



“Un proceso pertinente de
formación para la vida”

COLEGIO DE BACHILLERES

**Guía para presentar exámenes de
Recuperación o Acreditación Especial
(Apoya a Plan 92)**

**ESTADÍSTICA DESCRIPTIVA E
INFERENCIAL I**

Guía para presentar exámenes de
Recuperación o Acreditación Especial

Estadística Descriptiva e Inferencial I
(Versión preliminar)

Esta guía fue elaborada por la **Secretaría Académica**, a través de la **Dirección de Planeación Académica**.

Colaborador

Profr. Alejandro Rosas Snell

Colegio de Bachilleres, México
www.cbachilleres.edu.mx
Rancho Vista Hermosa No. 105
Ex-Hacienda Coapa,
04920, México, D.F.

La presente obra fue editada en el procesador de palabras Word 2002 (Office xp).

Word 2002 es marca registrada de Microsoft Corp.

Este material se utiliza en el proceso de enseñanza-aprendizaje del Colegio de Bachilleres, institución pública de educación media superior del Sistema Educativo Nacional.

Ninguna parte de esta publicación, incluido el diseño de la cubierta, puede reproducirse, almacenarse o transmitirse en forma alguna, ni tampoco por medio alguno, sea éste eléctrico, electrónico, químico, mecánico, óptico, de grabación o de fotocopia, sin la previa autorización escrita por parte del Colegio de Bachilleres, México.

SEPTIEMBRE 2004

ÍNDICE

PRESENTACIÓN	V
PRÓLOGO	VI
	1
UNIDAD 1. Introducción: Usos y funciones de la estadística	
1.1 Antecedentes de la estadística	3
Aplicación del conocimiento.....	4
Ejercicios.	5
Tabla de Comprobación	6
1.2 Conceptos preliminares	7
Aplicación del conocimiento.....	9
Ejercicios.	10
Tabla de Comprobación	12
Ejercicios de autoevaluación	13
Clave de respuesta	14
UNIDAD 2. Presentación de Datos	15
2.1 Distribución de frecuencias	17
Aplicación del conocimiento.....	20
Ejercicios.	22
Tabla de Comprobación	28
2.2 Gráficas	32
Aplicación del conocimiento.....	35
Ejercicios.	37
Tabla de Comprobación	39
Ejercicios de autoevaluación	41
Clave de respuesta	45
UNIDAD 3. Medidas descriptivas	49
3.1 Medidas de tendencia central	51
Aplicación del conocimiento.....	56
Ejercicios.	60
Tabla de Comprobación	63
3.2 Medidas de dispersión	64
Aplicación del conocimiento.....	66
Ejercicios.	69
Tabla de Comprobación	73
3.3 Medidas de tendencia central y de dispersión en curvas de frecuencia	74
Aplicación del conocimiento.....	76
Ejercicios.	77
Tabla de Comprobación	78
Ejercicios de autoevaluación	79
Clave de respuesta	81

UNIDAD 4. Correlación y regresión lineal.	83
4.1 Correlación lineal.	85
Aplicación del conocimiento.....	87
Ejercicios.	91
Tabla de Comprobación	94
4.2 Regresión lineal.	97
Aplicación del conocimiento.....	98
Ejercicios.	100
Tabla de Comprobación	103
Ejercicios de autoevaluación.	106
Clave de respuesta.	108
UNIDAD 5. Elementos de probabilidad.	110
5.1 y 5.2 Introducción a la probabilidad y frecuencia relativa.	112
Aplicación del conocimiento.....	114
Ejercicios.	115
Tabla de Comprobación	117
5.3 Nociones de probabilidad.	118
Aplicación del conocimiento.....	120
Ejercicios.	122
Tabla de Comprobación	125
5.4 Cálculo de probabilidades: procedimientos elementales de conteo.	126
Aplicación del conocimiento.....	128
Ejercicios.	130
Tabla de Comprobación	132
Ejercicios de autoevaluación.	133
Clave de respuesta.	135
BIBLIOGRAFÍA	136
SUGERENCIAS PARA PRESENTAR EXÁMENES DE RECUPERACIÓN O ACREDITACIÓN ESPECIAL.	137

PRESENTACIÓN

La evaluación de recuperación y la de acreditación especial son oportunidades extraordinarias que debes aprovechar para aprobar las asignaturas que, por diversas razones, reprobaste en el curso normal; pero ¡cuidado!, presentarse a un examen sin la preparación suficiente significa un fracaso seguro, es una pérdida de tiempo y un acto irresponsable que puedes evitar.

¿Cómo aumentar tu probabilidad de éxito en el examen mediante la utilización de esta guía? La respuesta es simple, observa las siguientes reglas:

- Convéncete de que tienes la capacidad necesaria para acreditar la asignatura. Recuerda que fuiste capaz de ingresar al Colegio de Bachilleres mediante un examen de selección.
- Sigue al *pie de la letra* las instrucciones de la guía.
- Procura dedicarte al estudio de este material, *durante 15 días al menos, tres horas diarias continuas*.
- Contesta toda la guía: es un requisito que la presentes resuelta y en limpio al profesor aplicador antes del examen correspondiente.

PROLOGO

En el marco del Programa de Desarrollo Institucional 2001-2006 el **alumno** tiene especial relevancia, por lo que el Colegio de Bachilleres Metropolitano se ha abocado a la elaboración de diversos materiales didácticos que apoyen al estudiante en los diversos momentos del proceso de enseñanza-aprendizaje.

Entre los materiales elaborados se encuentran las guías de estudio, las cuales tienen como propósito apoyar a los estudiantes que deben presentar exámenes de Recuperación o Acreditación Especial, con objeto de favorecer el éxito en los mismos.

En este contexto, la Guía para presentar exámenes de Recuperación o Acreditación Especial de **Estadística Descriptiva e Inferencial I**, se ha elaborado pensando en los estudiantes que por diversas causas reprobaron la asignatura en el curso normal y deben acreditarla a través de exámenes en periodos extraordinarios.

Esta guía se caracteriza por abordar, de manera sintética, los principales temas señalados en el programa de estudios, favorecer la ejercitación de los métodos, conceptos y modelos estadísticos en el manejo e interpretación cuantitativa y cualitativa de información diversa, además de los elementos básicos de la probabilidad estadística, así como proporcionar elementos de autoevaluación y sugerencias en caso de que se necesite mayor información para comprender dichos temas.

En la primera unidad de la guía, denominada **INTRODUCCIÓN: USOS Y FUNCIONES DE LA ESTADÍSTICA**, se abordan los aprendizajes relacionados con las características y conceptos básicos de la estadística (población, muestra, dato, estimación, etc.), además, se incluyen problemas donde se ejercita su uso adecuado.

En la segunda unidad, **PRESENTACIÓN DE DATOS**, se desarrollan aprendizajes sobre organización de información, tablas de frecuencia y gráficos estadísticos, También se ejercita su aplicación en diversos problemas.

La tercera unidad, **MEDIDAS DESCRIPTIVAS**, abarca el tema de las medidas de tendencia central: media, mediana y moda para datos agrupados y no agrupados, así como los procedimientos para calcular la desviación media, la varianza y la desviación estándar con datos agrupados.

La cuarta unidad, **CORRELACIÓN Y REGRESIÓN LINEAL**, contiene los aprendizajes sobre el coeficiente de correlación de Pearson, la obtención de la ecuación de regresión lineal y la aplicación de la correlación y regresión lineal en la solución de problemas.

En la quinta unidad, **ELEMENTOS DE PROBABILIDAD**, se desarrollan los elementos básicos de la probabilidad, el cálculo probabilístico de eventos excluyentes y no mutuamente excluyentes, se estudian los procedimientos para conocer el número de permutaciones y combinaciones con y sin repetición para datos ordenados y, como en las demás unidades, se presentan una serie de ejercicios en donde se aplican los conceptos y métodos estadísticos desarrollados.

Por último, se proporciona una bibliografía básica para consultar en fuentes originales los temas desarrollados en la guía.

UNIDAD 1
INTRODUCCIÓN:
USOS Y FUNCIONES
DE LA ESTADÍSTICA

1.1 ANTECEDENTES DE LA ESTADÍSTICA

APRENDIZAJES

- Identificar las funciones básicas de la estadística.

La palabra estadística proviene de una voz italiana, *statista*, que significa estadista. La acuñó G. Achenwall (1719-1772), un profesor de matemáticas. Su uso lo divulgó sir John Sinclair en su obra *Statistical account of scotland* (1791-1799); pero ya mucho antes del siglo XVIII se acostumbraba registrar y emplear datos.

La Estadística surge de la necesidad del hombre por interpretar el conjunto de datos que le proporciona la realidad, por medio de procedimientos que le permitan describir el comportamiento de un fenómeno en estudio, que bien pueden ser las edades de un grupo de personas, ingresos, gastos de alimentación, estaturas, etc.

Los juegos de azar también fueron motivo para que los matemáticos desarrollaran principios para aumentar la posibilidad (probabilidad) de ganar.

La Estadística es una rama de la ciencia Matemática que se encarga de la aplicación de ciertos procedimientos matemáticos a la información producida en el proceso de investigación; es decir, **la estadística es el conjunto de procedimientos matemáticos que permiten captar, clasificar, ordenar, procesar y analizar la información que se produce en el proceso de investigación, donde midiendo la regularidad de la misma se puede inferir la tendencia o comportamiento del fenómeno en estudio, construyendo modelos que se ajustan a su dinámica.**

La aplicación de la Estadística se da en varios niveles de complejidad; desde los que requieren técnicas muy sofisticadas hasta los que sólo necesitan, por ejemplo, la organización de un conjunto de datos en tablas, la construcción de algunas gráficas o el cálculo de ciertos promedios.

La Estadística son los **métodos** que se aplican a la **recolección, organización, presentación, análisis e interpretación de los datos numéricos**. Con base en lo anterior, las funciones de la Estadística Descriptiva e Inferencial (etapas de la investigación estadística) son:

- | | | |
|-------------------------|---|--|
| ESTADÍSTICA DESCRIPTIVA | { | <ol style="list-style-type: none"> 1) Recopilar o reunir los datos o información necesaria. 2) Organizar los datos o información obtenida. 3) Presentar los datos o información en forma ordenada (tablas y gráficas). 4) Analizar los datos o información presentada (describir). |
| ESTADÍSTICA INFERENCIAL | { | <ol style="list-style-type: none"> 5) Hacer inferencias de los datos o información analizada (obtener conclusiones). |

APLICACIÓN DEL CONOCIMIENTO

Utiliza los contenidos que acabas de leer sobre las funciones de la estadística y contesta las siguientes preguntas.

Se investiga el número de calificaciones aprobatorias de los seis semestres del Plan de Estudios del Plantel 4 del Colegio de Bachilleres en las materias de Matemáticas, Estadística Descriptiva e Inferencial y de Cálculo Diferencial e Integral. Para ello se requiere que ordenes las etapas de la investigación estadística, anotando en el paréntesis de la izquierda los números del 1 al 5 según corresponda.

- (2) Organizar la información obtenida.
- (4) Analizar la información obtenida.
- (5) Realizar inferencias de la información obtenida.
- (3) Presentar la información obtenida en forma ordenada.
- (1) Reunir o recopilar la información necesaria.

A continuación menciona ordenadamente las etapas de la investigación estadística para saber cuál es la población femenina activamente trabajadora en la delegación de Coyoacán.

EJERCICIOS

INSTRUCCIONES: Lee con atención los siguientes reactivos y anota en el paréntesis de la izquierda la letra que corresponda a la respuesta correcta.

1. () ¿Cuál es nombre que se le da a la rama de la Matemática que se encarga de la aplicación de ciertos procedimientos matemáticos a la información producida en el proceso de investigación?
 - a) Investigación de operaciones.
 - b) Aritmética.
 - c) Geometría Analítica.
 - d) Estadística.

2. () ¿Cuál es el nombre que se le da a los métodos que se aplican en la recolección, organización, presentación, análisis e inferencia de datos numéricos?
 - a) Cálculo numérico.
 - b) Álgebra.
 - c) Estadística.
 - d) Matemáticas aplicadas.

INSTRUCCIONES: Lee con atención el siguiente reactivo y contesta lo que se solicita.

3. Ordena las siguientes etapas de la investigación estadística, utilizando del 1 al 5, si se desea realizar un estudio con respecto a los trabajadores de la empresa TV Azteca que estudian y trabajan al mismo tiempo, y de sus edades correspondientes.
 - () Realizar inferencias de la información obtenida (conclusiones).
 - () Analizar la información presentada (describir).
 - () Presentar la información en forma ordenada (tablas y gráficas).
 - () Organizar la información obtenida.
 - () Recolectar o reunir la información necesaria.

TABLA DE COMPROBACIÓN

Número de pregunta	Respuesta correcta
1	d
2	c
3	5 4 3 2 1
Sugerencias	
<p>Repasa lo los temas sobre funciones de la estadística y las etapas de la investigación estadística en los siguientes textos.</p> <p>Estadística Elemental. Robert Johnson. Iberoamérica, México, p. 4-12</p> <p>Estadística Primer Curso. Enrique Portilla. Interamericana, México, p. 2-7</p>	

1.2. CONCEPTOS PRELIMINARES

APRENDIZAJES

- Diferenciar los rasgos esenciales de la estadística descriptiva e inferencial.
- Comprender los conceptos básicos de la estadística tales como: población, muestra, escala de medición, dato, encuesta y estimación.

El **campo de la Estadística** generalmente está dividido en **dos áreas**: Estadística Descriptiva o Deductiva y Estadística Inferencial o Inductiva.

Estadística descriptiva o deductiva: Es aquella **parte de la investigación estadística** que incluye la obtención, organización, presentación y descripción de información numérica.

Estadística inferencial o inductiva: Es una **técnica** mediante la cual se obtienen generalizaciones, o se toman decisiones con base en una información parcial o incompleta obtenida mediante técnicas descriptivas.

En resumen, la Estadística como disciplina o área de estudio, incluye tanto técnicas descriptivas como inferenciales. Incluye la obtención y tratamiento de datos numéricos y el empleo de métodos estadísticos con fines inferenciales, por lo tanto:

"La Estadística es el conjunto de procedimientos matemáticos que permiten captar, clasificar, ordenar, procesar y analizar la información que se produce en el proceso de investigación, donde midiendo la regularidad de la misma, se puede inferir la tendencia o comportamiento del fenómeno en estudio, construyendo modelos que se ajusten a su dinámica."

La estadística no solamente se dedica a la descripción; sus funciones son mucho más importantes, pues también podemos inferir propiedades de la población a partir del estudio de una parte de ella, es decir, de una muestra. Por lo que:

Población se define como la totalidad de todas las posibles observaciones.

Muestra se define como un conjunto de observaciones tomadas a partir de una población.

Por ejemplo, supongamos que se desea estudiar el coeficiente intelectual de los jóvenes de 2° semestre de bachillerato. Una forma de hacerlo es proporcionándoles a todos un test de inteligencia. Una vez realizado y calificado, se calcula la puntuación media del grupo (la media es un estadígrafo que refleja el promedio de las puntuaciones del test en dicha muestra). Como hay muchos alumnos, habrá de continuar el proceso; es decir, se obtiene muestra tras muestra.

Si cada uno de estos grupos es una muestra aleatoria, se pueden combinar todos los promedios o medias de las muestras para obtener una media total. Esta media es la mejor estimación del coeficiente intelectual de los jóvenes de 2° semestre de bachillerato.

Esto es, el promedio de las medias de las muestras se utiliza para conocer el valor de la población. Todos

los jóvenes de 2° semestre de bachillerato de México forman la población o universo, del cual se extraen muestras que pueden ser los planteles del Colegio de Bachilleres del DF o los planteles de las preparatorias del Estado de México.

Los valores que se refieren a la población se llaman **parámetros** y los que se refieren a las muestras se llaman **estadígrafos**.

En la ciencia se utilizan varios tipos de escalas para efectuar mediciones; estas escalas son:

Cardinal o nominal. Esta escala se utiliza como medida de identificación por ejemplo: los números que llevan los jugadores en la espalda; cuando se clasifica a una persona por sus ideas religiosas: protestante, católico, judío u otra; también se puede clasificar por sexo, color de ojos, partido político, zona urbana o rural, etc. Las estadísticas simples se realizan con datos cardinales.

Ordinal. En esta escala, los números reflejan el orden de las personas u objetos; estas medidas se disponen en orden de mayor a menor o viceversa.

Intervalos. Esta escala proporciona números que reflejan las diferencias entre particularidades por ejemplo: en termómetros, grados Celsius y Fahrenheit; en test, etc.

Razones. Esta escala tiene un cero absoluto, por ejemplo: las medidas de longitud, peso, capacidad, sonoridad, etc.

El dato. Es el valor de la variable asociado a un elemento de una población o una muestra. Por ejemplo, José Pérez ingresó a la Escuela Superior de Medicina a los 21 años, su cabello es oscuro, mide 1.89 metros y pesa 75 kilogramos. Cada una de estas cuatro mediciones es un valor individual o dato para cada una de las variables de respuesta medidas para José Pérez.

La característica que se puede analizar en un elemento de la población puede ser de tipo cualitativo o cuantitativo. El primer tipo de dato da lugar a la estadística de atributos, puesto que se está analizando el sexo, estado civil, las preferencias de un producto, etc., las cuales son cualificables.

Cuando la característica a estudiar de un dato es cuantitativa, daremos lugar a la estadística de variables, pues estas características toman valores distintos y pueden medirse; tal es el caso de la edad, ingresos, gasto, peso, estatura, etc. Así, los datos cualitativos o de atributo son el resultado de un proceso que categoriza o describe un elemento de una población; por ejemplo: el color es un atributo de un automóvil. Una muestra de los colores de los automóviles que están en un color es un atributo de un automóvil. Una muestra de los colores de los automóviles que están en un estacionamiento produciría datos como azul, rojo, amarillo, gris, etc. Los datos cuantitativos o numéricos, son el resultado de un proceso que cuantifica, es decir, que cuenta o mide.

La estadística de variables puede ser del tipo discreto (se obtiene contando) o del tipo continuo (se obtiene midiendo).

Los aspectos que se desean conocer de una población son, en general, la **estimación** de un promedio, de un porcentaje o la prueba de hipótesis. La estimación y la prueba de hipótesis son dos partes importantes de la estadística inferencial.

Las técnicas de estimación permiten, como su nombre lo indica, estimar parámetros mediante valores, por ejemplo; puede estimarse la resistencia promedio de ciertas piezas producidas en una fábrica o el porcentaje de electores que votarán para el candidato en futuras elecciones.

APLICACIÓN DEL CONOCIMIENTO

Observa cómo se aplican los conocimientos que acabas de leer sobre los conceptos estadísticos de muestra y población, para resolver el siguiente problema.

Un fabricante de medicamentos desea conocer la proporción de personas cuya hipertensión (alta presión sanguínea) puede ser controlada con un nuevo producto fabricado por la compañía. Al realizar un estudio en 5 000 personas hipertensas se encontró que 69% de ellos pudo controlar su hipertensión utilizando el nuevo medicamento. Suponiendo que estas personas son representativas del grupo de pacientes de hipertensión, contesta las siguientes preguntas.

a) ¿Cuál es la población?

Todas las personas que padecen la hipertensión arterial (es un número muy grande pero desconocido).

b) ¿Cuál es la muestra?

Son las 5 000 personas del estudio.

Aplicando los conocimientos anteriores responde la siguiente pregunta.

¿Cuál es el nombre que se les da a las variables que se obtienen contando?

_____.

EJERCICIOS

INSTRUCCIONES: Lee con atención los siguientes planteamientos y contesta lo que se solicita.

1. Un estudiante de estadística desea tener una idea aproximada acerca del valor (en pesos) del automóvil típico que poseen los profesores de su escuela. Para este caso, contesta:

I) ¿Cuál es la población? _____.

II) ¿Cuál es la muestra? _____.

2. Se desea saber el costo de la educación. Uno de los gastos que hace un estudiante es la compra de sus libros de texto. Sea x el costo de todos los libros comprados este semestre por cada estudiante de cierta escuela superior. Con base en lo anterior describe:

I) ¿Cuál es la población? _____.

II) ¿Cuál es la variable? _____.

3. El Colegio de Bachilleres tiene una población aproximada de 40 000 alumnos y en los primeros cinco planteles hay 18 000 alumnos. Con base en lo anterior describe:

I) ¿Cuál es el parámetro? _____.

II) ¿Cuál es el estadígrafo? _____.

INSTRUCCIONES: Lee con atención cada uno de los siguientes reactivos y anota en el paréntesis de la izquierda la letra de la opción de la respuesta correcta.

4. () El área de la estadística que estudia las características de un grupo de datos para conocer los valores que los describen, se llama:

- a) estadística descriptiva o deductiva.
- b) probabilidad.
- c) estadígrafo.
- d) cálculo numérico.

5. () ¿Cuál es el nombre del área de la estadística que analiza los datos de una muestra para conocer, a partir de estos datos, las características de la población de la cual se tomaron?

- a) Estadística descriptiva o deductiva.
- b) Probabilidad.
- c) Estadígrafo.
- d) Estadística inferencial o inductiva.

6. () ¿Cuál es el nombre del conjunto de elementos extraídos de un total para realizar un estudio?

- a) Estadígrafo.
- b) Población.
- c) Dato.
- d) Muestra.

7. () ¿Cuál es el nombre del total de todos los elementos que tienen una característica en común?
- a) Estadística.
 - b) Muestra.
 - c) Estadígrafo.
 - d) Población.
8. () ¿Cuál es el nombre de la variable asociada a un elemento de una población o una muestra?
- a) Estadígrafo.
 - b) Dato.
 - c) Población.
 - d) Parámetro.
9. () ¿Cuál es el nombre que se le da a la escala que tiene un cero absoluto?
- a) Nominal.
 - b) Intervalos.
 - c) Razones.
 - d) Dato.

TABLA DE COMPROBACIÓN

Número de pregunta	Respuesta correcta
1	I) La población son todos los vehículos de los profesores de la escuela. II) La muestra es una porción de la población, es decir, el número de automóviles de los profesores de matemáticas.
2	I) La población son todos los estudiantes inscritos en la escuela superior este semestre. II) La variable es la cantidad de dinero gastado por los estudiantes en la compra de libros para este semestre.
3	I) Los parámetros son los valores que se refieren a la población. II) Los estadígrafos son los valores que se refieren a la muestra.
4	a
5	d
6	d
7	d
8	b
9	c
Sugerencias	
Repasa hasta que tengas claro las diferencias y la relación entre población, muestra y estadígrafo.	

EJERCICIOS DE AUTOEVALUACIÓN

Cuentas con cuarenta minutos para resolverlos.

INSTRUCCIONES: Lee con atención los siguientes ejercicios y contesta lo que se solicita, escribiendo en el paréntesis de la izquierda la letra que corresponda a la respuesta correcta.

1. () Cuando no es conveniente considerar todos los elementos de la población, lo que se hace es estudiar una parte de esa población, a la cual se le llama:
 - a) estadígrafo.
 - b) probabilidad.
 - c) muestra.
 - d) estadística.

2. () ¿Cómo se llama la medida o valor susceptible de ser observado y contado?
 - a) Población.
 - b) Dato.
 - c) Muestra.
 - d) Paramétrica.

3. () La primera etapa de la investigación estadística es la que se encarga de recolectar:
 - a) muestras.
 - b) probabilidades.
 - c) poblaciones.
 - d) datos.

4. () ¿Cuál es la estadística que describe los datos y los analiza?
 - a) Inferencial o inductiva.
 - b) Financiera.
 - c) Descriptiva o deductiva.
 - d) No paramétrica.

5. () ¿Cómo se llama la estadística que permite realizar conclusiones?
 - a) Inferencial o inductiva.
 - b) Financiera.
 - c) Descriptiva o deductiva.
 - d) No paramétrica.

6. () ¿Qué tipo de escala utilizamos al clasificar a los alumnos de una escuela por el grado que cursan?
 - a) Cardinal.
 - b) Ordinal.
 - c) Razones.
 - d) Intervalos.

CLAVE DE RESPUESTAS

Número de pregunta	Respuesta correcta
1	c
2	b
3	d
4	c
5	a
6	b

UNIDAD 2

PRESENTACIÓN DE DATOS

2.1 DISTRIBUCIÓN DE FRECUENCIAS

APRENDIZAJES

- Organizar la información contenida en un conjunto de datos en la solución de problemas.
- Construir una tabla de frecuencias absolutas y relativas.

Una vez que se ha obtenido una muestra y se han analizado las observaciones de todas las unidades contenidas en ella, se ha concluido la tarea de obtención de los datos. Sin embargo, lo que se ha obtenido son los datos en bruto, es la **información antes de ser organizada y analizada**. Para manejar información no organizada se comienza con un **arreglo** de datos, es decir, las observaciones se colocan en orden por su magnitud (ascendente o descendente), por ejemplo:

Considérese una muestra de los ingresos ganados en cierto sábado por los estudiantes de bachillerato que trabajan. Suponga que se eligen aleatoriamente 20 de ellos y se obtienen los siguientes salarios:

30 11 42 8 30 18 25 25 17 30
29 21 23 25 15 35 26 13 21 36

El siguiente arreglo muestra las observaciones en forma ascendente.

8	18	25	30
11	21	25	30
13	21	26	35
15	23	29	36
17	25	30	42

Como observas, el arreglo anterior es una forma rudimentaria de organización de datos; sería tan engorrosa como poco práctica si el número de observaciones involucradas fuera mucho mayor. Cuando se trata de grandes cantidades de datos, es recomendable presentarlos en una distribución de frecuencias.

La organización de los datos generalmente implica un arreglo de las observaciones en clases. Al arreglo de los datos para expresar la frecuencia de ocurrencias (repeticiones) de las observaciones en cada una de estas clases se le conoce como **distribución de frecuencias**. Ésta consiste en una representación de las categorías (intervalos) numéricas de la variable junto con el número de entidades que se clasifican en cada categoría (intervalos); estas categorías que no se traslapan y son contiguas se denominan **intervalos de clase**.

Realicemos un ejemplo para mostrar el procedimiento a seguir para construir tablas de distribución de frecuencias absolutas (las frecuencias absolutas de clase son igual al número de observaciones o conteo que pertenecen a la clase tratada) y relativas, las cuales se calculan de la siguiente manera:

frecuencia relativa =

$$\frac{\text{frecuencia absoluta}}{\text{núm. total de observaciones}}$$

Con los datos del ejemplo de los ingresos de 20 estudiantes de bachillerato, construiremos una tabla que muestre la distribución de frecuencias.

30 11 42 8 30 18 25 25 17 30
29 21 23 25 15 35 26 13 21 36

Para organizar la información debemos pensar en cuántos intervalos de clase se necesitan para presentar los datos. El procedimiento es el siguiente:

Localizamos los valores máximo y mínimo de los datos: $X_{\text{máx}} = 42$ y $X_{\text{mín}} = 8$; a la diferencia de estos dos valores se le conoce como **rango o recorrido** de la variable, entonces:

$$(\text{Rango}) R = X_{\text{máx}} - X_{\text{mín}} = 42 - 8 = 34$$

Al rango lo dividimos entre el número de intervalos de clase deseados, obteniendo la **amplitud o longitud del intervalo**, por lo que:

$$\text{Amplitud} = \frac{\text{Rango}}{\text{Número de intervalos}}$$

Donde el número de intervalos se puede calcular de varias formas.

- 1) Eligiendo un número que al resolver el cociente el resultado sea un número cercano a un número entero, por ejemplo:

$$\text{Amplitud} = 34/6 = 5.6666; \quad \text{Amplitud} = 34/7 = 4.85; \quad \text{Amplitud} = 34/8 = 4.25;$$

la amplitud que seleccionamos es la de 4.85, que aproximamos a 5, porque es la más cercana a un número entero siguiente del resultado.

- 2) Calculando la raíz cuadrada de los datos en total, es decir, $\sqrt{20} = 4.47$ que lo aproximamos a 5 y con este dato observamos que el número de intervalos de clase es 7 ($5 = 34/\text{número de intervalos}$).

Recuerda que el número de intervalos de clase debe ser mayor que 5 y menor que 18, por ser un convenio internacional. Ahora bien, con base en la tabla del arreglo de los salarios de los 20 estudiantes de bachilleres, procederemos a elaborar la tabla de distribución de frecuencias correspondiente.

Para el primer intervalo de clase se le suma la amplitud al mínimo valor de los datos (es decir, $8 + 5 = 13$), por lo que el intervalo queda como 8 - 13; el siguiente, 13 - 18; después 18 - 23 y así sucesivamente. Lo anterior lo puedes comprobar examinando la siguiente tabla de las frecuencias.

Tabla: Distribución de las frecuencias absoluta y relativa de los salarios de 20 estudiantes de bachilleres.					
Intervalos de clase	Fronteras inf. sup.	Puntos medios	Conteo	Frecuencias absolutas	Frecuencias relativas
8-13	7.5-12.5	10	//	2	0.10
13-18	12.5-17.5	15	///	3	0.15
18-23	17.5-22.5	20	////	3	0.15
23-28	22.5-27.5	25	/////	5	0.25
28-33	27.5-32.5	30	////	4	0.20
33-38	32.5-37.5	35	//	2	0.10
38-43	37.5-42.5	40	/	1	0.10
Totales				20	1.00

Recuerda que para construir una distribución de frecuencias es necesario:

1) **Seleccionar el número de clases o intervalos de clase**; por lo general, este número es mayor que 5 y menor que 18. Para nuestro ejemplo fue de 7 intervalos de clase. El intervalo de clase casi siempre es un número non y los puntos medios de clase deberán tener el mismo número de dígitos que los datos en bruto; esto indicará tanto el límite superior como el inferior (fronteras) de cada clase, los cuales tendrán un dígito decimal extra (0.5); ninguno de los valores observados deberá caer en estos límites.

2) **Los puntos medios son el promedio de los límites inferior y superior** del intervalo de clase.

3) **Las frecuencias absolutas de clase son igual al número de observaciones** (conteo) que pertenecen a la clase tratada.

4) Las frecuencias relativas de clase se calculan de la siguiente manera:

$$\text{frecuencia relativa} = \frac{\text{frecuencia absoluta}}{\text{número total de observaciones}}$$

5) Las frecuencias absoluta y relativa acumuladas se calculan sumando sucesivamente las frecuencias absoluta y relativa.

APLICACIÓN DEL CONOCIMIENTO

Realicemos otro ejercicio recordando cómo se calcula el rango, la amplitud del intervalo, los límites (fronteras), las frecuencias absolutas y las frecuencias relativas, directamente.

Observa cómo se construye una tabla que muestre las distribuciones de frecuencias absoluta y relativa de una muestra de 50 calificaciones obtenidas en un examen final de estadística.

27 49 68 73 79 84 91 94 107 120
 43 50 71 74 80 84 91 96 108 122
 43 54 71 75 81 86 93 97 108 123
 44 58 71 76 82 88 94 103 116 127
 47 65 73 77 82 88 94 106 120 128

Como te habrás dado cuenta, la información presentada está arreglada, entonces localizamos los valores máximo y mínimo de los datos para calcular el rango:

$$X_{\text{máx}} = 128 \text{ y } X_{\text{mín}} = 27, \text{ por lo que: } R = 128 - 27 = 101$$

Para calcular la amplitud necesitamos un número de intervalos de clase, por lo que se sugiere calcules la raíz cuadrada del total de datos de la muestra, es decir:

$$\sqrt{50} = 7.071; \text{ este cálculo nos da el número aproximado de intervalos de clase, entonces:}$$

$$\text{Amplitud} = \frac{\text{Número de Intervalos}}{\text{Rango}} = 101/7 = 14.42857 \approx 15$$

Distribución de frecuencias absoluta y relativa de una muestra de 50 calificaciones de un examen final de estadística.					
Intervalos de clase	Fronteras inf. y sup.	Puntos medios	Conteo	Frecuencias absolutas	Frecuencias relativas
27-41	26.5-41.5	34	/	1	0.02
42-56	41.5-56.5	49	///////	7	0.14
57-71	56.5-71.5	64	//////	6	0.12
72-86	71.5-86.5	79	////////////////	14	0.28
87-101	86.5-101.5	94	//////////	10	0.20
102-116	101.5-116.5	109	//////	6	0.12
117-131	116.5-131.5	124	//////	6	0.12
Totales				50	1.00

Recuerda que no es necesario incrementar o disminuir el número decimal extra (0.5).

Con base en lo anterior, ahora construye una tabla de distribución de frecuencias absoluta y relativa de 20 estudiantes de una clase de Biología que obtuvieron las siguientes calificaciones.

5	9	4	4
4	2	6	5
3	4	5	6
6	3	7	7
2	8	5	5

Se sugiere utilices 8 intervalos de clase.

EJERCICIOS

INSTRUCCIONES: Analiza los conjuntos de datos de los reactivos 1 a 6, y contesta para cada uno de ellos, las siguientes preguntas:

- I) ¿Cuál es el arreglo ordenado en forma ascendente?
- II) ¿Cuál es el rango o recorrido de los datos?
- III) ¿Cuál es la amplitud o longitud del intervalo?
- IV) ¿Cuál es el número de intervalos adecuados?
- V) Elabora la tabla de distribuciones de frecuencias.

1. El siguiente conjunto de 60 datos representa las ventas de galletas por día en una tienda de autoservicio.

73 48 63 48 60 47 57 75 44 69 48 77 64 51 73 49
66 49 47 61 72 59 56 79 52 68 57 58 54 71 65 61
56 66 45 62 53 44 56 61 62 54 74 55 49 77 53 73
52 65 66 57 44 60 54 55 60 57 51 64

I)

II)

III)

IV)

V)

2. El siguiente conjunto de 150 datos representa el coeficiente intelectual de alumnos de la clase de estadística en el 5to. semestre.

88	91	104	113	125	101	114	104	93	91	96	91	100	94	85
119	91	106	120	129	120	109	106	109	121	126	122	112	92	109
93	89	124	96	105	95	91	114	108	113	107	97	128	125	128
99	120	101	108	118	118	113	124	115	121	120	118	111	121	88
106	106	97	104	105	122	112	103	114	115	115	100	105	108	119
102	127	121	116	100	95	89	108	93	107	118	106	98	119	118
108	89	108	114	102	96	99	105	125	126	100	115	113	116	116
109	104	113	118	110	129	124	92	88	113	100	110	101	103	113
114	106	105	115	98	112	103	101	101	89	109	99	108	111	122
108	114	125	121	122	117	105	112	88	104	97	85	116	113	126

I)

II)

III)

IV)

V)

3. El siguiente conjunto de 18 datos representa las llegadas tarde de ciertos trabajadores en un año: 40, 45, 46, 48, 46, 42, 50, 50, 41, 56, 51, 44, 52, 48, 43, 55, 49 y 47.

I)

II)

III)

IV)

V)

4. El siguiente conjunto de 20 datos representa las llamadas telefónicas en una oficina de gobierno en 20 días hábiles: 16, 22, 8, 5, 13, 17, 6, 11, 13, 24, 15, 5, 14, 15, 15, 18, 17, 20, 9, 7.

I)

II)

III)

IV)

V)

5. El siguiente conjunto de 16 datos representa los cortes de luz en la delegación Tlalpan en 30 días naturales: 19, 14, 8, 7, 18, 7, 24, 12, 15, 26, 12, 19, 15, 23, 15 y 11.

I)

II)

III)

IV)

V)

6. El siguiente conjunto de 20 datos representa las licencias médicas de ciertos trabajadores de un plantel escolar: 12, 3, 6, 5, 7, 6, 5, 1, 5, 4, 3, 4, 12, 7, 10, 8, 7, 9, 11 y 2.

I)

II)

III)

IV)

V)

TABLA DE COMPROBACIÓN

Número de pregunta	Respuesta correcta																																																												
1	<p>I) El arreglo es:</p> <p style="text-align: center;">44 44 44 45 47 47 48 48 48 49 49 49 51 51 52 52 53 53 54 54 54 55 55 56 56 56 57 57 57 57 58 59 60 60 60 61 61 61 62 62 63 64 64 65 65 66 66 66 68 69 71 72 73 73 73 74 75 77 77 79</p> <p>II) Rango = 35</p> <p>III) Amplitud = $4.375 \approx 5$</p> <p>IV) Número de intervalos = 8</p> <p>V)</p> <table border="1" data-bbox="456 1193 1273 1541" style="margin-left: auto; margin-right: auto;"> <caption>Distribución de frecuencias absolutas y relativas de 60 datos.</caption> <thead> <tr> <th>Intervalo de clase</th> <th>Fronteras inf. y sup.</th> <th>Punto medio</th> <th>Conteo</th> <th>Frecuencia absoluta</th> <th>Frecuencia relativa</th> </tr> </thead> <tbody> <tr> <td>44-48</td> <td>43.5-48.5</td> <td>46</td> <td>////////</td> <td>9</td> <td>0.1500</td> </tr> <tr> <td>49-53</td> <td>48.5-53.5</td> <td>51</td> <td>////////</td> <td>9</td> <td>0.1500</td> </tr> <tr> <td>54-58</td> <td>53.5-58.5</td> <td>56</td> <td>//////////</td> <td>13</td> <td>0.2166</td> </tr> <tr> <td>59-63</td> <td>58.5-63.5</td> <td>61</td> <td>////////</td> <td>10</td> <td>0.1666</td> </tr> <tr> <td>64-68</td> <td>63.5-68.5</td> <td>66</td> <td>////////</td> <td>8</td> <td>0.1333</td> </tr> <tr> <td>69-73</td> <td>68.5-73.5</td> <td>71</td> <td>//////</td> <td>6</td> <td>0.1000</td> </tr> <tr> <td>74-78</td> <td>73.5-78.5</td> <td>76</td> <td>////</td> <td>4</td> <td>0.0667</td> </tr> <tr> <td>79-83</td> <td>78.5-83.5</td> <td>81</td> <td>/</td> <td>1</td> <td>0.0167</td> </tr> <tr> <td>Totales</td> <td></td> <td></td> <td></td> <td>60</td> <td>0.9999</td> </tr> </tbody> </table>	Intervalo de clase	Fronteras inf. y sup.	Punto medio	Conteo	Frecuencia absoluta	Frecuencia relativa	44-48	43.5-48.5	46	////////	9	0.1500	49-53	48.5-53.5	51	////////	9	0.1500	54-58	53.5-58.5	56	//////////	13	0.2166	59-63	58.5-63.5	61	////////	10	0.1666	64-68	63.5-68.5	66	////////	8	0.1333	69-73	68.5-73.5	71	//////	6	0.1000	74-78	73.5-78.5	76	////	4	0.0667	79-83	78.5-83.5	81	/	1	0.0167	Totales				60	0.9999
Intervalo de clase	Fronteras inf. y sup.	Punto medio	Conteo	Frecuencia absoluta	Frecuencia relativa																																																								
44-48	43.5-48.5	46	////////	9	0.1500																																																								
49-53	48.5-53.5	51	////////	9	0.1500																																																								
54-58	53.5-58.5	56	//////////	13	0.2166																																																								
59-63	58.5-63.5	61	////////	10	0.1666																																																								
64-68	63.5-68.5	66	////////	8	0.1333																																																								
69-73	68.5-73.5	71	//////	6	0.1000																																																								
74-78	73.5-78.5	76	////	4	0.0667																																																								
79-83	78.5-83.5	81	/	1	0.0167																																																								
Totales				60	0.9999																																																								

Número de pregunta	Respuesta correcta																																																																								
2	<p>I) El arreglo es:</p> <p style="text-align: center;"> 85 85 88 88 88 88 89 89 91 91 91 91 91 92 92 93 93 93 94 95 95 96 96 96 97 97 97 98 98 98 98 99 99 99 100 100 100 100 100 101 101 101 101 101 102 102 103 103 103 104 104 104 104 104 105 105 105 105 105 105 106 106 106 106 106 106 107 107 108 108 108 108 108 108 108 108 109 109 109 109 109 110 110 111 111 112 112 112 112 113 113 113 113 113 113 113 113 114 114 114 114 114 114 115 115 115 115 115 116 116 116 116 117 118 118 118 118 118 118 119 119 119 120 120 120 120 121 121 121 121 121 122 122 122 122 124 124 124 125 125 125 125 126 126 126 127 128 128 129 129 </p> <p>II) Rango = 44</p> <p>III) Amplitud = 5</p> <p>IV) Número de intervalos = 9</p> <p>V)</p> <table border="1" data-bbox="512 1205 1444 1579"> <thead> <tr> <th colspan="6">Distribución de frecuencias absolutas y relativas de 150 datos.</th> </tr> <tr> <th>Intervalo de clase</th> <th>Fronteras Inf. y sup.</th> <th>Punto medio</th> <th>Conteo</th> <th>Frecuencia absoluta</th> <th>Frecuencia relativa</th> </tr> </thead> <tbody> <tr> <td>85-89</td> <td>84.5-89.5</td> <td>87</td> <td>////////</td> <td>8</td> <td>0.0533</td> </tr> <tr> <td>90-94</td> <td>89.5-94.5</td> <td>92</td> <td>//////////</td> <td>11</td> <td>0.0730</td> </tr> <tr> <td>95-99</td> <td>94.5-99.5</td> <td>97</td> <td>//////////</td> <td>15</td> <td>0.01000</td> </tr> <tr> <td>100-104</td> <td>99.5-104.5</td> <td>102</td> <td>//////////</td> <td>20</td> <td>0.1330</td> </tr> <tr> <td>105-109</td> <td>104.5-109.5</td> <td>107</td> <td>//////////</td> <td>27</td> <td>0.1800</td> </tr> <tr> <td>110-114</td> <td>109.5-114.5</td> <td>112</td> <td>//////////</td> <td>22</td> <td>0.1470</td> </tr> <tr> <td>115-119</td> <td>114.5-119.5</td> <td>117</td> <td>//////////</td> <td>19</td> <td>0.1270</td> </tr> <tr> <td>120-124</td> <td>119.5-124.5</td> <td>122</td> <td>//////////</td> <td>16</td> <td>0.1070</td> </tr> <tr> <td>125-129</td> <td>124.5-129.5</td> <td>127</td> <td>//////////</td> <td>12</td> <td>0.0800</td> </tr> <tr> <td colspan="4">Totales</td> <td>150</td> <td>1.0000</td> </tr> </tbody> </table>	Distribución de frecuencias absolutas y relativas de 150 datos.						Intervalo de clase	Fronteras Inf. y sup.	Punto medio	Conteo	Frecuencia absoluta	Frecuencia relativa	85-89	84.5-89.5	87	////////	8	0.0533	90-94	89.5-94.5	92	//////////	11	0.0730	95-99	94.5-99.5	97	//////////	15	0.01000	100-104	99.5-104.5	102	//////////	20	0.1330	105-109	104.5-109.5	107	//////////	27	0.1800	110-114	109.5-114.5	112	//////////	22	0.1470	115-119	114.5-119.5	117	//////////	19	0.1270	120-124	119.5-124.5	122	//////////	16	0.1070	125-129	124.5-129.5	127	//////////	12	0.0800	Totales				150	1.0000
Distribución de frecuencias absolutas y relativas de 150 datos.																																																																									
Intervalo de clase	Fronteras Inf. y sup.	Punto medio	Conteo	Frecuencia absoluta	Frecuencia relativa																																																																				
85-89	84.5-89.5	87	////////	8	0.0533																																																																				
90-94	89.5-94.5	92	//////////	11	0.0730																																																																				
95-99	94.5-99.5	97	//////////	15	0.01000																																																																				
100-104	99.5-104.5	102	//////////	20	0.1330																																																																				
105-109	104.5-109.5	107	//////////	27	0.1800																																																																				
110-114	109.5-114.5	112	//////////	22	0.1470																																																																				
115-119	114.5-119.5	117	//////////	19	0.1270																																																																				
120-124	119.5-124.5	122	//////////	16	0.1070																																																																				
125-129	124.5-129.5	127	//////////	12	0.0800																																																																				
Totales				150	1.0000																																																																				

Número de pregunta	Respuesta correcta																																																
3	<p>I) El arreglo es: 40 41 42 43 44 45 46 46 47 48 48 49 50 50 51 52 55 56</p> <p>II) Rango = 16</p> <p>III) Amplitud = 3</p> <p>IV) Número de intervalos = 6</p> <p>V)</p> <table border="1" data-bbox="512 813 1310 1104"> <caption>Distribución de frecuencias absolutas y relativas de 18 datos.</caption> <thead> <tr> <th>Intervalo de clase</th> <th>Fronteras inf. y sup.</th> <th>Punto medio</th> <th>Conteo</th> <th>Frecuencia absoluta</th> <th>Frecuencia relativa</th> </tr> </thead> <tbody> <tr> <td>40-42</td> <td>39.5-42.5</td> <td>41</td> <td>///</td> <td>3</td> <td>0.1666</td> </tr> <tr> <td>43-45</td> <td>42.5-45.5</td> <td>44</td> <td>///</td> <td>3</td> <td>0.1666</td> </tr> <tr> <td>46-48</td> <td>45.5-48.5</td> <td>47</td> <td>////</td> <td>5</td> <td>0.2777</td> </tr> <tr> <td>49-51</td> <td>48.5-51.5</td> <td>50</td> <td>////</td> <td>4</td> <td>0.2222</td> </tr> <tr> <td>52-54</td> <td>51.5-54.5</td> <td>53</td> <td>/</td> <td>1</td> <td>0.0555</td> </tr> <tr> <td>55-57</td> <td>54.5-57.5</td> <td>56</td> <td>//</td> <td>2</td> <td>0.1111</td> </tr> <tr> <td colspan="4">Totales</td> <td>18</td> <td>0.9999</td> </tr> </tbody> </table>	Intervalo de clase	Fronteras inf. y sup.	Punto medio	Conteo	Frecuencia absoluta	Frecuencia relativa	40-42	39.5-42.5	41	///	3	0.1666	43-45	42.5-45.5	44	///	3	0.1666	46-48	45.5-48.5	47	////	5	0.2777	49-51	48.5-51.5	50	////	4	0.2222	52-54	51.5-54.5	53	/	1	0.0555	55-57	54.5-57.5	56	//	2	0.1111	Totales				18	0.9999
Intervalo de clase	Fronteras inf. y sup.	Punto medio	Conteo	Frecuencia absoluta	Frecuencia relativa																																												
40-42	39.5-42.5	41	///	3	0.1666																																												
43-45	42.5-45.5	44	///	3	0.1666																																												
46-48	45.5-48.5	47	////	5	0.2777																																												
49-51	48.5-51.5	50	////	4	0.2222																																												
52-54	51.5-54.5	53	/	1	0.0555																																												
55-57	54.5-57.5	56	//	2	0.1111																																												
Totales				18	0.9999																																												
4	<p>I) El arreglo es: 5 5 6 7 8 9 11 13 13 14 15 15 15 16 17 17 18 20 22 24</p> <p>II) Rango = 19</p> <p>III) Amplitud = 4</p> <p>IV) Número de intervalos = 5</p> <p>V)</p> <table border="1" data-bbox="528 1509 1302 1776"> <caption>Distribución de frecuencias absolutas y relativas de 20 datos.</caption> <thead> <tr> <th>Intervalo de clase</th> <th>Fronteras inf. y sup.</th> <th>Punto medio</th> <th>Conteo</th> <th>Frecuencia absoluta</th> <th>Frecuencia relativa</th> </tr> </thead> <tbody> <tr> <td>5-8</td> <td>4.5-8.5</td> <td>6.5</td> <td>////</td> <td>5</td> <td>0.2500</td> </tr> <tr> <td>9-12</td> <td>8.5-12.5</td> <td>10.5</td> <td>//</td> <td>2</td> <td>0.1000</td> </tr> <tr> <td>13-16</td> <td>12.5-16.5</td> <td>14.5</td> <td>///////</td> <td>7</td> <td>0.3500</td> </tr> <tr> <td>17-20</td> <td>16.5-20.5</td> <td>18.5</td> <td>////</td> <td>4</td> <td>0.2000</td> </tr> <tr> <td>21-24</td> <td>20.5-24.5</td> <td>22.5</td> <td>//</td> <td>2</td> <td>0.1000</td> </tr> <tr> <td colspan="4">Totales</td> <td>20</td> <td>1.0000</td> </tr> </tbody> </table>	Intervalo de clase	Fronteras inf. y sup.	Punto medio	Conteo	Frecuencia absoluta	Frecuencia relativa	5-8	4.5-8.5	6.5	////	5	0.2500	9-12	8.5-12.5	10.5	//	2	0.1000	13-16	12.5-16.5	14.5	///////	7	0.3500	17-20	16.5-20.5	18.5	////	4	0.2000	21-24	20.5-24.5	22.5	//	2	0.1000	Totales				20	1.0000						
Intervalo de clase	Fronteras inf. y sup.	Punto medio	Conteo	Frecuencia absoluta	Frecuencia relativa																																												
5-8	4.5-8.5	6.5	////	5	0.2500																																												
9-12	8.5-12.5	10.5	//	2	0.1000																																												
13-16	12.5-16.5	14.5	///////	7	0.3500																																												
17-20	16.5-20.5	18.5	////	4	0.2000																																												
21-24	20.5-24.5	22.5	//	2	0.1000																																												
Totales				20	1.0000																																												

Número de pregunta	Respuesta correcta																																										
5	<p>I) El arreglo es: 7 7 8 11 12 12 14 15 15 15 18 19 19 23 24 26</p> <p>II) Rango = 19</p> <p>III) Amplitud = 4</p> <p>IV) Número de intervalos = 5</p> <p>V)</p> <table border="1" data-bbox="529 831 1305 1093"> <caption>Distribución de frecuencias absolutas y relativas de 16 datos.</caption> <thead> <tr> <th>Intervalo de clase</th> <th>Fronteras inf. y sup.</th> <th>Punto medio</th> <th>Conteo</th> <th>Frecuencia absoluta</th> <th>Frecuencia relativa</th> </tr> </thead> <tbody> <tr> <td>7-10</td> <td>6.5-10.5</td> <td>8.5</td> <td>///</td> <td>3</td> <td>0.1875</td> </tr> <tr> <td>11-14</td> <td>10.5-14.5</td> <td>12.5</td> <td>////</td> <td>4</td> <td>0.2500</td> </tr> <tr> <td>15-18</td> <td>14.5-18.5</td> <td>16.5</td> <td>////</td> <td>4</td> <td>0.2500</td> </tr> <tr> <td>19-22</td> <td>18.5-22.5</td> <td>20.5</td> <td>//</td> <td>2</td> <td>0.1250</td> </tr> <tr> <td>23-26</td> <td>22.5-26.5</td> <td>24.5</td> <td>///</td> <td>3</td> <td>0.1875</td> </tr> <tr> <td>Totales</td> <td></td> <td></td> <td></td> <td>16</td> <td>1.0000</td> </tr> </tbody> </table>	Intervalo de clase	Fronteras inf. y sup.	Punto medio	Conteo	Frecuencia absoluta	Frecuencia relativa	7-10	6.5-10.5	8.5	///	3	0.1875	11-14	10.5-14.5	12.5	////	4	0.2500	15-18	14.5-18.5	16.5	////	4	0.2500	19-22	18.5-22.5	20.5	//	2	0.1250	23-26	22.5-26.5	24.5	///	3	0.1875	Totales				16	1.0000
Intervalo de clase	Fronteras inf. y sup.	Punto medio	Conteo	Frecuencia absoluta	Frecuencia relativa																																						
7-10	6.5-10.5	8.5	///	3	0.1875																																						
11-14	10.5-14.5	12.5	////	4	0.2500																																						
15-18	14.5-18.5	16.5	////	4	0.2500																																						
19-22	18.5-22.5	20.5	//	2	0.1250																																						
23-26	22.5-26.5	24.5	///	3	0.1875																																						
Totales				16	1.0000																																						
6	<p>I) El arreglo es: 1 2 3 3 4 4 5 5 5 6 6 7 7 7 8 9 10 11 12 12</p> <p>II) Rango = 11</p> <p>III) Amplitud = 3</p> <p>IV) Número de intervalos = 5</p> <p>V)</p> <table border="1" data-bbox="529 1406 1305 1668"> <caption>Distribución de frecuencias absolutas y relativas de 20 datos.</caption> <thead> <tr> <th>Intervalo de clase</th> <th>Fronteras inf. y sup.</th> <th>Punto medio</th> <th>Conteo</th> <th>Frecuencia absoluta</th> <th>Frecuencia relativa</th> </tr> </thead> <tbody> <tr> <td>1-3</td> <td>0.5-3.5</td> <td>2</td> <td>////</td> <td>4</td> <td>0.2000</td> </tr> <tr> <td>4-6</td> <td>3.5-6.5</td> <td>5</td> <td>///////</td> <td>7</td> <td>0.3500</td> </tr> <tr> <td>7-9</td> <td>6.5-9.5</td> <td>8</td> <td>////</td> <td>5</td> <td>0.2500</td> </tr> <tr> <td>10-12</td> <td>9.5-12.5</td> <td>11</td> <td>////</td> <td>4</td> <td>0.2000</td> </tr> <tr> <td>13-15</td> <td>12.5-15.5</td> <td>14</td> <td></td> <td>0</td> <td>0.0000</td> </tr> <tr> <td>Totales</td> <td></td> <td></td> <td></td> <td>20</td> <td>1.0000</td> </tr> </tbody> </table>	Intervalo de clase	Fronteras inf. y sup.	Punto medio	Conteo	Frecuencia absoluta	Frecuencia relativa	1-3	0.5-3.5	2	////	4	0.2000	4-6	3.5-6.5	5	///////	7	0.3500	7-9	6.5-9.5	8	////	5	0.2500	10-12	9.5-12.5	11	////	4	0.2000	13-15	12.5-15.5	14		0	0.0000	Totales				20	1.0000
Intervalo de clase	Fronteras inf. y sup.	Punto medio	Conteo	Frecuencia absoluta	Frecuencia relativa																																						
1-3	0.5-3.5	2	////	4	0.2000																																						
4-6	3.5-6.5	5	///////	7	0.3500																																						
7-9	6.5-9.5	8	////	5	0.2500																																						
10-12	9.5-12.5	11	////	4	0.2000																																						
13-15	12.5-15.5	14		0	0.0000																																						
Totales				20	1.0000																																						
Sugerencias																																											
<p>Repasa cuidadosamente el procedimiento para elaborar una tabla de frecuencias en: Curso Práctico de Estadística de Portus G., Lincoyán. Mcgraw-Hill, México, pp. 56-60.</p>																																											

2.2 GRÁFICAS

APRENDIZAJES

- Construir gráficas estadísticas tales como: histogramas, polígonos, barras y circulares.

Comúnmente, el significado de los datos se puede comunicar más efectivamente por medio de gráficas que por medio de tablas de frecuencias. Existen cuatro formas para presentar los datos gráficamente: histogramas, polígonos, gráfico o diagrama de barras y gráficas circulares.

HISTOGRAMAS

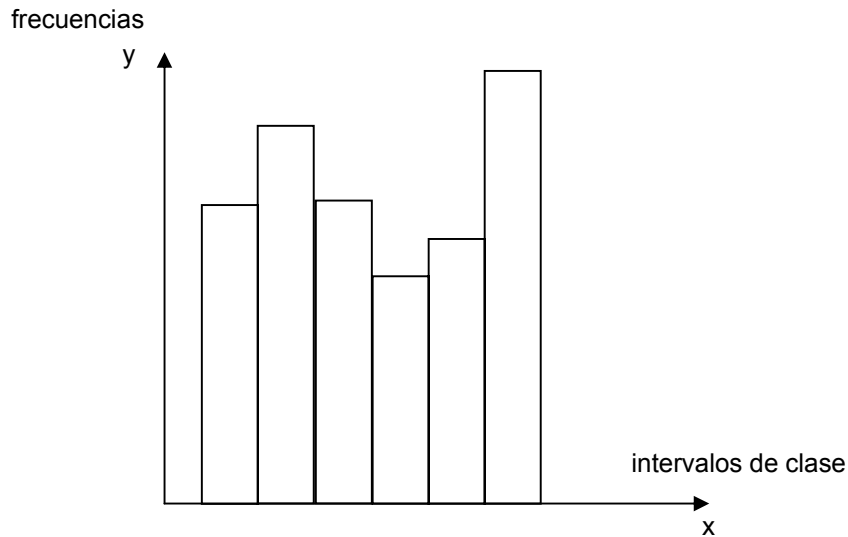
Es un tipo especial de gráfico de barras; se emplea para representar una distribución de frecuencias absoluta y relativa.

El histograma representa la información de las distribuciones de frecuencias, es decir, se representan las frecuencias de clase por medio de áreas de rectángulos (barras). **En un histograma las frecuencias quedan representadas por el área de los rectángulos y no por sus alturas**, por lo que las barras necesariamente se dibujan sin dejar espacios entre ellas.

Para dar generalidad a la impresión visual que proporciona un histograma, los profesionales en estadística recomiendan, para la elección de la longitud de los ejes, aplicar la regla de los tres cuartos, lo que quiere decir que el eje vertical debe ser los tres cuartos de la longitud del eje horizontal.

Para construir un histograma se observan las siguientes características:

- Los valores de las variables se disponen en el eje horizontal (eje x); las frecuencias con que ocurren los valores de la variable en el eje vertical (eje y).
- Cada intervalo de clase se representa por una barra del histograma; recuerda que estas barras tienen la misma amplitud que los intervalos de clase.
- La altura de cada barra corresponde a la frecuencia con que ocurren los valores en dichos intervalos de clase.
- Las barras se dibujan adyacentes entre sí y, recuerda, la proporción encerrada por una barra está representada por la frecuencia relativa correspondiente a dicho intervalo de clase.

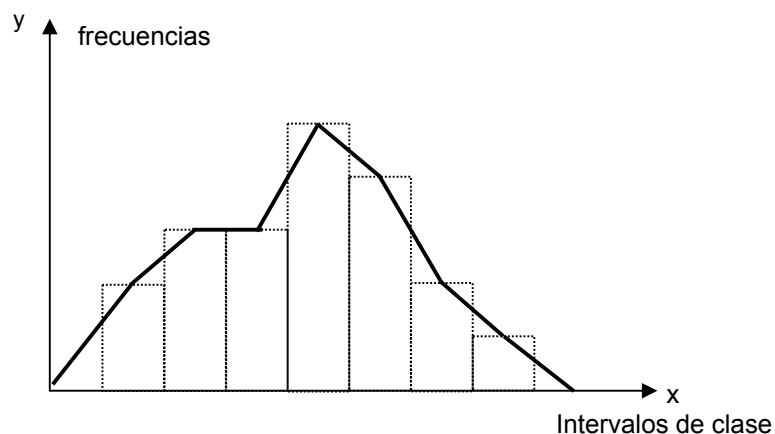


POLÍGONOS

Los datos de una distribución de frecuencias también se pueden graficar a través de un polígono de frecuencias. **Un polígono de frecuencias se construye, elaborando primero un histograma y luego conectando, por medio de líneas rectas, los puntos medios de la parte superior de cada una de las barras.**

Para construir un polígono se observan las siguientes características:

- ❖ Los puntos terminales del histograma se unen con el eje horizontal (eje x) en el punto que corresponde al punto medio de un intervalo de clase imaginaria de la misma amplitud y adyacente al histograma por la izquierda y por la derecha.
- ❖ El área total bajo la curva es igual al área total bajo el histograma correspondiente.
- ❖ Recuerda que la **poligonal**, conocida como **polígono de frecuencias**, se utiliza para mostrar de manera general el comportamiento de la información obtenida, sin la necesidad de realizar cálculos de áreas de rectángulos para verificar que el área bajo la **poligonal** es igual al área de las barras de un histograma.



GRÁFICAS CIRCULARES Y DIAGRAMA DE BARRAS

La entrega de información utilizando gráficas es un verdadero arte funcional, que no sólo sirve para presentar datos, sino también para expresar ideas (ideográficas) que se desea destacar.

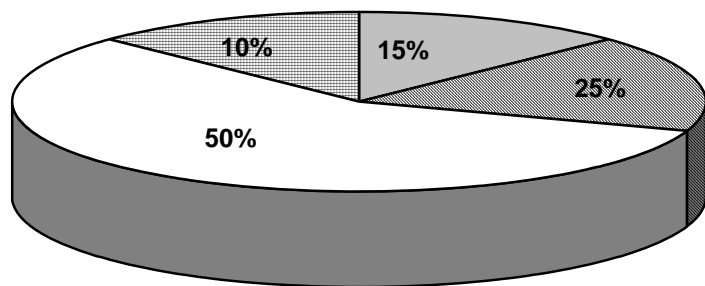
Regularmente, **las gráficas circulares o de pastel se utilizan para representar distribuciones porcentuales**; el círculo completo tiene un área que equivale al 100%, un sector representa un tanto por ciento equivalente a la razón entre el ángulo que forman los radios que limitan al sector y 360, que es el total en grados de la circunferencia; por ejemplo:

$$90/360 = 0.25 = 25\%$$

$$54/360 = 0.15 = 15\%$$

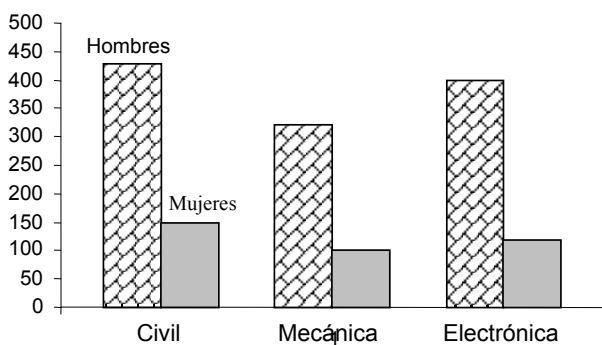
$$36/360 = 0.10 = 10\%$$

$$180/360 = 0.50 = 50\%$$

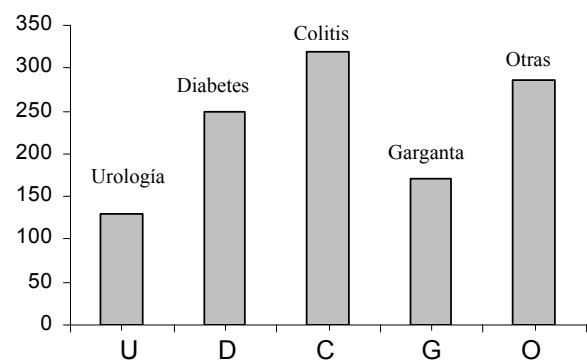


Las gráficas o diagramas de barras proporcionan más información y permiten una observación más rigurosa; **estas gráficas se utilizan para variables nominales y ordinales**; por ejemplo:

Matrícula de la E S I M E



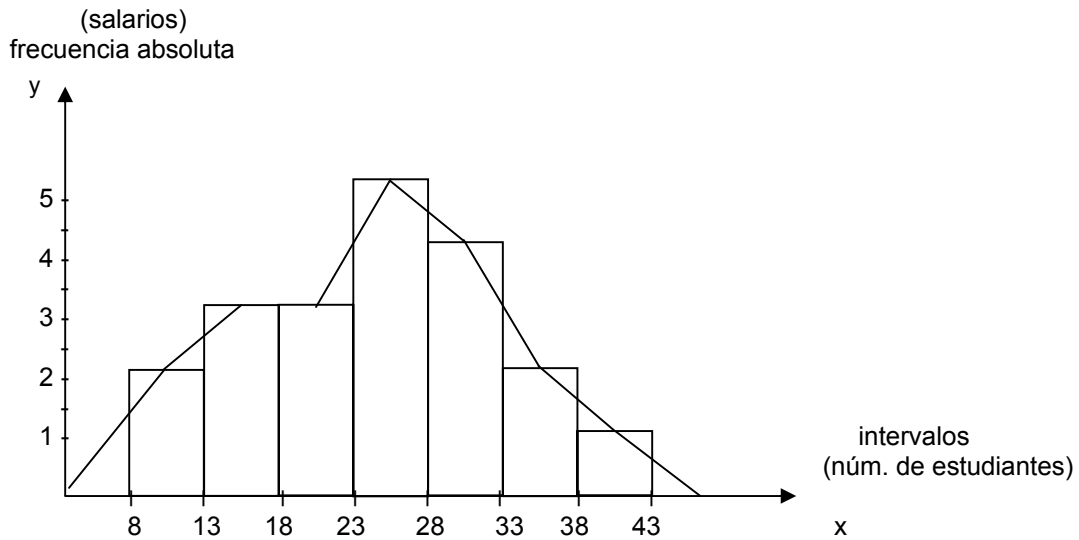
Consultas realizadas



APLICACIÓN DEL CONOCIMIENTO

Con base en la distribución de frecuencias del ejemplo de los 20 salarios de los estudiantes de bachillerato que trabajan, realizaremos la construcción del histograma correspondiente.

Recuerda que los valores de los intervalos se colocan en el eje horizontal o eje de las abscisas (eje x), y las frecuencias de ocurrencia se colocan en el eje vertical o eje de las ordenadas (eje y). Además, las alturas de cada barra representan las frecuencias de ocurrencia de cada intervalo de clase, entonces:



Distribución de frecuencias absoluta y relativa de los salarios de 20 estudiantes de bachilleres.					
Intervalos de clase	Fronteras inf. y sup.	Puntos medios	Conteo	Frecuencias absolutas	Frecuencias relativas
8-13	7.5-12.5	10	//	2	0.10
13-18	12.5-17.5	15	///	3	0.15
18-23	17.5-22.5	20	///	3	0.15
23-28	22.5-27.5	25	////	5	0.25
28-33	27.5-32.5	30	////	4	0.20
33-38	32.5-37.5	35	//	2	0.10
38-43	37.5-42.5	40	/	1	0.05
Totales				20	1.00

Construye un polígono de frecuencias relativas con los valores del ejercicio anterior.

Distribución de frecuencias absoluta y relativa de los salarios de 20 estudiantes de bachilleres.					
Intervalos de clase	Fronteras inf. y sup.	Puntos medios	Conteo	Frecuencias absolutas	Frecuencias relativas
8-13	7.5-12.5	10	//	2	0.10
13-18	12.5-17.5	15	///	3	0.15
18-23	17.5-22.5	20	///	3	0.15
23-28	22.5-27.5	25	////	5	0.25
28-33	27.5-32.5	30	////	4	0.20
33-38	32.5-37.5	35	//	2	0.10
38-43	37.5-42.5	40	/	1	0.05
Totales				20	1.00

EJERCICIOS

INSTRUCCIONES: De los siguientes enunciados, construye para cada uno de ellos:

- I) Un histograma de frecuencias absolutas.
- II) Un polígono de frecuencias relativas.

Recuerda que para cada ejercicio ya tienes elaborada la tabla de distribución de frecuencias.

1. El siguiente conjunto de 60 datos representa las ventas de galletas por día en una tienda de autoservicio.

Distribución de frecuencias absoluta y relativa de una muestra de 60 datos.					
Intervalos de clase	Fronteras inf. y sup.	Puntos medios	Conteo	Frecuencias absolutas	Frecuencias relativas
44-48	43.5-48.5	46	////////	9	0.1500
49-53	48.5-53.5	51	////////	9	0.1500
54-58	53.5-58.5	56	//////////	13	0.2166
59-63	58.5-63.5	61	////////	10	0.1666
64-68	63.5-68.5	66	////////	8	0.1333
69-73	68.5-73.5	71	//////	6	0.1000
74-78	73.5-78.5	76	////	4	0.0667
79-83	78.5-83.5	81	/	1	0.0167
Totales				60	0.9999

I)

II)

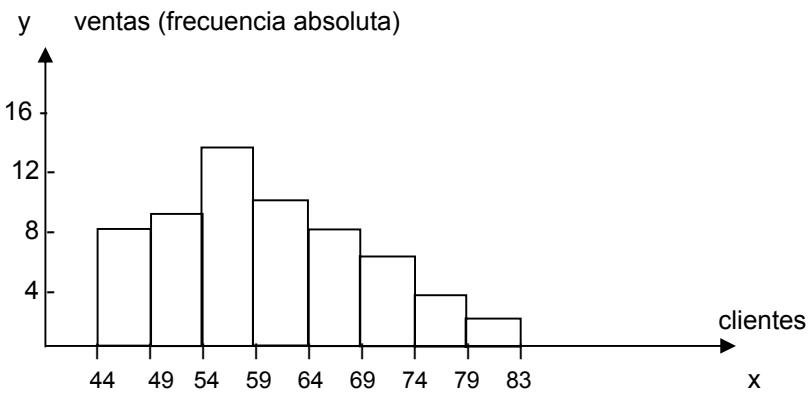
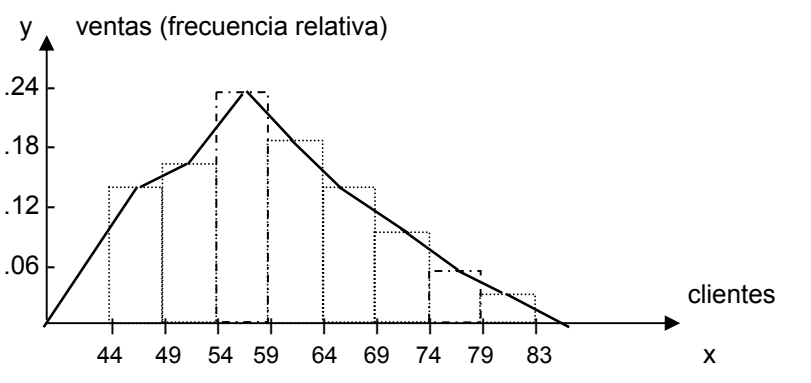
2. El siguiente conjunto de 150 datos representa el coeficiente intelectual de ciertos alumnos.

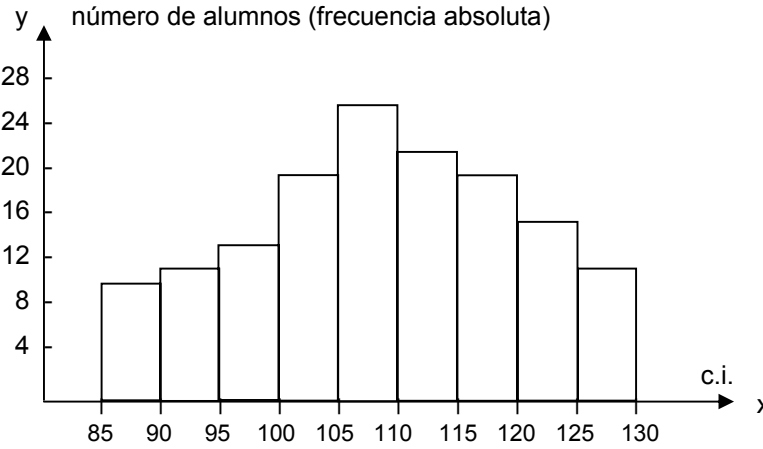
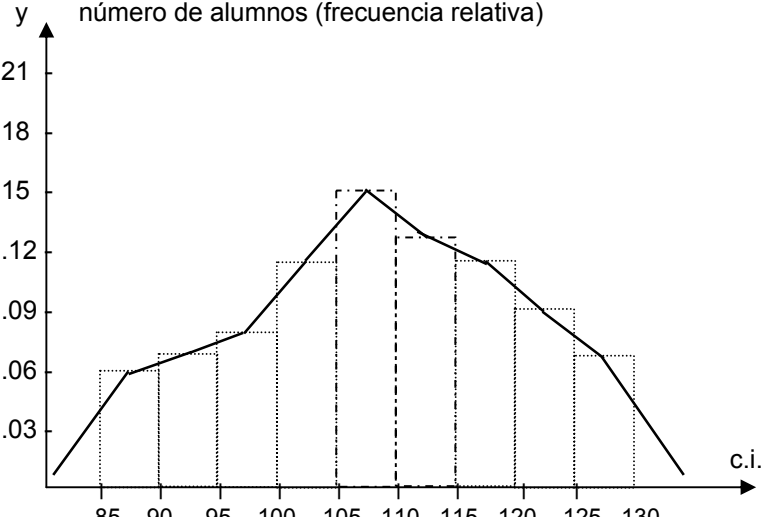
Distribución de frecuencias absoluta y relativa de una muestra de 150 datos.					
Intervalos de clase	Fronteras inf. y sup.	Puntos medios	Conteo	Frecuencias absolutas	Frecuencias relativas
85-89	84.5-89.5	87	////////	9	0.0600
90-94	89.5-94.5	92	//////////	11	0.0730
95-99	94.5-99.5	97	//////////	14	0.0930
100-104	99.5-104.5	102	//////////	20	0.1330
105-109	104.5-109.5	107	//////////	27	0.1800
110-114	109.5-114.5	112	//////////	22	0.1470
115-119	114.5-119.5	117	//////////	19	0.1270
120-124	119.5-124.5	122	//////////	16	0.1070
125-129	124.5-129.5	127	//////////	12	0.0800
Totales				150	1.0000

I)

II)

TABLA DE COMPROBACIÓN

Número de pregunta	Respuesta correcta
1	<p>I)</p>  <p>II)</p> 

Número de pregunta	Respuesta correcta
2	<p>I)</p>  <p>II)</p> 
Sugerencias	
<p>Recuerda que en un histograma las frecuencias están representadas por el área de los rectángulos (barras) y no por sus alturas.</p> <p>Para construir un polígono de frecuencias no es necesario calcular las áreas de los rectángulos (barras).</p>	

EJERCICIOS DE AUTOEVALUACIÓN

Cuentas con noventa minutos para resolverlos.

INSTRUCCIONES. Analiza los conjuntos de datos de los reactivos 1 a 3, y contesta para cada uno de ellos las siguientes preguntas.

I) Calcula el rango o recorrido de los datos.

II) ¿Cuál es la amplitud o longitud del intervalo?

III) Calcula el número de intervalos adecuados.

IV) Elabora la tabla de distribuciones de frecuencias.

1. El siguiente conjunto de datos representa las personas que ingresan sin depositar el boleto correspondiente, en un día, en 20 estaciones de la línea 2 del Metro.

3, 2, 5, 4, 3, 6, 0, 5, 4, 10, 1, 3, 0, 5, 11, 8, 9, 6, 3 y 4

I)

II)

III)

IV)

2. El siguiente conjunto de 30 datos representa las calificaciones obtenidas en un grupo al aplicar un examen de estadística descriptiva.

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

I)

II)

III)

IV)

3. Para comprobar la eficacia de los empleados encargados del llenado de paquetes con arroz, con peso nominal de 2 kilogramos, en una tienda de autoservicio se elige al azar una muestra de 30 paquetes, cuyos resultados en gramos fueron:

1930	1865	1977	2015	1943	1898
2120	1985	2053	2030	1972	2000
1907	1880	2005	1830	1966	1910
2075	2025	1954	1934	1988	1967
1946	1999	2061	2047	2093	1876

I)

II)

III)

IV)

INSTRUCCIONES: Analiza el conjunto de datos del reactivo 4, y contesta lo que se solicita:

4. El siguiente conjunto de datos muestra las 50 calificaciones obtenidas en un examen final de estadística.

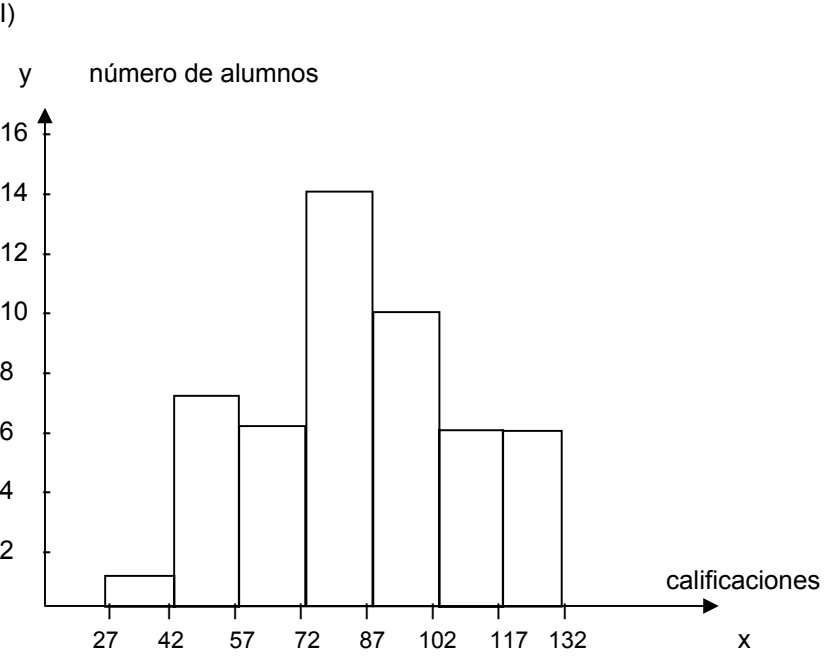
Distribución de frecuencias absoluta de una muestra de 50 calificaciones de un examen final de estadística.					
Intervalos de clase	Fronteras inf. sup.	Puntos medios	Conteo	Frecuencias absolutas	Frecuencias relativas
27-41	26.5-41.5	34	/	1	0.02
42-56	41.5-56.5	49	///////	7	0.14
57-71	56.5-71.5	64	//////	6	0.12
72-86	71.5-86.5	79	////////////////	14	0.28
87-101	86.5-101.5	94	////////////////	10	0.20
102-116	101.5-116.5	109	//////	6	0.12
117-131	116.5-131.5	124	//////	6	0.12
Totales				50	1.00

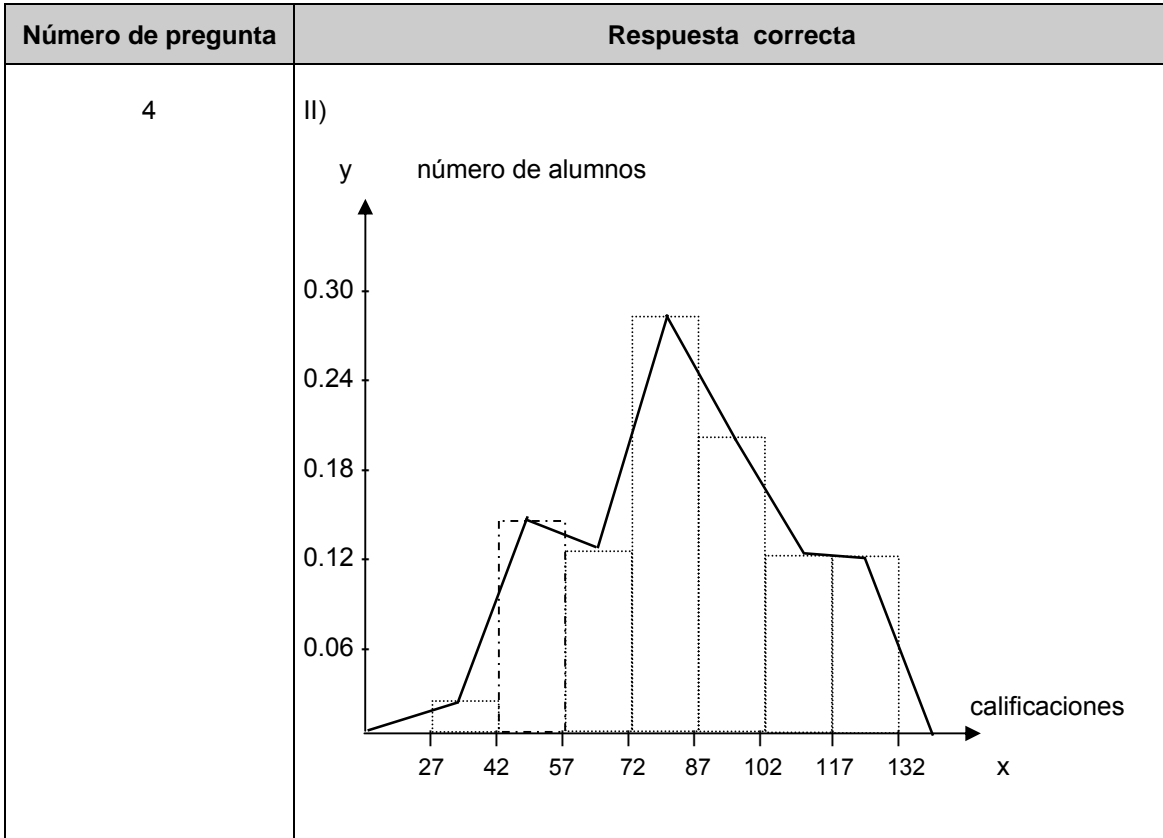
- I) Construye un histograma de frecuencias absolutas.

- II) Construye un polígono de frecuencias relativas.

CLAVE DE RESPUESTAS

Número de pregunta	Respuesta correcta																																													
1	<p>I) Rango = 11</p> <p>II) Amplitud = 3</p> <p>III) Número de intervalos = 5</p> <p>IV)</p> <table border="1" data-bbox="616 846 1414 1155"> <caption>Distribución de frecuencias absoluta y relativas de 20 estaciones del metro.</caption> <thead> <tr> <th>Intervalos de clase</th> <th>Fronteras inf. sup.</th> <th>Puntos medios</th> <th>Frecuencias absolutas</th> <th>Frecuencias relativas</th> </tr> </thead> <tbody> <tr> <td>0-2</td> <td>-0.5-2.5</td> <td>1</td> <td>4</td> <td>0.2000</td> </tr> <tr> <td>3-5</td> <td>2.5-5.5</td> <td>4</td> <td>10</td> <td>0.5000</td> </tr> <tr> <td>6-8</td> <td>5.5-8.5</td> <td>7</td> <td>3</td> <td>0.1500</td> </tr> <tr> <td>9-11</td> <td>8.5-11.5</td> <td>10</td> <td>3</td> <td>0.1500</td> </tr> <tr> <td>12-14</td> <td>11.5-14.5</td> <td>13</td> <td>0</td> <td>0.0000</td> </tr> <tr> <td colspan="3">Totales</td> <td>20</td> <td>1.0000</td> </tr> </tbody> </table>	Intervalos de clase	Fronteras inf. sup.	Puntos medios	Frecuencias absolutas	Frecuencias relativas	0-2	-0.5-2.5	1	4	0.2000	3-5	2.5-5.5	4	10	0.5000	6-8	5.5-8.5	7	3	0.1500	9-11	8.5-11.5	10	3	0.1500	12-14	11.5-14.5	13	0	0.0000	Totales			20	1.0000										
Intervalos de clase	Fronteras inf. sup.	Puntos medios	Frecuencias absolutas	Frecuencias relativas																																										
0-2	-0.5-2.5	1	4	0.2000																																										
3-5	2.5-5.5	4	10	0.5000																																										
6-8	5.5-8.5	7	3	0.1500																																										
9-11	8.5-11.5	10	3	0.1500																																										
12-14	11.5-14.5	13	0	0.0000																																										
Totales			20	1.0000																																										
2	<p>I) Rango = 6</p> <p>II) Amplitud = 0.9</p> <p>III) Número de intervalos = 7</p> <p>IV)</p> <table border="1" data-bbox="603 1498 1426 1868"> <caption>Distribución de frecuencias absolutas y relativas de 30 calificaciones de un examen de estadística.</caption> <thead> <tr> <th>Intervalos de clase</th> <th>Fronteras inf. sup.</th> <th>Puntos medios</th> <th>Frecuencias absolutas</th> <th>Frecuencias relativas</th> </tr> </thead> <tbody> <tr> <td>4-4.9</td> <td>3.95-4.85</td> <td>4.4</td> <td>3</td> <td>0.1000</td> </tr> <tr> <td>4.9-5.8</td> <td>4.85-5.75</td> <td>5.3</td> <td>4</td> <td>0.1333</td> </tr> <tr> <td>5.8-6.7</td> <td>5.75-6.65</td> <td>6.2</td> <td>7</td> <td>0.2333</td> </tr> <tr> <td>6.7-7.6</td> <td>6.65-7.55</td> <td>7.1</td> <td>6</td> <td>0.2000</td> </tr> <tr> <td>7.6-8.5</td> <td>7.55-8.45</td> <td>8.0</td> <td>5</td> <td>0.1666</td> </tr> <tr> <td>8.5-9.4</td> <td>8.45-9.35</td> <td>8.9</td> <td>4</td> <td>0.1333</td> </tr> <tr> <td>9.4-10</td> <td>9.35-10.25</td> <td>9.8</td> <td>1</td> <td>0.0333</td> </tr> <tr> <td colspan="3">Totales</td> <td>30</td> <td>0.9998</td> </tr> </tbody> </table>	Intervalos de clase	Fronteras inf. sup.	Puntos medios	Frecuencias absolutas	Frecuencias relativas	4-4.9	3.95-4.85	4.4	3	0.1000	4.9-5.8	4.85-5.75	5.3	4	0.1333	5.8-6.7	5.75-6.65	6.2	7	0.2333	6.7-7.6	6.65-7.55	7.1	6	0.2000	7.6-8.5	7.55-8.45	8.0	5	0.1666	8.5-9.4	8.45-9.35	8.9	4	0.1333	9.4-10	9.35-10.25	9.8	1	0.0333	Totales			30	0.9998
Intervalos de clase	Fronteras inf. sup.	Puntos medios	Frecuencias absolutas	Frecuencias relativas																																										
4-4.9	3.95-4.85	4.4	3	0.1000																																										
4.9-5.8	4.85-5.75	5.3	4	0.1333																																										
5.8-6.7	5.75-6.65	6.2	7	0.2333																																										
6.7-7.6	6.65-7.55	7.1	6	0.2000																																										
7.6-8.5	7.55-8.45	8.0	5	0.1666																																										
8.5-9.4	8.45-9.35	8.9	4	0.1333																																										
9.4-10	9.35-10.25	9.8	1	0.0333																																										
Totales			30	0.9998																																										

Número de pregunta	Respuesta correcta																																								
3	<p>I) Rango = 290</p> <p>II) Amplitud = 50</p> <p>III) Número de intervalos = 6</p> <p>IV)</p> <table border="1" data-bbox="475 730 1270 1055"> <caption>Distribución de frecuencias del llenado de paquetes con arroz.</caption> <thead> <tr> <th>Intervalos de clase</th> <th>Fronteras inf. sup.</th> <th>Puntos medios</th> <th>Frecuencias absolutas</th> <th>Frecuencias relativas</th> </tr> </thead> <tbody> <tr> <td>1830-1880</td> <td>1855</td> <td>1855</td> <td>3</td> <td>0.1000</td> </tr> <tr> <td>1880-1930</td> <td>1905</td> <td>1905</td> <td>4</td> <td>0.1333</td> </tr> <tr> <td>1930-1980</td> <td>1955</td> <td>1955</td> <td>9</td> <td>0.3000</td> </tr> <tr> <td>1980-2030</td> <td>2005</td> <td>2005</td> <td>7</td> <td>0.2333</td> </tr> <tr> <td>2030-2080</td> <td>2055</td> <td>2055</td> <td>5</td> <td>0.1666</td> </tr> <tr> <td>2080-2130</td> <td>2105</td> <td>2105</td> <td>2</td> <td>0.0666</td> </tr> <tr> <td colspan="3">Totales</td> <td>30</td> <td>0.9998</td> </tr> </tbody> </table>	Intervalos de clase	Fronteras inf. sup.	Puntos medios	Frecuencias absolutas	Frecuencias relativas	1830-1880	1855	1855	3	0.1000	1880-1930	1905	1905	4	0.1333	1930-1980	1955	1955	9	0.3000	1980-2030	2005	2005	7	0.2333	2030-2080	2055	2055	5	0.1666	2080-2130	2105	2105	2	0.0666	Totales			30	0.9998
Intervalos de clase	Fronteras inf. sup.	Puntos medios	Frecuencias absolutas	Frecuencias relativas																																					
1830-1880	1855	1855	3	0.1000																																					
1880-1930	1905	1905	4	0.1333																																					
1930-1980	1955	1955	9	0.3000																																					
1980-2030	2005	2005	7	0.2333																																					
2030-2080	2055	2055	5	0.1666																																					
2080-2130	2105	2105	2	0.0666																																					
Totales			30	0.9998																																					
4	<p>I)</p> 																																								



UNIDAD 3

MEDIDAS DESCRIPTIVAS

3.1 MEDIDAS DE TENDENCIA CENTRAL

APRENDIZAJES

- Calcular las medidas centrales; la moda, la mediana y la media aritmética para datos no agrupados.
- Calcular las medidas centrales; la moda, la mediana y la media aritmética para datos agrupados.

Las medidas de tendencia central son valores numéricos que **tienden a localizar, en algún sentido, la parte central de un conjunto de datos**. A menudo el término promedio se asocia a estas mediciones. Cada una de las diferentes medidas de tendencia central pueden recibir el nombre de valor promedio.

Con base en lo aprendido anteriormente, habrás observado una clara tendencia de agrupación en el centro de los valores más frecuentemente aparecidos. En general, la mayor densidad de frecuencia está en la parte central de las gráficas, de aquí deriva el nombre de medidas de tendencia central que se les da a la moda, mediana y media. Recuerda que las medidas de tendencia central se utilizan para describir y establecer comparaciones cuantitativas entre distribuciones.

MODA PARA DATOS NO AGRUPADOS

La moda o promedio típico de un conjunto de datos **es el valor que ocurre más frecuentemente en el conjunto**. Si un valor es seleccionado al azar del conjunto dado, el valor modal es aquél que es más probable de ser seleccionado. Así, la moda es generalmente considerada como el valor más típico de una serie de datos.

La moda para datos no agrupados es el valor que aparece con mayor frecuencia en un grupo de datos. Si existe una sola moda, se le llamará unimodal; si existen dos modas, se le llamará bimodal, y si son más de dos, se llamará multimodal.

MEDIANA PARA DATOS NO AGRUPADOS

La mediana es aquel valor que se encuentra a la mitad de una muestra o población cuyos valores están ordenados en magnitud. Si el número de valores es impar, la mediana para datos no agrupados es igual al valor de la mitad. Si el número de valores es par, la mediana es igual al promedio de los valores que quedan en la mitad.

MEDIA ARITMÉTICA PARA DATOS NO AGRUPADOS

La medida de tendencia central que se encuentra con más frecuencia es la media aritmética o simplemente media. **La media para datos no agrupados es el cociente de la suma de los valores divididos por el número de valores en el conjunto de datos**.

$$\text{Media} = \frac{\text{Suma de los valores}}{\text{Números de los valores}},$$

por lo tanto:

$$\bar{X} = \frac{\sum x_i}{n} \quad \text{ó} \quad \bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

donde \sum se utiliza para simbolizar la suma indicada o sumatoria de un conjunto de sumandos. $\sum x$ significa la sumatoria de los valores de la variable x.

MODA PARA DATOS AGRUPADOS

En una distribución de frecuencias de datos agrupados, **la moda es la marca del intervalo de clase que contiene la mayor frecuencia.**

La moda es la medida de tendencia central que no tiene propiedades que le permitan intervenir en desarrollos algebraicos como la media aritmética; por esto es de menor importancia teórica que ella; sin embargo, posee propiedades que ponen en evidencia ciertas cualidades de un colectivo, cosa que no ocurre con la media aritmética que promedia todos los valores igualando en un justo reparto todas las observaciones, es decir, suprimiendo sus individualidades. En cambio, la moda destaca los valores individuales, de lo que se desprende su utilidad e importancia en cierto tipo de análisis.

La moda para datos agrupados se puede calcular mediante la fórmula:

$$Mo = Lo + \left(\frac{d1}{d1 + d2} \right) w$$

donde:

- Lo** es el límite de la frontera inferior de la clase modal (recuerda que la clase modal es la clase con la más alta frecuencia en la distribución).
- d1** es la frecuencia de la clase modal menos la frecuencia de la clase que se encuentra inmediatamente debajo (antes) de ella.
- d2** es la frecuencia de la clase modal menos la frecuencia de la clase que se encuentra inmediatamente encima (después) de ella.
- w** es la amplitud de la clase modal (intervalo).

MEDIANA PARA DATOS AGRUPADOS

La mediana para datos agrupados será el valor que se encuentra a la mitad de una serie ordenada de valores. La mediana para una distribución de frecuencias es aquél punto del eje horizontal del histograma correspondiente en el cual, si se traza una línea vertical, el área comprendida bajo el histograma queda dividida en dos partes iguales.

La mediana para datos agrupados se puede calcular mediante la fórmula:

$$Med = Lo + \left(\frac{\frac{n}{2} - fa}{f} \right) i$$

donde:

- Lo** es el límite o frontera real inferior de la clase de la mediana.
- n** es el número total de datos.
- fa** es la frecuencia absoluta acumulada de la clase inmediata inferior a la clase de la mediana.
- f** es la frecuencia de la clase mediana.
- i** es la longitud o amplitud del intervalo.

MEDIA ARITMÉTICA PARA DATOS AGRUPADOS

La media aritmética ponderada (en aritmética, el concepto de media aritmética ponderada se aplica para calcular el valor promedio de cantidades a cada una de las cuales está asociado un número o peso que la pondera) de un conjunto de cantidades X_1, X_2, \dots, X_n ponderadas por los pesos m_1, m_2, \dots, m_n , **queda expresada por el cociente entre la suma de los productos de las cantidades por sus respectivas ponderaciones (pesos) y la suma de las ponderaciones.**

En una distribución de frecuencias agrupadas, todos los valores que caen dentro de un intervalo de clase se consideran de un mismo valor igual a la marca de clase; entonces las frecuencias son las ponderaciones de los valores que corresponden con las marcas de clase. De donde:

- ❖ **En una distribución de frecuencias agrupadas, las ponderaciones son las frecuencias y las marcas de clase son los valores que se ponderan.**

Recuerda que la media aritmética es la única de las medidas de tendencia central que puede intervenir en operaciones algebraicas. De las varias propiedades matemáticas que posee la media, únicamente mencionaremos dos de las más importantes.

- ❖ **En toda distribución, la suma de las desviaciones de sus variables con respecto a la media es cero** ($\sum x = X - \bar{X} = 0$).
- ❖ **La suma de los cuadrados de las desviaciones respecto a la media es siempre menor que la suma de los cuadrados de las desviaciones con respecto a cualquier otro valor.**

La propiedad anterior indica que la media aritmética es la medida de tendencia central que hace mínima la suma de los cuadrados de las desviaciones en torno de ella. Esta importante propiedad es el origen del llamado método de mínimos cuadrados para la búsqueda de la media, y es importante en estadística por su aplicación en el ajuste de curvas.

Entonces, la expresión para calcular la media para datos agrupados a partir de las ideas anteriores, es:

$$\bar{X} = \frac{\sum Xf}{\sum f}$$

donde:

- X** es la marca de clase o punto medio.
- f** es la frecuencia de cada clase.

Recuerda que las medidas de tendencia central son **promedios**, por lo tanto, mediante un promedio se puede tomar un puñado de cifras sobre cualquier tema (temperaturas, ingresos, velocidades, poblaciones, etc.) y calcular una cifra que representará al conjunto por completo.

Sin embargo, debes recordar que existen varias clases de mediciones conocidas comúnmente como promedios, y cada uno presenta un panorama diferente sobre las cifras que representa.

Por ejemplo, si consideramos los ingresos mensuales de diez trabajadores:

\$4500.00	\$3900.00	\$3750.00	\$3675.00	\$3225.00
\$3150.00	\$3150.00	\$3150.00	\$3150.00	\$2250.00

¿Cuál es el ingreso típico de estos trabajadores? El promediar sus ingresos podría dar la respuesta a la pregunta; entonces, el promedio de uso más frecuente es **la media**, por que no se especificó de qué tipo; éste promedio se obtiene sumando los elementos de una serie de números y dividiendo entre el número de elementos, para nuestro ejemplo, la media es \$3390.00.

La media es representativa de la serie de números en el sentido de que la suma de las cantidades según las cuales las cifras mayores exceden a la media es igual a la suma de las cantidades, según las cuales las cifras menores están por debajo de la media.

Para el caso de **la mediana**, como se observó, seis trabajadores ganan menos que la media, mientras que cuatro ganan más. Puede representarse este grupo con el ingreso de la familia que está situado exactamente en la parte central de todos los ingresos.

Para hacer esto se necesita determinar la mediana; sería fácil si hubiera once familias en el grupo. Al ordenar los ingresos de menor a mayor, la familia con el sexto ingreso de la ordenación tendría la mediana de los ingresos; sin embargo, con diez familias no se tiene una familia que esté en el centro de los ingresos; por tal razón, se suman los dos ingresos centrales (\$3150.00 y \$3225.00) se divide entre dos y da como resultado, que la mediana es \$3187.50, menor que la media.

Hasta ahora se han obtenido dos clases de promedios y ninguna familia tiene un ingreso igual a alguno de los valores hallados.

Ahora se quiere representar al grupo estableciendo el ingreso que ocurre con mayor frecuencia; esta clase de representatividad recibe el nombre de **moda**. En este ejemplo, el ingreso modal es \$3150.00, el cual es obtenido por un número mayor de familias, más que ningún otro.

Se han obtenido tres promedios diferentes, cada uno es válido, correcto e informativo a su manera; ¡pero qué distintos son!, éstos son:

Media	\$3390.00
Mediana	\$3187.50
Moda	\$3150.00

Estos valores de los ingresos podrían diferir más, si sólo una familia en el grupo fuera millonaria o muy pobre. Con base en lo anterior, podemos decir:

Primero, cuando se vea o se escuche algo sobre un promedio, debe averiguarse de qué clase se trata; entonces se sabrá qué clase de aspecto se considera.

Segundo, obsérvense las cifras que se promedian de manera que pueda juzgarse si el promedio es el adecuado.

Tercero, cada vez que alguien diga **promedio** no debe suponerse que se implica una cuantificación matemática; no es así, a menudo se habla de la persona promedio sin que ello se refiera a una media, mediana o moda; todo lo que se quiere significar es la idea de gente que en muchos aspectos se asemeja bastante al resto.

Recuerda que la **moda es aquel valor que se presenta más frecuentemente**, que la **mediana es la medida de tendencia central que divide el área de un histograma o el área bajo la curva de distribución en dos partes iguales**, de manera que el número de observaciones por debajo de la mediana es igual al número de observaciones por encima de ella y que la **media es la medida de tendencia central en la que la suma de las desviaciones respecto a ella es cero**.

Estas medidas que estás estudiando nos apoyan para visualizar en las gráficas donde se han trazado las ordenadas correspondientes, es decir:

- ❖ *La media aritmética es un punto de equilibrio (semejante a un centro de gravedad).*
- ❖ *La mediana tiene la propiedad de que su ordenada divide el área bajo la curva en dos partes iguales.*
- ❖ *La moda es la abscisa correspondiente a la mayor ordenada o pico de la curva.*

APLICACIÓN DEL CONOCIMIENTO

A continuación, calcularemos las medidas de tendencia central para **datos no agrupados**.

Once alumnos de primer año obtuvieron las siguientes calificaciones en una prueba de destreza: 70, 83, 74, 75, 81, 75, 92, 75, 90, 94 y 75. ¿Cuál es el valor de la moda?

El valor que se repite más frecuentemente es el 75, entonces 75 es la moda.

Los kilómetros recorridos por cinco estudiantes al venir a la escuela desde sus casas son: 1, 4, 10, 10 y 8. ¿Cuál es el valor de la mediana?

Núm. de estudiantes	Kilómetros Recorridos
1	1
2	4
3	8
4	10
5	10

Después de que se ordenan los valores, como se muestra en la tabla, se puede ver fácilmente que el valor de la mediana es el 8.

Para esta misma información, ahora calcularemos el valor de la media.

$$\bar{X} = \frac{1+4+10+8+10}{5} = 6.6 \text{ km} \quad \text{El valor de la media es } 6.6 \text{ km.}$$

Ahora, realiza los cálculos necesarios para encontrar la solución del siguiente planteamiento.

Las edades de los profesores de Taller de Lectura y Redacción son: 28, 30, 31, 30, 45, 25, 60, 48, 30, 24, 30, 39 y 30. Calcula el valor de la moda.

Ahora calcula el valor de la media.

En los ejercicios que se muestran a continuación calcularemos las medidas de tendencia central para **datos agrupados**.

La siguiente distribución de frecuencias corresponde a 150 puntajes del coeficiente de inteligencia (CI) de ciertos alumnos.

Distribución de frecuencias de 150 puntajes del CI			
Intervalos de clase	Fronteras inf. sup	Marca de clase	Frecuencia absoluta
85-89	84.5-89.5	87	9
90-94	89.5-94.5	92	11
95-99	94.5-99.5	97	14
100-104	99.5-104.5	102	20
*105-109	104.5-109.5	107	27
110-114	109.-114.5	112	22
115-119	114.5-119.5	117	19
120-124	119.5-124.5	122	16
125-129	124.5-129.5	127	12

Para calcular la moda se elige de la distribución de frecuencias, el intervalo donde la frecuencia de ocurrencia sea mayor; es decir, la clase modal:

$$Mo = Lo + \left(\frac{d1}{d1 + d2} \right) w$$

Observa de la tabla, que en el intervalo marcado (*) se encuentra la clase modal, entonces:

$$Mo = 104.5 + \left(\frac{7}{7 + 5} \right) 5 = 107.41$$

por lo tanto la moda es 107.41.

Para los datos del ejercicio anterior, calcularemos la mediana.

Distribución de frecuencias de 150 puntajes del CI			
Intervalo de clase	Fronteras inf. sup.	Frecuencia absoluta	Frecuencia abs. acum.
85-89	84.5-89.5	9	9
90-94	89.5-94.5	11	20
95-99	94.5-99.5	14	34
100-104	99.5-104.5	20	54
*105-109	104.5-109.5	27	81
110-114	109.5-114.5	22	103
115-119	114.5-119.5	19	122
120-124	119.5-124.5	16	138
125-129	124.5-129.5	12	150

Primero determinamos la clase que contiene a la mediana Esta clase se llama clase mediana y es la que contiene el valor que ocupa $n/2$, donde n es el número total de datos; después calculamos la frecuencia

acumulada absoluta correspondiente a la clase inmediata inferior a la clase de la mediana; determinamos la frecuencia de la clase de la mediana, el ancho del intervalo y la frontera real inferior de la clase de la mediana y finalmente aplicamos la fórmula:

$$Med = Lo + \left(\frac{\frac{n}{2} - fa}{f} \right) i$$

Observamos que $n/2 = 150/2 = 75$ y, por otra parte, la clase que contiene a la mediana es: 105 - 109, con una frecuencia de 27 y una frec. abs. acumulada de 81; la clase inmediata inferior de la mediana es de 54 y el ancho de clase o del intervalo es de 5 y el límite real inferior de la clase de la mediana es 104.5, entonces:

$$Med = 104.5 + \left(\frac{75 - 54}{27} \right) 5$$

por lo tanto la mediana es 108.388

Ahora, calcularemos la media.

Para aplicar la fórmula de la media aritmética para datos agrupados, es necesario realizar unos cálculos de productos y sumas, es decir: debemos multiplicar la marca de clase o punto medio ya calculado de cada intervalo por la frecuencia de ocurrencia de dicho intervalo, como se muestra a continuación:

Distribución de frecuencias de 150 puntajes del CI			
Intervalo de clase	Punto medio	Frecuencia absoluta	Xf
85-89	87	9	783
90-94	92	11	1012
95-99	97	14	1358
100-104	102	20	2040
105-109*	107	27	2889
110-114	112	22	2464
115-119	117	19	2223
120-124	122	16	1952
125-129	127	12	1524
Totales		150	$\sum Xf = 16245$

Ahora, si sustituimos los datos en la expresión de la media para datos agrupados, tenemos que:

$$\bar{X} = \frac{\sum Xf}{\sum f} = \frac{16245}{150} = 108.3 \quad \text{el valor de la media es de 108.3}$$

Siguiendo las ideas del ejemplo anterior realiza los cálculos necesarios para encontrar la solución del siguiente problema.

En esta distribución de frecuencias se presentan el número de retardos semestrales de los trabajadores en cierta empresa.

Intervalo de clase (retardos)	Número de empleados (frecuencia)
1-4	20
5-8	50
9-12	100
13-16	60
17-20	30

Encuentra el valor de la moda.

Encuentra el valor de la mediana.

Encuentra el valor de la media.

EJERCICIOS

INSTRUCCIONES: Lee con atención los siguientes enunciados y contesta lo que se solicita. Recuerda que estos ejercicios son de datos no agrupados.

1. El siguiente conjunto de datos representa el gasto diario en pesos (\$) de 15 señoras para la preparación de la comida: 60, 65, 80, 85, 70, 75, 75, 70, 69, 70, 68, 70,74, 70 y 70.

I. ¿Cuál es el valor que se presenta más frecuentemente (moda) en esta serie de datos?

II. ¿Cuál es el valor de la mediana?

III. ¿Cuál es el valor de la media?

2. Los siguientes valores representan los kilómetros recorridos por 5 estudiantes al venir a la escuela desde sus casas: 1, 4, 10, 8 y 10. ¿Cuál es el valor de la moda?

3. El número de nacimientos en el Estado de México en veinte días en diferentes hospitales son: 3, 2, 5, 4, 3, 6, 0, 5, 4, 10, 1, 3, 0, 5, 11, 8, 9, 6, 3 y 4.

Obtén:

I. La moda.

II. La mediana.

III. La media.

4. Las edades de los profesores de Taller de Lectura y Redacción son: 28, 30, 31, 30, 45, 25, 60, 48, 30, 24, 30, 39 y 30.

I. ¿Cuál es el valor de la mediana?

II. ¿Cuál es el valor de la media?

INSTRUCCIONES: Lee con atención los siguientes enunciados y contesta lo que se solicita. Recuerda que en estos ejercicios debes aplicar las fórmulas para datos agrupados.

5. La siguiente distribución de frecuencias representa los clientes de una tienda de abarrotes.

Intervalo de clase	Frecuencias
1-4	5
6-9	25
11-14	70
16-19	50
21-24	40
26-29	30
31-34	20
36-39	10

Calcula:

I. El valor de la moda.

II. El valor de la mediana.

III. El valor de la media.

6. La siguiente distribución de frecuencias representa las ventas en cientos de pesos en una tienda departamental.

Intervalo de clase (ventas)	Número de empresas (frecuencia)
1-5	2
6-10	8
11-15	30
16-20	26
21-25	4

Obtén:

I. La moda.

II. La mediana.

III. La media.

7. La siguiente distribución de frecuencias representa las edades de los trabajadores de una empresa.

Intervalo de clase (edades)	Número de empleados (frecuencia)
18-22	20
23-27	50
28-32	100
33-37	60
38-42	30

I. ¿Cuál es el valor de la moda?

II. ¿Cuál es el valor de la mediana?

III. ¿Cuál es el valor de la media?

TABLA DE COMPROBACIÓN

Número de pregunta	Respuesta correcta
1	I. Mo. = 70
	II. Med. = 70
	III. Media = 71.4
2	I. Mo. = 10
3	I. Mo. = 3
	II. Med. = 4
	III. Media = 4.6
4	I. Med. = 30
	II. Media = 34.615
5	I. 13.46
	II. 16.78
	III. 18.80
6	I. 14.73
	II. 14.67
	III. 14.57
7	I. 30.28
	II. 30.50
	III. 30.57
Sugerencias	
<p>Recuerda aplicar la fórmula adecuada para calcular la media.</p> <p>Para datos no agrupados:</p> $\bar{X} = \frac{\sum x_i}{n} \quad \text{ó} \quad \bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$ <p>para datos agrupados:</p> $\bar{X} = \frac{\sum Xf}{\sum f}$	

3.2 MEDIDAS DE DISPERSIÓN

APRENDIZAJES

- Calcular la desviación media de datos agrupados.
- Calcular la varianza de un conjunto de datos agrupados.
- Calcular la desviación estándar de un conjunto de datos agrupados.

Las medidas de dispersión abarcan el rango o recorrido (se estudió en las distribuciones de frecuencias), la desviación media, la desviación estándar o típica y la varianza. Estos valores numéricos describen el grado de dispersión o variabilidad de los datos.

Los valores de estas medidas de variabilidad o dispersión, serán mayores entre más dispersos se encuentran los datos, y serán menores cuando los datos se encuentren menos dispersos.

DESVIACIÓN MEDIA

La desviación media (DM) es la media aritmética de las desviaciones de los valores individuales con respecto al promedio de los datos. Al calcular la desviación media para datos agrupados, los valores absolutos de las desviaciones son utilizados, es decir, los signos (+) y (-) de las desviaciones se ignoran. La desviación media para datos agrupados se calcula con la siguiente expresión:

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

Recuerda que la desviación media (DM) es una medida de la dispersión bastante objetiva: cuanto mayor sea su valor, 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 la distribución. Por otra parte, al tomar los valores absolutos mide la desviación de una observación sin mostrar si está por encima o por debajo a la media aritmética.

VARIANZA

Recordemos, como se vio anteriormente, que para calcular la desviación media, fue necesario prescindir de los signos negativos tomando los valores absolutos de las desviaciones con respecto a la media aritmética. Si elevamos al cuadrado estas desviaciones, logramos con esta operación que todas las desviaciones den resultados positivos, sumando los cuadrados de las desviaciones y dividiendo por $n - 1$ (para muestras), se obtiene un estadístico llamado varianza.

La varianza para datos agrupados, denotada por S^2 , es la media aritmética de los cuadrados de las desviaciones respecto a la media y sirve de base para calcular la desviación estándar que es la más importante de todas las medidas de dispersión. Entonces:

$$S^2 = \frac{\sum f(X - \bar{X})^2}{n-1} \quad \text{donde } n \text{ es la suma de las frecuencias } (n = \sum f).$$

DESVIACIÓN ESTÁNDAR O TÍPICA

Al estudiar la varianza en el contenido anterior como medida del grado de variabilidad de las distribuciones, ésta tiene un inconveniente, ya que expresa la dispersión de una variable en unidades distintas a la que mide la variable; así, si estudiamos la dispersión de las estaturas de las personas respecto a la media de su grupo, el valor mide la desviación en centímetros, mientras que la varianza, mide la dispersión en centímetros cuadrados. Entonces, al extraer la raíz cuadrada regresamos a la unidad de medida de las variables, de lo cual nos resulta la desviación estándar o típica, denotada por "S".

La desviación estándar que estudiaremos es la más importante de todas las medidas de dispersión; en una distribución normal incluye aproximadamente el 68.27% de los términos, y por sus propiedades algebraicas se utiliza en los análisis estadísticos en los que intervienen relaciones matemáticas.

La desviación estándar es una forma refinada de la desviación media; la desviación estándar para datos agrupados es la raíz cuadrada de la varianza, por lo que:

$$S = \sqrt{\frac{\sum f(X - \bar{X})^2}{\sum f - 1}} = \sqrt{S^2}$$

APLICACION DEL CONOCIMIENTO

De la siguiente distribución de frecuencias, que muestran las ganancias de las empresas alimenticias en millones de pesos. ¿Cuál es el valor de la desviación media?

Ganancias (intervalo)	Número de empresas (frecuencia)
140-159	7
160-179	30
180-199	13
200-219	25
220-239	11
240-259	4

Observa cuidadosamente el procedimiento para resolver este ejemplo.

Es necesario mostrar con otras columnas las operaciones que nos permitan aplicar la fórmula de la desviación media, por lo tanto:

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{17355}{90} = 192.833$$

Ahora elaboremos la tabla que necesitamos para aplicar la fórmula de la desviación media, es decir, calculamos la marca de clase, el valor absoluto de las desviaciones y, finalmente, el producto de la frecuencia por las desviaciones; estos productos se suman como lo indica la fórmula, y tenemos que:

Marca de Clase (X)	Desviaciones $ x = X - \bar{X} $	$f x $
149.5	43.333	303.331
189.5	23.333	699.990
189.5	3.333	43.329
209.5	16.667	416.675
229.5	36.667	403.337
249.5	56.667	226.668
		$\sum f x = 2093.330$

Entonces, si sustituimos los datos en la expresión de la desviación media, tenemos que:

$$DM = \frac{\sum f|x|}{\sum f} = \frac{2093.330}{90} = 23.259$$

¿Cuál es la varianza de la siguiente distribución de frecuencias que representa las ventas de una tienda de autoservicio de refrescos durante un día?.

Intervalo	Frecuencia
27-29	1
30-32	10
33-35	14
36-38	33
39-41	14
42-44	7
45-47	3

Para calcular la varianza para esta distribución de frecuencias es necesario agregar algunas columnas para realizar el cálculo. No debes olvidar que para el cálculo de la varianza es necesario hacer el producto de las frecuencias por las correspondientes marcas de clase o punto medio (fX), calcular las desviaciones con respecto a la media (x), elevarlas al cuadrado (x²) y después, multiplicarlas por las frecuencias correspondientes (fx²), como se indica en la siguiente tabla:

Intervalo	Frecuencia	Marca de clase(X)	Xf	$ x = X - \bar{X} $	f $ X - \bar{X} $	$(x - \bar{x})^2$	f $(X - \bar{X})^2$
20-29	1	28	28	9	9	81	81
30-32	10	31	310	6	60	36	360
33-35	14	34	476	3	42	9	126
36-38	33	37	1221	0	0	0	0
39-41	14	40	560	3	42	9	126
42-44	7	43	301	6	42	36	252
45-47	3	46	138	9	27	81	243
			$\Sigma Xf=3034$				$\Sigma f(X - \bar{X})^2 = 1188$

De acuerdo con lo anterior la media es: $\bar{X} = \frac{\sum fX}{\sum f} = \frac{3034}{82} = 37$

Ahora, con base en la media y en la tabla anterior, procedemos a calcular la varianza:

$$S^2 = \frac{\sum f(X - \bar{X})^2}{\sum f - 1} = \frac{1188}{81} = 14.666$$

¿Cuál es la desviación estándar de la distribución de frecuencias para este mismo ejercicio? Para poder calcular la desviación estándar para esta distribución de frecuencias, es necesario agregar algunas columnas para realizar el cálculo, de la siguiente manera:

Si tenemos que la media es:
$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{3034}{82} = 37$$

la varianza es:
$$S^2 = \frac{\sum f(X - \bar{X})^2}{n-1} = \frac{1188}{81} = 14.666$$

por lo tanto:
$$S = \sqrt{14.666} = 3.8296$$

De acuerdo con el ejemplo anterior, resuelve el siguiente ejercicio. Los kilómetros recorridos por 20 estudiantes al venir a la escuela desde sus casas, se muestra a continuación:

Km recorridos (Intervalo)	Número de estudiantes (frecuencia)
0 y menos de 2	2
2 y menos de 4	5
4 y menos de 6	4
6 y menos de 8	8
8 y menos de 10	1

Calcula:

I. La desviación media.

II. El valor de la varianza.

III. La desviación estándar.

EJERCICIOS.

INSTRUCCIONES: Lee con atención los siguientes enunciados y contesta lo que se solicita.

1. La siguiente distribución de frecuencias muestra la estatura en cm de 108 alumnos.

Intervalo (estatura)	Frecuencia (alumnos)
123.5-128.5	1
128.5-133.5	4
133.5-138.5	9
138.5-143.5	24
143.5-148.5	29
148.5-153.5	22
153.5-158.5	14
158.5-163.5	5

Calcula:

I. La desviación media.

II. La varianza.

III. La desviación estándar.

2. Calcula la desviación media de la siguiente distribución de frecuencias que muestra los defectos de 180 productos de una línea de papelería.

Intervalo	Frecuencia
42-46	2
47-51	9
52-56	31
57-61	50
62-66	51
67-71	30
72-76	7

3. La siguiente distribución de frecuencias muestra la estatura de 108 alumnos de una escuela.

Intervalo (estatura)	Frecuencia (alumnos)
124.5-127.5	1
127.5-130.5	0
130.5-133.5	4
133.5-136.5	1
136.5-139.5	14
139.5-142.5	10
142.5-145.5	23
145.5-148.5	14
148.5-151.5	12
151.5-154.5	15
154.5-157.5	5
157.5-160.5	8
160.5-163.5	1

Calcula:

I. La desviación media.

II. La varianza.

III. La desviación estándar.

4. La siguiente distribución de frecuencias muestra las ventas en miles de pesos de 79 empresas.

Ventas	Número de empresas
10.5-15.5	14
15.5-20.5	23
20.5-25.5	30
25.5-30.5	8
30.5-35.5	4

Obtén:

I. El valor de la varianza.

II. La desviación estándar.

TABLA DE COMPROBACIÓN

Número de pregunta	Respuesta correcta
1	I. D.M.= 5.77 cm. II. $S^2 = 54.80$ III. $S = 7.403$
2	I. D.M. = 19.6572
3	I. D.M. = 5.84 II. $S^2 = 52.11$ III. $S = 7.2187$
4	I. $S^2 = 28.04$ II. $S = 5.2952$
Sugerencias	
<p>Para realizar los cálculos de las medidas de dispersión, es necesario calcular la media aritmética.</p> <p>Para el ejercicio 1 la media es 146.32 cm.</p> <p>Para el ejercicio 2 la media es 61.14.</p> <p>Para el ejercicio 3 la media es 146.39 cm.</p> <p>Para el ejercicio 4 la media es 20.784.</p>	

3.3 MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSIÓN EN CURVAS DE FRECUENCIA

APRENDIZAJES

- Comparar curvas simétricas y asimétricas a partir de la relación de medidas de dispersión.

MEDIDAS DE ASIMETRÍA

Recordemos que las medidas de tendencia central en una distribución de frecuencias simétricas se encuentran todas en el centro; es decir, las tres medidas (media, mediana y moda) coinciden. Cuando estos valores se apartan uno del otro, entonces la distribución se convierte en asimétrica o con sesgo.

El sesgo es el grado de asimetría o falta de simetría de una distribución. Si la curva de frecuencia de una distribución tiene cola más larga a la derecha del máximo central, se dice de la distribución está sesgada a la derecha o tiene sesgo positivo. Si es al contrario, se dice que está sesgada a la izquierda o que tiene sesgo negativo.

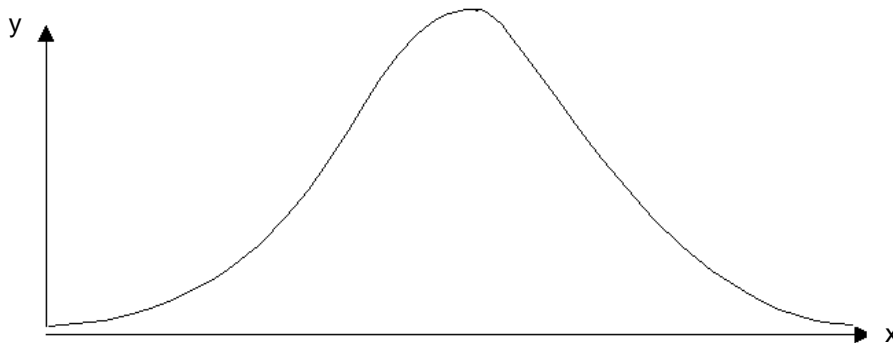
En distribuciones sesgadas, la media tiende a situarse con respecto a la moda al mismo lado que la cola más larga; así, una medida de asimetría nos es dada por la diferencia de la media con la moda, dividiéndola por la desviación estándar, cuya expresión es:

$$S_{k1} = \frac{\bar{X} - Mo}{S} \quad \text{Primer coeficiente de Pearson}$$

$$S_{k2} = \frac{3(\bar{X} - Med.)}{S} \quad \text{Segundo coeficiente de Pearson}$$

Los cuales se utilizan para medir el grado de asimetría.

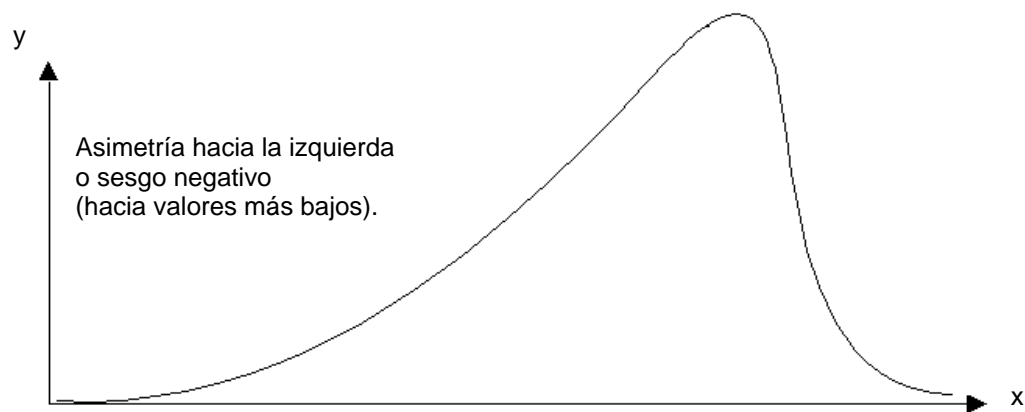
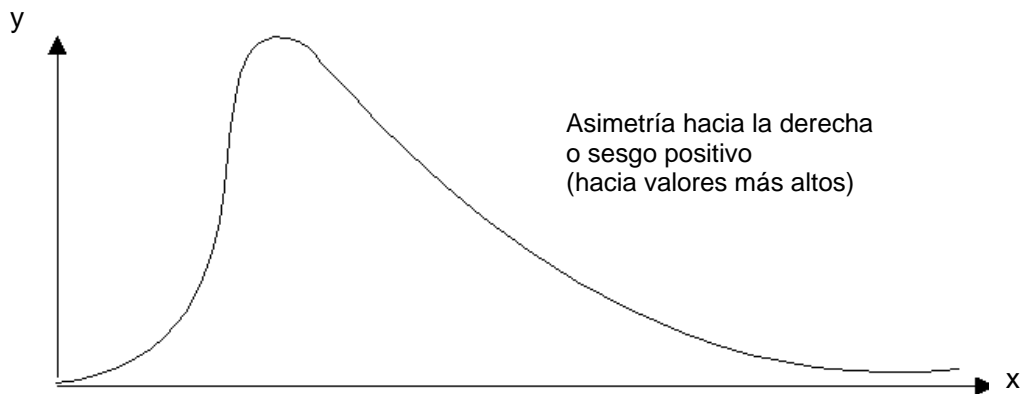
No debemos olvidar que una curva simétrica es como la que se muestra a continuación:



Los valores de las medidas de tendencia central se localizan en el centro de la curva, siendo esta curva simétrica (curva normal).

Si las frecuencias arriba de la clase modal no son las mismas que las frecuencias debajo de la clase modal, la distribución no es simétrica y los valores de los tres promedios de la distribución no son los mismos.

Cuando una distribución no es simétrica, la curva se torna asimétrica, ya sea hacia el lado derecho sobre el eje X o hacia el lado izquierdo sobre el eje X, como se muestra a continuación:



APLICACIÓN DEL CONOCIMIENTO

Una muestra de las importaciones de diez empresas reveló que en promedio sus importaciones fueron de 759, con una mediana de 701, una moda de 672 y una desviación estándar de 59. Calcula el primer coeficiente de asimetría de Pearson.

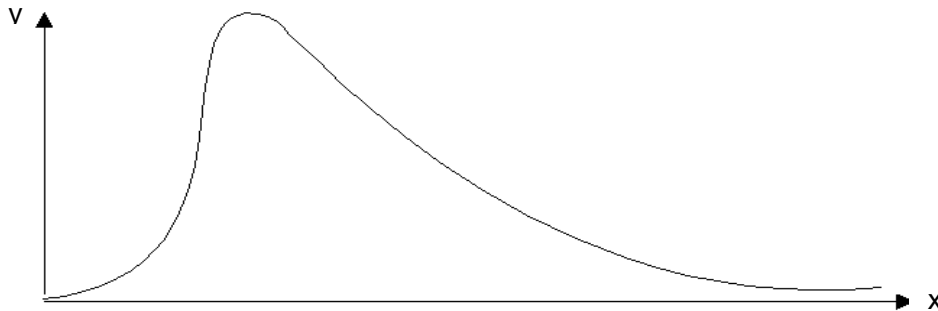
Al utilizar la fórmula del **primer coeficiente de Pearson**, la cual es:

$$S_{K1} = \frac{\bar{X} - Mo}{S}$$

$$\text{Tenemos que } S_{K1} = \frac{759 - 672}{59} = 1.474576$$

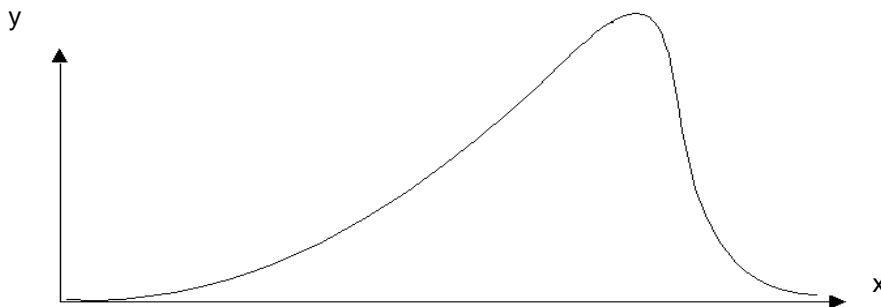
Una muestra de las ventas de doce tiendas departamentales reveló que en promedio sus ventas en un mes fueron de \$8000.00, con una mediana de \$7200.00, una moda de \$6900.00 y una desviación estándar de \$5700.00. Calcula el segundo coeficiente de asimetría de Pearson.

De la siguiente gráfica, indica qué tipo de asimetría (sesgo) tiene:



La curva presenta un sesgo hacia la derecha; por lo tanto, su asimetría es positiva.

De la siguiente gráfica, indica qué tipo de asimetría (sesgo) tiene:



EJERCICIOS.

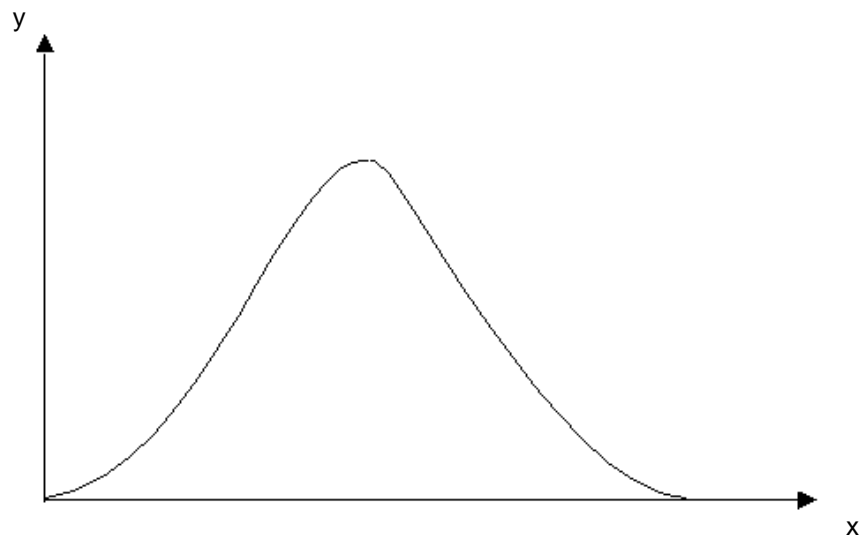
INSTRUCCIONES: Lee con atención los siguientes enunciados y contesta lo que se solicita.

1. La siguiente distribución muestra las edades de los trabajadores de una maquiladora para partes automotrices:

Edades	Número de empleados
18 - 22	8
23 - 27	45
28 - 32	60
33 - 37	68
38 - 42	23
43 - 47	16

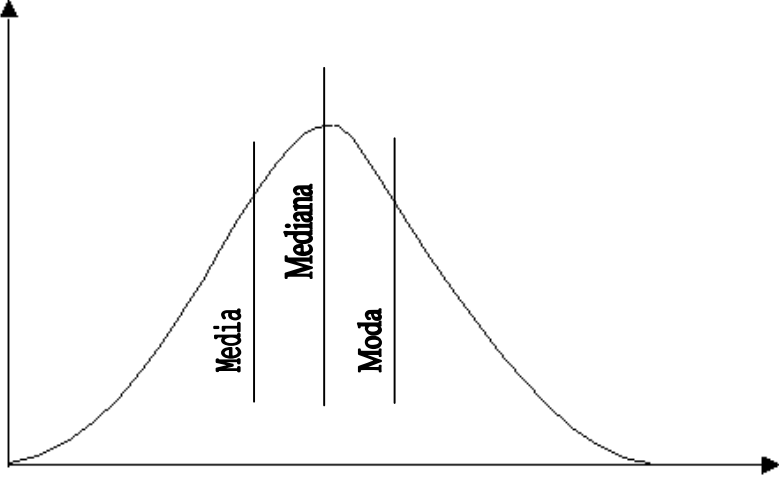
¿Cuál es el primer coeficiente de Pearson?

2. De la siguiente gráfica, señala las posiciones de las medidas de tendencia central en ella.



3. En una curva simétrica, ¿dónde se encuentran situadas las tres medidas de tendencia central?

TABLA DE COMPROBACIÓN

Número de pregunta	Respuesta correcta
1	$S_{k1} = -0.1556$
2	
3	En el centro de la curva simétrica, las tres medidas coinciden en ese punto de la curva.
Sugerencias	
<p>Para calcular el primer coeficiente de Pearson es necesario:</p> <p>Calcular la media aritmética, para nuestro ejercicio es de 32.2955.</p> <p>Calcular la moda, para este ejercicio es de 33.2955.</p> <p>Finalmente, calcular la desviación estándar, para este ejercicio de 6.166.</p>	

EJERCICIOS DE AUTOEVALUACIÓN

Cuentas con sesenta minutos para resolver estos ejercicios

INSTRUCCIONES: Lee con atención los siguientes reactivos y contesta lo que se pide.

1. La siguiente distribución de frecuencias muestra las ventas en miles de pesos de 90 empresas.

Intervalos	Frecuencia
140-159	7
160-179	30
180-199	13
200-219	25
220-239	11
240-259	4

¿Cuál es el valor de la desviación media?

2. Un licenciado en relaciones comerciales investiga sobre los gastos publicitarios de 48 empresas, obteniendo la siguiente distribución de frecuencias.

Gasto	Empresas
36-42	1
43-49	6
50-56	12
57-63	15
64-70	8
71-77	3
78-84	3

¿Cuál es el valor de la varianza?

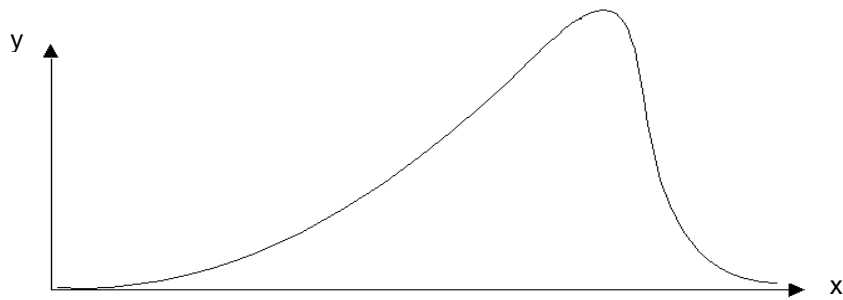
3. La siguiente distribución de frecuencias muestra las edades de los empleados de una empresa aduanal.

Edades	Empleados
18-22	8
23-27	45
28-32	60
33-37	68
38-42	23
43-47	16

I ¿Cuál es el valor de la desviación estándar?

II Calcula el primer coeficiente de asimetría de Pearson.

4. ¿Qué tipo de asimetría tiene la siguiente gráfica?



CLAVE DE RESPUESTAS

Número de pregunta	Respuesta correcta
1	D.M. = 23.259
2	$S^2 = 85.777$
3 I II	S = 6.1806 $S_{k1} = - 0.2361$
5	Negativo (sesgo hacia la izquierda)

NOTA.

Recuerda que para los tres primeros ejercicios debes calcular la media aritmética:
para el ejercicio 1, de 192.833
para el ejercicio 2, de 59.41
para el ejercicio 3, de 32.295

Para el ejercicio 4 los cálculos que debes de realizar son:
la media (32.2955)
la moda (33.7547) y
la desviación estándar (6.1806)

UNIDAD 4
CORRELACIÓN
Y REGRESIÓN LINEAL

4.1 CORRELACIÓN LINEAL

APRENDIZAJES

- Calcular la correlación lineal entre dos variables.
- Calcular el coeficiente de correlación de Pearson.

En las diferentes áreas del conocimiento existen problemas que requieren el análisis de más de una variable; por ejemplo:

- Un sociólogo puede estar interesado en saber qué clase de relación existe entre la tasa de delincuencia juvenil que hay en una comunidad y el grado de hacinamiento de los hogares que allí se encuentran.
- Un profesor de matemáticas puede estar interesado en conocer de qué manera se puede predecir el rendimiento en álgebra de un estudiante de bachillerato con base en el puntaje obtenido en una prueba de aptitud en dicha asignatura.
- Un psicólogo desea saber si existe alguna relación entre el concepto que tiene un estudiante de sí mismo y su promedio en el estudio.
- Un agrónomo desea conocer si existe relación entre la cantidad de lluvia caída y el rendimiento de ciertos productos agrícolas; es decir, si se afecta desfavorablemente tanto por la excesiva lluvia como por la sequía del suelo.

Estas relaciones y muchas otras se pueden investigar por medio del análisis de correlación y regresión lineal. Cuando el análisis involucra más de dos variables, la correlación y regresión será múltiple.

El objetivo principal de la correlación es medir la intensidad de una relación lineal entre dos variables; en esencia, la correlación es una medida de la relación entre dos variables.

La medida de correlación implica encontrar un valor numérico que exprese el grado de correspondencia o dependencia que existe entre dos variables. Al hablar de la correlación de dos variables, es necesario distinguir dos casos básicos:

Correlación positiva. Ocurre cuando al crecer o decrecer una de las variables, la otra crece o decrece paralelamente, por ejemplo: a medida que se eleva el nivel de vida de una población, tiende a aumentar el consumo de artículos que no son de primera necesidad.

Correlación negativa. Ocurre cuando al crecer alguna de las variables, la otra decrece; por ejemplo: a medida que se amplían los sistemas de salubridad y medicina preventiva, decrece el índice de mortalidad por enfermedades infecto-contagiosas.

Estas dos correlaciones y otras más, se pueden mostrar utilizando los diagramas de dispersión.

DIAGRAMAS DE DISPERSIÓN

La forma más sencilla para predeterminar si existe o no una correlación entre dos variables es construyendo un diagrama de dispersión. Al localizar los puntos sobre los ejes cartesianos, podemos ver lo siguiente:

Si los valores de las variables X y Y van aumentando, entonces existe una correlación positiva.

Por ejemplo: al correlacionar las edades del marido y de la esposa en las parejas conyugales; es decir, a mayor edad del marido, mayor edad de la esposa.

Si los valores de la variable X aumentan y ocasionan que los valores de la variable Y disminuyan, entonces existe una correlación negativa.

Por ejemplo: al correlacionar el número de accidentes de trabajo acaecidos en un espacio de tiempo, con el número de dispositivos de seguridad operantes en la planta de una industria; es decir, a mayor número de dispositivos de seguridad, menor número de accidentes de trabajo.

COEFICIENTE DE CORRELACIÓN DE PEARSON

De los diversos coeficientes de correlación que existen, el más conocido y utilizado es el Coeficiente de Pearson, el cual se simboliza con la letra minúscula r . **Se calcula dividiendo la suma de los productos de las desviaciones de cada variable de X y Y con respecto a sus medias (suma que se denomina covarianza¹ de X y Y), entre el producto de las desviaciones estándar de ambas variables;** de manera práctica, el coeficiente de correlación de Pearson se calcula con la siguiente fórmula:

$$r = \frac{N(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

donde:

N: es el número de pares de datos.

A través de algunos ejemplos, mostraremos cómo se utiliza la fórmula del coeficiente de Pearson, para ver la correlación que puede existir entre dos variables.

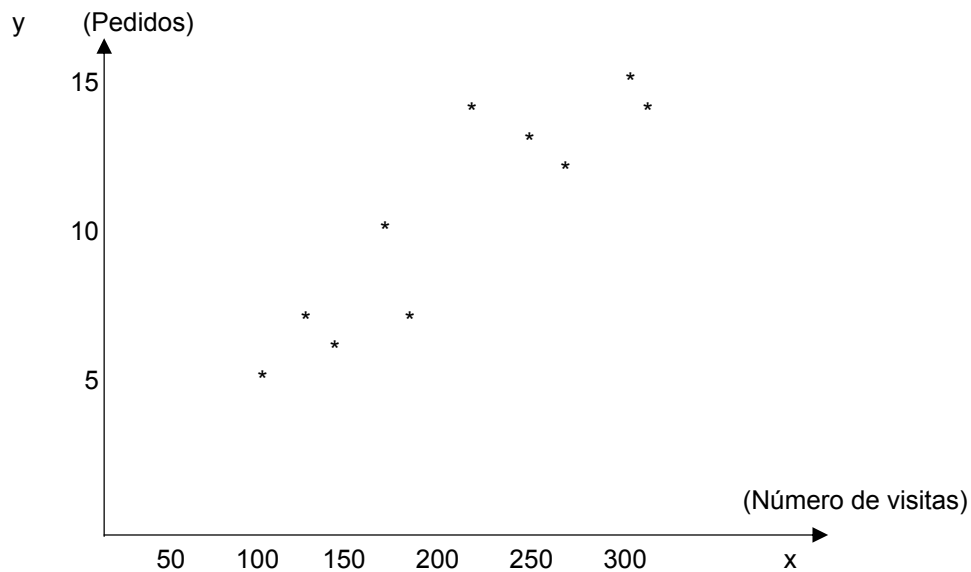
¹ La covarianza es una medida de la fuerza con la que dos variables x y y se encuentran linealmente relacionadas (correlación positiva o negativa), es decir, todos los puntos graficados deben pertenecer a una línea recta.

APLICACIÓN DEL CONOCIMIENTO

El departamento de ventas de una empresa realiza un análisis comparativo entre el volumen de pedidos levantados y el número de visitas efectuadas por sus diez vendedores en cierto lapso. Todos los vendedores trabajan en zonas similares, en lo referente al número de clientes que maneja cada uno y el potencial de compra de dichos clientes. Los resultados de la comparación fueron :

Vendedor	Pedidos	Visitas realizadas
1	13.4	245
2	10.3	172
3	15.1	291
4	6.9	124
5	7.3	191
6	14.2	218
7	5.2	101
8	11.8	259
9	14.3	307
10	5.5	142

Considerando al número de visitas como la primera variable (x) y al monto de los pedidos como la segunda variable (y), se construye un diagrama de dispersión y se indica el tipo de correlación existente.



El diagrama de dispersión indica que existe una correlación positiva.

Para darle su interpretación correcta es necesario conocer las siguientes características del coeficiente de correlación lineal, las cuales son:

- El valor de r es un número que satisface a la desigualdad: $-1 \leq r \leq 1$.
- Cuando la relación de dos variables es perfectamente positiva; o sea, cuando al variar la primera variable, la segunda variable varía en las mismas proporciones y en la misma dirección, el coeficiente de correlación es $+1$.

- c) Cuando la relación de dos variables es perfectamente negativa; o sea, al variar la primera variable, la segunda variable varía en las mismas proporciones pero en dirección contraria, el coeficiente de correlación es -1.
- d) Cuando no existe relación entre dos variables; o sea, cuando al variar la primera variable, las variaciones de la segunda variable no reflejan dependencia o conexión alguna con las variaciones de la primera variable, el coeficiente de correlación lineal es cero.

Lo anterior significa que, entre 0 y +1 cabe toda una gama de *correlaciones positivas*, que serán tanto más directamente proporcionales, cuanto más se acerquen a +1. Y entre -1 y 0 cabe toda una gama de *correlaciones negativas*, que serán tanto más inversamente proporcionales, cuanto más se acerquen a -1. Recuerda que en los coeficientes de correlación, cuanto más cerca de cero, indicarán ausencia de correlación.

Ahora analiza el siguiente ejemplo

Al efectuar un estudio sobre la marca de cierto producto se encontró que cincuenta personas habían usado anteriormente dicha marca y la habían cambiado. La relación entre el tiempo que había usado la marca, antes de sustituirla por otra, y el número de ex usuarios en cada caso, fue de:

Años de uso (X) :	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Núm. de ex usuarios (Y) :	8	6	9	4	6	5	3	2	4	3

Construye un diagrama de dispersión e indica el tipo de correlación.

A continuación se muestran las puntuaciones obtenidas por diez personas en una prueba de destreza manual para hacer una artesanía, con base en dos variables, X y Y:

X: 20 18 16 15 14 12 12 10 8 5
Y: 12 16 10 14 12 10 9 8 7 2

¿Cuál es el valor del coeficiente de correlación de Pearson y qué significa?

Se recomienda elaborar una tabla que muestre las variables para realizar los cálculos directos del coeficiente de correlación de Pearson, de la siguiente manera:

Columnas	1	2	3	4	5
VARIABLES	X	Y	X ²	Y ²	XY
	20	12	400	144	240
	18	16	324	256	288
	16	10	256	100	160
	15	14	225	196	210
	14	12	196	144	168
	12	10	144	100	120
	12	9	144	81	108
	10	8	100	64	80
	8	7	64	49	56
	5	2	25	4	10
	$\Sigma X = 130$	$\Sigma Y = 100$	$\Sigma X^2 = 1878$	$\Sigma Y^2 = 1138$	$\Sigma XY = 1440$

Con base en la tabla anterior, observa que en las columnas (1) y (2) están los datos de las variables X y Y del enunciado (pares de puntuaciones); en la columna (3) se obtuvieron los cuadrados de las puntuaciones de X, y en la columna (4) los cuadrados de las puntuaciones de Y. En la columna (5) están los productos de X por Y, finalmente se calcula la suma de los valores de las cinco columnas, los cuales se sustituyen en la fórmula del coeficiente de correlación de Pearson, de la siguiente forma:

$$r = \frac{N(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

por lo que tendremos

$$r = \frac{10(1440) - (130)(100)}{\sqrt{[10(1878) - (130)^2][10(1138) - (100)^2]}}$$

$$r = \frac{14400 - 13000}{\sqrt{(18780 - 16900)(11380 - 10000)}} = \frac{1440}{\sqrt{(1880)(1380)}}$$

$$r = \frac{1440}{\sqrt{2594400}} = \frac{1440}{1610} = 0.89$$

Con todas las características del coeficiente de correlación lineal, ya mencionadas, podemos interpretar el resultado del ejemplo anterior, donde calculamos el coeficiente de correlación de Pearson. Entonces, como $r = 0.89$, podemos concluir que la correlación que existe es fuerte y positiva entre estas dos variables.

Como observas, lo único tedioso es la elaboración de la tabla, pero ésta concentra todos los resultados que necesitas para aplicar la fórmula del coeficiente de correlación de Pearson de una manera sencilla y clara.

Ahora calcula el coeficiente de correlación de Pearson de la siguiente tabla de valores, que muestra los puntajes obtenidos por diez personas al iniciar estudios superiores en una prueba de aptitud (X) para la profesión y los puntajes obtenidos en satisfacción por el trabajo que desarrollan (Y).

X:	245	172	291	124	191	218	101	259	307	142
Y:	13.4	10.3	15.1	6.9	7.3	14.2	5.2	11.8	14.3	5.2

EJERCICIOS

INSTRUCCIONES: Lee con atención los siguientes reactivos y realiza lo que se solicita.

1. La siguiente información muestra la relación entre variables X (experiencia en ventas) y Y (cantidad vendida), de ciertos vendedores de una tienda departamental:

Experiencia en ventas (X):	3	4	4	5	5	6	6	7	7	7	8	9	9	10	10
Ventas (Y):	2	3	4	3	4	3	4	4	5	6	5	6	7	7	8

- I. Construye el diagrama de dispersión.

- II. Calcula el coeficiente de correlación de Pearson.

- III. Indica el tipo de correlación que existe.

2. Los siguientes datos muestran el número de horas (X) dedicadas a preparar un examen, y la calificación (Y) obtenida en dicha prueba.

Horas para el estudio (X):	2	3	3	4	4	5	5	6	6	6	7	7	7	8	8
Calificaciones (Y):	5	5	7	5	7	7	8	6	9	8	7	9	10	8	9

I. Construye el diagrama de dispersión.

II. Obtén el coeficiente de correlación de Pearson.

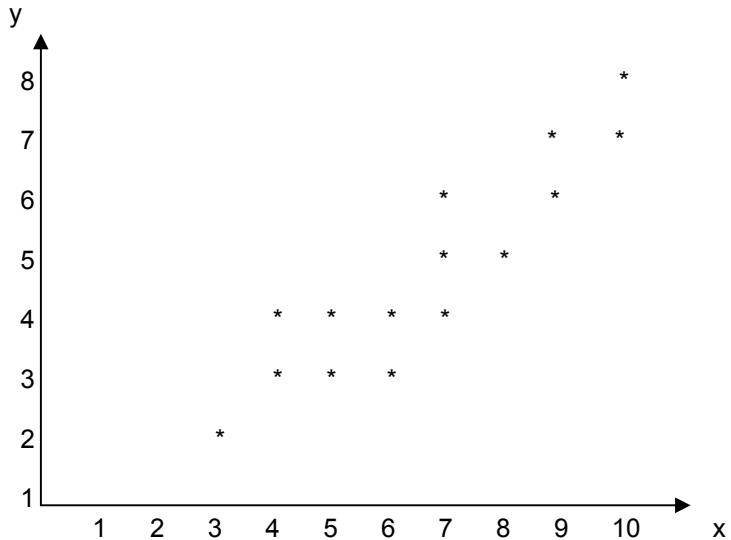
III. Señala cómo es la correlación entre las variables.

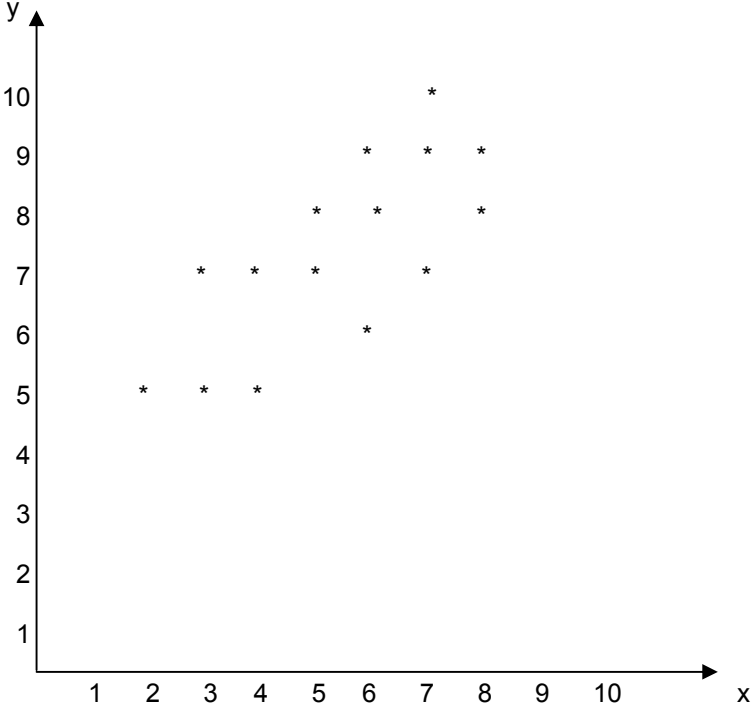
3. Los siguientes resultados muestran las aptitudes de seis estudiantes a través de exámenes para la materia de estadística, y para idiomas.

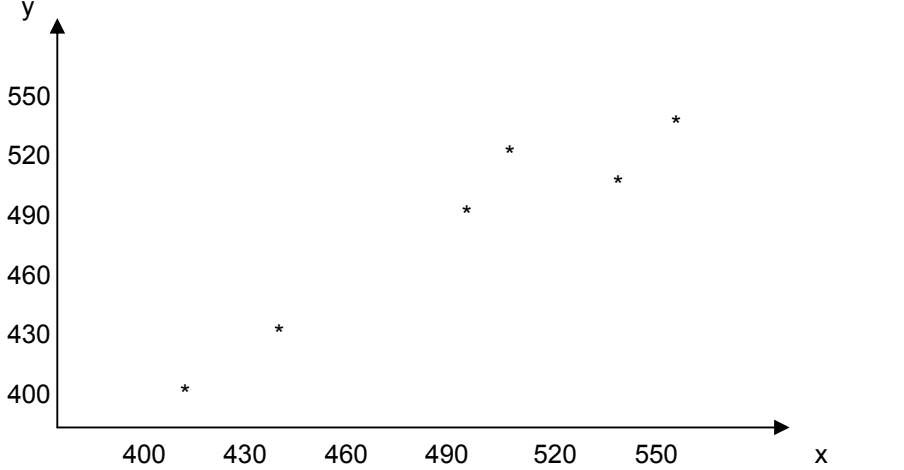
Idiomas (Y):	525	515	510	495	430	400
Estadística (X):	550	535	535	520	455	420

- I. Elabora un diagrama de dispersión.
- II. Calcula el coeficiente de correlación de Pearson.
- III. Indica qué tipo de correlación existe entre las variables.

TABLA DE COMPROBACIÓN

Número de pregunta	Respuesta correcta
1	<p>I. Diagrama de dispersión.</p>  <p>II. Los valores de los resultados al elaborar la tabla para aplicar la fórmula del coeficiente de correlación de Pearson son:</p> <p>$\sum X = 100$</p> <p>$\sum Y = 71$</p> <p>$\sum XY = 523$</p> <p>$\sum X^2 = 736$</p> <p>$\sum Y^2 = 379$;</p> <p>que al sustituir en la expresión de Pearson, tenemos que: $r = 0.91$</p> <p>III. La correlación entre estas dos variables es positiva, gráfica y algebraicamente.</p>

Número de pregunta	Respuesta correcta
2	<p>I. Diagrama de dispersión</p>  <p>II. El coeficiente de correlación de Pearson es 0.741</p> <p>III. La correlación es positiva gráfica y algebraicamente.</p>

Número de pregunta	Respuesta correcta														
3	<p>I. Diagrama de dispersión</p>  <table border="1" data-bbox="432 607 1339 1075"> <caption>Data points from the scatter plot</caption> <thead> <tr> <th>x</th> <th>y</th> </tr> </thead> <tbody> <tr> <td>410</td> <td>400</td> </tr> <tr> <td>440</td> <td>430</td> </tr> <tr> <td>490</td> <td>490</td> </tr> <tr> <td>510</td> <td>520</td> </tr> <tr> <td>540</td> <td>500</td> </tr> <tr> <td>560</td> <td>540</td> </tr> </tbody> </table> <p>II. El coeficiente de correlación de Pearson es 0.999</p> <p>III. La correlación es positiva gráfica y algebraicamente.</p>	x	y	410	400	440	430	490	490	510	520	540	500	560	540
x	y														
410	400														
440	430														
490	490														
510	520														
540	500														
560	540														
Sugerencias															
<p>Comprueba que seguiste adecuadamente los procedimientos para encontrar el coeficiente de correlación de Person y para construir los correspondientes diagramas de dispersión.</p>															

4.2 REGRESIÓN LINEAL

APRENDIZAJES

- Obtener la ecuación de regresión lineal.
- Aplicar los conceptos de correlación y regresión lineal en la solución de problemas.

Recuerda que la correlación informa sobre la intensidad de una relación lineal, nos dice cuál es la relación numérica exacta.

Por ejemplo, el coeficiente de correlación calculado para los datos del puntaje que obtuvieron al hacer artesanías, implica que hay una correlación lineal entre el número de los puntajes y las artesanías que realizaron los artesanos. Esto significa que el número de puntaje sirve para predecir el número de artesanías. Sin embargo, la correlación no informa cómo determinar un valor de “y” dado un valor de “x”. Lo anterior se hace con un exponencial y logarítmica.

Podemos recordar fácilmente que para trazar una línea recta es necesario tener dos valores para **x**, para obtener los valores correspondientes de **y**, y tener dos puntos o parejas ordenadas que al localizarlas en los ejes cartesianos y al unirlos resulte una gráfica lineal. Entonces podemos pensar que **para la relación entre dos variables X y Y, con un alto coeficiente de correlación lineal, puede suponerse una relación lineal del tipo que existe entre las variables de los ejercicios del cálculo del coeficiente de correlación de Pearson.**

A la ecuación $y = ax + b$, que describe la relación lineal entre las variables x y y , se le llama ecuación de regresión, y su gráfica, recta de regresión. La pendiente de la recta y la ordenada al origen de la recta de regresión se obtienen mediante las siguientes expresiones:

$$b = \frac{\sum X^2 \sum Y - \sum X (\sum XY)}{N \sum X^2 - (\sum X)^2}$$

$$a = \frac{N(\sum XY) - \sum X \sum Y}{N \sum X^2 - (\sum X)^2}$$

APLICACIÓN DEL CONOCIMIENTO

La siguiente tabla muestra los valores obtenidos en asistencia a juntas de trabajo para hombres como para mujeres. Con base en esta información, encuentra la ecuación de regresión lineal para estas dos variables.

X: 2 3 3 4 4 4 4 5 5 5 6 6 8
 Y: 1 2 3 1 2 3 4 2 3 4 3 5 5

Realicemos la tabla correspondiente donde concentramos los valores que necesitamos para aplicar las expresiones para calcular la pendiente de la línea recta (a) y la ordenada al origen (b); es decir, los valores $\sum X$, $\sum Y$, $\sum XY$ y $\sum X^2$.

X	Y	X ²	XY
2	1	4	2
3	2	9	6
3	3	9	9
4	1	16	4
4	2	16	8
4	3	16	12
4	4	16	16
5	2	25	10
5	3	25	15
5	4	25	20
6	3	36	18
6	5	36	30
8	5	64	40
$\sum X = 59$	$\sum Y = 38$	$\sum X^2 = 297$	$\sum XY = 190$

Con base en los datos obtenidos, los sustituimos en las fórmulas correspondientes que ya conoces, por lo tanto:

$$b = \frac{\sum X^2 \sum Y - \sum X (\sum XY)}{N \sum X^2 - (\sum X)^2}$$

$$b = \frac{(297)(38) - (59)(190)}{13(297) - (59)^2} = \frac{11286 - 11210}{3861 - 3481} = \frac{76}{380} = 0.2$$

$$a = \frac{\sum X^2 \sum Y - \sum X (\sum XY)}{N \sum X^2 - (\sum X)^2}$$

$$a = \frac{13(190) - (59)(38)}{380} = \frac{2470 - 2242}{380} = \frac{228}{380} = 0.6$$

Con base en los cálculos de **a** y de **b**, la ecuación de regresión es:

$$y = ax + b \Rightarrow y = 0.6x + 0.2$$

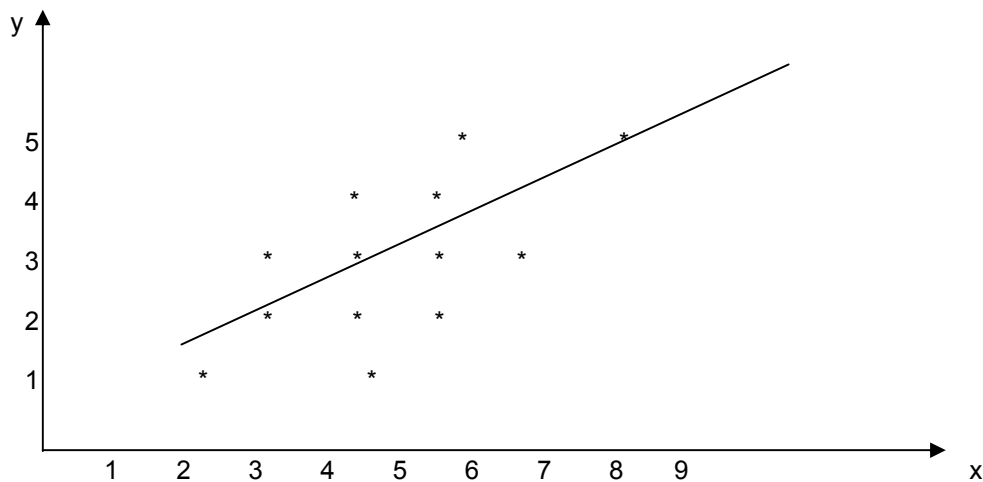
Si le damos valores a la ecuación de regresión, como por ejemplo:

si $x = 3$; $y = 2$, se obtiene:

$$y = 0.6(3) + 0.2 = 1.8 + 0.2 = 2, \quad \text{por lo tanto: } P_1(3, 2)$$

$$y = 0.6(8) + 0.2 = 4.8 + 0.2 = 5, \quad \text{por lo tanto: } P_2(8, 5)$$

Los puntos que acabamos de calcular, P_1 y P_2 , son los puntos por donde la recta de regresión pasa. Hagamos el diagrama de dispersión para comprobar esta afirmación; recuerda que para construir este diagrama, sólo necesitamos las parejas ordenadas de las variables de la tabla del enunciado y localizarlas en los ejes coordenados.



La ecuación de regresión permite tener un valor de la variable Y cuando se tiene un valor de la variable X; lo anterior permite realizar estimaciones si, por ejemplo, $X = 15$, la estimación se obtiene al sustituir el valor de X en la ecuación de regresión, la cual es $y = 0.6x + 0.2$, dando por resultado que $y = 9.2$.

Calcula la ecuación de regresión lineal y verifícala con el diagrama de dispersión correspondiente de la siguiente tabla de valores, que muestra los puntajes que obtuvieron diez personas al iniciar estudios superiores en una prueba de aptitud (X) para la profesión y los puntajes obtenidos en satisfacción por el trabajo que desarrollan (Y).

X:	245	172	291	124	191	218	101	259	307	142
Y:	13.4	10.3	15.1	6.9	7.3	14.2	5.2	11.8	14.3	5.2

EJERCICIOS

INSTRUCCIONES: Lee con atención los siguientes reactivos y realiza lo que se solicita.

1. La siguiente información muestra la relación entre variables X (experiencia en ventas) y Y (cantidad vendida), de ciertos vendedores de una tienda departamental:

Experiencia en ventas (X): 3 4 4 5 5 6 6 7 7 7 8 9 9 10 10
Cantidad vendida (Y): 2 3 4 3 4 3 4 4 5 6 5 6 7 7 8

- I. Calcula la ecuación de regresión lineal.

- II. Construye el diagrama de dispersión localizando los puntos P_1 y P_2 .

2. Los siguientes datos muestran el número de horas (X) dedicadas a preparar un examen, y la calificación (Y) obtenida en dicha prueba.

Horas para el estudio (X) : 2 3 3 4 4 5 5 6 6 6 7 7 7 8 8
Calificaciones (Y) : 5 5 7 5 7 7 8 6 9 8 7 9 10 8 9

I. Calcula la ecuación de regresión lineal.

II. Construye el diagrama de dispersión localizando los puntos P_1 y P_2 .

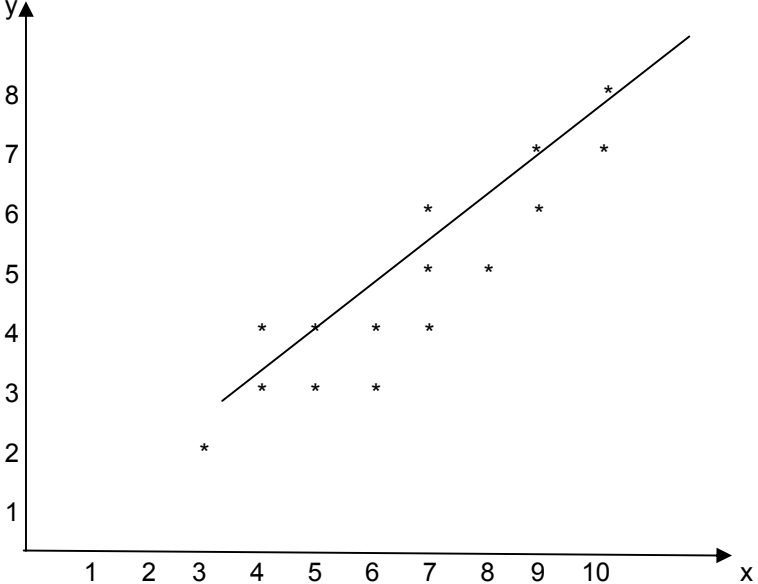
3. Los siguientes resultados muestran las aptitudes de seis estudiantes a través de exámenes para la materia de estadística y para idiomas.

Idiomas (Y):	525	515	510	495	430	400
Estadística (X):	550	535	535	520	455	420

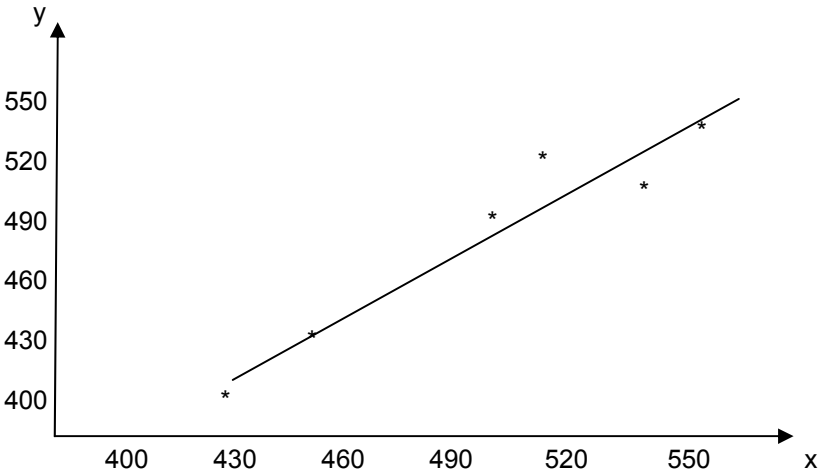
I. Calcula la ecuación de regresión lineal.

II. Construye el diagrama de dispersión localizando los puntos P_1 y P_2 .

TABLA DE COMPROBACIÓN

Número de pregunta	Respuesta correcta
1	<p>I. Para calcular la ecuación de regresión lineal, necesitamos de valores que al sustituirlos en las expresiones de a y b, encontremos los valores de la pendiente de la recta y su ordenada al origen:</p> <p>$a = 0.7163$ y $b = -0.0423$</p> <p>y la ecuación es $y = 0.7163x - 0.04231$</p> <p>II. El diagrama de dispersión es para $P_1(4, 2.82)$ y $P_2(10, 7.12)$:</p> 

Número de Pregunta	Respuesta correcta
2	<p>I. Para calcular la ecuación de regresión lineal, necesitamos de valores que al sustituirlos en las expresiones de a y b, encontremos los valores de la pendiente de la recta y su ordenada al origen, entonces:</p> <p>$a = 0.625$ y $b = 3.958$</p> <p>la ecuación es $y = 0.625x + 3.958$</p> <p>II. El diagrama de dispersión es $P_1(3, 5.833)$ y $P_2(8, 8.958)$</p> <p>The scatter plot displays 15 data points with a positive linear trend. The regression line is represented by the equation $y = 0.625x + 3.958$. Two points are highlighted: $P_1(3, 5.833)$ and $P_2(8, 8.958)$.</p>

Número de pregunta	Respuesta correcta
3	<p>I. Para calcular la ecuación de regresión lineal, necesitamos de valores que al sustituirlos en las expresiones de a y b, encontremos los valores de la pendiente de la recta y su ordenada al origen, entonces:</p> <p>$a = 0.9818$; y $b = -14.1886$</p> <p>la ecuación es : $y = 0.9818x - 14.1886$</p> <p>II. Diagrama de dispersión: $P_1(430, 407.98)$ y $P_2(550, 525.80)$</p> 
Sugerencias	
<p>Para encontrar la ecuación de regresión lineal es necesario calcular los siguientes valores, que al sustituirlos en las expresiones de a y b, encontremos los valores de la pendiente de la recta y su ordenada al origen.</p> <p>Ejercicio 1: $\sum X = 100$, $\sum Y = 71$, $\sum XY = 523$ y $\sum X^2 = 736$</p> <p>Ejercicio 2: $\sum X = 81$, $\sum Y = 110$, $\sum XY = 625$ y $\sum X^2 = 487$</p> <p>Ejercicio 3: $\sum X = 3015$, $\sum Y = 2875$, $\sum XY = 1458175$ y $\sum X^2 = 1528775$</p>	

EJERCICIOS DE AUTOEVALUACIÓN

Cuentas con sesenta minutos para resolver los problemas.

INSTRUCCIONES: Lee con atención los siguientes reactivos y contesta lo que se solicita.

1. La siguiente información muestra la relación entre las variables X (citas de mala conducta) y Y (visitas a la escuela), de ciertos padres de familia:

Citas de mala conducta (X):	1	2	2.5	3	4	4.5	5	6
Visitas a la escuela (Y):	1.5	2.2	2	2.5	4.5	4	5	5

I. Construye el diagrama de dispersión.

II. Indica el tipo de correlación que existe.

III. Calcula el coeficiente de correlación de Pearson.

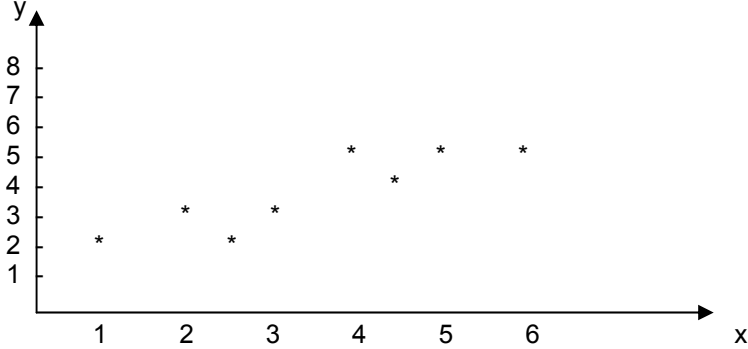
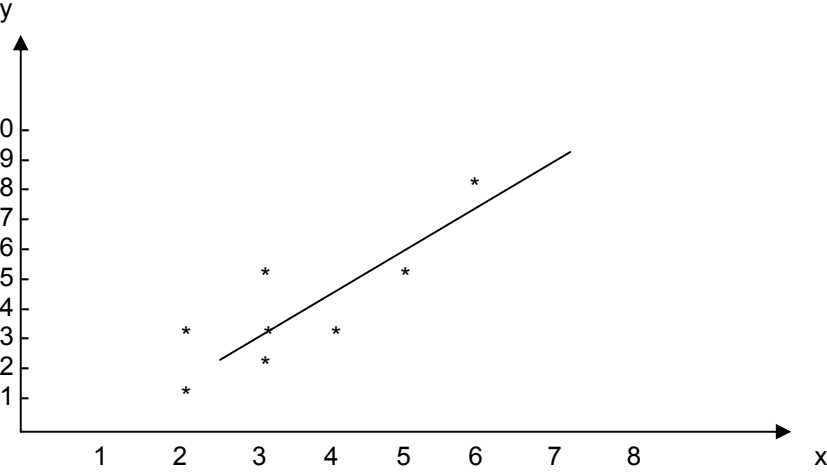
2. La siguiente información muestra la relación entre las variables X (llegadas temprano al trabajo) y Y (estímulos a la puntualidad), de ciertos trabajadores:

Llegadas temprano al trabajo (X):	2	3	2	3	4	3	5	6
Estímulos a la puntualidad (Y):	1	2	3	3	3	5	5	7

I. Calcula la ecuación de regresión lineal.

II. Construye el diagrama de dispersión localizando los puntos P_1 y P_2 .

CLAVE DE RESPUESTAS

Numero de pregunta	Respuesta correcta
1	<p>I. Diagrama de dispersión.</p>  <p>II. Correlación positiva.</p> <p>III. El coeficiente de correlación es de: $r = 0.9492$</p>
2	<p>I. Para calcular la ecuación de regresión lineal, necesitamos de valores que al sustituirlos en las expresiones de a y b, encontremos los valores de la pendiente de la recta y su ordenada al origen, entonces:</p> $a = 2.1071 \quad y \quad b = -3.75$ <p>la ecuación de regresión lineal es $y = 2.1071 x - 3.75$</p> <p>II. Diagrama de dispersión para $P_1(3, 3)$ y $P_2(5, 5.3)$ es:</p> 

UNIDAD 5

ELEMENTOS DE PROBABILIDAD

5.1-5.2 INTRODUCCIÓN A LA PROBABILIDAD Y FRECUENCIA RELATIVA

APRENDIZAJES

- Caracterizar la probabilidad estadística.
- Calcular la frecuencia relativa de valores de una variable aleatoria.
- Comprender los conceptos básicos de evento, muestra y población.

La **probabilidad** es el estudio de los fenómenos puramente aleatorios. La probabilidad apareció con base en los juegos de azar; cuando se utilizó la palabra probabilidad fue para indicar la posibilidad de que ocurra un evento o resultado.

El mundo en que vivimos está lleno de incertidumbre; las situaciones que implican incertidumbre varían de simples juegos de azar, como los dados y naipes, hasta problemas en campos tan variados e importantes como son las ciencias físicas, las sociales, la industria y los seguros, por mencionar algunos. Los problemas representativos de estos campos implican predicciones de lo que sucederá a futuro; es decir, qué probabilidad de ocurrencia existe para asegurar las predicciones.

Los primeros estudios sobre probabilidad fueron motivados por la posibilidad de acierto o de fracaso en los juegos de azar; es decir, qué ocurrencia tiene un suceso entre varios posibles.

La probabilidad según la frecuencia relativa, es el punto de vista más ampliamente sostenido, debido a que la frecuencia relativa es un cociente, como recuerdas, entre la frecuencia absoluta y el número de veces que se repite el experimento, entonces:

Consideremos un evento E que se produce en n repeticiones o ensayos de algún experimento; de acuerdo con el concepto de frecuencia relativa de ocurrencia ($f_a = NA/N$), la probabilidad del evento E, denotado por $P(E)$, es igual a la frecuencia relativa de ocurrencia del evento E, cuando n se aproxima al infinito (n repeticiones . . .).

Si hacemos nE que sea igual al número de veces que ocurre el evento E en n ensayos de un experimento, podemos estimar la probabilidad de E [$P(E)$] mediante la expresión:

$$P(E) = \frac{nE}{n}$$

La expresión anterior es la frecuencia relativa de ocurrencia del evento E. Esta interpretación de probabilidad depende de la idea de regularidad estadística, que establece que las frecuencias relativas tienden a estandarizarse y aproximarse a un valor fijo después de gran número de repeticiones de un experimento.

Con base en lo aprendido anteriormente sobre estadística descriptiva (frecuencia relativa en distribuciones de frecuencias), es necesario recordar algunos conceptos de probabilidad que se mencionaron anteriormente, cómo:

- Al conjunto de los posibles resultados de un experimento se le llama espacio muestral, espacio muestra o espacio de eventos.
- Al elemento o elementos que se encuentran en el espacio muestral se le llama evento.
- La población es la totalidad de todas las posibles observaciones.

APLICACIÓN DEL CONOCIMIENTO

Observa cuidadosamente el siguiente ejemplo. ¿Cuál es la probabilidad de obtener sol (cara) al lanzar una moneda?

En este ejemplo, observamos que la moneda sólo tiene dos resultados: águila (sello) o sol (cara), por lo que existen dos posibles casos de ocurrencia, por lo tanto sólo hay un resultado favorable, que aparezca sol, entonces:

$$P(E) = \frac{nE}{n} = \frac{1}{2} = 0.5$$

Veamos otro ejemplo. ¿Cuál es el espacio muestra del experimento de lanzar una moneda?

Recuerda que el espacio muestral es el formado por todos los posibles resultados del experimento, entonces:

$$S = \{a, s\}$$

Donde **S** representa al espacio muestral; **a** es el resultado de que aparezca águila y **s** es el resultado de que aparezca sol.

¿Cuál es la probabilidad de obtener un 3 al lanzar un dado?

¿Cuál es el espacio de eventos del experimento de lanzar un dado?

EJERCICIOS

INSTRUCCIONES: Lee con atención los siguientes reactivos y coloca en el paréntesis de la izquierda la letra de la opción correcta.

1. () ¿Cuál es el nombre que se le da al número asignado a un evento que mide la creencia de que ese evento pueda ocurrir?
 - a) Estadística.
 - b) Probabilidad.
 - c) Estadígrafo.
 - d) Medida de dispersión.

2. () ¿Cuál es el nombre que recibe el cociente de los casos favorables entre todos los casos posibles de un experimento dado?
 - a) Frecuencia absoluta.
 - b) Estadística.
 - c) Posibilidad.
 - d) Distribución.

INSTRUCCIONES: Lee con atención los siguientes enunciados y contesta lo que se te pide.

3. ¿Cuál es la probabilidad de sacar un “as” de una baraja de 52 cartas?

4. Una urna contiene tres bolas rojas, cinco bolas blancas y cuatro azules.
 - I. ¿Cuál es la probabilidad de extraer de la urna una bola roja?

 - II. ¿Qué probabilidad existe de que al extraer una bola, ésta sea blanca?

 - III. ¿Cuál es la probabilidad de extraer una bola azul?

5. En una urna hay dos mil bolas rojas y tres mil bolas blancas; si se extraen 100 bolas al azar, ¿cuál es la probabilidad que sean rojas?
6. Se lanzan dos dados, ¿cuál es la probabilidad de que aparezcan dos números iguales?
7. Se lanzan dos dados, ¿cuál es la probabilidad de que aparezcan dos números diferentes?
8. Escribe el espacio muestra del experimento de lanzar tres monedas.
9. ¿Cuál es el espacio de eventos del experimento de una urna que contiene dos bolas blancas y una negra, donde se extraen al azar dos bolas? El experimento se realiza con reemplazo.

TABLA DE COMPROBACIÓN

Número de pregunta	Respuesta correcta
1	b
2	c
3	$P (A) = 0.077$
4	I. $P (R) = 0.25$ II. $P (B) = 0.417$ III. $P (A) = 0.333$
5	$P (E) = 0.4$, entonces el 40% de las bolas son rojas.
6	P (números iguales) = 0.1667
7	P (números diferentes) = 0.8333
8	$S = \{(sss), (ssa), (sas), (saa), (aaa), (aas), (asa), (ass)\}$
9	$S = \{(b1, b1), (b1, b2), (b1, n), (b2, b1), (b2, b2), (b2,n), (n, b1), (n, b2), (n, n)\}$
Sugerencias	
<p>Si persisten las dudas, puedes consultar los libros cuyos títulos son:</p> <p>Estadística, Primer Curso. Portilla. Ed. Interamericana, 1987, México, p. 93-109.</p> <p>Curso Práctico de Estadística. Lincoyan Portus G. Ed. McGraw-Hill, 1988, México, p. 135-142.</p>	

5.3 NOCIONES DE PROBABILIDAD

APRENDIZAJES

- Calcular la probabilidad de eventos excluyentes y no mutuamente excluyentes.
- Calcular la probabilidad condicional de dos eventos.
- Calcular la probabilidad de eventos independientes.

La probabilidad de dos o más eventos puede ser mutuamente excluyente y no mutuamente excluyente.

EVENTOS MUTUAMENTE EXCLUYENTES

Estos eventos son también llamados eventos disjuntos; dos o más eventos son considerados mutuamente excluyentes, si éstos no pueden ocurrir conjuntamente; es decir, la ocurrencia de cualquiera de ellos excluye la ocurrencia de los otros. Sean dos eventos, A y B, mutuamente excluyentes o disjuntos, entonces las probabilidades del evento A o B es:

$$(P \cup B) = P(A) + P(B)$$

En este tipo de eventos no existe la intersección de los eventos A y B .

EVENTOS NO MUTUAMENTE EXCLUYENTES

Estos eventos llamados también eventos que se intersectan, son eventos que se traslapan parcialmente; es decir, si parte de uno de ellos y parte de otro de ellos ocurren conjuntamente. Sean los eventos A y B que se intersectan; es decir, parte del evento A es también parte del evento B , entonces:

$$(P \cup B) = P(A) + P(B) - P(A \cap B)$$

PROBABILIDAD CONDICIONAL

Si A y B son dos eventos dependientes, entonces la probabilidad de que ocurra tanto A como B es igual al producto de la probabilidad de A multiplicada por la probabilidad de B, con la condición de que A haya ocurrido, denotado por P (B|A) , entonces:

$$P(A \cap B) = P(A) P(B|A)$$

Por lo tanto, **la probabilidad condicional de cualquier evento es la probabilidad de que este evento ocurra con la condición de que el otro evento haya ocurrido, entonces:**

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{donde } P(A) \neq 0$$

SUCESOS INDEPENDIENTES

Recuerda que **si dos eventos son independientes, la ocurrencia de uno de ellos no afecta a la probabilidad de que el otro ocurra o deje de ocurrir.**

Sean A y B dos eventos independientes, entonces:

$$P(A \cap B) = P(A) P(B)$$

APLICACIÓN DEL CONOCIMIENTO

Observa con atención el siguiente ejemplo. ¿Cuál es la probabilidad de que al lanzar un dado se obtenga el número 4 ó 6?

Asignemos: evento $A = 4$ y evento $B = 6$, entonces sus probabilidades de ocurrencia son: $P(A) = 1/6$ y $P(B) = 1/6$, por lo tanto:

$$(P \cup B) = P(A) + P(B) \Rightarrow (P \cup B) = 1/6 + 1/6 = 2/6 = 1/3$$

Sea que se identifique S como el evento de que asistas a una escuela estatal y P el evento de que asistas a una escuela privada. Considera que no asistirás a ambas escuelas simultáneamente; si la probabilidad de que asistas a la estatal es 0.4 y a la privada es 0.25, ¿cuál es la probabilidad de que asistas a la estatal o a la privada, y de que no asistas a ninguna de ellas?

¿Cuál es la probabilidad de que al lanzar un dado, aparezca un número par o un número primo?

El espacio muestra de lanzar un dado es: $D = \{1, 2, 3, 4, 5, 6\}$, si el evento $A = (2, 4, 6)$ y $B = (2, 3, 5)$, entonces sus probabilidades son:

$$P(A) = 3/6 = 1/2, P(B) = 3/6 = 1/2 \text{ y } P(A \cap B) = 1/6, \text{ entonces:}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B), \text{ por lo tanto:}$$

$$P(A \cup B) = 1/2 + 1/2 - 1/6 = 5/6$$

Suponga que la probabilidad de que usted asista a un Politécnico es 0.6, la probabilidad de que trabaje es 0.7 y la probabilidad de que realice ambas actividades es 0.5. ¿Cuál es la probabilidad de que asista a un Politécnico o trabaje?

Se lanzan tres monedas, ¿cuál es la probabilidad de que todas sean soles, si la primera moneda es sol?

Si A es el evento, “la primera moneda es sol” y B es el evento, “las tres monedas son soles”, entonces:

$$A = \{(sss), (ssa), (sas), (saa)\}, \Rightarrow P(A) = 4/8$$

$$B = \{(sss)\}, \Rightarrow P(B) = 1/8, \text{ por lo que: } P(A \cap B) = 1/8, \text{ entonces:}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{1/8}{4/8} = \frac{1}{4} = 0.25$$

Se lanzan un par de dados, si los números que resultan son diferentes, ¿cuál es la probabilidad de que su suma sea par?

La probabilidad de que un hombre vivirá diez años más es $\frac{1}{4}$, la probabilidad de que su esposa vivirá diez años más es $\frac{1}{3}$; ¿cuál es la probabilidad de que ambos estén vivos dentro de diez años?

Como son eventos independientes, entonces:

$$P(A \cap B) = P(A) \cdot P(B) \Rightarrow P(A \cap B) = (1/4) \cdot (1/3) = 1/12 = 0.08333$$

En el tratamiento de cierta enfermedad se observa que cerca de la tercera parte de los enfermos se recupera. ¿Cuál es la probabilidad de que cuando menos dos de tres enfermos se recuperen. (1 = recuperación y 0 = muerte).

EJERCICIOS

INSTRUCCIONES: Lee con atención los siguientes enunciados y contesta lo que se te pide.

1. Se lanza un dado, si A es el evento "aparece un número menor que tres" y B es el evento, "aparece un número mayor que tres", ¿cuál es la probabilidad de que ocurra por lo menos uno de los eventos?
2. Se lanza un dado, ¿cuál es la probabilidad de que aparezca un número par o un número non?
3. Se lanza un dado, ¿cuál es la probabilidad de que aparezca un número primo o el número uno?
4. Las edades de los integrantes de un equipo de ocho jóvenes y cuatro mujercitas son: cinco jóvenes tienen menos de 16 años y tres tienen 18 años; dos mujercitas tienen menos de 15 años y dos tienen más de 18 años. Si se elige al azar una persona, ¿cuál es la probabilidad de que la persona elegida sea mujer o una persona menor de 16 años?

5. En el grupo 5X del Colegio de Bachilleres, 50 alumnos aprueban Estadística, 25 alumnos aprueban Contabilidad y 10 aprueban ambas asignaturas; si se elige a un alumno al azar, ¿cuál es la probabilidad de que haya aprobado Estadística o Contabilidad?

6. Se lanza un dado, si el número que resulta es impar, ¿cuál es la probabilidad de que sea primo?

7. Se lanza un dado, si el número que resulta es par, ¿cuál es la probabilidad de que sea primo?

8. La probabilidad de que un alumno repruebe Estadística es 18%, de que repruebe Costos es 16% y de que repruebe ambas asignaturas es 4%. Si se elige al azar un alumno y éste reprobó Costos, ¿cuál es la probabilidad de que haya reprobado también Estadística?

9. Al realizar una encuesta de hábitos alimenticios en una colonia del DF, se encontró que el 16% de las personas toman leche en sus comidas, el 80% toma refresco embotellado y el 4% toma leche y refresco embotellado. Si se elige a una persona al azar de estas colonias y ésta toma leche en sus comidas, ¿cuál es la probabilidad de que también tome refresco embotellado?
10. En la caja A se tienen quince artículos, de los cuales seis son defectuosos y en la caja B se tienen trece artículos de los cuales cinco son defectuosos. Si se extrae al azar un artículo de cada caja. ¿Cuál es la probabilidad de que ambos artículos no sean defectuosos?
11. En la papelería “EMY” hay dos urnas; la urna A contiene 18 lápices, de los cuales 7 son defectuosos y en la urna B hay 22 lápices, de los cuales 9 son defectuosos. Si se extrae al azar un lápiz de cada urna, ¿cuál es la probabilidad de que ambos lápices no sean defectuosos?
12. Se lanzan dos dados, si A es el evento, “el primer dado es par” y B es el Evento, “el segundo dado es dos o tres”, ¿cuál es la probabilidad de que ocurra A y B?

TABLA DE COMPROBACIÓN

Número de pregunta	Respuesta correcta
1	$(P \cup B) = 0.8333$
2	$(P \cup B) = 1$
3	$(P \cup B) = 0.6666$
4	$(P \cup B) = 0.75$
5	$(P \cup B) = 0.6923$
6	$P (B/A) = 0.6666$
7	$P (B/A) = 0.3333$
8	$P (E/C) = 0.25$
9	$P (R/L) = 0.25$
10	$P(A \cap B) = 0.3692$
11	$P(A \cap B) = 0.36111$
12	$P(A \cap B) = 0.1666$
Sugerencias	
<p>Recuerda que cuando los eventos son mutuamente excluyentes, utilizamos la fórmula: $(P \cup B) = P(A) + P(B)$, ejercicios 1, 2 y 3.</p> <p>Recuerda que cuando los eventos son no mutuamente excluyentes, utilizamos la fórmula: $(P \cup B) = P(A) + P(B) - P(A \cap B)$, ejercicios 4 y 5.</p> <p>Recuerda que la probabilidad condicional de cualquier evento es la probabilidad de que este evento ocurra con la condición de que el otro evento haya ocurrido, entonces: $P (B/A) = \frac{P(A \cap B)}{P(A)}$ = Se utiliza para los ejercicios 6, 7, 8 y 9.</p> <p>Recuerda que; sean A y B dos eventos independientes, entonces: $P(A \cap B) = P (A) P (B)$ se utiliza para los ejercicios 10, 11 y 12</p>	

5.4 CÁLCULO DE PROBABILIDADES: PROCEDIMIENTOS ELEMENTALES DE CONTEO

APRENDIZAJES

- Calcular el número de permutaciones sin repetición en un arreglo ordenado de datos.
- Calcular el número de permutaciones con repetición en un arreglo ordenado de datos.
- Calcular el número de combinaciones en un arreglo de datos.

PERMUTACIONES

Una permutación es un **arreglo** de todos o parte de los objetos dentro de un conjunto de objetos de **un orden definido**. El número total de permutaciones de un conjunto de objetos depende del número de objetos, tomados a la vez para cada permutación; es decir, puede ser, todos a la vez o parte de ellos.

PERMUTACIONES SIN REPETICIÓN

Las Permutaciones de diferentes objetos tomados todos a la vez, se calcula con la siguiente expresión:

$${}_n P_n = n (n - 1) (n - 2) \dots (3) (2)(1) = n!$$

donde: n : es el número de objetos, tomados a la vez para cada permutación.

${}_n P_n$: es el número total de permutaciones de n objetos, tomados n a la vez.

Las permutaciones de n objetos diferentes, tomados parte de ellos a la vez, se calcula con la siguiente expresión:

$${}_n P_r = n(n - 1)(n - 2)(n - 3) \dots (n - r + 1) = \frac{n!}{(n - r)!}$$

donde r : es el número de objetos, tomados a la vez para cada permutación.

${}_n P_r$: es el número total de permutaciones de n objetos, tomados r a la vez.

PERMUTACIONES CON REPETICIÓN

Con frecuencia se desea saber el número de permutaciones de objetos, de los cuales algunos son iguales. Para referirse al número de permutaciones con repeticiones de n objetos tomados de r en r , se utiliza la siguiente expresión:

$$n^r$$

Si se tienen n objetos de los cuales i son iguales, j son iguales y k son iguales, entonces el número de permutaciones se calcula con la siguiente expresión:

$${}_n P_{ijk} = \frac{n!}{i! j! k!}$$

COMBINACIONES

Una combinación es un arreglo de todos o parte de los objetos de un conjunto **sin considerar el orden** de éstos. El número total de combinaciones posibles de un conjunto de datos tomados todos a la vez es **uno**; por ejemplo: los arreglos posibles de las letras (a, b) son: ab y ba, ya que el orden no es considerado, por lo que el arreglo ab es el mismo que ba, por lo que hay sólo una combinación, entonces para calcular las combinaciones totales de n objetos tomados todos o parte de ellos a la vez, utilizaremos la siguiente expresión:

$${}_n C_r = C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)}$$

donde n : es el número total de objetos de un conjunto de datos.

r es el número de objetos, tomados a la vez para cada combinación.

${}_n C_r$ es el número total de combinaciones de n objetos, tomados r a la vez.

APLICACIÓN DEL CONOCIMIENTO.

Calcula el número de permutaciones del evento (a, b, c), tomados todos a la vez.

Como son todos a la vez, entonces: $n = 3$; por lo tanto;

$${}_n P_n = n(n-1)(n-2)\dots(3)(2)(1) = n!$$

$${}_n P_n = n! \Rightarrow {}_3 P_3 = 3! = 6$$

Calcula el número de permutaciones del evento (1, 3, 5, 7), tomados todos a la vez.

Calcula el número total de permutaciones del evento (a, b, c, d), tomados tres a la vez.

Si aplicamos la expresión anterior, debemos conocer el valor de n y de r ; entonces, del enunciado tenemos que:

$$n = 4 \text{ y } r = 3 \Rightarrow {}_4 P_3 = \frac{4!}{(4-3)!} = (4)(3)(2) = 24$$

Halla el total de permutaciones del evento (a, b, c, d), tomados dos a la vez.

¿Cuántos números de cuatro cifras se pueden formar con los elementos del evento (6, 9)?

Del enunciado observamos que $n = 2$ y $r = 4$, entonces aplicamos la expresión

$$n^r = 2^4 = 16$$

¿Cuántas placas de automóvil existen que consten de dos letras y tres números, si la primera letra es A y la segunda letra puede ser de la A a la F?

¿Cuántas permutaciones se pueden formar con los elementos: A, A, A, B, B, B, B, C y C?

Del enunciado observamos que: $n = 9$, $i = 3$, $j = 4$ y $k = 2$, entonces:

$${}_9 P_{3,4,2} = \frac{9!}{3!4!2!} = 1260$$

Los elementos A, B y C pertenecen a un evento, si $C = A$, ¿cuántas permutaciones existen?

¿Cuántas juntas directivas de cinco personas se pueden formar con doce miembros de una organización?

Del enunciado vemos que: $n = 12$ y $r = 5$, entonces:

$${}_{12}C_5 = \frac{12!}{5!(12-5)!} = 792$$

Un alumno del Plantel 4 del Colegio de Bachilleres tiene que contestar diez de doce preguntas de un examen de estadística, ¿de cuántas maneras puede contestar estas preguntas?

JERCICIOS

INSTRUCCIONES: Lee con atención los siguientes enunciados y contesta lo que se te pide.

1. ¿De cuántas formas se pueden sentar cinco alumnos en una banca?
2. En un grupo de treinta alumnos, el profesor va a elegir un representante, un suplente y un auxiliar, ¿de cuántas maneras puede hacerlo?
3. ¿Cuántas banderas tricolores se pueden formar con siete colores diferentes?
4. ¿Cuántos números telefónicos de seis cifras existen que inicien con los dígitos 1, 2, 3, 4 ó 6?
5. ¿Cuántas placas para automóvil existen que consten de dos letras y tres cifras, si el alfabeto consta de veintisiete letras?

6. ¿Cuántos números telefónicos de siete dígitos se pueden formar con los dígitos 1, 3, 5, 7 y 9?

7. Un estudiante de bachillerato debe contestar diez de doce preguntas de un examen de matemáticas, ¿cuántas maneras tiene para hacerlo, si tiene que contestar siete de las nueve primeras preguntas?

8. De un grupo de ocho personas, se va a elegir una comisión de cuatro personas, ¿de cuántas maneras se puede elegir la comisión?

9. ¿Cuál es el resultado de calcular ${}_{100}C_{98}$?

TABLA DE COMPROBACIÓN

Número de pregunta	Respuesta correcta
1	${}_5P_5 = 5! = 120$
2	${}_{30}P_3 = 24360$
3	${}_7P_3 = 210$
4	Existen cinco dígitos acompañados por 10^5 dígitos para formar el núm. telefónico, entonces: $5 (10^5) = 500000 (5 (n^r))$.
5	Existen 27 letras en el alfabeto, si se tienen dos letras, entonces: $27^2 = 729$, seguidas de tres cifras, entonces: $10^3 = 1000$, por lo tanto, existen 729000 placas ($729 (n^r)$).
6	$n^r \Rightarrow 5^7 = 78125$ números telefónicos.
7	${}_5C_3 = 10$ maneras.
8	${}_8C_4 = 70$ maneras.
9	${}_{100}C_{98} = 4950$
Sugerencias	
<p>No pierdas de vista que en las combinaciones no importa el orden de los objetos para llegar a el resultado.</p> <p>La fórmula a aplicar para las combinaciones con repetición es:</p> ${}_n P_{ijk} = \frac{n!}{i! j! k!}$	

EJERCICIOS DE AUTOEVALUACIÓN

Cuentas con sesenta minutos para resolver los siguientes ejercicios.

INSTRUCCIONES: Lee con atención los siguientes reactivos y coloca en el paréntesis de la izquierda la letra de la opción correcta.

1. () ¿Cuál es el nombre que recibe el conjunto de posibles resultados de un experimento?
 - a) Evento.
 - b) Espacio muestral.
 - c) Probabilidad.
 - d) Población.

2. () ¿Qué nombre recibe el número asignado a un evento que mide la creencia de que ese evento puede ocurrir?
 - a) Espacio muestral.
 - b) Población.
 - c) Probabilidad.
 - d) Experimento.

INSTRUCCIONES: Lee con atención los siguientes reactivos y contesta lo que se te pide.

3. La siguiente tabla muestra a un grupo de 300 estudiantes, de los cuales, 130 son mujeres y 110 son hombres, 240 son estudiantes de tiempo completo y 60 (40 mujeres y 20 hombres) son de tiempo parcial.

	Tiempo completo	Tiempo parcial	Totales
Mujeres	130	40	170
Hombres	110	20	130
Totales	240	60	300

- I. ¿Cuál es la probabilidad de que un estudiante elegido al azar sea mujer o de tiempo completo?

- II. ¿Cuál es la probabilidad de que un estudiante elegido al azar, sea de tiempo completo si se sabe que es mujer?

UNIDAD 5

4. En dos contenedores A y B, hay 5000 pantalones de los cuales el 10% están defectuosos y 4000 pantalones de los cuales el 15% están defectuosos, respectivamente. Si se extrae al azar un pantalón de cada contenedor, ¿cuál es la probabilidad de que ambos pantalones sean defectuosos?

5. ¿De cuántas maneras se pueden acomodar ocho libros diferentes en un estante?

6. Se disponen de cuatro lienzos de diferentes colores, ¿cuántas banderas bicolors se pueden formar?

7. El director de una empresa desea seleccionar a tres personas de un total de 123, para realizar cálculos aritméticos, ¿de cuántas maneras puede seleccionarlos?

8. Se lanzan tres monedas, ¿cuál es la probabilidad de que dos de las monedas caigan soles?

CLAVE DE RESPUESTAS.

Número de pregunta.	Respuesta correcta.
1	b
2	c
3	I. $P(M \cup Tc) = 0.93333$ II. $P\left(\frac{Tc}{M}\right) = 0.7647$
4	$P(A \cap B) = 0.015$
5	40320
6	12
7	1815726
8	3/8

BIBLIOGRAFÍA

1. JOHNSON, ROBERT. *Estadística Elemental*. Iberoamérica, México, 1990.
2. PORTUS G., LINCOYÁN. *Curso Práctico de Estadística*. McGraw-Hill, México, 1994.
3. PORTILLA C., ENRIQUE. *Estadística, Primer Curso*. Interamericana, México, 1987.
4. SÁNCHEZ C., OCTAVIO. *Probabilidad y Estadística*. McGraw-Hill, México, 1996.
5. NAIMAN, A. ROSENFELD, R. y ZIRKEL, G. *Introducción a la Estadística*. McGraw-Hill, México, 1987.

**SUGERENCIAS PARA PRESENTAR
EXÁMENES DE RECUPERACIÓN
O ACREDITACIÓN ESPECIAL**

Para evitar cualquier contratiempo al presentar el examen de recuperación o acreditación especial debes considerar las siguientes recomendaciones:

Organización:

- Preséntate al menos con 10 minutos de anticipación al salón indicado. Debes presentarle al profesor aplicador, esta Guía resuelta.
- Lleva el comprobante de inscripción al examen y tu credencial actualizada.
- Lleva dos lápices del No. 2 o 2 ½.
- No olvides una goma que no manche.

Durante el examen:

- Lee con atención tanto las instrucciones como las preguntas y si tienes alguna duda consúltala con el aplicador.
- Contesta primero las preguntas que te parezcan “fáciles” y después concentra toda tu atención en las “difíciles”.
- Si te solicitan explicar o desarrollar algún tema, identifica las ideas principales que quieras exponer y escríbelas de la manera más concreta y clara que puedas, evita el planteamiento de ideas innecesarias.
- Escribe tus respuestas con letra clara, legible y sin faltas de ortografía.
- Al terminar de contestar el examen, revísalo nuevamente para asegurarte que todas las preguntas estén contestadas.
- Centra tu atención en el examen, no trates de copiar, recuerda que el compañero de junto puede estar equivocado.

La Guía para presentar exámenes de
Recuperación o Acreditación Especial de
Estadística Descriptiva e Inferencial I
(versión preliminar)
se terminó de reimprimir en el mes de octubre de 2006
en los talleres del Colegio de Bachilleres.
Prolongación Rancho Vista Hermosa 105
Col. Ex Hacienda Coapa.
México, D.F. 04920

El tiraje fue de 271 ejemplares
más sobrantes para reposición