ESTADISTICA DESCRIPTIVA Y PROBABILIDAD

JAVIER GUILLÉN ANGUIANO
LOURDES ROMERO MIRANDA
FAUSTO GÓMEZ LORENCE

2018

PRESENTACIÓN

Estimado alumno:

Cuando a través de documentos especializados, Internet, periódicos, revistas y otros medios se reportan resultados de una investigación, suele utilizarse terminología propia de la estadística. Para comportarse como un *consumidor inteligente* de este tipo de información, cualquier persona con una preparación media tiene necesidad de poseer ciertos conocimientos elementales.

En el Cuaderno de Trabajo que ahora tienes en tus manos se tuvo presente lo anterior, por ello su contenido se relaciona con situaciones que pudieras encontrar al realizar un trabajo de investigación o en la vida cotidiana.

En 17 sesiones abordaremos la *Estadística Descriptiva* y algunos conceptos básicos de *Correlación* y *Probabilidad*.

Para estudiar este Cuaderno, debes:

- **Estar en alerta constante para comprender todo y contestar sin equivocarte.**
- **Tener a la mano constantemente un lápiz afilado, goma y calculadora; además, tener disponibles: computadora con impresora, escuadras, transportador y compás.**
- **Resolver todo el Cuaderno de Trabajo.**
- **Repasar diariamente y evitar interrupciones prolongadas al estudiar.**

La interacción permanente con este texto impulsará el desarrollo de tus capacidades de comprensión, análisis, crítica, aplicación e interpretación de la información. Los beneficios que puedas obtener al estudiarlo dependerán del trabajo sistemático y disciplinado que hagas con él.

De hacerlo así, al final podrás sentirte satisfecho del aprendizaje logrado.

Atentamente

Los autores

ÍNDICE

1ª SESIÓN
MÉTODO DE LA ESTADÍSTICA DESCRIPTIVA Y RECOPILACIÓN DE DATOS

COMPETENCIAS POR ADQUIRIR

- **Distingue los diferentes tipos de variables estadísticas.**
- **Diseña un procedimiento de selección aleatoria que le permita obtener datos de una población, con el fin de describir el comportamiento de alguna característica.**
- **Explica en qué consiste el método de la Estadística Descriptiva.**
- **Explica el propósito de la Estadística Inferencial y de la Teoría del Muestreo.**
- **Explica los conceptos de población, muestra, variable, dato, estadístico y parámetro.**
- **Identifica variables en una población o muestra dada.**
- **Explica lo que se puede averiguar mediante una encuesta.**
- **Menciona al menos tres recomendaciones para hacer una entrevista.**

A partir de lo anterior:

- Aprecia la importancia del muestreo.
- Reconoce que los datos estadísticos se obtienen por levantamiento o por experimentación.
- Concluye que el azar es causa de la variabilidad en los datos estadísticos.

MATERIAL NECESARIO: LÁPIZ, SACAPUNTAS, GOMA Y CALCULADORA.

LA ESTADÍSTICA DESCRIPTIVA Y SU MÉTODO

En nuestros días, la Estadística se sigue desarrollando para describir y relacionar entre sí diversos tipos de información: económica, política, social, psicológica, biológica, física etc.; es importante por ello familiarizarse con la manera de clasificarla y con el significado de ciertas palabras comúnmente utilizadas.

PARA EXPLICAR EL UNIVERSO OBJETIVA Y RACIONALMENTE

Gracias al trabajo de notables investigadores de diferentes épocas y disciplinas, la ciencia de nuestros días cuenta con una referencia general para obtener nuevos conocimientos: el **método científico**.

El método científico supone: definir el problema por resolver o analizar, formular soluciones tentativas (hipótesis) al problema, recolectar datos, analizarlos y en consecuencia confirmar o rechazar las hipótesis.

A la investigación que se hace empleando este método se le califica como científica y ha permitido descubrir hechos, relaciones o leyes en múltiples campos del conocimiento humano.

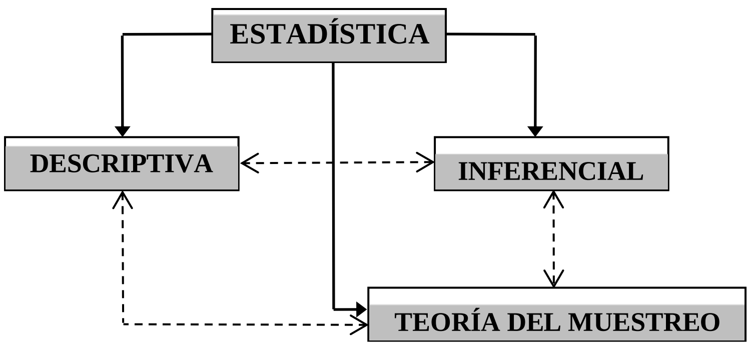
La Estadística tiene un gran valor como **auxiliar de las investigaciones científicas**, aunque **por sí misma** constituye un **método válido** para llegar a nuevos conocimientos, emitir juicios racionalmente sustentados y tomar decisiones fundamentadas; en ambas cosas se basa su importancia.

En general, **la estadística estudia la variabilidad de una característica de la población, considerando la homogeneidad o heterogeneidad en los valores observados.**

Por ejemplo, los focos LED que produce por mes una fábrica constituyen una “población” y tienen, entre sus diversas características, la duración en horas. ¿Cómo se puede saber cuál es la duración promedio de los focos en un mes? ¿Qué tanto varía esa duración entre los focos? ¿Qué tan confiable sería el valor de la duración promedio obtenido?

Para esas y otras preguntas la Estadística proporciona respuestas.

La Estadística para su estudio se divide en: Estadística Descriptiva, Teoría del Muestreo y Estadística Inferencial (o Inductiva).



En el estudio de un fenómeno, la **Estadística Descriptiva** comprende la recopilación de datos, su organización, presentación y cálculo de ciertas características numéricas.

De los procesos para seleccionar **muestras representativas** de una población se encarga la **Teoría del Muestreo.**

La Estadística **Inferencial** tiene como propósito hacer **inferencias** (predicciones) sobre determinadas características de una población, a partir del estudio de una muestra representativa.

MÉTODO

El *método* se refiere a la manera de hacer las cosas con un fin determinado.

Para estudiar los datos que se obtienen al estudiar un fenómeno, la Estadística Descriptiva constituye un método que comprende los siguientes pasos secuenciados:

1º Recopilación **2º Organización** **3º Presentación** **4º Análisis de datos**

TERMINOLOGÍA

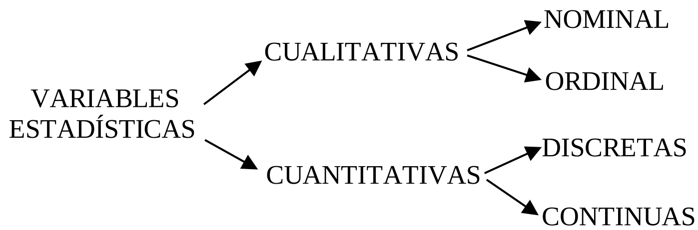
La primera etapa de la Estadística Descriptiva, la recopilación de datos, muchas veces se dificulta por el gran tamaño de la **población**.

Por **población** o **universo** se entiende al **conjunto** de elementos (objetos, individuos, conceptos o propiedades) con una o más características comunes.

A una característica **común,** pero que puede **cambiar su valor de un elemento a otro** en una población, se le denomina **variable;** por ejemplo, al estudiar los ríos de América la *longitud* es una característica común porque la tienen todos, pero **varía** de uno a otro.

Los valores que toma la variable en el fenómeno bajo estudio son los **datos**. Para el acopio de datos se emplean **escalas**, sean numéricas o no numéricas.

Esta es una manera de clasificar las variables en Estadística:



- Variables Cualitativas:
Los valores son “categorías” y ningún “valor” se puede decir que sea mayor o menor que otro.

- Ejemplos: para la variable *complexión física* los valores podrían ser *pequeña, mediana* y *grande*, para la variable *región de México en la que vive* los valores podrían ser *noroeste, noreste, occidente, oriente, centronorte, centrosur, suroeste* y *sureste*.

La variable **cualitativa nominal** es dicotómica (sí-no, falso-verdadero…) y sus categorías no son ordenables, mientras que la variable **cualitativa ordinal** sí tiene valores que se pueden ordenar según cierto criterio
Ejemplos: para la variable nominal *estudios de licenciatura* los valores podrían ser *tiene* y *no tiene*;
Para la variable ordinal *clase social*, los valores podrían ser *baja*, *media* y *alta*, para la variable *opinión sobre una propuesta política*, los valores podrían ser *muy en contra, más bien en contra, indiferente, más bien a favor* y *muy a favor*.
Para la variable cualitativa no podemos aplicar métodos aplicables a variables cuantitativas. El tratamiento estadístico que se hace de este tipo de variable es, en unas ocasiones, como variable cualitativa y en otras, como variable cuantitativa “cambiando” las etiquetas por números y “transformando” la variable en cuantitativa.
Por ejemplo, no se puede hablar del “promedio” del *estado civil*; o del *partido político de su preferencia*; a la inversa, la variable cuantitativa sí la podemos “transformar” en escala ordinal, o en cualitativa. Por ejemplo un conjunto de *edades en años* lo podemos agrupar, de acuerdo con algún criterio, en estas categorías de edad: *niños, jóvenes, adultos y ancianos*.

- Cuantitativas o numéricas:
Los valores de la variable cuantitativa o numérica son “números”, cada valor posible es menor, igual o mayor que otro valor.
Ejemplos de variables cuantitativas: edad, ingreso anual, calificación en un examen, número de años de educación, kilómetros de distancia entre dos poblaciones...
Hay números que son “etiquetas”, por ejemplo: el código postal; el número telefónico; la clave de una asignatura.

La variable **cuantitativa** se dice que es **discreta** cuando **no existen** valores posibles entre dos valores consecutivos de la variable, se trata de números enteros y son el resultado de **contar** (número de habitantes del pueblo, número de autopistas en Nayarit, etc.), en cambio, es continua cuando el número de valores posibles entre dos valores de la variable es infinito (son el resultado de **medir** temperatura, altura, peso, tamaño del árbol, edad, etc.).

ACOPIO DE DATOS

Cuando estudiar toda la población es un problema muy difícil (tal vez por insuficiencia de tiempo y de recursos), conviene estudiar sólo una porción de ella, escogiéndola de tal manera que la represente en cuanto a las características de interés para la investigación.

Al seleccionar una parte o subconjunto de una población se tiene una **muestra.**

Por ejemplo, si la población es el conjunto de automóviles de servicio particular de la ciudad de Florencia, en Italia, una muestra puede ser (subraya la opción correcta):

- Los automóviles de servicio público de Florencia.
- Los automóviles particulares modelo 2004 de Florencia.

Es evidente: entre mayor sea la muestra, más válidas para la población serán las conclusiones que resulten de su estudio.

Una muestra con las características de la población es una **muestra representativa**. Por lo tanto, cuando sea necesario tomar una muestra, es importante que sea **representativa de la población**, porque así las conclusiones obtenidas para la muestra se podrán aplicar con alto grado de validez a toda la población, acción que se conoce como **inferencia estadística**.

El modo de obtener una muestra representativa de la población es usar técnicas aleatorias.

Una muestra es **aleatoria** cuando sus elementos se seleccionan al azar, es decir, mediante un procedimiento que no señale con anticipación los elementos que serán escogidos. Un método aleatorio de seleccionar los datos de una población hace necesariamente que se refleje la **variabilidad** de los mismos.

Ejemplo: en un noticiero de televisión se quiere conocer el sentir de los ciudadanos acerca de la existencia de hambruna en México, para ello se invita a los televidentes a que envíen un correo electrónico opinando al respecto durante la duración del noticiero, para ello deben seleccionar alguna de las siguientes opciones:

Es muy alta ( ) *Es alta* ( ) *Es baja* ( ) No existe ( )

Al finalizar el noticiero, el locutor anuncia: ya tenemos el resultado de nuestra encuesta, *87% de la población mexicana opina que la hambruna en México no existe*…

¿Aceptarías como válida la conclusión del locutor? _____.

¿Crees que estuvo involucrada una muestra representativa? _____. ¿Por qué? ___________
____.

Señala otro factor que hace poco creíble la conclusión del locutor: ______________________

____

Otro ejemplo: de los 20 000 científicos, miembros del Sistema Nacional de Investigadores que había en 2014, en México, se seleccionó a 25 que habían hecho publicaciones recientes sobre astronomía, ¿la muestra fue aleatoria? ____ .

Si en una lista se hubiera numerado a los 7 200 científicos y elegido a los 25 cuyos números fueran obtenidos mediante un sorteo, ¿hubiera sido una muestra aleatoria? ____ .

Lo contrario de una muestra aleatoria es una muestra **viciada** o **no aleatoria**.

Al hacer un tratamiento matemático de los datos de una muestra se calculan *medidas* como su media aritmética y desviación estándar, a las cuales se les denomina **estadísticos** o **estadígrafos**; cuando esas mismas medidas se calculan incluyendo a todos los elementos de la población se les denomina **parámetros**.

RECOPILACIÓN DE DATOS

Después de precisar el problema y de diseñar el plan a seguir en la investigación estadística de un fenómeno, la *recopilación de datos* es el primer paso.

Cuando se decide iniciar el acopio de datos debe tenerse **claridad** sobre la **manera** en que habrá de hacerse y tener bien definidos los atributos o variables de interés en el fenómeno bajo estudio.

¿De qué manera se pueden recolectar los datos?

Los datos estadísticos se obtienen por levantamiento o por experimentación. Si desearas averiguar la velocidad promedio a la que viaja la pelota cuando un tenista determinado hace su primer saque, tendrías que buscar una forma para hacerte de los datos por "*levantamiento*", tomando en cuenta que el partido **no se puede repetir igual**; en cambio, si te interesara saber la velocidad a la que llega al piso una pelota nueva de tenis cuando se suelta desde una altura de 3 m, podrías *experimentar* (**repetir** el hecho) muchas veces.

Otro ejemplo de obtención de datos por levantamiento sería: __________________________.

Otro ejemplo de obtención de datos por experimentación sería: ______________________.

Los datos que se manejan por medio de la
estadística se obtienen al:
contar
o
medir

características de poblaciones, sin que necesariamente se trate de personas. . .

Una vez que se ha **precisado** el problema de interés y conformado el **plan** para realizar una investigación estadística, el primer paso de su ejecución es el acopio de información.

Pueden ser dos las maneras de recopilar datos:

I.- FORMA DIRECTA:

A. La información se obtiene *sin intermediarios* entre la persona y el objeto de estudio; se utilizan cuestionarios, entrevistas, programas de televisión, exámenes, reportes escritos, investigaciones ya realizadas, etcétera.

B. Por medio de experimentos científicos, medidas o conteos en condiciones controladas.

La observación directa es la forma **más confiable** en la recopilación de datos. Antes de aplicarla el observador debe tener claras sus necesidades y saber lo que busca.

Una manera de averiguar hechos, testimonios, actitudes y opiniones es efectuar una **encuesta**; ésta se puede hacer por medios orales (en cuyo caso se trata de una **entrevista**), o por medios escritos (entonces se trata de un **cuestionario**).

Las encuestas se clasifican de esta manera:

- **De conocimientos** o **hechos**, para averiguar lo que las personas **saben o hacen.**
- **De opinión**, para averiguar lo que las personas **piensan**
- **De actitud**, para averiguar lo que las personas **sienten**

En una **entrevista** el contacto puede ser personal o por medios electrónicos; en ambos casos se logra el acopio de testimonios orales. Al preparar una entrevista es importante familiarizarse con el tema a tratar y diseñar las preguntas por formular; al realizar la entrevista es recomendable hacer gala de cortesía para promover y facilitar la conversación, ser puntual en las citas, transcribir con fidelidad lo dicho por el entrevistado y ponerle toda la atención posible.

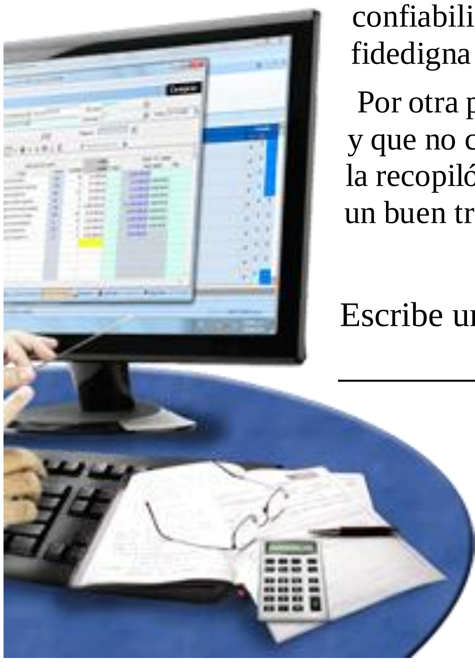
Los ítems del **cuestionario** deben prepararse teniendo en mente las posibles respuestas y previendo los cuadros y gráficas que se construirán y los valores numéricos que se obtendrán. Se debe aplicar en condiciones que permitan la concentración del encuestado y contestarse sin hacer pausas prolongadas.

Al preparar cuestionarios o entrevistas es recomendable considerar **únicamente preguntas que arrojen datos de interés para la investigación,** al aplicarlos tener en cuenta el grado de cultura de las personas que proporcionarán la información y evitar al máximo la influencia de los puntos de vista del aplicador. Sobre esa base, las preguntas formuladas en un lenguaje apropiado tendrán la interpretación debida por quien las conteste.

Si se trata de realizar **experimentos científicos**, lo más apropiado será guiarse por la orientación que proporciona el **Método Científico Experimental**.

II.- FORMA INDIRECTA:

Sucede cuando los datos se obtienen a través de registros, como archivos, bases de datos, Internet, almanaques, libros, tesis y otros medios escritos o electrónicos. En este caso es importante valorar la confiabilidad de la fuente de información, ya que dos fuentes de apariencia fidedigna pueden mostrar discrepancias entre sí.



Por otra parte, es indispensable que la información provenga de fuentes **confiables** y que no contenga datos inventados o agregados arbitrariamente por la persona que la recopiló. El cuidado de estos factores permitirá contar con cimientos sólidos para un buen trabajo de investigación.

Escribe una investigación estadística que te gustaría hacer:

____

____

¿Por qué sería importante esa investigación?

____

____

En tu investigación, menciona las variables bajo estudio.

____

¿De qué tipo son tus variables?

____

¿Tomarías muestra o abarcarías toda la población?

____

De ser el caso, ¿cómo tomarías la muestra?

____

¿Cómo recopilarías los datos? ¿En forma directa o indirecta? ¿Qué prepararías para recopilar los datos?

____.

____.

FICHA DE TRABAJO I.1

Contesta con claridad.

1. Explica el significado estadístico de:
Variable ____
Población ____
Muestra ____

2. Marca con *L* si los datos relacionados se obtienen por levantamiento y con *E* si se obtienen por experimentación.

El porcentaje de los mexicanos que usan tarjeta de crédito (   ).
El nivel de glucosa en la sangre de un mesero mayor de 18 años (   ).
Los gramos de pasiflora en té que requiere Pedro Bautista para quedar dormido (   ).
El tiempo de vida del gusano barrenador (*Cochliomyia hominivorax)* que ataca al ganado vacuno (   ).

3. El gobierno municipal de Nochixtlán, Oax., quiere investigar la superficie de los predios donde se asientan las viviendas de la entidad.

a. ¿Cuáles son las etapas del método de la Estadística Descriptiva que deben seguirse?

_________________________________   _____________________________________
_________________________________   _____________________________________

b. ¿Para qué se debe utilizar la Teoría del Muestreo?

____

c. ¿Cómo interviene la Estadística Inferencial en la investigación?

____

d. Indica:

- La población ____
- Una muestra ____
- La variable bajo estudio ____
- Un ejemplo de dato____

4. Identifica con *P* la población y con *M* las muestras:

a) Las palomas que anidan en el zócalo de Coyoacán, en la Cd. de México. (   )
b) Las palomas que habitan en la Cd. de México. (   )
c) Las palomas de la Cd. de México que viven en campanarios de iglesias. (   )

5. Menciona dos variables presentes en las siguientes poblaciones:

a) Los campos de cultivo de Jalisco. ____
b) Las computadoras producidas por la IBM. __________________________________

c) Las mediciones diarias durante un año de la contaminación ambiental en la ciudad de México.
____

6. Escribe “D” si la variable es discreta y “C” si es continua.

Número de habitantes de la Ciudad de México (   ).

Número de enfermedades que puede producir el sobrepeso en una persona (   ).

Presión de las llantas de un automóvil (   ).

Calificación que otorga la UNAM a sus estudiantes (   ).

7. Escribe “N” si la variable es nominal y “O” si es ordinal.

Grado de complacencia de los clientes de un hotel (   ).

Estado de ánimo (   ).

Estado civil (   ).

8. Supón que trabajas en una Notaría y te asignan averiguar la proporción de propietarios de vivienda que no cuentan con testamento en la Delegación.

a. ¿Intentarías trabajar con toda la población o seleccionarías una muestra?

b. Diseña, e indica con detalle, un procedimiento de selección aleatoria que te permita obtener los datos de la muestra.

a. ____________________________ **b.** ____________________________________
____
____
____
____

9. Da cuatro ejemplos de recopilación de datos acerca de las causas del cáncer:

a. Dos ejemplos de forma directa. ______________________________________
____
____
____

b. Dos ejemplos de forma indirecta ______________________________________
____
____

10. ¿Cuál es la clasificación de las encuestas?
____
____

11. Señala los medios por los que se puede efectuar una encuesta ______________________
____

12. Explica lo que se puede averiguar mediante una encuesta __________________________
____

13. Menciona tres recomendaciones válidas para diseñar entrevistas y para elaborar cuestionarios.
____
____
____

14. Menciona cuatro cuidados que deben tenerse al efectuar una entrevista.
____
____
____
____

15. Menciona cuatro cuidados que deben tenerse al formular un cuestionario.
____
____
____

RESUMEN

Al estudiar un fenómeno las etapas del método de la Estadística Descriptiva son la recopilación, organización, presentación y cálculo de medidas numéricas.
Una muestra es un subconjunto de una población. Se estudian cuando es imposible o inconveniente estudiar toda la población, procurando que sean representativas de toda la población, para ello se seleccionan en forma aleatoria.
Los elementos de una población tienen características cuyo valor varía de un elemento a otro, los valores de las variables son los datos.
Un muestreo correcto proporcionará una muestra representativa de la población (de esto se encarga la Teoría del Muestreo). El estudio de una muestra representativa conduce a establecer afirmaciones válidas para toda la población (Estadística Inferencial).
Los estadísticos son medidas numéricas que caracterizan a una muestra, lo parámetros a la población.
La recopilación de datos es directa cuando la información se obtiene sin utilizar intermediarios entre la persona y el objeto de estudio. Son medios para un acopio directo: los cuestionarios, entrevistas, experimentos científicos, programas de televisión, exámenes, reportes sobre el tiempo, etcétera.
Las encuestas pueden ser de conocimientos o hechos, de opiniones y de actitudes, para averiguar lo que las personas saben, piensan o sienten.
La encuesta para el acopio de datos se puede hacer mediante una entrevista o un cuestionario. La primera es una forma personal de reunir testimonios orales y el segundo, una forma escrita de obtener información; en las entrevistas y cuestionarios se deben considerar sólo preguntas que arrojen datos de interés para la investigación; al aplicarlos se recomienda tener en cuenta la cultura de los destinatarios y evitar la influencia del aplicador.
En investigaciones de las ciencias naturales conviene basarse en el Método Científico Experimental.
La recopilación de datos es indirecta cuando se hace por medio de registros escritos o electrónicos de diferente tipo.

En este resumen falta mencionar una clasificación importante, ¿cuáles son las clases? ___

____

____

____

____

____

____

2ª SESIÓN
ORGANIZACIÓN DE DATOS

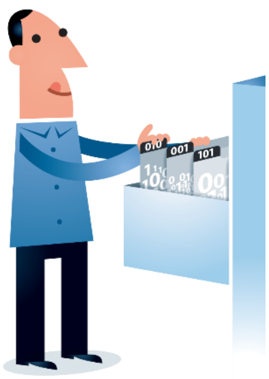
COMPETENCIAS POR ADQUIRIR

- **Identifica las peculiaridades de las tablas de frecuencias y de las tablas de frecuencias por intervalos.**
- **Calcula correctamente frecuencias relativas y frecuencias relativas porcentuales.**

En consecuencia:

- Valora la importancia de la recopilación y representación de datos en la investigación estadística.

MATERIAL NECESARIO: LÁPIZ, SACAPUNTAS Y GOMA.



ORGANIZACIÓN DE DATOS

Después de recopilar la información (o de manera paralela) se debe dar el siguiente paso del método estadístico*: organizar los datos*. Cuando se trata de un conjunto numeroso de valores conviene sintetizarlos en tablas donde podamos advertir mejor su comportamiento.

DESPUÉS DE COLECTAR. . . A ORGANIZAR

La Estadística es una disciplina de muchos datos y uno de sus quehaceres consiste en organizarlos con el fin de facilitar su manejo, vislumbrar mejor su comportamiento como conjunto y, por ende, el de las variables bajo estudio.

El análisis del comportamiento de los datos se facilita cuando están organizados en una Tabla de Frecuencias. Ésta puede ser de dos tipos: el primero resulta de registrar simplemente el número de veces que se repite cada dato, el otro se obtiene de una construcción un poco más elaborada al agruparlos en **clases**.

Al hacer una observación o al efectuar un experimento se obtienen valores de las variables **en juego, es decir,** datos**.**

¿Qué hacer con los datos para analizarlos mejor?

Lo primero es… **organizarlos**.

Ejemplo: Los tiempos redondeados a horas que hicieron 30 tráileres al trasladar mercancía de Veracruz, Ver., a Saltillo, Coah., fueron:

16	18	16	15	16	19	14	15	16	16
17	15	15	18	15	16	15	18	13	15
13	16	11	15	16	15	17	15	16	19

¿Están organizados de alguna manera esos datos?

Si se acomodaran de esta manera sería mejor porque así se puede visualizar qué tanto se repite cada dato:

11	14	15	15	15	16	16	16	17	18
13	15	15	15	15	16	16	16	18	19
13	15	15	15	16	16	16	17	18	19

Con el propósito de tener una idea más clara de sus características, a un conjunto de datos generalmente conviene organizarlo en cuadros o tablas.

Si el conjunto anterior se organiza en una **Tabla de Frecuencias** queda en esta forma:

Tiempo de traslado en horas	**Frecuencia**
11	1
13	2
14	1
15	10
16	9
17	2
18	3
19	2
	Total:

Tabla 2.1

En la primera columna se consignan los datos diferentes que tiene el conjunto, ordenados de menor a mayor (pudieran también ordenarse de mayor a menor), y en la segunda columna se indica la **frecuencia**, es decir, **el número de veces que se repitió** cada valor del tiempo.

Al organizar los datos en la tabla anterior tenemos mejor visión de algunas de sus características.

Por ejemplo, en la tabla 2.1 se observa que sólo 3 de los 30 (o sea, el 10%) de los tráileres hicieron tiempos menores de 14 horas, que el tiempo más frecuente fue de 15 horas, etc.

Contesta fijándote en la Tabla de Frecuencias:

¿Cuántos tráileres hicieron más de 14 horas pero menos de 17? ____ .

¿Algún tráiler hizo 12 horas de recorrido? ____ .

Si tuvieras que hacer un plan de traslado de mercancía de Veracruz a Saltillo, ¿qué tiempo considerarías para el mismo? ___________ .

Tiempo de traslado en horas	Frecuencia relativa
11	0.0333
13	
14	
15	
16	0.3000
17	
18	
19	
	Total:

Tabla 2.2

Las frecuencias también se pueden expresar de manera relativa.

En el primer renglón de la tabla 1.2 se advierte que *uno* de los *treinta* recorridos (1/30) duró 11 horas; el cociente (0.0333) es la **frecuencia relativa** de ese dato.

En *nueve* de *treinta* viajes (5° renglón de la tabla 1.1) se hicieron 16 horas, por lo tanto la frecuencia relativa de 16 es $\frac{9}{30} = 0.3000$

Calcula la frecuencia relativa de los demás datos y escríbelos en la tabla 2.2

¿Cuánto suman las frecuencias relativas? ______.

¿Cuánto deberían sumar si se pudieran tomar en cuenta todas las cifras decimales? ___ .

Si la frecuencia relativa se multiplica por cien se obtiene la **frecuencia relativa porcentual**: el 3.33% de los viajes duraron once horas; el 30% duraron 16 horas, etc.

Otro ejemplo: En una competencia de velocidad en 100 metros planos participaron 60 corredores. Los tiempos en segundos y con aproximación hasta décimas logrados por cada uno fueron los que se muestran abajo.

Observa que ahora los datos están aproximados a décimas de segundo.

12.6	12.8	13.1	14.6	12.0	13.1	12.6	10.6	13.5	11.0
10.1	15.2	12.6	11.6	13.1	15.4	10.3	13.7	11.2	13.4
11.7	13.1	13.8	13.4	12.9	13.4	12.9	15.8	12	13.8
14.6	11.8	12.8	12.1	10.4	13.7	14.5	13.0	14.6	12.1
12.1	15.5	10.5	11.1	13.2	15.6	13.9	15.9	11.4	11.9
13.7	13.6	13.8	14.3	12.5	14.8	11.2	10.4	10.7	14.1

Los datos, ¿están organizados de alguna manera? ____ . Al ordenarlos de menor a mayor tienen esta apariencia:

10.1	10.7	11.6	12.1	12.6	13.1	13.4	13.7	14.3	15.2
10.3	11	11.7	12.1	12.8	13.1	13.4	13.8	14.5	15.4
10.4	11.1	11.8	12.1	12.8	13.1	13.5	13.8	14.6	15.5
10.4	11.2	11.9	12.5	12.9	13.1	13.6	13.8	14.6	15.6
10.5	11.2	12	12.6	12.9	13.2	13.7	13.9	14.6	15.8
10.6	11.4	12	12.6	13	13.4	13.7	14.1	14.8	15.9

Cuando los registros anteriores se agrupan en seis "**intervalos reales de clase**" de tamaño 1, calculado este tamaño con un procedimiento que estudiarás más adelante y que consiste, básicamente, en dividir el rango del conjunto (que se obtiene restando el mayor valor menos el menor) entre el número de intervalos deseado.

El primer intervalo real de clase comprende los registros que van de 9.95 a 10.95, la segunda clase los que van de 10.95 a 11.95, etc. (tabla 2.3).

Tiempo logrado (s)	**Frecuencia (No. de Deportistas)**
9.95–10.95	7
10.95–11.95	9
11.95–12.95	13
12.95–13.95	18
13.95–14.95	7
14.95–15.95	6
	Total: ____

Tabla 2.3

Este es un ejemplo de una **Tabla de Frecuencias por Intervalos**; la columna de la izquierda contiene los intervalos donde quedan comprendidos los valores de los datos originales.

En este caso, el significado de la palabra frecuencia es diferente al de la tabla 2.1 del ejemplo anterior, ahora un valor de la frecuencia indica **el número de datos comprendidos en el intervalo respectivo**.

Preguntas:

1. Cada intervalo y el que sigue ¿tienen un límite común? ____ .

2. ¿Podría alguno de los datos originales coincidir con uno de esos límites comunes? ____ .

3. ¿Por qué?

____

____

Con referencia a la tabla 2.3, escribe las cantidades faltantes:

Tiempo logrado (s)	**Frecuencia relativa**
9.95–10.95	
10.95–11.95	
11.95–12.95	
12.95–13.95	
13.95–14.95	0.1167
14.95–15.95	
	Total: ______

Tiempo logrado (s)	**Frecuencia relativa porcentual**
9.95–10.95	
10.95–11.95	
11.95–12.95	
12.95–13.95	
13.95–14.95	11.67
14.95–15.95	
	Total: ______

FICHA DE TRABAJO 2.1

ANALIZA Y CONTESTA SIN EQUIVOCARTE:

1. En la tabla I de abajo se consigna el tiempo (en horas) de inconsciencia de 100 personas después de aplicarles una dosis de pentotal sódico (anestésico intravenoso); en la tabla II, la cantidad de hijos en 70 familias de una zona marginada en Bogotá, Colombia.

Tiempo (s)	No. de personas
6–10	1
11–15	1
16–20	2
21–25	12
26–30	60
31–35	15
36–40	6
41–45	3
	Total:

TABLA I

No. de hijos	No. de familias
1	2
2	4
3	10
4	12
5	18
6	15
7	6
8	3
	Total:

TABLA II

a. ¿Qué nombre recibe la tabla I? ____ .

b. ¿Qué nombre recibe la tabla II? ____ .

c. ¿Cuál es el significado de la frecuencia en la Tabla I? ___________________________
___________________________.

d. ¿Cuál es el significado de la frecuencia en la Tabla II?___________________________
___________________________.

2. Llena los espacios vacíos (los datos corresponden a las tablas I y II):

Tiempo (s)	Frecuencia relativa
6–10	
11–15	
16–20	
21–25	
26–30	
31–35	
36–40	
41–45	
	Total:

No. de hijos	Frecuencia relativa porcentual
1	
2	
3	
4	
5	
6	
7	
8	
	Total:

RESUMEN

Los datos se pueden organizar en alguno de estos tipos de tablas: de frecuencias y de frecuencias por intervalos. Estas tablas permiten visualizar mejor las características de la información y el comportamiento de las variables. Una *tabla de frecuencias* consta de dos columnas: en una se consigna el valor de cada dato y en la otra el número de veces que se repite (frecuencia). Una *tabla de frecuencias por intervalos* también consta de dos columnas: una contiene los intervalos de clase y la otra el número de datos dentro de cada uno de esos intervalos (frecuencia). Las frecuencias también se pueden expresar de manera relativa o relativa porcentual.

¿Qué se puede agregar a este resumen? ____

____

____

3ª SESIÓN
TABLAS DE FRECUENCIAS SIMPLES Y POR INTERVALOS

COMPETENCIAS POR ADQUIRIR

- **Organiza un conjunto de datos dado en una tabla de frecuencias simple.**
- **Organiza un conjunto de datos dado en una tabla de frecuencias por intervalos.**
- **Construye tablas de distribución de frecuencias, incorporando el uso de la computadora, para describir el comportamiento de una variable.**

MATERIAL NECESARIO: LÁPIZ, SACAPUNTAS Y GOMA.

TABLA DE FRECUENCIAS SIMPLE

La organización de los datos es necesaria cuando se quieren apreciar de manera más clara sus rasgos distintivos en conjunto. Si el número de datos es más de 30, una forma de organizarlos consiste en construir una **tabla de frecuencias simple**; otra es mediante una ***tabla de frecuencias por intervalos***.

INTERVALOS ABIERTO Y CERRADO

En la expresión

$2 < x < 6$ **(se lee: *x* es mayor que dos, pero menor que seis)**

Los valores de "x" están entre dos y seis, pero no pueden ser dos ni seis.

Por no incluir sus extremos, se dice que es un **intervalo abierto** y se denota como (2, 6), o así:

$$\underset{2}{(}\!\!\longrightarrow\!\!\underset{6}{)}$$

Los paréntesis redondos indican que **no se incluyen los valores extremos**.

En la expresión

$3 \leq x \leq 16$ **(se lee: *x* es mayor o igual que tres, pero menor o igual que dieciséis), donde *x* es el número de productos comprados**).

Los valores de "x" están entre tres y dieciséis, pero pueden ser también tres o dieciséis.

Por incluir sus extremos, se dice que es un **intervalo cerrado** y se denota como [3, 16], o así:

$$\underset{3}{[}\!\!\longrightarrow\!\!\underset{16}{]}$$

Los corchetes indican que **se incluyen los valores extremos.**

Nota: Un intervalo puede ser abierto en un extremo y cerrado en el otro.

Problema 1. Los médicos del hospital *Santa Fe* registraron los valores que aparecen abajo, acerca de la variable *tiempo de gestación*, en días, de una muestra de 60 mujeres. Se preguntan sobre la mejor manera de organizar esos datos, ¿cómo podrían hacerlo?

276	280	272	275	281	280	274	273	274	275
280	277	278	280	279	281	278	281	280	280
273	279	275	277	273	277	275	278	277	276
279	278	277	278	278	278	280	276	280	276
277	274	276	276	277	274	279	279	277	279
278	277	279	278	279	279	278	277	278	278

Un somero vistazo del conjunto anterior nos revela esto:

- **Casi todos** los datos se repiten y,
- Son relativamente **pocos** los datos diferentes entre sí.

Siendo así, una **tabla de frecuencias simple** es la mejor forma de presentarlos.

La primera columna de la tabla de frecuencias simple (tabla 2.1) se forma con los valores sucesivos de la variable *tiempo de gestación*; en este caso se trata de **números enteros**, de los cuales 272 es el menor y... ¿cuál es el valor del mayor? ______.

PERIODOS DE GESTACIÓN DE 60 MUJERES EN EL HOSPITAL SANTA FE		
TIEMPO DE GESTACIÓN (DÍAS)	CONTEO	FRECUENCIA
272	/	
273	///	
274	///	
275	///	
276	////	
277	~~////~~ ///	
278	~~////~~ ~~////~~	
279	~~////~~ ///	
280	~~////~~	
281	///	
		Total: 60

Tabla 3.1

La segunda columna de la tabla 3.1, con el encabezado "CONTEO", registra cuántas veces cada uno de los datos aparece en el conjunto, es decir, su frecuencia. Para contarlos se sugiere tachar el primer valor (276) y trazar una rayita inclinada en el renglón correspondiente de la tabla; tachar el segundo valor (280) y trazar otra rayita inclinada en el renglón correspondiente,... y así sucesivamente, hasta tachar todos los datos. La quinta rayita se traza cruzada.

En la tabla de frecuencias 2.1 no fueron registrados los 12 datos que están en las dos últi mas columnas de valores; regístralos con rayitas y cuando termines escribe los valores de las frecuencias a la derecha de las rayitas.

Problema 2. Un veterinario registró los valores del *peso al nacer* de 50 ratas (en gramos). ¿Cuál es la mejor manera de organizarlos para tener más claro su comportamiento?

4.9	4.9	5.0	4.6	5.1	4.7	4.9	4.8	4.8	4.7
5.0	4.9	**4.5**	4.7	4.9	5.0	4.8	5.0	**5.3**	5.1
4.9	5.0	5.0	4.8	5.0	4.6	5.0	4.8	5.0	4.9
5.1	4.9	4.9	5.0	4.8	5.1	4.7	4.7	4.6	4.8
4.9	4.7	4.6	4.9	4.6	4.8	4.9	4.9	5.0	5.0

El conjunto tiene dos características por las que se recomienda organizarlo mediante una tabla de frecuencias:

- **Casi todos** los valores **se repiten**.
- Son relativamente **pocos** los datos diferentes entre sí.

Puesto que la precisión de los datos está en **décimos** de gramo – y por ser el menor 4.5 y el mayor 5.3–, los números de la primera columna de la tabla comienzan en 4.5 y se deben suceder con una **décima** de diferencia, hasta llegar a 5.3

Completa la tabla 3.2. Para hacerlo, registra cada uno de los datos del conjunto de la manera mencionada y al final escribe los valores numéricos de las frecuencias.

PESOS DE 50 RATAS AL NACER		
PESO (GRAMOS)	FRECUENCIA	
	CONTEO	VALOR
4.5		
4.6		
4.7		
4.8		
4.9		
5.0		
5.1		
5.2		
5.3		
		Total: ____

Tabla 3.2

TABLA DE FRECUENCIAS POR INTERVALOS

EJEMPLO DE CRITERIO PARA REDONDEAR UN NÚMERO DECIMAL

■ Para redondear 5.08472 a milésimos, considero 5.084 como “respuesta potencial”.
Tomo las cifras después de los milésimos y las expreso como fracción decimal: .72
Por ser .72 mayor que 0.5, le sumo 1 milésimo a la respuesta potencial y obtengo: 5.085

■ Para redondear 5.08439 a milésimos, considero 5.084 como “respuesta potencial”.
Tomo las cifras después de los milésimos y las expreso como fracción decimal: .39
Por ser .39 menor que 0.5, la respuesta potencial la considero el resultado: 5.084

■ Para redondear 9.175000 a centésimos, considero 9.17 como “respuesta potencial”.
Tomo las cifras después de los centésimos y las expreso como fracción decimal: .500
Por ser .500 igual a 0.5, y como la última cifra de la respuesta potencial es impar, le sumo 1 centésimo para obtener 9.18

■ Para redondear 9.125000 a centésimos, considero 9.12 como “respuesta potencial”.
Tomo las cifras después de los centésimos y las expreso como fracción decimal: .500
Por ser .500 igual a 0.5, y como la última cifra de la respuesta potencial es par, la respuesta potencial la dejo sin cambio alguno y es el resultado: 9.12

Problema 3. Un cardiólogo calculó el *peso del corazón*, en gramos, de 80 adultos hospitalizados ¿Cómo los debe organizar para apreciar mejor su comportamiento?

352	349	355	348	356	354	350	352	355	353
336	339	349	352	342	350	348	344	352	342
360	346	354	353	351	344	343	353	335	339
357	350	352	349	359	352	358	348	348	350
354	338	343	354	346	353	352	351	344	343
352	351	348	344	357	349	338	354	351	**334**
345	348	338	350	349	341	350	355	358	350
346	345	346	345	347	347	346	347	342	347

El conjunto anterior está contenido en el intervalo [334, 360] y tiene 27 datos **distintos**.

Si se aplica el procedimiento visto, la tabla resultará demasiado extensa porque sería una tabla de 27 renglones. ¿Cómo conviene entonces organizar los datos?

En este conjunto destacan tres cosas:

1. El número de datos es relativamente **grande** (**más de 30**).
2. Los valores están muy **dispersos**, es decir, su "**rango**" es relativamente amplio. El rango, **r**, es el resultado de restar el menor de los datos del mayor, así:

   r = 360 –334 = 26

3. Hay ciertos **indicios de tendencia central**, ya que algunos datos con valores cercanos entre sí se repiten más que los otros.

Si sucede lo anterior se recomienda construir una **Tabla de Frecuencias por Intervalos**.

¿Cómo? El procedimiento consta de cuatro pasos.

Primer paso: Se calcula el **rango real, R**. Puesto que los datos mayor y menor son 360 y 334, el **rango real**, **R**, se calcula como la diferencia de 360.5 – 333.5

$$R = 360.5 - 333.5 = 27$$

Este resultado indica que 27 es el **número de datos diferentes** que como máximo **puede haber** en el conjunto anterior; comenzando por 334 y terminando en 360:

y se dice "*puede haber"* porque alguno de los 27 valores pudiera tener frecuencia cero.

¿El intervalo 334-360 es cerrado? ___, ¿por qué? _________________________________ .

***Segundo paso*:** Se decide **cuántos intervalos de clase** formar (se recomienda entre 5 y 15 inclusive). Para hacerlo es útil aplicar la "regla de Sturges":

$N = 3.322\log n + 1$ **(N indica el número de intervalos, n el número de datos)**

N = 3.322log 80 + 1

N = 3.322(1.903089987) + 1

N = 6.3222064937 + 1

N = 7.3220 ≈ 7 (se redondea al entero más próximo)

El rango real, **R**, se divide entre el número N de **intervalos de clase**. El resultado es la amplitud (A) del intervalo y nos indica el número de **datos diferentes** que podrá contener cada intervalo. Si decidimos formar 7 intervalos de clase:

$$A = \frac{R}{N}$$

Es decir: $A = \frac{\mathbf{27}}{7} = \mathbf{3}.8 \approx 4$

Tercer paso: se construyen los intervalos de clase**.** Para ello, el menor de los datos del conjunto original se toma como el **límite inferior** del primer intervalo de clase. El **límite superior** se obtiene sumándole $A - 1$. El límite inferior no siempre tiene que corresponder al menor de los datos, como se verá más adelante.

En nuestro ejemplo, el menor de los datos es 334 y el valor de $\boldsymbol{A}$ es 4 (entonces $\boldsymbol{A} - 1 = 3$). Por lo tanto, el primer intervalo de clase es:

334 – 337

El segundo intervalo comienza en el entero que sigue al límite superior del primer intervalo, o sea, en 338, éste será el límite inferior del segundo intervalo.

INTERVALOS
334 - 337
338 - 341

Tal como se hizo para el primer intervalo, el límite superior del segundo se obtiene al sumar $\boldsymbol{A} - 1$ (o sea 3) a su límite inferior:

338 – 341

De la misma manera se calculan los intervalos de clase restantes, hasta asegurarse de incluir al último de los datos.

En el ejemplo que nos ocupa, el menor de los datos es 334 y el mayor 360; entonces, ¿cuáles son los intervalos de clase? (Completa la tabla).

Cuarto paso. Se determinan las frecuencias, es decir, cuántos datos contiene cada intervalo. En la siguiente tabla realiza cuidadosamente este último paso. Sugerencia: tacha uno por uno los datos, al tiempo que los registras en la columna de conteo, mediante rayitas verticales, la quinta rayita se dibuja diagonal para formar grupos de 5. En la columna de Frecuencias escribe los totales correspondientes.

A. PESO EN GRAMOS DE LOS CORAZONES DE 80 ADULTOS HOSPITALIZADOS		
PESO	CONTEO	FRECUENCIAS
334 - 337		
338 - 341		
342 - 345		
346 - 349		
350 - 353		
354 - 357		
358 - 361		
		Total:

En una tabla de frecuencias por intervalos, ¿a qué debe ser igual la suma de las frecuencias? _________________________ .

Antes de continuar, recuerda dos cosas:

a) La tabla de frecuencias **por intervalos** debe tener un **título** y al menos dos columnas, en una se indican los **intervalos de clase** y en otra las **frecuencias. Frecuencia** significa el **número de datos en un intervalo determinado**. Nota: no se incluye la columna de conteo.

b) Con una tabla de frecuencias por intervalos se logra una impresión más clara de la **tendencia central** y de la **dispersión** de los datos. En el ejemplo anterior puede observarse una tendencia de los datos a aglutinarse alrededor del intervalo de clase 350 – 353 y a desvanecerse en los extremos.

Problema 4. El Jefe de Recursos Humanos de la empresa X aplicó un examen de conocimientos a los choferes que solicitaron su promoción. Las calificaciones fueron:

5.3	4.2	3.7	6.5	4.8	4.6	4.4	4.8	3.3	4.6
4.4	4.8	3.3	4.9	3.6	4.4	6.3	3.5	4.6	4.3
6.1	3.5	6.0	3.5	4.9	3.3	3.5	4.1	6.0	3.6
4.6	5.6	4.8	6.4	4.3	4.8	3.1	6.1	4.4	4.9
4.8	4.6	4.4	4.9	4.6	5.0	4.6	6.4	6.1	3.2
3.0	6.3	4.6	6.1	5.2	3.2	6.0	4.9	4.3	

¿Fueron altos los resultados del examen?. . . no, pero el Jefe de Recursos Humanos desea tener una imagen visual más clara de su **tendencia central** y de su **dispersión**, por lo que decide organizarlos en una tabla de frecuencias por intervalos.

¿Esta es una buena decisión? ____ , ¿por qué? ___________________________________

____

¿Qué pasos es recomendable que siga el Jefe de Recursos Humanos?

Primer paso (llena los espacios vacíos):

Calificación más baja: ____; calificación más alta: ____ Por lo tanto, el valor del **rango real** es:

R = 6.55 – 2.95 = 3.6

Segundo paso.

El rango real, **R**, se divide entre el número de intervalos, N, que se deseen formar.

Apliquemos la "regla de Sturges" para calcular el número de intervalos de clase:

N = 1 + 3.322 (log n) **donde N representa el número de intervalos de clase y n es el número de datos.**

En nuestro ejemplo, tenemos 59 datos, por lo tanto:

N = 1 + 3.322 (log 59)

N = 1 + 3.322 (1.770852012)

N = 1 + 5.88

N = 6.88

Con este resultado se decide si el número conveniente de intervalos es 6 ó 7.

Al calcular la amplitud para cada opción, tenemos:

$$A = \frac{3.6}{6} = 0.6$$

O bien: $$A = \frac{3.6}{7} = 0.514$$

Por el número de cifras en la parte decimal, resulta más práctico decidirse por 6 intervalos.

Como pudiste observar, el resultado que se obtiene aplicando la regla de Sturges no debe considerarse definitivo, sino sólo como una guía. El número de intervalos de clase obtenido por la regla puede aumentarse o disminuirse, según convenga.

Tercer paso: se construyen los intervalos de clase. Para eso, el límite inferior del primer intervalo de clase es 3.0 (calificación más baja). El límite **superior** se obtiene al sumarle $A - (\frac{1}{10})$:

$A - (\frac{1}{10})$ = 0.6 – 0.1 = 0.5 **(la precisión de los datos es de décimos, por eso se resta $\frac{1}{10}$ a A)**

Puesto que: 3.0 + 0.5 = 3.5

El primer intervalo es:

3.0 – 3.5

El límite **inferior** del **segundo** intervalo de clase se obtiene sumándole **un décimo** al límite superior del primer intervalo; se hace así porque los datos están expresados en **décimos**:

3.5 + .1 = 3.6

El límite superior del segundo intervalo se obtiene sumándole 0.5 a su límite inferior, como se hizo para el primer intervalo:

$$3.6 + .5 = 4.1$$

Por lo tanto, los dos primeros intervalos de clase son:

3.0 – 3.5
3.6 – 4.1

Así se continúa la construcción de los intervalos de clase, hasta incluir al dato más alto (6.5), ¿lo puedes hacer enseguida?

____

____

____

____

____

Enseguida se reproducen los datos, haz el recuento y completa la tabla:

5.3	4.2	3.7	6.5	4.8	4.6	4.4	4.8	3.3	4.6
4.4	4.8	3.3	4.9	3.6	4.4	6.3	3.5	4.6	4.3
6.1	3.5	6.0	3.5	4.9	3.3	3.5	4.1	6.0	3.6
4.6	5.6	4.8	6.4	4.3	4.8	3.1	6.1	4.4	4.9
4.8	4.6	4.4	4.9	4.6	5.0	4.6	6.4	6.1	3.2
3.0	6.3	4.6	6.1	5.2	3.2	6.0	4.9	4.3	

EMPRESA X CALIFICACIONES DEL EXAMEN DE CONOCIMIENTOS		
Intervalo	Conteo	Frecuencia
3.0 – 3.5		
3.6 – 4.1		
Total		

¿Coincide la suma de frecuencias con el número original de datos? ____.

¿Hacia cuál intervalo tienden a acumularse los datos? _______________.

LÍMITES REALES DE CLASE

La tabla de frecuencias por intervalos también se puede construir de tal manera que el límite superior de un intervalo **coincida** con el límite inferior del siguiente. Al **límite común** de dos intervalos se le denomina **límite real de clase**.

Fija tu atención en los límites de los intervalos de clase de la tabla 3.3

PESO EN KG DE LAS CALABAZAS COSECHADAS EN UNA HUERTA DE COMITÁN, CHIAPAS	
Intervalos	Frecuencia
0.5 – 0.8	1
0.9 – 1.2	6
1.3 – 1.6	3
1.7 – 2.0	9
2.1 – 2.4	11
2.5 – 2.8	48
2.9 – 3.2	13
3.3 – 3.6	7
3.7 – 4.0	5
Total	103

Tabla 3.3

¿En cuánto difieren el límite superior de un intervalo de clase y el límite inferior del siguiente? _______________ .

Para obtener el **límite real de clase** entre el primero y el segundo de los intervalos de clase, se suma el límite superior del primero (0.8) con el límite inferior del segundo (0.9) y el resultado se divide entre dos, se obtiene 0.85

Asimismo, el límite inferior del primer intervalo de clase se disminuye en cinco centésimas (la mitad de una décima); así:

0.5 – 0.8 es el intervalo **de clase**

0.45 – 0.85 es el intervalo **real de clase**

De la misma manera se calcula el límite real de clase entre el segundo y el tercer intervalo y se forma el segundo intervalo (su límite real inferior será 0.85)... y así sucesivamente.

Termina de construir la tabla con **límites reales de clase**.

PESO EN KG. DE LAS CALABAZAS COSECHADAS EN UNA HUERTA DE COMITÁN, CHIAPAS	
Intervalos	Frecuencia
0.45 – 0.85	1
0.85 – 1.25	6
	3
	9
	11
	48
	13
	7
	5
Total	103

FICHA DE TRABAJO III.1

Emplea la computadora para (imprime y engrapa las hojas aquí):

1. Construir la tabla de frecuencias simple. ¡No olvides ponerle título¡

MEDICIONES DE LA VELOCIDAD DE LA LUZ EN EL VACÍO (10^8 m/s)

2.97	2.99	2.98	2.99	2.97	2.97	2.95
2.95	2.98	2.99	2.90	2.99	2.96	2.98
2.98	2.96	2.99	2.96	2.96	2.95	2.98
2.98	2.97	2.98	2.93	2.93	2.97	2.97
2.96	2.93	2.95	2.97	2.99	2.99	2.94
2.99	2.96	2.97	2.99	2.98	2.96	2.94
2.95	2.91	2.95	2.94	2.96	2.94	2.93
2.99	2.96	2.98	2.99	2.99	2.97	2.99
2.96	2.98	2.98				

2. Organizar los siguientes datos en una tabla de frecuencias por intervalos.

TAMAÑO DE 60 VIRUS DE LA FAMILIA *TOGAVIRIDAE* EN NANÓMETROS (10^{-9} m)

55	74	67	64	57	53	58	73	52	39	61	46	46	53	59	55	59	65	50	58
63	50	47	51	49	64	69	58	59	77	45	44	53	65	84	52	58	64	57	51
79	45	55	53	58	54	54	48	39	66	35	62	60	58	72	78	43	62	71	68

Lee el siguiente resumen y escribe abajo algo que haya faltado:

RESUMEN

Los conjuntos relativamente numerosos de datos conviene organizarlos en una tabla de frecuencias, si casi todos los datos se repiten y hay pocos diferentes entre sí.

Cuando los datos estén expresados como números enteros, se indicarán los valores individuales en la tabla de frecuencias de tal modo que se sucedan con una unidad de diferencia. Si estuvieran expresados hasta décimos, se indicarán sucediéndose con una décima de diferencia, y así sucesivamente...

Una tabla de frecuencias simple consta de dos columnas, por lo menos, además del título. Una corresponde a los valores individuales de los datos y la otra se refiere a su frecuencia, es decir, al número de veces que se repite cada dato.

Otras tablas reciben el nombre de tablas de frecuencias por intervalos; en éstas el valor de la frecuencia indica el número de datos con valores comprendidos entre los extremos de un intervalo. Para construir este tipo de tablas se sigue un procedimiento de varios pasos en que se construyen intervalos "de clase" para los que se determina su frecuencia.

Otra forma de presentar la tabla de frecuencias por intervalos es utilizando límites reales de clase; en este caso coinciden los límites de dos intervalos contiguos.

____

4ª SESIÓN
ORGANIZACIÓN DE DATOS EN HISTOGRAMA, POLÍGONO DE FRECUENCIAS Y OJIVA

COMPETENCIAS POR LOGRAR

A partir de una tabla de frecuencias por intervalos dada, construir:

- **Un histograma.**
- **Un polígono de frecuencias.**
- **Una ojiva.**

MATERIAL NECESARIO: ESCUADRAS, LÁPIZ, SACAPUNTAS Y GOMA.

HISTOGRAMA

Una tabla de frecuencias por intervalos donde la variable independiente sea *continua* se puede representar mediante un histograma o un polígono de frecuencias. El histograma se parece a una gráfica de barras y el polígono de frecuencias es una gráfica lineal, pero ambos pueden representar lo mismo.

TÉRMINOS PARA RECORDAR. . .

Frecuencia: número de veces que se repite un valor de la variable (dato). En una tabla donde los datos estén agrupados en intervalos de clase: número de datos comprendidos en el intervalo.

Frecuencia acumulada: cuando los datos estén ordenados de menor a mayor, o viceversa, la frecuencia que se obtiene sumando la del dato actual más la acumulada del dato anterior.

Frecuencia relativa de un dato o intervalo: cociente de su frecuencia entre el total de datos.

Frecuencia relativa porcentual*:* La frecuencia relativa multiplicada por cien.

Marca de clase: punto medio de un intervalo, sea éste de clase o real de clase.

Variable continua: la que en teoría puede tomar cualquier valor dentro de un intervalo.

Variable discreta: la que sólo puede tomar ciertos valores dentro de un intervalo.

Variable dependiente: cuyos valores dependen de los que tome otra (la variable independiente).

Variable independiente: se le asignan valores a voluntad.

Problema. Una nutrióloga quiere construir una gráfica que le permita apreciar mejor las características de los datos contenidos en la tabla que aparece abajo, ¿cómo debe hacerla?

Una manera es elaborar un **histograma**.

El histograma es una gráfica de la tabla de frecuencias por intervalos, **siempre y cuando la variable independiente sea continua**.

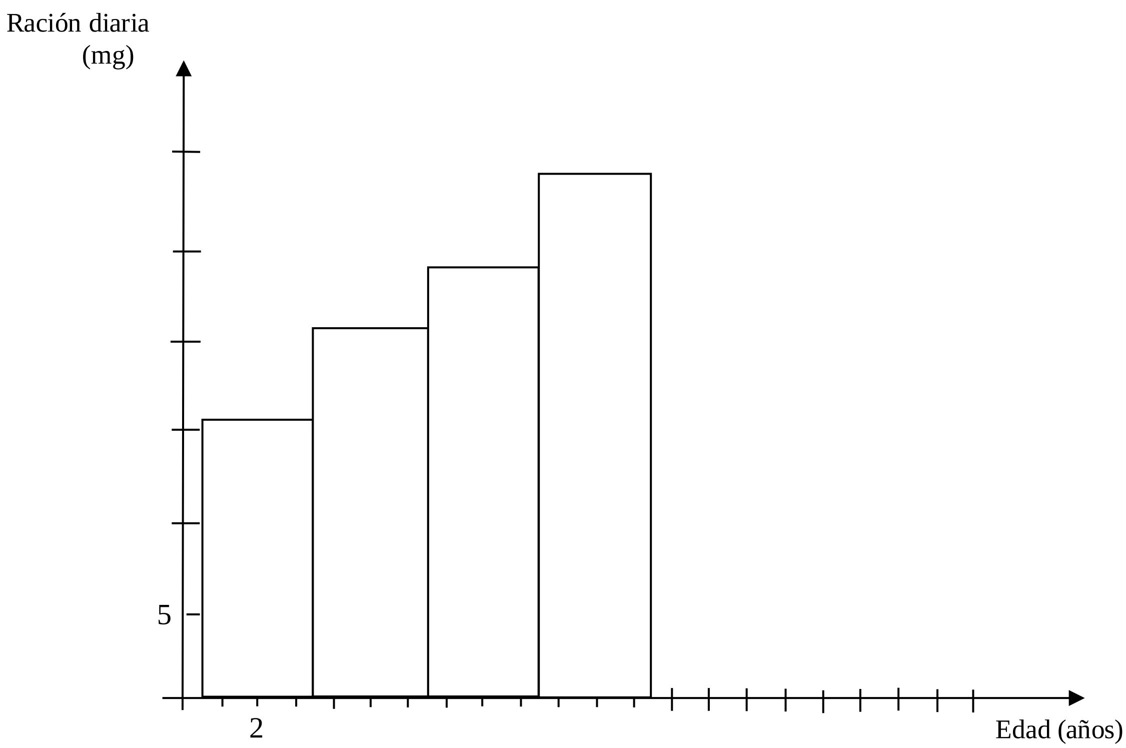
El histograma debe incluir **toda** la información de la tabla. En el eje de las abscisas se indican los **límites reales de clase** o las **marcas de clase** (puntos medios de los intervalos), con sus unidades. En el eje de las ordenadas **las frecuencias, las frecuencias relativas, los porcentajes** o la variable correspondiente con sus unidades.

NECESIDADES DIARIAS DE PROTEÍNA PARA LA MUJER SEGÚN EXPERTOS DE LA FAO	
Edad (años)	**Ración diaria (mg)**
1 – 3	15.5
4 – 6	21.5
7 – 9	24.5
10 – 12	29.5
13 – 15	31.5
16 – 18	30.5
19 – 21	29.5

Tabla 4.1

El histograma es una gráfica de barras adyacentes, cuyas bases son los intervalos de clase (se pueden indicar solamente las marcas de clase) y sus alturas las frecuencias, los porcentajes o los valores de la variable correspondiente.

Completa el histograma con escuadra y lápiz (no olvides ponerle título):



POLÍGONO DE FRECUENCIAS

Problema. Una enfermera pretende construir una gráfica **de líneas** para visualizar mejor el comportamiento de la temperatura, cuando aparece la erupción en enfermos de sarampión (ver tabla 4.2).

¿Qué le recomendarías?

La gráfica lineal que se utiliza comúnmente para casos como este se denomina **polígono de frecuencias.**

El polígono de frecuencias se puede trazar **sobre el histograma**; para hacerlo, en las bases superiores de cada rectángulo del histograma, señalándolas con puntos.

TEMPERATURAS DE 300 ENFERMOS DE SARAMPIÓN AL APARECER LA ERUPCIÓN		
Temperatura (°C)	**Frecuencia** (pacientes)	**Frecuencia Relativa**
37.0 - 37.2	5	.017
37.3 - 37.5	3	.010
37.6 - 37.8	11	.037
37.9 - 38.1	32	.107
38.2 - 38.4	84	.280
38.5 - 38.7	107	.357
38.8 - 39.0	38	.127
39.1 - 39.3	10	.033
39.4 - 39.6	5	.017
39.7 - 39.9	5	.017
Total	**300**	**1.000**

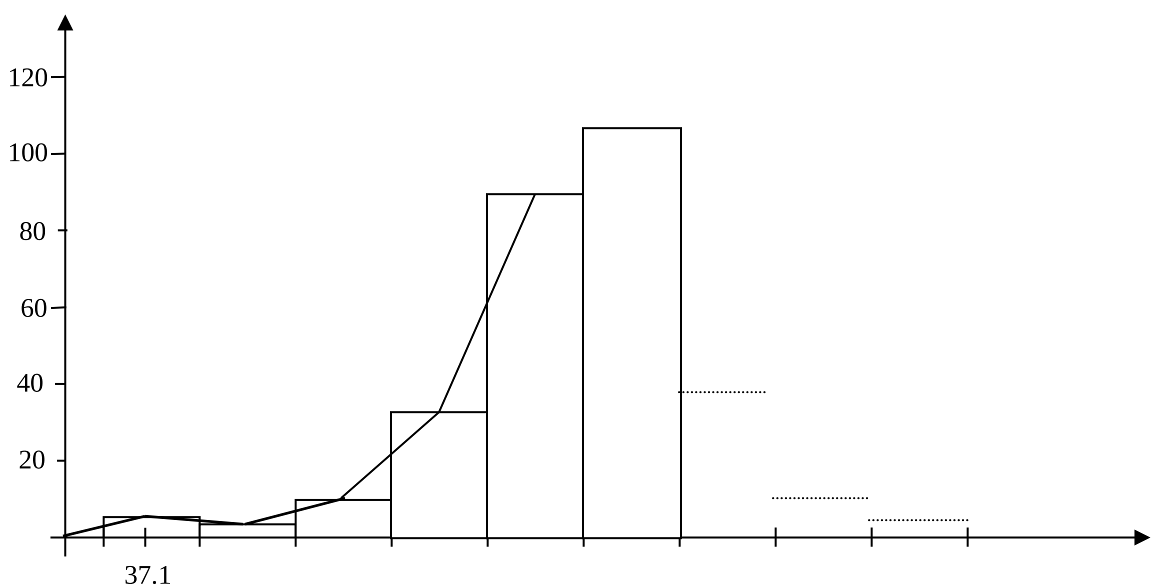
Tabla 4.2

Enseguida se unen los puntos consecutivos mediante líneas rectas.

El polígono empieza en la marca de clase de un intervalo real de clase, imaginario, con **frecuencia cero** (sobre el eje de las abscisas) y termina en otro intervalo igual (también sobre el eje de las abscisas).

Si en el eje vertical se expresan **frecuencias relativas,** entonces se trata de un **polígono de frecuencias relativas**.

Completa el **Polígono de Frecuencias** de la tabla 2.5 (ponle su título).



OJIVA

Después de organizar sus datos en una tabla de frecuencias acumuladas por intervalos, un médico procede a representarlos mediante una **ojiva.**

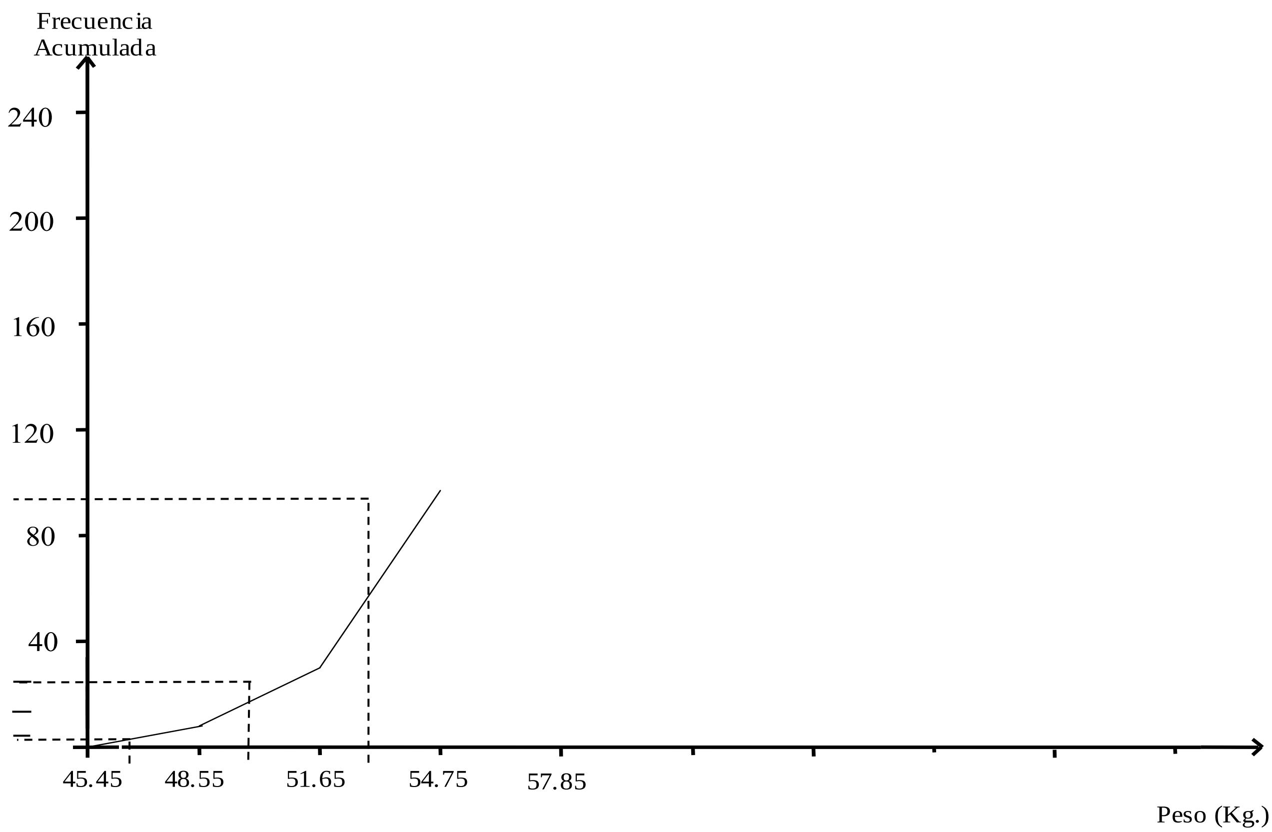
MEDICIONES DEL PESO DE 238 VARONES CON EDADES DE 17 AÑOS Y ESTATURAS ENTRE 1.60 y 1.80 M.		
Peso en Kg.	**Frecuencia**	**Frecuencia Acumulada**
45.45-48.55	9	9
48.55-51.65	21	30
51.65-54.75	67	97
54.75-57.85	72	169
57.85-60.95	32	201
60.95-64.05	21	222
64.05-67.15	11	233
67.15-70.25	3	236
70.25-73.35	2	238
	Total: 238	

Tabla 4.3

En el eje de las abscisas registra los **límites reales** de clase y en el de las ordenadas las frecuencias acumuladas (pueden ser también las frecuencias relativas acumuladas o los porcentajes acumulados). Los puntos los une por medio de líneas rectas.

Para empezar el trazo de su ojiva, al límite real inferior del primer intervalo lo ubica sobre el eje horizontal.

Termina la ojiva de la tabla 4.3 (con título).



Cuando la gráfica esta “anclada” al eje de las abscisas por la izquierda, se le llama **ojiva** “**menor que”**. ¿Cuántos varones tuvieron un peso **menor que** 54.75 kilogramos? ____ .

COMPLETA LAS ORACIONES Y TERMINA LA OJIVA “MAYOR QUE” DE LA TABLA 4.4

Se trata de una **Tabla de Porcentajes Acumulados por Intervalos**.

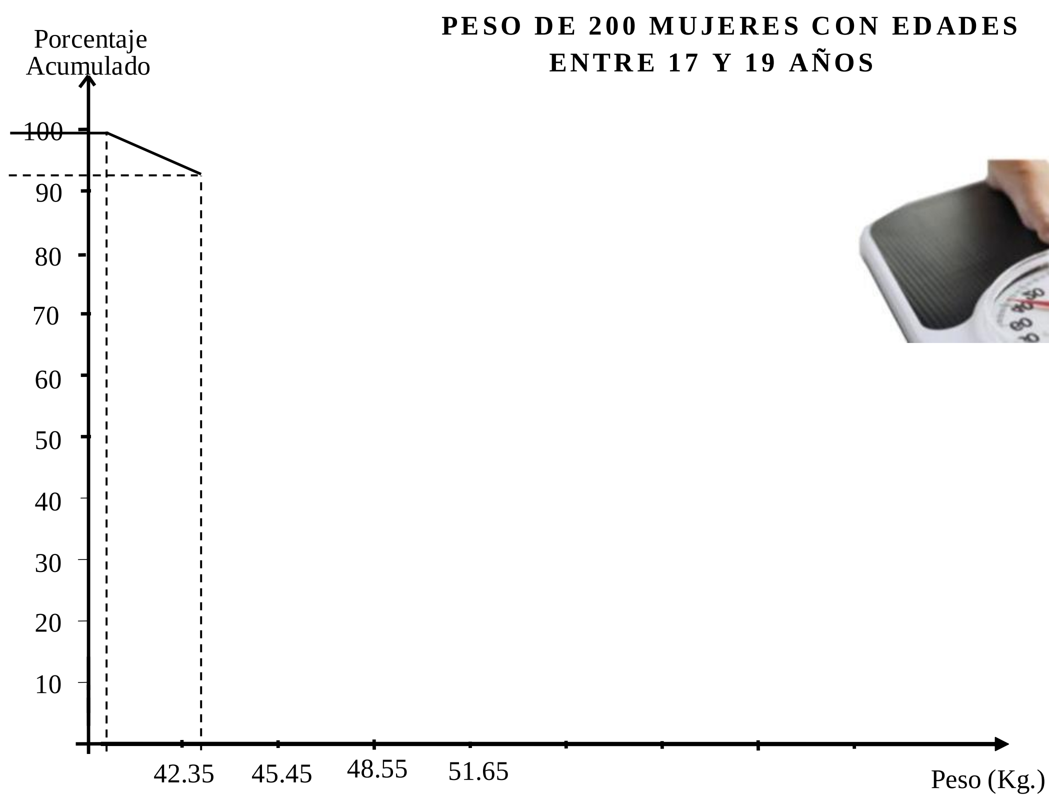
PESOS DE 200 MUJERES CON EDADES ENTRES 17 Y 19 AÑOS		
Peso en Kg	**Porcentaje**	**Porcentaje Acumulado**
42.35-45.45	7	100
45.45-48.55	23	93
48.55-51.65	38	70
51.65-54.75	20	32
54.75-57.85	7	12
57.85-60.95	4	5
60.95-64.05	1	1
Total	100	

TABLA 4.4

Al construir la ojiva “mayor que”, en el eje de las abscisas se marcan los límites reales de clase; en el de las ordenadas se grafican las frecuencias acumuladas, frecuencias relativas acumuladas o porcentajes _______________.

Cuando la ojiva es “mayor que” el último punto se localiza sobre el eje de las abscisas y corresponde al límite superior del último intervalo.

La frecuencia acumulada de un intervalo es la suma de su _______________ más las _______________ de los intervalos anteriores.



La ojiva **mayor que** siempre estará anclada al eje de las abscisas por la derecha. ¿Qué porcentaje de mujeres tuvieron un peso **mayor que** 48.55 kilogramos? ___________.

FICHA DE TRABAJO IV.1

PREPARA COMPUTADORA.

1.

PRODUCCIÓN DE TRIGO POR HECTÁREA EN 100 CAMPOS DE CULTIVO DE LA REPÚBLICA MEXICANA		
Producción (kg/ha)	**No. de campos de cultivo**	**Frecuencia relativa**
900 – 1200	2	.02
1300 – 1600	7	.07
1700 – 2000	9	.09
2100 – 2400	13	.13
2500 – 2800	24	.24
2900 – 3200	32	.32
3300 – 3600	10	.10
3700 – 4000	3	.03
Total	100	1.00

Tabla 4.5

Problema. En la Secretaría Agraria se conoce la producción de trigo del país en cierto año (derecha).

Representa esos datos mediante un histograma, un polígono de frecuencias y una ojiva. Hazlo por medio de computadora (usa la *frecuencia relativa*).

2. Para construir un histograma, ¿cómo debe ser la variable independiente? _____________.

RESULTADOS DE 100 CORREDORAS EN LA PRUEBA DE 400 M LIBRES	
Tiempo logrado (s)	**Corredoras**
49.50 — 50. 00	2
50.05 — 50. 55	7
50.60 — 51. 10	32
51.15 — 51.65	30
51.70 — 52. 20	22
52.25 — 52. 75	2
52.8 — 53.30	4
53.35 — 53.85	1
Total	100

3. Al tratar de igualar el récord de Ana Gabriela Guevara Espinoza, la campeona mundial en 400 m libres, cien corredoras obtuvieron los resultados consignados en la tabla de la izquierda. Por medio de una computadora construye el histograma, el polígono de frecuencias y la ojiva.

RESUMEN

El histograma es una forma de graficar los datos mediante rectángulos a partir de una tabla de frecuencias por intervalos, siempre y cuando la variable independiente sea continua. Sobre el eje de las abscisas se indican los límites reales de clase con sus unidades, o se pueden indicar las marcas de clase, es decir, los puntos medios de los intervalos. En el eje de las ordenadas las frecuencias.

Los rectángulos del histograma van unidos, sus bases son de un ancho igual al tamaño de los intervalos reales de clase y sus alturas equivalentes a las frecuencias o a las frecuencias relativas de cada intervalo (éstas se calculan dividiendo la frecuencia del intervalo entre la suma de las frecuencias de todos los intervalos).

El polígono de frecuencias es una gráfica lineal que comienza y termina sobre el eje de las abscisas. Se puede trazar sobre el histograma o en forma independiente; para ello se trazan líneas rectas que unan los puntos medios de los lados superiores de los rectángulos, a partir de un imaginario intervalo real de clase con frecuencia cero y concluyendo en otro con las mismas características.

Para construir una ojiva "menor que", en el eje horizontal se registran los límites reales de clase y en el vertical las frecuencias acumuladas (pueden ser también las frecuencias relativas acumuladas o los porcentajes acumulados) y los puntos se unen por medio de líneas rectas. El primer punto se localiza sobre el eje horizontal y corresponde al límite real inferior del primer intervalo. Otro tipo de ojiva es la denominada "mayor que".

5ª SESIÓN
GRÁFICAS LINEAL Y DE BARRAS

COMPETENCIAS POR ADQUIRIR

- **Seleccionar el tipo más conveniente de gráfica para presentar la información bajo estudio.**
- **Construir con toda la información pertinente una gráfica lineal, de barras o circular.**

MATERIAL NECESARIO: ESCUADRAS, LÁPIZ, GOMA Y SACAPUNTAS.

GRÁFICAS LINEAL Y DE BARRAS

La distribución de un conjunto de datos, sus tendencias y su regularidad, entre otros, se aprecian mejor por medio de una gráfica.

Las gráficas serán más útiles en la medida de su fidelidad para reflejar el comportamiento de los datos.

APRENDE HACIENDO. . .

Enseguida te ocuparás de aprender cómo representar datos por medio de una gráfica.

Tu aprendizaje será eficaz si sigues esta técnica:

1° **Estudia con mucha atención** la primera página, cuyo encabezado se marca con ►. Ahí se aborda la información esencial.

2° **Explica** mentalmente el contenido de la primera página. **Sólo después de hacerlo puedes permitirte pasar a la siguiente página.**

3° En las páginas subsecuentes a la primera aparecen espacios vacíos, llénalos con base en lo aprendido. **Está prohibido consultar la página anterior**, a menos que ya hubieses estudiado y resuelto todo el tema.

▶ GRÁFICA LINEAL

LEE CON ATENCIÓN:

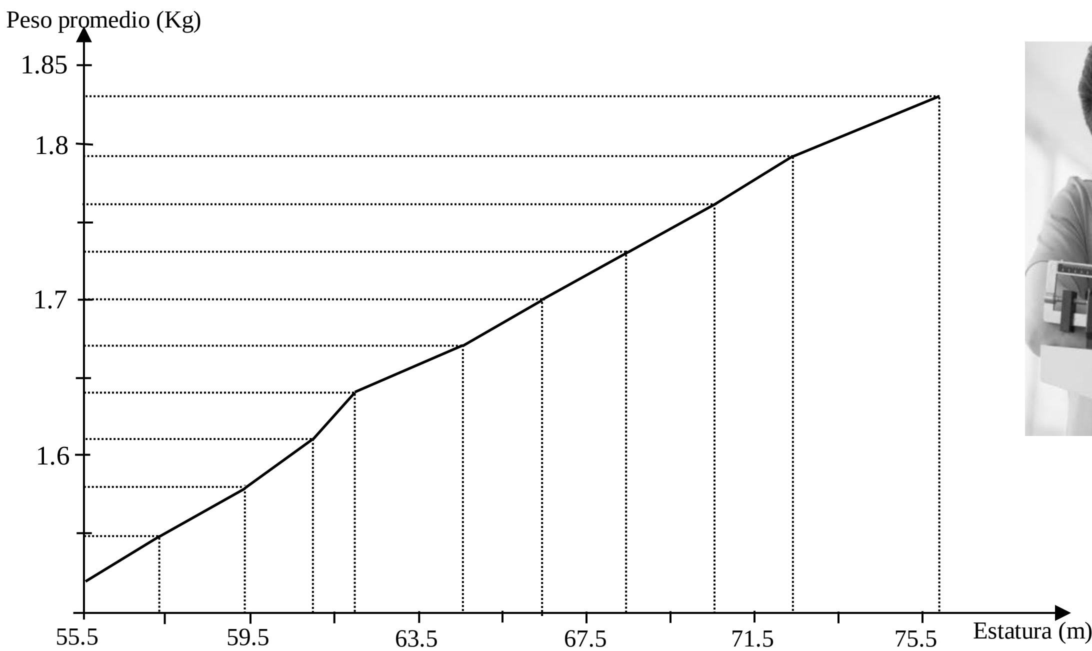
Sobre un plano cartesiano, la relación entre dos conjuntos de datos se puede representar por medio de puntos unidos por **líneas rectas**. A este tipo de representación se le conoce como **gráfica lineal**.

Estos son los pesos promedios recomendables para un hombre, según su estatura:

Estatura (m)	1.52	1.55	1.58	1.61	1.64	1.67	1.70	1.73	1.76	1.79	1.83
Peso Promedio (Kg)	55.5	57.2	59.3	61.1	62.2	64.6	66.5	68.5	70.8	72.4	76.0

La variable **independiente** -estatura-, se indica convencionalmente sobre el eje de las abscisas; la **dependiente**, -peso promedio-, sobre el eje de las ordenadas.
La información la podemos representar por medio de la siguiente gráfica lineal.

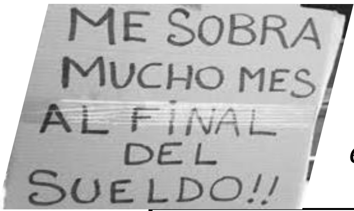
PESO PROMEDIO RECOMENDABLE PARA UN HOMBRE DE ACUERDO A SU ESTATURA



En la gráfica lineal se **interpola**, lo cual consiste en unir los puntos mediante líneas rectas. Al hacerlo se supone que los datos en puntos **intermedios** podrían ser reales.

La gráfica lineal debe incluir:

- Un **título** (donde se mencionen de preferencia las dos variables).
- El **nombre de la variable** en cada eje (**e**n este ejemplo: estatura y peso promedio) con sus **unidades (**metros y kilogramos).
- La **escala** en cada eje. Indicada mediante señales, con sus respectivos valores.



NECESITAS DOS ESCUADRAS GRADUADAS ¡TÉNLAS A LA MANO!

En la siguiente tabla se ordenaron los datos acerca de la *Inflación en México en el Periodo 1995-2003* (Autor: Ing. Manuel Aguirre Botello).

INFLACIÓN (%)	51.97	27.70	15.72	18.61	12.32	8.96	4.40	5.70	3.98
AÑO (PRESIDENTE)	1995 (CSG)	1996 (EZP)	1997 (EZP)	1998 (EZP)	1999 (EZP)	2000 (EZP)	2001 (VFQ	2002 (VFQ	2003 (VFQ

CSG: Carlos Salinas de Gortari; EZP: Ernesto Zedillo Ponce de León: VFQ: Vicente Fox Quezada

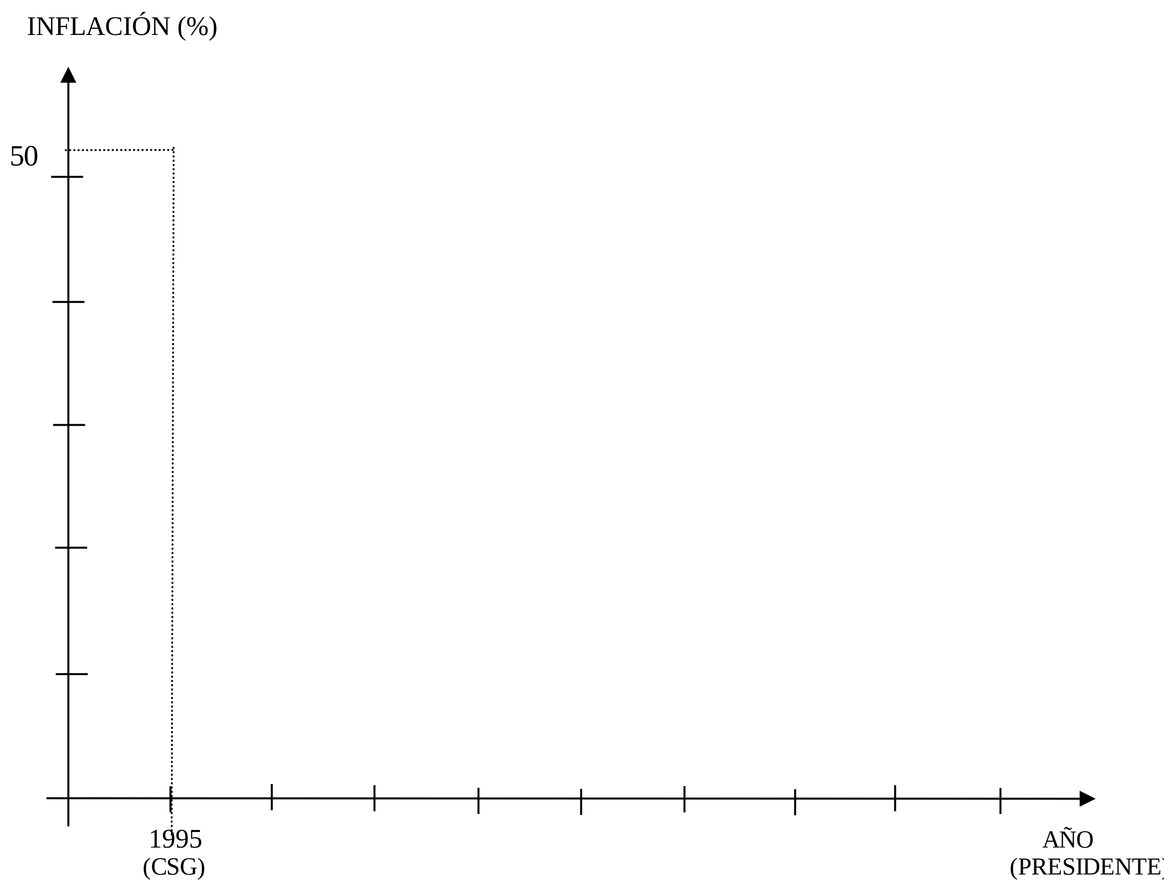
ESCRIBE LAS PALABRAS QUE FALTAN Y CONSTRUYE LA GRÁFICA:

Lo convencional en una gráfica lineal es que sobre el eje horizontal se represente la variable ____________ y sobre el eje vertical la variable _______________.

Mediante las líneas rectas que unen a los puntos, se supone el comportamiento de los datos en puntos intermedios. Es decir, se ___________.

Además del título (donde de preferencia se mencionan las dos variables) la gráfica lineal debe contener:

- Los nombres de las variables en cada eje. En este caso: tiempo e _____________ .
- Las unidades de las variables. En este caso: año y ___.
- La escala sobre cada eje, indicadas con señales y sus respectivos valores.



► GRÁFICA DE BARRAS

LEE CON ATENCIÓN:

Una **gráfica de barras** representa la relación entre dos conjuntos de datos, por medio de rectángulos cuyas **alturas** son los valores de la variable **dependiente**.

Las características **convencionales** de los rectángulos son:

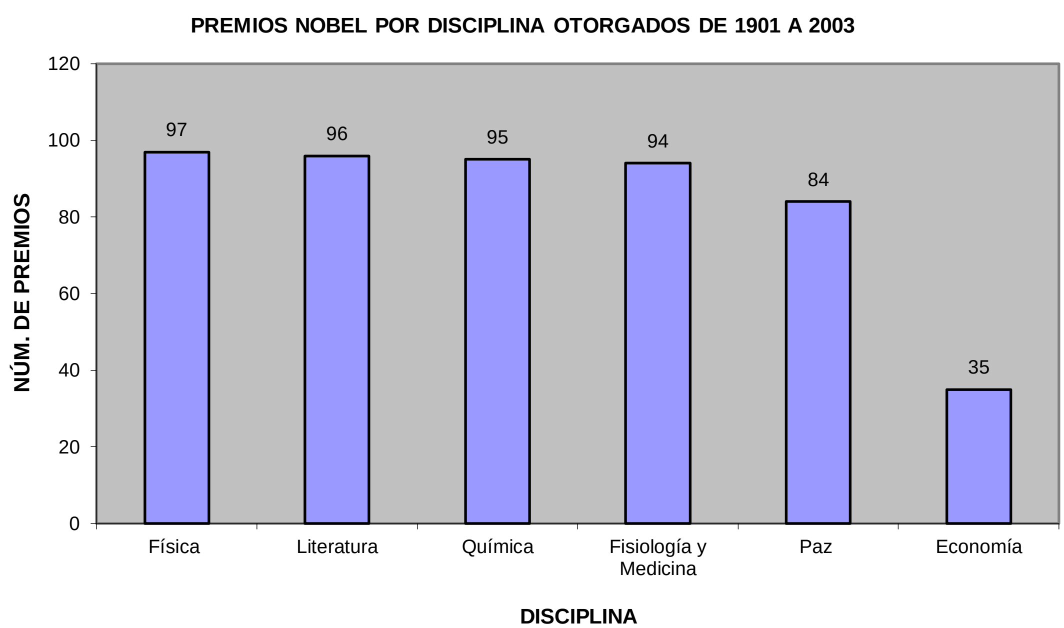
- 1ª. Su ancho se fija **libremente**, pero debe ser el mismo para todos los valores.
- 2ª. Se dibujan **separados** cuando la variable independiente es **discreta**.
- 3ª. Sus bases deben **centrarse** sobre los valores de la variable independiente.
- La gráfica estará completa sólo cuando incluya:
- Un **título,** donde se mencione por lo menos una de las variables.
- Los **nombres y las unidades** de las variables en cada eje.
- La **escala** sobre cada eje.

Por ejemplo, estos datos se pueden representar mediante una gráfica de barras:

PREMIOS NOBEL OTORGADOS DE 1901 A 2003

CATEGORÍA	Economía	Física	Fisiología y Medicina	Literatura	Paz	Química
PREMIOS	35	97	94	96	84	95

•



•

•

- Nota: Los valores en la parte superior de cada rectángulo, y el orden de mayor a menor como se organizaron, permiten apreciar mejor las diferencias (sobre todo cuando los valores de la variable dependiente son muy parecidos).

ESCRIBE LAS PALABRAS FALTANTES
ACERCA DE LA GRÁFICA DE BARRAS:

Los valores de la variable dependiente se representan por las ____________ de los rectángulos.
Las características convencionales de los rectángulos son:

- El ancho debe ser el mismo, pero se escoge ______________.
- Si la variable independiente es discreta, deben dibujarse ____________.
- Las bases se centran en los valores de la variable _______________.

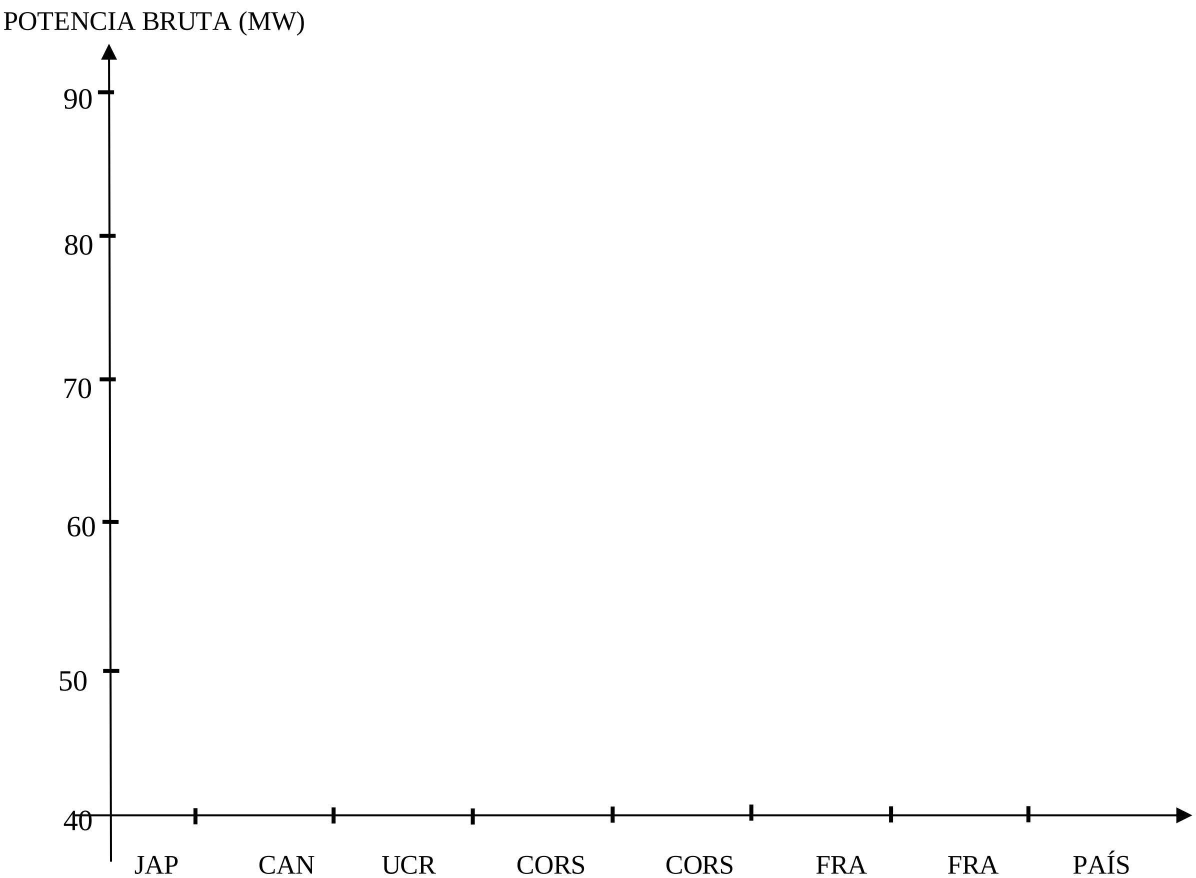
La gráfica de barras completa contiene:

- Un título donde se menciona al menos una de las ______________.

PAÍSES CON LAS CENTRALES NUCLEARES MÁS POTENTES DEL MUNDO EN 2013	
CENTRAL	**POTENCIA BRUTA**
JAPÓN: Central nuclear de Kashiwazaki-Kariwa.	8 212 Mw
CANADÁ: Central Nuclear Bruce.	6 234 MW
UCRANIA: Central Nuclear de Zaporizhia.	6 000 MW
COREA DEL SUR: Central Nuclear de Hanul.	5 908 MW
COREA DEL SUR: Central Nuclear de Hanbit.	5 875 MW
FRANCIA: Central Nuclear de Gravelines.	5 706 MW
FRANCIA: Central Nuclear de Paluel.	5 528MW

Fuente: http://elperiodicodelaenergia.com.

- La escala sobre cada ____. El nombre de la variable y sus unidades en cada eje.
- Construye la gráfica de barras (no olvides ponerle título).



► GRÁFICA DE BARRAS COMPUESTAS

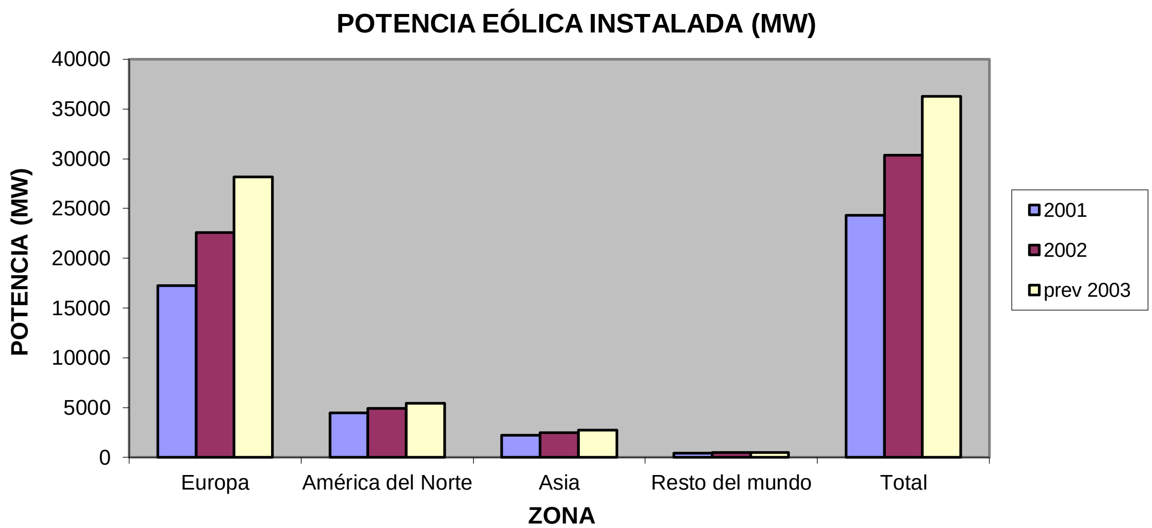
La gráfica de barras compuestas es útil para comparar a más variables dependientes con una relación entre sí.

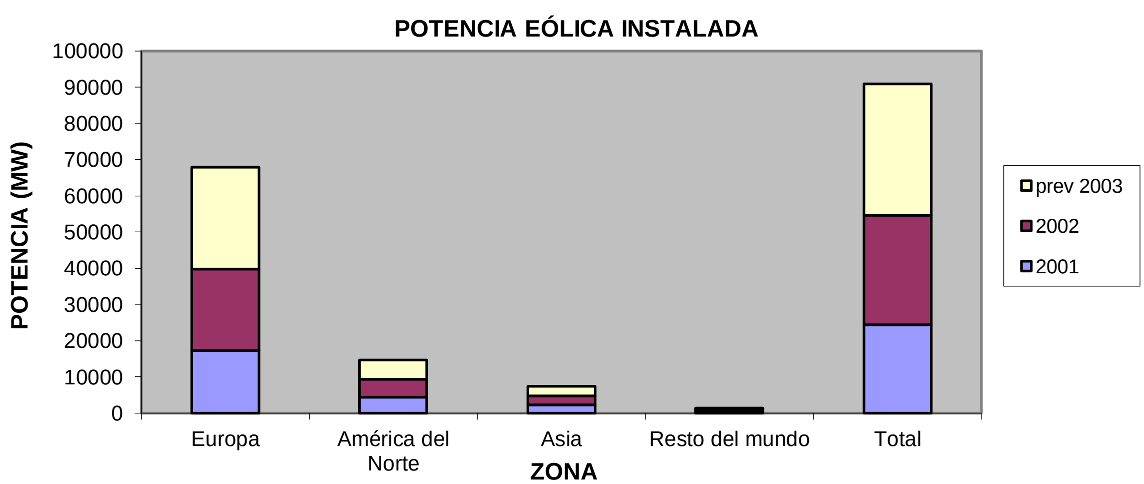
POTENCIA EÓLICA INSTALADA (MW)			
	2001	2002	prev 2003
Europa	17 250	22 560	28 200
América del Norte	4 450	4 900	5 400
Asia	2 220	2 470	2 700
Resto del mundo	400	450	500
Total	24 320	30 380	36 300

Prev= previsto   MW = megawats

Las recomendaciones para su construcción son las mismas que se aplican al construir una gráfica de barras simples, como las estudiadas antes; pero ahora, a cada valor de la variable independiente en el eje horizontal le corresponden más de uno de la dependiente (rectángulos), en el eje vertical.

Dos formas igualmente válidas de una gráfica de barras compuestas de la tabla anterior son las siguientes:





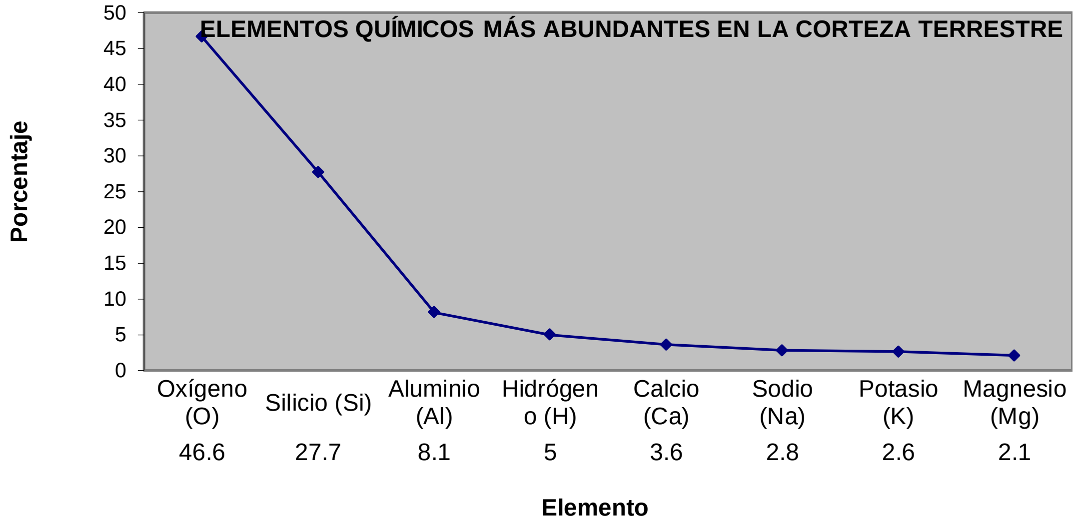
FICHA DE TRABAJO V.1

1. Utiliza hojas de papel milimétrico para construir la gráfica lineal de: *El aumento de la población en la República Mexicana desde 1910*. ¡Pon cuidado al localizar los datos en el eje de las abscisas!

Año	1910	1921	1930	1940	1950	1960	1970	1990	1995	2000
Habitantes (millones)	13.6	14.3	16.6	19.6	25.8	34.9	49.1	81.2	91.2	97.5

Fuentes: Instituto Nacional de Estadística, Geografía e Informática. Censo de Población y Vivienda 2000 y Tabulados Básicos. Tomo I, 2001. Consejo Nacional de Población (Conapo). México, 2002.

2. Revisa con detenimiento la siguiente gráfica lineal y su correspondiente tabla:



La gráfica nos indica que hay una disminución paulatina en el intervalo que va de un elemento químico al siguiente. ¿Qué valores existen entre dos valores consecutivos en el eje de las abscisas? __________ .

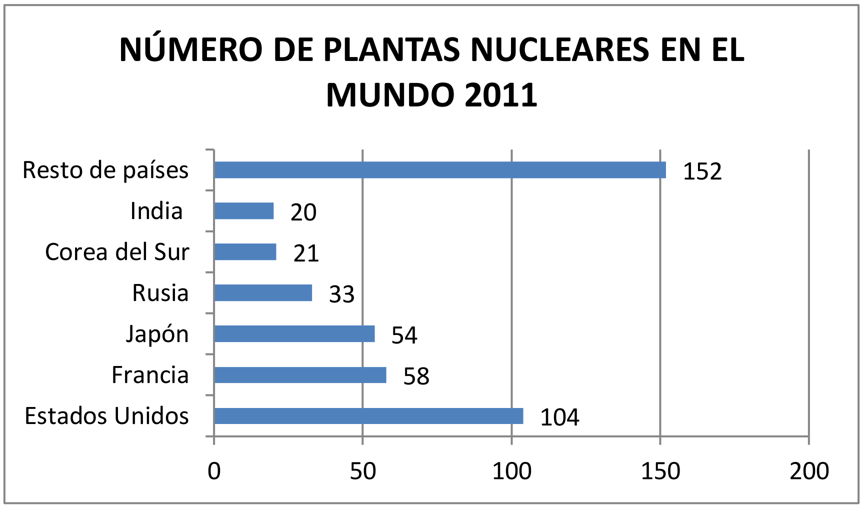
Por lo tanto, ¿es adecuado usar una gráfica lineal al representar este tipo de datos (discretos nominales)? _____ .

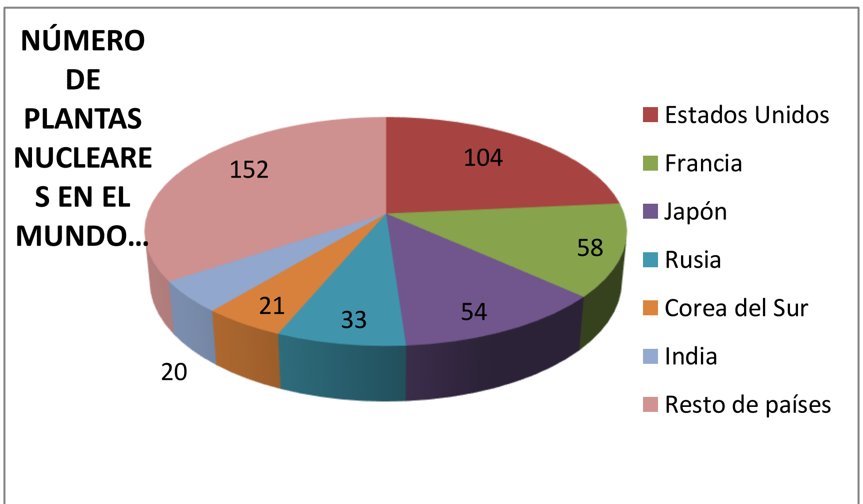
3. Para la tabla de abajo, ¿qué tipo de gráfica sería más conveniente? _______________

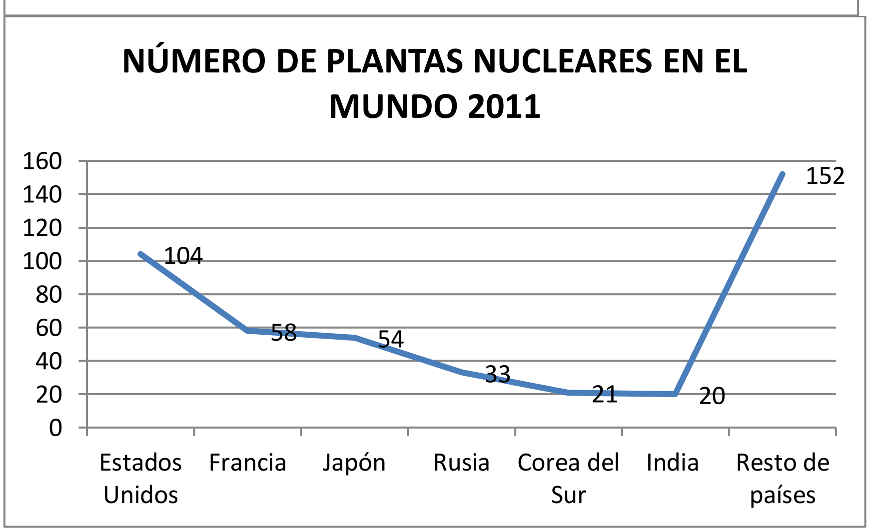
PAÍS	**NÚMERO DE PLANTAS NUCLEARES EN EL MUNDO AÑO 2010**	**PORCENTAJE**
ESTADOS UNIDOS	104	23.53
FRANCIA	58	13.12
JAPÓN	54	12.22
RUSIA	32	7.47
COREA DEL SUR	21	4.75
INDIA	20	4.52
RESTO DE PAÍSES	152	34.39
TOTAL	442	100.0

Fuente: CNN en Español.com. Datos de 2003.

4. La información de la tabla anterior se concentró en tres tipos de gráfica por medio de computadora. Indica por qué conviene, o no, cada una.







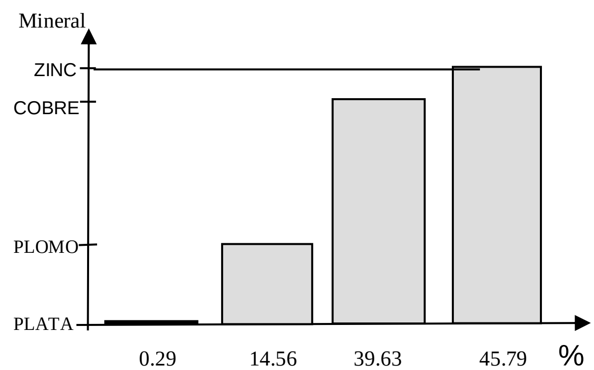
5. Encuentra lo correcto.

MÉXICO 2002 PRINCIPALES MINERALES PRODUCIDOS		
METAL	VALOR (MILLONES DE PESOS)	PORCENTAJE
Plata	3 991.931	0.2947
Plomo	1 334.091	14.5684
Cobre	5 533.593	39.6345
Zinc	3 872.222	45.7971
TOTAL	16 826.258	

Fuente: Dirección General de Minas. Secretaría de Economía.

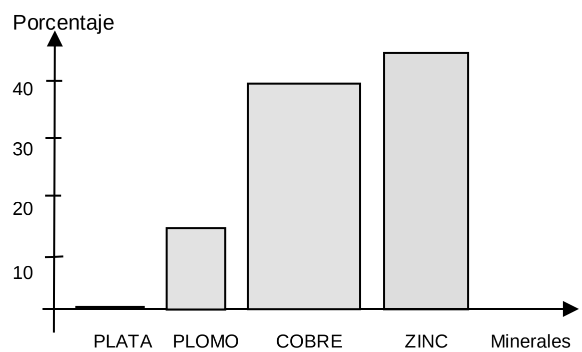
Abajo se muestran las gráficas de barras que cada uno de cuatro analistas obtuvo al representar los datos de la tabla de la izquierda.
En cada gráfica, encierra en un círculo el error cometido.

MÉXICO 2002
PRINCIPALES MINERALES PRODUCIDOS



Gráfica 1

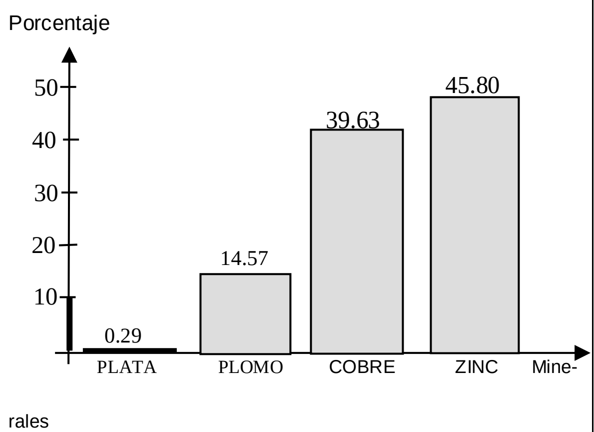
MÉXICO 2002
PRINCIPALES MINERALES PRODUCIDOS



Gráfica 2

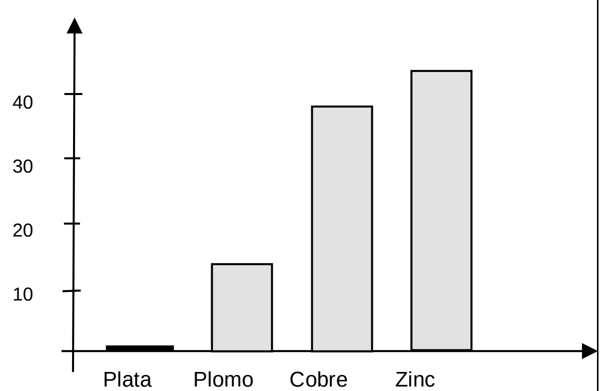
40

MÉXICO 2002
PRINCIPALES MINERALES PRODUCIDOS



Gráfica 3

MÉXICO 2002
PRINCIPALES MINERALES PRODUCIDOS



Gráfica 4

¿Cuál aceptarías como válida? ___________ .

RESUMEN

Las gráficas lineal, de barras y circular son tres formas básicas de representar conjuntos numéricos de datos.

Las gráficas lineal y de barras se construyen trazando dos ejes perpendiculares, por convención la variable dependiente se representa en el eje vertical, la independiente en el horizontal.

En la gráfica lineal se representan los pares ordenados de datos por medio de puntos y éstos se unen por medio de líneas rectas, es decir, se realiza una interpolación (se supone el comportamiento de los datos en puntos intermedios).

La de barras consiste en una serie de rectángulos con alturas equivalentes a los valores de la variable dependiente, cuyo ancho se selecciona de manera arbitraria (aunque debe ser el mismo para todos), con sus bases centradas en los valores de la independiente (marcados sobre el eje horizontal) y separados para indicar el carácter discreto de la variable independiente.

Tanto la gráfica lineal como la de barras se consideran completas si contienen:

- Título donde se haga referencia a las variables representadas.
- Nombre y unidades de las variables en cada eje.
- La escala en cada eje.

6ª SESIÓN
GRÁFICA CIRCULAR

COMPETENCIAS POR ADQUIRIR

- **Seleccionar el tipo más conveniente de gráfica para presentar la información en estudio.**
- **Construir con toda la información pertinente una gráfica circular.**

MATERIAL NECESARIO: ESCUADRAS, TRANSPORTADOR, LÁPIZ, GOMA Y SACAPUNTAS.

APRENDE HACIENDO. . .

Enseguida te ocuparás de aprender cómo representar datos por medio de una gráfica CIRCULAR.

Tu aprendizaje será eficaz si sigues esta técnica:

1° **Estudia con mucha atención** la primera página, cuyo encabezado se marca con ►. Ahí se aborda la información esencial.

2° **Explica mentalmente** el contenido de la primera página. **Sólo después de hacerlo puedes permitirte pasar a la siguiente página.**

3° En las páginas subsecuentes a la primera aparecen espacios vacíos, llénalos con base en lo aprendido. **Está prohibido consultar la página anterior**, a menos que ya hubieses estudiado y resuelto todo el tema.

► GRÁFICA CIRCULAR

LEE CON ATENCIÓN.

La gráfica circular se construye de preferencia cuando la variable, toma menos de nueve valores.

GRUPOS DE ALIMENTOS RECOMENDADOS PARA LA DIETA DEL MEXICANO	
Grupo	**Porcentaje**
Cereales y sus derivados	40
Frutas y verduras	30
Leguminosas y alimentos de origen animal	20
Grasas y aceites	8
Azúcares refinados	2
Total	100

Tabla VI.1

Observa la tabla de la izquierda.

En este caso, la variable independiente es el grupo alimenticio y sus "valores" son los que se enlistan en la primera columna: *Cereales y sus derivados*, *Frutas y verduras*, etc.

Los valores de la variable dependiente (segunda columna) suman 100. Este valor lo hacemos corresponder al número de grados de la circunferencia:

$$100 \to 360°$$

Por **regla de tres directa**, los grados para *Cereales y sus derivados* (40%) son:

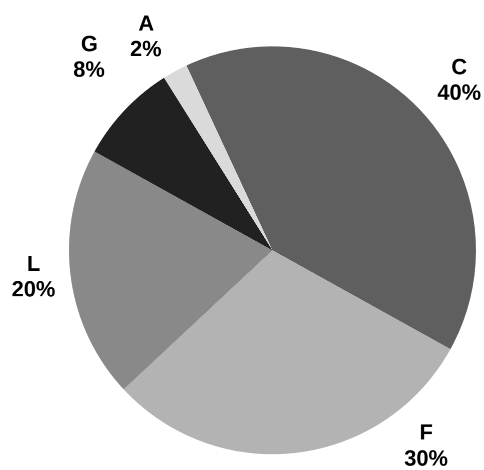
$$\begin{array}{c} 100 \to 360 \\ 40 \to x \end{array} \qquad x = \frac{360\,(40)}{100} \qquad x = 144°$$

Al aplicar la misma regla para *Frutas y verduras* se obtiene x = 108°, para *Leguminosas y alimentos de origen animal* resulta x = 72°; para *Grasas y aceites* x = 28.8° y para *Azúcares refinados* x = 7.2°

Enseguida trazamos una circunferencia de diámetro arbitrario y sobe ella marcamos los grados que corresponden a cada grupo alimenticio, a partir de un punto determinado. Las **áreas** de los **sectores circulares** obtenidos son **proporcionales** a los valores de la **variable**.

Una gráfica circular **completa** debe incluir, además de los sectores circulares, un **título** con el nombre de la variable, los **valores** de ésta o los de la variable dependiente.

ALIMENTOS RECOMENDADOS PARA LA DIETA DEL MEXICANO



C: cereales y sus derivados
F: frutas y verduras
L: leguminosas y alimentos de origen animal
G: grasas y aceites
A: azúcares refinadas

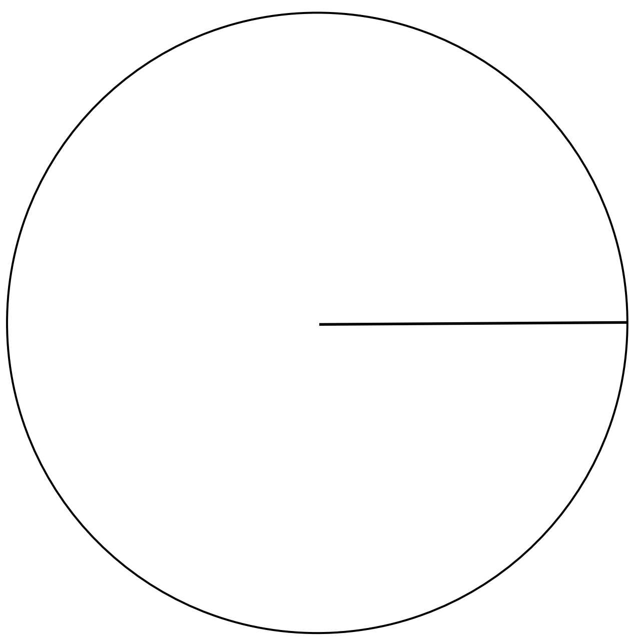
PREPARA TRANSPORTADOR, REGLA, LÁPIZ Y GOMA

Construye la gráfica circular de la siguiente tabla (los valores de la variable dependiente están en la 3ª columna):

Distribución porcentual acumulada de casos de SIDA por categoría de exposición en México al 31 de diciembre de 2002		
	Al 31 de dic	Porcentaje
Desconocido	7,019	10.30
Homosexual	12,368	18.15
Heterosexual	35,776	52.49
Bisexual	12,368	18.15
Otro (transfusión perinatal, drogas intravenosas, hemofílico).	614	0.90
Sumas	68,145	99.99

FUENTE: DGE. Notificación inmediata de casos de SIDA.

CASOS DE SIDA EN ADULTOS POR CATEGORÍA DE EXPOSICIÓN, EN MÉXICO, HASTA EL 31 DE DICIEMBRE DE 2002.



Una gráfica CIRCULAR se considera completa cuando incluye:________________
____

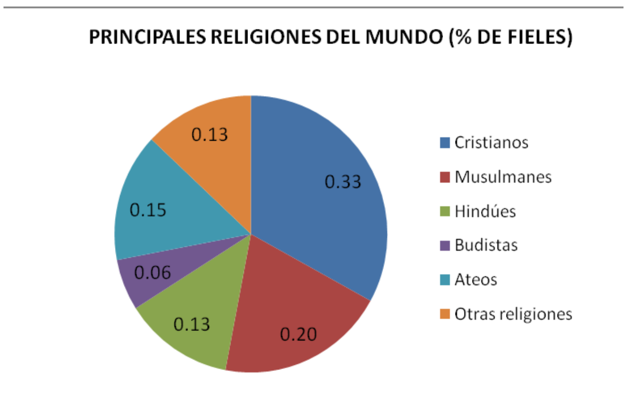
FICHA DE TRABAJO VI.1

1. Las que aparecen abajo son gráficas circulares en las que se han representado los datos de la tabla.

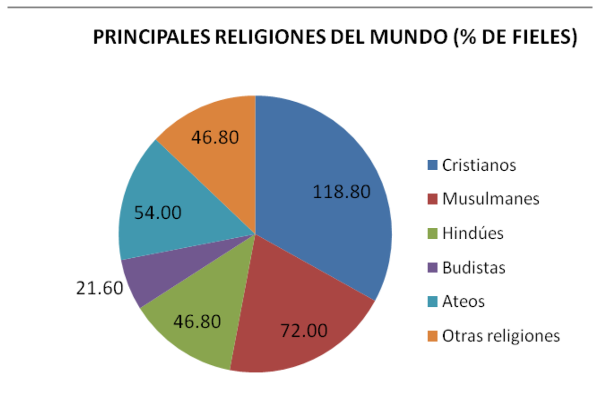
Indica en qué consisten los errores cometidos en ellas y selecciona la mejor por ser la más completa.

PRINCIPALES RELIGIONES DEL MUNDO	
RELIGIÓN	% DE FIELES
Cristianismo	33
Musulmana	20
Hindú	13
Budista	6
Ateos	15
Otra	13
	Total: 100

Fuente: news.bbc.co.uk/low/spanish/misc/newsid_4117000/4117552.stm

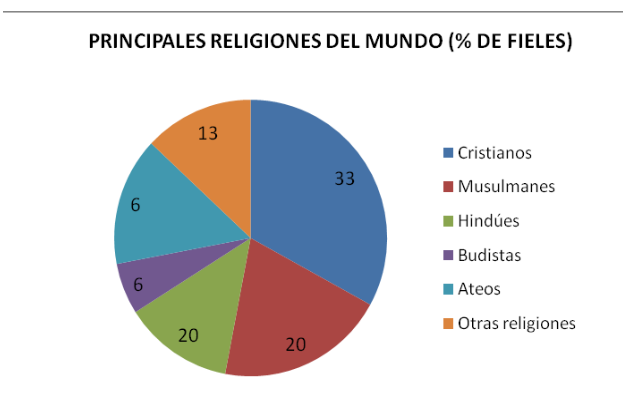


Gráfica 1

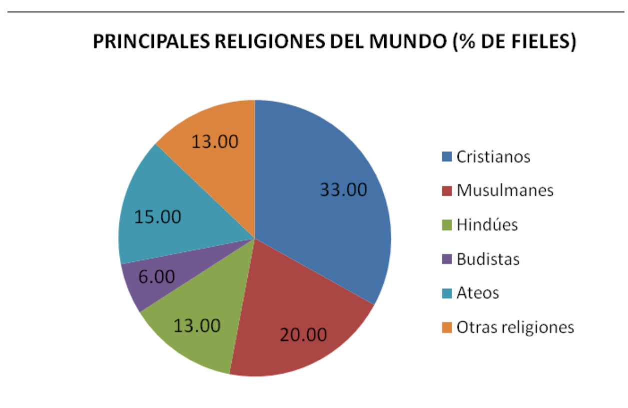


Gráfica 2

–



Gráfica 3



Gráfica 4

2. Decide el tipo de gráfica más adecuado para representar los datos de cada tabla.

A. ____________________

REPÚBLICA MEXICANA CASOS ACUMULADOS DE SIDA EN MARZO DE 2011 (*)	
Chiapas	5 928
Oaxaca	5 091
Puebla	6 858
Distrito Federal	23 811
Guerrero	6 376
Baja California	6 712
Jalisco	11 242
Veracruz	13 633
México	16 392

*Fuente: SS/DGE. Registro Nacional de Casos de SIDA. Procesó: SS/CENSIDA/DIO/SMI. www.aids-**sida**.org/estadist01.html*

B. ____________________

EVALUACIÓN DE LOS PRECIOS MUNDIALES DEL CAFÉ (DÓLARES POR 100 LIBRAS)	
CICLO	PRECIO
1992	63.6
1993	71.6
1994	149.7
1995	149.7
1996	122.2
1997	189.0
1998	135.3
1999	103.9
2000	88.3
2001	63.1

FUENTE: Banco Central de Reserva

C.____________________

SITUACIÓN LEGAL DEL ABORTO EN EL MUNDO EN PORCENTAJES DE LA POBLACIÓN Y NÚMERO DE PAÍSES		
Restricción legal	**Porcentaje de la población mundial**	**Número de países**
Sin restricción alguna (SR)	41.0	49
Permitido para salvar la vida de la mujer (PV)	25.0	52
Permitido por razones socioeconómicas (PS)	20.0	6
Permitido por razones de salud física (PF)	10.0	23
Permitido por razones de salud mental (PM)	4.0	20
Prohibido en todos los casos (NP)	.4	2
	Total: 100.4	Total: 152

Fuente: The Centre for Reproductive Law and Policy, Nueva York, 1998.

TAREA

Investiga cómo es una *gráfica de caja* y para qué sirve (no olvides citar la fuente).

RESUMEN

La gráfica circular también representa la relación entre dos conjuntos de datos (se recomienda en casos donde haya menos de nueve valores). El área total del círculo equivale al 100% de los valores y el área de cada sector a un valor de esa variable.
Al construir la gráfica circular primero se calcula el tamaño de los sectores mediante una regla de tres directa. Una vez determinado el número de grados para cada uno se inicia su trazo en el círculo a partir del radio colocado en una posición arbitraria.
La gráfica circular estará completa si incluye los valores de las variables (pudieran estar escritos en los sectores circulares) y el título donde se haga referencia a las variables representadas.

7ª SESIÓN
MEDIDAS DE TENDENCIA CENTRAL PARA DATOS NO AGRUPADOS

COMPETENCIAS POR ADQUIRIR

Dado un conjunto finito de datos no agrupados:

- **Determinar el valor de la medida más representativa de tendencia central.**
- **Determinar el valor de la medida de dispersión más apropiada.**

MATERIAL NECESARIO: LÁPIZ, GOMA, REGLA Y SACAPUNTAS.

MEDIDAS DE TENDENCIA CENTRAL

Muchas veces habrás escuchado expresiones donde se utiliza el término *promedio.*

En Estadística los promedios se denominan medidas de tendencia central. Las más conocidas y utilizadas son: la moda, la media aritmética y la mediana.

NOTACIÓN SIGMA…

La suma de los diez primeros números naturales la indicamos así:

$$1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 +9 + 10$$

Para **abreviar**sumas como la anterior se utiliza la letra griega **Σ** (sigma mayúscula). La forma de hacerlo consiste en representar cada número natural con la letra *n* y escribir:

$$\sum_{n=1}^{n=10} n$$ **Se lee: suma de *n*, desde n = 1 hasta *n* = 10**

Imagina un conjunto de 500 datos numéricos. Si representas un dato por medio de x_1, a otro por medio de x_2, a otro por x_3,... y así sucesivamente, hasta representar al último por medio de x_{500}, la suma quedaría indicada así:

$$x_1 + x_2 + x_3 + ... + x_{500}$$ ¿Qué varía en la notación de un sumando a otro?

Esa suma se puede abreviar aún más al fijarnos en lo que varía de un sumando a otro, el subíndice, al cual denotaremos con la letra ***j*** (subíndice mutante):

$$\sum_{j=1}^{j=500} x_j$$ **Se lee: suma de x_j, desde j =1 hasta j = 500**

¿Cómo se abrevia $y_1 + y_2 + y_3 + …. + y_{105}$? ____________ ¿Cómo se debe leer? ______________________

____

MEDIDAS DE TENDENCIA CENTRAL

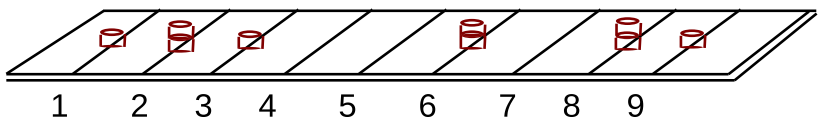
MEDIA ARITMÉTICA

"La Perica González" es entrenadora del equipo de basquetbol *Maravillas* y revisa los puntos anotados por las 9 jugadoras que participaron en el último partido:

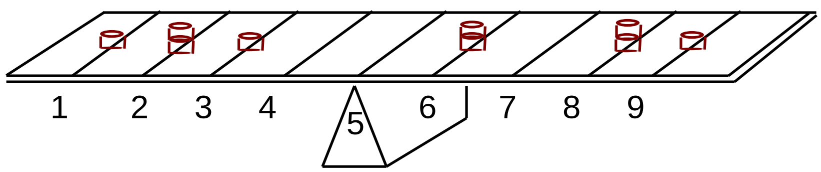
JUGADORA	Ilia Parres	Rosario Paz	Chela Díaz	Luisa Pineda	Esther Urgelles	Lupita Ramos	Dora Ponce	Olivia Suárez	Irene Ron
PUNTOS ANOTADOS	1	2	2	3	6	6	8	8	9

Tabla VII.1

Para tener una idea del promedio, "la Perica González" construye una tabla graduada de 0 a 10 unidades; enseguida, por cada jugadora coloca sobre la tabla un peso de un kilogramo en el puntaje logrado; de esta manera:



Después toma una base triangular (fulcro) y busca el punto donde los pesos queden en **equilibrio**. Ese punto resulta ser 5:



Hay **equilibrio** en el punto 5 porque los datos a la izquierda ejercen una fuerza hacia abajo (brazo de palanca) exactamente igual a la ejercida por los datos de la derecha.

"La Perica González" piensa que 5 fue el promedio de anotaciones de su equipo. Si a ti te hubieran preguntado cuál fue el promedio de puntos anotados, ¿qué hubieras hecho?

Lo que comúnmente hacemos es sumar los datos y dividir entre su número. Hazlo mentalmente a partir de la tabla 4.1 y escribe el resultado: _____.

Ambos, el promedio que obtuvo "la Perica González" y el tuyo, se conocen como **media aritmética**. A continuación descubre algunas de sus características, lee con cuidado, piensa y marca con ✓ las afirmaciones verdaderas y con ✗ las falsas **(Idea: ubica en tu mente la fórmula para calcular la media aritmética, de ahí surgen varias de las respuestas)**:

- Se calcula sumando todos los datos del conjunto y dividiendo lo que resulte entre el número de datos. ____
- El valor de cada dato influye en el resultado. ____

- En el partido anterior del equipo Maravillas la media aritmética fue 7 y tuvieron acción las mismas 9 jugadoras; esto significa que el total de puntos anotados fue 47. ____
- En su primer partido el equipo anotó 56 puntos y la media aritmética fue 8, lo cual quiere decir que participaron 7 jugadoras. ____
- Si todas y cada una de las jugadoras en el próximo partido anota 3 puntos más de los que anotó en esta ocasión, el valor de la media aritmética seguirá siendo 5. ____
- Las jugadoras del equipo contrario anotaron en promedio 7 puntos, 2 más que el promedio alcanzado por el equipo Maravillas, lo cual significa que ganaron el partido. ____
- Si todas y cada una de las jugadoras en el próximo partido anota un punto menos de los que anotó en esta ocasión, el valor de la media aritmética seguirá siendo 5. ____
- En otro partido anterior del equipo Maravillas, 8 de sus jugadoras tuvieron acción y la media aritmética de sus anotaciones fue 7. Del equipo contrario participaron también 8 jugadoras, pero su media aritmética fue 7.4, esto significa que las pupilas de "la Perica González" perdieron el partido. ____
- Si en el próximo partido todas y cada una de las jugadoras anota el doble de los puntos que anotó en esta ocasión, el valor de la media aritmética seguirá siendo 5. ____
- En cada partido futuro "La Perica González" quiere premiar a las jugadoras que logren puntos por encima de la media; por lo tanto, en cada partido deberá premiar a la mitad de las jugadoras que hayan anotado. ____
- Cuando las anotaciones se ordenan de menor a mayor, el valor que queda exactamente en medio coincide con la media aritmética. ____
- El valor de la media aritmética nunca coincidirá con uno de los datos del conjunto de anotaciones. ____
- Tricia Coss estuvo "en la banca" durante todo el partido, es decir, anotó 0 puntos. Si este valor se agrega a los puntajes de la tabla 4.1, el valor 5 de la media no se altera. ____

Problema. El Dr. Salazar registró los valores (derecha) de las temperaturas corporales (en °C) de 12 pacientes que acudieron a su consultorio por malestares estomacales. Quiere encontrar un valor representativo de esos valores, ¿qué debe hacer?

37	38	40	39
39	33	41	36
34	41	39	38

Decide organizar los datos en una tabla de frecuencias (tabla 4.2).

TEMPERATURAS CORPORALES DE LOS PACIENTES	
Temperatura (°C)	**Frecuencia**
33	1
34	1
35	0
36	1
37	1
38	2
39	3
40	1
41	2
	Total:

Tabla VII.2

El Dr. Salazar piensa que el valor ubicado **en medio**, o sea 37, puede ser el valor representativo del conjunto; pero de inmediato se percata de que tampoco varía cuando se quitan o se agregan ciertos valores; por ejemplo, si se agregan dos datos de valor 35, el valor central sigue siendo el mismo; además, es insensible a que por arriba de él hay menos datos que por abajo.

También repara en que 39 es el valor que **más se repite**, el de mayor frecuencia, por lo cual podría considerarlo como representativo; sin embargo, es un valor que no se alteraría si al conjunto se le agregaran otros datos, por ejemplo 37, 35, 40, etc.; por lo tanto, ya no resulta tan representativo.

Insatisfecho, finalmente, el Dr. Salazar concluye que para este caso sería más apropiado utilizar la **media aritmética**.

La media aritmética se obtiene al sumar los datos del conjunto y dividir el resultado entre el número de datos.

La temperatura corporal media, o media aritmética de las temperaturas corporales que registró el Dr. Salazar (representada con $\bar{t}$) la calcula así:

$$\bar{t} = \frac{37+38+40+39+39+33+41+36+34+41+39+38}{12}$$

Es decir: $\bar{t} = \frac{455}{12}$

$$\bar{t} = 37.916$$

Piensa en un conjunto de *n* datos (*n* representa un número natural) cuyos valores desconocemos. En tal caso los datos se pueden simbolizar por medio de una letra, *x* por ejemplo, con subíndices progresivos que sirvan para numerarlos de 1 a *n*:

$$x_1, x_2, x_3, \ldots, x_n$$

En estas circunstancias, la media aritmética del conjunto queda expresada como:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} \text{ o brevemente así: } \bar{x} = \frac{\sum_{i=1}^{i=n} x_i}{n}$$

Piensa en un conjunto de *k* datos cualesquiera (o sea, que *k* representa un número natural). Encierra la expresión para la media aritmética del conjunto:

A. $\bar{x} = \frac{\sum_{i=1}^{i=n} x_i}{k}$ B. $\bar{x} = \frac{\sum_{i=1}^{i=k} a_i}{k}$

A la media aritmética se le denomina también *centro de gravedad* de la distribución.

MEDIANA

Problema. A un electricista le piden instalar luz de intensidad *moderada* en una casa. Las lámparas que encuentra en la tlapalería tienen estas potencias en vatios:

100 75 40 60 110 20 120 30 200

Para decidir qué lámparas comprar, organiza los valores de menor a mayor:

20 30 40 60 **75** 100 110 120 200

Decide por lo tanto comprar lámparas de 75 vatios.

En un conjunto finito de datos, ordenados de menor a mayor, la mediana es el valor central

Lo anterior implica que la mitad de los datos tienen un valor menor o igual a su **mediana** y la otra mitad un valor mayor o igual a ésta.

Con la finalidad de localizar el lugar que ocupa la mediana y calcular su valor en un conjunto de muchos datos, se distingue: cuando el número de datos es impar y cuando es par. En cualquier caso, **primero** deben **ordenarse** los datos en forma creciente o decreciente.

Podemos encontrar la mediana incluso cuando nuestros datos son descripciones cualitativas, en lugar de números.

Aprendamos cómo se localiza el lugar que ocupa y cómo se calcula su valor para un conjunto relativamente grande, que contenga un número **impar** de datos.

Comenzamos con pocos datos: Este conjunto corresponde a las edades de 7 hermanos (número impar):

9 11 13 10 12 6 15

Ordenándolos de menor a mayor:

6 9 10 **11** 12 13 15

Inmediatamente nos damos cuenta de que la mediana es el dato situado en el cuarto lugar. Este lugar pudo haberse calculado dividiendo el número de datos (7) más uno, entre dos:

$$\frac{7+1}{2} = 4 \text{ (cuarto lugar)} \Rightarrow \text{mediana} = 11$$

Si tuviésemos, por ejemplo, un conjunto con 81 datos (número impar) **ordenados** según su valor, para calcular el lugar de la mediana procederíamos de la misma forma:

$$\frac{81+1}{2} = 41 \text{ (cuadragésimo primer } lugar)$$

Es decir, aun cuando el número de datos sea muy grande, tenemos una manera sencilla de determinar el lugar donde se encuentra la mediana.

Piensa en un conjunto de 1 875 datos. ¿Qué lugar ocupa la mediana? ______ .

Estudiemos la manera de localizar el lugar que ocupa y cómo se calcula su valor para un conjunto con un número **par** de datos.

Los siguientes datos son los salarios por hora, en pesos, de un grupo de seis empleados:

35 72 43 37 43 42

Ordenándolos:

35 37 42 43 43 47

La mediana se encuentra entre los salarios de los lugares 3º y 4º, es decir, entre 42 y 43.

Dicho en otra forma, la mediana se localiza entre el lugar dado por el número de datos entre dos ($\frac{6}{2} = 3°$) y el siguiente (el 4º).

El valor de la media aritmética de esos dos valores intermedios es la mediana:

$$\text{Mediana} = \frac{42+43}{2} = 42.5$$

Puedes advertir que para una cantidad **par** de datos, la mediana no necesariamente coincide con un dato; pero si la cantidad es **impar** la mediana sí coincide seguramente con uno de los datos.

Inventa algunos datos (un número par) de tal manera que la mediana coincida con uno de ellos:

____ .

La precipitación pluvial en Comitán, Chiapas, fue registrada, en milímetros, durante 120 días consecutivos. ¿Cómo se calcula el lugar que ocupa la mediana?

Primero: ______________________________

¿Entre cuáles datos se localiza la mediana? ________________________________.

Imagina un conjunto de 752 datos. Al ordenarlos, ¿qué lugar ocupa la mediana? _______.

MODA

Problema: Nestor Gómez trabaja en una tienda de ropa. Él quisiera independizarse y vender por su cuenta, pero quiere saber de qué talla le convendría hacer su primer pedido al proveedor. Con esa idea, durante un mes registra las tallas de las camisas que compran los clientes y las organiza en una tabla de frecuencias (abajo). ¿Cuál talla le conviene tomar como representativa?

Si selecciona la mediana, que es aproximadamente igual a 40, se da cuenta de que sus ventas no serían precisamente las mejores.

Por lo tanto, decide calcular la media aritmética y le resulta 38.66

TALLA	CAMISAS VENDIDAS
34	2
36	16
38	17
40	11
42	6
44	4
46	1
	Total:

Tabla VII.3

¿Debería pedir al sastre esa talla de camisas? ¿Las vendería más que las de tallas tradicionales? ¡Por supuesto que no!

¿Cuál es entonces la talla que debe encargar para su primer pedido? ¡Cierto!, la talla 38, es decir, la más frecuente.

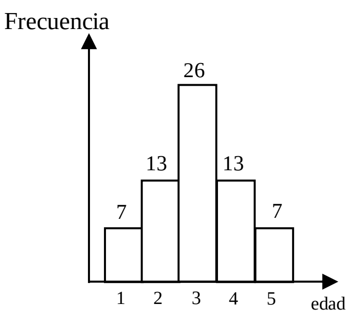
El dato con la mayor frecuencia en un conjunto es su moda..

Puede suceder que un conjunto tenga dos modas, o sea, dos valores con la máxima frecuencia; entonces se denomina **bimodal**. En general, si hay más de una moda el conjunto es **multimodal**.

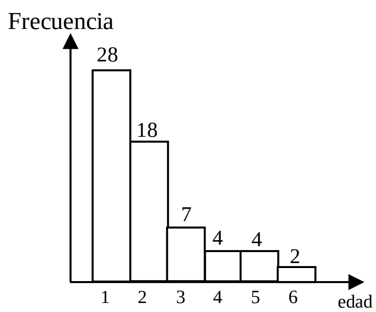
La moda siempre es un dato del conjunto; sin embargo, podemos encontrar conjuntos **sin moda**. Ejemplo: El conjunto: 1, 3, 9, 6, 7 es amodal; porque no tiene _________ .

UBICACIÓN DE MEDIA, MEDIANA Y MODA EN UN HISTOGRAMA

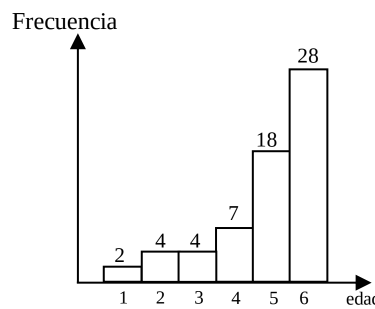
En los histogramas de abajo se representan las edades, en años cumplidos, de un grupo de niños. Indica con líneas verticales las posiciones de la media, moda y mediana.



I



II



III

Indica el histograma donde…La moda es mayor que la mediana _____.

La moda es menor que la mediana _____.

La variable es continua, simétrica y unimodal _____.

Coinciden la media, la mediana y la moda _____.

La media se localiza entre la moda y la mediana ___________.

FICHA DE TRABAJO VII.1

1. Expresa en forma abreviada la media aritmética del conjunto $e_1, e_2, e_3, \dots, e_m$

2. El número de hijos que dijeron tener 50 padres entrevistados en la colonia Miguel Hidalgo del estado de Guanajuato fue:

3 1 1 2 0 3 5 3 1 1 1 2 7
5 6 0 5 4 6 6 3 4 1 6 2 2
2 7 2 0 3 5 5 4 5 4 4 5
0 5 6 4 6 2 2 2 7 6 5 4

Construye la tabla de frecuencias simple, sobre la tabla calcula la media, la moda y la mediana y decide qué valor indicaría mejor el número típico de hijos que tienen esos padres. *Respuesta: el mejor valor lo da la ___________________ que es igual a _____.*

3. En el campeonato mundial de atletismo celebrado en Francia (agosto de 2003), en 20 km de marcha, las posiciones de los 10 competidores que llegaron primero a la meta fueron las que se aprecian a la derecha.

Sin hacer cálculos, ¿quiénes hicieron un mejor tiempo promedio, los españoles o los mexicanos? _______________

Compruébalo.

1. Jefferson Pérez (ECU) 1h 17.21 RM ORO
2. 'Paquillo' Fernández (ESP) 1h 18.00 PLATA
3. Roman Rasskazov (RUS) 1h 18.07 BRONCE
4. Noe Hernández (MÉX) 1h 18.14
5. Luke Adams (AUS) 1h 19.35
6. Ivan Trotskiy (BLR) 1h 19.40
7. David Márquez (ESP) 1h 19.46
8. Ilya Markov (RUS) 1h 20.14
9. José David Domínguez (ESP) 1h 20.15
10. Alejandro López (MÉX) 1h 20.24

Tabla VII.4

4. Estas son las edades, en años, a las que se titularon 12 estudiantes de veterinaria:

22, 23, 23, 23, 23, 23, 23, 23, 26, 28, 38, 51

¿Cuál es la edad más representativa de la titulación de los estudiantes? _____

5. Si en la serie datos:

11 16 13 17 11 13 38

se cambia el 38 por 50, ¿cuál de las medidas (media, moda y mediana) se ve afectada? ______________________.

6. El director de una escuela, entrevista a un candidato a profesor. Le ofrece \$10 000.00 mensuales, pero le promete que sólo será por un período de prueba, ya que luego su sueldo aumentará. Aquí se gana bien –le dice–, el salario promedio que tenemos los profesores es de \$30 000.00 mensuales. Después de un mes, el profesor vuelve a ver al director y le dice: Usted me

mintió. Les pregunté a otros profesores y ninguno gana más de $30 000.00 a la quincena. ¿Por qué me dijo que el salario promedio era de $30 000.00? El director le responde: No le mentí. Tome usted mismo la nómina y calcule: Yo gano $155 000.00; el subdirector $100 000.00; los nueve profesores de nivel "A" $10 000.00 cada uno; el profesor de nivel "B" $15 000.00 y los ocho profesores de nivel "C" $30 000.00 cada uno.
La nómina quincenal suma $600 000.00, y como somos 20 personas, el promedio de los salarios es de $30 000.00 ¿O me equivoco?

¿Quién tiene razón, el director o el profesor? ¿Por qué? ___________________________

____

____

7. En una investigación estadística se obtuvieron los siguientes datos acerca de las preferencias televisivas de los adultos:

PREFERENCIA	**NÚM. DE ADULTOS**
Películas	41 700
Noticieros	37 040
Musicales	12 548
Teleseries	40 710
Deportivos	23 006
Otro tipo	9 879

¿Cuál es la moda de la muestra? _____________

¿Tiene sentido calcular la media de la muestra?

____

¿Por qué? ____

____

RESUMEN

Las tres medidas más utilizadas para obtener un valor representativo de un conjunto de datos son: media aritmética, mediana y moda.
Ante un conjunto de valores, conviene reflexionar acerca del tipo de promedio más conveniente.
La media aritmética se obtiene al dividir la suma de los datos entre la cantidad que haya de los mismos en el conjunto en consideración.
La mediana se puede determinar siempre y cuando se organicen los valores, ya sea en orden creciente o decreciente.
Cuando se trata de pocos datos la mediana se encuentra de inmediato localizando el valor central; en caso de no existir éste, se obtiene tomando los dos valores centrales, sumándolos y dividiendo el resultado entre dos.
En conjuntos relativamente grandes se puede calcular el lugar que ocupa la mediana: si el número de datos es impar, basta con agregarle uno a ese número y dividir el resultado entre dos; si es par se divide el número de datos entre dos, la mediana se encontrará entre el lugar indicado por el número que resulte y su consecutivo.
La moda es el valor más frecuente en un conjunto. Se llegan a encontrar colecciones de datos sin moda (amodales), con dos (bimodales) o más de una (multimodales).

8ª SESIÓN
MEDIDAS DE TENDENCIA CENTRAL EN DATOS AGRUPADOS

COMPETENCIAS POR ADQUIRIR
- **Calcular la media ponderada.**
- **Encontrar los valores de la media aritmética, moda y mediana a partir de:**
  Una tabla de frecuencias simple.
  Una tabla de frecuencias por intervalos.

MEDIDAS DE TENDENCIA CENTRAL (DATOS AGRUPADOS)

MEDIA ARITMÉTICA

Problema: Salvador Nava es el jefe de Mantenimiento de la empresa "Jair S.A." y quiere obtener el voltaje promedio de salida en los contactos eléctricos de sus edificios. Con ese fin hace las mediciones, las redondea a números enteros y las organiza en la tabla VIII.1

VOLTAJE DE SALIDA DE LOS 75 CONTACTOS ELÉCTRICOS DE LOS EDIFICIOS DE LA EMPRESA JAIR S. A.									
Voltaje	110	112	115	117	123	125	136	141	145
Frecuencia	2	6	9	6	10	21	9	7	5

Tabla VIII.1

A partir de su tabla ¿cómo puede Salvador calcular el promedio que quiere?

El dato 110 se repite 2 veces, el dato 112 se repite 6 veces, etc.; es decir, se conocen los datos y cuánto "*pesan"* en la tabla. El más pesado es 125, porque es el que más se repite; ¿qué lugar ocupa la mediana? ______; ¿cuál es su valor? _____.

Pero a Salvador le interesa un promedio que sea "sensible" a todos los datos, por lo cual decide calcular la **media aritmética**. Para obtenerla de la Tabla 4.5, multiplica cada dato por su frecuencia, suma los resultados y, finalmente, el total lo divide entre el número de datos (éste lo obtiene sumando las frecuencias):

$$\bar{x} = \frac{110(2)+112(6)+115(9)+117(6)+123(10)+125(21)+136(9)+141(7)+145(5)}{2+6+9+6+10+21+9+7+5}$$

$$\bar{x} = \frac{9\,420}{75} \quad \rightarrow \quad \bar{x} = 125.6 \text{ voltios}$$

EL VALOR EXACTO DE LA MEDIA SE CALCULA MULTIPLICANDO CADA UNO DE LOS DATOS POR SU FRECUENCIA, SUMANDO LOS PRODUCTOS Y DIVIDIENDO EL RESULTADO ENTRE EL NÚMERO DE DATOS (LA SUMA DE LAS FRECUENCIAS).

CONSUMO DIARIO DE PROTEÍNAS DE 125 ADULTOS DE COENEO MICHOACÁN		
Proteínas (g)	**No. de personas**	**Marcas de clase**
0−15	13	7.5
16−31	15	23.5
32−47	65	39.5
48−63	20	55.5
64−79	12	71.5

Tabla VIII.2

Problema. En una clínica del Instituto Mexicano del Seguro Social, la dietista Camerina Cortez encuentra la tabla VIII.2.

Para saber cuál es el valor representativo del consumo proteínico, Camerina decide calcular la media aritmética. Pero el hecho de no conocer individualmente los datos le representa un obstáculo, ¿cómo lo puede resolver?

En el primer intervalo hay 13 datos; como no los conoce, los supone del mismo valor e iguales al valor medio del intervalo, es decir, a su marca de clase. Entonces, los 13 primeros datos son 7.5; los 15 datos del segundo intervalo son 23.5, y así sucesivamente.

En este supuesto, los números de la tercera columna son los “datos” que se repiten tantas veces como lo indica el valor de su frecuencia (segunda columna).

Ahora se puede calcular la media aritmética $\bar{x}$ multiplicando cada dato supuesto (marca de clase) por su frecuencia, sumando los resultados y dividiendo entre el total de datos:

$$\bar{x} = \frac{13(7.5)+15(23.5)+65(39.5)+20(55.5)+12(71.5)}{125}$$

$$\bar{x} = 39.884g$$

Así, Camerina obtuvo un valor **aproximado** de la media aritmética. No obstante, sería válido decir que en Coeneo se consumen en promedio 39.884 g. diarios de proteína.

PARA CALCULAR LA MEDIA ARITMÉTICA A PARTIR DE UNA TABLA DE FRECUENCIAS POR INTERVALOS, SE SUPONE QUE “TODOS LOS DATOS DE UN INTERVALO DETERMINADO TIENEN UN VALOR IGUAL A SU MARCA DE CLASE”.

MEDIANA

Problema. Ana Pérez aplica un examen de lectura en inglés a un grupo de aspirantes a ser contratados como traductores; organiza los resultados en la tabla VIII.3 y la entrega al gerente de recursos humanos.

El gerente quiere contratar primero a una persona que lea un número de palabras por minuto ni muy bajo ni muy alto. ¿Qué debe hacer?

El número total de datos (85) es

PALABRAS LEÍDAS EN UN MINUTO POR 85 ASPIRANTES A UN EMPLEO DE TRADUCTOR		
Palabras	**Aspirantes**	**Marcas de clase**
82−96	4	89
97−111	8	104
112−126	11	119
127−141	42	134
142−156	15	149
157−171	5	164
	Total: 85	

Tabla VIII.3

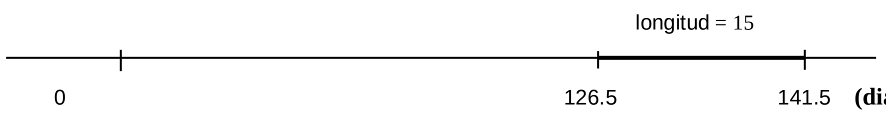
impar, por lo cual la mediana ocupa el lugar _____. Es decir, se localiza en el 4º intervalo, al que llamaremos “intervalo de la mediana”.

A partir de la tabla 4.7, ¿puede saber el gerente cuáles son los valores de los 42 datos que están en el intervalo donde se localiza la mediana? Evidentemente no, por lo tanto resulta imposible obtener con seguridad un valor exacto de la mediana, pero se puede calcular un valor **aproximado** a partir de esta hipótesis:

“A LO LARGO DE UN INTERVALO REAL DE CLASE LOS DATOS COMPRENDIDOS EN ÉL ESTÁN DISTRIBUIDOS UNIFORMEMENTE”

En el 4º intervalo (el de la mediana) hay 42 datos. Éstos se suponen distribuidos uniformemente a partir del 24º lugar (puesto que en los intervalos anteriores hay en total 23 datos).

La distancia del punto cero al origen del intervalo de la mediana está dada por el límite real inferior del intervalo de la mediana: 126.5

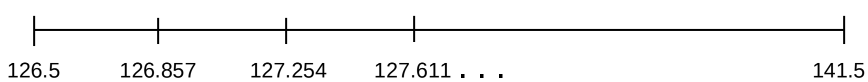


(diagrama sin escala)

El tamaño del intervalo de la mediana es: 141.5–126.5 = 15; para conocer la distancia entre dos datos consecutivos del intervalo de la mediana, se divide la longitud de ese intervalo entre el número de datos que contiene:

$$\frac{15}{42} = 0.357$$

O sea, la distancia entre dos datos consecutivos del intervalo de la mediana es 0.357:



(diagrama sin escala)

El valor de la mediana (43º dato, o 20º en el 4º intervalo) se obtiene agregándole 20 veces el valor 0.357 al límite real inferior del intervalo.

$$126.5 + 20(0.357) = 133.0643$$

MODA

En una Tabla de Frecuencias simple el valor **exacto** de la moda es el dato con mayor frecuencia. En una Tabla de Frecuencias <u>por Intervalos</u>:

LA MODA SE SUPONE IGUAL A LA MARCA DE CLASE DEL INTERVALO DE MAYOR FRECUENCIA

Es decir, sólo se puede conocer un valor **aproximado.**

En la tabla 4.7 anterior, la moda es 134, por ser éste la marca de clase del intervalo con mayor frecuencia.

FICHA DE TRABAJO VIII.1

RESUELVE APARTE

1. Observa la tabla VIII.4

Se nota esto: el dato x_1 se repite f_1 veces; el dato x_2 se repite f_2 veces, y así sucesivamente hasta el dato x_n, el cual se repite f_n veces.

La suma de las frecuencias es igual al total de datos.

Escribe la fórmula para calcular el valor **exacto** de la media aritmética $\bar{x}$.

Datos	Frecuencias
x_1	f_1
x_2	f_2
x_3	f_3
.	.
.	.
.	.
x_n	f_n

Tabla VIII.4

2. Calcula la media aritmética, moda y mediana del número de piezas en la tabla VIII.5. ¿Cuál es la medida más apropiada? ¿Por qué?

PIEZAS DENTALES EN NIÑOS CON EDADES ENTRE 12 Y 24 MESES									
Núm. de piezas	12	13	14	15	16	17	18	19	20
Núm. de niños	1	1	2	11	30	25	22	7	1

Tabla VIII.5

3. Resuelve mentalmente: *Diez estudiantes presentan un examen de Física, 4 mujeres y 6 hombres. El promedio de las calificaciones de las mujeres es de 6 y el de los hombres es de 8. ¿Cuál es el promedio de los 10 estudiantes?*

4. Resuelve mentalmente: *El profesor de Historia les dice a sus alumnos: Para la calificación final del curso, el peso de sus calificaciones en exámenes será 50%, el peso de las tareas 20% y el de sus exposiciones 30%. ¿Cuál es la calificación final de Aurora si obtuvo 6 en exámenes, 5 en tareas y 10 en exposiciones?*

5. Analiza la tabla VIII.6

DIÁMETROS DE LA CABEZA DE 75 NIÑOS	
Diámetro en cm	**No. de niños**
44.45 – 44.95	2
44.95 – 45.15	3
45.15 – 45.65	6
45.65 – 46.15	8
46.15 – 46.65	15
46.65 – 47.15	26
47.15 – 47.65	8
47.65 – 48.15	5
48.15 – 48.65	2

Tabla VIII.6

- ¿Cuál es el total de datos de la tabla?
- ¿En qué lugar se localiza la mediana?
- ¿Cuál es el intervalo de la mediana?
- ¿Cuántos datos hay en el intervalo de la mediana?
- ¿Cómo se supone que están distribuidos?
- ¿Qué lugar ocupa el primer dato del intervalo donde está la mediana?
- ¿En el intervalo de la mediana, cuál es la distancia desde el origen de ese intervalo al primer dato y entre datos consecutivos?
- ¿Cuál es el valor de la mediana?

TAREA

1. Decir “la media” es una manera ordinaria de referirse a la media aritmética. Además de ésta, existen otros dos tipos de media que reciben los nombres de *media geométrica* y *media armónica*. Investiga en qué casos se utilizan y la fórmula para calcular cada una. No olvides anotar la referencia.

2. Investiga la fórmula para obtener la mediana cuando los datos están agrupados en intervalos de clase. No olvides anotar la referencia.

3. Investiga la manera de obtener la moda cuando los datos están agrupados en intervalos de clase. No olvides anotar la referencia.

RESUMEN

Las medidas de tendencia central se pueden calcular a partir de tablas de frecuencias.

Tratándose de una tabla de frecuencias simples, los cálculos se pueden hacer con sólo tomar en cuenta que los datos se repiten tantas veces como su frecuencia lo indica.

En cambio, cuando la tabla de frecuencias por intervalos, los cálculos se hacen apoyándose en dos hipótesis debido al desconocimiento que en ese caso se tiene de los datos (sólo se sabe cuántos hay en cada intervalo). Las hipótesis son:

- Para calcular los valores de la media y la moda se supone que los valores de los datos en el intervalo correspondiente son las marcas de clase.
- La mediana se calcula suponiendo que, por intervalo, los datos se distribuyen uniformemente.

9ª SESIÓN
MEDIDAS DE DISPERSIÓN PARA DATOS NO AGRUPADOS

COMPETENCIAS POR ADQUIRIR

- Ante un problema dado calcularás el rango de los valores de la variable, su desviación media y su desviación típica.
- Definirás el concepto, construirás su fórmula y explicarás el significado de la desviación media y de la desviación estándar.
- Identificarás la relación entre desviación típica y varianza.

MATERIAL NECESARIO: CALCULADORA, LÁPIZ, SACAPUNTAS Y GOMA.

MEDIDAS DE DISPERSIÓN PARA DATOS NO AGRUPADOS

Las medidas de dispersión, como su nombre lo sugiere, dan idea de la variabilidad de los datos de un conjunto; dicho de otra manera, son indicadores de su grado de diferenciación.

Esta lección está dedicada al estudio de tres medidas de dispersión, su significado y las maneras de calcularlas.

En primer lugar se recuerda el rango (ya se utilizó con anterioridad), continuamos con la desviación media y terminamos con la desviación típica o estándar.

LA VARIACIÓN SE PUEDE EXPRESAR CON NÚMEROS. . .

El grado de variabilidad de un conjunto de datos es una información de interés. Al estudiar un fenómeno nos indica qué tan diferentes son los valores de la variable.

Entre más parecidos sean dichos valores, menos "dispersión" mostrarán y viceversa.

La dispersión es una característica de la información y, aun cuando hay varias maneras de obtener un valor numérico de la misma, la más utilizada por los científicos es la desviación estándar, también conocida como típica.

RANGO

Problema: En Tlacotalpan, Ver., la bióloga Alicia Rodríguez mide la longitud de 7 insectos voladores llamados *caballitos del diablo*, y obtiene estos valores en centímetros:

7, 8, 8, 12, 10, 5, 13

Para apreciar mejor la distribución de los datos Alicia consigue una tabla de madera, en uno de sus cantos le marca una escala de 4 a 14, encima de los valores correspondientes a los datos coloca pesos unitarios y todo sobre un fulcro en el punto de equilibrio (que corresponde al valor de la media aritmética de los datos):

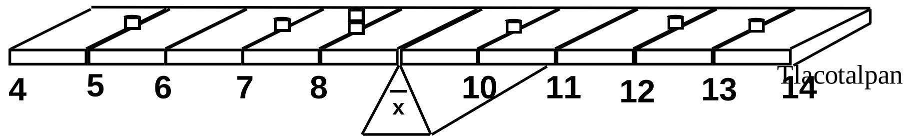


Figura IX.1

Alicia se interesa por saber **qué tan dispersos están los datos**. Para ello toma el mayor y el menor y calcula su diferencia:

13−5 = 8   Este valor se conoce como el **rango** de los datos

Posteriormente, en Alvarado, Ver., realiza otras 7 mediciones de la longitud del mismo insecto y obtiene:

12, 12, 13, 12, 12, 12, 4

Hace lo mismo que para el primer conjunto:

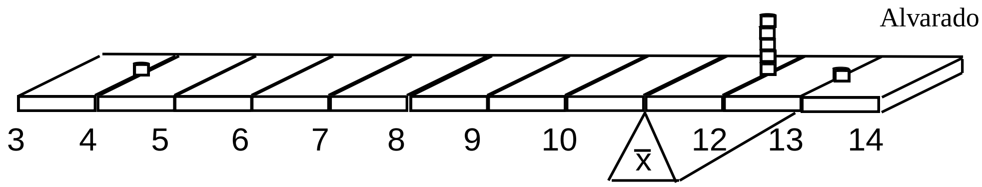


Figura IX.2

Y calcula el rango: 13−4 = 9

A primera vista Alicia percibe más dispersión en el conjunto de Tlacotalpan porque los datos se encuentran más "regados" que los de Alvarado; sin embargo, los valores de los rangos indican lo contrario:

Rango del conjunto de Tlacotalpan = 8
Rango del conjunto de Alvarado = 9 (mayor dispersión)

Esta inconsistencia se debe a que el rango **no toma en cuenta los valores intermedios**, lo cual es una desventaja.

El rango de un conjunto finito de datos numéricos es una medida cuantitativa de su variabilidad o dispersión. Al calcularlo sólo se toman en cuenta los valores extremos.

Lo anterior significa que Alicia debe buscar otra manera de medir la dispersión de los datos.

CÁLCULO DE LA DESVIACIÓN MEDIA

Alicia se concentra en las mediciones de Tlacotalpan y toma como referencia el punto de equilibrio, o sea, la media aritmética:

7, 8, 8, 12, 10, 5, 13 $\bar{x} = 9$

Se le ocurre que **el promedio del alejamiento de los datos con respecto a su media aritmética** podría ser una buena medida de su dispersión. En consecuencia, procede a obtener dichos alejamientos calculando el valor absoluto de la "desviación" (distancia) de cada dato con respecto a 9:

$|7 - 9| = 2$
$|8 - 9| = 1$
$|8 - 9| = 1$
$|12 - 9| = 3$
$|10 - 9| = 1$
$|5 - 9| = 4$
$|13 - 9| = 4$

(la diferencia entre dos rayas verticales se denomina **valor absoluto** y significa que el resultado debe ser **positivo**)

A continuación calcula la media aritmética de esas distancias, a lo que se denomina **desviación media, D. M.,**:

$$\text{D.M.} = \frac{2+1+1+3+1+4+4}{7} = 2.286$$

Calcula mentalmente: ¿Cuál hubiera sido la suma de las distancias si no se hubieran tomado valores absolutos? ____.

Alicia sigue los pasos anteriores con las mediciones que hizo en Alvarado y obtiene:

$$\text{D. M.} = \frac{1+1+2+1+1+1+7}{7} = 2$$

Estos dos últimos resultados la dejan más satisfecha, porque ahora las mediciones de Tlacotalpan (a primera vista más dispersas) tienen una desviación media mayor que las de Alvarado (que se aglutinan, sobre todo, en el valor 12).

La desviación media de un conjunto finito de datos numéricos es un promedio (media aritmética) de las distancias de los datos con respecto a su media aritmética.

En general, dado un conjunto $x_1, x_2, x_3, \ldots, x_n$ de n datos, la fórmula para calcular su desviación media se construye de la siguiente manera:

Primero se obtiene la media aritmética $\bar{x}$ de los n datos del conjunto mediante $\bar{x} = \dfrac{\sum_{i=1}^{i=n} x_i}{n}$

Luego se calcula la **distancia** de cada dato a la media aritmética $\bar{x}$, pero en valor absoluto para evitar los signos negativos: $|x_i - \bar{x}|$

Después se suman las n distancias:

$$\bar{x} = \sum_{i=1}^{i=n} |x_i - \bar{x}|$$

La desviación media D.M. se obtiene al dividir esa suma entre el número de distancias:

$$\text{D.M.} = \frac{\sum_{i=1}^{i=n} |x_i - \bar{x}|}{n}$$

Esta es la media aritmética del alejamiento de los datos con respecto a su media aritmética.

Problema. En los Viveros de Coyoacán, CdMx, el Dr. Matías Moreno mide la cantidad de O_2 (en centésimas de mililitro) que consumen 10 semillas de maíz durante su germinación (a 25 °C):

20, 24, 22, 21, 25, 24, 23, 20, 21, 20

Para conocer el índice de dispersión de sus datos, procede con orden a calcular su desviación media. Cada paso lo va registrando en las columnas de la tabla 5.1:

x_i	$\bar{x}$	$\lvert x_i - \bar{x}\rvert$	D. M.
20	22	|20 - 22| = 2	$1.6\,\frac{\text{ml}}{100}$
24		|24 - 22| = 2	
22		|22 - 22| = 0	
21		|21 - 22| = 1	
25		|25 - 22| = 3	
24		|24 – 22| = 2	
23		|23 - 22| = 1	
20		|20 - 22| = 2	
21		|21 - 22| = 1	
20		|20 - 22| = 2	
		$\sum_{i=1}^{i=n} \lvert x_i - \bar{x}\rvert = 16$	

Tabla IX.1

1er. paso: calcula la media aritmética (2a. columna):

2º paso: calcula la distancia de cada dato con respecto a la media aritmética (3a. columna).

3er. paso: suma las desviaciones (base de la 3a. columna):

4º paso: divide la suma de las desviaciones entre el número de datos (última columna).

A diferencia del rango, la desviación media toma en cuenta **todos** los valores.

CÁLCULO DE LA DESVIACIÓN TÍPICA Y DE LA VARIANZA

Problema. El médico Mauricio Kuri registra los tiempos de supervivencia, en meses, de once pacientes masculinos después de la aparición de metástasis pulmonar:

11, 13, 13, 13, 13, 13, 14, 14, 15, 15, 17

Para saber la variabilidad del tiempo de supervivencia, realiza seis pasos registrándolos en la tabla IX.2.

El Dr. Kuri primero calcula la media aritmética (2a. columna):

$$\bar{x} = \frac{11 + 5(13) + 2(14) + 2(15) + 17}{11} = \frac{151}{11}$$

Enseguida obtiene las desviaciones de los datos con respecto a su media (3a. columna). Aquí, el Dr. Kuri repara en que no tendría sentido sumar las desviaciones para calcular su media aritmética . ¿Por qué? ____

Él sabe que un número elevado al cuadrado siempre da un resultado positivo; así, para evitar los signos negativos eleva al cuadrado las desviaciones (4a. columna).

Luego suma los cuadrados de las desviaciones (base de la 4a. columna).

Después calcula la **varianza**, ***v***, dividiendo la suma anterior entre el número de desviaciones (media aritmética de los cuadrados de las desviaciones) (5a. columna, ¿la puedes completar?).

Por último, obtiene la DESVIACIÓN TÍPICA, ***s***, al extraer la raíz cuadrada de la varianza (escribe el resultado en la sexta columna).

x_i	$\bar{x}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$v = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$	s
11	13.727	-2.727	7.436	________	s = _______
13		-0.727	.528		
13		-0.727	.528		
13		-0.727	.528		
13		-0.727	.528		
13		-0.727	.528		
14		0.273	.074		
14		0.273	.074		
15		1.273	1.620		
15		1.273	1.620		
17		3.273	10.712		
			Σ= 24.176		

Tabla IX.2

La desviación típica (o estándar) es una medida cuantitativa de dispersión y constituye una tercer manera de medir el grado de dispersión de los datos. La dispersión la mide, al igual que la desviación media, **en relación con la media aritmética de los datos,** y también toma en cuenta a todos los valores.

A partir de un conjunto x_1, x_2, x_3, . . ., x_n de n datos, la fórmula para calcular la desviación típica o estándar se construye así:

1º. Se calcula la media aritmética del conjunto: $\bar{x} = \dfrac{\sum_{i=1}^{i=n} x_i}{n}$

2o. Se calcula la "desviación" de cada dato x_i del conjunto, con respecto a su media aritmética $\bar{x}$:

$$x_i - \bar{x}$$

3o. Para evitar los signos negativos se calcula el cuadrado de cada desviación: $(x_i - \bar{x})^2$.

4o. Se suman los cuadrados de las desviaciones: $\sum_{i=1}^{n} (x_i - \bar{x})^2$.

5o. Se divide la suma entre el número de datos, es decir, se calcula la media aritmética de los cuadrados de las desviaciones:

$$\frac{\sum_{i=1}^{i=n}(x_i - \bar{x})^2}{n}$$ A esta cantidad se le llama **varianza**

6o. Para compensar el haber elevado al cuadrado las desviaciones, se extrae la raíz cuadrada de la varianza. El resultado es la desviación típica ***s*** (o estándar):

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

FICHA DE TRABAJO IX.1

CONTESTA APARTE

1. ¿Cuál es el valor del rango para cada uno de estos dos conjuntos de valores?

3, 9, 9, 9, 9, 9, 9, 9   25, 27, 26, 30, 28, 29.3, 30.7, 29

En los valores de los rangos se nota una de sus desventajas como medida de variabilidad, ¿cuál? Para contestar compara a primera vista la variabilidad de los datos de un conjunto con respecto al otro.

2. En las farmacias de la colonia Educación, un jarabe para la tos producido por cierto laboratorio tiene los precios indicados a continuación. ¿Cuál es la desviación media del precio?

Indica cada paso para hacer el cálculo y organiza tus operaciones y resultados en una tabla.

\$46.00   \$47.10   \$46.50   \$48.60   \$46.10   \$47.00

En cambio, en las farmacias de la colonia Olímpica, el mismo jarabe tiene precios con una desviación media de \$2.34, ¿qué indica la diferencia de las desviaciones medias con respecto a sus precios en cada colonia?

3. ¿Qué relación existe entre la desviación típica y la varianza?

4. ¿Cuál de las siguientes afirmaciones es verdadera?

a. Entre mayor sea el valor de la media aritmética de un conjunto de datos, mayor será su dispersión.

b. Si se tienen dos conjuntos, el que tenga más datos tendrá mayor dispersión.

c. La desviación media siempre debe ser igual a la desviación estándar.

d. La desviación estándar y la desviación media son medidas de la dispersión que tienen los datos de un conjunto con respecto a su media aritmética.

5. El *Índice de Masa Corporal* (IMC) se utiliza para determinar si la persona tiene peso normal, peso excesivo o es obesa. Para los adultos de 20 años o más, un peso excesivo se considera cuando el IMC está entre 25 y 29.9 Kg/m^2; y obesidad cuando el IMC es de 30 Kg/m^2 o más. Al medir el IMC de 10 amas de casa, se obtuvo:

29.6, 33.2, 27.4, 30.8, 25.3, 40.0, 28.5, 22.3, 42.9, 28.8

Calcula, por un lado, la desviación estándar de las amas de casa que tienen peso excesivo, y por otro la desviación estándar de las que tienen obesidad (al realizar tus cálculos, indica cada paso y ordena los resultados en una tabla).

¿Cuáles IMC tienen mayor dispersión?

Calcula tu IMC. ¿Tienes obesidad?

RESUMEN

El rango es una medida cuantitativa del grado de variabilidad o dispersión de un conjunto de datos, donde únicamente se toman en cuenta el mayor y el menor de los valores. Se calcula obteniendo la diferencia entre esos valores.

Al comparar dos conjuntos puede suceder que aquél con rango mayor sea el menor disperso, contradicción debida a que los valores intermedios no afectan a esta medida de dispersión.

Con la desviación media medimos la variabilidad o dispersión de los datos de un conjunto. Al calcularla se toman en cuenta todos los valores.

Su cálculo se puede realizar en cuatro pasos:

a) Se calcula la media aritmética de los datos.
b) Se calculan las distancias (valores absolutos de las desviaciones) entre cada dato y la media aritmética.
c) Se suman las distancias.
d) Se divide la suma entre el número de datos.

La desviación media equivale a la media aritmética de las distancias de los datos con respecto a su propia media aritmética.

La desviación típica, o estándar, es otra manera de medir la dispersión de un conjunto de datos con respecto a su media aritmética. En su cálculo se toman en cuenta todos los valores del conjunto.

Pasos para obtener el valor de la desviación típica.

1o. Se calcula la media aritmética del conjunto.
2o. Se calcula la desviación de cada dato x_i del conjunto, con respecto a su media aritmética $\overline{x}$.
3o. Para evitar los signos negativos se calcula el cuadrado de cada desviación.
4o. Se suman los cuadrados de las desviaciones.
5o. Se divide la suma entre el numero de datos, es decir, se calcula la media aritmética de los cuadrados de las desviaciones. A esta cantidad se le llama varianza.
6o. Para compensar el haber elevado al cuadrado las desviaciones, extraemos la raíz cuadrada de la varianza. El resultado es la desviación típica o estándar.

10ª SESIÓN
MEDIDAS DE DISPERSIÓN PARA DATOS AGRUPADOS

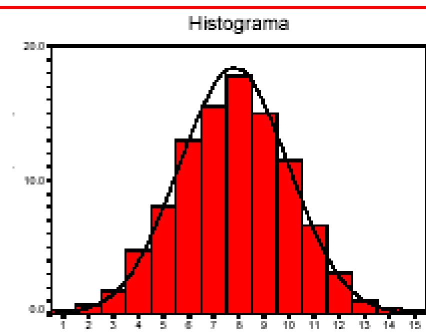
COMPETENCIAS POR ADQUIRIR

- **Calcularás la desviación media y la desviación típica para un conjunto dado de datos agrupados en clases.**

MEDIDAS DE DISPERSIÓN PARA DATOS AGRUPADOS

El cálculo de las desviaciones estándar o media, a veces requiere de supuestos con los que se sacrifica un poco de exactitud, pero se gana en simplicidad y rapidez; por otra parte, conviene tener presentes las denominaciones de mayor uso para distintas formas de una distribución de datos. Ambas cosas las podrás aprender enseguida.

¿PREFERIMOS LA SIMETRÍA. . . ?



Nos sorprendemos cuando la naturaleza nos muestra formas simétricas, como el cristal de nieve fotografiado con un microscopio de super alta resolución (abajo, copo de Wisconsin, capturado y fotografiado por K. Libbrecht y P. Rasmussen). *snowcrystals.com*.

En Estadística, al hacer estudios donde aparecen conjuntos muy grandes de datos, la primera búsqueda es la manera de agruparlos en clases; así medimos su tendencia central y su **grado de dispersión** (esto haremos aquí). Una vez organizados, tenemos la pauta para construir su histograma y asociado a éste su correspondiente polígono de frecuencias. . . y todavía más, "suavizamos" el polígono trazando una curva que no tenga picos o cambios abruptos (arriba).

Nos congratulamos cuando la curva resulta simétrica, y entonces la llamamos "**normal**". Cuando no lo es, con razón la denominamos sesgada, pero no obstante que sea normal, si los datos están muy **dispersos** la curva se "ensancha" y si son muy parecidos, la curva se hace más angosta.

DESVIACIÓN MEDIA

Problema. En una muestra formada por 100 jugadores de futbol de la liga *Olmeca* en Tuxtla Gutiérrez, Chiapas, sus edades fluctuaron entre 18 y 38 años, de esta manera:

Edad (años)	18 – 20	21 – 23	24 – 26	27 – 29	30 – 32	33 – 35	36 – 38
Frecuencia	5	25	33	17	8	4	8

Tabla X.1

Se quiere conocer la *desviación media* de las edades de los jugadores. Al no conocer los datos, en cada intervalo se considera a la marca de clase como el único valor y repetido tantas veces como indique su frecuencia.

Para calcular su *desviación media* se construye paso a paso la tabla X.2.

1er. paso: se calcula la media aritmética (3a. columna):

$$\frac{5(19)+25(22)+33(25)+17(28)+8(31)+4(34)+8(37)}{100}=26.26$$

2º paso: se calcula la *distancia* de cada marca de clase con respecto a la media aritmética (4a. columna).

3er. paso: se multiplica cada distancia por la frecuencia respectiva y se suman los resultados (5a. columna).

4º paso: la *desviación media* (D.M.) se obtiene al dividir la suma obtenida entre el número de datos (última columna).

x_i	f_i	$\bar{x}$	$\lvert x_i - \bar{x}\rvert$	$f_i \lvert x_i - \bar{x}\rvert$	$\text{D.M.}=\dfrac{\sum_{i=1}^{i=n} f_i \lvert x_i - \bar{x}\rvert}{n}$
19	5	26.26	|19 - 26.26| = 7.6	38	3.67
22	25		|22 - 26.26| = 4.6	115	
25	33		|25 - 26.26| = 1.6	52.8	
28	17		|28 - 26.26| = 1.4	23.8	
31	8		|31 - 26.26| = 4.4	35.2	
34	4		|34 – 26.26| = 7.4	29.6	
37	7		|37 - 26.26| = 10.4	72.8	
				$\sum_{i=1}^{i=n} f_i \lvert x_i - \bar{x}\rvert = 367.2$	

Tabla X.2

Para otro conjunto de jugadores se obtuvo una desviación media de 2.4, lo cual quiere decir que, en promedio, los valores están menos alejados de su media aritmética (menos dispersos) que los de la liga *Olmeca*.

PREPARA LÁPIZ, GOMA, CALCULADORA Y MUCHA CONCENTRACIÓN. . .

Al calcular la desviación media de los datos agrupados en una Tabla de Frecuencias por Intervalos, se supone que los datos de cada intervalo tienen el mismo valor que la marca de clase correspondiente.

Con esta hipótesis calcularás la **desviación media** de la tabla X.3.

- **Primer Paso**: Con las **marcas de clase** se obtiene la media aritmética $\bar{x}$:

$\bar{x} =$ ____ = ______________

$\bar{x} =$

- **Segundo Paso**: Se obtiene el valor absoluto de la desviación de cada dato (la marca de clase), con respecto a la media aritmética $|x_i - \bar{x}|$ (escribe los valores en la quinta columna de la tabla).
- **Tercer Paso**: Se multiplica cada desviación por la frecuencia $f_i\,|x_i - \bar{x}|$ (escribe los resultados en la sexta columna).
- **Cuarto Paso**: Se suman los resultados obtenidos y se divide entre el número de datos. Ese resultado es el valor de la desviación media, escríbelo en la última columna).

MARCAS ALCANZADAS EN UNA COMPETENCIA DE SALTO TRIPLE POR 100 ESTUDIANTES DEL INSTITUTO POLITÉCNICO NACIONAL						
Longitud (m)	**Frecuencia**	**Marca de Clase**	$\bar{x}$	$\lvert x_i - \bar{x}\rvert$	$f_i\lvert x_i - \bar{x}\rvert$	**D.M.**
4.7-5	1	4.85	________			________
5-5.3	0	5.15				
5.3-5.6	3	5.45				
5.6-5.9	15	5.75				
5.9-6.2	38	_____				
6.2-6.5	23	_____				
6.5-6.8	10	_____				
6.8-7.1	7	_____				
7.1-7.4	3	_____				

$\sum f_i\,|x_i - \bar{x}| =$

Tabla X.3

En otro grupo de 100 estudiantes en la Universidad de Guadalajara la desviación media resultó menor, ¿cuál competencia estuvo más reñida? ___________________________

DESVIACIÓN ESTÁNDAR O TÍPICA

Problema. En el estudio para introducir tractores a una región de Miahuatlán , Oaxaca, se midió la fuerza promedio de tiro empleada por 38 bueyes cuando se abrieron surcos con arado *vertedera* en igual número de campos de cultivo. Los resultados se organizaron en la tabla X.4 y se quiere saber el grado de dispersión que tiene la fuerza, es decir, su **desviación típica.**

FUERZA DE TIRO (NEWTONS) NÚMERO DE BUEYES	
1098-–1103	6
1104-–1109	7
1110-–1115	9
1116-–1121	12
1122-–1127	2
1128-–1133	2

Total: 38

Tabla X.4

Para calcularla, en cada intervalo suponemos que las fuerzas de tiro tienen el valor de la **marca de clase**. Por ejemplo, los seis bueyes del primer intervalo alcanzaron una potencia de 1100.5 newtons.

Con esta hipótesis daremos los pasos siguientes (registrados en tabla al final de esta página).

- **Primer paso**: Se calcula la media aritmética:

$$\bar{x} = \frac{1100.5(6) + 1106.5(7) + 1112.5(9) + 1118.5(12) + 1124.5(2) + 1130.5(2)}{38}$$

$$\bar{x} = 1112.97$$

- **Segundo paso**: Se calcula la desviación de cada dato (marca de clase) con respecto a la media (4a. columna de la tabla).
- **Tercer paso**: Se elevan al cuadrado las desviaciones (5a. columna).
- **Cuarto paso**: Se multiplica cada desviación elevada al cuadrado por su frecuencia correspondiente (6a. columna), y se suman los resultados (final de la 6a. columna).
- **Quinto paso**: Se obtiene la ***varianza, s***^{2}**,** dividiendo la suma resultante entre el número de datos (7a. columna).
- **Ultimo paso**: Calcular la desviación típica ***s*** extrayendo raíz cuadrada (7a. columna).

Marcas de Clase x_i	Frecuencias f_i	Media Aritmética $\bar{x}$	Desviaciones $x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i\,(x_i - \bar{x})^2$
1100.5	6	$\bar{x} = 1112.97$	-12.47	155.5009	933.0054
1106.5	7		-6.47	41.8609	293.0263
1112.5	9		-0.47	0.2209	1.9881
1118.5	12		5.53	30.5809	366.9708
1124.5	2		11.53	132.9409	265.8818
1130.5	2		17.53	307.3009	614.6018
					Σ = 2475.4742

Tabla X.5

$$s^2 = \frac{\sum f_i (x_i - \bar{x})^2}{n}$$

= 65.1440

$$s = \sqrt{\frac{\sum f_i (x_i \quad \bar{x})^2}{n}}$$

= 8.0711

PIENSA Y COMPLETA

Para calcular la Desviación Típica de los datos agrupados en una Tabla de Frecuencias por Intervalos, en cada intervalo vamos a suponer que **tienen el valor de la marca de clase correspondiente**.

Con esta hipótesis podremos calcular la Desviación Típica de los datos agrupados en la tabla que está al pie de esta página.

- **Primer paso**: Se calcula la media aritmética

$\overline{x} =$ ____

$\overline{x} =$ ________________________ = ______________

- **Segundo paso**: Se calcula la desviación de cada dato (marca de clase) con respecto a la media. (5a. columna de la tabla)
- **Tercer paso**: Se elevan al cuadrado las desviaciones (6a. columna)
- **Cuarto paso**: Se multiplican las desviaciones elevadas al cuadrado por la frecuencia correspondiente (2a. columna), y luego se suman los resultados (final de la 7a. columna)
- **Quinto paso**: Se divide la suma resultante entre el número de datos (8a. columna)
- **Ultimo paso**: Se extrae raíz cuadrada. (8a. columna)

LONGITUD EN SALTO TRIPLE LOGRADA POR 100 ESTUDIANTES DEL I.P.N.							
Longitud (m)	**Frecuencia**	**Marca de Clase**	$\overline{x}$	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$\mathbf{f_i}(x_i - \overline{x})^2$	$s^2 = \frac{\sum_{i=1}^{100} f_i (x_i - \overline{x})^2}{100}$
4.7-5	1	4.85					
5-5.3	0	5.15					
5.3-5.6	3	5.45					$s =$
5.6-5.9	15	5.75					
5.9-6.2	38	6.05					
6.2-6.5	23	6.35					
6.5-6.8	10	6.65					
6.8-7.1	7	6.95					
7.1-7.4	3	7.25					
Total							

Tabla X.6

FICHA DE TRABAJO X.1

1. Abajo, en una tabla, efectúa los pasos para calcular la desviación estándar de los datos.

ESTATURA DE 50 JUGADORES DE BASQUETBOL EN LA LIGA MAYOR DE CHIHUAHUA, MÉXICO		
Estatura (m)	Frecuencia	Marcas de Clase
1.60-1.65	1	1.625
1.65-1.70	0	1.675
1.70-1.75	1	1.725
1.75-1.80	0	1.775
1.80-1.85	15	1.825
1.85-1.90	18	1.875
1.90-1.95	9	1.925
1.95-2.00	3	1.975
2.00-2.05	2	2.025
2.05-2.10	1	2.075

TAREA: **a)** Con computadora, calcula la desviación estándar de la tabla anterior.

b) El rango intercuartílico y el coeficiente de variación son otras medidas de dispersión. Investiga su significado y la manera de calcularlos.

RESUMEN

Si se quiere calcular la desviación media, o la estándar, de una tabla de frecuencias por intervalos, se debe suponer que los datos en cada intervalo tienen un valor igual a la marca de clase del mismo y efectuar los pasos ya estudiados.

11ª SESIÓN
CORRELACIÓN LINEAL

COMPETENCIAS POR ADQUIRIR

- **Calcular el coeficiente de correlación lineal o de Pearson.**
- **Asociar un diagrama de dispersión con el valor del coeficiente de correlación.**
- **Asociar el error estándar de estimación con el valor del coeficiente de correlación.**
- **Calcular el error estándar de estimación para resolver un problema dado.**
- **Predecir valores de una variable empleando el modelo de regresión lineal.**

MATERIAL NECESARIO: CALCULADORA, ESCUADRAS, LÁPIZ, SACAPUNTAS Y GOMA.

CORRELACIÓN LINEAL POSITIVA

Al estudiar las siguientes páginas aprenderás la forma de medir la *relación lineal* que exista entre dos variables, manejarás una técnica para encontrar el modelo teórico de dicha relación y harás predicciones dentro de cierto margen de error.

Para ello es parte central y fundamental comprender los conceptos de **recta de regresión**, **error estándar de estimación** y **coeficiente de correlación**.

¿QUIÉN FUE PEARSON?

La fuerza de la relación entre dos variables se mide hasta nuestros días empleando el coeficiente **r** de **Pearson**.

Karl Pearson Smith (1857-1936) nació en Londres, Inglaterra, tuvo una vida muy productiva y versátil, pero sobre todo se le conoce como uno de los fundadores de la estadística moderna.

A los 36 años desarrolló métodos matemáticos para estudiar los procesos de la herencia y la evolución, a partir de entonces escribió 18 trabajos titulados *Contribuciones Matemáticas a la Teoría de la Evolución* e hizo valiosos aportes sobre *análisis de regresión* y *prueba de ji cuadrada para significancia estadística*, además, acuñó el término *desviación estándar*.

CORRELACIÓN LINEAL POSITIVA

Abundan situaciones en las que dos variables se relacionan (escribe una):

La *contaminación ambiental*	**se relaciona con**…	la *calidad del combustible* utilizado en los medios de transporte
La mayor *demanda de educación*	**se relaciona con**…	la *tasa de crecimiento poblacional*
____________________________	**se relaciona con**…	____________________________

El interés por saber si dos variables están relacionadas o asociadas es porque, de ser así, una de ellas pudiera ser causa de la otra. Se dice que una variable X es causa de otra variable Y, si una variación en X produce una variación en Y, siempre y cuando estén controladas las demás variables, lo cual en la práctica sucede sólo en algunos casos.

Que dos variables estén relacionadas entre sí no basta para demostrar su causalidad. Podemos estar seguros de que el incremento de temperatura en todo el planeta causa una mayor evaporación de los mares, pero no podríamos afirmar con certeza que como efecto de aumentar la inversión extranjera en negocios de hamburguesas, aumentan los obesos, menos aún aseverar que un aumento en las ventas de ositos de peluche al finalizar marzo fuera consecuencia del inicio de la primavera.

Cuando se tengan dos variables **escalares continuas**, averiguar su **correlación** o **grado de asociación** sirve para saber si existe una relación de interdependencia entre ellas, -aun cuando no sea de causa-efecto- y la magnitud y dirección de esa relación.

Otro interés frecuente es **predecir** resultados. A un inversionista de la Bolsa de Valores le importaría saber cómo cambiará el precio de las acciones en los próximos meses; a un comerciante le preocuparía saber si por hacer mayores gastos en publicidad de un producto aumentarán sus ventas; a un nutriólogo le gustaría predecir el efecto del Omega 3 en el nivel de colesterol de las personas. . . y así hay una infinidad de situaciones.

Problema. Jaime García está encargado de la capacitación del personal en un banco y debe resolver esta cuestión:

ANTIGÜEDAD (AÑOS)	CALIFICACIÓN (PUNTOS)
2.2	44
2.7	56
3.0	58
3.2	64
3.5	73
4.1	71
4.4	75
4.7	78
5.3	74
5.5	89

Tabla XI.1

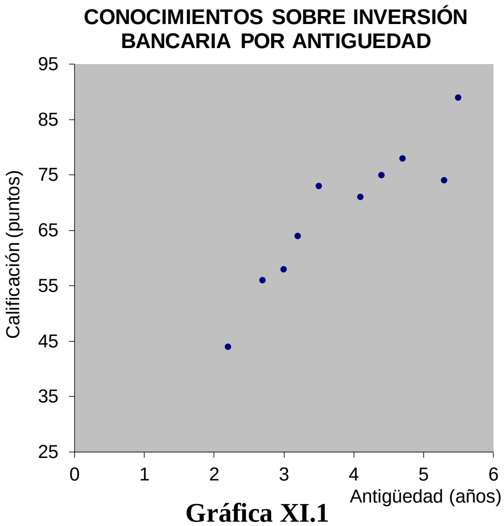
– ¿Cuál es el **grado de asociación** que hay entre las variables *conocimientos sobre inversión bancaria* y *antigüedad* de los empleados?

Con el fin de determinar el nivel de conocimiento de los empleados acerca de la inversión bancaria, Jaime toma una muestra de tamaño 10 y les aplica una prueba con escala de 0 a 100 puntos.

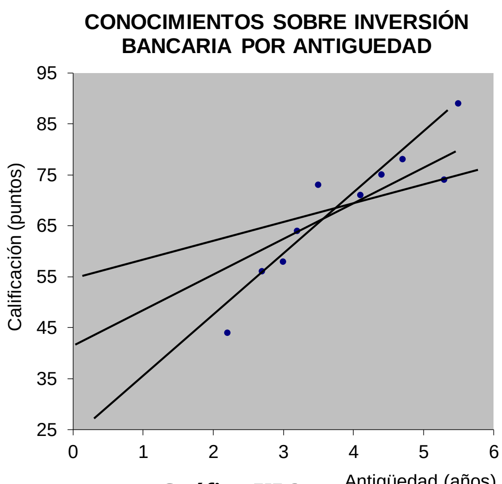
Después tabula los resultados de la prueba y la antigüedad respectiva de cada empleado (tabla XI.1).

Ahí puede observar inmediatamente que entre **mayor** es el número de años de antigüedad, **mayor** es la calificación obtenida por los empleados. Se dice que una relación entre dos variables es **positiva** cuando al aumentar el valor de una también aumenta el de la otra**.**

La representación gráfica de la tabla 6.1 tiene la apariencia de una “nube de puntos” y se le conoce con el nombre de diagrama o **gráfica de dispersión** (gráfica XI.1).



Gráfica XI.1



Gráfica XI.2

Por la ubicación que presentan los puntos de la gráfica de dispersión, se pueden **ajustar** a una **recta teórica,** la cual servirá como modelo de la **asociación lineal** entre los *conocimientos sobre inversión bancaria* y la *antigüedad* de los empleados.

En general, cuando los puntos provengan de dos variables continuas, se podrán trazar varias rectas que parezcan ajustarse a la nube (gráfica XI.2). La tarea es encontrar la **“recta de mejor ajuste”**.

De todas las rectas posibles, evidentemente ninguna puede pasar por todos los puntos, cabe preguntarse entonces cuál es la de **mejor ajuste**, o sea, la más representativa, la que pudiera considerarse como el modelo de la asociación entre las variables.

La recta de **mejor ajuste** es un modelo determinístico de utilidad para:

1.- **Describir la relación** entre las dos variables.

2.- **Predecir,** con cierto grado de aproximación el valor de una variable, dado un valor de la otra.

REGRESIÓN LINEAL

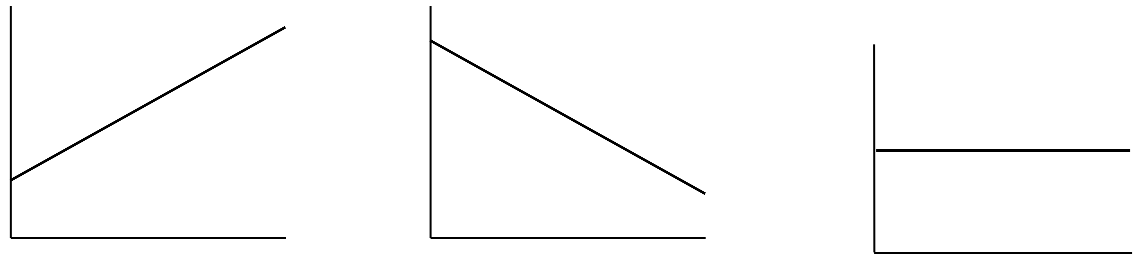
Para precisar **cómo** es la asociación entre las variables y determinar la recta de mejor ajuste, se utiliza una técnica estadística conocida como **regresión lineal**.

Primero recordemos la forma de la ecuación de una línea recta en un plano cartesiano:

$$y = mx + b$$

"y" es la variable dependiente, "x" la variable independiente, "m" la pendiente de la recta (nos indica cuánto cambia "y" cuando "x" cambia una unidad) y "b" la ordenada al origen (distancia desde el origen al punto de intersección de la recta con el eje de las ordenadas).

En el plano cartesiano, según el signo de su pendiente, una recta puede tomar distintas posiciones (diagramas XI.1).



Pendiente positiva (relación positiva) Pendiente negativa (relación negativa) Pendiente cero (no existe relación)

Diagramas XI.1

A la diferencia entre un valor observado (y_i) y el valor que predice la recta ($\hat{y}_i$), es decir, a (y_i - $\hat{y}_i$), se le denomina **error de predicción**, mismo que es negativo si el valor observado está por debajo del valor que predice la recta, o positivo si está por arriba.

En adelante nos referiremos a la recta de mejor ajuste como **recta de regresión.** Ésta se caracteriza por ser el eje de equilibrio entre los errores de predicción positivos y los negativos. Dicho de otra manera, debe ser tal que a su alrededor, en promedio, los errores positivos *pesen* lo mismo que los negativos.

Lo anterior sucede cuando la suma de los cuadrados de los "errores de predicción" es **mínima**. Condición que se cumple al emplear un método matemático (cuyo desarrollo omitiremos) denominado **método de mínimos cuadrados.**

El método de mínimos cuadrados da por resultado las fórmulas que permiten calcular la pendiente **m** de la **recta de regresión** y su ordenada al origen **b**, cuando se conocen los valores $\boldsymbol{x_i}$ de la variable independiente, los valores $\boldsymbol{y_i}$ de la variable dependiente y el número de datos **n**:

x_i	y_i	x_i^2	$x_i y_i$
2.2	44	4.84	96.8
2.7	56	7.29	151.2
3.0	58	9.00	174.0
3.2	64	10.24	204.8
3.5	73	12.25	255.5
4.1	71	16.81	291.1
4.4	75	19.36	330.0
4.7	78	22.09	366.6
5.3	74	28.09	392.2
5.5	89	30.25	489.5
Σ = 38.6 $\bar{x}$ = 3.86	Σ = 682 $\bar{y}$ = 68.2	Σ = 160.22	Σ = 2751.7

Tabla XI.2

$$m = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \bar{y} - m\bar{x}$$

Para aplicar estas fórmulas, Jaime García construye la tabla XI.2 (ahí los valores de la variable *antigüedad del empleado* se denotan por "x_i" y los valores de la variable *conocimientos del empleado sobre inversión bancaria* se denotan por y_i).

Al sustituir los resultados de la tabla 6.2 en las fórmulas, Jaime obtiene:

$$m = \frac{10\,(2751.7) - (38.6)(682)}{10\,(160.22) - (38.6)^2} = 10.618 \qquad b = 68.2 - 10.618(3.86) = 27.214$$

Por lo tanto, la **ecuación de la recta de regresión** es: $y = 10.618\,x + 27.214$

Con la ecuación anterior Jaime puede anticipar valores de la variable dependiente.

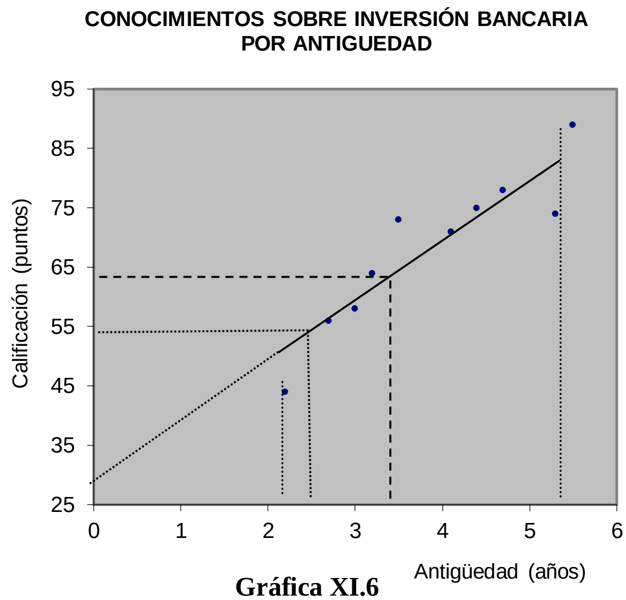
Puede predecir, por ejemplo, la calificación que teóricamente obtendría un empleado con 2.5 años de antigüedad ($x = 2.5$):

$y = 10.618(2.5) + 27.214$

$y = 53.76$ **(ver la gráfica XI.6)**

De acuerdo a la ecuación de la recta de regresión, la calificación esperada para un empleado con diez años de antigüedad es de 133.39 puntos. ¿Tiene sentido este valor? ¿Por qué?

En efecto, no tiene sentido, porque la calificación máxima es de 100 puntos.

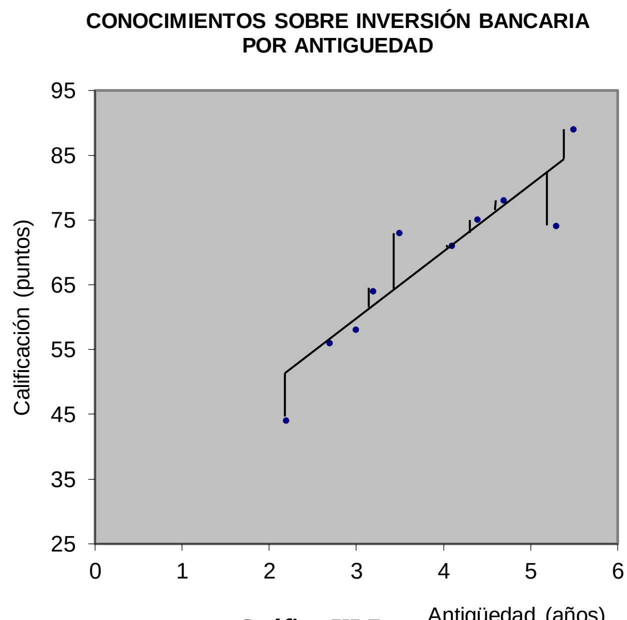


Gráfica XI.6

Las antigüedades reales están entre 2.2 y 5.5 años, fuera de este intervalo cerrado no es válido hacer predicciones. La recta de regresión es válida sólo en ese intervalo.

Al tomar como referencia la recta teórica de la gráfica XI.6, Jaime deduce que la calificación de un empleado con antigüedad de 3.5 años estaría cerca de 62 puntos; pero el valor real indicado en la tabla XI.2 es de 73 puntos, es decir, hay un **error de predicción** de 11 puntos.

En la gráfica de dispersión (XI.7) los errores de predicción se representan por medio de segmentos paralelos al eje de las ordenadas, conectando a cada punto con la recta de regresión. Los segmentos por encima de la recta representan los errores positivos, los que están por abajo, errores negativos.



Gráfica XI.7

ERROR ESTÁNDAR DE ESTIMACIÓN

Jaime quiere tener un **promedio de los errores de predicción**, o sea, del alejamiento entre las calificaciones reales y las que predice la recta de regresión.

El grado de dispersión de *y*, con respecto a su recta de regresión, se conoce por **error estándar de estimación *e*** y se calcula mediante:

$$e=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}}$$

(para el *i*'ésimo caso, y_i es el valor real de la variable dependiente, mientras $\hat{y}_i$ es su valor teórico sobre la recta de regresión).

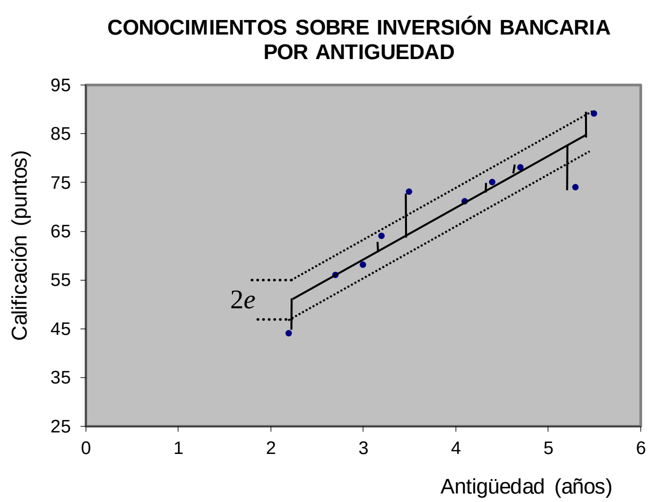
Esta expresión es la desviación estándar de la variable dependiente y mide el alejamiento vertical de los valores reales con respecto a los teóricos sobre la recta.

Si $\hat{y}_i$ representa el grado teórico de conocimientos de un empleado (obtenido con la ecuación de la recta), el error estándar de estimación de las calificaciones obtenidas por los empleados se puede calcular elaborando la tabla XI.3 ¿Puedes completarla y calcular *e*?

x_i	y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
2.2	44	50.5728	-6.5728	43.2029
2.7	56	55.8818	0.1181	0.01394
3.0	58	59.0672	-1.0672	1.1391
3.2	64	61.1908	2.8091	7.8910
3.5	73	64.3762	8.6237	74.3683
4.1	71	70.7470	0.2529	0.06396
4.4	75	73.9324	1.0675	1.1395
4.7	78	77.1178	0.8821	0.7781
5.3	74	83.4886	-9.4886	90.03529
5.5	89	85.6122	3.3877	11.4765
				$\sum$ =

e = __________

Tabla XI.3



Gráfica XI.8

El valor del error estándar de estimación (*e*) permite trazar dos rectas paralelas a las recta de regresión (izquierda).

En teoría, dentro de la franja de anchura 2*e* determinada por las rectas paralelas a la de regresión, se localiza el 68% de los datos. ¿Cuál es el porcentaje para este caso? _____ .

Si los puntos estuvieran más cerca de la recta de regresión, la franja sería más angosta, es decir, el valor de *e* sería menor.

Para la recta de regresión, la suma de los cuadrados de sus errores de predicción (Σe_i^2) es **menor** que para cualquier otra recta que se trace como modelo de la nube de puntos, es decir, la recta de regresión es la que nos hace equivocarnos menos al hacer predicciones.

COEFICIENTE DE CORRELACIÓN

Una vez obtenido el valor numérico del grado de alejamiento de las calificaciones con respecto a la recta de regresión lineal, Jaime se interesa por cuantificar la fuerza de la relación entre las dos variables bajo estudio: *conocimientos sobre inversión bancaria* por los empleados y *antigüedad* de los mismos.

La fuerza de la relación lineal entre dos variables se mide por medio de un indicador, se trata del **coeficiente de correlación** o coeficiente ***r*** de **Pearson**.

Para calcular ***r***, primero se determina la desviación estándar *s* de cada variable:

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \qquad s_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}} \qquad \dots\dots\dots\dots\dots\dots(1)$$

La **covarianza** entre las dos variables se define así:

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n} \qquad \dots\dots\dots\dots\dots\dots\dots\dots(2)$$

El coeficiente de correlación ***r*** se define como el cociente de la covarianza entre el producto de las desviaciones estándar:

$$r = \frac{s_{xy}}{s_x s_y} \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$$

Al substituir las expresiones (1) y (2) en la (3) se obtiene:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

El valor del coeficiente *r* se puede hallar por medio de una hoja de cálculo o programa de computadora.

Puesto que Jaime no tiene los conocimientos suficientes para manejar esos medios, utiliza un procedimiento por pasos que le facilita el trabajo.

Comienza calculando los factores de la fórmula de ***r***, ordenándolos así:

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
44	2.2	-24.2	-1.66	40.172	585.64	2.7556
56	2.7	-12.2	-1.16	14.152	148.84	1.3456
58	3.0	-10.2	-0.86	8.772	104.04	.7396
64	3.2	-4.2	-0.66	2.772	17.64	.4356
73	3.5	4.8	-0.36	-1.728	23.04	.1296
71	4.1	2.8	0.24	0.672	7.84	.0576
75	4.4	6.8	0.54	3.672	46.24	.2916
78	4.7	9.8	0.84	8.232	96.04	.7056
74	5.3	5.8	1.44	8.352	33.64	2.0736
89	5.5	20.8	1.64	34.112	432.64	2.6896
$\bar{x}$= 68.2	$\bar{y}$= 3.86			$\sum$=119.18	$\sum$=1495.6	$\sum$=11.224

Tabla XI.4

Con las sumas anteriores obtiene los valores de las desviaciones típicas para cada variable (s_x y s_y) y de la covarianza (s_{xy}):

$$s_x = \sqrt{\frac{1495.6}{10}} = 12.2294 \qquad s_y = \sqrt{\frac{11.224}{10}} = 1.0594 \qquad s_{xy} = \frac{119.18}{10} = 11.918$$

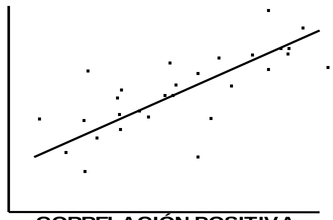
En consecuencia, el coeficiente de correlación resulta ser:

$$r = \frac{11.918}{(1.0594)(12.2294)} = 0.9199$$

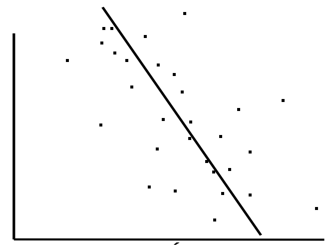
Entre más cercano a 1 (o a -1) esté el valor del coeficiente, la relación entre las variables será más fuerte. El valor de ***r*** obtenido por Jaime es cercano a 1, entonces la correlación entre la *antigüedad* de los empleados y sus *conocimientos sobre inversión bancaria* la podemos calificar como muy fuerte.

El valor absoluto del coeficiente de correlación lineal ***r*** nos indica la fuerza de la asociación lineal entre dos variables, aunque una no tenga nada que ver con la otra. Por eso, aun cuando la correlación sea buena, debe revisarse con cuidado si realmente existe dependencia entre las variables. Pudiéramos encontrar una correlación alta, por ejemplo, entre el coeficiente intelectual de un grupo de personas y el largo de su pelo, no obstante que estas variables son independientes entre sí.

El signo del coeficiente de correlación definirá la pendiente o inclinación de la recta:

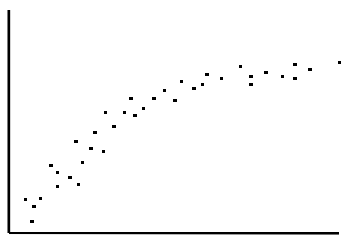


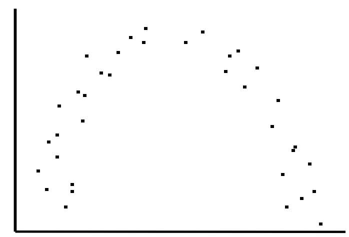
CORRELACIÓN POSITIVA



CORRELACIÓN NEGATIVA

En la práctica podemos encontrar asociaciones no lineales. En esos casos (que no trataremos aquí) los diagramas de dispersión tienen esta apariencia:





En cualquier caso, el valor del coeficiente de correlación tendrá un valor en el intervalo cerrado [1, -1].

Cuando ***r* = 1** ó ***r* = -1**, se habla de una correlación **perfecta** porque todos los puntos quedan alineados perfectamente; si ***r* = 0**, no existe ninguna asociación lineal entre las variables, por lo tanto su correlación es **nula**.

La correlación se puede calificar como **buena** si el valor del coeficiente está entre 0.7 y 1 (o entre -0.7 y -1 en caso de ser negativa), como **moderada** si su valor está entre 0.3 y 0.6 (entre -0.3 y -0.6 si fuera negativa) y **mala** si está entre 0 y 0.3 (0 y -0.3).

Las siguientes imágenes nos dan una mejor idea de la forma en que se distribuyen los puntos en el plano, según sea su grado de correlación.

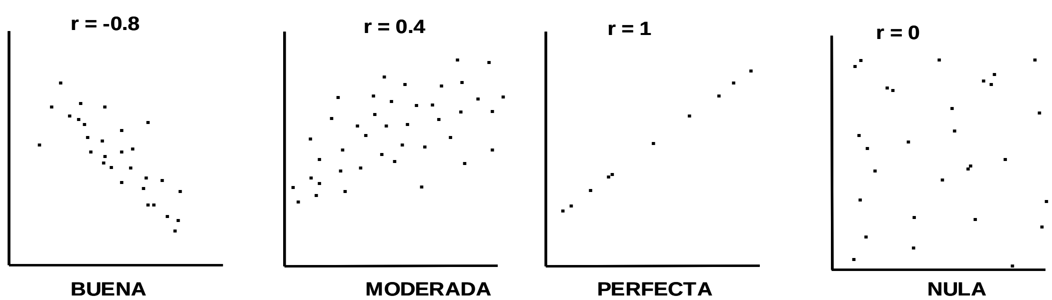


Diagrama XI.9

12ª SESIÓN
CORRELACIÓN LINEAL (CONTINÚA)

CORRELACIÓN LINEAL NEGATIVA

En lo que sigue abordarás casos en que la correlación resulta negativa o nula. Junto con los conceptos antes vistos podrás resolver algunos problemas para determinar el grado de asociación entre dos variables.

Si bien los análisis de correlación actualmente se hacen por medio de computadora, es útil comprender los conceptos haciendo cálculos "a pie" para algunos casos sencillos.

Problema. La licenciada en Administración, Alejandra Robledo, realiza estudios acerca del ausentismo del personal de una empresa y se interesa en averiguar:

> Cuál es el **grado de asociación** entre el *tiempo a la semana dedicado a convivir con la familia* y el *salario mensual* que perciben.

Con ese fin, Alejandra toma una muestra aleatoria de 12 empleados y elabora la tabla XII.1

Tiempo de convivencia semanal con la familia (hs.)	Salario mensual en números redondos (miles de $)
10	5
22	6
13	8
17	9
8	11
5	12
7	14
3	17
3	18
4	20
1	25
1	26

Tabla XII.1

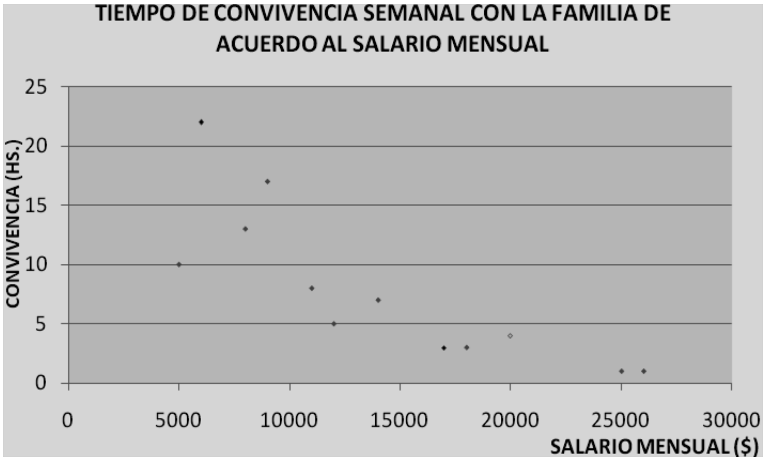
De inmediato aprecia que al incrementarse el valor de la variable independiente *salario mensual*, disminuye el valor de la variable dependiente *tiempo de convivencia semanal con la familia*.

La **correlación** entre dos variables es **negativa** si al aumentar el valor de una disminuye el de la otra**.**

Alejandra desea saber qué tan buena es la relación entre las variables.

Para encontrar la recta de regresión, es decir, la recta modelo de la asociación, decide aplicar el método de correlación lineal y empieza por representar gráficamente los valores de las variables (gráfica 6.10).

Si alrededor de la recta de ajuste debe haber un equilibrio de puntos, traza tú una recta



Gráfica XII.1

tentativa en la gráfica XII.1

En la recta que trazaste, ¿de qué signo es la pendiente? _________ y ¿cuál es el valor de su ordenada al origen? ______ .

La pendiente **m** y la ordenada al origen **b** de la recta de regresión se determinan aplicando las fórmulas que arroja el método de mínimos cuadrados.

Si con y_i se denota la variable dependiente *tiempo de convivencia semanal con la familia* y con x_i la variable independiente *salario mensual*, los valores de ***m*** y **b** se encuentran mediante las expresiones utilizadas anteriormente:

$$m = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad b = \bar{y} - m\bar{x}$$

Con ese propósito es útil auxiliarse de la tabla XII.2.

y_i	x_i (miles)	x_i^2 (millones)	$x_i y_i$ (miles)
10	5	25	50
22	6	36	132
13	8	64	104
17	9	81	153
8	11	121	88
5	12	144	60
7	14	196	98
3	17	289	51
3	18	324	54
4	20	400	80
1	25	625	25
1	26	676	26
Σ = 94	Σ = 171	Σ = 2 981	Σ = 921
$\bar{y}$ = 7.833	$\bar{x}$ = 14.25		

Tabla XII.2

Al substituir los resultados anteriores en las fórmulas, Alejandra obtuvo los valores de la pendiente *m* y la ordenada al origen:

$$m = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad b = \bar{y} - m\bar{x}$$

$$m = \frac{12\,(921\ 000) - 171\,000\,(94)}{12\,(2\,981\,000\,000) - (171\,000)^2}$$

$$= \frac{-5022}{6\,531000}$$

m = -0.000768

b = 7.833 + (-0.000768)(14250)

b = 7.833 + 10.944

b = 18.777 **¿b es igual a la ordenada al origen de la recta que trazaste? _____ .**

Por lo tanto, la recta de regresión tiene por ecuación:

$$y = -0.000768x + 18.777$$

Con esta ecuación, ¿puedes predecir cuántas horas a la semana convive con su familia una persona con un salario mensual de 15 mil pesos? (llena los espacios):

$$y = -0.000768\ (\ \_\_\_\_\_\_\_\_\ ) + 18.777$$

$$y = \_\_\_\_$$

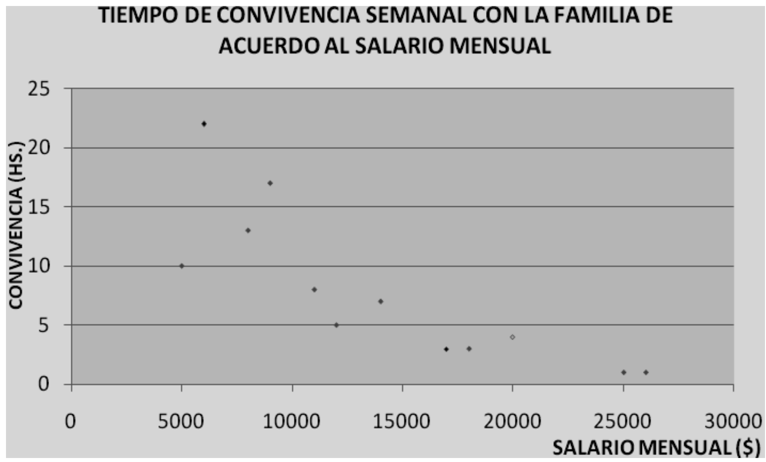
Con la ecuación de la recta de regresión, calcula mentalmente: Con un salario de 1 000 pesos al mes, ¿cuántas horas se espera que un empleado conviva con su familia? _____ . ¿Es válido hacer esta predicción? ____ . ¿Por qué? ______________________________

La recta de regresión se puede trazar determinando dos puntos: uno puede ser la ordenada al origen, sobre el eje *y*, (18.777); y otro puede hallarse, por ejemplo, con *y* = 0, el cual es un punto localizado sobre el eje *x* con valor igual a 24 449.20

Traza la recta de regresión en la gráfica XII.2.

La recta de regresión debe construirse sólo para el intervalo:

$$5\,000 < x < 26\,000$$



Gráfica XII.2

ERROR ESTÁNDAR DE ESTIMACIÓN

La fórmula para el **error estándar de estimación**

$$e=\sqrt{\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_t\right)^2}{n}}$$

nos permite calcular la dispersión de los datos con respecto a su recta de regresión.

Si con $\hat{y}_i$ se denota la variable dependiente *tiempo de convivencia semanal con la familia* (obtenido con la ecuación de la recta), el error estándar de estimación se puede calcular elaborando la tabla que sigue (las últimas cifras de la última columna se redondeó). ¿Puedes completarla y calcular *e*?

y_i	x_i (miles)	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
10	5	14.937	-4.937	24.3740
22	6	14.169	7.831	61.3246
13	8	12.633	0.367	0.1347
17	9	11.865	5.135	26.3682
8	11	10.329	-2.329	5.4242
5	12	9.561	-4.561	20.8027
7	14	8.025	-1.025	1.0506
3	17	5.721	-2.721	7.4038
3	18	4.953	-1.953	3.8142
4	20	3.417	0.583	0.3399
1	25	-0.423	1.423	2.0249
1	26	-1.191	2.191	4.8005
				$\sum =$

$e =$

Tabla XII.3

Con este valor de *e* traza tú, en la gráfica 6.11, las dos rectas paralelas a la recta de regresión. ¿Qué porcentaje de puntos se localiza en esa franja? ____ .

COEFICIENTE DE CORRELACIÓN

Para **medir** el grado de asociación lineal entre las dos variables, Alejandra emplea el **coeficiente de correlación,** o coeficiente **r** de **Pearson**, el cual es independiente de las escalas utilizadas al medir dichas variables:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Para facilitar su cálculo se auxilia de la siguiente tabla:

y_i	x_i (miles)	$y_i - \bar{y}$	$x_i - \bar{x}$ (miles)	$(x_i - \bar{x})(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})^2$ x10^6
10	5	2.167	-9.25	-20041.66	4.6944	85.56
22	6	14.167	-8.25	-116875	200.6944	68.06
13	8	5.167	-6.25	-32291.66	26.6944	39.06
17	9	9.167	-5.25	-48125.00	84.0277	27.56
8	11	0.167	-3.25	-541.66	.0279	10.56
5	12	-2.833	-2.25	6374.99	8.0277	5.06
7	14	-0.833	-0.25	208.33	.6944	0.06
3	17	-4.833	2.75	-13291.66	23.3611	7.56
3	18	-4.833	3.75	-18124.99	23.3611	14.06
4	20	-3.833	5.75	-22041.66	14.6944	33.06
1	25	-6.833	10.75	-73458.33	46.6944	115.56
1	26	-6.833	11.75	-80291.66	46.6944	138.06
$\bar{y} = 7.833$	$\bar{x} = 14.25$			Σ = -418500	Σ =479.6666	Σ = 544.25

Al sustituir las sumas en la fórmula, el coeficiente de correlación resulta ser ***r*** = -0.8191. ¿Es buena la correlación? _____.

CORRELACIÓN LINEAL NULA

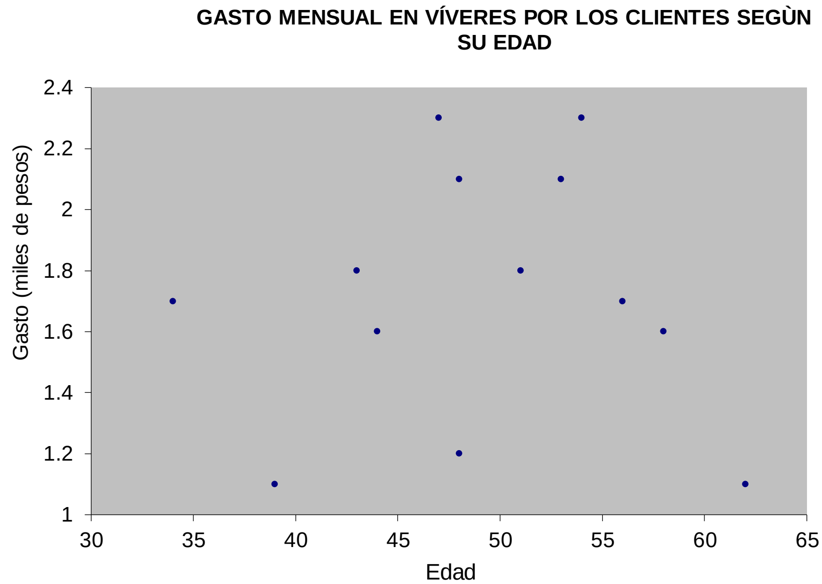
Problema. En un "minisuper" se quiere saber, acerca de sus clientes, qué tan fuerte es la relación entre sus *gastos mensuales en víveres* y su *edad*, para lo cual se recopilan y organizan los datos en la tabla XII.4.

A partir de la tabla es difícil visualizar con claridad si hay asociación entre las variables,

GASTO AL MES EN VÍVERES (MILES DE PESOS)	**EDAD (AÑOS)**
1.7	34
1.1	39
1.8	43
1.6	44
2.3	47
2.1	48
1.2	48
1.8	51
2.1	53
2.3	54
1.7	56
1.6	58
1.1	62

Tabla XII.4

Pero al representar todos los puntos en la gráfica 6.12…



Gráfica XII.3

Por la distribución de los puntos es difícil trazar una posible recta de regresión.

Ayudándote de la tabla que sigue y de estas relaciones vistas antes:

y_i	x_i (años)	x_i^2	$x_i y_i$
1.7	34	1156	57.8
1.1	39	1521	42.9
1.8	43	1849	77.4
1.6	44	1936	70.4
2.3	47	2209	108.1
2.1	48	2304	100.8
1.2	48	2304	57.6
1.8	51	2601	91.8
2.1	53	2809	111.3
2.3	54	2916	124.2
1.7	56	3136	95.2
1.6	58	3364	92.8
1.1	62	3844	68.2
Σ = 22.4	Σ = 637	Σ = 31949	Σ = 1098.5
$\bar{y} = 1.723$	$\bar{x} = 49$		

Tabla XII.5

$$m = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \bar{y} - m\bar{x}$$

calcula los valores de *m* y *b*.

m = ____________, b = ______

Con los valores anteriores, traza la recta de regresión en la gráfica 6.12.

En cuanto al coeficiente de correlación, ayudándote de la siguiente tabla:

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
34	1.7	-15	-0.023	0.345	225	0.000529
39	1.1	-10	-0.623	6.23	100	0.388129
43	1.8	-6	0.077	-0.462	36	0.005929
44	1.6	-5	-0.123	0.615	25	0.015129
47	2.3	-2	0.577	-1.154	4	0.332929
48	2.1	-1	0.377	-0.377	1	0.142129
48	1.2	-1	-0.523	0.523	1	0.273529
51	1.8	2	0.077	0.154	4	0.005929
53	2.1	4	0.377	1.508	16	0.142129
54	2.3	5	0.577	2.885	25	0.332929
56	1.7	7	-0.023	-0.161	49	0.000529
58	1.6	9	-0.123	-1.107	81	0.015129
62	1.1	13	-0.623	-8.099	169	0.388129
49	1.723			0.9	∑=736	∑=2.04307

Tabla XII.6

y de la fórmula $r = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$,calcula el valor del coeficiente de Pearson.

$r =$ _________

¿Cómo calificarías la correlación entre la variable *gastos mensuales en víveres* y *edad de los clientes*, perfecta, moderada, buena o mala? _________ .

RESUMEN

La correlación entre dos variables escalares continuas es su grado de relación o asociación.

No debe confundirse asociación con interdependencia o relación de causa-efecto. Pueden encontrarse variables con un alto grado de asociación pero que no guarden ninguna relación entre sí.

La correlación es positiva cuando los incrementos de una variable van acompañados de incrementos en la otra; es negativa si a incrementos en una le corresponden decrementos de la otra.

Para saber el grado de correlación o de asociación, en Estadística se emplea el coeficiente de correlación. Este es un número en [1, -1] que de acuerdo a su signo nos indica si la correlación es positiva o negativa y, de acuerdo a su valor si es perfecta, buena, moderada, mala o nula.

Los valores reales se pueden representar mediante puntos en el plano cartesiano y, dependiendo de la configuración que tenga la nube de puntos, se pueden ajustar mediante una recta o algún otro tipo de curva. La regresión lineal es la técnica estadística mediante la cual podemos determinar cómo es la relación lineal entre dos variables.

Cuando la configuración de puntos es lineal, se denomina recta de regresión a la que más se ajusta a los valores observados, y se expresa como: $y = mx + b$, donde "*m*" y "*b*" se pueden calcular a partir de los valores de las variables. La recta de regresión es un modelo útil porque también permite hacer predicciones acerca de valores de la variable dependiente.

La distancia vertical promedio de los puntos (parejas de datos reales) con respecto a la recta de regresión se conoce como el error estándar de estimación.

FICHA DE TRABAJO XII.2

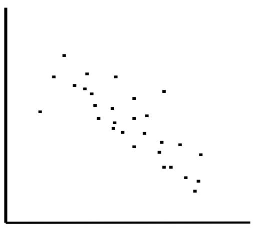
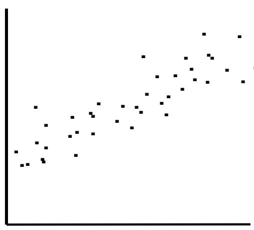
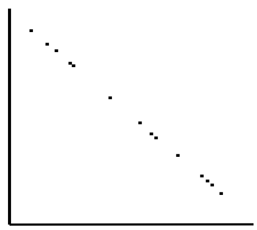
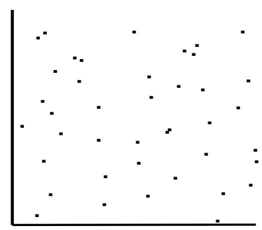
1.- Califica la correlación entre dos variables escalares continuas cuando el coeficiente de correlación es:

cero _______ uno ___________ entre 0.7 y 1 ________ entre -0.3 y -0.6 _________ .

2.- Calcula el valor del coeficiente de correlación lineal para estos valores:

$S_x = 2$ $S_y = 1.3$ $S_{xy} = 2.4$ Respuesta: _______.

3.- En cada diagrama escribe uno de estos valores del coeficiente de correlación, según corresponda: 0.8, -0.7, 0, -1:

4.- El coeficiente de correlación entre dos variables es 0.3; el coeficiente de otra pareja de variables diferente es 0.94. ¿A cuál le corresponde el mayor error estándar de estimación? Explica. ____

____

5.- Estos datos de productividad se obtuvieron en una empresa pequeña de productos de belleza:

Producción (x \$ 10 000.00)	3	4	5	6	7	8	9
Costos (x \$ 1000.00)	8	6.5	9.2	7.8	9	9.3	12.5

a) Indica si la correlación es buena o mala, positiva o negativa. Sugerencia: por pasos, registra los cálculos en una tabla, o bien utiliza una hoja de cálculo en la computadora. ____

b) Obtén la ecuación de la recta de regresión y haz la predicción de costos para producciones de \$ 65 000.00 y \$ 100 000.00 ¿son válidas?

____

c) **Calcula el error estándar de estimación ___________**

TAREA

1. Investiga en qué consiste el método de mínimos cuadrados (Elabora un resumen y cita la fuente).

13ª SESIÓN
PRINCIPIO DE LA MULTIPLICACIÓN. PERMUTACIONES

COMPETENCIAS POR ADQUIRIR

- **Representar mediante un diagrama de árbol las trayectorias y resultados posibles en una serie de sucesos.**
- **Aplicar el principio de la multiplicación al resolver un problema dado.**
- **Expresar la fórmula básica para el cálculo de permutaciones.**
- **Resolver problemas que requieran el cálculo de permutaciones.**

MATERIAL NECESARIO: CALCULADORA, LÁPIZ, GOMA Y SACAPUNTAS.

PRINCIPIO DE LA MULTIPLICACIÓN

A veces nos encontramos con preguntas del tipo *¿cuántos resultados son posibles?* o *¿de cuántas maneras se puede...?*

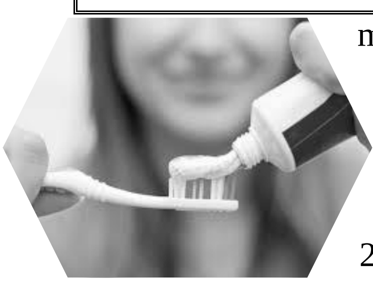
Por ejemplo, de 8 competidores en una carrera de autos se premiará a los 3 primeros lugares, ¿cuántos resultados son posibles?

Preguntas como la anterior se responden con sólo utilizar el **Principio de la Multiplicación**, de lo cual nos ocuparemos enseguida.

¿SABES CONTAR? . . .

Contar es relativamente sencillo: del conjunto de cosas por contar, a cada una le asociamos un número natural. Al primer elemento le asociamos el número natural 1, al segundo el 2, . . ., y así consecutivamente hasta finalizar, momento en el cual sabemos **cuántos** elementos tiene el conjunto.

Saber **cuántos** jitomates contiene una bolsa es sencillo, sin embargo, si quisiéramos saber **cuántas** quinielas diferentes de seis números se pueden hacer con los primeros 56 números naturales, la respuesta no se obtendría de manera tan directa…



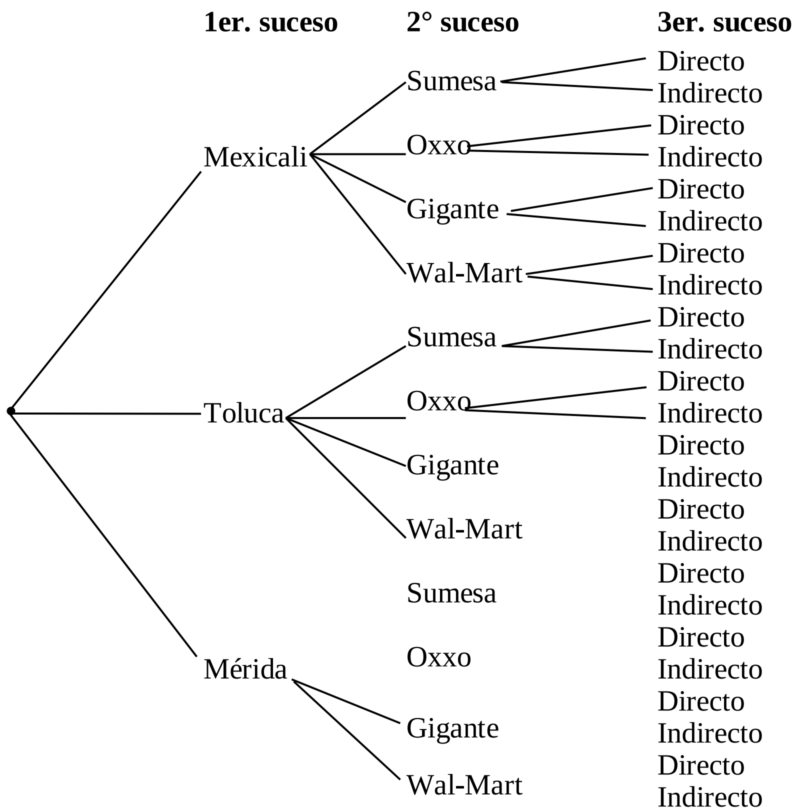
Luis es licenciado en mercadotecnia y quiere recopilar datos acerca del precio de venta de las pastas dentales en la República Mexicana. Para ello realiza tres pasos consecutivos (llamados **sucesos**):

1er. Suceso: Selecciona 3 ciudades al azar (resultan ser *Mexicali, Toluca y Mérida*).

2º. Suceso: En cada ciudad elige al azar 4 tiendas de autoservicio (obtiene *Sumesa, Oxxo, Gigante y Wal-Mart)*.

3er. Suceso: En cada tienda piensa emplear 2 métodos para conocer los precios (*directo* mediante la revisión de las etiquetas, e *indirecto* por medio de las listas de precios del Departamento de Ventas).

Para saber **de cuántas maneras** podría proceder para recopilar finalmente los datos, Luis construye un **diagrama de árbol.** ¿ Lo puedes completar?



La rama izquierda del árbol está constituida por tres **decisiones** para recopilar los datos: ir a Mexicali, ahí a un Sumesa y obtener los datos de manera directa. ¿Cuáles son las decisiones de la rama inferior?

____

En total son 24 las maneras (ramas) de recopilar los datos, las cuales se obtienen al multiplicar:

Número de decisiones posibles del 1er. suceso **X** **Número de decisiones posibles del 2º. suceso** **X** **Número de decisiones posibles del 3er. suceso**

Por cada una de las 3 decisiones posibles en el primer suceso, hay 4 en el segundo y, a su vez, 2 en el tercero; por esta razón el total de decisiones posibles se obtiene al multiplicar el número de decisiones que se pueden tomar en cada suceso por separado:

$$3 \times 4 \times 2 = \quad 24$$

Para generalizar lo anterior, llamemos d_1, d_2 y d_3 al número de decisiones que se pueden tomar respectivamente en el primer suceso (s_1), en el segundo (s_2) y en el tercero (s_3); entonces, las decisiones posibles en el suceso s_1 s_2 s_3 son:

$$d_1 \times d_2 \times d_3$$

Lo anterior se conoce como el **Principio de la Multiplicación**, de gran utilidad sobre todo cuando construir el diagrama de árbol resulta poco práctico o innecesario.

Problema. En la Dirección de Tránsito del Distrito Federal desean saber **cuántas** placas para automóvil se pueden elaborar con los 10 dígitos (del 0 al 9) y 26 letras del alfabeto. Cada placa consta de tres números (se pueden repetir pero el primero debe ser diferente de cero), y de tres letras que se pueden repetir.

¿Por qué sería impráctico elaborar un diagrama de *árbol*? ________________________

____

Si *colocar un número o una letra en cada lugar de la placa* se considera un suceso, el **número de decisiones posibles** en cada suceso es (escribe el número que falta):

1er. Núm.	**2º. Núm.**	**3er. Núm.**	**1ª. Letra**	**2ª. Letra**	**3ª. Letra**
9	10	10	26	26	____

Al aplicar el Principio de la Multiplicación se obtiene el total de placas:

$$9 \times 10^2 \times 26^3 = 15\,818\,400$$

Principio de la multiplicación

Si una primera **decisión** (operación o acción) puede tomarse de d_1 formas diferentes, una segunda **decisión** puede tomarse de d_2 formas diferentes, una tercera **decisión** puede tomarse de d_3 formas diferentes, . . . , y así sucesivamente hasta la enésima **decisión,** que puede tomarse de d_n formas diferentes, entonces el número total de formas diferentes en que pueden tomarse las n **decisiones** es:

$$d_1 \times d_2 \times d_3 \times \ldots \times d_n$$

Este principio también se conoce como **Principio Fundamental de Conteo**. ¿De qué otra forma se pudiera llamar al árbol? (para contestar observa las palabras resaltadas en negritas del cuadro anterior).____.

PERMUTACIONES

Al hacer "agua de limón" el resultado es diferente cuando el jugo de limón se mezcla con el azúcar, y después se le agrega el agua, que cuando al agua se le agrega el limón y luego el azúcar.

En otras situaciones, por ejemplo al elegir en una asociación a un presidente, un vicepresidente y un secretario de un grupo de 25 candidatos, el orden en que se haga la elección de los tres puede resultar relevante en cuanto al resultado.

Bajo ciertas condiciones, ¿cómo saber si es necesario tomar en cuenta el orden de los pasos o los elementos? Lo puedes aprender enseguida.

El factorial de un número

En ciertos conteos se utiliza la multiplicación abreviada de números naturales sucesivos, conocida como factorial de un número, y se representa por un signo de admiración.

Así, el factorial de 5 (o 5 factorial) se escribe 5! y es igual al producto de los primeros cinco números naturales:

$$5! = 5 \times 4 \times 3 \times 2 \times 1$$

De la misma manera el factorial de 4: $4! = 4 \times 3 \times 2 \times 1$

El factorial de 1: $1! = 1$

El factorial de k : $k! = k\,(k-1)\,(k-2),\ldots, 2 \times 1$

Por definición el factorial de 0 es 1: **$0! = 1$**

Para ser resueltos, muchos problemas requieren de la aplicación de **técnicas** que permitan contar grandes cantidades de manera rápida. Veamos algunos.

Problema. Diez personas se forman para abordar un autobús, ¿cuántas filas son posibles?

La respuesta se encuentra con facilidad cuando se piensa de manera inductiva, es decir, generalizando a partir de casos con pocos elementos.

Si **dos** personas (J y L) se forman, lo pueden hacer de estas maneras:

JL, LJ (son **2** maneras o ***arreglos sin repetición***, es decir **2!**)

Si se agrega E, las **tres** personas se pueden formar de estas maneras ¿cuál falta?:

EJL JEL JLE

ELJ LEJ _____ (son **6** maneras de formarse, es decir **3!**)

Si el grupo aumenta a **cuatro**, porque se une D, las filas posibles son ¿cuáles faltan?:

DEJL EDJL EJDL EJLD

DJEL JDEL JEDL ______

DJLE JDLE ______ JLED

DELJ _____ ELDJ ELJD

_____ LDEJ LEDJ LEJD

DLJE LDJE LJDE LJED (son **24** maneras, es decir **4!)**

Al seguir la secuencia anterior ya se puede saber de cuántas maneras se podrían formar las diez personas del problema original, o más; muéstralo completando la tabla siguiente:

NÚM. DE PERSONAS	**NÚM DE FILAS POSIBLES**
1	______
2	2! = 2
3	3! = 6
4	4! = 24
5	_______
6	________
10	_____________
n	_____________________

En el problema anterior, **el orden de los elementos es importante**: con una persona que se cambie de lugar ya se tiene una fila diferente.

Cuando el orden de los elementos importa, los arreglos se denominan ***permutaciones***. Además, los elementos en cada arreglo no se repiten, son ***ordenaciones sin repetición***.

Esta es otra manera de resolver el problema anterior: Son **10** personas, entonces el primer lugar lo puede ocupar cualquiera de las **10**, el segundo lugar lo puede ocupar alguna de las **9** restantes. . . y así, hasta el último, que lo podrá ocupar sólo una. O sea:

LUGAR	**1º**	**2º**	**3º**	**4º**	**5º**	**6º**	**7º**	**8º**	**9º**	**10º**
NÚMERO DE PERSONAS QUE LO PUEDEN OCUPAR	10	9	8	7	6	5	5	3	2	1

Al aplicar el **Principio de la Multiplicación**, el número total de filas posibles con 10 personas resulta ser:

$$10 \text{ x } 9 \text{ x } 8 \text{ x}7 \text{ x } 6 \text{ x}5 \text{ x } 4 \text{ x } 3 \text{ x } 2 \text{ x } 1 = 10!$$
$$= 3\,628\,800$$

El número de filas posibles con n elementos es:

$$\mathbf{n! = n(n - 1)\,(n - 2)\,(n - 3). . . (3)\,(2)\,(1)}$$

En adelante, a los arreglos, filas u ordenaciones sin repetición las denominaremos **permutaciones.** Por ejemplo, P_2^2 **(se lee "permutaciones de 2 elementos, tomados de 2 en 2")** denota el número de permutaciones de 2 elementos tomados de 2 en 2 y es igual a 2!, que a su vez es igual a 2.; P_3^3 **(se lee "permutaciones de 3 elementos, tomados de 3 en 3")**, etc.

P_n^n **= n! es el número de permutaciones que se pueden formar con n elementos tomados de n en n**.

Problema: Los cinco fundadores de una Asociación Civil necesitan integrar un Comité Directivo formado por un presidente y un tesorero, a los que seleccionarán de entre ellos mismos ¿cuántos comités directivos podrán formar?

El presidente podría ser cualquiera de los cinco, por lo tanto, el tesorero podría ser cualquiera de los cuatro restantes:

DESIGNACIÓN DEL PRESIDENTE	DESIGNACIÓN DEL TESORERO
5 posibilidades	4 posibilidades

Por el *Principio de la Multiplicación*, los posibles comités directivos, con un presidente y un tesorero seleccionados de entre ellos cinco, es: 5 x 4 = 20. Si además nombran a un secretario, ¿cuántas posibilidades habrían de designar al secretario? ____. Es decir:

NOMBRAMIENTO DEL PRESIDENTE	NOMBRAMIENTO DEL TESORERO	NOMBRAMIENTO DEL SECRETARIO
5 posibilidades	4 posibilidades	3 posibilidades

Por lo tanto, el número de comités directivos posibles, con un presidente, un tesorero y un secretario seleccionados de entre los cinco es 5 x 4 x 3 = 60. Si además decidieran nombrar un vocal, ¿cuántos comités sería posible integrar? Haz el cálculo mental: _____.

Vamos a generalizar mediante esta tabla (escribe lo que falta):

TOTAL DE ELEMENTOS	ELEMENTOS TOMADOS DEL TOTAL	RESULTADOS POSIBLES EXPRESADOS CON FACTORIALES
5	1	5! / (5 – 1)! = 5
	2	5! / (5 – 2)! = 20
	3	5! / (5 – 3)! = 60
	4	__________________
	5	__________________
6	**1**	___________
	2	**6! / (6 – 2)! = 30**
	3	_______________
	5	_______________
	6	_______________
7	**5**	_______________
n	**k**	_______________

Para resolver problemas en los que la manera de ***ordenar*** los elementos es importante, la última expresión de la tabla previa es una fórmula de gran utilidad.

El número de permutaciones posibles al tomar k elementos de un total de n es:

$$P_k^n = \frac{n!}{(n-k)!}$$ **con $k \le n$** **(Se lee: permutaciones de n elementos tomados de k en k).**

FICHA DE TRABAJO XIII.1

1. Un contador recién egresado de una universidad recibe una oferta de empleo por parte de una empresa. Para ello debe presentar su currículum vitae de manera escrita o en disco magnético; de ser aceptado se le asignaría en cualquiera de estas áreas: Recursos Humanos, Control Presupuestal o Finanzas, pudiendo ser contratado por tiempo indeterminado o por honorarios. ¿De cuántas maneras se puede incorporar a la empresa? Construye un diagrama de árbol para ilustrar todos los caminos posibles.

2. Al asignarle a una persona la *Clave Única de Registro Personal* (CURP) se utilizan: 4 letras (de 26), 6 dígitos y 8 letras (de 26). ¿Cuántas CURP diferentes se podrían asignar?

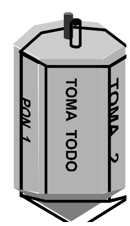
3. El menú de la comida en el restaurante Sep's ofrece para escoger una de: 3 bebidas, 3 sopas, 4 guisados y 5 postres. ¿De cuántas maneras es posible elegir una comida?

4. Resuelve mentalmente: ¿De cuántas maneras se pueden colocar en fila 5 personas?

5. Un examen de opción múltiple consta de 6 preguntas, cada una con 5 posibles respuestas de las cuales una es correcta; a) ¿de cuántas formas diferentes puede un estudiante resolver el examen si sólo puede seleccionar una respuesta para cada pregunta?, b) ¿de cuántas formas diferentes puede resolver el examen seleccionando una respuesta incorrecta para cada pregunta?

6. Escribe *A* en el paréntesis si el fenómeno es aleatorio, *D* si es determinista:

a) Girar una perinola con estos letreros en sus caras: *todos ponen*, *toma todo*, *pon 1*, *pon 2*, *pon todo*, *toma 1* y *toma 2*. ( )



b) La fecha de un temblor de tierra con epicentro en Michoacán. ( )

c) La trayectoria anual de la Tierra alrededor del Sol. ( )

d) El sorteo de la lotería nacional. ( )

e) El cambio del estado físico del agua al calentarse a una temperatura de 100 ºC. ( )

f) El número de personas que abandonarán el país mañana ( )

g) La trayectoria que sigue una bala de cañón al ser disparada ( )

7. Escribe dos ejemplos de fenómeno aleatorio y dos de fenómeno determinista, diferentes a los que leíste.

____

____

____

RESUMEN

Cuando hay varios sucesos consecutivos y cada uno implica varias posibilidades de decisión, el diagrama de árbol es de gran ayuda para visualizar todas las posibles ramas o secuencias de decisión.

Cuando el número de decisiones que pueden tomarse es muy grande, las posibles secuencias de decisión se pueden calcular aplicando el Principio de la Multiplicación, que dice:

Si una primera decisión (operación o acción) puede tomarse de d_1 formas diferentes, una segunda decisión puede tomarse de d_2 formas diferentes, una tercera decisión puede tomarse de d_3 formas diferentes, . . . , y así sucesivamente hasta la enésima decisión, que puede tomarse de d_n formas diferentes, entonces el número total de formas diferentes en que pueden tomarse estas n decisiones es:

$$d_1 \times d_2 \times d_3 \times . . . \times d_n$$

El diagrama de árbol también se conoce como árbol de decisiones, el Principio de la Multiplicación como Principio Fundamental de Conteo.

Las permutaciones de n objetos son los posibles órdenes en los que se pueden presentar los n objetos y se calculan mediante la fórmula: $P_n^n = n!$

Si de n elementos solamente se toman k, entonces el número de permutaciones que se pueden formar es:

$$P_k^n = \frac{n!}{(n-k)!}$$

Si en un conjunto de n elementos hay uno que se repite k veces, otro que se repite r veces, . . ., etcétera, el número de permutaciones se calcula así:

$$\mathbf{P_{k,r}^n = \frac{P_n^n}{P_k^k P_r^r \ldots} = \frac{n!}{k!\,r!}}$$

14ª SESIÓN
COMBINACIONES

COMPETENCIAS POR ADQUIRIR

- **Expresar la fórmula básica para el cálculo de combinaciones.**
- **Resolver problemas que requieran el cálculo de combinaciones.**

COMBINACIONES: OTRA CLASE DE ORDENACIONES. . .

En Estadística dos ordenaciones pueden ser diferentes cuando sus elementos se cambian de lugar. En tal caso se trata de **permutaciones**.

Cuando dos ordenaciones no se distinguen por el orden en que estén dispuestos sus elementos, sino porque éstos son diferentes, se trata de **combinaciones**.

Jorge (J), Paco (P) y Arturo (A) se pueden formar de estas maneras:

JPA PAJ AJP
JAP PJA APJ

Aquí tenemos **seis** permutaciones pero sólo **una** combinación porque se trata de los mismos elementos.

Problema. Los organizadores de un torneo de tenis femenil han publicado una Convocatoria y quieren estar prevenidos acerca del número de partidos que habrán de celebrarse en la primera ronda, sin que haya repeticiones. Cancelarían el torneo en caso de no inscribirse ninguna persona, sólo una o dos.

Si se inscribieran **3**, podríamos saber cuántos partidos se podrían celebrar, para ello representemos a las 3 jugadoras por t_1, t_2, t_3. En ese caso, los **partidos posibles** serían:

$\mathbf{t_1}t_2$, $\mathbf{t_1}t_3$ $\mathbf{t_2}t_3$

Con **3** jugadoras, el total de partidos de dos en dos sería: ____.

En caso de inscribirse **4** jugadoras, los partidos posibles serían:

$\mathbf{t_1}t_2$, $\mathbf{t_1}t_3$, $\mathbf{t_1}t_4$
$\mathbf{t_2}t_3$, $\mathbf{t_2}t_4$
$\mathbf{t_3}t_4$

¿Cuántos partidos (de dos en dos) se podrían celebrar con **4** jugadoras?: ____.

Si fueran **5** jugadoras, los **partidos posibles** serían:

$\mathbf{t_1}t_2$, $\mathbf{t_1}t_3$, $\mathbf{t_1}t_4$, $\mathbf{t_1}t_5$
$\mathbf{t_2}t_3$, $\mathbf{t_2}t_4$, $\mathbf{t_2}t_5$
____ ____
____

El total de partidos (de dos en dos) que podrían celebrar **5** jugadoras es ____.

¿Cuántos partidos se tendrían que organizar en caso de que se inscribieran, por ejemplo, 12 jugadoras en el torneo? Con el método que se resolvieron los casos anteriores se necesitaría de mucho tiempo, mejor busquemos una fórmula.

Para ese propósito COMPLETA la tabla:

NÚMERO DE JUGADORAS	**JUGADORAS POR PARTIDO**	**PERMUTACIONES**	**PARTIDOS**	**RELACIÓN ENTRE PERMUTACIONES Y PARTIDOS**
3	2	${}_3P_2 = 6$	3	$\frac{P_2^3}{2!} = 3$
4	2	${}_4P_2 = 12$	6	$\frac{P_2^4}{2!} = 6$
5	2	${}_5P_2 = 20$	10	$\frac{P_2^5}{2!} = 10$
6	2	${}_6P_2 =$ ____	15	________________
n	**2**	_______________	________	________________

Analicemos el resultado para n:

Puesto que $P_2^n = \frac{n!}{(n-2)!}$, entonces: $\frac{P_2^n}{2!} = \frac{n!}{(n-2)!\,2!}$

Al cociente del lado derecho de la igualdad se le conoce como el *número de combinaciones posibles con* ***n*** *elementos, tomados de dos en dos*, y se expresa así:

$$C_2^n = \frac{n!}{(n-2)!\,2!}$$ **(Se lee: combinaciones de n elementos tomados de 2 en 2)**

En general, si $k \le n$: $C_k^n = \frac{n!}{(n-k)!\,k!}$ **(Se lee: combinaciones de n elementos tomados de k en k)**. . .(I)

Combinación y permutación son dos conceptos distintos. Una combinación se utiliza cuando en un problema **no importa el orden** de los elementos, por ejemplo, $\mathbf{t_1}t_2$ en el ejemplo anterior es la misma combinación que $t_2\mathbf{t_1}$, porque es el mismo partido.

Problema: En una empresa se designarán tres gerentes que serán seleccionados de entre 10 candidatos, ¿cuántas posibilidades hay de hacer la designación?

Podemos empezar manejando números menores: Si fueran tres candidatos sólo habría una posibilidad de hacer la designación ¿qué pasaría si hubiera **4** candidatos?

Si se representan con c_1, c_2, c_3, c_4 a los candidatos, la terna se podría formar:

Al tomar a $\mathbf{c_1}$ con otros dos:

$\mathbf{c_1}c_2c_3$ $\mathbf{c_1}c_2c_4$ $\mathbf{c_1}c_3c_4$ (hay que notar que por ejemplo la terna $\mathbf{c_1}c_2c_3$ es la misma que $c_3\mathbf{c_1}c_2$ y que $c_2c_3\mathbf{c_1}$, es decir, **no importa el orden de los elementos**).

La terna faltante con $\mathbf{c_2}$ sería: $\mathbf{c_2}c_3c_4$

¿Faltarían otras ternas? ____. ¿Cuántas ternas resultaron con **4** elementos? ____.

En el caso de que fueran **5** candidatos:

Las ternas con $\mathbf{c_1}$ serían: $\mathbf{c_1}c_2c_3$ $\mathbf{c_1}c_2c_4$ $\mathbf{c_1}c_2c_5$ $\mathbf{c_1}c_3c_4$ $\mathbf{c_1}c_3c_5$ $\mathbf{c_1}c_4c_5$

Las restantes con $\mathbf{c_2}$ serían: $\mathbf{c_2}c_3c_4$ $\mathbf{c_2}c_3c_5$ $\mathbf{c_2}c_4c_5$

Sólo faltaría una terna, ¿cuál?:_________

¿Cuántas ternas (combinaciones) resultaron con **5** elementos? _____.

COMPLETA LA TABLA:

CANDIDATOS	DESIGNADOS	NÚM. DE TERNAS	COMBINACIONES
3	3	1	$\frac{3!}{(3-3)!3!}=1$
4	3	4	$\frac{4!}{(4-3)!3!}=4$
5	3	10	$\frac{5!}{(5-3)!3!}=10$
6	3	20	______________________________
10	3	____	______________________________
n	3	_______________	___________

La expresión de abajo a la derecha en la tabla es el número de combinaciones que se pueden formar de un conjunto de n elementos tomados de 3 en 3:

$$C_3^n=\frac{n!}{(n-3)!3!} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(II)}$$

Observemos las expresiones (I) y (II) que se obtuvieron al hacer las generalizaciones:

$$C_2^n=\frac{n!}{(n-2)!2!} \quad \ldots\ldots\ldots\ldots\ldots\text{(I)}$$

$$C_3^n=\frac{n!}{(n-3)!3!} \quad \ldots\ldots\ldots\ldots\ldots\text{(II)}$$

¿Cuál es la expresión para calcular el número de combinaciones de un total de

n elementos, tomados de **4** en **4**? Escríbela a continuación: ______________________.

¿Cuál es la expresión para **n** elementos tomados de **5** en **5**? ______________________.

Si hay n elementos en total, el número de combinaciones que se pueden formar tomando elementos de k en k es: $C_k^n=\frac{n!}{(n-k)!k!}$ **con k ≤ n**

La expresión anterior se aplica cuando se necesite formar arreglos donde el orden no interese.

Problema: Un jefe policiaco en el D. F. tiene bajo su mando a 30 policías y debe seleccionar de entre ellos a un grupo de 5 para una misión especial en una de las 16 Delegaciones. ¿Cuántos grupos podría formar?

Lo primero que se debe determinar es si el orden importa o no. En este caso no es relevante porque, por ejemplo, el conjunto $p_1p_7p_{11}p_{14}\ p_{18}$ es el mismo que otro en el que los elementos estén ordenados de diferente manera.

Enseguida conviene identificar **n** y **k**: n = 30, k = 5

Se sustituyen los valores en la fórmula y se calcula el resultado:

$$C_k^n = \frac{n!}{(n-k)!\,k!} = \frac{30!}{(30-5)!\,5!}$$

$$= \frac{30x29x28x27x26x\cancel{25x...2x1}}{\cancel{(25x24x23x22x21...2x1)}\,x5x4x3x2x1}$$

Antes de hacer operaciones conviene cancelar factores en el numerador y denominador:

$$= \frac{\cancel{30}\,x\,29\,x\,\cancel{28}\,x\,27\,x\,26}{\cancel{5}\,x\,\cancel{4}\,x\,\cancel{3}\,x\,\cancel{2}\,x\,\cancel{1}}$$

$$=29x7x27x26$$

$$= 142\ 506 \text{ grupos}$$

Problema. El coronel Ricardo Nettel es productor de cine y les pide a **tres** fotógrafos, que buscan trabajo, presentarse cierto día para someterse a las pruebas de capacidad. Ricardo duda que asistan todos y quiere saber cuántas son las posibilidades de asistencia para preparar el material necesario.

De haber citado a **cero** fotógrafos diríamos que a Ricardo no le funciona bien la cabeza; sin embargo matemáticamente sí tiene sentido:

Asisten: 0

Número de posibilidades: **1**

Si hubiera citado a **un** fotógrafo (f_1):

Asiste uno o ninguno:	1	0
	f_1	
Número de posibilidades:	**1**	**1**

Si hubiera citado a **dos** fotógrafos (f_1 y f_2):

Asisten dos, uno o ninguno:	2	1	0
	$f_1\ f_2$	f_1	
		f_2	
Número de posibilidades:	**1**	**2**	**1**

Ricardo en realidad citó a **tres** fotógrafos (f_1, f_2 y f_3):

Asisten tres, dos, uno o ninguno**:**	3	2	1	0
	$f_1\ f_2\ f_3$	$f_1\ f_2$	f_1	
		$f_1\ f_3$	f_2	
		$f_2\ f_3$	f_3	
Número de posibilidades:	**1**	**3**	**3**	**1**

Al ordenar en renglones las diferentes posibilidades obtenidas (marcadas con negritas) en el problema anterior, resulta un arreglo simétrico conocido como Triángulo de Pascal (identifica el patrón y completa):

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

__ __ __ __ __ __

Si de cada número en el triángulo de Pascal escribimos su equivalente número de combinaciones, se obtiene el mismo triángulo presentado de manera distinta (escribe las expresiones faltantes):

C_0^0

$C_0^1 \quad C_1^1$

$C_0^2 \quad C_1^2 \quad C_2^2$

$C_0^3 \quad C_1^3 \quad C_2^3 \quad C_3^3$

$C_0^4 \quad C_1^4 \quad C_2^4 \quad C_3^4 \quad C_4^4$

___ ___ ___ ___ ___ ___

Analiza el triángulo de Pascal y menciona tres de sus propiedades:

1. ____
2. ____
3. ____

Problema. Fausto Gómez participa en un concurso televisivo. Le presentan un tablero con 9 casilleros encubiertos y tiene derecho a seleccionar tres: a) ¿cuántas selecciones le es posible hacer?
Si uno de los casilleros contiene un premio, b) ¿cuántas selecciones de tres casilleros puede hacer de tal manera que uno sea el premiado?; c) ¿cuántas selecciones de tres casilleros puede hacer de tal manera que ninguno de los 3 sea el premiado?

a) El orden en que sean seleccionados los tres casilleros no importa, entonces el total de selecciones posibles lo sabemos al aplicar la relación $C_k^n = \frac{n!}{(n-k)!\,k!}$ con n = 9 y k = 3:

$$C_3^9 = \frac{9!}{(9-3)!\,3!} = \frac{9x8x7x6x\ldots x2x1}{6x5x4x\ldots 2x1\,x3x2x1} = \frac{9x8x7}{3x2x1}$$

$$= 84$$

b) Un casillero seleccionado debe ser el premiado, por lo tanto los otros dos se obtienen seleccionándolos de los restantes 8 casilleros:

POSIBLES SELECCIONES DE LOS 3 CASILLEROS	
1ª.SELECCION	2ª y 3ª SELECCIONES
1	$C_2^8 = \frac{8!}{(8-2)!\,2!}$ = 28

Por lo tanto, el número de maneras en que se pueden seleccionar 3 casilleros de tal manera que uno sea el premiado es: $1 \times 28 = 28$

c) Considerando los resultados anteriores, ¿de cuántas maneras puede Fausto seleccionar los 3 casilleros de manera que no obtenga el premiado? _____.

RESUMEN

Las combinaciones de n objetos tomados de m en m son los posibles subconjuntos de m términos tomados de un conjunto de n objetos, en donde el orden no es importante, y se calculan mediante la fórmula:

$$C_m^n = \frac{n!}{(n-m)!\,m!}$$

Las posibles combinaciones para 0, 1, 2, 3, . . ., elementos, de un total de n, se pueden organizar en un “triángulo de Pascal”, que es un arreglo simétrico con dos lados formados con números 1, y los elementos de un renglón intermedio se obtienen sumando los dos que quedan arriba.

FICHA DE TRABAJO XIII.1

1.- En un restaurante están interesados en conocer las costumbres de las personas en cuanto al orden en que ingieren los alimentos durante el desayuno. Si ofrecen paquetes que constan de pan de dulce, guisado, jugo, fruta y café, ¿de cuántas maneras podrían tomar esos alimentos, suponiendo que los ingieren uno tras otro?

2. Un egresado de la secundaria llena su solicitud de ingreso a una institución de educación media superior. Si debe manifestar 6 preferencias (sin repetir) y tiene 9 para elegir, ¿de cuántas maneras puede llenar su solicitud?, ¿de cuántas si puede repetir?

3. Se tienen 7 libros y sólo espacio para 3 en un librero. Calcula de cuántas maneras se pueden colocar 3 libros elegidos de los siete dados, suponiendo que no existen razones para preferir alguno.

4. En un hospital se cuenta con 10 cirujanos para organizar guardias de 5, ¿cuántas guardias se pueden organizar?

5. Una quiniela en México consta de seis números naturales tomados de un total de 56. Si no se pueden repetir números, ¿de cuántas maneras es posible hacer la quiniela?

6. Una pintura se forma mezclando al azar tres litros de pintura diferentes de un total de 8 colores disponibles. ¿Cuántos colores diferentes pueden crearse?

7. Calcula y saca una conclusión: a) C_4^9 y C_5^9 b) C_3^{13} y C_{10}^{13} c) C_7^{11} y C_4^{11}

8. a. ¿De cuántas maneras se puede contestar un examen de 3 preguntas, cada una con opciones de falso-verdadero? b. ¿ De cuántas si consta de 10 preguntas, cada una con cinco opciones?

TAREA

ESTUDIA SIN DISTRAERTE

TEORÍA DE CONJUNTOS

Conjunto

Un conjunto se puede definir de dos maneras: ***por comprensión*** y ***por extensión***.

DEFINICIÓN. Un conjunto se define ***por comprensión*** al dar una propiedad que caracterice a sus elementos.

NOTACIÓN. Un conjunto **A** definido por comprensión mediante la propiedad **P** se denota así:

A = {*x/x tiene la propiedad* P} **Se lee: A *es el conjunto de elementos x tal que x tiene la propiedad* P. Las *llaves* { } se leen "conjunto de los" la línea diagonal / se lee "tal que".**

Los conjuntos se denotan con letras mayúsculas. Los elementos, de ser el caso, con minúsculas.

EJEMPLO: E = {*x/x es entero primo mayor que* 6 *y menor que* 17}

P = {*x/x es el nombre del primer presidente de México*}

Todos lo elementos de E son: 7, 11 y 13. El único elemento de P es Guadalupe Victoria.

DEFINICIÓN. Un conjunto se define ***por extensión*** o ***enumeración*** si se da una lista explícita de todos sus elementos.

EJEMPLO. Un conjunto definido ***por extensión*** se denota colocando entre llaves los nombres de sus elementos, separados por comas. Los conjuntos E y P anteriores se denotan por extensión así:

E = {7, 11, 13} y P = {Guadalupe Victoria}

PERTENENCIA Y CONTENCIÓN

Si x es elemento del conjunto A se dice que ***x pertenece*** a A y se escribe $x \in$ A.

Para indicar que ***x no pertenece*** a A se escribe $x \notin$ A.

Por ejemplo, si M es el conjunto de los muralistas mexicanos del s. XX:

Diego Rivera $\in$ M, Luis Miguel $\notin$ M

DEFINICIÓN. Se dice que un conjunto **A** es **subconjunto** del conjunto **B**, se denota **A** $\subset$ **B**, si todos los elementos que pertenecen a **A** pertenecen también a B.

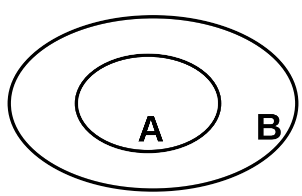


Fig. 1

Si **A** $\subset$ **B**, se dice también que el conjunto B ***incluye*** al conjunto A, y se denota B $\supset$ A (se lee B incluye a A, o bien B comprende a A).

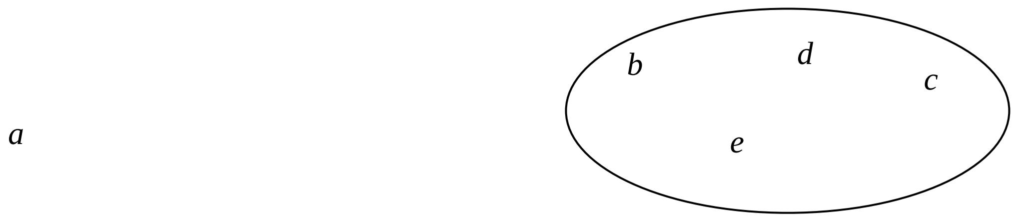
Para indicar que A ***no está incluido*** en B se escribe A $\not\subset$ B.

Las relaciones A $\subset$ B y B $\supset$ A son equivalentes por definición y se ilustran en este diagrama:

Los diagramas como los de la figura 1 se llaman ***diagramas de Venn***.[1]

Por ejemplo, para representar el conjunto de las vocales del alfabeto, se hace un dibujo como este:

[1] John Venn Venn (1834–1923) fue un lógico británico que se hizo famoso por sus diagramas.



LAS OPERACIONES DE BOOLE[2]

Intersección. Conjuntos ajenos, conjunto vacío.

DEFINICIÓN. Se llama ***intersección*** A ∩ B (se lee A intersección B) de dos conjuntos A y B, al conjunto de elementos que pertenecen **a la vez** a A y a B. **Se asocia con el conectivo "y".**

$$A \cap B = \{x/x \in A \ \ \boldsymbol{y} \ \ x \in B\}.$$

EJEMPLOS:

a. Si A = {2, 4, 6, 8} y B = {-5, 0, 4, 6, 9}, entonces A ∩ B = {4, 6}.

b. En una fábrica la intersección del conjunto de obreros que son casados C, con el conjunto de obreros que tienen hijos H, es el conjunto de obreros O que a la vez están casados **y** tienen hijos: C ∩ H = O.

c. En una oficina hay 4 secretarias de ojos cafés y 7 de pelo negro. Si hay 3 que tienen los ojos cafés y el pelo negro, representa esta situación mediante un diagrama de Venn.

DEFINICIÓN. Se llama intersección de dos o más conjuntos al conjunto cuyos elementos son los que pertenecen a todos ellos a la vez.

Por ejemplo, la intersección de los conjuntos:

A = {a, e, i, o, u}, B = {g, a, t, o}, C = {p, a, s, i, t, o}

que se denota A ∩ B ∩ C, es el conjunto: A ∩ B ∩ C = {a, o}.

Dos conjuntos son ***ajenos*** si no tienen ningún elemento en común. Dos conjuntos A y B son ajenos si y sólo si su intersección es el conjunto **vacío**: A ∩ B = ϕ.

El conjunto **vacío** es aquél que no tiene elementos, generalmente se denota con ϕ (fi) o mediante { }.

DEFINICIÓN. Se llama ***unión*** A ∪ B (se lee A unión B) de dos conjuntos A y B al conjunto cuyos elementos pertenecen o bien a A, o bien a B, o bien a A y a B. **Se asocia con el conectivo "o" inclusivo.**

$$A \cup B = \{x/x \in A \ \ \mathbf{o} \ \ x \in B \ \mathbf{o} \text{ en ambos}\}.$$

EJEMPLO. Si A = {1, 2, 4} y B = {2, 3}, entonces A ∪ B = {1, 2, 3, 4}.

DEFINICIÓN. La ***unión*** de varios conjuntos es un conjunto cuyos elementos son los que pertenecen a alguno (a uno por lo menos) de ellos.

EJEMPLO. Sean los conjuntos:

[2] George Boole: matemático inglés (1815 – 1864). Hizo aportaciones muy importantes a la lógica y a la teoría de probabilidades.

$P = \{1, 2, 4\} \quad Q = \{2, 3\}$ y $R = \{3, 4, 5, 6\}$

Entonces:

$$P \cup Q = \{1, 2, 3, 4\}$$

$$P \cup Q \cup R = \{1, 2, 3, 4, 5, 6\}$$

El **conjunto universal** U es el conjunto que incluye a todos los conjuntos que estén siendo considerados.

Al considerar conjuntos supondremos que todos ellos son subconjuntos de U.

DEFINICIÓN. Se llama ***diferencia*** A - B (se lee A menos B) de dos conjuntos A y B, al conjunto cuyos elementos pertenecen a A, pero no a B**.**

$$A - B = \{x/x \in A \ \textbf{y} \ x \notin B\}.$$

EJEMPLO. Si $A = \{1, 2, 3, 4\}$ y $B = \{2, 3\}$, entonces $A - B = \{1, 4\}$.

DEFINICIÓN. La ***diferencia*** B - A es el conjunto de elementos que pertenecen a B, pero no a A.

EJEMPLO. Si $A = \{1, 2, 3, 4\}$ y $B = \{2, 3\}$, entonces $B - A = \{\ \}$.

$$A - B = \{1, 4\}$$

¿Es B – A igual que A – B? ____.

DEFINICIÓN. Sea A un conjunto. Llamaremos ***complemento*** de A (respecto de U), y lo denotaremos con A′, al conjunto de los elementos de U que ***no pertenecen a* A**.

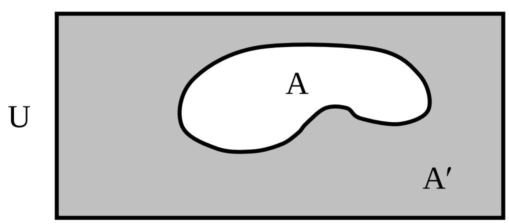


Fig. 2

En el diagrama Venn de la izquierda, el rectángulo representa al conjunto universal U, y la parte sombreada al conjunto A′, el complemento de A.

¿Es A′ = U – A? ____, ¿Es A = U - A′? ____.

Comprueba si comprendiste:

Dados los siguientes conjuntos:

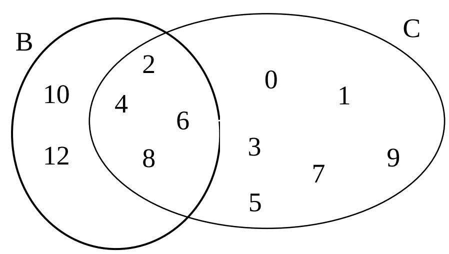
$U = \{x/x \ \varepsilon \ Z, -3 \le x \le 12\}$ (conjunto universal)

B = {x/x es par positivo, x ≤ 12}
C = {x/x es dígito}
D = {x/x ε Z, 8 ≤ x < 12}
E = {x/x ε Z, 0 < x < 6}

1. Expresa en forma extendida los conjuntos U, B, C, D y E.

____
____
____
____
____

2. Calcula y representa mediante un diagrama de Venn las siguientes operaciones (en la figura coloca los elementos que le correspondan a cada conjunto, como en el inciso **d.**

a. $A \cup A$: ______________________	**b.** $C \cup D$: ______________________
c. $B \cup E$: ______________________ 2 4 6 8	**d.** $B \cap C$: {2, 4, 6, 8} 

e. $C \cap D$: ______________________	**g.** $D \cap E$: ______________________

g. U – C: ______________________________	**h.** B – D: ______________________________
i. D – B: ______________________________	**j.** (U – C): ____________________________
k. B´: ________________________________	**l.** C´: ________________________________
m. (B $\cup$ E)´:___________________________	**n.** (B $\cap$ C)´: ___________________________
ñ. (B $\cap$ C) – (C $\cap$ D): ____________________	**o.** U´: ___ **p.** ϕ´: ____ **q.** B´$\cap$ B: ___ **r.** (A $\cap$ B) – (B $\cap$ A): ___

15ª SESIÓN
FENÓMENOS DETERMINISTAS Y ALEATORIOS
AXIOMAS DE LA PROBABILIDAD

COMPETENCIAS POR ADQUIRIR

- **Proponer ejemplos de fenómenos deterministas y aleatorios.**
- **Identificar un fenómeno por su carácter aleatorio o determinista.**
- **Expresar por escrito los axiomas básicos de la probabilidad y su significado.**
- **Aplicar los axiomas básicos de la probabilidad en la resolución de problemas.**

MATERIAL NECESARIO: ESCUADRAS, CALCULADORA, LÁPIZ, GOMA Y SACAPUNTAS.

FENÓMENOS DETERMINISTAS Y ALEATORIOS

Un propósito de la ciencia es descubrir las leyes que gobiernan a los fenómenos. Las leyes permiten hacer **predicciones** con precisión, tanto como lo permitan los instrumentos de medición disponibles.

Para fenómenos donde interviene en forma preponderante el **azar**, la predicción no siempre es posible. A fenómenos de este tipo se les llama **aleatorios**; su principal característica es la **imposibilidad de predecir con exactitud** su estado final.

Con excepción de unos cuantos sucesos, en realidad la mayoría de los fenómenos son aleatorios debido a nuestra imposibilidad de predecir el resultado final con total precisión; pero aun cuando esto fuera posible, en muchos casos no podríamos comprobarlo porque no siempre disponemos de instrumentos y técnicas apropiados.

¿Exactitud y **precisión** son sinónimos?

Precisión se refiere a la **unidad de medida** usada; si una medida es 35 metros se dice que se hizo con **precisión** de metros, si es de 35.2 m se hizo con precisión de décimas de metro o decímetros.

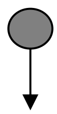
Cuando una medida consiste en contar es **exacta**, en otro caso sólo es una **aproximación**, buena o mala dependiendo del instrumento y del método que se utilice.

Podemos “medir” con exactitud el número de monedas que traemos en los bolsillos; en cambio sólo podremos dar una aproximación de nuestra estatura, sin importar el instrumento y el método empleado para obtenerla.

En la práctica no exigimos 100% de exactitud al medir, nos conformamos con cierto grado de aproximación. De esta manera las leyes se pueden aplicar en forma más o menos satisfactoria a ciertos fenómenos, que así dejan de considerarse **aleatorios** para denominarse **deterministas**.

A diferencia de los aleatorios, los fenómenos deterministas son aquellos para los cuales se han podido construir modelos que los describen en forma **satisfactoriamente aproximada**.

La caída de un cuerpo, por ejemplo, es un fenómeno determinista porque tenemos un modelo para describirlo:

FENÓMENO	MODELO DESCRIPTIVO DEL FENÓMENO
Caída de un cuerpo en el vacío 	$y = \frac{1}{2}gt^2$ Donde ***y*** es la altura del cuerpo sobre el piso en un tiempo ***t***; ***g*** es una constante con un valor aproximado de 9.81 m/s^2.

Un cuerpo de cierta masa, en el vacío, tarda un segundo en caer desde una altura de 4.905 metros; con el modelo anterior se puede predecir que otro cuerpo de diferente masa, soltado desde la misma altura, tardará en caer un tiempo igual, ¿por qué?

Entre los fenómenos aleatorios a partir de cuyo estudio se han obtenido elementos importantes para el desarrollo de la ciencia, están ciertos juegos de mesa: echar las cartas, lanzar un dado, una moneda, etc. Pero hay otros fenómenos aleatorios de interés más directo para los científicos, como la trayectoria que sigue una molécula en un gas, el sexo de un niño al engendrarse en el seno materno, el tiempo que tarda un automovilista en ir de un lugar a otro en la ciudad de Guanajuato, etcétera.

AXIOMAS DE LA PROBABILIDAD

¿QUÉ SON LOS AXIOMAS?. . .

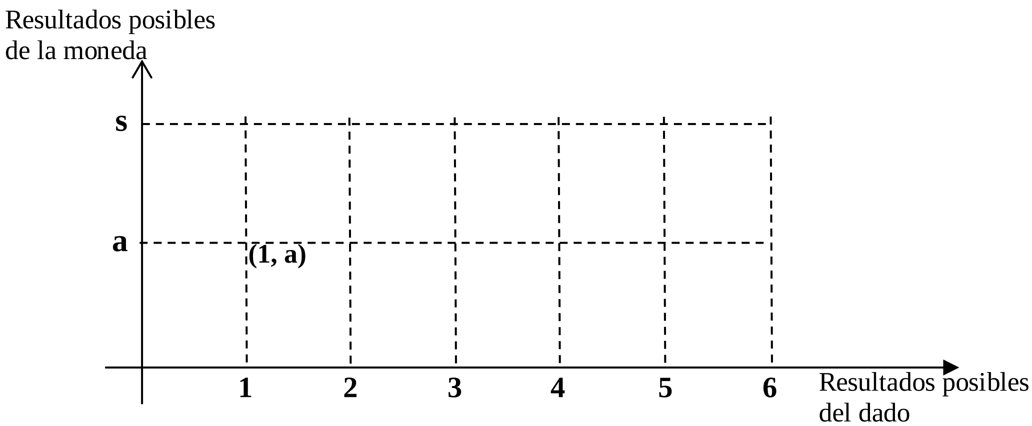
Una teoría matemática se puede imaginar como un edificio cuyos cimientos para sustentarla son, entre otros, los axiomas, entendidos como afirmaciones que sin requerir de una demostración se aceptan como verdaderas.

La teoría de la probabilidad se apoya en un entretejido de axiomas básicos que dan lugar a leyes probabilísticas y constituyen, a su vez, una herramienta para resolver problemas y la plataforma de partida hacia conocimientos de mayor complejidad.

LEE SIN INTERRUPCIONES Y CONTESTA

Experimento: *Se lanzan al mismo tiempo un dado de seis caras y una moneda.*

Los resultados se registran en el plano de abajo, donde las intersecciones de las líneas punteadas representan los resultados posibles del experimento, escribe los que faltan.



¿Cuántos resultados posibles tiene el experimento? ____. El conjunto de resultados posibles, o **espacio muestral**, lo representaremos con Ω.

¿Del total de resultados posibles, en cuántos se obtiene "águila"? ____ .

El conjunto de resultados donde se obtiene "águila" es sólo una parte o *subconjunto* del espacio muestral; cualquier **subconjunto** del espacio muestral se llama **evento.**

La probabilidad **clásica** o ***a priori*** de un evento E se define así:

$$P(E) = \frac{\text{número de resultados favorables}}{\text{número de resultados posibles}}$$

Los resultados favorables son los resultados que contiene el evento E.
Los resultados posibles son los resultados que contiene el espacio muestral Ω.

En el cociente anterior, al lanzar la moneda y el dado: ¿El número de resultados favorables (numerador), es siempre *positivo*? ____. ¿Podría ser *negativo*? ____ ¿Podría ser *cero?* ____.

El número de resultados posibles, los que contiene Ω (denominador), ¿es *positivo o negativo*? __________ ¿Podría ser *negativo?* ____ ¿Podría ser *cero?* ____ .

Ya que tanto el numerador como el denominador son siempre positivos, al dividir resulta que la probabilidad de cualquier evento siempre será *¿positiva o negativa?* ____________. Si el numerador fuera *cero*, la probabilidad del evento sería _____. Cuando la probabilidad de un evento es ***cero***, se dice que se trata de un evento **imposible.**

En símbolos: **$P(E) \geq 0$** **(la probabilidad de un evento E es mayor o igual a cero)** . . . . **(1er Axioma)**

En el caso anterior (lanzar un dado y una moneda), ¿cuántos *resultados posibles tiene* Ω? ____. Si el evento fuera el propio Ω, ¿cuántos *resultados posibles* tendría el evento? ____ . Al dividir estas dos cantidades según la definición de probabilidad de un evento, ¿qué resulta? ___ .

En notación matemática: **P(Ω) = ___** . . . . **(2º Axioma)**

Cuando la probabilidad de un evento es ***uno***, se dice que se trata de un evento ***seguro.*** Por lo tanto, el espacio muestral es un evento seguro.

Para que **P(E)** fuera mayor que 1, ¿cómo debería ser *el numerador comparado con el denominador, mayor, menor o igual?* _______. *¿Sería esto posible?* ___ .

1.- Contesta basándote en la definición clásica de probabilidad:

a) ¿Por qué la probabilidad de un evento no puede ser negativa? ________________________
____

b) ¿Qué debe pasar para que la probabilidad de un evento sea cero? ______________________
____

c) ¿Cual es el valor máximo de la probabilidad de cualquier evento? ___ .

d) ¿En qué caso la probabilidad de un evento es mayor que 1? __________________________

2.- Para un evento X, escribe en forma matemática los dos axiomas citados de la probabilidad.

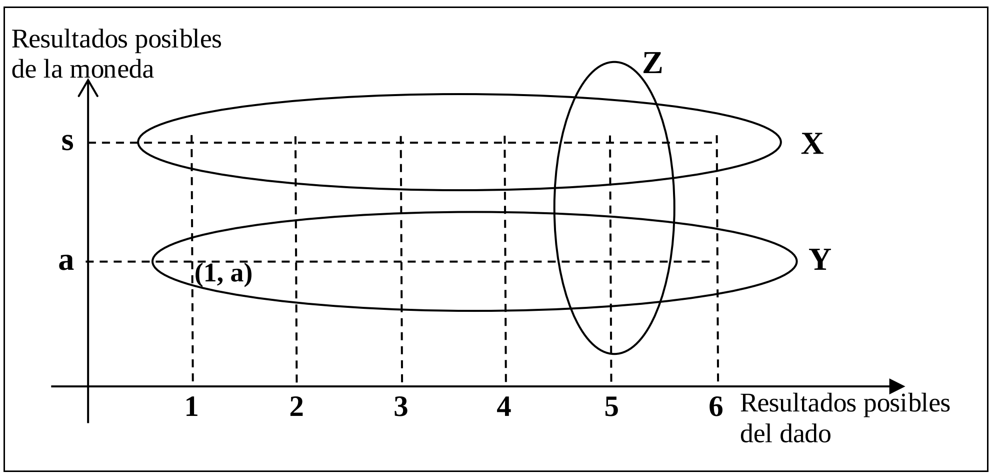
________ y ________

En el diagrama del experimento que nos ocupa se han encerrado los eventos X, Y y Z:

X = obtener "sol" (con cualquier número).
Y = obtener "águila" (con cualquier número).

Z = obtener cinco (con cualquier cara de la moneda)

Los eventos X y Y, ¿tienen elementos comunes? ___ .



Si dos eventos no tienen elementos en común, son **mutuamente excluyentes.**

Los eventos X y Z ¿son mutuamente excluyentes? ____.

El número de resultados posibles del evento X es ____ .Por lo tanto, $P(X) = \frac{\quad}{\quad} =$

El número de resultados posibles del evento Y es ___ .Entonces, su probabilidad es:

$$P(Y) = \frac{\quad}{\quad} =$$

Piensa en el evento X∪Y (X unión Y) formado tanto por los elementos de X como por los de Y. Este evento se asocia con la disyunción "**o**":

X∪Y = La moneda cae "sol" **o** "águila" (con cualquier número del dado).

¿Cuántos son los resultados del evento X∪Y, es decir, los resultados favorables? ____ . Entonces, la probabilidad de X∪Y es:

P(X∪Y) = _____ = _____

Compara los tres últimos resultados y escribe la relación de $P(X \cup Y)$ con $P(X)$ y $P(Y)$.

________________________________ .

Para dos eventos **mutuamente excluyentes** se cumple que $P(X \cup Y) = P(X) + P(Y)$

Para tres eventos mutuamente excluyentes X, Y, W:

$P(X \cup Y \cup W) =$ ____________________.

Para ***n*** **eventos mutuamente excluyentes**:

$P(E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_n) =$____________________________. . . . . . . (3er Axioma)

Los tres axiomas enunciados se conocen como ***Axiomas básicos de la probabilidad.***

Cada elemento contenido en un evento se denomina **evento elemental.** El espacio muestral **Ω** del lanzamiento de una moneda y un dado tiene 12 **eventos elementales**:

(1,a) (2,a) (3,a) (4,a) (5,a) (6,a) (1,s) (2,s) (3,s) (4,s) (5,s) (6,s)

La probabilidad de cada evento elemental es $\frac{1}{12}$, en consecuencia, sus probabilidades suman 1:

$$\underbrace{\frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \ldots + \frac{1}{12}}_{\text{(12 sumandos)}} = 1$$

Los eventos elementales del evento Y = *obtener "águila"* (con cualquier número) son: (1,a) (2,a) (3,a) (4,a) (5,a) (6,a) y tiene una probabilidad de ocurrencia $P(Y) = \frac{6}{12} = \frac{1}{2}$.

La probabilidad de cada evento elemental es $\frac{1}{12}$; por lo tanto la suma de las probabilidades de los eventos elementales del evento Y es:

$$\frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = \frac{6}{12} = \frac{1}{2} \text{ (igual a la probabilidad del evento Y)}$$

Lo anterior se puede generalizar así:

La probabilidad de un evento es igual a la suma de las probabilidades de sus eventos elementales. . . . . (4º Axioma)

Problema. En el auxilio prestado a una comunidad de 6 000 familias tabasqueñas, afectadas por inundaciones, se proporcionó: despensa a 3 000 familias, ropa a 1 500 familias y ambas cosas (despensa y ropa) a 1 200 familias.
En el diagrama de Venn de abajo, los eventos señalados son:

D: recibió despensa **R**: recibió ropa

Por lo tanto, D∩R: recibió despensa **y** ropa.

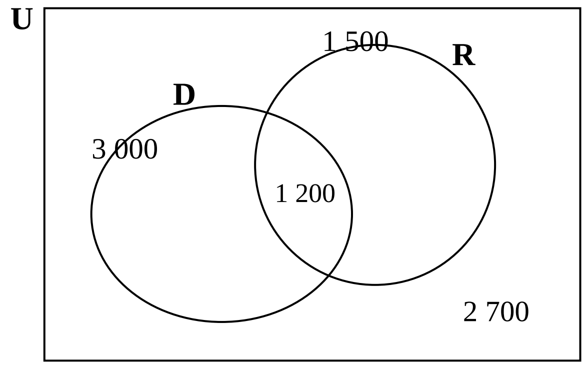


Diagrama de Venn

¿Cómo se obtuvo 2 700 y qué indica?

____

____

____

Calcula las probabilidades de ocurrencia de los siguientes eventos:

$P(D) = \frac{\quad}{\quad} =$ $P(R) = \frac{\quad}{\quad} =$ $P(D \cap R) = \frac{\quad}{\quad} =$ $P(D \cup R) = \frac{\quad}{\quad} =$

Con relación a los resultados anteriores, escribe + ó − para hacer válida la igualdad:

$$P(D \cup R) = P(D)\_\_\_P(R)\_\_\_P(D \cap R)$$

En general: **P(X ∪ Y) = P(X) + P(Y) – P(X ∩ Y). . . . .(5º Axioma)**

Se selecciona una familia al azar, ¿cuál es la probabilidad de que **no** haya recibido despensa?

Como D es el evento "recibió despensa", entonces su **negación** será el evento **complemento** D': ***no*** *recibió despensa*, cuya probabilidad de ocurrencia es (escribe lo que falta):

$$P(D') = \frac{\quad}{\quad}$$

$$=$$

Es decir: P(D) + P(D') = ___

o bien: P(D') = __ – P(D)

Como el evento R es "recibió ropa", entonces su **negación** es el evento **complemento,** R': *no* recibió *ropa* y la probabilidad de que ocurra es: $P(R') = \frac{4\,500}{6\,000}$

$= 0.75$

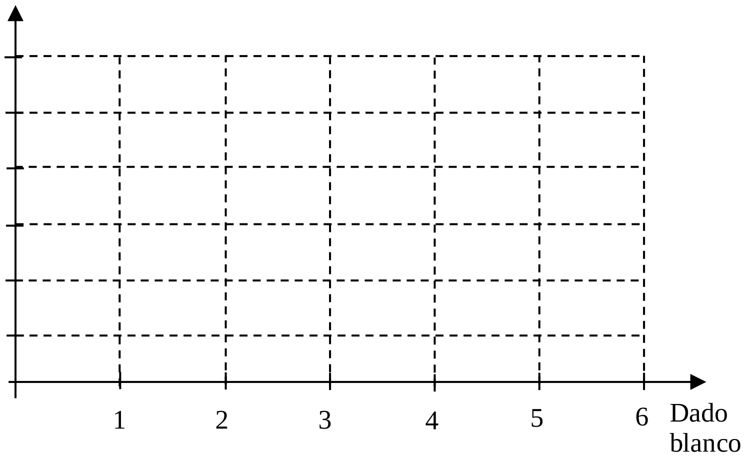
Por lo tanto: P(R) + P(R') = ___ o también: P(R') = __ – P(R)

En general: **P(X') = 1 – P(X)**. . . . . **(6º Axioma)**

¿EN QUÉ CONSISTEN LOS SEIS AXIOMAS DE LA PROBABILIDAD VISTOS?. . .

I. **Experimento**: *Se lanzan dos dados regulares (uno rojo y otro blanco) de seis caras.*

Completa el plano cartesiano de abajo y localiza en él, mediante puntos, todos los resultados posibles del experimento. Junto a cada punto escribe las respectivas parejas de coordenadas.



Sean los eventos: M: *los dados caen con el mismo número*; D: *el dado rojo cae 2;* C: *el dado blanco cae 4*; T: *el dado blanco cae 3* y Ω: espacio muestral. Enciérralos en el plano escribiendo la letra que los identifica.

Ahora contesta estas preguntas (expresa las probabilidades como fracciones simplificadas):

1. ¿Cuál es la probabilidad de que suceda M? ______ .
2. ¿Cuáles son los eventos elementales de M? ____________________________________ .
3. ¿Cuál es la probabilidad de que suceda cada evento elemental de M? ____ .
4. ¿Qué resulta al sumar las probabilidades de los eventos elementales de M? ________ .
5. ¿Cómo es la probabilidad de que suceda M comparada con la suma de las probabilidades de sus eventos elementales? _________________ .
6. ¿Cuántos eventos elementales tiene Ω? ____ .
7. ¿Cuál es la probabilidad de cada evento elemental de Ω? ____ .
8. ¿Cuánto suman las probabilidades de todos los eventos elementales de Ω? ___ .
9. ¿Cuál es la probabilidad de Ω? ___ .
10. ¿Cómo se denomina Ω (como evento) por tener una probabilidad de ocurrencia igual a 1? _________ ____________ .
11. ¿Cuál es la probabilidad de que suceda C? ____ .
12. ¿Cuál es la probabilidad de que suceda D? ____ .
13. ¿Cuál es la probabilidad de que suceda $C \cup D$? ____ .
14. ¿Cuál es la probabilidad de que el dado blanco caiga 4 o el rojo caiga 2? _____.
15. ¿Cuál es la probabilidad de que suceda T? ________ .
16. ¿Cuál es la probabilidad de que suceda $T \cap C$? ____ .
17. ¿Cuál es la probabilidad de que no resulte par? _____ .
18. ¿Cuál es la probabilidad de que suceda el evento complemento de M? _______ .
19. ¿A qué es igual P(M) + P(M’)? ____ .
20. ¿Cuál es la probabilidad de obtener par y que el dado rojo caiga 2? _____ .

FICHA DE TRABAJO XV.1

I. En un grupo de 75 personas, todos padecen hipertensión, de ellas: 4 tienen problemas cardiovasculares, 45 son diabéticos, 26 tienen sobrepeso y 8 son diabéticos y tienen sobrepeso; el resto sólo padece hipertensión.

En el cuadro de la derecha dibuja el diagrama de Venn que describa la situación.

(Expresa tus respuestas como fracciones simplificadas)

Se selecciona al azar a una persona.

1. ¿Para cada persona, cuál es la probabilidad de ser seleccionada? _____

2. Encuentra la probabilidad de que la persona seleccionada, además de hipertensa:

a. Sea diabética y tenga sobrepeso _____. **b.** Sea diabética o tenga sobrepeso _____. **c.** Sea diabética _____. **d.** Sólo sea diabética ____. **e.** Tenga sobrepeso ____. **f.** Sea hipertensa ____. **g.** Sólo tenga sobrepeso ____. **h.** Tenga problemas cardiovasculares ____. **i**. Sea diabética o tenga sobrepeso o tenga problemas cardiovasculares ____. **j.** A la vez sea diabética y tenga sobrepeso y tenga problemas cardiovasculares _____. **k.** No sea diabética _____.

3. Se escogen al azar dos personas. Calcula la probabilidad de que ambas, además de ser hipertensas:

a. Sean diabéticas y tengan sobrepeso _____. b. Sólo tengan un padecimiento _____.
c. Sean diabéticas, tengan sobrepeso y tengan problemas cardiovasculares _____. d. Sean diabéticas o tengan sobrepeso o problemas cardiovasculares _____.

II. Contesta haciendo sólo cálculos mentales.

1. Una moneda está hecha de tal manera que la probabilidad de que caiga águila es 3/7, ¿cuál es la probabilidad de que caiga sol? ____ .

2. La caja A contiene 10 bolas numeradas del 1 al 10; la caja B contiene 5 bolas numeradas del 1 al 5. Si al extraer una bola de cualquier caja resulta número impar obtengo un premio, ¿cuál es la caja de la que me conviene sacar la bola? ____ .

RESUMEN

Un fenómeno es *aleatorio* cuando se tiene *incertidumbre* al predecir sus resultados. En cambio, si sus resultados los podemos predecir con *cierto grado de certeza* lo llamamos *determinista*.

La base de la teoría de la probabilidad son sus axiomas. El primer axioma básico establece que la probabilidad de ocurrencia de un evento es mayor o igual a cero; el segundo establece que la probabilidad del espacio muestral es igual a 1; y el tercero, que se cumple sólo para eventos mutuamente excluyentes, afirma que la probabilidad de que ocurra el evento “A unión B” es igual a la probabilidad de que ocurra A más la probabilidad de que ocurra B.

Los tres axiomas constituyen los axiomas básicos de la probabilidad, pero además se reconocen otros: la probabilidad de un evento es igual a la suma de las probabilidades de sus eventos elementales; la probabilidad de la unión de dos eventos es igual a la suma de las probabilidades de cada evento menos la probabilidad de su intersección; y la probabilidad del complemento de un evento es igual a uno menos la probabilidad del evento.

16ª SESIÓN
PROBABILIDAD

COMPETENCIAS POR ADQUIRIR

- **Estimar el valor de la probabilidad de ocurrencia de un evento mediante la regularidad estadística.**
- **Ante un problema calcular la probabilidad de que ocurra determinado evento.**

PROBABILIDAD

La teoría de la probabilidad surgió en mesas de juego del s. XVII, pero al paso de los años se ha podido aplicar en la solución de problemas sociales y económicos. En la actualidad se incluye por lo general como parte de las carreras profesionales y en todos los estudios de posgrado de ciencias e ingeniería.

Mucho es el interés por la probabilidad. Las compañías de seguros requieren conocer sus riesgos de pérdida; los centros educativos contemplan en sus programas a la probabilidad como una herramienta para entender diversos fenómenos; los clubes deportivos estudian los porcentajes de desempeño de sus jugadores para determinar el valor de su contrato. . .

Pierre Fermat
1601- 1665

LOS PIONEROS. . .

En 1654, Antoine Chevalier de Méré, quien tenía un particular interés por los juegos de azar, planteó a Blaise Pascal y a Pierre de Fermat un dilema sobre un juego que consistía en lanzar un par de dados 24 veces. El problema era decidir si se debía apostar, o no, a que durante los 24 lanzamientos ocurriría al menos un par de seises.

Blaise Pascal
1623-1662

En 1657 Christian Huygens examinó la correspondencia que sostuvieron Fermat y Pascal; de ahí publicó el primer libro sobre Probabilidad titulado *De Ratonciniis in Ludo Aleae*, un tratado sobre problemas asociados con juegos de azar.

En el siglo XVII la Teoría de Probabilidad era muy popular, de manera que el tema se abordó mucho. Además de las mencionadas, destacaron las contribuciones de Jakob Bernoulli (1654-1705) y de Abraham de Moivre (1667-1705).

La única apuesta segura es la existencia de Dios. Si no existe, no se pierde nada. Pero de existir, el premio puede ser la salvación eterna. *Pascal.*

PROBABILIDAD SUBJETIVA

Cuando una persona común aventura un juicio sobre la probabilidad de que algo suceda se basa en sus creencias y experiencia; son estimaciones personales de su grado de seguridad sobre lo que puede pasar.

La **probabilidad subjetiva** se puede definir como la probabilidad asignada a un evento por un individuo, sustentada en una mezcla de sentimientos personales y evidencia disponible. La evidencia puede ser el número de veces que ocurrió el evento en el pasado o simplemente una creencia más o menos meditada.

Las asignaciones de probabilidad subjetiva suelen hacerse cuando los eventos se presentan sólo una vez o un número muy reducido de veces, o bien, cuando no se tienen registros que demuestren comportamientos cíclicos o de cierta regularidad.

Muchas decisiones sociales y administrativas se refieren a situaciones específicas y únicas, los responsables de tomar esas decisiones hacen un uso considerable de la probabilidad subjetiva. Un campesino afirma *hoy lloverá* porque observa los mismos signos que aparecieron antes de llover en días anteriores; un empleado compra dólares al finalizar un sexenio porque recuerda que subió su valor en situaciones similares del pasado; las autoridades ofrecen reubicar a familias radicadas abajo de un cerro porque creen que, debido a la inconsistencia de la tierra, la probabilidad de un desgajamiento es alta.

PROBABILIDAD EMPÍRICA

Mientras los fenómenos deterministas son predecibles con una seguridad total, los aleatorios sólo lo son con una seguridad parcial. La probabilidad de que algo suceda tiene que ver con la incertidumbre.

Es **seguro** que al combinarse dos átomos de hidrógeno con uno de oxígeno el resultado sea una molécula de agua.

En cambio, al lanzar un dado ideal de seis caras (perfectamente regular y de material totalmente uniforme) no podemos estar seguros de que caerá con la cara cinco hacia arriba, lo más que podemos afirmar es que hay cierta probabilidad de que así suceda. ¿Cuánto es el valor de esa probabilidad? ______.

Experimento. Prepara diez monedas del mismo valor. Vamos a interesarnos por las que caigan "águila". Se lanzarán 200 volados, pero no es necesario que tires una moneda 200 veces. Toma las diez a la vez en un recipiente y revuélvelas; lánzalas sobre una superficie y cuenta cuántas cayeron "águila". Al número de "águilas" lo llamamos **frecuencia del evento (F)** y lo denotamos por f_1, regístralo en la tabla 8.1

VOLADOS	F	FA	FRA
10	$f_1 =$	$f_{a1} =$	$f_{r1} =$
20	$f_2 =$	$f_{a2} =$	$f_{r2} =$
30	$f_3 =$	$f_{a3} =$	$f_{r3} =$
40	$f_4 =$	$f_{a4} =$	$f_{r4} =$
50	$f_5 =$	$f_{a5} =$	$f_{r5} =$
60	$f_6 =$	$f_{a6} =$	$f_{r6} =$
70	$f_7 =$	$f_{a7} =$	$f_{r7} =$
80	$f_8 =$	$f_{a8} =$	$f_{r8} =$
90	$f_9 =$	$f_{a9} =$	$f_{r9} =$
100	$f_{10} =$	$f_{a10} =$	$f_{r10} =$
110	$f_{11} =$	$f_{a11} =$	$f_{r11} =$
120	$f_{12} =$	$f_{a12} =$	$f_{r12} =$
130	$f_{13} =$	$f_{a13} =$	$f_{r13} =$
140	$f_{14} =$	$f_{a14} =$	$f_{r14} =$
150	$f_{15} =$	$f_{a15} =$	$f_{r15} =$
160	$f_{16} =$	$f_{a16} =$	$f_{r16} =$
170	$f_{17} =$	$f_{a17} =$	$f_{r17} =$
180	$f_{18} =$	$f_{a18} =$	$f_{r18} =$
190	$f_{19} =$	$f_{a19} =$	$f_{r19} =$
200	$f_{20} =$	$f_{a20} =$	$f_{r20} =$

Tabla XVI.1

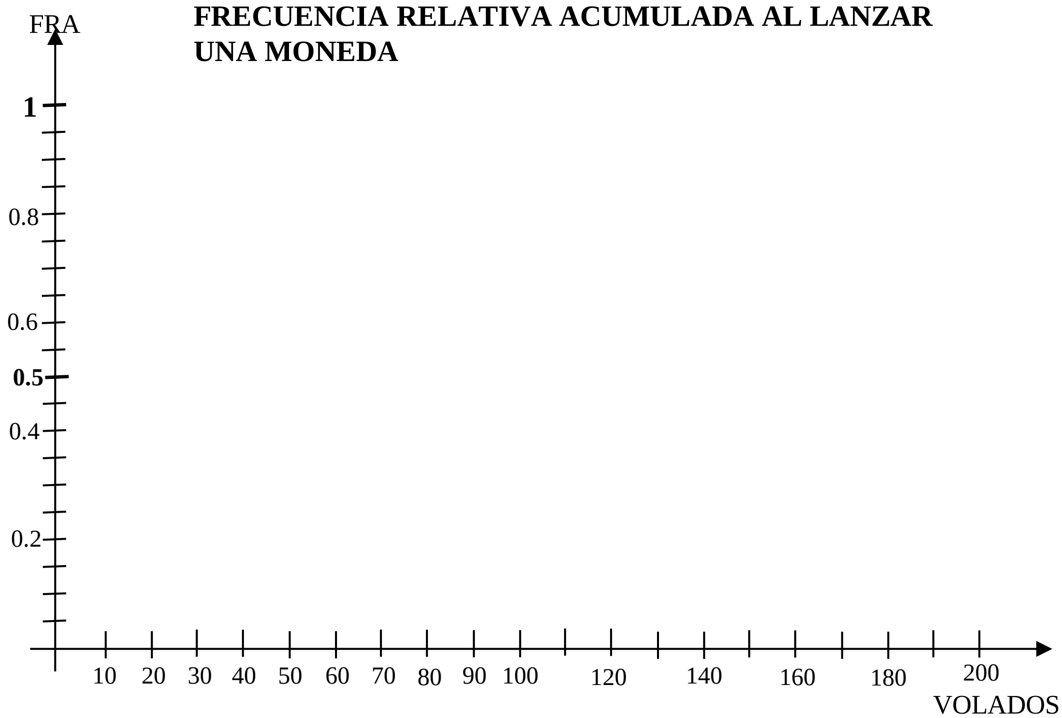
El lanzamiento de las 10 monedas lo repites diecinueve veces más para completar 200 volados, y al mismo tiempo registras los resultados de las frecuencias f_2, f_3 , f_4 . . . etc., en la tabla XVI.1

Enseguida calcula la **frecuencia acumulada** (**FA** en la 3ª columna) sumándole al valor de la frecuencia correspondiente el valor de todas las frecuencias anteriores.

La **frecuencia relativa acumulada** (**FRA** en la 4ª. columna) se obtiene al dividir la frecuencia acumulada entre el número acu-mulado de volados (1ª columna).

Construye la gráfica lineal de la FRA contra el número de lanzamientos:

FRECUENCIA RELATIVA ACUMULADA AL LANZAR UNA MONEDA



Gráfica XVI.1

En la **gráfica 16.1,** a medida que aumenta el número de lanzamientos, la **variación** de la frecuencia relativa ¿aumenta o disminuye? ____________. ¿A qué valor tiende la frecuencia relativa a medida que aumenta el número de lanzamientos? ____.

En efecto, al aumentar las repeticiones la frecuencia relativa tiende a estabilizarse alrededor de un valor fijo. Esta es una característica de los fenómenos aleatorios que se conoce como **regularidad estadística.**

El valor 0.5 es el "valor límite" o la **probabilidad empírica** de que la moneda caiga "águila" al ser lanzada. Matemáticamente esto se expresa así:

$\underset{n \to \infty}{Lím}\ FRA = 0.5$ (Se lee: el límite de la frecuencia relativa cuando el número de lanzamientos tiende a infinito es igual a 0.5).

PROBABILIDAD CLÁSICA O *A PRIORI*

Si existiera una moneda ideal, es decir, sin grosor, perfectamente regular y uniforme, los resultados al hacer un lanzamiento (al que llamaremos "experimento") sólo podrían ser "águila" (a) o "sol" (s).

Acuérdate: Un subconjunto cualquiera del espacio muestral Ω es un _________**.**

En el lanzamiento de una moneda los eventos son tres: $\{a\}$, $\{s\}$, $\{a, s\}$.

Definición: La probabilidad de que suceda un evento E se calcula así:
$P(E) = \dfrac{\qquad\qquad}{\qquad\qquad}$ **16.1**

El resultado favorable es aquél del cual nos interesa calcular su probabilidad de ocurrencia, así, la probabilidad de "águila" después de lanzar la moneda será:

$$P(\{a\}) = \frac{Número\ de\ resultados\ favorables}{Número\ de\ resultados\ posibles} = \frac{1}{2}$$

$$= 0.5$$ ¿Este valor es diferente del número al que tiende la gráfica 8.1?

Hay eventos totalmente seguros. Por ejemplo: al lanzar la moneda es seguro que el resultado sea "águila" o "sol". ¿Cuál es la probabilidad de que caiga "águila" o "sol"? P$\{a$ ó $s\}$ ___ .

Cualquier evento igual al espacio muestral es un **evento seguro** y su probabilidad de ocurrencia es igual a 1. Por el contrario, cualquier evento igual al conjunto vacío es un **evento imposible** y su probabilidad es igual a_____*: no puede suceder algo si es nada*.

La probabilidad clásica o a priori (16.1) de que suceda un evento; también se conoce como *Regla de Laplace* y para que sea válida existe una condición: todos los resultados posibles deben ser **equiprobables**, es decir, deben tener las mismas posibilidades de ocurrencia.

Un dado aparentemente homogéneo está hecho la mitad de plomo y la otra mitad de madera, es decir, está "cargado". Por lo tanto, sus números no tienen las mismas posibilidades de salir, los resultados posibles no son _________________.

Problema: Para supervisar la calidad de un lote de 20 radios "*Walkie Talkie*", Peter Stoll toma una muestra de 4 al azar. Si el lote contiene 3 defectuosos, ¿con qué probabilidad encontrará 2 defectuosos?
Para calcular la probabilidad es indispensable conocer tanto el numerador como el denominador de la definición 16.1, es decir: número de **resultados favorables** y de **resultados posibles.**

Conviene primero saber cuántos resultados son posibles. Las muestras que Peter puede tomar son combinaciones de 4 en 4 elementos, tomados de un total de 20:

$$C_4^{20} = \frac{20!}{(20-4)!\,4!} = 4\,845$$ (son los **resultados posibles).**

Un resultado es favorable si la muestra de 4 radios contiene 2 defectuosos. De las 4 845 muestras, veamos cuántas contienen 2 radios defectuosos.
Como la muestra es de 4 radios, supongamos que el primero es defectuoso (lo cual tiene 3 posibilidades porque hay 3 defectuosos), que el segundo también es defectuoso (lo cual tiene 2 posibilidades si ya el primero fue defectuoso); el tercero sería un radio no defectuoso (lo cual tiene 17 posibilidades) y el cuarto también sería no defectuoso. Completa:

1er. radio	2o. radio	3er. radio	4º. radio
3	2	17	_____

Por lo tanto, el número de muestras en las que hay 2 radios defectuosos es:

3 × 2 × 17 × ___ = 1 632 (estos son los **resultados favorables**)

Al aplicar la definición 8.1: $P(D) = \frac{1\,632}{4\,845} = 0.3368$ (D es el evento: *muestra con 2 radios defectuosos*)

Este resultado se puede expresar así: *La probabilidad de que al tomar una muestra aleatoria de 4 radios 2 sean defectuosos es 0.3368;* o bien: *Entre más muestras aleatorias se tomen (reemplazando los radios), el porcentaje que tenga dos radios defectuosos se acercará más a 33.68.*

¿Cuál es la probabilidad de que los 4 radios salgan defectuosos? ____. ¿Por qué? (Sugerencia: analiza lo que pasa en la definición 8.1).____.

RESUMEN

La probabilidad puede ser de tres tipos: subjetiva, empírica o axiomática. A esta última se le conoce también como probabilidad clásica o teórica.
La probabilidad subjetiva la emiten las personas basándose en su experiencia; la empírica se calcula realizando muchas veces un experimento hasta encontrar el valor al que tiende la frecuencia relativa; y la teórica se calcula mediante la "Regla de Laplace", que consiste en dividir el número de resultados favorables de un experimento entre el número de resultados posibles (el número de elementos que contiene el espacio muestral).
Un evento es un subconjunto del espacio muestral.
Para que la regla de Laplace se pueda aplicar es necesario que los resultados posibles sean equiprobables.

1

FICHA DE TRABAJO XVI.1

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

1. Contesta interpretando el diagrama de la izquierda. ¿Cuál es la probabilidad de que al lanzar dos dados de seis caras la suma de los puntos sea:

a. 5
b. 7
c. 9
d. número impar?

2. Tenemos en una urna 10 bolas del mismo tamaño pero de distintos colores: 1 azul, 3 rojas, 2 verdes, 2 amarillas y 2 blancas. Se realiza el experimento de sacar una bola al azar (sin mirar).

Dibuja aquí la urna con las bolas:	Calcula la probabilidad de que ocurran los siguientes eventos: a. Sale bola roja b. Sale bola verde c. Sale bola roja, amarilla o blanca d. Sale bola azul o verde e. Sale bola amarilla ¿Cuáles de los eventos anteriores son equiprobables?

3. En el problema anterior, supón que realizas la experiencia de sacar una bola al azar un millón de veces reponiéndola en cada ocasión.

a. ¿Cuántas veces crees que saldrá, aproximadamente, cada color de bola?
b. ¿Qué porcentaje del total representa cada número?

4. Marca con V (verdadero) o F (falso):

a. El valor de una probabilidad puede ser negativo. V ( ) F ( )
b. El valor máximo de una probabilidad es 100. V ( ) F ( )
c. El valor mínimo de una probabilidad es 0. V ( ) F ( )
d. Un evento siempre debe tener elementos. V ( ) F ( )
e. No hay eventos seguros. V ( ) F ( )
f. Hay eventos imposibles. V ( ) F ( )

5. Un lote contiene 20 estetoscopios, de los cuales 3 están defectuosos. ¿Cuál es la probabilidad de que al muestrear cuatro de ellos, dos estén defectuosos?

17ª SESIÓN
PROBABILIDAD CONDICIONAL

COMPETENCIAS POR ADQUIRIR

- Expresar el significado de probabilidad condicional.
- Expresar el significado de eventos independientes.
- Resolver problemas empleando los conceptos de probabilidad condicional y eventos independientes.

MATERIAL NECESARIO: CALCULADORA, LÁPIZ, GOMA Y SACAPUNTAS.

PROBABILIDAD CONDICIONAL

Calcular la probabilidad condicional de un evento implica reducir el espacio muestral y aplicar la definición clásica. A partir de problemas dados, aquí tendrás la oportunidad de comprenderlo mejor y por lo tanto de mejorar tus habilidades para resolver problemas de mayor complejidad.

SI Y SÓLO SI . . .

Una proposición condicional tiene la forma de este ejemplo:

Si el rosal crece, entonces florecerá.

El conectivo "si y sólo si" se utiliza para expresar la conjunción de dos condicionales recíprocas.

Este es un ejemplo de 2 condicionales recíprocas:

"Si te quiero, entonces te extraño"
"Si te extraño, entonces te quiero"

Al unirlas mediante el conectivo "y" se forma su conjunción o **bi**condicional:

"*Si te quiero, te extraño* ***y*** *si te extraño, te quiero****"*** **(se obvia el entonces).**

Afirmación que se puede resumir así:

"Te quiero ***si y sólo si*** *te extraño"*

¿Puedes expresar "Pienso si y sólo si existo" como una conjunción de dos condicionales? Escríbela enseguida.

Si _________, _________ y si _________, _________ .

DADO QUE . . .

En el lenguaje cotidiano solemos utilizar condiciones. Decimos, por ejemplo, “dado que tengo dinero, compro un reloj” que es equivalente a decir “compro un reloj **dado que** cumplo con la condición de tener dinero”.

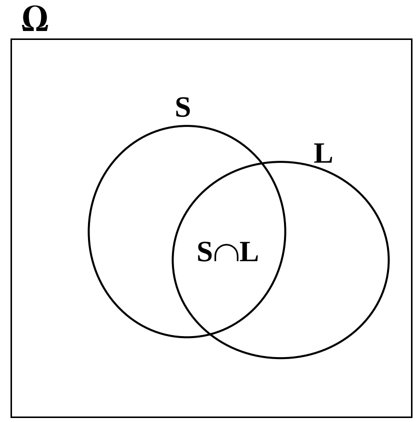


Diagrama XVII.1

Al afirmar “sabe mucho **dado que** lee buenos libros”, la condición que cumple la persona para saber mucho es leer buenos libros. Con la notación de conjuntos la situación puede representarse así (diagrama XVII.1):

S: sabe mucho.

L: lee buenos libros.

S/L: sabe mucho **dado que** lee buenos libros.

De esta manera **S/L** = **S∩L**

Problema. Al hacer un censo en la colonia Educación del D.F., ingeniero Daniel Cantó encontró que hay 608 casas habitación; en 2 éstas tienen televisor, en 327 se cuenta con computadora y en 1 ambas cosas.

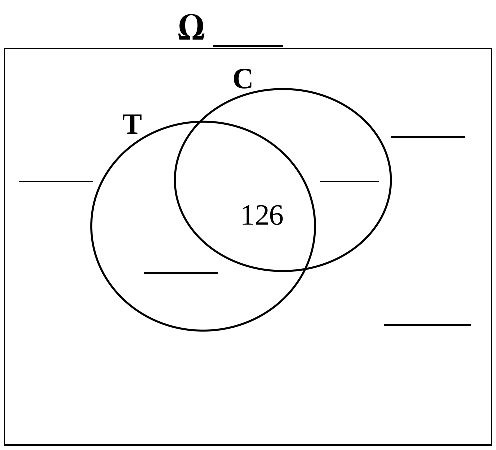


Diagrama XVII.2

A la izquierda se ilustra la situación mediante un diagrama de Venn (escribe los números que faltan).

Daniel quiere saber cuál es la probabilidad de que al seleccionar una casa al azar cuente con televisor, siempre y cuando tenga computadora.

El espacio muestral es Ω y como el evento **C** condiciona al evento **T,** a **C** se le denomina **espacio muestral reducido.** Los **resultados posibles** de T/C son únicamente los que se encuentran en T∩C.

Por lo tanto, la probabilidad de que ocurra T **dado que** ocurre C es:

$$P(T/C) = \frac{\text{Número de resultados posibles de } T \cap C}{\text{Número de resultados posibles de } C}$$

P(T/C) se lee: probabilidad de T dado que sucede C.

Es decir: $$P(T/C) = \frac{126}{327}$$

$$= 0.3853$$

Si un millón de veces se seleccionara al azar, con reemplazo, una casa, ¿cuál sería el porcentaje de casas que resultarían con televisor dado que tienen computadora? _______ .

Calcula: $P(C/T) = \frac{}{}$ (aquí T es el **espacio muestral reducido**)

$= \_\_\_\_\_$

¿Es P(T/C) = P(C/T)? ____.

Definición: $P(X/Y) = \dfrac{Número\ de\ resultados\ posibles\ de\ X \cap Y}{Número\ de\ resultados\ posibles\ de\ Y}$ **17.1**

Cuando la ocurrencia de un evento (X) no implica la ocurrencia de otro (Y), se dice que los eventos son **independientes**.

Definición: dos eventos , X y Y, son independientes si y sólo si:

P(X∩Y) = P(X)·P(Y) 17.2

Del censo del ingeniero Cantó resulta:

$P(C \cap T) = \frac{126}{608} = 0.2072$ $\quad P(C) = \frac{327}{608} = 0.5378$ $\quad P(T) = \frac{244}{608} = 0.4013$

$$P(C)P(T) = \frac{327}{608} \cdot \frac{244}{608}$$
$$= 0.2158$$

¿Son independientes C y T en el censo de Daniel? Si lo fueran se debería cumplir que P(C∩T) = P(C)·P(T), ¿se cumple? _____.

En las siguientes igualdades incompletas, T’ y C’ son los eventos complementarios de T y C respectivamente. Completa sin hacer cálculos por escrito ni emplear calculadora (dibuja un diagrama que ilustre las operaciones entre conjuntos, a partir del inciso **b.**:

a. P(Ω) = ___ .

b. P(C∩C’) = ___ .

c. P(T∩T’) = ___ .

d. P(T’∩C’) = ___ .

Problema. El biólogo Federico Caballero estudia las características de los salmones. En una laguna mantiene un banco con cuatro tipos de salmones, clasificados en tres tamaños, de esta manera:

TIPO / TAMAÑO	REY **(Y)**	ROSA **(R)**	PLATEADO **(P)**	CARNADA **(C)**	TOTAL
CHICO **(CH)**	48	12	27	9	96
MEDIANO **(M)**	34	13	16	7	70
GRANDE **(G)**	9	6	5	12	32
TOTAL	91	31	48	28	198

Tabla XVII.1

Al seleccionar un salmón al azar, la probabilidad de que sea rosa y mediano es:

$$P(R \cap M) = \frac{13}{198}$$

La probabilidad de que sea mediano: $P(M) = \frac{70}{198}$

La probabilidad de que sea de tipo rosa dado que es mediano:

$$P(R/M) = \frac{13}{70}$$ **(M se toma como el espacio muestral)**

Este último resultado es el cociente de los dos anteriores:

$$P(R/M) = \frac{13}{70}$$

$$= \frac{13/198}{70/198}$$

$$= \frac{P(R \cap M)}{P(M)}$$

Para dos eventos cualesquiera X y Y, la probabilidad de que suceda X dado que sucede Y se calcula así: $P(X/Y) = \frac{P(X \cap Y)}{P(Y)}$ **9.3**

A partir de la tabla XVII.1, indica la probabilidad de que un salmón escogido al azar sea:

a. Grande, dado que es plateado: P(G/P) = ___ = ___ = ___	**b.** Rey, dado que es rey: P(Y/Y) = ___ = ___

Calcula lo que falta (mantén el denominador 198 en los resultados):

$$P(M) = \frac{70}{198} \qquad P(R) = \frac{\quad}{\quad}$$

$$P(M \cap R) = \frac{13}{198}$$

Por lo tanto: $P(M) + P(R) - P(M \cap R) = \frac{70}{198} + \_\_\_\_ - \_\_\_\_$

$$= \_\_\_\_\_$$

Por otro lado: $P(M \cup R) = \frac{88}{198}$

Entonces: $P(M \cup R) = P(M) + \_\_\_\_\_ - \_\_\_\_\_\_\_$

Esta relación es válida para cualquier par de conjuntos X y Y:

$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$. 9.4

Calcula $P(G \cup C)$ aplicando la relación 9.4

____ .

____

Calcula la probabilidad de que al seleccionar un atún al azar sea:

- grande o mediano ____ .
- grande y mediano ____ .
- grande o mediano o chico __ .

FICHA DE TRABAJO XVII.1

1. El equipo Monarcas de futbol lleva 79 partidos ganados de 316 que ha jugado en su historia. En el actual torneo debe jugar 2 partidos más y un resultado no afecta al siguiente. Supón que se trata de un fenómeno aleatorio y calcula la probabilidad de que:

a. Gane los dos.

b. Pierda los dos.

c. Gane uno y pierda el otro.

2. El licenciado José Peralta les dice a sus clientes que gana 85 de cada 100 casos que defiende, mientras que la compañía de aviación "Azteca" afirma que el 96% de los pasajeros quedan satisfechos con el servicio. Armando Cortines le pide al Lic. Peralta que defienda su caso, mientras tanto hace un viaje por "Azteca". ¿Cuál es la probabilidad de que el Lic. Peralta gane su caso y quede satisfecho con el servicio de "Azteca"?

3. En Pochutla, Oax., 1 120 mujeres mayores de 21 años contestaron una encuesta acerca de su estado civil y escolaridad. Las frecuencias obtenidas se organizaron en la tabla de abajo. Escribe los totales.

ESTADO CIVIL / ESCOLARIDAD	CASADA (C)	SOLTERA (S)	VIUDA (V)	DIVORCIADA (D)	TOTAL
EDUC. BÁSICA (E)	213	28	16	5	
BACHILLERATO (B)	96	87	18	29	
LICENCIATURA (L)	64	103	2	81	
TOTAL					

Calcula la probabilidad de que una mujer seleccionada al azar:

a. Esté soltera, dado que tiene licenciatura.

b. Haya estudiado bachillerato dado que está divorciada.

c. Esté divorciada dado que estudió bachillerato.

d. Esté casada o haya tenido educación básica.

e. Esté viuda o casada.

f. Esté casada y divorciada. g. Haya estudiado bachillerato y esté viuda.

4. Al hacer un análisis de los comerciales transmitidos de 7 a 9 de la noche, en una semana, por cierto canal televisivo, se encontró que el 23% tenían un contenido claramente sexual. Por otra parte, se sabe que 3 de cada 10 mexicanos en edad escolar están inscritos en una escuela. Al seleccionar al azar a una persona que resida en la República Mexicana, ¿cuál es la probabilidad de que esté viendo un comercial de la TV con contenido sexual y esté inscrito en una escuela?

RESUMEN

La notación P(A/B) indica la probabilidad de que suceda A “dado que” (o “con la condición de que”) suceda B.

Al calcular la probabilidad condicional, el espacio muestral se reduce de tal manera que el valor de P(A/B) se encuentra dividiendo el número de resultados posibles de A (que son los contenidos en $A \cap B$) entre el número de resultados posibles de B (que se constituye en el espacio muestral reducido).

Dos eventos son independientes cuando la probabilidad de su intersección es igual al producto de las probabilidades de cada uno, y viceversa, si la probabilidad de su intersección es igual al producto de sus probabilidades, entonces son eventos independientes.

La probabilidad condicional P(A/B) se calcula mediante la igualdad

$$P(X/Y) = \frac{P(X \cap Y)}{P(Y)}$$

La probabilidad de que suceda A o B se calcula con esta fórmula: $P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$

BIBLIOGRAFÍA

- Asencio Ma. J., Romero J. A. y De Vicente, E. Estadística. Ed. Mc Graw Hill. España, 1999.
- Johnson, R. Estadística Elemental: Lo Esencial. Ed. ITP. México, 1999.
- Lopes P. A. Probabilidad y Estadística, conceptos, modelos, aplicaciones en Excel. Ed. Prentice Hall. Colombia, 2000.
- Mendenhall, W., Beaver R. J. y Beaver B. M. Introducción a la Probabilidad y Estadística para Administración y Economía. Ed. Thomson. México, 2002.
- Pagano R. R. Estadística para las ciencias del comportamiento. Ed. Thomson. México, 2006.
- Willoughby, S. Probabilidad y Estadística. Ed. PCSA. México, 1993.
- Weimer R. C. Estadística. Ed. CECSA. 2ª reimpresión. México, 1999.

FIN