GUÍA DE ESTUDIO DE ESTADÍSTICA Y PROBABILIDAD

BACHILLERATO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Dirección General de la Escuela Nacional Preparatoria
Colegio de Matemáticas
Jefatura de Producción Editorial de la Escuela Nacional Preparatoria

ESCUELA NACIONAL PREPARATORIA

COLEGIO DE MATEMÁTICAS

Sexto año Clave 1712 Plan 1996

ESTADÍSTICA Y PROBABILIDAD

Guía cuaderno de trabajo académico

Programa Actualizado

Aprobado por H. Consejo Técnico el 13 de abril 2018

Bachillerato

Coordinación y revisión
Martha Patricia Rodríguez Rosas

Autores
María Isabel Escalante Membrillo
María Fragoso Sandoval
Raúl Monroy Santana
Claudia América Serrano Liceaga

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Escuela Nacional Preparatoria
Dirección Genera: Biól. María Dolores Valle Martínez
Secretaria Académica: M. en C. María Josefina Segura Gortares
Departamento de Producción Editorial: Lic. María Elena Jurado Alonso

Diseño de Portada: DCG Edgar Rafael Franco Rodríguez
Diseño editorial: Martha Patricia Rodríguez Rosas
Corrección de estilo: Martha Patricia Rodríguez Rosas
Cuidado de edición: Jonathan Iván Jiménez Castellanos

Primera edición: febrero, 2020

Impreso en México.

PRESENTACIÓN

La Escuela Nacional Preparatoria, institución educativa con más de 150 años de experiencia formando jóvenes en el nivel medio superior, busca la constante actualización y mejora de sus materiales de apoyo a la docencia, así como la publicación de nuevos ejemplares, siempre teniendo en mente a nuestros alumnos y su aprovechamiento.

Después de varios años de trabajo, reflexión y discusión, se lograron dar dos grandes pasos: la actualización e implementación de los programas de estudios de bachillerato y la publicación de la nueva colección de Guías de Estudio. Sin embargo, los trabajos, resultado del espíritu crítico de los profesores, siguen dando fruto con publicaciones constantes de diversa índole, siempre en torno a nuestro quehacer docente y a nuestros programas actualizados.

Ciertamente, nuestra Escuela Nacional Preparatoria es una institución que no se detiene, que avanza con paso firme y constante hacia su excelencia académica, así como preocupada y ocupada por la formación integral, crítica y con valores de nuestros estudiantes, lo que siempre ha caracterizado a nuestra Universidad Nacional.

Aún nos falta más por hacer, por mejorarnos cada día, para que tanto nuestros jóvenes estudiantes como nuestros profesores seamos capaces de responder a esta sociedad en constante cambio y a la Universidad Nacional Autónoma de México, la Universidad de la Nación.

"POR MI RAZA HABLARÁ EL ESPÍRITU"

BIÓL. MARÍA DOLORES VALLE MARTÍNEZ

DIRECTORA GENERAL

ESCUELA NACIONAL PREPARATORIA

INTRODUCCIÓN

La guía de Estudios del Programa de Estadística y Probabilidad tiene como objetivo *brindar los contenidos básicos necesarios para presentar un examen extraordinario* acorde a los propósitos y enfoque principal del programa de esta asignatura.

El propósito de la asignatura de Estadística y Probabilidad es que los estudiantes consoliden sus conocimientos sobre estadística descriptiva e inicien el tránsito hacia el pensamiento estocástico, por medio de una enseñanza basada en el análisis de datos reales y en la resolución de problemas, de esta manera el programa permite que el alumno reconozca la importancia del análisis e interpretación de los números y datos en contextos reales, pero sobre todo brinda las herramientas necesarias para la toma de decisiones basadas en un pensamiento estocástico o probabilístico ya que muchas características asociadas a fenómenos naturales, sociales, del diario acontecer humano y del mundo productivo de las empresas ocurren de manera aleatoria o son objeto de un análisis estadístico, por lo que su estudio se convierte en una labor compleja.

En este sentido, la presente guía se basa en el análisis de situaciones contextualizadas que permiten que el alumno transite entre los contenidos y las habilidades de razonamiento para la resolución de las mismas. Los contenidos surgen para poder dar respuesta a cada interrogante planteada, así se abordan los contenidos conceptuales básicos. El programa de Estadística y Probabilidad está dividido en dos grandes unidades:

En la *Unidad 1, Estadística para analizar datos del entorno*, las situaciones pretenden desarrollar habilidades de análisis al tiempo que se aplican las técnicas de estadística descriptiva.

La Unidad 2, Probabilidad para estudiar la incertidumbre, se plantean situaciones donde interviene el azar o la incertidumbre con el fin de estimar la probabilidad de ocurrencia de eventos

La Guía de estudios de Estadística y probabilidad presenta en cada unidad, además de situaciones para introducir los contenidos, situaciones de autoevaluación, referencias bibliográficas, un modelo de examen extraordinario y al final las tablas de distribución normal y un formulario.

ÍNDICE PÁG.

UNIDAD I ESTADÍSTICA PARA ANALIZAR DATOS DEL ENTORNO

Objetivo específico.

El alumno:

> Desarrollará habilidades de investigación y análisis, a través del desarrollo de un proyecto en el cual delimite un problema de su interés, seleccione el medio para obtener datos reales (diseño de cuestionario o selección de una base de datos confiable), aplique técnicas de estadística descriptiva y utilice herramientas tecnológicas, para procesar la información, sistematizarla y analizarla, comunicar resultados con el lenguaje apropiado, obtener conclusiones y asumir una postura personal.

1.1. Componentes básicos de una investigación estadística

Importancia de la Estadística

La enseñanza de la Estadística, en los últimos años, ha tomado un notable impulso de acuerdo con Batanero (2001) “Estamos caminando hacia una sociedad cada vez más informatizada y una comprensión de las técnicas básicas de análisis de datos y su interpretación adecuada son cada día más importantes.” Por lo tanto, el estudiante debe comprender las principales técnicas para el análisis e interpretación de datos y ser capaz de procesar una gran cantidad de información.

La estadística tiene una gran utilidad en la vida cotidiana, como dice H.G. Wells:

> *“Llegará un día en que el pensamiento estadístico será tan necesario para ser un ciudadano eficiente como la capacidad de leer y escribir”*

Por ejemplo: si se desea comprar un auto, se requiere recolectar información: tipos de autos, precios, desempeño, color, tamaño, etc. y se debe procesar con el fin de elegir aquel que sea más adecuado a las necesidades.

Continuamente se presenta información, en los medios de comunicación, periódicos, televisión, radio, redes sociales, internet y se tiene que saber interpretar toda esa información y esto es una de las cosas que enseña la Estadística.

La **estadística** es la rama de las matemáticas, que se encarga de la recolección, análisis y descripción de datos con el objetivo de tomar decisiones lo más acertadas posibles.
Se clasifica en dos grandes grupos:

- **Estadística descriptiva,** recolecta y caracteriza un conjunto de datos.
- **Estadística inferencial,** estima las características de una población a partir de una muestra

1.1.1. Población y muestra. Estadístico y parámetro

Situación 1. Curiosidad.

A menudo aparecen ante nosotros una serie de interrogantes a las cuales quisiéramos dar una respuesta de una manera precisa, estas preguntas tienen que ver con nuestros intereses y con lo que ocurre es nuestra sociedad. Algunas de estas preguntas podrían ser:

¿Cuáles son las profesiones con mayores ingresos?

¿Cuáles son las carreras más demandadas?

¿Cuáles son los principales problemas de los jóvenes en México?

¿Cuántas mujeres sufren violencia durante su noviazgo?

¿Cuál es el número de embarazos en adolescentes en México? ¿y en la ciudad? ¿y en mi escuela?

¿Cuál es la red social más utilizada?

1.1 ¿Cómo podríamos dar respuesta a estas preguntas?, por ejemplo, si se quisiera saber ¿cuánto gastan los jóvenes de México, de entre 15 años y 20 años, en juegos de video? Diseña un procedimiento para dar respuesta a esta pregunta.

Etapas de un estudio estadístico

1. Selección de la población o muestra en estudio
2. Realización de una encuesta para recolectar los datos
3. Organización de la información (tablas o gráficas)
4. Tratamiento de la información
5. Conclusiones

Solución:

Para contestar a la pregunta que se plantea se requiere:

A) Seleccionar una muestra de jóvenes entre 15 y 20 años

B) Realizar una encuesta acerca de los juegos de video y la pregunta de interés

C) Organizar la información

D) Realizar el tratamiento de la información (calcular los principales estadísticos: media, mediana, moda, varianza y desviación estándar)

E) Establecer una conclusión

1.2 Un estudiante se pregunta ¿cuáles son las profesiones con mayores ingresos? No sabe si para hacer el estudio que conteste a su interrogante debe consultar a profesionistas recién egresados que han conseguido empleo o entrevistar a profesionistas que ya llevan más de 1 año en su trabajo, o entrevistar a ambos ¿en qué etapa del proceso de aplicación de la estadística se encuentra?

Solución:

Se encuentra en la selección de la población o muestra en estudio.

1.3 Ingresa a la página http://xurl.es/szq9f que muestra datos de un estudio realizado por la Asociación de Internet de México acerca de usuarios de internet. Identifica y ejemplifica en este estudio cada una de las etapas de un estudio estadístico.

Solución:

A) Población (usuarios de internet) de los que se tomó una muestra de 1968 +470 entrevistas de sobremuestra.

B) Se realizó una encuesta autoaplicada digital

C) Se organizó la información en tablas y gráficas

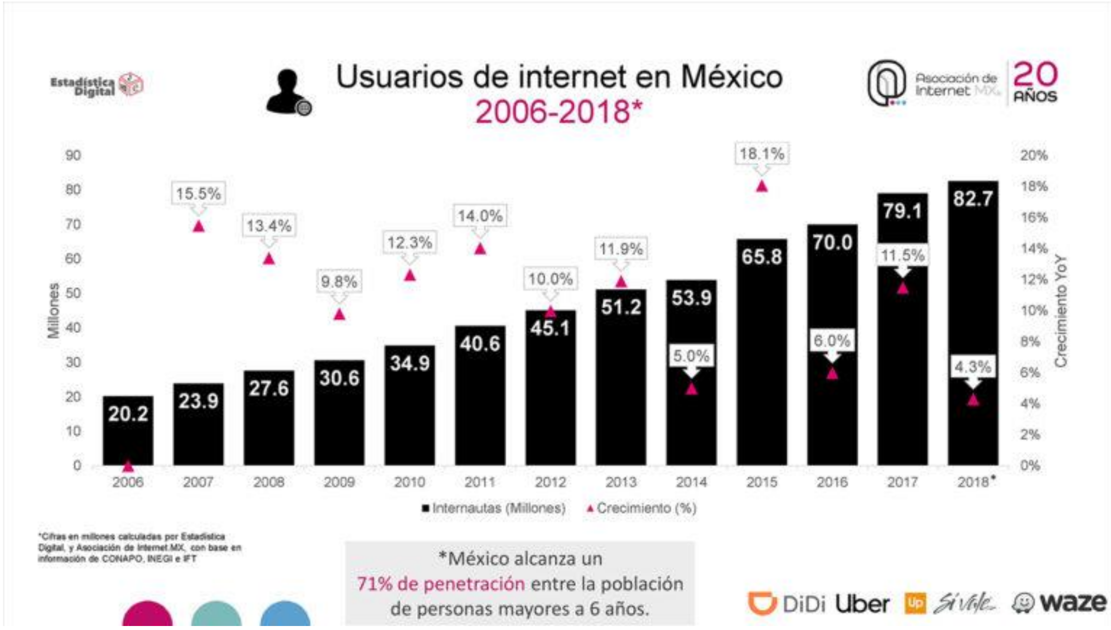


Figura 1. (Asociación de Internet). http://xurl.es/szq9f

D) Tratamiento de la información, algunos ejemplos son:

- **Usuarios de internet en México**: Lo primero que hay que decir es que el Internet está teniendo impacto sobre la población de personas mayores de 6 años. Esta estadística supone un 71% de la población nacional se integra al universo digital de las redes

- **Perfil del Internauta:** Se sigue comprobando que el uso del Internet, en materia de género, representa casi un 50/50%. El dato exacto es de 51% de mujeres y 49% de hombres. El mayor uso de Internet lo realizan las personas de 25 a 34 años.

- **Hábitos de Conexión:** El consumo de Internet por persona se encuentra en un periodo de hasta 8 horas con 20 minutos. El smartphone representa el equipo de mayor conectividad a las redes, destacándose en un 92%.

- **Actividades en Línea:** Más que todo, el consumo del Internet se realiza para Redes Sociales. El porcentaje es del 82% de la población. La jerarquía del consumo indica que más que todo se le da utilidad en materia de comunicación.

E) Conclusiones

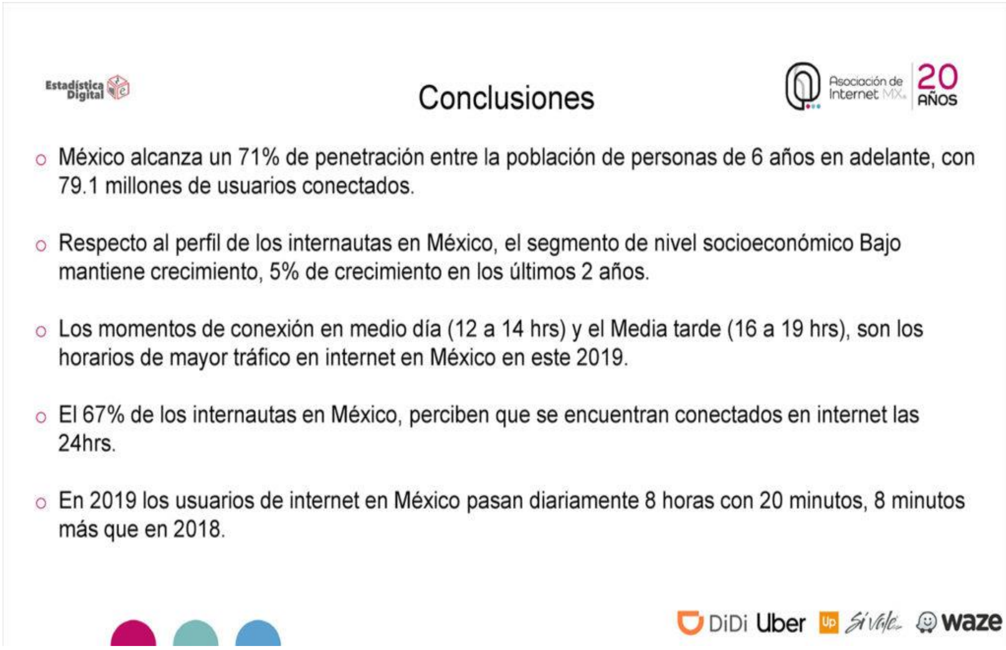


Figura 2. (Asociación de Internet). http://xurl.es/szq9f

Situación 2. Tiempo para llegar a la escuela.

La gente en la ciudad de México tarda hasta dos horas para llegar a su trabajo o escuela, esto implica una pérdida de tiempo considerable ya que pasan mucho tiempo en el trasporte que podrían dedicar a otra actividad. Para una mayor información ingresa a la siguiente liga: http://xurl.es/dtpeb

2.1 Si se realiza un estudio, similar al planteado en la nota, acerca del tiempo que tardan los estudiantes de tu escuela en el traslado hacia esta ¿cuál sería la población en estudio?

Población
Se llama **población** al conjunto de personas, elementos, artículos, etc. que forman la totalidad de un estudio.

Solución:

La población en estudio son los alumnos de tu escuela ya que forman la totalidad de los elementos en estudio.

2.2 ¿Cuál sería un ejemplo de una muestra para este estudio?

Muestra

Una **muestra** es cualquier parte de una población, en términos de conjuntos se llama muestra a un subconjunto de la población.

Solución:

Un ejemplo de una muestra para este estudio sería cualquier subconjunto de la población, en este caso, podrían ser los estudiantes de cuarto año de tu escuela, los de quinto o cualquier grupo.

2.3 Identifica en los casos siguientes si el estudio que se realiza corresponde a una muestra o una población.

A) Los participantes para la evaluación de un nuevo shampoo contra la caspa

B) El lote de 10,000 tabletas de ácido acetil salicílico fabricadas por un laboratorio químico

C) Las personas que fuman tabaco

D) Los autos negros en un estacionamiento
E) Los alumnos de sexto año de una escuela primaria

F) Los focos defectuosos en un lote de 5000 focos

Solución:

A) Muestra (se considera únicamente a los participantes los cuales son una parte de la población)

B) Población (Se toma a todo el lote es una población, se considera también que la población no son sólo personas)

C) Población (se habla de las personas que fuman tabaco, es decir la totalidad de individuos que tienen ese hábito)

D) Muestra (sólo se toma en cuenta a los autos negros que son una parte del total de autos en el estacionamiento)

E) Muestra (de todos los alumnos de la escuela se toman los de sexto)

F) Muestra (de los 5000 focos sólo se toman los defectuosos).

Situación 3. ¿A quién darías tu voto?

El primero de julio de 2018 hubo elecciones para elegir presidente de la república en México, a un año de este evento se encuestó a los ciudadanos para saber si volverían a votar por el candidato ganador, para más información consulta http://xurl.es/rr45s. La siguiente gráfica muestra si los ciudadanos volverían a votar por el actual presidente de la República Andrés Manuel López Obrador.

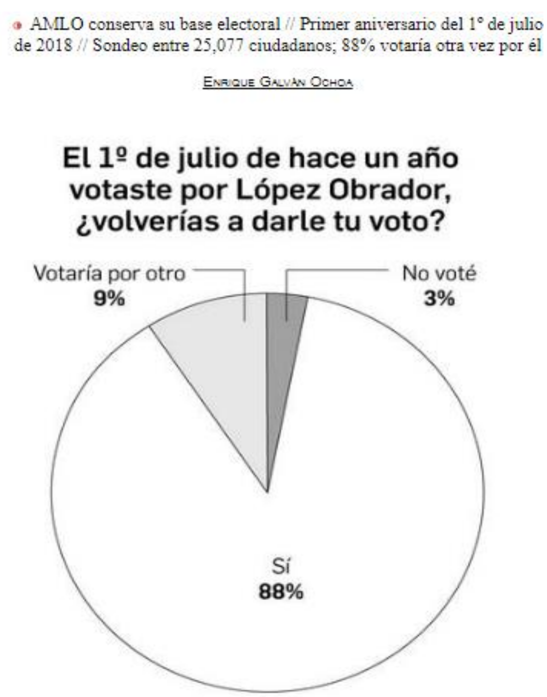


Figura 3. (Voto). http://xurl.es/rr45s

3.1 ¿Los datos mostrados en la gráfica se refieren a una población o son de una muestra?

Solución:

Los datos mostrados se refieren a una muestra, el gráfico menciona que es un sondeo realizado a 25 077 ciudadanos, es decir, una parte de la población.

3.2 ¿cuántos ciudadanos contestaron afirmativamente en la encuesta?

Solución:

22068, que corresponde al 88% de las personas que participaron en la encuesta.

3.3 ¿Cuántas personas contestaron que no votaron?

Solución:

753, lo cual representa el 3 % de los encuestados.

1.1.2. Estadístico y parámetro

Situación 4. México.

Juan es estudiante de preparatoria y a través de las redes sociales conoció a Matti un joven finlandés que tiene un gran interés por nuestro país. Matti, sabe que, en México, nos gustan los tacos, comemos chile, escuchamos a los mariachis y tenemos hermosas playas, sin embargo, desea conocer otros datos estadísticos acerca de los siguientes temas:

- **a)** Población
- **b)** Medio Ambiente
- **c)** Educación
- **d)** Precios
- **e)** Turismo

Parámetro

Es una cantidad numérica calculada sobre una población que permite caracterizarla.

Estadístico

Es una cantidad numérica calculada sobre una **muestra** que permite describirla. Cualquier valor numérico que diga algo acerca del conjunto de datos es un estadístico así, el valor máximo, mínimo y el rango son estadísticos ya que proporcionan información sobre el conjunto de datos.

4.1 Ayuda a Juan ingresando a https://www.inegi.org.mx/app/areasgeograficas/ consulta los datos y elabora una lista de aquellos que consideres de mayor interés para Matti.

Solución:

Algunos datos de interés podrían ser:
Inegi

A) Población total 119, 938, 473

B) Grado promedio de escolaridad 9.2

C) Población hablante de lengua indígena 6,695,228

D) Número de entidades federativas 32

E) Tasa de prevalencia delictiva por cada cien mil habitantes de 18 años y más, mujeres (Casos por cada 100 mil habitantes), 2018 27, 045

4.2 Para las siguientes situaciones indica si el estudio realizado se refiere a un estadístico o un parámetro.

A) Se hace una predicción sobre las preferencias electorales en el estado de Querétaro, entrevistando a 1000 personas.

B) Los datos obtenidos en el próximo censo de población 2020

C) Los resultados obtenidos por los alumnos que presentaron la prueba PISA

D) Un estudio de mercadotecnia para determinar las preferencias sobre un Nuevo jabón de tocador que se lanzará al Mercado.

Solución:

A) Se trata de un estudio muestral, por lo tanto, los resultados se refieren a un estadístico

B) Es un estudio de población, entonces, los resultados son parámetros.

C) Se toman todos los resultados, es decir, se considera a toda la población, los resultados son parámetros.

D) Se realiza un estudio a una muestra de la población que podría usar el jabón, por lo que se trata de un estadístico.

4.3 Investiga en el periódico, o redes sociales (Twitter, Facebook, Google, YouTube, etc.) tres notas que se relacionen con parámetros y tres notas que se refieran a estadísticos.

Solución:

Los siguientes son algunos ejemplos:

I. Estadísticos

a) “Día Mundial sin Tabaco: ¿qué países fuman más y menos en el mundo? (y en qué lugar se sitúan los de América Latina)” https://www.bbc.com/mundo/noticias-44311572
b) “El mexicano consume 163 litros de refresco al año, en promedio: Estudio”. https://noticieros.televisa.com/historia/consumo-refresco-mexico-datos-riesgos/
c) **“85% de la población piensa que podría tener diabetes en el futuro debido a su estilo de vida” http://fmdiabetes.org/estadisticas-en-mexico/**

II. Parámetros

a) “En el Buen Fin, seis de cada 10 combinaron compras en tiendas y online” https://www.jornada.com.mx/ultimas/economia/2019/12/06/seis-de-cada-10-consumidores-compraron-en-tiendas-fisicas-y-online-en-el-buen-fin-9277.html
b) En México se produjeron **300 mil 292 autos en noviembre**, 13 por ciento menos que lo registrado en el mismo mes del año pasado, reportó este viernes el **Instituto Nacional de Estadística y Geografía (Inegi).** https://elfinanciero.com.mx/empresas/produccion-de-autos-en-mexico-tiene-su-peor-noviembre-en-16-anos-cae-13
c) El promedio de goles por partido de Lionel Messi es de 0.87 https://www.goal.com/es-mx/noticias/cuantos-goles-lleva-lionel-messi-en-toda-su-carrera-entre/1janysxr71cok1xgydphg1yxof

1.1.3. Variables y escalas de medición

Situación 5. Violencia de género.

Luisa, una estudiante de bachillerato, está preocupada porque ha notado que el novio de su amiga la maltrata, ha escuchado en los medios de comunicación acerca de la violencia de género y desea tener más información. En esta búsqueda descubre que el Instituto Nacional de las Mujeres cuenta con información en http://xurl.es/atnb7 al leer este documento queda muy impresionada y decide hacer una encuesta con sus compañeros acerca de la violencia de género para tener una idea más clara del problema.

5.1 ¿Cuáles son las algunas de las variables que podría tener en cuenta para elaborar su encuesta?

Variable

Una **variable** es una característica que al ser medida en diferentes individuos puede tomar diferentes valores. Al mismo tiempo, una variable, en estadística, también es un atributo o carácter cualitativo que no es susceptible de medición.

Solución:

Algunas de las variables que podría tener en cuenta son:

a) Incidentes de agresión
b) Tipo de agresión
c) Lugar de la agresión
d) Vinculo del agresor
e) Hubo denuncia

Las **variables** se clasifican en dos grandes grupos:

Variables cualitativas*:* Son aquellas en las cuales el atributo es una *cualidad* del elemento se expresan generalmente mediante palabras.

Variables cuantitativas: Son aquellas que se expresan mediante una cantidad.

Las **variables cualitativas** a su vez se clasifican en:

a. **Nominal**: en este tipo de **variable** el atributo es una palabra que no se puede ordenar, por ejemplo, el color de un auto puede asumir distintos valores (palabras) rojo, azul, gris, negro, blanco, etc. Pero estos valores no se pueden ordenar.

b. **Ordinal**: en este tipo de **variable** el atributo igualmente es una palabra, pero se puede establecer un orden o intensidad del atributo, por ejemplo, el grado escolar, puede asumir los valores: primaria, secundaria, bachillerato, licenciatura, postgrado y todos ellos son palabras.

Las **variables cuantitativas** pueden ser:

a) **Discreta**: Los valores de una **variable cuantitativa** son necesariamente números, si estos valores numéricos solo pueden ser enteros, se dice que la **variable** es **discreta**. Por ejemplo, la **variable** número de hijos en una familia, sólo puede asumir valores enteros, 0, 1, 2, … dado que no existen pedazos de hijos.

b) **Continua**: Si los valores que toma la **variable** incluyen valores decimales se dice que la **variable** es **continua,** por ejemplo, la **variable** estatura, puede asumir un rango de valores que incluyen una parte entera y una decimal.

5.2 Ingresa a http://xurl.es/atnb7 lee el documento y elabora una encuesta sobre el tema de violencia hacia las mujeres que incluya tres variables de tipo cualitativo y tres de tipo cuantitativo.

Solución:

Algunas de las preguntas de la encuesta podrían ser:

A. Cualitativas
    1. ¿has sido agredida por ser mujer?
    2. En caso afirmativo ¿quién te agredió?
    3. ¿En qué lugar has sido agredida?

B. Cuantitativas
    1. ¿Cuántas veces has sido agredida?
    2. ¿Cuántas mujeres cercanas a ti han sido agredidas?
    3. ¿Cuántas personas te apoyaron cuando fuiste agredida?

5.3 Las siguientes preguntas fueron tomadas del cuestionario que aplicará el Inegi en el censo de población y vivienda 2020 indica el tipo de variable de cada una de estas. http://xurl.es/i5k3w

1) ¿Cuántas personas viven en la vivienda?
2) ¿La vivienda cuenta con drenaje?
3) ¿Qué religión profesan?
4) Nivel y grado de escolaridad
5) Número de hijos nacidos vivos

Solución:

1) Cuantitativa discreta (la respuesta a la pregunta es un número entero)
2) Cualitativa nominal (la respuesta es una palabra que no se puede ordenar)
3) Cualitativa nominal (la respuesta es una palabra que no se puede ordenar)
4) Cualitativa ordinal (la respuesta es una palabra que se puede ordenar)
5) Cuantitativa discreta (la respuesta a la pregunta es un número entero)

5.4 Realiza el ejercicio interactivo que sobre variables estadísticas se plantea en http://xurl.es/xeaz0

Situación 6. Restaurante.

María y Ernesto son dueños de un restaurante y quieren mejorar el servicio que proporcionan a sus clientes, así como aspectos de la administración, para esto deciden realizar una encuesta de satisfacción para de este modo identificar las fallas que tienen en la atención. Han decidido que las variables que van a considerar son:

a) Calidad del servicio
b) Tiempo de espera para ser atendido
c) Distancia a la que viven
d) Postre preferido
e) Calidad de los alimentos
f) Precio

6.1 ¿Qué sugerencias darías para medir las variables y recolectar la información?

La **medición** puede definirse como la asignación de números a objetos y eventos de acuerdo con ciertas reglas; la manera como se asignan esos números determina el tipo de **escala de medición**.

Escalas de medición

1. **Nominal:** escala de medición que únicamente cuantifica la presencia o ausencia del atributo.
2. **Ordinal:** en este tipo de escala se puede identificar la intensidad del atributo, ya que se puede establecer un orden, en esta escala, se habla de primero, segundo, tercero.
3. **Intervalo:** en esta escala el valor del cero no es absoluto, sino un cero arbitrario: no refleja ausencia de la magnitud medida, por lo que las operaciones aritméticas de multiplicación y división no son apropiadas.
4. **Razón:** corresponde al nivel de medición más completo. Tiene las mismas propiedades que la escala de intervalo, y además posee el cero absoluto. Aquí el valor del cero no es arbitrario, pues representa la ausencia total de la magnitud que se está midiendo.

Para distinguir las **escalas de intervalo** y de **razón** se debe identificar si el cero en la escala es arbitrario, en caso afirmativo, se trata de una **escala de intervalo** si no lo es sería de **razón**.

Solución:

El proceso de medición consiste en la asignación de un valor a la variable que se va a cuantificar, este valor depende de la escala de medición que se use. Las escalas de medición van de la más simple, que es la nominal, a la más compleja, la de razón.

A continuación, se explica el uso de cada una de ellas:

a. **Nominal**: se utiliza solo el conteo ya que no se puede establecer una intensidad entre los distintos valores, ejemplo: el género asume los valores, masculino o femenino y no podemos decir que por ejemplo femenino es más sexo que masculino, ya que no tiene sentido porque no se puede ordenar el valor de la variable.

b. **Ordinal**: en esta escala la variable también se cuantifica a través del conteo, pero en este caso se puede establecer un orden, ejemplo: nivel académico está variable puede asumir los valores (primaria, secundaria, bachillerato, licenciatura, maestría o doctorado) los cuales se pueden ordenar, aunque en este caso la magnitud en que se cuantifica la variable no es uniforme no podemos decir que alguien que estudió secundaria es lo doble de primaria, sólo podemos decir que estudio más.

c. **Intervalo**: está escala asume las características de las anteriores pero ahora se le agrega la característica de que el patrón de medida es uniforme, sin embargo, el cero es un valor arbitrario, ejemplo: la temperatura expresada en °C, puede asumir diversos valores, se pueden ordenar ya que 10°C es mayor que 3°C, la diferencia entre 6°C y 5°C es la misma que entre, por ejemplo, 27°C y 26°C, sin embargo 0°C no mide la ausencia total de temperatura la cual correspondería al cero absoluto que es de -273°C.

d. **Razón**: al igual que la anterior asume las características de las escalas mencionadas incluyendo la de intervalo y además posee la característica de que el cero no es un valor arbitrario, sino que refleja la ausencia total de la variable.

6.2 Identifica aquellas variables que sean de tipo nominal u ordinal justifica tu respuesta.

Solución:

Variable	Escala	Justificación
Calidad del Servicio	Ordinal	En la calidad del servicio se puede notar que la variable asume valores que se pueden ordenar, la calidad puede ser: mala, regular, buena, excelente; lo que implica una intensidad en la variable. Sin embargo, no se puede establecer que exista una diferencia uniforme por ejemplo entre mala y regular o regular y buena.
Postre preferido	Nominal	Esta variable puede asumir distintos valores pero estos no se pueden ordenar, no se puede decir por ejemplo, que *flan* tenga mayor intensidad en la variable que *fresas con crema*.
Calidad de los alimentos	Ordinal	La calidad de los alimentos toma valores diferentes que tienen distinta intensidad, por lo que, se pueden ordenar, pero no hay uniformidad entre los distintos valores que puede asumir la variable.

Tabla 1. (Elaboración propia)

6.3 Identifica aquellas variables que sean de tipo intervalo o de razón justifica tu respuesta.

Solución:

Variable	Escala	Justificación
Tiempo de espera para ser atendido	Razón	El tiempo de espera, se mide en minutos, utiliza una escala que se puede ordenar, hay un patrón uniforme entre cada unidad de medida y el cero no es arbitrario ya que indica la ausencia del atributo.
Distancia a la que viven	Intervalo	En este caso como la distancia esta referida al restaurante el cero es un valor arbitrario, por lo tanto, es una escala de intervalo.
Precio	Razón	En el precio, se puede establecer un orden, la escala es uniforme, y el cero no es arbitrario.

Tabla 2. (Elaboración propia)

Situación 7. Crecimiento bacteriano.

En la práctica cotidiana un químico farmacéutico biólogo utiliza Agar como medio de cultivo para el crecimiento bacteriano. ¿Sabes que es el Agar? ¿Sabes en donde se utiliza además de la cocina? Ingresa a http://xurl.es/kzlwd y lee la información que se presenta.

Al realizar un estudio acerca de Crecimiento bacteriano en distintos tipos de Agar, se identifican las siguientes variables:

a) Tipo de Agar

b) Bacteria

c) Número de bacterias por unidad de área en un tiempo determinado

d) Tiempo

e) Temperatura de crecimiento

f) Apariencia del agar después de un tiempo determinado

g) Grado de crecimiento (bajo, medio, alto)

h) Cambio de color en el agar

7.1 Identifica las variables que cuantificarías con la escala de medición nominal
Solución:

Variable	Tipo	Justificación
Tipo de Agar	Nominal	El tipo de agar asume valores que no se pueden ordenar.
Bacteria	Nominal	El tipo de bacteria no permite el ordenamiento
Apariencia del Agar después de un tiempo determinado	Nominal	La apariencia que presenta el agar no se pude ordenar.
Cambio de color en el Agar	Nominal	Podría asumir los valores: Si o No, los cuales no se pueden ordenar.

Tabla 3. (Elaboración propia)

7.2 Identifica las variables que cuantificarías con la escala de medición ordinal
Solución:

Variable	Escala	Justificación
Grado de crecimiento (bajo,medio,alto)	Ordinal	Los atributos bajo, medio,alto, se pueden ordenar.

Tabla 4. (Elaboración propia)

7.3 Identifica las variables que cuantificarías con la escala de medición de intervalo
Solución:

Variable	Escala	Justificación
Temperatura de crecimiento	Intervalo para °C	La temperatura en °C asume un valor de cero arbitrario.

Tabla 5. (Elaboración propia)

7.4 Identifica las variables que cuantificarías con la escala de medición de razón

Solución:

Variable	**Escala**	**Justificación**
Número de bacterias por unidad de área en un tiempo determinado	Razón	En este caso el cero indica la ausencia total de la variable y no es arbitrario.
Tiempo	Razón	En este caso el cero indica la ausencia total de la variable y no es arbitrario.

Tabla 6. (Elaboración propia)

1.2. Datos agrupados: Tablas de frecuencias, medidas de tendencia central, de dispersión, de posición, coeficiente de variación, histograma, polígono de frecuencias, ojiva

Situación 8. Redes sociales

Un estudiante de bachillerato desea conocer cuál es la red social que más utilizan sus compañeros, para ello pregunta a sus compañeros, ¿cuál es la red social que usan con mayor frecuencia? Los resultados se muestran a continuación

Red Social	
Facebook	卌 卌 卌 ||||
You Tube	卌 卌 卌
Twitter	卌 卌
Google+	卌	
Linkedin			
Instagram	卌 |||

Figura 4. (Elaboración propia)

8.1 ¿Cómo organizarías la información?

Tabla de frecuencias

Las **Tablas de frecuencias** son herramientas de Estadística donde se asocia a cada dato con el número de veces que ocurre.

Tipos de frecuencias

a) **Frecuencia absoluta** es el número de veces que aparece un determinado valor en un estudio estadístico. Se representa f_i

b) **Frecuencia relativa** es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. Se representa por f_r

c) **Frecuencia acumulada** es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado. Se representa por F_i

d) **Frecuencia relativa acumulada** es el cociente entre la frecuencia acumulada de un determinado valor y el número total de datos. Se puede expresar en tantos por ciento. Se representa por F_R

Solución:

La forma más adecuada de organizar la información es a través de una tabla de frecuencias.

8.2 Realiza una tabla de frecuencias para organizar la información

Solución:

Red social	Frecuencia
Facebook	19
You tube	15
Twitter	10
Google+	6
Linkedin	2
Instagram	8
Total	60

Tabla 7. (Elaboración propia)

8.3 Organiza la información en una tabla de frecuencias que contenga frecuencias absolutas, frecuencias acumuladas, frecuencias relativas y frecuencias relativas acumuladas.

Solución:

Red Social	Frecuencia absoluta f_i	Frecuencia acumulada F_i	Frecuencia relativa f_r	Frecuencia relativa acumulada F_R
Facebook	19	19	0.3167	0.3167
You Tube	15	34	0.2500	0.5667
Twitter	10	44	0.1667	0.7334
Google+	6	50	0.1000	0.8334
Linkedin	2	52	0.0333	0.8667
Instagram	8	60	0.1333	1
Total	60			

Tabla 8. (Elaboración propia)

8.4 Interpreta la información de la tabla anterior

Solución:

La mayoría de los encuestados usa Facebook un 31%, 73% usan o Facebook, You Tube o Twitter, las cuales son las redes sociales más usadas.

Situación 9. Número de hermanos.

Juan lee un artículo sobre un estudio en ratas, que muestra que el número de hermanos influye en la personalidad de la rata adulta (ingresa a http://xurl.es/gzmrb para más información), y siente curiosidad por saber si esta característica podría darse en humanos, así que pregunta a sus compañeros cuántos hermanos tienen**.**

Realiza una encuesta con ellos y obtiene los siguientes datos:

0	**1**	**2**	**0**	**1**	**1**	**0**	**3**	**1**	**2**	**2**	**1**	**0**	**1**	**1**
1	**1**	**0**	**2**	**1**	**0**	**1**	**1**	**1**	**1**	**2**	**1**	**3**	**0**	**1**
0	**0**	**1**	**1**	**1**	**1**	**2**	**2**	**0**	**1**	**1**	**1**	**0**	**2**	**3**
4	**0**	**1**	**1**	**2**	**1**	**1**	**2**	**1**	**3**	**0**	**0**	**1**	**1**	**0**

Tabla 9. (Elaboración propia)

9.1 Organiza la información en una tabla de frecuencias que contenga frecuencias absolutas, frecuencias acumuladas, frecuencias relativas y frecuencias relativas acumuladas.

Solución:

Número de hermanos	Frecuencia absoluta f_i	Frecuencia acumulada F_i	Frecuencia relativa f_r	Frecuencia relativa acumulada F_R
0	15	15	0.2500	0.2500
1	30	45	0.5000	0.7500
2	10	55	0.1667	0.9167
3	4	59	0.0667	0.9834
4	1	60	0.0166	1
Total	60			

Tabla 10. (Elaboración propia)

9.2 Interpreta la información de la tabla anterior

Solución:

El número de hermanos más frecuente es de 1, el 75 % de los encuestados tienen a lo más un hermano, los encuestados tienen de 0 a 4 hermanos.

9.3 Investiga ¿Cuál es el número promedio de hijos en México? ¿Consideras que los datos de la tabla anterior son representativos de ese dato?

Solución:

Según el Inegi el promedio de hijos en México es de 2.3 así en la encuesta anterior la mayoría de los estudiantes debería tener un hermano, por lo tanto, los datos son representativos de la población.

Situación 10. Tiempo de traslado

Mariana escucha decir a una de sus compañeras que tarda dos horas para llegar a la escuela y le parece un tiempo excesivo, ella llega a la escuela en 20 min, cree que esto afecta el desempeño académico. Decide encuestar a sus compañeros para saber el tiempo promedio que tardan los alumnos en llegar a la escuela, entrevista a 50 de sus compañeros. Los datos se presentan a continuación.

5	5	10	10	15	15	20	20	25	25
30	30	30	30	35	35	40	40	40	45
45	45	45	45	45	50	50	50	55	55
60	60	60	60	60	60	60	60	60	60
60	60	60	70	70	90	90	90	105	120

Tabla 11. (Elaboración propia)

Observa que existen muchos datos distintos y se pregunta qué estrategia usar para organizar los datos**.**

10.1 ¿Cómo organizarías la información?

Datos agrupados

Cuando se tiene una variable cuantitativa y son más de 50 los datos recolectados o existe una variación muy grande entre ellos lo aconsejable es agruparlos en intervalos y construir una tabla de frecuencias para datos agrupados.

Procedimiento para elaborar una tabla de frecuencia para datos agrupados

1. Calcula el rango o recorrido de los datos con la relación:

$$R = Máx - mín$$

2. Establece el número de clases o intervalos con la siguiente fórmula:

$$k > \frac{\log n}{\log 2}$$

3. Determina el ancho de los intervalos $t = \frac{R}{k}$

4. Forma los intervalos: Para formar la primera clase, se pone como límite inferior de la primera clase el valor menor de los datos y posteriormente se suma a este valor el ancho de clase c obteniendo de esta manera el límite superior de la primera clase, luego se procede a obtener los límites de la clase siguiente y así sucesivamente.

5. Contar los datos que caen dentro de cada intervalo

10.2 Realiza una tabla de frecuencias para datos agrupados

Solución:

1) Calcular el rango

$$R = 120 - 5$$
$$R = 115$$

2) Hallar el número de intervalos

$$k = \frac{\log 50}{\log 2}$$
$$k = 5.64$$

El número de intervalos, debe ser un número entero, en este caso puede ser 5 o 6, elegimos 5 debido a que al calcular el ancho de clase t se obtiene un valor entero que facilita la construcción de los intervalos.

3) Ancho de clase

$$t = \frac{115}{5}$$
$$t = 23$$

4) Construir los intervalos de clase: para construir los intervalos de clase se toma como límite inferior del primer intervalo el dato más pequeño, se usa corchete para indicar que este valor está incluido, posteriormente se suma a este valor el ancho de clase t para hallar el límite superior del intervalo, en este valor también se usará corchete para indicar que está incluido, este valor se usará como límite inferior de la siguiente clase, pero en este caso se usará paréntesis indicando que este valor no está incluido en este intervalo ya que se incluyó en el anterior, el límite superior se construye sumando el ancho de clase y usando corchete para indicar la inclusión del valor en el intervalo, los demás intervalos se construyen del mismo modo, hasta llegar al último que incluirá como límite superior el valor más grande.

$L_i - L_s$	f_i	F_i	f_r	F_R
[05-28]	10	10	0.2	0.2
(28-51]	18	28	0.36	0.56
(51-74]	17	45	0.34	0.90
(74-97]	3	48	0.06	0.96
(97-120]	2	50	0.04	1

Tabla 12. (Elaboración propia)

Situación 11. Índice Nacional de Precios al Consumidor.

Juan escucha decir a su abuelo que la vida es cada día más cara los precios están aumentando constantemente, Juan siente curiosidad por esta afirmación por lo que decide investigar si es correcto lo que dice su abuelo. En la escuela le han hablado del Índice Nacional de Precios al Consumidor (INPC), el cual según le explicó su maestro, es un indicador económico que mide cuánto se han incrementado los precios en un periodo de tiempo.

Ingresa a https://www.inegi.org.mx/temas/inpc/ y descarga la tabla de datos

11.1 ¿Cómo podría mostrar Juan la información a su abuelo para ilustrar que su afirmación es correcta?

Histograma

Es una gráfica que se emplea en datos agrupados para variables cuantitativas continuas; muestra la forma de la distribución de los datos. Se grafica de la siguiente manera en el eje x los intervalos y en el eje y las frecuencias absolutas, es muy importante destacar que tanto el eje y como el x son numéricos. En este tipo de gráfica tanto la altura del rectángulo como su ancho representan valores. Es una gráfica de columnas, en la cual una columna va pegada a la otra, es decir no hay espacio entre columnas.

Polígono de frecuencia

Es una gráfica para datos agrupados cuya variable es cuantitativa continua, es una gráfica lineal, se realiza localizando las coordenadas de las marcas de clase de cada intervalo con su respectiva frecuencia absoluta, una vez localizado cada uno de los puntos estos se unen obteniendo un polígono del cual toma el nombre la gráfica.

Solución:

Una de las formas más adecuadas para presentar información de manera rápida y procesarla en forma sencilla es a través de una gráfica. En este caso podríamos observar la evolución de los precios a través del tiempo.

Los datos se mostrarían a través de un gráfico, el más adecuado en este caso es una gráfica poligonal que ilustraría la evolución de los precios a través del tiempo.

Al descargar los datos y graficar se obtiene el gráfico que se muestra a continuación.

Figura 4. (Elaboración propia)

11.2 ¿Es correcta la afirmación del abuelo de Juan?

Con el gráfico anterior se puede concluir que no es correcta la afirmación del abuelo ya que el índice de precios y cotizaciones (IPC) tiene altibajos en septiembre de 2019, último dato de la gráfica, el (IPC) es similar al de diciembre de 2010 y a partir de enero de 2018 se observa una tendencia a la baja.

Situación 12. Tabaquismo.

De acuerdo con la organización mundial de la salud (OMS). El consumo de tabaco es uno de los principales factores de riesgo de varias enfermedades crónicas, como el cáncer y las enfermedades pulmonares y cardiovasculares. A pesar de ello, su consumo está muy extendido en todo el mundo. Varios países disponen de leyes que restringen la publicidad del tabaco, regulan quién puede comprar y consumir productos del tabaco, y dónde se puede fumar. https://www.who.int/topics/tobacco/es/

En México en menos de dos décadas el número de fumadores se incrementó de 9 a 13 millones de personas y las enfermedades asociadas al tabaquismo matan a más de 53000 personas cada año, es decir, 147 cada día.

Estas defunciones representan 10% de las muertes nacionales. Para mayor información ingresa a: https://www.paho.org/mex/index.php?option=com_content&view=article&id=96:situacion-tabaco-mexico&Itemid=387 lee el texto y da clic en Situación del Tabaco en México "tabaco en cifras"

Medidas de Tendencia Central

Las medidas de tendencia central también son conocidas como promedios, tienen como objetivo principal buscar un dato que represente al conjunto de datos. Es decir, un valor que resuma a todos los datos.

Las medidas de tendencia central son:

a) Media aritmética o promedio
b) Mediana
c) Moda

Dependiendo del tipo de datos o sus características se emplea cualquiera de estas, por ejemplo, si la variable es cualitativa nominal, únicamente se puede usar la moda como medida de tendencia central, si hay valores extremos que afecten a la media se usa la mediana.

12.1 Entre los datos que arroja este estudio se señala que "Difusión del hábito entre los adultos = Total = 8.8% Varones=12.9% Mujeres= 4.7% Intervalo de edades años = 18+", ¿Qué significa que la difusión del tabaco entre los adultos es del 8.8%? ¿Cómo consideras que se obtuvo este resultado?

Solución:

Esto quiere decir que entre un 8 y 9% de la población fuman, este resultado se obtuvo tomando una muestra representativa de la población a la cual se le aplicó un cuestionario y con los datos obtenidos se encontró la media o promedio que indica que el 8.8 % de la población fuma.

12.2 También se menciona que el consumo anual de cigarrillos por una persona es de 733, esto quiere decir que un fumador consume en promedio 2.008 cigarrillos diarios ¿qué significa este dato?

Solución:

Este dato significa que la población fuma en promedio entre dos y tres cigarrillos diariamente, este resultado podría parecer bajo, pero si consideramos que dentro de la muestra también se incluyen las personas que no fuman el resultado entonces es más preocupante.

Situación 13. Precio del dólar.

El precio del dólar es muy importante para cualquier nación ya que la mayoría de las importaciones de mercancías se realizan es esta moneda, la tasa de interés a nivel mundial también utilizan como base esta moneda, por lo tanto, para los países es básico tratar de predecir el precio del dólar para de este modo establecer los planes de desarrollo económico.

Ingresa a http://dolarpeso.mx/ descarga la tabla Pronostico Dólar: 2019, 2020, 2021, 2022 y 2023

Medidas de dispersión

Las medidas de dispersión tienen como objetivo medir que tanto se alejan los datos del valor central, es decir, cuanto se alejan de las medidas de tendencia central.

Las medidas de dispersión son el corazón de la estadística, ya que sin variación no habría estadística.

Las medidas de dispersión más comunes son:

a) Rango

b) Varianza

c) Desviación estándar

d) Coeficiente de variación

13.1 Explica qué significado tiene la columna min-máx.

Solución:

Esta columna indica que la predicción del valor del dólar se da dentro de un intervalo, el valor esperado del precio del dólar se encontraría dentro de ese rango. Mientras más grande es el intervalo la seguridad de que el precio del dólar este en ese valor es más alta y mientras más pequeño es el intervalo la confianza de que ese sea el valor del dólar disminuye.

13.2 ¿Cuál consideras que será el precio más probable del dólar en marzo de 2020 según los datos de la tabla?

Solución:

El precio del dólar se encontrará en un rango de 18.65 a 19.42.

13.3 ¿Cuánto podemos esperar que se aleje del precio promedio el valor del dólar en junio de 2021? ¿a qué consideras que se deba esto?

Solución:

El valor promedio es de 18.62, su valor se indica estará en un rango de 18.34-18.90, su variación será de ± 0.28 centavos, este valor, nos indica cuánto se espera que se aleje de la media el precio del dólar, es decir, mide la variación del precio.

Esta variación se puede deber a múltiples factores tanto políticos como económicos los cuales afectan el valor de la moneda.

13.4 Al realizar una predicción sobre el valor del dólar en un futuro ¿qué es más adecuado, dar un intervalo o un solo valor?

Solución:

Lo más adecuado es dar un intervalo ya que de este modo también se está cuantificando la posible variación del valor promedio o esperado.

Situación 14. Marihuana.

Existe un fuerte debate sobre la legalización del uso de la mariguana tanto para fines medicinales como para fines recreativos. En 2015, el Consejo de Europa publicó un informe titulado The European School Survey Project on Alcohol and Other Drugs. Entre otras cuestiones, la encuesta investigó los porcentajes de edad de los jóvenes de 16 años que habían consumido marihuana. Los datos son los resultados de 38 países europeos se presentan en http://xurl.es/6a3ja

14.1 ¿Cuál es el porcentaje promedio de jóvenes de 16 años que han consumido marihuana?

Solución:

Son 56 datos por lo tanto los datos se deben tratar como datos agrupados de acuerdo con el siguiente procedimiento.

1) Calcular el rango

$$R = 42 - 2$$
$$R = 40$$

2) Hallar el número de intervalos

$$k = \frac{\log 56}{\log 2}$$
$$k = 5.8073$$

El número de intervalos, debe ser un número entero, en este caso puede ser 5 ó 6, elegimos 5 debido a que al calcular el ancho de clase *t* se obtiene un valor entero que facilita la construcción de los intervalos.

3) Ancho de clase t

$$t = \frac{40}{5}$$
$$t = 8$$

4) Construir los intervalos de clase de acuerdo con el procedimiento indicado en la situación 11. El valor más pequeño es 2 y el mayor 42 se comenzará en 2 como límite inferior del primer intervalo y se debe terminar en 42 como límite superior del último intervalo con un ancho de clase *t* de 8. La tabla de frecuencias absolutas y acumuladas queda:

$L_i - L_s$	f_i	F_i
$[02-10]$	21	21
$(10-18]$	10	31
$(18-26]$	17	48
$(26-34]$	4	52
$(34-42]$	4	56
Total	56	

Tabla 13. (Elaboración propia)

5) A continuación, se muestra la manera de calcular las medidas de tendencia central. Mediante las siguientes fórmulas:

Medidas de Tendencia Central

Las **medidas de tendencia central** también son conocidas como promedios, tienen como objetivo principal buscar un dato que represente al conjunto de datos.

Las medidas de tendencia central son:

a) Media aritmética o promedio

$$\overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

b) Mediana:

$$Md = L + \left(\frac{\frac{n}{2} - \sum f_i}{f_{Md}} \right)(t)$$

c) Moda:

$$Mo = L + \left(\frac{d_1}{d_1 + d_2} \right)(t)$$

Para el cálculo se usa la tabla de frecuencias y se completan las columnas de acuerdo con las fórmulas.

Se requiere calcular las marcas de clase x_i, así como el producto $f_i x_i$

Para calcular x_i se emplea la fórmula:

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

Para cada intervalo.

A continuación, se ilustra lo mencionado en el párrafo anterior.

No es necesario calcular todas las marcas de clase para cada intervalo, basta con calcular la primera y a esta sumarle el ancho de clase t a cada marca de clase subsecuente para hallar las demás.

$L_i - L_s$	f_i	F_i	x_i	$f_i x_i$
$[02-10]$	21	21	6	126
$(10-18]$	10	31	14	140
$(18-26]$	17	48	22	374
$(26-34]$	4	52	30	120
$(34-42]$	4	56	38	152
Total	56			$\sum = 912$

Tabla 14. (Elaboración propia)

Las medidas de tendencia central son entonces:

a) Media

$$\overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

$$\overline{x} = \frac{912}{56}$$

$$\overline{x} = 16.29$$

b) Mediana

$$Md = L + \left(\frac{\frac{n}{2} - \sum f_i}{f_{Md}} \right)(t)$$

El significado de cada una de las expresiones de la fórmula es el siguiente:

Md = mediana

L = límite inferior del intervalo que contiene a la mediana

El intervalo que contiene a la mediana es aquel donde se ubica la mediana, es decir, donde se halla el 50 % de los datos.

n = número de datos

$\sum f_i$ = frecuencia acumulada anterior al intervalo que contiene a la mediana

f_{Md} = frecuencia absoluta del intervalo que contiene a la mediana

t = ancho de clase

Para nuestro ejercicio es:

$$Md = 10 + \left(\frac{\frac{56}{2} - 21}{10} \right)(8)$$

$$Md = 10 + \left(\frac{28 - 21}{10} \right)(8)$$

$$Md = 10 + \left(\frac{7}{10} \right)(8)$$

$$Md = 10 + 5.6$$

$$Md = 15.6$$

c) Moda

$$Mo = L + \left(\frac{d_1}{d_1 + d_2} \right)(t)$$

El significado de cada de las expresiones de la fórmula es el siguiente:

Mo = moda

L = límite inferior del intervalo que contiene a la moda

El intervalo que contiene a la moda es aquel que tiene mayor frecuencia absoluta.

d_1 = diferencia entre la frecuencia absoluta del intervalo que contiene a la moda y la frecuencia absoluta del intervalo anterior

d_2 = diferencia entre la frecuencia absoluta del intervalo que contiene a la moda y la frecuencia absoluta del intervalo posterior

t = ancho de clase

Para nuestro ejercicio es:

El intervalo con mayor frecuencia absoluta es: $[02-10]$ ya que contiene 21 datos que es la frecuencia absoluta más grande, por lo tanto, L es el límite inferior de este intervalo, es decir, 2.

Para calcular d_1, como el intervalo que contiene a la moda es el primero antes de este no hay datos y

$$d_1 = 21-0$$
$$d_1 = 21$$

En el intervalo siguiente hay 10 datos, por lo tanto:

$$d_2 = 21-10$$
$$d_2 = 11$$

Sustituyendo en la fórmula:

$$Mo = 2+\left(\frac{21}{21+11}\right)(8)$$
$$Mo = 2+\left(\frac{21}{32}\right)(8)$$
$$Mo = 2+5.25$$
$$Mo = 7.25$$

14.2 ¿Cuál es el coeficiente de variación de los datos? ¿qué significa este valor?

Para hallar el coeficiente de variación se requiere calcular la varianza y la desviación estándar como se indica a continuación.

Para ello es necesario continuar con la tabla de frecuencias y agregarle las columnas que requiere la fórmula de la varianza.

$L_i - L_s$	f_i	F_i	x_i	$f_i x_i$	$x_i - \bar{x}$	$f_i(x_i - \bar{x})$	$f_i(x_i - \bar{x})^2$
$[02-10]$	21	21	6	126	-10.29	-216.09	2223.57
$(10-18]$	10	31	14	140	-2.29	-22.9	52.44
$(18-26]$	17	48	22	374	5.71	97.07	554.27
$(26-34]$	4	52	30	120	13.71	54.84	751.86
$(34-42]$	4	56	38	152	21.71	86.84	1885.30
Total	56			$\sum = 912$			$\sum = 5467.44$

Tabla 15. (Elaboración propia)

Medidas de dispersión

Las **medidas de dispersión** tienen como objetivo medir que tanto se alejan los datos del valor central, es decir, cuanto se alejan de las medidas de tendencia central. Las medidas de dispersión son el corazón de la estadística.

Las medidas de dispersión más comunes son:

a) Rango:

$$\mathrm{R} = \mathrm{M} - \mathrm{m}$$

b) Varianza:

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

c) Desviación estándar:

$$s = \sqrt{s^2}$$

d) Coeficiente de variación:

$$C.V. = \frac{s}{\bar{x}}$$

Sustituyendo los valores en la fórmula:

a) Varianza:

$$s^2 = \frac{\sum_{i=1}^{n} f_i \left(x_i - \overline{x}\right)^2}{n-1}$$

$$s^2 = \frac{5467.44}{56-1}$$
$$s^2 = 99.408$$

b) Desviación estándar

$$s = \sqrt{99.408}$$
$$s = 9.97$$

c) Coeficiente de variación

$$C.V. = \frac{s}{\overline{x}}$$

$$C.V. = \frac{9.97}{16.29}$$
$$C.V. = 0.61$$

Esto significa que la dispersión de los datos es de 61%, es decir, los datos están muy dispersos.

14.3 Realiza un histograma que represente los datos e interprétalo.

Histograma

Es una gráfica que se emplea en datos agrupados para **variables cuantitativas continuas**; muestra la forma de la distribución de los datos. Se grafica de la siguiente manera en el eje *x* los intervalos y en el eje *y* las frecuencias absolutas, es muy importante destacar que tanto el eje *y* como el *x* son numéricos. En este tipo de gráfica tanto la altura del rectángulo como su ancho representan valores.

Es una gráfica de columnas, en la cual una columna va pegada a la otra, es decir, no hay espacio entre columnas.

Solución:

El histograma es un gráfico de rectángulos, todos los rectángulos tienen la misma base, su base tiene como longitud el ancho del intervalo y su altura es la frecuencia del intervalo. Los intervalos se grafican en el eje *x* mientras que las frecuencias en el eje *y*.

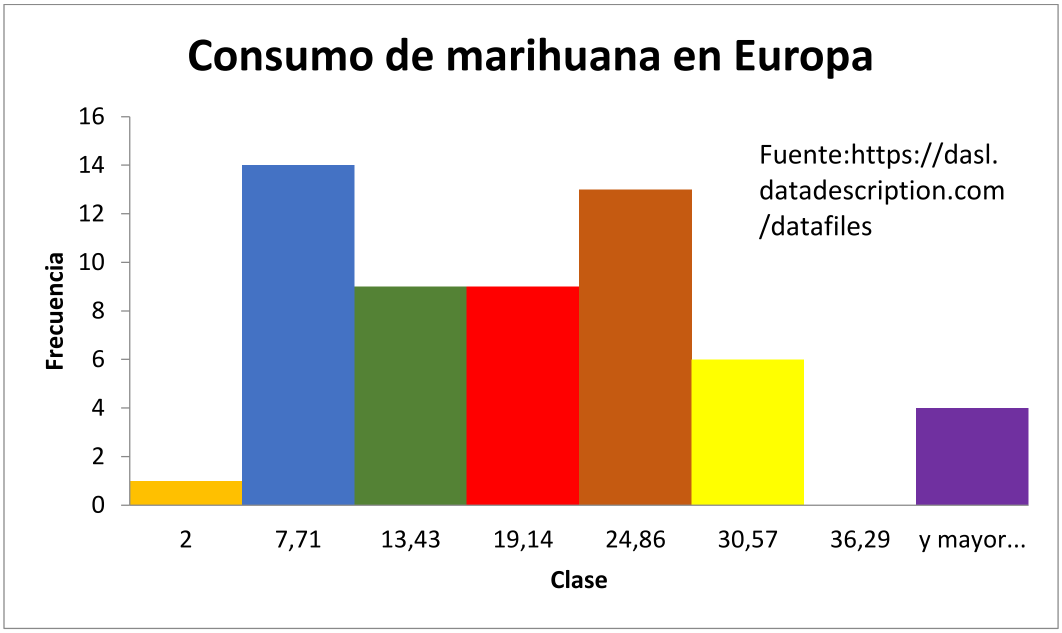


Figura 5. Consumo de Marihuana (Elaboración propia)

1.3. Datos bivariados: Tabulación, representación gráfica de datos de dos variables, correlación lineal, regresión lineal.

Situación 15. Nutrición.

El primer año de vida es la etapa de mayor crecimiento de los niños. Su medición bimensual permite a los padres y al personal de salud saber si el niño tiene un crecimiento y desarrollo óptimos. Algunas enfermedades pueden afectar el crecimiento, siendo una señal para acudir al médico, con el fin de descartar alguna enfermedad.
Investiga en la siguiente liga http://xurl.es/l6fco página 38 y 39, cual es el peso normal de un niño en los doce primeros meses de vida.

15. 1 ¿Consideras que existe una relación entre la edad y el peso del niño? ¿De qué tipo?

Solución:

Si existe una relación entre la edad y el peso del niño y el peso del niño y esta relación es directamente proporcional a mayor edad mayor peso del niño.

15. 2 Copia los datos de la tabla que se muestra en http://xurl.es/l6fco página 39, indica cuales son las variables y gráfica los datos por medio de un diagrama de dispersión o nube de puntos, finalmente realiza una breve descripción de la información en la gráfica.

Variable
Es una magnitud que puede tomar diversos valores.

Las **variables** se pueden clasificar en:

a) **variable independiente**
b) **variable dependiente**

Solución:

Las variables son edad y peso, al copiar los datos se obtiene la siguiente tabla:

Edad (meses)	Al nacer	1	2	3	4	5	6	7	8	9	10	11	12
Peso (kg)	3.2	4.2	5.1	5.8	6.4	6.9	7.3	7.6	7.9	8.2	8.5	8.9	9.4

Tabla 16. Nutrición niñas (IMSS). (Elaboración propia)

Diagrama de dispersión

Los datos se localizan en el plano cartesiano a la **variable independiente** se le asigna el eje x, mientras que a la **variable dependiente** le corresponde el eje y. A este gráfico se le llama **diagrama de dispersión** y tiene como fin observar el tipo de relación que existe entre los puntos.

Se gráfica edad contra peso y se obtiene la gráfica que se muestra a continuación.

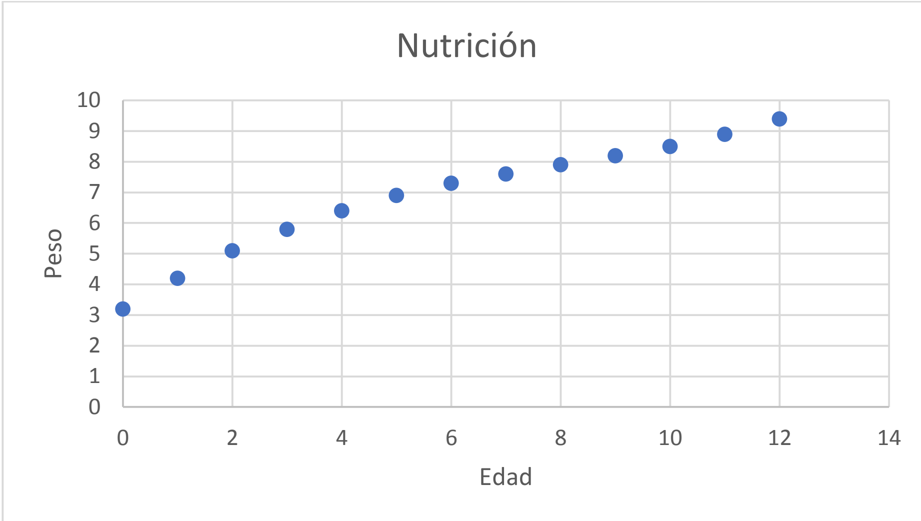


Figura 6. Diagrama de dispersión nutrición (Elaboración propia)

Descripción de la gráfica:

Se puede observar que al aumentar la edad aumenta el peso, esto ocurre en todos los puntos. El peso más pequeño es de 3 kg para un recién nacido y el mayor es de 9.5 kg a los 12 meses.

15.3 Reflexiona si una de las variables depende de la otra e indica ¿cuál es?

Variables

La variable que se va a predecir se llama **dependiente**, y la variable que se emplea para predecir se nombra como **independiente**. Antes de cualquier análisis se identifica el tipo de variable.

Solución:

Al analizar los datos se observa que el peso depende de la edad a mayor peso mayor edad y a menor edad menor peso. En este caso lo que se va a predecir es el peso por lo tanto es la variable dependiente y la variable que predice es la edad la cual es la variable independiente.

15.4 ¿Consideras que podría existir una relación entre los datos de tipo lineal? ¿Cómo podrías saber que efectivamente existe una relación de este tipo entre los datos**?**

Datos bivariados

En ocasiones se miden dos variables a un mismo individuo, por ejemplo, peso y estatura. Es decir, para un mismo individuo se observarán dos características, a este tipo de **datos** se les llama **bivariados**.

Si las dos variables de las que provienen los datos son cuantitativas resulta interesante establecer si existe una relación entre estas variables y de qué tipo es.

Esta relación puede ser: lineal, cuadrática, polinomial, trigonométrica, exponencial, etc.

Así pues, en los **datos bivariados** deseamos conocer si existe una relación funcional entre las variables (recta de regresión), con el fin de que se pueda predecir el valor de una respecto a la otra.

Coeficiente de correlación

Cuando se tienen datos bivariados debemos establecer si existe una relación lineal entre estos, esto se puede cuantificar a través del **coeficiente de correlación lineal** también llamado de **Pearson**.

Solución:

Al observar el diagrama de dispersión se puede notar que para el periodo considerado existe una tendencia lineal en los datos. Para poder establecer con mayor precisión si existe una relación lineal entre las variables se calcula el coeficiente de correlación lineal o de Pearson. El coeficiente de Pearson es un índice, que independiente de las unidades de las variables, mide el grado de relación entre las dos variables.

15.5. ¿Cuál es el valor del coeficiente de correlación?

Coeficiente de correlación

Cuando se tienen **datos bivariados** debemos establecer si existe una **relación lineal** entre estos, esto se puede cuantificar a través del **coeficiente de correlación lineal** también llamado de **Pearson**.

El **coeficiente de correlación lineal** o de **Pearson** se calcula mediante la siguiente relación:

$$r = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \sum (y_i - \overline{y})^2}}$$

En donde:

$$\sum (x_i - \overline{x})(y_i - \overline{y})$$

Es un valor que indica el grado de variación conjunta de dos variables respecto a sus medias, se conoce como **covarianza.**

La **covarianza** es un dato básico para determinar si existe una **relación lineal** entre dos variables y por lo tanto es determinante para el cálculo del **coeficiente de Pearson.**

Este coeficiente toma valores en el intervalo $\left[-1 \leq r \leq 1\right]$ y de su valor se puede decir lo siguiente:

1) Si $r = 1$ se dice que existe una correlación total (perfecta) lineal de tipo directa.

2) Si $r = -1$ se dice que existe una correlación total (perfecta) lineal de tipo inversa.

3) Si $r = 0$ la correlación es nula

4) Los valores de $r = -1, 1$ ó 0 son muy difíciles de encontrar en la práctica por ello se dice que valores que se acerquen a -1 ó 1 indican que la correlación lineal entre las variables es muy fuerte.

Solución:

Para calcular el coeficiente de correlación lineal se usa la fórmula:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Entonces para su cálculo se construye una tabla como la de abajo con las columnas que se indican, cuyas sumas o totales se usan para establecer el valor de *r.*

Edad (meses)	Peso (kg)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
0	3.2	-6	-3.68	22.08	36	13.5424
1	4.2	-5	-2.68	13.4	25	7.1824
2	5.1	-4	-1.78	7.12	16	3.1684
3	5.8	-3	-1.08	3.24	9	1.1664
4	6.4	-2	-0.48	0.96	4	0.2304
5	6.9	-1	0.02	-0.02	1	0.0004
6	7.3	0	0.42	0	0	0.1764
7	7.6	1	0.72	0.72	1	0.5184
8	7.9	2	1.02	2.04	4	1.0404
9	8.2	3	1.32	3.96	9	1.7424
10	8.5	4	1.62	6.48	16	2.6244
11	8.9	5	2.02	10.1	25	4.0804
12	9.4	6	2.52	15.12	36	6.3504
$\sum = 78$	$\sum = 89.4$			85.2	182	41.8232
$\bar{x} = 6$	$\bar{y} = 6.88$					

Tabla 17. (Elaboración propia)

$$r = \frac{85.2}{\sqrt{[182][41.8232]}}$$

$$r = \frac{85.2}{\sqrt{7611.8224}}$$

$$r = \frac{85.2}{87.2457}$$

$$r = 0.9765$$

El valor 0.9765 indica que existe una fuerte correlación lineal de tipo directa entre las variables.

15.6. ¿Si existe una correlación lineal, halla la recta que modela la relación entre las variables?

Recta de mejor ajuste o recta de regresión lineal

Una vez que se establece que hay una **correlación lineal** entre las variables, determinando el coeficiente de **Pearson**, se puede hallar el modelo lineal al que se ajustan los datos. A la recta que se obtiene se le llama **recta de mejor ajuste** y el procedimiento se conoce como **regresión lineal**. La **regresión lineal** plantea el siguiente **modelo lineal**:

$$\hat{y} = a + bx$$

$\hat{y}$ Significa, que el valor que se obtiene es un valor de predicción o ajustado.
En donde:

$$b = \frac{\sum(x_i y_i) - n\,\bar{x}\,\bar{y}}{\sum(x_i^{\,2}) - n\,\bar{x}^2}$$

Y $\quad a = \bar{y} - b\,\bar{x}$

Solución:

Para la recta de mejor ajuste se usa

$$\hat{y} = a + bx$$

Para calcular b se usa la siguiente tabla:

Edad (meses)	Peso (kg)	$x_i y_i$	x_i^2
0	3.2	0	0
1	4.2	4.2	1
2	5.1	10.2	4
3	5.8	17.4	9
4	6.4	25.6	16
5	6.9	34.5	25
6	7.3	43.8	36
7	7.6	53.2	49
8	7.9	63.2	64
9	8.2	73.8	81
10	8.5	85	100
11	8.9	97.9	121
12	9.4	112.8	144
78	89.4	621.6	650
6	6.88		

Tabla 18. Recta de mejor ajuste (Elaboración propia)

$$b = \frac{\sum(x_i y_i) - n\,\bar{x}\,\bar{y}}{\sum(x_i^2) - n\,\bar{x}^2}$$

$$b = \frac{621.6 - (13)(6)(6.88)}{650 - (13)(36)}$$

$$b = \frac{621.6 - 536.64}{650 - 468}$$

$$b = \frac{84.96}{182}$$

$$b = 0.4668$$

Para el valor de a

$$a = \overline{y} - b\,\overline{x}$$

$$a = 6.88 - (0.4668)(6)$$

$$a = 4.0792$$

Por lo tanto, la recta de mejor ajuste es:

$$\hat{y} = a + bx$$

$$\hat{y} = 4.0792 + 0.4668x$$

15.7. ¿Cuál será el peso probable de una niña de 15 meses?

Solución:

Para conocer cuál será el peso de una niña de 15 meses, sabemos que $x = 15$

$$y = 4.0792 + (0.4668)(15)$$

$$y = 4.0792 + 7.002$$

$$y = 11.0812 \text{ kg}$$

15.8. ¿Cuántos meses se espera que tenga una niña que pesa 7 kg?

Solución:

Al sustituir y despejar en la recta de mejor ajuste.

$$\hat{y} = 4.0792 + 0.4668x$$

$$7 = 4.0792 + 0.4668x$$

$$7 - 4.0792 = 0.4668x$$

$$2.9208 = 0.4668x$$

$$\frac{2.9208}{0.4668} = x$$

$$6.257 \text{ meses} = x$$

Situación 16. Compras en línea.

Las compras en línea se han vuelto cada vez más comunes, sin embargo, existe cierto sector de la población que no se siente seguro al realizar una compra por este medio, los jóvenes tienden a hacer un mayor uso de la tecnología mientras que las personas mayores tienen más desconfianza. Ingresa a http://xurl.es/48h77 en donde se muestran datos acerca del tiempo que pasa una persona haciendo compras en línea y la edad que tiene.

16.1 Descarga los datos y realiza un diagrama de dispersión

Solución:

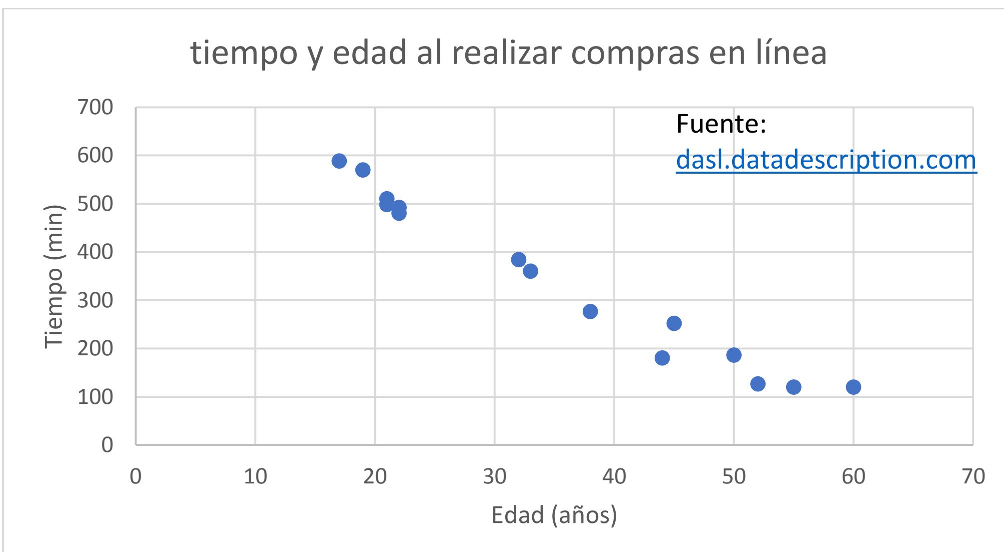


Figura 7. Compras en línea (Elaboración propia)

16.2 ¿Consideras que existe una relación entre las compras en línea y la edad de las personas? ¿De qué tipo?

Solución:

Con base en el gráfico de nube de puntos o dispersión se observa que existe una relación entre la edad de las personas y el tiempo que pasan realizando compras en línea, está relación es inversamente proporcional, a menor edad mayor tiempo realizando compras en línea mientras que a mayor edad menor tiempo para esta actividad.

16.3 Identifica cuál es la variable independiente y cuál la dependiente.

Solución:

La variable independiente es la edad y la dependiente el tiempo que pasan realizando compras en línea.

16.4 ¿Cómo podrías determinar la intensidad de la relación entre las variables?

Solución:

Calculando el coeficiente de correlación lineal

16.5 Si existe una relación lineal de tipo inversa ¿cómo debe ser el coeficiente de correlación lineal?

Solución:

Como la relación entre las variables es inversamente proporcional el coeficiente de correlación debe ser negativo.

16.6 Calcula el coeficiente de correlación lineal

Solución:

El valor del coeficiente de correlación lineal es $r = -0.9838$

16.7 Encuentra la ecuación de la recta que modela la situación

Solución:

La ecuación de la recta que modela la situación es $\hat{y} = -11.503x + 750.02$

16.8 ¿Cuántos minutos se espera que haga compras en línea una persona de 37 años?

Solución:

$$\hat{y} = -11.503x + 750.02$$
$$\hat{y} = -11.503(37) + 750.02$$
$$\hat{y} = -425.611 + 750.02$$
$$\hat{y} = 324.409$$

16.9 Si una persona estuvo 275 minutos haciendo compras en línea ¿Qué edad se espera que tenga?

$$\hat{y} = -11.503x + 750.02$$
$$275 = -11.503x + 750.02$$
$$275 - 750.02 = -11.503x$$
$$-475.02 = -11.503x$$
$$\frac{-475.02}{-11.503} = x$$
$$41.29 = x$$

EXAMEN DE AUTOEVALUACIÓN

Situación 1. Comer en la escuela.

Los siguientes datos fueron obtenidos de una encuesta a 50 estudiantes de la Escuela Nacional Preparatoria, acerca de cuánto gastaban para comer en la escuela.

0	15	20	30	38
0	15	20	30	38
0	17	25	30	38
0	17	25	35	40
10	18	25	35	40
10	18	25	35	45
10	18	28	35	45
15	20	28	35	50
15	20	28	35	60
15	20	30	35	70

Tabla 19. Dinero gastado para comer en la escuela (Elaboración propia)

1. **¿De qué tipo es la variable cantidad de dinero gastada para comer?**
   A) Cualitativa nominal
   B) Cualitativa ordinal
   C) Cuantitativa discreta
   D) Cuantitativa continua

2. **¿Qué escala de medición se empleó para medir los datos?**
   A) Nominal
   B) Ordinal
   C) Intervalo
   D) Razón

3. **¿Cuánto gastan en promedio para comer los estudiantes?**

   A) 25
   B) 26.12
   C) 30
   D) 35.50

4. **¿Cuál es el valor de la desviación estándar?**
   A) 13.91
   B) 14.67
   C) 193.39
   D) 215.21

5. ¿Cuánto vale el coeficiente de variación?

A) 0.5617
B) 0.4633
C) 1.8801
D) 1.9094

Situación 2. Nutrición.

La siguiente tabla muestra el peso normal de una niña desde recién nacida hasta un año, según datos tomados de la guía de salud del IMSS.

Edad (meses)	**Peso (kg)**
Al nacer	3.2
1	4.2
2	5.1
3	5.8
4	6.4
5	6.9
6	7.3
7	7.6
8	7.9
9	8.2
10	8.5
11	8.9
12	9.4

Tabla 16. Nutrición niñas (IMSS)
http://www.imss.gob.mx/sites/all/statics/salud/guias_salud/2018/guia-salud-ninas-ninos-2018.pdf

6. **La variable peso se identifica como:**
   A) nominal
   B) discreta
   C) independiente
   D) dependiente

7. **¿Cuál es el valor coeficiente del coeficiente de correlación lineal?**
   A) 0.9537
   B) 0.9765
   C) 0.8625
   D) 0.8732

8. **¿cuál es la recta de regresión?**
   A) $y = 4.0681x - 0.4681$
   B) $y = -0.4681x + 4.0681$
   C) $y = 0.4681x + 4.0681$
   D) $y = 4.0681x + 0.4681$

9. **¿Cuál será el peso esperado de una niña de 15 meses?**
   A) 10.52
   B) 11.09
   C) 11.59
   D) 13.01

10. **¿Cuántos meses se esperaría que tenga una niña que pesa 7 kg?**

    A) 5.8
    B) 6.0
    C) 6.3
    D) 6.4

RESPUESTAS DE EXAMEN DE AUTOEVALUACIÓN

Número de Pregunta	1	2	3	4	5	6	7	8	9	10
Respuesta	D	D	B	B	A	D	B	C	B	C

REFERENCIAS BIBLIOGRAFICAS

Batanero, C. (2001). *Didáctica de la estadística*. Granada, España: Grupo de Investigación en Educación Estadística Departamento de Didáctica de la Matemática Universidad de Granada.

Córdoba, M. (2003). *Estadística: Descriptiva e Inferencial*. Lima, Perú: Distribuidora Imprenta Editorial, pp. 3-9

Orlandoni, G. (2010). *Escalas de medición en Estadística*. TELOS. Revista de Estudios Interdisciplinarios en Ciencias Sociales, Vol. 12 No. 2, 243-246.

Sánchez, E. (2013). *Elementos de estadística y su didáctica a nivel bachillerato*. México, D.F., Secretaría de Educación Pública pp-14-32

INEGI. (2015). *Encuesta intercensal 2015. Principales resultados*. Marzo 12, 2019, de INEGI recuperado de sitio web: https://www.inegi.org.mx/contenidos/programas/intercensal/2015/doc/eic_2015_presentacion.pdf

INEGI. (2019). INEGI. (2019). Temas. marzo 18, 2019, de INEGI recuperado de sitio web: https://www.inegi.org.mx/temas/educacion/

IMSS. (2018). IMSS. Guía de salud 2018, mayo 6, 2019, recuperado de sitio web: http://xurl.es/l6fco

INEGI. (2019). INEGI. (2019). Temas. marzo 24, 2019, de INEGI recuperado de sitio web: https://www.inegi.org.mx/temas/natalidad/

Tamayo, J. (2019). Estudio de Hábitos de Usuarios de Internet en México 2019. noviembre 23, 2019, de webmarketingtips recuperado de sitio web: http://xurl.es/szq9f.

Publimetro. (2018). Capitalinos y mexiquenses tardan hasta dos horas en llegar a su trabajo, marzo 22, 2019, de Publimetro recuperado de sitio web: http://xurl.es/dtpeb

Galván, E. (2019). Dinero, noviembre 17, 2019, de La Jornada recuperado de sitio web: http://xurl.es/rr45s.

INEGI. (2019). México en Cifras, agosto 16, 2019, de INEGI, recuperado de sitio web: https://www.inegi.org.mx/app/areasgeograficas/

Khawaja, A. (2018). Día Mundial sin Tabaco: ¿qué países fuman más y menos en el mundo? (y en qué lugar se sitúan los de América Latina), noviembre 16, 2019, de BBC News, recuperado de sitio web: https://www.bbc.com/mundo/noticias-44311572

Noticieros Televisa. (2019). El mexicano consume 163 litros de refresco al año, en promedio: Estudio, noviembre 11, 2019, de Noticieros Televisa, recuperado de sitio web: https://noticieros.televisa.com/historia/consumo-refresco-mexico-datos-riesgos/

Federación Mexicana de Diabetes AC. (2018). Estadísticas en México. octubre 19, 2019, de Federación Mexicana de Diabetes AC, recuperado de sitio web: http://fmdiabetes.org/estadisticas-en-mexico/

Alegría, A. (2019). En el Buen Fin, seis de cada 10 combinaron compras en tiendas y online, diciembre 11, 2019, de La Jornada, recuperado de sitio web: https://www.jornada.com.mx/ultimas/economia/2019/12/06/seis-de-cada-10-consumidores-compraron-en-tiendas-fisicas-y-online-en-el-buen-fin-9277.html

Sánchez, A. (2019). Producción automotriz mexicana registra su mayor caída en 16 años, diciembre 13, 2019, de El financiero, recuperado de sitio web: https://elfinanciero.com.mx/empresas/produccion-de-autos-en-mexico-tiene-su-peor-noviembre-en-16-anos-cae-13

Goal. (2019). Cuántos goles lleva Lionel Messi en toda su carrera, diciembre 11, 2019, de Goal.com, recuperado de sitio web: https://www.goal.com/es-mx/noticias/cuantos-goles-lleva-lionel-messi-en-toda-su-carrera-entre/1janysxr71cok1xgydphg1yxof

Instituto nacional de las Mujeres. (2013). Violencia contra las mujeres, noviembre 14, 2019, de Instituto Nacional de las Mujeres, recuperado de sitio web: http://xurl.es/atnb7

INEGI. (2019). Censo de Población y Vivienda 2020. Metodología en Consulta Pública Ficha Técnica del Cuestionario Básico, noviembre 26, 2019, de INEGi, recuperado de sitio web: http://xurl.es/i5k3w

Superprof (2018). Ejercicios interactivos de variables estadísticas, octubre 25, 2019, de Superprof, recuperado de sitio web: http://xurl.es/xeaz0

Universidad de Granada. (2015). Preparación de medios de cultivo, octubre 26, 2019, de Universidad de Granada, recuperado de sitio web: http://xurl.es/kzlwd

Nadal, R. (2008). Las ratas con más hermanos son menos ansiosas de mayores, noviembre 5, 2019, de Universidad Autónoma de Barcelona, recuperado de sitio web: http://xurl.es/gzmrb

INEGI. (2018). Índice Nacional de Precios al Consumidor (INPC), noviembre 12, 2019, de INEGI, recuperado de sitio web: https://wwwinegi.org.mx/temas/inpc/

Organización Mundial de la Salud. (2018). Tabaquismo, octubre 27, 2019, de Organización Mundial de la Salud, recuperado de sitio web: https://www.who.int/topics/tobacco/es/

OPS/OMS México. (2008). Situación del Tabaco en México, noviembre 17, 2019, de OPS/OMS México, recuperado de sitio web: https://www.paho.org/mex/index.php?option=com_content&view=article&id=96:situacion-tabaco-mexico&Itemid=387

La agencia pronóstico económico. (2019). Pronóstico dólar 2020, 2021, 2022, 2023. Precio del dólar hoy, noviembre 15, 2019, de La agencia pronóstico económico, recuperado de sitio web: http://dolarpeso.mx/

DASL. (2015). Datafiles, diciembre 11, 2019, de DASL, recuperado de sitio web: http://xurl.es/6a3ja

DASL (2018). Online Shopping, diciembre 10, 2019, de DASL, recuperado de sitio web: http://xurl.es/48h77

UNIDAD II PROBABILIDAD PARA ESTUDIAR LA INCERTIDUMBRE

Objetivos específicos.

El alumno:

- Desarrollará el pensamiento probabilístico, al plantear proyectos de investigación que permitan resolver problemas en situaciones donde interviene el azar o la incertidumbre con el fin de estimar la probabilidad de ocurrencia de eventos, mediante la realización de experimentos aleatorios, el uso de recursos tecnológicos para su simulación, y la aplicación de los resultados de la probabilidad clásica.

- Explicará los fenómenos aleatorios cuya probabilidad puede estimarse mediante las distribuciones aleatorias binomial y normal para medir la incertidumbre, a través del cálculo de probabilidades, por medio de tablas, simulaciones o analíticamente.

2.1. Fenómenos determinísticos y aleatorios

Situación 1. Aula de clases.

En el aula de una institución educativa, un profesor les dice a sus alumnos que harán unos experimentos.

1.1. Encontrar las soluciones de la ecuación $x^2 - 9 = 0$, ¿qué tipo de experimento será?

Experimento

Un **experimento** o **fenómeno** es el proceso de obtener una observación, datos o resultados. Se clasifica en:

- **Fenómeno regular o determinístico**. Tiene la característica de que, al llevarse a cabo bajo condiciones específicas, su resultado está determinado de manera única.
- **Fenómeno aleatorio o no determinístico**. Es aquél que al ocurrir o efectuarse bajo las mismas condiciones no siempre da el mismo resultado.

Solución:

Como las soluciones de la ecuación son siempre las mismas y no cambian $(x_1 = 3 \text{ y } x_2 = -3)$, se trata de un experimento determinístico.

1.2. Indicar el número de alumnos que usan un celular de cierta marca, ¿qué tipo de fenómeno será?

Solución:

Se trata de un experimento aleatorio, dado que la respuesta no puede darse con certeza y si se repite bajo las mismas condiciones, la respuesta cambiará.

2.1.1. Espacio muestral de fenómenos aleatorios

1.3. Si el profesor repite el experimento anterior cada año, en un grupo de 40 alumnos, ¿cuál es el espacio muestral de éste?

Espacio Muestral

El **espacio muestral (o probabilístico)** de un experimento aleatorio es el conjunto de todos los posibles resultados de interés del fenómeno. Se denota con **Ω**, **S** o **U**.

Solución:

Puede que ningún alumno tenga un celular de esa marca, o que solo un estudiante posea un celular de ese tipo, o dos, y así sucesivamente, hasta los 40 alumnos. Así el espacio muestral es $\Omega = \{0, 1, 2, 3, \cdots, 40\}$.

2.1.2. Eventos de un experimento: simples, compuestos; nulos, seguros, mutuamente excluyentes, no excluyentes entre sí, independientes y dependientes

1.4. Si el experimento es el número de alumnos que usan un celular de cierta marca en un grupo de 40 alumnos, entonces al subconjunto del espacio muestral $A=\{0, 2, 4, \cdots, 38, 40\}$ se le llama:

> **Evento**
>
> Es cualquier subconjunto de un espacio muestral, también se le conoce como **suceso**, es la mínima unidad de análisis para efectos de cálculo de probabilidades. Se denota con una letra mayúscula.

Solución:

Evento, este caso es un evento compuesto.

1.5. El evento $B=\{5\}$ del anterior experimento, se le conoce como:

> **Tipos de eventos**
>
> Un **evento elemental** o **simple** es el formado con un solo resultado.
>
> Un **evento compuesto** es el formado con dos o más eventos simples.

Solución:

Evento elemental o simple, porque tiene un solo elemento.

1.6. Considerando el experimento de la pregunta 1.4., sea el evento C de que a lo más 40 alumnos tienen celular de una cierta compañía, es decir, $C=\Omega$. Se le llama:

> Un **evento seguro** es el que contiene todos los elementos del espacio muestral.
>
> Un **evento imposible** o **nulo** es el que no tiene elementos de interés para el fenómeno.

Solución:

Evento seguro, porque tiene todos los elementos del espacio muestral.

1.7. Considerando el experimento de la pregunta 1.4., sea el evento *D* en el que al menos 41 alumnos tienen celular de una cierta compañía, se le llama:

Solución:

Evento imposible, porque *D* no tiene elementos, es decir, $D = \varnothing$.

1.8. Nuevamente considerando el experimento de la pregunta 1.4., sea el evento *E* en el que el número de alumnos que tienen celular de una cierta compañía es impar, es decir $E = \{1, 3, 5, \cdots, 37, 39\}$ y sea $F = \{4, 16, 32\}$. ¿Cómo se les llama a los eventos *E* y *F*?

> Dos eventos son **mutuamente excluyentes** si uno y sólo uno de ellos puede tener lugar a la vez. Su intersección es $\varnothing$.
>
> Dos eventos son **no excluyentes entre sí**, cuando suceden los dos al mismo tiempo. Es decir, su intersección es diferente del vacío.

Solución:

Eventos mutuamente excluyentes, porque no tienen elementos en común, es decir su intersección es el vacío.

1.9. Finalmente considerando el experimento y el evento de la pregunta 1.4., entonces los eventos *A* y *F* (pregunta anterior), se llaman:

Solución:

Eventos no excluyentes entre sí, porque tienen elementos en común y $F \subset A$.

2.2. Técnicas de conteo: notación factorial, principio fundamental del conteo, permutaciones, ordenaciones, ordenaciones con repetición, y combinaciones

Situación 2. Sistema de alarma.

Una compañía deberá instalar un nuevo sistema de alarma de seguridad, el cual se activa y desactiva introduciendo el código numérico de 5 dígitos apropiado en el orden correcto en un tablero digital.

2.1. Si el sistema se puede comprar a 4 posibles distribuidores y cada uno de estos acepta tres formas de pago, ¿cuántas formas distintas tendrá la compañía de adquirir el nuevo sistema de seguridad?

Multiplicación de opciones o Regla general de conteo

Si una selección de k pasos, de los cuales el primero se puede efectuar en n_1 formas diferentes, el segundo de n_2, ... y así sucesivamente, hasta la k–ésima que se puede realizar de n_k maneras distintas, entonces la selección total de formas diferentes que se puede hacer es el producto

$$(n_1)(n_2)\cdots(n_k)$$

Solución:

Una representación gráfica de las diferentes formas de adquirir y pagar el nuevo sistema ayudará a contarlas. Esta representación, llamada **árbol de decisión o diagrama de árbol**, se muestra en la figura 1. En el punto de partida hay cuatro opciones (distribuidores) para comprar el sistema y por cada uno de estas, se tendrán tres opciones para pagar (forma de pago); así, el árbol de decisión muestra claramente que existen 4(3) =12 formas distintas de adquirir el nuevo sistema.

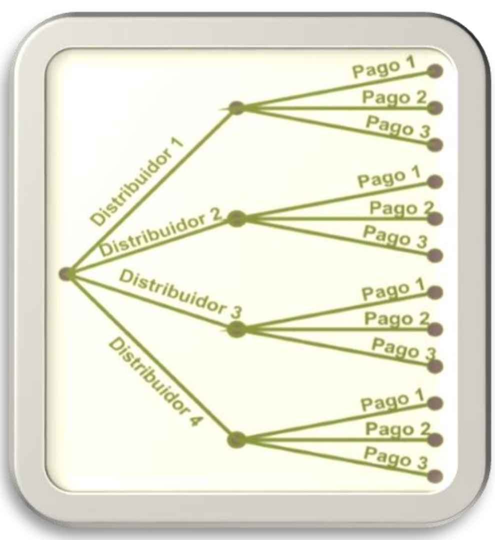


Figura 1. Diagrama de árbol. (Elaboración propia)

2.2. La compañía desea ahora saber el número total de posibles códigos que se pueden generar para este sistema, si sólo se pueden utilizar los dígitos del 1 al 5, una sola vez.

Factorial de un número

Dado un número entero positivo n, se denota y se define el factorial de éste como:

$$n! = n(n-1)(n-2)\cdots(3)(2)(1)$$

También por definición, 0!=1

Permutación

Dados n objetos, una **permutación** de ellos es cualquiera de las diferentes maneras en que se pueden acomodar en orden dichos objetos.
Se denota con P_n y su fórmula es $P_n = n!$

Solución:

Es posible elegir el primer número del código de entre 5 dígitos, el segundo número de entre 4, el tercero de 3, el cuarto de 2 y el último de un dígito, ya que no se pueden repetir, teniendo así 120 códigos diferentes. Así se tiene que es una permutación de 5 elementos, $P_5 = 5! = 120$.

2.3. Si ahora pueden utilizarse los dígitos del 0 al 9 solo una vez, ¿cuál será el número total de posibles códigos que se pueden generar para este sistema?

Ordenación

Dados n objetos y $r \leq n$, una **ordenación** de n objetos, tomados de r en r, es cualquiera de las diferentes maneras en que se pueden elegir en orden estos r objetos.
Se denota con O_r^n o con ${}_nO_r$

Su fórmula es $O_r^n = \dfrac{n!}{(n-r)!} = n(n-1)(n-2)\cdots(n-(r-1))$

Solución:

Es posible elegir el primer número del código de entre 10 dígitos, el segundo número de entre 9, el tercero de 8, el cuarto de 7 y el último de entre 6 dígitos, ya que no se pueden repetir e interesa el orden en como den, teniendo así 10(9)(8)(7)(6) = 30,240 códigos diferentes. En este caso es una ordenación de 10 dígitos, seleccionando solo 5 de ellos, es decir, ${}_{10}O_5 = \dfrac{10!}{5!} = 30{,}240$.

2.4. ¿Cuál será el número total de posibles códigos que se pueden generar para este sistema de alarma de seguridad, si los dígitos del 0 al 9 pueden utilizarse más de una vez?

Ordenación con repetición

Dados n objetos y cualquier número entero positivo r, se llama una **ordenación con repetición** de n objetos, tomados de r en r a cualquiera de las diferentes maneras en que se pueden llenar r lugares con los n objetos, permitiendo repetir dichos objetos.

Se denota con OR_r^n o con ${}_nOR_r$

Su fórmula es $OR_r^n = n^r$

Solución:

Como ahora es posible que se repitan dígitos, entonces el primer número del código se puede elegir de entre los 10 dígitos, el segundo de 10, el tercero de 10, el cuarto de 10 y el último de entre 10 dígitos, teniendo así 10(10)(10)(10)(10)=100,000 códigos diferentes. En este caso se tiene una ordenación con repetición de 10 dígitos que se pueden repetir hasta 5 veces, es decir, ${}_{10}OR_5 = 10^5 = 100{,}000$.

2.5. ¿Cuál será el número total de posibles códigos que se pueden generar para este sistema de alarma de seguridad, si los dígitos del 0 al 9 pueden utilizarse sólo una vez, sin importar el orden en que se den?

Combinación

Dados n objetos y $r \leq n$, una **combinación** de n objetos, tomados de r en r es cualquiera de las diferentes maneras en que se pueden elegir r objetos de los n disponibles sin importar el orden en que se presentan.

Se denota con C_r^n , ${}_nC_r$ o la más utilizada $\binom{n}{r}$

Su fórmula es $\binom{n}{r} = \dfrac{n!}{r!\,(n-r)!}$

Solución:

En este caso es posible elegir el primer número del código de entre 10 dígitos, el segundo número de entre 9, el tercero de 8, el cuarto de 7 y el último de entre 6 dígitos, ya que no se pueden repetir los dígitos y como no interesa el orden en como den, se tendrá que los códigos obtenidos se repiten 5!=120 veces, tiendo que el número de códigos diferentes es de $\frac{(10)(9)(8)(7)(6)}{120}=252$. En este caso es una combinación de 10 dígitos, seleccionando sólo 5 de ellos, es decir, ${}_{10}C_5=\frac{10!}{5!\,5!}=252$.

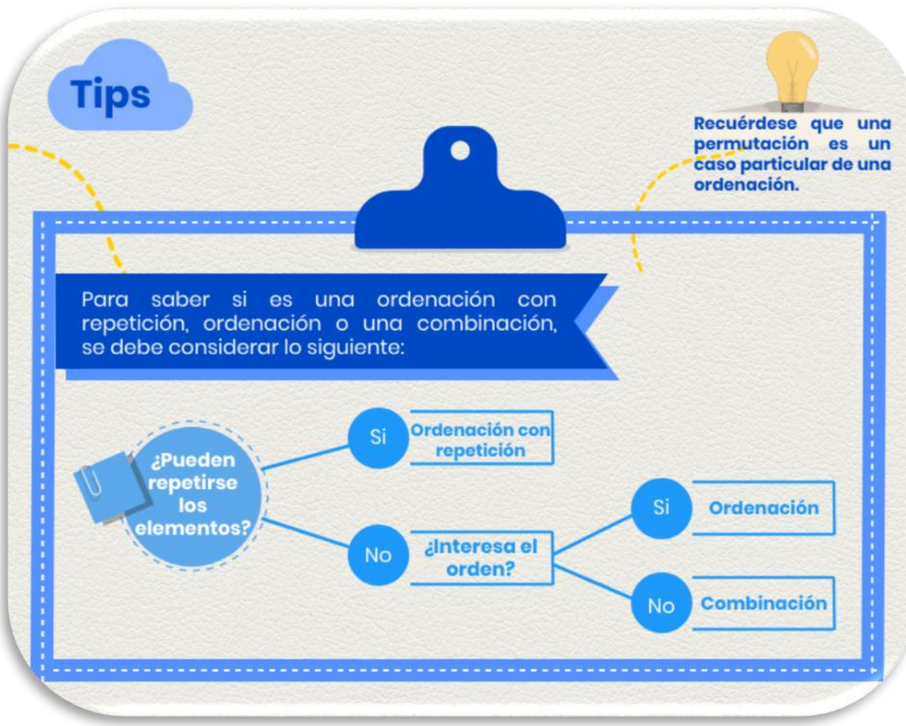


Figura 2. Tips. Elaboración propia

La figura anterior, es una herramienta para determinar cuándo se tiene una permutación, una ordenación, una ordenación con repetición o una combinación de un conjunto de objetos.

2.3 Probabilidad de eventos

Probabilidad de un evento

Si E es un evento de un espacio muestral finito Ω, se denota la probabilidad de la ocurrencia de este evento mediante $P(E)$, la cual se puede dar de diferentes formas: en fracción, con decimales o en porcentaje (multiplicando el resultado por 100%). Además, se cumplen los siguientes postulados:

1) $0 \leq P(E) \leq 1$
2) $P(\Omega) = 1$
3) $P(\varnothing)=0$
4) $P(E^c) = 1 - P(E)$

2.3.1 Enfoques de la Probabilidad

2.3.1.1. Objetivo (frecuencial y clásico)

Situación 3. Color de ojos.

En un grupo de clases de inglés hay 15 niñas de las cuales 3 tienen ojos azules y una de ojos verdes. Se escogen dos niñas al azar.

3.1. ¿Cuál es la probabilidad de que las dos niñas tengan ojos azules?

Probabilidad bajo el enfoque clásico

En la probabilidad bajo el enfoque clásico los eventos son equiprobables, por lo que la probabilidad de que suceda un evento E es:

$$P(E) = \frac{\text{Número de elementos de } E}{\text{Número de elementos de } \Omega}$$

Solución:

Se tienen 3 niñas con ojos azules en el grupo, de las cuales solo dos deben escogerse al azar sin importar el orden en que sean seleccionadas (ya se eligen las dos al mismo tiempo). Llámese A al evento que se desea.

El total de elementos que tiene el evento A es igual a la combinación de 3 objetos tomados de 2 en 2 y el número de elementos de Ω o casos totales, es la combinación de 15 objetos tomados de 2 en 2, así:

$$P(A)=\frac{\binom{3}{2}}{\binom{15}{2}}=\frac{\frac{3!}{2!\,1!}}{\frac{15!}{2!\,13!}}=\frac{3}{105}=\frac{1}{35}=0.0286 \text{ ó } 2.86\%$$

3.2. ¿Cuál es la probabilidad en el que ninguna de las niñas tenga ojos azules?

Solución:

Se sabe que hay 3 niñas con ojos azules en el grupo y 12 que no tienen ojos azules. Sea *N* el evento de que ninguna de las dos niñas elegidas tenga ojos azules, así el total de elementos que tiene este suceso es igual a la combinación de 12 objetos tomados de 2 en 2 y el número de elementos de Ω, es la combinación de 15 objetos tomados de 2 en 2, entonces:

$$P(N)=\frac{\binom{12}{2}}{\binom{15}{2}}=\frac{\frac{12!}{2!\,10!}}{\frac{15!}{2!\,13!}}=\frac{66}{105}=\frac{22}{35}=0.6286 \text{ ó } 62.86\%$$

3.3. ¿Cuál es la probabilidad de que una de las niñas tenga ojos azules y la otra, ojos verdes?

Solución:

Sea *E* el evento en el que de las dos niñas que se escogen al azar, una tenga ojos azules y la otra ojos verdes.

Retomando el enunciado de la situación 3, hay tres niñas con ojos azules y una de ojos verdes, entonces el total de elementos que tiene el suceso *E* es igual a la combinación de 3 objetos tomados de 1 en 1 por la combinación de 1 objeto tomado de 1 en 1, y el número de elementos de Ω no cambia, es C_2^{15}. Por lo tanto

$$P(E)=\frac{C_1^3\,C_1^1}{C_2^{15}}=\frac{\left(\frac{3!}{1!\,2!}\right)\left(\frac{1!}{1!\,0!}\right)}{\frac{15!}{2!\,13!}}=\frac{3}{105}=\frac{1}{35}=0.0286 \text{ ó } 2.86\%$$

3.4. ¿Cuál es la probabilidad en la que al menos una de las niñas tenga ojos azules?[1]

Solución:

Sea *D* es el evento en el que al menos una de las dos niñas tenga ojos azules. Este suceso tiene dos posibles resultados: elegir una niña que tenga ojos azules y una con ojos de otro color, o elegir dos niñas que tengan ojos azules del grupo de 15 niñas.

Si D_1 es el evento que una de las dos niñas una tenga ojos azules, por lo que el número de elementos de D_1 o casos favorables es la combinación de 3 objetos tomados de 1 en 1 por una combinación de 12 objetos tomados de 1 en 1.

Si D_2 es evento en el que las dos niñas seleccionadas tengan ojos azules, el número de elementos de este son las combinaciones de 3 objetos tomados de dos en dos.

Así el evento *D* es $D_1 \cup D_2$ y cómo D_1 y D_2 son sucesos mutuamente excluyentes (ver sección 2.3.2), entonces $P(D) = P(D_1) + P(D_2)$. Por consiguiente:

$$P(D) = \frac{\binom{3}{1}\binom{12}{1} + \binom{3}{2}}{\binom{15}{2}} = \frac{\left(\frac{3!}{1!\,2!}\right)\left(\frac{12!}{1!\,11!}\right) + \frac{3!}{2!\,1!}}{\frac{15!}{2!\,13!}} = \frac{36+3}{105} = \frac{13}{35} = 0.3714 \text{ ó } 37.14\%$$

3.5. ¿Cuál es la probabilidad de que las dos niñas tengan ojos verdes?

Solución:

Sea *H* el evento en el que las dos niñas seleccionadas tengan ojos verdes, como en el grupo de la clase hay 15 niñas 3 con ojos azules y una de ojos verdes este evento es imposible ya que solo hay una niña con ojos verdes. No se pueden elegir dos de ojos verdes por lo que el número de elementos de *H* es cero. Entonces:

$$P(H) = \frac{0}{{}_{15}C_2} = \frac{0}{\frac{15!}{2!\,13!}} = \frac{0}{105} = 0$$

[1] Cabe señalar que la palabra "***al menos***" significa tomar mínimamente desde ese número y "***a lo más***" es considerar como máximo ese número.

3.6. Suponiendo que el experimento de seleccionar a las dos niñas lo realizó el docente de la clase de inglés 120 veces, obteniéndose los resultados de la tabla 1. ¿Cuál es la probabilidad en qué al realizar nuevamente el experimento, las dos niñas tengan ojos azules?

Nota: Las niñas con ojos de otro color son aquellas que no tienen ojos azules ni verdes.

Resultado	Frecuencia
Una niña con ojos verdes y una con ojos azules	23
Una niña con ojos verdes y una con ojos de otro color	24
Una niña con ojos azules y una con ojos de otro color	22
Dos niñas con ojos azules	15
Dos niñas con ojos de otro color	36
Total	**120**

Tabla 1. Distribución de frecuencia. (Elaboración propia)

Probabilidad bajo el enfoque de frecuencia relativa

En la probabilidad bajo el enfoque de frecuencia relativa, la probabilidad de que suceda un evento *E* es:

$$P(E)=\frac{\text{Número de casos favorables o número de éxitos}}{\text{Número total de intentos}}$$

Solución:

Sea *A* el evento en que las dos niñas que se escogen al azar tengan los ojos azules, entonces bajo el enfoque de frecuencia relativa, la probabilidad de este evento es

$$P(A)=\frac{15}{120}=\frac{1}{8}=0.125 \text{ ó } 12.5\%$$

Situación 4. Torneo de ajedrez.

La prepa 8 y la prepa 3 juegan doce partidos de ajedrez, donde la prepa 8 gana cinco, la prepa 3 gana cuatro y en tres hacen tablas. Acuerdan ahora jugar un torneo de tres partidos. Se denotan los sucesos de la siguiente manera:
W_1, W_2, W_3 son los sucesos de que prepa 8 gana el primer, segundo y tercer partido, respectivamente.
R_1, R_2, R_3 son los sucesos de que prepa 3 gana el primer, segundo y tercer partido, respectivamente.
T_1, T_2, T_3 son los sucesos de que ambas prepas hacen tablas el primer, segundo y tercer partido, respectivamente.

De acuerdo con el enunciado y con base en la experiencia se sabe que:

$P(\text{Prepa 8 gane}) = P(W) = \frac{5}{12}$; $P(\text{Prepa 8 no gane}) = P(W^C) = \frac{7}{12}$;

$P(\text{Prepa 3 gane}) = P(R) = \frac{4}{12} = \frac{1}{3}$; $P(\text{Prepa 3 no gane}) = P(R^C) = \frac{8}{12} = \frac{2}{3}$;

$P(\text{Tablas}) = P(T) = \frac{3}{12} = \frac{1}{4}$; $P(\text{No tablas}) = P(T^C) = \frac{9}{12} = \frac{3}{4}$.

Además, los eventos se consideran independientes, porque se acuerda jugar un torneo de tres partidos, cada partido es independiente ya que la probabilidad no se afecta.

Dos eventos son **independientes**, si el que suceda uno de ellos no afecta la realización del otro evento. Es decir,

$$P(A|B) = P(A) \quad \text{o bien} \quad P(B|A) = P(B)$$

Además $P(A\,y\,B) = P(A \cap B) = P(A)P(B)$

Dos eventos son **dependientes** cuando uno de ellos afecta la ocurrencia de que suceda el otro.

4.1. Determina la probabilidad de que la prepa 8 gane los tres partidos.

Solución:

La probabilidad de que la prepa 8 gane los tres partidos bajo el enfoque de frecuencia relativa, es:

$$P(W_1 \text{ y } W_2 \text{ y } W_3) = P(W_1 \cap W_2 \cap W_3) = \left(\frac{5}{12}\right)\left(\frac{5}{12}\right)\left(\frac{5}{12}\right) = \frac{125}{1728} = 0.0723 \text{ ó } 7.23\%$$

4.2. ¿Cuál es la probabilidad de que hagan tablas en dos partidos las prepas?

Solución:

Esta probabilidad se calcula de considerando que en dos partidos quedan tablas y en otro no, este último se puede dar en el tercer partido, o en el segundo o en el primer partido, así:

$$P(\text{tablas en dos partidos}) = P\left(T_1 \cap T_2 \cap T_3^C\right) + P\left(T_1 \cap T_2^C \cap T_3\right) + P\left(T_1^C \cap T_2 \cap T_3\right)$$
$$= \left(\frac{3}{12}\right)\left(\frac{3}{12}\right)\left(\frac{9}{12}\right) + \left(\frac{3}{12}\right)\left(\frac{9}{12}\right)\left(\frac{3}{12}\right) + \left(\frac{9}{12}\right)\left(\frac{3}{12}\right)\left(\frac{3}{12}\right)$$
$$= \frac{243}{1728} = \frac{9}{64} = 0.1406 \text{ ó } 14.06\%$$

4.3. Encuentra la probabilidad de que ganen alternadamente las prepas.

Solución:

La probabilidad de que ganen alternadamente es que primero gane la prepa 8, después la prepa 3 y en el tercer partido gana nuevamente la prepa 8, o bien, gane primero la prepa 3, luego la prepa 8 y nuevamente la prepa 3. Entonces:

$$P(W_1 \cap R_2 \cap W_3) + P(R_1 \cap W_2 \cap R_3) = \left(\frac{5}{12}\right)\left(\frac{4}{12}\right)\left(\frac{5}{12}\right) + \left(\frac{4}{12}\right)\left(\frac{5}{12}\right)\left(\frac{4}{12}\right)$$
$$= \frac{180}{1728} = \frac{5}{48} = 0.1042 \text{ ó } 10.42\%$$

4.4. ¿Cuál es la probabilidad en el que la prepa 3 gane al menos un partido?

Solución:

$$P(\text{al menos la prepa 3 gana un partido}) = 1 - (\text{prepa 3 no gana ningún partido})$$
$$= 1 - P\left(R_1^C \cap R_2^C \cap R_3^C\right) = 1 - \left(\frac{8}{12}\right)\left(\frac{8}{12}\right)\left(\frac{8}{12}\right) = \frac{1216}{1728} = \frac{19}{27} = 0.7037 \text{ ó } 70.37\%$$

2.3.1.2. Simple, conjunta y condicional

4.5. La probabilidad de que la prepa 3 gane el primer partido, ¿de qué tipo es?

Tipos de probabilidad

$$\text{Tipos de probabilidad} \begin{cases} \text{Simple} \\ \text{Conjunta o compuesta} \\ \text{Condicional} \end{cases}$$

La proporción de resultados exitosos respecto al total de resultados, conocida como **probabilidad simple**, es la forma más sencilla del cálculo de probabilidades.

La **probabilidad conjunta** o **compuesta** es la probabilidad correspondiente a un evento compuesto por dos o más eventos.

La **probabilidad condicional** es la probabilidad correspondiente a un evento, si se sabe que ya ocurrió otro evento del cual depende el primero.

Solución:

Es una probabilidad simple, puesto que se refiere a la probabilidad de un solo evento y además se calcula bajo el enfoque de frecuencia relativa. En este caso se tiene:

$$P(\text{Prepa 3 gana el primer partido}) = \frac{\text{Número de veces que gana prepa 3}}{\text{Número total de partidos ganados}}$$

4.6. La probabilidad de que la prepa 3 gane el primer partido o la prepa 8 el segundo, ¿de qué tipo es?

Solución:

Es una probabilidad conjunta, porque que se refiere a la probabilidad de dos eventos. En este caso se quiere:

$$P(\text{Prepa 3 gana el primer partido o prepa 8 gana el segundo partido})$$
$$= \frac{\text{No. de veces que gana prepa 3} + \text{No. de veces que gana prepa 8}}{\text{Número total de partidos ganados}}$$

4.7. La probabilidad de que la prepa 8 gane el segundo partido dado que ganó el primero, ¿de qué tipo es?

Solución:

Es una probabilidad condicional, dado que es la probabilidad de un evento sabiendo que ya aconteció anteriormente otro suceso.

$$P(\text{Prepa 8 gana el segundo partido dado que prepa 3 ganó el primer partido})$$
$$= P(\text{Prepa 8 gana el segundo partido} \mid \text{prepa 3 ganó el primer partido})$$

2.3.2. Unión, intersección y complemento de eventos

Situación 5. Reprobación en una preparatoria.

En la preparatoria "América Latina" el 25% reprobaron Matemáticas, 15% reprobaron Química y el 10% reprobaron las dos materias. Se selecciona un estudiante al azar, se denotan los siguientes eventos y sus probabilidades:

$P(M)=0.25$, probabilidad de reprobar Matemáticas

$P(Q)=0.15$, probabilidad de reprobar Química

Operaciones con eventos

Sean *A* y *B* dos eventos cualesquiera de un mismo espacio muestral, entonces:

- La **unión** de ellos es el evento formado por todos los elementos de *A* o todos los elementos de *B*, o de ambos.

$$A \cup B = \{x \mid x \in A \text{ o } x \in B\}$$

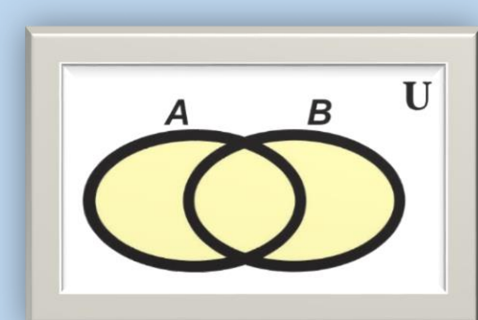


Figura 3. Unión. (Elaboración propia)

- La **intersección** de ellos es el evento formado por todos los elementos de *A* y todos los elementos de *B*.

$$A \cap B = \{x \mid x \in A \text{ y } x \in B\}$$

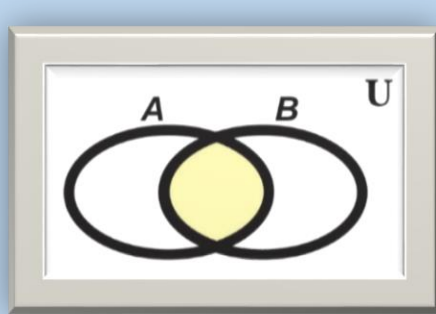


Figura 4. Intersección. (Elaboración propia)

- La **diferencia** de *A*–*B* es el evento formado por todos los elementos de *A* que no estén en *B*.

$$A - B = A \cap B^C = \{x \mid x \in A \text{ y } x \notin B\}$$

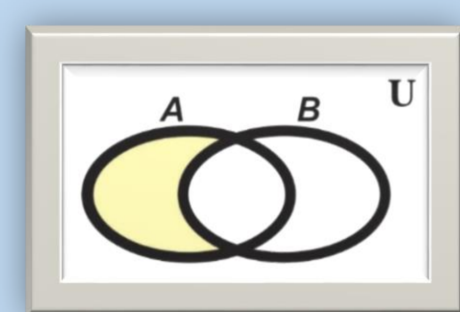


Figura 5. Diferencia. (Elaboración propia)

- El **complemento** de A^C es el evento formado por todos los elementos que no pertenecen al conjunto *A* pero pertenecen al espacio muestral *S*, Ω o *U*.

$$A^C = \{x \mid x \notin A \text{ y } x \in U\}$$

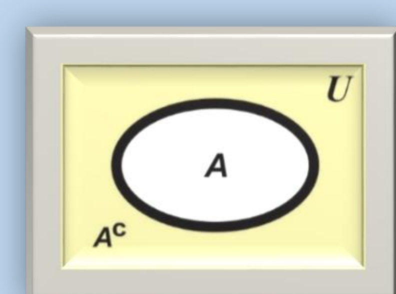


Figura 6. Complemento. (Elaboración propia)

Regla aditiva general de eventos

Sean *A* y *B* dos eventos cualesquiera de un mismo espacio muestral, entonces la probabilidad de que suceda A o B, o ambos es:

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

Regla aditiva para eventos mutuamente excluyentes

Sean *A* y *B* dos eventos mutuamente excluyentes de un mismo espacio muestral, entonces la probabilidad de que suceda A o B, o ambos es:

$$P(A\text{ o }B)=P(A\cup B)=P(A)+P(B)$$

Regla multiplicativa general de eventos

Sean *A* y *B* dos eventos cualesquiera de un mismo espacio muestral, entonces la probabilidad de que suceda A y B es:

$$P(A\text{ y }B)=P(A\cap B)=P(A)P(B|A)=P(B)P(A|B)$$

Regla multiplicativa para eventos independientes

Sean *A* y *B* dos eventos independientes de un mismo espacio muestral, entonces la probabilidad de que suceda A y B, es:

$$P(A\text{ y }B)=P(A\cap B)=P(A)P(B)=P(B)P(A)$$

Teoremas:

1. La probabilidad del complemento de un evento es:

$$P\left(A^C\right)=1-P(A)$$

2. La probabilidad de un evento imposible es: $P(\varnothing)=0$

3. Si el espacio muestral *S* se obtuvo de un experimento aleatorio, entonces $P(S)=1$

4. La probabilidad de la diferencia A relativa a B es:

$$P(A-B)=P\left(A\cap B^{C}\right)=P(A)-P(A\cap B)$$

Análogamente la diferencia de B relativa a A es:

$$P(B-A)=P\left(B\cap A^{C}\right)=P(B)-P(A\cap B)$$

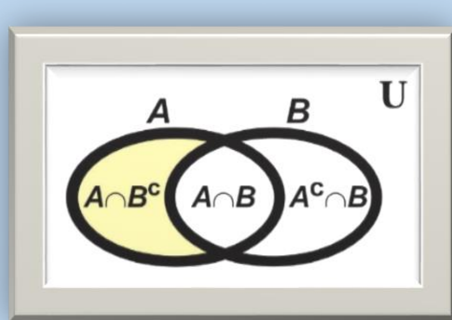


Figura 7. (Elaboración propia

Utilizando los anteriores teoremas y operaciones, se tiene que

$P(M\cap Q)=0.10$, es la probabilidad en que el estudiante elegido al azar repruebe Matemáticas y Química.

$P\left(M^{C}\right)=1-P(M)=1-.25=0.75$, es la probabilidad en que el estudiante seleccionado no repruebe (apruebe) Matemáticas.

$P\left(Q^{C}\right)=1-P(Q)=1-0.15=0.85$, es la probabilidad en que el alumno elegido al azar no repruebe (apruebe) Química.

5.1. ¿Cuál es la probabilidad de que el estudiante elegido adeude Matemáticas o Química?

Solución:

La probabilidad en que el alumno seleccionado repruebe Matemáticas o Química se calcula utilizando la regla general de la adición, así:

$$P(M\cup Q)=P(M)+P(Q)-P(M\cap Q)=0.25+0.15-0.10=0.30 \text{ ó } 30\%$$

5.2. ¿Cuál es la probabilidad que el alumno escogido no adeude Matemáticas ni Química?

Solución:

Está probabilidad se calcula más fácilmente utilizando la probabilidad del complemento de un evento, es decir

$$P\left(M^C \cap Q^C\right) = P\left(\left(M \cup Q\right)^C\right) = 1 - P\left(M \cup Q\right) = 1 - 0.30 = 0.70 \text{ ó } 70\%$$

Lo anterior, se puede representar mediante el diagrama de Venn (figura 8).

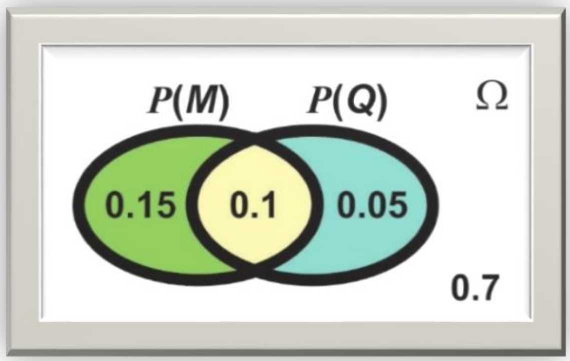


Figura 8. Probabilidades. (Elaboración propia)

5.3. ¿Cuál es la probabilidad en que el estudiante seleccionado adeude Química pero no Matemáticas?

Solución:

La probabilidad buscada se puede observar en la figura 7 (lado derecho), es decir, se quiere la probabilidad en que el estudiante únicamente repruebe Química. Así,

$$P\left(Q - M\right) = P\left(Q \cap M^C\right) = P\left(Q\right) - P\left(M \cap Q\right) = 0.15 - 0.10 = 0.05 \text{ ó } 5\%$$

5.4. Si la preparatoria tiene en total 2500 estudiantes, ¿cuántos estudiantes no adeudan ninguna de las dos asignaturas?

Solución:

Son (0.7)(2500) = 1750 alumnos, puesto que hay 2500 estudiantes y la probabilidad de que no adeuden ninguna de las dos asignaturas es 0.70 (pregunta 5.2.).

5.5. ¿Cuál es la probabilidad en que el estudiante seleccionado adeude Matemáticas, si se sabe que aprobó Química?

Fórmula para la probabilidad condicional

Para determinar la probabilidad condicional de que el evento *A* suceda, dado que ya ocurrió el evento *B*, se utiliza la fórmula

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

Análogamente, para determinar la probabilidad de que suceda *B* puesto que ya ocurrió el evento *A* se utiliza

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

Solución:

La probabilidad buscada es una probabilidad condicional $P\left(M \middle| Q^C\right)$, aplicando la fórmula anterior, se obtiene:

$$P\left(M \middle| Q^C\right) = \frac{P\left(M \cap Q^C\right)}{P\left(Q^C\right)} = \frac{P(M - Q)}{P\left(Q^C\right)} = \frac{0.25 - 0.1}{0.85} = \frac{0.15}{0.85} = 0.18 \text{ ó } 18\%$$

2.4. Teorema de Bayes

Situación 6. Visita al museo.

En la Ciudad de México se recibe alumnos como visitantes en sus museos de Historia Natural, Historia de la Ciudad y Ciencia y Tecnología, en una proporción de 18%, 32% y 50%, respectivamente, de los cuales se ha tenido información de que se les ha dado un mal servicio en un 3%, 1% y 4% de cada uno de los museos, respectivamente. Se selecciona un visitante al azar y se definen los eventos:

M: Al visitante se le dio un mal servicio
M^C: Al visitante se le dio un buen servicio
N: El visitante fue al museo de Historia Natural
C: El visitante fue al museo de Historia de la Ciudad
T: El visitante que fue al museo de Ciencia y Tecnología

Con base a lo anterior, se tienen las siguientes probabilidades:

$$P(N)=0.18,\quad P(N^C)=0.82,\quad P(M|N)=0.03,\quad P(M^C|N)=0.97$$
$$P(C)=0.32,\quad P(C^C)=0.68,\quad P(M|C)=0.01,\quad P(M^C|C)=0.99$$
$$P(T)=0.50,\quad P(T^C)=0.50,\quad P(M|T)=0.04,\quad P(M^C|T)=0.96$$

6.1. ¿Cuál es la probabilidad de que se le haya dado un buen servicio al visitante?

Teorema de Bayes

Una forma de calcular una probabilidad condicional es empleando el teorema de Bayes, que se utiliza cuando se desea determinar la probabilidad de que ocurra un evento A_i dado que B ya ocurrió, es decir:

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1)+P(A_2)P(B|A_2)+\cdots+P(A_n)P(B|A_n)}$$
$$=\frac{P(A_i)P(B|A_i)}{P(B)}$$

donde los eventos A_i son mutuamente excluyentes y exhaustivos.

Solución:

Utilizando la fórmula anterior, se tiene que

$$P\left(M^C\right)=P(N)P\left(M^C|N\right)+P(C)P\left(M^C|C\right)+P(T)P\left(M^C|T\right)$$
$$=0.18(0.97)+0.32(0.99)+0.5(0.96)=0.9714 \text{ ó } 97.14\%$$

6.2. Determina la probabilidad de que no se le haya dado un buen servicio al visitante.

Solución:

Haciendo uso del teorema de Bayes, se obtiene

$$P(M)=P(N)P(M|N)+P(C)P(M|C)+P(T)P(M|T)$$
$$=0.18(0.03)+0.32(0.01)+0.5(0.04)=0.0286 \text{ ó } 2.86\%$$

6.3. Si el visitante seleccionado no se quejó del servicio prestado, ¿cuál es la probabilidad de que haya visitado el museo de Historia Natural?

Solución:

Usando la fórmula de Bayes, se tiene que

$$P\left(N\middle|M^C\right)=\frac{P(N)P\left(M^C|N\right)}{P(N)P\left(M^C|N\right)+P(C)P\left(M^C|C\right)+P(T)P\left(M^C|T\right)}$$
$$=\frac{(0.18)(0.97)}{(0.18)(0.97)+(0.32)(0.99)+(0.5)(0.96)}=0.1797 \text{ ó } 17.97\%$$

6.4. ¿Cuál es la probabilidad de que el visitante elegido aleatoriamente haya visitado el museo de ciencia y tecnología, dado que se quejó del servicio prestado?

Solución:

Utilizando la fórmula anterior, se tiene que

$$P(T|M)=\frac{P(T)P(M|T)}{P(N)P(M|N)+P(C)P(M|C)+P(T)P(M|T)}$$
$$=\frac{(0.5)(0.04)}{(0.18)(0.03)+(0.32)(0.01)+(0.5)(0.04)}=0.6993 \text{ ó } 69.93\%$$

2.5. Variables aleatorias

Situación 7. Estacionamiento público.

Una empresa desea conocer las preferencias de la gente de la ciudad en cuanto a automóviles. Para llevar a cabo lo anterior, se elige un estacionamiento público para observar cuidadosamente la entrada de los vehículos a éste, durante una semana.

7.1. ¿Qué tipo de variable aleatoria es el número de autos de color blancos estacionados en ese lugar?

Variable aleatoria

Una **variable aleatoria** X es una función con valor numérico definida sobre un espacio muestral. Se clasifica en:

- **Variable aleatoria discreta**. Es aquella que puede asumir un número finito o infinito numerable de posibles valores.

- **Variable aleatoria continua**. Es aquella que puede asumir un número infinito no numerable de valores, es decir, puede tomar un número infinito de valores correspondientes a los puntos de uno o más intervalos de la recta real.

-

Solución:

Es una variable aleatoria discreta, ya que el número de autos blancos es contable, es decir, se estacionaron 0, 1, 2, …, etc.

7.2. ¿Qué tipo de variable aleatoria es el tiempo que estuvo estacionado un automóvil durante esa semana?

Solución:

Es una variable aleatoria continua, porque puede asumir un número infinito de valores, si se consideran horas y minutos.

2.5.1. Distribución Binomial

Situación 8. Futbol escolar.

En la preparatoria “América Latina” se formó el equipo de futbol “Leopardos”, la probabilidad de ganar el partido cuando juega se sabe que es de $\frac{2}{3}$. Se acuerda jugar 5 partidos, donde solo hay dos posibilidades: ganar o perder.
8.1. Determinar la probabilidad de al menos que gane tres partidos el equipo.

Distribución Binomial

Si p es la probabilidad de que ocurra un suceso en un solo intento (llamada probabilidad de éxito) y $q = 1 - p$ es la probabilidad de que no ocurra en un solo intento (llamado probabilidad de fracaso), entonces la probabilidad de que un suceso ocurra exactamente X veces en N intentos (o sea X éxitos y $N - X$ fracasos) viene dada por:

$$P(X = x) = \binom{N}{x} p^x q^{N-x} = \frac{N!\, p^x q^{N-x}}{x!\,(N-x)!}$$

donde $X = 0, 1, 2, 3, \ldots, N$

Media	$\mu = Np$
Varianza	$\sigma^2 = Npq$
Desviación típica	$\sigma = \sqrt{Npq}$

Además, se dice que X tiene una **distribución Binomial**.

Solución:

Sea X el número de partidos ganados (éxitos) por el equipo, N es el total de partidos, así se tiene que el equipo tiene una distribución binomial con:

$$X = 0, 1, 2, 3, 4, 5; \; p = \frac{2}{3}; \; q = \frac{1}{3}.$$

Realizando la tabla de distribución de la variable aleatoria, se tiene la tabla 2:

X	$P(X=x)$
0	${}_5C_0\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^5=\frac{1}{243}\approx 0.004$
1	${}_5C_1\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^4=\frac{10}{243}\approx 0.041$
2	${}_5C_2\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^3=\frac{40}{243}\approx 0.165$
3	${}_5C_3\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2=\frac{80}{243}\approx 0.329$
4	${}_5C_4\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^1=\frac{80}{243}\approx 0.329$
5	${}_5C_5\left(\frac{2}{3}\right)^5\left(\frac{1}{3}\right)^0=\frac{32}{243}\approx 0.132$
$\sum_{i=0}^{5} P(x_i)$	**1**

Tabla 2. Distribución Binomial. (Elaboración propia)

Entonces la probabilidad a determinar es que $X \geq 3$, es decir:

$$\begin{aligned}
P(X\geq 3) &= P(X=3)+P(X=4)+P(X=5) \\
&= \binom{5}{3}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2+\binom{5}{4}\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^1+\binom{5}{5}\left(\frac{2}{3}\right)^5\left(\frac{1}{3}\right)^0 \\
&= \frac{80+80+32}{243}=\frac{192}{243}=0.7901 \text{ ó } 79.01\%
\end{aligned}$$

8.2. ¿Cuál es la probabilidad de que gane a lo más dos partidos?

Solución:

Ahora se desea la probabilidad de que $X \leq 2$, es decir

$$P(X \le 2) = P(X=0)+P(X=1)+P(X=2)$$
$$= \binom{5}{0}\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^5 + \binom{5}{1}\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^4 + \binom{5}{2}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^3$$
$$= \frac{1+10+40}{243} = \frac{51}{243} = 0.2099 \text{ ó } 20.99\%$$

2.5.2. Distribución Normal

Situación 9. Deporte en la preparatoria

Los alumnos de la preparatoria Fray Pedro de Gante tienen marcas en ejecución en una prueba de atletismo, siguiendo un programa de adiestramiento se sabe que este tiene aproximadamente una distribución Normal con una media igual a 14 seg. y una desviación estándar igual a 2 seg. Si los atletas tienen una marca por debajo de 11 seg., deben ser retirados.

9.1. ¿Cuál es la probabilidad de qué al elegir un estudiante al azar, este sea retirado?

Distribución Normal

La **distribución Normal** es una de las distribuciones de probabilidad continua más importantes, la curva normal o gaussiana está definida por la ecuación:

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(X-\mu)^2} \qquad \text{(ecuación 1)}$$

donde:

- μ es la media
- σ^2 es la varianza
- σ es la desviación típica o estándar

El área total delimitada por la curva (figura 9) y el eje X es 1; por tanto, el área bajo la curva cuando de $X=a$ a $X=b$, con $a<b$, representa la probabilidad de que X esté entre a y b. Esta probabilidad se denota por $P(a \le X \le b)$.

Cuando se expresa la variable X en unidades estándar $(\mu = 0 \text{ y } \sigma = 1)$ se tiene que el estadístico

$$z = \frac{X-\mu}{\sigma} \qquad \text{(ecuación 2)}$$

Distribución Normal (continuación)

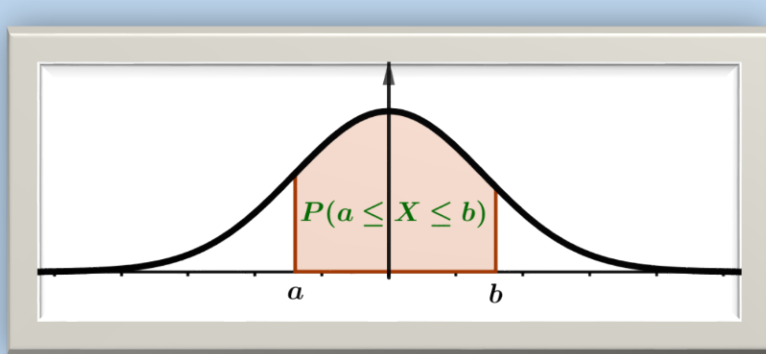


Figura 9. Distribución Normal. (Elaboración propia)

Y si la ecuación 1 se sustituye por la expresión 2, se obtiene

$$Y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad \text{(ecuación en forma canónica)}$$

En este caso se dice que z esta normalmente distribuida con media 0 y varianza 1.

En la gráfica de la figura 10 se visualiza un gráfico de la forma canónica; en ella se muestra que en los intervalos comprendidos entre z=–1 y z=1; z=–2 y z=2; así como, z=–3 y z=3 hay aproximadamente a 68.27%, 95.45% y 99.73%, respectivamente del total de datos que conforman la curva (ver también el ANEXO 1).

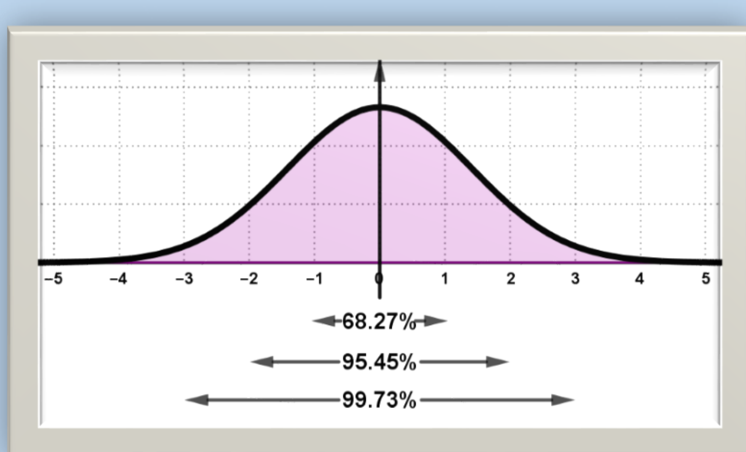


Figura 10. Distribución Normal Estándar. (Elaboración propia)

Solución:

De acuerdo con el enunciado, se tiene que $\bar{X} = 14$, $s = 2$ y $x = 11$, por lo que la unidad estándar es: $z = \dfrac{x - \bar{X}}{s} = \dfrac{11-14}{2} = \dfrac{-3}{2} = -1.5$.

El área bajo la curva es a la izquierda de z = –1.5 (ver figura 11), y usando la tabla del Anexo 1[2] (ver tabla 3), se obtiene que:

$$P(z \le -1.5) = 0.5 - 0.4332 = 0.0668 \text{ ó } 6.68\%$$

z	**0**	**0.01**
1.0	0.3413	0.3438
1.1	0.3643	0.3665
1.2	0.3849	0.3869
1.3	0.4032	0.4049
1.4	0.4192	0.4207
1.5	0.4332	0.4345
1.6	0.4452	0.4463
1.7	0.4554	0.4564
1.8	0.4641	0.4649
1.9	0.4713	0.4719

Tabla 3. Tabla Normal Estándar. (Elaboración propia)

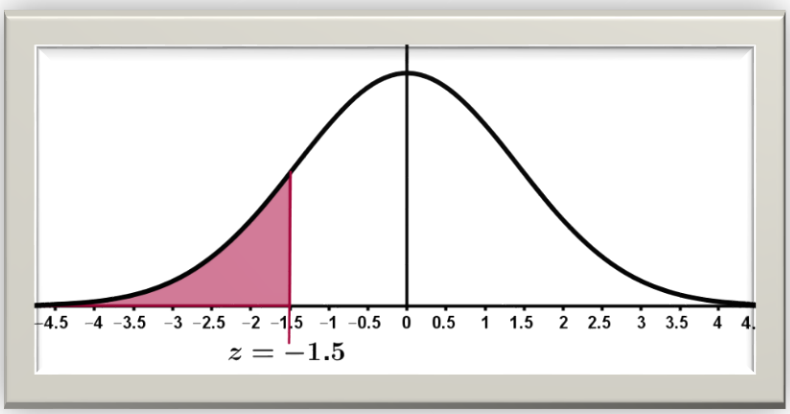


Figura 11. (Elaboración propia)

El anterior resultado indica que la probabilidad de elegir un estudiante y que éste sea retirado o que la marca que obtuvo esté por debajo de 11 seg., es del 6.68%.

[2] En el ANEXO 1 se puede encontrar el área bajo la curva Normal Estándar comprendida entre 0 y un valor positivo z. Recuérdese que la curva es simétrica.

9.2. Determina la probabilidad de qué al elegir un alumno al azar, éste tenga una marca mayor a 12.5 seg. y menor a 16 seg.

Solución:

Se tiene que $\bar{X}=14$, $s=2$, $x_1=12.5$ y $x_2=16$, y se quiere obtener $P(12.5 \le X \le 16)$, por lo que se tiene que encontrar los valores de x estandarizados, así: $z_1=\dfrac{12.5-14}{2}=\dfrac{-1.5}{2}=-0.75$ y $z_2=\dfrac{16-14}{2}=\dfrac{2}{2}=1$.

Entonces se tiene que calcular el área bajo la curva de –0.75 a 1 (ver figura 12).

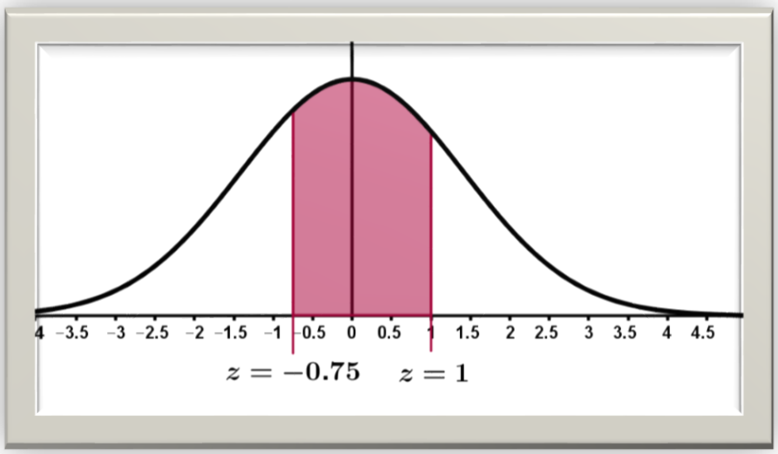


Figura 12. (Elaboración propia)

Usando la tabla del Anexo 1 (ver tabla 4), se tiene que calcular dos áreas, éstas se pueden ver claramente en las figuras 13 y 14, y la distribución Normal es simétrica se tiene $P(-0.75 \le z)=P(z \le 0.75)=0.2734$ y $P(z \le 1)=0.3413$

z	**0**	**0.05**
0.5	0.1915	0.2088
0.6	0.2257	0.2422
0.7	0.2580	0.2734
0.8	0.2881	0.3023
0.9	0.3159	0.3289
1.0	0.3413	0.3531
1.1	0.3643	0.3749
1.2	0.3849	0.3944
1.3	0.4032	0.4115
1.4	0.4192	0.4265

Tabla 4. Tabla Normal Estándar. (Elaboración propia)

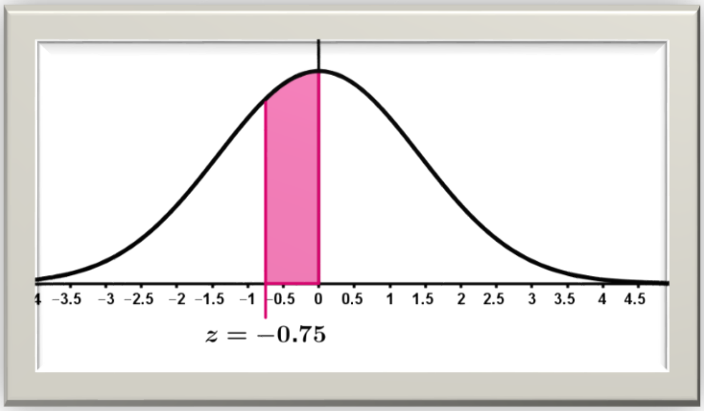


Figura 13. (Elaboración propia)

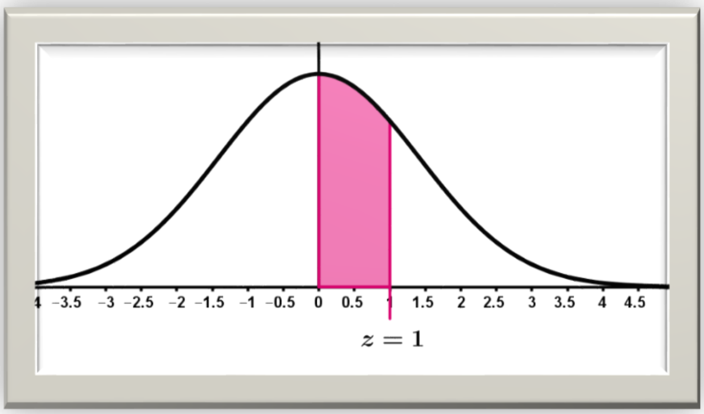


Figura 14. (Elaboración propia)

Por lo tanto, la probabilidad que se requiere es:
$P(-0.75 \leq z \leq 1) = P(-0.75 \leq z) + P(z \leq 1) = 0.2734 + 0.3413 = 0.6147$ ó 61.47%,
es decir, la probabilidad de elegir un estudiante y que éste obtenga una marca mayor a 12.5 seg. y menor a 16 seg. es del 61.47%.

EXAMEN DE AUTOEVALUACIÓN

Situación 1. Sistema de cómputo.

El sistema de cómputo de una universidad ha dejado de dar servicio porque necesita reparaciones. Las reparaciones del servicio previas han sido causadas por fallas de hardware, fallas de software o fallas de alimentación (eléctrica). Cuando el sistema se suspende, 69% del tiempo es por problemas de hardware, 14% es por problemas de software y 17% del tiempo porque el problema es eléctrico. Los ingenieros de mantenimiento han determinado que las probabilidades de los problemas de hardware, software y electrónicos son 1%, 5% y 3%, respectivamente.

1. ¿Qué probabilidad hay en el que una suspensión actual se deba a una falla de hardware? [3 puntos]

A) $\frac{69}{190}$

B) $\frac{7}{19}$

C) $\frac{51}{190}$

D) $\frac{1}{99}$

2. ¿Cuál es probabilidad en el que una suspensión actual se deba a una falla de software? [3 puntos]

A) $\frac{133}{981}$

B) $\frac{7}{19}$

C) $\frac{69}{190}$

D) $\frac{51}{190}$

Situación 2. Estudiantes extranjeros

Cierta universidad de Estados Unidos informa que 66% de los estudiantes de posgrado que obtienen el grado de doctorado en Administración de negocios en ese país son ciudadanos de otros países. Considérese una muestra aleatoria de 20 estudiantes extranjeros de esa universidad que recientemente obtuvieron su doctorado.

3. Determina la probabilidad de que haya 7 extranjeros en la muestra que hayan obtenido su doctorado en administración de negocios. [2 puntos]

A) 0.9966
B) 0.1836
C) 0.8164
D) 0.0034

4. Determina la probabilidad de que haya al menos 6 extranjeros en la muestra que hayan obtenido su doctorado en administración de negocios. [3 puntos]

A) 0.00022
B) 0.99978
C) 0.00018
D) 0.99982

Situación 3. Enfermedades de alumnos.

En la preparatoria "América Latina", se tiene que 30% de los alumnos padecen gastritis, el 15% padece hipertensión y el 5% ambos padecimientos. Se selecciona un estudiante al azar.

5. ¿Cuál es la probabilidad de que sea hipertenso y no tenga gastritis? [1 punto]

A) 60%
B) 10%
C) 25%
D) 30%

6. ¿Cuál es la probabilidad de que no padezca ninguna de las dos enfermedades? [1 punto]

A) 85%
B) 70%
C) 95%
D) 60%

7. ¿Cuál es la probabilidad de que solo padezca una enfermedad? [2 puntos]

A) 45%
B) 65%
C) 35%
D) 55%

8. Si la preparatoria tiene en total 2500 estudiantes, ¿cuántos estudiantes son hipertensos y no padecen gastritis? [1 punto]

A) 250 alumnos
B) 125 alumnos
C) 375 alumnos
D) 750 alumnos

RESPUESTAS DEL EXAMEN DE AUTOEVALUACIÓN

Número de pregunta	1	2	3	4	5	6	7	8
Respuesta	A	B	D	B	B	D	C	A

BIBLIOGRAFÍA

De Oteyza, E., Lam E., Hernández, C. y Carrillo, A. (2015). *Probabilidad y estadística*. Ciudad de México, México: Pearson Educación.

Mendenhall, W., Beaver, R. y Beaver, B. (2015). *Introducción a la probabilidad y estadística*. Ciudad de México, México: CENGAGE Learning: 14ª edición.

Salazar, L., Bahena H. y Martínez, R. (2014). *Probabilidad y estadística*. Ciudad de México, México: Grupo Editorial Patria.

Spiegel, R. y Lindstrom, P. (2000). *Estadística*. México: Mc Graw–Hill.

EXAMEN TIPO EXTRAORDINARIO

Situación 1. Embarazos en madres adolescentes.

En México el embarazo en madres adolescentes representa un problema de salud pública y es un riesgo tanto para la salud de la mujer como para la del producto.
La siguiente tabla fue tomada del INEGI y muestra el porcentaje de nacimientos registrados de madres adolescentes (menores de 20 años). Desde 1994, hasta 2017.

Abarca los periodos presidenciales siguientes:

1) Ernesto Zedillo Ponce de León1994-1999
2) Vicente Fox Quesada 2000-2005
3) Felipe Calderón Hinojosa 2006-2011
4) Enrique Peña Nieto 2012-2017

Periodo	Porcentaje
1994	16.9
1995	16.5
1996	16.4
1997	16.3
1998	16.5
1999	16.9
2000	17.1
2001	17.2
2002	17.2
2003	16.8
2004	17.2
2005	17.4
2006	17.2
2007	17.8
2008	18.3
2009	18.8
2010	18.8
2011	19.2
2012	19.4
2013	19.4
2014	19.2
2015	18.2
2016	17.8
2017	17.9

Tabla 20. Porcentaje de embarazos en madres adolescentes en México (INEGI)
https://www.inegi.org.mx/temas/natalidad/

1. **Es todo aquel valor que permite caracterizar o describir a la población de madres adolescentes: [1 punto]**

   A) estadístico
   B) variación
   C) parámetro
   D) atributo

2. **¿Qué escala de medición se usó para medir la variable porcentaje de madres adolescentes? [2 puntos]**

   A) nominal
   B) ordinal
   C) intervalo
   D) razón

3. **¿En qué periodo presidencial el promedio del porcentaje de embarazos en madres adolescentes fue mayor? [2 puntos]**

   A) Ernesto Zedillo
   B) Vicente Fox
   C) Felipe Calderón
   D) Enrique Peña Nieto

4. **¿Cuál es valor de la mediana del porcentaje de embarazos de madres adolescentes? [2 puntos]**

   A) 17.20
   B) 17.25
   C) 17.30
   D) 17.40

5. **¿En qué periodo presidencial hubo una mayor variación en el porcentaje de embarazos de madres adolescentes? [2 puntos]**

   A) Ernesto Zedillo
   B) Vicente Fox
   C) Felipe Calderón
   D) Enrique Peña Nieto

6. **¿Cuál fue el valor del coeficiente de variación durante el periodo de Felipe Calderón? [2 puntos]**

   A) 0.6770
   B) 0.7422
   C) 0.0368
   D) 0.0404

Situación 2. Estaturas niños.

La siguiente tabla muestra el peso y estatura que deben tener los niños cuyas edades van de uno a cinco años, según datos del IMSS.

Peso	Estatura
9.6	75.7
10.9	82.3
12.2	87.8
13.3	91.9
14.3	96.1
15.3	99.9
16.3	103.3
17.3	106.7

Tabla 21. Estatura y peso niños
http://xurl.es/l6fco

7. **¿Cuál es el valor del coeficiente de correlación lineal y que indica? [3 puntos]**

   A) 0.9981 indica una baja relación lineal entre las variables
   B) 0.9963 indica una baja relación lineal entre las variables
   C) 0.9981 indica una fuerte relación lineal entre las variables
   D) 0.9963 indica una fuerte relación lineal entre las variables

8. **La ecuación de la recta de mejor ajuste o de regresión lineal es: [3 puntos]**

   A) $y = 0.2498x + 9.5742$
   B) $y = 0.2498x - 9.5742$
   C) $y = 9.5742x - 0.2498$
   D) $y = 9.5742x + 0.2498$

9. **¿Cuál será el peso de un niño cuya estatura es de 120 cm? [2 puntos]**

   A) 18.3
   B) 19.3
   C) 20.4
   D) 21.4

10. **¿Qué estatura debe tener un niño que pesa 15 kg? [2 puntos]**

   A) 95.75
   B) 96.38
   C) 97.75
   D) 98.38

Situación 3. Médicos de un hospital.

En el hospital "Santo Niño de Atocha" se presentan las siguientes problemáticas:
Tres médicos A, B, C que laboran en el hospital, son candidatos para formar una comisión para verificar la eficacia de una vacuna. Las probabilidades para que los médicos A, B, C sean aceptados en dicha comisión es de 0.4, 0.6, 0.8, respectivamente. (Considere que la probabilidad de que un médico sea aceptado en la comisión no influye en las probabilidades de que las otras acepten)

11. **¿Cuál es la probabilidad de que al menos uno de estos médicos sea aceptado en la comisión? [3 puntos]**
   A) 0.952
   B) 0.192
   C) 0.808
   D) 0.048

12. **¿Cuál es la probabilidad de que dos de estos médicos sean aceptados en la comisión? [3 puntos]**
   A) 0.048
   B) 0.464
   C) 0.288
   D) 0.128

Los médicos determinaron que la probabilidad de éxito en la eficacia de la vacuna (H21) es de 0.7.

13. Calcula la probabilidad que una vez administrada a 15 pacientes, ninguno sufra la enfermedad (HN1). [2 puntos]

A) $P(X=0)=\binom{15}{0}(0)^{0.7}(0.15)^{0.3}$

B) $P(X=15)=\binom{15}{15}(0.15)^{0.3}(0)^{0.7}$

C) $P(X=15)=\binom{15}{15}(0.7)^{15}(0.3)^{0}$

D) $P(X=0)=\binom{15}{0}(0.7)^{0}(0.3)^{15}$

14. Determina la probabilidad que una vez administrada a 15 pacientes, dos de ellos contraigan la enfermedad (HN1) [2 puntos]

A) $P(X=2)=\binom{15}{2}(0.3)^{2}(0.7)^{13}$

B) $P(X=2)=\binom{15}{2}(0.7)^{2}(0.3)^{13}$

C) $P(X=1)+P(X=2)=\binom{15}{1}(0.7)^{1}(0.3)^{14}+\binom{15}{2}\square(0.7)^{2}\square(0.3)^{13}$

D) $P(X=1)+P(X=2)=\binom{15}{1}(0.3)^{1}(0.7)^{14}+\binom{15}{2}(0.3)^{2}(0.7)^{13}$

Situación 4. En la cafetería de la preparatoria.

En la cafetería de una preparatoria hay una máquina de café, la cual está regulada de tal modo que descargue un promedio de 200 mililitros (ml) por vaso. Si la cantidad de líquido está distribuida normalmente con una desviación estándar igual a 15ml.

15. ¿Qué porcentaje de vasos contendrán más de 230ml? [2 puntos]

A) 97.72%
B) 47.72%
C) 2.28%
D) 95.44%

16. ¿Cuál es la probabilidad de que un vaso contenga entre 191 y 209 ml? [2 puntos]
A) 0.2257
B) 0.7257
C) 0.2743
D) 0.4514

Si en la cafetería se están solicitando dos personas para laborar como mesera(o) y como cajera(o); y cinco personas se presentaron a solicitar los puestos, ya que les es indistinto ocupar cualquiera de estos, dado que cualquiera es apto para cualquiera de los puestos.

17. ¿De cuántas formas distintas puede el dueño de la cafetería contratar a las personas para los puestos? [1 punto]
A) ${}_5O_2$
B) ${}_5C_2$
C) 2^5
D) 5!

18. La primera tarea que tiene la mesera es recoger de la mesa un salero, dos salseras, cuatro platos y cuatro tazas. ¿De cuántas formas distintas puede colocarlos en fila en un estante, si los objetos son distinguibles entre sí y cada clase debe estar junto? [2 puntos]
A) $(1!)(2!)(4!)(4!)$
B) $\frac{9!}{(1!)(2!)(4!)(4!)}$
C) $(1!)(2!)(4!)(4!)(4!)$
D) $\frac{13!}{(1!)(2!)(4!)(4!)(4!)}$

Si se sabe que, en la cafetería, 70% de los clientes pide una bebida caliente, 15% solicita un bocadillo (pastel, galleta, etc.) y 8% compra tanto una bebida caliente como un bocadillo para acompañar ésta.

19. Si una persona al azar pide un café, ¿cuál es la probabilidad de que también compre un bocadillo? [2 puntos]
A) $(0.70)(0.15)$
B) $\frac{0.08}{0.70}$
C) $\frac{0.08}{0.15}$
D) $\frac{0.15}{0.70}$

20. Si un cliente solicita un bocadillo, ¿cuál es la probabilidad de que pida además un café? [2 puntos]

A) $(0.70)(0.15)$

B) $\frac{0.15}{0.70}$

C) $\frac{0.08}{0.70}$

D) $\frac{0.08}{0.15}$

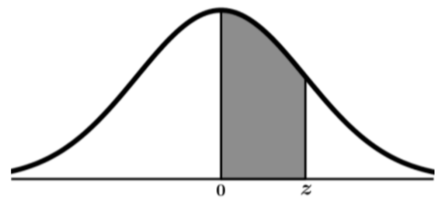
RESPUESTAS DEL EXAMEN TIPO EXTRAORDINARIO

Número de Pregunta	Respuesta
1	C
2	D
3	D
4	C
5	A
6	C
7	D
8	B
9	C
10	D
11	A
12	B
13	D
14	B
15	C
16	D
17	A
18	C
19	B
20	D

ANEXO 1. Tabla de la Normal Estándar

Área de la distribución Normal Estándar

Las entradas de esta tabla son las probabilidades de que una variable aleatoria, con una distribución normal estándar, tome un valor entre 0 y z; la probabilidad está representada por el área sombreada bajo la curva de la figura. El área para valores negativos de z se obtiene por simetría.



z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997

Datos no agrupados	Datos agrupados	Datos no agrupados	Datos agrupados
Media aritmética		*Varianza*	
$\overline{X}=\frac{\sum_{i=1}^{n}(x_i)}{n}$	$\overline{X}=\frac{\sum_{i=1}^{m}(f_i\cdot x_i)}{n}=\frac{\sum_{i=1}^{m}(f_i\cdot x_i)}{\sum_{i=1}^{m}(f_i)}$	$s^2=\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}{n-1}$ $\sigma^2=\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}{N}$	$s^2=\frac{\sum_{i=1}^{m}\left(f_i\left(x_i-\overline{X}\right)^2\right)}{n-1}$ $\sigma^2=\frac{\sum_{i=1}^{m}\left(f_i\left(x_i-\overline{X}\right)^2\right)}{N}$
Moda		**Estadísticos de posición (cuartiles)**	
$Moda=Mo=L+\left(\frac{d_1}{d_1+d_2}\right)\cdot t$		$Q_i=x_{\frac{i\cdot(n+1)}{4}}$ $i=1,2,3$	$Q_i=L+\left(\frac{\frac{i\cdot n}{4}-\sum f_i}{f_{Q_i}}\right)\cdot t$ $i=1,2,3$
Mediana		**Coeficiente de correlación**	**Ecuación de regresión**
Si n es impar $Md=x_{\frac{n+1}{2}}$ Si n es par $Md=\frac{x_{\frac{n}{2}}+x_{\frac{n}{2}+1}}{2}$	$Md=L+\left(\frac{\frac{n}{2}-\sum f_i}{f_{mediana}}\right)\cdot t$	$r=\frac{\sum_{i=1}^{n}\left[\left(x_i-\overline{X}\right)\left(y_i-\overline{Y}\right)\right]}{\sqrt{\left[\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2\right]\left[\sum_{i=1}^{n}\left(y_i-\overline{Y}\right)^2\right]}}$	$y=a+bx$ $a=\overline{Y}-b\overline{X}$ $b=\frac{n\sum_{i=1}^{n}(x_i\cdot y_i)-\sum_{i=1}^{n}(x_i)\sum_{i=1}^{n}(y_i)}{n\sum_{i=1}^{n}(x_i)^2-\left(\sum_{i=1}^{n}(x_i)\right)^2}=\frac{\sum_{i=1}^{n}(x_i\cdot y_i)-n\overline{X}\cdot\overline{Y}}{\sum_{i=1}^{n}\left(x_i^2\right)-n\overline{X}^2}$
Medidas de dispersión		**Cálculo Combinatorio**	
Desviación media		**Permutación** $P_n=n!$	**Ordenación con repetición** ${}_nOR_r=OR_r^n=n^r$
$D_M=\frac{\sum_{i=1}^{n}\lvert x_i-\overline{X}\rvert}{n}$	$D_M=\frac{\sum_{i=1}^{m}\left[f_i\cdot\lvert x_i-\overline{X}\rvert\right]}{n}$	**Ordenación** ${}_nO_r=O_r^n=\frac{n!}{(n-r)!}$	**Combinación** ${}_nC_r=C_r^n=\binom{n}{r}=\frac{n!}{r!(n-r)!}$
Desviación estándar		**Estandarización**	**Distribución Binomial**
$s=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}{n-1}}$ $\sigma=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}{n}}$	$s=\sqrt{\frac{\sum_{i=1}^{m}\left(f_i\left(x_i-\overline{X}\right)^2\right)}{n-1}}$ $\sigma=\sqrt{\frac{\sum_{i=1}^{m}\left(f_i\left(x_i-\overline{X}\right)^2\right)}{n}}$	$z=\frac{x_i-\overline{X}}{s}$	$P(X=x)={}_NC_r\,p^xq^{N-x}=\binom{N}{r}p^x(1-p)^{N-x}$

NOTAS