Colegio de Ciencias y Humanidades
Plantel Naucalpan

PROGRAMA DE APOYO AL EGRESO

ESTADÍSTICA Y PROBABILIDAD I

Coordinador:
Héctor Gabriel Rivera Vargas

Autores:
Aquilino Zecua Fernández
Beatriz Arellano Sánchez
Carlos Alberto Álvarez García
Héctor Gabriel Rivera Vargas
Mauricio Albuerne Sánchez

ENERO 2019

Presentación

El presente material está diseñado para el Programa de Apoyo y Egreso (PAE), tratando de apoyar al profesor y a los estudiantes, por medio de ejemplos y ejercicios, sin pretender sustituir al profesor ni a la vasta gama de materiales bibliográficos que si profundizan en los temas.

El material está basado en el Programa de Estudio de Estadística y Probabilidad I y II, por lo tanto, se organiza en 3 unidades: Unidad 1 Obtención, descripción e interpretación de información estadística, Unidad 2 Obtención e interpretación de información estadística con datos bivariados y Unidad 3 Azar, modelación y toma de decisiones.

Los comentarios, criticas o sugerencias ayudarán en gran medida a mejorar.

Esperamos que este sencillo, pero esmerado trabajo, sea un producto que sirva tanto a los compañeros como a los estudiantes.

Los autores

Enero 2019

Contenido Temático por unidades

UNIDAD	NOMBRE
1	Obtención, descripción e interpretación de información estadística
2	Obtención e interpretación de información estadística con datos bivariados
3	Azar, modelación y toma de decisiones

En los Programas de Estudio de Estadística y Probabilidad I y II, se proponen los siguientes materiales bibliográficos y en red:

Para el profesor

1. Bonet, J. (2003). *Lecciones de estadística. Estadística descriptiva y probabilidad.* España: Editorial Club Universitario.
2. Castillo, J., Gómez, J. (1998). *Estadística inferencial básica*. México: Grupo Editorial Iberoamérica.
3. De Canales, F. H., E. L. de Alvarado y E. B. Pineda. (1986). *Metodología de la investigación. Manual para el desarrollo de personal de salud.* Primera edición. Limusa.
4. Kahneman, D., A., Tversky y P. Slovic. (1982). *Judgement under uncertainty: heuristics and biases*. Cambridge.
5. Kerlinger, Fred N. y Howard B. Lee. (2000). *Investigación del comportamiento. Métodos de investigación en ciencias sociales,* 4a edición. México: McGraw Hill
6. Lander, J. P. (2014). *R for everyone*. Addison Wesley Data & Analytic Series.
7. Lohr, S. (1999). *Teoría del muestreo*. Thomson.
8. Méndez, I. (2000). *El protocolo de investigación*. Trillas.
9. Peat, J., y B. Barton. (2005). *Medical statistics*. Blackwell Publishing.
10. Prieto, L., I. Herranz. (2005) ¿Qué significa "*Estadísticamente significativo"?.* Díaz de Santos.
11. R. Isaac. (1995). *The pleasures of probability*. Springer.
12. Sutherland, D. (1992). *Irracionalidad. El enemigo interior*. Alianza.
13. Scheaffer, R., Gnanadesikan, M., Watkins, A. y Witmer, J. (1996). *Activity based statistics.* Instructor Resources. Springer.
14. Triola, M. (2009). *Estadística*, décima edición. México: Pearson Addison Wesley.

15. Willoughby, S. (2000). *Probabilidad y estadística*. México: Publicaciones Cultural.

Para el estudiante

1. Ávila, R. *et al*. (2005) *Paquete didáctico para estadística y probabilidad*. México: cch–unam.
2. Ávila, R. *et al*. (2005). *Paquete didáctico para estadística y probabilidad II.* México: cch–unam.
3. Bonet, J. (2003). *Lecciones de estadística. Estadística descriptiva y probabilidad*. España: Editorial Club Universitario.
4. Carrascal, U. (2009). *Estadística descriptiva con Microsoft Excel 2007*. Alfaomega.
5. Castillo, J., Gómez, J. (2000). *Estadística inferencial básica*. México: Grupo Editorial Iberoamérica.
6. Chao, L. (1989). *Introducción a la estadística*. cecsa.
7. Cintas, P., Almagro, L y Llabrés, J. *Estadística práctica con minitab*. Pearson–Prentice Hall
8. Haigh, J. (2003). *Matemáticas y juegos de azar*. Metatemas.
9. Mochón, S. (2004). *Desarrollando conceptos de probabilidad y estadística*. McGraw–Hill.
10. Paulos, J. A. (1998). *Un matemático lee el periódico*. Metatemas.
11. Paulos, J. A. (1988). *El hombre anumérico*. Metatemas.
12. Pérez, A.; Rodríguez, D., *et al*. (2005). *Paquete de evaluación de Estadística y Probabilidad I*. México: cch–unam.
13. Rossman. (2006). *Workshop statistics*. Key Curriculum Press.
14. Scheaffer, R., Ggnanadesikan, M., Watkins, A. y Witmer, J. (1996). *Activity based statistics.* Student Guide. Springer.
15. Triola, M. (2009). *Estadística*. Décima edición. México: Pearson Addison Wesley.
16. Willoughby, S. (2000). *Probabilidad y estadística*. México: Publicaciones Cultural.

Mesografía

<http://www.inegi.org.mx> al 10 de diciembre de 2015.
<http://consulta.mx> al 10 de diciembre de 2015.
<http://estepais.com> al 10 de diciembre de 2015.
<http://www.censo2010.org.mx> al 10 de diciembre de 2015.
<http://math.exeter.edu/rparris/> al 10 de diciembre de 2015.
<http://portalacademico.cch.unam.mx/materiales/prof/matdidac/sitpro/mate/estad/estad1/Estadistica/index.html> al 10 de diciembre de 2015.

Propuesta de videos

Para Estadística Descriptiva, historia y fundamentos

<https://www.youtube.com/watch?v=gs2CSCr4hGQ> al 10 de diciembre de 2015.
<https://www.youtube.com/watch?v=po37moq9eYg&list=PL2fMb9MMS3eaOrJQYYQiUOPVkOMBmVUsa> al 10 de diciembre de 2015.

Representación gráfica

<https://www.youtube.com/watch?v=pT3OfSsdXC8> al 10 de diciembre de 2015.

Gráficos en Excel (datos agrupados y datos no agrupados)

<https://www.youtube.com/watch?v=04pGYGNxRZY> al 10 de diciembre de 2015.

<https://www.youtube.com/watch?v=s0PP3v_YcCU> al 10 de diciembre de 2015.

<https://www.youtube.com/watch?v=2pifUf0Y4ts> al 10 de diciembre de 2015.

Medidas de tendencia central y de dispersión. Deciles, percentiles.

<https://www.youtube.com/watch?v=XDUndiON7fk> al 10 de diciembre de 2015.

<https://www.youtube.com/watch?v=GgXi00GbCnQ> al 10 de diciembre de 2015.

Hombre vitruviano y proporción áurea (útil en el tema de correlación)

<https://www.youtube.com/watch?v=xQIfNlI4U0M> al 10 de diciembre de 2015.

Regresión lineal con calculadora

<https://www.youtube.com/watch?v=IXho6TxHj88> al 10 de diciembre de 2015.
<https://www.youtube.com/watch?v=2Kb9ppoej-4> al 10 de diciembre de 2015.

MTC y gráficas con Minitab

<https://www.youtube.com/watch?v=bvwNNDxJEmc> al 10 de diciembre de 2015.

<https://www.youtube.com/watch?v=DoftNzC6wQI> al 10 de diciembre de 2015.

Probabilidad. Regla de multiplicación

<https://www.youtube.com/watch?v=ifTWwKH8AT0> al 10 de diciembre de 2015.

Independencia

<https://www.youtube.com/watch?v=d4yIg-nEk-M> al 10 de diciembre de 2015.

Probabilidad condicional

<https://www.youtube.com/watch?v=lk9NPdNgXhQ> al 10 de diciembre de 2015.

Probabilidad total y teorema de Bayes

<https://www.youtube.com/watch?v=3h29_gTdZGQ> al 10 de diciembre de 2015.

<https://www.youtube.com/watch?v=kd_F14sNZPY> al 10 de diciembre de 2015.

Otras sugerencias

https://www.youtube.com/watch?v=YV1Yp5tlLvo&feature=youtu.be

https://www.youtube.com/watch?v=2V6Y_4Qymtc

https://www.youtube.com/watch?v=IXho6TxHj88

https://es.khanacademy.org/math/statistics-probability/probability-library?fbclid=IwAR0cmdMGtgML8vJXtuu4_tB78NmPeC3Swx-jGYqwgDNWHp0jBTGpvvWFKw4

Unidad 1. Obtención, descripción e interpretación de información estadística

Propósito. Al finalizar la unidad el alumno realizará inferencias informales acerca del comportamiento de una característica de interés en una población definida dentro de su entorno, a partir del análisis de su tendencia, variabilidad y distribución, en una muestra obtenida de dicha población, para contribuir a la formación de su pensamiento estadístico.

Aprendizajes. El alumno:

- Discute que la estadística estudia la variabilidad de una característica de la población, considerando la homogeneidad o heterogeneidad en los valores observados.
- Explica las nociones de variable, población y muestra estadísticas.
- Aprecia la importancia del muestreo.
- Reconoce que los datos estadísticos se obtienen por levantamiento o por experimentación.
- Valora la importancia de la recopilación y representación de datos en la investigación estadística.
- Concluye que el azar es causa de la variabilidad en los datos estadísticos.
- 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.
- Construye tablas de distribución de frecuencias, incorporando también el uso de la computadora, para describir el comportamiento de una variable.
- Construye gráficas, incorporando también el uso de la computadora para describir el comportamiento de una variable.
- Calcula medidas de tendencia central, de dispersión y de posición, incorporando también el uso de la computadora o la calculadora para describir el comportamiento de una variable.
- Concluye que el comportamiento de una colección de datos se manifiesta a partir de su tendencia, dispersión y distribución, dentro de algún contexto.
- Escoge la medida de tendencia más adecuada para describir el comportamiento de una colección de datos.
- Infiere el comportamiento de la variable, a partir de la descripción del comportamiento de los datos.
- Argumenta la validez de las inferencias informales que realice, a partir del comportamiento de una colección de datos.
- Compara la variabilidad entre dos muestras de dos distintas poblaciones por medio de sus coeficientes de variación.
- Discute la Regla Empírica y sus limitaciones, en términos también de relación entre tendencia, dispersión y distribución.

Examen diagnóstico

Nombre: ____

Grupo: ___________

Contesta las siguientes preguntas:

1. ¿Qué es la estadística? ____

____

2. ¿Qué es el azar? ____

____

3. ¿Qué es la probabilidad? ____

____

4. ¿Qué hace que un evento sea más probable que otro? ______________________

____

____

5. ¿Qué es una proporción? ____

____

____

6. ¿Qué es una población? ____

____

____

7. ¿Qué una muestra? ____

____

____

8. ¿Para qué se realizan muestreos? ____

____

____

Cuestionario

Consideraciones: En las tres primeras afirmaciones del cuestionario, se habla de lanzamientos de una moneda, considérala común y corriente. La letra "a" representa obtener águila en un lanzamiento determinado y la letra "s" representa obtener sol. Observa, la secuencia se conforma por el número de veces que se lanza la moneda, por ejemplo, la secuencia "a s a", representa una secuencia de resultados al lanzarse 3 veces una moneda y haber obtenido águila en el primero y tercer lanzamientos, y sol en el segundo lanzamiento.

Instrucciones. Lee cada una de las siguientes afirmaciones, toma tu tiempo para pensar y tacha la V si consideras la afirmación como verdadera o tacha la F si la consideras falsa:

a) Se lanza 6 veces una moneda al aire y los resultados obtenidos fueron "a s a s a s".

   Afirmación: La secuencia anterior es aleatoria V F

   Justifica tu respuesta:

   ____

b) Se lanza 4 veces una moneda y los resultados han sido "a a a a".

   Afirmación: Dada la secuencia de resultados anterior, sabemos que si lanzamos de nuevo la moneda, el resultado será "a" V F

   Justifica tu respuesta:

   ____

c) Se lanza 6 veces una moneda.

   Afirmación: La secuencia "a s a s s s" es más probable que haya sucedido que la secuencia "s s s s s s" V F

   Justifica tu respuesta:

   ____

Por último, considera el juego del "melate", donde en cada sorteo hay una combinación ganadora de 6 números:

a) Afirmación: Es más probable que la combinación ganadora sea 4, 5, 18, 12, 15, 33 que la combinación 1, 2, 3, 4, 5, 6 V F

   Justifica tu respuesta:

   ____

b) La combinación 12, 17, 24, 29, 32, 36 nunca ha salido como ganadora y todas las demás combinaciones ya lo han sido.

   Afirmación: Dicho lo anterior, la combinación tiene mayor probabilidad de ser la ganadora que todas las demás V F

   Justifica tu respuesta:

   ____

Actividad

Identifica entre variables cualitativas y variables cuantitativas:

1. Nombre: ________________________________ Variable ________________
2. No. de cuenta: _____________________ Variable ____________________
3. Edad: ________________Variable: ________________________
4. Peso: _______________ Variable ______________________________
5. Estatura: ______________ Variable __________________
6. Color de ojos: ______________ Variable __________________
7. No. de materias reprobadas: ______________ Variable _________________
8. Gasto por transporte y alimentación por semana: ____________
   Variable _______________
9. No de hermanos: ______________ Variable __________________
10. Carrera que piensas estudiar: ________________ Variable _______________
11. Promedio: ________________ Variable __________________
12. Color preferido: ____________ Variable __________________
13. Asignatura preferida: ______________ Variable __________________

Menciona 5 variables de los cuadernos, menciona entre variables cualitativas y cuantitativas:

1. ____
2. ____
3. ____
4. ____
5. ____

Menciona 5 variables de las computadoras, menciona entre variables cualitativas y cuantitativas:

1. ____
2. ____
3. ____
4. ____
5. ____

Actividad

Identifica una variable que podría ser de interés de cada una de las siguientes poblaciones y proporciona 2 de sus posibles valores de la variable.

a) bolsas de papas Sabritas chicas
   Ejemplo de variable de interés: ____________________
   Posibles valores de la variable de interés: _______________, _____________
b) estudiantes del CCH Naucalpan
   Ejemplo de variable de interés: ____________________
   Posibles valores de la variable de interés: _______________, ____________
c) teléfonos celulares
   Ejemplo de variable de interés: ____________________
   Posibles valores de la variable de interés: _______________, ____________
d) focos de un mismo tipo de una producción realizada en un día específico en una fábrica
   Ejemplo de variable de interés: ____________________
   Posibles valores de la variable de interés: _______________, ____________
e) vacas de un rancho
   Ejemplo de variable de interés: ____________________
   Posibles valores de la variable de interés: _______________, ____________
f) microbuseros del Edo. de México
   Ejemplo de variable de interés: ____________________
   Posibles valores de la variable de interés: _______________, ____________
g) comerciantes ambulantes
   Ejemplo de variable de interés: ____________________
   Posibles valores de la variable de interés: _______________, ____________
h) políticos
   Ejemplo de variable de interés: ____________________
   Posibles valores de la variable de interés: _______________, ____________
i) bebés recién nacidos
   Ejemplo de variable de interés: ____________________
   Posibles valores de la variable de interés: _______________, ____________
j) monto en las recargas de tiempo aire de Telcel
   Ejemplo de variable de interés: ____________________
   Posibles valores de la variable de interés: _______________, ____________
k) desempleados en el área metropolitana
   Ejemplo de variable de interés: ____________________
   Posibles valores de la variable de interés: _______________, ____________

Población. Cualquier colección de unidades que puedan interesar en un estudio. Esta colección debe estar bien definida, de tal forma que se pueden distinguir entre sus miembros aquellos que lo son y los que no lo son.

Muestra. Cualquier subconjunto de la población que estudiamos.

Unidad de observación. Un solo miembro de la población que estudiamos.

Medición. Esta puede ser cualitativa y cuantitativa. Una medida es un número o denominación que podemos asignar a la unidad de observación. Si este número expresa dimensiones o capacidades, etc., se le llama medición cuantitativa. Si la denominación registra características, atributos o actitudes, la llamaremos medición cualitativa.

Inferencia estadística. Una inferencia estadística es una conclusión obtenida acerca de una población completa, desde la información tomada de una muestra.

Ejercicios.

1. Varias veces durante el día un ingeniero de control de calidad, en una fábrica textil, selecciona diferentes muestras de metros cuadrados de telas, las examina y registra el número de imperfecciones que encuentra por cada metro cuadrado.

La población es ____
____

La muestra es ____

La unidad de observación es ____
____

La medición es ____
____

La medición es cualitativa o cuantitativa __________________

Si encuentra demasiadas imperfecciones en una muestra, ¿Qué decisión podría tomar? ____
____

En este caso, ¿Cuál es la importancia de utilizar una muestra y no utilizar a toda la población (Si es que aún existe o ya existe)? ______________________________
____
____

2. Un investigador médico examina los efectos de un agente cancerígeno en las ratas. Tres semanas después de inyectado el agente en 7 ratas de una población, el investigador realiza una intervención quirúrgica para extraer y pesar los tumores de cada una de ellas.
La población es ____

La muestra es ____

La unidad de observación es _____________________________________

La medición es ____

La medición es cualitativa o cuantitativa _______________

Si encuentra un promedio muy alto del peso de los tumores, ¿Qué podría pensar el investigador? ____

En este caso, ¿Cuál es la importancia de utilizar una muestra y no utilizar a toda la población (Si es que aún existe o ya existe)? _____________________________
____
____

3. Se consideró una escuela de niños que reciben educación especial. Algunos niños fueron escogidos para recibir un nuevo programa de instrucción. Se aplicó un examen antes y después de recibir la instrucción para las diferencias.
La población es ____
____

La muestra es ____
____

La unidad de observación es _____________________________________
____

La medición es ____

____
La medición es cualitativa o cuantitativa _________

Si obtenemos un promedio de las diferencias relativamente alto, ¿Qué podríamos decidir? ____
____

En este caso, ¿Cuál es la importancia de utilizar una muestra y no utilizar a toda la población (Si es que aún existe o ya existe)? _____________________________

____
____

4. El encargado de control de calidad de una empresa, hará una inspección de 300 artículos de todos los producidos en una semana, en cada uno de ellos su inspección será para determinar si el producto es defectuoso o no.
La población es ____
____

La muestra es ____
____

La unidad de observación es ____
____

La medición es ____
____

La medición es cualitativa o cuantitativa __________________

Si este encargado encuentra un porcentaje de producto defectuoso relativamente alto, ¿Qué tendrá que decidir? ____
____

En este caso, ¿Cuál es la importancia de utilizar una muestra y no utilizar a toda la población (Si es que aún existe o ya existe)? _______________________________
____
____
____

5. Un gerente desea conocer si aquellos empleados que reciben 25 días de vacaciones son más productivos durante el año, que los que reciben sólo 15 días. El gerente selecciona una muestra de 40 trabajadores, y registra su rendimiento.
Aquí podríamos considerar dos poblaciones, ¿Cuáles serían? _________________
____
____
____

Y comparar promedios de rendimiento para tomar una decisión.
En este caso, ¿Cuál es la importancia de utilizar una muestra y no utilizar a toda la población (Si es que aún existe o ya existe)? _______________________________
____
____
____

6. Profeco realizó un muestreo de 200 pilas Duracell de una producción de 10000, lo anterior con la finalidad de conocer si cumple con lo estipulado en el producto, de que "por lo menos cada pila tiene una vida útil de 20 horas"

La población es ____
____

La muestra es ____
____

La unidad de observación es ____
____

La medición es ____
____

La medición es cualitativa o cuantitativa ________________

Si el promedio de vida útil es significativamente menor que 20 horas, ¿Qué tendrá que hacer Profeco? ____

En este caso, ¿Cuál es la importancia de utilizar una muestra y no utilizar a toda la población (Si es que aún existe o ya existe)? _______________________________
____
____
____

Actividad. Lee la siguiente parte del artículo y realiza un resumen.

Febrero 2007 Biomédicas

Estadística, biomedicina y método científico

Tzipe Govezensky, Grupo de Biología teórica

Un paso fundamental y característico del método científico es la experimentación. El experimento puede consistir en recolectar datos (observacional), puede consistir en un conjunto de simulaciones en computadora, o bien puede tratarse de experimentos controlados (experimentos diseñados). La estadística está íntimamente ligada al método científico, ya que nos proporciona una herramienta matemática y, por lo tanto, objetiva para recoger, organizar, resumir y analizar los datos provenientes de la experimentación.

En cuanto a la recolección de datos, por razones de tiempo y costos, casi nunca existe la posibilidad de medir la variable que queremos conocer en toda la población que nos interesa; es decir, tenemos que tomar una muestra. Este hecho plantea varios problemas: ¿Cómo debemos escoger las unidades experimentales que incluiremos en el experimento? ¿Cuántas unidades se requieren? La estadística sugiere diferentes métodos de muestreo para diferentes situaciones; en el caso de experimentos en el laboratorio, debemos cuidar que las unidades experimentales sean tomadas al azar. También existen ecuaciones estadísticas para calcular el número mínimo de unidades experimentales que se requieren en un experimento determinado. Es importante saber que este número depende de la variabilidad del fenómeno en cuestión, de qué tan preciso queremos que sea nuestro resultado, de cuál es la diferencia mínima que nos interesa detectar y de cuántos grupos queremos comparar. Si la heterogeneidad de la variable en estudio es desconocida, será necesario hacer un experimento preliminar para evaluarla. En esta etapa de planeación, la estadística ha aportado muchas ideas acerca del diseño de experimentos.

Organizar y resumir los resultados de un experimento puede hacerse de muchas maneras, no necesariamente requerimos del apoyo del método estadístico. Cuando los resultados son deterministas, por ejemplo: la secuencia de una proteína, la cinética resultante de un modelo matemático, etcétera, la descripción de los resultados es directa. Pero por lo general, los resultados en biomedicina tienen la heterogeneidad característica de los seres vivos, aunado al error experimental y a la dificultad derivada de experimentar con muestras, casi siempre pequeñas, de la población en estudio. Con este tipo de resultados, aun la descripción de los resultados puede ser difícil y, en estos casos, podemos hacer uso de métodos propuestos por la estadística descriptiva. Los datos pueden resumirse como el promedio o la mediana de las mediciones realizadas. La recomendación de la estadística es incluir siempre alguna medida de variabilidad, como el rango de las mediciones o su desviación estándar. Si queremos dar una descripción completa de los datos obtenidos, la recomendación será presentar la distribución obtenida ya sea en forma de tabla o en forma gráfica; por ejemplo, una posible salida del citofluorómetro (FACS) es la distribución de las células analizadas. Otra técnica de estadística descriptiva utilizadas en biomedicina es el análisis de conglomerados (*cluster*) que se utiliza, por ejemplo, en la elaboración de árboles filogenéticos. Ensayos que tienen como resultado una gran cantidad de datos, como los microarreglos, también presentan dificultades a nivel de organización y resumen de los datos; a estos resultados se les ilustra a través de la técnica de componentes principales, y también se emplea el análisis de conglomerados.

Una vez que tenemos nuestros resultados ¿Cómo podemos obtener conclusiones válidas para toda la población basándonos en los datos de una muestra? ¿Cómo evitar conclusiones subjetivas a pesar de la heterogeneidad biológica sumada al error experimental? Es en esta etapa de análisis e interpretación de los datos donde se aplica la **inferencia estadística** cuya característica principal es la de poder asociar a la conclusión obtenida un valor de confiabilidad. El desarrollo de la teoría de la probabilidad ha fortalecido el alcance de las aplicaciones de la estadística, y se basa en ella para calcular dicho valor de confiabilidad. En principio nos gustaría y nos facilitaría el análisis de nuestros datos, el saber que estamos 100 por ciento seguros del resultado obtenido, pero el estudio de variables heterogéneas nos obliga a tener que pensar en términos de probabilidades, las cuales debemos saber interpretar adecuadamente.

En los casos más sencillos, en los que estamos interesados en evaluar solamente una variable, un resultado estadístico no será de la forma: "el promedio de la variable en la población es *M*", sino de la forma "hay una probabilidad *P* de que el promedio real de la variable en la población se encuentre entre *A* y *B*"; y aunque la información parece más difusa, en realidad resulta ser más precisa. Estamos familiarizados con este tipo de respuestas en el caso de los análisis clínicos, en donde se nos informa el "rango normal" de la variable que medimos. En este ejemplo el "rango normal" no abarca al 100 por ciento de las personas sanas, el médico usa su criterio y quizá valores de otras variables para decidir si un paciente con valores fuera de dicho rango puede considerarse "normal".

Las pruebas estadísticas para comparar entre un valor fijo y uno experimental, o entre varios grupos experimentales, ya sean pruebas paramétricas o no-paramétricas, se basan en la idea siguiente: derivar la distribución de una medida numérica calculada utilizando los datos experimentales suponiendo que los tratamientos aplicados en el experimento no tienen ningún efecto, esto es lo que se conoce como hipótesis nula (H_0); de esta manera, no se requiere conocer los resultados para proponer la prueba estadística. Asimismo, contando con la distribución mencionada, podemos saber la probabilidad de obtener determinados resultados si la hipótesis nula se cumple, y podemos fijar cuáles resultados consideraremos muy poco probables, pensando por lo tanto, que no pertenecen a la

Continúa en la página 11

Resumen:

____

____

____

____

____

____

____

____

____

____

Actividad.

1. Busca otro artículo cuyo experimento o estudio mencione variable, población y muestra. Sugerencia: En la revista del consumidor puedes encontrar varios ejemplos.
2. Anota la variable de estudio, la población y la muestra que mencione el artículo.
3. Anota la referencia bibliográfica del artículo.

____

____

____

____

____

____

____

____

Recopilación de datos

A todo conjunto de n observaciones tomado a partir de una población se le llama muestra. El tipo de muestra que resulta de interés principal para los especialistas en estadística es aquel que es verdaderamente representativo de la población a partir de la cual se selecciona. Por lo que una muestra aleatoria se espera tenga esta característica[1]

Frecuencia. Se refiere al número de veces que ocurre un valor particular o fenómeno.

Frecuencia de un intervalo. Se refiere al número de valores que caen dentro del intervalo.

Frecuencia relativa de un intervalo. Se refiere a la proporción de todos los valores dados que caen dentro del intervalo.

Tablas de frecuencias. Llamada también distribución de frecuencias es un arreglo sistemático de los valores agrupados en el intervalo de clase. Se usan para resumir datos de tal modo que la frecuencia de cada intervalo esté claramente mostrada y pueda calcularse fácilmente la frecuencia relativa de cada intervalo.

Construcción de una tabla de frecuencias para datos cualitativos. Para este tipo de datos, las clases están inherentemente definidas. Por esa razón, para construir una tabla de frecuencias simplemente hacemos un conteo del número de puntos de los datos que caen dentro de cada clase y entonces calculamos la frecuencia relativa.

Ejemplo

Se registró la marca de los coches en el estacionamiento de la escuela

Marca	**Número de coches (f_i)**	**Porcentaje (fr_i)**
Ford	*26*	
Chrysler	*30*	
Volkwagen	*28*	
Nissan	*18*	
Toyota	*8*	
Total		

[1] Chao, L. (1989). *Introducción a la estadística*. cecsa

Actividad

Con tu profesor construye la tabla de frecuencias para datos cuantitativos para el siguiente conjunto de observaciones:

22	*08*	*15*	*19*	*13*	*23*	*23*	*09*
20	*17*	*11*	*11*	*13*	*17*	*11*	*10*
19	*26*	*17*	*23*	*14*	*24*	*21*	*17*
15	*14*	*21*	*20*	*10*	*26*	*13*	*11*
05	*21*	*13*	*15*	*13*	*07*	*16*	*15*

Ejercicio

Construye una tabla de frecuencias para cada uno de los siguientes conjuntos de datos

a) Los siguientes son los tiempos (en minutos) al pinchar 32 dedos para comprobar la coagulación

1.42	*1.38*	*1.42*	*1.46*	*1.21*	*1.49*	*1.41*	*1.66*
1.42	*1.40*	*1.37*	*1.39*	*1.45*	*1.23*	*1.48*	*1.43*
1.42	*1.57*	*1.46*	*1.41*	*1.36*	*1.40*	*1.37*	*1.40*
1.37	*1.38*	*1.34*	*1.32*	*1.33*	*1.42*	*1.27*	*1.36*

b) Los siguientes corresponden a las calificaciones obtenidas en un examen

6.5	*7.3*	*9.4*	*5.4*	*6.4*	*7.4*	*4.2*	*3.4*	*8.9*	*4.3*	*0.4*
2.4	*9.8*	*9.4*	*6.7*	*7.6*	*6.5*	*5.4*	*3.4*	*6.3*	*8.7*	*9.9*
6.7	*2.6*	*5.6*								

Representaciones gráficas

Una vez que hemos sabido valorar la importancia de la recopilación de datos en el proceso de una investigación y sabemos construir tablas de distribución de frecuencias para representar el comportamiento de las variables cualitativas y cuantitativas, estamos listos para poder representar nuestra información gráficamente.

Histograma. Un histograma es una representación gráfica para datos cuantitativos, por ejemplo: estatura de los estudiantes, peso corporal, contenido de llenado de refresco, etc. En un histograma los límites reales de cada uno de los intervalos de clase que constituyen la agrupación de datos están marcados sobre el eje horizontal y las frecuencias o frecuencias relativas están marcadas sobre el eje vertical. El histograma se construye por medio de rectángulos unidos cuyas bases son las de los intervalos de clase que conforman los límites reales que ellos representan y cuyas alturas representan a las frecuencias o a las frecuencias relativas. Ejemplifiquemos lo anterior considerando la siguiente distribución de frecuencias que corresponde al Número de minutos que utilizan los alumnos para trasladarse de su casa a la escuela:

Li	*Ls*	*Frecuencia*
10	14	10
15	19	12
20	24	7
25	29	3
30	34	2
35	39	1

Los límites reales de cada uno de los intervalos anteriores son:

Li Real	*Ls Real*

En el histograma podemos observar que existe una concentración mayor de las observaciones en los primeros 3 intervalos, con lo que podemos interpretar de que los alumnos no requieren de mucho tiempo para trasladarse de su casa a la escuela.

Ejercicio

Elabora el histograma de la siguiente distribución de frecuencias, la cual corresponde al número de cigarrillos que solían fumar al día 28 personas hospitalizadas por enfisema pulmonar:

Li	Ls	f
2	4	3
5	7	0
8	10	10
11	13	10
14	16	5

Ejercicio

Elabora el histograma de la siguiente distribución de frecuencias, la cual corresponde a la duración en minutos de 26 conversaciones en teléfono celular:

Li	Ls	f
0.2	0.4	10
0.5	0.7	10
0.8	1.0	3
1.1	1.3	2
1.4	1.6	1

Polígono de frecuencias. Un polígono de frecuencias es una representación gráfica para datos cuantitativos, para trazar un polígono de frecuencias usamos segmentos de recta que conecten los puntos medios de los techos de los rectángulos que representan los intervalos de clase, cerramos el polígono por los extremos, prolongando los segmentos de recta de cada uno de ellos de tal forma que toquen el eje horizontal en el punto medio del siguiente intervalo de clase.

Con lo que podemos observar una mayor concentración de los datos entre los 9.5 minutos y los 24.5 minutos.

Ejercicio

Elabora el polígono de la distribución de frecuencias, la cual corresponde al número de cigarrillos que solían fumar al día 28 personas hospitalizadas por enfisema pulmonar.

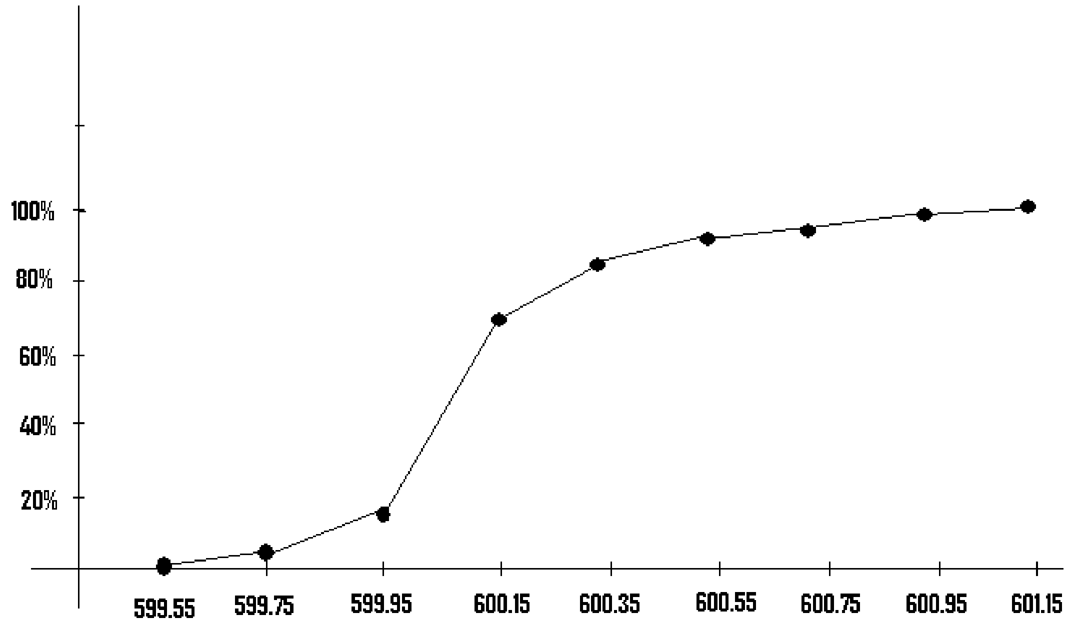
Ejercicio

Elabora el histograma de la distribución de frecuencias, la cual corresponde a la duración en minutos de 26 conversaciones en teléfono celular

Ojiva. Una ojiva es una representación gráfica de las distribuciones de frecuencias acumuladas de los datos cuantitativos, la cual por lo regular se utiliza para índices y porcentajes de variación. Para construir una ojiva, debemos colocar el límite inferior de la clase más baja sobre el eje x, para mostrar que no existe una observación de esa magnitud o menor. Después directamente encima del límite superior de la clase más baja, colóquese un punto en la altura igual a la frecuencia de la clase. Ahora, en el límite superior de la clase después de la más baja, colóquese un punto un punto igual a la frecuencia acumulada. Continuamos de esta forma hasta que se coloque un punto que represente la frecuencia acumulada total en un lugar inmediatamente encima del límite superior de la clase más alta. Por último, unimos estos puntos por medio de líneas rectas. Es indistinto utilizar la frecuencia acumulada o la frecuencia relativa acumulada._Ejemplifiquemos esto construyendo la ojiva de la siguiente distribución de frecuencias referente al contenido de envases de refresco de la marca Coca-Cola en su presentación de 600 mililitros:

Li	*Ls*	*Li Real*	*Ls*Real	*Marca de clase*	*Frecuencia*	*Frecuencia Relativa*	*Frecuencia Relativa Acumulada*
599.6	*599.7*				*1*	*1.52*	*1.52*
599.8	*599.9*				*12*		*19.70*
600	*600.1*			*600.05*	*35*	*53.03*	
600.2	*600.3*			*600.25*	*12*	*18.18*	*90.91*
600.4	*600.5*				*4*		*96.97*
600.6	*600.7*				*0*	*0.00*	*96.97*
600.8	*600.9*				*1*		
601	*601.1*			*601.05*	*1*	*1.52*	*100.00*

La ojiva correspondiente a esta distribución de frecuencias es:



Interpretación. Observa que una vez que ha llegado a los 600.35 mililitros existe una cierta estacionalidad del valor de la variable, esto significa que casi ha llegado a abarcar el 100% de las observaciones.

Ejercicio

Elabora la ojiva de la distribución de frecuencias, la cual corresponde al número de cigarrillos que solían fumar al día 28 personas hospitalizadas por enfisema pulmonar.

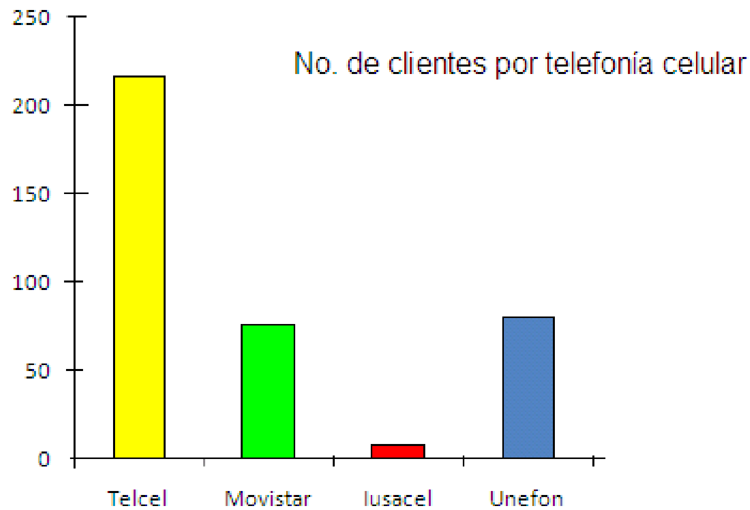
Ejercicio

Elabora la ojiva de la distribución de frecuencias, la cual corresponde a la duración en minutos de 26 conversaciones en teléfono celular.

Gráfica de barras. La gráfica de barras se utiliza muy a menudo para representar datos cualitativos, por ejemplo: marca de celular, tipo de sangre, género, estado civil, etc. En esta gráfica la anchura de cada una de las barras no tiene importancia, todas deben ser idénticas y las barras no se deben de solapar y tampoco se deben de unir, deben estar separadas procurando que el espacio que las separa sea el mismo en cada par de barras. De hecho, es irrelevante unir las barras, ya que no existe ningún orden lógico. La construcción de este tipo de gráfica es muy simple, se colocan sobre el eje horizontal marcas de acuerdo con las frecuencias correspondientes, de esta forma cada renglón de la tabla de frecuencias se convierte en una barra de la gráfica. Por ejemplo, se tiene la siguiente distribución de frecuencias:

Telefonía celular	Número de clientes
Telcel	216
Movistar	76
Iusacel	8
Unefon	80

La gráfica correspondiente a la distribución anterior es:



Ejercicio

Utiliza la información que te pueden brindar tus compañeros para obtener una distribución de frecuencias para representarla por medio de una gráfica de barras. Interpreta la gráfica.

<u>Gráfica circular.</u> La gráfica circular también se le conoce como gráfica de sectores o gráfica de pastel. Esta gráfica sirve para representar datos cualitativos, por ejemplo: partido político, raza, religión, etc.

La construcción de este tipo de gráfica requiere de la distribución de frecuencias relativas, ya que cada una de estas frecuencias nos proporcionará el porcentaje de "pastel" que le corresponde a cada uno de los valores de la variable. Sabemos que el círculo consta de 360° y la totalidad del círculo estará representado por el 100% y cada uno de los valores estará representado por un sector. Entonces, una frecuencia relativa igual a 30% ¿a cuántos grados equivale? Recurramos a una regla de 3 para obtener la respuesta:

$$360° \rightarrow 100\%$$
$$\square° \rightarrow 30\%$$

El número de grados que equivale al 30% es $\frac{(360°)(30\%)}{100\%} = 108°$.

Otro ejemplo, ¿a cuántos grados equivale una frecuencia relativa igual a 58.5%?

Entonces:

$$360° \rightarrow 100\%$$
$$\square° \rightarrow 58.5\%$$

El número de grados que equivale al 30% es $\frac{(360°)(58.5\%)}{100\%} = 210.6°$.

Ejercicio.

a) ¿a cuántos grados equivale una frecuencia relativa igual a 60.9%?

b) ¿a cuántos grados equivale una frecuencia relativa igual a 12.5%?

c) ¿a cuántos grados equivale una frecuencia relativa igual a 50%?

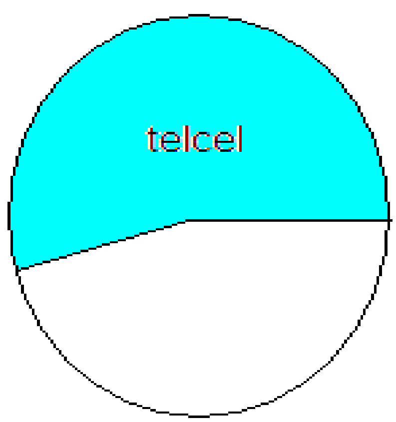
Ahora estamos listos para realizar un ejemplo de gráfica circular.

Requerimos de la distribución de frecuencias relativas y la equivalencia en grados de cada una de las frecuencias que componen dicha distribución:

Telefonía celular	**Número de clientes**	**Porcentaje (Frecuencia Relativa)**	**Equivalencia en grados**
Telcel	216	$216/380x100 = 56.84\%$	$\frac{(360°)(56.84\%)}{100\%} = 204.6° \approx 205°$
Movistar	76		
Iusacel	8		$\frac{(360°)(2.11\%)}{100\%} = 7.596° \approx 8°$
Unefon	80	$80/380x100 = 21.05\%$	
Total	$\mathbf{380}$	$\mathbf{100}\%$	$\mathbf{360°}$

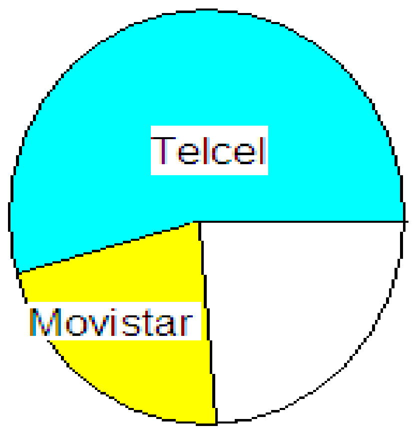
Comencemos a construir los sectores:

El primer sector representa al porcentaje del número de clientes de Telcel, correspondiéndole 205° del círculo (del pastel):

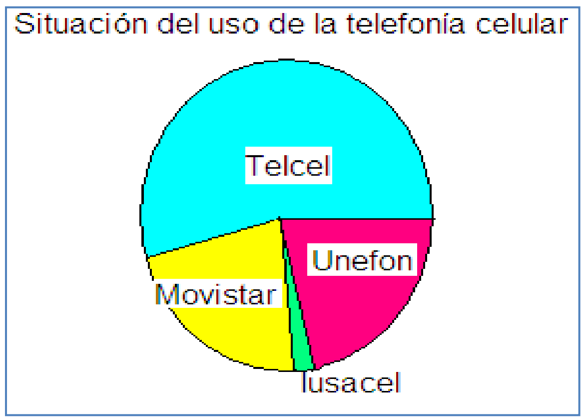


Sugiero considerar los grados como habitualmente lo hacemos, es decir los primeros 90% en el primer cuadrante.

El siguiente sector representa el porcentaje del número de clientes de Movistar, correspondiéndole 72°:



El siguiente sector representa el porcentaje del número de clientes de Iusacel, correspondiéndole 8°, y, lo restante a Unefón correspondiéndole 76°.
Con esto y con un título a la gráfica que informe lo que estamos representando finalizamos la gráfica:



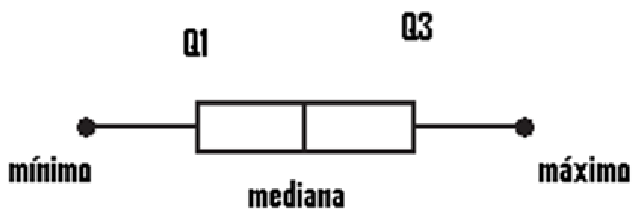
Interpretación. Observa en la gráfica como la empresa de telefonía Telcel abarca con la mayor parte del pastel, es decir de los clientes.

Ejercicio

Pregunta a cada uno de tus compañeros ¿Qué televisora abierta prefieres? Recaba la información y representa la distribución de frecuencias por medio de una gráfica circular. Interpreta la gráfica.

Gráfica de caja. La gráfica de caja es una herramienta útil para mostrar cómo se distribuyen los datos con relación a la posición de la medida de tendencia central y la dispersión de los datos. La gráfica de caja se elabora con 5 elementos: dato mínimo, dato máximo, mediana, primer cuartil (Q1) y tercer cuartil (Q3). La línea que divide a la caja es la mediana, los extremos de la caja serán Q1 y Q3 y los "bigotes" que salen de la caja serán el dato mínimo y el máximo.
Con lo que podemos concluir que:

Los datos simétricos están equilibrados o casi equilibrados en el centro. Los datos sesgados están dispersos más hacia un lado del centro que hacia el otro lado.



Esta gráfica de caja muestra un conjunto de datos simétricos

Los datos sesgados hacia la derecha implican que están dispersos más hacia la derecha del centro que hacia la izquierda

Este conjunto de datos está sesgado hacia la izquierda.

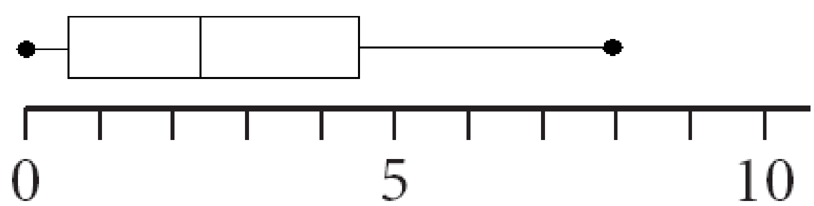
Para ejemplificar considera las calificaciones de 18 estudiantes que presentaron el examen extraordinario de Matemáticas 1 del periodo EB2017: 4.5, 3.2, 5, 4.2, 2.1, 1.5, 0, 0, 0, 2.5, 6.5, 5.5, 3.6, 1, 0.5, 8, 2.1, 0.5

Ordenando los datos de menor a mayor tenemos:

0, 0, 0, 0.5, 0.5, 1, 1.5, 2.1, 2.1, 2.5, 3.2, 3.6, 4.2, 4.5, 5, 5.5, 6.5, 8

Encontremos los 5 elementos base para construir la gráfica:

$$Mínimo = 0, Máximo = 8, Mediana = (2.1 + 2.5)/2 = 2.3,$$
$$Q1 = 0.5, Q3 = 4.5$$



Podemos observar que existe simetría en la variable, ya que la línea que divide a la caja se encuentra al centro de ella. También observamos que existe un sesgo hacia la derecha que implica que los datos estén dispersos más hacia la derecha del centro que hacia la izquierda.

Ejemplo. Considerando las calificaciones de 13 estudiantes que presentaron el examen extraordinario de Matemáticas 1 del periodo EB2017, con la diferencia de que estos 13 tuvieron 20 horas de asesoría antes de presentar dicho extraordinario:

4 4 .5, 5, 3, 3.2, 7, 9.5, 0.8, 6.5, 6.5, 0, 7, 6

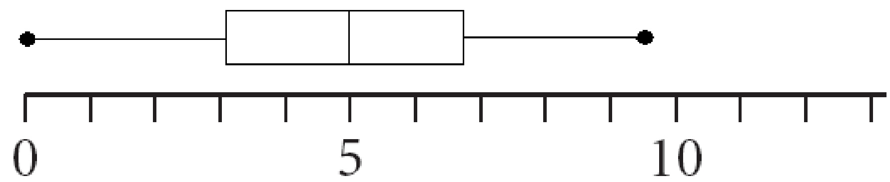
Ordenando los datos de menor a mayor tenemos:

0, 0.8, 3, 3.2, 4, 4.5, 5, 6, 6.5, 6.5, 7, 7, 9.5

Encontremos los 5 elementos base para construir la gráfica:

$$Mínimo = 0 \quad Máximo = 9.5 \quad Mediana = 5$$

$$Q1 = (3 + 3.2)/2 = 3.1 \qquad Q3 = (6.5 + 7)/2 = 6.75$$



Interpretación. Podemos observar que existe simetría en la variable, ya que la línea que divide a la caja se encuentra al centro de ella. También observamos que no existe ningún sesgo.

Ejercicio

Las siguientes observaciones corresponden al diámetro de ciertos tornillos que se utilizan en el sector automotriz, las cantidades proporcionadas en milímetros:
3.4, 3.25, 3.16, 3.24, 3.18, 3.40, 3.19, 3.19, 3.20

Elabora la gráfica de caja correspondiente.

En las gráficas de caja:

¿Qué porcentaje de datos está representado por la caja? ____________________

¿Qué porcentaje representa cada uno de los bigotes? ______________________

¿Se encuentra la mediana siempre en el centro de la caja? __________________

Medidas de tendencia central

La presentación gráfica proporciona una descripción visual general de los datos obtenidos. Sin embargo, hay limitaciones para su empleo. Por una parte, no siempre resulta posible presentar visualmente los datos; por otra, por lo común la descripción gráfica no se presta al tratamiento matemático necesario para el análisis estadístico. Por estas razones, en estadística se prefiere utilizar las medidas tendencias central.

Una medida de tendencia central es un valor típico o representativo de un conjunto de datos. Como tales suelen situarse hacia el centro del conjunto de datos ordenados por magnitud. Las medidas de tendencia central más comunes son: la media, la moda y la mediana.

Media aritmética

La media aritmética, es la suma de los valores de todas las observaciones dividida entre el número de observaciones realizadas.

La media aritmética se expresa simbólicamente de la siguiente forma:

Ejemplo: Obtener la media de los siguientes datos:

19 2 43 12 34 32 25 23 56

Solución: Se calcula a media de la siguiente forma:

Mediana

La mediana de un conjunto de observaciones es el valor que queda en la parte central de un grupo de observaciones ordenadas por su magnitud.

Ejemplo: Obtener la mediana de los siguientes datos:

19 2 43 12 34 32 25 23 56

Solución: Primero ordenamos de mayor a menor o de menor a mayor:

Otro ejemplo: Obtener la mediana de los siguientes datos:

19 2 43 12 34 32 25 23 56 58

Moda

La moda se define como el valor o clase que tiene la mayor frecuencia, en un conjunto de observaciones. Ejemplo, consideremos la siguiente distribución de frecuencias:

Telefonía que emplea	**No. de clientes**
Telcel	234
Iusacel	23
Unefon	45
Movistar	178

La moda es Telcel, de acuerdo con el uso de telefonía.

Para datos cuantitativos podemos obtener la moda. Ejemplo: Obtener la moda de los siguientes datos:

19 2 19 12 19 32 25 23 56

La moda resulta sumamente útil para expresar la tendencia central de observaciones correspondientes a características cualitativas tales como color, estado civil, materia reprobada, etc.

Ejercicios.

1. Obtenga la media, la mediana y la moda de los siguientes conjuntos de datos:
   a) 2, 4, 6, 7, 2, 1, 0
   b) 1, 1, 1, 5, 2, 0, 0, 2
   c) 7.6, 7.3, 7.0, 5.7, 6.0, 10, 9.5, 6.9, 8.5
   d) 0, 0, 0, 1, 1, 1, 2, 2, 4, 3, 3, 5, 6
2. El conjunto de datos 5, ___, 5, 6, 8, 4 tiene media igual a 7, ¿Cuál es el procedimiento que debes seguir para encontrar el dato faltante?
3. El conjunto de datos 10, ___, 5, 6, 8, 4 tiene mediana igual a 7, ¿Cuál es el procedimiento que debes seguir para encontrar el dato faltante?
4. Proporciona un conjunto de datos de 6 elementos con variación (es decir, no deben ser iguales todos) y media igual a 4.
5. Proporciona un conjunto de datos de 8 elementos con variación (es decir, no deben ser iguales todos) y con mediana igual a 10.
6. Proporciona un conjunto de datos de 5 elementos con variación (es decir, no deben ser iguales todos) y con mediana igual a 10.
7. Proporciona un conjunto de 4 elementos con variación (es decir, no deben ser iguales todos) con media, mediana y moda iguales.
8. Proporciona un conjunto de 6 elementos con variación (es decir, no deben ser iguales todos) con media, mediana y moda distintas.
9. Hay 10 personas en un ascensor, 4 mujeres y 6 hombres. La media del peso de las mujeres es de 60 kilos y el de los hombres de 80. ¿Cuál es la media de las 10 personas del ascensor? ___________________

   Con esta información anota, ¿Cuál podría ser el peso de las 4 mujeres si sabemos que no todas tienen el mismo peso? ________

   ¿Y el de los 6 hombres, si sabemos que no todos tienen el mismo peso? ____.

   De acuerdo con tus datos que propusiste, ¿Cuál es la moda de los 10 datos? ________ ¿Y la mediana? _______

Medidas de Tendencia central para datos agrupados

Consideremos la siguiente distribución de frecuencias

Li	Ls	fi
10	14	9
15	19	10
20	24	8
25	29	5
30	34	3
35	39	2

La media estará dada por: $\bar{x} = \frac{\sum fixi}{n}$

Por lo tanto:

Li	Ls	fi		
10	14	9		
15	19	10		
20	24	8		
25	29	5		
30	34	3		
35	39	2		
		$\boxed{}$		$\boxed{}$

$$\bar{x} = \frac{\sum fixi}{n} = \frac{\boxed{}}{\boxed{}} = \boxed{}$$

Para obtener la mediana, utilizamos la fórmula: $x = Li + \left(\frac{\frac{n}{2} - S}{f_{med}} \right) i$

Requerimos de los límites reales:

Li	Ls	$LiReal$	$LsReal$	fi
10	*14*			*9*
15	*19*			*10*
20	*24*			*8*
25	*29*			*5*
30	*34*			*3*
35	*39*			*2*

$$\tilde{x} = LiReal + \left(\frac{\frac{n}{2} - S}{f_{med}}\right) i \;=\; \boxed{} + \left(\frac{\frac{\boxed{}}{2} - \boxed{}}{\boxed{}}\right)\boxed{} = \boxed{}$$

Para obtener la moda, requerimos de la fórmula:

$$\hat{x} = LiReal + \left(\frac{d1}{d1+d2}\right) i$$

Li	Ls			fi
10	14			9
15	19			10
20	24			8
25	29			5
30	34			3
35	39			2

$$\hat{x} = LiReal + \left(\frac{d1}{d1+d2}\right) i \;=\; \boxed{} + \left(\frac{\boxed{}}{\boxed{} + \boxed{}}\right)\boxed{} = \boxed{}$$

Ejercicios. Obtén la media, mediana y moda para las siguientes distribuciones de frecuencias:

Li	*Ls*	*fi*
2	3	10
4	5	12
6	7	8
8	9	10
10	11	5
12	13	1

Li	*Ls*	*fi*
20	29	1
30	39	1
40	49	1
50	59	10
60	69	13

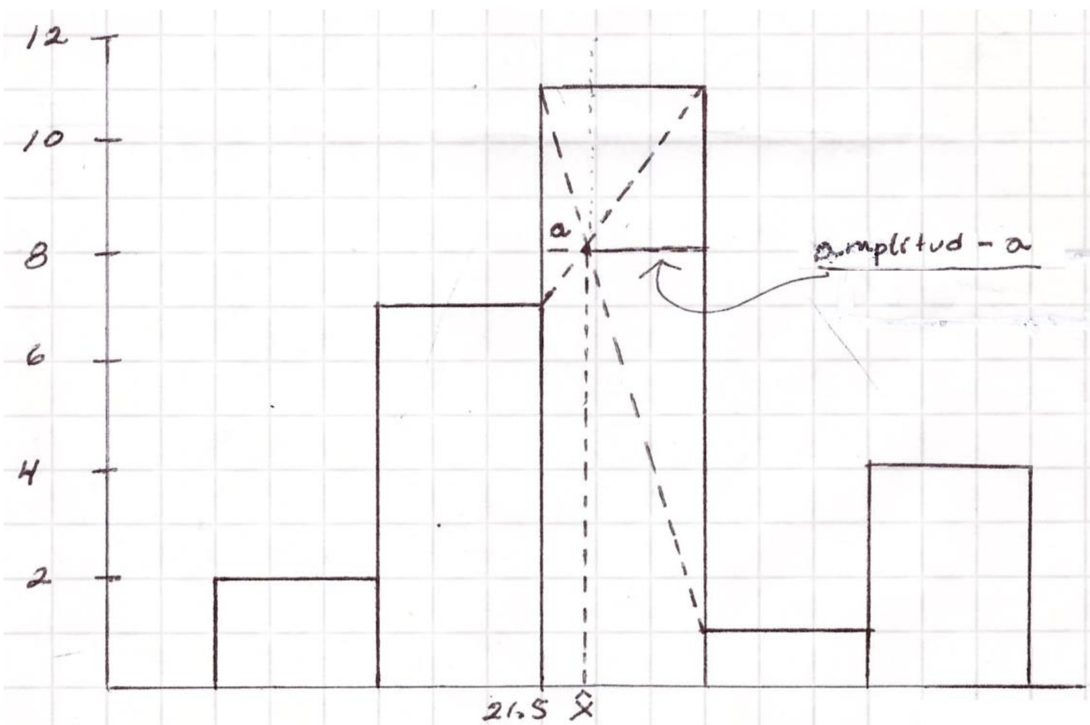
Li	*Ls*	*f*
10	15	2
16	21	7
22	27	11
28	33	1
34	39	4

Mediana y moda para datos agrupados usando el histograma para obtener su valor

Ejemplo. Consideremos la siguiente distribución de frecuencias y obtengamos la moda y la mediana

Li	Ls	f	x_i	LiR	LsR
10	15	2			
16	21	7			
22	27	11			
28	33	1			
34	39	4			

Para obtener la moda:



$$\frac{a}{d_1} = \frac{\text{amplitud} - a}{d_2}$$

En nuestro caso:

$$\frac{a}{4} = \frac{6-a}{10}$$

$$10a = 4(6-a)$$

$$10a = 24 - 4a$$

$$10a + 4a = 24$$

$$14a = 24$$

$$a = \frac{24}{14} = 1.71$$

$$\frac{24}{14} = \frac{4}{4+10}(6)$$

Lo cual equivale a

$$\left(\frac{d_1}{d_1+d_2}\right)(\text{amplitud})$$

$$\hat{X} = li + \left(\frac{d_1}{d_1+d_2}\right)\text{amplitud}$$

$$= 21.5 + 1.71$$

$$= 23.21$$

Para obtener la mediana:

Del histograma, ¿cuál es el valor del área de cada rectángulo?

$AR1 =$ ______ $AR2 =$ ______ $AR3 =$ ______ $AR4 =$ ______ $AR5 =$ ______

$ATotal =$ _________ entonces la mitad del área total es $ATotal/2 =$ _________

En este caso, sumando las áreas de los dos primeros rectángulos, ¿cuánta área resulta? ________

¿cuánta área faltaría para alcanzar la mitad del área total? ________________

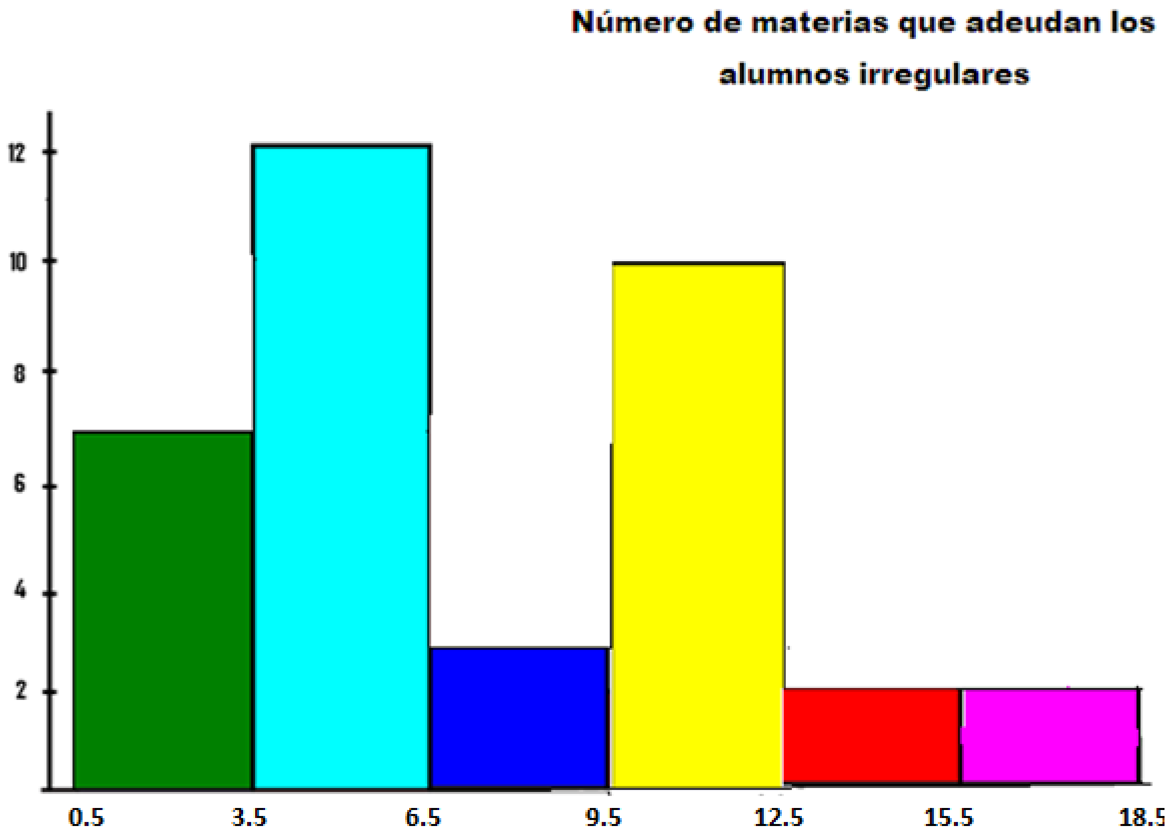
Para que el área del siguiente rectángulo sea $21\ u^2$, el cual tiene una altura (h) de $11\ u$, ¿Cuánto debe medir la base (b)sabiendo que $b\ x\ h\ = 21\ u^2$? __________

Por lo que el valor de la mediana es

$$\tilde{x} = \_\_\_\_\_\_\_\_\_\_$$

Usando la fórmula tenemos que:

Ejercicio 1. Considera el siguiente histograma que representa una distribución de frecuencias y obtén la moda y la mediana



La distribución de frecuencias es:

LiR	LsR	f	x_i
0.5	3.5		2.5
3.5	6.5		
6.5		3	
		2	

Para obtener la moda:

Para obtener la mediana:

Del histograma, ¿cuál es el valor del área de cada rectángulo?

$AR1 =$ ______ $AR2 =$ ______ $AR3 =$ ______

$AR4 =$ ______ $AR5 =$ ______ $AR6 = \underline{6u^2}$

$ATotal =$ _________ entonces la mitad del área total es $ATotal/2 =$ _________

En este caso, considerando el área del primer rectángulo tenemos $21\ u^2$, ¿cuánta área faltaría para alcanzar la mitad del área total? ________________. Toma en cuenta que si tomaras otro rectángulo nos pasaríamos de la mitad.

Para que el área del siguiente rectángulo sea $33\ u^2$, el cuál tiene una altura (h) de $12\ u$, ¿Cuánto debe medir la base (b)sabiendo que $b\ x\ h\ = 33\ u^2$? __________

Por lo que el valor de la mediana es

$$\tilde{x} = \_\_\_\_\_\_\_\_\_\_$$

Usando la fórmula tenemos que:

Ejercicio 2. Considera el siguiente histograma que representa una distribución de frecuencias y obtén la moda y la mediana

Para la moda:

Para la mediana:

Comparación de media, mediana y moda.

1. Los datos siguientes representan el número de niños en cada una de las 10 familias, escogidas aleatoriamente: 2, 9, 1, 1, 4, 5, 2, 3, 0, 1. Obtén media, mediana y moda:

¿Cuál de ellas tres es la que mejor describe el comportamiento de la variable? ___

¿Por qué? ____

____

2. Los datos siguientes representan los resultados de un examen final, aplicado a ocho estudiantes en un curso de estadística: 87, 79, 93, 69, 75, 91, 84, 77. Calcula la media y la mediana.

3. Se tiene la siguiente información referente al gasto semanal por transporte semanalmente de 9 alumnos del CCH-N:

0, 40, 60, 70, 80, 90, 120, 120, 500.

Obtén media, mediana y moda.

El dato con valor de 500 es llamado dato atípico, ya que, de acuerdo con el conjunto de datos, es un valor extremo, y también de acuerdo con el contexto, no es común que un alumno gaste $500 por transporte a la semana. Supón que la persona que está realizando el estudio, se equivocó, y en vez de anotar el 500, debió anotar 1200. ¿Cuál de las medidas de tendencia central se afectaron? ____
Anota sus valores: ____

4. Un estudio de control de calidad, de una producción de papas Sabritas, se basa en un muestreo de 7 bolsas de 40 gramos. La siguiente información corresponde al peso neto de esas bolsas: 3, 39, 40, 40, 40, 40, 40. ¿Cuál de estos pesos, puede considerarse dato atípico? ________ ¿A cuál de las medidas de tendencia central afecta? _______ Obtén el valor de cada una de ellas:

¿Crees que, en este caso, la media puede considerarse como una medida de tendencia central representativa? ______ ¿Por qué? ______________________

____

El encargado de este estudio debe de tomar una decisión: detener el proceso de llenado o continuar trabajando. ¿Qué crees que debe hacer si elige a la mediana como medida de tendencia central representativa? _______

5. Un profesor seleccionó la calificación del primer examen de 10 de sus estudiantes, la información es la siguiente: 0, 0, 1, 1, 1, 2, 9, 10, 10, 10. Obtén media, mediana y moda:

¿Cuál de las medidas de tendencia central consideras más representativa? ______

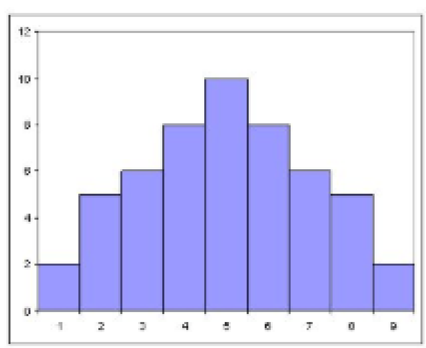
Conclusiones: La media o promedio aritmético, es la medida de posición o tendencia central más comúnmente usada. Sin embargo, no es siempre ideal usarla como un promedio, porque es muy sensible a los valores extremos que causan que una distribución de valores sea oblicua (asimétrica), estos valores extremos también son llamados _______________. Relativamente, el cálculo de la ______ es más difícil de calcular, ya que requiere de sumar todos los números de un conjunto y dividirlos entre el número de elementos del conjunto. La ________ requiere únicamente de ordenarlos. Para la _______ el conteo es todo lo necesario. A pesar de lo anterior, la media es la medida de posición más usada, ya que la problemática de sumar todos los valores para obtenerla, la tecnología la ha reducido en gran medida. Además, la media considera todos los valores del conjunto de observaciones.

La localización de la mediana no está afectada por los _______________. Por ejemplo, el siguiente conjunto de datos que corresponde al número de materias adeudadas de 5 alumnos: 0, 0, 1, 2, 17, contiene un dato atípico, ¿Cuál es? _____ Es decir, si esta observación tuviera un valor de 2 por ejemplo, la mediana seguiría teniendo el mismo valor. Así, cuando los valores extremos están incluidos en los datos y ellos presentan una forma oblicua, frecuentemente usamos la mediana, en lugar de la media. Para cuando los datos son simétricos, los valores de la media y la mediana no difieren mucho. Las desventajas de la mediana son: el ordenamiento de los datos para obtener su valor y la dificultad de manejarla para definirla algebraicamente, al carecer de una fórmula.

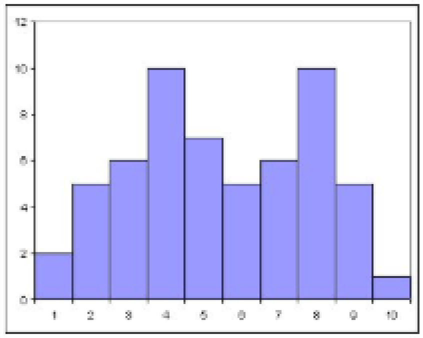
La moda no es tan usada como la ______ y la _________. El concepto de una clase modal puede ser descriptivo, pero rara vez necesitamos localizar a la moda con un solo valor. Para encontrarla necesitamos organizar los datos y además el concepto es difícil de definir en términos algebraicos.

Consideremos las siguientes figuras para visualizar las relaciones entre las tres medidas de posición o de tendencia _________.

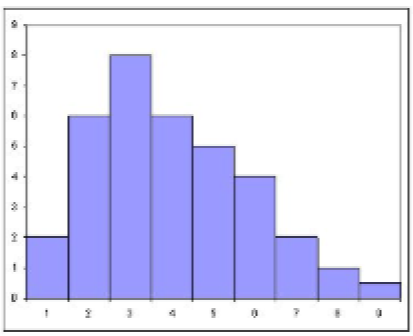
En un suave, simétrico, unimodal polígono de frecuencias, la media, la moda y mediana son iguales.

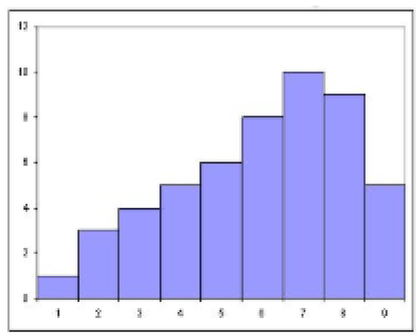


Si el polígono es simétrico bimodal, la media y la moda son iguales, pero hay dos modas.



Cuando el polígono de frecuencias es asimétrico hacia la izquierda o a la derecha, la media, la moda y la mediana producen valores diferentes.





Medidas de dispersión

Analicemos los siguientes 4 conjuntos de datos, los cuales corresponden al número de materias con calificación de DIEZ en los primeros cuatro semestres del CCH de 5 estudiantes:

7, 7, 7, 7, 7 → la media es ____

6, 7, 7, 7, 8 → la media es ____

5, 5, 7, 7, 11 → la media es ____

4, 5, 6, 10, 10 → la media es ____

La media es ___ en los cuatro conjuntos de datos, pero estos son completamente distintos. De hecho, no hay variabilidad alrededor de la media en el primer conjunto, es decir, como la media es 7, no hay ninguna variación entre esta y cada uno de los datos. Pero no sucede lo mismo en los demás conjuntos, de hecho, en el último conjunto, de cierta forma podemos decir que existe una gran variabilidad alrededor de la media.

Existen medidas que reflejan la variabilidad de los datos. El rango es una de esas medidas.

En el primer conjunto el rango es $R = M - m$ = ______

En el segundo conjunto el rango es ______

En el tercer conjunto el rango es ______

En el cuarto conjunto el rango es ______

Sin embargo, observe que el tercer conjunto y el cuarto, tienen el mismo rango, pero son conjuntos que muestran distinta variación con respecto a la media.

En las siguientes líneas representa los datos por medio de un cuadrito para cada uno de los conjuntos:

Para el conjunto 1: ____

Para el conjunto 2: ____

Para el conjunto 3: ____

Para el conjunto 4: ____

¿Qué conjuntos de datos tienen la misma media y el mismo rango? _______

A pesar de estas características en común entre el tercer conjunto y el cuarto, son muy diferentes, observa tus representaciones anteriores. ¿Cuál de los conjuntos consideras que tiene mayor variabilidad, alrededor de la media? ___________ __________ ¿Por qué? ____________________________________

El rango únicamente nos proporciona la extensión del conjunto de datos, pero no proporciona información de la distribución de los datos dentro del conjunto, ya que este considera únicamente el dato más pequeño y el más grande. Existe una medida de variabilidad que considera todas las medidas del conjunto de datos llamada VARIANZA, la cual toma en cuenta la diferencia entre cada dato respecto a la media. Dicha diferencia es llamada desviación. Si la diferencia cae a la izquierda de la media, la llamaremos desviación negativa y si cae a la derecha de la media la llamaremos desviación positiva.

Pero existe un problema, consideremos el siguiente conjunto de datos, los cuales corresponden al contenido de refresco de 6 envases de 600 mililitros:

$$598,\ \ 599,\ \ 600,\ \ 600,\ \ 601,\ \ 602$$

¿Qué valor tiene la media? $\bar{x} =$ _____

Y, ¿Cuáles son las diferencias entre cada dato con respecto a la media?

_____, _____, _____, _____, _____, _____

Si sumamos estos valores obtenemos un valor de _______.

La suma de las desviaciones negativas invalida a la suma de las desviaciones positivas, resultando con ello una medida inútil de variabilidad.

Para evitar este problema, elevamos cada desviación al cuadrado, propiciando con ello que todos estos sean positivos. Sumamos los cuadrados de las desviaciones y dividimos entre el número de datos para promediar. Lo anterior, produce la VARIANZA, que indica la desviación, con respecto a la media.

Debido a que elevamos al cuadrado cada una de las desviaciones de los datos con respecto a la media, tomamos la raíz cuadrada de la varianza, para tener de nuevo las mismas dimensiones y esta se asemeje más a la medida promedio de las desviaciones. La raíz cuadrada positiva de la VARIANZA la llamaremos DESVIACIÓN ESTÁNDAR.

Para datos no agrupados, la fórmula de la varianza que utilizaremos es la siguiente:

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

La desviación estándar es la raíz cuadrada de la varianza.

Ejemplo: Consideremos el siguiente conjunto de datos que corresponde al contenido neto de 7 bolsas de papas Sabritas: $40, 39, 38, 40, 43, 39, 41$. Las cantidades anteriores dadas en gramos. Para realizar el cálculo de la varianza puede convenir usar una tabla:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Suma		

Sustituyendo en la fórmula, tenemos que:

$s^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n-1} = \underline{\hspace{5cm}} =$

Entonces la desviación estándar: $s =$ ________

Con una "s" denotaremos a la desviación estándar.

Ahora, calcula la varianza y desviación estándar del siguiente conjunto de datos, que también corresponde al contenido neto de 7 bolsas de papas Sabritas: $10, 10, 10, 30, 50, 70, 100$. Cantidades dadas en gramos. Usa la siguiente tabla:

¿Qué puedes concluir al comparar las dos desviaciones estándares?

____

____

Otra medida de dispersión, no tan usada como la varianza o desviación estándar es la DESVIACIÓN MEDIA, la cual consiste en calcular el promedio del valor absoluto de las desviaciones a partir de la media. La fórmula para su cálculo es la siguiente:

$$D.M. = \frac{\sum_i^n |x_i - \bar{x}|}{n}$$

Ejemplo: Calcular la D.M. del siguiente conjunto de datos que corresponde a los asaltos denunciados en el Municipio de Naucalpan los últimos 5 días:

$2, 10, 18, 20, 50$ Previamente debemos calcular la media, $\bar{x} =$ ______

Para obtener D.M. conviene usar la siguiente tabla:

Por lo tanto, $D.M. = \frac{\sum_i^n |x_i - \bar{x}|}{n} = \underline{\hspace{4cm}} =$

Ejercicios. Obtén varianza, desviación estándar y desviación media de los siguientes conjuntos de datos:

a) 23, 24, 25, 27
b) 1.78, 1.80, 1.80, 1.81, 1.81
c) -5, 0, 5, -4, 9

Menciona dos conjuntos de datos de 4 elementos que representen calificaciones de dos grupos, con media igual a 7 pero diferente desviación estándar (muestra el valor de *s* de cada conjunto).

Ejercicios para el primer examen

Realiza los siguientes ejercicios (primera parte)

1. El Departamento de control de calidad de una empacadora de manzanas está encargada de inspeccionar 200 manzanas de las que le llegan diariamente de los huertos de donde las adquieren, lo anterior con la finalidad de pesar a cada una de ellas.

a) La población es ________________________________
b) La muestra es ____
c) La unidad de observación es __________________________________
d) La medición es ____
e) La medición es ¿cualitativa o cuantitativa? ______________________

2. Menciona 2 variables de la población de profesores del CCH Naucalpan y menciona 2 posibles valores de cada una de esas variables

Variable: ________Dos posibles valores de esta variable son: _____________
Variable: ___________ Dos posibles valores de esta variable son: __________

3. Se registraron los tiempos de una carrera de 100 metros de 26 estudiantes varones del CCH N los cuales practican atletismo. Los tiempos registrados son los siguientes:

10.9, 12.3, 11.0, 11.2, 11.9, 12.5, 13.0, 12.4, 12.2, 11.8, 11.0, 11.5, 12.6, 13.0, 11.4, 12.0, 11.8, 14.4, 12.5, 12.7, 11.4, 10.8, 11.0, 10.8, 11.3, 14.0

a) Obtén los límites reales y las marcas de clase
b) Obtén la distribución de los datos mediante el procedimiento de agrupación visto en clase
c) Dibuja e interpreta el histograma correspondiente.
d) Dibuja la ojiva correspondiente.

4. Para las elecciones del 2018, una empresa realizó una encuesta a 2000 electores para preguntarles ¿Quién considera es el más corrupto para ocupar la Presidencia? Los resultados fueron los siguientes:

Candidato	Número de personas
José Antonio Meade	500
Andrés Manuel López Obrador	1280
Ricardo Anaya	180
El Bronco	40

Elabora una gráfica de pastel para representar la información.

5. Se registraron durante 10 días los grados centígrados a las 12 horas en el Municipio de Naucalpan: $21, 20, 21, 23, 19, 20, 19, 23.5, 21, 22$. Elabora e interpreta la gráfica de caja correspondiente a los datos registrados.

Realiza los siguientes ejercicios (segunda parte)

I. Los siguientes datos corresponden al número de hojas utilizadas después de finalizar el año escolar del cuaderno profesional de 100 hojas destinado para una sola materia de 10 alumnos de primaria: 94, 90, 98, 12, 94, 91, 94, 94, 88, 85

1. Obtén media, mediana y moda
2. ¿Cuál de las medidas de tendencia central consideras NO consideras representativa y por qué?
3. Sustituyendo el dato que tiene el valor de 12, ¿Qué valor debe de tener para que la media tenga el valor de 94?

II. Los siguientes datos corresponden al total de inasistencias por día de los empleados de una empresa: 0, 1, 1, 0, 2, 0, 0, 0, 0, 16

4) Obtén media, mediana y moda
5) ¿Cuál de las medidas de tendencia central NO consideras representativa y por qué?
6) Sustituyendo el dato que tiene el valor de 16, ¿Qué valor debe de tener para que la media tenga el valor de 1?

III. Del siguiente conjunto de datos: 23, 27, 25, 26, 24, los cuales corresponden a los grados centígrados registrados en Naucalpan, a las 13 horas los días 13, 14, 15, 16 y 17 de septiembre respectivamente.

7. Obtén media, mediana, moda,
8. Obtén desviación media
9. Obtén varianza por medio de la que llamamos primera fórmula
10. Obtén varianza por medio de la que llamamos segunda fórmula
11. Obtén desviación estándar

IV. Considera el siguiente conjunto de datos: 10, 15, 15, 20

12. Obtén varianza por medio de la que llamamos primera fórmula
13. Obtén varianza por medio de la que llamamos segunda fórmula
14. Obtén s
15. Propón otro conjunto de 4 datos, con media y rango igual al conjunto anterior, pero con s (desviación estándar) mayor al primer conjunto. Obtén su valor.

V. Se tiene el siguiente conjunto de datos ordenados de menor a mayor, correspondientes al número de materias que adeudan 6 estudiantes:
_____, 0, 5, 5, 5, ______

16. ¿Qué datos son los que van en cada "rayita" si la mediana, media y moda tienen un valor de 5?

VI. Se tiene el siguiente conjunto de datos correspondiente a las denuncias realizadas en 4 días en un Municipio del Estado de México: 50, 55, 55, ____

17. ¿Qué dato es el que falta si la media es igual a 55?
18. ¿Qué dato es el que falta si la mediana es igual a 52.5?

VII. Considera las siguientes dos muestras correspondientes al peso de contenido neto de dos máquinas que llenan las bolsas de papas Sabritas de 40 gramos:
Máquina 1: 40, 41, 39.9, 39.9, 40, 40, 40, 40, 40.2, 39
Máquina 2: 35, 41, 39, 39.9, 44, 36, 40.9, 41, 40.2, 45

19. Obtén media y desviación estándar
20. Considerando el valor de la media y la desviación estándar de cada muestra, ¿Qué máquina hay que ajustar o dejar de usar? ¿Por qué?

VIII. La siguiente distribución de frecuencias corresponde al número de cigarrillos que solían fumar al día 28 personas hospitalizadas por enfisema pulmonar:

Li	Ls	f_i	x_i	$f_i x_i$	Li	Ls	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$	$\lvert x_i - \bar{x}\rvert$	$f_i\lvert x_i - \bar{x}\rvert$	x_i^2	$f_i x_i^2$
2	4	3											
5	7	0											
8	10	10											
11	13	10											
14	16	5											
		28											

21. Obtén $\bar{x}$ (media),
22. Realiza el histograma, traza el polígono de frecuencias, “suaviza” el polígono de frecuencias e interpreta las gráficas de acuerdo al contexto.
23.
24. Menciona si el polígono de frecuencias es simétrico unimodal, simétrico bimodal, asimétrico con sesgo a la derecha o asimétrico con sesgo a la izquierda
25. Obtén $\tilde{x}$ (mediana) utilizando el histograma
26. Obtén $\tilde{x}$ (mediana) utilizando la fórmula
27. Obtén $\hat{x}$ (moda) utilizando el histograma
28. Obtén $\hat{x}$ (moda) utilizando la fórmula
29. Obtén s^2 (varianza) por medio de la que llamamos primera fórmula
30. Obtén s^2 (varianza) por medio de la que llamamos segunda fórmula y obtén s(Desviación estándar)
31. Obtén D.M. (Desviación media)

IX. La siguiente distribución de frecuencias corresponde a la duración en minutos de 26 conversaciones en teléfono celular:

Li	Ls	f_i	x_i	$f_i x_i$	Li	Ls	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$	$\lvert x_i - \bar{x}\rvert$	$f_i\lvert x_i - \bar{x}\rvert$	x_i^2	$f_i x_i^2$
0.2	0.4	10											
0.5	0.7	10											
0.8	1.0	3											
1.1	1.3	2											
1.4	1.6	1											
		26											

32. Obtén $\bar{x}$ (media),
33. Realiza el histograma, traza el polígono de frecuencias, "suaviza" el polígono de frecuencias e interpreta las gráficas de acuerdo al contexto.
34. Menciona si el polígono de frecuencias es simétrico unimodal, simétrico bimodal, asimétrico con sesgo a la derecha o asimétrico con sesgo a la izquierda
35. Obtén $\tilde{x}$ (mediana) utilizando el histograma
36. Obtén $\tilde{x}$ (mediana) utilizando la fórmula
37. Obtén $\hat{x}$ (moda) utilizando el histograma
38. Obtén $\hat{x}$ (moda) utilizando la fórmula
39. Obtén s^2 (varianza) por medio de la que llamamos primera fórmula
40. Obtén s^2 (varianza) por medio de la que llamamos segunda fórmula y obtén s(Desviación estándar)
41. Obtén D.M. (Desviación media)

Unidad 2

Obtención e interpretación de información estadística con datos bivariados

Propósito. Al finalizar la unidad el alumno analizará la relación entre dos variables estadísticas y realizará predicciones, a partir del reconocimiento y la modelación de dicha relación, evaluando el grado de intensidad en ella, con la finalidad de elevar su capacidad de interpretar y evaluar críticamente la información estadística en dos variables aparejadas.

Aprendizajes. El alumno:

- Distingue que entre dos variables puede existir alguna relación.
- Construye tablas de contingencia para presentar la información correspondiente a dos variables cualitativas aparejadas.
- Examina la información vertida en una tabla de contingencia, en términos de la relación entre dos variables cualitativas, dentro del contexto de una investigación o un problema.
- Valora la importancia de las tablas de contingencia en la presentación y análisis del comportamiento de dos variables cuantitativas aparejadas.
- Examina la información vertida en un diagrama de dispersión, en términos de la correlación entre dos variables, dentro del contexto de una investigación estadística o un problema.
- Descubre que la recta de mínimos cuadrados es la que mejor modela la correlación entre dos variables, cuando ésta se presenta de manera aproximadamente lineal.
- Analiza los estimadores de los parámetros de la recta de mejor ajuste y el coeficiente de correlación, interpretándolos dentro del contexto.
- Identifica que existen otros tipos de relación entre dos variables cuantitativas, además de la lineal.
- Valora, con el apoyo de la computadora, el comportamiento del coeficiente de correlación entre dos variables cuantitativas, ajustando algunos puntos en el plano bidimensional a una relación aproximadamente lineal y/o a una serie de datos no relacionados, dentro del contexto de una investigación o problema.
- Estima, considerando las limitaciones del dominio, el valor de una variable regresora de un valor de la variable de respuesta, por medio de la recta de mejor ajuste entre dos variables cuantitativas, con el apoyo de la computadora.

Introducción

Hasta el momento los análisis se han ocupado de problemas relacionados únicamente con una sola variable. Traten de recordar que han obtenido de esas variables y escríbelas a continuación: media, ________________________

Sin embargo, existen muchas situaciones en las cuales se trabaja con pares de mediciones y debe decidirse si existe relación entre ambas.

Por ejemplo: puede resultar de gran interés la relación entre las estaturas y pesos de estudiantes universitarios, entre los promedios de puntos de calificaciones de los estudiantes en bachillerato y sus promedios de puntos de calificaciones en la universidad.

Menciona otros 2 pares de variables que creas se relacionen: _____________

____

Si existe una relación entre las dos variables que estás considerando, también sería deseable determinar la fuerza de esa relación o dependencia, es decir,

¿crees que todas las parejas de variables se relacionen con una misma fuerza o intensidad? _____

Si existe una relación entre las variables puede suceder que se desee predecir el valor de una variable a partir del valor de la otra.

Por ejemplo, conociendo la estatura de una persona, podemos predecir su peso corporal; y esta predicción, puede ser más efectiva si existe una fuerte relación entre las variables y una mala predicción sino lo es.

¿Crees que existan variables que dependan, no solamente de una variable sino de 2 o más variables? ___________.

Por ejemplo, el peso corporal de una persona, ¿de qué otras variables puede ser que dependa? ___________________. La regresión múltiple es la extensión de la regresión simple y estudia el efecto de más de una variable independiente *X* sobre la dependiente Y.

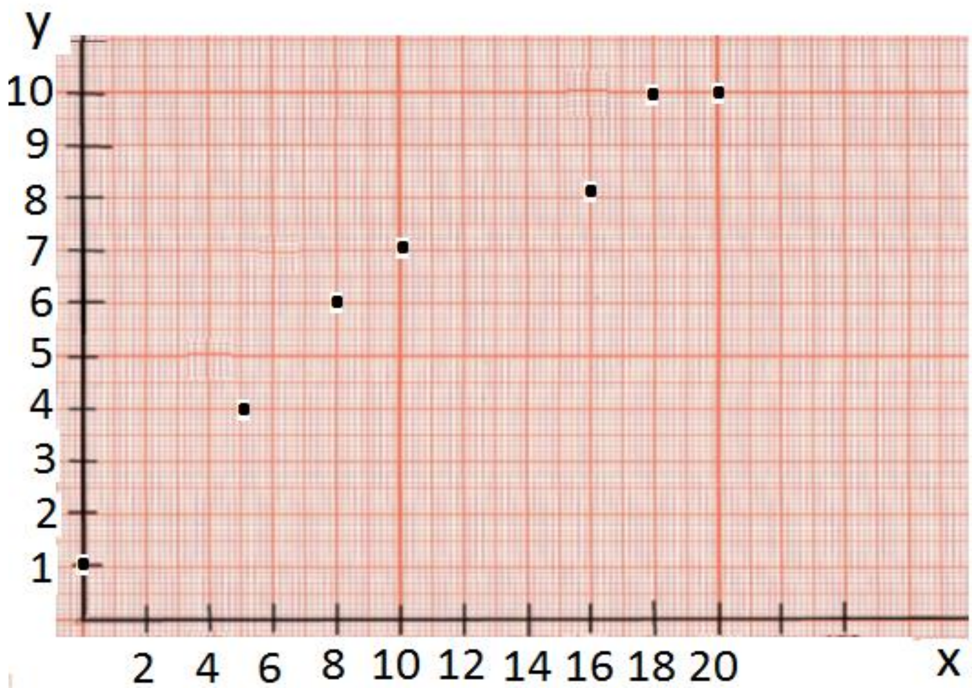
Nosotros únicamente abordaremos la regresión lineal simple.

Ejemplo.

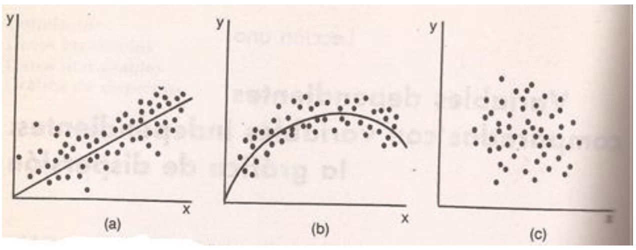
Sean ***X*** el número horas recibidas por parte del PIA para preparar el examen extraordinario de Matemáticas I y ***Y*** la calificación obtenida en dicho examen. Los siguientes datos corresponden a la información que presentaron 7 alumnos que presentaron dicho examen.

Nombre	X	Y
Ricardo Mora	0	1
Javier Gómez	5	4
Leticia Roa	18	10
Mauricio Nava	20	10
Gabriel Pineda	8	6
Karla Vargas	10	7
Miriam Balbuena	16	8

Diagrama de dispersión



2. La siguiente imagen muestra 3 diagramas de dispersión: a) indica una relación lineal, b) indica una relación curvilínea y c) no indica ninguna relación:



Y, en nuestro diagrama de dispersión, ¿Qué tipo de relación será? ¿O no existe relación? ____________________________

Actividad

1. En el diagrama de dispersión, traza la recta que mejor se ajuste a los puntos del diagrama.
2. Dibuja una línea vertical entre cada punto y la recta.
3. Utiliza la siguiente tabla para anotar:
   a) Las mediciones con regla de cada línea vertical trazada
   b) El cuadrado de cada medición
   c) La suma de esas mediciones al cuadrado.

Medición	Medición al cuadrado

Mínimos cuadrados. La técnica de regresión se refiere al procedimiento de obtener una ecuación con fines de estimación o predicción. La variable por estimar o predecir se denomina variable dependiente; y la otra variable, aquella que proporciona la base para la estimación, se denomina variable independiente. En nuestro caso, la variable independiente es: ____________ y la variable dependiente es ________________. Pero entonces, ¿Cómo obtener dicha recta? El método de ajustar una línea recta a pares de observaciones en base a este criterio se conoce como **método de mínimos cuadrados**. Brevemente, el criterio de mínimos cuadrados implica que la recta elegida para ajustar los puntos del diagrama de dispersión sea tal que la suma de los cuadrados de las distancias verticales entre los puntos y la recta sea lo más pequeña posible. Vamos a obtener la recta que mejor se ajusta a nuestros datos mediante el método de mínimos cuadrados y veamos que tan buenos resultaron para calcular.

Hablando en forma más general, si hay n pares de observaciones en la muestra, el criterio de mínimos cuadrados exige que $\sum[y-(mx+b)]^2$ sea la suma mínima. Recuerda que en la actividad anterior realizaste algo que se relaciona con esto, ¿Qué fue? _______________________________________

Entonces, cualquier recta que minimice esta cantidad recibe el nombre de recta de mínimos cuadrados. Esto explica por qué el ajuste de mínimos cuadrados generalmente se considera como el "mejor" ajuste; minimiza la suma de las desviaciones entre los puntos y la recta, elevadas al cuadrado. En la recta de regresión $y = mx + b$ los valores de $\boldsymbol{x}$ y $\boldsymbol{y}$ se dan en la muestra. La intersección con el eje $\boldsymbol{y}$**,** es decir $\boldsymbol{b}$ y la pendiente $\boldsymbol{m}$, son las incógnitas y tienen que calcularse a partir de los datos muestrales. Las fórmulas para obtener los coeficientes son:

$$m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \qquad b = \frac{\sum y - m\sum x}{n}$$

x	y	xy	x^2	y^2
$\sum X =$	$\sum Y =$	$\sum XY =$	$\sum X^2 =$	$\sum Y^2 =$

Para obtener los valores de ***m*** y ***b***, ¿Qué suma no utilizas? __________.

Será de gran utilidad para una fórmula que usaremos posteriormente.

Ahora utilicemos las fórmulas para obtener la recta de mínimos cuadrados:

$$m = \frac{7(\qquad\qquad)-(\qquad\qquad)(\qquad\qquad)}{7(\qquad\qquad\qquad\qquad)-(\qquad\qquad)^2} = \boxed{\qquad\qquad}$$

$$b = \frac{(\qquad\qquad)-(\qquad\qquad)(\qquad\qquad)}{7} = \boxed{\qquad\qquad}$$

La recta que mejor se ajusta a los datos es $y =$ __________$x +$ ___________

Actividad

1. Utilicemos dos puntos por donde pasa la recta, ejemplo si *x = 10*,

   $y =$

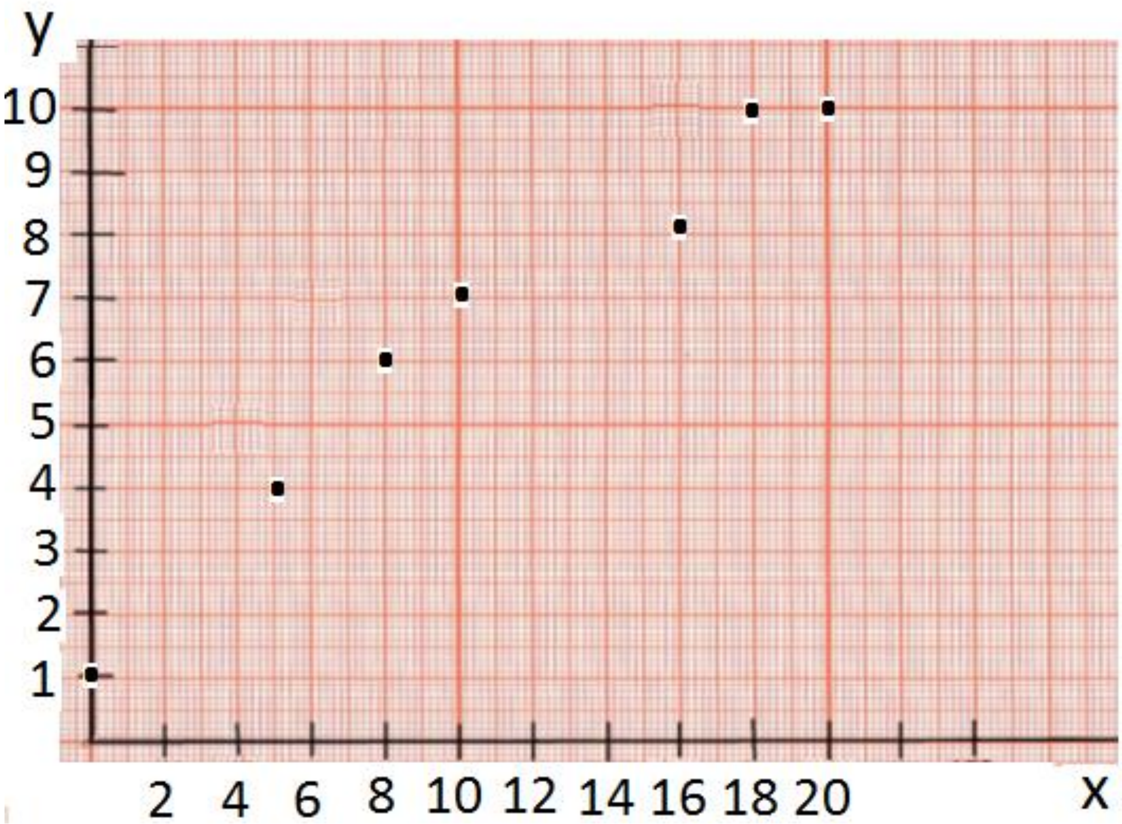
   Entonces un punto por donde pasa la recta es *(,)*

   Otro punto por donde pasa la recta es por ejemplo si *x = 0*

   $y =$

   Entonces otro punto por donde pasa la recta es *(,)*

2. Localicemos ambos puntos en el diagrama de dispersión y tracemos la recta que mejor se ajusta a nuestros datos:



3. Dibuja una <u>línea vertical</u> entre cada punto y la recta.

4. Utiliza la siguiente tabla para anotar:
   a) Las mediciones con regla de cada línea vertical trazada
   b) El cuadrado de cada medición
   c) La suma de esas mediciones al cuadrado

.

Medición	Medición al cuadrado

5. Compara la suma anterior con esta. ¿Si obtuvimos una suma menor? ______

Es esa la razón por la cual se le llama recta de mínimos cuadrados.

Pero ¿Cómo puedo usar la recta?

Si *x*= *12* ¿Qué valor estiman o predicen para Y?

$$y =$$

La interpretación es: ____

____

____

Correlación. El objetivo de estudio de correlación es determinar la consistencia de una relación entre observaciones por pares. El término correlación literalmente significa relación mutua, ya que indica el grado en el que los valores de una variable se relacionan con los valores de otra. La medida usual del grado de correlación basándose en una muestra de n pares de observaciones es el coeficiente de correlación, comúnmente denotado r.

La fórmula para el coeficiente de correlación es:

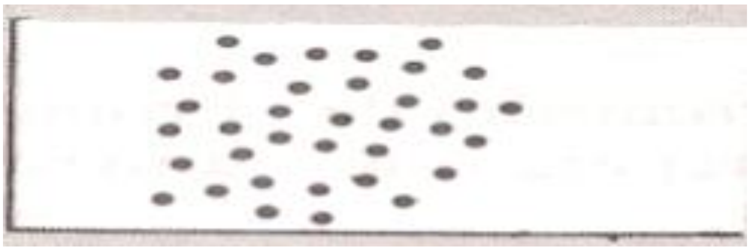
$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2\ ][n\sum y^2 - (\sum y)^2]}}$$

El coeficiente de correlación presenta dos propiedades que establecen la naturaleza de una relación entre dos variables. Una es su signo (+ o -) y la otra, es su magnitud. El signo es igual al de la pendiente de una recta que podría “ajustarse” a los datos si éstos se graficarán en un diagrama de dispersión, en la recta que obtuvimos con el método de mínimos cuadrados, ¿Qué signo tiene? _____ Esto quiere decir que el valor de ***r*** también tendrá este signo. La magnitud de ***r*** indica cuán cerca están de la “recta” tales puntos. Por ejemplo, los valores próximos a -1.00 o +1.00 indican que los valores están bastante cerca de la recta o sobre ella, mientras que los valores próximos a 0 sugieren mayor dispersión. A continuación, se muestran diversos diagramas de dispersión para proponer ejemplos de la interpretación que se le puede dar al coeficiente de correlación:

Relación positiva perfecta

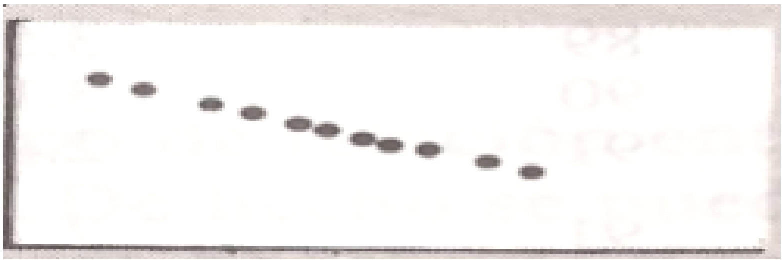
Relación positiva moderada

No existe relación



Relación negativa moderada

Relación negativa perfecta



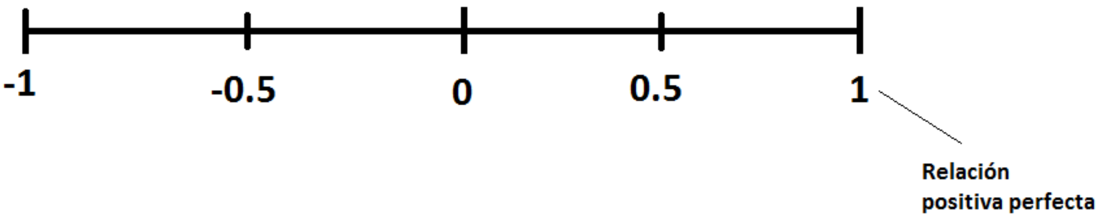
De acuerdo con la información anterior, ¿Cuál piensas que sea el grado de correlación entre las variables ***X*** y ***Y*** de nuestro caso?

____

____

¡Claro tu respuesta anterior fue con base a la observación del diagrama de dispersión!

Completa la información del siguiente esquema para representar lo que indica la magnitud de r, toma en cuenta solamente para la relación positiva y negativa perfecta es un valor específico:



¿Cuál es el valor de r en nuestro ejemplo? ________________________________

¿Cuál es la interpretación de este valor? _________________________________

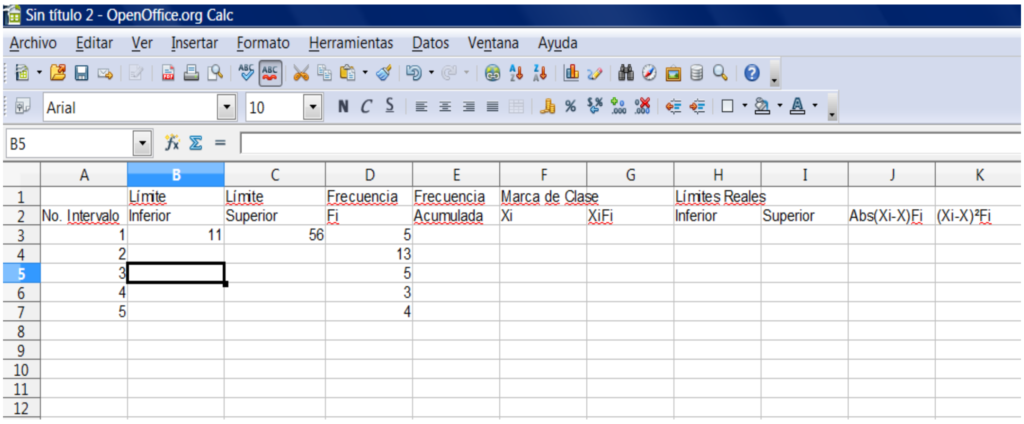
Práctica de estadística en Open Office principalmente

Ésta práctica puede utilizarse para Excel Office, los comandos son los mismos, sólo que la sintaxis de las fórmulas es diferente; por ejemplo en lugar de poner punto y coma (;) se ponen una coma (,).

El objetivo de la siguiente práctica tiene el propósito de aplicar los conocimientos de estadística descriptiva y regresión lineal. Se solicita que sigas los siguientes pasos y hagas las anotaciones correspondientes. Al terminarla se pide que elabores una reflexión donde incluyas conceptos y funciones de Excel que hayas aprendido al final de la práctica.

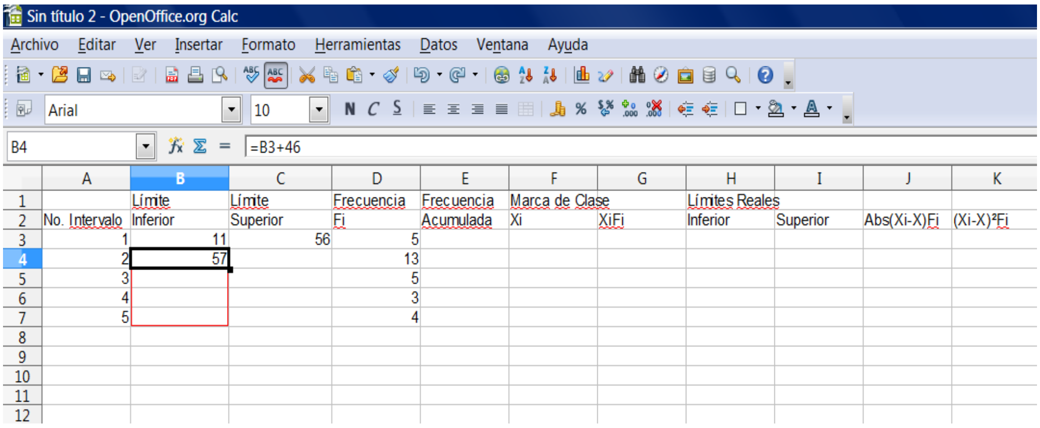
Estadística descriptiva: Tabla de Distribución de Frecuencia de robos en una Delegación

1. Abre open office Calc (hoja de cálculo parecida a Excel). A continuación teclea los siguientes datos en la columna correspondiente



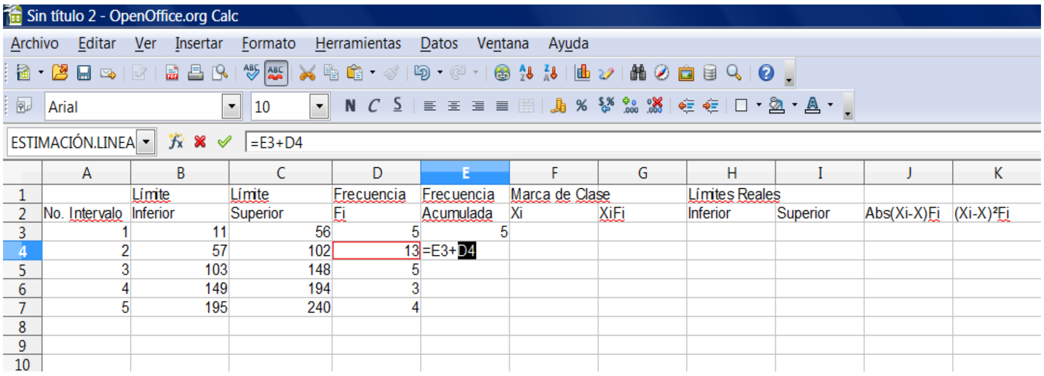
	A	B	C	D	E	F	G	H	I	J	K
1		Límite	Límite	Frecuencia	Frecuencia	Marca de Clase		Límites Reales			
2	No. Intervalo	Inferior	Superior	Fi	Acumulada	Xi	XiFi	Inferior	Superior	Abs(Xi-X)Fi	(Xi-X)²Fi
3	1	11	56	5							
4	2			13							
5	3			5							
6	4			3							
7	5			4							

2. Para calcular el siguiente límite inferior es necesario incluir la siguiente fórmula en B4 =B3+46 enseguida dar enter, el siguiente límite será de 57. A continuación poner el cursor en la orilla de la celda B4, al observar que se pone una cruz arrastrar el mouse hasta llegar a la celda B7 y a continuación soltarlo para copiar la fórmula en las siguientes celdas.



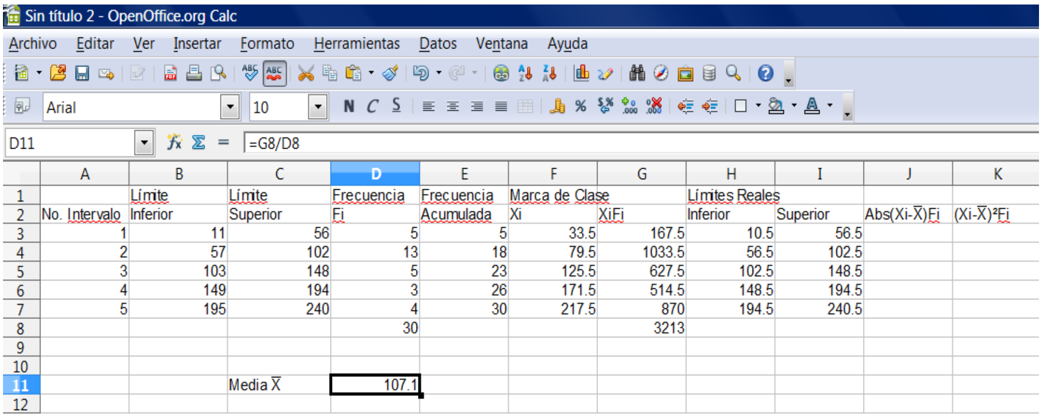
	A	B	C	D	E	F	G	H	I	J	K
1		Límite	Límite	Frecuencia	Frecuencia	Marca de Clase		Límites Reales			
2	No. Intervalo	Inferior	Superior	Fi	Acumulada	Xi	XiFi	Inferior	Superior	Abs(Xi-X)Fi	(Xi-X)²Fi
3	1	11	56	5							
4	2	57		13							
5	3			5							
6	4			3							
7	5			4							

3. Ahora se hará lo mismo en la columna C de límite superior. Se selecciona B4, se oprime Ctrl C al mismo tiempo, para copiarla, enseguida se selecciona C4 y se oprime al mismo tiempo Ctrl V. Presentándose el número 102. A continuación se puede arrastrar el mouse o bien, copiar las siguientes fórmulas para obtener todos los límites. Para calcular la frecuencia acumulada en E3 se copia el valor de D3. En E4 se utilizará la fórmula =E3+D4
En E5 la fórmula será =E4+D5 y así sucesivamente hasta obtener toda la columna.



	A	B	C	D	E	F	G	H	I	J	K
1		Límite	Límite	Frecuencia	Frecuencia	Marca de Clase		Límites Reales			
2	No. Intervalo	Inferior	Superior	Fi	Acumulada	Xi	XiFi	Inferior	Superior	Abs(Xi-X)Fi	(Xi-X)²Fi
3	1	11	56	5	5						
4	2	57	102	13	=E3+D4						
5	3	103	148	5							
6	4	149	194	3							
7	5	195	240	4							
8											
9											
10											

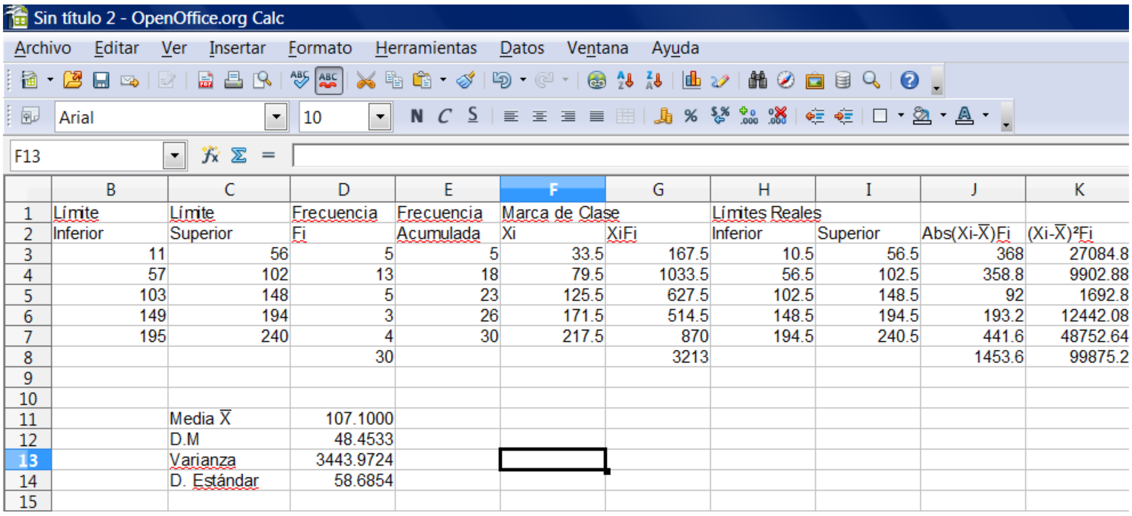
4. Para obtener Xi, se puede utilizar la función promedio, en la celda F3, se incluirá   =PROMEDIO(B3;C3)   y así sucesivamente hasta completar la columna.
5. Para la columna G, en G3 se utilizará =F3*D3 , completa la columna. En la celda H3 se tecleará la fórmula =B3-.5 y en la celda I3 la fórmula a usar será = C3+.5 Ahora completa las columnas
6. Para la columa J primero se debe calcular la Media $\bar{X}$, por lo que se procede a sumar la frecuencia en D8 =suma(D3: D7) , posteriormente copiarla en G8 . Ahora en D11 incluir la fórmula =G8/D8   para calcular $\bar{X}$. La tabla de distribución de frecuencias debe quedar de la siguiente forma:



	A	B	C	D	E	F	G	H	I	J	K
1		Límite	Límite	Frecuencia	Frecuencia	Marca de Clase		Límites Reales			
2	No. Intervalo	Inferior	Superior	Fi	Acumulada	Xi	XiFi	Inferior	Superior	Abs(Xi-$\bar{X}$)Fi	(Xi-$\bar{X}$)²Fi
3	1	11	56	5	5	33.5	167.5	10.5	56.5		
4	2	57	102	13	18	79.5	1033.5	56.5	102.5		
5	3	103	148	5	23	125.5	627.5	102.5	148.5		
6	4	149	194	3	26	171.5	514.5	148.5	194.5		
7	5	195	240	4	30	217.5	870	194.5	240.5		
8				30			3213				
9											
10											
11			Media $\bar{X}$	107.1							
12											

7. Para desviación media en J3 incluir la fórmula =abs(f3-d$11)*d3 A continuación copiar la fórmula en las siguientes celdas hasta completar la columna y en J8 incluir la suma.
8. Para calcular la columna K (varianza) se utilizará la fórmula =POTENCIA(f3-d$11;2)*d3

9. Enseguida copiar la fórmula en las siguientes celdas y hacer las sumas correspondientes para calcular la desviación media, varianza y desviación estándar.
10. Se solicita que en D12 se calcule la desviación media, en D13 la varianza y en D14 la desviación estándar. Al finalizar la tabla de distribución de frecuencias quedará así:



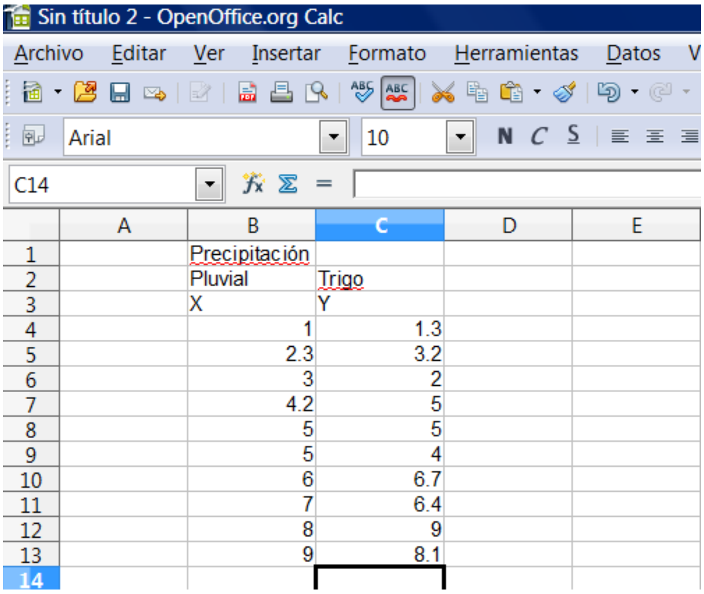
	B	C	D	E	F	G	H	I	J	K
1	Límite	Límite	Frecuencia	Frecuencia	Marca de Clase		Límites Reales			
2	Inferior	Superior	Fi	Acumulada	Xi	XiFi	Inferior	Superior	Abs(Xi-X̄)Fi	(Xi-X̄)²Fi
3	11	56	5	5	33.5	167.5	10.5	56.5	368	27084.8
4	57	102	13	18	79.5	1033.5	56.5	102.5	358.8	9902.88
5	103	148	5	23	125.5	627.5	102.5	148.5	92	1692.8
6	149	194	3	26	171.5	514.5	148.5	194.5	193.2	12442.08
7	195	240	4	30	217.5	870	194.5	240.5	441.6	48752.64
8			30			3213			1453.6	99875.2
9										
10										
11		Media X̄	107.1000							
12		D.M	48.4533							
13		Varianza	3443.9724							
14		D. Estándar	58.6854							
15										

Para practicar, puedes incluir los siguientes datos de botellas de tequila vendidas

No.I	Li	Ls	fi
1	11	56	6
2	57	102	12
3	103	148	5
4	149	194	3
5	195	240	4

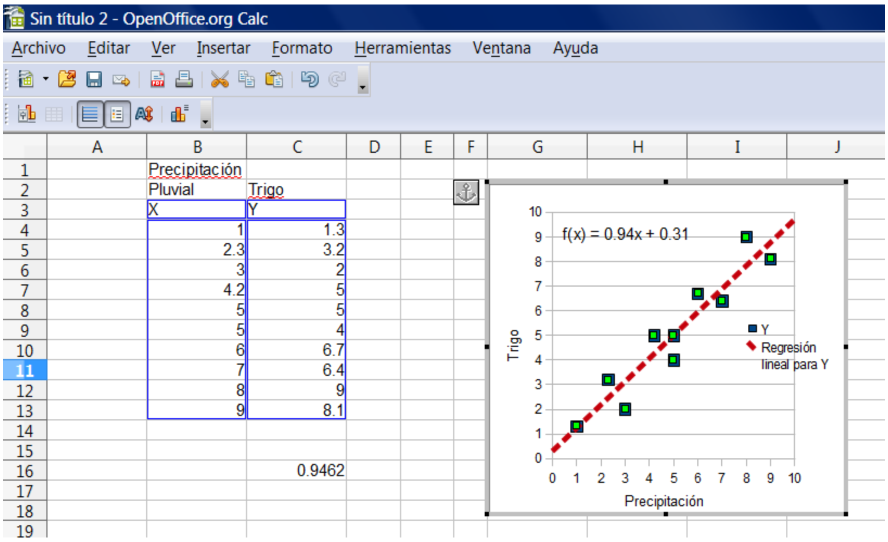
Correlación y Regresión Lineal de precipitación pluvial y rendimiento de trigo

1. Capturar los siguientes datos en las columnas respectivas



2. En C16 calcular el coeficiente de correlación con la siguiente fórmula: =COEF.DE.CORREL(B4:B13;C4:C13)

3. A continuación sombrear con el mouse las celdas de los datos B4:C13, ir al menú en insertar gráfico, seleccionar xy (dispersión). Una vez que se tiene el gráfico dar click en alguno de los puntos y éstos cambiarán de color. Posteriormente dar click derecho y seleccionar línea de tendencia, enseguida aparecerá una ventana con dos pestañas, en donde dice Tipo verificar que este marcado lineal y seleccionar Mostrar Ecuación. En la gráfica aparecerá la ecuación de regresión junto con la línea, como se muestra a continuación.

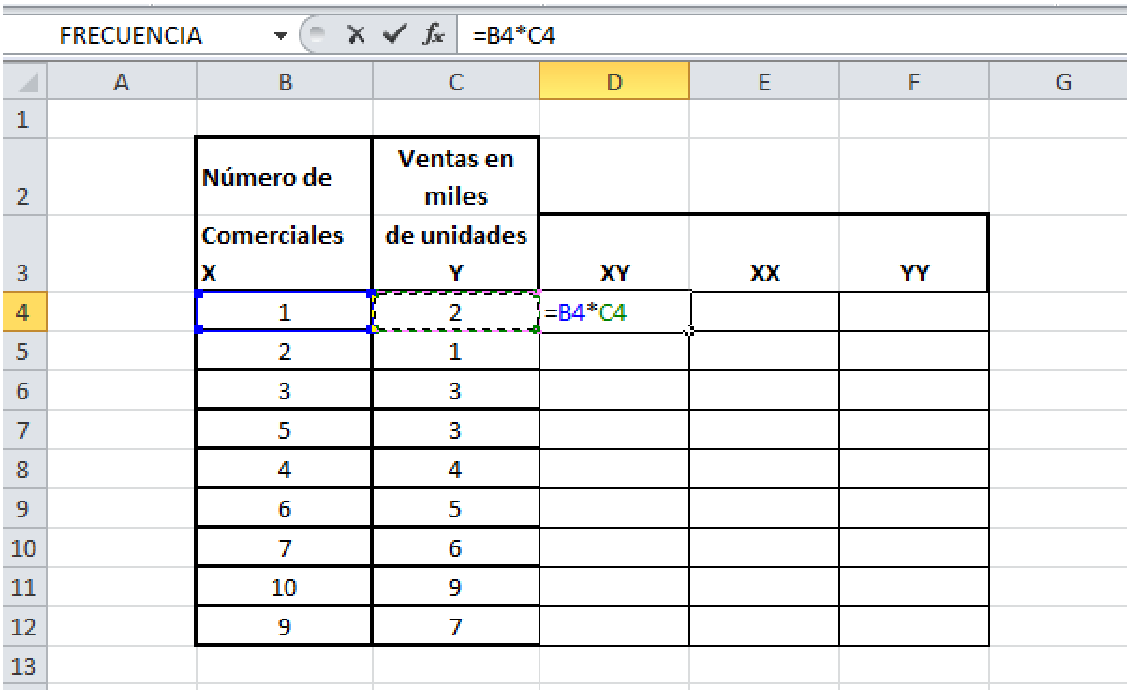


Para practicar puedes utilizar los datos de comerciales y ventas en cientos de un producto.

Número de Comerciales	Ventas en miles de unidades
1	2
2	1
3	3
5	3
4	4
6	5
7	6
10	9
9	7

La tabla en donde se calculan los valores de las sumas utilizadas para determinar el coeficiente de Pearson o correlación y la recta de regresión por mínimos cuadrados, se construye utilizando las fórmulas de =SUMA() ; = POTENCIA(número, potencia) y la multiplicación.

Iniciando: se ponen títulos y la primera fórmula para después arrastrarla a las demás celdas, como se muestra en las siguientes pantallas.



	A	B	C	D	E	F	G
1							
2		Número de	Ventas en miles				
3		Comerciales X	de unidades Y	XY	XX	YY	
4		1	2	=B4*C4			
5		2	1				
6		3	3				
7		5	3				
8		4	4				
9		6	5				
10		7	6				
11		10	9				
12		9	7				
13							

FRECUENCIA =B4*B4

	A	B	C	D	E	F
1						
2		**Número de**	**Ventas en miles**			
3		**Comerciales X**	**de unidades Y**	**XY**	**XX**	**YY**
4		1	2	2	=B4*B4	4
5		2	1	2		
6		3	3	9		
7		5	3	15		
8		4	4	16		
9		6	5	30		
10		7	6	42		
11		10	9	90		
12		9	7	63		
13						

Para la suma de las columnas

FRECUENCIA =SUMA(B4:B12)

	A	B	C	D	E	F
1						
2		**Número de**	**Ventas en miles**			
3		**Comerciales X**	**de unidades Y**	**XY**	**XX**	**YY**
4		1	2	2	1	4
5		2	1	2		
6		3	3	9		
7		5	3	15		
8		4	4	16		
9		6	5	30		
10		7	6	42		
11		10	9	90		
12		9	7	63		
13	Total	=SUMA(B4:B12)				
14		SUMA(**número1**, [número2], ...)				

FRECUENCIA =(D13-(B13*C13/9))/(RAIZ(E13-(POTENCIA(B13,2)/9))*RAIZ(F13-(POTENCIA(C13,2)/9)

	A	B	C	D	E	F	G	H	I
1									
2		**Número de**	**Ventas en miles**						
3		**Comerciales X**	**de unidades Y**	**XY**	**XX**	**YY**			
4		1	2	2	1	4			
5		2	1	2	4	1			
6		3	3	9	9	9			
7		5	3	15	25	9			
8		4	4	16	16	16			
9		6	5	30	36	25			
10		7	6	42	49	36			
11		10	9	90	100	81			
12		9	7	63	81	49			
13	Total	47	40	269	321	230			
14									
15									
16			=(D13-(B13*C13/9))/(RAIZ(E13-(POTENCIA(B13,2)/9))*RAIZ(F13-(POTENCIA(C13,2)/9)))						
17									

Determinando las proyecciones con la recta de regresión

FRECUENCIA =(0.7956*B4)+0.2897

	A	B	C	D	E	F	G	H	I
1									
2		**Número de**	**Ventas en miles**					Proyección con Recta de regresión	
3		**Comerciales X**	**de unidades Y**	**XY**	**XX**	**YY**		y = 0.7956x + 0.2897	
4		1	2	2	1	4		=(0.7956*B4)+0.2897	
5		2	1	2	4	1			
6		3	3	9	9	9			
7		5	3	15	25	9			
8		4	4	16	16	16			
9		6	5	30	36	25			
10		7	6	42	49	36			
11		10	9	90	100	81			
12		9	7	63	81	49			
13	Total	47	40	269	321	230			
14									
15									
16			0.956960209						

Distancia al cuadrado entre lo que da de venta la ecuación de regresión y los datos de la muestra

FRECUENCIA =POTENCIA(H4-C4,2)

	A	B	C	D	E	F	G	H	I
1									
2		**Número de**	**Ventas en miles**					Proyección con Recta de regresión	Distancia ^{2}
3		**Comerciales X**	**de unidades Y**	**XY**	**XX**	**YY**		y = 0.7956x + 0.2897	**Proyección vs muestra**
4		1	2	2	1	4		1.0853	=POTENCIA(H4-C4,2)
5		2	1	2	4	1		1.8809	
6		3	3	9	9	9		2.6765	
7		5	3	15	25	9		4.2677	
8		4	4	16	16	16		3.4721	
9		6	5	30	36	25		5.0633	
10		7	6	42	49	36		5.8589	
11		10	9	90	100	81		8.2457	
12		9	7	63	81	49		7.4501	
13	Total	47	40	269	321	230			
14									
15									
16			0.956960209						

FRECUENCIA =RAIZ(I13)

	A	B	C	D	E	F	G	H	I
1									
2		**Número de**	**Ventas en miles**					Proyección con Recta de regresión	Distancia ^{2}
3		**Comerciales X**	**de unidades Y**	**XY**	**XX**	**YY**		y = 0.7956x + 0.2897	**Proyección vs muestra**
4		1	2	2	1	4		1.0853	0.83667609
5		2	1	2	4	1		1.8809	0.77598481
6		3	3	9	9	9		2.6765	0.10465225
7		5	3	15	25	9		4.2677	1.60706329
8		4	4	16	16	16		3.4721	0.27867841
9		6	5	30	36	25		5.0633	0.00400689
10		7	6	42	49	36		5.8589	0.01990921
11		10	9	90	100	81		8.2457	0.56896849
12		9	7	63	81	49		7.4501	0.20259001
13	Total	47	40	269	321	230			4.39852945
14									=RAIZ(I13)

Cambiando la ecuación de regresión

FRECUENCIA =(0.8*B4)+0.2897

	A	B	C	D	E	F	G	H	I	J
1										
2		**Número de**	**Ventas en miles**					Proyección con Recta de regresión	Distancia ^{2}	Cambiando la "m" en la ecuación de la recta de regresión
3		**Comerciales X**	**de unidades Y**	**XY**	**XX**	**YY**		y = 0.7956x + 0.2897	**Proyección vs muestra**	y = 0.8x + 0.2897
4		1	2	2	1	4		1.0853	0.83667609	=(0.8*B4)+0.2897
5		2	1	2	4	1		1.8809	0.77598481	1.8897
6		3	3	9	9	9		2.6765	0.10465225	2.6897
7		5	3	15	25	9		4.2677	1.60706329	4.2897
8		4	4	16	16	16		3.4721	0.27867841	3.4897
9		6	5	30	36	25		5.0633	0.00400689	5.0897
10		7	6	42	49	36		5.8589	0.01990921	5.8897
11		10	9	90	100	81		8.2457	0.56896849	8.2897
12		9	7	63	81	49		7.4501	0.20259001	7.4897
13	Total	47	40	269	321	230			4.39852945	
14									2.097267138	
15										
16			0.956960209							

Calculando la distancia

	A	B	C	D	E	F	G	H	I	J	K
1											
2		**Número de**	**Ventas en miles**					Proyección con Recta de regresión	Distancia ^{2}	Cambiando la "m" en la ecuación de la recta de regresión	Distancia ^{2}
3		**Comerciales X**	**de unidades Y**	**XY**	**XX**	**YY**		y = 0.7956x + 0.2897	**Proyección vs muestra**	y = 0.8x + 0.2897	**Proyección vs muestra**
4		1	2	2	1	4		1.0853	0.83667609	1.0897	=POTENCIA(J4-C4,2)
5		2	1	2	4	1		1.8809	0.77598481	1.8897	0.79156609
6		3	3	9	9	9		2.6765	0.10465225	2.6897	0.09628609
7		5	3	15	25	9		4.2677	1.60706329	4.2897	1.66332609
8		4	4	16	16	16		3.4721	0.27867841	3.4897	0.26040609
9		6	5	30	36	25		5.0633	0.00400689	5.0897	0.00804609
10		7	6	42	49	36		5.8589	0.01990921	5.8897	0.01216609
11		10	9	90	100	81		8.2457	0.56896849	8.2897	0.50452609
12		9	7	63	81	49		7.4501	0.20259001	7.4897	0.23980609
13	Total	47	40	269	321	230			4.39852945		4.40477481
14									2.097267138		2.098755538
15											
16			0.956960209								
17											

Como se muestra la distancia al cuadrado cambiando por una recta cercana a la recta de regresión lineal es mayor a la calculada por mínimos cuadrados.

Ejercicios.

Instrucciones. Realicen lo que se indica:

1. Se seleccionan aleatoriamente seis estudiantes para evaluar la relación entre sus horas de estudio (X) y las calificaciones que obtienen (Y) en un curso de estadística. Se obtienen los siguientes datos:

Estudiante	x	y			
A	9	38			
B	11	67			
C	14	49			
D	4	12			
E	6	10			
F	6	24			

a. Tracen el diagrama de dispersión
b. Obtén la ecuación de regresión
c. Obtén el valor de y, si se sabe que x vale 15
d. Obtén el coeficiente de correlación r e interpreta su valor

2. Se han obtenido datos experimentales para el rendimiento de frijoles de soya y la cantidad de agua utilizada para irrigación en seis parcelas.

Cantidad usada de agua (X)	Rendimiento (Y)			
0	20			
2	28			
4	32			
1	25			
3	30			
5	31			

a. Obtén la ecuación de regresión, con la cantidad usada de agua como variable independiente.

b. Calcula el coeficiente de correlación de la muestra e interpreta su valor.

c. Si $x = 6$, ¿Cuál es el valor de y?

d. ¿Cuál es la interpretación del valor obtenido en el inciso anterior?

3. La siguiente información fue proporcionada por el Ing. Lara:

No. de albañiles (X)	20	50	15	30	45	70
No. de días en concluir la obra (Y)	15	10	17	12	11	7

a. Traza el diagrama de dispersión

b. Observando el diagrama de dispersión, ¿Crees que hay relación entre las variables? ¿Qué tipo de relación tendrán? ______________________________

c. Obtén la ecuación de regresión

d. Obtén el valor de y, si se sabe que $x = 25$

e. Obtén el coeficiente de correlación r e interpreta su valor

f. ¿Cuál es la interpretación del valor obtenido en el inciso anterior?

Unidad 3

Azar: modelación y toma de decisiones

Propósito. Al finalizar la unidad el alumno continuará el desarrollo de su pensamiento estadístico, a través del conocimiento y modelación de los fenómenos aleatorios, desde los tres enfoques de probabilidad, incluyendo la toma de decisiones.

- Reafirma las diferencias entre fenómenos deterministas y fenómenos aleatorios.
- Identifica a la probabilidad como la medida de la posibilidad de ocurrencia de un evento.
- Infiere la probabilidad de ocurrencia de algún resultado de un fenómeno aleatorio de manera subjetiva.
- Mide la probabilidad aproximada de ocurrencia de algún resultado de un fenómeno aleatorio a partir de una serie lo suficientemente grande de observaciones experimentales o de simulaciones físicas o por computadora.
- Valora la importancia de la conceptualización de la definición clásica de probabilidad.
- Distingue la relación entre la probabilidad frecuencial y clásica, a partir de analizar la aproximación a un valor, modelando fenómenos aleatorios de manera física o con la computadora.
- Define los conceptos de espacio muestral y evento.
- Construye espacios muestrales, eventos simples y eventos compuestos por medio de la disyunción, la conjunción o la negación.
- Explica el concepto de mutua exclusividad entre eventos.
- Calcula probabilidades de eventos.
- Construye tablas de contingencia para representar las relaciones entre dos eventos.
- Construye la expresión para el cálculo de la probabilidad condicional entre dos eventos a partir de la información contenida en una tabla de contingencia.
- Calcula probabilidades condicionales utilizando la expresión correspondiente.
- Reconoce el concepto de independencia.
- Calcula la probabilidad conjunta de eventos independientes.
- Concluye que la información obtenida a través del cálculo de probabilidades es importante en la toma de decisiones.

Práctica de probabilidad (Introducción)

Esta práctica tiene la intención de dar una idea e iniciarlos en el estudio de la probabilidad. Por favor vayan haciendo las anotaciones pertinentes, sigan los pasos en orden y respondan las preguntas, además investiguen en varias fuentes los conceptos que se indican.

Se solicita que cada uno de los alumnos tenga al menos una mandarina, una naranja y una manzana.

11. De las frutas, cuenten los gajos y número de semillas en cada uno, posteriormente registren sus resultados en una tabla como ésta:

	Número de gajos	Número de semillas
Lo que pensé que saldría		
Lo que salió		

12. Contando gajos vs semillas. Llena la tabla como se muestra en el ejemplo y después calcula la media de semillas por gajo.

No. gajos	1	2	3					n
No. semillas	8	12	3					

13. Ahora mide cada una de las semillas y registra los datos en una tabla como la siguiente. La idea es notar la dispersión del tamaño de las semillas. Para ello elabora la gráfica, calcula la media y desviación estándar.

Semilla	1	2	3	4					n
cms	.9	1.2	1.5	1					

14. Con base al paso anterior responde ¿En cuál de las frutas se presenta más variación en el tamaño de las semillas?
15. Elabora un diagrama de dispersión del número de gajos vs número de semillas y fundamenta si hay o no correlación entre ellas por cada una de las frutas.
16. Responde lo siguiente en base a lo que crees y lo que observaste: ¿Qué tan probable fue:
    a) que encontraras en la naranja más de diez semillas?
    b) que hubiera más de tres gajos?
    c) que la manzana no tuviera semillas?
    d) que en la mandarina hubiera por lo menos 12 semillas?
    e) que tuviera menos de cinco gajos?

17. Junten en una bolsa todas las semillas y revuelvan. ¿Cuántas semillas hay en total?
    ¿Cuántas de cada fruta? ¿Qué probabilidad de ser seleccionada tiene cada una?
18. Selecciona una semilla sin ver y registra el resultado (20 veces).

	Lo que pensé			Lo que salió			Adiviné
	Manzana	Mandarina	Naranja	Manzana	Mandarina	Naranja	
1	X			X			Si
2		X			X		No
20		X				X	No

19. Responde las siguientes preguntas:
    ¿Qué tan probable es
    a. que adivinaras todas las veces?
    b. que saliera más de 6 veces semillas de mandarina?
    c. que sólo seleccionaras una semilla de naranja en las veinte veces?
    d. que no se seleccionara una semilla de manzana?

    ¿Influyó en algo que regresaras o no la semilla seleccionada a la bolsa?

20. Tarea
    Investiga los conceptos: azar, probabilidad, sus enfoques, eventos (tipos), espacio muestral.
    Después indica cuáles y en qué parte de la practica anterior se utilizaron.

Fenómenos deterministas y aleatorios

Existen fenómenos, en donde existe con certeza lo que va a ocurrir al efectuarse o llevarlos a cabo, por ejemplo: caída libre, la duración del movimiento de rotación de la Tierra, etc. A este tipo de fenómenos se les llama deterministas.

Pero la incertidumbre se encuentra en una gran variedad de fenómenos, unos ocurren en el trabajo, casa, escuela, e inclusive en las empresas y en el gobierno a los cuales les llamamos fenómenos aleatorios. Por esta razón, en la actualidad la probabilidad está teniendo un crecimiento notable en sus aplicaciones dentro de los diversos campos de la actividad humana, ya que ésta es un valor que le adjudicamos a los fenómenos y que nos indica que tan posible es que ocurra.

Otros grandes campos, en donde se presentan los fenómenos aleatorios es en el clima y en los juegos, por ejemplo: lanzamiento de dados, monedas, urnas, ruletas, loterías, etc.

A continuación, se enlista una serie de fenómenos, escriban en la línea, si corresponde a un fenómeno determinista o aleatorio.

a) Valor del peso en la bolsa mexicana de valores para el día de mañana _______________
b) El día exacto en el que se registrará el siguiente sismo en la Cd. de México ______________
c) Tiempo en el que tardará en caer un gis con un peso de 3 gramos a una altura de 3 metros __________
d) El lado de encima al dejar caer una moneda en donde ambos lados son águilas ______________

Piensen en dos fenómenos deterministas, en dos aleatorios y escríbanlos a continuación:

Fenómeno determinista es ____

____

Fenómeno determinista es ____

____

Fenómeno aleatorio es ____

____

Fenómeno aleatorio es ____

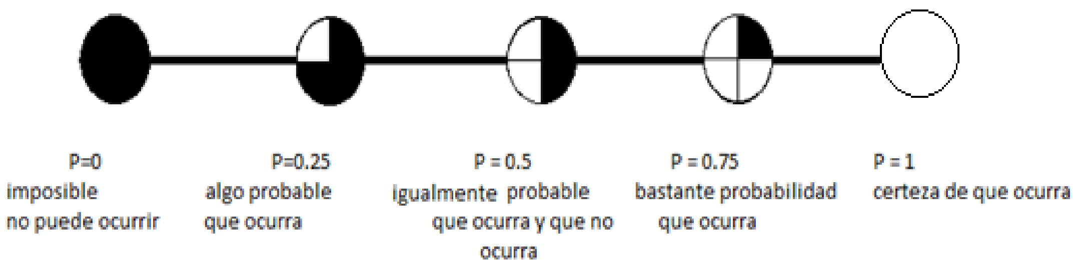
____

A los fenómenos deterministas se les adjudica una probabilidad de ocurrencia de 1, y también se les denomina eventos seguros, ya que, tenemos la seguridad de lo que ocurrirá al efectuarse. Por ejemplo, sabemos que desde hace millones de años, después de que pasa la noche, amanece; entonces, seguro que mañana después de que haya pasado la noche, amanecerá; o también, al lanzar un dado común y corriente, es seguro que, al lanzarlo, obtengamos un número menor a 10. ¿Qué otro ejemplo se les ocurre? ____

Por el contrario, cuando el evento no ocurrirá o es imposible que suceda, le adjudicamos la probabilidad de ocurrencia de 0, y se les denomina eventos imposibles. Por ejemplo, la probabilidad de que gané la lotería sin tener ningún número; o también, esperar a obtener tres águilas al lanzar dos monedas. ¿Qué otro ejemplo se les ocurre? ____________________________________

____

Pero ¿a todos los eventos se les adjudica la probabilidad de 0 o 1? ¡Claro que no!, ¡imagínense la situación con respecto a su calificación!, ¡pensar que es imposible que aprueben!, o, por el contrario, ¡pensar que es seguro que aprueben!, es decir, que la probabilidad de la situación sea 0 o 1, ¡qué padre! si tienen la seguridad de aprobar, o, ¡qué desgracia! si es imposible aprobar. Por lo tanto, debemos pensar que la probabilidad de los eventos debe fluctuar entre 0 y 1, como lo indica el siguiente esquema:



Sabemos que, la probabilidad de que llueva el día de hoy o que no llueva el día de hoy es 1, ya que, estamos seguros de que una de estas dos situaciones tiene que ocurrir. Por lo tanto, si se ha informado que la probabilidad de que llueva es 0.6, de que no llueva debe ser 0.4, para que la probabilidad de que suceda cualquiera de estas dos situaciones sea 1, ya que estamos seguros que una de esas dos situaciones tiene que ocurrir.

De la misma forma, si consideramos el evento aprobar el curso y le adjudicamos la probabilidad de 0.8, la probabilidad del evento no aprobar será de _______, ya que estamos seguros de que una de estas dos ocurrirá.

Consideren, ahora los siguientes enunciados:

a) Ya que no hay nubes. La probabilidad de que hoy llueva más tarde es de -0.90
b) La probabilidad de que una muestra mineral contenga cobre es de 0.28 y la probabilidad de que no contenga cobre es de 0.25
c) La probabilidad de que un abogado gane un caso es de 0.30 y la probabilidad de que lo pierda es cinco veces más alta.

¿Cuáles no manejan un valor apropiado para la probabilidad y por qué?
____
____
____

La probabilidad de que cierto equipo de futbol gane su próximo partido es 0.2 y la podemos representar por $P(G) = 0.2$, de que empate es de 0.5 y la podemos representar por $P(E) = 0.5$. ¿Cuál es la probabilidad de que pierda?

____

Un amigo me dijo que era imposible, que su novia esté embarazada, de hecho, dice que la probabilidad de que lo esté es de 0.1. De acuerdo con lo que el amigo dice, ¿es apropiado el valor que se le asigna a la probabilidad de que esté embarazada? ________ ¿Por qué? ______________________________

____

Es igualmente probable que apruebe o repruebe el curso, de hecho, cada una de ellas tiene una probabilidad de que suceda de 0.6. ¿Es válido lo anterior? __ ¿por qué? ____

La probabilidad de que salga en tres años es casi imposible, de hecho, la probabilidad de que eso suceda es 0.5. ¿Es correcto el valor que se maneja? ___ ¿por qué? ____

El enfoque clásico de probabilidad.

Laplace en 1812 dio la definición que se conoce como clásica de probabilidad**:** “La teoría del azar consiste en reducir todos los acontecimientos del mismo tipo a un cierto número de casos favorables al acontecimiento cuya probabilidad se busca. La proporción entre este número y el de todos los casos posibles es la medida de esta probabilidad, que no es, pues, más que una fracción cuyo numerador es el número de casos favorables y cuyo denominador el de todos los posibles.”

Con la definición de Laplace, la probabilidad de un evento es una medida de la posibilidad que el evento pueda ocurrir. La probabilidad es una medida en una escala de 0 a 1. Entonces, de acuerdo con Laplace, para hallar la probabilidad de que un evento pueda ser “favorable”, dividimos el número de maneras que puede ocurrir favorablemente entre el total de maneras que puede ocurrir el evento, es decir

$$P(evento) = \frac{número\ de\ maneras\ que\ puede\ ocurrir\ favorablemente}{total\ de\ maneras\ que\ puede\ ocurrir\ el\ evento}$$

Si contamos los favorables a cierto evento, ¿desde qué valor empezamos a contar? __________. Entonces, si no hay elementos favorables a un evento o suceso, su probabilidad será de _________. Y le llamamos evento ________. Si contamos los favorables a cierto evento, ¿podrá ser que haya más favorables de los que pueden ocurrir? _______ En general, el número máximo de elementos favorables a cierto evento o suceso será igual al número de elementos de _______. Por lo tanto, la probabilidad de este evento será igual a _____. A este evento le llamamos evento ________.

El enfoque de probabilidad como frecuencia relativa.

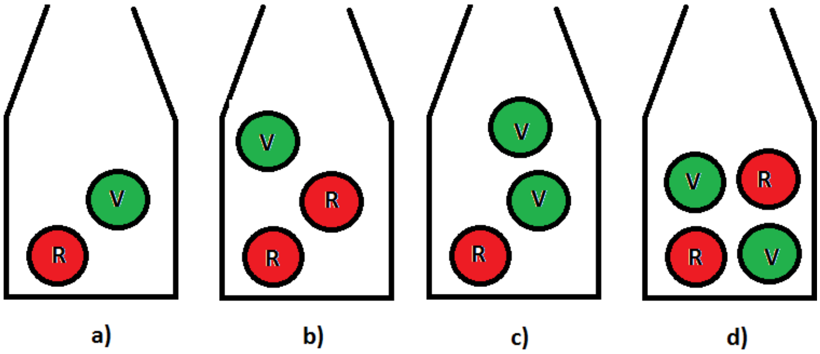
La probabilidad como frecuencia relativa se calcula a partir de las frecuencias relativas observadas de los resultados de los experimentos o pruebas repetidas. Anteriormente, ya habías estudiado este concepto, ¿Cómo obtenías estas frecuencias? ____

Si lanzamos una moneda 100 veces, ¿Cuántos soles esperarías? ______

Si efectuamos el experimento de lanzar un dado común y corriente 300 veces, ¿Cuántos cuatros esperaríamos? ______________

Otros experimentos aleatorios de los cuales podamos generar resultados aleatorios podrán ser los siguientes: observar el género de 200 alumnos, el semestre escolar de 150 alumnos, los aciertos y fallos de tiros libres de un jugador de basquetbol, ¿Qué otro se te ocurre? ________________________

Observa la siguiente figura, que representan 4 bolsas que contienen bolas de color rojo (R) y verde (V):



Contesta V (verdadero) o F (falso) a las siguientes preguntas:

a) Es más fácil obtener R en el inciso (a) que en el inciso (b) ______

b) Es más fácil obtener R en el inciso (b) que en el inciso (d) ______

c) Es más fácil obtener R en el inciso (a) que en el inciso (d) ______

d) Es más fácil obtener R en el inciso (a) que en el inciso (c) ______

e) Es más fácil obtener R en el inciso (b) que en el inciso (c) ______

Andrés tomó una de las bolsas para hacer extracciones de bolas y obtuvo el siguiente resultado: **R R V R R R V R**

¿Con qué bolsa piensas que estaba jugando? _______ ¿Por qué? __________

____

El enfoque de probabilidad Subjetiva.

La probabilidad subjetiva es un grado de creencia o percepción personal, este concepto no se basa en la repetitividad del evento, sino más bien en la experiencia personal de quien adjudica la probabilidad, de esta manera, la probabilidad representa un juicio personal de un fenómeno impredecible. Por ejemplo, ¿quién crees que va a ganar el Super Bowl?, al jugar el Melate, sabemos que la probabilidad de que cada uno de los seis números sea seleccionado es la misma, pero nosotros creemos que los que van a ser seleccionados como el conjunto ganador son los míos. ¿Qué otro ejemplo puedes proporcionar? ____

____

Observa la siguiente ilustración para tener una idea del contexto y lee lo que tres de los personajes de la ilustración piensan.

Contesta las siguientes preguntas de acuerdo a la ilustración y las actividades realizadas en las sesiones 1 y 2

1. ¿Cuál es la diferencia entre un fenómeno determinista y aleatorio? ____

2. ¿Cuál es la escala de probabilidad? ______________________________

3. Dibuja la escala de probabilidad

4. ¿A qué enfoque de probabilidad recurre Homero, el clásico, frecuencial o subjetivo? _________________________ ¿Por qué? _________________ ____

5. Kent piensa que perderá, ya que Homero ha perdido 8 de los 10 juegos que ha jugado, ¿qué enfoque de probabilidad ha utilizado para pensar o creer eso? ____

6. Carlson piensa que casi es imposible que salga un as, ¿qué enfoque de probabilidad ha utilizado para pensar eso? ________________________

7. Carlson, pensó algo utilizando la frase "casi imposible", ¿qué hubiera hecho que considerara que fuera "imposible" y no "casi imposible"? ____ ____

8. Si sale el as ganará Homero, ¿crees que sea imposible que gané Homero? ______ ¿Por qué? _____________________________________ ____ ____

9. Kent piensa que Homero perderá, ¿qué probabilidad asigna a este evento? ________________________

10. Kent afirma "perderá Homero", es decir está seguro de que así será, ¿qué probabilidad se les asigna a los eventos seguros? __________

Espacio muestral

Para el enfoque clásico y para el enfoque de probabilidad como frecuencia relativa, es necesario contabilizar el total de posibilidades o el número total de experimentos observados o realizados. Al **conjunto total de posibilidades** le llamamos **Espacio Muestral** y la denotaremos por $\boldsymbol{S}$.

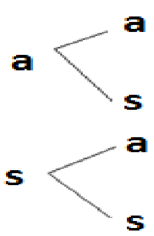
Por ejemplo, al lanzar una moneda al aire, ¿qué resultados podemos obtener? ____________. Y entonces el espacio muestral podrá escribirse de la siguiente manera: $S = \{a, s\}$ Y al lanzar dos monedas al aire, ¿qué resultados podemos obtener? ______________ .

Para representar los posibles resultados, podemos utilizar la siguiente tabla de doble entrada, complétala:

	a	s
a	aa	
s		

Entonces $S = \{ \qquad \qquad \qquad \qquad \}$

O también los podemos representar por medio de un diagrama de árbol, observen la siguiente figura:



Si lanzamos 3 monedas al aire, ¿cuáles son los posibles resultados del experimento? ____

Utiliza el siguiente espacio y lo que creas conveniente para obtenerlos.

Entonces $S = \{ \qquad\qquad\qquad\qquad \}$

Y si lanzamos 4 monedas, ¿cuántos son los posibles resultados del experimento? ________

Si lanzamos al aire una moneda y un dado común y corriente, ¿qué resultados podemos obtener? ___________. Si lo crees conveniente utiliza el siguiente espacio para crear tu diagrama de árbol y puedas obtener los resultados posibles del experimento aleatorio:

Entonces $S = \{ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \}$

Como el espacio muestral es un conjunto podemos formar subconjuntos suyos, a los que le llamamos **eventos**. Por ejemplo, al lanzar un dado, obtenemos el siguiente espacio muestral:

$S = \{ 1, 2, 3, 4, 5, 6 \}$

Podemos formar eventos, por ejemplo:

$$A = \{obtener\ par\} = \{2, 4, 6\}$$

Escribe los elementos de los siguientes subconjuntos de S:

$$B = \{obtener\ número\ impar\} = \{ \qquad\qquad \}$$

$$C = \{número\ primo\} = \{2, 3, 5\}$$

$$D = \{múltiplo\ de\ 3\} = \{ \qquad \}$$

$$E = \{número\ mayor\ a\ 5\}$$

$$F = \{número\ menor\ o\ igual\ a\ 5\}$$

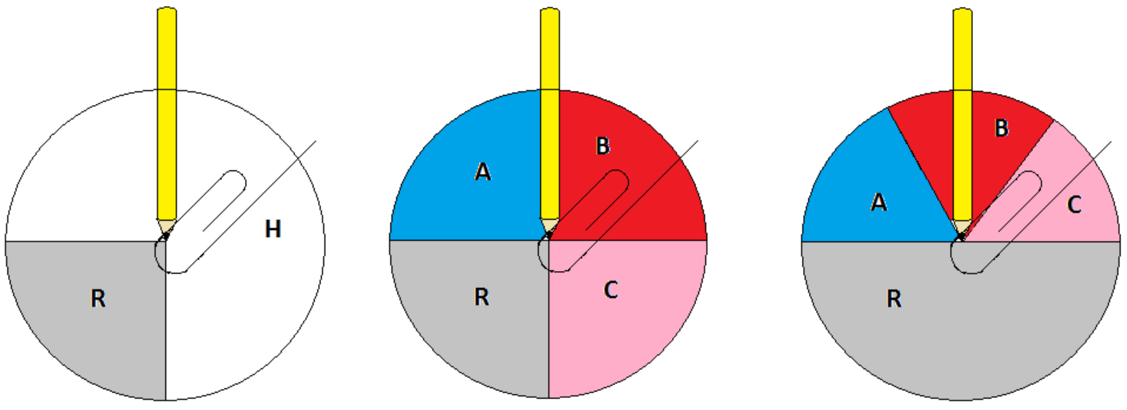
$$G = \{ \qquad\qquad\qquad\qquad\qquad \} = \{5, 6\}$$

$$H = \{múltiplos\ de\ 7\} = \underline{\qquad\qquad\qquad}$$

La probabilidad de A es $P(A) = \frac{3}{6} = 0.5$

¿Cuál es la probabilidad de cada uno de los demás eventos?

Ruletas. Considera las ruletas:



RULETA DE ANA **RULETA DE BETO** **RULETA DE CARMEN**

a) Con respecto a la ruleta de Ana:

Si diéramos un empujón al clip y observamos en que zona apunta, ¿Cuál es la probabilidad de que apunte en la zona R? _____ ¿Cuál es la probabilidad de que no apunte en la zona R? _______ ¿Cuál es la probabilidad de que apunte en la zona H? _____ ¿Cuál es el espacio muestral? __________________

Obtén las siguientes probabilidades

$$P(R) = \_\_\_\_$$

$$P(H) = \_\_\_\_$$

$$P(R) + P(H) = \_\_\_\_$$

----

b) Con respecto a la ruleta de Beto:

Si diéramos un empujón al clip y observamos en que zona apunta, ¿cuál es la probabilidad de que apunte en la zona R? _____ ¿Cuál es la probabilidad de que no apunte en la zona R? _______ ¿Cuál es la probabilidad de que apunte en la zona H? _____ ¿Cuál es la probabilidad de que apunte en la zona B? ____ ¿Cuál es la probabilidad de que apunte en la zona A o la zona R? ______ ¿Cuál es el espacio muestral? ______________

Obtén las siguientes probabilidades

$$P(R) = \_\_\_\_$$

$$P(H) = \_\_\_\_$$

$$P(A) + P(B) + P(C) = \_\_\_\_$$

<u>c) Con respecto a la ruleta de Carmen:</u>

Si diéramos un empujón al clip y observamos en que zona apunta, ¿cuál es la probabilidad de que apunte en la zona R? _____ ¿Cuál es la probabilidad de que no apunte en la zona R? _______ ¿Cuál es la probabilidad de que apunte en la zona C? _____ ¿Cuál es la probabilidad de que apunte en la zona B? ____ ¿Cuál es la probabilidad de que apunte en la zona A o la zona R? ______ ¿Cuál es el espacio muestral? ______________

Obtén las siguientes probabilidades

$$P(R) = \_\_\_\_$$

$$P(H) = \_\_\_\_$$

$$P(A) + P(B) + P(C) = \_\_\_\_$$

$$P(R) + P(A) + P(B) + P(C) = \_\_\_\_$$

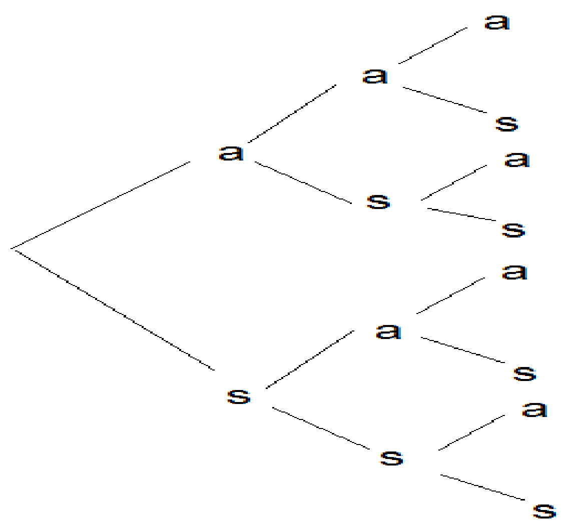
<u>d) Con respecto a la ruleta de Ana, Beto y Carmen:</u>

¿En qué ruleta hay más probabilidad de que el clip apunte en la zona R? ______

¿En qué ruletas hay idéntica probabilidad de que el clip apunte en la zona R? _______

Retomando el lanzamiento de 3 monedas, el conjunto que conformaba el espacio muestral es $S = \{aaa, aas, asa, ass, saa, sas, ssa, sss\}$

El diagrama de árbol muestra sus elementos:



Entonces si lanzamos 3 monedas al aire, la probabilidad de obtener por lo menos 1 sol es $7/8$ y podemos representar al evento y su respectiva probabilidad como:

$$A = \{obtener\ por\ lo\ menos\ 1\ sol\}$$

y está constituido por los elementos

$$A = \{aas, asa, ass, saa, sas, ssa, sss\}$$

Y la probabilidad del evento A la podemos representar por $P(A) = \frac{7}{8}$ y se lee "probabilidad de A igual a siete octavos"

Aprovechemos el ejemplo para obtener las siguientes probabilidades:

Consideremos los siguientes eventos:

$$B = \{obtener\ tres\ resultados\ iguales\} = \{\qquad\qquad\qquad\}$$

$$P(B) = \underline{\qquad}$$

$$C = \{obtener\ exactamente\ 1\ águila\} = \{\qquad\qquad\qquad\}$$

$$P(C) = \underline{\qquad}$$

$$D = \{obtener\ a\ lo\ más\ 2\ soles\} = \{\qquad\qquad\qquad\}$$

$$P(D) = \underline{\qquad}$$

$$E = \{\qquad\qquad\qquad\qquad\} = \{aaa, aas, asa, ass\}$$

$$P(E) = \underline{\qquad}$$

$$F = \{\qquad\qquad\qquad\qquad\} = \{\,\}$$

$P(F) = 0$; F es un evento imposible

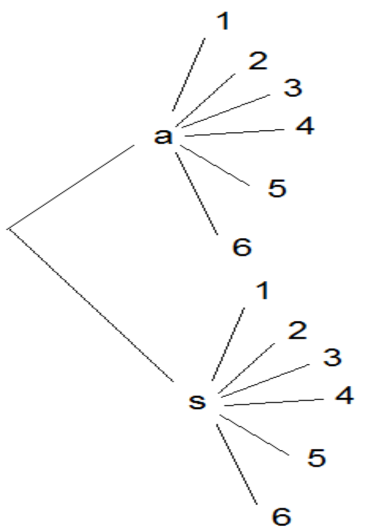
$$G = \{obtener\ un\ águila\ o\ un\ sol\ por\ lo\ menos\} = \{\qquad\qquad\qquad\}$$

$P(G) = 1$; G es un evento seguro

También analizamos los posibles resultados al lanzar un dado común y corriente y una moneda, obtuvimos que

$$S = \{a1, a2, a3, a4, a5, a6, s1, s2, s3, s4, s5, s6\}$$

El diagrama de árbol muestra sus elementos:



Consideremos los siguientes eventos:

Recuerda que los eventos son subconjuntos del espacio muestral,

$$A = \{obtener\ un\ número\ par\} = \{ \qquad\qquad\qquad\qquad\qquad\qquad \}$$

$$P(A) = 6/12$$

$$B = \{obtener\ \ águila\} = \{ \qquad\qquad\qquad\qquad \}$$

$$P(B) = \underline{\qquad}$$

$$C = \{obtener\ 2\ soles\} = \{\ \}$$

$P(C) = 0;\ C$ es un evento imposible

$$D = \{obtener\ un\ número\ menor\ a\ 10\} = \{ \qquad\qquad\qquad\qquad\qquad\qquad \}$$

$P(D) = 1;\ D$ es un evento seguro

Los paraguas. Una señora ha invitado a tres amigas a tomar el té. Como llovía bastante, cada una de las tres invitadas trajo un paraguas que colocaron en la entrada de la casa. Terminada la merienda, la dueña de la casa ha dado un paraguas, al azar, a cada una de sus amigas. De los siguientes eventos, ¿Cuál crees que sea el más fácil que ocurra?:

- Cada una de las señoras recibió el paraguas correcto.
- Todas recibieron un paraguas cambiado.
- Sólo una de las visitantes recibió su propio paraguas.

Realiza el diagrama de árbol para darte una idea de lo anterior:

¿Cuál es la probabilidad de que cada una de las señoras reciba el paraguas correcto? ______

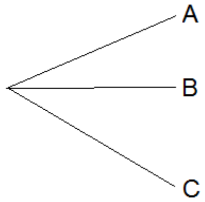
¿Cuál es la probabilidad de que todas reciban un paraguas que no sea el suyo? _______

¿Cuál es la probabilidad de que sólo una de las visitantes reciba su propio paraguas? ________

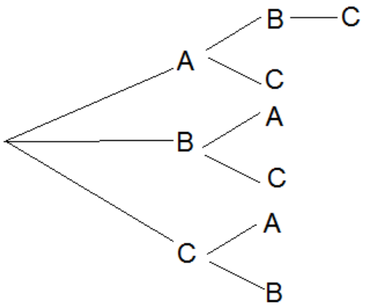
Escribe todas las ordenaciones distintas en que las letras A, B, C han resultado: ____

¿Crees que existan otras? __________________

Para estar seguros vamos a utilizar un diagrama en árbol. En primer lugar, escribimos la primera letra posible en cada ordenación:



Después, escribimos la segunda letra posible para cada una de la primera elegida:

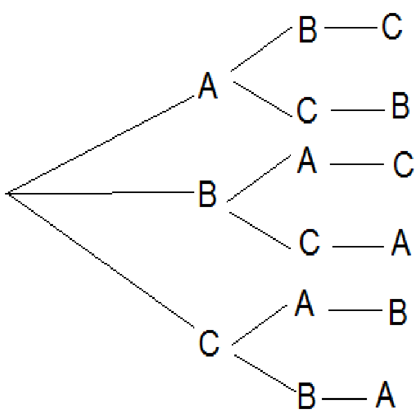


La tercera letra ya no puede elegirse, por lo que obtenemos todas las posibles ordenaciones de las tres letras.

Completa el árbol.

¿Cuántas ordenaciones diferentes han resultado? _______

El experimento aleatorio descrito puede considerarse como de tres extracciones sucesivas de los paraguas, que se hacen sin reemplazamiento. Los resultados posibles se pueden esquematizar en el siguiente diagrama de árbol:

1ra. Extracción **2da. Extracción** **3ra. Extracción**	**Orden final**	**No. de paraguas correctos**	**Probabilidad**
	ABC	3	1/6
	ACB	1	1/6
	BAC	1	1/6
	BCA	0	1/6
	CAB	0	1/6
	CBA	1	1/6

En la primera extracción tenemos 3 resultados. Para cada uno de ellos, en la segunda tenemos 2. Para cada uno, en la tercera tenemos 1. Luego el número de ordenaciones distintas de 3 objetos es: $P_3 = (3)(2)(1) = 6$.
Puesto que hay 6 resultados posibles podemos esperar que cada resultado ocurra 1 de cada 6 veces cuando el experimento se repite muchas veces. Este hecho se expresa con la notación: $P(ABC) = \frac{1}{6}$ y se lee “probabilidad de obtener ABC es igual a $1/6$”.

Supongamos que el domingo siguiente la señora invita a cuatro amigas A, B, C y D a tomar el té. Escribe todas las formas posibles de entregar al azar los abrigos a las invitadas. Para ello realiza un diagrama de árbol:

a) ¿De cuántas formas diferentes pueden entregarse al azar 4 abrigos a 4 señoras? __________

b) ¿De cuántas formas distintas pueden ordenarse las 4 letras A, B, C y D? _______

c) ¿Cuál es la probabilidad de que la señora A reciba su abrigo? _______

d) ¿Cuál es la probabilidad de que todas reciban su abrigo? ________

Ejercicios.

Retomando las ideas anteriores, contesten lo siguiente:

a) En un examen con preguntas de opciones múltiples sobre historia se pide que el alumno ordene por fechas tres acontecimientos históricos. Un alumno que no ha estudiado este tema decide responder al azar para tratar de acertar por casualidad. ¿Cuál es la probabilidad de que ordene todos los acontecimientos correctamente? _______

¿Cuál es la probabilidad de acertar una sola fecha por casualidad? _____

b) ¿De cuántas formas se pueden colocar en una estantería 3 libros? ____ ¿Y cuatro? ____

c) ¿Cuántas palabras distintas, aunque carezcan de significado, se pueden formar con todas las letras de la palabra AMOR sin que ninguna letra se repita? __________

d) Sin repetir dígitos ¿cuántos números se pueden formar con los dígitos 2, 3, 4, 5? _______

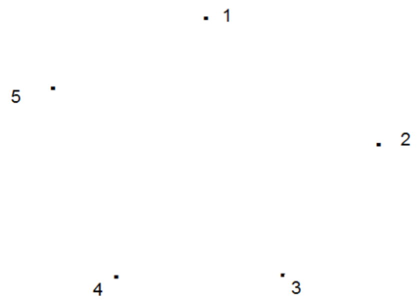
e) Una persona olvidó los 3 últimos dígitos de un número telefónico de 8 dígitos, esta persona recuerda que ninguno de esos 8 dígitos se repite. De hecho, el recuerda que el número comienza con 53789_ _ _. Marcó todos los números que podían ser, atinándole hasta el último posible. ¿Cuántos números tuvo que marcar? ______ Puedes ayudarte con un diagrama de árbol.

f) Marcos y Enrique intervienen en un torneo de tenis. La primera persona que gane 2 juegos seguidos o que complete 3 gana el torneo. Con ayuda de un diagrama de árbol, obtengan cuales son los posibles resultados ____________

____

¿Cuántas rectas se pueden trazar?

Como sabes por dos puntos pasa una sola recta. ¿Podrías indicar cuántas rectas pasan 5 puntos entre los cuales no hay 3 alineados? En el esquema adjunto hemos numerado los puntos de 1 a 5, dibuja todas las rectas que puedas:



¿Cuántas has dibujado? ________

En el siguiente cuadro hemos anotado todas las rectas posibles:

2do punto

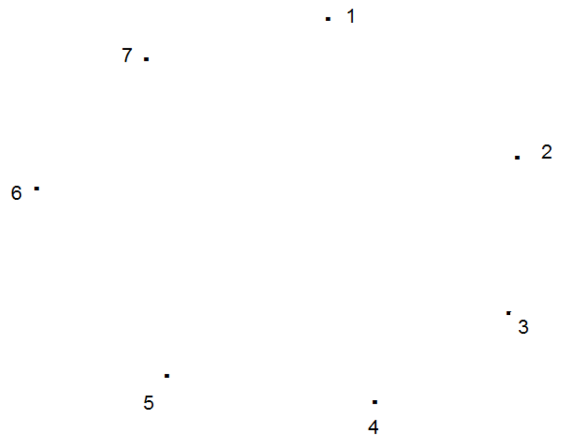
1er punto

	2	3	4	5
1	+	+	+	+
2	0	+	+	+
3	0	0	+	+
4	0	0	0	+
5	0	0	0	0

Fíjate en las casillas. Si es posible trazar desde el primer punto al segundo una recta que no se haya dibujado anteriormente se anotó una cruz; en caso contrario se anotó un 0.

Como ves el número de cruces dibujadas toma forma de triángulo. ¿Cuántas cruces aparecen dibujadas? _____ ¿Cuántos 0? _____

Añadimos 2 puntos más a los 5 que teníamos. Dibuja todas las rectas que puedas.



Construye una tabla como la anterior para determinar el número de rectas que se pueden trazar uniendo los puntos.

Responde las siguientes preguntas:

¿Cuántas casillas tiene la tabla? _____ ¿Cuántas cruces aparecen? ______

¿Cuántos ceros? ______ ¿Cuántas rectas aparecen por 7 puntos, de los cuales no hay 3 alineados? _____ ¿Y por 20? ______

Ejercicios

Retomando las ideas anteriores, resuelvan los siguientes ejercicios:

a) Ana tiene que realizar un examen sobre 10 temas, pero sólo ha estudiado 8. El examen consta de 2 temas.

- ¿Qué probabilidad de aprobar tiene si ha de contestar bien a los 2? ____
- ¿Y si basta con responder 1? ______

b) En una reunión de 5 presidentes, cada uno de ellos saludo de mano a los demás, ¿cuántos saludos de mano se realizaron? _______

Probabilidad de eventos compuestos

Antes de comenzar, debemos aclarar que en matemáticas y en particular en probabilidad trabajamos con la lógica de la O incluyente, es decir que al referirnos a dos características o más, puede ocurrir que se satisfagan 1, 2 o más. Por ejemplo, esto sucede cuando decimos, "debe ser estudiante o trabajador", y debemos entender que puede reunir 1 o inclusive las dos características.

Debe de quedar claro lo anterior, debido a que habitualmente, se usa, la lógica del O exclusivo, es decir, es uno o el otro, pero no pueden ser los dos. Por ejemplo, cuando decimos es de primero o de segundo, queda claro que es imposible que puedan ser las dos.

Podemos decir que la palabra "o" la podemos simbolizar por ____, que en la lógica de conjuntos representa la unión. La "y" la podemos simbolizar por ____, que en la lógica de conjuntos representa la intersección. Y, la palabra "no", la podemos simbolizar por ____, que en la lógica de conjuntos representa el complemento.

La probabilidad de la unión de dos eventos puede obtenerse por medio de la siguiente expresión:

$$P(R \cup I) = \quad P(R) + P(I) - P(R \cap I)$$

$$= \quad \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

¿Cuál crees que sea la razón por la que se resta la $P(R \cap I)$? ______________

____

Con la ayuda de tu profesor, propongan eventos para ejemplificar la razón.

Ejercicio. En una evaluación estudiantil del personal docente, V es el evento que un profesor es muy capaz en su área, D es el evento de que aplica pruebas difíciles y R es el evento de que califica en forma estricta, enuncia con palabras las probabilidades que se expresan como:

a) $P(\bar{D})$

____

b) $P(D \cup R)$

____

c) $P(\bar{V} \cap \bar{R})$

____

d) $P(V \cap R)$

____

¿Cómo representarías los siguientes enunciados?

a) No califique en forma estricta.

____

b) No aplique pruebas difíciles, pero califique en forma estricta.

____

c) No sea muy capaz en su área y/o no aplique pruebas difíciles.

____

Ejercicio

Un profesor realizó una encuesta a todos sus alumnos del CCH Naucalpan para obtener la siguiente información referente al tipo de persona que dicen ser (tranquilo, estudioso o inteligente):

- 20 dicen ser tranquilos, pero no son ni estudiosos ni inteligentes.
- 15 dicen ser tranquilos y estudiosos, pero no inteligentes.
- 5 dicen ser tranquilos, estudiosos e inteligentes.
- 20 dicen ser estudiosos e inteligentes, pero no tranquilos.
- 10 dicen no tener ninguna de esas características.
- 55 dicen tener la característica de ser estudioso (no se especifica que tengan esta característica únicamente).

- 38 dicen tener la característica de ser inteligentes (no se especifica que tengan esta característica únicamente).
- 15 dicen ser tranquilos e inteligentes (no se especifica que tengan esas dos características nada más).

Vamos a representar a los siguientes eventos como:

$$I \ = \ \{alumno\ inteligente\}$$

$$T \ = \ \{alumno\ tranquilo\}$$

$$E \ = \ \{alumno\ estudioso\}$$

Para indicar el evento "el alumno es inteligente y estudioso" escribimos $I \cap E$, usando el símbolo de la intersección de conjuntos. Entonces, ¿cómo indicarías que el alumno tiene las tres características? __________ ¿Sabrías indicar que significan los eventos siguientes?

a) $I \cap \bar{E}$ ____

b) $\bar{T} \cap E$ ____

c) $\bar{E} \cap \bar{I}$ ____

Para simbolizar el evento "el alumno al menos una de las tres características la tiene" escribimos $I \cup T \cup E$, usando el símbolo de la unión de conjuntos.

¿Cómo simbolizas que "el alumno al menos tiene una de estas dos características, Inteligente o Estudioso? _________

¿Sabrías simbolizar que significan los eventos siguientes?

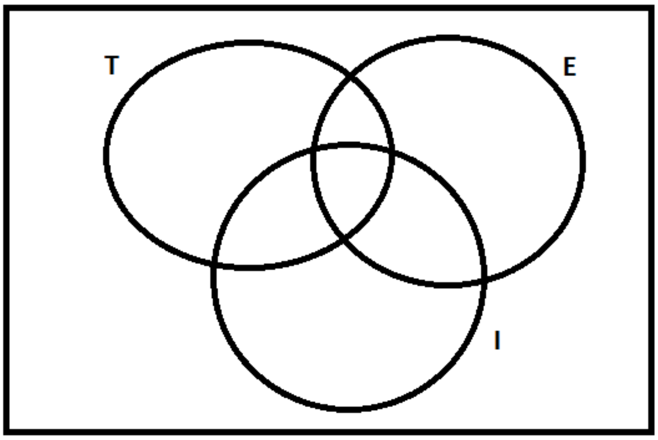
a) $I \cup \bar{E}$ ____

b) $\bar{T} \cup E$ ____

¿Cómo podríamos indicar el evento "no ocurre que el alumno al menos tenga una característica"? ___________

Retomemos la información que se proporciona, si el profesor consideró a todos sus alumnos, ¿cuántos tiene? ________

Para responder a la pregunta anterior, debemos pensar que algunos alumnos fueron considerados dos o inclusive tres veces, lo cual podría dificultar dar una respuesta. Para ello podemos recurrir a un diagrama de Venn, dicho esquema servirá para representar la información:



Observa en el esquema que un círculo representa al evento T, otro al evento I y otro al evento E. ¿Qué representa que no estén separados los círculos? ____

¿En qué región del rectángulo, representarías a los 10 alumnos que no tienen ninguna característica? ____

Uno de los datos que se nos proporcionan como información, es que 55 dicen tener la característica de ser estudioso (nota que no se especifica que tengan esta característica únicamente), lo cual quiere decir que dentro del círculo que representa al evento E, debe de representar en total a 55 alumnos.

Con tus compañeros de equipo, discutan y llenen todas las regiones del diagrama de Venn con el número de alumnos correcto, para contestar de manera acertada a la pregunta: ¿Cuántos alumnos tiene el profesor? ____________

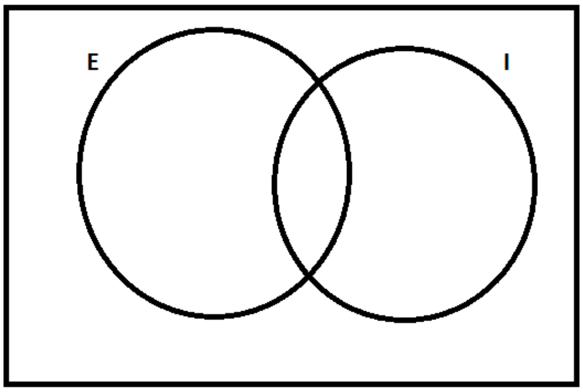
Ilumina de color verde la región $\bar{T} \cap E$

Ilumina de amarillo la región $\bar{T} \cap \bar{E} \cap I$

Utilizando el diagrama de Venn, el cual fue de gran ayuda para saber el TOTAL de alumnos del profesor, obtén las siguientes probabilidades al seleccionar al azar a un alumno:

a) La probabilidad de que sea tranquilo ____________
b) La probabilidad de que no sea estudioso ___________
c) La probabilidad de que sea inteligente y estudioso ___________
d) La probabilidad de que al menos tenga una de las tres características __________
e) La probabilidad de que no tenga ninguna de esas tres características __________
f) $P(T \cup E)$= ____________
g) $P(\bar{T} \cap E)$=____________
h) $P(\bar{T} \cap \bar{E})$=_____________

Ya sabemos cuántos alumnos tiene el profesor, y consideremos que la característica de ser tranquilo no importa, ¿cómo quedaría representada la información por medio de un diagrama de Venn?



La información del diagrama anterior, la puedes representar por medio de una tabla de contingencia, llena los espacios correspondientes:

	I	$\bar{I}$
E		
$\bar{E}$		

Si sumamos los alumnos estudiosos y los no estudiosos, ¿cuántos alumnos debe de haber? _______

Elabora una tabla de doble entrada para representar I y T

Elabora una tabla de doble entrada para representar T y E

¿Cuál es $P(T \cap E)$? __________

¿Cuál es $P(T \cup E)$? ___________

Probabilidad condicional y eventos independientes

Considera la siguiente situación hipotética:

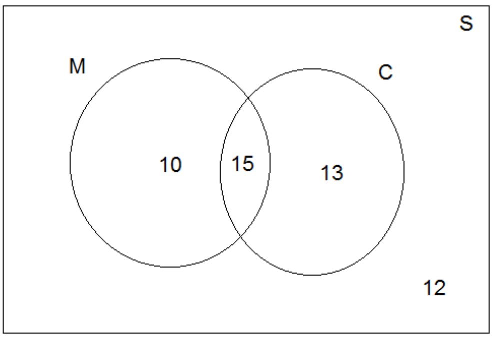
"La calificación de cada uno de los 50 estudiantes (35 son mujeres), será reprobatoria excepto 1. La persona acreedora de esa calificación aprobatoria será seleccionada al azar"

Ana Gabriela es una integrante de ese grupo, ¿cuál es la probabilidad de que sea la acreedora de la calificación aprobatoria? _____. Pero, si la persona que apruebe tiene que ser mujer ¿cuál es la probabilidad de que apruebe Ana Gabriela **dada** esta condición? _____ Bueno, te percataste que hubo una reducción del número de elementos del espacio muestral, ¿a cuántos elementos se redujo? _______.

Ahora consideremos otra situación:

El grupo está compuesto por 50 alumnos, de los cuales 25 de ellos utilizan el metro para llegar a la escuela, 28 combi y 15 ambos. Si seleccionamos a un alumno al azar, ¿cuál es la probabilidad de que no utilice ninguno de estos dos transportes para llegar a la escuela? Vamos a representar la información por medio de un diagrama de Venn, pero antes, definamos a los eventos $M = \{Utilice\ metro\}$, $C = \{Utilice\ combi\}$. Recuerda que $C \cap M$ simboliza a, ¿qué enunciado? ________________________________. Y la probabilidad es igual a $P(C \cap M) =$ ______.

Entonces, el diagrama de Venn:



a) ¿Cuántos alumnos no utilizan ni metro ni combi para llegar a la escuela? ___

b) ¿Cuántos alumnos utilizan metro o combi para llegar a la escuela? ____

c) ¿Cuántos no toman metro para llegar a la escuela? _____

Simboliza y calcula sus probabilidades:

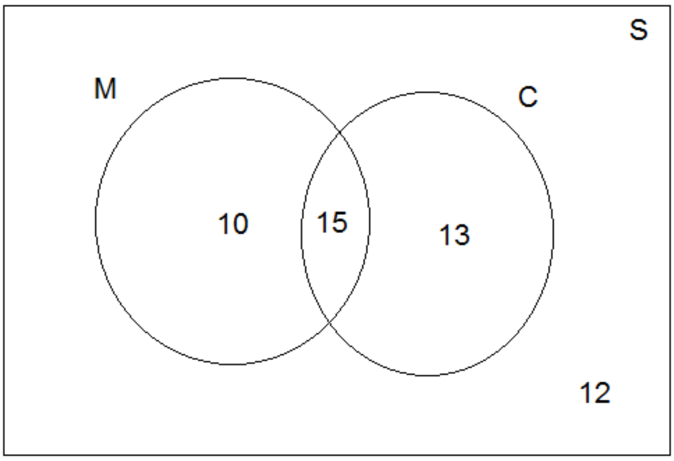
Podemos también representar la información por medio de una tabla de doble entrada, llena los espacios correspondientes:

	M	$\overline{M}$
C		
$\overline{C}$		

Puedes observar, que la tabla de doble entrada también es muy útil para calcular probabilidades, ¿podrías calcular las probabilidades anteriores haciendo uso de ella? ____Si seleccionamos a un alumno al azar, ¿cuál es la probabilidad de que utilice el metro para llegar a la escuela? _________
Con símbolos: _______________

Vamos a condicionar la situación, por ejemplo: si el evento C ocurre, es decir, ¿cuál es la probabilidad de que el alumno seleccionado utilice metro **dado** que utiliza combi para llegar a la escuela? Con la condición dada, el espacio muestral que estaba constituido de 50 alumnos se redujo a, ¿cuántos alumnos? ____, pues efectivamente, ya que sólo ellos son los que reúnen la característica de utilizar combi para ir a la escuela.

Colorea de azul la región del diagrama de Venn que paso la condición, es decir, el círculo que representa al evento C.



Realiza lo mismo en la tabla de contingencia, colorea de azul los cuadros que pasaron la condición:

	M	$\overline{M}$
C		
$\overline{C}$		

Con esa condición, ¿cuál es la probabilidad de que ocurra M, dado C? ___________

La anterior probabilidad la podemos simbolizar como $P(M\backslash C)$ y se lee "probabilidad de M dado C" y de acuerdo al contexto de la situación significa "probabilidad de que el alumno seleccionado utilice metro dado que utiliza combi para llegar a la escuela".

¿Cuál será la probabilidad de que **no** ocurra el evento M, dado que C ocurrió? Es decir, $P(\overline{M}\backslash C)$=_________

Observa el diagrama de Venn o la tabla de doble entrada y date cuenta que para obtener este tipo de probabilidades, podemos hacer uso de la siguiente expresión: $P(A|B) = \frac{P(A\cap B)}{P(B)}$. Haciendo uso de la expresión corrobora tu resultado:

$P(\bar{M}|C) = \frac{}{\boxed{}}$

También podemos condicionar la situación por medio del evento M, es decir, el evento M ocurrirá o suponemos que ocurrirá, ¿cuál es la probabilidad de que C dado M? Es decir, $P(C|M) =$_________

En muchas ocasiones nos encontraremos con eventos independientes, eventos que no influyen la probabilidad de otro si ocurren o no. Si un evento A y un evento B son independientes, entonces la probabilidad de que sucedan simultáneamente se obtendrá multiplicando la $P(A)$ por la $P(B)$. Esta probabilidad se simboliza por $P(A \cap B) = P(A)P(B)$ y se lee "probabilidad de que sucedan A y B al mismo tiempo es igual a la probabilidad de A por la probabilidad de B".

Ejercicio.

Consideremos las ruedas de una bicicleta

Para que la bicicleta pueda servir, tanto la llanta A y B deben estar en perfectas condiciones, o de menos no deben de estar ponchadas, por lo que, si falla una, o las dos, la bicicleta estará defectuosa.

Si una llanta no tiene ningún defecto, diremos que es buena; de lo contrario, que es defectuosa. Para representar escribiremos:

$A = \{la\ llanta\ A\ está\ buena\}$ $\quad$ $B = \{la\ llanta\ B\ está\ buena\}$

$\bar{A} = \{ \qquad\qquad\qquad\qquad \}$ $\quad$ $\bar{B} = \{$

Supongamos que el 10 por 100 de las llantas utilizadas para cierta marca de bicicletas son defectuosas. Es decir, que 90 por 100 son buenas.

a) Calcula las probabilidades:

$P(A) =$ $P(B) =$ $P(\bar{A}) =$ $P(\bar{B}) =$

b) Calcula también las siguientes expresiones:

$P(A) + P(\bar{A}) =$ $P(B) + P(\bar{B}) =$

c) Para indicar el evento $\{las\ llantas\ A\ y\ B\ son\ buenas\}$ escribimos $A \cap B$, usando el símbolo de la intersección de conjuntos. ¿Sabrías indicar que significan los eventos siguientes?

$A \cap \bar{B}$ ____

$\bar{B} \cap A$ ____

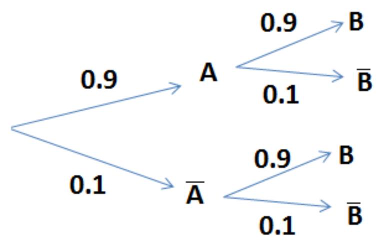
$\bar{B} \cap \bar{A}$ ____

d) Para indicar el evento "al menos una de las llantas A o B son buenas" escribimos $A \cup B$, usando el símbolo de la unión de conjuntos. Observa el siguiente diagrama en el que representamos como conjuntos los eventos $A, \bar{A}, B$ y $\bar{B}$.

	A	$\bar{A}$
B	$A \cap B$	$\bar{A} \cap B$
$\bar{B}$	$A \cap \bar{B}$	$\bar{A} \cap \bar{B}$

e) Raya los rectángulos que representan el evento $A \cup B$

Mediante el siguiente diagrama en árbol vamos a representar todas las situaciones que pueden producirse:



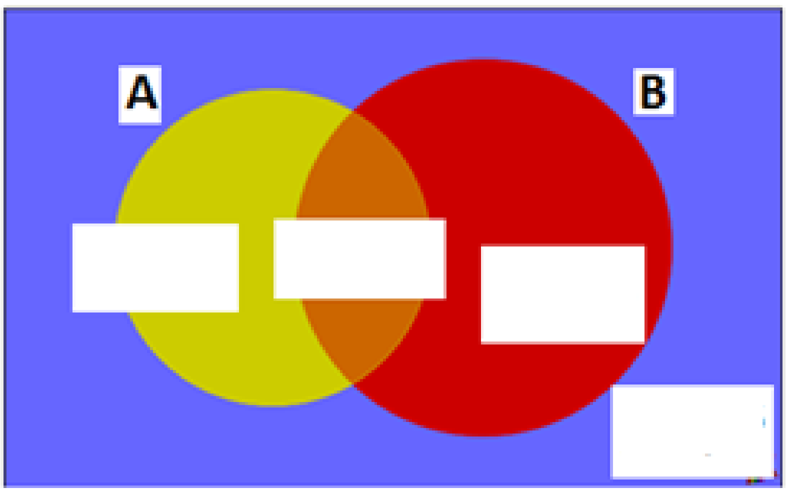
suceso		probabilidad
$A \cap B$	...	0.9x0.9= 0.81
$A \cap \bar{B}$	...	
$\bar{A} \cap B$	...	
$\bar{A} \cap \bar{B}$	...	

Calca las probabilidades de los siguientes eventos:

- Sólo A sirva.
- Tanto A como B son llantas buenas.
- Al menos una de las llantas es buena.

Calcula la probabilidad del evento $A \cup B$ y compárala con cada una de las sumas siguientes: $P(A) + P(B) - P(A \cap B)$ = ____________________

Llena los espacios del siguiente diagrama de Venn, con la respectiva probabilidad de cada región, recuerda que la suma de todas esas probabilidades debe ser 1, puesto que el rectángulo representa el espacio muestral, es decir, todas las posibilidades:



Ejercicio.

Considera a los eventos $A = \{Alberto\ llegue\ temprano\}$,

$B = \{Beatriz\ llegue\ temprano\}$ y $C = \{Carolina\ llegue\ temprano\}$

Considera también a lo tres eventos independientes entre sí.

De acuerdo con las estadísticas la $P(A) = 0.3$, $P(B) = 0.4$ y $P(C) = 0.5$. Encuentre la probabilidad de que:

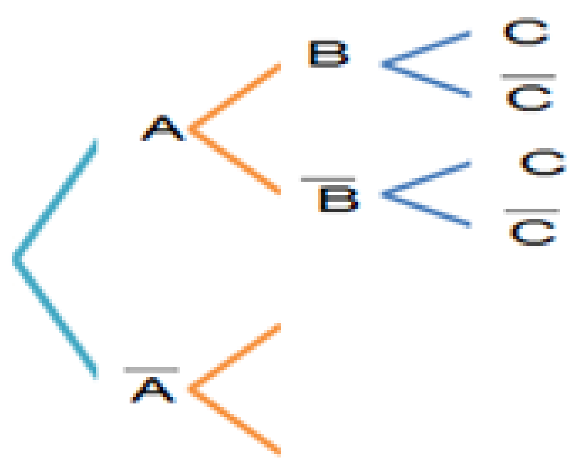
a) Los tres lleguen temprano

b) Ocurra exactamente uno de esos eventos, es decir, A llegue temprano y los otros 2 no, B llegue temprano y los otros 2 no, o, C llegue temprano y los otros 2 no.

Completa lo siguiente:

$$P(A) = 0.3;\ P(B) = 0.4;\ P(C) = \_\_;\ P(\bar{A}) = \_\_;\ P(\bar{B}) = \_\_\ y\ P(\bar{C}) = 0.5$$

Con un diagrama de árbol también podemos representar la información, con la ayuda de tus compañeros de equipo, lo pueden lograr, utilicen el siguiente esquema y complétenlo:



Ejercicio.

A continuación, se presenta una situación para que obtengas algunas probabilidades, obtenlas por medio de un diagrama de árbol, diagrama de Venn, tabla de doble entrada y por expresión matemática.

Dos ambulancias se mantienen listas para emergencias. Debido a la demanda y a la posibilidad de fallas mecánicas, la probabilidad de que una ambulancia específica esté disponible cuando se necesite es .8. La disponibilidad de una ambulancia es independiente de la otra.

a) En caso de catástrofe, ¿cuál es la probabilidad de que ambas ambulancias estén disponibles?
b) ¿Cuál es la probabilidad de que ninguna esté disponible?
c) Si se necesita una ambulancia en una emergencia, ¿cuál es la probabilidad de que al menos haya una disponible?
d) Si se necesita una ambulancia en una emergencia, ¿cuál es la probabilidad de que haya solamente una disponible?

Ejercicios

1. Sea P(A) = 0.2, P(B) = 0.3 y A y B eventos independientes. Obtenga:

$a)P(A \cap B)$

$b)P(A \cup B)$

$c)P(A^C \cap B^C)$

2. Sea P(A) = 0.4 y P(B) = 0.2, y A y B eventos mutuamente excluyentes. Obtenga:

$a)P(A \cup B)$

$b)P(A \cap B)$

$c)P(B^C)$

3. Sea P(A) = 0.4, P(B) = 0.2 y $P(A \cap B)$ = 0.1. Obtenga:

$a)P(A \cup B)$

$b)P(A^C \cap B^C)$

a) la representación por medio de un diagrama de Venn
b) la probabilidad de que suceda A o B
c) la probabilidad de que no suceda ninguno de los dos eventos A y B

Los eventos A y B, ¿son independientes? Justifica tu respuesta.

4. Sea P(A) = 0.5, P(B) = 0.3 y $P(A \cup B)$ = 0.6. Obtenga:

$a)P(A \cap B)$

$b)P(A^C \cap B^C)$

$c)P((A \cap B)^C)$

Represente por medio de una tabla de doble entrada.

5. Un aparato contiene siete sistemas electrónicos complejos. Sin que el comprador lo sepa, hay tres defectuosos. Se seleccionan dos de siete para

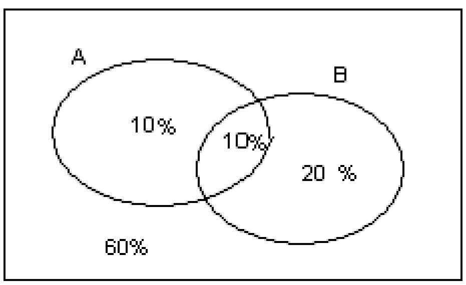
someterlos a pruebas exhaustivas y se clasifican como sistema defectuoso o no defectuoso.

a) Haga una lista de los puntos muestrales para este experimento. Sol. Habrá 21 puntos muestrales, enlístelos.
b) Sea A el suceso de que la selección no incluye defectuosos. Haga una lista de los puntos muestrales en A. Sol. Habrá 6 elementos en el evento A, enlístelos.
c) Asigne probabilidades a los puntos muestrales y encuentre P(A).

6. Un dado se lanza dos veces. ¿Cuál es la probabilidad de que la suma de los números observados sea mayor que 9?

7. Se debe examinar un grupo de personas respecto a dos sistemas comunes de cierta enfermedad. Se considera que 20% de las personas presentan el síntoma A, 30% tiene el síntoma B, 10% tiene ambos síntomas y el resto no tiene síntoma alguno. Para una persona escogida al azar de este grupo, encuentre las probabilidades de los eventos siguientes:

a) Que la persona no presente síntoma alguno.
b) Que la persona presente al menos un síntoma.
c) Que la persona presente ambos síntomas, dado que presenta el síntoma B.



Represente los eventos por medio de una tabla de doble entrada

8. De 10 niñas de clase, 3 tienen ojos azules. Si se escogen dos niñas al azar, ¿cuál es la probabilidad de que...

(i) las dos tengan ojos azules?

(ii) ninguna tenga ojos azules?

(iii) una por lo menos tenga ojos azules?

9. En cierta ciudad, 50% de la población tienen cabellos castaños, 20% tiene ojos castaños y 10% tiene cabellos y ojos castaños. Se escoge una persona al azar

i) Si tiene cabellos castaños, ¿cuál es la probabilidad de que también tenga ojos castaños?
ii) Si tiene ojos castaños, ¿cuál es la probabilidad de que no tenga cabellos castaños?
iii) ¿Cuál es la probabilidad de que no tenga cabellos castaños ni ojos castaños?

10. En una escuela particular de bachillerato se tiene la siguiente información con respecto a los idiomas que estudian sus 125 alumnos:

1 solamente estudia los tres idiomas que se imparten en la escuela.

5 estudian francés y alemán.

10 estudian inglés y alemán.

8 estudian francés e inglés.

9 estudian solamente inglés.

22 estudian francés.

Y 75 de esos 125 que son el total de los alumnos en la escuela no estudian algún idioma.

Si se selecciona a un alumno al azar de esta escuela,

a) ¿Cuál es la probabilidad de que estudie francés?
b) ¿Cuál es la probabilidad de que no estudie algún idioma?
c) ¿Cuál es la probabilidad de que estudie alemán?

d) ¿Cuál es la probabilidad de que estudie de menos algún idioma?
e) ¿Cuál es la probabilidad de que estudie solamente alemán?

11. Un vendedor de autos caros tiene 23 clientes de los cuales 17 son millonarios, 8 jubilados, incluidos 4 que también son millonarios. Si seleccionamos de manera aleatoria a uno de sus clientes, ¿Cuál es la probabilidad de que el cliente...

a) ¿Cuál es la probabilidad de que el cliente sea millonario o jubilado?
b) ¿Cuál es la probabilidad de que el cliente no sea millonario ni jubilado?

Sugerencia: Resuélvelo por medio de una tabla de doble entrada.

12. De un grupo de 48 personas que asisten a una fiesta, 30 fuman, 25 consumen bebidas alcohólicas y 10 ni fuman ni consumen bebidas alcohólicas. Al seleccionar a una persona de manera aleatoria de esta fiesta, ¿Cuál es la probabilidad de que...

a) fume e ingiera bebidas alcohólicas?
b) únicamente fume?
c) únicamente ingiera bebidas alcohólicas?

13. Al SILADIN han asistido 100 alumnos a presentar los exámenes de Matemáticas y Física. Sabiendo que las matemáticas la han aprobado 54 alumnos en total, Física 75 alumnos y 40 han aprobado ambas asignaturas, ¿cuál es la probabilidad de que al seleccionar a un alumno al azar...

a) no haya aprobado ninguna asignatura?
b) haya aprobado una de las dos asignaturas?
c) haya aprobado las dos asignaturas?

14. En una encuesta de 100 estudiantes se observó la siguiente información: 39 alumnos estudian inglés, 39 francés, 36 portugués, 15 estudian inglés y francés, 18 francés y portugués, 19 inglés y portugués y 5 estudian los tres idiomas.

a) Construya el diagrama de Venn correspondiente.
b) ¿Cuál es la probabilidad de que al seleccionar de manera aleatoria a un estudiante de esos 100 éste estudie 2 idiomas exactamente?
c) ¿Cuál es la probabilidad de que al seleccionar de manera aleatoria a un estudiante de esos 100 éste estudie 1 idioma exactamente?
d) ¿Cuál es la probabilidad de que al seleccionar de manera aleatoria a un estudiante de esos 100 éste estudie francés y portugués?

Con la ayuda de tu profesor, trata de simbolizar con símbolos de la lógica de conjuntos las anteriores preguntas.

15. En un grupo de 60 personas, se sabe que 40 trabajan, 24 estudian y 27 tiene coche. Si 55 trabajan o tienen coche, 49 trabajan o estudian, 40 estudian o tiene coche y 2 de esas personas trabajan, estudian y tienen coche. Si seleccionamos a una persona de este grupo,

a) ¿Cuál es la probabilidad de que trabaje, estudie y tenga coche?
b) ¿Cuál es la probabilidad de que tenga coche, pero no estudie ni trabaje?
c) ¿Cuál es la probabilidad de que no trabaje, ni estudie y tampoco tenga coche?
d) ¿Cuál es la probabilidad de que no trabaje, pero si estudie y tenga coche?

16. Los empleados de una compañía se encuentran separados en tres divisiones: Administración (A), Ventas (V) y Bodega (B). La siguiente tabla indica el número de empleados en cada división clasificados por sexo:

	Mujer (M)	Hombre (H)	Total
Administración (A)	120		
Ventas (V)	90		180
Bodega (B)		100	
Total		220	500

Si se elige aleatoriamente un empleado:

a) ¿Cuál es la probabilidad de que sea mujer?
b) ¿Cuál es la probabilidad de que sea mujer dado que no es un empleado de Bodega?
c) ¿Cuál es la probabilidad de que sea mujer y trabaje en Ventas?
d) ¿Cuál es la probabilidad de que sea hombre o trabaje en Administración?
e) ¿Cuál es la probabilidad de que no sea mujer o trabaje en Ventas o trabaje en Bodega?

Obtenga:

f) $P(AUM^{C})$

g) $P(A^{C} \cap M^{C})$

h) $P((AUM)^{C})$

i) $P(H \setminus B)$

j) $P(A^{C} \setminus H^{C})$

k) $P(A \setminus V)$

l) $P(V^{C} \setminus H)$

17. En una investigación que se realizó en los antros se obtuvo la siguiente información:

	Fuma (F)	No Fuma	Total
Mayor de edad (M)			42%
Menor de edad (N)		5%	
Total	82%		100%

Obtenga:

a) $P(F \cup M^C)$

b) $P(F^C \cap M^C)$

c) $P((F \cup M)^C)$

d) $P(N \setminus F)$

e) $P(N^C \setminus F^C)$

f) $P(F \setminus N)$

Instrumento de evaluación

1. ¿Qué característica tienen los eventos mutuamente excluyentes? ________

____

2. ¿Cuál es la probabilidad de que sucedan al mismo tiempo dos eventos mutuamente excluyentes? ____

3. ¿Cómo se observan los círculos que representan dos eventos mutuamente excluyentes en un diagrama de Venn? ________________________________

4. ¿Cómo obtienes y cómo simbolizas la probabilidad de que sucedan A o B?

____

5. ¿Qué significa que dos eventos sean independientes? _________________

____

6. ¿Cómo obtienes probabilidades condicionales? _______________________

7. La probabilidad de que Karina pase su examen de aritmética es de 0.7 y de que Carlos pase ese mismo examen de 0.6. Consideremos que estos dos eventos son independientes. Por medio de una tabla de doble entrada, un diagrama de Venn y un diagrama de árbol, encuentra la probabilidad de que...

a) Ninguno pase el examen.
b) Los dos pasen el examen.
c) De menos uno pase el examen.
d) Solamente uno pase el examen.

8. En una urna colocaron una bola azul con el número 3, una bola roja con el número 4, otra roja con el número 5 y una roja más con el número 7. Se definen los eventos:

$$R = \{bola\ roja\}$$

$$A = \{bola\ azul\}$$

$$I = \{bola\ con\ número\ impar\}$$

Si selecciona al azar a una bola de la urna, simboliza y obtén la probabilidad de que…

a) La bola sea roja.

b) La bola tenga un número impar.

c) La bola sea roja y tenga un número impar.

d) De que sea roja y azul.

e) Los eventos R y A, son mutuamente excluyentes.

f) Si la bola es roja, ¿cuál es la probabilidad de que tenga número impar?

g) Si la bola tiene un número impar, ¿cuál es la probabilidad de que sea azul?