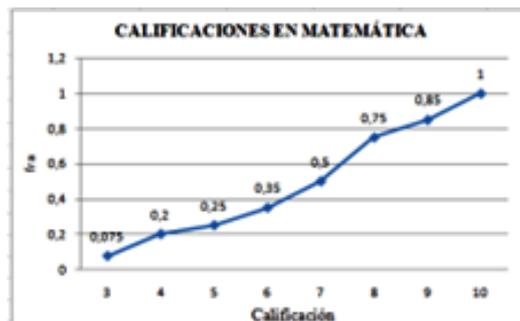
2.2) Elaborar un diagrama de barras verticales en 3D con la frecuencia relativa empleando Excel.



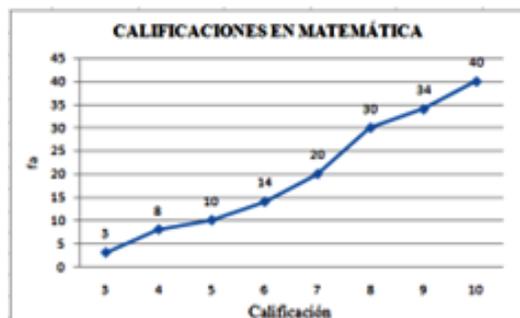
2.3) Elaborar un diagrama de barras horizontales en 3D con la frecuencia porcentual empleando Excel.



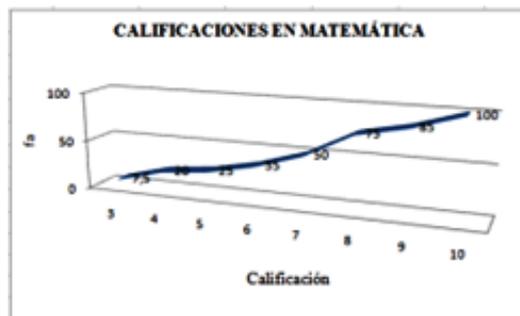
2.4) Elaborar un polígono de frecuencias en 2D con la frecuencia relativa acumulada de manera manual y empleando Excel.



2.5) Elaborar una ojiva en 2D de manera manual y empleando Excel.

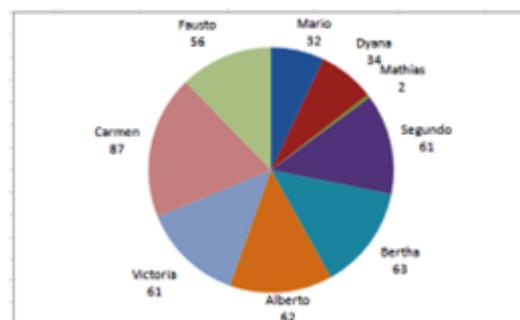


2.6) Elaborar una ojiva de porcentajes en 3D empleando Excel.



3) Elaborar un diagrama de sectores en 2D de manera manual y empleando Excel con los siguientes datos corresponde a las edades de un grupo de personas:

Nombre	Mario	Dyana	Mathías	Segundo	Bertha	Alberto	Victoria	Carmen	Fausto
Edad	32	34	2	61	63	62	61	87	56



4) Elaborar un diagrama de sectores en 2D y 3D con las edades de 10 familiares suyos empleando Excel.

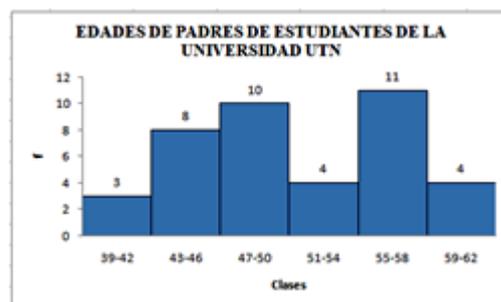
5) En una encuesta efectuada a los estudiantes de Segundo Semestre de la Universidad UTN sobre la edad de sus padres, se obtuvieron los siguientes resultados:

40	45	56	60	62	48	56	52
54	44	43	58	49	54	46	57
40	45	56	48	44	48	57	53
48	50	47	45	56	47	47	56
58	44	47	58	41	59	55	60

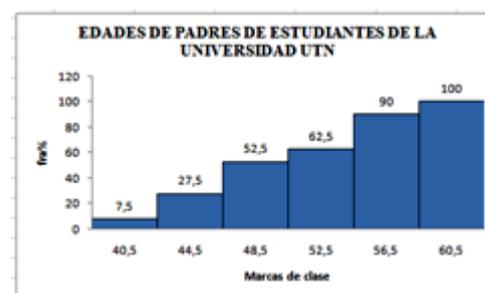
5.1) Terminar de llenar la siguiente tabla de manera manual y empleando Excel.

Clases	f	xm	fr	fa	$f\%$	fra	$fra\%$
39-42	3		0,075				7,5
43-46		44,5				0,275	
47-50	10			21	25		
51-54			0,1			0,625	62,5
55-58	11	56,5					
59-62			0,1	40	10	1	100
	40		1		100		

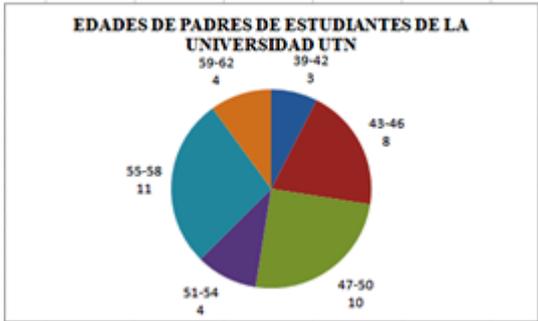
5.2) Elaborar un histograma para la frecuencia absoluta de manera manual y empleando Excel, ubicando la clases en el eje horizontal del gráfico.



5.3) Elaborar un histograma para la fra% de manera manual y empleando Excel, ubicando las marcas de clase en el eje horizontal del gráfico.



5.4) Elaborar un diagrama de sectores en 2D para la frecuencia absoluta de manera manual y empleando Excel.



6) Elabore de manera manual un polígono de frecuencias uniendo los puntos medios de las bases superiores de los rectángulos en un diagrama de barras creado por usted con datos de cualquier tema de su interés.

7) Elabore de manera manual un polígono de frecuencias uniendo los puntos medios (marcas de clase) de las bases superiores en un histograma creado por usted con datos de cualquier tema de su interés.

8) Investigue sobre un tema que a usted le interese y elabore un diagrama de su predilección de manera manual y empleando Excel.

9) Investigue sobre un tema de su gusto y elabore un diagrama de tallos y hojas.

10) Elabore un pictograma sobre un tema de su agrado.

CAPÍTULO II

MEDIDAS DE TENDENCIA CENTRAL

COMPETENCIAS:

- ✓ Interpreta adecuadamente las características y propiedades de las medidas de tendencia central, y comprende sus aplicaciones.
- ✓ Emplea correctamente algoritmos matemáticos para calcular medidas de tendencia central de manera manual y empleando Excel.
- ✓ Realiza eficientemente diagramas de caja y bigotes de manera manual y empleando Graph.
- ✓ Crea y resuelve correctamente ejercicios de aplicación sobre las medidas de tendencia central.

CONTENIDOS:

- ✓ Media Aritmética
- ✓ Media Geométrica
- ✓ Media Armónica
- ✓ La Mediana
- ✓ Medidas de Posición: Cuartiles, Deciles y Percentiles
- ✓ Moda

2.1) MEDIA ARITMÉTICA

Las medidas de tendencia central son medidas representativas que como su nombre lo indica, tienden a ubicarse hacia el centro del conjunto de datos, es decir, una medida de tendencia central identifica el valor del dato central alrededor de cual se centran los demás datos, siendo la media aritmética una de aquellas medidas.

La medida aritmética, al igual que cualquier otra medida de datos estadísticos, cuando se calcula a nivel de toda la *población*, se denominan *parámetro*, como por ejemplo, la calificación promedio en el examen de admisión de todos los estudiantes que ingresan a la Universidad UTN al primer semestre del presente año lectivo. Pero si se calcula basada en *muestras*, se denomina *estadígrafo o estadístico*, como por ejemplo, la calificación promedio en el examen de admisión de estudiantes de colegios fiscales que ingresan a la Universidad UTN al primer semestre del presente año lectivo.

2.1.1) MEDIA ARITMÉTICA SIMPLE

2.1.1.1) Definición

Es la medida de tendencia central más utilizada por lo general se ubica hacia el centro de distribución estadística.

2.1.1.2) Métodos de Cálculo

a) Para Datos sin Agrupar

La media de una población es el parámetro μ (que se pronuncia miu). Si hay N observaciones en el conjunto de datos de la población, la media se calcula así:

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N} = \frac{\sum x_i}{N}$$

La media de una muestra es un estadístico \bar{x} (que se lee “x barra”). Con n observaciones en el conjunto de datos de la muestra (x_1, x_2, \dots), la media se determina así:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum x_i}{n}$$

b) Para Datos Agrupados en Tablas de Frecuencias.- Cuando una serie se la agrupa en *serie simple con frecuencias* para obtener la media aritmética, se multiplica la variable por la frecuencia respectiva (f), luego se obtiene la suma de todos estos productos y luego a este valor se lo divide para el número de elementos (n). Todo esto puede representarse mediante una fórmula matemática, así:

$$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + f_3 \cdot x_3 + \dots + f_n \cdot x_n}{f_1 + f_2 + f_3 + \dots + f_n} = \frac{\sum f_i \cdot x_i}{\sum f} = \frac{\sum fx}{n}$$

Donde $n = \sum f$ es la frecuencia total (o sea, el número total de casos)

c) Para Datos Agrupados en Intervalos.- Cuando una serie se la agrupa en *intervalos* para obtener la media aritmética, se multiplica la marca de clase de intervalo (xm) por la frecuencia respectiva (f), luego se obtiene la suma de todos estos productos y luego a este valor se lo divide para el número de elementos. Todo esto puede representarse mediante una fórmula matemática, así

$$\bar{x} = \frac{f_1 \cdot xm_1 + f_2 \cdot xm_2 + f_3 \cdot xm_3 + \dots + f_n \cdot xm_n}{f_1 + f_2 + f_3 + \dots + f_n} = \frac{\sum f_i \cdot xm_i}{\sum f} = \frac{\sum f \cdot xm}{n}$$

Ejemplo ilustrativo

Calcular la media aritmética de las siguientes calificaciones de Estadística tomadas de una muestra de 20, sin agrupar, agrupando en tablas de frecuencias y agrupando en intervalos.

4, 8, 10, 10, 5, 10, 9, 8, 6, 8, 10, 8, 5, 7, 4, 4, 8, 8, 6 y 6

Solución:

1) Sin agrupar

$$\bar{x} = \frac{\sum x_i}{n}$$
$$\bar{x} = \frac{4 + 8 + 10 + 10 + 5 + 10 + 9 + 8 + 6 + 8 + 10 + 8 + 5 + 7 + 4 + 4 + 8 + 8 + 6 + 6}{20}$$
$$\bar{x} = \frac{144}{20} = 7,2$$

En Excel se calcula así:

a) Se escriben los números y se inserta la función PROMEDIO como se muestra en la siguiente figura:

	A	B	C	D
1	4	8	10	10
2	5	10	9	8
3	6	8	10	8
4	5	7	4	4
5	8	8	6	6
6				
7	Media aritmética	=		
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				

Insertar función

Buscar una función:
Escriba una breve descripción de lo que desea hacer y, a continuación, haga clic en Ir

O seleccionar una categoría: Estadísticas

Seleccionar una función:

- PERMUTACIONES
- POISSON
- PROBABILIDAD
- PROMEDIO**
- PROMEDIO.SI
- PROMEDIO.SI.CONJUNTO
- PROMEDIOA

PROMEDIO(número1;número2;...)
Devuelve el promedio (media aritmética) de los argumentos, los cuales pueden ser números, nombres, matrices o referencias que contengan números.

[Ayuda sobre esta función](#)

Aceptar Cancelar

b) Seleccionar las celdas (Rango A₁:D₅)

The screenshot shows an Excel spreadsheet with the following data in columns A to D, rows 1 to 5:

	A	B	C	D
1	4	8	10	10
2	5	10	9	8
3	6	8	10	8
4	5	7	4	4
5	8	8	6	6

Below the data, cell C7 contains the formula `=PROMEDIO(A1:D5)`. The 'Argumentos de función' dialog box is open, showing the range A1:D5 and the result 7,2. The dialog text reads: 'Devuelve el promedio (media aritmética) de los argumentos, los cuales pueden ser números, nombres, matrices o referencias que contengan números. Número1: número1;número2;... son entre 1 y 255 argumentos numéricos de los que se desea obtener el promedio. Resultado de la fórmula = 7,2'.

c) Pulsar en Aceptar.

	A	B	C	D	E
1	4	8	10	10	
2	5	10	9	8	
3	6	8	10	8	
4	5	7	4	4	
5	8	8	6	6	
6					
7	Media aritmética		7,2	=PROMEDIO(A1:D5)	

2) Agrupando en tablas de frecuencias

Además presentar los datos en un diagrama de sectores.

x	f
4	3
5	2
6	3
7	1
8	6
9	1
10	4
Total	20

$$\bar{x} = \frac{\sum fx}{n}$$

$$\bar{x} = \frac{3 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 + 1 \cdot 7 + 6 \cdot 8 + 1 \cdot 9 + 4 \cdot 10}{3 + 2 + 3 + 1 + 6 + 1 + 4} = \frac{144}{20} = 7,2$$

En Excel se calcula insertando la función:

$$=SUMAPRODUCTO(A12:A18;B12:B18)/SUMA(B12:B18)$$

	A	B	C	D	E	F	G	H
11	x	f						
12	4	3	=CONTAR.SI(A2:D6;4)					
13	5	2	=CONTAR.SI(A2:D6;5)					
14	6	3	=CONTAR.SI(A2:D6;6)					
15	7	1	=CONTAR.SI(A2:D6;7)					
16	8	6	=CONTAR.SI(A2:D6;8)					
17	9	1	=CONTAR.SI(A2:D6;9)					
18	10	4	=CONTAR.SI(A2:D6;10)					
19		20						
20								
21								
22	7,2		=SUMAPRODUCTO(A12:A18;B12:B18)/SUMA(B12:B18)					

3) Agrupando en intervalos

Intervalos	f	xm
4- 5	5	4,5
6 -7	4	6,5
8- 9	7	8,5
10-11	4	10,5

$$\bar{x} = \frac{\sum f \cdot xm}{n}$$

$$\bar{x} = \frac{5 \cdot 4,5 + 4 \cdot 6,5 + 7 \cdot 8,5 + 4 \cdot 10,5}{5 + 4 + 7 + 4} = \frac{150}{20} = 7,5$$

Nota: Cuando se agrupa en intervalos los cálculos son sólo aproximaciones

En Excel se calcula insertando la función: SUMAPRODUCTO (C27:C30;D27:D30)/SUMA(C27:C30) como se muestra en la siguiente figura:

26	x	f	xm					
27	4	5	4,5	=PROMEDIO(A27:B27)				
28	6	7	6,5	=PROMEDIO(A28:B28)				
29	8	9	8,5	=PROMEDIO(A29:B29)				
30	10	11	10,5	=PROMEDIO(A30:B30)				
31								
32	7,5			=SUMAPRODUCTO(C27:C30;D27:D30)/SUMA(C27:C30)				

Nota: La principal propiedad de la media aritmética es:

La suma algebraica de las desviaciones de un conjunto de datos respecto de su media aritmética es cero. Si x es un dato, su desviación respecto a \bar{x} es la diferencia $x - \bar{x}$. La suma de estas diferencias es 0.

Para datos sin agrupar: $\sum(x - \bar{x}) = 0$

Para datos agrupados en tablas de frecuencias: $\sum f(x - \bar{x}) = 0$

Para datos agrupados en intervalos: $\sum f(xm - \bar{x}) = 0$

Empleando los datos del ejemplo anterior se comprueba la principal propiedad de la media aritmética:

x	f	xm	$f(xm - \bar{x})$
4	5	4,5	$5(4,5 - 7,5) = -15$
6	7	6,5	$4(6,5 - 7,5) = -4$
8	9	8,5	$7(8,5 - 7,5) = 7$
10	11	10,5	$4(10,5 - 7,5) = 12$
Suma			0

2.1.2) MEDIA ARITMÉTICA PONDERADA

Cuando los números $x_1, x_2, x_3, \dots, x_k$ se les asocia ciertos factores peso (o pesos) $w_1, w_2, w_3, \dots, w_k$, dependientes de la relevancia asignada a cada número, en tal caso se requiere calcular la media aritmética ponderada, la cual se calcula así:

$$\bar{x} = \frac{w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + \dots + w_k \cdot x_k}{w_1 + w_2 + w_3 + \dots + w_k} = \frac{\sum w \cdot x}{\sum w}$$

Ejemplo ilustrativo: Se tiene una información acerca de las utilidades por pan y cantidades vendidas de panes de tres tiendas. Calcular la media aritmética promedio de la utilidad por pan.

Tienda	Utilidad/pan	Cantidad vendida
1	1	2000
2	0,8	1800
3	0,9	2100

Solución:

$$\bar{x} = \frac{w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + \dots + w_k \cdot x_k}{w_1 + w_2 + w_3 + \dots + w_k} = \frac{\sum w \cdot x}{\sum w}$$

$$\bar{x} = \frac{2000 \cdot 1 + 1800 \cdot 0,8 + 2100 \cdot 0,9}{2000 + 1800 + 2100} = \frac{5330}{5900} = 0,90339$$

En Excel se calcula insertando de la siguiente manera:

Se insertar la función: $\text{SUMAPRODUCTO}(B36:B38;C36:C38)/\text{SUMA}(C36:C38)$

	A	B	C	D	E	F	G
34							
35	Tienda	Utilidad/pan	Cantidad vendida				
36	1	1	2000				
37	2	0,8	1800				
38	3	0,9	2100				
39							
40							
41	0,90339			=SUMAPRODUCTO(B36:B38;C36:C38)/SUMA(C36:C38)			

TAREA DE INTERAPRENDIZAJE N° 5

- 1) Defina con sus propias palabras lo que entiende por medidas de tendencia central
- 2) ¿Cuál es la diferencia entre parámetro y estadígrafo?. Mediante un ejemplo ilustre su respuesta.
- 3) ¿Qué entiende por media aritmética simple?
- 4) ¿Qué entiende por media aritmética ponderada?
- 5) Calcule la media aritmética de las siguientes calificaciones de Estadística tomadas de una muestra en forma manual y empleando Excel.

10	8	9	7	6
5	4	8	6	3
8	3	6	9	10
8	10	10	9	8

5.1) Sin agrupar.

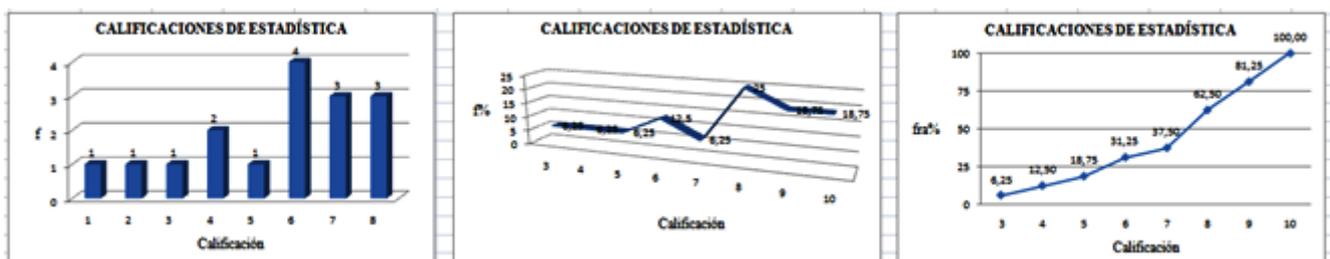
7,35

5.2) Agrupando en frecuencias.

7,35

6) Compruebe la propiedad principal de la media aritmética con los datos del ejercicio anterior sin agrupar y con los datos agrupados en frecuencias de manera manual y empleando Excel.

7) Presente los datos del ejercicio 5 en un diagrama de barras verticales en 3D empleando Excel, en un polígono de frecuencias para $f\%$ en 3D empleando Excel, y en una ojiva de porcentajes en 2D elaborada en forma manual y empleando Excel.



8) Calcule la media aritmética de las siguientes calificaciones de Matemática tomadas de una muestra en forma manual y empleando Excel.

10	8	9	7	6	3	7	10	6
5	4	8	8	3	4	8	9	5
8	3	8	9	10	5	9	8	4
8	10	10	9	8	6	10	7	3

8.1) Sin agrupar.

7,0833

8.2) Agrupando en frecuencias.

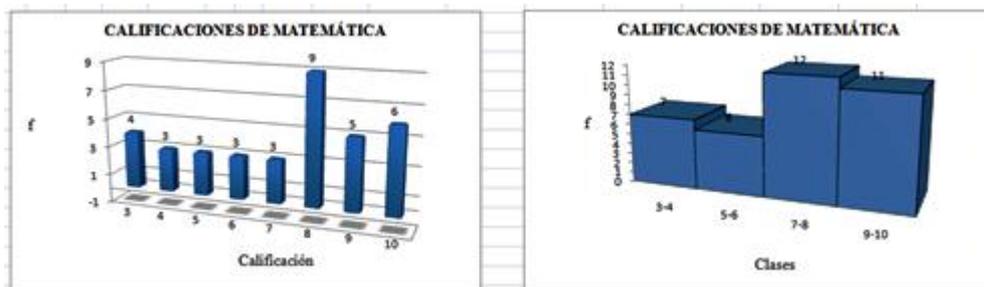
7,0833

8.3) Agrupando en intervalos de ancho 2.

7

9) Compruebe la propiedad principal de la media aritmética con los datos del ejercicio anterior agrupados en intervalos de manera manual y empleando Excel.

10) Presente los datos del ejercicio 8 en un diagrama de barras verticales e histograma en 3D elaborados empleando Excel.



11) Cree y resuelva un ejercicio similar al N° 8 con datos de cualquier tema de su interés.

12) Para construir un edificio se contrataron 30 obreros con un sueldo mensual de \$ 300 cada uno. Calcule el sueldo promedio.

\$ 300

13) En una investigación sobre la población en 4 barrios de la ciudad de Ibarra, se encontró que el número de habitantes es: 2000, 3000, 4500, 5000. Se supone que en 10 años la población se duplicará. Calcule la población promedio dentro de 10 años.

7250 habitantes

14) Cuatro personas ganan mensualmente: \$400, \$300, \$500, \$700. Calcule el salario promedio si a cada uno le aumentan \$80.

\$555

15) Un grupo de estudiantes obtuvieron las siguientes calificaciones evaluadas sobre 10 como se indica en la siguiente tabla:

Asignatura	Calificación					
	7	8	6	6	5	10
Matemática	7	8	6	6	5	10
Estadística	8	9	6	4	10	8
Inglés	9	10	8	8	7	6

Calcule la calificación promedio del grupo

7,5
72

16) Un estudiante en la asignatura de Estadística en los tres aportes parciales evaluados sobre 10 tiene: 4, 6 y 10. ¿Cuánto debe obtener en el cuarto aporte para que su promedio exacto sea 7?

8

17) Cree y resuelva un ejercicio similar anterior.

18) Los tres primeros aportes de un estudiante en la asignatura de Matemática son: el primer aporte es el doble del segundo, y éste es cuatro unidades menos que el tercer aporte, y el cuarto aporte es 2 unidades más que el tercer aporte. Si el promedio exacto es 5, ¿cuáles fueron los aportes?

$$x_1 = 4, x_2 = 2, x_3 = 6 \text{ y } x_4 = 8$$

19) Cree y resuelva un ejercicio similar anterior.

20) Si el examen final de Estadística cuenta tres veces más que una evaluación parcial, y un estudiante tiene 8 en el examen final, 7 y 9 en las dos parciales. Calcule la calificación media en forma manual y empleando Excel.

8

21) Cree un ejercicio de aplicación sobre la media aritmética ponderada y resuélvalo forma manual y empleando Excel.

2.2) MEDIA GEOMÉTRICA

2.2.1) PROPIEDADES

- La media geométrica proporciona una medida precisa de un cambio porcentual promedio en una serie de números.
- Se utiliza con más frecuencia para calcular la tasa de crecimiento porcentual promedio de series de datos, a través del tiempo.
- Es una medida de tendencia central por lo general menor que la media aritmética salvo en el extraño caso en que todos los incrementos porcentuales sean iguales, entonces las dos medias serán iguales.
- Se le define como la raíz enésima del producto de “n” valores. Cuando los datos son bastantes o cantidades grandes, para facilitar el cálculo se lo debe simplificar pero sin alterar su naturaleza, para lo cual se puede utilizar los logaritmos de base 10.

2.2.2) MÉTODOS DE CÁLCULO

2.2.2.1) Para Datos No Agrupados

Se emplea la ecuación:

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$$

O aplicando logaritmos la ecuación:

$$\log G = \frac{\log x_1 + \log x_2 + \log x_3 + \cdots \log x_n}{n}$$

Ejemplo ilustrativo N° 1

La media geométrica es útil en el cálculo de tasas de crecimiento; por ejemplo, si el crecimiento de las ventas en un pequeño negocio son 3%, 4%, 8%, 9% y 10%, hallar la media de crecimiento.

Solución:

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$$

$$G = \sqrt[5]{3 \cdot 4 \cdot 8 \cdot 9 \cdot 10} = 6,128$$

Respuesta: 6,128%

O utilizando logaritmos:

$$\log G = \frac{\log x_1 + \log x_2 + \log x_3 + \cdots + \log x_n}{n}$$

$$\log G = \frac{\log 3 + \log 4 + \log 8 + \log 9 + \log 10}{5}$$

$$\log G = \frac{0,4771 + 0,6021 + 0,9031 + 0,9542 + 1}{5}$$

$$\log G = \frac{3,9365}{5}$$

$$\log G = 0,7873$$

$$G = \text{antilog } 0,7873$$

$$G = 6,128$$

Empleando Excel se calcula de la siguiente manera:

a) Insertar la función Media Geométrica y pulsar en Aceptar.

PROBABILIDAD

	A	B	C	D	E	F	G	H
1	3							
2	4							
3	8							
4	9							
5	10							
6								
7	G	=						
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								

Insertar función

Buscar una función:

Escriba una breve descripción de lo que desea hacer y, a continuación, haga clic en Ir

O seleccionar una categoría: Estadísticas

Seleccionar una función:

- MAXA
- MEDIA.ACOTADA
- MEDIA.ARMO
- MEDIA.GEOM**
- MEDIANA
- MIN
- MINA

MEDIA.GEOM(número1;número2;...)
Devuelve la media geométrica de una matriz o rango de datos numéricos positivos.

[Ayuda sobre esta función](#)

Aceptar Cancelar

b) Seleccionar las celdas (Rango A1:A5)

The screenshot shows an Excel spreadsheet with the following data in cells A1:A5:

1	3
2	4
3	8
4	9
5	10

Cell B7 contains the formula `=MEDIA.GEOM(A1:A5)`. The 'Argumentos de función' dialog box is open, showing the function name 'MEDIA.GEOM' and the arguments:

- Número1: A1:A5 = {3\4\8\9\10}
- Número2: = número

The dialog box also displays the result: `= 6,127774126` and the description: 'Devuelve la media geométrica de una matriz o rango de datos numéricos positivos. Número1: número1;número2;... son de 1 a 255 números, argumentos, matrices o referencias que contienen números cuya media se desea calcular. Resultado de la fórmula = 6,127774126'. Buttons for 'Aceptar' and 'Cancelar' are visible.

c) Pulsar en Aceptar.

	A	B	C	D
1	3			
2	4			
3	8			
4	9			
5	10			
6				
7	G	6,1277741	=MEDIA.GEOM(A1:A5)	

Ejemplo ilustrativo N° 2

Calcular la tasa de crecimiento promedio a la que ha variado las ventas de cierto producto con base a la siguiente tabla:

Mes	Enero	Febrero	Marzo	Abril	Mayo	Junio
Ventas	500	550	600	700	800	850

Solución:

Es necesario calcular el porcentaje que las ventas de cada mes representan respecto de los obtenidos el mes anterior.

Mes	Ventas	Porcentaje del mes anterior
Enero	500	
Febrero	550	$550/500=1,100$
Marzo	600	$600/550=1,091$
Abril	700	$700/600=1,167$
Mayo	800	$800/700=1,143$
Junio	850	$850/800=1,063$

Calculando la media geométrica se obtiene:

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$$

$$G = \sqrt[5]{1,100 \cdot 1,091 \cdot 1,167 \cdot 1,143 \cdot 1,063}$$

$$G = 1,112$$

Restando 1 para convertirlo a un incremento mensual promedio da $1,112 - 1 = 0,112$, o un incremento promedio de 11,2% para el período de 6 meses.

Comprobación:

Mes	Ventas	Ventas calculadas con G
Enero	500	
Febrero	550	$500 \times 1,112 = 556,000$
Marzo	600	$556 \times 1,112 = 618,272$
Abril	700	$618,272 \times 1,112 = 687,518$
Mayo	800	$687,518 \times 1,112 = 764,52$
Junio	850	$764,52 \times 1,112 = 850,146$

Se puede observar que el valor de 850,146 calculado con la media geométrica es semejante al valor de venta real de 850, por lo tanto el valor calculado para la media geométrica está correcto.

2.2.2.2) Para Datos Agrupados en Tablas de Frecuencias

Se emplea la siguiente ecuación:

$$\log G = \frac{\sum \log x_i \cdot f_i}{n}$$

Donde:

f_i = frecuencia absoluta de cada dato x_i

Ejemplo ilustrativo N° 3

Calcular la media geométrica para las siguientes calificaciones de Estadística:

x_i	f_i
4	5
6	8
8	9
9	10
10	8

Solución:

Se llena la siguiente tabla, realizando los cálculos respectivos:

x_i	f_i	$\log x_i$	$\log x_i \cdot x_i$
4	5	0,602	3,010
6	8	0,778	6,225
8	9	0,903	8,128
9	10	0,954	9,542
10	8	1,000	8,000
Total	40		34,906

Se aplica la siguiente ecuación para obtener la respuesta.

$$\log G = \frac{\sum \log x_i \cdot f_i}{n}$$

$$\log G = \frac{34,906}{40} = 0,873$$

$$G = \text{anti log } 0,873 = 7,458$$

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G
1	x_i	f_i					
2	4	5					
3	6	8					
4	8	9					
5	9	10					
6	10	8					
7							
8	G	7,45836483	=10^(SUMAPRODUCTO(LOG10(A2:A6);B2:B6)/SUMA(B2:B6))				

2.2.2.3) Para Datos Agrupados en Intervalos

Se emplea la ecuación:

$$\log G = \frac{\sum \log xm \cdot f_i}{n}$$

Donde:

xm = marca de clase

TAREA DE INTERAPRENDIZAJE N° 6

- 1) Realice un organizador gráfico sobre la media geométrica
- 2) Cree y resuelva un problema similar al ejemplo ilustrativo N° 2 para el cálculo de la media geométrica con datos sin agrupar. Resuelva manualmente empleando las dos ecuaciones presentadas y empleando Excel

3) Calcular la media geométrica para las siguientes calificaciones de Estadística de manera manual y con Excel

x_i	f_i
1	3
2	5
3	8
4	8
5	7
6	6
7	8
8	9
9	6
10	10

G = 5,23

4) Cree y resuelva un ejercicio similar al anterior.

5) Dado los siguientes datos:

19, 20, 21, 20, 19, 20, 21, 22, 22, 23, 24, 25, 26, 27, 28, 28, 29, 30, 31 y 33

5.1) Agrupe en intervalos de ancho 3.

5.2) Calcule la media geométrica manera manual y empleando Excel.

G = 24,15

6) Cree y resuelva un ejercicio similar al anterior.

2.3) MEDIA ARMÓNICA

La media armónica de una serie de números es el recíproco, o inverso, de la media aritmética de los recíprocos de dichos números, entendiéndose como recíproco al número que multiplicado por este nos da la unidad.

2.3.1) PROPIEDADES

- Es un promedio que se utiliza para el cálculo del costo promedio y todo tipo de variables expresadas en tasas o porcentajes.
- La media armónica no está definida en el caso de la existencia en el conjunto de valores nulos.
- Cuando la unidad constante o unidad de evaluación es igual a la unidad del numerador de una razón, se usa el promedio armónico, y si es igual a la unidad del denominador se usa el promedio aritmético.

2.3.2) MÉTODOS DE CÁLCULO

2.3.2.1) Para Datos No Agrupados

Sean los números x_1, x_2, \dots, x_n . La media armónica H se obtiene con la siguiente ecuación:

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

O con la siguiente ecuación:

$$H = \frac{1}{\frac{1}{n} \left(\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right)}$$

Ejemplo ilustrativo: La velocidad de producción de azúcar de tres máquinas procesadoras son 0,5, 0,3 y 0,4 minutos por kilogramo. Hallar el tiempo promedio de producción después de una jornada de 4800 minutos del proceso.

Solución:

Como en la razón minutos/kilogramos (min/kg) cada máquina trabaja 4800 min, la razón constante es el tiempo de trabajo (4800 min), es decir la constante es la unidad del numerador, por lo tanto se debe emplear el promedio armónico.

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{3}{\frac{1}{0,5} + \frac{1}{0,3} + \frac{1}{0,4}} = 0,383$$

O empleando la otra ecuación:

$$H = \frac{1}{\frac{1}{n} \left(\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right)} = \frac{1}{\frac{1}{3} \left(\frac{1}{0,5} + \frac{1}{0,3} + \frac{1}{0,4} \right)} = 0,383$$

El tiempo promedio de producción es 0,383 minutos por kilogramo de azúcar.

Empleando Excel se calcula de la siguiente manera:

a) Insertar la función MEDIA.ARMO y pulsar en Aceptar.

	A	B	C	D	E	F	G	H
1	0,5							
2	0,3							
3	0,4							
4	H	=						
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								

b) Seleccionar las celdas (Rango A₁:A3).

c) Pulsar en Aceptar.

	A	B	C	D
1	0,5			
2	0,3			
3	0,4			
4	H	0,38297872	=MEDIA.ARMO(A1:A3)	

2.3.2.2) Para Datos Agrupados en Tablas de Frecuencias

Se emplea cualquiera de las siguientes ecuaciones:

$$H = \frac{n}{\sum_{i=1}^n \frac{f_i}{x_i}} = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_n}{x_n}} \quad H = \frac{1}{\frac{1}{n} \left(\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_n}{x_n} \right)}$$

Ejemplo ilustrativo: En la siguiente tabla se presentan los datos sobre el tiempo en horas que se demoran en realizar la misma obra determinados obreros. Calcular el tiempo promedio que se demora en realizar la obra un obrero tipo (un obrero promedio).

Tiempo	Obreros
4	4
5	5
6	7
7	2
9	2

Solución:

$$H = \frac{n}{\sum_{i=1}^n \frac{f_i}{x_i}} = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_n}{x_n}} = \frac{20}{\frac{4}{4} + \frac{5}{5} + \frac{7}{6} + \frac{2}{7} + \frac{2}{9}} = \frac{20}{\frac{463}{126}} = \frac{2520}{463} = 5,44$$

En Excel se calcula de la siguiente manera:

Insertar la función = SUMA(B2:B6)/SUMAPRODUCTO((1/A2:A6);B2:B6) como se muestra en la siguiente figura:

	A	B	C	D	E	F	G
1	Tiempo	Obreros					
2	4	4					
3	5	5					
4	6	7					
5	7	2					
6	9	2					
7							
8	H	5,4427646	=SUMA(B2:B6)/SUMAPRODUCTO((1/A2:A6);B2:B6)				

2.3.2.3) Para Datos Agrupados en Intervalos

Se emplea la siguiente ecuación:

$$H = \frac{n}{\sum_{i=1}^n \frac{f_i}{xm_i}} = \frac{n}{\frac{f_1}{xm_1} + \frac{f_2}{xm_2} + \dots + \frac{f_n}{xm_n}}$$

Ejemplo ilustrativo: En la siguiente tabla se presentan los datos sobre el tiempo en minutos que se demoran para resolver una prueba de Estadística determinados estudiantes. Calcular el tiempo promedio que se demora en resolver la prueba un estudiante tipo.

Tiempo	Estudiantes
[40-50)	4
[50-60)	8
[60-70)	10
[70-80)	7
[80-90]	11

Solución:

Realizando los cálculos respectivos se obtiene:

x_i	f_i	xm_i	f_i/xm_i
[40-50)	4	45	0,089
[50-60)	8	55	0,145
[60-70)	10	65	0,154
[70-80)	7	75	0,093
[80-90]	11	85	0,129
Total	40		0,611

Aplicado la ecuación se obtiene:

$$H = \frac{n}{\sum_{i=1}^n \frac{f_i}{xm_i}} = \frac{n}{\frac{f_1}{xm_1} + \frac{f_2}{xm_2} + \dots + \frac{f_n}{xm_n}} = \frac{40}{0,611} = 65,47$$

En Excel se calcula insertando la siguiente función:

= SUMA(C2:C6)/SUMAPRODUCTO(1/(D2:D6);C2:C6), como se muestra en la siguiente figura:

	A	B	C	D	E	F	G	
1	x		f	xm				
2	40	50	4	45	=PROMEDIO(A2:B2)			
3	50	60	8	55	=PROMEDIO(A3:B3)			
4	60	70	10	65	=PROMEDIO(A4:B4)			
5	70	80	7	75	=PROMEDIO(A5:B5)			
6	80	90	11	85	=PROMEDIO(A6:B6)			
7								
8	H	65,4734	=SUMA(C2:C6)/SUMAPRODUCTO(1/(D2:D6);C2:C6)					

TAREA DE INTERAPRENDIZAJE N° 7

- 1) Realice un organizador gráfico sobre la media armónica.
- 2) Calcule la media armónica de manera manual y empleando Excel de los siguientes números:
2, 4, 6, 8, 9 y 10
H= 4,789
- 3) Cree y resuelva un ejercicio similar al anterior.
- 4) En una empresa se ha controlado el tiempo que tardan tres obreros en realizar una obra. Uno demora 8 horas, otro 6 horas y un tercero 4 horas.
 - 4.1) Halle de manera manual y empleando Excel el rendimiento de un obrero tipo (obrero promedio).
H= 5,534
 - 4.2) ¿Para qué le serviría a la empresa saber el rendimiento promedio de un obrero tipo?
- 5) Cree y resuelva un ejercicio similar al anterior.
- 6) Cree y resuelva dos ejercicios similares al ejemplo resuelto para el cálculo de la media armónica con datos agrupados en tablas de frecuencias.
- 7) En la siguiente tabla se presentan los datos sobre el tiempo en minutos que se demoran para resolver una prueba de Estadística determinados estudiantes.

Intervalo de tiempo	Nº de estudiantes
[45-50)	2
[50-55)	2
[55-60)	7
[60-65)	4
[65-70)	5
[75-80)	7
[85-90]	13

7.1) Calcule el tiempo promedio que se demora en resolver la prueba un estudiante tipo. Resolver de manera manual y empleando Excel.

$$H = 69,096$$

7.2) ¿Para qué le serviría al profesor saber el tiempo promedio que se demora en realizar la prueba un estudiante tipo?

8) Cree y resuelva un ejercicio similar al anterior

8.1) Realice los cálculos de manera manual y empleando Excel

8.2) Compruebe que la media geométrica es menor o igual que la media aritmética, y mayor o igual que la media armónica, es decir, en símbolos:

$$H \leq G \leq \bar{x}$$

9) ¿En qué caso ocurriría que la media geométrica sea igual a la media aritmética e igual a la media armónica?. Ponga un ejemplo y resuélvalo manera manual y empleando Excel.

2.4) LA MEDIANA

La mediana, llamada algunas veces media posicional, es el valor del término medio que divide una distribución de datos ordenados en dos partes iguales, es decir, el 50% de los datos se ubican sobre la mediana o hacia los puntajes altos y el 50% restante hacia los puntajes bajos.

2.4.1) PROPIEDADES

-La Mediana no tiene propiedades que le permite intervenir en desarrollos algebraicos como la media aritmética, sin embargo, posee propiedades que ponen en evidencia ciertas cualidades de un conjunto de datos, lo cual no ocurre con la media aritmética que promedia todos los valores y suprime sus individualidades. En cambio, la mediana destaca los valores individuales.

- Tiene la ventaja de no estar afectada por las observaciones extremas, ya que no depende de los valores que toma la variable, sino del orden de las mismas.

-Para el cálculo de la mediana interesa que los valores estén ordenados de menor a mayor.

- Su aplicación se ve limitada, ya que solo considera el orden jerárquico de los datos y no alguna propiedad propia de los datos, como en el caso de la media aritmética.

2.4.2) MÉTODOS DE CÁLCULO

2.4.2.1) Para Datos No Agrupados

a) Si el número n de datos es impar, la mediana es el dato que se encuentra a la mitad de la lista. Para calcular su posición se aplica la siguiente ecuación:

$$Md = x_{\frac{n+1}{2}} = x_{\frac{n+1}{2}}$$

Ejemplo ilustrativo:

Calcular la mediana de las siguientes calificaciones del curso de Estadística evaluadas sobre diez: 10, 8, 6, 4, 9, 7, 10, 9 y 6

Solución:

1) Se ordena los datos de menor a mayor:

4	6	6	7	8	9	9	10	10
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9

2) Se aplica la ecuación:

$$Md = x_{\frac{n+1}{2}}$$

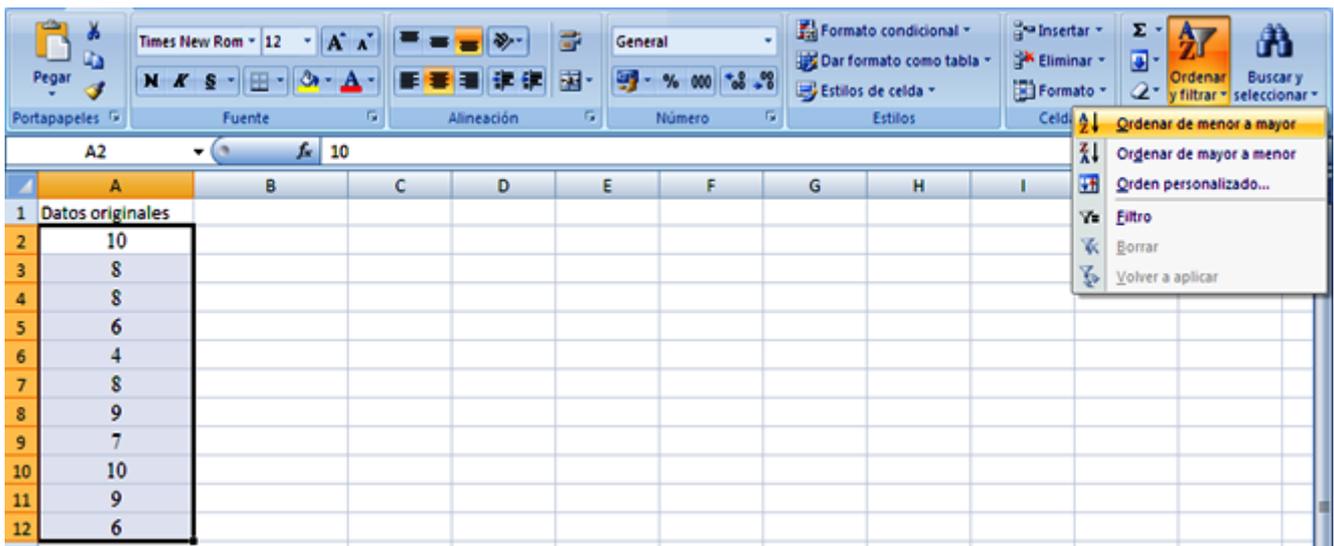
$$Md = x_{\frac{9+1}{2}} = Md = x_5$$

La mediana es el valor de x_5 (quinto dato), es decir, $Md=8$

En Excel se calcula así:

a) Ordenar los datos:

Se selecciona los datos y se pulsa en ordenar de menor a mayor.



	A
1	Datos ordenados de menor a mayor
2	4
3	6
4	6
5	7
6	8
7	8
8	8
9	9
10	9
11	10
12	10

b) Se inserta la función Mediana. Pulsar en Aceptar.

c) Seleccionar las celdas (Rango A2:A10).

d) Pulsar en Aceptar.

MEDIANA			
	A	B	C
1	Datos ordenados de menor a mayor		
2	4		
3	6		
4	6		
5	7		
6	8		
7	9		
8	9		
9	10		
10	10		
11			
12	Md	8	=MEDIANA(A2:A10)

O sin ordenar los datos, insertar la función MEDIANA(A1:I1) como se muestra en la siguiente figura:

MEDIANA									
	A	B	C	D	E	F	G	H	I
1	10	8	6	4	9	7	10	9	6
2									
3	Md		8						=MEDIANA(A1:I1)

b) Si el número n de datos es par, la mediana es la media aritmética de los dos datos que se encuentran a la mitad de la lista. Para calcular su posición se aplica la siguiente ecuación:

$$Md = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

Ejemplo ilustrativo: Calcular la mediana de las siguientes calificaciones del curso de Matemática evaluadas sobre diez: 10, 8, 9, 6, 4, 8, 9, 7, 10 y 9

Solución:

1) Se ordena los datos de menor a mayor:

4	6	7	8	8	9	9	9	10	10
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}

2) Se aplica la ecuación

$$Md = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_5 + x_6}{2} = \frac{8 + 9}{2} = 8,5$$

En Excel se calcula de la siguiente manera:

Se ordena los datos de menor a mayor e insertar la función MEDIANA(A1:A10) como se muestra en la siguiente figura:

MEDIANA				
	A	B	C	D
1	4			
2	6			
3	7			
4	8			
5	8			
6	9			
7	9			
8	9			
9	10			
10	10			
11	Md	8,5	=MEDIANA(A1:A10)	

2.4.2.2) Para Datos Agrupados en Tablas de Frecuencia

Para calcular la posición de la mediana se aplica la siguiente ecuación:

$$Md = \frac{n + 1}{2}$$

Ejemplo ilustrativo:

Dados los siguientes 20 números:

1, 3, 3, 5, 5, 5, 5, 2, 2, 2, 6, 6, 4, 4, 4, 4, 5, 5, 5, 5

1) Agrupar los datos en tabla de frecuencia.

Solución:

x	f
1	1
2	3
3	2
4	4
5	8
6	2
Total	20

2) Calcular la mediana.

Solución:

Calculando la posición de la mediana se obtiene:

$$Md = \frac{n + 1}{2} = \frac{20 + 1}{2} = 10,5$$

Como la posición de la mediana es 10,5, su valor es el promedio de los datos décimo y undécimo. Para observar con claridad cuáles son los datos décimo y undécimo se aconseja calcular la frecuencia acumulada.

x	f	fa
1	1	1
2	3	4
3	2	6
4	4	10
5	8	18
6	2	20
Total	20	

Se observa que el décimo dato es 4 y el undécimo es 5, por lo tanto:

$$Md = \frac{4 + 5}{2} = 4,5$$

2.4.2.3) Para Datos Agrupados en Intervalos

a) Por interpolación

Ejemplo ilustrativo: Calcular la mediana de los pesos de un grupo de 50 personas que se distribuyen de la siguiente manera:

Intervalos	f
[45,55)	6
[55,65)	10
[65,75)	19
[75,85)	11
[85,95)	4

Solución:

Primero se calcula $n/2$ y después se averigua el intervalo en el que está la mediana, este intervalo recibe el nombre de intervalo o clase de la mediana. Para averiguar el intervalo en el que está la mediana se aconseja calcular la frecuencia acumulada.

$$\frac{n}{2} = \frac{50}{2} = 25$$

Intervalos	f	fa
[45,55)	6	6
[55, 65)	10	16
[65, 75)	19	35
[75, 85)	11	46
[85, 95)	4	50

En este ejemplo el intervalo de la media es [65,75). Se observa que 16 valores están por debajo del valor 65. Los 9 que faltan para llegar a 25 se interpolan en el ancho del intervalo de la mediana que en este ejemplo es 10.

19 corresponde a 10

1 corresponde a 10/19

$$9 \text{ corresponde a } 9 \cdot \frac{10}{19} = \frac{90}{19} = 4,737$$

Por lo tanto la Mediana es igual $65 + 4,737 = 69,737$

b) Empleando la ecuación

$$Md = Li_{md} + \left(\frac{\frac{n}{2} - Fa}{f_{md}} \right) \cdot c$$

En donde:

Li_{md} = Límite inferior del intervalo de clase de la mediana.

n = número total de datos.

Fa = Frecuencia acumulada del intervalo de clase que antecede al intervalo de clase de la mediana.

f_{md} = Frecuencia absoluta del intervalo de clase de la mediana.

c = Ancho del intervalo de clase de la mediana.

Ejemplo ilustrativo: Calcular la mediana del ejemplo anterior y representarla mediante un histograma de frecuencias acumuladas.

Se calcula la frecuencia acumulada como se muestra en la siguiente tabla:

Intervalos	f	fa
[45,55)	6	6
[55,65)	10	16
[65,75)	19	35
[75,85)	11	46
[85,95)	4	50

Solución:

Se calcula la posición de la mediana de la siguiente manera:

$$\frac{n}{2} = \frac{50}{2} = 25$$

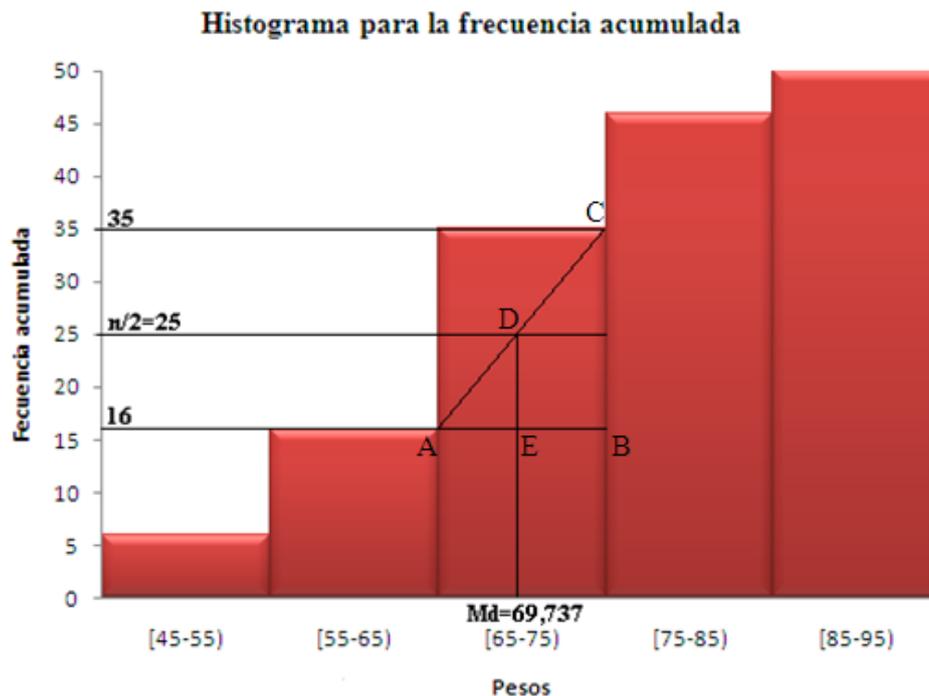
Por lo tanto el intervalo o clase de la mediana es [65,75).

Al aplicar la ecuación respectiva se obtiene:

$$Md = Li_{md} + \left(\frac{\frac{n}{2} - Fa}{f_{md}} \right) \cdot c \Rightarrow Md = 65 + \left(\frac{\frac{50}{2} - 16}{19} \right) \cdot 10 = 65 + \left(\frac{9}{19} \right) \cdot 10 = 65 + \frac{90}{19} = 69,737$$

c) Resolviendo de manera gráfica

A continuación se presenta un histograma para la frecuencia acumulada.



Observando el gráfico se determina que $Md = 65 + AE$

Los triángulos ABC y AED son semejantes, por lo que se cumple:

$$\frac{AB}{CB} = \frac{AE}{DE}$$

$$\frac{75 - 65}{35 - 16} = \frac{AE}{25 - 16} \Rightarrow \frac{10}{19} = \frac{AE}{9}$$

Despejando AE se obtiene:

$$\frac{10}{19} \cdot 9 = AE \Rightarrow AE = \frac{90}{19} = 4,737$$

Entonces, $Md = 65 + AE = 65 + 4,737 = \rightarrow Md = 69,737$

TAREA DE INTERAPRENDIZAJE N° 8

- 1) Escriba 3 diferencias entre media aritmética y mediana.
- 2) Realice un organizador gráfico sobre la mediana.
- 3) Calcule la mediana de los números 6, 6, 5, 2, 3, 4, 4, 5, 5, de manera manual y empleando Excel.
Md= 5
- 4) Cree y resuelva un ejercicio similar al anterior
- 5) Calcule la mediana de los números 11, 12, 9, 10, 7, 8, de manera manual y empleando Excel.
Md= 9,5
- 6) Cree y resuelva un ejercicio similar al anterior.

7) Dados los siguientes 35 números:

2, 2, 2, 2, 2, 4, 4, 4, 4, 4, 6, 6, 6, 6, 6, 4, 4, 4, 4, 8, 8, 8, 6, 6, 6, 6, 6, 10, 10, 10, 8, 8, 8, 10, 10 y 10

7.1) Calcule la mediana sin agrupar los datos de manera manual y empleando Excel.

Md=6

7.2) Calcule la mediana agrupando los datos en una tabla de frecuencias.

Md=6

8) Cree y resuelva un ejercicio similar al anterior.

9) Calcule la mediana de las siguientes edades de personas y representarla mediante un histograma para la frecuencia acumulada.

Intervalos	f
[60, 63)	5
[63, 66)	18
[66, 69)	42
[69, 72)	27
[72, 75)	8

Md= 67,93

10) Cree y resuelva un ejercicio similar al anterior.

11) Dados los siguientes números:

50, 55, 59, 60, 69, 65, 66, 69, 63, 64, 70, 72, 77, 78, 79, 79, 77, 78, 71, 72, 73, 75, 77, 74, 73, 73, 74, 77, 80, 82, 85, 88, 89, 89, 85, 81, 82, 83, 82, 81, 90, 91, 92, 93, 94, 95, 96, 99, 100 y 109

11.1) Agrupe los datos en intervalos de ancho 10.

11.2) Calcule la media aritmética de manera manual y empleando Excel.

78,7

11.3) Calcule la media geométrica de manera manual y empleando Excel.

77,77

11.4) Calcule la media armónica de manera manual y empleando Excel.

76,81

11.5) Calcule la mediana por interpolación, empleando la ecuación y empleando un histograma para la frecuencia acumulada.

78,33

12) Cree y resuelva un ejercicio similar al anterior.

2.5) MEDIDAS DE POSICIÓN

Son similares a la mediana en que también subdividen una distribución de mediciones de acuerdo con la proporción de frecuencias observadas. Mientras que la mediana divide a una distribución en mitades, los cuartiles (Q) la dividen en cuartos, los deciles (D) la dividen en décimos y los puntos percentiles (P) la dividen en centésimos.

Colectivamente, cuartiles, deciles y percentiles se denominan cuantiles. Puesto que sirven para ubicar datos particulares dentro de ciertas porciones de una distribución de datos, toman el nombre de medidas de posición.

2.5.1) CUARTILES.- Son cada uno de los 3 valores Q_1, Q_2, Q_3 que dividen a la distribución de los datos en 4 partes iguales.

2.5.1.1) Propiedades

Los cuartiles son un caso particular de los percentiles. Hay 3 cuartiles:

Primer cuartil: $Q_1 = P_{25}$, segundo cuartil: $Q_2 = D_5 = P_{50} = \text{Mediana}$, tercer cuartil: $Q_3 = P_{75}$

2.5.1.2) Métodos de Cálculo

a) Para Datos No Agrupados

La posición o ubicación de los cuartiles se encuentra aplicando la siguiente ecuación:

$$Q_k = X_{\left[\frac{n \cdot k + 1}{4}\right]} = X_{\left[\frac{n \cdot k + 2}{4}\right]}$$

Donde:

n = número total de datos

k = número del cuartil

Ejemplo ilustrativo:

Encuentre los cuartiles dada la siguiente distribución, y representelos gráficamente mediante un diagrama de caja y bigotes: 6, 9, 9, 12, 12, 12, 15 y 17

Solución:

Para calcular los cuartiles se ordena los datos de menor a mayor

6	9	9	12	12	12	15	17
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8

Aplicando la ecuación para el cuartil uno se obtiene:

$$Q_k = X_{\left[\frac{n \cdot k + 2}{4}\right]}$$

$$Q_1 = X_{\left[\frac{n+2}{4}\right]} = X_{\left[\frac{8+2}{4}\right]} = X_{\left[\frac{10}{4}\right]} = X_{2,5}$$

Como la posición del cuartil 1 es 2,5, su valor es el promedio de los datos segundo y tercero

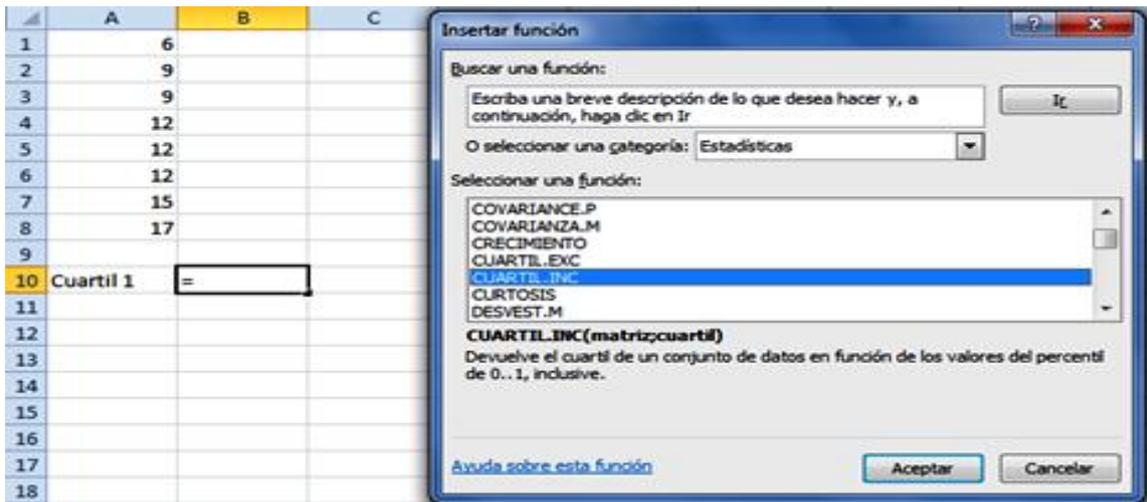
$$Q_1 = X_{2,5} = \frac{x_2 + x_3}{2} = \frac{9 + 9}{2} = 9$$

O también la posición 2,5 dice que el cuartil 1 está ubicado al 50% del trayecto comprendido entre el segundo dato, que es 9 y el tercer dato que es 9, es decir, $Q_1 = 9 + 0,5(9-9) = 9$

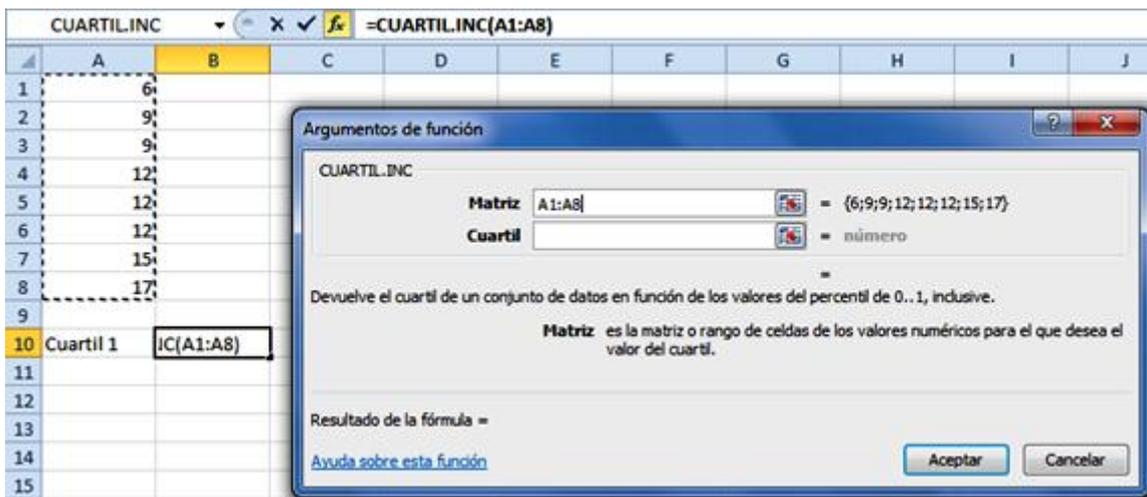
Interpretación: Este resultado indica que el 25% de los datos es inferior a 9

En Excel se calcula de la siguiente manera:

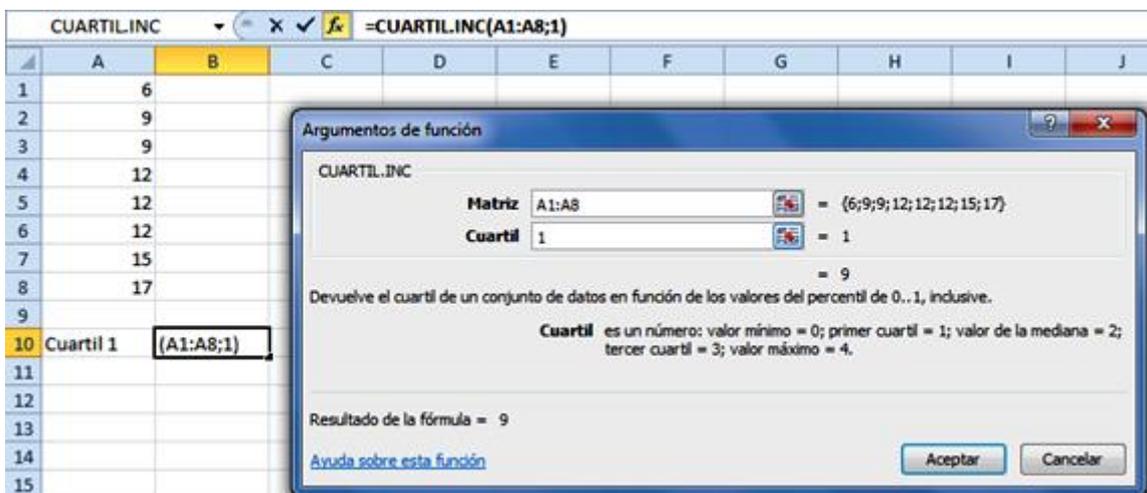
a) Se inserta la función CUARTIL.INC y se pulsa en Aceptar.



b) Seleccionar las celdas (Rango A1:A8).



c) Escribir 1 en la opción Cuartil en la ventana de los argumentos la función.



d) Pulsar en Aceptar.

	A	B	C	D
1	6			
2	9			
3	9			
4	12			
5	12			
6	12			
7	15			
8	17			
9				
10	Cuartil 1	9	=CUARTIL.INC(A1:A8;1)	

Aplicando la ecuación para el cuartil dos se obtiene:

$$Q_k = X_{\left[\frac{n \cdot k + 2}{4}\right]}$$

$$Q_2 = X_{\left[\frac{n \cdot 2 + 2}{4}\right]} = X_{\left[\frac{2n + 2}{4}\right]} = X_{\left[\frac{2 \cdot 8 + 2}{4}\right]} = X_{\left[\frac{16 + 2}{4}\right]} = X_{4,5} = \frac{x_4 + x_5}{2} = \frac{12 + 12}{2} = 12$$

O también la posición 4,5 dice que el cuartil 2 está ubicado al 50% del trayecto comprendido entre el cuarto dato, que es 12 y el quinto dato que también es 12, es decir,

$$Q_2 = 12 + 0,5(12 - 12) = 12$$

Interpretación: Este resultado indica que el 50% de los datos es inferior a 12

En Excel se calcula de la siguiente manera:

a) Repetir los pasos para el cuartil 1, y en la opción de cuartil escribir 2



b) Pulsar en Aceptar

	A	B	C	D
1	6			
2	9			
3	9			
4	12			
5	12			
6	12			
7	15			
8	17			
9				
10	Cuartil 2	12	=CUARTIL.INC(A1:A8;2)	

Aplicando la ecuación para el cuartil tres se obtiene:

$$Q_k = X_{\left[\frac{n \cdot k + 2}{4}\right]}$$

$$Q_3 = X_{\left[\frac{3n+2}{4}\right]} = X_{\left[\frac{3 \cdot 8 + 2}{4}\right]} = X_{\left[\frac{24+2}{4}\right]} = X_{\frac{26}{4}} = X_{6,5} = \frac{x_6 + x_7}{2} = \frac{12 + 15}{2} = 13,5$$

O también la posición 6,5 dice que el cuartil 2 está ubicado al 50% del trayecto comprendido entre el doceavo dato, que es 12 y el quinceavo dato que es 15, es decir, $Q_3 = 12 + 0,5(15 - 12)$

$$Q_3 = 12 + 0,5(3) = 12 + 1,5 = 13,5$$

Interpretación: Este resultado indica que el 75% de los datos es inferior a 13,5

En Excel se calcula de la siguiente manera:

Repetir los pasos para el cuartil 1, y en la opción de cuartil escribir 3. Pulsar en Aceptar.

CUARTIL.INC				
	A	B	C	D
1	6			
2	9			
3	9			
4	12			
5	12			
6	12			
7	15			
8	17			
9				
10	Cuartil 3	12,75	=CUARTIL.INC(A1:A8;3)	

Notas importantes:

-Los cálculos en Excel para un número impar de datos coinciden con los cálculos realizados con las ecuaciones.

-Para un número par de datos, aunque en ciertas ocasiones coinciden, suele existir diferencias en los cálculos del Q_1 y Q_3 realizados con Excel. Este error de cálculo es: $e = 0,25d$, en donde d es la distancia de separación de los datos

-Para el Q_1 se resta el error al valor obtenido con Excel

-Para el Q_3 se suma el error al valor obtenido con Excel

En nuestro ejemplo $e = 0,25(x_7 - x_6) = 0,25(15 - 12) = 0,25(3) = 0,75$. Al suma el error al valor Q_3 inicialmente calculado con Excel se obtiene el valor correcto como se muestra en la siguiente figura:

CUARTIL.INC X ✓ fx =				
	A	B	C	D
1	6			
2	9			
3	9			
4	12			
5	12			
6	12			
7	15			
8	17			
9				
10	Cuartil 3	13,5	=CUARTIL.INC(A1:A8;3)+0,75	

Para elaborar un diagrama de caja y bigotes es necesario saber:

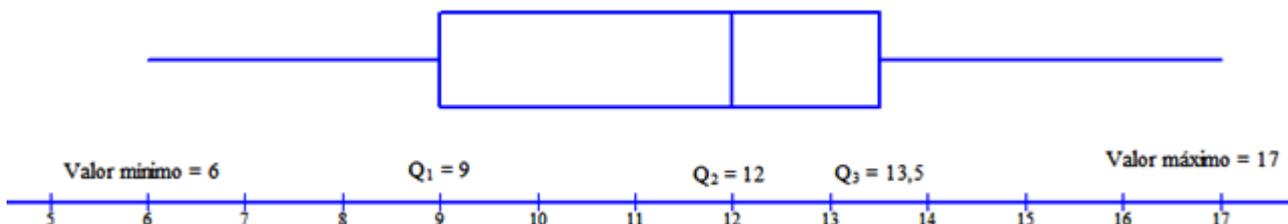
Un **diagrama de caja y bigotes** es una representación gráfica que ayuda a visualizar una distribución de datos: caja desde Q_1 a Q_3 (50% de los datos), y bigotes el recorrido (distancia desde valor mínimo hasta el valor máximo).

Para elaborar un diagrama de caja se procede de la siguiente manera:

- Se marca los valores de la serie de datos sobre el eje horizontal o vertical.
- Se ubica sobre el eje el valor mínimo, primer cuartil, mediana o segundo cuartil, tercer cuartil y el valor máximo.
- Se construye un rectángulo (caja) paralelo al eje, de longitud desde Q_1 a Q_3 y anchura arbitraria.

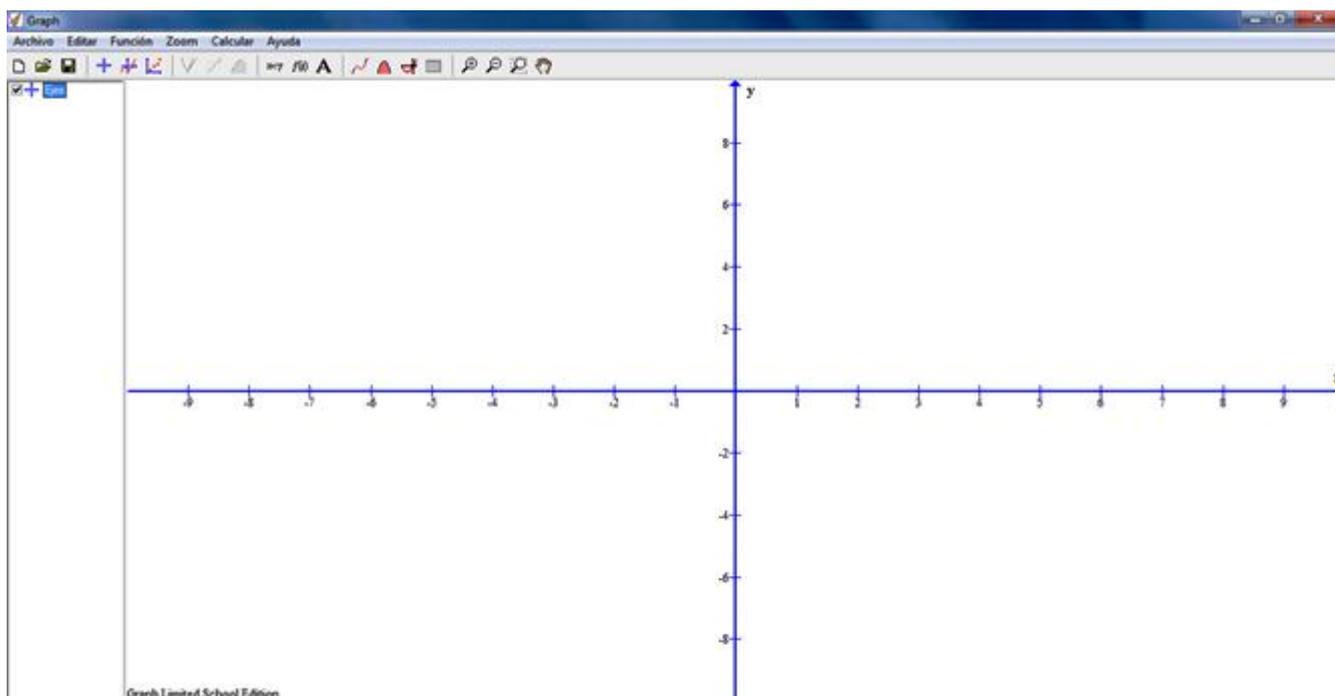
De acuerdo al ejemplo ilustrativo se tiene:

Valor mínimo = 6
 $Q_1 = 9$
 $Q_2 = 12$
 $Q_3 = 13,5$
 Valor máximo = 17

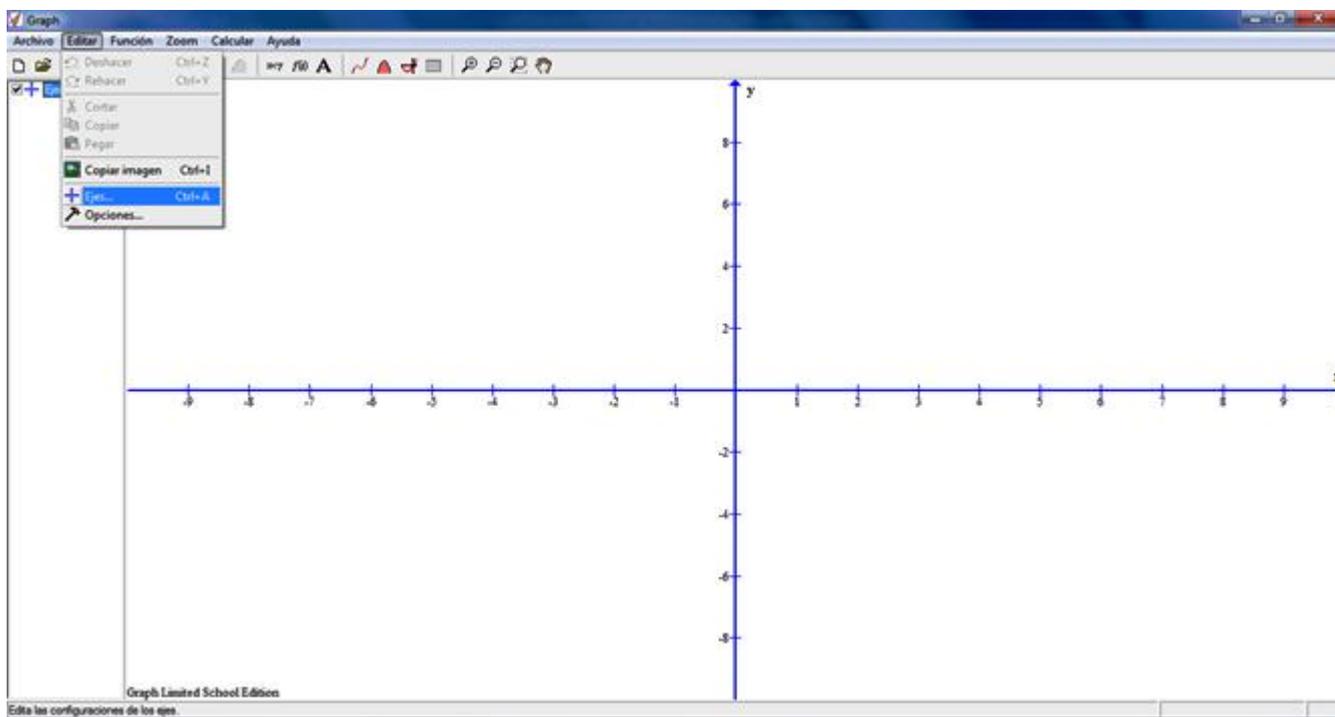


El diagrama de caja y bigotes en Graph se elabora de la siguiente manera:

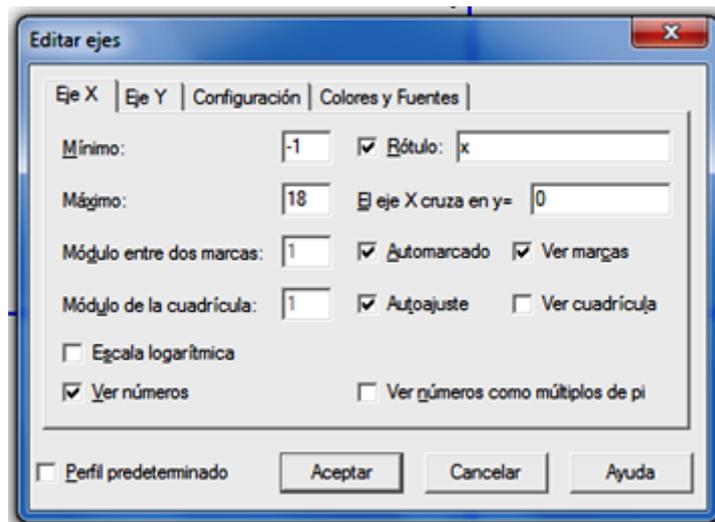
a) Ingresar al programa



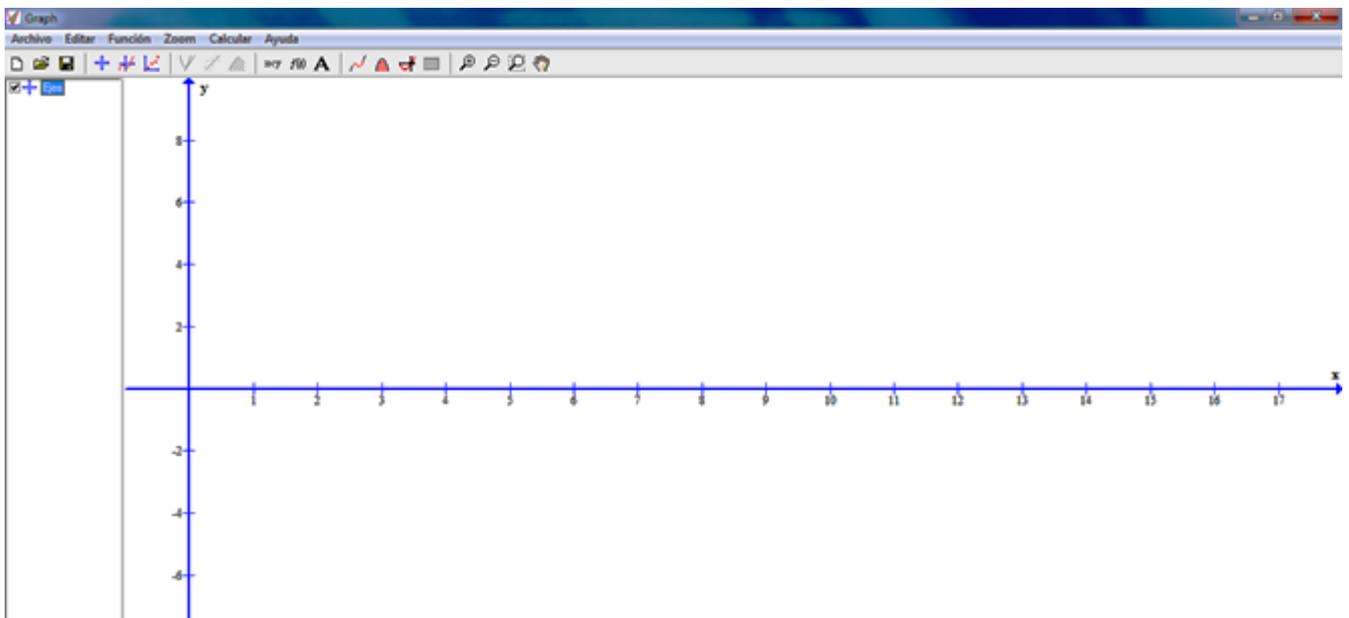
b) Clic en Editar



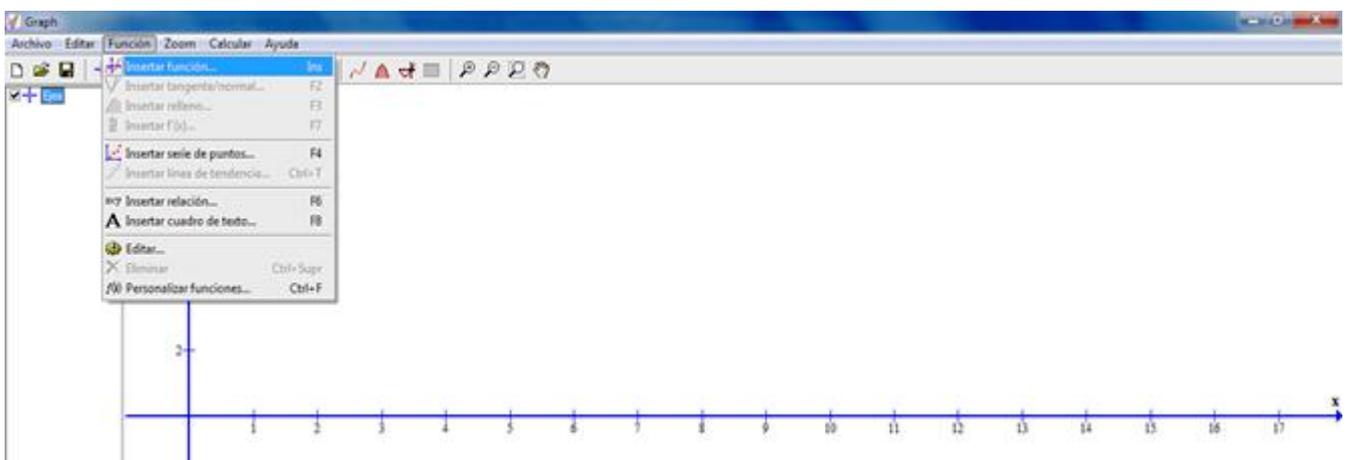
c) Clic en Ejes. En el eje X, en Mínimo escribir -1, y en Máximo 18



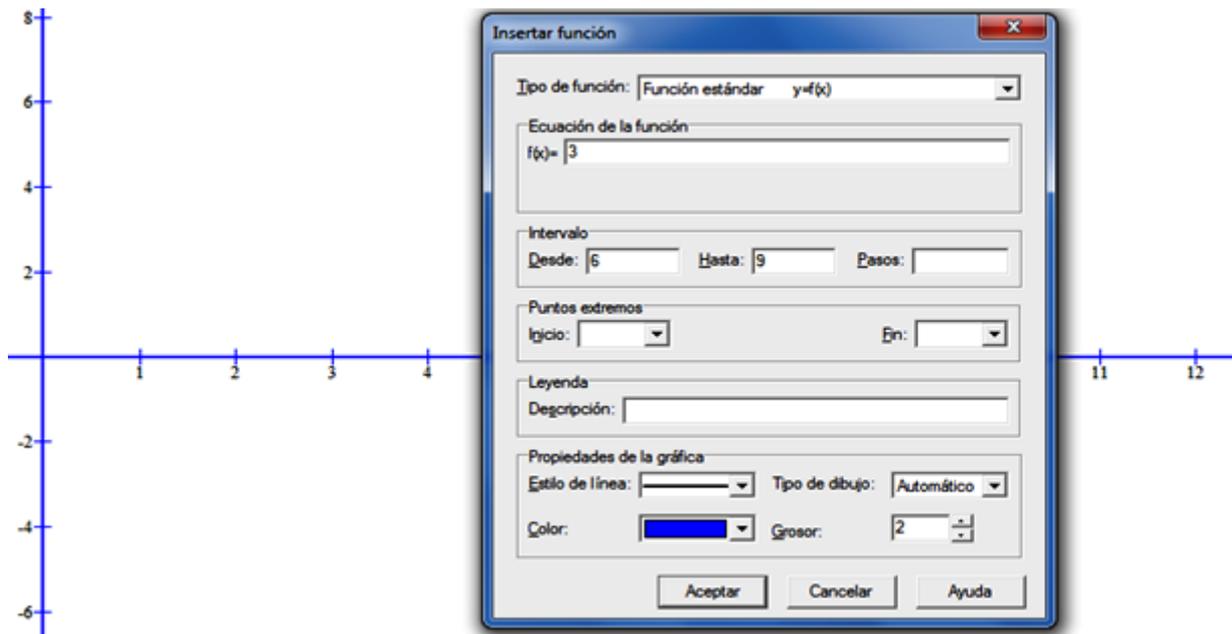
d) Clic en Aceptar



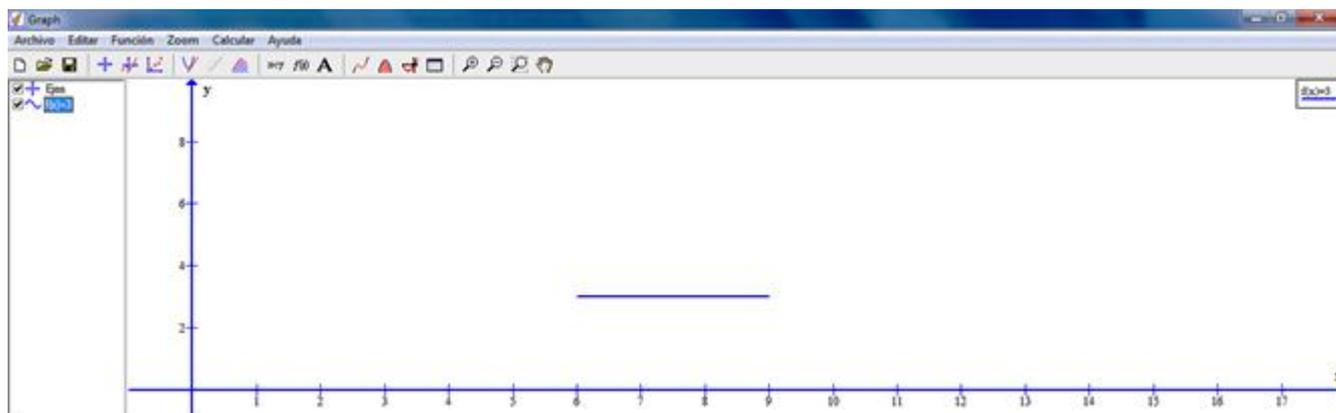
e) Clic en Función



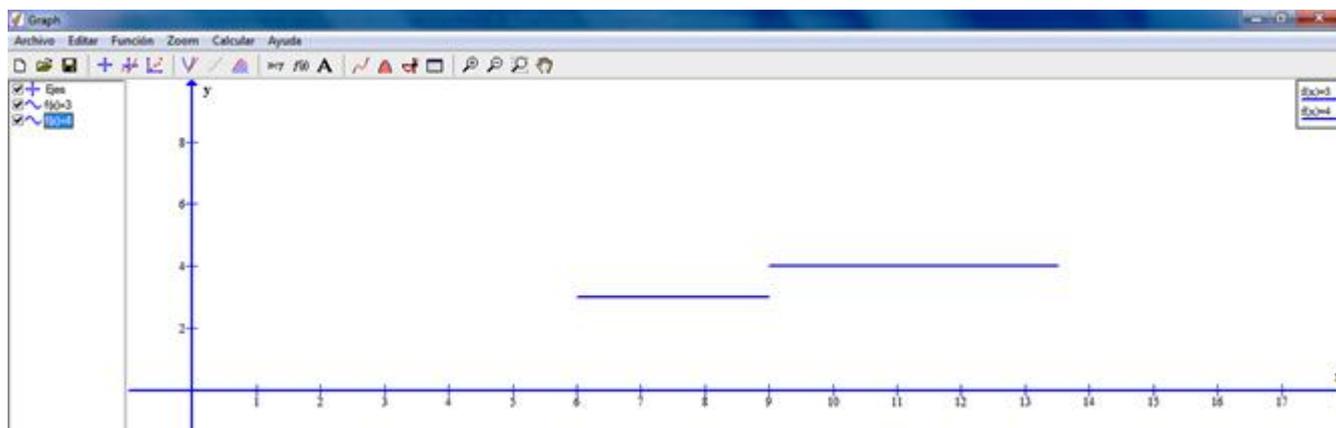
f) Clic en Insertar Función. En Ecuación de la función $f(x)$, escribir 3. En Intervalo, en Desde, escribir 6, y en Hasta, escribir 9.



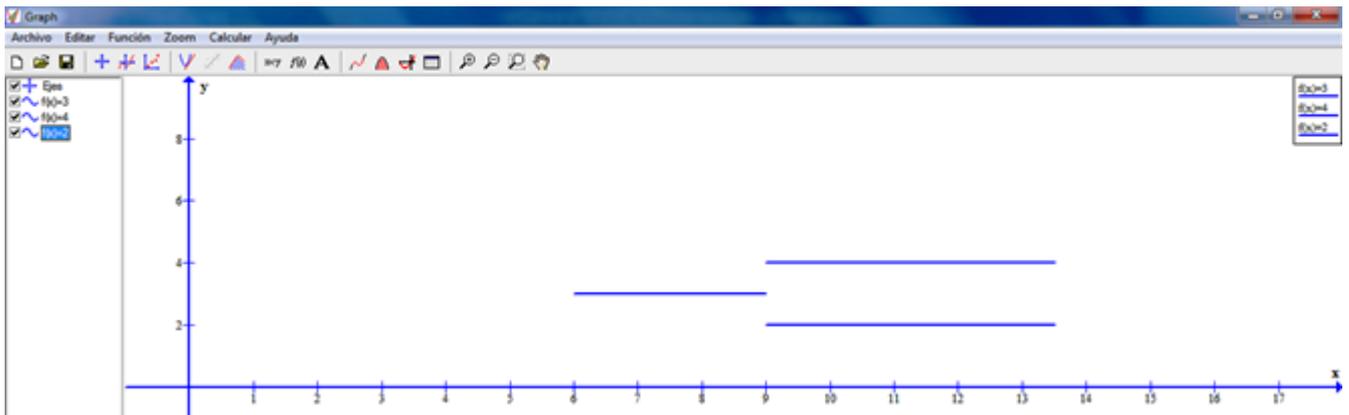
g) Clic en Aceptar



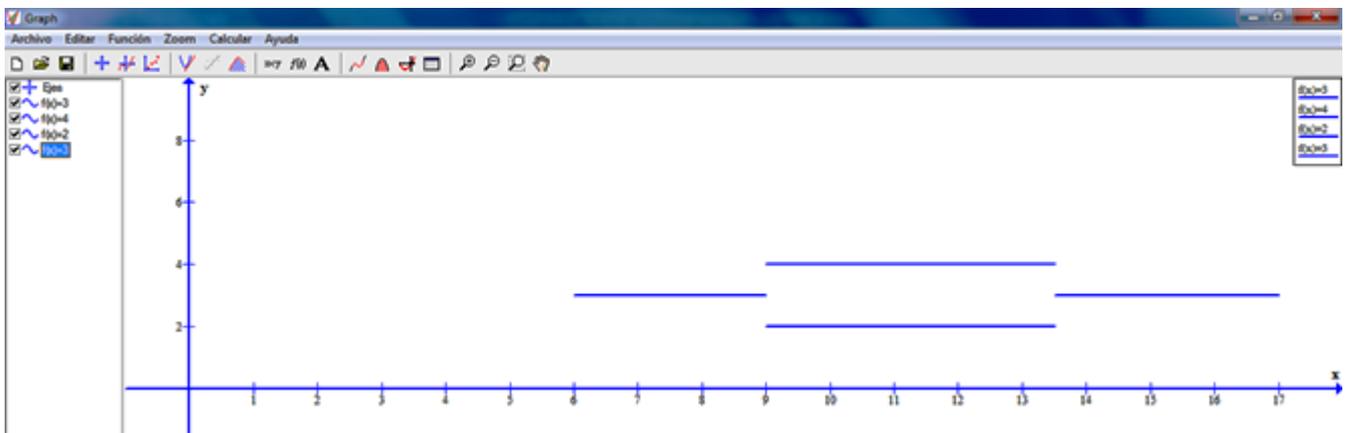
h) Clic en Función. Clic en Insertar Función. En Ecuación de la función $f(x)$, escribir 4. En Intervalo, en Desde, escribir 9, y en Hasta, escribir 13.5. Clic en Aceptar.



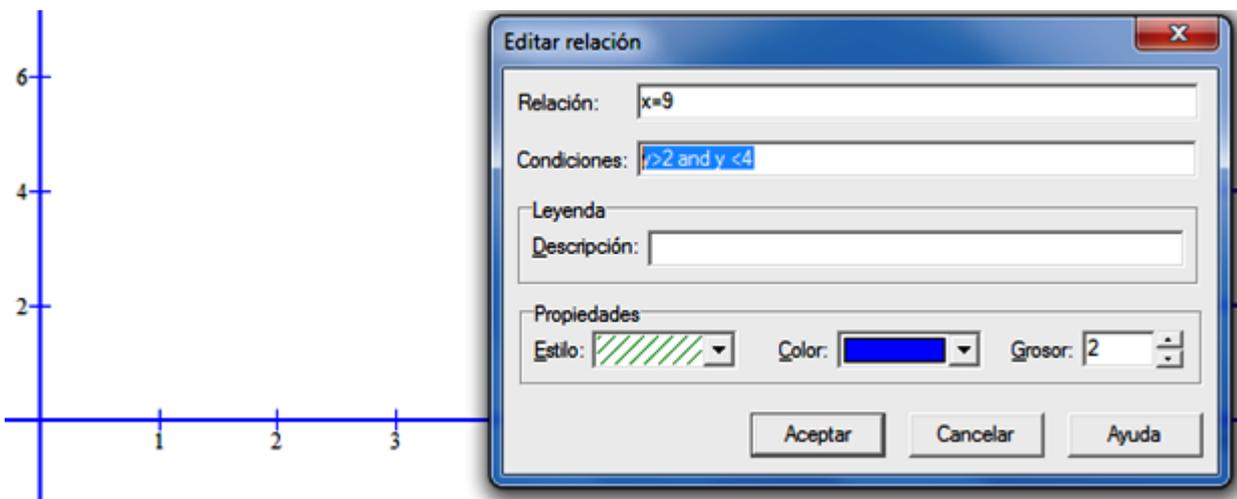
i) Clic en Función. Clic en Insertar Función. En Ecuación de la función $f(x)$, escribir 2. En Intervalo, en Desde, escribir 9, y en Hasta, escribir 13.5. Clic en Aceptar.



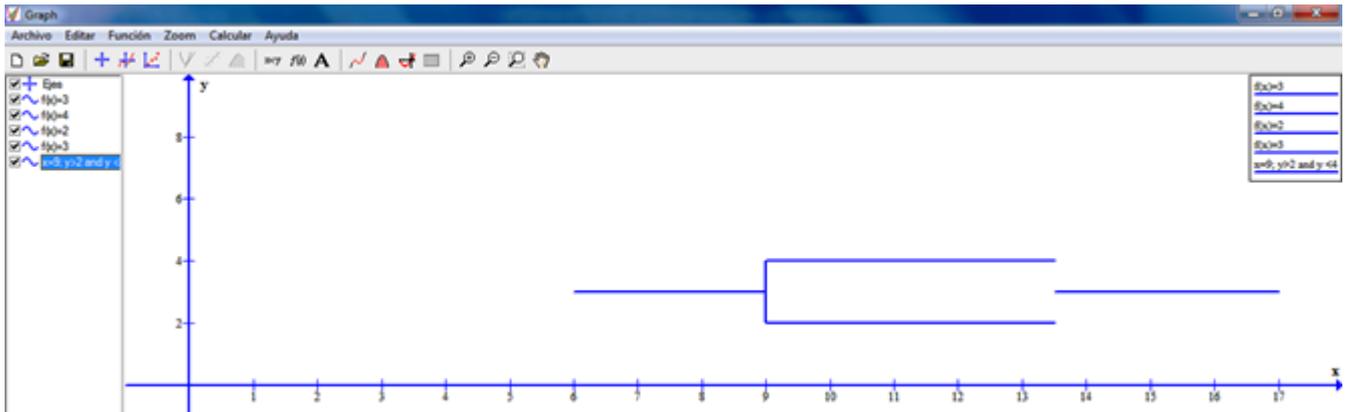
j) Clic en Función. Clic en Insertar Función. En Ecuación de la función $f(x)$, escribir 3. En Intervalo, en Desde, escribir 13.5, y en Hasta, escribir 17. Clic en Aceptar.



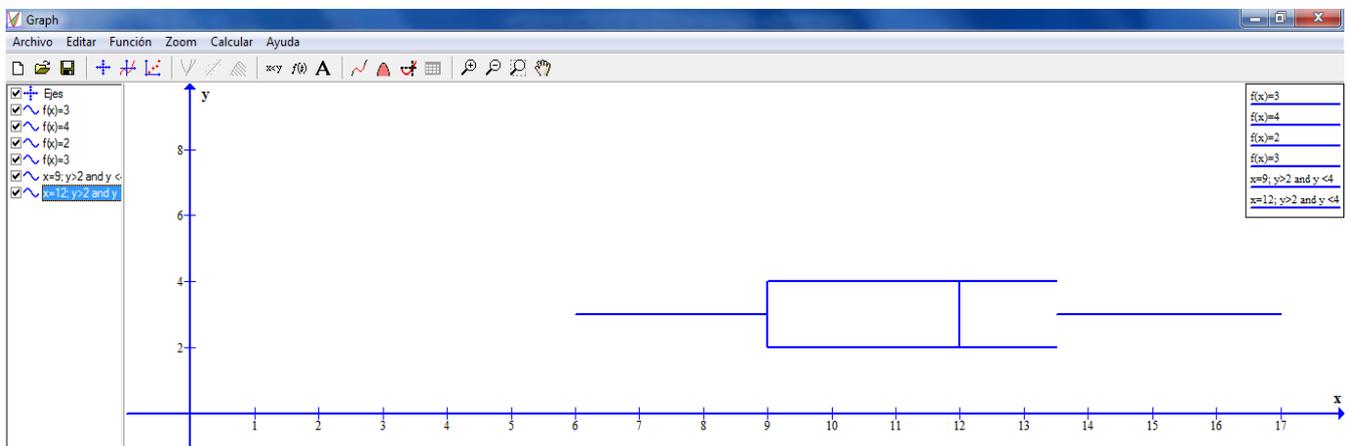
k) Clic en Función. Clic en Insertar Relación. En Relación, escribir $x=9$. En Condiciones, escribir $y > 2$ and $y < 4$



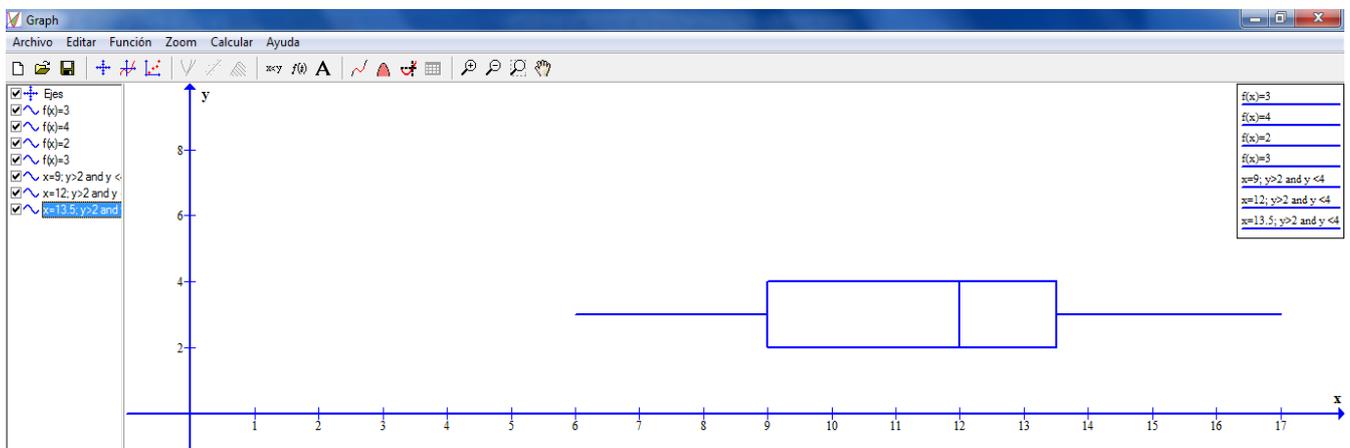
l) Clic en Aceptar



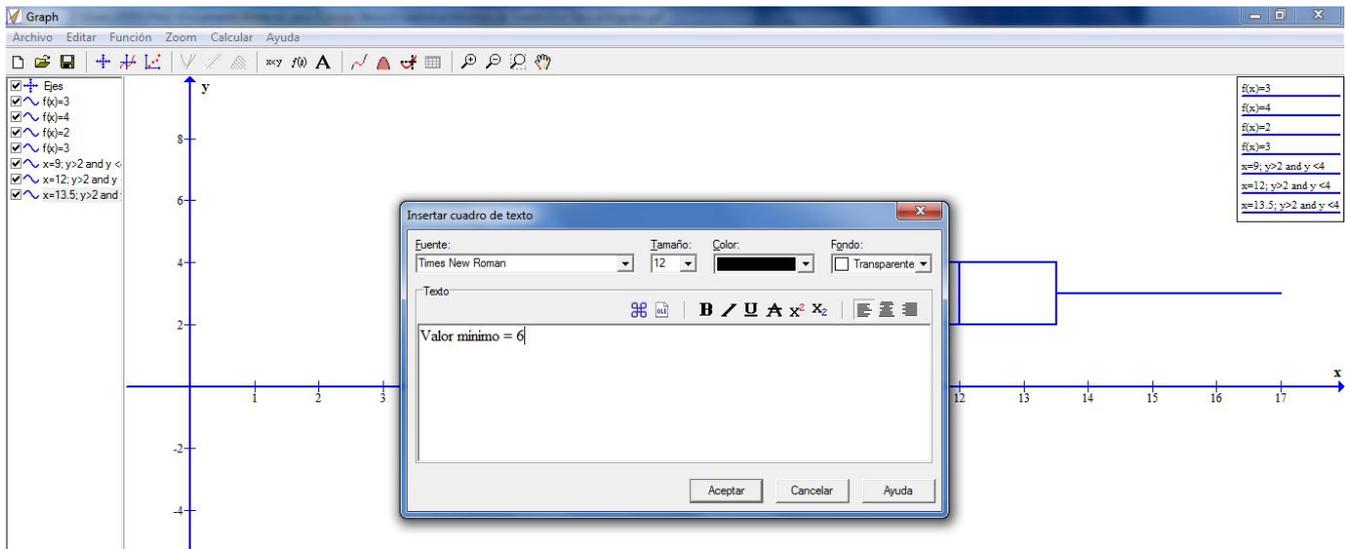
m) Clic en Función. Clic en Insertar Relación. En Relación, escribir $x=12$. En Condiciones, escribir $y>2$ and $y <4$. Clic en Aceptar



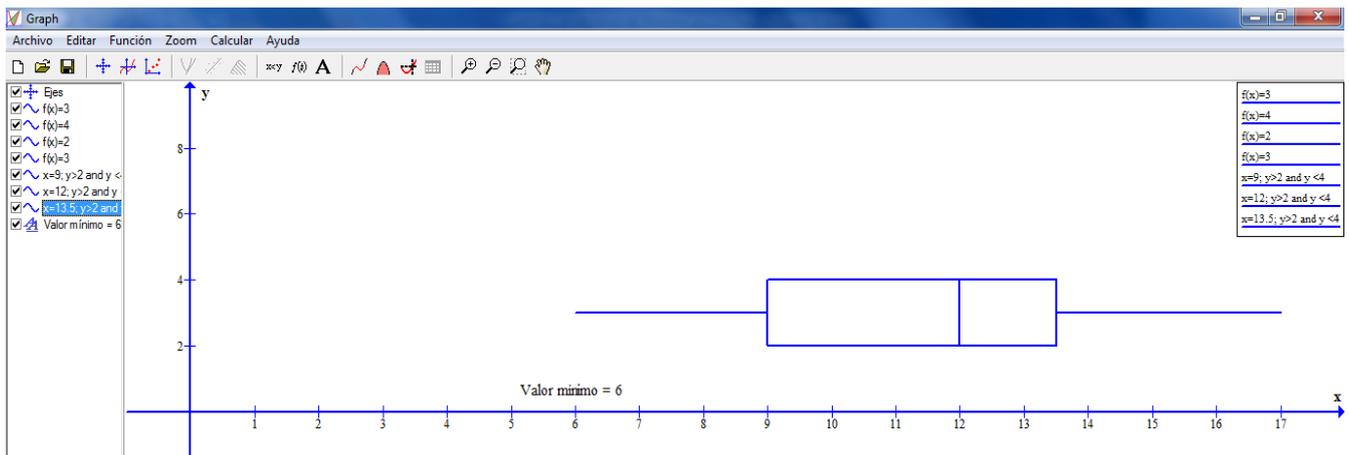
n) Clic en Función. Clic en Insertar Relación. En Relación, escribir $x=13.5$. En Condiciones, escribir $y>2$ and $y <4$. Clic en Aceptar



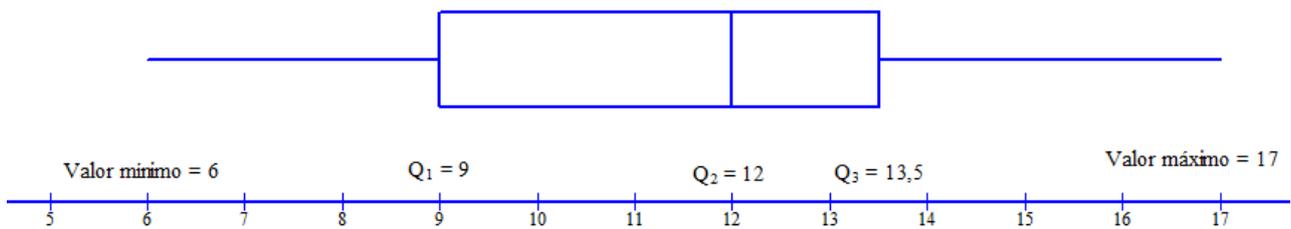
ñ) Clic en Función y luego clic en insertar cuadro de texto, o teclear (F8). En el cuadro de texto escribir Valor mínimo = 6



o) Clic en Aceptar. Arrastrar el texto escrito al lugar deseado.



p) Repetir los pasos ñ) y o) para escribir $Q_1 = 9$, $Q_2 = 12$, $Q_3 = 13,5$ y Valor máximo = 17. De esta manera queda elaborado el diagrama de caja y bigotes.



b) Para Datos Agrupados en Tablas de Frecuencias

Se aplica la misma ecuación empleada para el cálculo en los datos no agrupados

Ejemplo ilustrativo: Dada la siguiente tabla:

x	f
6	1
9	2
12	3
15	1
17	1

1) Calcular el cuartil 2

2) Representar los cuartiles en un histograma para la fra(%) (Frecuencia relativa acumulada medida en porcentajes). Determinar gráficamente el valor de los cuartiles

Solución:

1) Cálculo del cuartil 2

Aplicando la primera ecuación para el cuartil dos se obtiene:

$$Q_k = X_{\left[\frac{n \cdot k + 2}{4}\right]}$$
$$Q_2 = X_{\left[\frac{n \cdot 2 + 2}{4}\right]} = X_{\left[\frac{2(n+1)}{4}\right]} = X_{\left[\frac{n+1}{2}\right]} = X_{\left[\frac{8+1}{2}\right]} = X_{\left[\frac{9}{2}\right]} = X_{4,5}$$

Como la posición del cuartil 2 es 4,5, su valor es el promedio de los datos cuarto y quinto

Para observar con claridad cuáles son los datos cuarto y quinto se aconseja calcular la frecuencia acumulada

x	f	fa
6	1	1
9	2	3
12	3	6
15	1	7
17	1	8

Se observa que el cuarto dato es 12 y el quinto dato es 12, por lo tanto

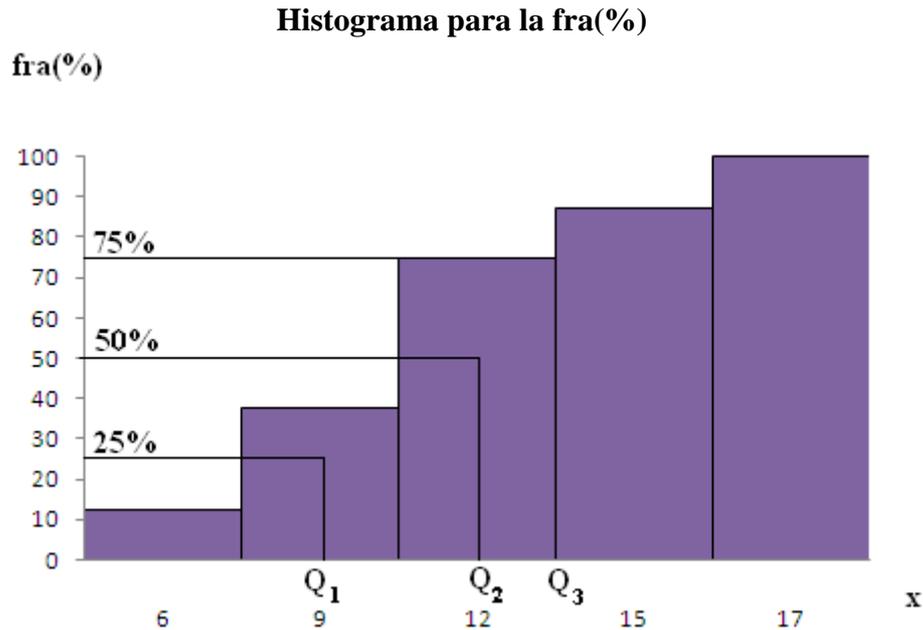
$$Q_2 = X_{4,5} = \frac{x_4 + x_5}{2} = \frac{12 + 12}{2} = 12$$

2) Representando los cuartiles en un histograma para la fra(%)

Calculando la fra(%) se obtiene:

x	f	fa	fr	fra	$fra (\%)$
6	1	1	0,125	0,125	12,5
9	2	3	0,25	0,375	37,5
12	3	6	0,375	0,75	75
15	1	7	0,125	0,875	87,5
17	1	8	0,125	1	100
n	8				

A continuación se presenta el gráfico solicitado elaborado en Excel y Paint:



Observando en el gráfico anterior se observa que $Q_1 = 9$, $Q_2 = 12$ y $Q_3 = (12 + 5)/2 = 13$,

c) Para Datos Agrupados en Intervalos

Se emplea la siguiente ecuación:

$$Q_k = Li_Q + \left(\frac{\frac{nk}{4} - Fa}{f_Q} \right) \cdot c$$

Donde

Li_p = Límite inferior del intervalo de clase del cuartil

n = número total de datos

Fa = Frecuencia acumulada del intervalo de clase que antecede al intervalo de clase del cuartil

f_Q = Frecuencia absoluta del intervalo de clase del cuartil

c = Ancho del intervalo de clase del cuartil

Ejemplo ilustrativo: Dado los siguientes datos sobre pesos de un grupo de 50 personas:

Intervalos	f
45- 55	6
55- 65	10
65- 75	19
75- 85	11
85- 95	4

1) Calcular los cuartiles empleando la ecuación

2) Calcular los cuartiles empleando un histograma para fra(%) (Frecuencia relativa acumulada mediada en porcentajes)

Solución:

1) Cálculo de los cuartiles empleando la ecuación

1.1) Cálculo del primer cuartil

Primero se calcula $nk/4$ y después se averigua el intervalo en el que está el cuartil, este intervalo recibe el nombre de intervalo o clase del primer cuartil. Para averiguar el intervalo en el que están los cuartiles se aconseja calcular la frecuencia acumulada

$$\frac{n \cdot k}{4} = \frac{50 \cdot 1}{4} = 12,5$$

Intervalos	f	fa
45- 55	6	6
55- 65	10	16
65- 75	19	35
75- 85	11	46
85- 95	4	50
n	50	

Por lo tanto en este ejemplo:

El intervalo del segundo cuartil es 55-65.

El número total de datos es $n=10$

Se observa que 6 valores están por debajo del valor 55, es decir $Fa=6$.

La frecuencia absoluta f_Q del intervalo del cuartil es 10

El ancho del intervalo del cuartil es $c=65-55=10$.

Al aplicar la ecuación se obtiene:

$$Q_k = Li_Q + \left(\frac{\frac{nk}{4} - Fa}{f_Q} \right) \cdot c$$

$$Q_1 = 55 + \left(\frac{\frac{50 \cdot 1}{4} - 6}{10} \right) \cdot 10 = 55 + \left(\frac{\frac{50}{4} - 6}{10} \right) \cdot 10 = 55 + \left(\frac{13}{20} \right) \cdot 10 = 55 + 6,5$$

$$Q_1 = 61,5$$

1.2) Cálculo del segundo cuartil

Primero se calcula $nk/4$ y después se averigua el intervalo en el que está el cuartil, este intervalo recibe el nombre de intervalo o clase del cuartil.

$$\frac{n \cdot 2}{4} = \frac{50 \cdot 2}{4} = 25$$

Por lo tanto para el segundo cuartil se tiene:

Intervalo: 65-75

$n=10$

$Fa=16$

$f_Q=19$

$c=75-65=10$

Al aplicar la ecuación se obtiene:

$$Q_k = Li_Q + \left(\frac{\frac{nk}{4} - Fa}{f_Q} \right) \cdot c$$

$$Q_2 = 65 + \left(\frac{\frac{50 \cdot 2}{4} - 16}{19} \right) \cdot 10 = 65 + \left(\frac{\frac{100}{4} - 16}{19} \right) \cdot 10 = 65 + \left(\frac{9}{19} \right) \cdot 10 = 65 + 4,737$$

$$Q_2 = 69,737$$

1.3) Cálculo del tercer cuartil

Primero se calcula $nk/4$ y después se averigua el intervalo en el que está el cuartil, este intervalo recibe el nombre de intervalo o clase del cuartil.

$$\frac{n \cdot 3}{4} = \frac{50 \cdot 3}{4} = 37,5$$

Por lo tanto para el segundo cuartil se tiene:

Intervalo: 75-85

$n=10$

$Fa=35$

$f_Q=11$

$c=85-75=10$

Al aplicar la ecuación se obtiene:

$$Q_k = Li_Q + \left(\frac{\frac{nk}{4} - Fa}{f_Q} \right) \cdot c$$

$$Q_3 = 75 + \left(\frac{\frac{50 \cdot 3}{4} - 35}{11} \right) \cdot 10 = 75 + \left(\frac{\frac{150}{4} - 35}{11} \right) \cdot 10 = 75 + \left(\frac{5}{22} \right) \cdot 10 = 75 + 2,273$$

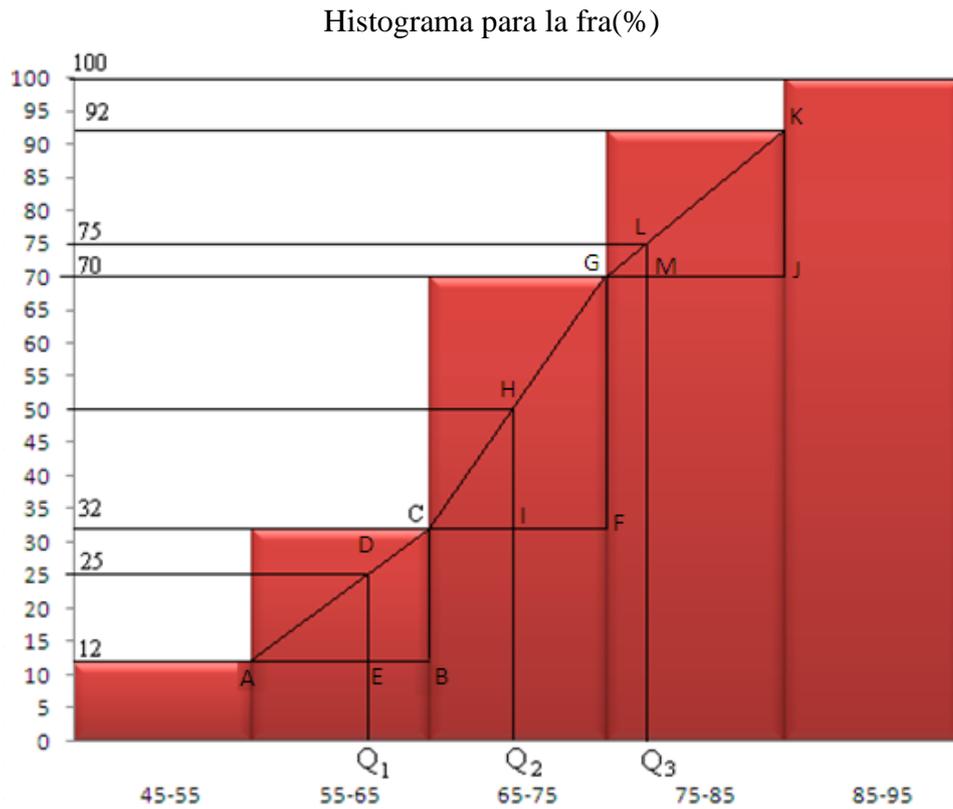
$$Q_3 = 77,273$$

2) Cálculo de los cuartiles empleando un histograma para fra(%)

2.1) Calculando la fra(%) se obtiene:

Intervalos	f	fa	fr	fra (%)
45- 55	6	6	0,12	12
55- 65	10	16	0,20	32
65- 75	19	35	0,38	70
75- 85	11	46	0,22	92
85- 95	4	50	0,08	100
n	50			

2.2) Elaborando el histograma en Excel y en Paint se obtiene la siguiente figura:



2.3) Cálculo del primer cuartil

Observando en gráfico tenemos que el $Q_1 = 55 + AE$

Los triángulos ABC y AED son semejantes, por lo que se cumple:

$$\frac{AB}{CB} = \frac{AE}{DE}$$

$$\frac{65 - 55}{32 - 12} = \frac{AE}{25 - 12} \Rightarrow \frac{10}{20} = \frac{AE}{13}$$

Despejando AE se obtiene:

$$\frac{10}{20} \cdot 13 = AE \Rightarrow AE = 6,5$$

Entonces, $Q_1 = 55 + 6,5 = 61,5$

2.3) Cálculo del segundo cuartil

Observando en gráfico tenemos que el $Q_2 = 65 + CI$

Los triángulos CFG y CIH son semejantes, por lo que se cumple:

$$\frac{CF}{FG} = \frac{CI}{HI}$$

$$\frac{75 - 65}{70 - 32} = \frac{CI}{50 - 32} \Rightarrow \frac{10}{38} = \frac{CI}{18}$$

Despejando CI se obtiene:

$$\frac{10}{38} \cdot 18 = AE \Rightarrow AE = 4,737$$

$$\text{Entonces, } Q_2 = 65 + 4,737 = 69,737$$

2.3) Cálculo del tercer cuartil

Observando en gráfico tenemos que el $Q_3 = 75 + GM$

Los triángulos GJK y GML son semejantes, por lo que se cumple:

$$\frac{GJ}{JK} = \frac{GM}{ML}$$

$$\frac{85 - 75}{92 - 70} = \frac{CI}{75 - 70} \Rightarrow \frac{10}{22} = \frac{CI}{5}$$

Despejando CI se obtiene:

$$\frac{10}{22} \cdot 5 = CI \Rightarrow CI = 2,273$$

$$\text{Entonces, } Q_3 = 75 + 2,273 = 77,273$$

2.5.2) DECILES

2.5.2.1) Definición

Son cada uno de los 9 valores $D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9$ que dividen a la distribución de los datos en 10 partes iguales.

El primer decil es igual al décimo percentil ($D_1 = P_1$), el segundo decil es igual al veinteavo percentil ($D_2 = P_{20}$), y así sucesivamente.

2.5.2.2) Métodos de Cálculo

a) Para Datos No Agrupados

La posición o ubicación de los deciles se encuentra aplicando la siguiente ecuación:

$$D_k = X_{\left[\frac{n \cdot k + 1}{10}\right]} = X_{\left[\frac{n \cdot k + 5}{10}\right]}$$

Donde:

n = número total de datos.

k = número del decil.

Ejemplo ilustrativo:

Calcular el quinto decil de la siguiente distribución: 6, 9, 9, 12, 12, 12, 15 y 17

Solución:

Para calcular los deciles se ordena los datos de menor a mayor.

6	9	9	12	12	12	15	17
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8

Aplicando la ecuación para el quinto decil se obtiene:

$$D_k = X_{\left[\frac{n \cdot k + 5}{10}\right]}$$

$$D_5 = X_{\left[\frac{n \cdot 5 + 5}{10}\right]} = X_{\left[\frac{5n + 5}{10}\right]} = X_{\left[\frac{5 \cdot 8 + 5}{10}\right]} = X_{\left[\frac{40 + 5}{10}\right]} = X_{4,5} = \frac{x_4 + x_5}{2} = \frac{12 + 12}{2} = 12$$

O también la posición 4,5 dice que el decil 5 está ubicado al 50% del trayecto comprendido entre el cuarto dato, que es 12 y el quinto dato que también es 12, es decir,
 $D_5 = 12 + 0,5(12 - 12) = 12$

En Excel se calcula de la siguiente manera:

Como D_5 es igual a P_{50} se introduce la función PERCENTIL.INC(A1:A8;0,5) como se muestra en la siguiente figura:

PERCENTIL.INC				
	A	B	C	D
1	6			
2	9			
3	9			
4	12			
5	12			
6	12			
7	15			
8	17			
9				
10	D5=P50	12	=PERCENTIL.INC(A1:A8;0,5)	

b) Para Datos Agrupados en Tablas de Frecuencia

Se emplea la misma ecuación utilizada en el cálculo de los deciles para datos sin agrupar.

c) Para Datos Agrupados en Intervalos

Se emplea la siguiente ecuación:

$$D_k = Li_D + \left(\frac{\frac{nk}{10} - Fa}{f_D} \right) \cdot c$$

Donde:

Li_D = Límite inferior del intervalo de clase del decil.

n = número total de datos.

Fa = Frecuencia acumulada del intervalo de clase que antecede al intervalo de clase del decil.

f_D = Frecuencia absoluta del intervalo de clase del decil.

c = Ancho del intervalo de clase del decil.

2.5.3) PERCENTILES O CENTILES

2.5.3.1) Definición

Son cada uno de los 99 valores $P_1, P_2, P_3, \dots, P_{99}$ que dividen atribución de los datos en 100 partes iguales.

2.5.3.2) Métodos de Cálculo

a) Para Datos No Agrupados

La posición o ubicación de los percentiles se encuentra aplicando la siguiente ecuación:

$$P_k = X_{\left[\frac{n \cdot k}{100} + \frac{1}{2} \right]} = X_{\left[\frac{n \cdot k + 50}{100} \right]}$$

Donde:

n = número total de datos

k = número del percentil

Ejemplo ilustrativo:

Calcular los percentiles de orden 20 y 33 del peso de diez personas que pesan (en kg)

80, 78, 65, 73, 65, 67, 72, 68, 70 y 72

Solución:

Se ordena los datos de menor a mayor se tiene:

65	65	67	68	70	72	72	73	78	80
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}

1) Cálculo del percentil de orden 20 se obtiene:

$$P_k = X_{\left[\frac{n \cdot k + 50}{100}\right]}$$

$$P_{20} = X_{\left[\frac{n \cdot 20 + 50}{100}\right]} = X_{\left[\frac{10 \cdot 20 + 50}{100}\right]} = X_{\left[\frac{250}{100}\right]} = X_{2,5} = \frac{x_2 + x_3}{2} = \frac{65 + 67}{2} = 66$$

En Excel se obtiene un valor aproximado insertando la función PERCENTIL.INC(A1:A10;0,2) como se muestra en la siguiente figura:

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E
1	65				
2	65				
3	67				
4	68				
5	70				
6	72				
7	72				
8	73				
9	78				
10	80				
11					
12	P20	66,6	=PERCENTIL.INC(A1:A10;0,2)		

2) Cálculo del percentil de orden 33 se obtiene:

$$P_k = X_{\left[\frac{n \cdot k + 50}{100}\right]}$$

$$P_{33} = X_{\left[\frac{n \cdot 33 + 50}{100}\right]} = X_{\left[\frac{10 \cdot 33 + 50}{100}\right]} = X_{\left[\frac{380}{100}\right]} = X_{3,8} = \frac{x_3 + x_4}{2} = \frac{67 + 68}{2} = 67,5 = 68$$

En Excel se obtiene un valor aproximado insertando la función PERCENTIL.INC(A1:A10;0,33) como se muestra en la siguiente figura:

PERCENTIL.INC					
	A	B	C	D	E
1	65				
2	65				
3	67				
4	68				
5	70				
6	72				
7	72				
8	73				
9	78				
10	80				
11					
12	P20	67,97	=PERCENTIL.INC(A1:A10;0,33)		

b) Para Datos Agrupados en Tablas de Frecuencia

Se emplea la misma ecuación utilizada en el cálculo de los percentiles para datos sin agrupar.

c) Para Datos Agrupados en Intervalos

Se emplea la ecuación:

$$P_k = Li_p + \left(\frac{\frac{nk}{100} - Fa}{f_p} \right) \cdot c$$

Donde:

Li_p = Límite inferior del intervalo de clase del percentil.

n = número total de datos.

Fa = Frecuencia acumulada del intervalo de clase que antecede al intervalo de clase del percentil.

f_p = Frecuencia absoluta del intervalo de clase del percentil.

c = Ancho del intervalo de clase del percentil.

TAREA DE INTERAPRENDIZAJE N° 9

- 1) ¿El valor de la mediana con qué valor del cuartil, decil y del percentil coincide?. Plantee y resuelva un ejercicio para ilustrar su respuesta.
- 2) ¿Por qué a los cuartiles, deciles y percentiles se les considera como medidas de posición?
- 3) Realice un organizador gráfico sobre las medidas de posición.
- 4) Calcule los 3 cuartiles de las siguientes distribuciones de datos de manera manual y empleando Excel. Realice un diagrama de caja y bigotes de manera manual y empleando Graph.

4.1) 5, 2, 6, 4, 1 y 3

$$Q_1 = 2; Q_2 = 3; Q_3 = 5$$

4.2) 5, 2, 8, 4, 1, 6, 7 y 3

$$Q_1 = 2,5; Q_2 = 4,5; Q_3 = 6,5$$

4.3) 9, 2, 8, 4, 5, 6, 7, 3 y 1

$$Q_1 = 3; Q_2 = 5; Q_3 = 7$$

4.4) 36, 8, 12, 32, 24, 28, 16 y 4

$$Q_1 = 10; Q_2 = 20; Q_3 = 30$$

4.5) 80, 70, 40, 60, 50, 30, 20 y 10

$$Q_1 = 25; Q_2 = 45; Q_3 = 65$$

5) Dada la siguiente tabla:

x	6	9	12	15	17
f	1	2	3	1	1

5.1) Calcule el primero y tercer cuartil.

$$Q_1=9; Q_3=13,5$$

5.2) Calcule el segundo cuartil empleando un histograma para la frecuencia absoluta acumulada.

$$Q_2=12$$

6) Cree y resuelva un ejercicio similar al presentado en el cálculo de los cuartiles para datos agrupados en intervalos.

7) Emplee los datos del ejercicio anterior y calcular los cuartiles empleando un histograma para la frecuencia absoluta acumulada.

8) Calcule el quinto decil de 1, 3, 6, 9, 12, 15, 18 y 21 de manera manual y empleando Excel.

$$D_5=10,5$$

9) Cree y resuelva un ejercicio sobre el cálculo del decil 3 y del decil 7 para datos agrupados en tablas de frecuencias.

10) Cree y resuelva un ejercicio sobre el cálculo de los deciles de orden 4 y 8 para datos agrupados en intervalos empleando las ecuaciones y a través de un histograma para la fra(%).

11) Calcule el percentil de orden 25 de 2, 4, 6, 8, 10, 12, 14, 16, 18, 20 y 22 de manera manual y empleando Excel.

$$P_{25}=6$$

12) Calcule el percentil de orden 75 de 10, 20, 40, 50, 60, 70, 80, 90, 100, 120 y 140.

$$P_{75}=95$$

13) Plantee y resuelva un ejercicio sobre el cálculo de los percentiles 35 y 60 para datos agrupados en tablas de frecuencias.

14) Plantee y resuelva un ejercicio sobre el cálculo de percentiles 45 y 85 para datos agrupados en intervalos empleando las ecuaciones y a través de un histograma para la fra(%).

2.6) MODA

La moda de un conjunto de datos es el valor que aparece con mayor frecuencia.

2.6.1) PROPIEDADES

- No es afectada por valores muy altos o muy bajos.
- La moda, al igual que la mediana, no se presta para tratamientos algebraicos como la media aritmética.
- La moda puede no existir, e incluso no ser única en caso de existir.
- Cuando en un conjunto de datos hay tres o más datos diferentes con la misma frecuencia mayor, esta información a menudo no resulta útil (demasiadas modas tienden a distorsionar el significado de moda). Por lo que en estos casos se considera que el conjunto de datos no tiene moda.

Para un conjunto de datos unimodales existe la siguiente relación empírica:

$$\text{Media aritmética} - \text{moda} = 3 (\text{media aritmética} - \text{mediana})$$

2.6.2) MÉTODOS DE CÁLCULO

2.6.2.1) Para Datos No Agrupados

Se observa el dato que tiene mayor frecuencia

Ejemplo ilustrativo N° 1

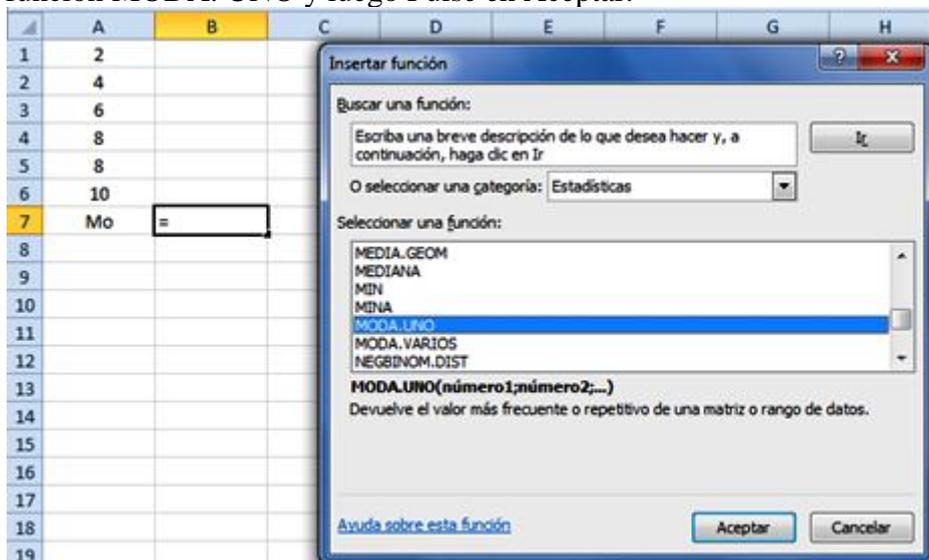
Determinar la moda del conjunto de datos 2, 4, 6, 8, 8 y 10

Solución:

$Mo = 8$, porque es el dato que ocurre con mayor frecuencia. A este conjunto de datos se le llama unimodal

En Excel se calcula de la siguiente manera:

a) Se inserta la función MODA. UNO y luego Pulse en Aceptar.



b) Seleccionar las celdas (Rango A1:A6).

Argumentos de función

MODA.UNO

Número1: A1:A6 = {2;4;6;8;8;10}

Número2: = matriz

= 8

Devuelve el valor más frecuente o repetitivo de una matriz o rango de datos.

Número1: número1;número2;... son de 1 a 255 números, nombres, matrices o referencias que contienen números para los que desea la moda.

Resultado de la fórmula = 8

[Ayuda sobre esta función](#)

Aceptar Cancelar

c) Pulsar en Aceptar.

	A	B	C	D
1	2			
2	4			
3	6			
4	8			
5	8			
6	10			
7	Mo	8	=MODA.UNO(A1:A6)	

Ejemplo ilustrativo N° 2

Determinar la moda del conjunto de datos: 2, 4, 6, 8 y 10

Solución:

Este conjunto de datos no tiene moda, porque todos los datos tienen la misma frecuencia.

En Excel se calcula de la siguiente manera:

Se inserta la función MODA.UNO, se selecciona las celdas respectivas y se pulsa en Aceptar.

	A	B	C	D
1	2			
2	4			
3	6			
4	8			
5	10			
6				
7	Mo	#N/A	=MODA.UNO(A1:A5)	

Ejemplo ilustrativo N° 3

Determinar la moda del conjunto de datos: 8, 4, 6, 6, 8, 2 y 10

Solución:

Este conjunto de datos tiene dos modas, 8 y 6, y se llama bimodal.

En Excel se calcula de la siguiente manera:

Se inserta la función MODA.VARIOS, la cual debe especificarse como fórmula de matriz, para lo cual se selecciona las celdas donde aparecerá la respuesta (B9:B10). Luego se inserta la función MODA.VARIOS, se selecciona las celdas respectivas (A1:A7)

MODA.VARIOS

Argumentos de función

MODA.VARIOS

Número1 A1:A7 = {8;4;6;6;8;2;10}

Número2 = matriz

= {8;6}

Devuelve una matriz vertical de los valores más frecuente o repetitivos de una matriz o rango de datos. Para una matriz horizontal, use =TRANSPONER(MODA.VARIOS(número1,número2,...)).

Número1: número1;número2;... son de 1 a 255 números, nombres, matrices o referencias que contienen números para los que desea la moda.

Resultado de la fórmula = 8

[Ayuda sobre esta función](#) Aceptar Cancelar

Finalmente, se presiona Ctrl+Mayús+Enter.

	MODA.VARIOS		
	A	B	C
1	8		
2	4		
3	6		
4	6		
5	8		
6	2		
7	10		
8			
9	Mo	8	
10		6	

2.6.2.2) Para Datos Agrupados en Tablas de Frecuencia

Se observa el dato tiene mayor frecuencia

Ejemplo ilustrativo: Calcule la moda o modas (si las hay) de los siguientes datos:

x	f
2	1
4	2
6	3
8	1
10	1

Solución:

Se observa que el dato con mayor frecuencia es 6, por lo tanto $M_o = 6$

2.6.2.3) Para Datos Agrupados en Intervalos

Se halla en el intervalo o clase que tenga la frecuencia más alta, llamada intervalo o clase modal. Se emplea la siguiente ecuación:

$$M_o = L_{M_o} + \left(\frac{D_a}{D_a + D_b} \right) \cdot c$$

L_{M_o} = Límite inferior de la clase modal.

D_a = Diferencia entre la frecuencia absoluta de la clase modal y la clase que la antecede.

D_b = Diferencia entre la frecuencia absoluta de la clase modal y la clase que le sigue.

c = ancho de la clase modal.

Ejemplo ilustrativo: Calcule la moda o modas (si las hay) de los siguientes datos:

Intervalo o Clase	f
10-19	3
20-29	7
30-39	15
40-49	12
50-59	8

Solución:

Se observa que la clase modal es 30-39, ya que es el intervalo con la mayor frecuencia.

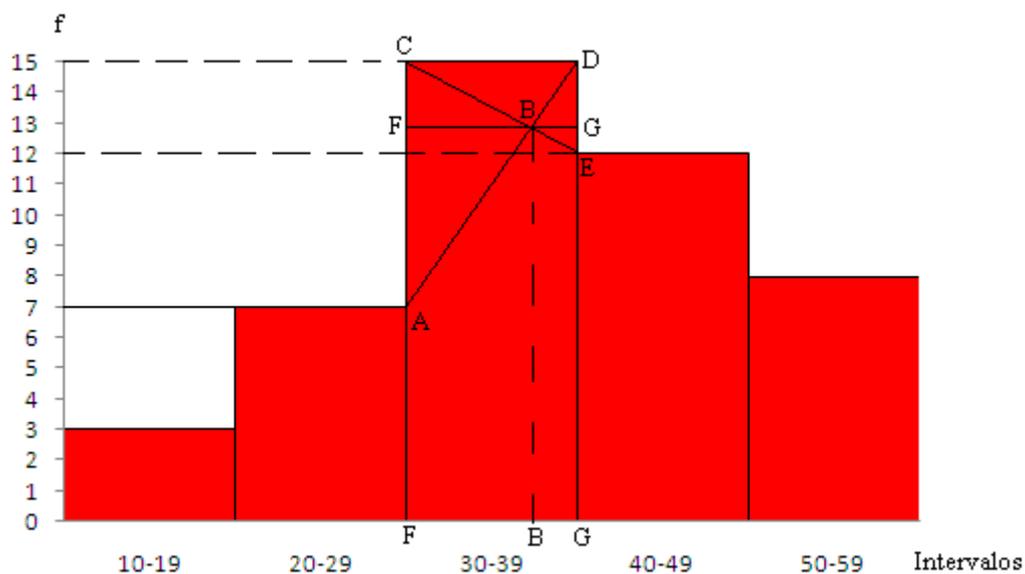
Aplicando la ecuación

$$M_o = L_{M_o} + \left(\frac{D_a}{D_a + D_b} \right) \cdot c$$

Se tiene:

$$M_o = 30 + \left(\frac{15 - 7}{(15 - 7) + (15 - 12)} \right) \cdot 10 = 30 + \left(\frac{8}{8 + 3} \right) \cdot 10 = 30 + \frac{80}{11} = 37,27$$

Gráficamente empleando un histograma se calcula la moda de la siguiente manera:



La clase modal es 30-39, ya que es el intervalo con la mayor frecuencia

Observando el histograma se tiene que $Mo = 30 + FB$

Los triángulos ABC y EBD son semejantes, por lo que se cumple:

$$\frac{FB}{AC} = \frac{BG}{DE}$$

Donde:

AC = Diferencia entre la frecuencia absoluta de la clase modal y la clase que la antecede.

BG es igual al ancho del intervalo 30-39 menos FB.

DE = Diferencia entre la frecuencia absoluta de la clase modal y la clase que le sigue.

Reemplazando valores y despejando FB se tiene:

$$\frac{FB}{15 - 7} = \frac{10 - FB}{15 - 12} \Rightarrow \frac{FB}{8} = \frac{10 - FB}{3} \Rightarrow 3FB = 8(10 - FB) \Rightarrow 3FB = 80 - 8FB$$

$$3FB + 8FB = 80 \Rightarrow 11FB = 80 \Rightarrow FB = \frac{80}{11} = 7,27$$

Por lo tanto $Mo = 30 + FB = 30 + 7,27 = 37,27$

TAREA DE INTERAPRENDIZAJE N° 10

1) Realice un organizador gráfico sobre la moda.

2) Para una tienda de modas o para un diseñador de autos, ¿de qué le serviría saber el valor de la moda?.

3) Se está estudiando el ingreso diario de un grupo de personas y se tiene los siguientes valores en dólares: 350, 400, 500, 350, 550, 1500 y 2000.

3.1) Calcule manualmente y empleando Excel la media aritmética, la mediana y la moda.

$$\bar{x} = \$ 807,14; Md = \$ 500; Mo = \$ 350$$

3.2) ¿Qué valor es más representativo del ingreso promedio?. Argumente su respuesta.

4) Plantee un resuelva un ejercicio con datos sin agrupar y compruebe la relación empírica entre la media aritmética, mediana y moda.

5) Averigüe a 30 compañeros de su clase sobre el número de hermanas y hermanos.

5.1) Elabore una tabla de frecuencias.

5.2) Calcule la media aritmética, mediana y moda.

6) Dados los siguientes datos: 50, 52, 59, 60, 60, 63, 64, 65, 69, 69, 70, 70, 72, 72, 74, 74, 75, 75, 76, 75, 74, 70, 77, 78, 78, 79, 79, 75, 80, 80, 81, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 98, 99, 100 y 109

6.1) Calcule manualmente y empleando Excel la media aritmética, la mediana y la moda con los datos sin agrupar.

$$\bar{x} = 78,42; Md=78; Mo=75$$

6.2) Agrupe los datos en intervalos de ancho 10. Complete la siguiente tabla:

Intervalo	<i>f</i>
50-59	
	7
70-79	
	12
90-99	
	2

6.3) A partir de la tabla anterior calcule media aritmética, la mediana y la moda.

$$\bar{x} = 78,7; Md = 78,33; Mo = 76,47$$

6.4) Calcule la media aritmética empleando Excel.

$$\bar{x} = 78,7$$

6.5) Calcule la moda empleando un histograma.

$$Mo = 76,47$$

6.6) Calcule la mediana a través de un histograma para la fra(%)

$$Md = 78,33$$

6.7) ¿Por qué varían los resultados de los datos sin agrupar con los datos agrupados en intervalos?

7) Plantee y resuelva un ejercicio similar al anterior.

CAPÍTULO III

MEDIDAS DE DISPERSIÓN

COMPETENCIAS:

- ✓ Interpreta adecuadamente las características y propiedades de las medidas de dispersión y comprende sus aplicaciones.
- ✓ Emplea correctamente algoritmos matemáticos para calcular medidas de dispersión de manera manual y empleando Excel.
- ✓ Crea y resuelve correctamente ejercicios de aplicación sobre las medidas de dispersión.

CONTENIDOS:

- ✓ Desviación Media o Desviación Promedio
- ✓ Varianza y Desviación Estándar
- ✓ Otras Medidas de Dispersión
- ✓ Dispersión Relativa o Coeficiente de Variación

MEDIDAS DE DISPERSIÓN

Las medias de tendencia central o posición nos indican donde se sitúa un dato dentro de una distribución de datos. Las medidas de dispersión, variabilidad o variación nos indican si esos datos están próximos entre sí o si están dispersos, es decir, nos indican cuán esparcidos se encuentran los datos. Estas medidas de dispersión nos permiten apreciar la distancia que existe entre los datos a un cierto valor central e identificar la concentración de los mismos en un cierto sector de la distribución, es decir, permiten estimar cuán dispersas están dos o más distribuciones de datos.

Estas medidas permiten evaluar la confiabilidad del valor del dato central de un conjunto de datos, siendo la media aritmética el dato central más utilizado. Cuando existe una dispersión pequeña se dice que los datos están dispersos o acumulados cercanamente respecto a un valor central, en este caso el dato central es un valor muy representativo. En el caso que la dispersión sea grande el valor central no es muy confiable. Cuando una distribución de datos tiene poca dispersión toma el nombre de distribución homogénea y si su dispersión es alta se llama heterogénea.

3.1) DESVIACIÓN MEDIA O DESVIACIÓN PROMEDIO

La desviación media o desviación promedio es la media aritmética de los valores absolutos de las desviaciones respecto a la media aritmética.

3.1.1) PROPIEDADES

Guarda las mismas dimensiones que las observaciones. La suma de valores absolutos es relativamente sencilla de calcular, pero esta simplicidad tiene un inconveniente: Desde el punto de vista geométrico, la distancia que induce la desviación media en el espacio de observaciones no es la *natural* (no permite definir ángulos entre dos conjuntos de observaciones). Esto hace que sea muy engorroso trabajar con ella a la hora de hacer inferencia a la población.

Cuando mayor sea el valor de la desviación media, mayor es la dispersión de los datos. Sin embargo, no proporciona una relación matemática precisa entre su magnitud y la posición de un dato dentro de una distribución.

La desviación media al tomar los valores absolutos mide una observación sin mostrar si la misma está por encima o por debajo de la media aritmética.

3.1.2) MÉTODOS DE CÁLCULO

3.1.2.1) Para Datos No Agrupados

Se emplea la fórmula:

$$DM = \frac{\sum |x - \bar{x}|}{n}$$

Ejemplo ilustrativo:

Calcular la desviación media de la distribución: 3, 8, 8, 8, 9, 9, 9, 18

Solución:

Se calcula la media aritmética.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{3 + 8 + 8 + 8 + 9 + 9 + 9 + 18}{8} = 9$$

Se calcula la desviación media.

$$DM = \frac{\sum |x - \bar{x}|}{n}$$

$$DM = \frac{|3 - 9| + |8 - 9| + |8 - 9| + |8 - 9| + |9 - 9| + |9 - 9| + |9 - 9| + |18 - 9|}{8}$$

$$DM = \frac{6 + 1 + 1 + 1 + 0 + 0 + 0 + 9}{8} = \frac{18}{8} = 2,25$$

Empleando Excel se calcula de la siguiente manera:

a) Se inserta la función DESVPROM como se muestra en la siguiente figura:

The image shows a screenshot of an Excel spreadsheet and the 'Insertar función' (Insert Function) dialog box. The spreadsheet has a column with the following values: 3, 8, 8, 8, 9, 9, 9, 18. The cell in row 9, column A, is labeled 'DM' and contains an equals sign. The 'Insertar función' dialog box is open, showing the 'Estadísticas' category selected. The list of functions includes DESVESTA, DESVESTP, DESVESTPA, DESVIA2, DESVPROM (highlighted), DIST.WEIBULL, and DISTR.BETA. The dialog box also displays the function's syntax: DESVPROM(número1;número2;...) and its description: Devuelve el promedio de las desviaciones absolutas de la media de los puntos de datos. Los argumentos pueden ser números, nombres, matrices o referencias que contienen números. There are buttons for 'Aceptar' (Accept) and 'Cancelar' (Cancel) at the bottom of the dialog box.

b) Pulsar en Aceptar y seleccionar las celdas.

The screenshot shows an Excel spreadsheet with the following data in column B:

	A	B
1		3
2		8
3		8
4		8
5		9
6		9
7		9
8		18
9	DM	=DESVPROM(A1:A8)

The dialog box 'Argumentos de función' for DESVPROM is open, showing:

- Número1: A1:A8 = {3\8\8\8\9\9\9\18}
- Número2: = número
- Resultado de la fórmula = 2,25

c) Pulsar en Aceptar.

	A	B	C	D
1		3		
2		8		
3		8		
4		8		
5		9		
6		9		
7		9		
8		18		
9	DM	2,25	=DESVPROM(A1:A8)	

3.1.2.2) Para Datos Agrupados en Tablas de Frecuencia

Se emplea la fórmula:

$$DM = \frac{\sum f|x - \bar{x}|}{n}$$

Ejemplo ilustrativo: Calcular la desviación media en base a la siguiente tabla sobre las calificaciones de un estudiante en 12 asignaturas evaluadas sobre 10.

Calificación	Cantidad de asignaturas
6	4
7	2
8	3
9	2
10	1
Total	12

Solución:

Se calcula la media aritmética.

$$\bar{x} = \frac{\sum fx}{n} = \frac{4 \cdot 6 + 2 \cdot 7 + 3 \cdot 8 + 2 \cdot 9 + 1 \cdot 10}{12} = \frac{24 + 14 + 24 + 18 + 10}{12} = \frac{90}{12} = 7,5$$

Se llena la siguiente tabla:

x	f	$ x - \bar{x} $	$f x - \bar{x} $
6	4	1,5	6
7	2	0,5	1
8	3	0,5	1,5
9	2	1,5	3
10	1	2,5	2,5
Total	12		14

Se emplea la ecuación de la desviación media.

$$DM = \frac{\sum f|x - \bar{x}|}{n} = \frac{14}{12} = 1,167$$

3.1.2.3) Para Datos Agrupados en Intervalos

Se emplea la fórmula:

$$DM = \frac{\sum f|xm - \bar{x}|}{n}$$

Donde xm es la marca de clase.

Ejemplo ilustrativo:

Calcular la desviación media de un curso de 40 estudiantes en la asignatura de Estadística en base a la siguiente tabla:

Calificación	Cantidad de estudiantes
2-4	6
4-6	8
6-8	16
8-10	10
Total	40

Solución:

Para calcular la media aritmética se llena la siguiente tabla:

Intervalo	f	xm	$f \cdot xm$
2-4	6	3	18
4-6	8	5	40
6-8	16	7	112
8-10	10	9	90
Total	40		260

Calculando la media aritmética se obtiene:

$$\bar{x} = \frac{\sum f \cdot x m}{n} = \frac{260}{40} = 6,5$$

Para calcular la desviación media se llena la siguiente tabla:

Intervalo	f	xm	$ xm - \bar{x} $	$f xm - \bar{x} $
2-4	6	3	3,5	21
4-6	8	5	2,5	12
6-8	16	7	0,5	8
8-10	10	9	2,5	25
Total	40			66

$$DM = \frac{66}{40} = 1,65$$

TAREA DE INTERAPRENDIZAJE N° 11

1) Conteste a las siguientes preguntas.

1.1) ¿Cuál es la diferencia entre medidas de tendencia central y medidas de dispersión?

1.2) ¿Qué permiten o que se logra con las medidas de dispersión?

1.3) ¿En qué caso una distribución de datos toma el nombre de homogénea?. Explique con un ejemplo.

1.4) ¿En qué caso una distribución de datos toma el nombre de heterogénea?. Explique con un ejemplo.

2) Realice un organizador gráfico sobre la desviación media.

3) Calcule la desviación media de las siguientes distribuciones empleando la ecuación y mediante Excel.

3.1) 6, 8, 7, 2, 4, 5, 8 y 9

MD= 1,875

3.2) 10, 4, 8, 9, 6, 10, 8 y 10

MD= 1,625

4) Crear y resolver un ejercicio similar al anterior.

5) Calcule la desviación media empleando los datos de la siguiente tabla:

x	f
20	2
19	2
18	3
17	4
16	6
15	3
14	3
Total	23

DM = 1,46

6) Cree y resuelva un ejercicio similar al ejemplo presentado en el cálculo de la desviación media para datos agrupados en tablas de frecuencias.

7) La siguiente tabla muestra la cantidad de dinero que se gasta por semana un grupo de 50 personas. Calcular la desviación media.

Intervalo	f
10-20	8
20-30	15
30-40	12
40-50	15
Total	50

DM= 9,456

8) Cree y resuelva un ejercicio con datos agrupados en intervalos y calcular

8.1) La media geométrica.

8.2) La media armónica.

8.3) La mediana.

8.4) El cuartil N°1 y el cuartil N°3 empleando las ecuaciones y mediante un histograma para la fra(%).

8.5) La Moda empleando la ecuación y mediante un histograma.

8.6) La desviación media.

3.2) VARIANZA Y DESVIACIÓN ESTÁNDAR

La varianza es la media aritmética de los cuadrados de las desviaciones respecto a la media aritmética, es decir, es el promedio de las desviaciones de la media elevadas al cuadrado. La desviación estándar o desviación típica es la raíz de la varianza.

La varianza y la desviación estándar proporcionan una medida sobre el punto hasta el cual se dispersan las observaciones alrededor de su media aritmética.

3.2.1) PROPIEDADES

- La varianza y desviación estándar (o cualquier otra medida de dispersión) indican el grado en que están dispersos los datos en una distribución. A mayor medida, mayor dispersión.
- La varianza es un número muy grande con respecto a las observaciones, por lo que con frecuencia se vuelve difícil para trabajar.
- Debido a que las desviaciones son elevadas al cuadrado y la varianza siempre se expresa en términos de los datos originales elevados al cuadrado, se obtiene unidades de medida de los datos que no tiene sentido o interpretación lógica. Por ejemplo, si se calcula la varianza de una distribución de datos medidos en metros, segundos, dólares, etc., se obtendrá una varianza mediada en metros cuadrados, segundos cuadrados, dólares cuadrados, respectivamente, unidades de medida que no tienen significado lógico respecto a los datos originales.
- Para solucionar las complicaciones que se tiene con la varianza, se halla la raíz cuadrada de la misma, es decir, se calcula la desviación estándar, la cual es un número pequeño expresado en unidades de los datos originales y que tiene un significado lógico respecto a los mismos.

A pesar de lo anterior, es difícil describir exactamente qué es lo que mide la desviación estándar. Sin embargo, hay un resultado útil, que lleva el nombre del matemático ruso Pafnuty Lvovich Chebyshev, y se aplica a todos los conjuntos de datos. Este teorema de Chebyshev establece que para todo conjunto de datos, por lo menos $1 - 1/k^2$ de las observaciones están dentro de k desviaciones estándar de la media, en donde k es cualquier número mayor que 1. Este teorema se expresa de la siguiente manera:

$$1 - \frac{1}{k^2}$$

Así por ejemplo, si se forma una distribución de datos con $k = 3$ desviaciones estándar por debajo de la media hasta 3 desviaciones estándar por encima de la media, entonces por lo menos

$$1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{9-1}{9} = \frac{8}{9} = 0,8889 = 88,89\%$$

Interpretación: El 88,89% de todas las observaciones estarán dentro ± 3 desviaciones de la media.

3.2.2) MÉTODOS DE CÁLCULO

3.2.2.1) Para Datos No Agrupados

La varianza para una población se calcula con:

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$$

Donde:

x_i = *observaciones individuales de la población*

μ = *media aritmética poblacional*

N = *número de observaciones de la población*

La desviación estándar poblacional se calcula con:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

La varianza de la muestra se calcula con:

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

Donde:

x_i = *observaciones individuales de la muestra*

\bar{x} = *media aritmética de la muestra*

n = *número de observaciones de la muestra*

La desviación estándar de una muestra se calculó con:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

Notas:

1) Para el cálculo de la varianza de una muestra se divide por $n-1$ en lugar de N , debido a que se tiene $n-1$ grados de libertad en la muestra. Otra razón por la que se divide por $n-1$ es debido a que una muestra generalmente está un poco menos dispersa que la población de la cual se tomó. Al dividir para $n-1$ en lugar de N se cumple con la tendencia y sentido lógico de que la varianza y desviación estándar de la muestra deben tener un valor más pequeño que la varianza y desviación estándar de la población.

2) En la realidad, salvo indicación expresa, no se calcula la varianza y la desviación estándar de la población, ya que para ahorrar tiempo, esfuerzo, dinero, etc. es mejor trabajar con datos que representan a la muestra.

Ejemplo ilustrativo N° 1

Considere que los siguientes datos corresponden al sueldo de una población: \$350, \$400, \$500, \$700 y \$1000

1) Calcular la desviación estándar.

2) ¿Cuál es el intervalo que está dentro de $k = 2$ desviaciones estándar de la media?. ¿Qué porcentaje de las observaciones se encuentran dentro de ese intervalo?

Solución:

1) Para la calcular la desviación estándar se sigue los siguientes pasos:

a) Se calcula la media aritmética.

$$\mu = \frac{\sum x_i}{N} = \frac{350 + 400 + 500 + 700 + 1000}{5} = \frac{2950}{5} = \$ 590$$

$$\mu = \frac{2950}{5}$$

$$\mu = \$ 590$$

b) Se aplica la respectiva fórmula para calcular la varianza

$$\begin{aligned}\sigma^2 &= \frac{\sum(x_i - \mu)^2}{N} \\ \sigma^2 &= \frac{(350 - 590)^2 + (400 - 590)^2 + (500 - 590)^2 + (700 - 590)^2 + (1000 - 590)^2}{5} \\ \sigma^2 &= \frac{57600 + 36100 + 8100 + 12100 + 168100}{5}\end{aligned}$$

$$\sigma^2 = \frac{282000}{5}$$

$$\sigma^2 = \$^2 56400$$

c) Se calcula la desviación estándar.

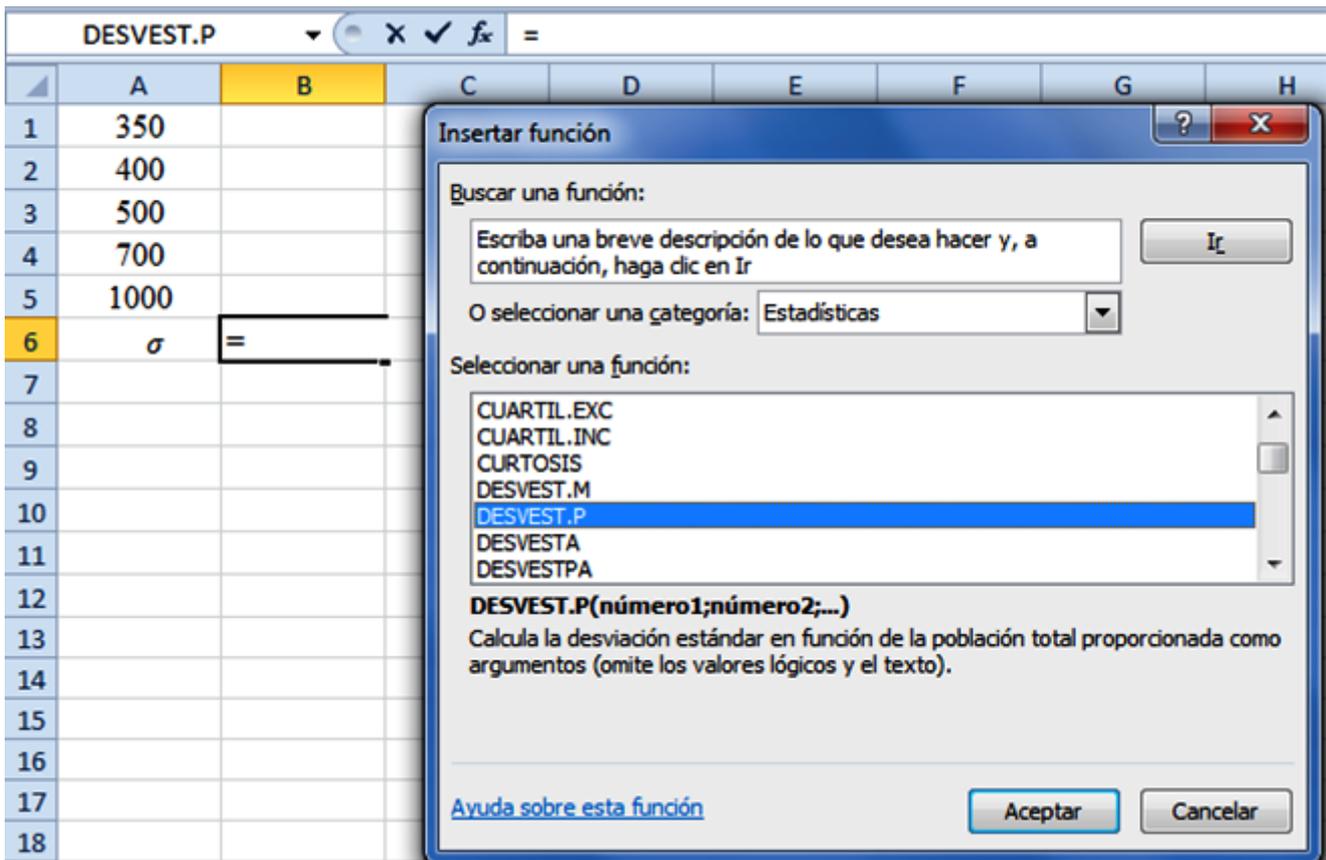
$$\sigma = \sqrt{\sigma^2}$$

$$\sigma = \sqrt{\$^2 56400}$$

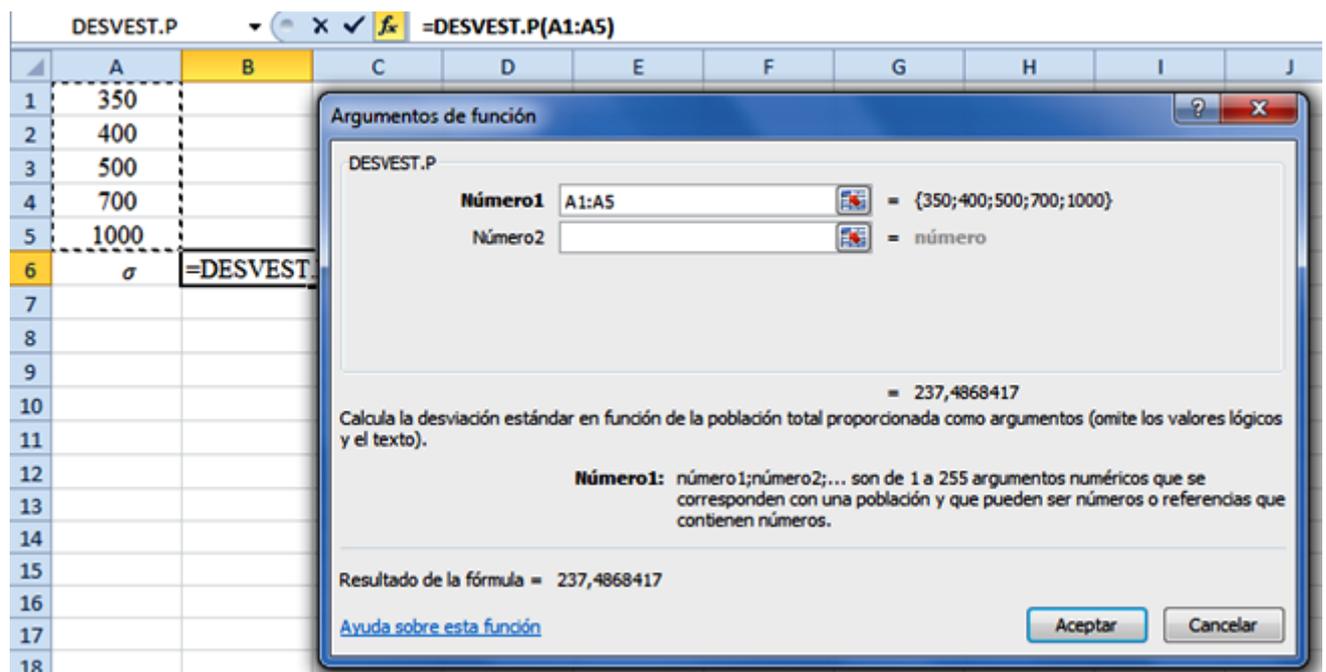
$$\sigma = \$237,4868$$

Empleando Excel se calcula de la siguiente manera:

a) Se inserta la función DESVEST.P. Clic en Aceptar



b) Pulsar en Aceptar y seleccionar las celdas.



c) Pulsar en Aceptar.

DESVEST.P				
	A	B	C	D
1	350			
2	400			
3	500			
4	700			
5	1000			
6	σ	237,48684	=DESVEST.P(A1:A5)	

2) Cálculo del intervalo de $k = 2$ desviaciones estándar de la media.

Se transportan 2 desviaciones estándar ($2 \times \$ 237,4868$) = \$ 474,97 por encima y por debajo de la media $\mu = \$590$

Por lo tanto se tiene un intervalo desde $\$ 590 - \$474,97 = \$ 115,03$ hasta $\$ 590 + \$474,97 = \$ 1064,97$

Aplicando el Teorema de Chebyshev

$$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{4 - 1}{4} = \frac{3}{4} = 0,75 = 75\%$$

Interpretación: Se puede afirmar de que por lo menos el 75% los sueldos están entre \$ 115,03 y \$ 1064,97

Ejemplo ilustrativo N° 2

Dos empresas, A y B, venden sobres de café instantáneo de 350 gramos. Se seleccionaron al azar en los mercados cinco sobres de cada una de las compañías y se pesaron cuidadosamente sus contenidos. Los resultados fueron los siguientes.

A	B
350,14	350,09
350,18	350,12
349,98	350,20
349,99	349,88
350,12	349,95

1) ¿Qué empresa proporciona más café en sus sobres?

2) ¿Qué empresa llena sus sobres de manera más consistente?

Solución:

a) Se calcula las medias aritméticas.

$$\bar{x}_A = \frac{\sum x_i}{n} = \frac{350,14 + 350,18 + 349,98 + 349,99 + 350,12}{5}$$

$$\bar{x}_A = \frac{1750,41}{5} = 350,082$$

$$\bar{x}_B = \frac{\sum x_i}{n} = \frac{350,09 + 350,12 + 350,20 + 349,88 + 349,95}{5} = \frac{1750,24}{5} = 350,048$$

$$\bar{x}_B = \frac{1750,24}{5} = 350,048$$

Interpretación: Como la media aritmética de la empresa A es mayor que la de la empresa B, por lo tanto la empresa A proporciona más café en sus sobres.

b) Se calcula las desviaciones estándar.

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

A	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
350,14	0,058	0,003364
350,18	0,098	0,009604
349,98	-0,102	0,010404
349,99	-0,092	0,008464
350,12	0,038	0,001444
Total		0,03328

B	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
350,09	0,042	0,001764
350,12	0,072	0,005184
350,20	0,152	0,023104
349,88	-0,168	0,028224
349,95	-0,098	0,009604
		0,06788

$$s_A = \sqrt{\frac{0,03328}{5 - 1}} = \sqrt{\frac{0,03328}{4}} = 0,0912$$

$$s_B = \sqrt{\frac{0,06788}{5 - 1}} = \sqrt{\frac{0,06788}{4}} = 0,13$$

Interpretación: Como la desviación estándar de la empresa A es menor a la desviación estándar de la empresa B, por lo tanto la empresa A es más consistente al llenar los sobres de café.

Empleando Excel se calcula de la siguiente manera:

Se sigue los pasos para el cálculo de la desviación estándar de una población pero utilizando la función DESVEST.M como se muestra en las siguientes figuras:

DESVEST.M				
	A	B	C	D
1	A			
2	350,14			
3	350,18			
4	349,98			
5	349,99			
6	350,12			
7	S_A	0,091214	=DESVEST.M(A2:A6)	

DESVEST.M				
	A	B	C	D
1	B			
2	350,09			
3	350,12			
4	350,2			
5	349,88			
6	349,95			
7	S_B	0,130269	=DESVEST.M(A2:A6)	

3.2.2.2) Para Datos Agrupados en Tablas de Frecuencia

La varianza para una población se calcula con:

$$\sigma^2 = \frac{\sum f(x_i - \mu)^2}{N}$$

Donde:

f = frecuencia absoluta.

La desviación estándar poblacional se calcula con:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum f(x_i - \mu)^2}{N}}$$

La varianza de la muestra se calcula con:

$$s^2 = \frac{\sum f(x_i - \bar{x})^2}{n - 1}$$

La desviación estándar de una muestra se calcula con:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum f(x_i - \bar{x})^2}{n - 1}}$$

Ejemplo ilustrativo

Calcular la desviación estándar de los siguientes datos correspondientes a una muestra.

Calificaciones	f
4	3
5	6
6	4
7	13
8	7
10	6
Total	39

Solución:

a) Se llena la siguiente tabla:

Calificaciones	f	fx
4	3	12
5	6	30
6	4	24
7	13	91
8	7	56
10	6	60
Total	39	273

b) Se calcula la media aritmética.

$$\bar{x} = \frac{\sum f \cdot x_i}{n} = \frac{273}{39} = 7$$

c) Se llena la siguiente tabla:

Calificaciones	f	fx_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f(x_i - \bar{x})^2$
4	3	12	-3	9	27
5	6	30	-2	4	24
6	4	24	-1	1	4
7	13	91	0	0	0
8	7	56	1	1	7
10	6	60	3	9	54
Total	39	273			116

d) Se calcula la desviación estándar

$$s = \sqrt{\frac{116}{39 - 1}} = \sqrt{\frac{116}{38}} = \sqrt{3,0526} = 1,747$$

3.2.2.3) Para Datos Agrupados en Intervalos

La varianza para una población se calcula con:

$$\sigma^2 = \frac{\sum f(xm_i - \mu)^2}{N}$$

Donde:

f = frecuencia absoluta

xm = marca de clase

La desviación estándar poblacional se calcula con:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum f(xm_i - \mu)^2}{N}}$$

La varianza de la muestra se calcula con:

$$s^2 = \frac{\sum f(xm_i - \bar{x})^2}{n - 1}$$

La desviación estándar de una muestra se calcula con:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum f(xm_i - \bar{x})^2}{n - 1}}$$

Ejemplo ilustrativo

Calcular la desviación estándar de los siguientes datos correspondientes a una muestra.

Intervalo	f
60-65	5
65-70	20
70-75	40
80-85	27
85-90	8
Total	100

Solución:

a) Se llena la siguiente tabla:

Intervalo	f	xm	$f \cdot xm$
60-65	5	62,5	312,5
65-70	20	67,5	1350
70-75	40	72,5	2900
80-85	27	82,5	2227,5
85-90	8	87,5	700
Total	100		7490

b) Se calcula la media aritmética.

$$\bar{x} = \frac{\sum f \cdot xm}{n} = \frac{7490}{100} = 74,9$$

c) Se llena la siguiente tabla:

Intervalo	f	xm	$f \cdot xm$	$xm_i - \bar{x}$	$(xm_j - \bar{x})^2$	$f(xm_j - \bar{x})^2$
60-65	5	62,5	312,5	-12,4	153,76	768,8
65-70	20	67,5	1350	-7,4	54,76	1095,2
70-75	40	72,5	2900	-2,4	5,76	230,4
80-85	27	82,5	2227,5	7,6	57,76	1559,52
85-90	8	87,5	700	12,6	158,76	1270,08
Total	100		7490			4924

d) Se calcula la desviación estándar.

$$s = \sqrt{\frac{4924}{100 - 1}}$$

$$s = \sqrt{\frac{4924}{99}}$$

$$s = \sqrt{49,737}$$

$$s = 7,052$$

TAREA DE INTERAPRENDIZAJE N° 12

- 1) Elabore un organizador gráfico sobre la varianza y desviación estándar.
- 2) Consulte sobre la biografía de Pafnuty Lvovich Chebyshev y elabore un organizador gráfico sobre la misma.
- 3) Calcule la desviación estándar de las siguientes distribuciones correspondientes a una población empleando la ecuación y mediante Excel.
 - 3.1) 10, 12, 14, 16, 18, 20 y 40 9,3
 - 3.2) 30, 20, 50, 40, 60, 80 y 90 23,73
- 4) Cree y resuelva un ejercicio similar al anterior.
- 5) Cree y resuelva un ejercicio similar al ejemplo presentado para el cálculo de la desviación estándar de una muestra para datos sin agrupar.

6) Calcule la desviación estándar de una muestra dada la siguiente tabla:

x_i	f
14	6
15	5
18	4
20	12
30	7
40	6

8,903

7) El atraso diario al trabajo en la empresa D & M en el año pasado tuvo un promedio de 78 empleados con una desviación estándar de 13. Se recolectó una muestra de datos para el año en curso y se ubicaron en la siguiente tabla:

Número de empleados atrasados	Días en los que ese número estuvo atrasado
50 – 59	2
60 – 69	4
70 – 79	5
80 – 89	6
90 – 99	3
Total	20

7.1) Calcule la media aritmética.

76,5

7.2) Calcule la mediana empleando la fórmula y mediante un histograma para la fra (%).

78

7.3) Calcule la moda utilizando la fórmula y mediante un histograma para la frecuencia absoluta.

82,5

7.4) Calcule la desviación estándar.

12,4

8) Cree y resuelva un ejercicio similar al anterior.

3.3) OTRAS MEDIDAS DE DISPERSIÓN

3.3.1) RANGO, AMPLITUD TOTAL O RECORRIDO

Dada una serie de valores x_1, x_2, \dots, x_n , su recorrido es la diferencia aritmética entre el máximo y el mínimo de estos valores.

$$Re = x_{\text{máx}} - x_{\text{mín}}$$

Es la medida de dispersión más sencilla y también, por tanto, la que proporciona menos información. Además, esta información puede ser errónea, pues el hecho de que no influyan más de dos valores del total de la serie puede provocar una deformación de la realidad.

Ejemplo ilustrativo:

Calcula el rango de las siguientes distribuciones:

- 1) 4, 6, 8, 10, 12, 14, 16
- 2) 5, 10, 13, 13, 14, 15, 17

Solución:

$$Re_1 = 16 - 4 = 12$$

$$Re_2 = 17 - 5 = 12$$

En Excel se inserta la función $\text{MAX}(\text{Celdas}) - \text{MIN}(\text{Celdas})$ como muestra la siguiente figura:

	A	B	C	D	E	F	G	H
1	4				5			
2	6				10			
3	8				13			
4	10				13			
5	12				14			
6	14				15			
7	16				17			
8								
9	Recorrido	12	=MAX(A1:A7)-MIN(A1:A7)		Recorrido	12	=MAX(E1:E7)-MIN(E1:E7)	

Ambas series tienen rango 12, pero están desigualmente distribuidas, pues mientras la primera se distribuye uniformemente a lo largo de todo el recorrido, la segunda tiene una mayor concentración en el centro.

La amplitud es una medida de dispersión cuya ventaja es la facilidad con que se calcula. Tiene en cambio las siguientes desventajas:

- En su cálculo sólo intervienen dos elementos del conjunto.
- Al aumentar el número de observaciones, puede esperarse que aumente la variabilidad. Puesto que la amplitud no tiene en cuenta el tamaño del conjunto, no es una medida adecuada para comparar la variabilidad de dos grupos de observaciones, a menos que éstos sean del mismo tamaño.

Nota: Cuando los datos están agrupados en intervalos se calcula la amplitud sacando la diferencia entre la marca de clase mayor y la marca de clase menor.

3.3.2) AMPLITUD INTERCUARTÍLICA

La amplitud intercuartílica es la distancia entre el tercer cuartil Q_3 y el primer cuartil Q_1 .
Amplitud intercuartílica = tercer cuartil - primer cuartil = $Q_3 - Q_1$

3.3.3) RANGO SEMI-INTERCUARTIL O DESVIACIÓN CUARTÍLICA

La desviación cuartílica es la mitad de la distancia entre el tercer cuartil y el primero

$$DQ = \frac{Q_3 - Q_1}{2}$$

Ejemplo ilustrativo:

Si el tercer cuartil = 24 y el primer cuartil = 10. ¿Cuál es la desviación cuartílica?

Solución:

La amplitud intercuartílica es $24 - 10 = 14$;

Por lo tanto, la desviación cuartílica es:

$$DQ = \frac{14}{2}$$

3.3.4) RANGO PERCENTIL O AMPLITUD CUARTÍLICA

Cada conjunto de datos tiene 99 percentiles, que dividen el conjunto en 100 partes iguales.

La amplitud cuartílica es la distancia entre dos percentiles establecidos.

El rango percentil o amplitud cuartílica 10 a 90 es la distancia entre el 10° y 90° percentiles, definida por

$$\text{Rango percentil } 10 - 90 = P_{90} - P_{10}$$

TAREA DE INTERAPRENDIZAJE N° 13

- 1) Realice un organizador gráfico sobre las otras medidas de dispersión.
- 2) Plantee una distribución de datos sin agrupar y calcule el rango, la amplitud intercuartílica, el rango semi-intercuartil y el rango percentil de manera manual y empleando Excel.
- 3) Plantee una distribución de datos agrupados en tablas de frecuencia y calcule el rango, la amplitud intercuartílica, el rango semi-intercuartil y el rango percentil.
- 4) Plantee una distribución de datos agrupados en intervalos y calcule el rango, la amplitud intercuartílica, el rango semi-intercuartil y el rango percentil.

3.4) DISPERSIÓN RELATIVA O COEFICIENTE DE VARIACIÓN

Las medidas de dispersión anteriores son todas medidas de variación absolutas. Una medida de dispersión relativa de los datos, que toma en cuenta su magnitud, está dada por el coeficiente de variación.

El Coeficiente de variación (CV) es una medida de la dispersión relativa de un conjunto de datos, que se obtiene dividiendo la desviación estándar del conjunto entre su media aritmética y se expresa generalmente en términos porcentuales.

3.4.1) PROPIEDADES

- Puesto que tanto la desviación estándar como la media se miden en las unidades originales, el CV es una medida independiente de las unidades de medición.
- Debido a la propiedad anterior el CV es la cantidad más adecuada para comparar la variabilidad de dos conjuntos de datos.

3.4.2) MÉTODOS DE CÁLCULO

Para una población se emplea la siguiente fórmula:

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Donde:

$CV =$ Coeficiente de variación

$\sigma =$ desviación estándar de la población

$\mu =$ media aritmética de la población

Para una muestra se emplea la siguiente fórmula:

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Donde:

$CV =$ Coeficiente de variación

$s =$ desviación estándar de la muestra

$\bar{x} =$ media aritmética de la muestra

Ejemplo ilustrativo N° 1

Mathías, un estudiante universitario, tiene las siguientes calificaciones en las 10 asignaturas que recibe en su carrera: 8, 7, 10, 9, 8, 7, 8, 10, 9 y 10. Josué, un compañero de Mathías, tiene las siguientes calificaciones: 8, 9, 8, 7, 8, 9, 10, 7, 8 y 10. ¿Cuál estudiante tiene menor variabilidad en sus calificaciones?

Solución:

Como se está tomando en cuenta todas las asignaturas, se debe calcular el coeficiente de variación poblacional.

Sin agrupar los datos empleando Excel se calcula el coeficiente de variación tal como se muestra en la siguiente figura:

	A	B	C	D	E	F	G	H
1	Mathias				Josué			
2	8				7			
3	7				7			
4	10				8			
5	9				8			
6	8				8			
7	7				8			
8	8				9			
9	10				9			
10	9				10			
11	10				10			
12	μ	8,6	=PROMEDIO(A2:A11)		μ	8,4	=PROMEDIO(E2:E11)	
13	σ	1,1135529	=DESVEST.P(A2:A11)		σ	1,0198039	=DESVEST.P(E2:E11)	
14	CV	12,948289	=(B13/B12)*100		CV	12,140523	=(F13/F12)*100	

Agrupando los datos en tablas de frecuencias se calcula así:

a) Se agrupa las calificaciones y se realiza el cálculo la media aritmética

Para Mathías se obtiene:

Calificaciones (x_i)	f	fx_i
7	2	14
8	3	24
9	2	18
10	3	30
Total	10	86

$$\mu = \frac{\sum fx_i}{N} = \frac{86}{10} = 8,6$$

Para Josué se obtiene:

Calificaciones (x_i)	f	fx_i
7	2	14
8	4	32
9	2	18
10	2	20
Total	10	84

$$\mu = \frac{\sum fx_i}{N} = \frac{84}{10} = 8,4$$

b) Se calcula la desviación estándar

Para Mathías se obtiene:

Calificaciones (x_i)	f	fx_i	$(x_i - \mu)^2$	$f(x_i - \mu)^2$
7	2	14	2,56	5,12
8	3	24	0,36	1,08
9	2	18	0,16	0,32
10	3	30	1,96	5,88
Total	10	86		12,4

$$\sigma = \sqrt{\frac{\sum f(x_i - \mu)^2}{N}} = \sqrt{\frac{12,4}{10}} = 1,1136$$

Para Josué se obtiene:

Calificaciones (x_i)	f	fx_i	$(x_i - \mu)^2$	$f(x_i - \mu)^2$
7	2	14	1,96	3,92
8	4	32	0,16	0,64
9	2	18	0,36	0,72
10	2	20	2,56	5,12
Total	10	84		10,4

$$\sigma = \sqrt{\frac{\sum f(x_i - \mu)^2}{N}} = \sqrt{\frac{10,4}{10}} = 1,0198$$

c) Se calcula el coeficiente de variación

Para Mathías se obtiene:

$$CV = \frac{\sigma}{\mu} = \frac{1,1136}{8,6} = 0,129 = 12,9\%$$

Empleando Excel es como muestra la siguiente figura:

	A	B	C	D	E	F	G
1	x_i	f	$f(x_i - \mu)^2$				
2	7	2	5,12	=B2*(A2-SB\$7)^2			
3	8	3	1,08	=B3*(A3-SB\$7)^2			
4	9	2	0,32	=B4*(A4-SB\$7)^2			
5	10	3	5,88	=B5*(A5-SB\$7)^2			
6	Total	10	12,4	=SUMA(C2:C5)			
7	μ	8,6	=SUMAPRODUCTO(A2:A5;B2:B5)/SUMA(B2:B5)				
8							
9	σ	1,1135529	=RCUAD(C6/SUMA(B2:B5))				
10							
11	CV	12,948289	=(B9/B7)*100				

Para Josué se obtiene:

$$CV = \frac{\sigma}{\mu} = \frac{1,0198}{8,4} = 0,121 = 12,1\%$$

Empleando Excel es como muestra la siguiente figura:

	A	B	C	D	E	F
1	x_i	f	$f(x_i - \mu)^2$			
2	7	2	3,92	=B2*(A2-\$B\$7)^2		
3	8	4	0,64	=B3*(A3-\$B\$7)^2		
4	9	2	0,72	=B4*(A4-\$B\$7)^2		
5	10	2	5,12	=B5*(A5-\$B\$7)^2		
6	Total	10	10,4	=SUMA(C2:C5)		
7	μ	8,4	=SUMAPRODUCTO(A2:A5;B2:B5)/SUMA(B2:B5)			
8						
9	σ	1,0198039	=RCUAD(C6/SUMA(B2:B5))			
10						
11	CV	12,140523	=(B9/B7)*100			

Interpretación: Por lo tanto, Josué tiene menor variabilidad en sus calificaciones

Ejemplo ilustrativo N° 2

Se saca una muestra de un curso de la Universidad UTN sobre las calificaciones en las asignaturas de Matemática y Estadística, resultados que se presentan en las siguientes tablas. ¿En qué asignatura existe mayor variabilidad?. Realice los cálculos empleando Excel

Matemática	
Intervalos	f
2 - 4	8
5 - 7	12
8 - 10	20
Total	40

Estadística	
Intervalos	f
2 - 4	8
5 - 7	14
8 - 10	18
Total	40

Solución:

Los cálculos para la asignatura de Matemática empleando Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H	I
1	Intervalos		f	xm			$f(xm_i - \bar{x})^2$		
2	2	4	10	3	=PROMEDIO(A2:B2)		152,1	=C2*(D2-SB\$7)^2	
3	5	7	8	6	=PROMEDIO(A3:B3)		6,48	=C3*(D3-SB\$7)^2	
4	8	10	22	9	=PROMEDIO(A4:B4)		97,02	=C4*(D4-SB\$7)^2	
5	Total		40	=SUMA(C2:C4)			255,6	=SUMA(G2:G4)	
6									
7	\bar{x}	6,9	=SUMAPRODUCTO(D2:D4;C2:C4)/C5						
8									
9	s	2,56	=RCUAD(G5/(C5-1))						
10									
11	CV	37,1	=(B9/B7)*100						

Los cálculos para la asignatura de Estadística empleando Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H	I
1	Intervalos		f	xm			$f(xm_i - \bar{x})^2$		
2	2	4	8	3	=PROMEDIO(A2:B2)		112,5	=C2*(D2-SB\$7)^2	
3	5	7	14	6	=PROMEDIO(A3:B3)		7,875	=C3*(D3-SB\$7)^2	
4	8	10	18	9	=PROMEDIO(A4:B4)		91,125	=C4*(D4-SB\$7)^2	
5	Total		40	=SUMA(C2:C4)			211,5	=SUMA(G2:G4)	
6									
7	\bar{x}	6,75	=SUMAPRODUCTO(D2:D4;C2:C4)/C5						
8									
9	s	2,33	=RCUAD(G5/(C5-1))						
10									
11	CV	34,5	=(B9/B7)*100						

Interpretación: Por lo tanto el curso presenta mayor variabilidad en la asignatura de Matemática.

TAREA DE INTERAPRENDIZAJE N° 14

1) Calcule el coeficiente de variación de las siguientes distribuciones de datos referentes a poblaciones. Realice los cálculos empleando la fórmula respectiva y utilizando Excel.

1.1) 6, 8, 10, 4, 7, 8, 9, 8, 4 y 6

27,105 %

1.1) 6, 6, 8, 8, 8, 8, 9, 9, 5 y 7

17,31 %

2) Calcule el coeficiente de variación de las siguientes distribuciones de datos referentes a muestras. Realice los cálculos empleando la fórmula respectiva y utilizando Excel.

2.1) 6, 8, 10, 9, 8, 7, 8, 10, 9 y 10

15,93 %

2.2) 6, 6, 8, 10, 8, 10, 9, 9, 5 y 7

22,45 %

3) Calcule el coeficiente de variación de manera manual y empleando Excel empleando los datos de la siguiente tabla correspondientes a una población.

x_i	f
7	4
8	8
9	12
10	6
Total	30

10,88%

4) Cree y resuelva un ejercicio similar al presentado en el ejemplo 1 con datos sin agrupar y agrupando en tablas de frecuencias. Resolver de manera manual y empleando Excel.

5) Calcule el coeficiente de variación de manera manual y empleando Excel utilizando los datos de la siguiente tabla correspondientes a una muestra.

Intervalos	f
2 - 4	10
5 - 7	8
8 - 10	22
Total	40

37,1 %

6) Cree y resuelva un ejercicio similar al presentado en el ejemplo 2. Resuelva de manera manual y empleando Excel.

CAPÍTULO IV

MEDIDAS DE FORMA

COMPETENCIAS:

- ✓ Interpreta adecuadamente las características y tipos de asimetría y curtosis.
- ✓ Emplea correctamente algoritmos matemáticos para calcular medidas de asimetría y curtosis en forma manual y empleando Excel.
- ✓ Crea y resuelve correctamente ejercicios de aplicación sobre el cálculo de medidas de asimetría y curtosis.

CONTENIDOS:

- ✓ Asimetría: Tipos de Asimetría y Medidas de Asimetría
- ✓ Curtosis o Apuntamiento: Tipos de Curtosis y Medidas de Curtosis.

4.1) ASIMETRÍA

Es una medida de forma de una distribución que permite identificar y describir la manera como los datos tiende a reunirse de acuerdo con la frecuencia con que se hallen dentro de la distribución. Permite identificar las características de la distribución de datos sin necesidad de generar el gráfico.

2.4.1) TIPOS DE ASIMETRÍA

La asimetría presenta las siguientes formas:

2.4.1.1) Asimetría Negativa o a la Izquierda

Se da cuando en una distribución la minoría de los datos está en la parte izquierda de la media. Este tipo de distribución presenta un alargamiento o sesgo hacia la izquierda, es decir, la distribución de los datos tiene a la izquierda una cola más larga que a la derecha.

También se dice que una distribución es simétrica a la izquierda o tiene sesgo negativo cuando el valor de la media aritmética es menor que la mediana y éste valor de la mediana a su vez es menor que la moda, en símbolos $\bar{x} < Md < Mo$.

Nota: Sesgo es el grado de asimetría de una distribución, es decir, cuánto se aparta de la simetría.

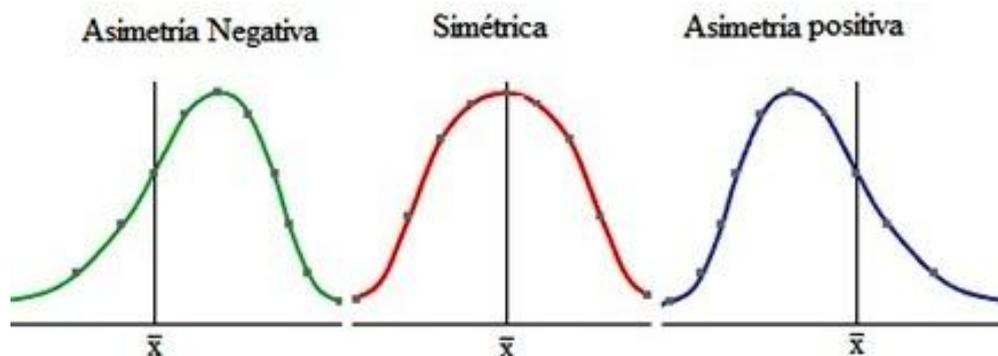
2.4.1.2) Simétrica

Se da cuando en una distribución se distribuyen aproximadamente la misma cantidad de los datos a ambos lados de la media aritmética. No tiene alargamiento o sesgo. Se representa por una curva normal en forma de campana llamada campana de Gauss (matemático Alemán 1777-1855) o también conocida como de Laplace (1749-1827). También se dice que una distribución es simétrica cuando su media aritmética, su mediana y su moda son iguales, en símbolos $\bar{x} = Md = Mo$

2.4.1.3) Asimetría Positiva o a la Derecha

Se da cuando en una distribución la minoría de los datos está en la parte derecha de la media aritmética. Este tipo de distribución presenta un alargamiento o sesgo hacia la derecha, es decir, la distribución de los datos tiene a la derecha una cola más larga que a la izquierda.

También se dice que una distribución es simétrica a la derecha o tiene sesgo positivo cuando el valor de la media aritmética es mayor que la mediana y éste a valor de la mediana a su vez es mayor que la moda, en símbolos $\bar{x} > Md > Mo$.



2.4.2) MEDIDAS DE ASIMETRÍA

4.4.2.1) Coeficiente de Karl Pearson

$$As = \frac{3(\bar{x} - Md)}{s}$$

Donde:

\bar{x} = media aritmética.

Md = Mediana.

s = desviación típica o estándar.

Nota:

El Coeficiente de Pearson varía entre -3 y 3

Si $As < 0 \rightarrow$ la distribución será asimétrica negativa.

Si $As = 0 \rightarrow$ la distribución será simétrica.

Si $As > 0 \rightarrow$ la distribución será asimétrica positiva.

4.4.2.2) Medida de Yule Bowley o Medida Cuartílica

$$As = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1}$$

Donde:

Q_1 = Cuartil uno; Q_2 = Cuartil dos = Mediana; Q_3 = Cuartil tres.

Nota:

La Medida de Bowley varía entre -1 y 1

Si $As < 0 \rightarrow$ la distribución será asimétrica negativa.

Si $As = 0 \rightarrow$ la distribución será simétrica.

Si $As > 0 \rightarrow$ la distribución será asimétrica positiva.

4.4.2.3) Medida de Fisher

Para datos sin agrupar se emplea la siguiente fórmula:

$$As = \frac{\sum(x_i - \bar{x})^3}{n\sigma^3}$$

Para datos agrupados en tablas de frecuencias se emplea la siguiente fórmula:

$$As = \frac{\sum f(x_i - \bar{x})^3}{n\sigma^3}$$

Para datos agrupados en intervalos se emplea la siguiente fórmula:

$$As = \frac{\sum f(xm - \bar{x})^3}{n\sigma^3}$$

Donde:

x_i = cada uno de los valores

n = número de datos

\bar{x} = media aritmética

f = frecuencia absoluta

σ^3 = cubo de la desviación estándar poblacional

xm = marca de clase

Nota:

Si $As < 0$ → Indica que existe presencia de la minoría de datos en la parte izquierda de la media, aunque en algunos casos no necesariamente indicará que la distribución sea asimétrica negativa

Si $As = 0$ → la distribución será simétrica

Si $As > 0$ → Indica que existe presencia de la minoría de datos en la parte derecha de la media, aunque en algunos casos no necesariamente indicará que la distribución sea asimétrica positiva

Ejemplo ilustrativo:

Calcular el Coeficiente de Pearson, Medida Cuartílica y la Medida de Fisher dada la siguiente distribución: 6, 9, 9, 12, 12, 12, 15 y 17

Solución:

Calculando la media aritmética se obtiene:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{6 + 9 + 9 + 12 + 12 + 12 + 15 + 17}{8} = \frac{92}{8} = 11,5$$

Para calcular los cuartiles se ordena los datos de menor a mayor

6	9	9	12	12	12	15	17
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8

Calculando el cuartil uno se obtiene:

$$Q_k = X_{\left[\frac{n \cdot k + 2}{4}\right]}$$

$$Q_1 = X_{\left[\frac{n+2}{4}\right]} = X_{\left[\frac{8+2}{4}\right]} = X_{\left[\frac{10}{4}\right]} = X_{2,5} = \frac{x_2 + x_3}{2} = \frac{9 + 9}{2} = 9$$

Calculando el cuartil dos se obtiene:

$$Q_k = X_{\left[\frac{n \cdot k + 2}{4}\right]}$$

$$Q_2 = X_{\left[\frac{n \cdot 2 + 2}{4}\right]} = X_{\left[\frac{2n+2}{4}\right]} = X_{\left[\frac{2 \cdot 8 + 2}{4}\right]} = X_{\left[\frac{16+2}{4}\right]} = X_{4,5} = \frac{x_4 + x_5}{2} = \frac{12 + 12}{2} = 12$$

Calculando el cuartil tres se obtiene:

$$Q_k = X_{\left[\frac{n \cdot k + 2}{4}\right]}$$

$$Q_3 = X_{\left[\frac{3n+2}{4}\right]} = X_{\left[\frac{3 \cdot 8 + 2}{4}\right]} = X_{\left[\frac{24+2}{4}\right]} = X_{\frac{26}{4}} = X_{6,5} = \frac{x_6 + x_7}{2} = \frac{12 + 15}{2} = 13,5$$

Calculando la desviación estándar muestral se obtiene:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

$$s = \sqrt{\frac{(6 - 11,5)^2 + (9 - 11,5)^2 + (9 - 11,5)^2 + (12 - 11,5)^2 + (12 - 11,5)^2 + (12 - 11,5)^2 + (15 - 11,5)^2 + (17 - 11,5)^2}{8 - 1}}$$

$$s = 3,505$$

Calculando el Coeficiente de Pearson se obtiene:

$$As = \frac{3(\bar{x} - Md)}{s} = \frac{3(11,5 - 12)}{3,505} = \frac{-1,5}{3,505} = -0,428$$

Calculando la Medida de Bowley se obtiene

$$As = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1} = \frac{9 + 13,5 - 2 \cdot 12}{13,5 - 9} = -0,333$$

Calculando la desviación estándar poblacional se obtiene:

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

$$\sigma = \sqrt{\frac{(6 - 11,5)^2 + (9 - 11,5)^2 + (9 - 11,5)^2 + (12 - 11,5)^2 + (12 - 11,5)^2 + (12 - 11,5)^2 + (15 - 11,5)^2 + (17 - 11,5)^2}{8}}$$

$$\sigma = 3,279$$

Calculando la Medida de Fisher se obtiene

Datos	$(x_i - \bar{x})^3$
6	-166,375
9	-15,625
9	-15,625
12	0,125
12	0,125
12	0,125
15	42,875
17	166,375
Total	12

$$As = \frac{\sum(x_i - \bar{x})^3}{n\sigma^3} = \frac{12}{8(3,279)^3} = 0,035$$

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G
1	Datos	$(x_i - \bar{x})^3$					
2	6	-166,375					
3	9	-15,625					
4	9	-15,625					
5	12	0,125					
6	12	0,125					
7	12	0,125					
8	15	42,875					
9	17	166,375					
10	Total	12	=SUMA(B2:B9)				
11	n	8	=CONTAR(A2:A9)				
12	Media aritmética	11,5	=PROMEDIO(A2:A9)				
13	Desviación estándar	3,5050983	=DESVEST.M(A2:A9)				
14	Desviación poblacional	3,2787193	=DESVEST.P(A2:A9)				
15	Cuartil 1	9	=CUARTIL.INC(A2:A9;1)				
16	Cuartil 2	12	=CUARTIL.INC(A2:A9;2)				
17	Cuartil 3	13,5	=CUARTIL.INC(A2:A9;3)+0,25*(A8-A7)				
18	Coefficiente de Pearson						
19	$As = \frac{3(\bar{x} - Md)}{s}$	-0,427948	=3*(B12-B16)/B13				
20							
21	Medida de Bowley						
22	$As = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1}$	-0,333333	=(B15+B17-2*B16)/(B17-B15)				
23							
24	Medida de Fisher						
25	$As = \frac{\sum(x_i - \bar{x})^3}{n\sigma^3}$	0,0425577	=B10/(B11*B14^3)				
26							
27	Coefficiente de Asimetría en Excel	0,0530788	=COEFICIENTE.ASIMETRIA(A2:A9)				

Nota: El COEFICIENTE.ASIMETRIA(A2:A9) es un valor que tiene consideraciones semejantes a la Medida de Fisher

TAREA DE INTERAPRENDIZAJE N° 15

- 1) Realice un organizador gráfico sobre la asimetría.
- 2) Consulte y realice un organizador gráfico para cada una de las biografías de Gauss, Laplace, Pearson, Bowley y Fisher.
- 3) Calcule empleando las fórmulas y mediante Excel el Coeficiente de Pearson, Medida de Bowley y la Medida de Fisher dadas las siguientes distribuciones.
 - 3.1) 4, 4, 8, 14, 14, 16, 18 y 20
-0,85; -0,45; -0,31
 - 3.1) 6, 8, 10, 12, 14, 16, 18 y 20
0
- 4) Cree y resuelva un ejercicio para el cálculo del Coeficiente de Pearson, Medida de Bowley y la Medida de Fisher para datos agrupados en tablas de frecuencias, y otro ejercicio para datos agrupados en intervalos. Emplee los conocimientos adquiridos en los anteriores capítulos.

4.2) CURTOSIS O APUNTAMIENTO

La curtosis mide el grado de agudeza o achatamiento de una distribución con relación a la distribución normal, es decir, mide cuán puntiaguda es una distribución.

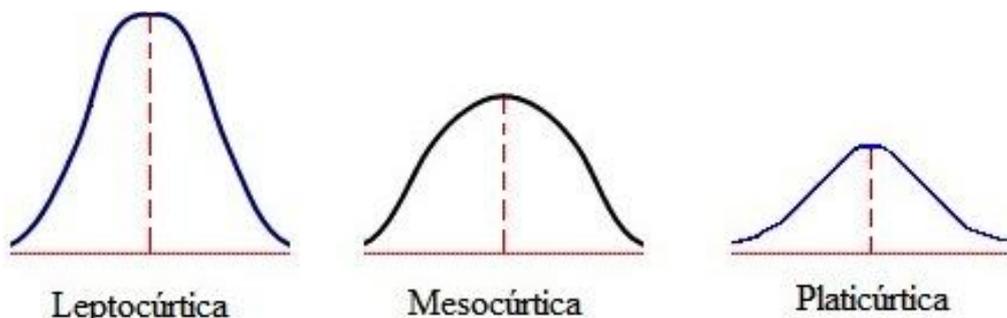
4.2.1) TIPOS DE CURTOSIS

La curtosis determina el grado de concentración que presentan los valores en la región central de la distribución. Así puede ser:

4.2.1.2) Leptocúrtica.- Existe una gran concentración.

4.2.1.3) Mesocúrtica.- Existe una concentración normal.

4.2.1.4) Platicúrtica.- Existe una baja concentración.



4.2.2) MEDIDAS DE CURTOSIS

4.2.2.1) Medida de Fisher

Para datos sin agrupar se emplea la siguiente fórmula:

$$\alpha = \frac{\sum(x_i - \bar{x})^4}{n\sigma^4}$$

Para datos agrupados en tablas de frecuencias se emplea la siguiente fórmula:

$$\alpha = \frac{\sum f(x_i - \bar{x})^4}{n\sigma^4}$$

Para datos agrupados en intervalos se emplea la siguiente fórmula:

$$\alpha = \frac{\sum f(xm - \bar{x})^4}{n\sigma^4}$$

Donde:

x_i = cada uno de los valores

n = número de datos

\bar{x} = media aritmética

σ^4 = Cuádruplo de la desviación estándar poblacional

f = frecuencia absoluta

xm = marca de clase

Nota:

Si $\alpha < 3$ → la distribución es platicúrtica

Si $\alpha = 3$ → la distribución es normal o mesocúrtica

Si $\alpha > 3$ → la distribución es leptocúrtica

4.2.2.2) Medida basada en Cuartiles y Percentiles

$$\kappa = \frac{\text{Desviación cuartílica}}{\text{Amplitud cuartílica}} = \frac{\frac{Q_3 - Q_1}{2}}{P_{90} - P_{10}} = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

κ (letra griega minúscula kappa) = Coeficiente percentil de curtosis

Nota:

Si $\kappa < 0,263$ → la distribución es platicúrtica

Si $\kappa = 0,263$ → la distribución es normal o mesocúrtica

Si $\kappa > 0,263$ → la distribución es leptocúrtica

Esta medida no es muy utilizada.

Ejemplo ilustrativo: Determinar qué tipo de curtosis tiene la siguiente distribución: 6, 9, 9, 12, 12, 12, 15 y 17. Emplear la medida de Fisher y el coeficiente percentil de curtosis.

Solución:

Calculando la media aritmética se obtiene

$$\bar{x} = \frac{\sum x_i}{n} = \frac{6 + 9 + 9 + 12 + 12 + 12 + 15 + 17}{8} = \frac{92}{8} = 11,5$$

Calculando la desviación estándar poblacional se obtiene:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

$$\sigma = \sqrt{\frac{(6 - 11,5)^2 + (9 - 11,5)^2 + (9 - 11,5)^2 + (12 - 11,5)^2 + (12 - 11,5)^2 + (12 - 11,5)^2 + (15 - 11,5)^2 + (17 - 11,5)^2}{8}}$$

$$\sigma = 3,279$$

Calculando la Medida de Fisher se obtiene:

Datos	$(x_i - \bar{x})^4$
6	915,0625
9	39,0625
9	39,0625
12	0,0625
12	0,0625
12	0,0625
15	150,0625
17	915,0625
Total	2058,5

$$\alpha = \frac{\sum (x_i - \bar{x})^4}{n\sigma^4} = \frac{2058,5}{8 \cdot (3,279)^4} = 2,23$$

Para calcular los cuartiles y percentiles se ordena los datos de menor a mayor:

6	9	9	12	12	12	15	17
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8

Calculando el cuartil uno se obtiene:

$$Q_k = X_{\left[\frac{n \cdot k + 2}{4}\right]}$$

$$Q_1 = X_{\left[\frac{n+2}{4}\right]} = X_{\left[\frac{8+2}{4}\right]} = X_{\left[\frac{10}{4}\right]} = X_{2,5} = \frac{x_2 + x_3}{2} = \frac{9 + 9}{2} = 9$$

Calculando el cuartil tres se obtiene:

$$Q_k = X_{\left[\frac{n \cdot k + 2}{4}\right]}$$

$$Q_3 = X_{\left[\frac{3n+2}{4}\right]} = X_{\left[\frac{3 \cdot 8 + 2}{4}\right]} = X_{\left[\frac{24+2}{4}\right]} = X_{\frac{26}{4}} = X_{6,5} = \frac{x_6 + x_7}{2} = \frac{12 + 15}{2} = 13,5$$

Calculando el percentil 90 se tiene:

$$P_k = X_{\left[\frac{n \cdot k + 50}{100}\right]}$$

$$P_{90} = X_{\left[\frac{n \cdot 90 + 50}{100}\right]} = X_{\left[\frac{8 \cdot 90 + 50}{100}\right]} = X_{\left[\frac{770}{100}\right]} = X_{7,7} = \frac{x_7 + x_8}{2} = \frac{15 + 17}{2} = 16$$

Calculando el percentil 10 se tiene:

$$P_k = X_{\left[\frac{n \cdot k + 50}{100}\right]}$$

$$P_{10} = X_{\left[\frac{n \cdot 10 + 50}{100}\right]} = X_{\left[\frac{8 \cdot 10 + 50}{100}\right]} = X_{\left[\frac{130}{100}\right]} = X_{1,3} = x_1 = 6$$

Calculando el coeficiente percentil de curtosis se obtiene:

$$\kappa = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{13,5 - 9}{2(16 - 6)} = 0,225$$

Como $\alpha = 2,23$ y $\kappa = 0,225$, la distribución es platicúrtica

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F
1	Datos	$(x_i - \bar{x})^4$				
2	6	915,06250				
3	9	39,06250				
4	9	39,06250				
5	12	0,06250				
6	12	0,06250				
7	12	0,06250				
8	15	150,06250				
9	17	915,06250				
10	Total	2058,5	=SUMA(B2:B9)			
11	n	8	=CONTAR(A2:A9)			
12	Media aritmética	11,5	=PROMEDIO(A2:A9)			
13	Desviación poblacional	3,2787193	=DESVEST.P(A2:A9)			
14	Cuartil 1	9	=CUARTIL.INC(A2:A9;1)			
15	Cuartil 3	13,5	=CUARTIL.INC(A2:A9;3)+0,25*(A8-A7)			
16	$\alpha = \frac{\sum(x_i - \bar{x})^4}{n\sigma^4}$	2,226609	=B10/(B11*B13^4)			
17						
18	Percentil 10	7,4	=PERCENTIL.INC(A2:A9;0,1)-0,25*(A3-A2)			
19	Percentil 90	16,350	=PERCENTIL.INC(A2:A9;0,9)+0,25*(A8-A7)			
20	$\kappa = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$	0,2500	=(B15-B14)/(2*(B19-B18))			
21						
22						
23	Curtosis en Excel	-0,224121	=CURTOSIS(A2:A9)			
24	Valor semejante a la α	2,7758789	=B23+3			

TAREA DE INTERAPRENDIZAJE N° 16

- 1) Realice un organizador gráfico sobre la curtosis.
- 2) Cree y resuelva un ejercicio similar al presentado para el cálculo de las medidas de curtosis con datos sin agrupar.
- 3) Resuelva el ejercicio anterior empleando Excel.
- 4) Cree y resuelva un ejercicio para el cálculo de las medidas de curtosis con datos agrupados en tablas de frecuencia, y otro ejercicio con datos agrupados en intervalos. Emplee los conocimientos adquiridos en los anteriores capítulos.

CAPÍTULO V

CORRELACIÓN Y RELACIÓN

COMPETENCIAS:

- ✓ Describe adecuadamente con sus propias palabras el significado de correlación y regresión.
- ✓ Emplea correctamente algoritmos matemáticos para resolver ejercicios de aplicación sobre correlación y regresión de manera manual y empleando Excel.
- ✓ Elabora eficientemente diagramas de dispersión y líneas de regresión de manera manual, empleando Excel y Graph.
- ✓ Crea y resuelve correctamente ejercicios de aplicación sobre correlación y regresión.

CONTENIDOS:

- ✓ Análisis de Correlación: Diagrama de Dispersión, Clasificación de la Correlación, Coeficientes de Correlación y Coeficiente de Determinación.
- ✓ Análisis de Regresión: Principio de los Mínimos Cuadrados y Error Estándar de Estimación.

Cuando se estudian en forma conjunta dos características (variables estadísticas) de una población o muestra, se dice que estamos analizando una variable estadística bidimensional. La correlación es el grado de relación que existe entre ambas características, y la regresión es la forma de expresar matemáticamente dicha relación.

5.1) ANÁLISIS DE CORRELACIÓN

Dado dos variables, la correlación permite hacer estimaciones del valor de una de ellas conociendo el valor de la otra variable.

5.1.1) DIAGRAMA DE DISPERSIÓN

Los diagramas de dispersión son planos cartesianos en los que se marcan los puntos correspondientes a los pares ordenados (X,Y) de los valores de las variables.

5.1.2) CLASIFICACIÓN DE LA CORRELACIÓN

5.1.2.1) Según la relación entre variables

- **Correlación lineal:** Se representa mediante una línea recta.
- **Correlación no lineal:** Se representa con una línea curva.

5.1.2.2) Según el número de variables

- **Correlación simple:** La variable dependiente actúa sobre la variable independiente.
- **Correlación múltiple:** Cuando la variable dependiente actúa sobre varias variables independientes.
- **Correlación parcial:** Cuando la relación que existe entre una variable dependiente y una independiente es de tal forma que los demás factores permanezcan constantes.

5.1.2.3) Según el valor cuantitativo

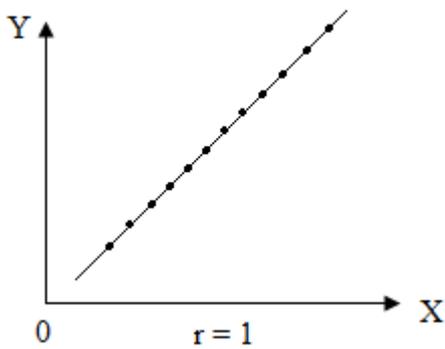
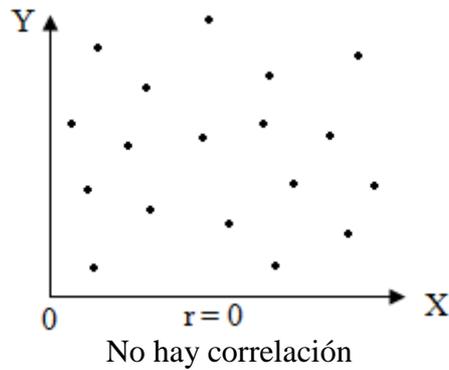
- **Correlación perfecta:** El valor del coeficiente de correlación es 1
- **Correlación imperfecta:** El coeficiente de correlación es menor a 1 sea en sentido positivo o negativo.
- **Correlación nula:** El coeficiente de correlación es 0. No existe correlación entre las variables. Ejemplo: Número de calzado de una persona y su cociente intelectual.

5.1.2.4) Según el signo

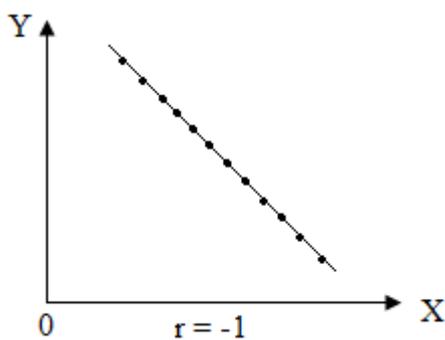
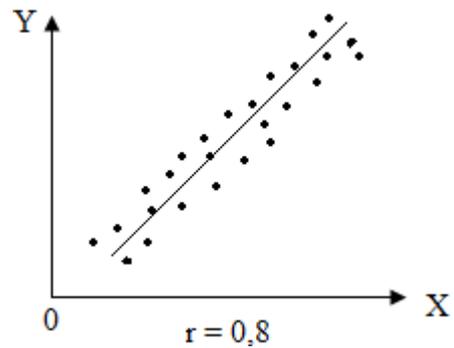
- **Correlación positiva.-** Dos variables tiene correlación positiva cuando al aumentar o disminuir el valor de una de ellas entonces el valor correspondiente a la otra aumentará o disminuirá respectivamente, es decir, cuando las dos variables aumentan en el mismo sentido. Ejemplo: Peso de una persona y su talla.
- **Correlación negativa.-** Dos variables tiene correlación negativa cuando al aumentar o disminuir el valor de una de ellas entonces el valor de la otra disminuirá o aumentará respectivamente, es decir, una variable aumenta y otra disminuye o viceversa. Ejemplo: Número de partidos ganados por un equipo en una temporada y su posición final en la tabla.

5.1.3) COEFICIENTES DE CORRELACIÓN

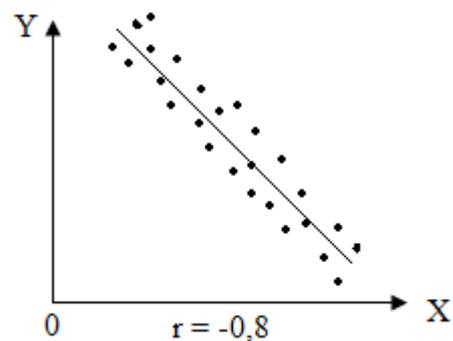
Los coeficientes de correlación son medidas que indican la situación relativa de los mismos sucesos respecto a las dos variables, es decir, son la expresión numérica que nos indica el grado de relación existente entre las 2 variables y en qué medida se relacionan. Son números que varían entre los límites +1 y -1. Su magnitud indica el grado de asociación entre las variables; el valor $r = 0$ indica que no existe relación entre las variables; los valores ± 1 son indicadores de una correlación perfecta positiva (al crecer o decrecer X, crece o decrece Y) o negativa (Al crecer o decrecer X, decrece o crece Y).



Correlación Positiva



Correlación Negativa



Para interpretar el coeficiente de correlación utilizamos la siguiente escala:

Valor	Significado
-1	Correlación negativa grande y perfecta
-0,9 a -0,99	Correlación negativa muy alta
-0,7 a -0,89	Correlación negativa alta
-0,4 a -0,69	Correlación negativa moderada
-0,2 a -0,39	Correlación negativa baja
-0,01 a -0,19	Correlación negativa muy baja
0	Correlación nula
0,01 a 0,19	Correlación positiva muy baja
0,2 a 0,39	Correlación positiva baja
0,4 a 0,69	Correlación positiva moderada
0,7 a 0,89	Correlación positiva alta
0,9 a 0,99	Correlación positiva muy alta
1	Correlación positiva grande y perfecta

5.1.3.1) COEFICIENTE DE CORRELACIÓN DE KARL PEARSON

Llamando también coeficiente de correlación producto-momento.

a) Para datos no agrupados se calcula aplicando la siguiente ecuación:

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

r = Coeficiente producto-momento de correlación lineal

$$x = X - \bar{X}; y = Y - \bar{Y}$$

Ejemplo ilustrativo:

Con los datos sobre las temperaturas en dos días diferentes en una ciudad, determinar el tipo de correlación que existe entre ellas mediante el coeficiente de PEARSON.

X	18	17	15	16	14	12	9	15	16	14	16	18	$\Sigma X = 180$
Y	13	15	14	13	9	10	8	13	12	13	10	8	$\Sigma Y = 138$

Solución:

Se calcula la media aritmética

$$\bar{x} = \frac{\sum x_i}{n}$$

Para X:

$$\bar{X}_X = \frac{180}{12} = 15$$

Para Y:

$$\bar{Y}_Y = \frac{138}{12} = 11,5$$

Se llena la siguiente tabla:

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	xy	y^2
18	13	3	1,5	9	4,5	2,25
17	15	2	3,5	4	7	12,25
15	14	0	2,5	0	0	6,25
16	13	1	1,5	1	1,5	2,25
14	9	-1	-2,5	1	2,5	6,25
12	10	-3	-1,5	9	4,5	2,25
9	8	-6	-3,5	36	21	12,25
15	13	0	1,5	0	0	2,25
16	12	1	0,5	1	0,5	0,25
14	13	-1	1,5	1	-1,5	2,25
16	10	1	-1,5	1	-1,5	2,25
18	8	3	-3,5	9	-10,5	12,25
180	138			72	28	63

Se aplica la fórmula:

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} = \frac{28}{\sqrt{(72)(63)}} = 0,416$$

Existe una correlación moderada

En Excel se calcula de la siguiente manera:

a) Se inserta la función COEF.DE.CORREL y pulsar en Aceptar.

The screenshot shows an Excel spreadsheet with two columns of data: Column A (X) and Column B (Y). The data points are: (18, 13), (17, 15), (15, 14), (16, 13), (14, 9), (12, 10), (9, 8), (15, 13), (16, 12), (14, 13), (16, 10), (18, 8). The formula bar shows the function 'COEF.DE.CORREL' is being entered. The 'Insertar función' dialog box is open, displaying a list of functions under the 'Estadísticas' category. 'COEF.DE.CORREL' is highlighted in the list. The description for this function reads: 'COEF.DE.CORREL(matriz1;matriz2) Devuelve el coeficiente de correlación de dos conjuntos de datos.'

b) En el cuadro de argumentos de la función, en el recuadro de la Matriz 1 seleccionar las celdas de X, y en el recuadro de la Matriz 2 seleccionar las celdas de Y.

The screenshot shows an Excel spreadsheet with the following data in columns A and B:

	A	B
1	X	Y
2	18	13
3	17	15
4	15	14
5	16	13
6	14	9
7	12	10
8	9	8
9	15	13
10	16	12
11	14	13
12	16	10
13	18	8

The 'Argumentos de función' dialog box displays the following information:

- Function: COEF.DE.CORREL
- Matriz1: A2:A13 = {18\17\15\16\14\12\9\15\16\14\16\18}
- Matriz2: B2:B13 = {13\15\14\13\9\10\8\13\12\13\10\8}
- Result: = 0,41573971
- Description: Devuelve el coeficiente de correlación de dos conjuntos de datos.
- Note: **Matriz2** es un segundo rango de celdas de valores. Los valores deben ser números, nombres, matrices o referencias que contengan números.
- Result of the formula: = 0,41573971
- Buttons: Ayuda sobre esta función, Aceptar, Cancelar

c) Pulsar en Aceptar.

	A	B	C	D	E
1	X	Y			
2	18	13			
3	17	15			
4	15	14			
5	16	13			
6	14	9			
7	12	10			
8	9	8			
9	15	13			
10	16	12			
11	14	13			
12	16	10			
13	18	8			
14					
15		0,41573971	=COEF.DE.CORREL(A2:A13;B2:B13)		

El diagrama de dispersión en Excel se realiza de la siguiente manera:

a) Seleccionar los datos e insertar diagrama de dispersión.

Correlación - Microsoft Excel

Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista Complementos

Tabla dinámica Tabla Imagen Imágenes prediseñadas Formas SmartArt Columna Línea Circular Barra Área Dispersión Otros gráficos Hipervínculo Cuadro Encabezado de texto pie de página

A2 fx 18

	A	B	C	D	E	F	G	J	K
1	X	Y							
2	18	13							
3	17	15							
4	15	14							
5	16	13							
6	14	9							
7	12	10							
8	9	8							
9	15	13							
10	16	12							
11	14	13							
12	16	10							
13	18	8							

Dispersión sólo con marcadores
Compara pares de valores.
Utilícelo cuando los valores no están en el orden del eje X o cuando representan medidas separadas.

b) En diagrama dispersión, escoger el primero.

Cambiar tipo de gráfico Guardar como plantilla Tipo
Cambiar entre filas y columnas Seleccionar datos Datos
Diseños de gráfico

3 Gráfico fx

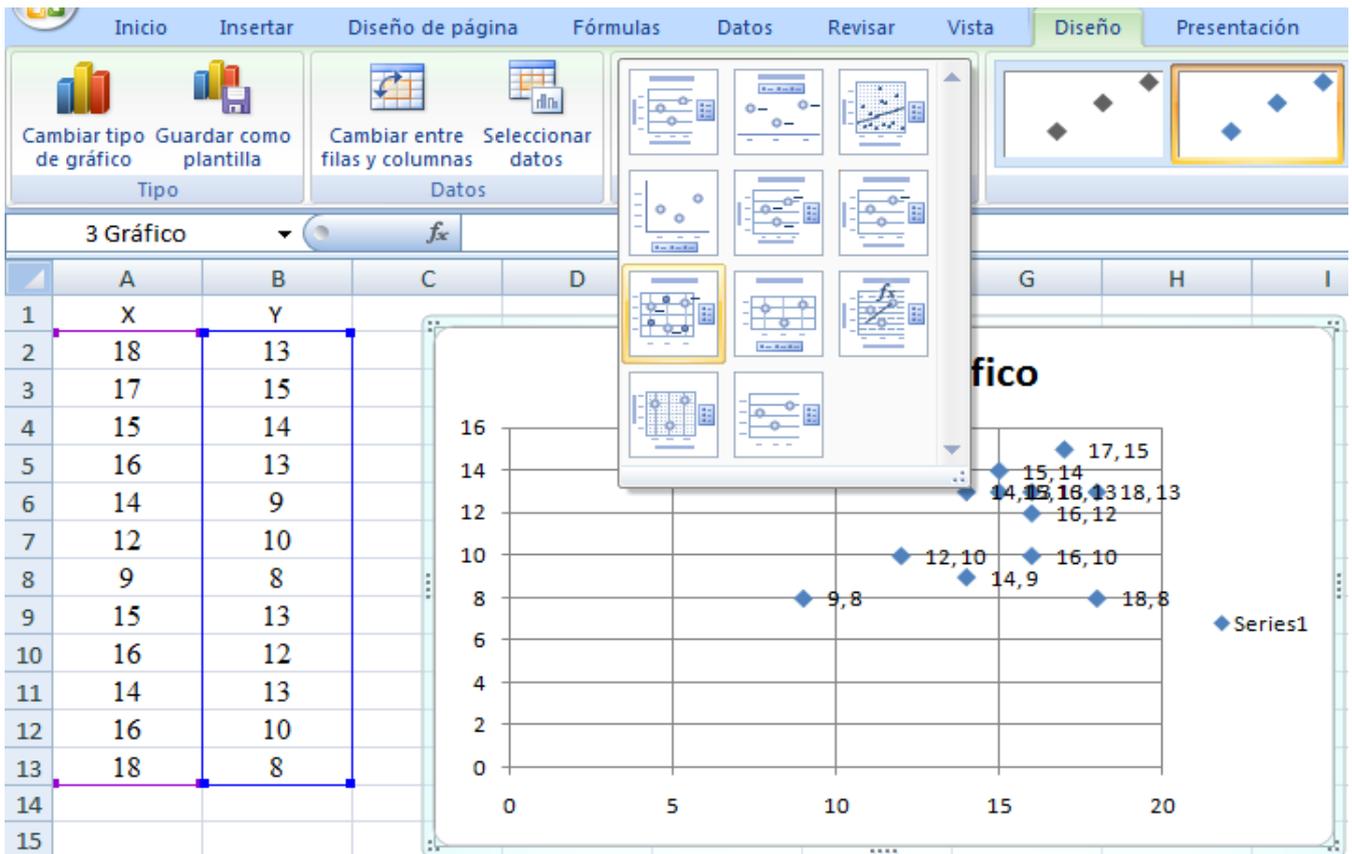
	A	B	C	D	E	F	G	H	I
1	X	Y							
2	18	13							
3	17	15							
4	15	14							
5	16	13							
6	14	9							
7	12	10							
8	9	8							
9	15	13							
10	16	12							
11	14	13							
12	16	10							
13	18	8							
14									
15									

16
14
12
10
8
6
4
2
0

0 5 10 15 20

◆ Series1

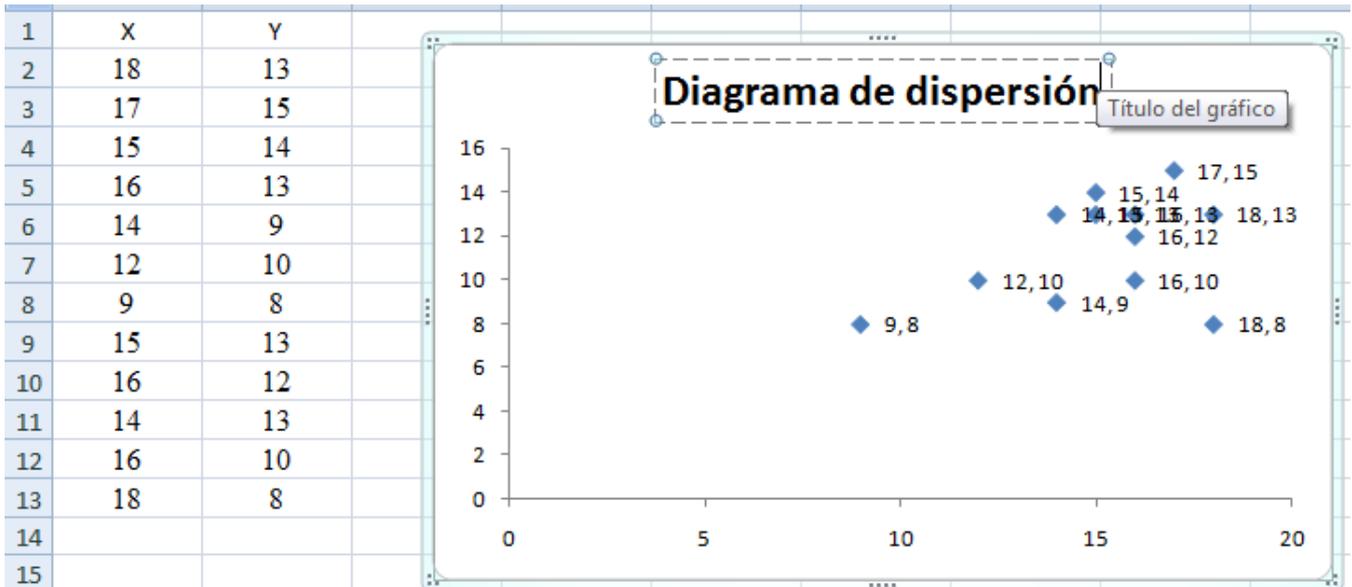
c) Para que ver las coordenadas escoger el diseño N° 7.



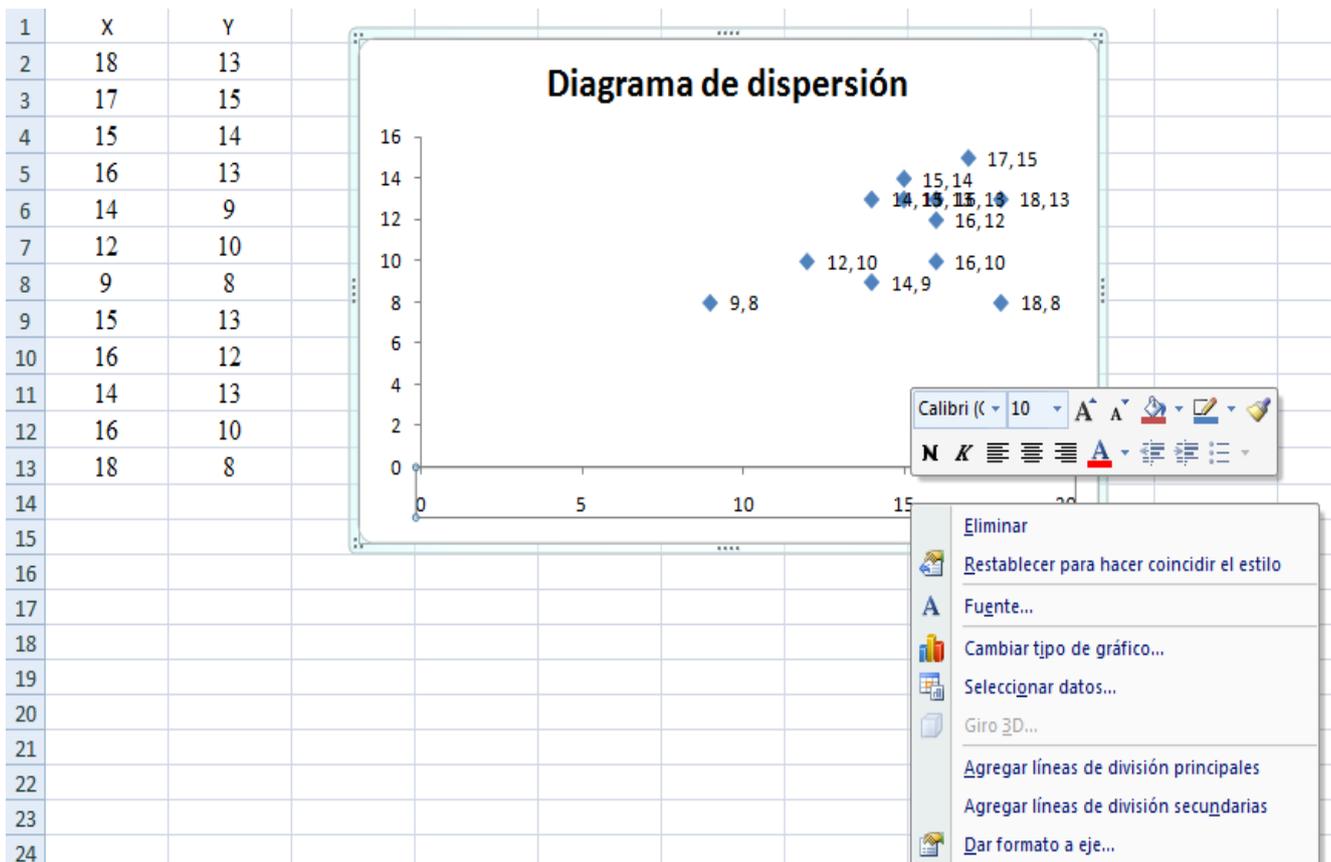
d) Borrar Serie 1, las líneas horizontales y verticales (haciendo clic y suprimir en cada objeto).



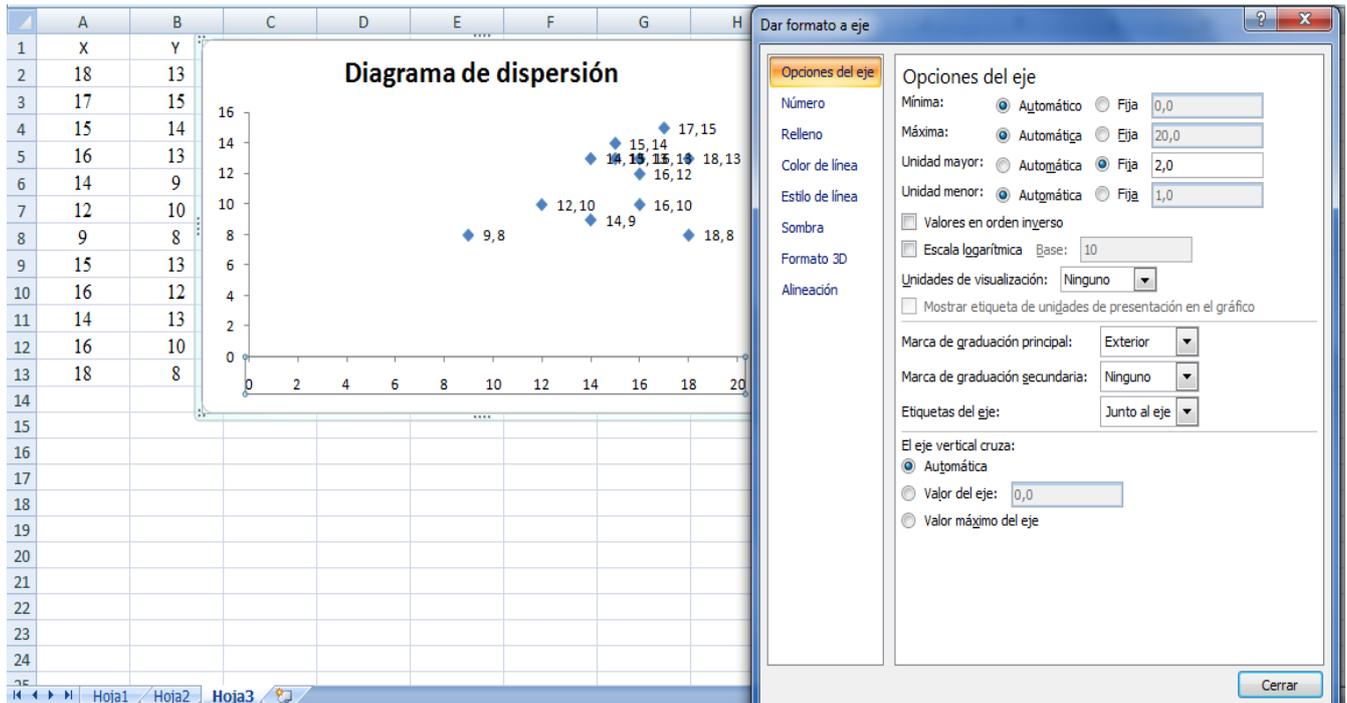
e) En título del gráfico escribir Diagrama de dispersión.



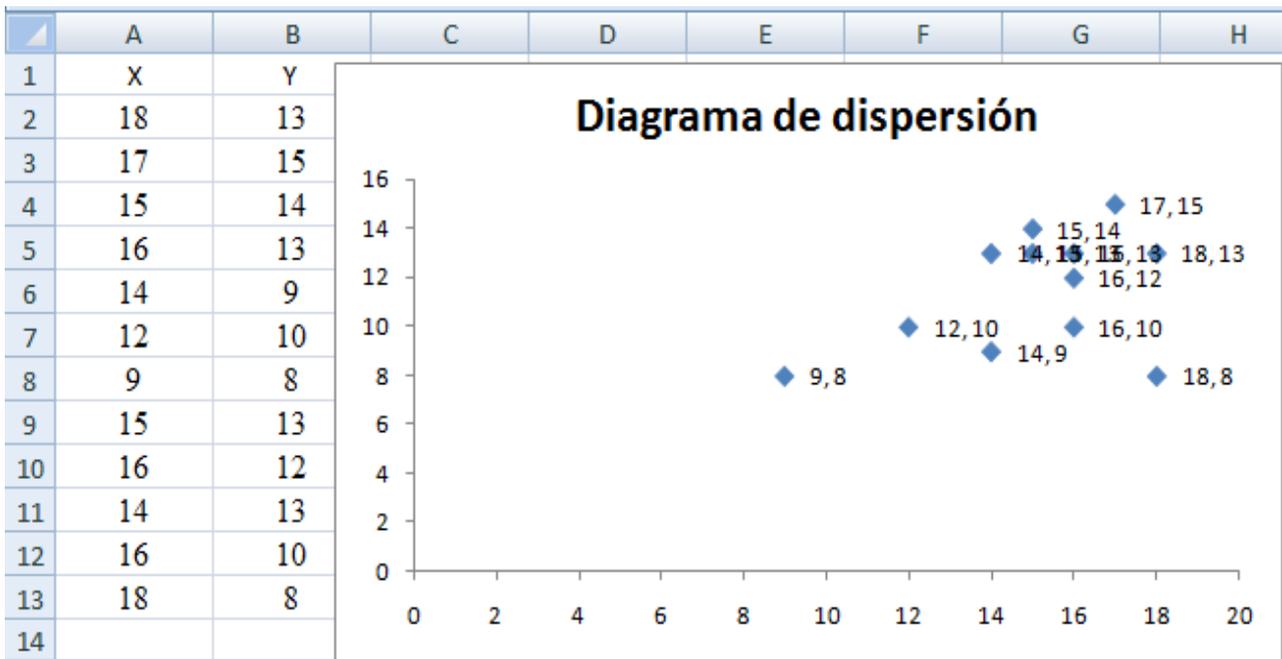
f) Clic en el eje x, y luego clic derecho para dar formato al eje.



g) Poner 2 en la casilla unidad mayor para ver los números de 2 en 2 en el eje x.

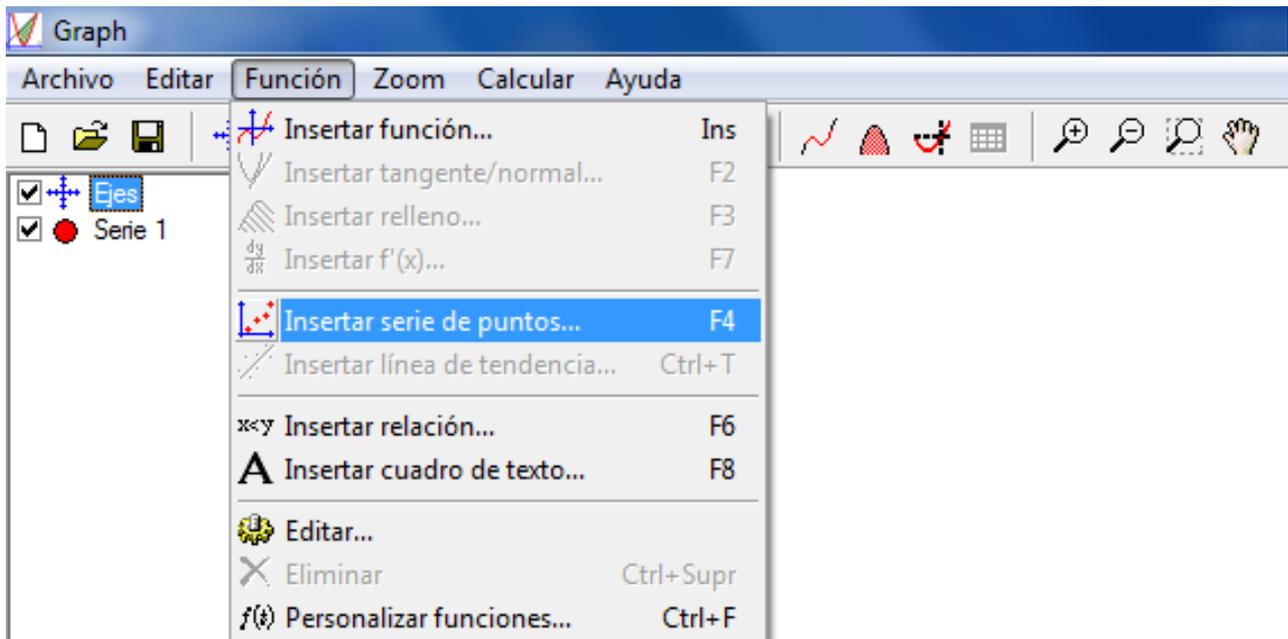


h) Clic en Cerrar para culminar la elaboración del diagrama de dispersión, aunque se le puede seguir haciendo más mejoras.

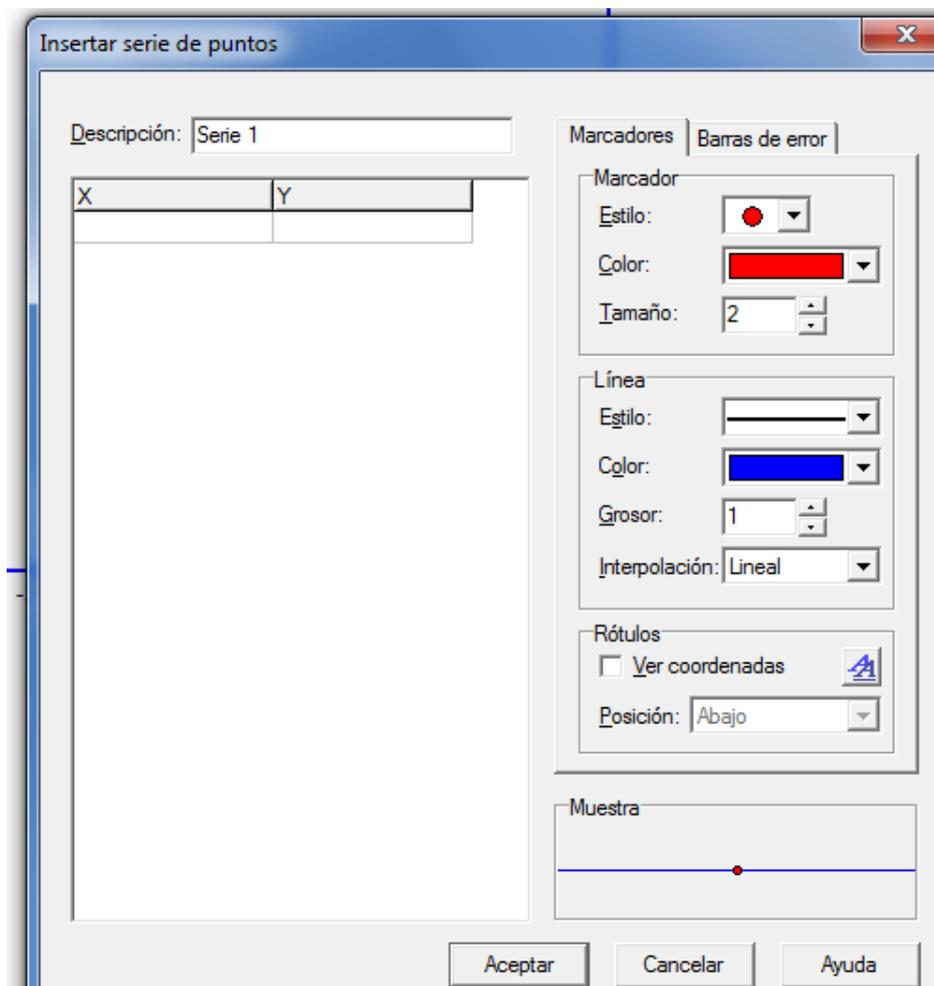


Para realizar el diagrama de dispersión en el programa Graph se procede de la siguiente manera:

a) Clic en Función.



b) Clic en Insertar serie de puntos.



c) Escribir los puntos, y en estilo de línea, escoger sin línea. En rótulos poner en ver coordenadas a la derecha. Pulsar en Aceptar.

Descripción: Serie 1

X	Y
18	13
17	15
15	14
16	13
14	9
12	10
9	8
15	13
16	12
14	13
16	10
18	8

Marcadores | Barras de error

Marcador
 Estilo: 
 Color: 
 Tamaño: 2

Línea
 Estilo: 
 Color: 
 Grosor: 1
 Interpolación: Lineal

Rótulos
 Ver coordenadas 
 Posición: Derecha

Muestra
 Derecha
 •(2.37,9.53)

Aceptar Cancelar Ayuda

d) Para editar los ejes, hacer clic en Editar y luego en Ejes.

Graph

Archivo Editar Función Zoom Calcular Ayuda

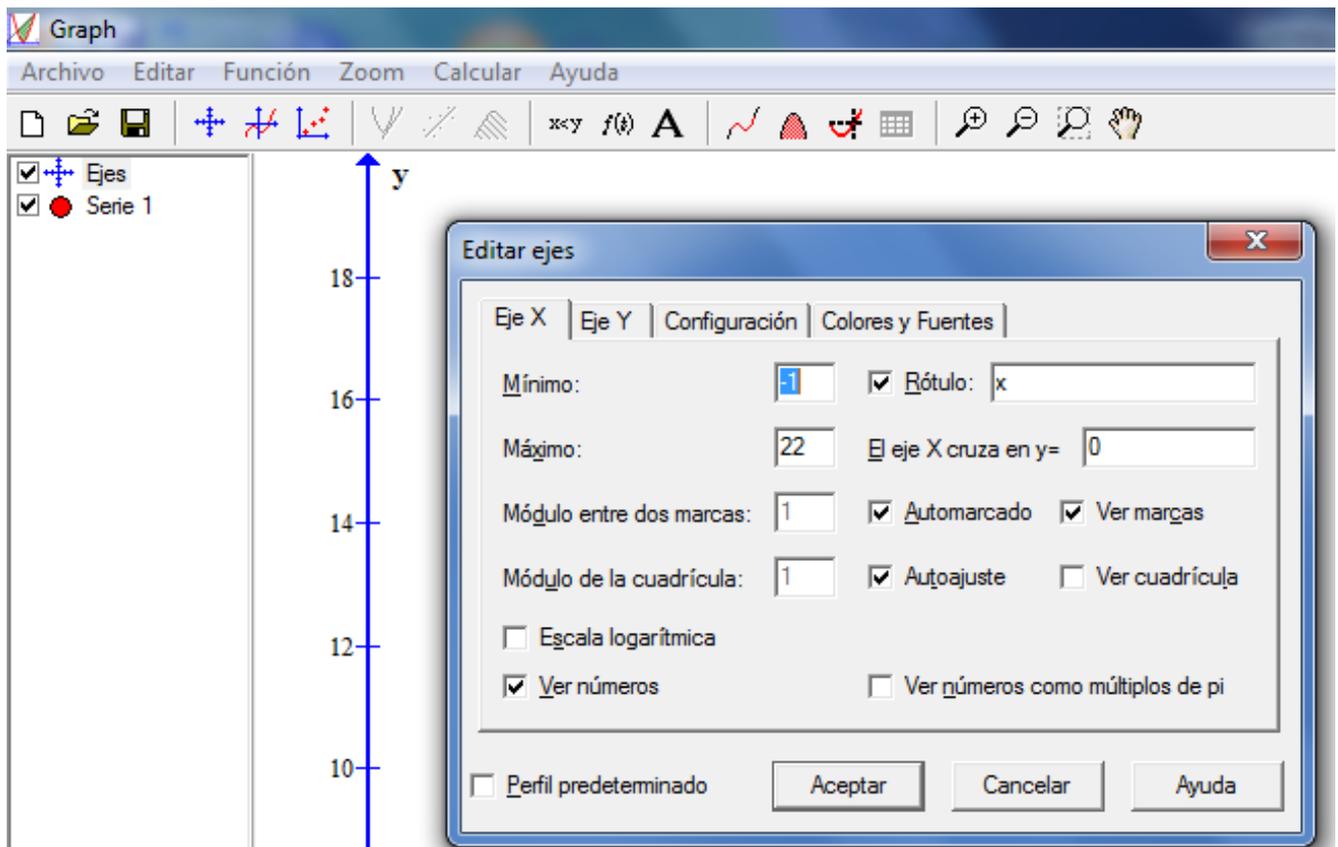
Eje
 Serie

Deshacer Ctrl+Z
 Rehacer Ctrl+Y
 Cortar
 Copiar
 Pegar
 Copiar imagen Ctrl+I
Ejes... Ctrl+A
 Opciones...

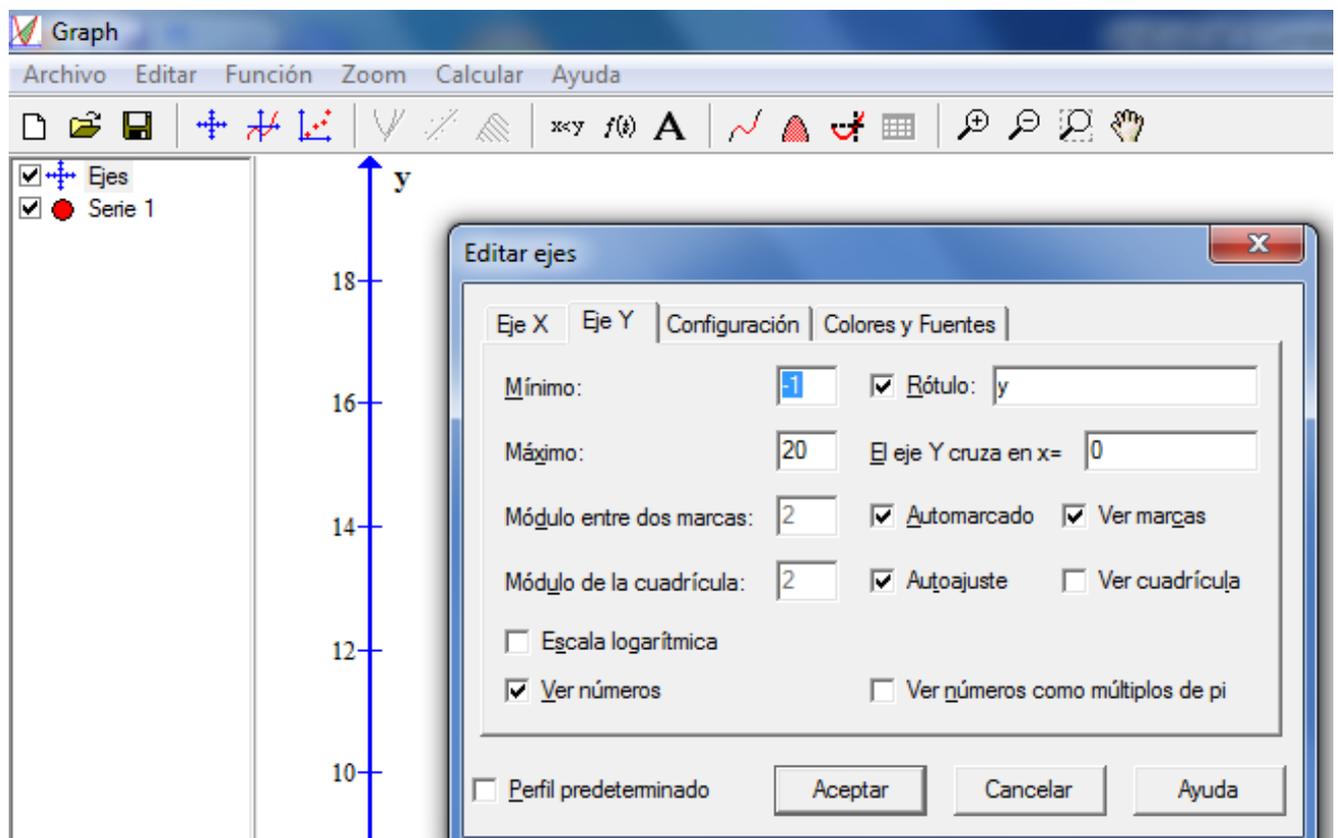
$x < y$ $f(x)$ A       

14

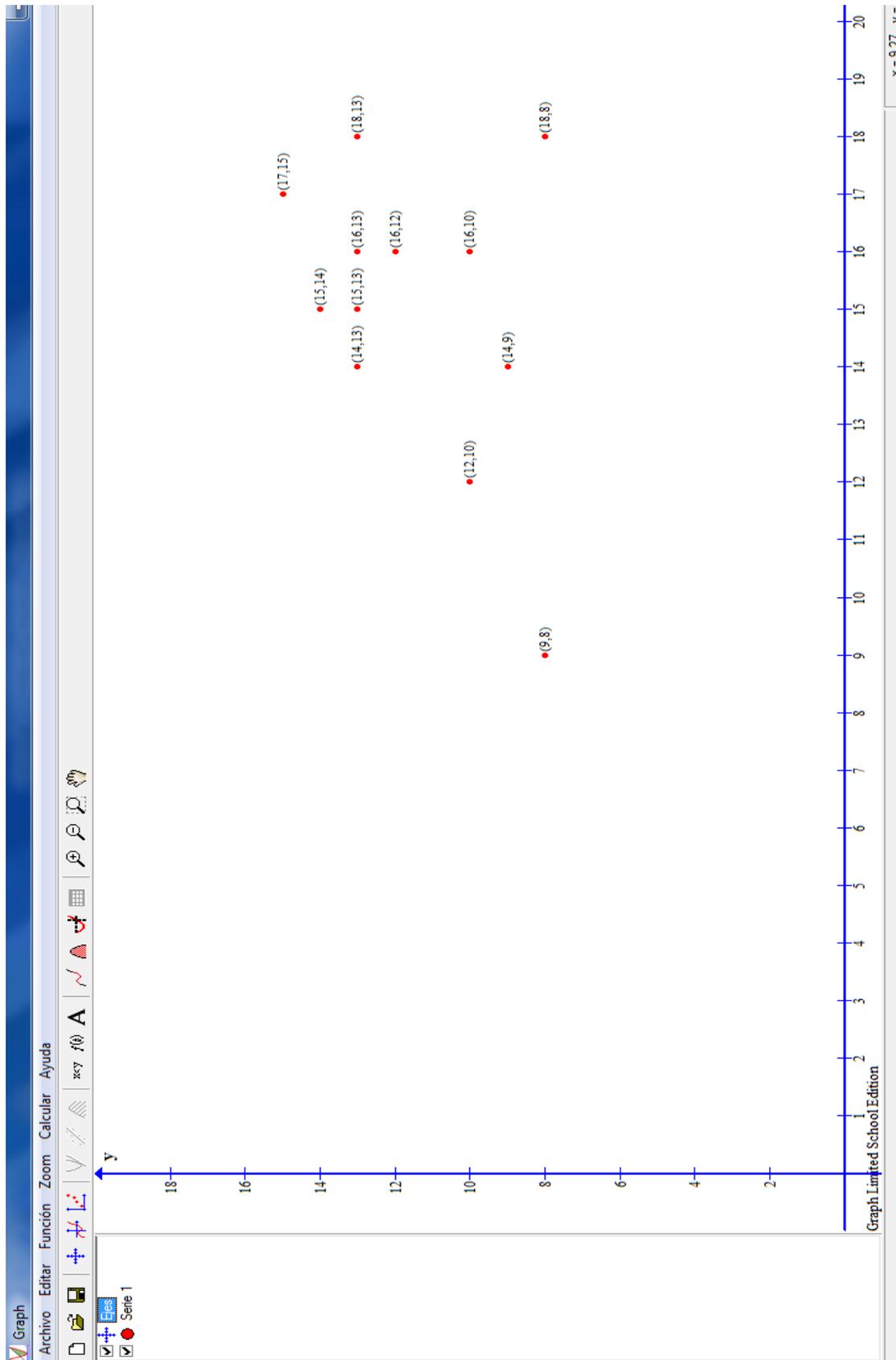
e) Llenar las casillas del Eje X de acuerdo a los datos del ejercicio.



f) Llenar las casillas del Eje Y de acuerdo a los datos del ejercicio.



g) Pulsar en Aceptar para dar por culminado la elaboración del diagrama de dispersión, el cual se presenta en la siguiente figura:



b) Para datos agrupados, el coeficiente de Correlación de Pearson se calcula aplicando la siguiente fórmula:

$$r = \frac{n \cdot \sum f \cdot dx \cdot dy - (\sum fx \cdot dx) (\sum fy \cdot dy)}{\sqrt{[n \cdot \sum fx \cdot dx^2 - (\sum fx \cdot dx)^2][n \cdot \sum fy \cdot dy^2 - (\sum fy \cdot dy)^2]}}$$

Donde:

n = número de datos.

f = frecuencia de celda.

fx = frecuencia de la variable X.

fy = frecuencia de la variable Y.

dx = valores codificados o cambiados para los intervalos de la variable X, procurando que al intervalo central le corresponda $dx = 0$, para que se hagan más fáciles los cálculos.

dy = valores codificados o cambiados para los intervalos de la variable X, procurando que al intervalo central le corresponda $dy = 0$, para que se hagan más fáciles los cálculos.

Ejemplo ilustrativo:

Con los siguientes datos sobre los Coeficientes Intellectuales (X) y de las calificaciones en una prueba de conocimiento (Y) de 50 estudiantes:

N° de estudiante	X	Y	N° de estudiante	X	Y
1	76	28	26	88	40
2	77	24	27	88	31
3	78	18	28	88	35
4	79	41	29	88	26
5	79	43	30	89	30
6	80	45	31	89	24
7	80	34	32	90	18
8	80	18	33	90	11
9	82	40	34	90	15
10	82	35	35	91	38
11	83	30	36	92	34
12	83	21	37	92	31
13	83	22	38	93	33
14	83	23	39	93	35
15	84	25	40	93	24
16	84	11	41	94	40
17	84	15	42	96	35
18	85	31	43	97	36
19	85	35	44	98	40
20	86	26	45	99	33
21	86	30	46	100	51
22	86	24	47	101	54
23	86	16	48	101	55
24	87	20	49	102	41
25	88	36	50	102	45

- 1) Elaborar una tabla de dos variables
- 2) Calcular el coeficiente de correlación

Solución:

1) En la *tabla de frecuencias de dos variables*, cada recuadro de esta tabla se llama una *celda* y corresponde a un par de intervalos, y el número indicado en cada celda se llama *frecuencia de celda*. Todos los totales indicados en la última fila y en la última columna se llaman *totales marginales o frecuencias marginales*, y corresponden, respectivamente, a las frecuencias de intervalo de las distribuciones de frecuencia separadas de la variable X y Y.

Para elaborar la tabla se recomienda:

- Agrupar las variables X y Y en un igual número de intervalos.
- Los intervalos de la variable X se ubican en la parte superior de manera horizontal (fila) y en orden ascendente.
- Los intervalos de la variable Y se ubican en la parte izquierda de manera vertical (columna) y en orden descendente.

Para elaborar los intervalos se procede a realizar los cálculos respectivos:

En la variable X:

Calculando el Rango se obtiene:

$$R = x_{\max} - x_{\min} = 102 - 76 = 26$$

Calculando el número de intervalos se obtiene:

$$n_i = 1 + 3,32 \cdot \log(n) = 1 + 3,32 \cdot \log 50 = 6$$

Calculando el ancho se obtiene:

$$i = \frac{R}{n_i} = \frac{26}{6} = 4,33$$

En la variable Y:

Calculando el Rango se obtiene:

$$R = y_{\max} - y_{\min} = 55 - 11 = 44$$

Calculando el número de intervalos se obtiene:

$$n_i = 1 + 3,32 \cdot \log(n) = 1 + 3,32 \cdot \log 50 = 6$$

Calculando el ancho se obtiene:

$$i = \frac{R}{n_i} = \frac{44}{6} = 7,33$$

Nota: Para la variable X se tomará un ancho de intervalo igual a 5 y para la variable Y un ancho de intervalo igual a 8 para obtener un número de intervalos igual a 6 para cada variable.

Contando las frecuencias de celda para cada par de intervalos de las variables X y Y se obtiene la siguiente tabla de frecuencias de dos variables:

		Coeficientes Intelectuales (X)					f_y	
		76-80	81-85	86-90	91-95	96-100		101-105
Calificaciones (Y)	51-58					1	2	3
	43-50	2					1	3
	35-42	1	3	3	3	3	1	14
	27-34	2	2	3	3	1		11
	19-26	1	4	5	1			11
	11-18	2	2	4				8
	f_x	8	11	15	7	5	4	50

Interpretación:

- El número 5 es la frecuencia de la celda correspondiente al par de intervalos 86-90 en Coeficiente Intelectual y 19-26 en Calificación obtenida en la prueba de conocimiento.
- El número 8 en la fila de f_x es el total marginal o frecuencia marginal del intervalo 76-80 en Coeficiente Intelectual.
- El número 14 en la columna de f_y es el total marginal o frecuencia marginal del intervalo 35-42 en Calificación obtenida en la prueba de conocimiento.
- El número 50 es total de frecuencias marginales y representa al número total de estudiantes.

2) Realizando los cálculos respectivos se obtiene la siguiente tabla:

		X	76-80	81-85	86-90	91-95	96-100	101-105				
Y	dx		-2	-1	0	1	2	3	f_y	$f_y \cdot dy$	$f_y \cdot dy^2$	$f \cdot dx \cdot dy$
	dy											
51-58	3						1	2	3	9	27	24
							6	18				
43-50	2		2					1	3	6	12	-2
			-8					6				
35-42	1		1	3	3	3	3	1	14	14	14	7
			-2	-3	0	3	6	3				
27-34	0		2	2	3	3	1		11	0	0	0
			0	0	0	0	0					
19-26	-1		1	4	5	1			11	-11	11	5
			2	4	0	-1						
11-18	-2		2	2	4				8	-16	32	12
			8	4	0							
f_x			8	11	15	7	5	4	50	2	96	46
$f_x \cdot dx$			-16	-11	0	7	10	12	2			
$f_x \cdot dx^2$			32	11	0	7	20	36	106			
$f \cdot dx \cdot dy$			0	5	0	2	12	27	46			

Nota:

Los números de las esquinas de cada celda en la anterior tabla representan el producto $f \cdot dx \cdot dy$, así por ejemplo, para obtener el número -8 de los intervalos 76-80 en X y 43-50 en Y se obtiene multiplicando $2 \cdot (-2) \cdot (2) = -8$. Para obtener el número 6 de los intervalos 96-100 en X y 51-58 en Y se obtiene multiplicando $1 \cdot 2 \cdot 3 = 6$.

Los números de la última columna (24, -2, 7, 0, 5 y 12) se obtienen sumando los números de las esquinas en cada fila, así por ejemplo, para obtener el número 24 se suma $6 + 18 = 24$.

Los números de la última fila (0, 5, 0, 2, 12 y 27) se obtienen sumando los números de las esquinas en cada columna, así por ejemplo, para obtener el número 27 se suma $18 + 6 + 3 = 27$.

Para obtener el número 2 de la antepenúltima columna se obtiene sumando los resultados de $fy \cdot dy$, es decir, representa la $\sum fy \cdot dy$

Para obtener el número 2 de la antepenúltima fila se obtiene sumando los resultados de $fx \cdot dx$, es decir, representa la $\sum fx \cdot dx$

Para obtener el número 96 de la penúltima columna se obtiene sumando los resultados de $fy \cdot dy^2$, es decir, representa $\sum fy \cdot dy^2$

Para obtener el número 106 de la penúltima fila se obtiene sumando los resultados de $fx \cdot dx^2$, es decir, representa $\sum fx \cdot dx^2$

Para obtener último número de la última columna se obtiene sumando los resultados de la última columna ($46=24-2+7+0+5+12$), es decir, representa $\sum fx \cdot dx \cdot dy$

Para obtener último número de la última fila se obtiene sumando los resultados de la última fila ($46=0+5+0+2+12+27$), y tiene que ser igual al último número de la última columna como comprobación que los cálculos de la tabla han sido correctos.

Observando los datos en la tabla anterior se reemplaza los valores en la ecuación del Coeficiente de Correlación de Pearson para datos agrupados, obteniéndose:

$$r = \frac{n \cdot \sum f \cdot dx \cdot dy - (\sum fx \cdot dx) (\sum fy \cdot dy)}{\sqrt{[n \cdot \sum fx \cdot dx^2 - (\sum fx \cdot dx)^2][n \cdot \sum fy \cdot dy^2 - (\sum fy \cdot dy)^2]}}$$

$$r = \frac{50 \cdot 46 - (2)(2)}{\sqrt{[50 \cdot 106 - (2)^2][50 \cdot 96 - (2)^2]}} = \frac{2300 - 4}{\sqrt{[5300 - 4][4800 - 4]}} = \frac{2296}{\sqrt{[5296][4796]}}$$

$$r = \frac{2296}{\sqrt{25399616}} = \frac{2296}{5039,803} = 0,456$$

Existe una correlación positiva moderada

TAREA DE INTERAPRENDIZAJE N° 17

1) Elabore un organizador gráfico de los tipos de correlación.

2) Con los datos de la siguiente tabla sobre las temperaturas del día X y del día Y en determinadas horas en una ciudad

X	9	10	12	14	16	18	20	22	24	26	28	30
Y	12	14	15	16	17	20	22	23	26	28	31	32

2.1) Calcule el coeficiente de correlación de Pearson empleando la fórmula y mediante Excel.

0,99

2.2) Elabore el diagrama de dispersión de manera manual.

2.3) Elabore el diagrama de dispersión empleando Excel.

2.4) Elabore el diagrama de dispersión empleando el programa Graph.

3) Cree y resuelva un ejercicio similar al anterior.

4) Dada la siguiente tabla de frecuencias de dos variables, con los datos sobre las calificaciones obtenidos en un curso de 50 estudiantes en la asignatura de Matemática (X) y en la asignatura de Estadística (Y), determinar el tipo de correlación que existe entre ellas mediante el coeficiente de Pearson.

		X					fy
		1-2	3-4	5-6	7-8	9-10	
Y	9-10				7	8	15
	7-8				6		6
	5-6			3	4		7
	3-4	5	5	1			11
	1-2	7	4				11
	fx	12	9	4	17	8	50

Correlación positiva muy alta de 0,91

5) Dada la siguiente tabla de frecuencias de dos variables, con los datos sobre los pesos en kilogramos en dos barrios diferentes en una ciudad, determinar el tipo de correlación que existe entre ellas mediante el coeficiente de Pearson.

		X					fy	
		40-49	50-59	60-69	70-79	80-89		90-99
Y	90-99				3	3	4	10
	80-89			8	2	2	4	16
	70-79			5	10	8	1	24
	60-69	8	1	2	5	2		18
	50-59	3	10	6	2			21
	40-49	4	6	1				11
	fx	15	17	22	22	15	9	100

Correlación positiva moderada de 0,688

6) Dada la siguiente tabla de frecuencias de dos variables, con los datos sobre las calificaciones obtenidos en un curso de 100 estudiantes en la asignatura de Matemática (X) y en la asignatura de

Estadística (Y), determinar el tipo de correlación que existe entre ellas mediante el coeficiente de Pearson para datos agrupados.

N° de estudiante	X	Y									
1	40	60	26	57	73	51	71	86	76	84	83
2	41	50	27	58	78	52	72	88	77	84	84
3	42	55	28	60	79	53	72	89	78	85	86
4	43	59	29	61	60	54	72	70	79	86	88
5	44	40	30	62	61	55	73	71	80	86	89
6	45	42	31	63	62	56	74	72	81	86	70
7	45	49	32	64	63	57	74	73	82	87	78
8	45	60	33	64	64	58	74	74	83	87	79
9	45	62	34	65	65	59	75	75	84	88	78
10	48	66	35	65	66	60	76	76	85	88	77
11	49	69	36	66	67	61	76	77	86	88	79
12	50	50	37	66	69	62	77	78	87	88	78
13	50	52	38	66	50	63	77	79	88	89	78
14	56	54	39	66	52	64	78	60	89	89	60
15	56	56	40	67	55	65	78	67	90	89	69
16	56	59	41	68	56	66	78	65	91	90	90
17	56	59	42	68	57	67	78	68	92	91	96
18	56	40	43	68	59	68	79	69	93	92	97
19	57	45	44	69	40	69	79	50	94	93	99
20	57	47	45	69	45	70	79	59	95	94	80
21	57	48	46	69	47	71	80	90	96	95	81
22	57	49	47	69	49	72	81	94	97	96	82
23	57	80	48	70	90	73	82	96	98	97	83
24	57	70	49	70	99	74	82	99	99	98	89
25	57	72	50	70	80	75	83	80	100	99	70

	X	40-48						94-102	
Y	$\begin{matrix} dx \\ dy \end{matrix}$	-3	-2	-1	0	1	1	3	f_y
94-102	3								7
	2								
	1								
	0								
	-1	5							
	-2	3							
40-48	-3	2							9
	f_x	10							100

Correlación positiva moderada de 0,62

7) Crear y resolver un ejercicio similar al anterior sobre cualquier tema de su preferencia.

5.1.3.2) COEFICIENTE DE CORRELACIÓN POR RANGOS DE SPEARMAN

Este coeficiente se emplea cuando una o ambas escalas de medidas de las variables son ordinales, es decir, cuando una o ambas escalas de medida son posiciones. Ejemplo: Orden de llegada en una carrera y peso de los atletas.

Se calcula aplicando la siguiente ecuación:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

r_s = Coeficiente de correlación por rangos de Spearman

d = Diferencia entre los rangos (X menos Y)

n = Número de datos

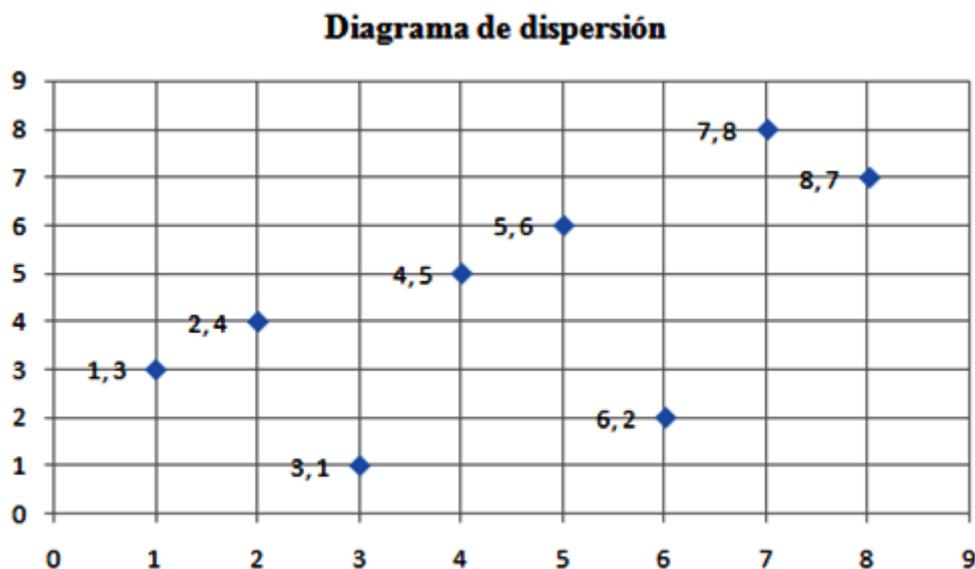
Nota: Los datos hay que traducirlos u ordenarlos en rangos. A los puntajes más elevados le asignamos el rango 1 al siguiente el rango 2 y así sucesivamente. Si se repiten dos puntajes o más se calculan las medias aritméticas.

Ejemplo ilustrativo N° 1: La siguiente tabla muestra el rango u orden obtenido en la primera evaluación (X) y el rango o puesto obtenido en la segunda evaluación (Y) de 8 estudiantes universitarios en la asignatura de Estadística. Realizar el diagrama de dispersión y calcular el coeficiente de correlación por rangos de Spearman.

Estudiante	X	Y
Dyanita	1	3
Elizabeth	2	4
Mario	3	1
Orlando	4	5
Mathías	5	6
Josué	6	2
Anita	7	8
Lucía	8	7

Solución:

El diagrama de dispersión hecho en Excel se muestra en la siguiente figura:



Para calcular el coeficiente de correlación por rangos de Spearman se llena la siguiente tabla:

Estudiante	X	Y	$d = X - Y$	$d^2 = (X - Y)^2$
Dyanita	1	3	-2	4
Elizabeth	2	4	-2	4
Mario	3	1	2	4
Orlando	4	5	-1	1
Mathías	5	6	-1	1
Josué	6	2	4	16
Anita	7	8	-1	1
Lucía	8	7	1	1
				$\Sigma d^2 = 32$

Se aplica la fórmula:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 32}{8(8^2 - 1)} = 1 - \frac{192}{504} = \frac{504 - 192}{504} = \frac{312}{504} = 0,619$$

Por lo tanto existe una correlación positiva moderada entre la primera y segunda evaluación de los 8 estudiantes.

En Excel se calcula de la siguiente manera:

a) Se inserta la función COEF.DE.CORREL y pulsar en Aceptar.

The image shows an Excel spreadsheet with the following data:

	A	B	C
1	X	Y	
2	1	3	
3	2	4	
4	3	1	
5	4	5	
6	5	6	
7	6	2	
8	7	8	
9	8	7	
10			
11	Coeficiente de correlación	=	

The 'Insertar función' dialog box is open, showing the 'COEF.DE.CORREL' function selected in the list. The description of the function is: 'Devuelve el coeficiente de correlación de dos conjuntos de datos.'

b) En el cuadro de argumentos de la función, en el recuadro de la Matriz 1 seleccionar las celdas de X, y en el recuadro de la Matriz 2 seleccionar las celdas de Y.

The screenshot shows an Excel spreadsheet with columns A and B. Column A contains values 1 through 8, and column B contains values 3, 4, 1, 5, 6, 2, 8, 7. A dialog box titled 'Argumentos de función' is open, showing the function 'COEF.DE.CORREL' with arguments 'Matriz1: A2:A9' and 'Matriz2: B2:B9'. The result is displayed as 'Resultado de la fórmula = 0,619047619'. The dialog box also includes a description of the function and buttons for 'Aceptar' and 'Cancelar'.

c) Pulsar en Aceptar.

	A	B	C	D	E	F
1	X	Y				
2	1	3				
3	2	4				
4	3	1				
5	4	5				
6	5	6				
7	6	2				
8	7	8				
9	8	7				
10						
11	Coeficiente de correlación		0,6190476	=COEF.DE.CORREL(A2:A9;B2:B9)		

Ejemplo ilustrativo N° 2

La siguiente tabla muestra las calificaciones de 8 estudiantes universitarios en las asignaturas de Matemática y Estadística. Calcular el coeficiente de correlación por rangos de Spearman y realizar el diagrama de dispersión.

N°	Estudiante	Matemática	Estadística
1	Dyana	10	8
2	Elizabeth	9	6
3	Mario	8	10
4	Orlando	7	9
5	Mathías	7	8
6	Josué	6	7
7	Anita	6	6
8	Lucía	4	9

Solución:

Para calcular el coeficiente de correlación por rangos de Spearman se procede a clasificar u ordenar los datos en rangos (X para Matemática y Y para Estadística) tomando en cuenta las siguientes observaciones:

En la asignatura de Matemática se observa:

- Dyana tiene la más alta calificación, ocupando el primer puesto, por lo que su rango es 1
- Elizabeth ocupa el segundo puesto, por lo que su rango es 2
- Mario se encuentra ubicado en el tercer lugar, por lo que su rango es 3
- Orlando y Mathías ocupan el cuarto y quinto puesto, por lo que su rango es la media aritmética de 4 y 5 que da por resultado 4,5
- Josué y Anita ocupan el sexto y séptimo lugar, por lo que su rango es la media aritmética de 6 y 7 que da por resultado 6,5
- Lucía se encuentra ubicada en el octavo lugar, por lo que su rango es 8

En la asignatura de Estadística se observa:

- Mario tiene la más alta calificación, ocupando el primer puesto, por lo que su rango es 1
- Orlando y Lucía ocupan el segundo y tercer puesto, por lo que su rango es la media aritmética de 2 y 3 que da por resultado 2,5
- Dyana y Mathías ocupan el cuarto y quinto puesto, por lo que su rango es la media aritmética de 4 y 5 que da por resultado 4,5
- Josué se encuentra ubicado en el sexto lugar, por lo que su rango es 6
- Elizabeth y Anita ocupan el séptimo y octavo lugar, por lo que su rango es la media aritmética de 7 y 8 que da por resultado 7,5

Los rangos X y Y se presentan en la siguiente tabla:

N°	Estudiante	Matemática	Estadística	X	Y
1	Dyana	10	8	1	4,5
2	Elizabeth	9	6	2	7,5
3	Mario	8	10	3	1
4	Orlando	7	9	4,5	2,5
5	Mathías	7	8	4,5	4,5
6	Josué	6	7	6,5	6
7	Anita	6	6	6,5	7,5
8	Lucía	4	9	8	2,5

Calculando d , d^2 y Σd^2 se obtiene los siguientes resultados:

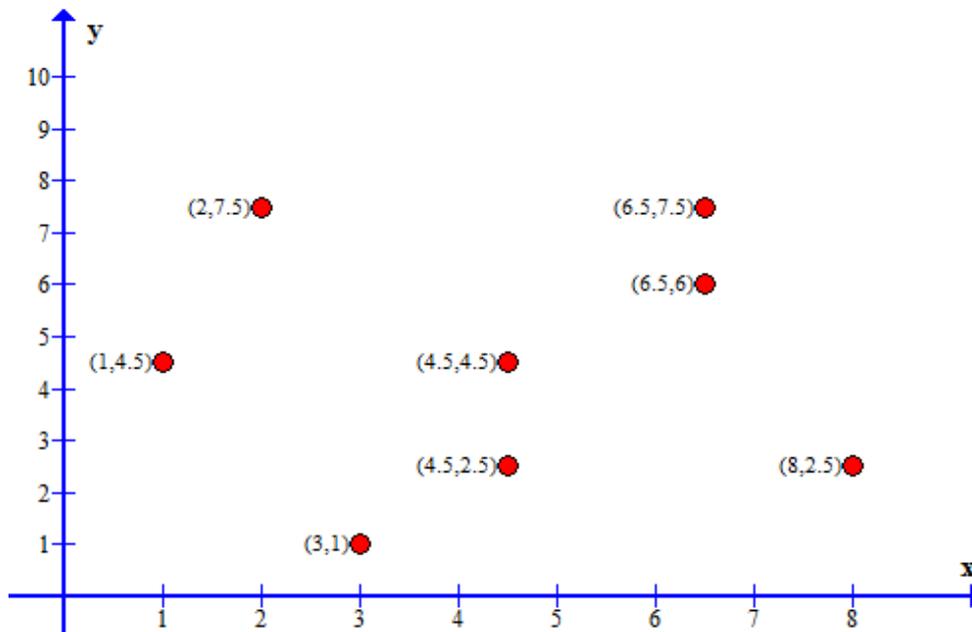
N°	Estudiante	Matemática	Estadística	X	Y	$d = X - Y$	$d^2 = (X - Y)^2$
1	Dyana	10	8	1	4,5	-3,5	12,25
2	Elizabeth	9	6	2	7,5	-5,5	30,25
3	Mario	8	10	3	1	2	4
4	Orlando	7	9	4,5	2,5	2	4
5	Mathías	7	8	4,5	4,5	0	0
6	Josué	6	7	6,5	6	0,5	0,25
7	Anita	6	6	6,5	7,5	-1	1
8	Lucía	4	9	8	2,5	5,5	30,25
							$\Sigma d^2 = 82$

Aplicando la fórmula se obtiene:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 82}{8(8^2 - 1)} = 1 - \frac{492}{504} = \frac{504 - 492}{504} = \frac{12}{504} = 0,024$$

Por lo tanto existe una correlación positiva muy baja

El diagrama de dispersión hecho en Graph se muestra en la siguiente figura:



TAREA DE INTERAPRENDIZAJE N° 18

- 1) Consulte sobre la biografía de Spearman y realice un organizador gráfico de la misma.
- 2) La siguiente tabla muestra el rango u orden obtenido en la primera evaluación (X) y el rango o puesto obtenido en la segunda evaluación (Y) de 8 estudiantes universitarios en la asignatura de Matemática.

X	Y
1	4
2	5
3	6
4	8
5	3
6	2
7	1
8	7

- 2.1) Realice el diagrama de dispersión en forma manual.
- 2.2) Realice el diagrama de dispersión empleando Excel.

2.3) Realice el diagrama de dispersión empleando el programa Graph.

2.4) Calcule el coeficiente de correlación por rangos de Spearman empleando la ecuación.

-0,19

2.5) Calcule el coeficiente de correlación empleando Excel.

-0,1905

3) Cree y resuelva un ejercicio similar al anterior.

4) La siguiente tabla muestra las calificaciones de 18 estudiantes universitarios en las asignaturas de Matemática y Estadística.

N°	Estudiante	Matemática	Estadística
1	Dyanita	10	3,5
2	Elizabeth	9	1
3	Mario	8	6
4	Orlando	8	8
5	Mathías	7	7
6	Benjamín	6	10
7	Segundo	6	4
8	Bertha	6	3,5
9	Alberto	5	1
10	Victoria	4	3
11	Sandra	4	9
12	Ximena	3	5
13	Darío	3	2,5
14	Santiago	2	0,7
15	José	1	2
16	Tomás	0,7	1,5
17	Paola	0,5	2,5
18	Kevin	0,5	0,5

4.1) Realice el diagrama de dispersión en forma manual.

4.2) Realice el diagrama de dispersión empleando Excel.

4.3) Realice el diagrama de dispersión empleando el programa Graph.

4.4) Calcule el coeficiente de correlación por rangos de Spearman empleando la ecuación.

0,49

4.5) Calcule el coeficiente de correlación empleando Excel.

0,49

5) Cree y resuelva un ejercicio similar al anterior.

5.1.4) COEFICIENTE DE DETERMINACIÓN

Revela qué porcentaje del cambio en Y se explica por un cambio en X. Se calcula elevando al cuadrado el coeficiente de correlación.

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

$$x = X - \bar{X}$$

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

r = Coeficiente de correlación de Pearson

r² = Coeficiente de determinación

La ecuación del coeficiente producto-momento (Coeficiente de Pearson) $r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$ puede escribirse en la forma equivalente:

$$\text{Coeficiente de Pearson} = r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

De donde coeficiente de determinación = r² = (Coeficiente de Pearson)²

Ejemplo ilustrativo

Con los datos de la siguiente tabla sobre las temperaturas, calcular el coeficiente de determinación empleando la ecuación obtenida de la forma equivalente del coeficiente de Pearson.

X	18	17	15	16	14	12	9	15	16	14	16	18
Y	13	15	14	13	9	10	8	13	12	13	10	8

Solución:

Se calcula el coeficiente de Pearson llenando la siguiente tabla:

X	Y	XY	X ²	Y ²
18	13	234	324	169
17	15	255	289	225
15	14	210	225	196
16	13	208	256	169
14	9	126	196	81
12	10	120	144	100
9	8	72	81	64
15	13	195	225	169
16	12	192	256	144
14	13	182	196	169
16	10	160	256	100
18	8	144	324	64
ΣX = 180	ΣY = 138	ΣXY = 2098	ΣX² = 2772	ΣY² = 1650

Se aplica la ecuación para calcular el coeficiente de Pearson.

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} = \frac{12 \cdot 2098 - 180 \cdot 138}{\sqrt{[12 \cdot 2772 - (180)^2][12 \cdot 1650 - (138)^2]}}$$

$$r = \frac{25176 - 24840}{\sqrt{[33264 - 32400][19800 - 19044]}} = \frac{336}{\sqrt{[864][756]}} = \frac{336}{\sqrt{653184}} = \frac{336}{808,198} = 0,4157$$

Elevando al cuadrado coeficiente de Pearson queda calculado el coeficiente de determinación.

$$\text{Coeficiente de determinación} = r^2 = (0,4157)^2 = 0,1728$$

Esto establece que 17,28% del cambio en Y se explica mediante un cambio en X.

Nota:

El r^2 tiene significado sólo para las relaciones lineales. Dos variables pueden tener $r^2 = 0$ y sin embargo estar relacionadas en sentido curvilíneo. El valor de r^2 no se interpreta como si la variable Y fuera causado por un cambio de la variable X, ya que la correlación no significa causa.

En Excel se calcula elevando al cuadrado el coeficiente de correlación o insertando la función =COEFICIENTE.R2 como muestra la siguiente figura:

COEFICIENTE.R2						
	A	B	C	D	E	
1	X	Y				
2	18	13				
3	17	15				
4	15	14				
5	16	13				
6	14	9				
7	12	10				
8	9	8				
9	15	13				
10	16	12				
11	14	13				
12	16	10				
13	18	8				
14						
15	r	0,4157397	=COEF.DE.CORREL(A2:A13;B2:B13)			
16	r ²	0,1728395	=B15^2			
17	r ²	0,1728395	=COEFICIENTE.R2(B2:B13;A2:A13)			

TAREA DE INTERAPRENDIZAJE N° 19

1) La siguiente tabla muestra el dinero en miles de dólares gastado en publicidad por una empresa (X) para vender sus productos, y el número en miles de clientes (Y) que compran los productos de la empresa.

X	15	17	14	13	18	20	17	18	16	14	20	18
Y	30	34	28	26	32	40	34	36	32	25	40	36

1.1) Realice el diagrama de dispersión en forma manual.

1.2) Realice el diagrama de dispersión empleando Excel.

1.3) Realice el diagrama de dispersión empleando el programa Graph.

1.4) Calcule el coeficiente de Pearson empleando las dos fórmulas.

0,96015

1.5) Calcule el coeficiente de determinación empleando las dos fórmulas y mediante Excel.

0,9219

2) La siguiente tabla muestra el tiempo en minutos dedicado al estudio y la calificación sobre 10 obtenida.

X	140	150	130	120	170	190	180	160	200	110	100	90
Y	7	8	7	6	8	10	9	8	10	6	5	4

2.1) Realice el diagrama de dispersión en forma manual.

2.2) Realice el diagrama de dispersión empleando Excel.

2.3) Realice el diagrama de dispersión empleando el programa Graph.

2.4) Calcule el coeficiente de Pearson empleando las dos fórmulas.

0,9817

2.5) Calcule el coeficiente de determinación empleando las dos fórmulas y mediante Excel.

0,9638

3) Cree y resuelva un ejercicio similar a los anteriores.

5.2) ANÁLISIS DE REGRESIÓN

Los primeros y más importantes estudios al respecto se deben a los científicos Francis Galton (1822-1911) y Karl Pearson (1857-1936). Fue Galton quien utilizó por primera vez el término regresión para indicar que, aunque influida por la estatura de sus padres, la estatura de los hijos “regresaba” a la media general.

La regresión examina la relación entre dos variables, pero restringiendo una de ellas con el objeto de estudiar las variaciones de una variable cuando la otra permanece constante. En otras palabras, la regresión es un método que se emplea para predecir el valor de una variable en función de valores dados a la otra variable. En estadística la palabra predecir no se utiliza en el sentido empleado por los astrólogos, futurólogos y mentalistas, sino mas bien en un sentido lógico como es el de utilizar el conocimiento del comportamiento de una variable para obtener información sobre otra variable. Por ejemplo, puede predecirse el resultado que obtendrá un estudiante en su examen final, basados en 

conocimiento de las calificaciones promedio de sus exámenes parciales, o predecir la preferencia de los estudiantes por profesiones científicas, conociendo los promedios de sus calificaciones en los estudios escolares.

En todos los casos de regresión existe una dependencia funcional entre las variables. En el caso de dos variables, siendo una de ellas (X) variable independiente y la otra (Y) la dependiente, se habla de regresión de Y sobre X; Por ejemplo, los ingenieros forestales utilizan la regresión de la altura de los árboles sobre su diámetro, lo cual significa que midiendo el diámetro (variable independiente) y reemplazando su valor en una relación definida según la clase de árbol se obtiene la altura, y aun sin necesidad de cálculos aprecian la altura utilizando gráficas de la función de dependencia, altura = función del diámetro.

5.2.1) PRINCIPIO DE LOS MÍNIMOS CUADRADOS

5.2.1.1) LA RECTA DE LOS MÍNIMOS CUADRADOS

Se llama línea de mejor ajuste y se define como la línea que hace mínima la suma de los cuadrados de las desviaciones respecto a ella de todos los puntos que corresponden a la información recogida.

La recta de los mínimos cuadrados que aproxima el conjunto de puntos $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots, (X_N, Y_N)$ tomando en cuenta a Y como variable dependiente tiene por ecuación

$$Y = a_0 + a_1X$$

A esta ecuación suele llamarse recta de regresión de Y sobre X, y se usa para estimar los valores de Y para valores dados de X

Si a la recta de regresión $Y = a_0 + a_1X$ se le suma en ambos lados $\sum Y = \sum(a_0 + a_1X)$ se obtiene $\sum Y = a_0N + a_1 \sum X$

Si a la recta de regresión $Y = a_0 + a_1X$ se multiplica por X a ambos lados y luego se suma $\sum XY = \sum X(a_0 + a_1X)$ se obtiene $\sum XY = a_0 \sum X + a_1 \sum X^2$

Las constantes a_0 y a_1 quedan fijadas al resolver simultáneamente las ecuaciones anteriormente encontradas, es decir, al resolver el siguiente sistema de ecuaciones:

$$\begin{cases} \sum Y = a_0N + a_1\sum X \\ \sum XY = a_0\sum X + a_1\sum X^2 \end{cases}$$

Que se llaman las ecuaciones normales para la recta de mínimos cuadrados.

Las constantes a_0 y a_1 de las anteriores ecuaciones también se pueden calcular empleando las siguientes fórmulas:

$$a_0 = \frac{\sum Y \cdot \sum X^2 - \sum X \cdot \sum XY}{N \sum X^2 - (\sum X)^2} \qquad a_1 = \frac{N \sum XY - \sum X \cdot \sum Y}{N \sum X^2 - (\sum X)^2}$$

Otra ecuación para los mínimos cuadrados para $x = X - \bar{X}$, $y = Y - \bar{Y}$ de la recta de regresión de Y sobre X es:

$$y = \left(\frac{\sum xy}{\sum x^2} \right) x$$

La recta de los mínimos cuadrados que aproxima el conjunto de puntos $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots \dots (X_N, Y_N)$ tomando en cuenta a X como variable dependiente tiene por ecuación:

$$X = b_0 + b_1 Y$$

A esta ecuación suele llamarse recta de regresión de X sobre Y, y se usa para estimar los valores de X para valores dados de Y. Las constantes b_0 y b_1 quedan fijadas al resolver el siguiente sistema de ecuaciones:

$$\begin{cases} \Sigma X = b_0 N + b_1 \Sigma Y \\ \Sigma XY = b_0 \Sigma Y + b_1 \Sigma Y^2 \end{cases}$$

Las constantes b_0 y b_1 del sistema de ecuaciones anterior se pueden calcular empleando las siguientes fórmulas:

$$b_0 = \frac{\Sigma X \cdot \Sigma Y^2 - \Sigma Y \cdot \Sigma XY}{N \Sigma Y^2 - (\Sigma Y)^2} \qquad b_1 = \frac{N \Sigma XY - \Sigma X \cdot \Sigma Y}{N \Sigma Y^2 - (\Sigma Y)^2}$$

Otra ecuación para los mínimos cuadrados para $x = X - \bar{X}$, $y = Y - \bar{Y}$ es:

$$x = \left(\frac{\Sigma xy}{\Sigma y^2} \right) y$$

El punto de intersección entre las rectas $Y = a_0 + a_1 X$ con $X = b_0 + b_1 Y$ se simboliza (\bar{X}, \bar{Y}) y se llama centroide o centro de gravedad

Ejemplo ilustrativo

Con los datos de la siguiente tabla sobre la altura en centímetros (X) y los pesos en kilogramos (Y) de una muestra de 8 estudiantes varones tomada al azar del segundo semestre de una universidad.

X	152	157	162	167	173	178	182	188
Y	56	61	67	72	70	72	83	92

1) Ajustar la recta de mínimos cuadrados para Y como variable dependiente resolviendo el sistema:

$$\begin{cases} \Sigma Y = a_0 N + a_1 \Sigma X \\ \Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 \end{cases}$$

2) Ajustar la recta de mínimos cuadrados para Y como variable dependiente empleando las fórmulas:

$$a_0 = \frac{\Sigma Y \cdot \Sigma X^2 - \Sigma X \cdot \Sigma XY}{N \Sigma X^2 - (\Sigma X)^2} \qquad a_1 = \frac{N \Sigma XY - \Sigma X \cdot \Sigma Y}{N \Sigma X^2 - (\Sigma X)^2}$$

3) Ajustar la recta de mínimos cuadrados para Y como variable dependiente empleando la fórmula:

$$y = \left(\frac{\Sigma xy}{\Sigma x^2} \right) x$$

4) Ajustar la recta de mínimos cuadrados para X como variable dependiente resolviendo el sistema:

$$\begin{cases} \Sigma X = b_0 N + b_1 \Sigma Y \\ \Sigma XY = b_0 \Sigma Y + b_1 \Sigma Y^2 \end{cases}$$

5) Calcular el punto centroide.

6) Calcular el coeficiente de determinación.

7) Elaborar el diagrama de dispersión. Y en el mismo diagrama graficar las dos rectas de mínimos cuadrados obtenidas en los pasos anteriores.

8) Estimar el valor de Y cuando X = 200 en el diagrama de dispersión de Y como variable dependiente.
R: 8,2

9) Estimar el valor de X cuando Y= 100 en el diagrama de dispersión X como variable dependiente.

Solución:

Para comenzar a resolver el ejercicio se llena la siguiente tabla:

X	Y	XY	X ²	Y ²
152	56	8512	23104	3136
157	61	9577	24649	3721
162	67	10854	26244	4489
167	72	12024	27889	5184
173	70	12110	29929	4900
178	72	12816	31684	5184
182	83	15106	33124	6889
188	92	17296	35344	8464
$\Sigma X = 1359$	$\Sigma Y = 573$	$\Sigma XY = 98295$	$\Sigma X^2 = 231967$	$\Sigma Y^2 = 41967$

1) Reemplazando valores en el sistema se tiene:

$$\begin{cases} \Sigma Y = a_0 N + a_1 \Sigma X \\ \Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 \end{cases}$$

$$\begin{cases} 573 = a_0 \cdot 8 + a_1 \cdot 1359 \\ 98295 = a_0 \cdot 1359 + a_1 \cdot 231967 \end{cases} \Rightarrow \begin{cases} 8a_0 + 1359a_1 = 573 \\ 1359a_0 + 231967a_1 = 98295 \end{cases}$$

Resolviendo el sistema por determinantes (regla de Cramer) se obtiene:

$$a_0 = \frac{\Delta a_0}{\Delta} = \frac{\begin{vmatrix} 573 & 1359 \\ 98295 & 231967 \end{vmatrix}}{\begin{vmatrix} 8 & 1359 \\ 1359 & 231967 \end{vmatrix}} = \frac{573 \cdot 231967 - 98295 \cdot 1359}{8 \cdot 231967 - 1359 \cdot 1359} = \frac{-665814}{8855} = -75,191$$

$$a_1 = \frac{\Delta a_1}{\Delta} = \frac{\begin{vmatrix} 8 & 573 \\ 1359 & 98295 \end{vmatrix}}{8855} = \frac{8 \cdot 98295 - 1359 \cdot 573}{8855} = \frac{7653}{8855} = 0,864$$

Interpretación:

- El valor $a_1 = 0,864$ indica que la recta tiene una pendiente positiva aumentando a razón de 0,864
- El valor de $a_0 = -75,191$ indica el punto en donde la recta interseca al eje Y cuando $X = 0$

En Excel el sistema se resuelve de la siguiente manera:

a) Insertar la función MDETERM.

The screenshot shows an Excel spreadsheet with the following data:

1											
2	{	$8a_0 + 1359a_1 = 573$									
3	{	$1359a_0 + 231967a_1 = 98295$									
4											
5		573	1359								
6		98295	231967								
7											
8		Δa_0	$=MDETERM(B5:B7)$								
9											
10											
11											
12											

The dialog box 'Argumentos de función' for MDETERM shows:

- Matriz: B5:B7 = {1359;231967;0}
- Resultado de la fórmula = MDETERM(B5:B7)

b) Seleccionar las celdas del Δa_0 .

The screenshot shows an Excel spreadsheet with the following data:

1											
2	{	$8a_0 + 1359a_1 = 573$									
3	{	$1359a_0 + 231967a_1 = 98295$									
4											
5		573	1359								
6		98295	231967								
7											
8		Δa_0	$=MDETERM(A5:B6)$								
9											
10											
11											
12											

The dialog box 'Argumentos de función' for MDETERM shows:

- Matriz: A5:B6 = {573;1359;98295;231967}
- Resultado de la fórmula = -665814

c) Pulsar en Aceptar para culminar con el cálculo del Δa_0 .

	A	B	C	D
1				
2	$8a_0 + 1359a_1 = 573$			
3	$1359a_0 + 231967a_1 = 98295$			
4				
5	573	1359		
6	98295	231967		
7				
8	Δa_0	-665814	=MDETERM(A5:B6)	

d) Repetir los pasos anteriores para calcular el Δ y el Δa_1 , para luego calcular a_0 y a_1 como indica la siguiente figura:

	A	B	C	D	E	F	G	H
1								
2	$8a_0 + 1359a_1 = 573$							
3	$1359a_0 + 231967a_1 = 98295$							
4								
5	573	1359						
6	98295	231967						
7								
8	Δa_0	-665814	=MDETERM(A5:B6)					
9								
10						$a_0 = \frac{\Delta a_0}{\Delta}$	-75,19074	=B8/B14
11	8	1359						
12	1359	231967						
13								
14	Δ	8855	=MDETERM(A11:B12)			$a_1 = \frac{\Delta a_1}{\Delta}$	0,8642575	=B20/B14
15								
16								
17	8	573						
18	1359	98295						
19								
20	Δb_0	7653	=MDETERM(A17:B18)					

Reemplazando valores en la ecuación respectiva se obtiene:

$$Y = a_0 + a_1X \Rightarrow Y = -75,191 + 0,864X$$

2) Con los datos de la tabla anterior se substituye valores en las siguientes ecuaciones:

$$a_0 = \frac{\sum Y \cdot \sum X^2 - \sum X \cdot \sum XY}{N \sum X^2 - (\sum X)^2} = \frac{573 \cdot 231967 - 1359 \cdot 98295}{8 \cdot 231967 - (1359)^2} = \frac{-665814}{8855} = -75,191$$

$$a_1 = \frac{N \sum XY - \sum X \cdot \sum Y}{N \sum X^2 - (\sum X)^2} = \frac{8 \cdot 98295 - 1359 \cdot 573}{8 \cdot 231967 - (1359)^2} = \frac{7653}{8855} = 0,864$$

Reemplazando valores en la ecuación respectiva se obtiene:

$$Y = a_0 + a_1X \Rightarrow Y = -75,191 + 0,864X$$

3) Se calcula las medias aritméticas de X y Y para llenar la siguiente tabla:

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{X} = \frac{1359}{8} = 169,875$$

$$\bar{Y} = \frac{573}{8} = 71,625$$

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x^2	y^2
152	56	-17,88	-15,625	279,297	319,516	244,141
157	61	-12,88	-10,625	136,797	165,766	112,891
162	67	-7,875	-4,625	36,422	62,016	21,391
167	72	-2,875	0,375	-1,078	8,266	0,141
173	70	3,125	-1,625	-5,078	9,766	2,641
178	72	8,125	0,375	3,047	66,016	0,141
182	83	12,125	11,375	137,922	147,016	129,391
188	92	18,125	20,375	369,297	328,516	415,141
$\Sigma X = 1359$	$\Sigma Y = 573$			$\Sigma xy = 956,625$	$\Sigma x^2 = 1106,875$	$\Sigma y^2 = 925,875$

Reemplazando valores en la fórmula respectiva se obtiene:

$$y = \left(\frac{\sum xy}{\sum x^2} \right) x \Rightarrow y = \frac{956,625}{1106,875} x \Rightarrow Y - \bar{Y} = \frac{956,625}{1106,875} (X - \bar{X})$$

$$Y - 71,625 = \frac{956,625}{1106,875} (X - 169,875) \Rightarrow 1106,875(Y - 71,625) = 956,625(X - 169,875)$$

$$1106,875Y - 79280,20838 = 956,625X - 162510,4984$$

$$1106,875Y = 956,625X - 162510,4984 + 79280,20838$$

$$1106,875Y = 956,625X - 83230,29$$

$$Y = \frac{956,625X - 83230,29}{1106,875} \Rightarrow Y = \frac{956,625X}{1106,875} - \frac{83230,29}{1106,875} \Rightarrow Y = 0,864X - 75,19$$

$$Y = -75,19 + 0,864X$$

4) Reemplazando valores en sistema respectivo se obtiene:

$$\begin{cases} \Sigma X = b_0 N + b_1 \Sigma Y \\ \Sigma XY = b_0 \Sigma Y + b_1 \Sigma Y^2 \end{cases}$$

$$\begin{cases} 1359 = b_0 \cdot 8 + b_1 \cdot 573 \\ 98295 = b_0 \cdot 573 + b_1 \cdot 41967 \end{cases} \Rightarrow \begin{cases} 8b_0 + 573b_1 = 1359 \\ 573b_0 + 41967b_1 = 98295 \end{cases}$$

Resolviendo el sistema se obtiene:

$$b_0 = 95,871$$

$$b_1 = 1,033$$

Reemplazando valores en la ecuación de la recta de mínimos cuadrados se obtiene:

$$X = b_0 + b_1 Y$$

$$X = 95,871 + 1,033Y$$

Interpretación:

- El valor $b_1 = 1,033$ indica que la recta tiene una pendiente positiva aumentando a razón de 1,033

- El valor de $b_0 = 95,871$ indica el punto en donde la recta interseca al eje X cuanto $Y = 0$

5) Para calcular el centroide (\bar{X}, \bar{Y}) se resuelve el sistema formado por las dos rectas de los mínimos cuadrados en donde X es \bar{X} y Y es \bar{Y} .

$$\begin{cases} Y = -75,191 + 0,864X \\ X = 95,871 + 1,033Y \end{cases}$$

Al resolver el sistema se obtiene el centroide: $X = 169,3$ y $Y = 71,092$

6) Se aplica la ecuación para calcular el coeficiente de Pearson.

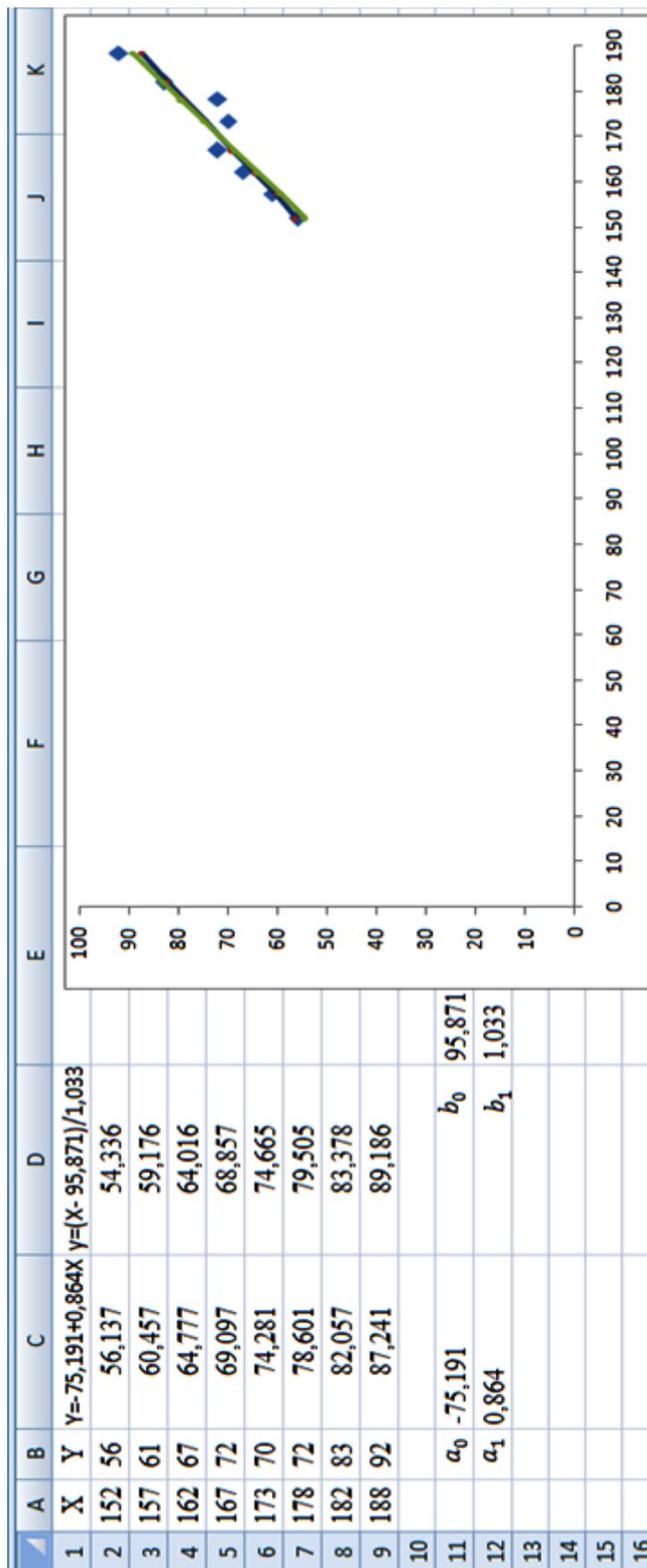
$$r = \frac{N \Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N \Sigma X^2 - (\Sigma X)^2][N \Sigma Y^2 - (\Sigma Y)^2]}} = \frac{8 \cdot 98295 - 1359 \cdot 573}{\sqrt{[8 \cdot 231967 - (1359)^2][8 \cdot 41967 - (573)^2]}}$$

$$r = 0,94497$$

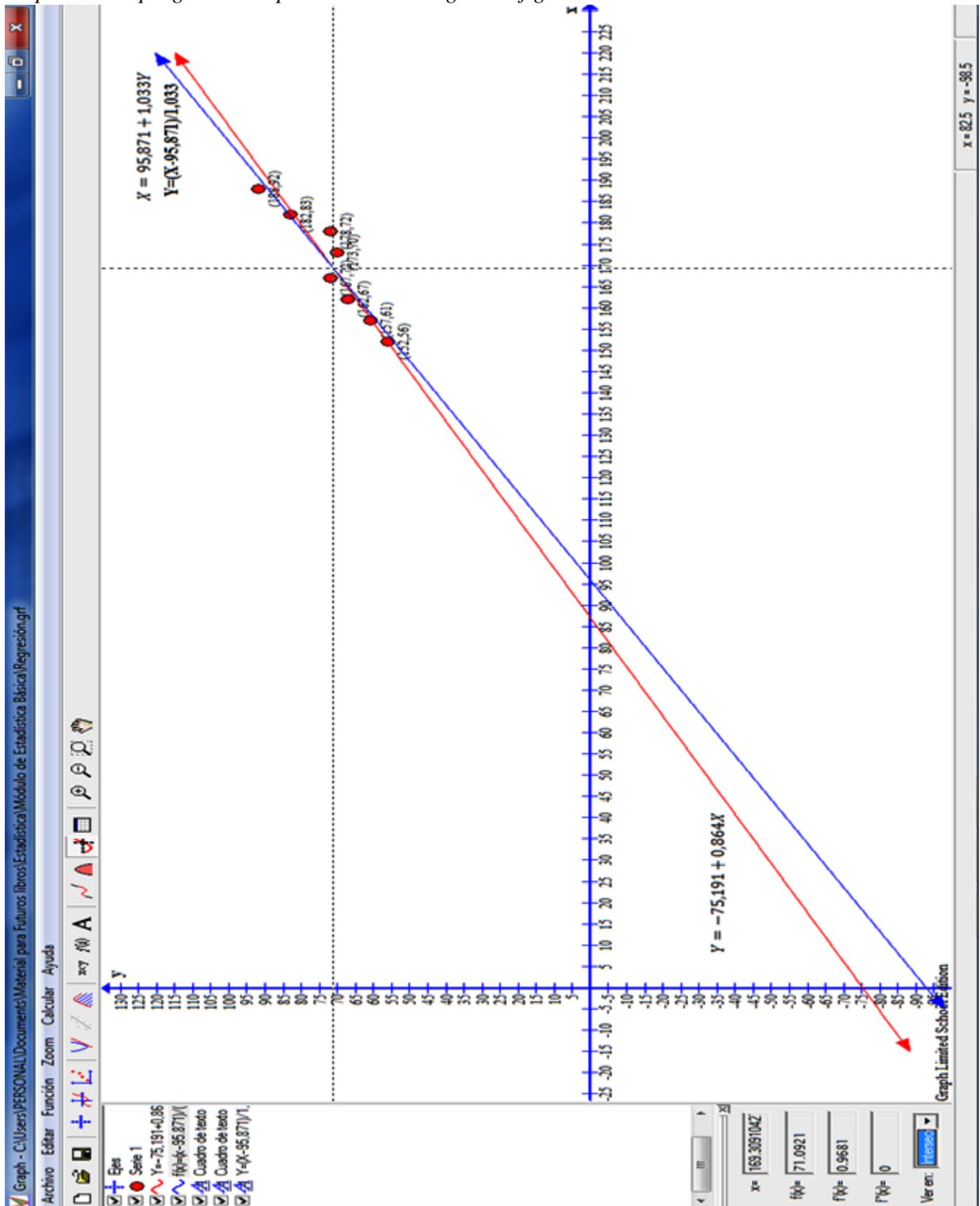
Elevando al cuadrado coeficiente de Pearson queda calculado el coeficiente de determinación.

$$\text{Coeficiente de determinación} = r^2 = (0,94497)^2 = 0,893$$

7) En Excel, insertando gráfico de dispersión se obtiene la siguiente figura:



Empleando el programa Graph se obtiene la siguiente figura:



8) Reemplazando $X = 200$ en la ecuación solicitada se obtiene:

$$Y = -75,191 + 0,864X = -75,191 + 0,864 \cdot 200 = -75,191 + 172,8 = 97,609$$

9) Reemplazando $Y = 100$ en la ecuación solicitada se obtiene:

$$X = 95,871 + 1,033Y = X = 95,871 + 1,033 \cdot 100 = X = 95,871 + 103,3 = 199,171$$

TAREA DE INTERAPRENDIZAJE N° 20

1) Consulte sobre la biografía de Francis Galton y de Cramer, y realice un organizador gráfico de cada una.

2) Dada la siguiente tabla sobre la altura en centímetros (X) y los pesos en kilogramos (Y) de una muestra de 8 estudiantes varones tomada al azar del segundo semestre de una universidad.

X	150	155	160	165	170	175	180	185
Y	55	60	63	67	70	74	79	85

2.1) Ajuste la recta de mínimos cuadrados para Y como variable dependiente resolviendo el sistema. El sistema resuelva de manera manual y mediante Excel.

$$\begin{cases} \Sigma Y = a_0 N + a_1 \Sigma X \\ \Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 \end{cases}$$

$$Y = -66,869 + 0,812X$$

2.2) Ajuste la recta de mínimos cuadrados para Y como variable dependiente empleando las fórmulas

$$a_0 = \frac{\Sigma Y \cdot \Sigma X^2 - \Sigma X \cdot \Sigma XY}{N \Sigma X^2 - (\Sigma X)^2}$$

$$a_1 = \frac{N \Sigma XY - \Sigma X \cdot \Sigma Y}{N \Sigma X^2 - (\Sigma X)^2}$$

$$a_0 = -66,869; \quad a_1 = 0,812$$

2.3) Ajuste la recta de mínimos cuadrados para Y como variable dependiente empleando la fórmula

$$y = \left(\frac{\Sigma xy}{\Sigma x^2} \right) x$$

$$Y = -66,869 + 0,812X$$

2.4) Ajuste la recta de mínimos cuadrados para X como variable dependiente resolviendo el sistema. El sistema resuelva de manera manual y mediante Excel.

$$\begin{cases} \Sigma X = b_0 N + b_1 \Sigma Y \\ \Sigma XY = b_0 \Sigma Y + b_1 \Sigma Y^2 \end{cases}$$

$$X = 83,18 + 1,22Y$$

2.5) Ajuste la recta de mínimos cuadrados para X como variable dependiente empleando las fórmulas

$$b_0 = \frac{\Sigma X \cdot \Sigma Y^2 - \Sigma Y \cdot \Sigma XY}{N \Sigma Y^2 - (\Sigma Y)^2}$$

$$b_1 = \frac{N \Sigma XY - \Sigma X \cdot \Sigma Y}{N \Sigma Y^2 - (\Sigma Y)^2}$$

$$b_0 = 83,18; \quad b_1 = 1,22$$

2.6) Ajuste la recta de mínimos cuadrados para X como variable dependiente empleando la fórmula

$$x = \left(\frac{\Sigma xy}{\Sigma y^2} \right) y$$

$$X = 83,18 + 1,22Y$$

2.7) Calcule el punto centroide de manera manual y empleando Excel.

$$\bar{X} = 170,9; \quad \bar{Y} = 71,9$$

2.8) Calcule el coeficiente de determinación.

$$0,99$$

2.9) Elabore el diagrama de dispersión. Y en el mismo diagrama graficar las dos rectas de mínimos cuadrados obtenidas en los pasos anteriores. Elabore de manera manual, empleando Excel y el programa Graph.

2.10) Estime el valor de Y cuando X = 173 en el diagrama de dispersión de Y como variable dependiente.

73,6

2.11) Estime el valor de X cuando Y = 73 en el diagrama de dispersión de Y como variable dependiente.

172,2

3) Cree y resuelva un ejercicio similar al anterior con datos obtenidos de 10 amigas suyas.

5.2.1.2) LA PARÁBOLA DE LOS MÍNIMOS CUADRADOS

La parábola de mínimos cuadrados que aproxima el conjunto de puntos $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots (X_N, Y_N)$ tiene ecuación dada por $Y = a_0 + a_1X + a_2X^2$, donde las constantes a_0, a_1 y a_2 se determinan al resolver simultáneamente el sistema de ecuaciones que se forma al multiplicar la ecuación $Y = a_0 + a_1X + a_2X^2$ por 1, X, Y sucesivamente, y sumando después.

$$\begin{cases} \Sigma Y = a_0N + a_1\Sigma X + a_2\Sigma X^2 \\ \Sigma XY = a_0\Sigma X + a_1\Sigma X^2 + a_2\Sigma X^3 \\ \Sigma X^2Y = a_0\Sigma X^2 + a_1\Sigma X^3 + a_2\Sigma X^4 \end{cases}$$

Ejemplo ilustrativo

La siguiente tabla muestra la población de un país en los años 1960-2010 en intervalos de 5 años.

Año	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010
Población (millones)	4,52	5,18	6,25	7,42	8,16	9,12	10,92	11,62	12,68	13,12	13,97

1) Ajustar una parábola de mínimos cuadrados de la forma $Y = a_0 + a_1X + a_2X^2$

2) Calcular los valores de tendencia para los años dados.

3) Estimar la población para los años 2015 y 2020.

4) Calcular el coeficiente de determinación.

5) Elaborar un diagrama de dispersión, y en el mismo diagrama graficar la parábola de los mínimos cuadrados.

Nota: Se recomienda codificar o cambiar la numeración de los años, eligiendo X de modo que el año central, 1985, corresponda a X= 0, para que se hagan más fáciles los cálculos.

Solución:

1) Para ajustar una parábola de mínimos cuadrados se llena la siguiente tabla:

Año	X	Y	X ²	X ³	X ⁴	XY	X ² Y
1960	-5	4,52	25	-125	625	-22,6	113
1965	-4	5,18	16	-64	256	-20,72	82,88
1970	-3	6,25	9	-27	81	-18,75	56,25
1975	-2	7,42	4	-8	16	-14,84	29,68
1980	-1	8,16	1	-1	1	-8,16	8,16
1985	0	9,12	0	0	0	0	0
1990	1	10,92	1	1	1	10,92	10,92
1995	2	11,62	4	8	16	23,24	46,48
2000	3	12,68	9	27	81	38,04	114,12
2005	4	13,12	16	64	256	52,48	209,92
2010	5	13,97	25	125	625	69,85	349,25
Σ	0	102,96	110	0	1958	109,46	1020,66

Se reemplaza valores en el sistema y se obtiene:

$$\begin{cases} \Sigma Y = a_0 N + a_1 \Sigma X + a_2 \Sigma X^2 \\ \Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 + a_2 \Sigma X^3 \\ \Sigma X^2 Y = a_0 \Sigma X^2 + a_1 \Sigma X^3 + a_2 \Sigma X^4 \end{cases}$$

$$\begin{cases} 102,96 = a_0 \cdot 11 + a_1 \cdot 0 + a_2 \cdot 110 \\ 109,46 = a_0 \cdot 0 + a_1 \cdot 110 + a_2 \cdot 0 \\ 1020,66 = a_0 \cdot 110 + a_1 \cdot 0 + a_2 \cdot 1958 \end{cases} \Rightarrow \begin{cases} 11a_0 + 0a_1 + 110a_2 = 102,96 \\ 0a_0 + 110a_1 + 0a_2 = 109,46 \\ 110a_0 + 0a_1 + 1958a_2 = 1020,66 \end{cases}$$

Resolviendo el sistema empleando determinantes (regla de Cramer) se obtiene:

$$a_0 = \frac{\Delta a_0}{\Delta} = \frac{\begin{vmatrix} 102,96 & 0 & 110 \\ 109,46 & 110 & 0 \\ 1020,66 & 0 & 1958 \end{vmatrix}}{\begin{vmatrix} 11 & 0 & 110 \\ 0 & 110 & 0 \\ 110 & 0 & 1958 \end{vmatrix}} = \frac{22175524,8 + 0 + 0 - 12349986 - 0 - 0}{2369180 + 0 + 0 - 1331000 - 0 - 0} = \frac{9825538,8}{1038180} = 9,464$$

$$a_1 = \frac{\Delta a_1}{\Delta} = \frac{\begin{vmatrix} 11 & 102,96 & 110 \\ 0 & 109,46 & 0 \\ 110 & 1020,66 & 1958 \end{vmatrix}}{1038180} = \frac{\begin{vmatrix} 11 & 102,96 & 110 \\ 0 & 109,46 & 0 \\ 110 & 1020,66 & 1958 \end{vmatrix}}{1038180}$$

$$a_1 = \frac{23577549,48 + 0 + 0 - 1324466 - 0 - 0}{1038180} = \frac{2357549,48}{1038180} = 0,995$$

$$a_2 = \frac{\Delta a_2}{\Delta} = \frac{\begin{vmatrix} 11 & 0 & 102,96 \\ 0 & 110 & 109,46 \\ 110 & 0 & 1020,66 \end{vmatrix}}{1038180} = \frac{\begin{vmatrix} 11 & 0 & 102,96 \\ 0 & 110 & 109,46 \\ 110 & 0 & 1020,66 \end{vmatrix}}{1038180}$$

$$a_2 = \frac{1234998,6 + 0 + 0 - 1245816 - 0 - 0}{1038180} = \frac{-10817,4}{1038180} = -0,01$$

El sistema resuelto en Excel se muestra en la siguiente figura:

	A	B	C	D	E	F	G	H
16						a_0	a_1	a_2
17		$11a_0 + 0a_1 + 110a_2 = 102,96$			102,96	11	0	110
18		$0a_0 + 110a_1 + 0a_2 = 109,46$			109,46	0	110	0
19		$110a_0 + 0a_1 + 1958a_2 = 1020,66$			1020,66	110	0	1958
20								
21	102,96	0	110					
22	109,46	110	0		$a_0 = \frac{\Delta a_0}{\Delta}$	9,464	=B24/B40	
23	1020,66	0	1958					
24	Δa_0	9825538,8	=MDETERM(A21:C23)					
25								
26								
27	11	102,96	110					
28	0	109,46	0		$a_1 = \frac{\Delta a_1}{\Delta}$	0,995	=B30/B40	
29	110	1020,66	1958					
30	Δa_1	1033083,5	=MDETERM(A27:C29)					
31								
32	11	0	102,96					
33	0	110	109,46					
34	110	0	1020,66					
35	Δa_2	-10817,4	=MDETERM(A32:C34)					
36								
37	11	0	110					
38	0	110	0		$a_2 = \frac{\Delta a_2}{\Delta}$	-0,010	=B35/B40	
39	110	0	1958					
40	Δ	1038180	=MDETERM(A37:C39)					

Reemplazando los valores encontrados se obtiene la ecuación de la parábola de mínimos cuadrados:

$$Y = a_0 + a_1X + a_2X^2 \Rightarrow Y = 9,464 + 0,995X - 0,01X^2$$

2) Los valores de tendencia se obtienen al reemplazar los valores de X en la ecuación de la parábola de mínimos cuadrados, los cuales se presenta en la siguiente tabla:

Año	X	Y	Valores de tendencia $Y = 9,464 + 0,995X - 0,01X^2$
1960	-5	4,52	4,24
1965	-4	5,18	5,32
1970	-3	6,25	6,39
1975	-2	7,42	7,43
1980	-1	8,16	8,46
1985	0	9,12	9,46
1990	1	10,92	10,45
1995	2	11,62	11,41
2000	3	12,68	12,36
2005	4	13,12	13,28
2010	5	13,97	14,19

3) Para estimar la población de los años 2015 y 2020 se transforma estos años a X siguiendo la secuencia de la tabla anterior, siendo X = 6 para el año 2015 y X= 7 para el 2020

Entonces para el 2015 se tiene:

$$Y = 9,464 + 0,995X - 0,01X^2 = 9,464 + 0,995(6) - 0,01(6)^2 = 9,464 + 5,97 - 0,36 = 15,074$$

Para el 2020 se tiene:

$$Y = 9,464 + 0,995X - 0,01X^2 = 9,464 + 0,995(7) - 0,01(7)^2 = 9,464 + 6,965 - 0,49 = 15,939$$

4) Se llena la siguiente tabla y se aplica la ecuación para calcular el coeficiente de Pearson

Año	X	Y	X ²	XY	Y ²
1960	-5	4,52	25	-22,6	20,430
1965	-4	5,18	16	-20,72	26,832
1970	-3	6,25	9	-18,75	39,063
1975	-2	7,42	4	-14,84	55,056
1980	-1	8,16	1	-8,16	66,586
1985	0	9,12	0	0	83,174
1990	1	10,92	1	10,92	119,246
1995	2	11,62	4	23,24	135,024
2000	3	12,68	9	38,04	160,782
2005	4	13,12	16	52,48	172,134
2010	5	13,97	25	69,85	195,161
Σ	0	102,96	110	109,46	1073,490

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} = \frac{11 \cdot 109,46 - 0 \cdot 102,96}{\sqrt{[11 \cdot 110 - (0)^2][11 \cdot 1073,490 - (102,96)^2]}}$$

$$r = 0,996$$

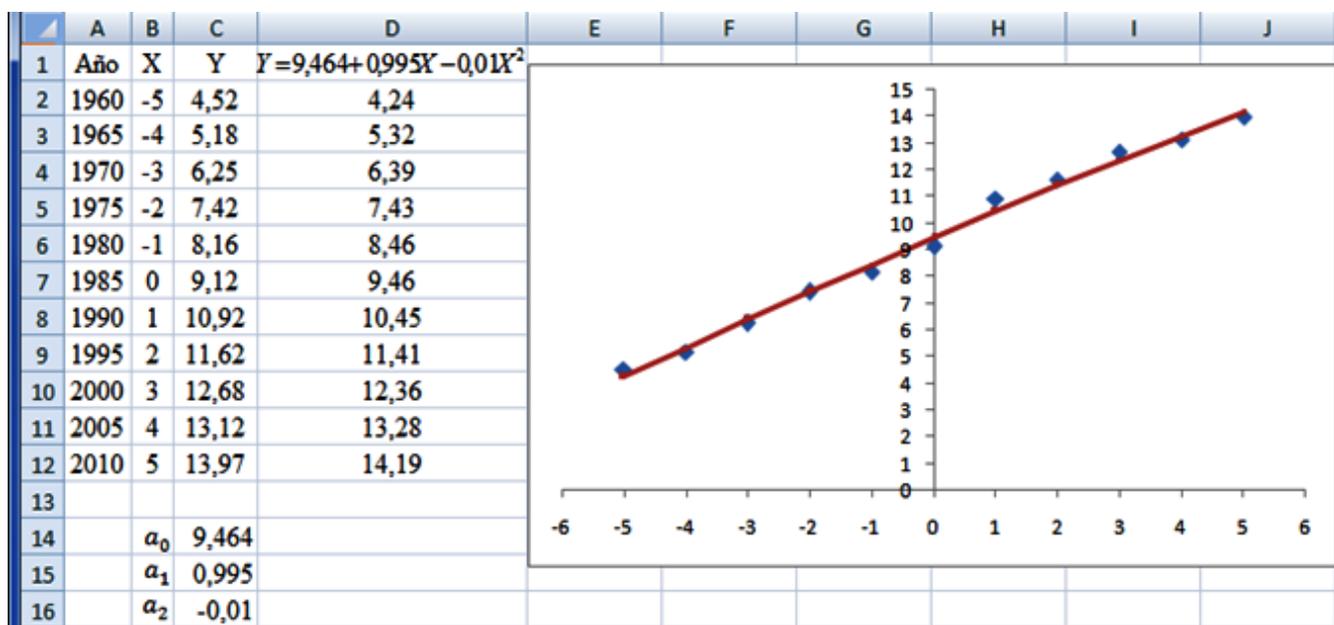
Elevando al cuadrado coeficiente de Pearson queda calculado el coeficiente de determinación.

$$\text{Coeficiente de determinación} = r^2 = (0,996)^2 = 0,992$$

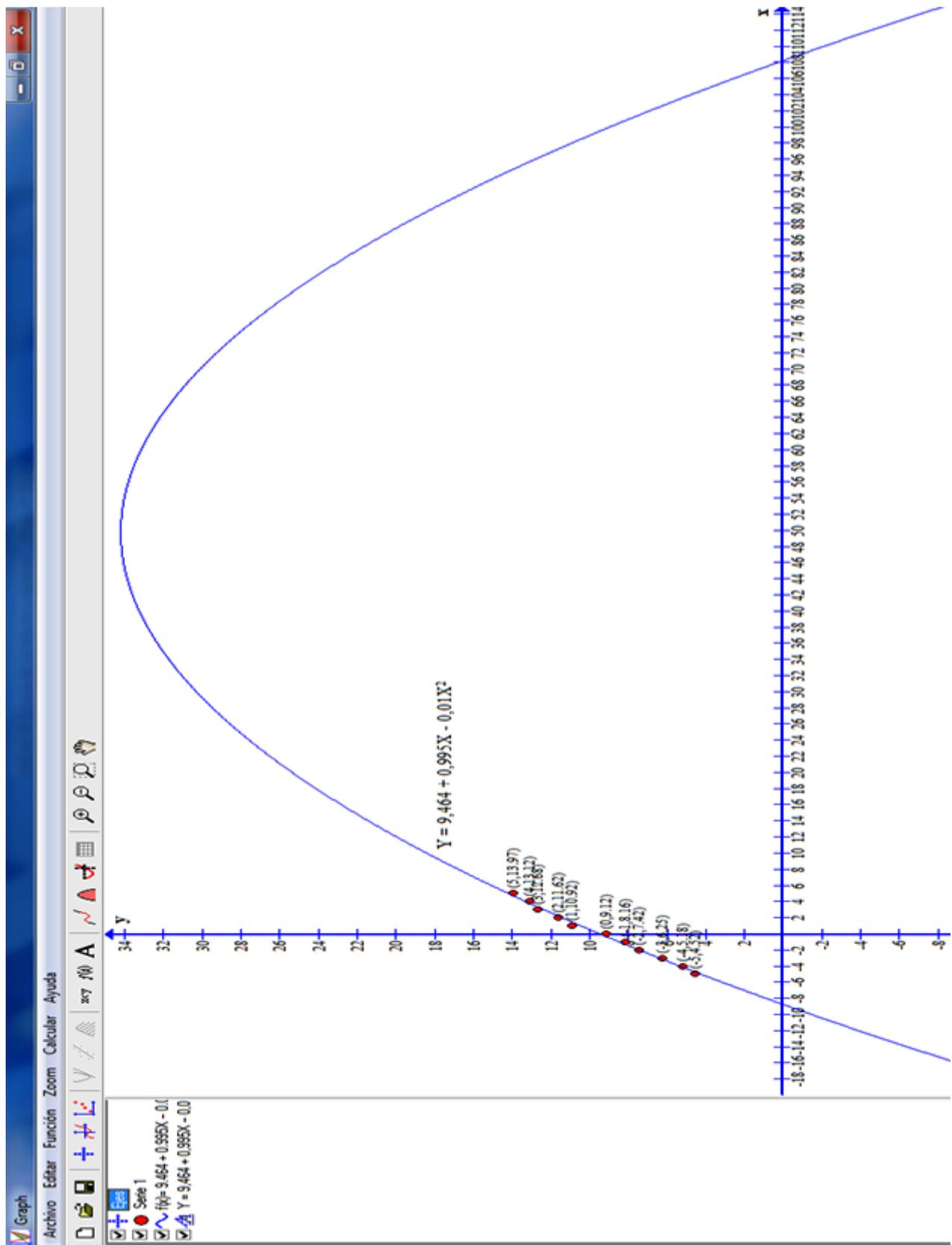
El coeficiente de determinación calculado en Excel se muestra en la siguiente figura:

	A	B	C	D	E
1	Año	X	Y		
2	1960	-5	4,52		
3	1965	-4	5,18		
4	1970	-3	6,25		
5	1975	-2	7,42		
6	1980	-1	8,16		
7	1985	0	9,12		
8	1990	1	10,92		
9	1995	2	11,62		
10	2000	3	12,68		
11	2005	4	13,12		
12	2010	5	13,97		
13					
14	r	0,9960666	=COEF.DE.CORREL(B2:B12;C2:C12)		
15		0,9960666	=COEF.DE.CORREL(A2:A12;C2:C12)		
16	r ²	0,9921487	=E15^2		
17		0,9921487	=COEFICIENTE.R2(C2:C12;B2:B12)		

5) El diagrama de dispersión y la parábola de los mínimos cuadrados mediante Excel se muestra en la siguiente figura:



Empleando el programa Graph se obtiene la siguiente figura:



TAREA DE INTERAPRENDIZAJE N° 21

1) La siguiente tabla muestra la población aproximada de la Provincia de Imbabura en los años 1960-2010 en intervalos de 5 años.

Año	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010
Población (miles)	123	140	170	201	221	247	296	315	344	356	379

1.1) Ajuste una parábola de mínimos cuadrados de la forma $Y = a_0 + a_1X + a_2X^2$
 $Y = 256,464 + 26,991X - 0,265X^2$

1.2) Calcule los valores de tendencia para los años dados de manera manual y empleando Excel.

Año	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010
Valor de tendencia	114,88	144,26	173,11	201,42	229,21	256,46	283,19	309,39	335,05	360,19	384,79

1.3) Estime la población para los años 2015 y 2020

Año 2015 = 408,87 miles de habitantes
 Año 2020 = 432,42 miles de habitantes

1.4) Calcule el coeficiente de determinación de manera manual y empleando Excel.

0,992

1.5) Elabore un diagrama de dispersión, y en el mismo diagrama graficar la parábola de los mínimos cuadrados de manera manual, empleando Excel y empleando Graph.

2) Cree y resuelva un ejercicio de aplicación de la parábola de los mínimos cuadrados con datos de la población del Ecuador o de cualquier otro país de manera manual, empleando Excel y Graph.

5.2.1.3) REGRESIÓN EXPONENCIAL

Cuando la curva de regresión de y sobre x es exponencial, es decir para cualquier x considerada, la media de la distribución está dada por la siguiente ecuación predictora:

$$Y = \alpha \cdot \beta^X$$

Tomando logaritmos en ambos miembros:

$$\log Y = \log \alpha + X \cdot \log \beta$$

Y se puede estimar ahora $\log Y$ y $\log \beta$, y de ahí obtener α y β , aplicando *los métodos de los mínimos cuadrados*.

Donde las constantes α y β quedan fijadas al resolver simultáneamente las ecuaciones:

$$\begin{cases} \Sigma \log Y = \log \alpha \cdot N + \log \beta \cdot \Sigma X \\ \Sigma X \cdot \log Y = \log \alpha \cdot \Sigma X + \log \beta \cdot \Sigma X^2 \end{cases}$$

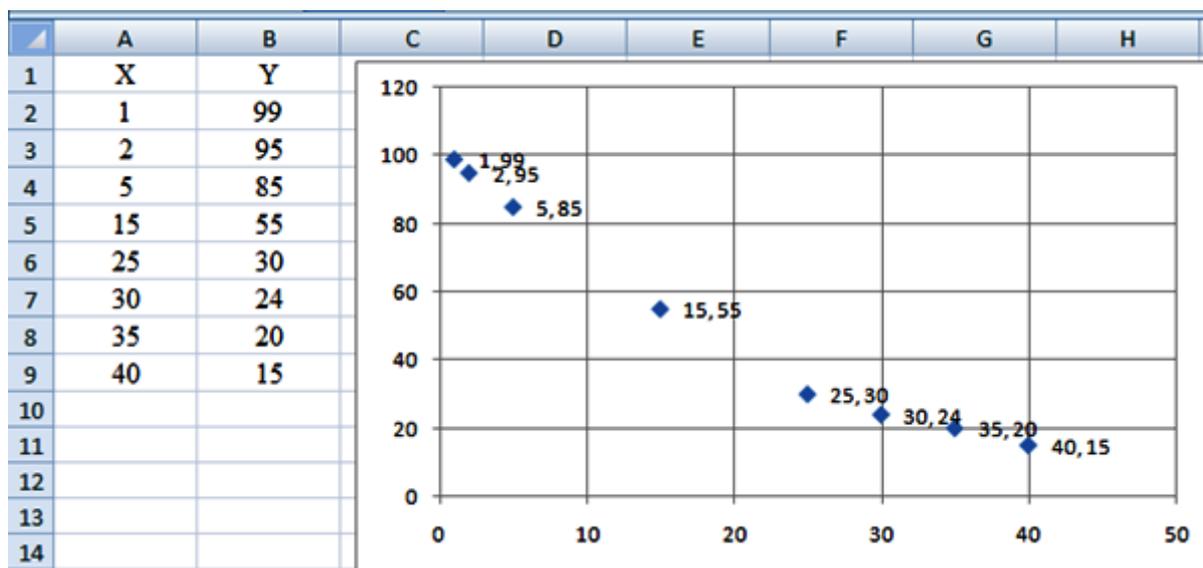
Ejemplo ilustrativo: Las cifras siguientes son datos sobre el porcentaje de llantas radiales producidas por cierto fabricante que aún pueden usarse después de recorrer cierto número de millas:

Miles de Millas recorridas (X)	1	2	5	15	25	30	35	40
Porcentaje útil (Y)	99	95	85	55	30	24	20	15

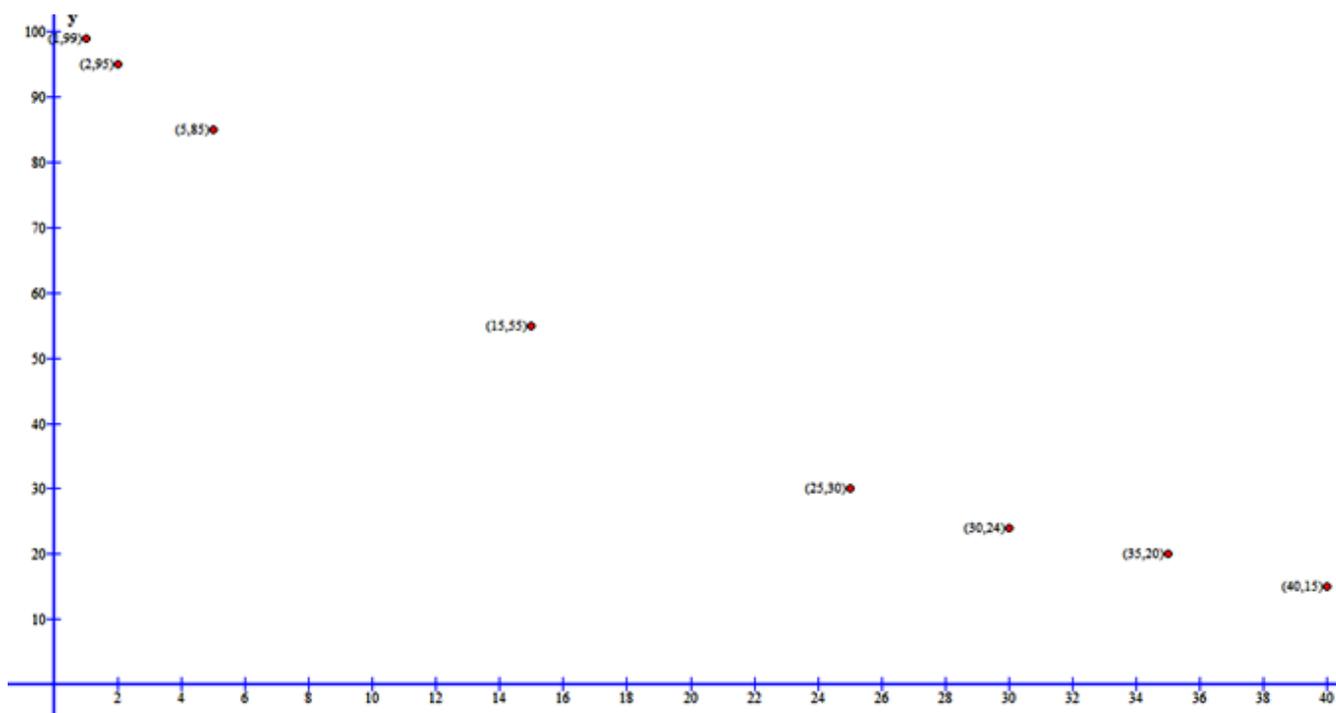
- 1) Elaborar el diagrama de dispersión.
- 2) Ajustar una curva exponencial aplicando el método de mínimos cuadrados.
- 3) Calcular la ecuación predictora.
- 4) Graficar la ecuación predictora.
- 5) Estimar qué porcentaje de las llantas radiales del fabricante durarán 50000 millas.

Solución:

1) *Elaborando el diagrama de dispersión empleando Excel se obtiene la siguiente figura:*



Empleando el programa Graph se obtiene la siguiente figura:



2) Se llena la siguiente tabla:

X	Y	logY	X ²	X · logY
1	99	1,996	1	1,996
2	95	1,978	4	3,955
5	85	1,929	25	9,647
15	55	1,740	225	26,105
25	30	1,477	625	36,928
30	24	1,380	900	41,406
35	20	1,301	1225	45,536
40	15	1,176	1600	47,044
ΣX = 153		ΣlogY = 12,97759	ΣX ² = 4605	ΣX · logY = 212,61769

Resolviendo empleando Excel se muestra en la siguiente figura:

	A	B	C	D	E	F	G	H	I
1		X	Y	log Y		X ²		X logY	
2		1	99	1,99564	=LOG10(C2)	1	=B2^2	1,99564	=B2*D2
3		2	95	1,97772	=LOG10(C3)	4	=B3^2	3,95545	=B3*D3
4		5	85	1,92942	=LOG10(C4)	25	=B4^2	9,64709	=B4*D4
5		15	55	1,74036	=LOG10(C5)	225	=B5^2	26,10544	=B5*D5
6		25	30	1,47712	=LOG10(C6)	625	=B6^2	36,92803	=B6*D6
7		30	24	1,38021	=LOG10(C7)	900	=B7^2	41,40634	=B7*D7
8		35	20	1,30103	=LOG10(C8)	1225	=B8^2	45,53605	=B8*D8
9		40	15	1,17609	=LOG10(C9)	1600	=B9^2	47,04365	=B9*D9
10	Σ	153		12,97759		4605		212,61769	
11				=SUMA(D2:D9)		=SUMA(F2:F9)		=SUMA(H2:H9)	
12	N	8	=CONTAR(A2:A9)						

Reemplazando valores en el sistema se obtiene:

$$\begin{cases} \Sigma \log Y = \log \alpha \cdot N + \log \beta \cdot \Sigma X \\ \Sigma X \cdot \log Y = \log \alpha \cdot \Sigma X + \log \beta \cdot \Sigma X^2 \end{cases}$$

$$\begin{cases} 12,97759 = \log \alpha \cdot 8 + \log \beta \cdot 153 \\ 212,61769 = \log \alpha \cdot 153 + \log \beta \cdot 4605 \end{cases} \Rightarrow \begin{cases} 8 \log \alpha + 153 \log \beta = 12,97759 \\ 153 \log \alpha + 4605 \log \beta = 212,61769 \end{cases}$$

Al resolver el sistema se obtiene:

$$\log \alpha = \frac{\Delta_{\alpha}}{\Delta} = \frac{\begin{vmatrix} 12,97759 & 153 \\ 212,61769 & 4605 \end{vmatrix}}{\begin{vmatrix} 8 & 153 \\ 153 & 4605 \end{vmatrix}} = \frac{59761,80195 - 32530,50657}{36840 - 23409} = \frac{27231,295389}{13431} = 2,027495747$$

$$\log \beta = \frac{\Delta_{\beta}}{\Delta} = \frac{\begin{vmatrix} 8 & 12,97759 \\ 153 & 212,61769 \end{vmatrix}}{13431} = \frac{1700,944152 - 1985,57127}{13431} = \frac{-284,627118}{13431} = -0,02119180389$$

Reemplazando valores se obtiene:

$$\log Y = \log \alpha + X \cdot \log \beta$$

$$\log Y = 2,027496 - 0,02119X$$

Aplicando el antilogaritmo se obtiene:

$$\alpha = \text{anti log } 2,027495747 = 106,536$$

$$\beta = \text{anti log}(-0,02119180389) = 0,952$$

Resolviendo empleando Excel se muestra en la siguiente figura:

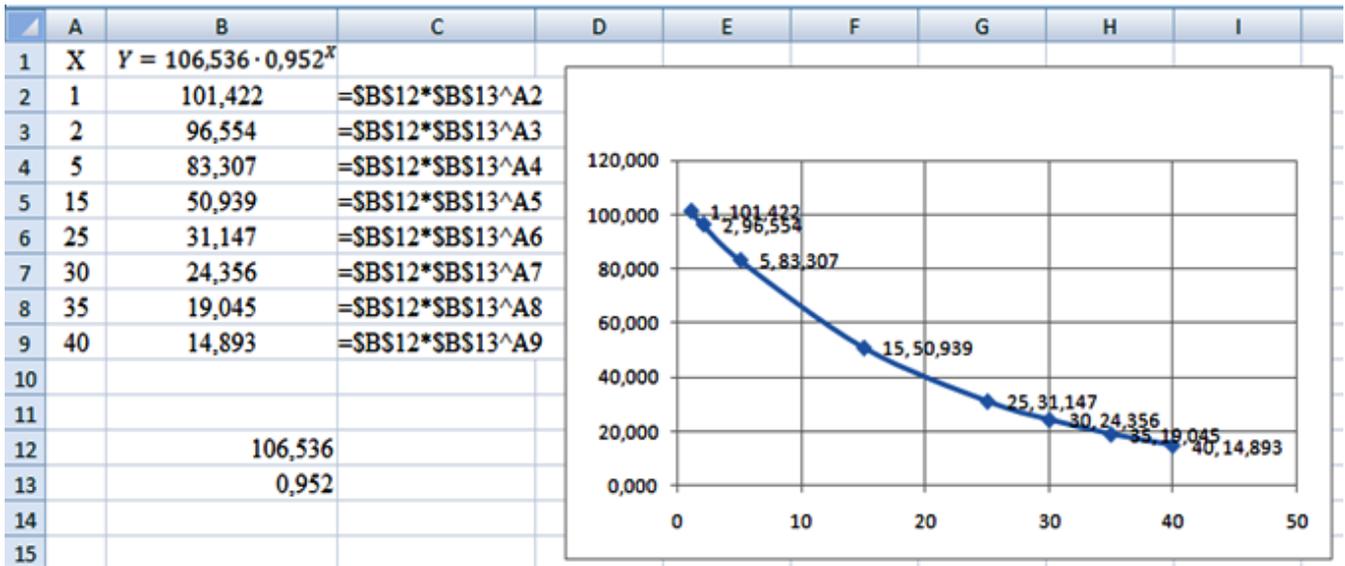
	A	B	C	D	E	F
13						
14	}	8log α + 153log β = 12,97759				
15		153log α + 4605log β = 212,61769				
16						
17	8,000	153	12,9776			
18	153	4605	212,6177			
19						
20	12,97759	153				
21	212,61769	4605				
22			27231,3151460	=MDETERM(A20:B21)		
23						
24	8	12,97759				
25	153	212,61769				
26			-284,6304178	=MDETERM(A24:B25)		
27						
28	8	153				
29	153	4605				
30			13431	=MDETERM(A28:B29)		
31						
32	log α	2,02749722	=C22/C30		α	106,536 =10^G21
33						
34	log β	-0,02119205	=C26/C30		β	0,952 =10^G23

3) Reemplazando en la ecuación predictora se obtiene:

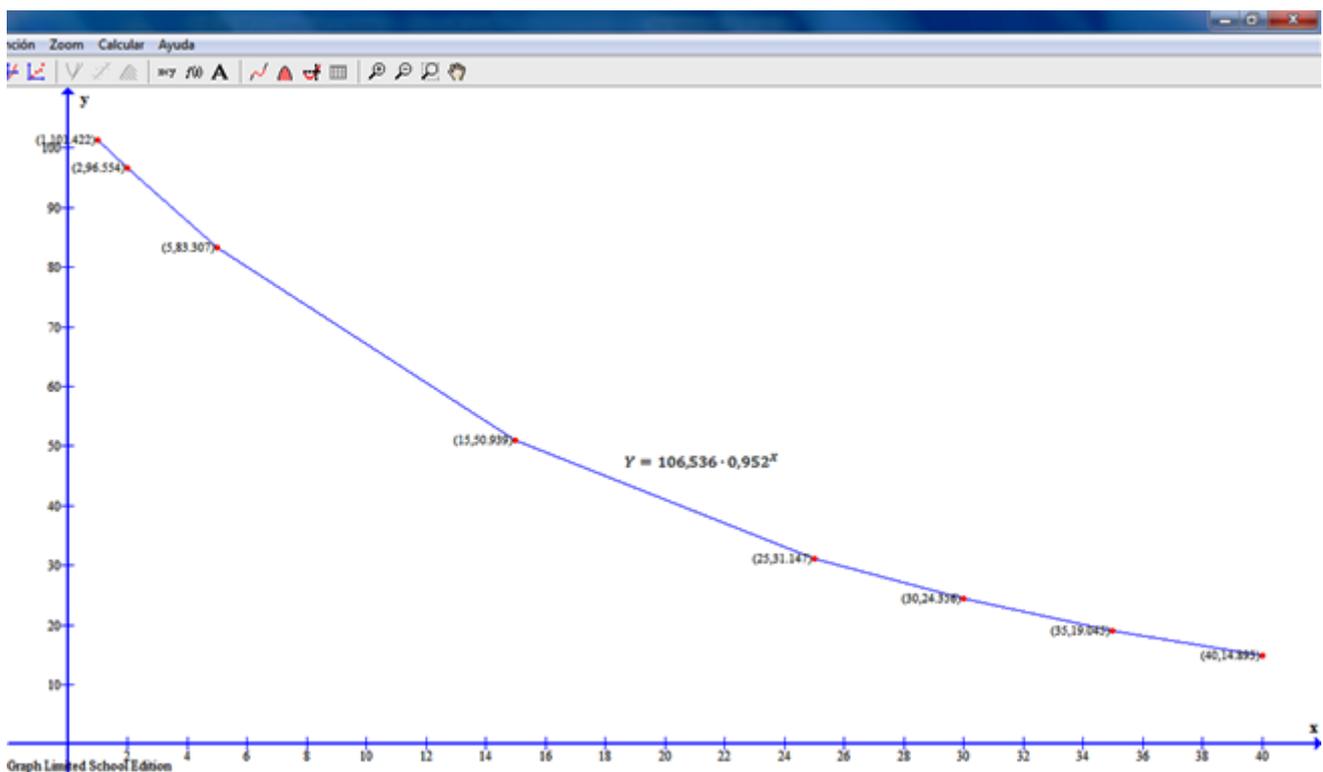
$$Y = \alpha \cdot \beta^X$$

$$Y = 106,536 \cdot 0,952^X$$

4) Graficando la ecuación predictora empleando Excel se obtiene la siguiente figura:



En Graph se obtiene la siguiente figura:



5) La estimación del porcentaje de llantas radiales que durarán 50000 millas se obtiene reemplazando en la ecuación predictora el valor de $X = 50$

$$Y = 106,536 \cdot 0,952^X$$

$$Y = 106,536 \cdot 0,952^{50} = 9,106$$

Entonces el porcentaje sería de 9,106%

TAREA DE INTERAPRENDIZAJE N° 22

- 1) Elabore un organizador gráfico sobre la regresión exponencial.
- 2) Las cifras siguientes son datos sobre el porcentaje de llantas radiales producidas por cierto fabricante que aún pueden usarse después de recorrer cierto número de millas:

Miles de Millas recorridas (X)	1	2	5	10	20	30	40	50
Porcentaje útil (Y)	98	92	80	64	36	32	17	11

2.1) Ajuste una curva exponencial aplicando el método de mínimos cuadrados. Resolver manualmente y empleando Excel. Realizar los cálculos empleando la mayor cantidad de decimales.

$$\log Y = 1,9988 - 0,0189X$$

2.2) Calcule la ecuación predictora.

$$Y = 99,72 \cdot 0,9574^X$$

2.3) Grafique la ecuación predictora de manera manual, empleando Excel y el programa Graph.

2.4) Estime qué porcentaje de las llantas radiales del fabricante durarán 35000 millas.

21,7%

3) Cree y resuelva un ejercicio empleando los conocimientos de la regresión exponencial de manera manual, empleando Excel y Graph.

5.2.1.4) REGRESIÓN POTENCIAL

La regresión potencial tiene por ecuación predictora:

$$Y = \alpha \cdot X^\beta$$

Y la regresión recíproca es:

$$Y = \frac{1}{\alpha + \beta \cdot X}$$

Para el primer caso los valores siguen una ley potencial. Si la ecuación predictora está dada por: $Y = \alpha \cdot X^\beta$, tomando logaritmos en ambos miembros, queda:

$$\log Y = \log \alpha + \beta \cdot \log X$$

Donde las constantes α y β quedan fijadas al resolver simultáneamente las ecuaciones:

$$\begin{cases} \Sigma \log Y = \log \alpha \cdot N + \beta \cdot \Sigma \log X \\ \Sigma \log X \cdot \log Y = \log \alpha \cdot \Sigma \log X + \beta \cdot \Sigma (\log X)^2 \end{cases}$$

Para el segundo caso, si la ecuación predictora está dada por $Y = 1/(\alpha + \beta \cdot X)$, entonces invirtiendo, la misma expresión se puede escribir $1/Y = (\alpha + \beta \cdot X)/1$, o sea:

$$Y = \frac{1}{\alpha + \beta \cdot X} \Rightarrow \frac{1}{Y} = \alpha + \beta \cdot X$$

Donde las constantes α y β quedan fijadas al resolver simultáneamente las ecuaciones:

$$\begin{cases} \Sigma \frac{1}{Y} = \alpha \cdot N + \beta \cdot \Sigma X \\ \Sigma X \cdot \frac{1}{Y} = \alpha \cdot \Sigma X + \beta \cdot \Sigma X^2 \end{cases}$$

Ejemplos ilustrativo N° 1

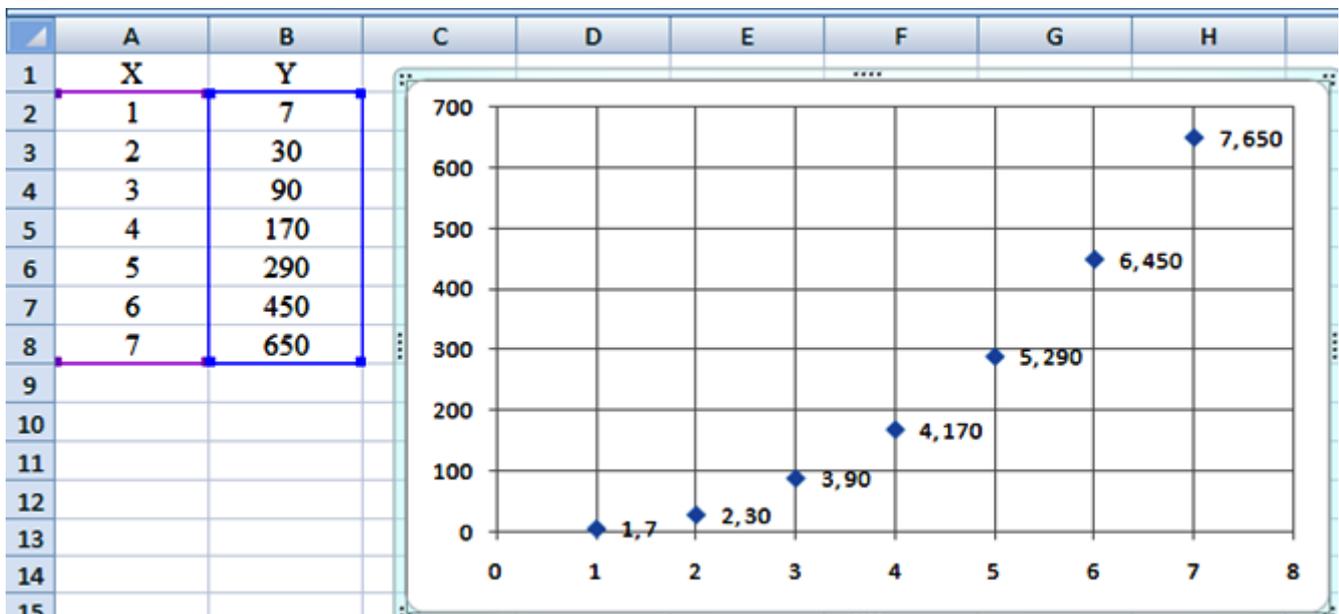
Sea el siguiente conjunto de valores, las lecturas de un experimento donde X es el volumen (variable independiente) e Y es la presión de una masa dada de gas (variable resultante).

X	1	2	3	4	5	6	7
Y	7	30	90	170	290	450	650

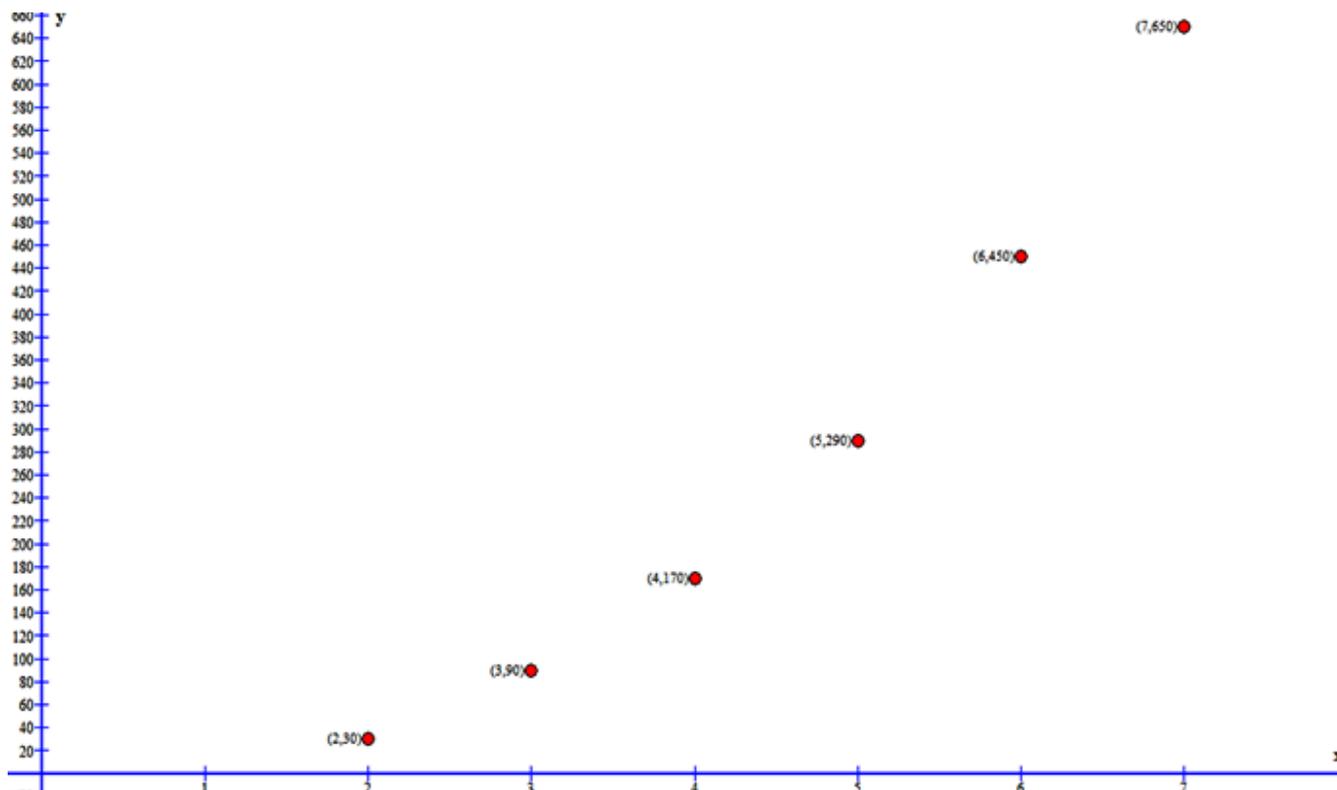
- 1.1) Elaborar el diagrama de dispersión.
- 1.2) Ajustar una curva exponencial aplicando el método de mínimos cuadrados.
- 1.3) Calcular la ecuación predictora.
- 1.4) Graficar la ecuación predictora.
- 1.5) Estimar la presión de la masa de gas de volumen 9.

Solución:

1.1) El diagrama de dispersión elaborado en Excel se presenta en la siguiente figura:



El diagrama de dispersión elaborado en Graph se presenta en la siguiente figura:



1.2) Para ajustar una curva exponencial aplicando el método de mínimos cuadrados se llena la siguiente tabla:

X	Y	$\log X$	$\log Y$	$\log X \cdot \log Y$	$(\log X)^2$
1	7	0,0000	0,8451	0,0000	0,0000
2	30	0,3010	1,4771	0,4447	0,0906
3	90	0,4771	1,9542	0,9324	0,2276
4	170	0,6021	2,2304	1,3429	0,3625
5	290	0,6990	2,4624	1,7211	0,4886
6	450	0,7782	2,6532	2,0646	0,6055
7	650	0,8451	2,8129	2,3772	0,7142
$\Sigma X = 28$		$\Sigma \log X = 3,7024$	$\Sigma \log Y = 14,4354$	$\Sigma \log X \cdot \log Y = 8,8829$	$\Sigma (\log X)^2 = 2,4890$

Reemplazando valores en el sistema de ecuaciones se obtiene:

$$\begin{cases} \Sigma \log Y = \log \alpha \cdot N + \beta \cdot \Sigma \log X \\ \Sigma \log X \cdot \log Y = \log \alpha \cdot \Sigma \log X + \beta \cdot \Sigma (\log X)^2 \end{cases}$$

$$\begin{cases} 14,4354 = \log \alpha \cdot 7 + \beta \cdot 3,7024 \\ 8,8829 = \log \alpha \cdot 3,7024 + \beta \cdot 2,4890 \end{cases} \Rightarrow \begin{cases} 7 \log \alpha + 3,7024 \beta = 14,4354 \\ 3,7024 \log \alpha + 2,4890 \beta = 8,8829 \end{cases}$$

Al resolver el sistema se obtiene: $\log \alpha = 0,819$; $\beta = 2,351$

Reemplazando valores en la ecuación predictora expresada en logaritmos se tiene:

$$\log Y = \log \alpha + \beta \cdot \log X$$

$$\log Y = 0,819 + 2,351 \cdot \log X$$

1.3) Para calcular la ecuación predictora, primero se calcula el valor de α de la siguiente manera:

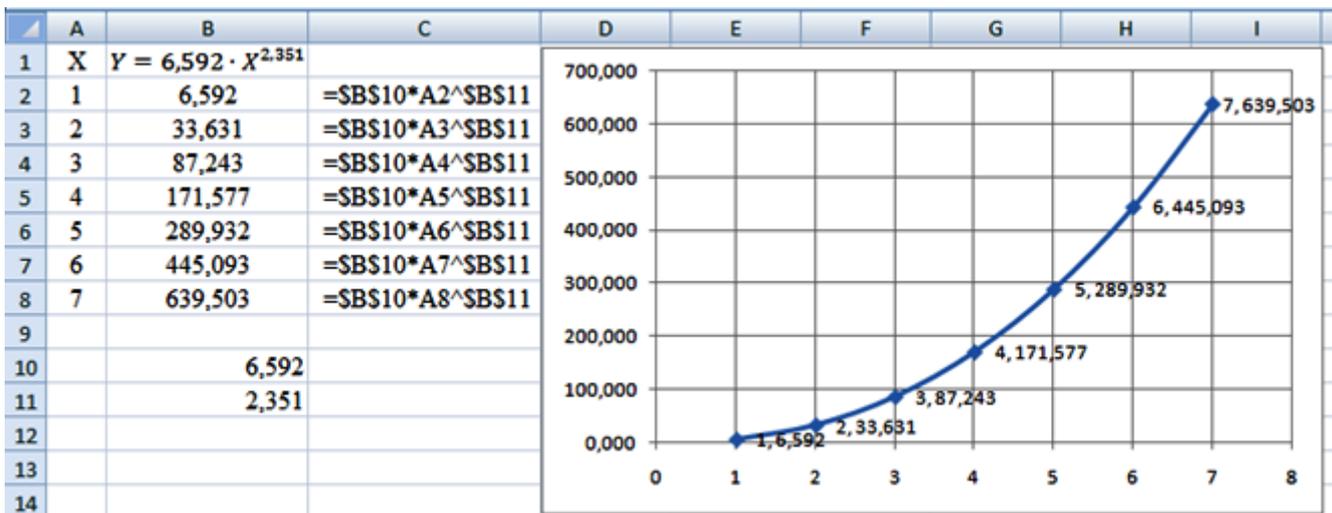
$$\log \alpha = 0,819 \Rightarrow \alpha = \text{antilog } 0,819 = 6,592$$

Reemplazando en la ecuación predictora se obtiene:

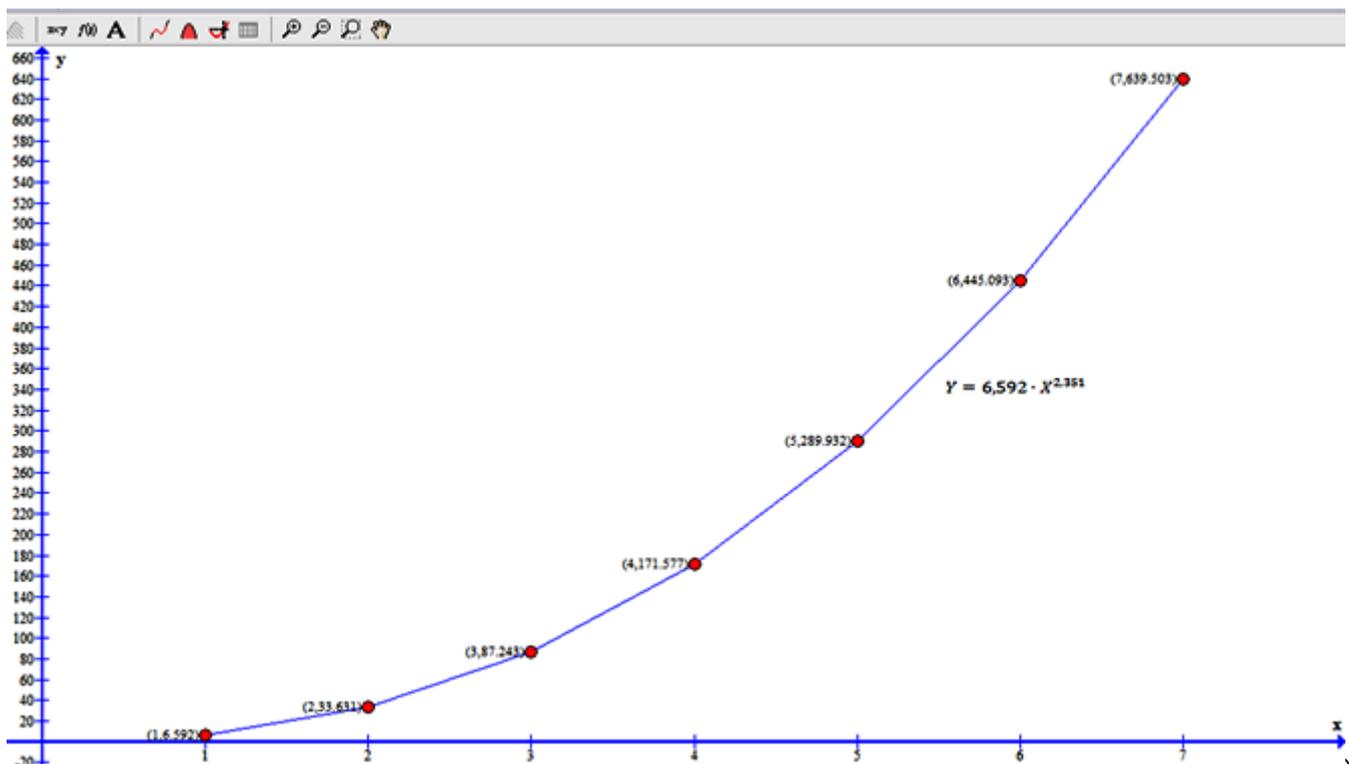
$$Y = \alpha \cdot X^\beta$$

$$Y = 6,592 \cdot X^{2,351}$$

1.4) Graficando la ecuación predictora mediante Excel se muestra en la siguiente figura:



Empleando Graph se obtiene la siguiente figura:



1.5) Para estimar la presión de la masa de gas de volumen 9 se reemplaza el valor $X = 9$ en la ecuación predictora

$$Y = 6,592 \cdot X^{2,351}$$

$$Y = 6,592 \cdot 9^{2,351} = 1154,63$$

Ejemplo ilustrativo N° 2

Sea el siguiente conjunto de valores, las lecturas de un experimento donde X es la variable independiente e Y la variable resultante.

X	1	2	3	4	5	6	7
Y	1,4	1	0,9	0,7	0,6	0,55	0,5

2.1) Elaborar el diagrama de dispersión.

2.2) Calcular las constantes α y β , aplicando el método de mínimos cuadrados.

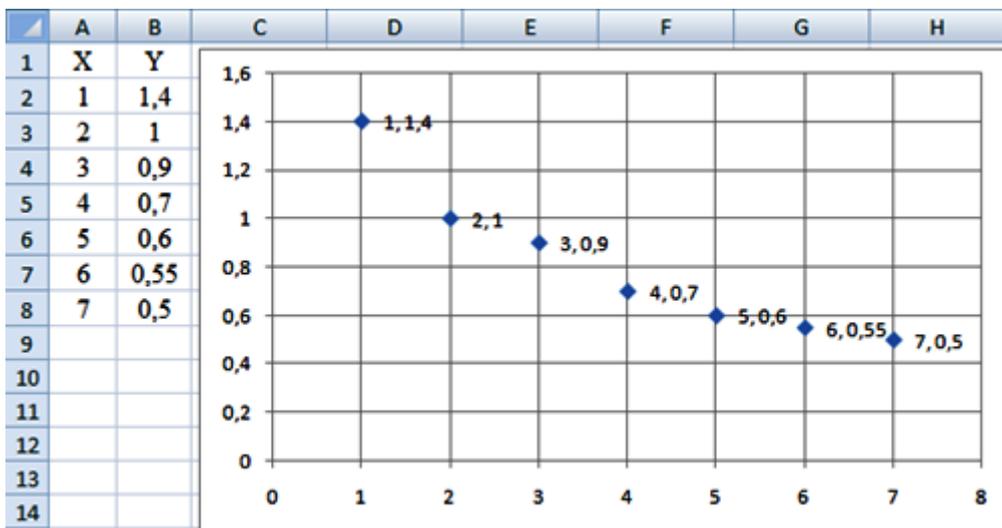
2.3) Calcular la ecuación predictora.

2.4) Graficar la ecuación predictora.

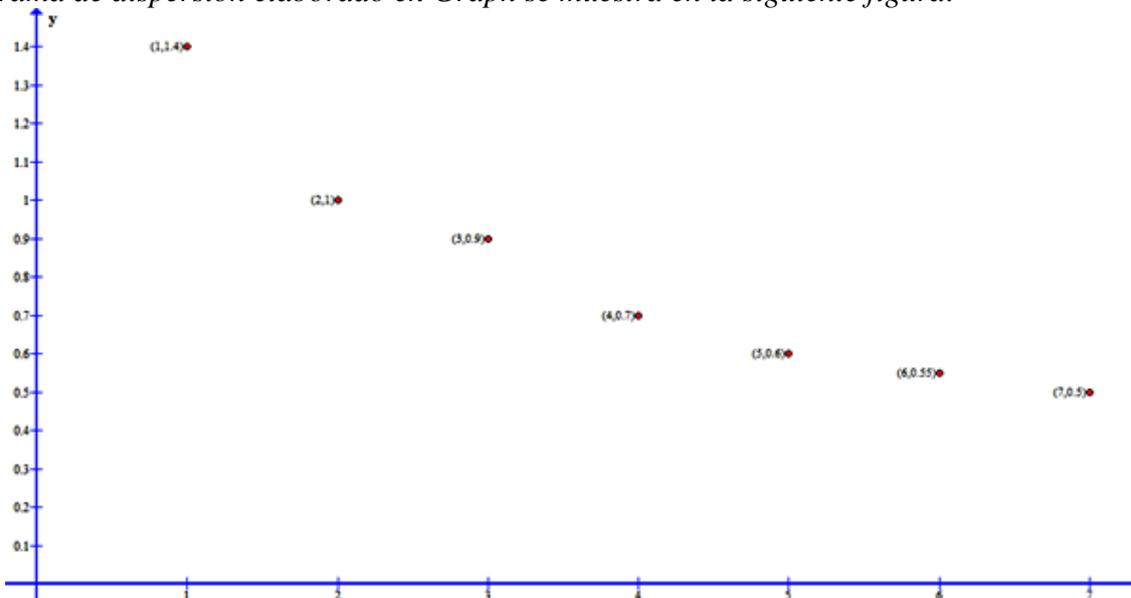
2.5) Estimar el valor de Y para $X = 9$

Solución:

2.1) El diagrama de dispersión elaborado en Excel se muestra en la siguiente figura:



El diagrama de dispersión elaborado en Graph se muestra en la siguiente figura:



2.2) Para calcular las constantes α y β , aplicando el método de mínimos cuadrados se llena la siguiente tabla:

X	Y	$1/Y$	$X(1/Y)$	X^2
1	1,4	0,7143	0,7143	1
2	1	1,0000	2,0000	4
3	0,9	1,1111	3,3333	9
4	0,7	1,4286	5,7143	16
5	0,6	1,6667	8,3333	25
6	0,55	1,8182	10,9091	36
7	0,5	2,0000	14,0000	49
$\Sigma X = 28$		$\Sigma (1/Y) = 9,7388$	$\Sigma X(1/Y) = 45,0043$	$\Sigma X^2 = 140$

Reemplazando valores en el siguiente sistema se obtiene:

$$\begin{cases} \Sigma \frac{1}{Y} = \alpha \cdot N + \beta \cdot \Sigma X \\ \Sigma X \cdot \frac{1}{Y} = \alpha \cdot \Sigma X + \beta \cdot \Sigma X^2 \end{cases}$$

$$\begin{cases} 9,7388 = \alpha \cdot 7 + \beta \cdot 28 \\ 45,0043 = \alpha \cdot 28 + \beta \cdot 140 \end{cases} \Rightarrow \begin{cases} 7\alpha + 28\beta = 9,7388 \\ 28\alpha + 140\beta = 45,0043 \end{cases}$$

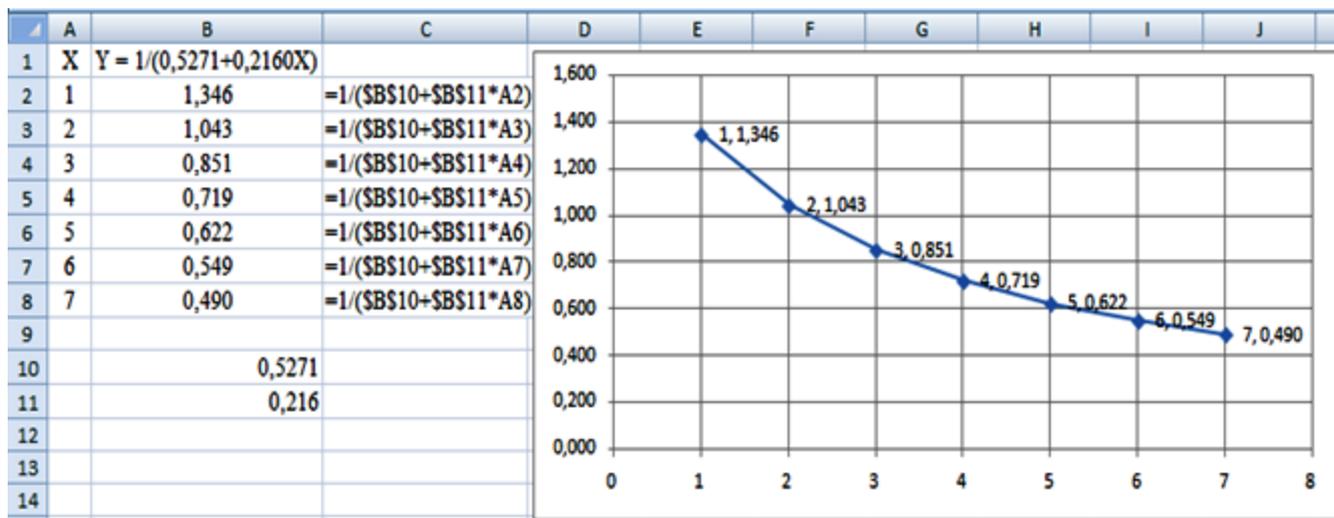
Al resolver el sistema se obtiene:

$$\alpha = 0,5271; \beta = 0,2160$$

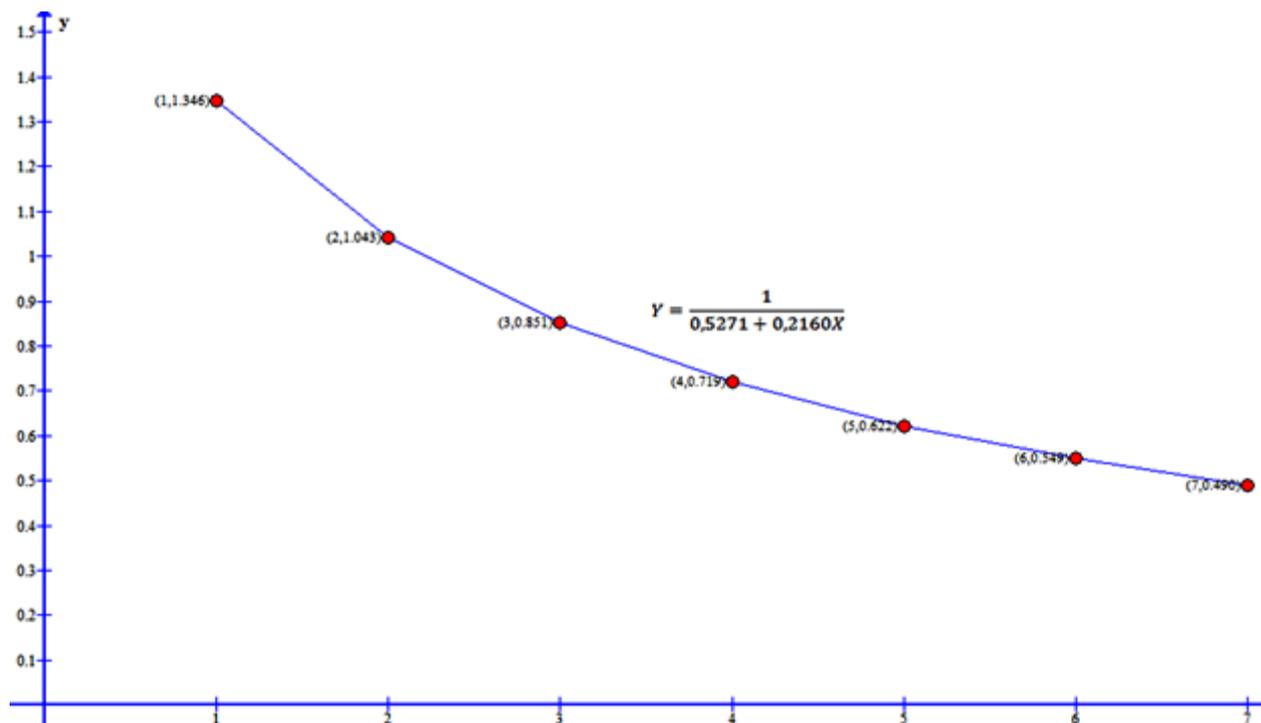
2.3) Para calcular la ecuación predictora se reemplaza los valores encontrados de α y β , y se obtiene:

$$Y = \frac{1}{\alpha + \beta \cdot X} \Rightarrow Y = \frac{1}{0,5271 + 0,2160X}$$

2.4) La gráfica la ecuación predictora elaborada en Excel se muestra en la siguiente figura:



La gráfica la ecuación predictora elaborada en Graph se muestra en la siguiente figura:



2.5) Para estimar el valor de Y para X = 9 se reemplaza el valor de X en la ecuación predictora.

$$Y = \frac{1}{0,5271 + 0,2160X}$$

$$Y = \frac{1}{0,5271 + 0,2160 \cdot 9} = 0,405$$

TAREA DE INTERAPRENDIZAJE N° 23

- 1) Elabore un organizador gráfico sobre la regresión potencial.
- 2) Sea el siguiente conjunto de valores, las lecturas de un experimento donde X es el volumen (variable independiente) e Y es la presión de una masa dada de gas (variable resultante).

X	1	2	3	4	5	6	7
Y	5	35	90	180	300	460	670

- 2.1) Elabore el diagrama de dispersión de manera manual, empleando Excel y Graph
- 2.2) Ajuste una curva exponencial aplicando el método de mínimos cuadrados empleando por lo menos 4 decimales para los cálculos.

$$\log Y = 0,7437 + 2,4883 \cdot \log X$$
- 2.3) Calcule la ecuación predictora.

$$Y = 5,5424 \cdot X^{2,4883}$$
- 2.4) Grafique la ecuación predictora de manera manual, empleando Excel y Graph.
- 2.5) Estime la presión de la masa de gas de volumen 8.

979,17

3) Cree y resuelva un ejercicio similar al anterior.

4) Sea el siguiente conjunto de valores, las lecturas de un experimento donde X es la variable independiente e Y la variable resultante.

X	1	2	3	4	5	6	7
Y	1,5	1	0,8	0,9	0,5	0,4	0,3

4.1) Elabore el diagrama de dispersión de manera manual, empleando Excel y Graph.

4.2) Calcule las constantes α y β , aplicando el método de mínimos cuadrados de manera manual y empleando Excel.

$$\alpha = 0,0159; \beta = 0,4196$$

4.3) Calcule la ecuación predictora.

$$Y = \frac{1}{0,0159 + 0,4196X}$$

4.4) Grafique la ecuación predictora de manera manual, empleando Excel y Graph.

4.5) Estime el valor de Y para X = 8

$$0,2965$$

5) Cree y resuelva un ejercicio similar al anterior de manera manual, empleando Excel y el programa Graph.

5.2.2) ERROR ESTÁNDAR DE ESTIMACIÓN

Es el grado de dispersión de los datos con respecto a la recta de regresión $Y = a_0 + a_1X$
El error estándar de estimación se calcula con la fórmula:

$$s_e = \sqrt{\frac{\sum(Y_i - Y_{est})^2}{N - 2}}$$

Donde:

Y_i = cada valor de Y

Y_{est} = valor estimado de Y a partir de la recta de regresión

N = número de datos

Nota: Como se puede observar, el error estándar de estimación es un cálculo de la desviación estándar de la muestra de datos con respecto a la recta de regresión, en la que Y_{est} sustituye a la media de la muestra, y con n-2 en el denominador en vez de n-1. La razón de que sea n-2, es debido a que se pierde 2 grados de libertad al calcular las 2 constantes, a_0 y a_1 en la recta de regresión.

Otras ecuaciones para calcular el error estándar de estimación son:

$$s_e = \sqrt{\frac{\sum Y^2 - a_0 \sum Y - a_1 \sum XY}{N - 2}} \quad s_e = \sqrt{\frac{\sum y^2 - a_1 \sum xy}{N - 2}}$$

Donde:

a_0 = ordenada en el origen (punto de intersección de la recta con el eje y)

a_1 = pendiente de la recta (tangente del ángulo de inclinación de la recta)

$x = X - \bar{X}$

$y = Y - \bar{Y}$

Ejemplo ilustrativo

Calcular error estándar de estimación empleando las 3 fórmulas dadas, utilizando los datos de la tabla del ejemplo para ajustar la recta de mínimos cuadrados para Y como variable dependiente.

X	Y
152	56
157	61
162	67
167	72
173	70
178	72
182	83
188	92

Solución:

Para comenzar a resolver este ejemplo recordemos que ya se obtuvo los valores respectivos al resolver el ejemplo para ajustar la recta de mínimos cuadrados, los cuales fueron:

$$\Sigma X = 1359; \Sigma Y = 573; \Sigma XY = 98295; \Sigma X^2 = 231967; \Sigma Y^2 = 41967; \Sigma xy = 956,625$$
$$\Sigma x^2 = 1106,875; \Sigma y^2 = 925,875; a_0 = -75,191; a_1 = 0,864; Y = -75,191 + 0,864X$$

1) Para emplear la primera fórmula se llena la siguiente tabla:

X	Y	$Y_{est} = 75,191 + 0,86X$	Y_{est}	$(Y - Y_{est})^2$
152	56	$-75,191+0,86(152)$	55,529	0,222
157	61	$-75,191+0,86(157)$	59,829	1,371
162	67	$-75,191+0,86(162)$	64,129	8,243
167	72	$-75,191+0,86(167)$	68,429	12,752
173	70	$-75,191+0,86(173)$	73,589	12,881
178	72	$-75,191+0,86(178)$	77,889	34,680
182	83	$-75,191+0,86(182)$	81,329	2,792
188	92	$-75,191+0,86(188)$	86,489	30,371
Σ				103,312

Se reemplaza valores en la primera fórmula se obtiene:

$$s_e = \sqrt{\frac{\Sigma(Y_i - Y_{est})^2}{N - 2}}$$

$$s_e = \sqrt{\frac{103,312}{8 - 2}}$$

$$s_e = 3,842$$

Realizando los cálculos de los componentes de la fórmula empleando Excel se obtiene un valor más exacto, ya que Excel utiliza una mayor cantidad de decimales al realizar los cálculos. Estos cálculos se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H
1		X	Y	XY	X ²	Y ²	$Y_{est} = a_0 + a_1X$	$(Y - Y_{est})^2$
2		152	56	8512	23104	3136	56,17639752	0,0311161
3		157	61	9577	24649	3721	60,49768492	0,2523204
4		162	67	10854	26244	4489	64,81897233	4,7568817
5		167	72	12024	27889	5184	69,14025974	8,1781144
6		173	70	12110	29929	4900	74,32580463	18,712586
7		178	72	12816	31684	5184	78,64709204	44,183833
8		182	83	15106	33124	6889	82,10412196	0,8025975
9		188	92	17296	35344	8464	87,28966685	22,187238
10	Σ	1359	573	98295	231967	41967		99,104687
11		N	8	=CONTAR(B2:B9)				
12								
13		$a_0 = \frac{\sum Y \cdot \sum X^2 - \sum X \cdot \sum XY}{N \sum X^2 - (\sum X)^2}$			-75,190740	=(C10*E10-B10*D10)/(C11*E10-(B10)^2)		
14								
15								
16								
17		$a_1 = \frac{N \sum YX - \sum X \cdot \sum Y}{N \sum X^2 - (\sum X)^2}$			0,8642575	=(C11*D10-B10*C10)/(C11*E10-(B10)^2)		
18								
19								
20		$s_e = \sqrt{\frac{\sum (Y_i - Y_{est})^2}{N - 2}}$			4,0641663	=RCUAD(H10/(C11-2))		
21								
22								

2) Reemplazando valores en la segunda fórmula se obtiene:

$$s_e = \sqrt{\frac{\sum Y^2 - a_0 \sum Y - a_1 \sum XY}{N - 2}}$$

$$s_e = \sqrt{\frac{41967 - (-75,191)(573) - 0,864(98295)}{8 - 2}}$$

$$s_e = \sqrt{\frac{41967 + 43084,443 - 84926,88}{6}}$$

$$s_e = \sqrt{\frac{124,563}{6}}$$

$$s_e = 4,556$$

Los cálculos de los componentes de la fórmula empleando Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H
1		X	Y	XY	X ²	Y ²		
2		152	56	8512	23104	3136		
3		157	61	9577	24649	3721		
4		162	67	10854	26244	4489		
5		167	72	12024	27889	5184		
6		173	70	12110	29929	4900		
7		178	72	12816	31684	5184		
8		182	83	15106	33124	6889		
9		188	92	17296	35344	8464		
10	Σ	1359	573	98295	231967	41967		
11		N	8	=CONTAR(B2:B9)				
12								
13		$a_0 = \frac{\sum Y \cdot \sum X^2 - \sum X \cdot \sum XY}{N \sum X^2 - (\sum X)^2}$			-75,190740	=(C10*E10-B10*D10)/(C11*E10-(B10)^2)		
14								
15								
16								
17		$a_1 = \frac{N \sum YX - \sum X \cdot \sum Y}{N \sum X^2 - (\sum X)^2}$			0,8642575	=(C11*D10-B10*C10)/(C11*E10-(B10)^2)		
18								
19								
20								
21		$s_e = \sqrt{\frac{\sum Y^2 - a_0 \sum Y - a_1 \sum XY}{N - 2}}$			4,0641663	=RCUAD((F10-E13*C10-E17*D10)/(C11-2))		
22								
23								

3) Reemplazando valores en la tercera fórmula se obtiene:

$$s_e = \sqrt{\frac{\sum y^2 - a_1 \sum xy}{N - 2}}$$

$$s_e = \sqrt{\frac{925,875 - 0,864(956,625)}{8 - 2}}$$

$$s_e = \sqrt{\frac{99,351}{6}}$$

$$s_e = 4,069$$

Los cálculos de los componentes de la fórmula empleando Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H	I	J	K
1		X	Y	XY	X ²	Y ²	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x ²	y ²
2		152	56	8512	23104	3136	-17,875	-15,625	279,297	319,516	244,141
3		157	61	9577	24649	3721	-12,875	-10,625	136,797	165,766	112,891
4		162	67	10854	26244	4489	-7,875	-4,625	36,422	62,016	21,391
5		167	72	12024	27889	5184	-2,875	0,375	-1,078	8,266	0,141
6		173	70	12110	29929	4900	3,125	-1,625	-5,078	9,766	2,641
7		178	72	12816	31684	5184	8,125	0,375	3,047	66,016	0,141
8		182	83	15106	33124	6889	12,125	11,375	137,922	147,016	129,391
9		188	92	17296	35344	8464	18,125	20,375	369,297	328,516	415,141
10	Σ	1359	573	98295	231967	41967			956,625	1106,875	925,875
11		N	8								
12											
13	$a_1 = \frac{N\sum YX - \sum X \cdot \sum Y}{N\sum X^2 - (\sum X)^2}$				0,8642575			\bar{X}	169,875	=PROMEDIO(B2:B9)	
14								\bar{Y}	71,625	=PROMEDIO(C2:C9)	
15											
16	$s_e = \sqrt{\frac{\sum y^2 - a_1 \sum xy}{N - 2}}$				4,0641663						
17											

Empleando exclusivamente Excel para calcular el error estándar de estimación se procede de la siguiente manera:

a) Se inserta la función ERROR.TIPICO.XY

The screenshot shows the Excel interface with the 'ERROR.TIPICO.XY' function dialog box open. The dialog box contains the following text and fields:

- Function name: ERROR.TIPICO.XY
- Conocido_y: =matriz
- Conocido_x: =matriz
- Description: Devuelve el error típico del valor de Y previsto para cada X de la regresión.
- Additional info: Conocido_y es una matriz, o rango de puntos de datos dependientes, formada por: números, nombres, matrices o referencias que contengan números.
- Buttons: Ayuda sobre esta función, Aceptar, Cancelar

b) Se selecciona las celdas respectivas.

ERROR.TIPICO.XY

Conocido_y B2:B9 = {56|61|67|72|70|72|83|92}

Conocido_x A2:A9 = {152|157|162|167|173|178|182|188}

= 4,064166307

Devuelve el error típico del valor de Y previsto para cada X de la regresión.

Conocido_x es una matriz, o rango de puntos de datos independientes, formada por: números, nombres, matrices o referencias que contengan números.

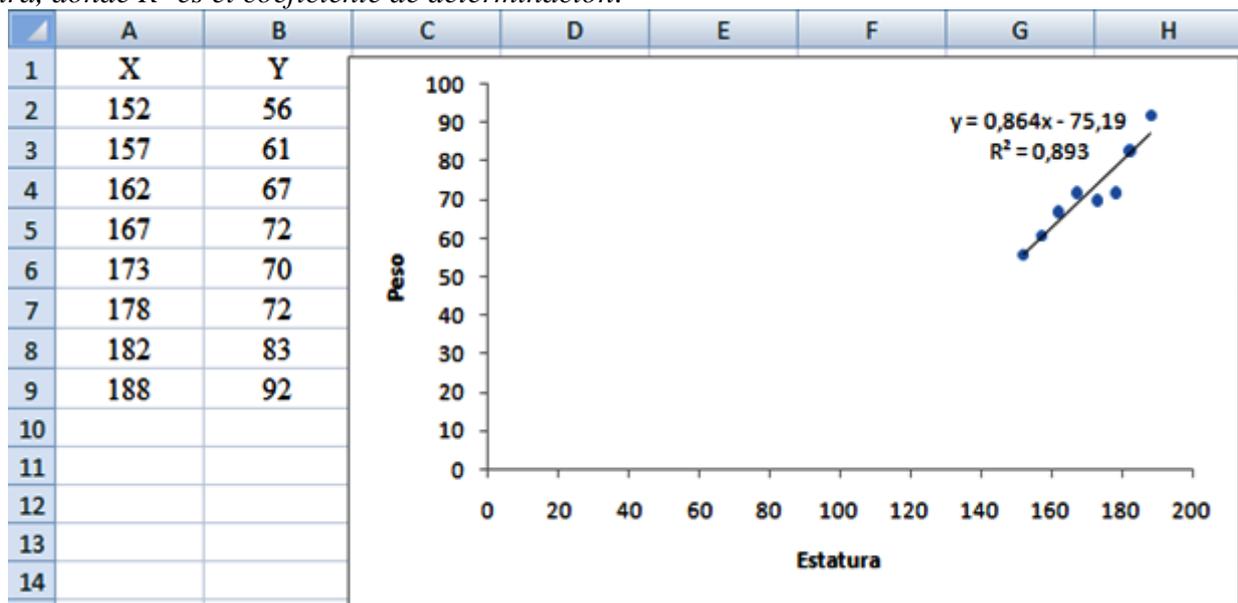
Resultado de la fórmula = 4,064166307

Aceptar Cancelar

c) Pulsar en Aceptar.

	A	B	C	D
1	X	Y		
2	152	56		
3	157	61		
4	162	67		
5	167	72		
6	173	70		
7	178	72		
8	182	83		
9	188	92		
10				
11	4,06416631	=ERROR.TIPICO.XY(B2:B9;A2:A9)		

El diagrama dispersión y la recta de mínimos cuadrados elaborado en Excel se muestra en la siguiente figura, donde R^2 es el coeficiente de determinación:



Interpretación: El valor de $s_e = 4,064$, significa que los puntos están dispersos a una distancia de 4,064 de la recta de regresión.

TAREA DE INTERAPRENDIZAJE N° 24

Dada la siguiente tabla sobre la altura en centímetros (X) y los pesos en kilogramos (Y) de una muestra de 8 estudiantes varones tomada al azar del segundo semestre de una universidad.

X	150	155	160	165	170	175	180	185
Y	56	61	64	68	72	75	80	90

- 1) Calcule el coeficiente de determinación de manera manual y empleando Excel. 0,97

- 2) Calcule el error estándar de estimación empleando la primera fórmula. Utilice 5 decimales para los cálculos. Los elementos de la fórmula calcule empleando Excel, tal como se indicó en el ejemplo. 2,1

- 3) Calcule el error estándar de estimación empleando la segunda fórmula. Utilice 5 decimales para los cálculos. Los elementos de la fórmula calcule empleando Excel, tal como se indicó en el ejemplo. 2,1

- 4) Calcule el error estándar de estimación empleando la tercera fórmula. Utilice 5 decimales para los cálculos. Los elementos de la fórmula calcule empleando Excel, tal como se indicó en el ejemplo. 2,1

- 5) Calcule el error estándar de estimación empleando exclusivamente Excel. 2,1

- 6) Elabore el diagrama de dispersión, y en el mismo diagrama graficar la recta de regresión. Realice de manera manual, empleando Excel y Graph.

CAPÍTULO VI

SERIES DE TIEMPO

COMPETENCIAS:

- ✓ Analiza adecuadamente el uso de los datos de series de tiempo y su aplicación para hacer pronósticos.
- ✓ Emplea correctamente algoritmos matemáticos para resolver ejercicios de aplicación sobre series de tiempo de manera manual y empleando Excel.
- ✓ Elabora eficientemente diagramas de dispersión y líneas de tendencias de manera manual, empleando Excel y Graph.
- ✓ Crea y resuelve correctamente ejercicios de aplicación sobre series de tiempo.

CONTENIDOS:

- ✓ Definición
- ✓ Movimientos o Componentes
- ✓ Modelos de Series de Tiempo
- ✓ Métodos de Suavizamiento y Pronóstico
- ✓ Análisis de Tendencia
- ✓ Análisis de Movimientos Estacionales
- ✓ Análisis de Movimientos Cíclicos e Irregulares

6.1) DEFINICIÓN

Las series de tiempo llamadas también series cronológicas o series históricas son un conjunto de datos numéricos que se obtienen en períodos regulares y específicos a través del tiempo, los tiempos pueden ser en años, meses, semanas, días u otra unidad adecuada al problema que se esté trabajando. Ejemplos de series de tiempo son: Ventas mensuales de un producto en una empresa, producción total anual de petróleo en Ecuador durante un cierto número años o las temperaturas anunciadas cada hora por el meteorólogo para un aeropuerto.

Matemáticamente, una serie de tiempo se define por los valores Y_1, Y_2, Y_3, \dots de una variable Y (ventas mensuales, producción total, etc.) en tiempos t_1, t_2, t_3, \dots . Si se reemplaza a X por la variable tiempo, estas series se definen como distribuciones de pares ordenados (X, Y) en el plano cartesiano, siendo Y una función de X ; esto se denota por:

$$Y = f(t) \rightarrow Y = f(X)$$

El principal objetivo de las series de tiempo es hacer proyecciones o pronósticos sobre una actividad futura, suponiendo estables las condiciones y variaciones registradas hasta la fecha, lo cual permite planear y tomar decisiones a corto o largo plazo. Después, con base en esa situación ideal, que supone que los factores que influyeron en la serie en el pasado lo continuarán haciendo en el futuro, se analizan las tendencias pasadas y el comportamiento de las actividades bajo la influencia de ellas; por ejemplo, en la proyección de ventas de un producto o de un servicio de una empresa se calculan los posibles precios, la reacción del consumidor, la influencia de la competencia, etc.

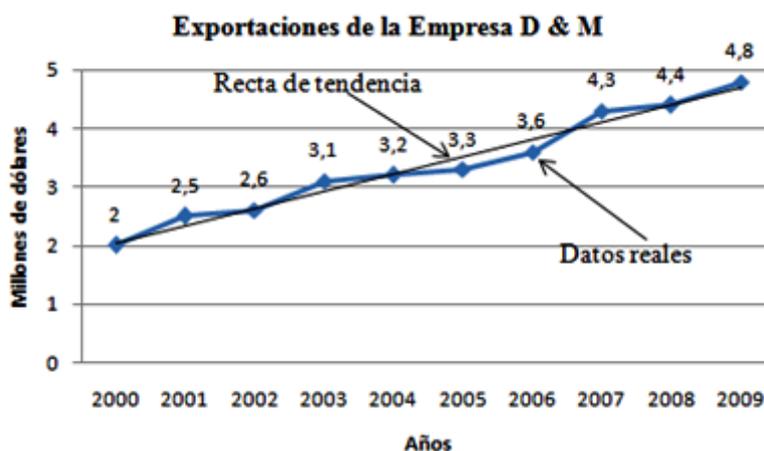
6.2) MOVIMIENTOS O COMPONENTES

El modelo clásico o de descomposición, considera que los datos de series de tiempo están compuestas de los siguientes cuatro patrones básicos:

6.2.1) TENDENCIA SECULAR

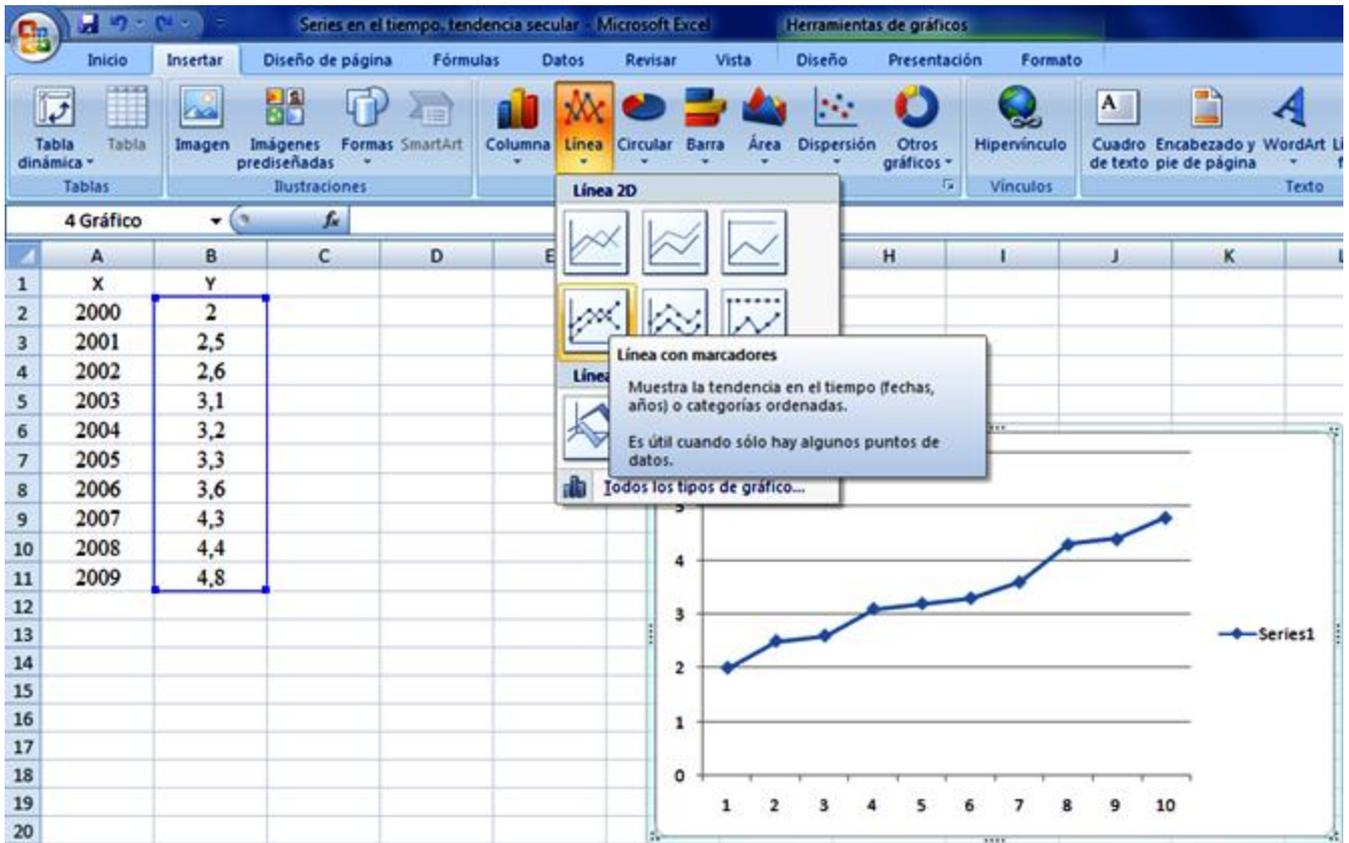
La tendencia secular o simplemente tendencia, son movimientos o variaciones continuas de la variable de modo uniforme y suave, por encima o por debajo, que se observan en el largo plazo durante un período de longitud prolongada. Representan el comportamiento predominante o dirección general de la serie de tiempo como ascendente o descendente. La gráfica de la tendencia suele ser una curva suave y aun una línea recta que muestra la tendencia de las variaciones. Ejemplos de tendencia secular son las ventas, exportaciones, producción y el empleo.

La siguiente gráfica muestra la tendencia de exportaciones de la Empresa D & M en período 2000-2009. Aunque los datos muestran ciertas variaciones están por encima y por debajo de la recta de tendencia, la tendencia secular es ascendente.

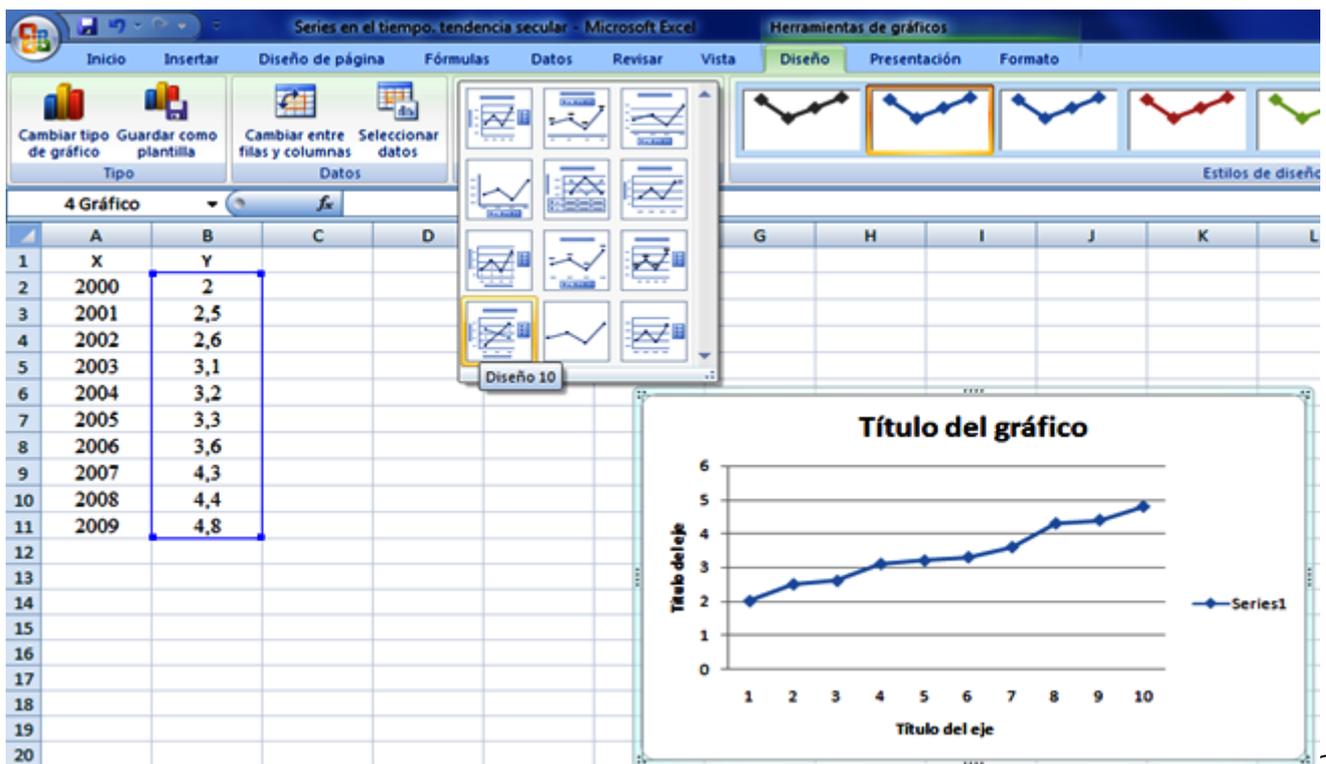


Empleando Excel para realizar la gráfica anterior se procede de la siguiente manera:

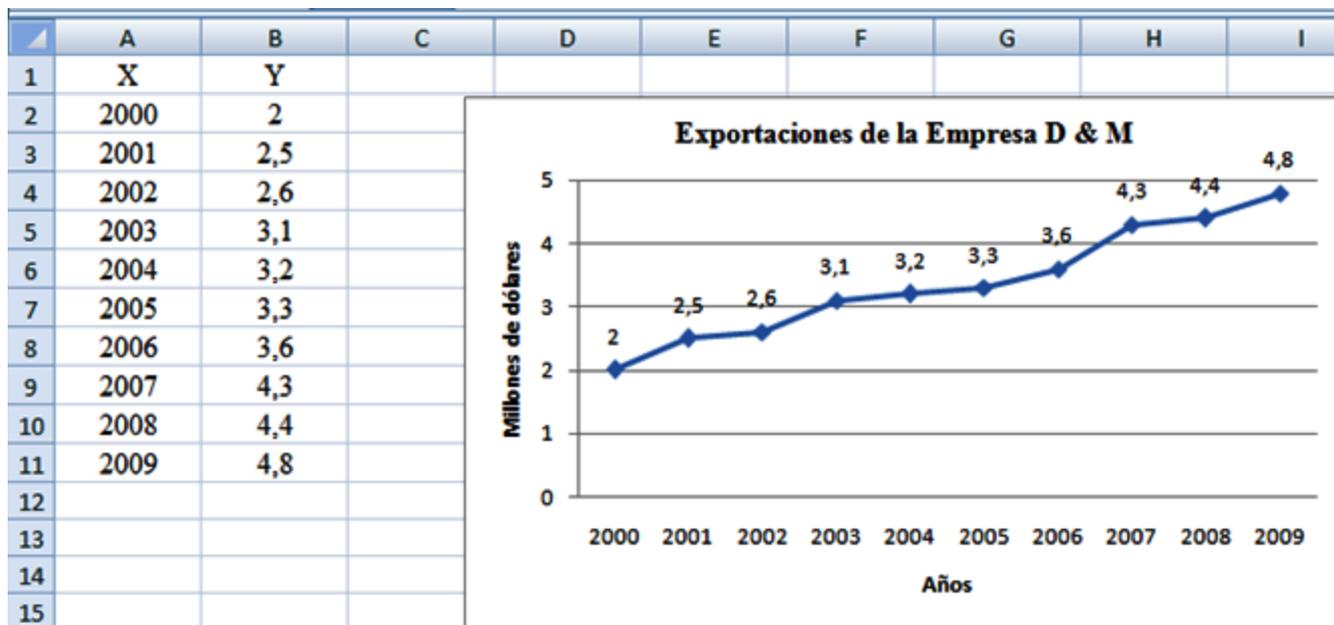
a) Escribir los valores de X y Y en la hoja de cálculo. Seleccionar B2:B13. Clic en Insertar Gráfico de Línea con marcadores.



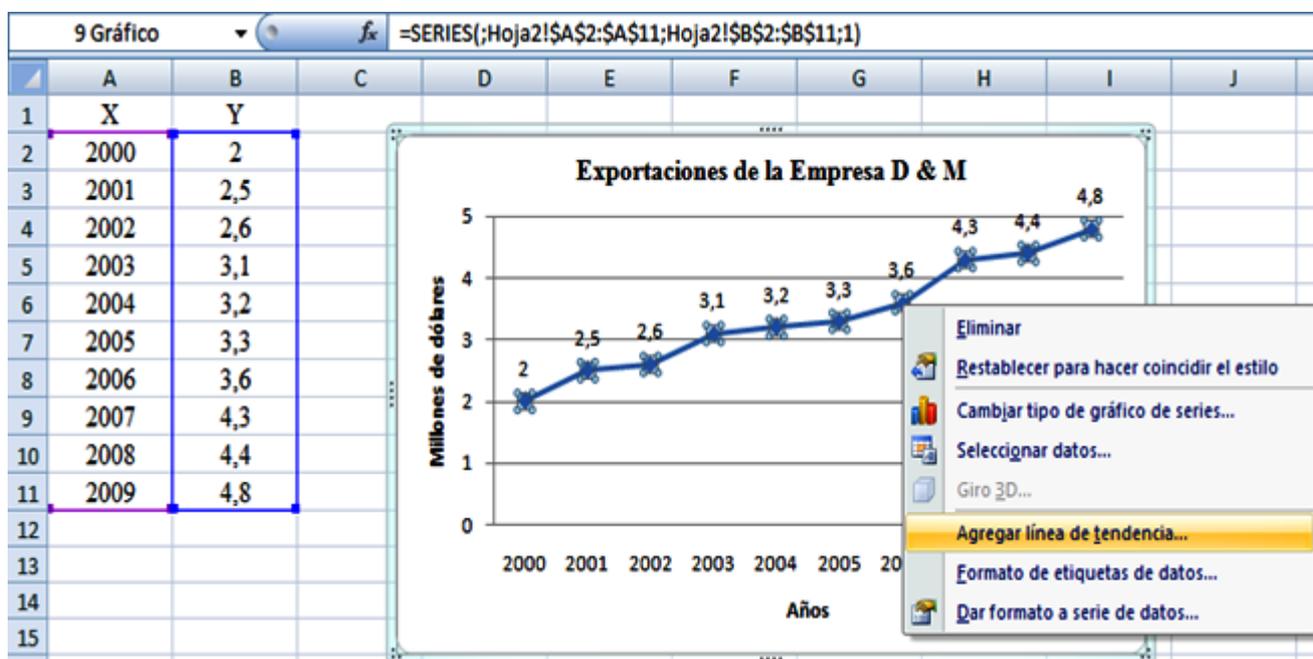
b) Seleccionar el gráfico. Clic en Diseño 10 de Diseños de Gráfico.



c) Borrar el texto Series 1 del gráfico. Escribir Exportaciones de la Empresa D & M en título del gráfico. Escribir Años en título del eje horizontal. Escribir Millones de dólares en título del eje vertical. Poner los años de las celdas A2:A11 en los números del eje horizontal. Agregar etiquetas. Dar formato al eje vertical con un máximo de 5, tal como ya se indicó en capítulos anteriores.

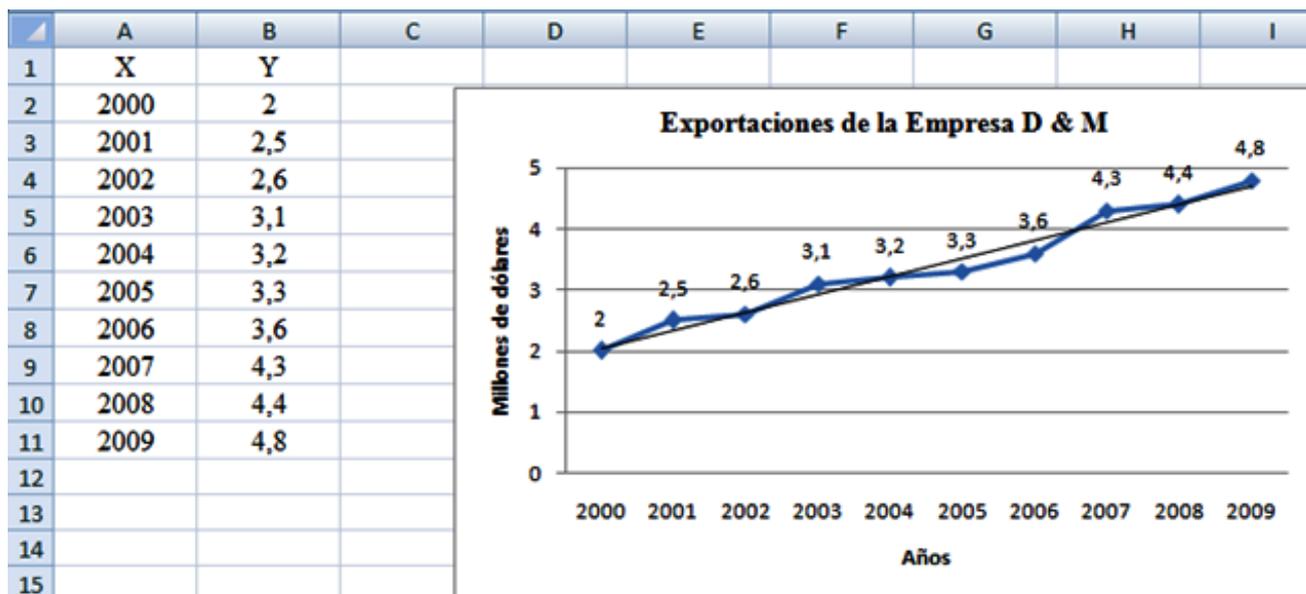


d) Clic derecho en serie de datos (en cualquier punto).



e) Clic en Agregar línea de tendencia. Elegir Lineal en Opciones de línea de tendencia.

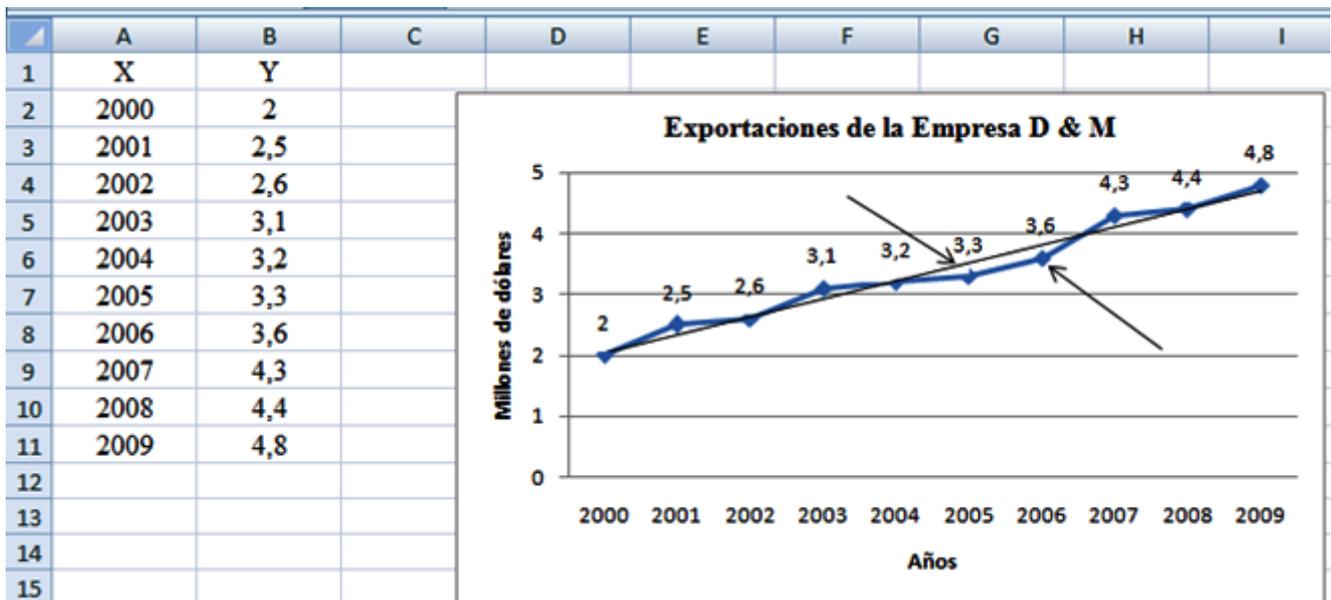
f) Clic en Cerrar en la ventana de Formato de línea de tendencia.



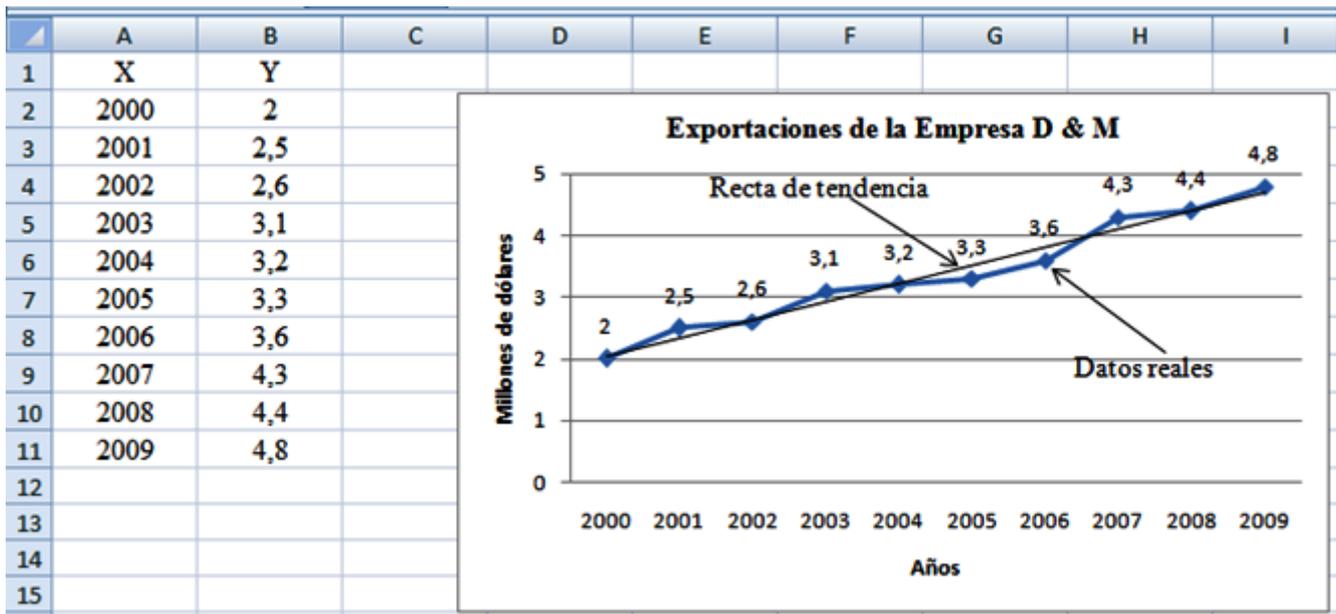
g) Clic en insertar Formas.

	A	B	C
1	X	Y	
2	2000	2	
3	2001	2,5	
4	2002	2,6	
5	2003	3,1	
6	2004	3,2	
7	2005	3,3	
8	2006	3,6	
9	2007	4,3	
10	2008	4,4	
11	2009	4,8	

h) Seleccionar Flecha y ubicar en el gráfico en dos ocasiones.



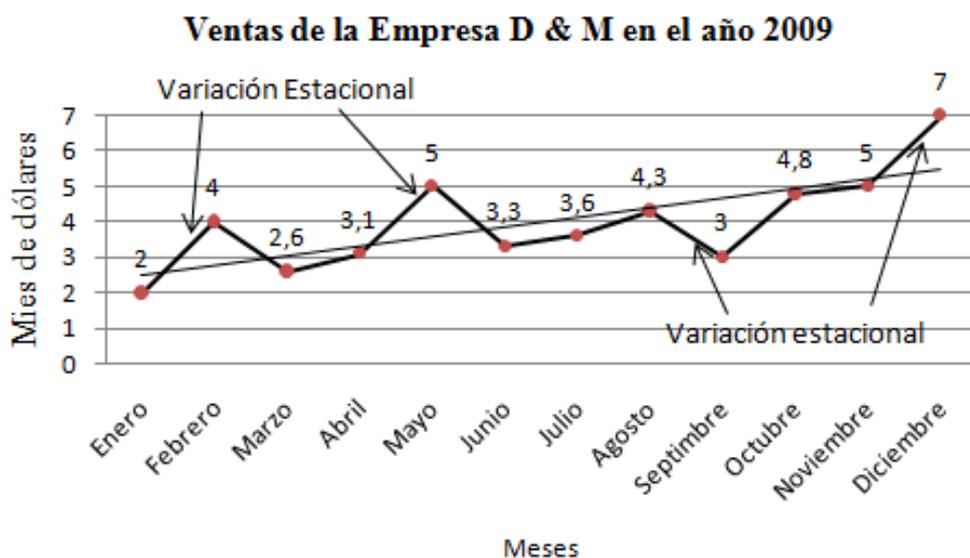
i) Insertar Cuadro de texto. Escribir Recta de tendencia en un cuadro de texto, y en otro escribir Datos reales.



6.2.2) MOVIMIENTOS ESTACIONALES

Representa un movimiento periódico que se producen en forma similar cada año por la misma época, en correlación con los meses o con las estaciones del año y aun con determinadas fechas. Si los sucesos no se repiten anualmente, los datos deben recolectarse trimestral, mensual o incluso semanalmente. Ejemplos de movimientos estacionales son la variación de precios de ciertos productos, incremento de ventas de juguetes y disminución de ventas de útiles Navidad, incremento de ventas de flores por el día del amor y la amistad, etc.

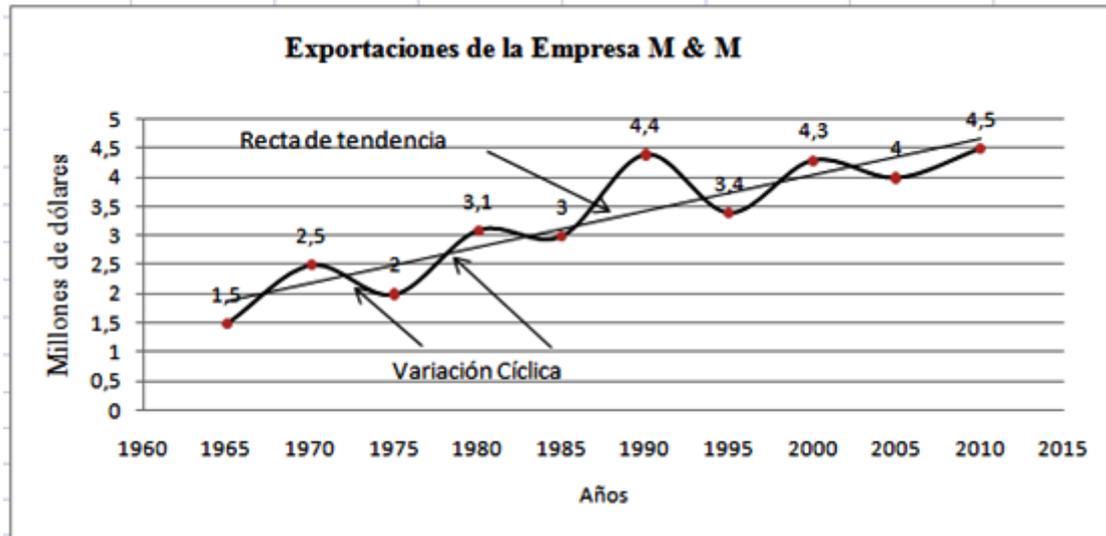
A continuación se muestra un ejemplo de gráfica que representa este tipo de movimientos estacionales:



6.2.3) MOVIMIENTOS CÍCLICOS

Son variaciones hacia arriba y hacia abajo de la tendencia que se presentan cada cierto número de intervalos, en forma periódica de manera ondular a modo de oscilaciones más o menos regulares durante un período relativamente prolongado, que por lo general abarca tres o más años de duración. La producción, empleo, promedio industrial, etc. son ejemplos de este tipo de movimientos.

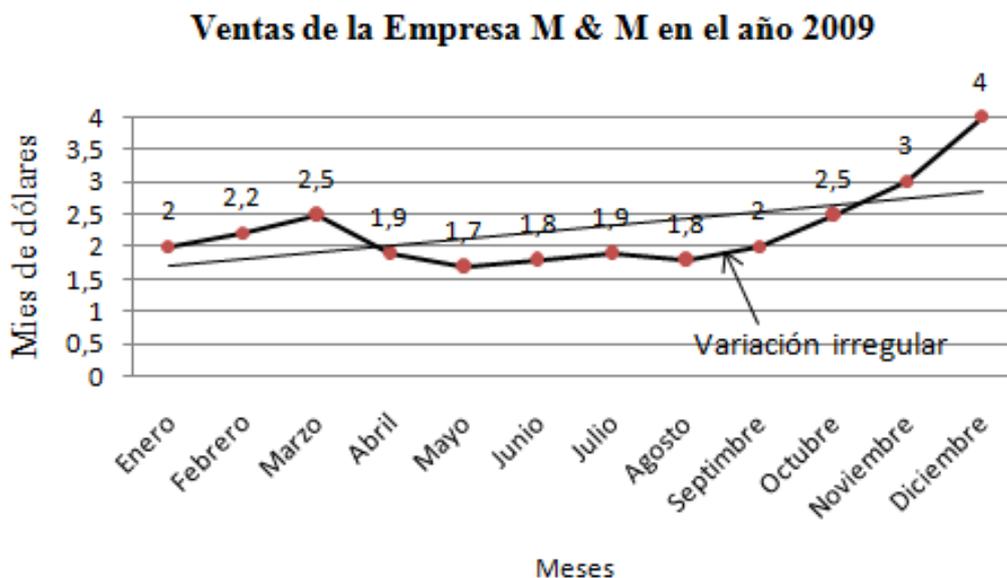
A continuación se muestra un ejemplo de gráfica que representa este tipo de movimientos cíclicos:



6.2.4) MOVIMIENTOS IRREGULARES O ALEATORIOS

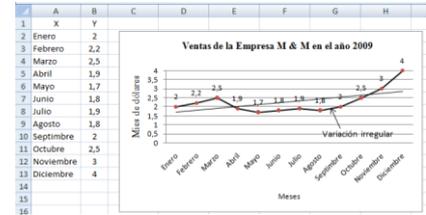
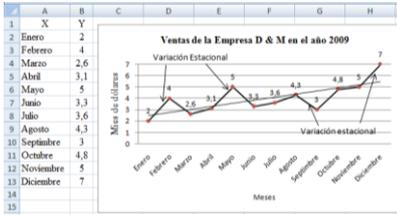
Son aquellas variaciones producidas por sucesos de ocurrencia imprevisible o accidental que producen movimientos sin un patrón discernible; así por ejemplo, las exportaciones de una empresa pueden ser afectadas por sucesos inusuales no previsible tales como huelgas, guerras, terremotos, inundaciones, etc. Estas variaciones irregulares son de corta duración y de magnitud muy variable.

A continuación se muestra un ejemplo de gráfica que representa este tipo de movimientos irregulares:



TAREA DE INTERAPRENDIZAJE N° 25

- 1) Realice un organizador gráfico sobre las series de tiempo.
- 2) Elabore empleando Excel las gráficas de los ejemplos presentados en los movimientos estacionales, cíclicos e irregulares.



- 3) Cree y elabore una gráfica que represente a cada uno de los movimientos de las series de tiempo de manera manual y empleando Excel.

6.3) MODELOS DE SERIES DE TIEMPO

Son expresiones matemáticas de relación entre los movimientos de tendencia secular (T), movimientos cíclicos (C), movimientos estacionales (E) y movimientos irregulares (I) que generan la variable Y. Hay dos modelos para la definición de Y, los cuales son:

6.3.1) MODELO MULTIPLICATIVO

En el que Y queda definida por el producto de las variaciones.

$$Y = T \cdot C \cdot E \cdot I$$

6.3.2) MODELO ADITIVO

En el que Y queda definida por la suma de las variaciones.

$$Y = T + C + E + I$$

En el modelo multiplicativo, las variaciones se expresan en términos relativos o porcentuales de la tendencia, en tanto que en el modelo aditivo las variaciones se expresan como residuos en las mismas unidades originales. El modelo aditivo sufre el supuesto irreal de que los movimientos o componentes son independientes uno de otro, algo que difícilmente se da en el caso de la vida real. El modelo multiplicativo supone que los movimientos o componentes interactúan entre sí y no se mueven independientemente, por lo que este modelo es más utilizado que el aditivo. Sin embargo, el criterio fundamental que se debe seguir en el caso de una situación dada es emplear el modelo que mejor se ajuste a los datos.

6.4) MÉTODOS DE SUAVIZAMIENTO Y PRONÓSTICO

Estos métodos eliminan las fluctuaciones aleatorias de la serie de tiempo, proporcionando datos menos distorsionados del comportamiento real de misma.

6.4.1) MÉTODO DE LOS PROMEDIOS MÓVILES

El movimiento medio de orden N de una serie de valores Y_1, Y_2, Y_3, Y_n se define por la sucesión de valores correspondientes a las medias aritméticas:

$$\frac{Y_1 + Y_2 + \dots + Y_N}{N}; \frac{Y_2 + Y_3 + \dots + Y_{N+1}}{N}; \frac{Y_3 + Y_4 + \dots + Y_{N+2}}{N}; \dots$$

Por ejemplo: Dados los valores 4, 6, 8, 10, 12 tendríamos para el movimiento medio de orden 2

$$\frac{4 + 6}{2}; \frac{6 + 8}{2}; \frac{8 + 10}{2}; \frac{10 + 12}{2}$$

O sea los valores 5; 7; 9; 11

Para el movimiento medio de orden 3 se tiene la serie

$$\frac{4 + 6 + 8}{3}; \frac{6 + 8 + 10}{3}; \frac{8 + 10 + 12}{3}$$

O sea los valores 6; 8; 10

Para el movimiento de orden 4

$$\frac{4 + 6 + 8 + 10}{4}; \frac{6 + 8 + 10 + 12}{4}$$

O sea los valores 7, 12

Nota:

Utilizando adecuadamente estos movimientos medios se eliminan los movimientos o variaciones estacionales, cíclicas e irregulares, quedando sólo el movimiento de tendencia. Este método presenta el inconveniente de que se pierden datos iniciales y finales de la serie original. También se puede observar que a medida que N crece, la cantidad de nuevos datos se reduce.

*Si se emplean medias aritméticas ponderadas en el método de los promedios móviles, el método toma de nombre **Promedios Móviles Ponderados de Orden N**.*

Ejemplo ilustrativo

Con los siguientes datos acerca de la ventas en miles de dólares de la Empresa D & M durante los últimos 3 años tomados en períodos de trimestres:

Trimestre	Ventas
1	12
2	16
3	20
4	34
5	23
6	19
7	20
8	35
9	11
10	19
11	24
12	36

- 1) Suavizar los datos empleando el método de los promedios móviles de orden 3 (longitud de 3 períodos).
- 2) Pronosticar las ventas para el trimestre número 13.
- 3) Suponga que para el Gerente de Ventas la última venta realizada es el doble de importante que la penúltima, y la antepenúltima venta tiene la mitad de importancia que la penúltima. Realizar el pronóstico de ventas para el trimestre número 13 empleando el método de los promedios móviles ponderados de orden 3.
- 4) Elaborar un gráfico en el que consten las ventas y los promedios móviles (ventas suavizadas).

Solución:

- 1) El cálculo de los promedios móviles de orden 3 se presentan en la siguiente tabla:

Trimestre	Ventas	Pronóstico (Promedios móviles)
1	12	
2	16	$(12+16+20)/3 = 16,00$
3	20	$(16+20+34)/3 = 23,33$
4	34	$(20+34+23)/3 = 25,67$
5	23	$(34+23+19)/3 = 25,33$
6	19	$(23+19+20)/3 = 20,67$
7	20	$(19+20+35)/3 = 24,67$
8	35	$(20+35+11)/3 = 22,00$
9	11	$(35+11+19)/3 = 21,67$
10	19	$(11+19+24)/3 = 18,00$
11	24	$(19+24+36)/3 = 26,33$
12	36	

Empleando Excel se muestra en la siguiente figura:

	A	B	C	D
1	Trimestre	Ventas	Pronóstico (Promedios móviles)	
2	1	12		
3	2	16	16,00	=PROMEDIO(B2:B4)
4	3	20	23,33	=PROMEDIO(B3:B5)
5	4	34	25,67	=PROMEDIO(B4:B6)
6	5	23	25,33	=PROMEDIO(B5:B7)
7	6	19	20,67	=PROMEDIO(B6:B8)
8	7	20	24,67	=PROMEDIO(B7:B9)
9	8	35	22,00	=PROMEDIO(B8:B10)
10	9	11	21,67	=PROMEDIO(B9:B11)
11	10	19	18,00	=PROMEDIO(B10:B12)
12	11	24	26,33	=PROMEDIO(B11:B13)
13	12	36		

2) El último valor del promedio móvil, que en este ejemplo es 26,33, representa el pronóstico de las ventas para el trimestre número 13, y teóricamente para todo trimestre futuro.

3) Para resolver lo planteado se toma en cuenta las 3 últimas ventas con sus respectivos pesos o ponderaciones. Estos datos se presentan en la siguiente tabla:

Trimestre	Ventas	Pesos (w)
10	19	0,5
11	24	1
12	36	2

Reemplazando valores en la fórmula de la media aritmética ponderada se obtiene:

$$\text{Pronóstico} = \bar{x} = \frac{w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + \dots + w_k \cdot x_k}{w_1 + w_2 + w_3 + \dots + w_k} = \frac{\sum w \cdot x}{\sum w}$$

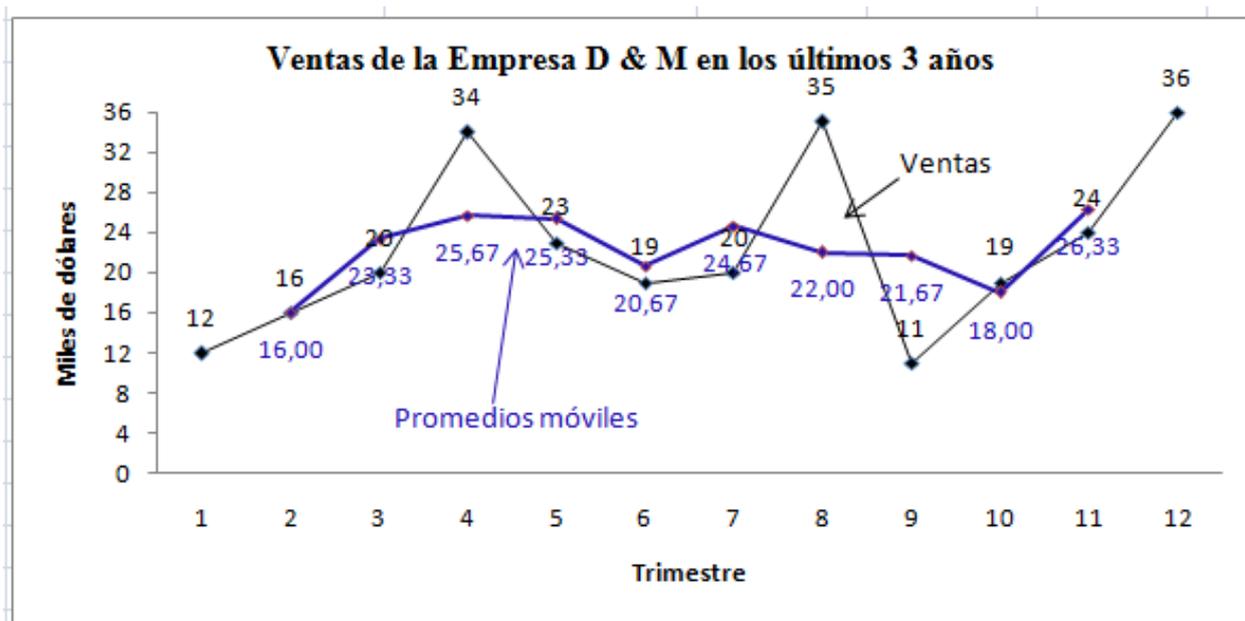
$$\text{Pronóstico} = \frac{0,5 \cdot 19 + 1 \cdot 24 + 2 \cdot 36}{0,5 + 1 + 2} = \frac{105,5}{3,5} = 30,14$$

El valor 30,14 es el pronóstico de ventas para el trimestre número 13.

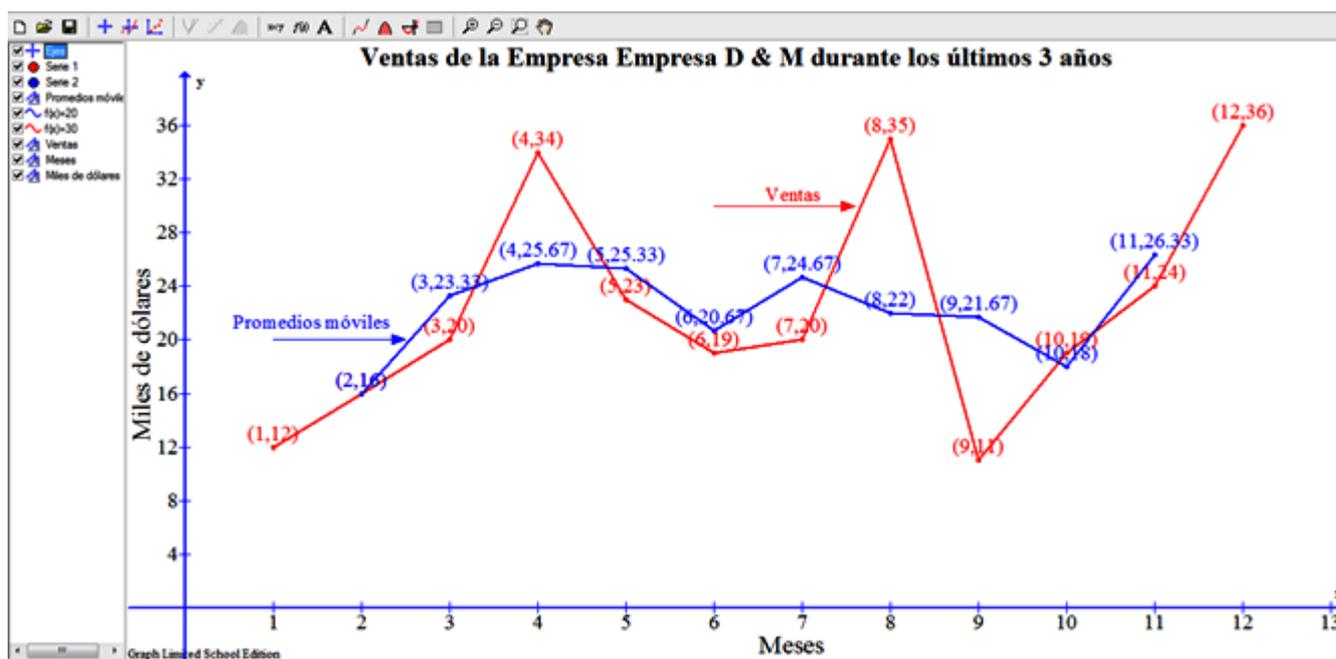
Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F
1	Trimestre	Ventas	Pesos (w)			
2	10	19	0,5			
3	11	24	1			
4	12	36	2			
5						
6	Pronóstico	30,143	=SUMAPRODUCTO(C2:C4;B2:B4)/SUMA(C2:C4)			

4) El gráfico en el que constan las ventas y los promedios móviles se muestra en la siguiente figura elaborado empleando Excel:



Empleando Graph se muestra en la siguiente figura:



6.4.2) SUAVIZACIÓN EXPONENCIAL

Este método contiene un mecanismo de autocorrección que ajusta los pronósticos en dirección opuesta a los errores pasados. Es un caso particular de promedios móviles ponderados de los valores actuales y anteriores en el cual las ponderaciones disminuyen exponencialmente. Se emplea tanto para suavizar como para realizar pronósticos. Se emplea la siguiente fórmula:

$$Y_{t+1} = \alpha \cdot X_t + (1 - \alpha) \cdot Y_t$$

Donde:

Y_{t+1} = pronóstico para cualquier período futuro.

α = constante de suavización, a la cual se le da un valor entre 0 y 1.

X_t = valor real para el período de tiempo.

Y_t = pronóstico hecho previamente para el período de tiempo

Cuando exista menos dispersión en los datos reales respecto a los datos pronosticados entonces será más confiable el método empleado. Para saber cuan preciso es el método empleado en la realización del pronóstico se utiliza la siguiente fórmula del **cuadrado medio del error (CME)** como *indicador de precisión del pronóstico*:

$$CME = \frac{\sum(Y_t - X_t)^2}{n}$$

Siendo n el número de errores

Ejemplo ilustrativo

Con los siguientes datos acerca de la ventas en miles de dólares de la Empresa D & M durante los últimos 12 meses:

Meses	Ventas
Septiembre	6
Octubre	7
Noviembre	6
Diciembre	12
Enero	7
Febrero	10
Marzo	6
Abril	4
Mayo	9
Junio	7
Julio	8
Agosto	6

- 1) Suavizar los datos empleando el método de suavización exponencial con $\alpha = 0,5$. Pronosticar las ventas para el mes de septiembre. Calcular el cuadrado medio del error. Elaborar un gráfico en el que consten las ventas y los pronósticos.
- 2) Suavizar los datos empleando el método de los promedios móviles de orden 3. Pronosticar las ventas para mes de septiembre. Calcular el cuadrado medio del error. Elaborar un gráfico en el que consten las ventas y los promedios móviles.
- 3) ¿Qué método es el más preciso?

Solución:

1) Realizando los cálculos de suavizamiento se obtienen los resultados respectivos de pronóstico, los cuales se presentan en la siguiente tabla:

Meses	Ventas (X_t)	Pronóstico con $\alpha = 0,5$ $Y_{t+1} = \alpha \cdot X_t + (1 - \alpha) \cdot Y_t$		
Septiembre	6			
Octubre	7	$Y_{Oct.} = X_{sep.} = 6$		
Noviembre	6	$Y_{Nov.} = \alpha \cdot X_{Oct.} + (1 - \alpha) \cdot Y_{Oct.}$	$= 0,5 \cdot 7 + (1 - 0,5) \cdot 6$	$= 6,5$
Diciembre	12	$Y_{Dic.} = \alpha \cdot X_{Nov.} + (1 - \alpha) \cdot Y_{Nov.}$	$= 0,5 \cdot 6 + (1 - 0,5) \cdot 6,5$	$= 6,25$
Enero	7	$Y_{Ene.} = \alpha \cdot X_{Dic.} + (1 - \alpha) \cdot Y_{Dic.}$	$= 0,5 \cdot 12 + (1 - 0,5) \cdot 6,25$	$= 9,125$
Febrero	10	$Y_{Feb.} = \alpha \cdot X_{Ene.} + (1 - \alpha) \cdot Y_{Ene.}$	$= 0,5 \cdot 7 + (1 - 0,5) \cdot 9,125$	$= 8,063$
Marzo	6	$Y_{Mar.} = \alpha \cdot X_{Feb.} + (1 - \alpha) \cdot Y_{Feb.}$	$= 0,5 \cdot 10 + (1 - 0,5) \cdot 8,063$	$= 9,032$
Abril	4	$Y_{Abr.} = \alpha \cdot X_{Mar.} + (1 - \alpha) \cdot Y_{Mar.}$	$= 0,5 \cdot 6 + (1 - 0,5) \cdot 9,032$	$= 7,516$
Mayo	9	$Y_{May.} = \alpha \cdot X_{Abr.} + (1 - \alpha) \cdot Y_{Abr.}$	$= 0,5 \cdot 4 + (1 - 0,5) \cdot 7,516$	$= 5,758$
Junio	7	$Y_{Jun.} = \alpha \cdot X_{May.} + (1 - \alpha) \cdot Y_{May.}$	$= 0,5 \cdot 9 + (1 - 0,5) \cdot 5,758$	$= 7,379$
Julio	8	$Y_{Jul.} = \alpha \cdot X_{Jun.} + (1 - \alpha) \cdot Y_{Jun.}$	$= 0,5 \cdot 7 + (1 - 0,5) \cdot 7,379$	$= 7,189$
Agosto	6	$Y_{Agt.} = \alpha \cdot X_{Jul.} + (1 - \alpha) \cdot Y_{Jul.}$	$= 0,5 \cdot 8 + (1 - 0,5) \cdot 7,189$	$= 7,595$
		$Y_{t+1} = \alpha \cdot X_{Agt.} + (1 - \alpha) \cdot Y_{Agt.}$	$= 0,5 \cdot 6 + (1 - 0,5) \cdot 7,595$	$= 6,798$

Observando la tabla anterior se tiene que el pronóstico de ventas para el mes de septiembre es de 6,798, o para cualquier período futuro, ya que los datos no presentan una tendencia sino que se supone que varían o fluctúan a largo plazo alrededor de este valor promedio.

Calculando el cuadrado medio del error se obtienen los siguientes resultados, los cuales se presentan en la siguiente tabla:

Meses	Ventas (X_t)	Pronóstico Y_t	Error $(Y_t - X_t)^2$
Septiembre	6		
Octubre	7	6	1
Noviembre	6	6,5	0,25
Diciembre	12	6,25	33,063
Enero	7	9,125	4,516
Febrero	10	8,063	3,752
Marzo	6	9,032	9,193
Abril	4	7,516	12,362
Mayo	9	5,758	10,511
Junio	7	7,379	0,144
Julio	8	7,189	0,658
Agosto	6	7,595	2,544
		Total	77,993

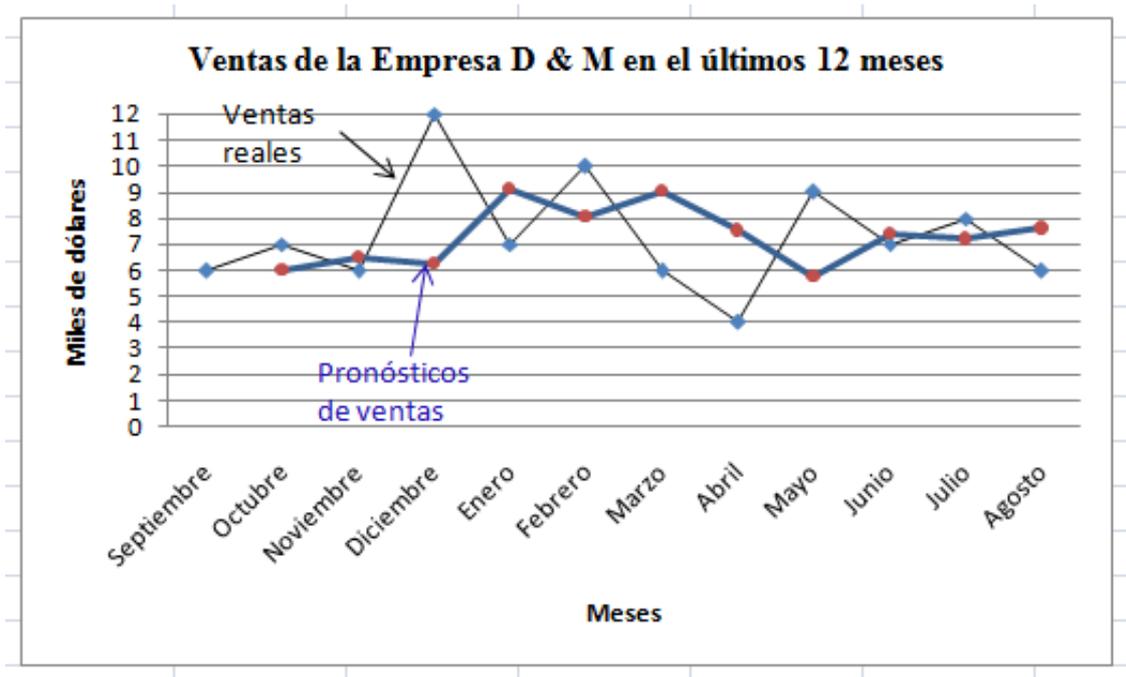
Aplicando la fórmula se obtiene el cuadrado medio del error:

$$CME = \frac{\sum(Y_t - X_t)^2}{n} = \frac{77,993}{11} = 7,09$$

Los cálculos realizados en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F
1	Meses	Ventas (X_t)	Pronóstico con $\alpha=0,5$	$Y_{t+1} = \alpha \cdot X_t + (1 - \alpha) \cdot Y_t$	Error	$(Y_t - X_t)^2$
2	Septiembre	6				
3	Octubre	7	6,000	=B2	1,000	=(B3-C3)^2
4	Noviembre	6	6,500	=SBS15*B3+(1-SBS15)*C3	0,250	=(B4-C4)^2
5	Diciembre	12	6,250	=SBS15*B4+(1-SBS15)*C4	33,063	=(B5-C5)^2
6	Enero	7	9,125	=SBS15*B5+(1-SBS15)*C5	4,516	=(B6-C6)^2
7	Febrero	10	8,063	=SBS15*B6+(1-SBS15)*C6	3,754	=(B7-C7)^2
8	Marzo	6	9,031	=SBS15*B7+(1-SBS15)*C7	9,188	=(B8-C8)^2
9	Abril	4	7,516	=SBS15*B8+(1-SBS15)*C8	12,360	=(B9-C9)^2
10	Mayo	9	5,758	=SBS15*B9+(1-SBS15)*C9	10,512	=(B10-C10)^2
11	Junio	7	7,379	=SBS15*B10+(1-SBS15)*C10	0,144	=(B11-C11)^2
12	Julio	8	7,189	=SBS15*B11+(1-SBS15)*C11	0,657	=(B12-C12)^2
13	Agosto	6	7,595	=SBS15*B12+(1-SBS15)*C12	2,543	=(B13-C13)^2
14		Y_{t+1}	6,797	=SBS15*B13+(1-SBS15)*C13		
15	α	0,5		CME	7,090	=PROMEDIO(E3:E13)

La gráfica de las ventas y los pronósticos con el método de suavización exponencial elaborada en Excel se muestra en la siguiente figura:



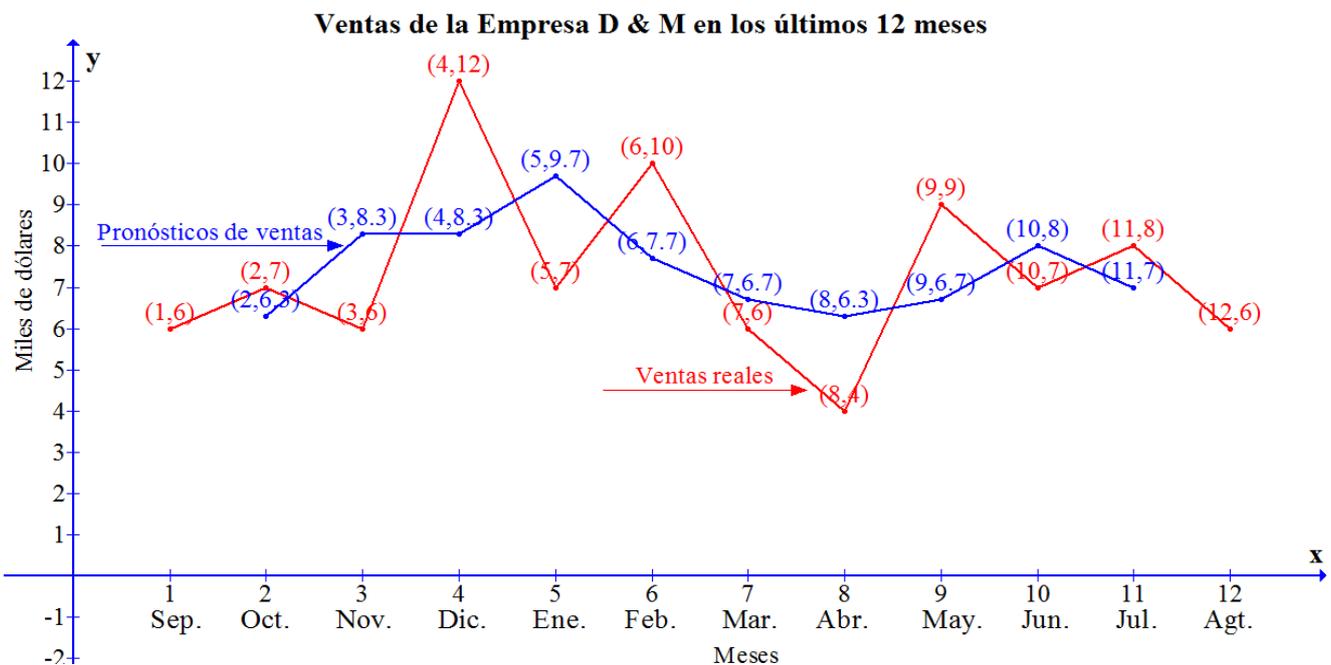
2) Suavizando los datos empleando el método de los promedios móviles de orden 3 elaborado en Excel se muestra en la siguiente figura:

	A	B	C	D	E	F
1	Meses	Ventas (X_t)	Pronósticos (promedios móviles)		Error	$(Y_t - X_t)^2$
2	Septiembre	6				
3	Octubre	7	6,333	=PROMEDIO(B2:B4)	0,444	=(B3-C3)^2
4	Noviembre	6	8,333	=PROMEDIO(B3:B5)	5,444	=(B4-C4)^2
5	Diciembre	12	8,333	=PROMEDIO(B4:B6)	13,444	=(B5-C5)^2
6	Enero	7	9,667	=PROMEDIO(B5:B7)	7,111	=(B6-C6)^2
7	Febrero	10	7,667	=PROMEDIO(B6:B8)	5,444	=(B7-C7)^2
8	Marzo	6	6,667	=PROMEDIO(B7:B9)	0,444	=(B8-C8)^2
9	Abril	4	6,333	=PROMEDIO(B8:B10)	5,444	=(B9-C9)^2
10	Mayo	9	6,667	=PROMEDIO(B9:B11)	5,444	=(B10-C10)^2
11	Junio	7	8	=PROMEDIO(B10:B12)	1,000	=(B11-C11)^2
12	Julio	8	7	=PROMEDIO(B11:B13)	1,000	=(B12-C12)^2
13	Agosto	6				
14				CME	4,522	=PROMEDIO(E3:E12)

Observando el gráfico anterior se tiene que el último pronóstico calculado es de 7, por lo que el pronóstico para septiembre es de 7.

Observando el gráfico anterior se tiene que el cuadrado medio del error es de 4,522.

La gráfica de las ventas y los pronósticos con el método de los promedios móviles elaborada en Graph se muestra en la siguiente figura:



3) Como CME en el método de suavización exponencial es de 7,09 y con el método de los promedios móviles es de 4,52, se concluye que el método de los promedios móviles es el más preciso para este ejemplo ilustrativo.

TAREA DE INTERAPRENDIZAJE N° 26

- 1) Realice un organizador gráfico sobre los modelos de series de tiempo.
- 2) Realice un organizador gráfico sobre los métodos de suavizamiento y pronóstico.
- 3) Con los siguientes datos acerca de la ventas en miles de dólares de la Empresa D & M durante los últimos 20 trimestres:

Trimestre	Ventas	Trimestre	Ventas
1	12	11	24
2	16	12	36
3	20	13	22
4	34	14	18
5	23	15	24
6	19	16	34
7	20	17	15
8	35	18	23
9	11	19	25
10	19	20	38

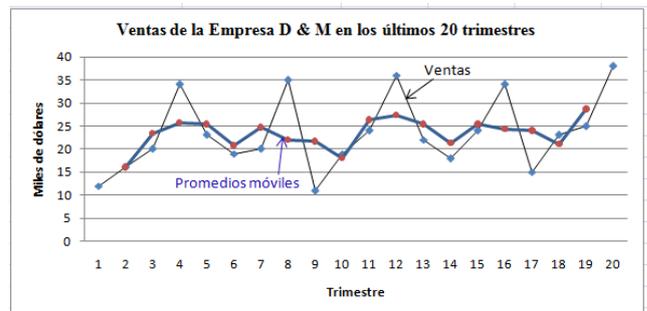
3.1) Suavice los datos empleando el método de los promedios móviles de orden 3 de manera manual y empleando Excel. Pronostique las ventas para el trimestre número 21.

28,67

3.2) Suponga que para el Gerente de Ventas, la venta realizada en el trimestre número 20 es el cuádruplo de importante que la realizada en el trimestre número 17, la venta del trimestre número 19 es el triple de importante que la del trimestre número 17, la venta del trimestre número 18 es la mitad de importante que la del trimestre número 20. Realizar el pronóstico de ventas de manera manual y empleando Excel para el trimestre número 21 utilizando el método de los promedios móviles ponderados para un orden o longitud de 4.

28,8

3.3) Elaborar un gráfico en el que consten las ventas y los promedios móviles elaborado de manera manual y empleando Excel.



4) Cree y resuelva un ejercicio de aplicación de promedios móviles con datos reales sobre cualquier tema de su preferencia.

5) Con los siguientes datos acerca de la ventas en miles de dólares de la Empresa M & M durante los últimos 12 meses:

Meses	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Enero	Febrero	Marzo	Abril	Mayo	Junio
Ventas(X_t)	5	6	7	9	8	12	7	10	7	8	11	8

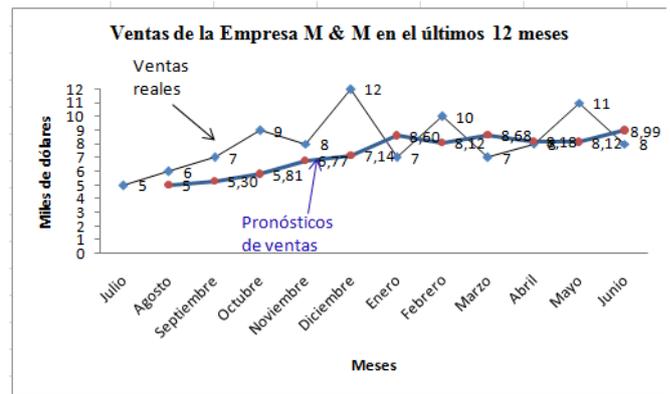
5.1) Suavice los datos utilizando el método de suavización exponencial con $\alpha = 0,3$ de manera manual y empleando Excel. Pronostique las ventas para el mes de julio.

8,7

5.2) Calcule el cuadrado medio del error del método anterior de manera manual y empleando Excel

5,2

5.3) Elabore un gráfico en el que consten las ventas y los pronósticos de suavización exponencial de manera manual y empleando Excel.



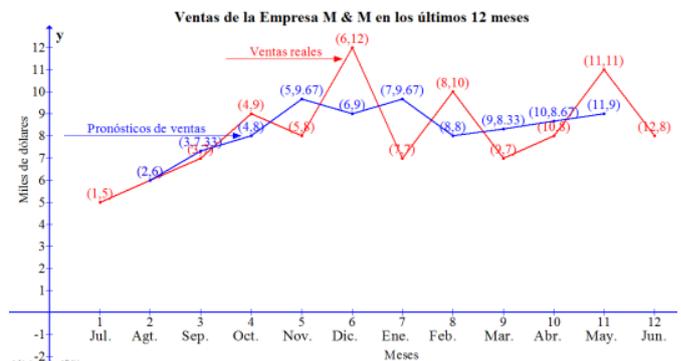
5.4) Suavice los datos empleando el método de los promedios móviles de orden 3 de manera manual y utilizando Excel. Pronosticar las ventas para mes de julio.

9

5.5) Calcular el cuadrado medio del error del método anterior de manera manual y utilizando Excel.

3,02

5.6) Elabore un gráfico en el que consten las ventas y los promedios móviles de manera manual y empleando Graph.



5.7) ¿Qué método es el más preciso para este ejercicio?

El método de los promedios móviles

6) Cree y resuelva un ejercicio similar al anterior (5.1 a 5.7) con datos reales sobre cualquier tema de su preferencia.

6.5) ANÁLISIS DE TENDENCIA

Es necesario describir la tendencia ascendente o descendente a largo plazo de una serie cronológica por medio de alguna línea, y la más adecuada será la que mejor represente los datos y sea útil para desarrollar pronósticos. Para lograr la estimación de la tendencia se utilizan con más frecuencia los siguientes métodos:

6.5.1) MÉTODO DE LOS MÍNIMOS CUADRADOS

Este método ya se estudió en el capítulo anterior, en el que se indicó las formas para hallar la ecuación de una recta de mínimos cuadrados. Con esta recta se obtendrán los valores de tendencia.

Ejemplo ilustrativo:

Con los siguientes datos acerca de las ventas en millones de dólares de la Empresa M & M:

Año (X)	Ventas (Y)
1995	3,4
1996	3,1
1997	3,9
1998	3,3
1999	3,2
2000	4,3
2001	3,9
2002	3,5
2003	3,6
2004	3,7
2005	4
2006	3,6
2007	4,1
2008	4,7
2009	4,2
2010	4,5

- 1) Hallar la ecuación de tendencia por el método de los mínimos cuadrados.
- 2) Pronosticar la tendencia de exportación para el 2011.
- 3) Elaborar la gráfica para los datos y la recta de tendencia.

Solución:

1) Para hallar la ecuación de tendencia por el método de los mínimos cuadrados se llena la siguiente tabla, codificando la numeración de los años 1995 como 1, 1996 como 2, y así consecutivamente para facilitar los cálculos.

Año (X)	X	Y	XY	X ²	Y ²
1995	1	3,4	3,40	1	11,56
1996	2	3,1	6,20	4	9,61
1997	3	3,9	11,70	9	15,21
1998	4	3,3	13,20	16	10,89
1999	5	3,2	16,00	25	10,24
2000	6	4,3	25,80	36	18,49
2001	7	3,9	27,30	49	15,21
2002	8	3,5	28,00	64	12,25
2003	9	3,6	32,40	81	12,96
2004	10	3,7	37,00	100	13,69
2005	11	4	44,00	121	16,00
2006	12	3,6	43,20	144	12,96
2007	13	4,1	53,30	169	16,81
2008	14	4,7	65,80	196	22,09
2009	15	4,2	63,00	225	17,64
2010	16	4,5	72,00	256	20,25
Total	136	61	542,3	1496	235,86

Reemplazando valores en las siguientes fórmulas se obtiene los valores de a_0 y a_1 :

$$a_0 = \frac{\sum Y \cdot \sum X^2 - \sum X \cdot \sum XY}{N \sum X^2 - (\sum X)^2} = \frac{61 \cdot 1496 - 136 \cdot 542,3}{16 \cdot 1496 - (136)^2} = \frac{17503,2}{5440} = 3,2175 = 3,22$$

$$a_1 = \frac{N \sum XY - \sum X \cdot \sum Y}{N \sum X^2 - (\sum X)^2} = \frac{16 \cdot 542,3 - 136 \cdot 61}{16 \cdot 1496 - (136)^2} = \frac{380,8}{5440} = 0,07$$

Interpretación:

- El valor $a_1 = 0,07$ al ser positiva indica que existe una tendencia ascendente de las exportaciones aumentando a un cambio o razón promedio de 0,07 millones de dólares por cada año.

- El valor de $a_0 = 3,22$ indica el punto en donde la recta interseca al eje Y cuando $X = 0$, es decir indica las exportaciones estimadas para el año 1996 igual a 3,22.

Reemplazado los valores anteriores en la recta de tendencia se obtiene:

$$Y = a_0 + a_1 X$$

$$Y = 3,22 + 0,07X$$

2) Para pronosticar la tendencia de exportación para el 2011 se reemplaza $X = 17$ en la recta de tendencia, obteniendo el siguiente resultado:

$$Y = 3,22 + 0,07X$$

$$Y = 3,22 + 0,07 \cdot 17 = 4,41$$

Los cálculos en Excel se muestran en la siguiente figura:

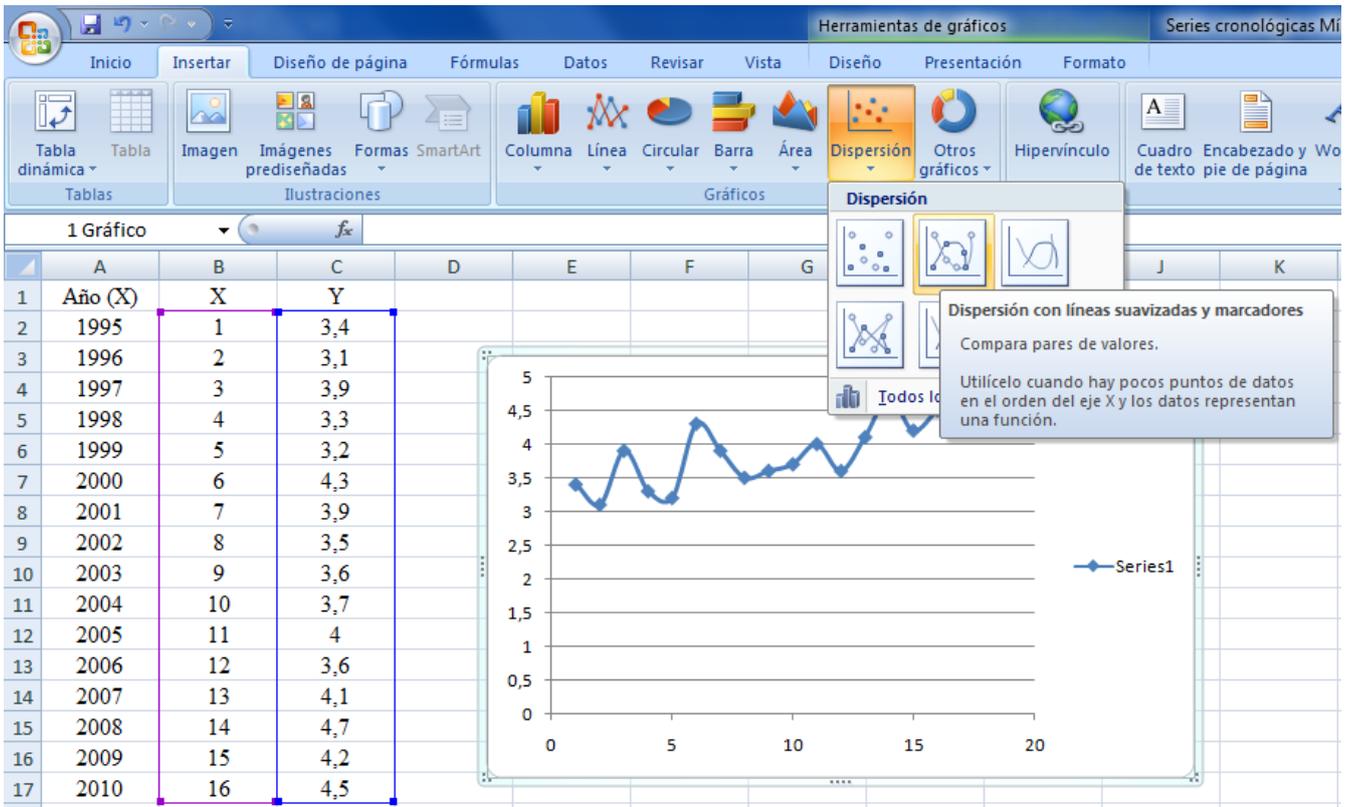
	A	B	C	D	E	F	G	H	I	J	K	L
1	Año (X)	X	Y	XY		X ²		Y ²		Pronóstico		
2	1995	1	3,4	3,40	=B2*C2	1	=B2^2	11,56	=C2^2			
3	1996	2	3,1	6,20	=B3*C3	4	=B3^2	9,61	=C3^2			
4	1997	3	3,9	11,70	=B4*C4	9	=B4^2	15,21	=C4^2			
5	1998	4	3,3	13,20	=B5*C5	16	=B5^2	10,89	=C5^2			
6	1999	5	3,2	16,00	=B6*C6	25	=B6^2	10,24	=C6^2			
7	2000	6	4,3	25,80	=B7*C7	36	=B7^2	18,49	=C7^2			
8	2001	7	3,9	27,30	=B8*C8	49	=B8^2	15,21	=C8^2			
9	2002	8	3,5	28,00	=B9*C9	64	=B9^2	12,25	=C9^2			
10	2003	9	3,6	32,40	=B10*C10	81	=B10^2	12,96	=C10^2			
11	2004	10	3,7	37,00	=B11*C11	100	=B11^2	13,69	=C11^2			
12	2005	11	4	44,00	=B12*C12	121	=B12^2	16,00	=C12^2			
13	2006	12	3,6	43,20	=B13*C13	144	=B13^2	12,96	=C13^2			
14	2007	13	4,1	53,30	=B14*C14	169	=B14^2	16,81	=C14^2			
15	2008	14	4,7	65,80	=B15*C15	196	=B15^2	22,09	=C15^2			
16	2009	15	4,2	63,00	=B16*C16	225	=B16^2	17,64	=C16^2			
17	2010	16	4,5	72,00	=B17*C17	256	=B17^2	20,25	=C17^2			
18	2011									4,41	=SDS23+SDS26*(B17+1)	
19	Total	136	61	542,30		1496		235,86				
20		=SUMA(B2:B18)	=SUMA(C2:C17)	=SUMA(D2:D17)		=SUMA(F2:F18)		=SUMA(H2:H17)				
21	N	16	CONTAR(B2:B17)									
22												
23	$a_0 = \frac{\sum Y \cdot \sum X^2 - \sum X \cdot \sum XY}{N \sum X^2 - (\sum X)^2}$			3,22				= (C19*F19-B19*D19)/(B21*F19-(B19)^2)				
24												
25	$a_1 = \frac{N \sum XY - \sum X \cdot \sum Y}{N \sum X^2 - (\sum X)^2}$			0,070				= (B21*D19-B19*C19)/(B21*F19-(B19)^2)				
26												
27												

3) La gráfica de los datos y la recta de tendencia elaborada en Excel se muestran en la siguiente figura:

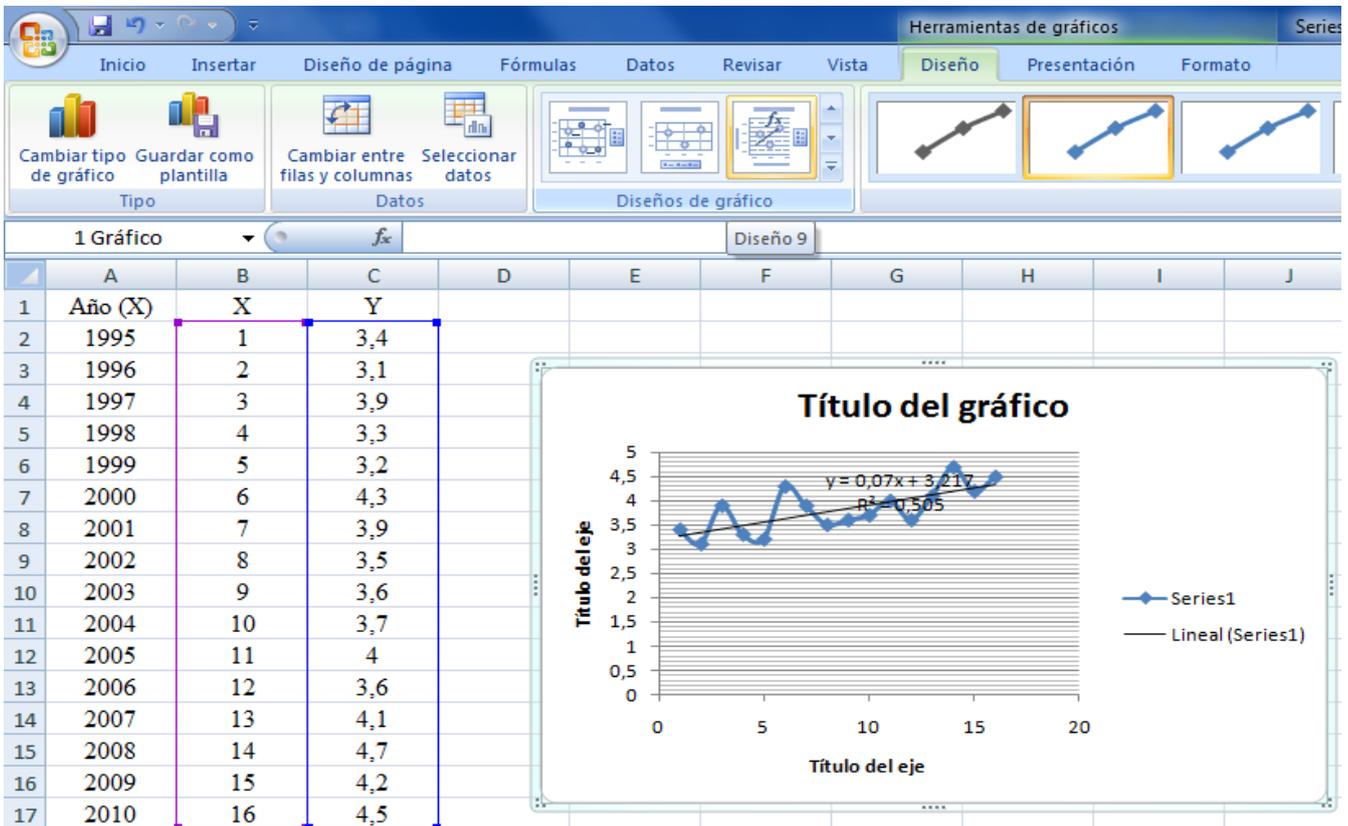


Para realizar la gráfica anterior empleando Excel se procede de la siguiente manera:

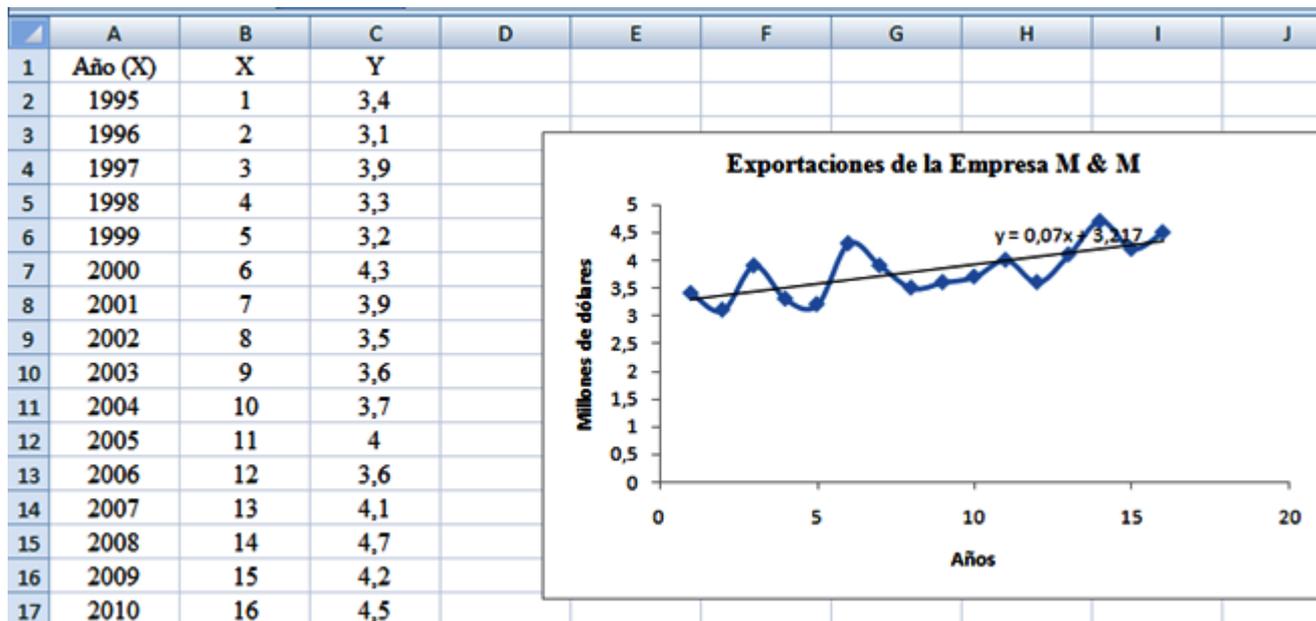
a) Escribir los datos en la hoja de cálculo. Seleccionar los datos. Insertar Gráfico de Dispersión con líneas suavizadas y marcadores.



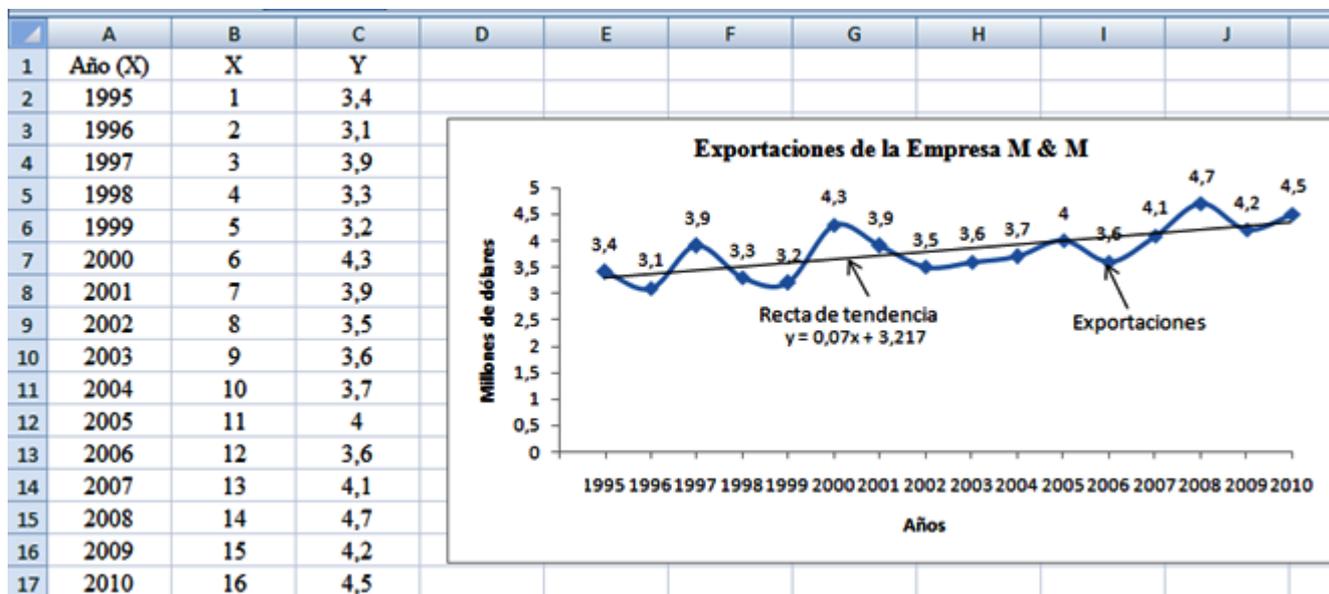
b) Seleccionar el gráfico. Escoger el Diseño 9 en Diseño de gráfico.



c) Borrar las palabras Series1 y Lineal (Series1). Borrar las líneas de división. Borrar $R^2 = 0,505$ que representa el coeficiente de determinación que este caso no es necesario. Escribir Exportaciones de la Empresa M & M en el título del gráfico. Escribir Años en el título del eje horizontal. Escribir millones de dólares en el eje vertical.



d) Arrastrar el cuadro la ecuación $Y = 0,07X + 3,217$ hacia abajo, esta ecuación es la recta de tendencia que Excel calcula al realizar el gráfico. Agregar etiquetas. Editar el eje x de 1995 al 2010. Insertar los textos Recta de tendencia y Exportaciones. Insertar formas de Flecha.



6.5.2) MÉTODO DE LOS SEMIPROMEDIOS

Este método se aplica con el objeto de simplificar los cálculos y consiste en:

- a) Agrupar los datos en dos grupos iguales
- b) Obtener el valor central (mediana) de los tiempos y la media aritmética de los datos de cada grupo, consiguiéndose así dos puntos de la recta de tendencia (X_1, Y_1) y (X_2, Y_2) .
- c) Estos valores se reemplazan en el siguiente sistema:

$$\begin{cases} Y_1 = a_0 + a_1 X_1 \\ Y_2 = a_0 + a_1 X_2 \end{cases}$$

- d) Resolviendo el sistema se encuentran los valores de a_0 y a_1 , los cuales se reemplazan en la ecuación de la recta de tendencia, la cual es:

$$Y = a_0 + a_1 X$$

Con esta recta de tendencia se puede realizar pronósticos, los cuales son menos exactos que los obtenidos con el método de los mínimos cuadrados, sin embargo, su diferencia es mínima.

Ejemplo ilustrativo N° 1

Con los siguientes datos sobre las ventas en millones de dólares de la Empresa D & M

Año (X)	Ventas (Y)
2000	1,5
2001	1,8
2002	2
2003	1,5
2004	2,2
2005	2
2006	3
2007	2,8
2008	2,4
2009	2,9
2010	3

- 1) Hallar la ecuación de tendencia por el método de los semipromedios.
- 2) Pronosticar la tendencia de ventas para el 2011.
- 3) Elaborar la gráfica para los datos y la recta de tendencia.

Solución:

1) Se codifica la numeración de los años 2000 como 1, 2001 como 2, y así consecutivamente para facilitar los cálculos. Se agrupa en dos grupos iguales.

Año	X	Y	Valor central X	Semipromedio Y
2000	1	1,5	3	1,8
2001	2	1,8		
2002	3	2		
2003	4	1,5		
2004	5	2,2		
2005	6	2		
2006	7	3	9	2,82
2007	8	2,8		
2008	9	2,4		
2009	10	2,9		
2010	11	3		

El año 2005 se dejó por fuera para tener grupos con el mismo número de años. El valor central de 3 corresponde a la mediana del primer grupo 1, 2, 3, 4 y 5. El valor central de 9 corresponde a la mediana del segundo grupo 7, 8, 9, 10 y 11. El semipromedio 1,8 corresponden a la media aritmética del primer grupo. El semipromedio 2,82 corresponden a la media aritmética del segundo grupo. De esta manera se obtienen dos puntos (3, 1.8) y (9, 2.82) de la recta de tendencia.

Reemplazando los puntos en el siguiente sistema se obtiene:

$$\begin{cases} Y_1 = a_0 + a_1 X_1 \\ Y_2 = a_0 + a_1 X_2 \end{cases} \Rightarrow \begin{cases} 1,8 = a_0 + 3a_1 \\ 2,82 = a_0 + 9a_1 \end{cases}$$

Resolviendo el sistema empleando la regla de Cramer se obtiene:

$$a_0 = \frac{\Delta_{a_0}}{\Delta} = \frac{\begin{vmatrix} 1,8 & 3 \\ 2,82 & 9 \end{vmatrix}}{\begin{vmatrix} 1 & 3 \\ 1 & 9 \end{vmatrix}} = \frac{7,74}{6} = 1,29$$

$$a_1 = \frac{\Delta_{a_1}}{\Delta} = \frac{\begin{vmatrix} 1 & 1,8 \\ 1 & 2,82 \end{vmatrix}}{\begin{vmatrix} 1 & 3 \\ 1 & 9 \end{vmatrix}} = \frac{1,02}{6} = 0,17$$

Como a_1 es positiva, la recta tiene una tendencia ascendente (pendiente positiva).

Reemplazando los valores calculados se tiene la recta de tendencia, la cual es:

$$Y = a_0 + a_1 X$$

$$Y = 1,29 + 0,17X$$

2) Para pronosticar la tendencia de exportación para el 2011 se reemplaza $X = 12$ en la recta de tendencia, obteniendo el siguiente resultado:

$$Y = 1,29 + 0,17X$$

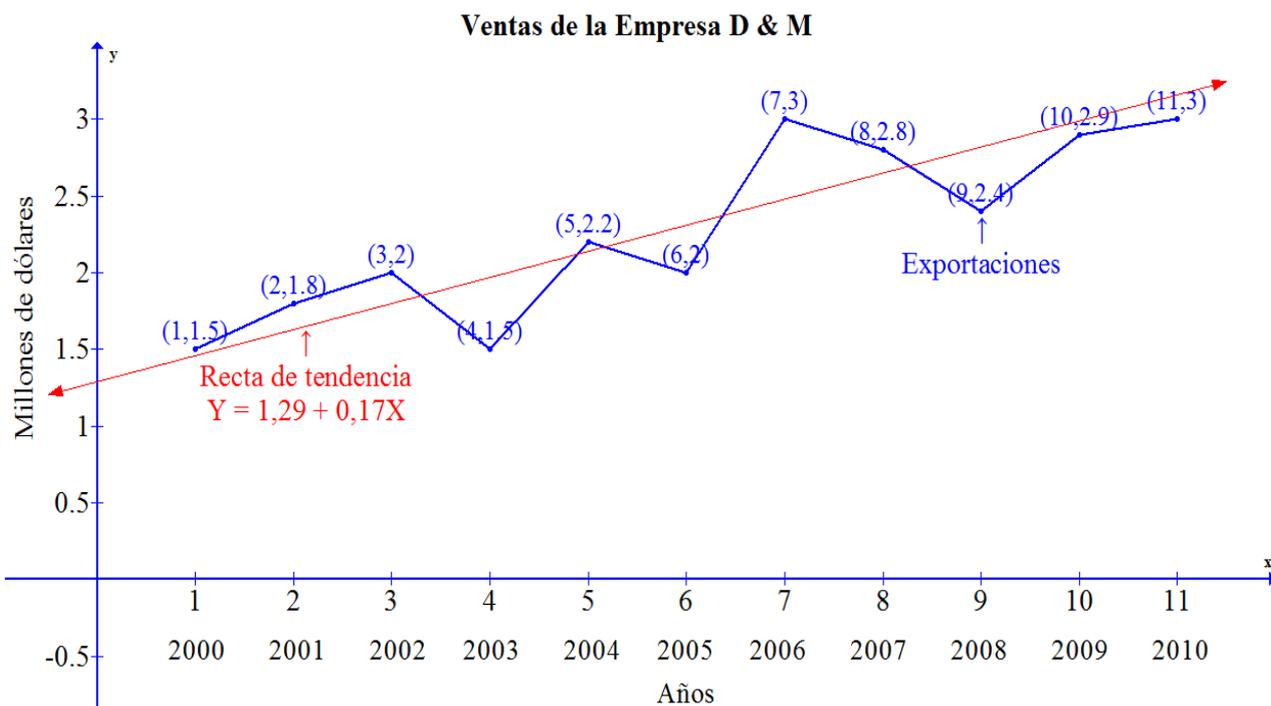
$$Y = 1,29 + 0,17 \cdot 12 = 3,33$$

Interpretación: Existe una tendencia ascendente a un cambio promedio de 0,17 millones de dólares por cada año, por lo que el Gerente de ventas de la empresa debe seguir aplicando las políticas necesarias para mantener la tendencia ascendente y mejorar la tasa de crecimiento.

Los cálculos realizados en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H	I
1	Año (X)	X	Y	Valor central X		Semipomedio Y		Pronóstico	
2	2000	1	1,5						
3	2001	2	1,8						
4	2002	3	2	3	=MEDIANA(B2:B6)	2	=PROMEDIO(C2:C6)		
5	2003	4	1,5						
6	2004	5	2,2						
7	2005	6	2						
8	2006	7	3						
9	2007	8	2,8						
10	2008	9	2,4	9	=MEDIANA(B8:B12)	2,82	=PROMEDIO(C8:C12)		
11	2009	10	2,9						
12	2010	11	3						
13	2011	12						3,33	=H19+H22*B13
14									
15	$Y = a_0 + a_1 X$								
16	1,8	1	3		6	=MDETERM(B16:C17)			
17	2,82	1	9						
18									
19		1,80	3		7,74	=MDETERM(B19:C20)		1,29	=E19/E16
20		2,82	9						
21									
22		1	1,80		1,02	=MDETERM(B22:C23)		0,17	=E22/E16
23		1	2,82						

3) La gráfica de los datos y la recta de tendencia elaborada en Graph se muestran en la siguiente figura:



Ejemplo ilustrativo N° 2

Con los siguientes datos acerca de las ventas en miles de dólares de la Empresa D & M durante los últimos 12 meses:

Meses (X)	Ventas (Y)
Septiembre	6
Octubre	7
Noviembre	6
Diciembre	12
Enero	7
Febrero	10
Marzo	6
Abril	4
Mayo	9
Junio	7
Julio	8
Agosto	6

- 1) Hallar la ecuación de tendencia por el método de los semipromedios.
- 2) Pronosticar la tendencia de ventas para el mes de septiembre.
- 3) Elaborar la gráfica para los datos y la recta de tendencia.

Solución:

1) Se codifica la numeración de los años 2000 como 1, 2001 como 2, y así consecutivamente para facilitar los cálculos. Se agrupa en dos grupos iguales.

Meses (X)	X	Y	Valor central X	Semipromedio Y
Septiembre	1	6	3,5	8
Octubre	2	7		
Noviembre	3	6		
Diciembre	4	12		
Enero	5	7	9,5	6,667
Febrero	6	10		
Marzo	7	6		
Abril	8	4		
Mayo	9	9		
Junio	10	7		
Julio	11	8		
Agosto	12	6		

Reemplazando los valores centrales y los semipromedios puntos en el siguiente sistema se obtiene:

$$\begin{cases} Y_1 = a_0 + a_1 X_1 \\ Y_2 = a_0 + a_1 X_2 \end{cases} \Rightarrow \begin{cases} 8 = a_0 + 3,5a_1 \\ 6,667 = a_0 + 9,5a_1 \end{cases}$$

Resolviendo el sistema se obtiene:

$$a_0 = 8,778$$

$$a_1 = -0,222$$

Como a_1 es negativa, la recta tiene una tendencia descendente (pendiente negativa).

Reemplazando los valores calculados se tiene la recta de tendencia, la cual es:

$$Y = a_0 + a_1X$$

$$Y = 8,778 - 0,222X$$

2) Para pronosticar la tendencia de ventas para el mes de septiembre se reemplaza $X = 13$ en la recta de tendencia, obteniendo el siguiente resultado:

$$Y = 8,778 - 0,222X$$

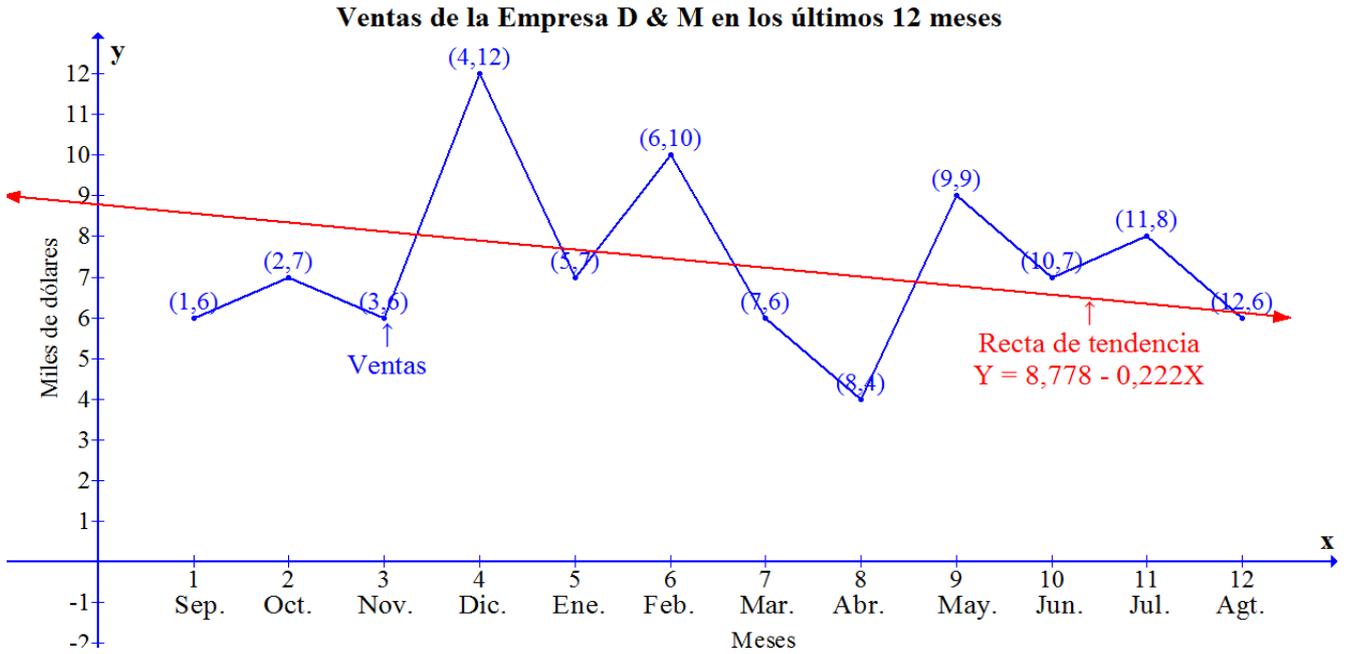
$$Y = 8,778 - 0,222 \cdot 13 = 5,89$$

Interpretación: Existe una tendencia descendente a un cambio promedio de 0,222 miles de dólares por cada mes, por lo que el Gerente de ventas de la empresa debe aplicar los correctivos pertinentes para salir de esta situación.

Los cálculos realizados en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H	I
4	Noviembre	3	6						
5	Diciembre	4	12	3,5	=MEDIANA(B2:B7)	8	=PROMEDIO(C2:C7)		
6	Enero	5	7						
7	Febrero	6	10						
8	Marzo	7	6						
9	Abril	8	4						
10	Mayo	9	9						
11	Junio	10	7	9,5	=MEDIANA(B8:B13)	6,667	=PROMEDIO(C8:C13)		
12	Julio	11	8						
13	Agosto	12	6						
14	Septiembre	13						5,89	=I19+I22*B13
15									
16		$Y = a_0 + a_1X$							
17	8,0	1	3,5		6	=MDETERM(B16:C17)			
18	6,667	1	9,5						
19									
20		8,00	3,5		52,667	=MDETERM(B19:C20)	a_0	8,778	=E19/E16
21		6,67	9,5						
22									
23		1	8,00		-1,333	=MDETERM(B22:C23)	a_1	-0,222	=E22/E16
24		1	6,67						

3) La gráfica de los datos y la recta de tendencia elaborada en Graph se muestran en la siguiente figura:



TAREA DE INTERAPRENDIZAJE N° 27

- 1) Realice un organizador gráfico del análisis de tendencia
- 2) ¿Qué interpretación tiene el valor de a_0 y a_1 en la recta de tendencia?
- 3) Con los siguientes datos sobre las exportaciones en millones de dólares de la Empresa M & M

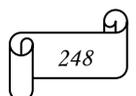
Año (X)	Exportaciones (Y)
2000	4,3
2001	3,9
2002	3,5
2003	3,6
2004	3,7
2005	4
2006	3,6
2007	4,1
2008	4,7
2009	4,2
2010	4,5

3.1) Halle la recta de tendencia por el método de los mínimos cuadrados de manera manual y empleando Excel. Codifique la numeración de los años 2000 como 1, 2001 como 2, y así consecutivamente. Realice la interpretación respectiva.

$$Y = 3,644 + 0,061X$$

3.2) Pronostique la tendencia para el 2011

4,38



3.3) Grafique los datos y la recta de tendencia empleando Excel



3.4) Halle la recta de tendencia por el método de los semipromedios de manera manual y empleando Excel. Codifique la numeración de los años 2000 como 1, 2001 como 2, y así consecutivamente. Realice la interpretación respectiva.

$$Y = 3,59 + 0,07X$$

3.5) Pronostique la tendencia para el 2011.

4,43

3.6) Grafique los datos y la recta de tendencia empleando Graph.



4) Con los siguientes datos acerca de las exportaciones en millones de dólares de la Empresa M & M

Año (X)	Exportaciones (Y)
1995	3,4
1996	3,1
1997	3,9
1998	3,3
1999	3,2
2000	4,3
2001	3,9
2002	3,5
2003	3,6
2004	3,7
2005	4
2006	3,6
2007	4,1
2008	4,7
2009	4,2
2010	4,5

4.1) Halle la recta de tendencia por el método de los semipromedios de manera manual y empleando Excel. Realice la interpretación respectiva.

$$Y = 3,308 + 0,059X$$

4.2) Pronostique la tendencia para el 2011

4,3

4.3) Grafique los datos y la recta de tendencia empleando Graph.



5) Cree un ejercicio de estimación de tendencia con datos reales sobre cualquier tema de su preferencia. Resuélvalo empleando los dos métodos aprendidos de manera manual, empleando Excel y Graph. Realice las interpretaciones y pronósticos respectivos.

6.6) ANÁLISIS DE MOVIMIENTOS ESTACIONALES

Para analizar el movimiento estacional debemos estimar cómo varían los datos de la serie cronológica en el período de tiempo. Un conjunto de números que muestra los valores relativos de una variable durante los períodos de tiempo se llama un *índice estacional* para la variable. El índice estacional medio del año ha de ser 100%; esto es, la suma de los números índice de los 12 meses suman 1200%, o de los cuatro trimestres suman el 400%, en caso contrario ha de corregirse multiplicado por el factor de ajuste, el mismo que es:

$$\text{Factor de ajuste mensual} = \frac{120}{\text{suma de medias mensuales}}$$

$$\text{Factor de ajuste trimestral} = \frac{400}{\text{suma de medias trimestrales}}$$

6.6.1) CÁLCULO DEL ÍNDICE ESTACIONAL POR EL MÉTODO DEL PORCENTAJE MEDIO

Este método consiste en calcular los índices estacionales como porcentajes de los períodos de tiempo (mensual o trimestral). Para lo cual se calcula de cada año la media mensual o trimestral, según sea el caso, luego se divide el dato de cada mes o trimestre por la media mensual o trimestral del correspondiente año y se multiplica por 100, y luego se calcula la media de cada mes o trimestre, obteniéndose el índice estacional.

6.6.2) DESESTACIONALIZACIÓN DE LOS DATOS O AJUSTE DE LOS DATOS A LA VARIACIÓN ESTACIONAL

Una vez obtenidos los índices estacionales es posible eliminar el movimiento estacional de los datos, para lo cual se divide todos los datos originales por el índice estacional del período de tiempo (mes o trimestre) correspondiente. Los valores desestacionalizados reflejan cómo sería la variable si se corrigiera la influencia estacional.

Ejemplo ilustrativo:

Con los datos de la siguiente tabla que muestra las exportaciones en millones de dólares de la Empresa D & M.

Trimestre \ Año	I	II	III	IV
2008	20	32	22	40
2009	25	35	30	45
2010	28	38	36	44

- 1) Calcular el índice estacional
- 2) Desestacionalizar los datos

Solución:

1) Se calcula la media trimestral, la cual se presenta en la siguiente tabla:

Trimestre \ Año	I	II	III	IV	Media trimestral
2008	20	32	22	40	28,5
2009	25	35	30	45	33,75
2010	28	38	36	44	36,5

Se divide el dato de cada trimestre por la media trimestral del correspondiente año y se multiplica por 100, como se muestra en la siguiente tabla:

Trimestre \ Año	I	II	III	IV
2008	70,175	112,281	77,193	140,351
2009	74,074	103,704	88,889	133,333
2010	76,712	104,110	98,630	120,548

Se calcula la media de cada trimestre como se muestra en la siguiente tabla:

Trimestre \ Año	I	II	III	IV
2008	70,175	112,281	77,193	140,351
2009	74,074	103,704	88,889	133,333
2010	76,712	104,110	98,630	120,548
<i>Media</i>	<i>73,654</i>	<i>106,698</i>	<i>88,237</i>	<i>131,411</i>

Se suma las medias de cada trimestre, las cuales deben dar como resultado 400. Al sumar $73,654 + 106,698 + 88,237 + 131,411$ se obtiene 399,999, por lo que no existe la necesidad de multiplicar la media trimestral por el factor de ajuste trimestral. Por lo tanto las medias trimestrales representan el índice estacional, como se muestra en la siguiente tabla:

Trimestre \ Año	I	II	III	IV
2008	70,175	112,281	77,193	140,351
2009	74,074	103,704	88,889	133,333
2010	76,712	104,110	98,630	120,548
<i>Media</i>	<i>73,654</i>	<i>106,698</i>	<i>88,237</i>	<i>131,411</i>
Índice estacional	73,654%	106,698%	88,237%	131,411%

Interpretación:

El índice estacional de 73,654% para el primer trimestre significa que las exportaciones de empresa D & M son de 73,654% del promedio del año total. Las exportaciones son $100\% - 73,654\% = 26,346\%$ por debajo del promedio trimestral del año.

El índice estacional de 106,698% para el segundo trimestre significa que las exportaciones de empresa D & M son de 106,698% del promedio del año total. Las exportaciones son $100\% - 106,698\% = 6,698\%$ por encima del promedio trimestral del año.

2) Dividiendo los valores reales por sus índices estacionales respectivos se obtienen los valores desestacionalizados también denominados *corregidos estacionalmente*. En la siguiente tabla se muestra los valores desestacionalizados:

2008 - I: $20/0,73654 = 27,15$, y así sucesivamente

Trimestre \ Año	I	II	III	IV
2008	27,15	29,99	24,93	30,44
2009	33,94	32,80	33,99	34,24
2010	38,02	35,61	40,80	33,48

Interpretación: El valor de 27,15 significa que si las exportaciones de la empresa D & M no estuvieran sujetas a la variación estacional, las exportaciones para el primer trimestre del año 2008 hubieran sido de 27,15 millones de dólares.

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H
1	Trimestre	I	II	III	IV	Media trimestral		
2	Año							
3	2008	20	32	22	40	28,50	=PROMEDIO(B3:E3)	
4	2009	25	35	30	45	33,75	=PROMEDIO(B4:E4)	
5	2010	28	38	36	44	36,50	=PROMEDIO(B5:E5)	
6	Total	73	105	88	129			
7	Media	24,333	35	29,333	43,000			
8								
9	Trimestre	I	II	III	IV			
10	Año							
11	2008	70,175	112,281	77,193	140,351	=(E3/\$F\$3)*\$A\$17		
12	2009	74,074	103,704	88,889	133,333	=(E4/\$F\$4)*\$A\$17		
13	2010	76,712	104,110	98,630	120,548	=(E5/\$F\$5)*\$A\$17		
14	Media	73,654	106,698	88,237	131,411	=PROMEDIO(E11:E13)	400	=SUMA(B14:E14)
15	Índice	73,654	106,698	88,237	131,411	=E14*\$C\$18		
16								
17	100							
18	Factor de ajuste		1	=400/G14				
19								
20		Datos desestacionalizados						
21	Trimestre	I	II	III	IV			
22	Año							
23	2008	27,154	29,991	24,933	30,439	=(E3/\$E\$15)*\$A\$17		
24	2009	33,943	32,803	33,999	34,244	=(E4/\$E\$15)*\$A\$17		
25	2010	38,016	35,615	40,799	33,483	=(E5/\$E\$15)*\$A\$17		

TAREA INTERAPRENDIZAJE N° 28

1) Realice un organizador gráfico sobre el análisis de movimientos estacionales.

2) Con los datos de la siguiente tabla que muestra las exportaciones trimestrales en millones de dólares de la empresa M & D durante los años 2008-2009.

Año \ Trimestre	I	II	III	IV
2008	24	31	21	42
2009	27	32	26	41
2010	28	27	35	44

2.1) Calcule el índice estacional de manera manual y empleando Excel.

Índice	83,551	95,756	86,068	134,625
--------	--------	--------	--------	---------

2.2) Desestacionalice los datos de manera manual y empleando Excel.

Año \ Trimestre	I	II	III	IV
2008	28,725	32,374	24,399	31,198
2009	32,316	33,418	30,209	30,455
2010	33,513	28,197	40,666	32,683

3) Con los datos de la siguiente tabla que muestra las exportaciones mensuales en millones de dólares de la empresa M & D durante los años 2005-2009.

Mes \ Año	En	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ag.	Sep.	Oct.	Nov.	Dic.
2005	4	8	6	4	8	6	4	6	8	4	4	10
2006	3	9	7	6	7	8	7	5	8	6	7	11
2007	5	8	10	6	8	10	8	7	8	6	8	12
2008	6	10	12	9	11	9	8	7	10	7	7	12
2009	6	10	12	8	10	12	10	10	10	8	10	14

3.1) Calcule el índice estacional de manera manual y empleando Excel.

Índice	59,74	114,60	115,67	81,48	111,11	111,86	9,11	87,34	111,75	77,03	88,89	149,43
--------	-------	--------	--------	-------	--------	--------	------	-------	--------	-------	-------	--------

3.2) Desestacionalice los datos de manera manual y empleando Excel.

Mes \ Año	En	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ag.	Sep.	Oct.	Nov.	Dic.
2005	6,70	6,98	5,19	4,91	7,20	5,36	4,39	6,87	7,16	5,19	4,50	6,69
2006	5,02	7,85	6,05	7,36	6,300	7,15	7,68	5,73	7,16	7,79	7,88	7,36
2007	8,37	6,99	8,65	7,36	7,20	8,94	8,78	8,02	7,16	7,79	9,00	8,03
2008	10,04	8,73	10,38	11,05	9,90	8,05	8,78	8,02	8,95	9,09	7,88	8,03
2009	10,04	8,73	10,38	9,82	9,00	10,73	10,98	11,45	8,95	10,39	11,25	9,37

4) Cree y resuelva un ejercicio similar al anterior sobre cualquier tema de su preferencia.

6.7) ANÁLISIS DE MOVIMIENTOS CÍCLICOS E IRREGULARES

Los movimientos cíclicos son de tipo periódico y presentan más de un año de duración. Comúnmente, tales movimientos o variaciones no se pueden apartar de la naturaleza irregular, por lo que se analizarán juntas.

Recordemos que $Y = T \cdot C \cdot E \cdot I$ de donde $C \cdot I = Y/T \cdot E$. Por lo que los movimientos cíclicos e irregulares se obtienen dividiendo los datos originales entre el valor de tendencia estimado, y este cociente multiplicando por 100% de la siguiente manera:

$$CI = \frac{Y}{Y_{est.}} \cdot 100\%$$

Donde:

Y = Variable Y

$Y_{est.}$ = Valor de tendencia estimado

CI = Movimientos cíclicos e irregulares

El cociente se *multiplica por 100* a fin de que la media cíclica sea 100. Un valor cíclico relativo de 100 indicará la ausencia de toda influencia cíclica en el valor de la serie de tiempo anual.

Para facilitar la interpretación de relativos ciclos suele elaborarse una *gráfica de ciclos*, en el que se describen los ciclos relativos según el año correspondiente. Las cumbres y valles asociados con el componente cíclico de las series de tiempo pueden resultar más evidentes por medio de la elaboración de una gráfica de este tipo.

Ejemplo ilustrativo

Con los siguientes datos acerca de las ventas en millones de dólares de la Empresa M & M:

Año (X)	Ventas (Y)
1995	3,4
1996	3,1
1997	3,9
1998	3,3
1999	3,2
2000	4,3
2001	3,9
2002	3,5
2003	3,6
2004	3,7
2005	4
2006	3,6
2007	4,1
2008	4,7
2009	4,2
2010	4,5

- 1) Determinar el componente cíclico de cada uno de los valores de la serie cronológica usando la ecuación de tendencia
- 2) Elaborar una gráfica de ciclos

Solución:

1) La ecuación de tendencia lineal obtenida empleando el método de los mínimos cuadrados es:

$$Y = 3,22 + 0,07X$$

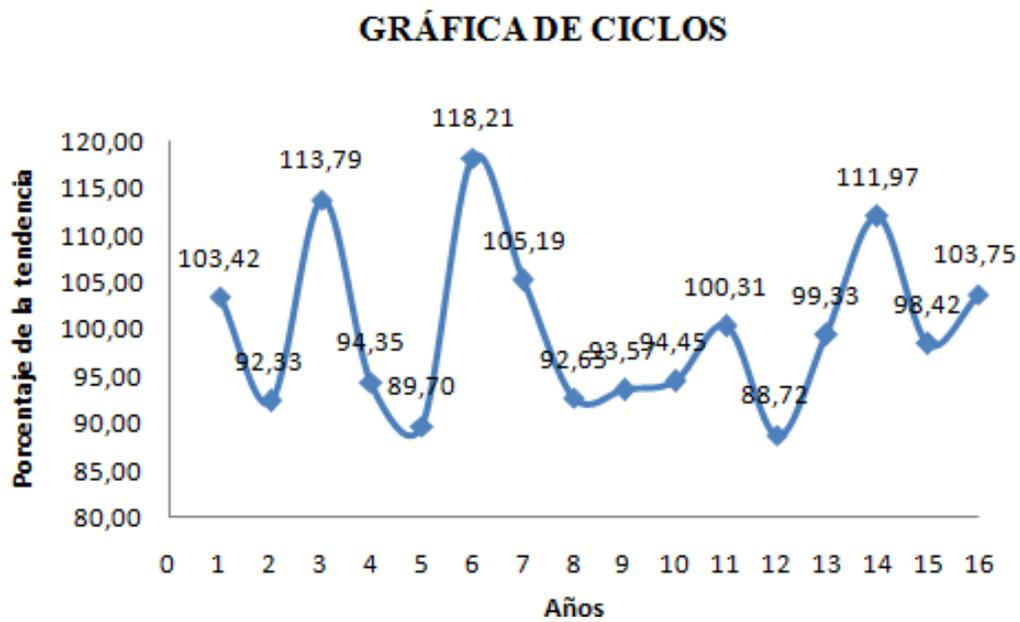
Con esta ecuación se calcula los valores estimados de Y reemplazando los valores de X en la recta de tendencia. Luego se divide los datos originales Y entre el valor de tendencia estimado, y este cociente se multiplica por 100%, como se muestra en la siguiente tabla:

Año (X)	Ventas (Y)	$Y_{est} = 3,22 + 0,07X$	$C \cdot I = (Y/Y_{est}) \cdot 100\%$
1995	3,4	3,29	103,42
1996	3,1	3,36	92,33
1997	3,9	3,43	113,79
1998	3,3	3,50	94,35
1999	3,2	3,57	89,70
2000	4,3	3,64	118,21
2001	3,9	3,71	105,19
2002	3,5	3,78	92,65
2003	3,6	3,85	93,57
2004	3,7	3,92	94,45
2005	4	3,99	100,31
2006	3,6	4,06	88,72
2007	4,1	4,13	99,33
2008	4,7	4,20	111,97
2009	4,2	4,27	98,42
2010	4,5	4,34	103,75

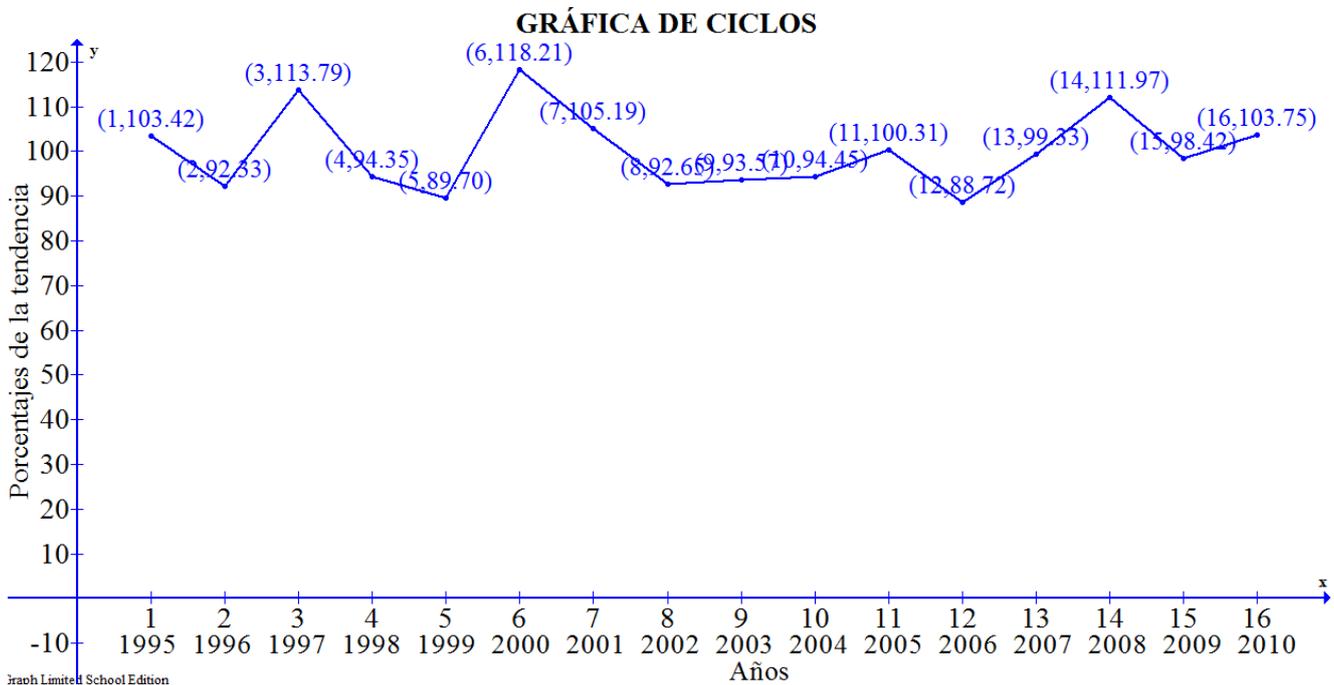
Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Año (X)	X	Y	XY		X ²		Y ²		$Y_{est} = 3,22 + 0,07X$	$C \cdot I = (Y/Y_{est}) \cdot 100$	
2	1995	1	3,4	3,40	=B2*C2	1	=B2^2	11,56	=C2^2	3,29	103,42	=(C2/J2)*100
3	1996	2	3,1	6,20	=B3*C3	4	=B3^2	9,61	=C3^2	3,36	92,33	=(C2/J2)*101
4	1997	3	3,9	11,70	=B4*C4	9	=B4^2	15,21	=C4^2	3,43	113,79	=(C2/J2)*102
5	1998	4	3,3	13,20	=B5*C5	16	=B5^2	10,89	=C5^2	3,50	94,35	=(C2/J2)*103
6	1999	5	3,2	16,00	=B6*C6	25	=B6^2	10,24	=C6^2	3,57	89,70	=(C2/J2)*104
7	2000	6	4,3	25,80	=B7*C7	36	=B7^2	18,49	=C7^2	3,64	118,21	=(C2/J2)*105
8	2001	7	3,9	27,30	=B8*C8	49	=B8^2	15,21	=C8^2	3,71	105,19	=(C2/J2)*106
9	2002	8	3,5	28,00	=B9*C9	64	=B9^2	12,25	=C9^2	3,78	92,65	=(C2/J2)*107
10	2003	9	3,6	32,40	=B10*C10	81	=B10^2	12,96	=C10^2	3,85	93,57	=(C2/J2)*108
11	2004	10	3,7	37,00	=B11*C11	100	=B11^2	13,69	=C11^2	3,92	94,45	=(C2/J2)*109
12	2005	11	4	44,00	=B12*C12	121	=B12^2	16,00	=C12^2	3,99	100,31	=(C2/J2)*110
13	2006	12	3,6	43,20	=B13*C13	144	=B13^2	12,96	=C13^2	4,06	88,72	=(C2/J2)*111
14	2007	13	4,1	53,30	=B14*C14	169	=B14^2	16,81	=C14^2	4,13	99,33	=(C2/J2)*112
15	2008	14	4,7	65,80	=B15*C15	196	=B15^2	22,09	=C15^2	4,20	111,97	=(C2/J2)*113
16	2009	15	4,2	63,00	=B16*C16	225	=B16^2	17,64	=C16^2	4,27	98,42	=(C2/J2)*114
17	2010	16	4,5	72,00	=B17*C17	256	=B17^2	20,25	=C17^2	4,34	103,75	=(C2/J2)*115
18												
19	Total	136	61	542,30		1496		235,86				
20		=SUMA(B2:B18)	=SUMA(C2:C17)	=SUMA(D2:D17)		=SUMA(F2:F18)		=SUMA(H2:H17)				
21	N	16	=CONTAR(B2:B17)									
22												
23	$a_0 = \frac{\sum Y - \sum X^2 - \sum X \cdot \sum XY}{N \sum X^2 - (\sum X)^2}$			3,22								
24												
25	$a_1 = \frac{N \sum YX - \sum X \cdot \sum Y}{N \sum X^2 - (\sum X)^2}$			0,070								
26												
27												

2) La gráfica de ciclos elaborada empleando Excel se muestra en la siguiente figura:



La gráfica de ciclos elaborada empleando Graph se muestra en la siguiente figura:



Graph Limited School Edition

TAREA DE INTERAPRENDIZAJE N° 29

- 1) Elabore un organizador gráfico sobre el análisis de movimientos cíclicos e irregulares
- 2) Con los siguientes datos sobre las exportaciones en millones de dólares de la Empresa M & D

Año (X)	Ventas (Y)
2000	4,3
2001	3,9
2002	3,5
2003	3,6
2004	3,7
2005	4
2006	3,6
2007	4,1
2008	4,7
2009	4,2
2010	4,5

2.1) Determine el componente cíclico de cada uno de los valores de la serie cronológica usando la ecuación de tendencia calculada por el método de los mínimos cuadrados de manera manual y empleando Excel.

$C \cdot I = (Y/Y_{est}) \cdot 100\%$
116,07
103,57
91,47
92,61
93,71
99,77
88,45
99,25
112,12
98,76
104,32

2.2) Elabore una gráfica de ciclos de manera manual, empleando Excel y Graph.



3) Cree un ejercicio de aplicación de análisis de movimientos cíclicos e irregulares con datos reales sobre cualquier tema de su preferencia. Resuélvalo de manera manual, empleando Excel y Graph. También elabore la gráfica de ciclos.

REFERENCIAS BIBLIOGRÁFICAS

- BENALCÁZAR, Marco, (2002), Unidades para Producir Medios Instruccionales en Educación, SUÁREZ, Mario Ed. Graficolor, Ibarra, Ecuador.
- DAZA, Jorge, (2006), Estadística Aplicada con Microsoft Excel, Grupo Editorial Megabyte, Lima, Perú.
- GOVINDEN, Lincoyán, (1985), Introducción a la Estadística, Ed. McGraw Hill. Interamericana Editores. S.A., Bogotá, Colombia.
- JOHNSON, Robert, (2003), Estadística Elemental, Ed. Math Learning, Ed. Tercera, México DF.
KUBY, Patricia.
- KAZMIER, J. Leonard, (1990). Estadística Aplicada a la Administración y la Economía, Ed. McGrawHill, Ed. Segunda, Bogotá, Colombia.
- LIND, Marchal, (2005), Estadística Aplicada a los Negocios y a la Economía, Ed. McGraw- Hill, MASON Ed. Décima., Mexico DF.
- MARTINEZ, Bencardino, (1981), Estadística Comercial, Ed. Norma, Bogotá, Colombia.
- MORENO, Francis, (1993), Estadística Inferencial, Universidad Particular de Loja, Loja, Ecuador.
- SÁNCHEZ, Jesús, (2007), Introducción a la Estadística Empresarial, Madrid, España.
- SALTOS, Hector, (1986), Estadística de Inferencia, Ed. Pío XII, Ambato, Ecuador.
- SHAO, Stephen, (1980), Estadística para Economistas y Administradores de Empresas, Ed. Herrero Hnos, México DF.
- SPIEGEL, Murray, (2000), Estadística, Serie de Compendios Schaum, Ed. McGraw-Hill, México.
- SUÁREZ, Mario, (2004), Interaprendizaje Holístico de Matemática, Ed. Gráficas Planeta, Ibarra, Ecuador.
- SUÁREZ, Mario, (2004), Hacia un interaprendizaje de Álgebra y Geometría, Ed. Gráficas Planeta, Ibarra, Ecuador.
- SUÁREZ Mario, (2011), Distribución de frecuencias para datos agrupados en intervalos, www.monografias.com/trabajos87/.
- SUÁREZ, Mario, (2011), Cálculo del tamaño de la muestra, www.monografias.com/trabajos87/
- SUÁREZ, Mario, (2011), Gráficos estadísticos básicos, www.monografias.com/trabajos88/

SUÁREZ, Mario, (2011), Guía didáctica para el interaprendizaje de medidas de tendencia central, www.monografias.com/trabajos85/

SUÁREZ, Mario, (2011), Media aritmética, www.monografias.com/trabajos85/

SUÁREZ, Mario, (2011), Ejemplos ilustrativos resueltos de la Moda, www.monografias.com/trabajos85/

SUÁREZ, Mario, (2011), La mediana para datos no agrupados y agrupados, www.monografias.com/trabajos87/

SUÁREZ, Mario, (2011), Medidas de posición, www.monografias.com/trabajos87/

SUÁREZ, Mario, (2011), Medidas de dispersión, www.monografias.com/trabajos89/

SUÁREZ, Mario, (2011), Medidas de forma: asimetría y curtosis, www.monografias.com/trabajos87/

SUÁREZ, Mario, (2011), Coefficiente de correlación de Pearson para datos agrupados en intervalos, www.monografias.com/trabajos86/

SUÁREZ, Mario, (2011), Coefficiente de Correlación por Rangos de Spearman, www.monografias.com/trabajos85/

SUÁREZ, Mario, (2011), Dispersión relativa o coeficiente de variación, www.monografias.com/trabajos88/

SUÁREZ, Mario, (2011), La recta de los mínimos cuadrados, www.monografias.com/trabajos85/

SUÁREZ, Mario (2011), Análisis de regresión mediante la parábola de los mínimos cuadrados, www.monografias.com/trabajos86/

SUÁREZ, Mario (2011), Regresión potencial mediante el método de los mínimos cuadrados, www.monografias.com/trabajos89/

SUÁREZ, Mario (2011), Regresión exponencial mediante el método de los mínimos cuadrados, www.monografias.com/trabajos89/

SUÁREZ, Mario, (2011), Análisis de tendencia para series de tiempo, www.monografias.com/trabajos87/

SUÁREZ, Mario, (2011), Métodos de suavizamiento y pronóstico para series de tiempo, www.monografias.com/trabajos87/

STEVENSON, William, (1981), Estadística para Administración y Economía, Ed. Harla S.A de C.V. México D.F.

WEBSTER, Allen, (2000), Estadística Aplicada a los Negocios y a la Economía, Ed. McGraw Hill. Interamericana Editores S.A. Bogotá, Colombia.