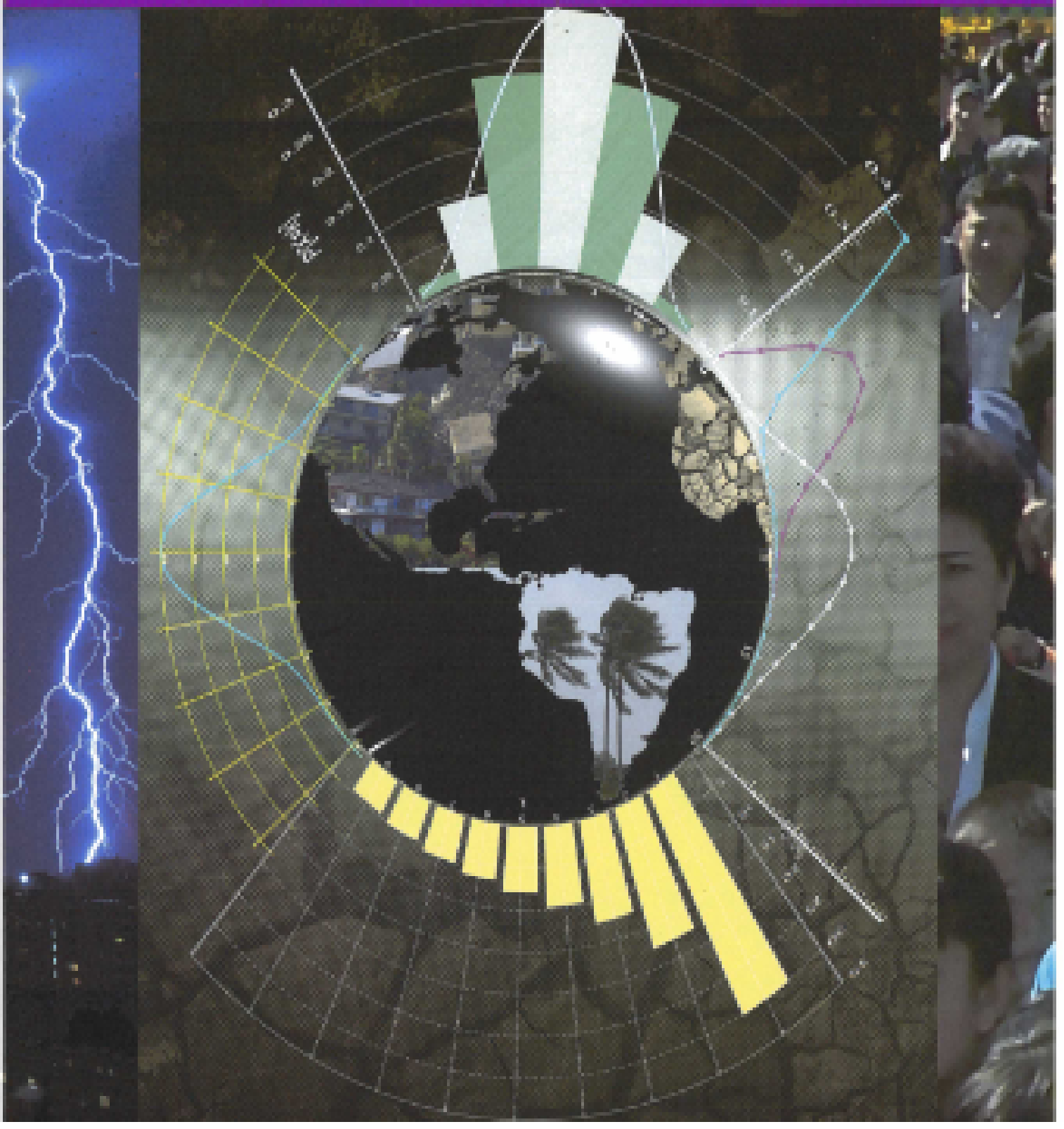
Estadística en Fenómenos Naturales y Procesos Sociales



ESTADÍSTICA EN FENÓMENOS NATURALES Y PROCESOS SOCIALES

Secretaría de Educación Pública
Emilio Chuayffet Chemor

Subsecretaría de Educación Media Superior
Rodolfo Tuirán Gutiérrez

Dirección General del Bachillerato
Carlos Santos Ancira

Autor
Felipe Jesús Cueva del Castillo Mendoza

Asesoría académica
Demetrio Garmendia Guerrero

Apoyo técnico pedagógico
Carolina L. de Marcellin

Revisión técnico pedagógica
Luciano Morales Ramírez

Coordinación y servicios editoriales
Edere S. A. de C. V.
José Ángel Quintanilla D'Acosta
Mónica Lobatón Díaz

Diseño y diagramación
Visión Tipográfica Editores, S. A. de C.V.

Material fotográfico e iconografía
Shutterstock Images, LLC
Martín Córdova Salinas
Isabel Gómez Caravantes

Primera edición, 2012
D.R. © Secretaría de Educación Pública, 2012
Argentina 28, Centro,
06020, México, D. F.

ISBN: 978-607-8229-42-0

Impreso en México

Presentación general 7
Como utilizar este material 10
Tu plan de trabajo 13
¿Con qué saberes cuento? 15

UNIDAD 1 LA ESTADÍSTICA DESCRIPTIVA Y LOS FENÓMENOS NATURALES Y PROCESOS SOCIALES

¿Qué voy a aprender y cómo? 21
La estadística y sus principios básicos 26
Fenómenos naturales y procesos sociales: sus características 26
Fenómenos naturales y procesos sociales en tu contexto 28
Tipos de eventos: determinísticos y aleatorios 32
Fenómenos naturales y procesos sociales determinísticos y aleatorios 33
Concepto de probabilidad 34
Asignación de la probabilidad a un evento aleatorio 35
¡Ahora te toca a ti! 38
Las variables como instrumento de medición de la incertidumbre y la variabilidad de un fenómeno aleatorio 42
Variables cuantitativas 43
Variables cualitativas 44
Variables categóricas nominales y ordinales 45
Clasificación de las variables por su grado de asociación 46
¡Ahora te toca a ti! 47
Muestreo: población, muestra y técnicas 50
Técnicas de muestreo y su importancia en el análisis de la información 50
Censos y muestras 51
Métodos de muestreo 53
¡Ahora te toca a ti! 57
Las distribuciones de probabilidad y su relación con las variables relacionadas con fenómenos naturales y procesos sociales 60
Distribución de frecuencias para variables categóricas 61
Histogramas de frecuencias 62
Histogramas para variables numéricas discretas 63
Histograma para una variable continua 65
Distribuciones de frecuencias para variables numéricas discretas 68
Distribuciones de frecuencias para variables numéricas continuas 68
Distribución de probabilidad continua 69
Distribuciones de probabilidades teóricas 70

Distribución binomial 71
Distribución de Poisson 73
¡Ahora te toca a ti! 74
Evalúa tu aprendizaje 80

UNIDAD 2 TRATAMIENTO ESTADÍSTICO DE LA INFORMACIÓN DE FENÓMENOS NATURALES Y PROCESOS SOCIALES

¿Qué voy a aprender y cómo? 85
La estadística descriptiva y su aplicación en la explicación de los fenómenos naturales y los procesos sociales 91
Medidas de tendencia central: media y mediana 92
Medidas de dispersión 98
¡Ahora te toca a ti! 103
Las distribuciones de probabilidad como herramientas para estimar las probabilidades de eventos relacionados con fenómenos naturales y procesos sociales 106
Distribución de probabilidad normal 106
Distribución de probabilidad binomial 117
Distribución de probabilidad de Poisson 119
¡Ahora te toca a ti! 121
El modelo de regresión y el de correlación lineal como medidas para describir la asociación entre variables 125
Correlación 125
Regresión lineal: ajuste de una recta a los datos bivariados 134
¡Ahora te toca a ti! 140
Evalúa tu aprendizaje 146
¿Ya estoy preparado(a)? 149
Apéndices
Apéndice 1. Clave de respuestas 164
Apéndice 2. Guía de uso de una hoja de cálculo para la aplicación de herramientas estadísticas 202
Apéndice 3. Bases de datos 221
Apéndice 4. Tabla de distribución normal estándar 231
Apéndice 5. Formas estandarizadas para dividir los intervalos de clase. 233
Apéndice 6. La consulta de fuentes de información por Internet 237
Apéndice 7. Mi ruta de aprendizaje 239
Fuentes de consulta 240

Este libro fue elaborado para ayudarte a estudiar el módulo *Estadística en fenómenos naturales y procesos sociales* del plan de estudios de la Preparatoria Abierta que ha establecido la Secretaría de Educación Pública (SEP), pero también está diseñado para utilizarse en otras modalidades no escolarizadas y mixtas. Sabiendo que trabajarás de manera independiente la mayor parte del tiempo este libro te brinda orientaciones muy precisas sobre lo que tienes que hacer y te proporciona la información que requieres para aprender.

Los estudios que iniciarás tienen como sustento un enfoque de educación por competencias, por lo que se busca que trabajes en adquirir nuevos conocimientos, habilidades y valores, así como en recuperar otros para transformarlos en capacidad para desempeñarte de forma eficaz y eficiente en diferentes ámbitos de tu vida profesional, laboral y personal.

Para facilitar tu estudio es importante que tengas muy claro qué implica aprender competencias, cómo estudiar en una modalidad no escolarizada y cómo utilizar este material.

¿Qué es una competencia?

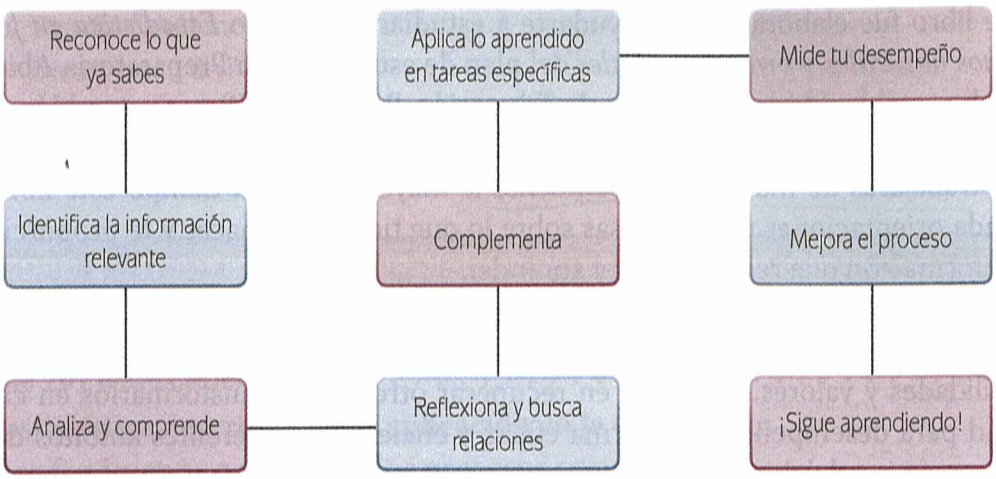
En el contexto educativo, hablar de "competencias" no es hacer referencia a una contienda o a una justa deportiva. El *Acuerdo 442* de la Secretaría de Educación Pública define competencia como la integración de habilidades, conocimientos, actitudes y valores en un contexto específico.

La meta de la formación como bachiller es que tú desarrolles las competencias que han sido definidas por la SEP como perfil de egreso para la Educación Media Superior.[1] No se pretende que te dediques a memorizar información o que demuestres habilidades aisladas. El objetivo es que logres aplicar de manera efectiva tus conocimientos, habilidades, actitudes y valores en situaciones o problemas concretos.

La cantidad de información de la que se dispone en la época actual nos impulsa a buscar formas diferentes de aprender ya que memorizar contenidos resulta insuficiente. Ahora se requiere que puedas analizar la información y te apropies de los conocimientos haciéndolos útiles para ti y tu entorno.

Por eso, cuando estudies no orientes tus esfuerzos solamente a identificar los conceptos más importantes, sino a analizarlos con detenimiento para comprenderlos y reflexionar sobre cómo se relacionan con otros términos. Busca información adicional. Pero no te quedes allí, aprende cómo aplicarlos en situaciones y contextos propuestos en las actividades. De manera concreta, para aprender es recomendable que sigas estos pasos:

1 De acuerdo con el Marco Curricular Común, el estudiante de bachillerato deberá desarrollar competencias genéricas y disciplinares básicas.



En este libro además de leer y estudiar textos y procedimientos, encontrarás problemas para resolver, casos para analizar y proyectos a ejecutar. Estos te ofrecerán evidencias sobre las capacidades que desarrollarás y podrás valorar tus avances.

Para acreditar el módulo *Estadística en fenómenos naturales y procesos sociales* es importante que demuestres que eres capaz de analizar y resolver situaciones, problemas y casos que te exigen la articulación de conocimientos, habilidades, actitudes y valores.

Estudiar en una modalidad no escolarizada

Una modalidad educativa no escolarizada como la que estás cursando tiene como ventaja una gran flexibilidad. Tú decides a qué hora y dónde estudias así como que tan rápido avanzas. Puedes adecuar tus horarios a otras responsabilidades cotidianas, como el trabajo, la familia o cualquier proyecto personal.

Pero en esta modalidad educativa, se requiere que tú:

- Seas capaz de dirigir tu proceso de aprendizaje. Es decir que:
  - Definas tus metas personales, considerando los objetivos de aprendizaje de los cursos.
  - Asignes tiempos para el estudio y procures contar con todos los recursos necesarios en un espacio apropiado.
  - Regules tu ritmo de avance.
  - Aproveches los materiales didácticos que la SEP ha preparado para apoyarte.
  - Utilices otros recursos que puedan ayudarte a profundizar tu aprendizaje.
  - Identifiques cuándo enfrentas dificultades para aprender y busques ayuda para superarlas.

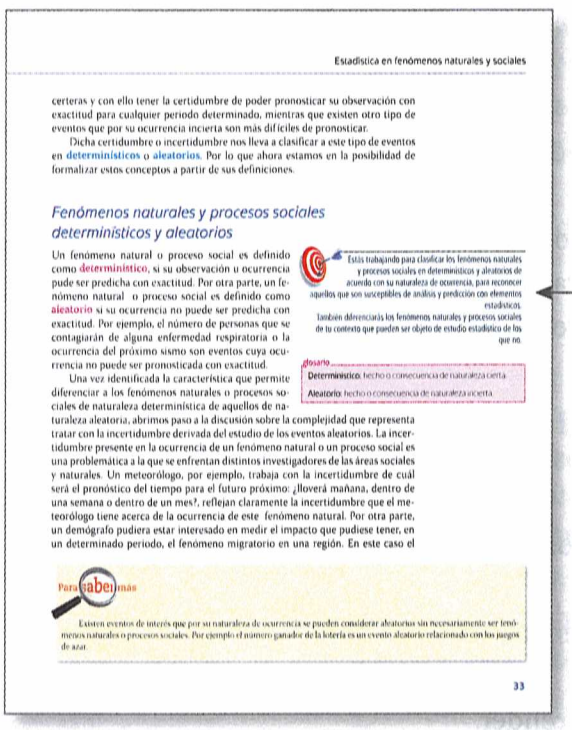
- Te involucres de manera activa en tu aprendizaje. Es decir que:
  - Leas para comprender las ideas que se te presentan y para construir significados.
  - Recurras a tu experiencia como punto de partida para el aprendizaje.
  - Realices las actividades propuestas y revises los productos que generes.
  - Reconozcas tus fortalezas y tus debilidades como estudiante.
  - Selecciones las técnicas de estudio que funcionen mejor para ti.
  - Emprendas acciones para enriquecer tus capacidades para aprender y subsanar tus limitaciones.
- Asumas una postura crítica y propositiva. Es decir que:
  - Analices de manera crítica los conceptos presentados.
  - Investigues sobre los temas que estudias y explores distintos planteamientos en torno a ellos.
  - Plantees alternativas de solución a los problemas.
  - Explores formas diversas de enfrentar las situaciones.
  - Adoptes una postura personal en los distintos debates.
- Seas honesto(a) y te comprometas contigo mismo(a). Es decir que:
  - Realices tú mismo(a) las actividades.
  - Consulta las respuestas después de haber llevado a cabo las actividades.
  - Busques asesoría en los Centros de Servicios de Preparatoria Abierta.
  - Destines el tiempo de estudio necesario para lograr los resultados de aprendizaje.
  - Evalúes tus logros de manera constante. Es decir que:
  - Analices tu ejecución de las actividades y los productos que generes utilizando la retroalimentación que se ofrece en el libro.
  - Identifiques los aprendizajes que alcances utilizando los referentes que te ofrece el material.
  - Reconozcas las limitaciones en tu aprendizaje y emprendas acciones para superarlas.
  - Aproveches tus errores como una oportunidad para aprender.
- Reflexiones sobre tu propio proceso de aprendizaje. Es decir que:
  - Te preguntes de manera constante: ¿Qué estoy haciendo bien?, ¿qué es lo que no me ha funcionado?
  - Realices ajustes en las estrategias que utilizas para mejorar tus resultados de aprendizaje.

Como puedes ver, el estudio independiente es una tarea que implica el desarrollo de muchas habilidades que adquirirás y mejorarás a medida que avances en tus estudios. El componente principal es que estés comprometido con tu aprendizaje.

Cómo utilizar este material

Este libro te brinda los elementos fundamentales para apoyar tu aprendizaje. Está constituido por diversas secciones en las que se proponen los pasos que debes seguir para estudiar.

1. En la sección *Tu plan de trabajo* encontrarás el propósito general del módulo, las competencias que deberás desarrollar y una explicación general de las unidades. Es importante que la leas.
2°. En la sección *¿Con qué saberes cuento?* hay una primera actividad de evaluación con la que puedes valorar si posees los saberes requeridos para el estudio del módulo. Es oportuno que identifiques desde el inicio si necesitas aprender o fortalecer algún conocimiento o habilidad antes de comenzar.
3°. Después de la sección anterior, se presentan las unidades en el orden sugerido para su estudio. Cada una de ellas contiene actividades de aprendizaje e información necesaria para realizarlas; sin embargo, continuamente se te sugiere que consultes fuentes adicionales.



Estadística en fenómenos naturales y sociales

certeras y con ello tener la certidumbre de poder pronosticar su observación con exactitud para cualquier periodo determinado, mientras que existen otro tipo de eventos que por su ocurrencia incierta son más difíciles de pronosticar.

Dicha certidumbre o incertidumbre nos lleva a clasificar a este tipo de eventos en **determinísticos** o **aleatorios**. Por lo que ahora estamos en la posibilidad de formalizar estos conceptos a partir de sus definiciones.

Fenómenos naturales y procesos sociales determinísticos y aleatorios

Un fenómeno natural o proceso social es definido como **determinístico**, si su observación u ocurrencia pude ser predicha con exactitud. Por otra parte, un fenómeno natural o proceso social es definido como **aleatorio** si su ocurrencia no puede ser predicha con exactitud. Por ejemplo, el número de personas que se contagiarán de alguna enfermedad respiratoria o la ocurrencia del próximo sismo son eventos cuya ocurrencia no puede ser pronosticada con exactitud.

Una vez identificada la característica que permite diferenciar a los fenómenos naturales o procesos sociales de naturaleza determinística de aquellos de naturaleza aleatoria, abrimos paso a la discusión sobre la complejidad que representa tratar con la incertidumbre derivada del estudio de los eventos aleatorios. La incertidumbre presente en la ocurrencia de un fenómeno natural o un proceso social es una problemática a la que se enfrentan distintos investigadores de las áreas sociales y naturales. Un meteorólogo, por ejemplo, trabaja con la incertidumbre de cuál será el pronóstico del tiempo para el futuro próximo: ¿llloverá mañana, dentro de una semana o dentro de un mes?, reflejan claramente la incertidumbre que el meteorólogo tiene acerca de la ocurrencia de este fenómeno natural. Por otra parte, un demógrafo pudiera estar interesado en medir el impacto que pudiese tener, en un determinado periodo, el fenómeno migratorio en una región. En este caso el

Estás trabajando para clasificar los fenómenos naturales y procesos sociales en determinísticos y aleatorios de acuerdo con su naturaleza de ocurrencia, para reconocer aquellos que son susceptibles de análisis y predicción con elementos estadísticos.

También deberás diferenciar los fenómenos naturales y procesos sociales de tu contexto que pueden ser objeto de estudio estadístico de los que no.

Glosario

Determinístico: hecho o consecuencia de naturaleza cierta.

Aleatorio: hecho o consecuencia de naturaleza incierta.

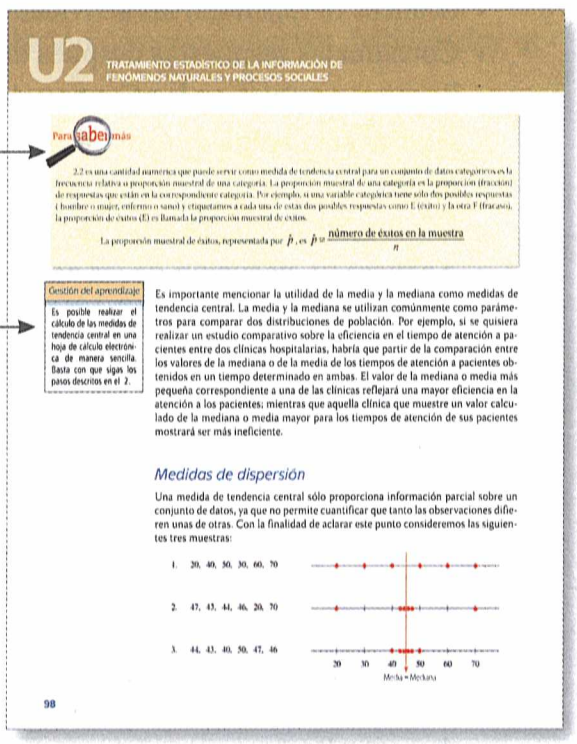
Para saber más

Existen eventos de interés que por su naturaleza de ocurrencia se pueden considerar aleatorios sin necesariamente ser fenómenos naturales o procesos sociales. Por ejemplo el número ganador de la lotería es un evento aleatorio relacionado con los juegos de azar.

33

Indicador de desempeño. Muestra las acciones que realizarás en un periodo determinado. Al conjuntar los diversos desempeños enunciados lograrás el propósito formativo de la unidad. Utilízalos como un referente para valorar de manera continua tu desempeño.

Para saber más. Brinda información interesante, curiosa o novedosa sobre el tema que se está estudiando.



U2 TRATAMIENTO ESTADÍSTICO DE LA INFORMACIÓN DE FENÓMENOS NATURALES Y PROCESOS SOCIALES

Para saber más

2.2 es una cantidad numérica que puede servir como medida de tendencia central para un conjunto de datos categóricos es la frecuencia relativa o proporción muestral de una categoría. La proporción muestral de una categoría es la proporción (fracción) de respuestas que están en la correspondiente categoría. Por ejemplo, si una variable categórica tiene solo dos posibles respuestas (hombre o mujer, enfermo o sano) y etiquetamos a cada una de estas dos posibles respuestas como E (éxito) y la otra F (fracaso), la proporción de éxitos (E) es llamada la proporción muestral de éxitos.

La proporción muestral de éxitos, representada por $\hat{p}$, es $\hat{p} = \frac{\text{número de éxitos en la muestra}}{n}$

Gestión del aprendizaje

Es posible realizar el cálculo de las medidas de tendencia central en una hoja de cálculo electrónica de manera sencilla. Basta con que sigas los pasos descritos en el 2.

Es importante mencionar la utilidad de la media y la mediana como medidas de tendencia central. La media y la mediana se utilizan comúnmente como parámetros para comparar dos distribuciones de población. Por ejemplo, si se quisiera realizar un estudio comparativo sobre la eficiencia en el tiempo de atención a pacientes entre dos clínicas hospitalarias, habría que partir de la comparación entre los valores de la mediana o de la media de los tiempos de atención a pacientes obtenidos en un tiempo determinado en ambas. El valor de la mediana o media más pequeña correspondiente a una de las clínicas reflejará una mayor eficiencia en la atención a los pacientes; mientras que aquella clínica que muestre un valor calculado de la mediana o media mayor para los tiempos de atención de sus pacientes mostrará ser más ineficiente.

Medidas de dispersión

Una medida de tendencia central sólo proporciona información parcial sobre un conjunto de datos, ya que no permite cuantificar que tanto las observaciones difieren unas de otras. Con la finalidad de aclarar este punto consideremos las siguientes tres muestras:

1. 20, 40, 50, 30, 60, 70
2. 47, 43, 44, 46, 20, 70
3. 44, 43, 40, 50, 47, 46

20 30 40 50 60 70

Media = Mediana

98

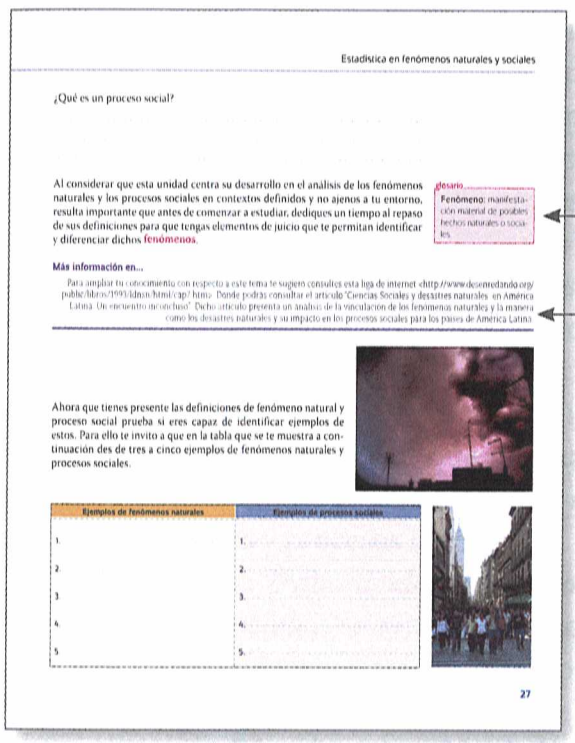
Gestión del aprendizaje. Ofrece información que orienta tu estudio. Puede presentar explicaciones de carácter teórico, estrategias de aprendizaje y técnicas de estudio.

4º. Para evaluar las actividades que realices está el primer apéndice del libro. En él encontrarás la clave de respuestas a las actividades y la retroalimentación; es importante que lo consultes para confrontar tus resultados y reafirmar la correcta resolución de las mismas.

5º. También encontrarás una sección de evaluación final titulada *¿Ya estoy preparado(a)?* Su resolución te permitirá valorar si ya lograste los aprendizajes propuestos y si estás en condiciones de presentar tu examen para acreditar el módulo ante la SEP. Califica honestamente tus respuestas y una vez que tengas los resultados piensa en las estrategias de estudio que te funcionaron y en las que no para que mejores tu proceso de aprendizaje.

Con frecuencia, en el libro se recomienda buscar información en Internet por medio de un buscador o acceder a páginas electrónicas. Por la velocidad con que se actualiza la información en Internet encontrarás que algunas ya no están disponibles, por lo

Glosario. Define términos que pueden ser desconocidos o de difícil comprension.


Estadística en fenómenos naturales y sociales

¿Qué es un proceso social?

Al considerar que esta unidad centra su desarrollo en el análisis de los fenómenos naturales y los procesos sociales en contextos definidos y no ajenos a tu entorno, resulta importante que antes de comenzar a estudiar, dediques un tiempo al repaso de sus definiciones para que tengas elementos de juicio que te permitan identificar y diferenciar dichos **fenómenos**.

Glosario
Fenómeno: manifestación material de posibles hechos naturales o sociales.

Más información en...

Para ampliar tu conocimiento con respecto a este tema te sugiero consultes esta liga de internet <http://www.desenredando.org/public/libros/1999/ldnaa/html/cap7.htm>. Donde podrás consultar el artículo "Ciencias Sociales y desastres naturales en América Latina. Un encuentro inconcluso". Dicho artículo presenta un análisis de la vinculación de los fenómenos naturales y la manera como los desastres naturales y su impacto en los procesos sociales para los países de América Latina.

Ahora que tienes presente las definiciones de fenómeno natural y proceso social prueba si eres capaz de identificar ejemplos de estos. Para ello te invito a que en la tabla que se te muestra a continuación des de tres a cinco ejemplos de fenómenos naturales y procesos sociales.

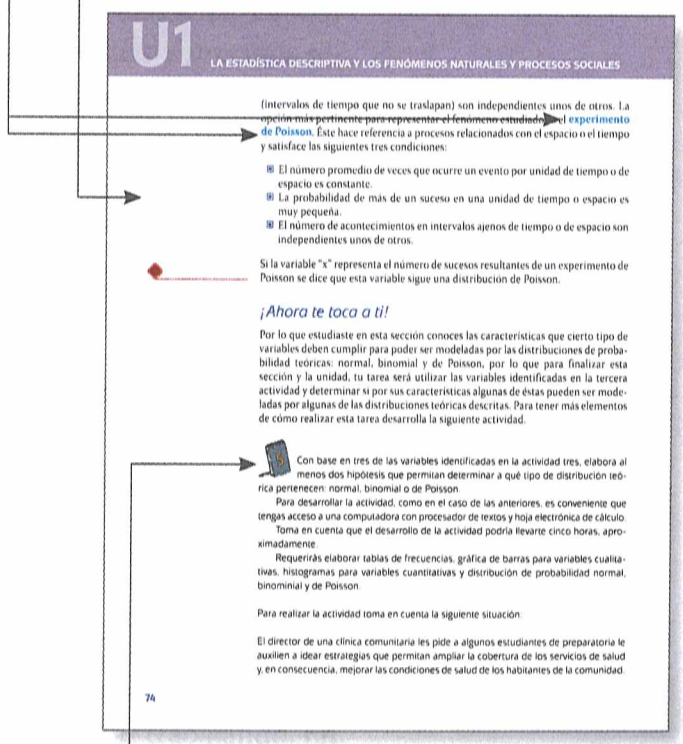
Ejemplos de fenómenos naturales	Ejemplos de procesos sociales
1.	1.
2.	2.
3.	3.
4.	4.
5.	5.

27

Más información en. Presenta sugerencias de documentos, en soporte impreso o digital, a los que puedes recurrir para ampliar tus conocimientos.

Concepto clave. A lo largo del libro se resaltan con azul los términos esenciales para la comprensión de la situación o el tema que estás analizando.

Alto. Te sugiere momentos para interrumpir el estudio sin dejar un proceso de aprendizaje incompleto.


U1 LA ESTADÍSTICA DESCRIPTIVA Y LOS FENÓMENOS NATURALES Y PROCESOS SOCIALES

(intervalos de tiempo que no se traslapan) son independientes unos de otros. La opción más pertinente para representar el fenómeno estudiado es el **experimento de Poisson**. Éste hace referencia a procesos relacionados con el espacio o el tiempo y satisface las siguientes tres condiciones:

- El número promedio de veces que ocurre un evento por unidad de tiempo o de espacio es constante.
- La probabilidad de más de un suceso en una unidad de tiempo o espacio es muy pequeña.
- El número de acontecimientos en intervalos ajenos de tiempo o de espacio son independientes unos de otros.

Si la variable "x" representa el número de sucesos resultantes de un experimento de Poisson se dice que esta variable sigue una distribución de Poisson.

¡Ahora te toca a ti!

Por lo que estudiaste en esta sección conoces las características que cierto tipo de variables deben cumplir para poder ser modeladas por las distribuciones de probabilidad teóricas: normal, binomial y de Poisson, por lo que para finalizar esta sección y la unidad, tu tarea será utilizar las variables identificadas en la tercera actividad y determinar si por sus características algunas de éstas pueden ser modeladas por algunas de las distribuciones teóricas descritas. Para tener más elementos de cómo realizar esta tarea desarrolla la siguiente actividad.

Con base en tres de las variables identificadas en la actividad tres, elabora al menos dos hipótesis que permitan determinar a qué tipo de distribución teórica pertenecen: normal, binomial o de Poisson.

Para desarrollar la actividad, como en el caso de las anteriores, es conveniente que tengas acceso a una computadora con procesador de textos y hoja electrónica de cálculo.

Toma en cuenta que el desarrollo de la actividad podría llevarte cinco horas, aproximadamente.

Requerirás elaborar tablas de frecuencias, gráfica de barras para variables cualitativas, histogramas para variables cuantitativas y distribución de probabilidad normal, binominial y de Poisson.

Para realizar la actividad toma en cuenta la siguiente situación:

El director de una clínica comunitaria les pide a algunos estudiantes de preparatoria le auxilien a idear estrategias que permitan ampliar la cobertura de los servicios de salud y, en consecuencia, mejorar las condiciones de salud de los habitantes de la comunidad.

74

Actividad. Encontrarás una gran diversidad de actividades con las que desarrollarás tus competencias. Lee las instrucciones con atención y ejecútalas.

que saber buscar (navegar) te será muy útil. Si tienes alguna duda sobre cómo hacerlo, consulta el Apéndice 6. "La consulta de fuentes de información en Internet".

A lo largo del texto también encontrarás una serie de elementos gráficos que pretenden ser ayudas en la gestión de tu aprendizaje.

Conforme avances en tu estudio del módulo *Estadística en fenómenos naturales y procesos sociales* identificarás cuáles de estos recursos son útiles según tus capacidades para aprender y tu estilo de aprendizaje.

¡Aprovéchalos para sacar el mayor beneficio de este libro!

Los saberes del módulo fueron seleccionados para que aprendas a registrar, organizar y analizar información mediante herramientas estadísticas –tablas, gráficas y modelos matemáticos– con el fin de estimar la probabilidad de ocurrencia de eventos relacionados con fenómenos naturales y procesos sociales. El objetivo es que ese conocimiento facilite la toma de decisiones individuales y colectivas en tu comunidad, región, país y mundo.

Las competencias que desarrollarás con el estudio de *Estadística en fenómenos naturales y procesos sociales* se relacionan con la aplicación de los principios y el manejo de las herramientas propias de la estadística descriptiva. También incrementarás tus conocimientos para buscar y recopilar datos así como para clasificarlos de acuerdo con sus características. De manera paralela, incrementarás tu capacidad de expresión escrita al reportar o informar los resultados de tus análisis.

Debido al carácter interdisciplinario del módulo trabajarás aquellas competencias que posibiliten tu interpretación y análisis de fenómenos naturales y procesos sociales. Te allegarás de herramientas matemáticas tales como gráficas, tablas y modelos. Te proveerás de los saberes que posibiliten el estudio de sequías, huracanes, enfermedades, sismos y crecimiento de las poblaciones; y harás uso de las competencias que refuercen tu conocimiento sobre la migración, la pobreza, la estructura poblacional y el acceso a la educación, entre otros.

Los saberes necesarios para tu proceso de aprendizaje se disponen en dos unidades. En la primera, *La estadística en los fenómenos naturales y procesos sociales*, se desarrollan los principios básicos de la estadística; las características de los fenómenos naturales y los procesos sociales así como los eventos determinísticos y aleatorios en ellos. Además, se identifican diferentes tipos de variables y sus posibles relaciones; también se estudian distintas técnicas de muestreo y algunas distribuciones de probabilidad que por sus características resultan ser modelos teóricos adecuados para la modelación de eventos aleatorios relacionados con fenómenos naturales y procesos sociales.

En la unidad 2, *Tratamiento estadístico de la información de fenómenos naturales y procesos sociales*, se retoman elementos estudiados en la unidad anterior para aplicar herramientas estadísticas que permitan ordenar, procesar e interpretar la información. Se profundiza el estudio de las distribuciones de probabilidad para su uso como modelos de estimación de la probabilidad de ocurrencia de eventos aleatorios. Se estudia el grado de asociación entre variables.

Además de la información sobre los temas anteriores, encontrarás actividades de aplicación de los saberes, no dejes de resolverlas y corroborar su resolución en el Apéndice 1 pues ello te dará seguridad y confianza para continuar trabajando. En ocasiones, se sugerirá el uso de programas ofimáticos y la localización de información a través de Internet por eso es recomendable que tengas acceso a una computadora con conexión.

No estás solo en tu tarea de aprender. En el libro encontrarás ejemplos y procedimientos a seguir para auxiliarte en el desarrollo de competencias diversas. Sin embargo, es importante aclarar que en el camino hallarás obstáculos que tendrás que enfrentar pues el conocimiento no se logra en el primer intento, necesariamente. Reflexiona de manera continua en lo que estás haciendo y cómo lo estás llevando a cabo.

Analiza de forma constante tu proceso de enseñanza-aprendizaje para determinar el grado de desarrollo que durante el estudio del módulo hayas tenido de las competencias correspondientes. Para auxiliarte a hacerlo, tanto en diversas actividades como al final de cada una de las unidades te serán propuestos algunos instrumentos de evaluación. Si tu desempeño no es el óptimo, pide asesoría para detectar aciertos y errores y convertir los últimos en oportunidad de mejoramiento permanente.

El diálogo es un componente esencial en la construcción del conocimiento por lo que si tienes oportunidad acude a los Centros de Servicios de Preparatoria Abierta y estudia con otros compañeros y con los asesores. Si puedes, también dialoga con las fuentes de información que consultes ya que con su lectura generarás nuevas ideas, podrás construir argumentos y sustentar interpretaciones; acostúmbrate a cuestionar los libros y documentos que leas y a buscar en ellos respuestas pero también a plantear nuevas interrogantes.

El módulo está diseñado para 75 horas de estudio, destinarás 30 para la primera unidad y las 45 restantes a la segunda. Es recomendable que organices tu tiempo y le dediques dos horas diarias de lunes a viernes, con el fin de terminarlo en ocho semanas, aproximadamente.

Como te habrás dado cuenta posees saberes que te permiten avanzar en la construcción de nuevos conocimientos pero también habrás observado que si no tienes los elementos necesarios para emprender nuevos estudios el inicio se dificulta. Comprueba con la resolución de la siguiente prueba si ya estás preparado para comenzar.

I. Lee el siguiente caso y responde las situaciones descritas.

> A un alumno de primer semestre de la licenciatura en Biología se le pide estudiar el dengue ya que formará parte del equipo de divulgación de la campaña sobre la prevención de dicho padecimiento, en Veracruz.

1. Al buscar información sobre el tema, encuentra el siguiente párrafo en un buscador de Internet.

> "El dengue es una de las Enfermedades Transmitidas por Vectores (ETV) con mayor presencia en el mundo. Su aparición se asocia al ambiente urbano doméstico, a los hábitos de la población y a la carencia de servicios básicos como el suministro de agua, falta de recolección de basura y desechos de la vivienda".

Al leer se da cuenta que es la introducción a:

a) Un reportaje periodístico de la sección "Ciencia".
b) Un artículo de divulgación científica de una revista.
c) Un reporte de investigación de un biólogo.
d) Una nota informativa sobre un avance científico.

2. Por la lectura del párrafo anterior, el estudiante infiere que los factores naturales y sociales que originan un fenómeno como el dengue son:
   a) Urbanización sin planear, prevalencia de lluvias, falta de servicios públicos adecuados.
   b) Falta de servicios públicos adecuados, pobreza e ignorancia y un mosquito o vector.
   c) Falta de servicios públicos adecuados, urbanización sin planear, falta de hábitos de higiene.
   d) Presencia del mosquito o vector, mala urbanización y prevalencia de lluvias.

3. Al continuar con su investigación, el estudiante encuentra la siguiente información.

El dengue es una enfermedad infecciosa aguda de etiología viral, transmitida por la picadura de la hembra de mosquitos del género Aedes en donde el único reservorio del virus es el hombre. El agente etiológico es el dengue virus con cuatro vertientes o serotipos: 1, 2, 3 y 4. La infección puede manifestarse desde un cuadro asintomático, fiebre indiferenciada, fiebre por dengue clásico, dengue hemorrágico y hasta el síndrome de choque por dengue.

El mosquito transmisor del dengue es un ejemplo de adaptación de una especie al ámbito humano, con criaderos, hábitats, fuente de alimentación y desplazamiento ligados al entorno domiciliario. Aunque la enfermedad depende de la presencia y abundancia de los mosquitos vectores la transmisión es función de los diferentes serotipos del virus y de la cantidad de personas susceptibles o inmunes. Las condiciones para su proliferación se presentan en una sociedad donde prevalecen la falta de conciencia, el desconocimiento y la actitud de las familias en el control y eliminación de criaderos, así como la carencia de prácticas de auto protección, como el uso de mosquiteros y de insecticidas.

Con base en la información anterior, el estudiante deduce que la forma de erradicar o controlar la infección podría ser:

a) Impedir los encharcamientos y las reservas inadecuadas de agua en la fase reproductiva del mosquito así como el uso frecuente de insecticidas.
b) Evitar la picadura por el mosquito y el uso frecuente de insecticidas.
c) Procurar hábitos que no generen al vector un hábitat de adaptación.
d) Establecer dos o más acciones que eviten la proliferación del mosquito.

4. Como un paso más en su investigación, encuentra el siguiente párrafo relacionado con las características de la enfermedad del dengue:

La enfermedad con manifestación no grave o moderada concluye en la recuperación del paciente de 5 a 10 días, después del inicio. La manifestación severa DH se caracteriza por una permeabilidad vascular ocasionada por liberación de citocinas cuando los linfocitos T atacan las células infectadas con el virus del dengue. De no atenderse oportunamente las formas graves de la enfermedad, el paciente puede presentar un desenlace fatal. El dengue hemorrágico y sus complicaciones son cada vez más frecuentes y pueden surgir a partir de respuestas secundarias de las células T, es decir, cuando se presenta una segunda infección es común observarla en infantes con bajas concentraciones de anticuerpos maternos.

¿Cuáles de las siguientes acciones propias de un estudio científico identificó el estudiante para llegar a las conclusiones consideradas en el texto?

a) Observación, hipótesis, experimentación, reflexión, comprobación, conclusiones.
b) Planteamiento de un problema, elaboración de hipótesis, comprobación de la hipótesis, obtención de resultados.
c) Observación del fenómeno, elaboración de hipótesis, comprobación de hipótesis, reflexión de la comprobación, formulación de principios y leyes.

d) Hipótesis, comprobación, observación, reflexión, obtención de resultados.

5. Con la finalidad de ahondar más sobre los factores ambientales que inciden directamente en la prevalencia de la enfermedad del dengue, el estudiante tuvo acceso a la siguiente información:

> Dentro de los factores que favorecen la presencia y el aumento de los casos de dengue destacan los relacionados con el cambio climático que influyen en la intensidad y la duración de las temporadas de lluvia y frecuencia de huracanes; deficientes servicios urbanos como dotación de agua e inadecuada disposición final de la basura; así como la falta de percepción de riesgos y nulas medidas de autoprotección y prevención de la población. Otros factores específicos, dependientes de la población y los propios responsables de la aplicación de medidas antivectoriales, como fallas logísticas y operativas, cobertura insuficiente y falta de una estrategia de manejo integrado de vectores, se relacionan también con la persistencia de la enfermedad.

Con esa información el estudiante concluye que el cambio climático influye directamente en la proliferación de dengue debido a:

a) La falta de estrategias para el manejo de los vectores.
b) El deshielo de los polos que provoca corrientes frías que favorecen la formación de huracanes.
c) La intensidad y duración de lluvias causadas por la presencia de huracanes que favorecen los encharcamientos donde el vector se reproduce.
d) La duración de la temporada de lluvias.

6. Dada la naturaleza del entorno social y ambiental en el que la enfermedad se desarrolla, selecciona cuáles de las siguientes instituciones gubernamentales pudieran considerarse como las idóneas para intervenir en una campaña de tratamiento y prevención de la enfermedad.

a) Los gobiernos federal, estatal, municipal y la población civil.
b) La Secretaría de Salud, la SEMARNAP y la Secretaría de Educación Pública.
c) La Secretaría de Salud, Conagua y la Secretaría de Educación Pública.
d) Los gobiernos federal, estatal y municipal.

7. Al investigar sobre la magnitud de la enfermedad, el estudiante encuentra los siguientes datos:

> El número de casos de dengue aumenta a medida que la enfermedad se propaga a nuevas zonas ya que se producen brotes epidémicos de carácter explosivo. Cada año aproximadamente 500 000 personas que padecen dengue grave –niños en una gran proporción— necesitan hospitalización y fallecen alrededor de 2.5%.

¿Qué cantidad de las 500 000 personas citadas que padecen dengue grave representa el 2.5%?

a) 250 000
b) 50 000
c) 10 000
d) 12 500

8. Con la finalidad de cuantificar la incidencia de dengue en el estado de Veracruz, el estudiante recurre a datos de prevalencia de la enfermedad, los cuales muestran que la tendencia de este padecimiento en los últimos años ha ido a la alza. El siguiente cuadro presenta los datos anuales de la tasa de incidencia por cada 100 000 habitantes, el número de casos de dengue clásico y la tendencia del mismo de continuar el cambio climático y no aplicarse las medidas de control.

AÑO	CASOS	TASA POR 100 MIL HABITANTES
2000	578	8.10
2001	2 228	31.10
2002	5 094	70.70
2003	1 274	17.57
2004	5 761	79.00
2005	11 761	161.70
2006	8 537	116.70
2007	10 037	141.24

Nota: Hasta la semana epidemiológica número 44.

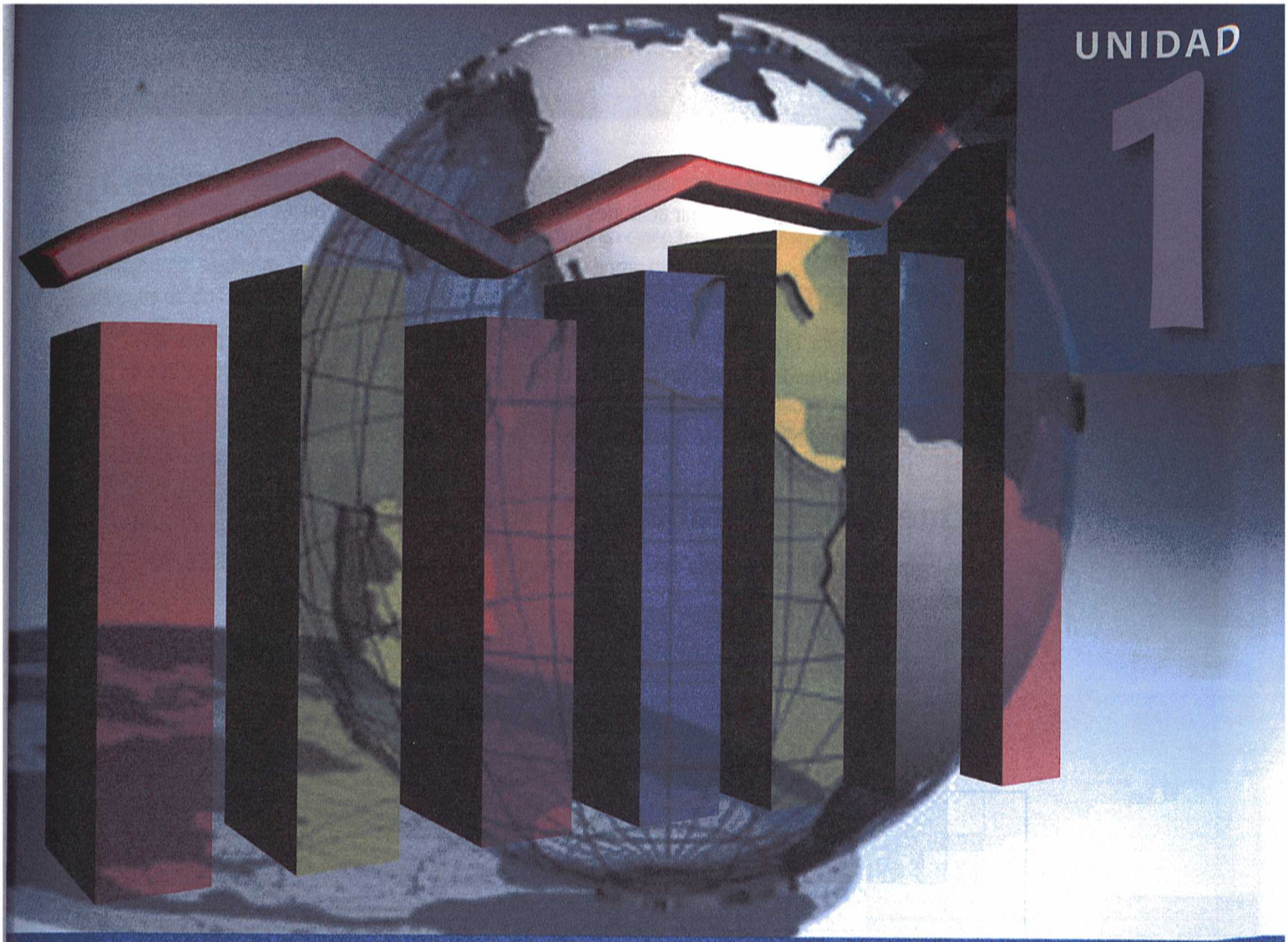
Fuente: Servicios de salud del Gobierno Veracruz.
Departamento de Control de Enfermedades Transmitidas por vector.

Grafica en el plano cartesiano las parejas ordenadas correspondientes al año y la tasa por habitantes, con base en la información de la tabla. Considera al eje de la x como el año y al de la y como la tasa por habitantes.

9. Con base en los datos de la tabla anterior responde: ¿Cuál fue el promedio de casos de dengue clásico en el periodo de ocho años, de 2000 a 2007?
   a) 2421.22
   b) 1351.29
   c) 2236.60
   d) 5658.75

10. Si en la tabla el año 2000 se representa x = 0 y si se considera a y como el número de casos de dengue clásico. Entonces la pareja ordenada de los casos de dengue clásico presentes en 2000 se pude representar por la pareja ordenada (0,578). De esta misma forma la pareja ordenada correspondiente al número de casos de dengue clásico observados en 2007 puede ser representado por la pareja ordenada (7,10037). Con base en estas dos parejas ordenadas aplica la fórmula de la ecuación de la recta punto pendiente para determinar la ecuación de la recta en su forma pendiente-intersección que pasa por estos dos puntos del plano cartesiano.
   a) $y = -1321.29x - 578$
   b) $y = 1321.29x + 578$
   c) $y = 578x + 1321.29$
   d) $y = -578x - 1321.29$

En el diseño de la actividad anterior se tomaron en cuenta saberes (conceptuales, procedimentales y actitudinales) de las áreas de Comunicación, Matemáticas, Ciencias experimentales y Ciencias sociales y humanidades. Verifica tus respuesta en el Apéndice 1 y si no obtuviste más de siete aciertos es recomendable que repases los saberes en los que no hayas obtenido un desempeño satisfactorio. Puedes consultar los materiales de módulos anteriores como Representaciones simbólicas y algoritmos, Universo natural, Ser social y sociedad y El lenguaje en la relación del hombre con el mundo.



UNIDAD 1

La estadística descriptiva y los fenómenos naturales y procesos sociales

¿Qué voy a aprender y cómo?

¿Crees que es factible predecir con cierto nivel de certeza el clima para la próxima semana?, ¿conoces algún método que permita medir la probabilidad de que llueva el próximo jueves?, ¿podrías nombrar algún instrumento que posibilite medir el impacto económico o migratorio derivado de una sequía prolongada?

Las respuestas a las preguntas anteriores te dan un panorama de qué tan familiarizado estás con las técnicas de medición de la incertidumbre asociada a la ocurrencia de un evento que puede afectar la vida de los seres humanos. Si no sabes mucho sobre el tema… no te preocupes. En esta unidad aprenderás a usar los métodos estadísticos y el cálculo de las probabilidades, herramientas básicas para el manejo de la incertidumbre generada por la falta de certeza sobre el lugar o el momento de ocurrencia de un fenómeno natural y un proceso social.

Te podrás preguntar, ¿por qué centrarnos en la interpretación y el análisis de los fenómenos naturales y procesos sociales? La razón es simple, porque están presentes en el espacio donde tú y yo nos movemos, donde vivimos, trabajamos, paseamos y realizamos cualquier actividad de convivencia cotidiana; este entorno al que pueden afectar la lluvia, las inundaciones, los terremotos, las erupciones volcánicas, etcétera. En este espacio en el que pueden estar presentes la pobreza, la marginación, el desempleo, la migración, etcétera.

Es en la cotidianidad de lo natural y lo social en donde nos desenvolvemos y de la cual no somos ajenos. Su entendimiento conlleva no sólo a la búsqueda de información documental sino a la medición de su impacto a partir de la aplicación de métodos estadísticos y el cálculo de probabilidades. Su conocimiento profundo además de sensibilizarnos, permite nuestra participación de

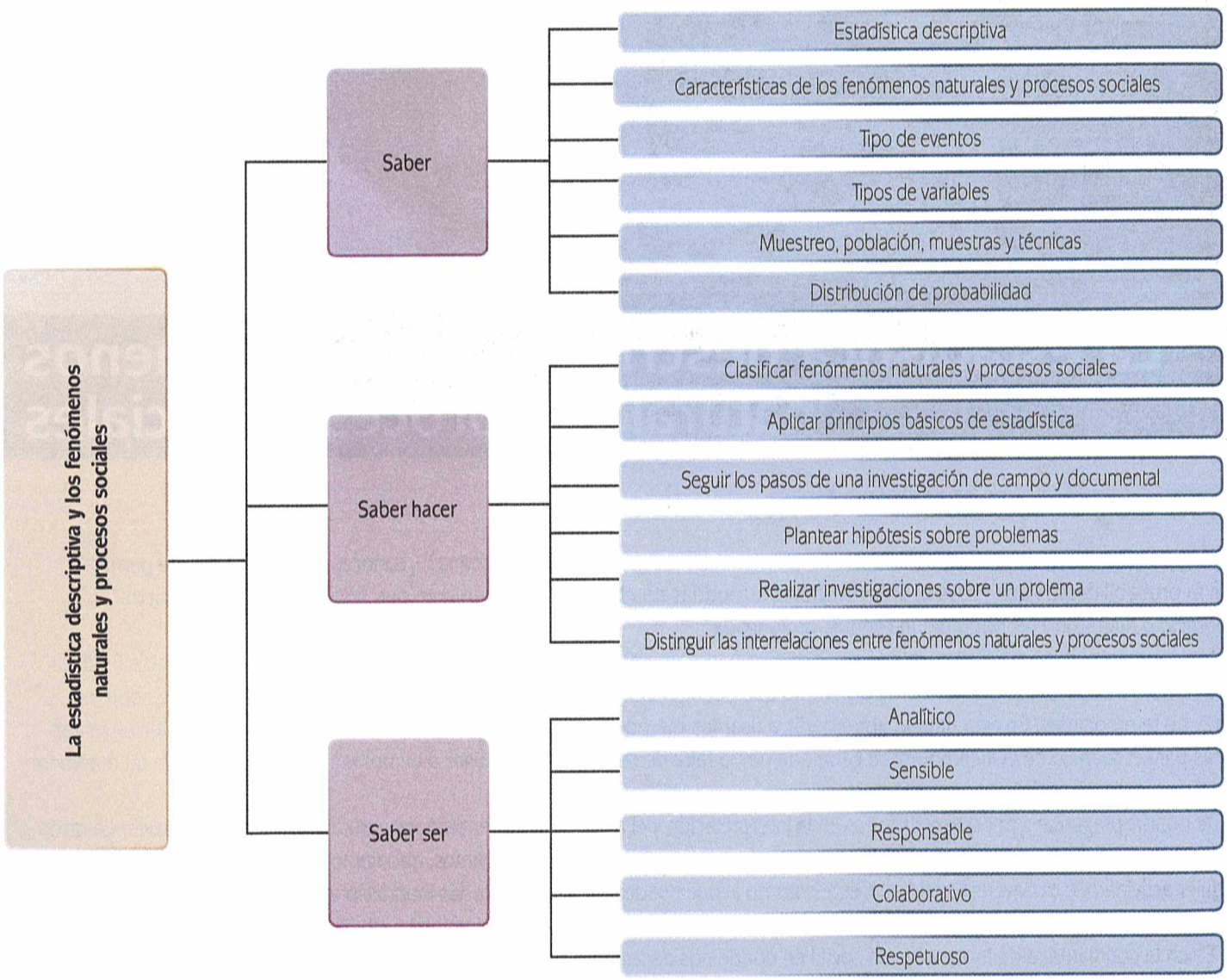
manera colaborativa ante la ocurrencia y el comportamiento de fenómenos naturales y procesos sociales. Para ello, debemos asumir actitudes favorables para el trabajo y convivencia social.

¿Con qué propósito?

De manera concreta, el propósito de la unidad es que analices fenómenos naturales y procesos sociales de tu entorno mediante los conceptos básicos de la probabilidad y la estadística (tipos de eventos, variables, muestreo, gráficas) para reconocer los diferentes tipos de distribución de la información y con ello explicar su comportamiento en un contexto determinado.

¿Qué saberes trabajaré?

Para lograr el propósito formativo descrito es importante que adquirieras y manejes los saberes relacionados a continuación:



¿Cómo organizaré mi estudio?

Con el fin de facilitar el estudio de la unidad se estructuró en cinco temas relacionados en el cronograma siguiente. Es recomendable que planees 30 horas de estudio organizando en 15 sesiones de dos horas cada una. De esa manera si decides estudiar diariamente o en periodos semanales de lunes a viernes, podrás concluir esta unidad en el transcurso de tres semanas.

Tema	1ª semana					2ª semana					3ª semana				
Fenómenos naturales y procesos sociales															
Fenómenos naturales y procesos sociales determinísticos y aleatorios en tu contexto															
Las variables y su importancia como instrumentos de medición para el estudio de eventos aleatorios y sus relaciones															
Los métodos de muestreo y su importancia en el análisis de información															
Las distribuciones de probabilidad y su relación con los fenómenos naturales y procesos sociales															

¿Cuáles serán los resultados de mi trabajo?

Para lograr el propósito de la unidad es importante que aprendas a:

- Clasificar un conjunto de eventos en fenómenos naturales y procesos sociales de acuerdo con sus características específicas con la finalidad de reconocer los de mayor incidencia en tu contexto.
- Identificar los fenómenos naturales y procesos sociales en determinísticos y aleatorios, de acuerdo con la naturaleza de su ocurrencia para reconocer aquellos que se puedan analizar y predecir mediante elementos estadísticos.
- Determinar y analizar los tipos de variables en la interpretación y explicación de fenómenos naturales y procesos sociales.
- Identificar y situar fenómenos naturales y procesos sociales de tu contexto que pueden ser objeto de estudio estadístico y aquellos que no pueden ser estudiados a través de este método.
- Identificar fenómenos naturales y procesos sociales de tu contexto y utilizar técnicas de muestreo para la obtención de datos que permitan interpretar y explicar los fenómenos que ocurren en tu entorno.
- Identificar y explicar los distintos tipos de distribución de probabilidad a través de gráficas.

Con el fin de fortalecer los conocimientos adquiridos durante el estudio de las secciones se elaboraron cinco actividades. Para su desarrollo es conveniente que cuentes con una computadora con procesador de texto, hoja de cálculo y software de presentación.

Antes de empezar

Lo conocido es el cimiento para la adecuada construcción de nuevos saberes, varios de los conceptos que estudiarás a continuación los aprendiste con anterioridad. Comienza por reconocer que tanto dominas los saberes enumerados en la siguiente tabla. Marca con una paloma (✓) aquellos que estás seguro(a) de manejar y con una cruz (✗) los que no. En la última columna anota qué fuente consultarías para su estudio, en caso de requerir reforzar su manejo.

	Sí (✓)	No (✗)	Fuente de consulta para su estudio
Sé			
Cuáles son los tipos de lenguaje.			
Qué es desigualdad social.			
Qué es distribución de la riqueza.			
Definir crecimiento poblacional.			
Definir pobreza.			
El significado de la simbología matemática.			
Cuáles son los métodos de investigación científica para análisis de la sociedad.			
Definir las leyes de la herencia y el principio de Ardí-Weinberg.			
Distinguir la jerarquía de las operaciones matemáticas.			
Sé			
Aplicar los principios de análisis para explicar e interpretar la sociedad de la que formo parte, las instituciones que la integran y cómo es la interacción entre ellas.			
Distinguir las funciones exponenciales y logarítmicas.			
Sumar, restar, multiplicar y dividir.			
Calcular potencias y raíces cuadradas.			
Calcular desigualdades lineales.			
Calcular el valor de ciertas variables.			

	Sí (✓)	No (✗)	Fuente de consulta para su estudio
Aplicar el método científico en cualquier tipo de investigación.			
Soy			
Independiente en mi aprendizaje.			
Autónomo en la toma de decisiones.			
Responsable al asumir las consecuencias de mis actos.			
Honesto porque actúo correctamente ante las circunstancias y no altero los datos.			
Reflexivo por cuestionar los diversos hechos de la realidad que me rodea.			
Crítico al actuar y reaccionar ante cualquier situación.			
Respetuoso ante la diversidad de creencias, costumbres y formas de expresión.			
Solidario al trabajar con otros compañeros.			

¡Consciente de tus fortalezas y debilidades, comenzamos!

INICIO

Los pobladores de la comunidad de Santiago, ubicada en la costa del Pacífico del estado de Guerrero, tienen una alta propensión a las enfermedades transmitidas por mosquitos y moscas; por ello, el Departamento de Seguridad Social y Prevención de Enfermedades y Desastres Naturales (DESPEDEN) decidió elaborar un programa que permita mejorar la calidad de vida de los habitantes de la comunidad.

Para desarrollar el programa, en el plazo de un año, el Departamento necesita recabar información cuyo análisis e interpretación se traduzca en acciones concretas para lograr: (a) una mayor cobertura de los servicios de salud, (b) diseñar campañas de prevención que mitiguen el impacto de las enfermedades dentro de la comunidad (identificando los factores del entorno que pudieran incidir directamente en su prevalencia); (c) establecer medidas preventivas para disminuir el impacto de eventos que por su naturaleza de ocurrencia resulten inciertos de pronosticar y traigan secuelas a las condiciones de vida de los habitantes de Santiago.

El DESPEDEN lanza una convocatoria entre los estudiantes de preparatoria para que participen en el desarrollo del programa. El requisito único es que manejen herramientas estadísticas. Interesado en la convocatoria reflexionas: ¿dónde puedo desarrollarlas? Estudiando esta unidad.[1]

La estadística y sus principios básicos

Hablar de fenómenos naturales como la lluvia o las tormentas y de procesos sociales como el crecimiento de la población, el desempleo o el analfabetismo es común. Habrás oído explicaciones y comentarios sobre ellos en los medios de comunicación de masas o entre la gente. Tal vez hayas escuchado: "El pronóstico del tiempo para hoy será de 23° C" o quizá leíste en un periódico: "Para el cierre de 2012 habrá más de 700 000 nuevos empleos". Pero, ¿cuál es el sustento de tales aseveraciones?, ¿cómo pronosticar el comportamiento de los huracanes, las lluvias, los temblores que afectan a las comunidades? Por medio del estudio científico y las herramientas matemáticas. De manera específica, mediante la **Estadística**, ciencia matemática que se sirve de los conjuntos de datos para obtener, a partir de ellos, conclusiones basadas en el cálculo de probabilidades.

El uso de la Estadística es común en diversas ciencias y situaciones cotidianas pues permite obtener información detallada de todo tipo de variables sociales y naturales.

Fenómenos naturales y procesos sociales: sus características

Recuerdas...

¿Qué es un fenómeno natural?

Para saber más

La etimología del término Estadística proviene de status, término latino que significa estado. El concepto se asocia desde la Antigüedad con la necesidad de los Estados de recopilar datos con diversos propósitos: militar, fiscal, agrícola, entre otros.

1 Si prefieres puedes diseñar un caso distinto a éste. Te sugerimos plantearlo con variables propias del contexto de tu comunidad. Si prefieres esta opción entonces sustituye las referencias a la comunidad de Santiago por las observaciones de tu comunidad pero mantén las indicaciones y los lineamientos que se plantean.

¿Qué es un proceso social?

Al considerar que esta unidad centra su desarrollo en el análisis de los fenómenos naturales y los procesos sociales en contextos definidos y no ajenos a tu entorno, resulta importante que antes de comenzar a estudiar, dediques un tiempo al repaso de sus definiciones para que tengas elementos de juicio que te permitan identificar y diferenciar dichos **fenómenos**.

glosario

Fenómeno: manifestación material de posibles hechos naturales o sociales.

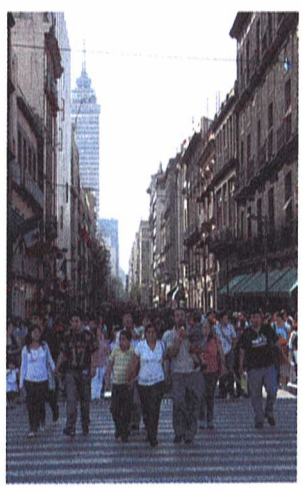
Más información en...

Para ampliar tu conocimiento con respecto a este tema te sugiero consultes esta liga de internet <http://www.desenredando.org/public/libros/1993/ldnsn/html/cap7.htm>. Donde podrás consultar el artículo "Ciencias Sociales y desastres naturales en América Latina: Un encuentro inconcluso". Dicho artículo presenta un análisis de la vinculación de los fenómenos naturales y la manera como los desastres naturales y su impacto en los procesos sociales para los países de América Latina.



Ahora que tienes presente las definiciones de fenómeno natural y proceso social prueba si eres capaz de identificar ejemplos de estos. Para ello te invito a que en la tabla que se te muestra a continuación des de tres a cinco ejemplos de fenómenos naturales y procesos sociales.

Ejemplos de fenómenos naturales	Ejemplos de procesos sociales
1.	1.
2.	2.
3.	3.
4.	4.
5.	5.



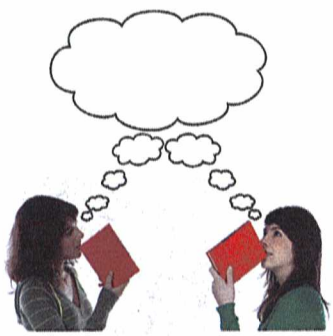
Fenómenos naturales y procesos sociales en tu contexto

Estás trabajando para clasificar un conjunto de eventos en fenómenos naturales y procesos sociales de acuerdo con sus características específicas con la finalidad de reconocer los de mayor incidencia en tu contexto.

glosario

Eventos: hechos o consecuencias de la ocurrencia de un fenómeno.

A partir de las definiciones y los ejemplos de fenómenos naturales y procesos sociales que has identificado, es momento de que comiences con un ejercicio de reflexión que te permita identificar qué tipo de fenómenos naturales y procesos sociales inciden directamente en tu contexto. Para facilitar tu ejercicio de reflexión es importante te respondas las siguientes preguntas relacionadas con tu entorno:



- ¿Cuáles son las características geográficas y sociales más características de mi comunidad?
- ¿Qué fenómenos naturales y procesos sociales se presentan con mayor frecuencia?
- ¿Cuáles de ellos la afectan más, tanto de manera positiva como de manera negativa?
- ¿Con qué tanta frecuencia se presentan estos fenómenos y procesos?

Estas preguntas están encaminadas al desarrollo de competencias que te permitan identificar las características de los procesos sociales y los fenómenos naturales que más influyen en tu contexto.

Una vez que has realizado el ejercicio de reflexión escribe en la tabla siguiente ejemplos de fenómenos naturales y procesos sociales que consideres influyentes en tu contexto.

Ejemplos de fenómenos naturales que por su naturaleza puedan influir en tu contexto	Ejemplos de procesos sociales que por su naturaleza pueden influir en tu contexto



Supongamos por un momento que a uno de los estudiantes de preparatoria que respondió a la convocatoria del DESPEDEN se le encomienda la tarea de identificar un cierto número de fenómenos naturales y de procesos sociales que por sus características pudieran considerarse significativos para su contexto. Además se le pide que jerarquice de mayor a menor (por su grado de incidencia en su entorno) a aquéllos fenómenos naturales y procesos sociales que pudieran presentar una mayor frecuencia de ocurrencia en su contexto.

A partir de una lista de 16 fenómenos característicos de la comunidad costera, elabora una tabla en la que clasifiques fenómenos naturales y procesos sociales. Posteriormente, a partir de esta lista selecciona cuatro fenómenos naturales y cuatro procesos sociales que por las características del entorno geográfico puedas considerar son los de mayor influencia en el espacio geográfico planteado. Una vez hecho esto, enlista los fenómenos naturales y procesos sociales en orden ascendente para establecer la jerarquía que por guardan en relación con el grado de influencia dentro del contexto señalado al comienzo.[2]

Para el desarrollo de esta actividad es conveniente contar con una computadora con un programa de procesador de textos y un programa de hoja electrónica de cálculo.

El tiempo estimado para su desarrollo es de dos horas y los pasos para llevarla a cabo son los siguientes:

1. En la tabla se muestra una lista de fenómenos que por sus características pueden ser ubicados en el entorno geográfico de una comunidad costera. En la columna de la derecha clasifica cada uno de estos como fenómeno natural o proceso social.

2 Como se te mencionó al inicio de esta unidad, si prefieres realizar tu propio proyecto, identifica y clasifica dentro de tu comunidad cuatro fenómenos naturales y cuatro procesos sociales que incidan directamente en ti y tus conocidos.

Fenómeno	Clasificación
Huracanes	
Inseguridad	
Lluvias	
Escolaridad en la comunidad	
Obesidad en la comunidad	
Inundaciones	
Dengue en la comunidad	
Temperatura	
Pobreza y marginación	
Sequía	
Crecimiento demográfico	
Epidemias	
Migración	
Inmigración	
Enfermedades en la comunidad	
Natalidad en la comunidad	
Servicios de salud en la comunidad	

2. A partir de la tabla de clasificación de eventos selecciona cuatro fenómenos naturales y cuatro procesos sociales que por sus características consideres que pueden incidir en el contexto referido.
3. Redacta un informe con las siguientes características y recuerda que éste debe ser guardado en tu portafolio de evidencias.

Extensión. Una cuartilla como máximo.
Introducción. En esta apartado describe brevemente qué son los fenómenos naturales y qué los procesos sociales.
Desarrollo. Anota los argumentos que te llevaron a clasificar los fenómenos naturales y los procesos sociales en el contexto considerado; también expón las razones de la jerarquía que utilizaste para clasificarlos por su grado de influencia en el contexto referido.
Conclusión. Explica por qué consideras importante identificar y jerarquizar por su grado de impacto en tu entorno a los fenómenos naturales y los procesos sociales.[3]

3 Como una actividad alternativa puedes realizar ésta basándote en la identificación de 16 fenómenos característicos de tu comunidad. Posteriormente debes desarrollar los pasos 1 a 3 de la presente actividad.

Para terminar, en la siguiente tabla marca con una cruz (✗) si consideras que tu desempeño en cada una de las respuestas a evaluar fue muy bueno (MB), regular (R) o malo (M). Asimismo, de ser necesario, indica que se debe mejorar en cada uno de los aspectos que se evalúan.

Aspecto a evaluar	MB	R	M	¿Qué debo mejorar?
Expresé argumentos claros y precisos para definir los fenómenos naturales y los procesos sociales.				
Clasifiqué adecuadamente los fenómenos naturales y los procesos sociales.				
Definí argumentos claros y precisos sobre el criterio que me llevó a identificar, clasificar y jerarquizar adecuadamente los fenómenos naturales y los procesos sociales que considero inciden directamente en mi contexto.				
Expresé argumentos claros con respecto a la importancia que tiene identificar y jerarquizar por su grado de impacto en mi entorno los fenómenos naturales y los procesos sociales.				

La identificación de los fenómenos naturales y los procesos sociales que afectan un contexto auxilian a tomar medidas preventivas que eliminen o al menos mitiguen su impacto dentro de una comunidad. Por ejemplo, si vives en una ciudad ubicada en una región geográfica de alto riesgo sísmico es fácil percatarte que este fenómeno tiene gran probabilidad de que ocurra, permitiéndote con ello tomar medidas preventivas como: asistir a un curso de primeros auxilios, participar activamente en simulacros de evacuación, tener a la mano los números telefónicos de los cuerpos de socorro, etcétera. De esta forma estarás bien preparado para disminuir los riesgos derivados de la realización de un evento de esta naturaleza. Es importante mencionar que el impacto derivado de la ocurrencia de fenómenos naturales o procesos sociales también puede darse en un sentido positivo. Piensa en algunos de estos tipos de fenómenos o procesos que resultan benéficos para tu comunidad y redacta una cuartilla describiéndolos y anotando su influencia benéfica.

No olvides verificar tus respuestas en el Apéndice 1.

Tipos de eventos: determinísticos y aleatorios



En su proceso reflexivo, el estudiante de preparatoria se dio cuenta que algunos de los eventos asociados con cambios en la naturaleza son diarios o ocurren en periodos específicos; por ejemplo, observó que eventos de la naturaleza como el día y la noche, las estaciones del año, se suscitan en periodos específicos, siendo completamente predecibles en cuanto a la hora, día, mes o año en que estarán presentes. Por otra parte, se percató que existen otro tipo de eventos asociados con los fenómenos naturales cuya característica de ocurrencia es esporádica o no ocurre en periodos específicos, lo que los convierte en eventos completamente impredecibles o difíciles de precisar; por ejemplo, los desastres naturales: tsunamis, maremotos, lluvias prolongadas, tornados, terremotos, etcétera.

En el desarrollo de sus reflexiones, los procesos sociales también estuvieron presentes. Observó que la ocurrencia de este tipo de eventos mantiene una lógica parecida; por ejemplo, el ingreso de un trabajador asalariado es un evento totalmente predecible, la naturaleza de ocurrencia de éste lo hace totalmente pronosticable (siempre y cuando no se considere un periodo demasiado extenso donde seguramente por incrementos salariales o por ascenso de puesto el trabajador aumentaría sus percepciones económicas) pues al tener el trabajador un salario fijo, se sabe con toda certeza cuánto será su salario mañana, pasado mañana o dentro de un mes. Por otra parte, el preparatoriano reflexionó sobre el proceso social que considera como evento al número de inmigrantes y migrantes hacia o desde cierta región en un tiempo determinado, observó que por su ocurrencia los flujos de personas que migran hacia o desde determinada región no se pueden anticipar de manera exacta ni en un periodo determinado.

Las observaciones con respecto a la ocurrencia tanto de los eventos naturales como de los procesos sociales nos conduce a realizar una nueva caracterización de dichos fenómenos basada en la ocurrencia de sus eventos asociados. Podemos observar que la ocurrencia de estos eventos está caracterizada por la **certidumbre** o **incertidumbre** de su presencia; por un lado podemos obtener predicciones

certeras y con ello tener la certidumbre de poder pronosticar su observación con exactitud para cualquier periodo determinado, mientras que existen otro tipo de eventos que por su ocurrencia incierta son más difíciles de pronosticar.

Dicha certidumbre o incertidumbre nos lleva a clasificar a este tipo de eventos en **determinísticos** o **aleatorios**. Por lo que ahora estamos en la posibilidad de formalizar estos conceptos a partir de sus definiciones.

Fenómenos naturales y procesos sociales deterministicos y aleatorios

Un fenómeno natural o proceso social es definido como **determinístico**, si su observación u ocurrencia pude ser predicha con exactitud. Por otra parte, un fenómeno natural o proceso social es definido como **aleatorio** si su ocurrencia no puede ser predicha con exactitud. Por ejemplo, el número de personas que se contagiarán de alguna enfermedad respiratoria o la ocurrencia del próximo sismo son eventos cuya ocurrencia no puede ser pronosticada con exactitud.



Estás trabajando para clasificar los fenómenos naturales y procesos sociales en determinísticos y aleatorios de acuerdo con su naturaleza de ocurrencia, para reconocer aquellos que son susceptibles de análisis y predicción con elementos estadísticos.

También diferenciarás los fenómenos naturales y procesos sociales de tu contexto que pueden ser objeto de estudio estadístico de los que no.

glosario

Determinístico: hecho o consecuencia de naturaleza cierta.

Aleatorio: hecho o consecuencia de naturaleza incierta.

Una vez identificada la característica que permite diferenciar a los fenómenos naturales o procesos sociales de naturaleza determinística de aquellos de naturaleza aleatoria, abrimos paso a la discusión sobre la complejidad que representa tratar con la incertidumbre derivada del estudio de los eventos aleatorios. La incertidumbre presente en la ocurrencia de un fenómeno natural o un proceso social es una problemática a la que se enfrentan distintos investigadores de las áreas sociales y naturales. Un meteorólogo, por ejemplo, trabaja con la incertidumbre de cuál será el pronóstico del tiempo para el futuro próximo: ¿lloverá mañana, dentro de una semana o dentro de un mes?, reflejan claramente la incertidumbre que el meteorólogo tiene acerca de la ocurrencia de este fenómeno natural. Por otra parte, un demógrafo pudiera estar interesado en medir el impacto que pudiese tener, en un determinado periodo, el fenómeno migratorio en una región. En este caso el



Existen eventos de interés que por su naturaleza de ocurrencia se pueden considerar aleatorios sin necesariamente ser fenómenos naturales o procesos sociales. Por ejemplo el número ganador de la lotería es un evento aleatorio relacionado con los juegos de azar.

glosario

Factor: término utilizado para referirse a la causa o causas que propicia un cambio en la observación de una respuesta. Por ejemplo, la baja precipitación pluvial es la causa que propicia una disminución en la producción agrícola.

En este caso el factor es la precipitación pluvial.

demógrafo está consciente de que los flujos migratorios están sujetos a muchos **factores** que los hacen difíciles de predecir.

La dificultad de eliminar la incertidumbre sobre la ocurrencia de eventos relacionados con fenómenos naturales y procesos sociales aleatorios, hace importante que los estudiosos cuenten con herramientas que les otorguen elementos para poder entender y de ser posible, pronosticar el comportamiento de un fenómeno natural o un proceso social, a partir del estudio del comportamiento de la incertidumbre relacionada con un evento aleatorio. A este conjunto de herramientas, de métodos y técnicas cuyo objetivo es analizar y predecir dichos eventos, es a lo que a partir de este momento denominaremos **Estadística** y cuyos métodos y técnicas serán descritos a lo largo de esta unidad.

Estadística

Se puede definir como la ciencia referente a la recolección, análisis e interpretación de datos, que busca explicar condiciones regulares en fenómenos de tipo aleatorio.

Para dar paso a la descripción de dichos métodos y técnicas comenzamos con la definición del concepto de probabilidad, cuya comprensión da paso al desarrollo de un conjunto de técnicas encaminadas a cuantificar la incertidumbre asociada con un evento aleatorio.

Concepto de probabilidad

En cualquier evento aleatorio relacionado con un fenómeno natural o un proceso social siempre existe la incertidumbre pues no se sabe si ocurrirá o no. Por eso es importante contar con una medida que permita cuantificar la oportunidad o posibilidad de ocurrencia de un evento aleatorio. La medida que cumple con esta característica es la **probabilidad**, aquella que asigna a la oportunidad o posibilidad de ocurrencia de un evento aleatorio un valor numérico que se encuentra entre 0 y 1 (o entre 0 y 100 si hablamos en términos porcentuales). Dicha medida funciona de la siguiente manera: si el valor numérico asociado a la ocurrencia de un evento aleatorio es 1, representa la certeza de que es totalmente seguro que éste ocurra, o que su probabilidad de ocurrencia sea del 100%; mientras que un valor de 0 indica que es imposible su realización o equivalentemente que se tiene un 0% de probabilidad de que ocurra. Por ejemplo, si la probabilidad de ocurrencia de un evento aleatorio es 1/5, diríamos entonces que hay un 20% de probabilidad de que ocurra.



Asignación de la probabilidad a un evento aleatorio

Uno de los procedimientos para asignar una probabilidad a un evento aleatorio consiste en considerar a ésta como la razón entre el número de casos (sucesos) favorables y el número total de casos (sucesos posibles). Esto es, si denotamos a "A" como un evento aleatorio, la probabilidad del evento aleatorio "A" a la que denotamos por P(A) se obtiene como esta razón, donde "k" es el número de casos favorables que caracterizan el evento y "n" es el número de casos posibles que caracterizan el fenómeno bajo estudio.

A esta técnica de asignación de la probabilidad a un evento aleatorio se le conoce como **enfoque clásico** y es válida siempre y cuando no existan evidencias que lleven a creer que algunos de los eventos aleatorios asociados con el fenómeno bajo estudio puedan tener mayores oportunidades o posibilidades de ocurrencia sobre los demás. Cuando todos los eventos de un espacio muestral tienen la misma probabilidad de ocurrencia se denominan como **eventos equiprobables**.

glosario

Enfoque clásico: enfoque que determina la probabilidad de un evento con base en el supuesto de que todos los eventos asociados con la ocurrencia de un fenómeno tienen la misma probabilidad de ocurrir.

A manera de ejemplo retomemos el caso del meteorólogo. Supongamos que quiere calcular la probabilidad asociada con el evento de que llueva una sola vez durante tres días consecutivos. La combinación de todos los eventos posibles se muestran en la siguiente tabla.

Evento	Día 1	Día 2	Día 3
1	Llueve	Llueve	Llueve
2	Llueve	Llueve	No llueve
3	Llueve	No llueve	Llueve
4	No llueve	Llueve	Llueve
5	Llueve	No llueve	No llueve
6	No llueve	Llueve	No llueve
7	No llueve	No llueve	Llueve
8	No llueve	No llueve	No llueve

Partiendo del enfoque clásico, como el número de casos favorables que corresponden al evento de que llueva una sola vez durante tres días consecutivos es 3 (los correspondientes a los casos 5, 6 y 7) y el número total de casos (como lo muestra la tabla anterior) son 8, entonces la probabilidad de este evento queda determinada por 3/8, o equivalentemente se tiene un 37.5% de probabilidad de que llueva una vez durante tres días consecutivos.

Es importante comentar que el ejemplo anterior es una simplificación de la realidad y que sólo se planteó para ejemplificar el uso de la asignación de probabilidades de eventos desde el enfoque o paradigma clásico de la probabilidad. Por



Para saber más El origen del estudio de la probabilidad se remonta al análisis de los resultados de los juegos de azar realizado durante los siglos XVI y XVII por los matemáticos Blaise Pascal y Pierre de Fermat.

su complejidad la modelación de los fenómenos meteorológicos requiere la aplicación de otros artificios probabilísticos que van más allá de los utilizados por el enfoque clásico. Existen otros dos enfoques para asignación de la probabilidad: el **enfoque frecuentista** y el **enfoque subjetivo**. La naturaleza de estos enfoques se deja como actividad de investigación para el final de esta sección.

Como ya hemos mencionado, hay cierto tipo de fenómenos naturales y procesos sociales cuya ocurrencia puede tener un impacto directo en nuestro contexto, es por ello que una vez identificados seamos capaces de cuantificar su impacto hacia nuestro entorno, generando conciencia acerca de las implicaciones que su ocurrencia tiene hacia nuestro contexto. Si bien es cierto que la simple percepción posibilita que tengamos un aceptable nivel de discernimiento para intuir de alguna manera el grado de ocurrencia de un evento, es importante contar con una o varias medidas objetivas que además de cuantificar su impacto nos sirvan como instrumentos para validar o refutar nuestras percepciones.

Desde el punto de vista intuitivo es lógico pensar que para medir el impacto de un fenómeno natural o un proceso social sobre el entorno sólo es necesario estudiar la frecuencia con la que se presenta. En este sentido, resulta razonable considerar como medidas objetivas del impacto de un evento a aquellas que involucren el cálculo de su frecuencia observada. Dichas medidas son conocidas como frecuencia absoluta y frecuencia relativa.

Más información en...

En la página web <http://www.youtube.com/watch?v=J-B-Enaphw4> podrás encontrar un video con un ejemplo de cálculo de la probabilidad de un evento desde el punto de vista clásico.

Para definir los términos frecuencia absoluta y frecuencia relativa consideremos que tenemos un fenómeno aleatorio bajo estudio (que puede ser un fenómeno natural o un proceso social) del que contamos con todos los registros de ocurrencia de sus eventos asociados dentro de un tiempo específico o un espacio geográfico determinado.

Frecuencia absoluta

Se define como **frecuencia absoluta** de un evento al número de veces que éste es observado dentro de un periodo específico o en un espacio geográfico determinado.

Como un ejemplo de este concepto podríamos considerar que al observar la precipitación pluvial durante 30 días en cierta región observamos que sólo llovió 10 días. Por lo tanto la frecuencia absoluta de este fenómeno queda determinada por el número de días en los que se observó la precipitación pluvial, que para este ejemplo es de 10 días.

Frecuencia relativa

Se define como **frecuencia relativa** al cociente entre la frecuencia absoluta del evento y el número total de eventos observados dentro de un periodo específico o un espacio geográfico determinado. Así en términos simbólicos la frecuencia relativa de un evento se define como **n/N** (expresión que se lee como el número de casos observados entre el número total de casos), donde **n** y **N** representan respectivamente el número de veces que el evento fue observado y el número total de eventos observados en un periodo específico de tiempo o en un espacio geográfico determinado.

A partir del ejemplo anterior la frecuencia relativa para este evento la obtendríamos como resultado de dividir la frecuencia absoluta (que para este ejemplo es 10) entre el número de días que el fenómeno fue observado (que para este ejemplo es de 30), obteniéndose con ello una frecuencia relativa de 0.333 o equivalentemente del 33.3%

Para una mejor comprensión de ambos términos veamos el siguiente ejemplo:

En cierta comunidad rural se inicia un programa para la prevención del contagio y transmisión de enfermedades respiratorias entre hombres y mujeres de la tercera edad de 70 o más años durante el periodo invernal. Se pretende informar a dicha población sobre cuáles son los cinco tipos de enfermedades respiratorias (resfriado común, gripe o influenza, neumonía, alveolitus y alveolitos fibrosa) más frecuentes. Para ello se cuenta en el hospital comunitario con un registro del número de individuos en este rango de edades a los que se les diagnosticó alguna de las cinco enfermedades respiratorias consideradas. En la tabla se muestra el registro.

Enfermedad	Número de casos registrados (n)
Resfriado común	36
Gripe o influenza	24
Neumonía	26
Alveolosis	33
Alveolosis fibrosa	31
Número Total de Casos "N"	150

Para medir el impacto de cada una de estas enfermedades se recurre al cálculo de sus respectivas frecuencias relativas, las cuales se muestran en la siguiente tabla junto con sus correspondientes frecuencias absolutas.

Enfermedad	Frecuencia absoluta	Frecuencia relativa
Resfriado común	36	0.24
Gripe o influenza	24	0.16
Neumonía	26	0.17
Alveolosis	33	0.22
Alveolosis fibrosa	31	0.21
Total	150	1

Estos datos se pueden representar en una gráfica de pastel como la siguiente.



glosario
Prevalencia: número total de casos de enfermos para un tipo específico de enfermedad en un momento y lugar particular.

Como ejemplo de cálculo de la frecuencia relativa, se muestra la manera en que se obtuvo el valor correspondiente a la frecuencia relativa asociada con la **prevalencia** del resfriado común. Para ello se consideró a N = 150 y n = 36, de donde se obtiene que la frecuencia relativa correspondiente es de 36/100 o del 24%. Dicha cantidad demuestra que durante un año el resfriado común fue la enfermedad respiratoria que más impacto tuvo entre los adultos mayores de 70 o más años. Tal demostración significa que es alta la probabilidad de que el fenómeno se repita.

El ejemplo anterior deja claro que el hecho de contar con registros de la ocurrencia de los eventos relacionados con un fenómeno natural nos permite poder medir su impacto.

¡Ahora te toca a ti!

En esta sección comenzaste a estudiar los elementos básicos para diferenciar entre un evento aleatorio y un evento determinístico. Asimismo, se presentaron algunas herramientas estadísticas útiles en la cuantificación del grado de impacto que un

fenómeno natural o proceso social de naturaleza aleatoria pude tener en tu contexto. Prosigue tu estudio con la siguiente actividad.



Retoma la lista de los fenómenos naturales y procesos sociales que identificaste en la actividad anterior (1) y clasifícalos en determinísticos y aleatorios, argumentando las razones que te llevaron a realizar esta clasificación. Con base en la problemática planteada al inicio de la unidad o de la actividad alternativa que desarrollaste en la actividad 1, obtén registros de ocurrencia de dos fenómenos y calcula su frecuencia relativa asociada. Investiga en tres fuentes la definición término estadística, así como de los conceptos frecuentista y subjetivo de la probabilidad.

Para el desarrollo de esta actividad es pertinente que cuentes con computadora con procesador de palabras y hoja de cálculo electrónica.

El tiempo estimado para su desarrollo es de tres horas y los pasos para llevarla a cabo son los siguientes:

1. Clasifica los cuatro fenómenos naturales y los cuatro procesos sociales considerados en la actividad 1 en determinísticos y aleatorios.[4] En la siguiente tabla anota tus respuestas.

Fenómeno	Clasificación

4 Si decidiste desarrollar tu propio proyecto entonces de los fenómenos clasificados en el paso 1 deberás seleccionar dos y obtener en cada caso sus registros de ocurrencia con la finalidad de medir su impacto en tu entorno a partir del cálculo de su frecuencia relativa asociada.

2. Supón que para mejorar las condiciones de salud de los habitantes de la comunidad de Santiago, estás interesado en conocer los principales padecimientos de salud pública en la comunidad. Para realizar su medición solicitas al director de la clínica de la localidad los datos correspondientes al tipo y número de casos de los padecimientos más comunes en ella. En la tabla siguiente se muestra el registro de ocurrencia anual para los distintos tipos de padecimientos presentes.

Tipo de enfermedad	Casos registrados (n)
Enfermedades respiratorias	245
Dengue	675
Infecciones intestinales	300
Cáncer	50
Diabetes	270
Número total de casos (N)	1540

3. Como puedes observar en la tabla, el dengue es el padecimiento que más prevalencia tuvo a lo largo de un año. Ahora, para continuar tu estudio estás interesado en medir el impacto de los cuarto diferentes tipos dengue (serotipo 1, 2, 3 y 4). En la tabla siguiente se muestra el registro anual del número de casos.

Tipos de dengue Serotipo	Casos registrados (n)
1	235
2	145
3	200
4	95
Número total de casos (N)	675

Con base en estos registros calcula la frecuencia relativa asociada.

Redacta un informe escrito con las siguientes características, y recuerda guardarlo en el portafolio de evidencias si consideras que demuestra tu aprendizaje.

Gestión del aprendizaje
Es posible realizar gráficas en Excel de manera sencilla. Basta con que sigas los pasos descritos en el Apéndice 2.

Extensión. Dos cuartillas como máximo.

Introducción. En este apartado debes describir brevemente la naturaleza del azar así como la naturaleza de los eventos aleatorios y determinísticos.

Desarrollo. Explica cuál es la diferencia que caracteriza a un fenómeno determinístico de un fenómeno aleatorio, justificando tu respuesta. Con los datos de las tablas correspondientes al tipo de enfermedad y tipo de dengue serotipo, realiza una gráfica de pastel

para ambos casos. Representa la frecuencia relativa de cada uno de sus eventos e identifica el que tiene más impacto.

Conclusión. Describe brevemente por qué consideras importante identificar los fenómenos naturales y procesos sociales en determinísticos y aleatorios. Explica la relevancia que tiene el cálculo de la frecuencia relativa, así como el uso de la probabilidad y estadística para modelar la incertidumbre relacionada con los eventos aleatorios.

En la siguiente tabla marca con una cruz (✗) si consideras que tú desempeño en cada una de las respuestas a evaluar fue muy bueno (MB), regular (R) o malo (M). Asimismo, de ser necesario, indica aquello que debes mejorar en cada uno de los aspectos a evaluar.

Aspecto a evaluar	MB	R	M	¿Qué debo mejorar?
Entendí adecuadamente el concepto del azar y fui capaz de definir adecuadamente qué son los fenómenos aleatorios y los fenómenos determinísticos.				
Clasifiqué adecuadamente los fenómenos en determinísticos y aleatorios y pude argumentar adecuadamente cuál es su diferencia.				
Realicé adecuadamente los cálculos de la frecuencia relativa y elaboración de los diagramas de pastel.				
Entendí los conceptos de probabilidad frecuentista y subjetiva y supe indicar las diferencias entre estos y el enfoque clásico.				
Comprendí la importancia que tiene el cálculo de la frecuencia relativa.				
Comprendí la importancia que tiene la probabilidad y estadística como instrumentos de medición de la incertidumbre asociada con los eventos aleatorios.				

No olvides verificar tus respuestas en el Apéndice 1.

Las variables como instrumento de medición de la incertidumbre y la variabilidad de un fenómeno aleatorio

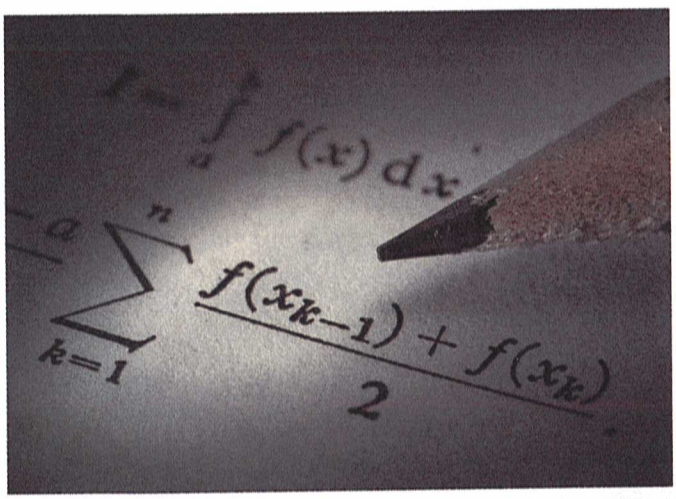


Estás trabajando para determinar y analizar los tipos de variables en la interpretación y explicación de fenómenos naturales y procesos sociales.

Los fenómenos naturales y los procesos sociales poseen características que por su naturaleza pueden ser medidas y estudiadas. Por ejemplo, si consideramos el fenómeno natural de la precipitación pluvial en México, las características (a las cuales también denominaremos **variables**) que podrían ser objeto de medición y estudio podrían ser la precipitación pluvial mensual observada durante un año, o bien, la distribución de la precipitación pluvial anual en México por entidad federativa. Por otra parte, si consideramos un proceso social como el analfabetismo, entonces pudiéramos estar interesados en estudiar y obtener mediciones a partir de las características relacionadas con el número de personas analfabetas por entidad federativa, así como de la distribución del analfabetismo en México entre un grupo de personas de o durante un cierto tiempo.

glosario

Variable: símbolo que representa el valor de una cantidad cuyo posible valor observado no es fijo por que puede variar.



A partir de la naturaleza aleatoria de estos fenómenos es fácil observar que las características consideras están sujetas a variaciones en el tiempo o en el espacio; esto es, si hubiese la posibilidad de tomar distintas mediciones de una misma característica en distintos periodos y espacios, nos percataríamos que debido a la incertidumbre relacionada con la ocurrencia del fenómeno asociado, éstas no resultarían ser idénticas. Por ejemplo, cantidades como los niveles de precipitación pluvial, así como el número de personas analfabetas no son iguales si son medidas en distintos momentos de tiempo y espacio. Una **variable** es entonces cualquier característica cuyo valor puede cambiar al ser medida en distintos periodos del espacio o tiempo.

La incertidumbre derivada de la aleatoriedad de estos fenómenos dificulta la obtención de medidas idénticas para cada una de sus características, para distintos periodos y diferentes espacios. Para estudiar, entonces, esa variabilidad se recurre a los **métodos estadísticos**, herramientas y técnicas que también posibilitan controlar y predecir el comportamiento de tales fenómenos.

glosario

Métodos estadísticos: secuencia de procedimientos útiles para el manejo de datos cualitativos y cuantitativos de la investigación.

El desarrollo de los métodos estadísticos depende en gran medida de las **características métricas** de las variables en consideración, para lo cual es básico distinguir los distintos tipos de variables definidas en función de sus características métricas.

glosario

Características métricas: elementos que forman parte de objetos, seres vivos o fenómenos que pueden ser sujetos de medición.

Variables cuantitativas

Antes de comenzar la exposición de las herramientas que permitirán estudiar la variabilidad de las mediciones es importante considerar la naturaleza de dichas mediciones. Por ejemplo, para el caso de las mediciones de las variables relacionadas con la precipitación pluvial mensual observada durante un año en México y con la distribución de la precipitación pluvial anual por entidad federativa, la escala de medición es en milímetros cúbicos de agua, por lo que esto nos lleva a considerar que las mediciones para la precipitación pluvial no tomarán valores enteros, necesariamente; por ejemplo, podemos contar con registros de que la precipitación media mensual para cierta región alcanzó los 775 milímetros cúbicos de agua mientras que para otra región su medición media mensual alcanzó los 866.7 milímetros cúbicos de agua.

En el estudio de los fenómenos naturales y procesos sociales aleatorios es muy común encontrarse con este tipo de mediciones, definidas como **variables continuas**, y que cumplen la propiedad de que entre dos valores observados hay una infinidad de posibles valores observables. Por ejemplo, si existiese la posibilidad de realizar un número infinito de mediciones de la precipitación media mensual seguramente nos encontraríamos con una infinidad de registros cuyos valores oscilarían entre los 775 y 866.7 milímetros cúbicos de agua. Otra característica de las variables continuas es que, en teoría, se pueden obtener una infinidad de valores distintos asociados con su medición. Siempre es posible identificar el rango de variación de todos sus valores posibles, dicho rango de variación es denominado **intervalo de variación** de todas las posibles mediciones asociadas con una variable continua. De aquí la razón por la cual a las variables continuas también se les defina como **variables intervalares**.

Por ejemplo, para el caso la precipitación pluvial, su intervalo de variación puede comprender cualquier valor que esté en un rango de valores positivos, donde claramente este rango de valores es infinito.

Como ejemplos de variables continuas podemos considerar el peso y la estatura de una persona, el porcentaje de personas desempleadas en cierta comunidad, la cantidad de energía eléctrica consumida por una comunidad en cierto intervalo de tiempo.

Así como existen mediciones de naturaleza continua, existen otro tipo de ellas que no son de tipo continuo y que aparecen con mayor frecuencia en áreas sociales

y humanísticas. Por ejemplo, si pensamos en la variable que mide el número de personas que asisten a un hospital comunitario, es claro que su naturaleza métrica dista de ser continua. Lo anterior es patente al observar la naturaleza de sus mediciones, pues ésta sólo considera valores enteros, lo que imposibilita el poder observar un valor para la variable entre dos valores sucesivos como serían 12 y 13.

glosario

Variable discreta: variable que sólo puede tomar valores dentro de un conjunto numerable, generando con ello una separación entre sus valores observables sucesivos.

A partir de las observaciones antes expuestas, ahora estamos en posibilidad de dar una definición de **variable discreta**. Se denomina como variable discreta a aquella que sólo puede adoptar valores numéricos enteros, de tal modo que entre dos valores consecutivos hay por lo menos un valor no observable, dicho de otro modo, los valores intermedios que pudieran ser considerados entre dos valores enteros consecutivos, carecen de sentido.

Por último es importante mencionar que a diferencia de las variables continuas, las variables discretas pueden tomar sólo un número finito de valores. Por ejemplo, si consideramos a la variable que mide el número de alumnos que cursan la materia de Probabilidad y Estadística cada día de la semana, es claro que este número no puede ser infinito.

Entre algunos ejemplos de variables discretas podríamos citar: el número de personas obsesas en una comunidad, el número de contribuyentes morosos que no han pagado sus impuestos, el número de personas que padecen cierta enfermedad y que están internadas en un hospital comunitario.



Algunas cosas más que debes saber sobre las variables continuas. Un atributo esencial de una variable continua es que, a diferencia de una variable discreta, nunca se le puede medir exactamente. Con una variable continua debe haber inevitablemente un error de medición. Esto como consecuencia de lo inexactos que pueden llegar a ser sus instrumentos de medición como: básculas, balanzas, cronómetros, relojes voltímetros, etcétera.

Variables cualitativas

glosario

Medición cuantitativa: medición de naturaleza numérica obtenida a partir de la observación directa de alguna característica específica de un objeto, ser vivo o fenómeno.

Hasta ahora hemos tratado con ejemplos de variables cuya naturaleza de medición es **cuantitativa**; esto es, las características de medición de los fenómenos naturales y procesos sociales expuestos nos llevaron a obtener mediciones de características numéricas continuas o discretas. Sin embargo, existen otro tipo de mediciones sobre las cuales no se puede asignar de manera natural un valor numérico. Por ejemplo, supongamos que la variable bajo estudio es la que representa la característica asociada al nivel socioeconómico al que pertenece una familia de cierta

comunidad. La naturaleza de esta variable es claramente **cualitativa**, ya que en términos sociales queda especificado que por su nivel de ingresos una familia puede ser clasificada en tres niveles: bajo, medio y alto. Con esto en mente notamos que, si utilizamos esta variable para medir a una familia en particular, el valor que obtendremos no será de naturaleza numérica, pues al tomar la medición obtendremos como resultado una característica y no una medida numérica. A partir de estas consideraciones estamos en condiciones de poder definir el concepto de variable categórica. Se define como **variable categórica** a aquella variable cuyos **niveles de respuesta** son de tipo cualitativo pues están relacionados con características de personas u objetos como: color, raza, credo, filiación política, nivel educativo, entre otros.

glosario

Medición cualitativa: medición de naturaleza no numérica obtenida a partir de la observación directa de alguna característica específica de un objeto, ser vivo o fenómeno y que busca describir alguna o algunas de sus cualidades.

Variable categórica: variable que no se puede cuantificar numéricamente.

Niveles de respuesta: conjunto de posibles valores que puede tomar una variable.

Variables categóricas nominales y ordinales

Dado que la naturaleza de las variables categóricas impide considerar sus mediciones como numéricas, es necesario asignar un valor numérico a cada uno de sus niveles de respuesta. Por ejemplo, lo que procedería en el ejemplo anterior sería asignar los valores de 1, 2 y 3 a cada uno de los tres niveles de respuesta considerados (por supuesto que también se le pudo haber asignado cualesquier otra terna de números; por ejemplo 2, 4 y 8) asignando por ejemplo el valor 1 al nivel socioeconómico bajo y 3 al nivel alto; esto con la finalidad de que estas cantidades rescaten de alguna manera la naturaleza jerárquica de estos niveles socioeconómicos. Es importante destacar que la asignación de valores numéricos a los niveles de respuesta de una variable categórica tomará sentido más allá de considerarlos simples etiquetas siempre y cuando por su naturaleza las categorías de su respuesta guarden de manera natural una jerarquía. Si este es el caso, entonces la **variable categórica** se le denomina **ordinal**.

Sin embargo, hay ocasiones en las que la característica de la respuesta no es de tipo ordinal; por ejemplo, supóngase que una universidad está realizando un estudio de deserción escolar por lo que considera que la variable asociada con la escuela de procedencia de un estudiante pudiera explicar el fracaso escolar de esta universidad. Si pensamos que la escuela de procedencia del estudiante inscrito en la universidad se puede clasificar como bachillerato tecnológico y comercial, bachillerato público y bachillerato privado; entonces esta variable cuenta con tres niveles de respuesta cuya naturaleza no es de tipo jerárquico. Así entonces si pretendiéramos asignar valores numéricos a sus categorías, el ordenamiento en que lo hagamos sería indiferente, pues los números que asignemos sólo representarían

simples etiquetas sin ningún sentido jerárquico. Por tanto, una **variable categórica** se define como **nominal** si sus niveles de respuestas no guardan ninguna relación de jerarquía entre ellos.

Clasificación de las variables por su grado de asociación

La interacción del hombre con su medio ha impactado negativamente el entorno; fenómenos como la industrialización y la explosión demográfica han propiciado cambios sustanciales en el ambiente. Problemáticas actuales como la contaminación del aire y de los mantos acuíferos, el calentamiento global y la desertificación de bosques y selvas, son claros ejemplos de ello.

Del mismo modo en que la interacción del hombre con la naturaleza tiene consecuencias negativas para el medio ambiente, la ocurrencia de cierto tipo de fenómenos naturales —que por su naturaleza van más allá de su control—afectan el contexto social. Fenómenos como la prevalencia de ciertas enfermedades infecciosas, así como la migración y la desnutrición por causa de falta de alimentos, son ocasionados por cambios abruptos en la naturaleza (sequías, terremotos e inundaciones).

glosario
Hipótesis: supuesto al que se le otorga un cierto grado de posibilidad para extraer de ello un efecto o consecuencia.

Estos vínculos entre los fenómenos naturales y los procesos sociales abren paso al planteamiento de **hipótesis** sobre las posibles relaciones de causa y efecto que pudieran existir entre variables asociadas con fenómenos naturales y variables asociadas con procesos sociales, dando lugar a la identificación de variables que se puedan definir como dependientes o independientes.

La dependencia o independencia entre dos variables está basada en la relación de causa efecto entre ambas; esto es, si el valor observado de una de éstas tiene efecto directo sobre el valor observado de la otra o viceversa.

A manera de ejemplo consideremos la hipótesis sobre la relación de causa efecto entre la prevalencia mensual de las enfermedades respiratorias en cierta comunidad con la temperatura mensual registrada en dicho lugar a lo largo de un año. Por las características del fenómeno bajo estudio, es natural pensar que la vulnerabilidad de una persona a padecer una enfermedad de tipo respiratorio se ve acentuada por la presencia de bajas temperaturas. Así, el efecto de la temperatura se puede reflejar como causa de la aparición de una enfermedad relacionada con el aparato respiratorio. De lo anterior podemos concluir que la variable dependiente está representada por la prevalencia mensual de las enfermedades respiratorias en la comunidad y la variable independiente por la temperatura mensual anual en la comunidad.

Siguiendo el método científico sabemos que toda hipótesis planteada debe ser sujeta a comprobación; en este sentido es importante mencionar que dentro de los métodos estadísticos existen herramientas que permiten validar este tipo de hipótesis.

Una **variable independiente** se define como la característica que se supone es la causa del fenómeno estudiado. Mientras que, la **variable dependiente** lo hace como la característica cuyos valores están afectados por los valores que pueda tomar la variable independiente.



La importancia de las variables en el método científico es básica ya que la finalidad del trabajo científico no es otra que descubrir la existencia de las variables y su magnitud, y probar las relaciones que las unen entre si. De las hipótesis se derivan las variables bajo estudio y se deben fundar en ellas la determinación del campo de la investigación, de las informaciones a recoger y de los métodos a emplear.

¡Ahora te toca a ti!

Con base en lo expuesto ya tienes un panorama de las características que definen al tipo de variables utilizadas para el desarrollo de los análisis estadísticos. Cabe destacar que el identificar las variables en un fenómeno natural o proceso social permite obtener mediciones para cuantificar su impacto e interrelación con otras variables de tu contexto.



Con base en los fenómenos naturales y procesos sociales aleatorios identificados en la actividad 2, deberás identificar y clasificar variables asociadas a estos fenómenos en cualitativas y cuantitativas, así como plantear hipótesis sobre las posibles relaciones que pudieran existir entre éstas.

Para el desarrollo de esta actividad debes contar con computadora con procesador de textos y hoja electrónica de cálculo.

El tiempo estimado para el desarrollo es de10 horas y los pasos a seguir son los siguientes:

1. Con base en las características del entorno en el cual se encuentra ubicada la comunidad de Santiago, en el siguiente cuadro se presenta un conjunto de variables relacionadas con la naturaleza del entorno de esta comunidad.

Variables identificadas con eventos aleatorios relacionados con el entorno de la comunidad de Santiago
Sexo de los habitantes de Santiago
Escolaridad de los habitantes mayores de 18 años
Temperatura media anual en la comunidad
Tipo de enfermedades más frecuentes
Porcentaje anual de habitantes que se contagiaron de dengue durante el verano
Porcentaje de habitantes que se contagiaron de alguna enfermedad respiratoria durante el invierno
Número de altas semanales en la clínica de la comunidad
Número diario de camas disponibles en la clínica de la comunidad
Porcentaje de familias cuyo sustento económico es obtenido en actividades derivadas de la comercialización de productos del mar
Número anual de huracanes y tormentas tropicales que han impactado la comunidad
Porcentaje anual de la población que ha migrado hacia la comunidad
Porcentaje anual de la población que ha emigrado fuera de la comunidad
Índice de analfabetismo en la comunidad
Número diario de personas detectadas con diabetes en el hospital de la comunidad.
Presión sanguínea sistólica registrada en las personas que acuden a la clínica de la comunidad.
Tipo de accidente por el que ingresan los pacientes a la sala de emergencia en la clínica de la comunidad

A partir de las variables registradas en el cuadro anterior, clasifícalas en variables cuantitativas (continuas o discretas) y cualitativas (ordinales y nominales). Utiliza para ello la tabla que se muestra a continuación:

Variables cuantitativas		Variables cualitativas	
Continuas	Discretas	Ordinales	Nominales

2. De las variables que clasificaste en la tabla anterior identifica aquellas que puedan tener una relación de causa-efecto. Utiliza la siguiente tabla para registrar cada variable independiente identificada con su correspondiente variable dependiente.[5]

Variables independientes	Variable dependientes

Al terminar tu análisis, escribe un informe con las siguientes características. Al igual que en las dos actividades anteriores no debes olvidar guardar el producto de esta actividad en tu portafolio de evidencias.

Extensión. Una cuartilla como máximo.

Introducción. En esta apartado describe que es una variable y su uso como instrumento de medición.

Desarrollo. Describe cuántos tipos de variables existen, sus características y argumenta las razones de tu clasificación. A partir de las variables identificadas como dependientes e independientes, establece las hipótesis que te llevaron a considerar la existencia de una posible relación de causa-efecto.

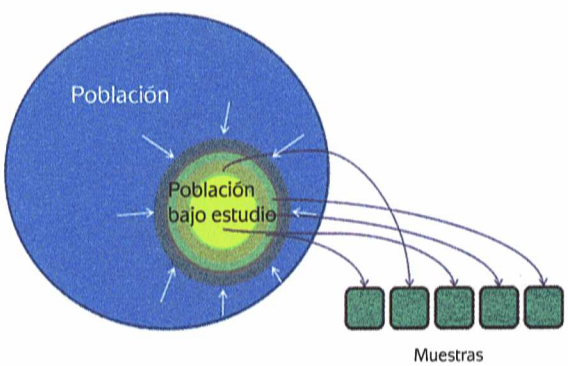
Conclusión. Describe brevemente cuál es la importancia de las variables como instrumentos de medición para el estudio de eventos aleatorios y sus relaciones.

5 Si decidiste seguir la actividad alternativa planteada en la actividad 1, entonces tendrás que seleccionar 16 variables identificadas con eventos aleatorios relacionados con el entorno de tu comunidad. A partir de ellas desarrollarás los puntos 2 y 3 de la presente actividad.

En la siguiente tabla marca con una cruz (✗) si consideras que tu desempeño en cada una de las respuestas a evaluar fue muy bueno (MB), regular (R) o malo (M). Asimismo, de ser necesario, indica aquello que debes mejorar en cada uno de los aspectos a evaluar.

Aspecto a evaluar	MB	R	M	¿Qué debo mejorar?
Di una descripción adecuada del concepto de variable y su uso como instrumento de medición.				
Clasifiqué adecuadamente las variables en cuantitativas y cualitativas.				
Identifiqué adecuadamente las variables como independientes y dependientes.				
Plantee adecuadamente las hipótesis que me llevaron a identificar las posibles relaciones de causa-efecto entre las variables identificadas como dependientes e independientes.				
Expresé con claridad mis ideas con relación a la importancia de las variables como instrumento de medición para el estudio de eventos aleatorios y sus relaciones.				

No olvides verificar tus respuestas en el Apéndice 1.



Muestreo: población, muestra y técnicas

Técnicas de muestreo y su importancia en el análisis de la información

Hasta ahora hemos aprendido la utilidad que tiene la identificación y caracterización de variables como instrumentos de medición para el estudio de fenómenos naturales y procesos



Estás trabajando para Identificar los fenómenos naturales y procesos sociales de su contexto y utiliza técnicas de muestreo para la obtención de datos que permitan interpretar y explicar dichos fenómenos que ocurren en su entorno.

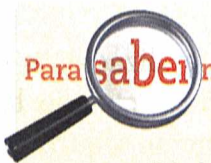
sociales y sus relaciones. Pero un instrumento de medición carece de utilidad si no se tiene posibilidad de utilizarse para obtener información. Por ejemplo, si eres dueño de una báscula y estás interesado en conocer el peso de tus familiares y amigos cercanos, pero no cuentas con la disposición de ellos para pesarlos, tu instrumento de medición entonces no te retribuye ninguna utilidad, pues tu objetivo no fue cumplido ya que sigues sin conocer el peso de tus familiares y amigos cercanos. Algo similar pasa con las variables una vez que han sido definidas si no es viable obtener información de ellas, su utilidad no va más allá del conocimiento del tipo de características que son capaces de medir. Por lo tanto, se requieren las fuentes de información de donde obtener sus valores o mediciones.

Como podemos ver, las **fuentes de información** son el insumo sin el cual el estudio de los fenómenos naturales y sociales no se podría llegar a buen término. Es por ello que una vez definida una variable de interés, es importante cerciorarnos si tenemos la posibilidad de acceder a la información que nos permita obtener las mediciones relacionadas con ella.

Ya que se identificaron las fuentes de información, el paso natural es la **recolección de mediciones o datos** que permita estudiar el comportamiento del fenómeno. En este punto, la calidad de la información recolectada juega un papel muy importante, pues de ésta dependerá la confiabilidad de los resultados que obtengamos del fenómeno a estudiar.

Censos y muestras

Una de las fuentes más confiables para obtener información es el censo, el cual se define como el recuento de una población (entendiendo por población al conjunto de todos los elementos o individuos que son objeto del estudio) con la finalidad de conocer cuestiones inherentes a ella. El **censo** arroja información valiosa para realizar mediciones sobre las variables de interés. Sin embargo, en algunas ocasiones, los recursos económicos y el tiempo con los que cuenta un investigador o estu-



La utilidad de un censo radica en la posibilidad de examinar a cada persona o elemento de la población que deseamos describir. Sin embargo, un censo es demasiado costoso por la inversión en recursos humanos, financieros y materiales. Además de requerir una considerable inversión de tiempo para su realización lo que puede derivar en demora para la obtención de sus resultados.

dioso para realizar un estudio son limitados; por lo que medir una cierta característica a todos los individuos o elementos de una población puede ser inviable, aun más si se piensa que la población de estudio es muy numerosa. Por ejemplo, si un investigador quiere saber la estatura y el peso de la población de todos los mexicanos mayores de 20 años, la cantidad de recursos en dinero y tiempo sería tal que sería virtualmente imposible que lo haga.

glosario
Muestra: colección de mediciones seleccionadas de una población de interés.

Por supuesto que existen alternativas metodológicas que permiten sortear esta dificultad, que son los **métodos de muestreo**. Tales metodologías permiten al investigador obtener información confiable de una variable sin tener que realizar un censo. Estas técnicas se basan en la recolección de información de una fracción de la población de estudio (con la evidente disminución de costos por la recolección de información, así como un ahorro significativo en tiempo y otros recursos).

Antes de dar la definición y descripción de las técnicas de muestreo es importante que partamos de la definición del concepto de **muestra**, mismo que nos será de gran utilidad para entender la naturaleza de los métodos de muestreo expuestos en esta unidad.

Concepto de muestra



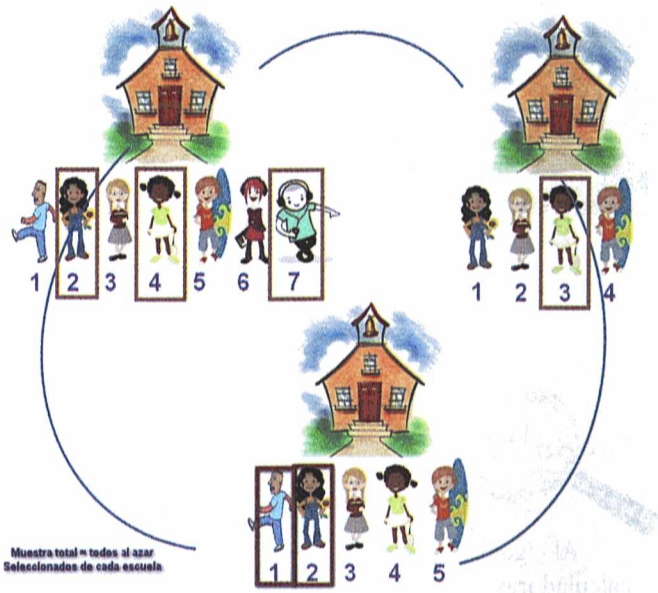
Se define a una muestra como el subconjunto extraído de la población mediante la aplicación de algún método de muestreo, que permite obtener conclusiones validas que se pueden generalizar hacia la población. Retomando nuestro ejemplo de la estatura y el peso de la población de todos los mexicanos mayores de 20 años, si deseamos obtener una muestra, debemos seleccionar sólo un subconjunto o fracción de hombres mexicanos en ese rango de edad a partir de la cual obtendríamos información del peso y la estatura que puede ser generalizable hacia toda la población masculina.

Representatividad de una muestra

glosario
Métodos de muestreo: conjunto de técnicas diseñadas para la obtención de una muestra

La naturaleza de los **métodos de muestreo** permite encontrar generalizaciones a partir de una muestra de la población correspondiente siempre y cuando la muestra sea representativa de la población. Para estar razonablemente seguro de esto debemos considerar cuidadosamente la forma en que se selecciona la muestra. A veces es tentador tomar el camino fácil y recopilar datos de una manera casual, pero si una muestra es elegida sobre una base de conveniencia o comodidad, será imposible interpretar los datos resultantes con confianza. Por

ejemplo, supongamos que un investigador está interesado en medir el ingreso por hogar en cierta ciudad, para ello tendría que censar a todos los hogares de la localidad, lo que implicaría una importante inversión de recursos y tiempo. Dicha situación lo obliga a tomar una muestra para obtener la información requerida. Si por comodidad decidiera seleccionar una muestra con hogares de un solo sector de la ciudad, entonces la información podría carecer de representatividad ya que muchos hogares estarían fuera de la muestra. Una muestra mal tomada puede introducir un **sesgo** en los resultados, el sesgo en una muestra se define como la tendencia a diferir en sus resultados de la correspondiente población.



Métodos de muestreo

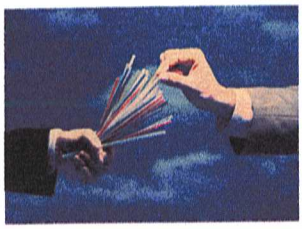
El método más sencillo que estudiaremos se denomina **muestreo aleatorio simple**. Una muestra aleatoria simple es una muestra seleccionada mediante un método que asegura que cada muestra posible del tamaño deseado tiene la misma oportunidad de ser elegida. Por ejemplo, supongamos que el investigador del ejemplo anterior quiere elegir una muestra aleatoria simple de 1000 hogares. Para que esta muestra sea una muestra aleatoria simple cada uno de los muchos posibles subconjuntos diferentes de 1000 hogares deben tener la misma probabilidad de ser seleccionado. Una muestra solamente de hogares que pertenezcan a un sólo sector de la ciudad no sería una muestra aleatoria simple de todos los hogares, ya que en este caso un hogar que pertenece a otro sector de la ciudad no tiene ninguna posibilidad de ser seleccionado.

glosario

Muestreo aleatorio simple: muestreo en el que todas las muestras tienen la misma probabilidad de ser seleccionadas.

Muestra aleatoria simple

Si representamos por "n" al número de objetos o individuos en la muestra, entonces una muestra aleatoria simple de tamaño "n" se define como aquella que se selecciona a partir de un procedimiento que asegura que, todas las muestras posibles diferentes del tamaño deseado tienen la misma probabilidad de ser seleccionadas. La definición de muestra aleatoria simple implica que cada uno de los miembros de la población tiene la misma probabilidad de ser seleccionado en la muestra.



Selección de una muestra aleatoria simple

Se pueden utilizar diferentes métodos para seleccionar una muestra aleatoria simple. Una forma es registrar el nombre o distintivo de cada miembro de la población en pedazos de papel, diferentes e indistinguibles. El proceso consiste en mezclar los pedazos de papel y luego seleccionar "n" de estos, uno por uno. Este método es el más fácil de entender, pero tiene desventajas obvias. La mezcla de los papeles debe ser adecuada, y la generación del número necesario de papeles para realizar la elección puede resultar tediosa, incluso para poblaciones relativamente pequeñas. Un método muy utilizado para la obtención de una muestra aleatoria simple es hacer primero una lista. Cada elemento de ella puede ser identificado con un número, y utilizar una tabla de números aleatorios o un generador de números aleatorios para seleccionar la muestra. Un generador de números aleatorios es un procedimiento que produce una secuencia de números que satisface las propiedades asociadas con la noción de aleatoriedad.

Para ejemplificar la representatividad de una muestra y de un muestreo aleatorio simple, supóngase que un estudiante de medicina de cierta localidad está interesado en aplicar lo estudiado en su última lección de bioestadística. Su profesor le enseñó las características que una muestra aleatoria simple debe tener para ser representativa y algunos mecanismos para su selección. Además, le habló de una técnica exploratoria útil en la búsqueda de posibles relaciones de dependencia entre dos variables.

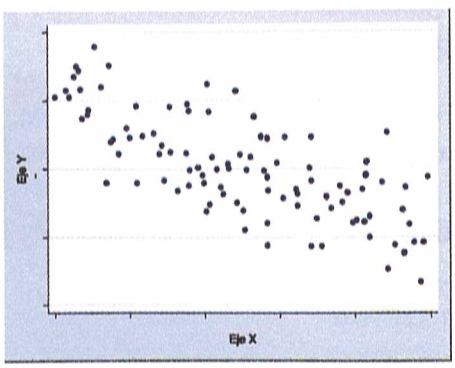
Conocida como **diagrama de dispersión**, esta técnica utiliza un conjunto de registros apareados de la forma (X,Y), con X como variable independiente e Y como variable dependiente, a partir de los que se elabora una gráfica conformada por parejas ordenadas en el plano cartesiano con base en la identificación del valor de la variable X respecto al eje horizontal, mientras que el valor de la variable Y se identifica con el eje vertical. La forma que toma la nube de puntos resultante es la que permite identificar si existe alguna tendencia que indique algún grado de asociación entre las variables. Las siguientes graficas muestran las tres posibilidades de asociación entre variables que pueden ser identificadas a partir de este método.

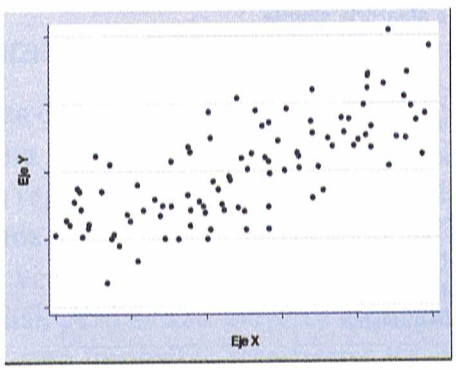
Para saber más

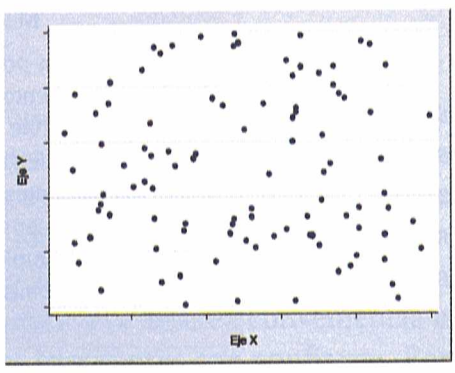
Al igual que las calculadoras científicas, la mayoría de los paquetes de software de estadística incluyen un generador de números aleatorios

glosario

Diagrama de dispersión: representación gráfica de la relación entre dos variables.

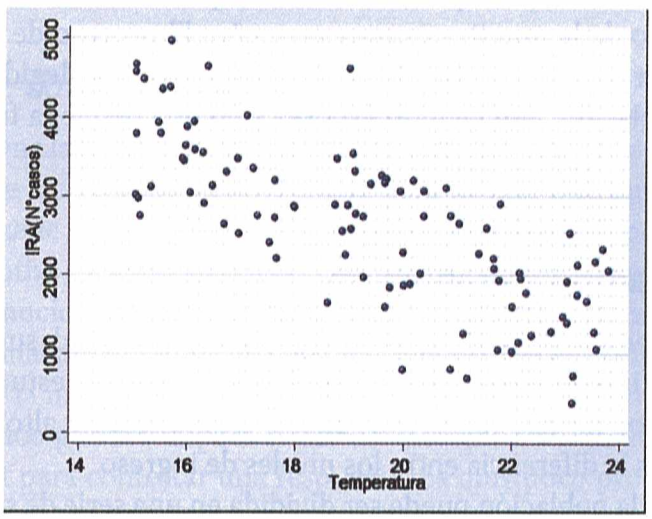






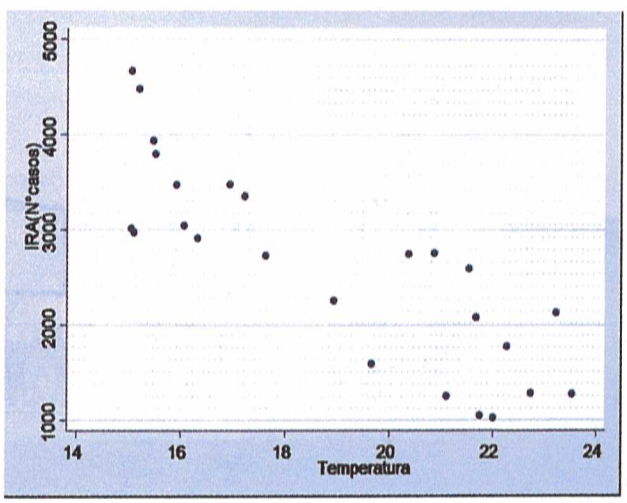
La primera de estas gráficas muestra que no hay relación entre las variables, mientras que la segunda y la tercera muestran la asociación entre ellas, ya que en el primer caso el valor de la variable Y aumenta conforme lo hace el valor de la variable X, mientras que en el segundo éste disminuye.

Ahora supongamos que el estudiante cuenta con 100 registros apareados de temperatura y número asociado de casos de infecciones respiratorias agudas (IRA) recolectados durante 100 semanas consecutivas en el hospital de la localidad. Por su práctica médica y por los reportes de investigación que ha consultado sabe que existe una tendencia que relaciona la temperatura con el aumento o disminución de la presencia de casos de IRA. Con esta información está dispuesto a aplicar los conocimientos de estadística adquiridos en su última lección. Para ello elabora un diagrama de dispersión que considera a la variable X como la temperatura y a la variable Y como el número de casos de IRA registrados y obtiene la siguiente gráfica de dispersión.



Al observar con detenimiento la gráfica anterior identifica la tendencia esperada pues los datos recolectados muestran que conforme mayor es la temperatura tiende a haber una menor presencia de IRA y viceversa.

Por último, aplica un muestreo aleatorio simple, para lo cual decide elegir por una muestra aleatoria simple de tamaño n= 25 a partir de la población de los 100 registros apareados de temperatura y número de casos de infecciones respiratorias agudas (IRA). Con estas observaciones obtiene la siguiente gráfica de dispersión.



Advierte que a pesar de contar con un número menor de observaciones como resultado de la selección de una muestra aleatoria simple, es posible identificar que los datos muestreados representan adecuadamente la tendencia observada por los datos totales. Con ello concluye que es posible generalizar el comportamiento de la población con el simple hecho de haber observado el comportamiento de la muestra.

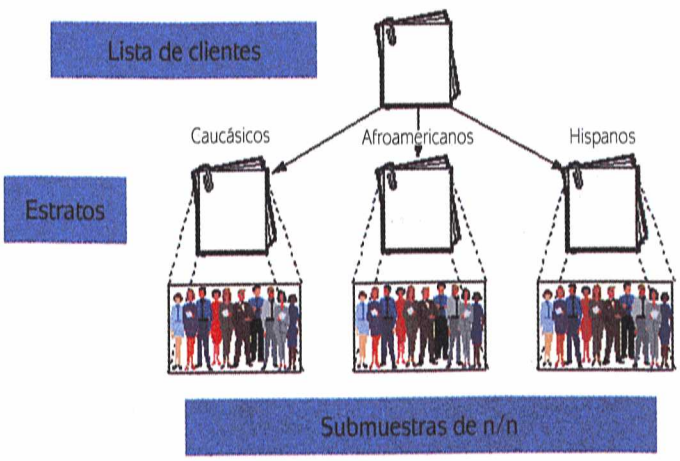
Si bien es cierto que en este ejemplo bastaba con ver el comportamiento de la muestra y de la población, la mayoría de las veces no se cuenta con la información del comportamiento de la población estudiada, por lo que no es posible contrastar los resultados obtenidos de la muestra con el comportamiento de los datos de la población y con ello corroborar su representatividad. Por lo tanto, es el método de muestreo adecuadamente aplicado lo único que nos dará la confianza, más no la certeza, de haber producido una muestra representativa de la población.

A pesar de que es posible obtener una muestra aleatoria que no se desempeñe de manera razonable como representante de la población, esto es más probable cuando el tamaño de la muestra es muy pequeño. El proceso de selección al azar nos permite tener una alta probabilidad de que la muestra elegida refleja adecuadamente a la población, aun cuando ésta sea sólo una pequeña fracción de la población.

Muestreo aleatorio estratificado

Retomemos el ejemplo del ingreso por hogar en cierta ciudad, supongamos que el investigador está interesado en dividir la población objeto de estudio en subgrupos que consideren los niveles socioeconómicos bajo, medio y alto, para identificar qué tan grande es la diferencia entre los niveles de ingreso.

Cuando toda la población puede ser dividida en una serie de subgrupos que no se superponen, el método más adecuado para el análisis es el conocido como **muestreo aleatorio estratificado**. En el **muestreo aleatorio estratificado**, se seleccionan diferentes muestras aleatorias de cada subgrupo, de manera independiente. Para el ejemplo referido, el investigador en lugar de tomar al azar una muestra simple de todos los hogares, tomará tres diferentes muestras aleatorias simples, una de hogares con nivel socioeconómico bajo, una de hogares con nivel socioeconómico medio y otra de hogares con un nivel socioeconómico alto. Estas tres muestras le proporcionarán la información acerca de los tres subgrupos, así como sobre la población total de hogares en la ciudad. Cuando la población se divide de esta



manera, los subgrupos se llaman **estratos** y cada subgrupo individual se llama **estrato** (el singular de los estratos). El muestro estratificado implica seleccionar una muestra aleatoria simple por separado en cada estrato. El muestreo estratificado se puede utilizar en lugar del muestreo aleatorio simple siempre y cuando sea importante obtener información sobre las características de los estratos individuales, así como de toda la población.

La verdadera ventaja de un muestreo estratificado es que a menudo nos permite hacer **inferencias** más concretas de una población que un muestreo aleatorio simple. En general, es mucho más fácil hacer estimaciones más específicas de la población que con un muestreo aleatorio simple, ya que es mucho más fácil hacer estimaciones de las características de un grupo homogéneo que de un grupo heterogéneo. Por ejemplo, es más preciso estimar el impacto económico del trabajo de la mujer en hogares de un mismo estrato social que tratar de medir el impacto desde la población de hogares, en general.

glosario

Inferencia: proceso que puede ligar dos o más proposiciones.

Muestreo de conveniencia

A menudo se tiene la tentación de recurrir al muestreo por "conveniencia"; es decir, con un grupo de fácil acceso o conveniente para formar una muestra. ¡Esta es una receta que te puede llevar al desastre! Los resultados de este tipo de muestras rara vez son informativos y generalizables. Una forma común de muestreo por conveniencia es la conocida **muestra de respuesta voluntaria**. Este tipo de muestras se basa en la información obtenida de personas que se ofrecen como voluntarios para formar parte de la muestras, a menudo respondiendo a un anuncio, llamando a un número telefónico publicado para registrar una opinión o al iniciar sesión en un sitio de Internet para contestar una respuesta. Es muy poco probable que los individuos que participan en tales encuestas voluntarias sean representativos de una población más grande de interés.

¡Ahora te toca a ti!

Recordarás que tu tarea en la sección anterior consistió en identificar variables y plantear hipótesis referentes a posibles relaciones de dependencia entre ellas. Es momento de utilizar hipótesis para seleccionar dos variables y medir la posible relación de causa-efecto entre dos fenómenos.

El objetivo de esta tarea es que comprendas la utilidad que un diagrama de dispersión tiene para



apoyar la verificación de hipótesis planteadas con respecto a la posible relación de causa-efecto entre dos variables. Los elementos para hacerlo te fueron dados durante el desarrollo de la presente sección.



Con base en un conjunto de observaciones pareadas a dos variables que pudieran tener algún nivel de asociación, construye un diagrama de dispersión e identifica si existe evidencia gráfica para concluir que las variables estudiadas están relacionadas.

Realiza un muestreo aleatorio simple para la obtención de dos muestras de las observaciones pareadas. Elabora para ambas un diagrama de dispersión comparando sus respectivas gráficas con la gráfica obtenida para la totalidad de los datos.

Con base en el muestreo concluye sí es posible que las gráficas de dispersión para los datos muestrales exhiban la misma tendencia que la mostrada en la gráfica de los datos completos.

Argumenta la representatividad de ambas muestras con base en este resultado.

Como en el caso de las actividades anteriores sería conveniente que tuvieses acceso a una computadora con procesador de textos y hoja de cálculo electrónica.

El tiempo estimado para desarrollar esta actividad es de cinco horas. Para comenzar a trabajar toma en cuenta la siguiente situación:

Por estar inmersa en una región tropical, en la comunidad de Santiago se presentan temperaturas extremadamente altas durante los meses de verano. Las temperaturas altas en combinación con la precipitación pluvial, derivada de la presencia de huracanes o tormentas tropicales en la región, se convierten en factores ambientales que pueden propiciar el aumento en la incidencia de los casos de dengue durante los meses correspondientes a la temporada veraniega.

Partiendo de esta hipótesis y con la finalidad de validarla, en la comunidad se cuenta con un conjunto de 100 observaciones pareadas sobre la precipitación pluvial media, observada en los meses de mayo, junio, julio, agosto y septiembre, correspondientes a los registros obtenidos en los últimos 20 años (véase Apéndice 3). Acompañan a dichas observaciones, sus respectivas mediciones del porcentaje de casos de dengue registrados en los meses correspondientes.

Con los datos considerados:

1. Aplica alguna de las técnicas de muestreo aleatorio simple expuestas y genera dos muestras aleatorias simples del conjunto total de las observaciones pareadas. Una de ellas de tamaño n=15 y la otra de tamaño n=25.
2. Elabora tres gráficas de dispersión, una correspondiente a la totalidad de las observaciones pareadas y las otras dos a las dos muestras de observaciones pareadas que obtuviste.

3. Mediante el análisis de la gráfica de dispersión de todas las observaciones, concluye si es posible observar alguna tendencia que te pueda llevar a suponer que las variables consideradas tienen algún grado de asociación. De existir ésta, índica en qué sentido se presenta.
4. Compara la gráfica de dispersión de todos los datos con las gráficas de dispersión obtenidas para ambas muestras. Con base en esta comparación, reflexiona si ¿hay razones para pensar que estas muestras son representativas? Para contestar esta pregunta recuerda que tienes que fijarte si la tendencia en las gráficas de dispersión muestral es la misma que la tendencia observada en la gráfica de dispersión que incluye todos los datos[6]
5. Realiza una presentación, en computadora, con las siguientes características:

   Inicio. Diseña una diapositiva para explicar los conceptos de población, censo, muestra, muestreo aleatorio simple, muestreo estratificado, muestreo por conveniencia, muestra representativa, así como gráfica de dispersión.

   Desarrollo. Diseña varias diapositivas más para las técnicas de muestreo, tanto las que fueron expuestas en esta sección como aquellas que hayas podido investigar en otras fuentes como libros impresos y electrónicos. Documenta las tres gráficas de dispersión que realizaste durante el desarrollo de la actividad y acompáñalas con las conclusiones obtenidas a partir de su análisis, respondiendo las siguientes preguntas: ¿Las gráficas de dispersión muestran evidencia observable de que exista algún tipo de relación de causa-efecto entre las variables consideradas?, ¿hubo o no pérdida de información acerca del comportamiento de la población como resultado del muestreo?

   Cierre. Reúne las diapositivas para hacer una presentación en la que expliques la importancia que tienen los métodos de muestreo como herramientas de análisis de datos, cuál es la importancia de obtener una muestra representativa para el análisis de información y las ventajas y desventajas que pudiera tener la aplicación de algún método de muestreo sobre un censo y viceversa.

Gestión del aprendizaje

Es posible realizar gráficas de dispersión en un programa de hoja de cálculo de maneara sencilla. Basta con que sigas los pasos descritos en el Apéndice 2.

Si consideras que tu presentación es una muestra del avance que has tenido en el conocimiento y el manejo de estadística, guárdala en tu portafolio de evidencias.

Para concluir esta actividad, en la siguiente tabla marca con una cruz (✗) si lograste (L) o no lograste (NL) desarrollar cada uno de los aspectos considerados. Asimismo en cada caso indica el nivel desempeño que tuviste en el desarrollo de cada uno de estos aspectos, y de ser necesario índica que debes mejorar.

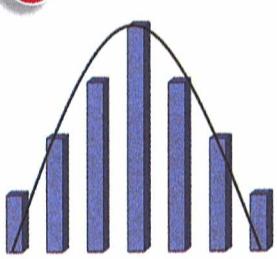
6 Si decidiste seguir la actividad alternativa para desarrollar tu propio proyecto, entonces tú serás el encargado de seleccionar dichas variables, para lo cual es necesario que primero verifiques si existen fuentes de información de las cuales se puedan obtener datos para medir dichas variables y entonces puedas desarrollar los puntos planteados en los incisos de 1 a 4.

Escala de evaluación del nivel de desempeño: 0 Nulo 1 Deficiente 2 Aceptable 3 Satisfactorio				
Aspectos a evaluar	**L**	**NL**	**Escala de evaluación**	**¿Qué debo mejorar?**
Realicé la descripción de los conceptos de población, censo, muestra, muestreo aleatorio simple, muestreo estratificado, muestreo por conveniencia, muestra representativa así como de gráfica de dispersión.				
Documenté los métodos de muestreo que no fueron explicados en la sección.				
Realicé e interpreté las gráficas de dispersión.				
Identifiqué si hubo un cambio en la tendencia de las gráficas de dispersión obtenidas a partir de los datos recabados mediante la aplicación de un muestreo aleatorio simple en comparación con la gráfica desarrollada con todos los datos.				
Expliqué con mis palabras qué es una muestra representativa, así como su importancia en el análisis de información.				
Identifiqué cuál es la importancia de los métodos de muestreo como herramienta para el análisis de información.				
Identifiqué cual es la ventaja de contar con una muestra representativa para el análisis de información.				
Identifiqué las ventajas y desventajas que tiene aplicar algún método de muestreo sobre el uso de un censo para la obtención de información.				

Verifica tus respuestas en el Apéndice 1 para asegurar la correcta resolución de la actividad.

Las distribuciones de probabilidad y su relación con las variables relacionadas con fenómenos naturales y procesos sociales

Estás trabajando para identificar y explicar los diversos tipos de distribución de probabilidad a través de gráficas.



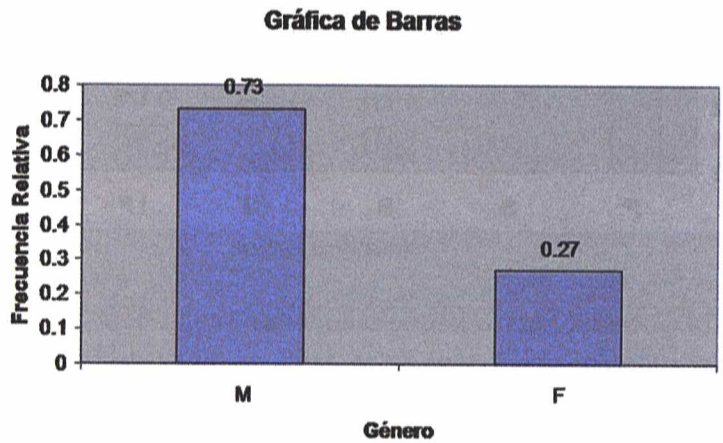
En la sección anterior estudiamos los métodos de muestreo y su utilidad para obtener muestras representativas cuya información permitan generalizar resultados hacia una población objeto de estudio. Conjeturar que se cuenta con la posibilidad de medir cualquier característica de interés a cada uno de los elementos de la población, implica que se pueden obtener mediciones poblacionales de variables numéricas (cuantitativas) o categóricas (cualitativas) y con ello conocer la manera cómo se distribuyen dichas medidas dentro de la población. Veamos por qué.

Distribuciones de frecuencias para variables categóricas

Un investigador realizará un estudio sociodemográfico sobre equidad y género con los datos censales de todos los hogares de una ciudad. Está interesado en medir el impacto del género en la economía familiar a partir de la variable que clasifica el estatus de género de la manutención económica del hogar en masculino y femenino. Lo primero que deberá hacer es recurrir a la descripción de cómo se distribuyen (**distribución de frecuencias**) ambas categorías (masculino y femenino) entre la población de hogares, de tal manera que el impacto sea medido a partir de la frecuencia relativa que ambas respuestas tienen entre el número total de hogares.

glosario
Distribución de frecuencias: instrumento tabular que presenta los valores de los datos y la frecuencia con que se presentan facilitando con ello su comprensión e interpretación.

La **frecuencia relativa**, como recordarás, se calcula dividiendo la frecuencia absoluta entre el número total de observaciones en el conjunto de los datos. Por ejemplo, si el investigador encontró que la frecuencia relativa de esta variable para género masculino y femenino fue de 73% y 27% respectivamente, entonces la distribución de frecuencias puede resumirse en una gráfica de barras cuya frecuencia relativa (proporción) queda representada por la altura de cada rectángulo correspondiente a cada valor de la variable; tal y como lo muestra la siguiente gráfica:



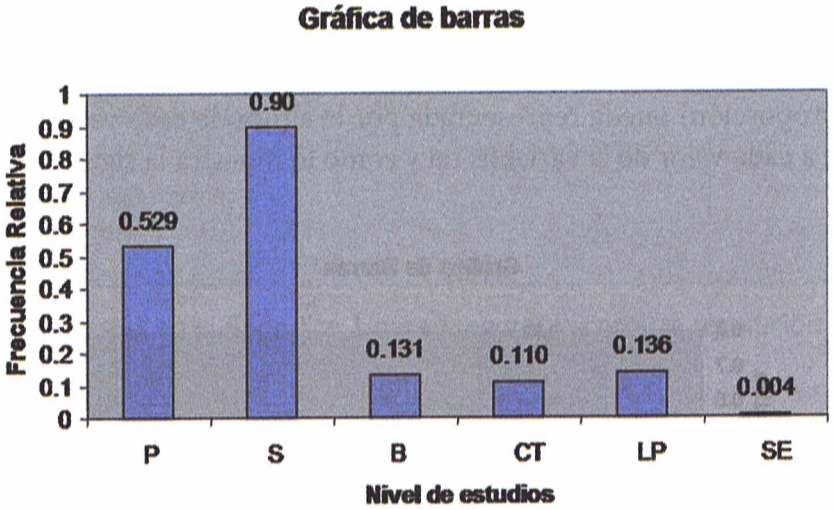
glosario
Frecuencia relativa: cociente entre la frecuencia absoluta y el tamaño de la muestra.

Si un hogar se selecciona aleatoriamente de esta población, la **frecuencia relativa** de ambas categorías puede ser interpretada como la probabilidad de que este hogar pertenezca a alguna de las dos categorías correspondientes al género. Lo que representa que en el largo plazo en un muestreo repetitivo un hogar cuyo sustento económico es aportado por un hombre sería seleccionado alrededor del 73% de las veces, y un hogar cuyo sustento económico es aportado por una mujer sería seleccionado alrededor del 27% de las veces. Si definimos a la variable categórica "x" como x=masculino (M) o femenino (F), en términos de notación probabilística la anterior condición se puede describir como:

$$P(x = \text{masculino}) = .73,$$
$$P(x = \text{femenino}) = .27.$$

Cuya expresiones se leen como: "la probabilidad de que al seleccionar un hogar al azar el valor de la variable "x" sea masculino es 0.73 y la probabilidad de que el valor de la variable "x" sea femenino es 0.27", respectivamente.

Ahora supongamos que el investigador pretende conocer la distribución de frecuencias de la variable "nivel de escolaridad del jefe del hogar", considerando jefe del hogar a aquel individuo que, sin importar el género (masculino o femenino), es el sostén económico. Si definimos a "x" como la variable objeto de estudio, entonces ésta es de tipo categórico con los siguientes niveles de respuesta: primaria (P), secundaria (S), bachillerato (B), carrera técnica (CT), licenciatura y posgrado (LP) y sin estudios (SE). La distribución de frecuencias para esta variable se muestra en la siguiente gráfica:



Con base en esta distribución de frecuencias, si un hogar es seleccionado al azar la probabilidad de que este hogar tenga asociado un nivel de respuesta para la variable "x" de primaria es de 0.529. Simbólicamente las probabilidades asociadas con esta distribución de frecuencias quedan expresadas de la siguiente manera.

$$P(x = \text{bachillerato}) = .110$$
$$P(x = \text{carrera técnica}) = .131$$

Histogramas de frecuencias

Antes de comenzar a exponer las distribuciones de frecuencias para variables numéricas continuas y discretas, haremos un breve paréntesis para describir una

técnica gráfica que resulta de gran utilidad para la representación de la distribución de frecuencias para este tipo de variables. Conocida como **histograma de frecuencias**, esta gráfica adopta ciertas particularidades en su elaboración dependiendo de la característica de la variable bajo estudio, por eso explicaremos de manera independiente los métodos de construcción de dichas gráficas para variables discretas y continuas.

glosario

Histograma: gráfica de barras que muestra el número de datos observados o su frecuencia.

Histograma de frecuencias: gráfica de barras que se utiliza para representar la frecuencia de los datos.

Histogramas para variables numéricas discretas

Es común que en ciertas poblaciones rurales las mujeres tengan una cantidad mayor de hijos que en las poblaciones urbanas. Un investigador social está interesado en estudiar la distribución de frecuencias relacionada con el número de hijos que cada mujer procreó en cierta comunidad rural. Para ello toma una muestra aleatoria de 30 mujeres y registra para cada una el número de hijos vivos que ha procreado. En la siguiente tabla se muestran los registros obtenidos para las 30 mujeres.

Número de hijos															
12	2	4	6	6	7	8	7	8	11	8	3	5	6	7
10	1	9	7	6	9	7	5	4	7	4	6	7	8	10

Como paso previo a la elaboración de un histograma es necesario generar una tabla de distribución de frecuencias. Con este tipo de variables es posible crear distintos tipos de tablas; uno de ellos consiste en listar los valores posibles del 1 al 12 junto con sus correspondientes frecuencias absoluta y relativa, tal y como se muestra en la siguiente tabla.

Distribución de frecuencias para el número de hijos		
Número de hijos	Frecuencia absoluta	Frecuencia relativa
1	1	0.033
2	1	0.033
3	1	0.033
4	3	0.100
5	2	0.067
6	5	0.167
7	7	0.233
8	4	0.133
9	2	0.067
10	2	0.067
11	1	0.033
12	1	0.033

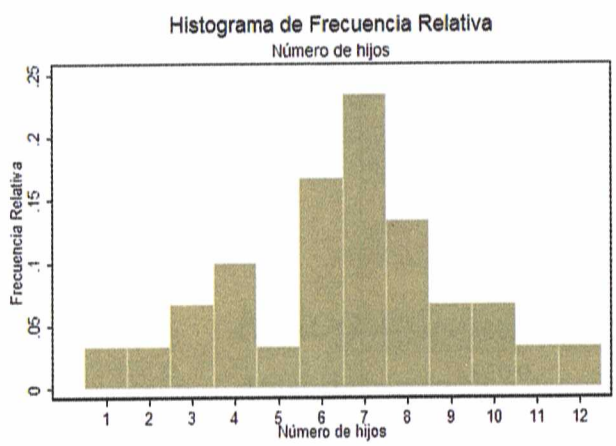
Otra forma de construir una tabla de frecuencias es elaborar una distribución de frecuencias más compacta agrupando los datos y algunos de los posibles valores en intervalos; por ejemplo, se pueden reunir los valores 1, 2 y 3 para formar un intervalo de 1 a 3, junto con su correspondiente frecuencia de 3. Un agrupamiento similar de los otros valores da como resultado la siguiente tabla de distribución de frecuencias relativas.

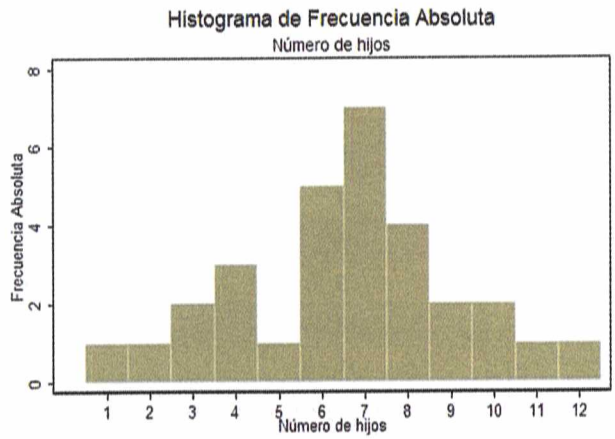
Distribución de frecuencias del número de hijos usando intervalos		
Número de hijos	**Frecuencia absoluta**	**Frecuencia relativa**
1-3	3	0.100
4-6	10	0.333
7-9	13	0.433
10-12	4	0.133

Elaborada la tabla de distribución de frecuencias se desarrollan los siguientes pasos para la elaboración de un histograma:

1. Se dibuja una escala horizontal y se marcan los valores posibles de la variable.
2. Se establece una escala vertical con la marca de la frecuencia o frecuencia relativa.
3. Por encima de cada posible valor se dibuja un rectángulo centrado en el valor mismo, de modo que el rectángulo para 1 este centrado en 1, el del 5 lo esté en 5 y así sucesivamente. La altura de cada rectángulo está determinada por la respectiva frecuencia o frecuencia relativa. A menudo, los posibles valores son números enteros consecutivos, en cuyo caso el ancho de la base de cada rectángulo es 1.

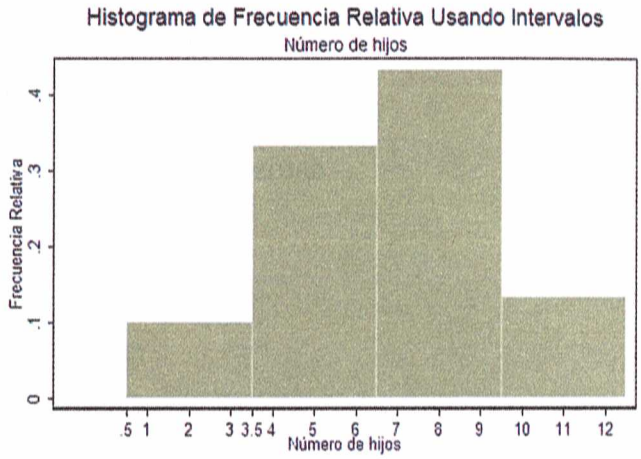
La representación para la primera tabla de frecuencias de los histogramas de frecuencia relativa y de frecuencia absoluta son los siguientes:





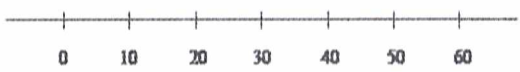
En ambos histogramas, cada rectángulo está centrado sobre su correspondiente valor numérico. Cuando la frecuencia relativa se utiliza en lugar de la frecuencia absoluta para la escala vertical (eje de las "x"), la escala en el eje vertical (eje de las "y") es diferente, pero no se modifican todas las características esenciales de la gráfica (forma localización y extensión).

Se puede construir de manera similar un histograma, con base en la segunda tabla de frecuencias agrupadas. En este caso el rectángulo representa la frecuencia absoluta o frecuencia relativa de cada intervalo. Para el intervalo 1-3 el rectángulo se extiende de 0.5 a 3.5 por lo que no hay diferencias entre los rectángulos del histograma.



Histograma para una variable continua

Cuando la característica a medir es de tipo continua, el proceso de construcción del histograma es más laborioso pues, en la mayoría de los casos, ciertas medidas —como la estatura de un individuo (en metros) o el peso de una persona (en kilogramos)— son tan distintas que no es posible asignar la misma medida a un grupo de individuos de una población. La salida a este problema es definir la tabla de frecuencias con nuestras propias categorías; por ejemplo, para el caso del peso de un individuo, si suponemos que para la población objeto de estudio éste puede variar entre los 30 y 90 kilos se tiene un rango de variación de 60 kilogramos, por lo que una forma de agrupar los pesos podría ser en 6 intervalos de 10 kilogramos de longitud como se muestra en la siguiente figura, en la cual cada observación pude clasificarse en alguno de estos intervalos, denominados, en ocasiones, intervalos de clase.



Pero, ¿dónde podemos colocar observaciones que caen en el límite entre clases? Habría que ubicar dicha observación en el intervalo de clase superior, cuyos valores varían entre 50 y 60, y no en el intervalo inferior de clase cuyos valores varían entre 40 y 50. Así en una distribución de frecuencias una clase puede ser de 40 a <50, donde el símbolo < es sustituto de la frase menor que. Esta clase contendrá todas las observaciones que son mayores que o iguales a 40. La observación 50 podría entonces caer en la clase de 50 a <60, tal y como lo ejemplificamos a continuación.

El gobierno de un estado busca planear la distribución de recursos que en materia de educación se destinarán a cada uno de los 50 municipios que conforman su territorio, en el año venidero. Para hacerlo, el gobierno busca el porcentaje de estudiantes que están inscritos en las escuelas públicas, en cada municipio. Los datos recabados se muestran en la siguiente tabla:

Porcentaje de estudiantes matriculados en instituciones de educación pública por municipio										
96	86	81	84	77	90	73	53	90	93	76
86	78	76	88	86	87	64	89	86	80	66
70	90	89	82	73	72	56	55	75	77	82
83	79	75	43	50	64	80	82	75	96	60
81	59	73	58	73	59					

La observación más pequeña es 43 y las más grande 96. Es razonable comenzar el primer intervalo de clase en 40 y tomar cada intervalo de ancho 10. Esto genera intervalos de clase de 40 a <50, 50 a <60, 60 a<70, 70 a<80, 80 a<90 y 90 a<100.

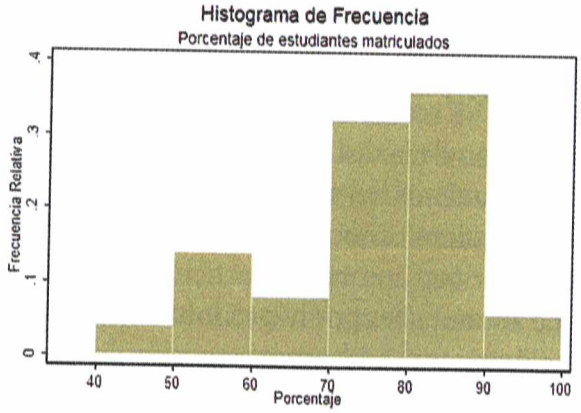
La siguiente tabla despliega una tabla de distribución de frecuencias resultante junto con las frecuencias relativas.

Distribución de frecuencias para el porcentaje de estudiantes matriculados en instituciones de educación pública por municipio		
Intervalos de clase	**Frecuencia absoluta**	**Frecuencia relativa**
40 a <50	1	0.02
50 a <60	7	0.14
60 a <70	4	0.08
70 a <80	15	0.30
80 a <90	17	0.34
90 a <100	6	0.12

Con base en la información de la tabla anterior se construye un histograma de frecuencias desarrollado al seguir los pasos que a continuación se enumeran:

1. Marcar los límites de los intervalos de clase en un eje horizontal.
2. Utilizar la frecuencia o frecuencia relativa como escala del eje vertical.
3. Dibujar un rectángulo para cada clase directamente por encima del intervalo correspondiente (de modo que los bordes se encuentren en los límites de clase del intervalo). La altura de cada rectángulo es la frecuencia o la frecuencia relativa del intervalo de clase correspondiente.

El histograma resultante es el siguiente:



No hay reglas para construir el número de intervalos de clase o la longitud de los intervalos; se pueden utilizar pocos intervalos relativamente amplios para agrupar demasiados datos en una cuantas categorías o un gran número de intervalos relativamente estrechos que pueden diseminar los datos de más, por lo que habría intervalos conteniendo muy pocas observaciones. Ninguna de estas dos estrategias de agrupación dará información de cómo los valores se distribuyen en todo el rango de medición, y ciertas características podrían perderse. En general, con una pequeña cantidad de datos, tal vez entre 5 y 10 intervalos de clase deberían ser utilizados. Cuando existe una gran cantidad de datos, una distribución basada en 15 a 20 (o más) intervalos de clase debería ser usada. La cantidad

$$\sqrt{\text{número de observaciones}},$$

es frecuentemente utilizada como un estimador del número apropiado de intervalos de clase: 5 intervalos para 25 observaciones, 10 intervalos cuando el número de observaciones es 100, y así sucesivamente.

Es claro que el histograma de frecuencias es una herramienta gráfica de gran utilidad porque permite percibir de manera visual, a partir de la frecuencia relativa, el impacto que tiene la presencia de cada variable dentro de una población. Tal percepción es posible a través de la descripción gráfica de la frecuencia relativa asociada con cada uno de los valores de la variable bajo estudio.

Para saber más

Hasta el momento hemos utilizado formas un tanto arbitrarias para dividir los intervalos de clase. Existen otras medidas conocidas como cuantiles las cuales agrupan la distribución de frecuencias en partes iguales; entendidas éstas como intervalos de clase que comprenden la misma proporción de valores. Para una descripción más detallada de la manera como operan estas medidas puedes consultar el Apéndice 2.

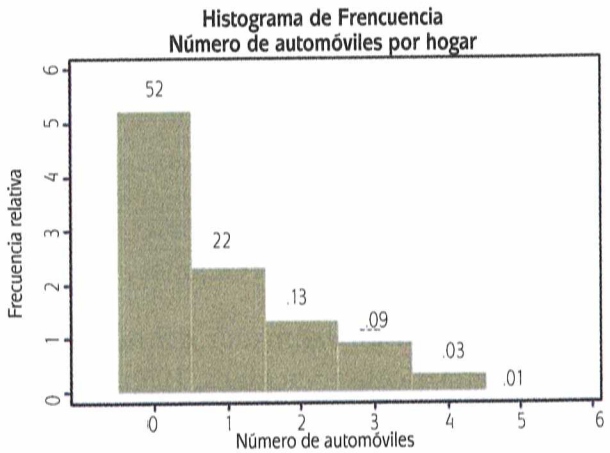
Para saber más

Existen distintos programas de estadística que te pueden ayudar con la labor de diseñar un histograma. Paquetes como el SPSS, Minitab o STATA tienen cargadas funciones que te permiten realizar un histograma de forma rápida y sencilla.

Distribuciones de frecuencias para variables numéricas discretas

Para dar paso a la exposición de la distribución de frecuencias para las variables numéricas de tipo discreto partimos del siguiente ejemplo. Supongamos que en un estudio socioeconómico se quiere conocer la posibilidad que tienen los habitantes de cierta comunidad de contar con automóvil particular. Para tal efecto se considera la variable x = número de automóviles por hogar, los valores posibles para x son 0, 1, 2, 3, 4 y 5 (en esta población no existen hogares con más de cinco vehículos).

Para resumir la distribución de frecuencias para esta variable utilizaremos un **histograma de frecuencias relativas**. Para ello supóngase que se cuenta con registros censales del número total de vehículos para cada hogar de la localidad, por lo que se tiene la posibilidad de calcular la frecuencia relativa asociada con cada una de las respuestas de la variable "x". Esta frecuencia relativa es representada en el siguiente histograma de frecuencias, el cual representa la distribución de frecuencias para la variable x = número de automóviles por hogar.

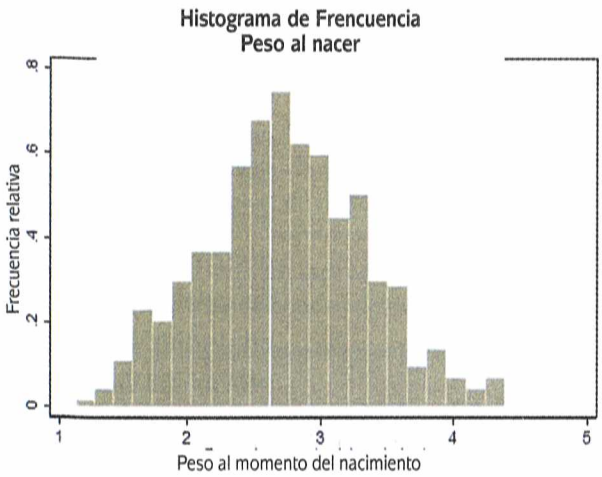


A partir de esta gráfica se puede observar que el valor más común de la población es x = 0. En términos probabilísticos la probabilidad asociada con x = 0 es $P(x=0)=.52$ esto es, el 52% de los hogares no cuentan con automóvil propio. Mientras que a partir de la misma gráfica se puede observar que sólo el 1% de los hogares tiene x = 5 automóviles de su propiedad; es decir, este valor tiene una probabilidad asociada de $P(x=5)=0.01$.

Distribuciones de frecuencias para variables numéricas continuas

Supongamos que se quiere estudiar el peso al nacer (en kilogramos) para todos los bebés nacidos en el año 2010 en cierta localidad. La variable x = peso de recién

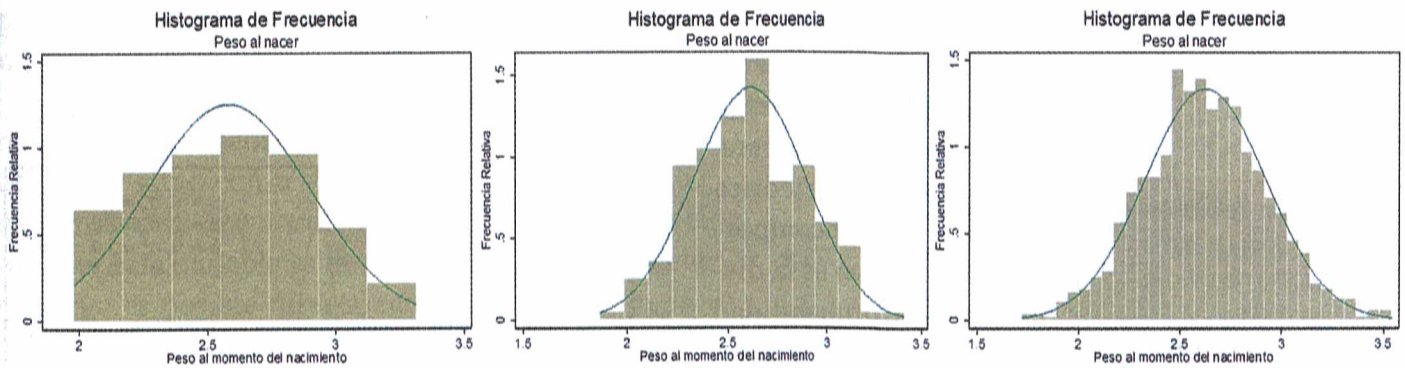
nacido para la población de todos los bebés recién nacidos en esta localidad es un ejemplo de una variable continua. Una forma de describir la distribución de los valores de "x" para esta población es la construcción de un histograma, la escala de medición (en este caso, es el rango que incluye a todos los posibles pesos al momento del nacimiento) se divide en intervalos de clase. Cada valor del conjunto de datos (en este caso, cada peso al nacer) es clasificado en uno de los intervalos. La gráfica correspondiente a este histograma está representada por la siguiente figura.



Este histograma representa la distribución del peso de los recién nacidos en esta localidad, y puede ser visto como la distribución de población de la variable x=peso de un recién nacido. La probabilidad de observar un valor de la variable x está asociada con el área de su rectángulo correspondiente.

Distribuciones de probabilidad continua

Si en el ejemplo anterior tomamos un número mayor de intervalos como consecuencia de un aumento en el número de observaciones de la variable "x", las longitudes de los intervalos de clase se vuelven más estrechas y los histogramas resultantes adquieren una forma más suave en comparación con el primer histograma, tal y como se puede observar en los siguientes histogramas.



A partir del estudio de los histogramas anteriores se puede observar que conforme aumenta el número de intervalos, los histogramas tienden a parecerse más a una curva suave. Si además partimos de la idea de que el área de cualquier rectángulo en un histograma de frecuencias se puede interpretar como la probabilidad de observar un valor de la variable en el intervalo correspondiente cuando un individuo es seleccionado al azar de la población, y si además consideramos que la suma total de las áreas de los rectángulos que conforman el histograma es 1, entonces el área

total bajo esta curva suave también deber ser igual a 1. Si pudiéramos identificar esta curva, podríamos tener un modelo adecuado para representar la distribución de frecuencias de la población para esta variable continua. A esta curva se le conoce con el nombre de **distribución de probabilidad continua**; además, conserva las características importantes de la población (forma general, centro, rango de variación de sus valores, etcétera). Por lo tanto puede servir como modelo para la distribución de frecuencias de los valores de la población y sustituir el cálculo de las probabilidades de eventos, debido a qué el área bajo la curva y por encima de cualquier intervalo particular se puede interpretar como la probabilidad de observar un valor en ese intervalo cuando un individuo es seleccionado al azar de la población.

glosario
Distribución de probabilidad continua: distribución de probabilidad asociada con una variable de tipo continuo.

Entonces concluimos que una **distribución de probabilidad continua** es una curva suave, a la cual llamaremos **curva de densidad**, que sirve como modelo para la distribución de frecuencias de la población para una variable continua y que cuenta con las siguientes propiedades:

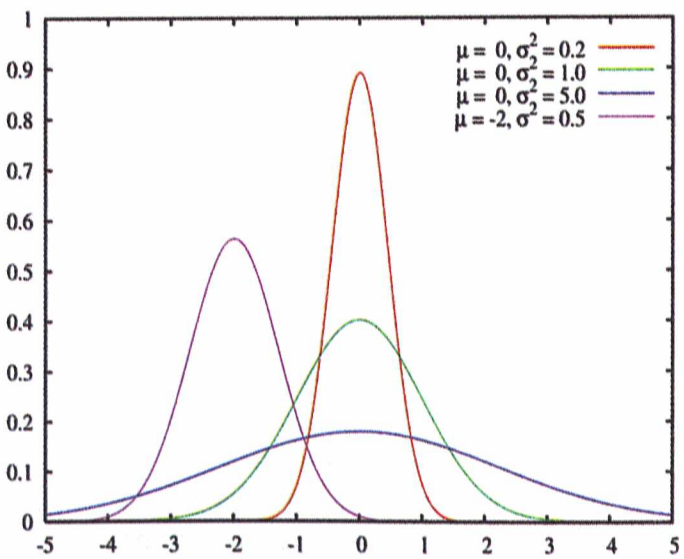
- El área total bajo la curva es igual a 1.
- El área bajo la curva y por encima de cualquier intervalo particular se interpreta como la probabilidad de observar un valor en el intervalo correspondiente, cuando una persona u objeto es seleccionado al azar de la población.

Distribuciones de probabilidades teóricas

La curva suave, como la de los ejemplos anteriores, a la que se aproximaron los histogramas conforme se aumentó el número de observaciones y el número de intervalos de clase, es simétrica. Así como en el caso de los pesos de los recién nacidos, en la naturaleza existen fenómenos cuyas mediciones continuas tienden a comportarse de esta manera.

glosario
Distribución de probabilidad: función que asigna a cada suceso definido sobre una variable la probabilidad de que dicho suceso ocurra.

Una distribución teórica que formaliza este comportamiento continuo es la **distribución de probabilidad** normal ya que en ella la curva suave tiene forma de campana y es simétrica. Las distribuciones normales también son referidas como curvas normales que se caracterizan al representarse en gráficas como las siguientes:



Cuando hay necesidad de medir alguna variable de tipo continuo es recomendable organizar las mediciones en un histograma de frecuencias, ya que éste brinda evidencia visual de un comportamiento de tipo normal. Esto es verificar si el histograma que hemos elaborado muestra tendencia a ser simétrico y tener forma de campana.

El uso de la distribución normal es frecuente para el estudio de fenómenos reales. En la unidad 2 describiremos con más detalle sus características y la utilizaremos como herramienta para el cálculo de probabilidades de eventos naturales y sociales asociados con variables que supondremos siguen una **distribución de probabilidad normal**.

Para saber más

Conocidas con el nombre de pruebas de bondad de ajuste permiten con cierto nivel de certeza darnos información sobre la normalidad en un conjunto de mediciones. Entre algunas de estas pruebas están la prueba de Kolmogórov-Smirnov así como el contrate de bondad de ajuste ji-cuadrado.

glosario

Distribución de probabilidad normal: distribución de probabilidad continua cuya gráfica se asemeja a una campana.

Distribución binomial

Mientras que el modelo teórico de distribución normal sirve para modelar adecuadamente la distribución de frecuencias de cierto tipo de variables continuas pero no para las variables discretas teóricas que por sus características pueden ser utilizadas como modelos para describir adecuadamente la distribución de frecuencias de cierto tipo de variables discretas. Estudiemos un ejemplo.

En cierto hospital comunitario se afirma que la prevalencia de diabetes entre los individuos de una comunidad es de 40%. Para verificar este supuesto un investigador decide realizar el siguiente experimento: durante 15 días consecutivos selecciona una muestra aleatoria de seis personas a las cuales se les haya realizado la prueba de la diabetes. Posteriormente para cada una de las 15 muestras cuenta el número de personas que resultaron con un diagnóstico positivo de la enfermedad.

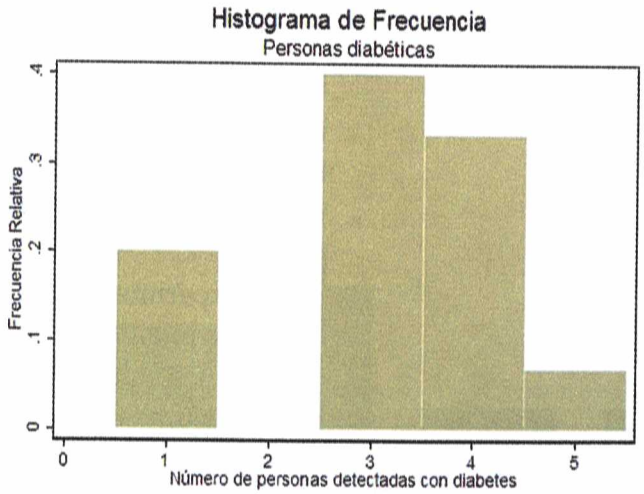
glosario

Discretizar: categorizar una variable cuantitativa o convertirla en cualitativa u ordinal.

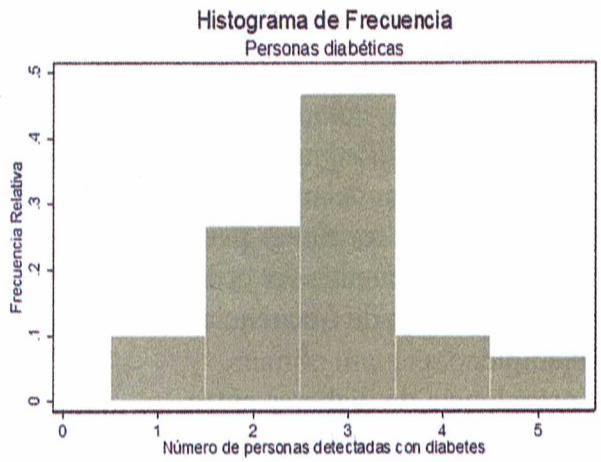
Por la naturaleza del fenómeno a estudiar el investigador considera como medida experimental a la variable que considera el estado de salud con respecto a este padecimiento de cada individuo que acudió a la clínica ha realizarse el estudio. Esta variable es de naturaleza categórica, sin embargo es posible **discretizarla** al definir 0=no enfermo y 1=enfermo, obteniendo con ello la variable discreta numérica x=0.1.

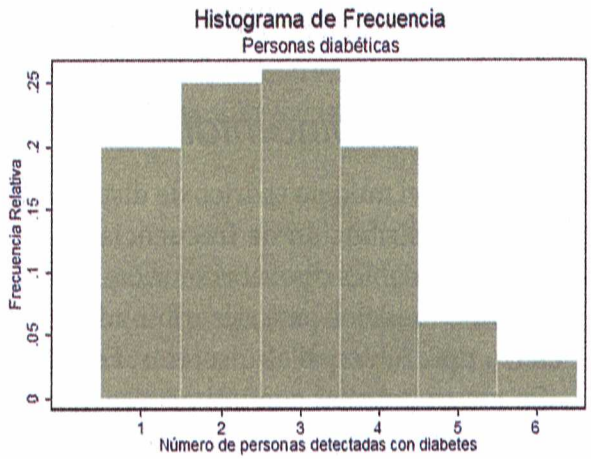
La distribución de frecuencias para los individuos que dieron positivo en la prueba de diabetes se muestra a continuación.

Al observar la distribución de frecuencias representada en el histograma, el investigador concluye que ésta otorga evidencia que refuta el planteamiento, pues a partir de esta gráfica se puede observar que



para la mayoría de las 15 muestras el dato observado más frecuente fue el de tres pacientes detectados con la enfermedad. Por lo que después de lo concluído, pensó que si de cada seis pacientes diagnosticados, en la mayoría de los casos se observaron tres pacientes detectados con el padecimiento (50%), entonces hay suficiente evidencia para afirmar que la prevalencia real del padecimiento entre la población es del 50% y no del 40% como le habían informado. Para confirmar sus sospechas el investigador decidió continuar con su experimento, así que replicó el experimento dos veces más, pero ahora considerando periodos de 30 y 100 días para recolectar la información. Los histogramas de frecuencias para amabas muestras presentaron el siguiente comportamiento.





En ambas gráficas la tendencia se confirmó, pues para ambos casos se puede observar que la mayoría de los casos estuvieron relacionados con la observación de tres pacientes detectados. Situación que lo llevó a concluir que la prevalencia real del padecimiento es de 50% y no del 40% como se había pensado.

En el experimento realizado por el investigador es claro que en la detección de un individuo con diabetes no influyen los resultados obtenidos de los otros individuos analizados, si a este hecho agregamos el supuesto que indica que la probabilidad "p" de detectar un individuo con diabetes permanece constante de un intento a otro, entonces el experimento que acabamos de describir es un ejemplo de una clase llamada **experimentos binomiales**. En general, un experimento binomial es aquel que posee las cuatro propiedades siguientes:

1. El experimento consiste de "n" intentos idénticos.
2. Cada intento da lugar a exactamente a dos resultados, llamados éxito y fracaso (estos resultados pueden ser medidos por la variable x=0 correspondiente al fracaso y x=1 correspondiente al éxito).
3. Los "n" intentos son independientes.
4. La probabilidad "p" de un éxito permanece constante de un intento a otro.

Si partimos del hecho de que el experimento realizado por el investigador es un experimento binomial, entonces los tres histogramas obtenidos representan aproximaciones a la distribución de probabilidad teórica para el número de éxitos de la variable "x". Por lo tanto si "x" es una variable discreta que toma solo los valores 0 y 1 y el número de éxitos (x=1) es obtenido a partir de la realización de un **experimento binomial,** entonces a la distribución de probabilidad para el número de éxitos se le denomina **distribución binomial.**

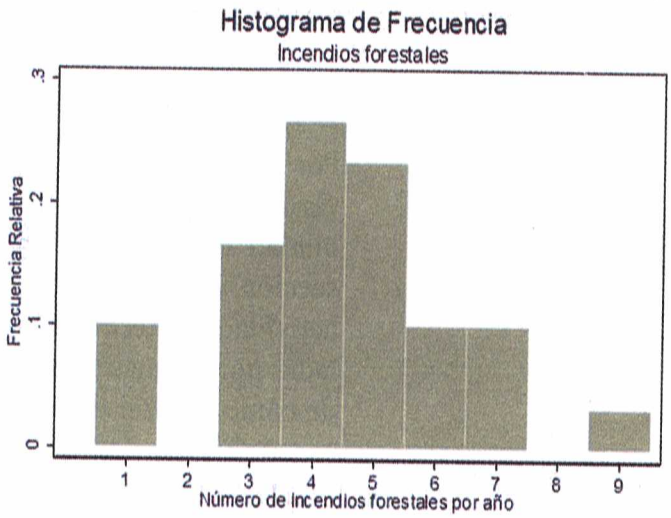
glosario

Experimento binomial: experimento que se caracteriza por la realización de un número finito de ensayos considerados independientes, para cada uno de los cuales sólo se tienen dos posibles resultados.

Distribución binomial: distribución de probabilidad asociada con la realización de un experimento binomial

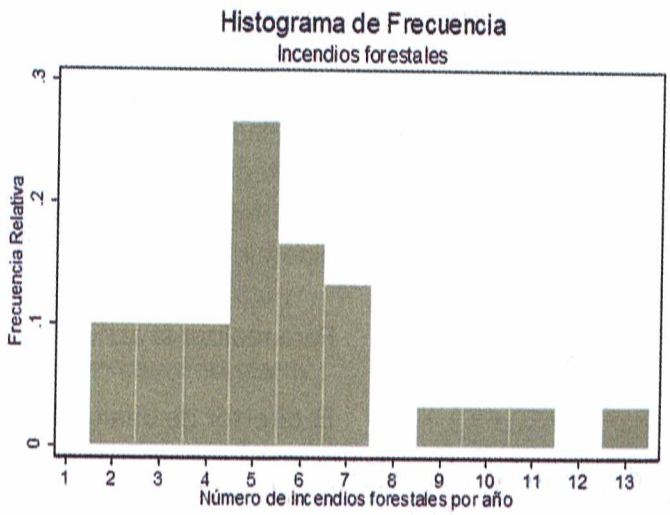
Distribución de Poisson

Un investigador forestal está alarmado porque debido al calentamiento global y la deforestación la tasa media anual de incendios forestales se ha incrementado en los últimos 30 años. Para saber si su consternación tiene razón de ser decide comparar los registros anuales del número de incendios ocurridos en los 30 años previos al año en que sospecha comenzó el incremento en el número de incendios. Revisa los registros correspondientes al número de incendios anuales ocurridos durante los 30 años contados a partir del año en que el investigador sospecha empezó el incremento. Define a la variable "x" como el número total de incendios observados en la región durante un periodo de un año y elabora un histograma de frecuencias de esta variable para los dos periodos considerados.



El investigador observa en ambos histogramas que hubo un incremento en la aparición de incendios y que el número de estos alcanzó la cifra record de 10 o más por año. Después de su análisis concluye que tanto el calentamiento global como la deforestación influyeron para el aumento de los incendios, en la región, en los últimos 30 años.

Si en el ejemplo anterior suponemos que, para ambos periodos observados, el número promedio de veces que ocurrió un incendio durante un año permaneció constante, la probabilidad de haber observado uno en un periodo reducido es muy pequeña y el número de incendios en intervalos ajenos de tiempo



(intervalos de tiempo que no se traslapan) son independientes unos de otros. La opción más pertinente para representar el fenómeno estudiado es el **experimento de Poisson**. Éste hace referencia a procesos relacionados con el espacio o el tiempo y satisface las siguientes tres condiciones:

- El número promedio de veces que ocurre un evento por unidad de tiempo o de espacio es constante.
- La probabilidad de más de un suceso en una unidad de tiempo o espacio es muy pequeña.
- El número de acontecimientos en intervalos ajenos de tiempo o de espacio son independientes unos de otros.

Si la variable "x" representa el número de sucesos resultantes de un experimento de Poisson se dice que esta variable sigue una distribución de Poisson.

¡Ahora te toca a ti!

Por lo que estudiaste en esta sección conoces las características que cierto tipo de variables deben cumplir para poder ser modeladas por las distribuciones de probabilidad teóricas: normal, binomial y de Poisson, por lo que para finalizar esta sección y la unidad, tu tarea será utilizar las variables identificadas en la tercera actividad y determinar si por sus características algunas de éstas pueden ser modeladas por algunas de las distribuciones teóricas descritas. Para tener más elementos de cómo realizar esta tarea desarrolla la siguiente actividad.



Con base en tres de las variables identificadas en la actividad tres, elabora al menos dos hipótesis que permitan determinar a qué tipo de distribución teórica pertenecen: normal, binomial o de Poisson.

Para desarrollar la actividad, como en el caso de las anteriores, es conveniente que tengas acceso a una computadora con procesador de textos y hoja electrónica de cálculo.

Toma en cuenta que el desarrollo de la actividad podría llevarte cinco horas, aproximadamente.

Requerirás elaborar tablas de frecuencias, gráfica de barras para variables cualitativas, histogramas para variables cuantitativas y distribución de probabilidad normal, binominial y de Poisson.

Para realizar la actividad toma en cuenta la siguiente situación:

El director de una clínica comunitaria les pide a algunos estudiantes de preparatoria le auxilien a idear estrategias que permitan ampliar la cobertura de los servicios de salud y, en consecuencia, mejorar las condiciones de salud de los habitantes de la comunidad.

Les solicitó centraran el estudio en las tres variables identificadas en la actividad tres. Dichas variables corresponden a la presión sanguínea de los pacientes que acuden a la clínica de la comunidad, el número diario de camas disponibles y el número de altas semanales.

El interés manifestado por el estudio de la variable presión arterial se explica porque en los últimos meses los médicos subalternos han percibido un incremento inusual de pacientes con presión arterial anormal, lo que podría acarrear problema de salud pública.

El director de la clínica también está preocupado por medir la eficiencia de la capacidad instalada del número de camas así como de la eficiencia en la aplicación de los servicios hospitalarios. Para ello pidió el estudio de las variables correspondientes al registro diario del número de camas disponibles y del número de altas médicas semanales.

Para realizar el estudio con base en lo solicitado, los estudiantes piden al director de la clínica, los datos necesarios para llevar a cabo el estudio que sustente las propuestas para la mejora de la salud de los habitantes de la comunidad y de la prestación de los servicios médicos.

Los datos muestrales para cada una de las tres variables se presentan en las siguientes tablas.

Para saber más

La presión sanguina sistólica de una persona es aquella que se alcanza en el sístole o contracción miocárdica durante la cual el corazón expulsa la sangre que hay en su interior.

Presión sanguínea sistólica									
98	141	127	112	107	111	116	104	87	130
131	124	126	129	65	126	139	136	137	102
128	152	115	124	96	114	116	119	108	113
126	116	107	134	82	141	85	126	131	141
128	85	145	150	97	125	109	119	159	114
117	131	122	104	145	119	123	138	109	102
115	109	143	128	136	118	125	124	109	133
131	127	100	110	109	112	125	166	93	92
113	119	116	136	93	131	141	153	85	119
92	112	140	121	108	92	98	159	91	125

Número diario de camas disponibles en la clínica de la comunidad					
4	7	7	3	4	3
7	5	5	3	5	5
3	6	4	3	5	0
3	1	4	5	4	5
6	3	4	3	5	3

Número de altas médicas semanales en la clínica de la comunidad					
10	15	11	9	21	13
17	17	17	15	18	15
18	10	16	13	14	9
15	9	18	20	12	16
16	15	12	14	20	13

Trabaja con la información disponible y siguiendo los pasos que a continuación se enumeran:

1. Para cada una de las variables descritas identifica cada una de ellas como continuas o discretas.

2. Plantea hipótesis sobre la naturaleza de la información que dio origen a cada una de las variables para que puedas determinar qué tipo de distribución de probabilidad (normal, binomial o de Poisson) puede modelar más adecuadamente la distribución de frecuencias en cada una de las tres variables.
3. Elabora una presentación de la información con las siguientes características:
   Inicio. Diseña las diapositivas para los conceptos de histograma, distribución de probabilidad teórica, distribución de probabilidad normal, binomial y de Poisson, Para el caso de las distribuciones discretas, incluye una explicación de cada una de las hipótesis que le dan sustento a su aplicación.
   Desarrollo. Elabora una o dos diapositivas para cada uno de los tipos de modelo distribucional: normal, binomial o de Poisson y que asignaste a cada una de las tres variables consideradas. Argumenta en cada caso que supuestos o hipótesis estableciste sobre los experimentos que dieron origen a tales mediciones y que te permitieron decidir sobre que distribución de probabilidad asignar. Para cada conjunto de datos presenta una tabla de distribución de frecuencias así como la gráfica de su correspondiente histograma.
   Si identificaste alguna de estas variables como continua, con base en su histograma de frecuencias correspondiente, expón los argumentos por los cuales consideras a la distribución de probabilidad normal como el modelo adecuado o inadecuado para representar la distribución de frecuencias así como la gráfica de su correspondiente histograma.
   Cierre. Elabora algunas diapositivas señalando, con base en los tres histogramas de frecuencia, en cada caso el evento que tiene la mayor probabilidad de ocurrir. Una vez identificado lo anterior desarrolla la conclusión sobre los tres planteamientos trabajados en la actividad:
   a) Para la presión arterial sistólica, y con base en los registros obtenidos, determina si hay suficientes pruebas para alarmarse, tomando en cuenta que la presión arterial sistólica normal de un paciente es de 120.
   b) Con base en los registros obtenidos para el número diario de camas disponibles y el número de altas en la clínica responde qué capacidad instalada de la clínica es necesaria y qué modificaciones habría que hacer para la eficiencia en la atención.[7]

Para finalizar la actividad, evalúa tu desempeño. En la siguiente tabla marca con una cruz (✗) si logré (L) o no logré (NL) desarrollar cada uno de los aspectos considerados. Así mismo en cada caso indica el nivel desempeño que tuviste en el desarrollo de cada uno de estos aspectos. De ser necesario índica que debes mejorar para elevar tu nivel de desempeño de cada aspecto a evaluar.

7 Si vienes desarrollando la actividad alternativa te recomiendo que a partir de las variables que identificaste en la actividad 3 selecciones dos variables de tipo discreto y una de tipo continuo a partir de las cuales puedas obtener información, y con esta información desarrolles todos los pasos de esta actividad.

Escala de evaluación del nivel de desempeño: 0 Nulo 1 Deficiente 2 Aceptable 3 Satisfactorio				
Aspectos a evaluar	L	NL	Escala de evaluación	¿Qué debo mejorar?
Desarrollé los conceptos de histograma, distribución de probabilidad teórica, distribución de probabilidad normal, binomial y de Poisson así como de las hipótesis que dan sustento a su aplicación				
Identifiqué las variables en continuas y discretas				
Argumenté la asignación de la distribución de probabilidad para cada una de las tres variables.				
Para cada variable elaboré una tabla de frecuencias y un histograma de frecuencias.				
Identifiqué alguna variable continua y proporcioné argumentos validos para aceptar o desechar la distribución normal como un modelo adecuado para representar su distribución de frecuencias.				
Calculé probabilidades de eventos asociados con las distribuciones de probabilidad utilizadas.				

Es conveniente que consultes el Apéndice 1 para verificar las respuestas y revisar los procedimientos que llevaste a cabo en su resolución.

CIERRE

Ahora que has concluido el estudio de esta unidad es importante que hagas una pausa para recapitular los conocimientos adquiridos. Partiendo del eje conductor, la comprensión de los fenómenos naturales y los procesos sociales, se expusieron los saberes básicos para que pudieras identificarlos y estudiar su comportamiento. Asimismo, aprendiste a distinguir los elementos para diferenciar este tipo de fenómenos en eventos determinísticos y eventos aleatorios. Utilizaste herramientas y conceptos de tipo probabilísticos y estadístico con el fin de analizar la naturaleza de incertidumbre asociada con los eventos aleatorios. Te acercaste a elementos y conceptos de probabilidad y estadística como cálculo de probabilidades de eventos, frecuencia relativa, definición de variables, variables dependientes e independientes, métodos de muestreo, diagramas de dispersión, histogramas de frecuencias así como distribuciones de probabilidad teórica normal, binomial y de Poisson, que te permitieron estudiar la incertidumbre asociada con la naturaleza de los eventos aleatorios.

Una vez asimilados los saberes es momento para que soluciones el problema planteado al inicio de la unidad.

Si partimos del eje principal del problema, *la búsqueda de medidas que permitan implementar acciones encaminadas a mejorar las condiciones de vida y de salud de*

los pobladores en la comunidad de Santiago, entonces una ruta de acción para la resolución del problema es la selección de los eventos aleatorios relacionados con las enfermedades y la cobertura de servicios hospitalarios en la comunidad.

La utilidad de la selección de estos eventos se basa en la obtención de información relacionada con las condiciones sanitarias y de cobertura de servicios de salud, obteniendo con ello elementos de juicio para identificar las medidas que permitan el mejoramiento de las condiciones de vida y de salud de los pobladores de esta comunidad.

En cuanto a condiciones de salud se refiere, durante el desarrollo de las actividades de la unidad se reconocieron las siguientes:

1. Se identificó al dengue como el mayor problema de salud pública pues fue el padecimiento que mayor frecuencia relativa presentó (véase el Apéndice 3).
2. El análisis de los datos generó información sobre la relación entre el porcentaje de personas enfermas de dengue y los niveles de precipitación pluvial promedio mensual registrados durante los meses más lluviosos del año. A partir de ello se concluyó que a mayor precipitación pluvial promedio mensual aumenta el porcentaje de casos de esta enfermedad entre los habitantes (véase el Apéndice 3).
3. Se determinó que de la muestra aleatoria de pacientes que acudieron a la clínica de la comunidad la mayoría tuvo registros de presión sanguínea sistólica que rondaron el valor normal de 120 (véase el Apéndice 3).

Con referencia a los servicios de salud de la comunidad se identificaron las siguientes condiciones de operación:

1. El número de camas disponibles para la atención de enfermos en el hospital de la comunidad durante el periodo observado, la mayoría de las veces, estuvo entre las tres y las cinco camas disponibles por día y sólo en una ocasión no se tuvo disponibilidad de camas (véase el Apéndice 3).
2. El número más frecuente de altas médicas semanales en el hospital comunitario fue de 15 (véase el Apéndice 3).

A partir de la información recopilada se pueden proponer las siguientes medidas para mejorar las condiciones de vida y salud en la comunidad:

- Durante la época de lluvias habría que tomar medidas preventivas en relación con el dengue, pues es éste el principal problema de salud pública en la comunidad y su prevalencia aumenta conforme lo hace la intensidad en la precipitación pluvial.
  Las medidas están asociadas a la higiene, principalmente con la eliminación de los posibles lugares de depósito de los huevos del mosquito que causa la enfer-

medad tales como la eliminación del agua de lluvia de los recipientes, los huecos de árboles, las rocas, paredes y tapias, las letrinas abandonadas, la depresiones de terreno, los blocks de construcción y los objetos.

- En cuanto a los servicios de salud es de destacar que la eficiencia en el servicio de la clínica de la comunidad es el adecuado.
- La frecuencia en el número de altas médicas semanales fue lo suficientemente alta como para asegurar que siempre hay camas disponibles. Sin embargo, como medida preventiva se sugiere aumentar la capacidad instalada de camas de hospital de 10 a 12. Pues en los registros de camas disponibles por día se encontró que, en dos ocasiones, la capacidad de atención en la clínica se redujo a ninguna o a lo más una cama disponible.

Esta ruta que seguimos para dar solución a la problemática planteada es una de las posibles rutas que pudimos haber abordado ya que los fenómenos naturales y los procesos sociales identificados como influyentes en la comunidad, pueden derivar en problemáticas especificas que una vez identificadas y analizadas conducen al planteamiento de medidas cuya aplicación se traduce en un beneficio para los habitantes de esta comunidad.

Así que, si decidiste realizar la actividad alternativa, planteada desde el principio, te invito a que con base en su desarrollo identifiques y analices problemáticas específicas, y que a partir de esto propongas medidas que den solución o bien mitiguen el impacto de éstas en tu entorno.

Identificación de diferentes escenarios donde aplicar lo aprendido

Si bien es cierto que las características de los fenómenos naturales y de los procesos sociales nos sirvieron de eje conductor para el desarrollo de esta unidad es importante aclarar que la aplicación de las metodologías descritas es mucho más amplia.

La naturaleza incierta en su ocurrencia caracteriza a la mayoría de los fenómenos naturales y procesos sociales pero no es privativa de este tipo de eventos, ya que la incertidumbre está en una gran cantidad de actividades que desarrolla el ser humano. Por ejemplo, existe incertidumbre sobre los resultados en los juegos de azar como la lotería y los pronósticos deportivos; en el comercio para saber cuál será el monto de las ventas para mañana o el próximo mes; en el cálculo del tiempo que tardaremos en trasladarnos a nuestra escuela o trabajo, así como en otras tantas actividades que realizamos. Luego entonces la incertidumbre en la ocurrencia de eventos es un suceso cotidiano y es por ello que los conceptos y métodos estudiados son aplicables a escenarios que van más allá de los eventos aleatorios relacionados con los fenómenos naturales y los procesos sociales.

Evalúa tu aprendizaje

Terminaste de estudiar la primera unidad del módulo. Es importante que hagas un recuento de lo que has logrado en tu proceso de aprendizaje. Por eso propongo la siguiente evaluación.

I. Relaciona las dos columnas de tal forma que, delante de cada enunciado de la columna A escribas el número de la columna B que complete correctamente el enunciado.

Columna A
Son ejemplos de fenómenos naturales ( )
Son ejemplos de procesos sociales ( )
Son ejemplos de eventos determinísticos ( )
Son ejemplos de eventos aleatorios ( )
Definición de fenómeno natural ( )
Definición de proceso social ( )
Definición de censo ( )
Definición de muestra ( )
Definición de muestra representativa ( )
Definición de evento determinístico ( )
Definición de evento aleatorio ( )
Definición de Estadística ( )
Es una característica de los eventos aleatorios ( )
Es una característica de los eventos determinísticos ( )
Ventajas de trabajar con una muestra ( )
Ventaja de trabajar con un censo ( )
Son ejemplos de variables cualitativas ( )
Son ejemplos de variables cuantitativas ( )
Definición de histograma ( )
Definición de frecuencia relativa ( )
Utilidad de un diagrama de dispersión ( )
Es una característica de las variables dependientes ( )
Son tipos de variables cualitativas ( )
Son tipos de variables cuantitativas ( )
Característica de un muestreo aleatorio simple ( )
Característica de un muestreo aleatorio estratificado ( )
Característica de un muestreo por conveniencia ( )
Definición de población ( )
Son característica de una curva de densidad ( )
Son las características de un experimento binomial ( )
Son las características de un experimento Poisson ( )
Característica de una curva que sigue una distribución teórica normal ( )
Son algunos tipos de distribuciones de probabilidad teóricas ( )
Es un ejemplo de una distribución de probabilidad continua ( )
Son ejemplos de distribuciones de probabilidad discretas ( )
Es un elemento que sirve para elaborar un histograma ( )

Columna B
1. El experimento consiste de n intentos idénticos. • Cada intento da lugar a exactamente a dos resultados, llamados éxito y fracaso (estos resultados pueden ser medidos por la variable x=0 correspondiente al fracaso y x=1 correspondiente al éxito). • Los n intentos son independientes. • La probabilidad "p" de un éxito permanece constante de un intento a otro.
2. Recuento de una población con la finalidad de conocer cuestiones inherentes a ella.
3. Se define como el cociente entre la frecuencia absoluta del evento y el número total de eventos observados dentro de un periodo específico de tiempo o un espacio geográfico determinado.
4. Distribuciones de probabilidad normal, binomial y de Possion.
5. Conjunto de todos los elementos o individuos que son objeto del estudio.
6. Incertidumbre.
7. Variables categóricas nominales y ordinales.
8. Intervalos de clase.
9. Todo evento en la naturaleza que sucede por si solo sin ser propiciado por la intervención directa del hombre.
10. Distribución de probabilidad normal.
11. Es una muestra seleccionada mediante un método que asegura que cada muestra posible del tamaño deseado tiene la misma oportunidad de ser elegida.
12. Porcentaje de población desocupada, peso, estatura.
13. Se puede predecir su ocurrencia con exactitud.
14. Terremotos, erupciones volcánicas, sequías.
15. Tiene forma de campana y es simétrica.
16. El número promedio de veces que ocurre un evento por unidad de tiempo o de espacio es constante. • La probabilidad de más de un suceso en una unidad de tiempo o espacio es muy pequeña. • El número de acontecimientos en intervalos ajenos de tiempo o de espacio son independientes unos de otros.
17. Analfabetismo, migración, marginación.
18. Tiene la ventaja de requerir para su realización de una menor inversión de recursos humanos, financieros y materiales.
19. Distribución binomial y de Poisson.
20. Relación de causa efecto.
21. Tiene la ventaja de poder medir la característica de interés a cada individuo o elemento de la población objeto de estudio.
22. El área total bajo la curva es igual a 1. • El área bajo la curva y por encima de cualquier intervalo particular se interpreta como la probabilidad de observar un valor en el intervalo correspondiente, cuando una persona u objeto es seleccionado al azar de la población.
23. Variables continuas y variables discretas.
24. Divide la población objeto de estudio en subgrupos, y selecciona muestras aleatorias simples hacia dentro de cada subgrupo.
25. Es un diagrama que sirve para determinar visualmente la posible relación de causa efecto entre dos variables.
26. Terremotos, inundaciones, migración epidemias.
27. Proceso que genera un cambio en cualquier ámbito de la sociedad, se caracteriza por la intervención directa del hombre, pues a diferencia de la naturaleza éste siempre tiene la posibilidad de interferir directamente en su contexto social para modificarlo.
28. Subconjunto extraído de la población mediante la aplicación de algún método de muestreo.
29. Representación gráfica de una distribución de frecuencias.
30. El fenómeno del día y la noche, el salario diario de un trabajador, la duración de una jornada laboral.
31. Ciencia referente a la recolección, análisis e interpretación de datos, que busca explicar condiciones regulares en fenómenos de tipo aleatorio.
32. Nivel Socioeconómico, sexo, religión, filiación política.
33. Técnica de muestreo que recurre a un grupo de fácil acceso o conveniente para formar una muestra.
34. Muestra caracterizada por incluir representantes de toda la población objeto de estudio.
35. Es aquel evento cuya naturaleza de ocurrencia es incierta y no se puede predecir con exactitud.
36. Es aquel evento cuya naturaleza de ocurrencia se puede predecir con exactitud.

II. Lee con atención el siguiente texto:

La sequía, pobreza y destrucción de los recursos naturales

La desertificación es un proceso complejo de los ecosistemas de las tierras secas, se le define como la degradación de la tierra en zonas áridas, semiáridas y subhúmedas secas, resultante de factores diversos como variaciones climáticas y las actividades humanas, está presente en todos los continentes (excepto en la Antártida) y afecta a millones de personas. La desertificación se produce como resultado del desequilibrio, provocado muchas veces por el ser humano, ocurre a largo plazo a causa de la demanda de recursos de los ecosistemas, de los que se obtienen beneficios como: cosechas, forrajes, madera, alimentación, combustible. La combinación de los cambios climáticos y las acciones del hombre (agricultura, ganadería, pesca, silvicultura y crecimiento demográfico) suponen una gran amenaza para las tierras secas y a su vez, afecta a las sociedades humanas y a todos los seres vivos del planeta.

III. En la siguiente tabla se enumeran los fenómenos de los que habla el texto, clasifícalos en fenómenos naturales y procesos sociales.

Fenómenos y procesos
Desertificación / Migración / Sequia/ Pobreza/ Marginación/ Inundaciones/ Deforestación Contaminación ambiental/ Incendios/ Erosión del suelo/ Calentamiento global/ Desigualdad social Precipitación pluvial/ Desnutrición/ Estaciones del año

Fenómenos naturales	Procesos sociales

IV. Ahora clasifica los fenómenos y procesos en eventos aleatorios y determinísticos.

Eventos aleatorios	Eventos determinísticos

V. El siguiente cuadro muestra un conjunto de variables que pueden ser medidas a partir del contexto de sequía, pobreza y destrucción de los recursos naturales. Estas variables han se clasificaron en variables de acidez o sequía, sociales, económicas, demográficas y de presencia del Estado.

Variables de acidez o sequía	Variables sociales	Económicas	Demográficas	Presencia del Estado
• Precipitación • pluvial • Ocurrencia de la precipitación pluvial • Evapotranspiración potencial • Índice de aridez	• Tasa de migración neta • Porcentaje de mujeres jefas del hogar • Densidad demográfica • Tasa media de crecimiento anual de la población. • Estructura de la población por edad y percepción económica • Incidencia de pobreza • Acceso a la educación • Acceso a servicios de salud • Enfermedades de mayor incidencia • Brecha de pobreza • Tasa de migración neta • Intensidad de la pobreza • Índice de desarrollo humano • Mortalidad infantil • Escolaridad • Porcentaje de • personas o familias con energía eléctrica • Porcentaje de personas o familias con agua potable • Porcentaje de personas con drenaje	• Destino de los recursos para la recuperación de tierras • Ingreso agrícola de las familia/Ingreso total por familia • Producción para la subsistencia/Producción total • Expansión de la frontera agropecuaria • Tamaño de las propiedades rurales • Porcentaje de la población que usa leña y carbón como combustible • Tasa de endeudamiento • Extracción de productos forestales • Estructura de la propiedad • Población económicamente activa por género.	• Población • Densidad(por km2) • Tasa de crecimiento de la población • Tasa media anual de migración neta • Población mayor de 75 años	• Número de políticas destinadas al área • Coherencia y articulación en las políticas de estado de aplicación a la zona • Tipo de política destinada • (asistencias, • promocional) • Fase de proyecto en que participa la población(en todo el proceso, en la ejecución, ausencia de participación) • Aportes efectivos de capital • Áreas atendidas sobre áreas no atendidas

En el cuadro siguiente, y con base en sus características, clasifícalas en cuantitativas (continuas y discretas) y cualitativas (ordinales y nominales).

Variables cuantitativas		Variables cualitativa	
Continuas	Discretas	Ordinales	Nominales



UNIDAD 2

Tratamiento estadístico de la información de fenómenos naturales y procesos sociales

¿Qué voy a aprender y cómo?

¿Sabías que la estadística se puede aplicar a la literatura y a la música?, ¿alguna vez habías oído que también puede ser útil en la pintura? Aunque algunas personas lo consideren poco creíble, está comprobado que la estadística tiene un papel fundamental en el avance del conocimiento en muchos ámbitos, desde la investigación sobre medicamentos, salud pública mercados, control de calidad y hasta de las artes.

La aplicación de la estadística, como verás en esta unidad, en campos diversos de la ciencia y la vida posibilita tomar decisiones de manera individual y colectiva.

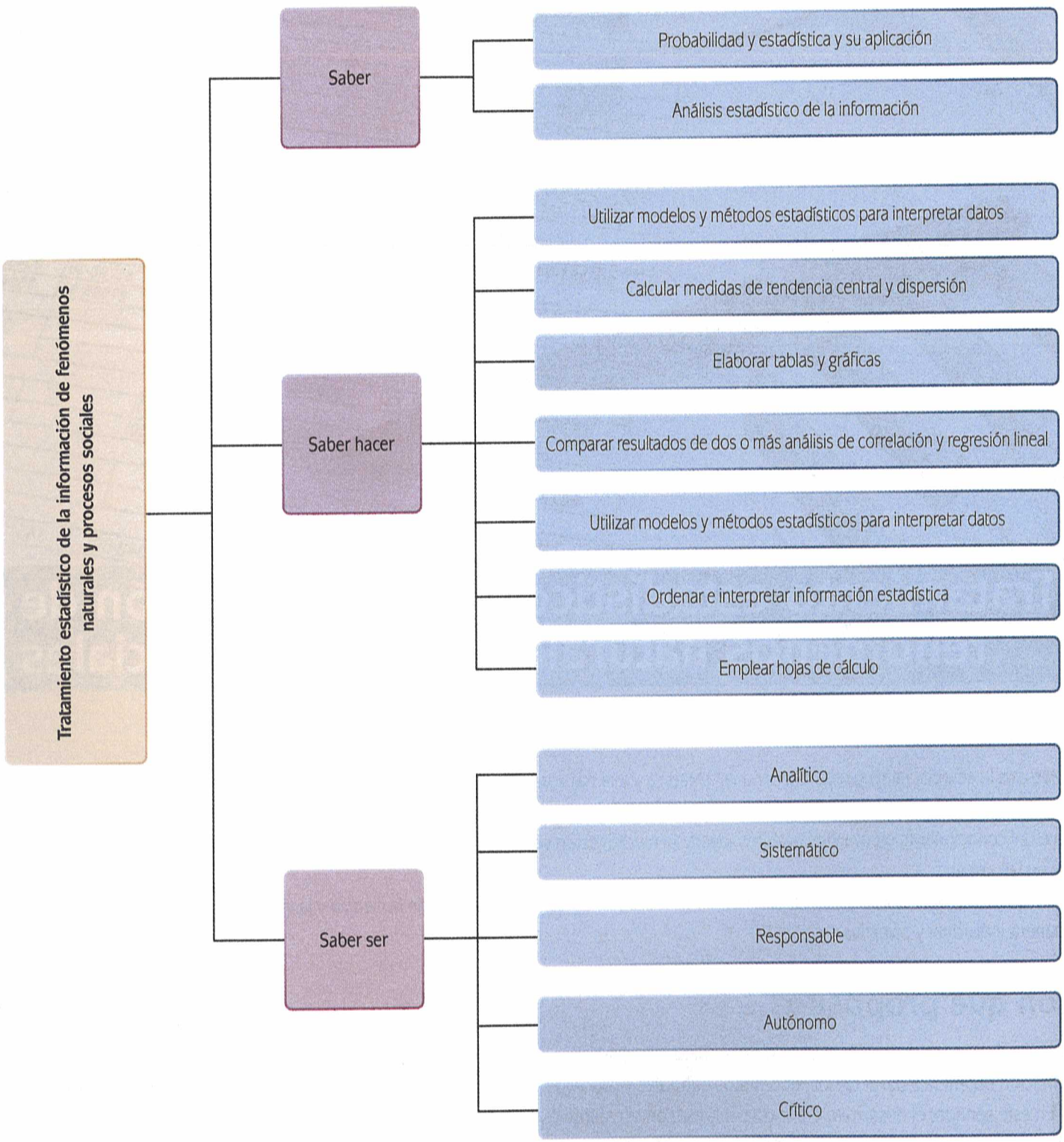
¿Con qué propósito?

El objetivo de esta unidad es el desarrollo de modelos de estadística descriptiva que permitan analizar, ordenar, interpretar y explicar la información obtenida a partir de los fenómenos naturales y procesos sociales para predecir y entender estos hechos en su contexto.

En este sentido, el desarrollo del texto continua con el espíritu establecido desde la primera unidad de aplicar tus aprendizajes en cualquier entorno específico, permitiendo así la contextualización dentro de todo ámbito, ofreciendo de esta manera un criterio para la toma de decisiones individuales, de modo que puedas incidir en tu comunidad, región y país.

¿Qué saberes trabajaré?

Trabajarás los siguientes saberes:



Con la finalidad de aplicar los saberes se han elaborado tres actividades presentadas al final de cada sección o tema. Para su realización es aconsejable que tengas acceso a una computadora que cuente con procesador de textos, hoja de cálculo y programa para elaboración de presentaciones.

¿Cómo organizaré mi estudio?

Tres secciones estructuran la unidad. Es recomendable que planees su estudio en 45 horas organizando 22 sesiones de dos horas y una sesión de una hora. De esa manera si decides estudiar diariamente o en periodos semanales de lunes a viernes, podrás concluir esta unidad en cinco semanas, aproximadamente.

Tema	1ª semana	2ª semana	3ª semana	4ª semana	5ª semana
La estadística descriptiva y su aplicación en la explicación de los fenómenos naturales y los procesos sociales					
Las distribuciones de probabilidad como herramientas para estimar las probabilidades de eventos relacionados con fenómenos naturales y procesos sociales					
El modelo de regresión y el de correlación lineal como medidas para describir la asociación entre variables					

¿Cuáles serán los resultados de mi trabajo?

A medida que avances habrás desarrollado más saberes, habilidades y actitudes por lo que serás capaz de:

- Argumentar la posibilidad de ocurrencia de algunos fenómenos naturales y procesos sociales de su entorno mediante el uso de los tipos de distribución de probabilidad.
- Interpretar tablas y gráficas que representen la información de los hechos o fenómenos naturales y procesos sociales vinculados dentro de tu contexto.
- Utilizar las TIC para obtener información y procesarla a través de una hoja de cálculo.
- Calcular las medidas de tendencia central y de dispersión de varios conjuntos de datos de fenómenos naturales y procesos sociales de interés en tu contexto local, regional y nacional.
- Interpretar y explicar el funcionamiento de fenómenos naturales y procesos sociales dentro de su contexto local y regional a través de modelos de estadística descriptiva.
- Correlacionar las variables involucradas entre fenómenos naturales y procesos sociales de tu contexto

- Evaluar y determinar la probabilidad de ocurrencia y el comportamiento de fenómenos naturales y procesos sociales de importancia en tu contexto para orientar sus acciones en el ámbito individual y propiciar estrategias de mejora en tu localidad, región, país, mundo.
- Analizar y sintetizar las observaciones estadísticas para llegar a conclusiones lógicas.

Antes de empezar

Para comenzar de manera exitosa el estudio de un tema nuevo es importante la recapitulación de los conocimientos previos. A continuación se describe el conjunto de saberes que consideramos prerrequisito indispensable para facilitar tu tránsito hacia la construcción de los saberes de la presente unidad.

Marca con una paloma (✓) aquellos que estás seguro(a) de manejar y con una cruz (✘) los que no. En la última columna anota qué fuente consultarías para su estudio, en caso de requerir reforzar su manejo.

	Sí (✓)	No (✘)	Fuente de consulta para su estudio
Sé			
Acerca de mi entorno social y natural: fenómeno natural y proceso social, desigualdad social, distribución de la riqueza, crecimiento poblacional y funciones vitales de los seres vivos.			
Simbología matemática y jerarquía de operaciones, plano cartesiano, ecuación de la recta, desigualdad lineal, variación proporcional, variación inversamente proporcional, variación polinomial.			
Tipos de lenguaje.			
Métodos de investigación científica y análisis de la sociedad.			
Relacionar los tipos de fenómenos naturales y procesos sociales en mi entorno.			
Sé			
Sumar, restar, multiplicar, dividir, jerarquizar operaciones, identificar simbología matemática, calcular el valor de ciertas variables, ubicar puntos en el plano cartesiano, conocer los elementos que permiten elaborar la ecuación de una línea recta en su forma pendiente ordenada al origen, distinguir las funciones exponenciales y logarítmicas.			
Identificar los distintos tipos de lenguaje de cada campo disciplinar.			
Aplicar un método de investigación.			
Aplicar los principios de análisis de las teorías sociales para explicar e interpretar la sociedad en que vivo, las instituciones que la integran y cómo se da la interacción entre ellas.			

	Sí (✓)	No (✗)	Fuente de consulta para su estudio
Soy			
Independiente: al gestionar su aprendizaje.			
Autogestivo: cuando toma decisiones.			
Responsable: al asumir las consecuencias de mis actos.			
Honesto: al actuar correctamente ante cada circunstancia y no alterar los datos.			
Reflexivo: antes diversos hechos de la realidad.			
Crítico: al actuar y reaccionar ante cualquier situación.			
Respetuoso: a la diversidad de creencias, costumbres y formas de expresión.			
Solidario: para el trabajo en equipo, en el trato con mis compañeros.			

¡Consciente de tus fortalezas y debilidades, comenzamos!

INICIO

Antes de dar comienzo con el estudio de la unidad es importante mencionar que hasta ahora has aprendido a utilizar herramientas de probabilidad y estadística cuya aplicación te han ayudado a resolver distintas problemáticas, continua usándolas para el análisis de la siguiente situación:

La comunidad de El Encino es uno de los principales pasos fronterizos entre México y Estados Unidos. Está enclavada en una zona árida con escasa precipitación pluvial, sin embargo su posición geográfica la ha convertido en polo de atracción para la **inmigración**. Personas de distintas partes del país y de los países centroamericanos ven en El Encino la oportunidad de acceder a mejores puestos laborales, mayores ingresos y **movilidad social**.

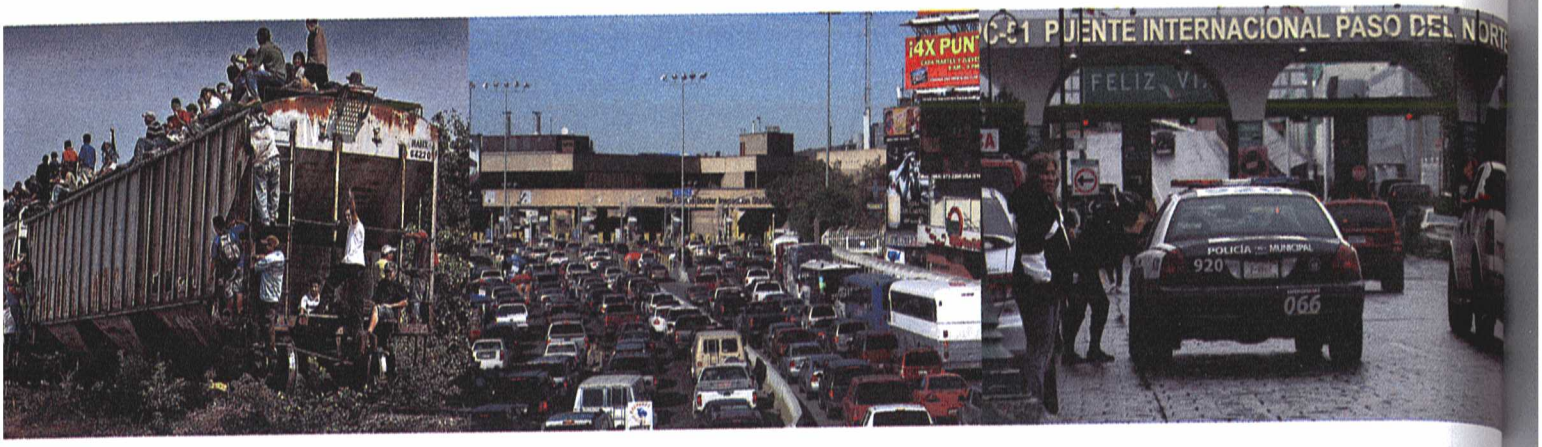
Un investigador de la Universidad de Sonora está interesado en medir los beneficios y perjuicios que el fenómeno **migratorio** ha tenido en El Encino. Enfocará parte de la investigación en la medición del impacto económico del fenómeno migratorio pues algunas características de la población inmigrante –estructura por edad, nivel de especialización para el trabajo y niveles de ingreso– influyen el desarrollo económico. Algunos estudios sobre el tema afirman que una población inmigrante joven, bien preparada y de ingresos con ingresos medios o altos, puede dar mayor dinamismo a las distintas áreas

glosario

Migración: movimiento poblacional que se realiza del campo a la ciudad y viceversa, de una población a otra o de un país a otro, provocado por cataclismos, guerras, violentos cambios sociales o persecuciones, o bien motivado por razones de tipo económico o cualquier otra que induzcan a los habitantes de un país o región a mudar su residencia.

Inmigración: es la llegada de personas a un país o región procedentes de otro país o región para residir en él.

Movilidad social: movimientos o desplazamientos que efectúan los individuos, las familias o los grupos dentro de un determinado sistema socioeconómico.



productivas de la comunidad mientras que una población inmigrante con escasa preparación, bajo poder adquisitivo y demasiado joven origina problemas de pobreza, desocupación y desempleo. Es por ello que aspectos sociales como la pobreza y el índice delictivo también formarán parte de la investigación.

Aunque pudiese resultar un estudio amplio, el investigador pretende relacionar también el entorno natural. Las escasa precipitación pluvial impacta negativamente la distribución del agua a la comunidad, por eso habría que incluir ese factor en la investigación.

¿Trabajas con el investigador?, ¿estudias los efectos que puede tener la migración sobre la comunidad de El Encino? La lectura de los temas y la resolución de las actividades planteadas te ayudará a solucionar el problema ya que cuantificarás el impacto económico de los movimientos migratorios en la comunidad de El Encino. Lo harás a partir de un análisis comparativo entre ciertas características económicas y demográficas de la población oriunda y la foránea. Realizarás cálculos probabilísticos asociados con los siguientes eventos:

- Flujos migratorios futuros hacia y desde la comunidad.
- Número de trabajadores de origen foráneo encontrados en una muestra de cien que se dedican a labores de un alto nivel de especialización.
- Distribución del ingreso de los trabajadores oriundos y de origen foráneo.
- Distribución por grupos de edad de la **población económicamente activa** de trabajadores de origen foráneo y oriundos.
- Cálculo del coeficiente de correlación que permitirá medir el sentido y el grado de asociación entre la variable **tasa de inmigración neta anual** contra los indicadores anuales representados por las variables: **producto interno bruto**, **índice de pobreza humana**, índice de delincuencia y nivel medio anual de almacenamiento de la represa que surte de agua a la comunidad de El Encino.

Quizás por las características del lugar donde resides te podría resultar más interesante desarrollar una actividad alternativa a la planteada pero que considere los

glosario

Población económicamente activa (PEA): conjunto de personas de más de 12 años que desempeñan una ocupación.

Tasa de inmigración neta anual: personas que entran a un país o una región durante un año.

Producto Interno Bruto (PIB): valor de todos los bienes y servicios finales producidos dentro de una nación o región en un año determinado.

Índice de pobreza humana: indicador que mide el nivel de vida en un país o en una región.

elementos aquí descritos. Si decides tomar ejemplos de tu contexto, durante el desarrollo de las tres actividades que conforman la unidad, se te irán otorgando los elementos necesarios que te llevarán a dar soluciones pertinentes a tus planteamientos.

¡Ahora si estamos listos para comenzar!

DESARROLLO

La estadística descriptiva y su aplicación en la explicación de los fenómenos naturales y los procesos sociales



Estás trabajando para interpretar y explicar el funcionamiento de los fenómenos naturales y procesos sociales dentro de tu contexto local y regional a través de los modelos de estadística descriptiva.

Para interpretar tablas y gráficas que representen la información de los hechos o fenómenos naturales y procesos sociales vinculados dentro de tu contexto.

Para calcular las medidas de tendencia central y de dispersión de varios conjuntos de datos de fenómenos naturales y procesos sociales de interés en tu contexto local, regional y nacional.

Para analizar y sintetizar las observaciones estadísticas para llegar a conclusiones lógicas.

Para utilizar las TIC con el fin de obtener información y procesarla a través de una hoja de cálculo.

En cierto año el gobierno de Oaxaca inició un programa de cuidados médicos encaminado a la distribución de medicamentos gratuitos a personas de bajos recursos que no cuentan con algún tipo de seguridad social. Registros de la Secretaria de Salud estatal señalan que sólo 24% los beneficiarios en la capital del Estado se habían apuntado a sólo dos semanas de la fecha límite de inscripción. Ante tal panorama, la Secretaría de Salud decide revisar el funcionamiento del programa en los 49 municipios de la entidad. Los datos sobre el porcentaje de registro por municipio se presentan en la siguiente tabla.

Porcentaje de personas elegibles que se han inscrito al programa de medicamentos gratuitos en cada uno de los municipios del Estado											
24	60	12	38	21	26	23	33	19	19	26	60
16	21	28	20	21	41	22	16	29	26	22	16
48	11	19	13	22	22	30	20	21	34	26	20
25	19	17	21	27	19	27	60	20	52	20	12
14	18										

Al analizar el cuadro, se observa la gran variabilidad en el porcentaje de personas elegibles que se han registrado en el programa en cada uno de los municipios. La observación la lleva a cuestionarse si el 24% es una cifra representativa y por lo tanto concentra la mayoría de los valores de los porcentajes de registro a su alrededor? De no ser así, ¿cuál sería un valor típico para este conjunto de datos?,

glosario

Variabilidad numérica: término utilizado para indicar la heterogeneidad presente en un conjunto de datos numéricos.

Dispersión: grado de distanciamiento de un conjunto de valores respecto a un valor representativo.

¿cómo podrían resumir la **variabilidad numérica**? Los porcentajes de registro varían mucho de municipio a municipio pues van desde un mínimo del 11% hasta un máximo del 60%.

Con el estudio de esta sección se podrá responder a estas interrogantes al plantear medidas numéricas que describan con mayor precisión el valor numérico representativo de la información; asimismo, se podrán generar medidas "resumen" que cuantifiquen el grado de **dispersión** en un conjunto de datos. Se expondrán con detalle las medidas representativas del centro de los datos —la media y la mediana (las dos medidas más utilizadas del centro de distribución)— así como las medidas resumen de la variabilidad de los datos, tales como la varianza y la desviación estándar.

Medidas de tendencia central: media y mediana

Al describir un conjunto de datos numéricos es común que haya un valor representativo de las observaciones. Tal número describe de manera aproximada donde están localizados o centrados los datos a lo largo de una recta numérica. A este valor se le denomina **medida de tendencia central** y las dos medidas de tendencia central más utilizadas son la media y la mediana.

La media

La **media** o **promedio muestral** de un conjunto de datos numéricos es el resultado de la suma de las observaciones dividido entre el número de observaciones mismas. Es útil contar con una notación concisa para la variable sobre la cual fueron hechas las observaciones, para el número de observaciones en el conjunto de datos y para cada una de las observaciones individuales en la forma siguiente:

x = variable a partir de la cual hemos obtenido los datos muestrales.
n = número de observaciones en el conjunto de datos (el tamaño de la muestra).
x_1 = primera observación en el conjunto de datos.
x_2 = segunda observación en el conjunto de datos.
xn = n-ésima (última) observación en el conjunto de datos.

Por ejemplo, podemos tener una muestra de tamaño $n = 4$ observaciones de la variable x = peso de un recién nacido (en kilogramos):

$$x_1 = 3.8 \quad x_2 = 2.9 \quad x_3 = 2.6 \quad x_4 = 3.4$$

glosario

Magnitud: propiedad física que puede ser medida; por ejemplo, la temperatura, el peso, etcétera.

Nótese que el subíndice de la x no tiene relación con la **magnitud** de la observación. En este ejemplo, x_1 es tan sólo la primera observación en el conjunto de datos y no, necesariamente, la observación más pequeña, y x_n es la última observación pero no necesariamente la más grande.

La suma de $x_1, x_2, \ldots, x_n$ se puede representar como $x_1 + x_2 + \ldots + x_n$ pero esta notación puede ser engorrosa. La letra griega Σ es la que tradicionalmente se utiliza en matemáticas para representar una suma. En particular, Σx representa la suma de todos los valores de x en los datos bajo consideración.

La media muestral de un conjunto de observaciones $x_1, x_2, \ldots, x_n$ representada por $\bar{x}$ se define como

$$\bar{x} = \frac{\text{suma total de observaciones en la muestra}}{\text{número de observaciones en la muestra}} = \frac{x_1 + x_2 + \cdots, x_n}{n} = \frac{\Sigma x}{n}$$

Por lo tanto, el peso promedio para los cuatro recién nacidos queda determinado como el resultado de la suma de sus correspondientes cuatro valores muestrales $3.8 + 2.9 + 2.6 + 3.4 = 12.7$ dividido entre el tamaño de la muestra $n = 4$; esto es

$$\bar{x} = \frac{\sum x}{n} = \frac{12.7}{4} = 3.18$$

Ejemplo

Los ortopedistas de un hospital comunitario que atienden a personas de la tercera edad con pérdida de movilidad en la rodilla, realizaron un estudio para investigar el efecto que una nueva terapia tiene sobre la movilidad de la rodilla. Para ello, midieron la capacidad de extender la rodilla. Los valores para la extensión de la rodilla (en grados) para cada uno de los 10 participantes en el estudio se dan a continuación:

$$x_1 = 59 \quad x_2 = 46 \quad x_3 = 64 \quad x_4 = 49 \quad x_5 = 56$$
$$x_6 = 70 \quad x_7 = 45 \quad x_8 = 52 \quad x_9 = 63 \quad x_{10} = 52$$

La suma de estos valores muestrales es $59 + 46 + 64 + \ldots + 52 = 556$ y la **media muestral** de la extensión de la rodilla es:

$$\bar{x} = \frac{\sum x}{n} = \frac{556}{10} = 55.6$$

glosario

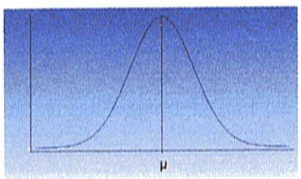
Media muestral: promedio de una variable tomada de una muestra.

Es importante mencionar que aunque los valores de los datos en el ejemplo son todos números enteros, podemos considerar a 55.6 grados como un valor representativo para las medidas de la extensión de la rodilla para esta muestra, ya que en torno a este valor se centran los valores muestrales observados. Así como el de este ejemplo, existen otros conjuntos de mediciones cuyo valor promedio calculado no es parte del conjunto de valores observados, más aún, por la naturaleza de la variable su valor ni siquiera tiene alguna posibilidad de ser observado, sin embargo, se puede considerar representativo por que la mayoría de las cantidades obser-

vadas se encuentran concentradas alrededor de su valor. Por ejemplo, el número medio de hijos por familia puede ser 1.8, mientras que ninguna familia puede tener 1.8 hijos.

La media de la muestra $\bar{x}$ se calcula a partir de observaciones de la muestra, lo que la convierte en una característica de esa muestra. Es costumbre utilizar letras del alfabeto latino para indicar las características de la muestra, como lo hemos hecho con $\bar{x}$. Las características de la población se suelen designar por letras griegas. Una de las más importantes de estas características es la media poblacional.

> La media poblacional, representada por μ (*se lee mu*) es el promedio de todos los valores de x de la población entera.



Para el caso de una distribución de probabilidad normal, como la que se ve en la figura, el valor de la **media** μ es el que divide en dos partes iguales (0.5 del área bajo la curva a cada lado de μ) a esta distribución de probabilidad.

Por ejemplo, el peso promedio al nacer de los 3000 niños nacidos vivos en un hospital durante un año (si consideramos a este número de niños nacidos como la población) podría ser de $\mu = 3.2$ kilogramos. Por otra parte, una muestra de n = 5 niños nacidos en ese año podría haber presentado pesos de 2.8, 3.2, 2.6, 3.7 y 3.4 kg, de la cual obtenemos $\bar{x} = 3.14$ para esta muestra en particular (un valor algo menor que μ). Sin embargo, una segunda muestra puede dar $\bar{x} = 3.52$, la tercera $\bar{x} = 2.9$, y así sucesivamente. El valor de $\bar{x}$ varía de una muestra a otra, mientras que sólo hay un valor de μ.

Una posible desventaja potencial de la media, como medida de tendencia central de un conjunto de datos, es que su valor puede verse afectado por la presencia de valores extremos o atípicos (observaciones inusualmente grandes o pequeñas) en el conjunto de datos. Estudiemos el siguiente ejemplo:

Cuarenta estudiantes de preparatoria abierta se inscribieron en un curso de educación a distancia. El instructor hizo los materiales del curso, las calificaciones y notas de clase para los estudiantes en un sitio web. El software de gestión del curso lleva la cuenta de la frecuencia con que cada estudiante accede a cualquiera de las páginas web en el sitio de la clase. Un mes después de que el curso comenzó, el instructor solicitó un informe que indica las veces que cada alumno ha accedido a una página web en el sitio de la clase. Las 40 observaciones fueron las siguientes:

Número de veces que cada estudiante accede al sitio de la clase											
20	37	4	20	0	84	14	36	5	331	19	0
0	22	3	13	14	36	4	0	18	8	0	26
4	0	5	23	19	7	12	8	13	16	21	7
13	12	8	42								

La media muestral para este conjunto de datos es de $\bar{x} = 23.10$. Muchos argumentarían que 23.10 no es un valor muy representativo de esta muestra, ya que 23.10 es mayor que la mayoría de las observaciones en el conjunto de datos; sólo 7 de 40 observaciones o el 17.5% son mayores que 23.10. Los dos valores extremos de 84 y 331 tienen un impacto sustancial en el valor de $\bar{x}$.

Ahora dirigimos nuestra atención a una medida de tendencia central que no es tan sensible a valores extremos, la mediana.

La mediana

Se define a la mediana como la cantidad que divide al conjunto de datos en dos partes iguales. Una vez que los valores de los datos se han enumerado de menor a mayor, la **mediana** es el valor medio en la lista y es el que divide la lista en dos partes iguales. Dependiendo de si el tamaño de la muestra n es par o impar, el proceso de determinación de la mediana es ligeramente diferente. Cuando n es un número impar, (por ejemplo, 5), la mediana de la muestra es el valor medio único, que en ese caso sería el valor correspondiente al tercer número listado de izquierda a derecha. Sin embargo, cuando n es par (por ejemplo, 6), hay dos valores centrales que dividen al conjunto de datos en dos partes iguales en la lista ordenada, por lo que se necesita promediar estos dos valores centrales para obtener la mediana de la muestra.

La mediana de la muestra se obtiene al ordenar un conjunto de n observaciones tomando como criterio de orden el iniciar con el valor más pequeño y terminando con el más grande (con los valores repetidos incluidos, de tal forma que cada muestra aparece en una lista ordenada). La mediana muestral se define como el valor medio si n es impar y como el promedio de los dos valores medio si n es par.

Estudiemos el siguiente ejemplo:

Ejemplo

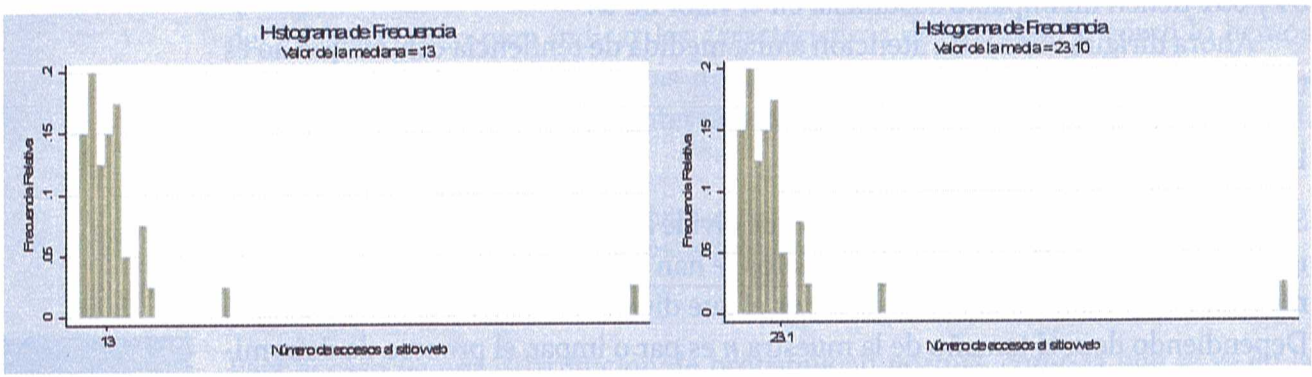
El tamaño de la muestra para los datos del acceso al sitio web del ejemplo anterior, n = 40, un número par. La **mediana** es el promedio de los valores para las observaciones 20 y 21 (el centro de los datos) en la lista ordenada. La organización de menor a mayor genera la siguiente lista (con los dos valores centrales resaltados):

Lista ordenada de los datos											
0	0	0	0	0	0	3	4	4	4	5	5
7	7	8	8	8	12	12	13	13	13	14	14
16	18	19	19	20	20	21	22	23	26	36	36
37	42	84	331								

La mediana se determina como:

$$\text{mediana} = \frac{13+13}{2} = 13$$

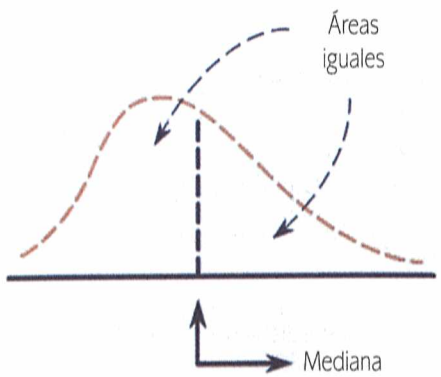
Al mirar los siguientes histogramas para este conjunto de datos —que muestran los valores marcados sobre el eje x para la mediana y la media respectivamente— el valor de la mediana parece ser más típico que el valor de la media $\bar{x}. = 23.10$. Pues se encuentra más hacia el centro del conjunto de los datos.



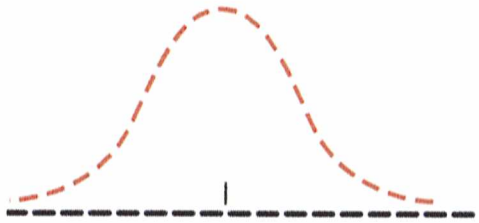
La media de la muestra puede ser sensible incluso a un único valor que se encuentra muy por encima o por debajo del resto de los datos. La mediana, por el contrario, es bastante insensible a los valores atípicos. Por ejemplo, la observación más grande de la muestra (331) en el ejemplo anterior se puede incrementar en cualquier cantidad sin cambiar el valor de la mediana. Asimismo, un aumento en las observaciones de segundo o tercer lugar no afecta a la mediana, ni una reducción en varias de las más pequeños observaciones. Esta estabilidad de la mediana es lo que a veces justifica su uso como una medida de tendencia central.

Comparación de la media y la mediana

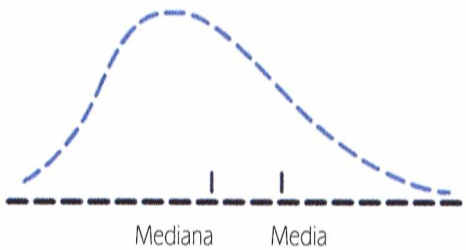
La siguiente figura muestra un histograma en forma de una curva suave que podría representar la distribución de valores de una muestra, o bien la de una población. Gráficamente, la mediana es el valor en el eje de medición que separa a la curva en dos partes, con 0,5 (50%) del área bajo la curva en cada parte.



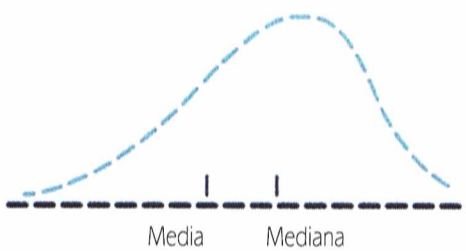
Cuando el histograma es simétrico, como es el caso de la curva de una distribución normal, como la que se muestra en la siguiente figura, el punto de simetría es el punto de división de la curva en dos áreas iguales y además es el punto de equilibrio. En este caso la media y la mediana son iguales.



Por otra parte, cuando el histograma tiene un solo pico (es unimodal), con una cola superior más larga —como se muestra en la siguiente gráfica (a esta característica de la gráfica se le conoce como **sesgo positivo**)— los valores extremos de la cola superior jalan a la media, por lo que generalmente ésta es mayor que la mediana. Por ejemplo, en un examen, una calificación inusualmente alta, eleva la media pero no afecta a la mediana.



Del mismo modo, cuando un **histograma unimodal** está negativamente sesgado (con una cola inferior más larga), como se muestra en la siguiente gráfica, la media es generalmente menor que la mediana.



Puede darse el caso de que la moda no exista: por ejemplo el conjunto de datos 3, 5, 8, 10, 12, 15 y 16 pero también que ésta no sea única; por ejemplo el conjunto de datos 2, 3, 4, 4, 4, 5, 5, 7, 7, 7 y 9 tiene dos modas, 4 y 7, y se llama **bimodal**.

Para saber más

La moda muestral de un conjunto de números es otra medida alternativa de tendencia central. Se define como el valor que ocurre con mayor frecuencia en un conjunto de datos, es decir, el valor más frecuente. Por ejemplo, el conjunto de observaciones 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12 y 18 tiene moda 9.

Para saber más

2.2 es una cantidad numérica que puede servir como medida de tendencia central para un conjunto de datos categóricos es la frecuencia relativa o proporción muestral de una categoría. La proporción muestral de una categoría es la proporción (fracción) de respuestas que están en la correspondiente categoría. Por ejemplo, si una variable categórica tiene sólo dos posibles respuestas (hombre o mujer, enfermo o sano) y etiquetamos a cada una de estas dos posibles respuestas como E (éxito) y la otra F (fracaso), la proporción de éxitos (E) es llamada la proporción muestral de éxitos.

La proporción muestral de éxitos, representada por $\hat{p}$, es $\hat{p} = \dfrac{\text{número de éxitos en la muestra}}{n}$

Gestión del aprendizaje

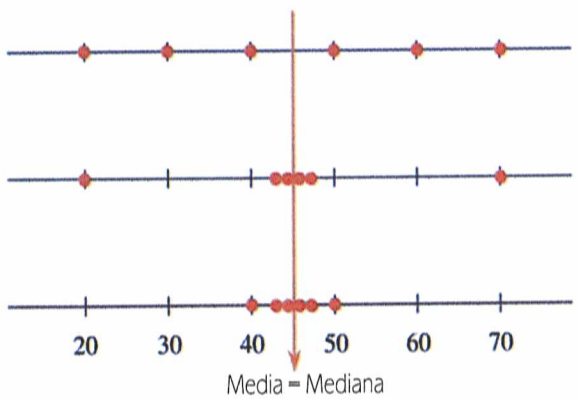
Es posible realizar el cálculo de las medidas de tendencia central en una hoja de cálculo electrónica de manera sencilla. Basta con que sigas los pasos descritos en el 2.

Es importante mencionar la utilidad de la media y la mediana como medidas de tendencia central. La media y la mediana se utilizan comúnmente como parámetros para comparar dos distribuciones de población. Por ejemplo, si se quisiera realizar un estudio comparativo sobre la eficiencia en el tiempo de atención a pacientes entre dos clínicas hospitalarias, habría que partir de la comparación entre los valores de la mediana o de la media de los tiempos de atención a pacientes obtenidos en un tiempo determinado en ambas. El valor de la mediana o media más pequeña correspondiente a una de las clínicas reflejará una mayor eficiencia en la atención a los pacientes; mientras que aquella clínica que muestre un valor calculado de la mediana o media mayor para los tiempos de atención de sus pacientes mostrará ser más ineficiente.

Medidas de dispersión

Una medida de tendencia central sólo proporciona información parcial sobre un conjunto de datos, ya que no permite cuantificar que tanto las observaciones difieren unas de otras. Con la finalidad de aclarar este punto consideremos las siguientes tres muestras:

1. 20, 40, 50, 30, 60, 70
2. 47, 43, 44, 46, 20, 70
3. 44, 43, 40, 50, 47, 46



De acuerdo con sus correspondientes gráficas podemos observar que, a pesar de que tres conjuntos tienen una media y mediana de 45, la dispersión de los datos en los tres casos es completamente distinta. Hay una gran variabilidad en la primera muestra en comparación con la tercera. La segunda se observa menos variable que la primera y más que la tercera. La variabilidad en la segunda muestra se debe a que los dos valores extremos están demasiado lejos del centro.

El **rango muestral** de un conjunto de datos es definido como:
Rango = observación más grande − la observación más pequeña

En general, una mayor variabilidad reflejará un rango muestral más amplio. Sin embargo, la variabilidad es una característica de cualquier conjunto de datos. Las dos primeras muestras del ejemplo anterior presentan rangos calculados de 70 - 20 = 50, pero las gráficas nos permiten observar que hay menos variabilidad en la segunda muestra. Con base en el análisis de las gráficas es posible concluir que si no se consideran todos los datos para el cálculo del rango, la información sobre la verdadera naturaleza de la variación de un conjunto de datos puede ser poco confiable. El no considerar todos los datos limita el uso del rango muestral como una medida confiable para comparar la variabilidad entre dos conjuntos de datos.

Esta limitación del rango hace necesario contar con otras medidas de dispersión con capacidad para incorporar la contribución que tiene cada observación a la variabilidad total del conjunto de datos. He aquí algunas de las medidas de dispersión con esta característica.

Desviación de la media

La **desviación de la media** es una de las medidas de variabilidad más utilizadas. Mide el grado en que cada observación de la muestra se desvía de su correspondiente media muestral $\bar{x}$. Al restar $\bar{x}$ de cada observación se proporciona un conjunto de desviaciones de la media.

Dada una muestra de tamaño n, las n desviaciones de la media muestral son las diferencias

$$(x_1 - \bar{x}), (x_2 - \bar{x}), \ldots, (x_n - \bar{x})$$

Una desviación en particular será positiva si el valor de x correspondiente es mayor que $\bar{x}$ y negativo si el valor de x es menor que $\bar{x}$. Estudiemos el ejemplo siguiente:

Un laboratorio multinacional produce vacunas y las distribuye en diversos países. El año pasado produjo una vacuna contra el virus de la influenza, cuyo costo es variable según el país. En la tabla siguiente se muestran los datos correspon-

dientes al costo de una dosis de la vacuna (convertido a dólares estadounidenses) para algunos países de América Latina.

País	Precio de la vacuna por dosis
Argentina	3.02
Brasil	4.67
Chile	3.28
Colombia	3.51
Costa Rica	3.42
Perú	2.70
Uruguay	2.87

Al estudiar la tabla observamos que hay poca variabilidad en los precios; por lo tanto, para este conjunto de datos $\bar{x}$ = $3.36.

La siguiente tabla muestra los datos junto con sus desviaciones correspondientes, obtenidas al restar 3.36 de cada observación. Tres de las desviaciones son positivas ya que las observaciones son mayores que $\bar{x}$. Las desviaciones negativas corresponden a las que son más pequeñas que $\bar{x}$. Algunas de las desviaciones son muy grandes en **magnitud absoluta** (1.31 y 0.60, por ejemplo), indicando con esto las observaciones que están más lejos de la media muestral.

glosario
Magnitud absoluta: es la medición que no considera o ignora el signo negativo de la cantidad asignada. Por ejemplo la magnitud absoluta de -2 es 2.

País	Precio de la vacuna por dosis	Desviaciones de la media
Argentina	3.02	-0.34
Brasil	4.67	1.31
Chile	3.28	-0.08
Colombia	3.51	0.15
Costa Rica	3.42	0.06
Perú	2.76	-0.60
Uruguay	2.87	-0.49

A mayor variabilidad en la muestra, mayores serán las magnitudes (ignorando los signos) de las desviaciones. Ahora consideraremos cómo combinar las desviaciones en una sola medida de variabilidad. Un primer intento podría ser calculando la desviación promedio al sumar todas las desviaciones medias (esta suma puede representarse mediante $\sum(x-\bar{x})$) y luego dividirla por n.

glosario
Contribución: se define como tal al valor numérico asociado con una desviación media. Por ejemplo una contribución de -3 significa que la diferencia correspondiente a la desviación media entre un valor x y su media $\bar{x}$ es -3.

En la tabla anterior se puede observar que el valor de la suma de las siete desviaciones de la media es aproximadamente cero; es decir que, en la mayoría de los casos, el valor calculado de la suma $\sum(x-\bar{x})$ estará muy cercano a cero o incluso será cero. Ésta es la razón que limita la aplicabilidad de la desviación media como una medida general de la variabilidad de un conjunto de datos muestrales, ya que las **contribuciones** que tengan la misma magnitud absoluta pero diferente signo como -2 y 2 serán anuladas al momento de sumarlas, por lo que su contribución a la variabilidad no se considerará.

Varianza y desviación estándar

Una manera de evitar que se eliminen las contribuciones negativas con las positivas es elevarlas al cuadrado antes de sumarlas. Al hacer esto, las contribuciones con diferente signo pero con la misma magnitud (como -2 y +2) no se anularán al sumarse porque bajo esta medida tendrán idéntica contribución a la variabilidad (ambas tendrán 4). En general, elevar las contribuciones al cuadrado equivale a elevar al cuadrado cada una de las desviaciones de la media. El cuadrado de estas desviaciones para muestras de tamaño n se representa como $(x_1-\bar{x})^2, (x_2-\bar{x})^2, \ldots, (x_n-\bar{x})^2$, así que su respectiva suma es

$$(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2=\sum(x-\bar{x})^2.$$

Al dividir el resultado de la suma por el tamaño de la muestra n se obtiene la desviación cuadrada promedio. Aunque ésta parece ser una medida $\sum(x-\bar{x})^2$ razonable de la variabilidad utilizaremos un divisor ligeramente menor a n, como lo sugiere la teoría de la estimación estadística.

La **varianza muestral**, representada como S^2, es la suma de los cuadrados de las desviaciones de la media divididas entre n-1. Esto es:

$$S^2=\frac{\sum(x-\bar{x})^2}{n-1}$$

La **desviación estándar muestral** es la raíz cuadrada positiva de la varianza muestral y se representa como S.



La teoría de estimación estadística estudia como obtener información acerca de un parámetro desconocido de una población, mediante muestras extraidas de ella. Esto a partir de la aplicación de un conjunto de técnicas que permiten dar un valor aproximado al parámetro desconocido (estimación del parámetro) a partir de los datos proporcionados por una muestra. Por ejemplo, una estimación de la media μ desconocida de una determinada población podría ser la media $\bar{x}$ obtenida para una muestra de la población de tamaño n.

La teoría de la estimación estadística se divide en tres grandes bloques, cada uno de los cuales tiene distintos métodos que se usan en función de las características y propósitos del estudio:

Estimación puntual: 2

- Método de los momentos.
- Método de la máxima verosimilitud.
- Método de los mínimos cuadrados.
- Estimación por intervalos.
- Estimación bayesiana.

Una mayor variabilidad en la muestra da como resultado un valor relativamente grande de S^2 o S, mientras que un valor de S^2 o S cercano a cero indica una menor variabilidad. Debes tomar en cuenta que cualquier unidad utilizada para medir la variable x (por ejemplo, kilogramos o segundos) y el cuadrado de sus desviaciones —y por lo tanto S^2— se encuentran en unidades al cuadrado. Por lo que al tomar la raíz cuadrada, ésta tendrá las mismas unidades que la variable x. Por ejemplo, si se considera a x como la estatura en metros de hombres adultos entonces para una muestra de estatura de 50 hombres, las unidades de desviación estándar de estos datos ($S = 1.75$) también estarán representadas en metros.

Por otra parte si la variable x representa el precio de un libro entonces para una muestra de 30 libros si obtuviéramos una desviación estándar de $S = 125.3$, las unidades de ésta también estarían dadas en pesos ($).

Retomando el ejemplo sobre las dosis de la vacuna contra la influenza y específicamente sobre el costo de la dosis utilizamos las desviaciones de la media calculadas en el ejemplo anterior para calcular la desviación estándar y la varianza muestral. El cálculo de cada una de las desviaciones de la media así como su suma, se proporcionan en la última columna del siguiente cuadro.

País	Precio de la vacuna por dosis	Desviaciones de la media	Desviaciones de la media cuadrada
Argentina	3.02	-0.34	0.1156
Brasil	4.67	1.31	1.7161
Chile	3.28	-0.08	0.0064
Colombia	3.51	0.15	0.0225
Costa Rica	3.42	0.06	0.0036
Perú	2.76	-0.6	0.3600
Uruguay	2.87	-0.49	0.2401
			$\sum(x+\bar{x})^2 = 2.4643$

A partir de este valor obtenemos los valores de la varianza y la desviación estándar de la siguiente forma:

$$S^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{2.4643}{7-1} = \frac{2.4643}{6} = 0.4107$$

y

$$S = \sqrt{0.4107} = 0.641$$

Hay medidas de variabilidad para toda la población, que son análogas a S^2 y S para una muestra. Estas medidas reciben el nombre de **varianza de la población** y la

Gestión del aprendizaje

Es posible realizar el cálculo de las medidas de dispersión en una hoja de cálculo electrónica de manera. Basta con que sigas los pasos descritos en el Apéndice 2.

desviación estándar de la población y se simbolizan σ^2 y σ (léase sigma cuadrada y sigma), respectivamente.

S^2 Varianza muestral
σ^2 Varianza poblacional
S^2 Desviación estándar muestral
σ Desviación estándar poblacional

En muchos procedimientos estadísticos nos gustaría utilizar el valor de σ pero desafortunadamente suele desconocerse. Por lo tanto, en su lugar usamos un valor muestral calculado que esperamos sea cercano a σ; es decir, un valor que sea un buen estimador de σ. Por esta razón se debe calcular el valor de S^2 a partir del cociente cuyo denominador es n-1; ya que en un muestreo repetitivo, en lugar de n, el valor S^2 resultante tiende a estar más cerca de σ^2 en promedio.

Las medidas de dispersión o variabilidad informan sobre cuanto se alejan del centro los valores de dispersión. La medición de la variabilidad es un parámetro para comparar dos distribuciones de población. Por ejemplo, si pretendiéramos comparar la propagación de una enfermedad por edades entre dos grupos poblacionales, calcular una medida de dispersión para una muestra de edades de individuos enfermos de cada población mostrará una aproximación de la variabilidad por edades que el padecimiento exhibe en ambas poblaciones. Por lo tanto, la muestra cuya medida de dispersión calculada sea mayor, mostrará evidencia de que en su correspondiente población existe una mayor variabilidad de edades en las que se ha propagado la enfermedad, mostrando con ello una mayor vulnerabilidad a la enfermedad entre un mayor grupo de edades, que aquella población cuya muestra tuvo una menor dispersión. Esto indicaría la presencia de un problema de salud mayor en la población que mostró más variabilidad muestral.

¡Ahora te toca a ti!

Ahora que te acercaste a los fundamentos y conceptos de la estadística descriptiva es el momento de aplicar los nuevos conocimientos. Retoma el caso planteado al inicio de la unidad y empieza a medir el impacto del fenómeno migratorio en la comunidad de El Encino. Emplea los métodos y conceptos de estadística descriptiva estudiados en esta sección.

Si decidiste resolver un caso alternativo al planteado al principio de la unidad, es momento de que recopiles información para realizarlo. Ubica fuentes de información donde puedas obtener dos muestras representativas de 100 mediciones cada una, correspondientes a dos poblaciones distintas. Esta información te servirá para el cálculo de las medidas de tendencia central y dispersión que te permitan interpretar los resultados obtenidos.



Realiza un análisis descriptivo de dos poblaciones utilizando las medidas de tendencia central y dispersión. Mediante el cálculo e interpretación de estos valores, identifica si existe evidencia suficiente para suponer diferencias sustanciales entre las dos poblaciones.

Es recomendable que tengas acceso a una computadora con procesador de texto, hoja electrónica de cálculo y elaboración de presentaciones.

La actividad está diseñada para que trabajes en ella 10 horas, aproximadamente, y los pasos a seguir para llevarla a cabo son los siguientes:

1. Considera las variables, nivel de ingreso y edad, como parámetros de comparación para obtener las mediciones de dos grupos: los oriundos del lugar y los inmigrantes.
   Con dicha variable podrás identificar si existe diferencia entre lo niveles de ingreso de ambos grupos y medir su contribución a la economía de la comunidad.
2. Los datos se obtienen por medio de dos muestras aleatorias representativas de tamaño n = 100 individuos pertenecientes a los dos grupos de residentes y que forman parte de la población económicamente activa (PEA) de El Encino. Se obtienen para ambas muestras 100 mediciones de cada variable involucrada en el análisis. (Consultar la base de datos en el Apéndice 3).
3. Para cada muestra realiza lo siguiente:
   - Calcula las medidas de tendencia central (mediana y media) así como las medidas de dispersión (rango, varianza y desviación estándar), con las mediciones de ambas variables .
   - Elabora los histogramas de frecuencia para cada variable.
4. Analiza los valores calculados de las medidas de tendencia central y de dispersión, y compara los niveles de ingreso y la estructura por edad entre los grupos que son objeto del estudio. Reflexiona si la comparación de los valores calculados para ambas muestras aporta alguna evidencia que sustente si el flujo migratorio ha sido positivo o no para la comunidad de El Encino.
5. Utilizando histogramas de frecuencia compara la distribución del nivel de ingreso y la de distribución de la edad entre las dos muestras. Reflexiona si la información que obtuviste mediante la comparación te da elementos para concluir si el flujo migratorio ha tenido algún nivel de impacto en la comunidad.
6. Sintetiza tus conclusiones elaborando una presentación en computadora. Para que sea comprensible, deberá contar con los siguientes apartados:

   Diapositivas de inicio o presentación en las que expliques los conceptos de medidas de tendencia central y medidas de dispersión. Incluye en ellas las fórmulas para calcular la media, mediana, rango, varianza y desviación estándar.

   Diapositivas de desarrollo en las que presentes los valores calculados de las medidas de tendencia central, dispersión y los histogramas de frecuencias correspondientes a las

variables: nivel de ingreso y edad para cada una de las dos muestras y las conclusiones obtenidas con relación al impacto que el flujo migratorio tiene sobre la economía de la comunidad. Sustenta tu análisis con la comparación de ambas muestras (cantidades y gráficas).

Para cerrar la actividad, en una o varias diapositivas explica la importancia que tiene el cálculo de las medidas de tendencia central y de dispersión para el análisis de la información.

Al terminar la actividad, y con el fin de evaluar tu desempeño, en la siguiente tabla marca con una cruz (✗) si lograste (L) o no (NL) desarrollar cada uno de los aspectos considerados. Asimismo, en cada caso indica el nivel desempeño que tuviste en el desarrollo de cada uno de estos aspectos. De ser necesario índica que debes mejorar.

Escala de evaluación del nivel de desempeño: 0 Nulo 1 Deficiente 2 Aceptable 3 Satisfactorio				
Aspectos a evaluar	**L**	**NL**	**Escala de evaluación**	**¿Qué debo mejorar?**
Describí los conceptos de medidas de tendencia central y de dispersión.				
Realicé los cálculos de las medidas de tendencia central y de dispersión.				
Realicé la interpretación de las medidas de tendencia central y de dispersión así como de los histogramas de frecuencia.				
Expliqué la importancia que tienen las medidas de tendencia central y de dispersión para el análisis de información.				

La variabilidad de la información es común en diversas situaciones cotidianas. Su presencia dificulta la comparación entre dos cantidades cuyos valores presentan cambios recurrentes. Por eso es que la aplicación de las medidas resumen de la variabilidad —como las medidas de tendencia central y de dispersión— toman sentido.

Por ejemplo, si estás interesado en realizar la comparación de tu desempeño académico entre dos semestres consecutivos, bastará con que calcules la media o promedio correspondiente a las calificaciones de ambos semestres. Obtendrás una medida que además de resumir la variabilidad de tus calificaciones semestrales, te provee con un parámetro que te permite comparar tu desempeño en ambos semestres. Por otra parte, si estuvieras interesado en comparar las estabilidad entre el precio del kilogramo del jitomate y del aguacate durante un año, bastará con que calcules la desviación estándar para el precio de ambos productos durante ese lapso; aquel producto que resulte con menos desviación estándar en el precio, se considerará como el que se comportó de manera más estable durante el año.

No olvides verificar tus respuestas en el Apéndice 1.



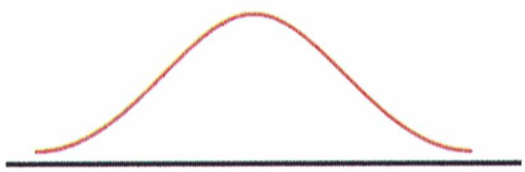
Las distribuciones de probabilidad como herramientas para estimar las probabilidades de eventos relacionados con fenómenos naturales y procesos sociales

Estás trabajando para argumentar la posibilidad de ocurrencia de algunos fenómenos naturales y procesos sociales de tu entorno, mediante el uso de los tipos de distribución de probabilidad. Para evaluar y determinar la probabilidad de ocurrencia y el comportamiento de fenómenos naturales y procesos sociales de importancia en tu contexto para orientar sus acciones en el ámbito individual y proponer estrategias de mejora en tu localidad, región, país, mundo. Para utilizar las TIC en la obtención y procesamiento de la información a través de hojas de cálculo.

En la última sección de la unidad anterior estudiaste las distribuciones de probabilidad teórica normal, binomial y de Poisson, modelos adecuados para el análisis de la distribución de frecuencias de cierto tipo de variables. En esta sección profundizarás su estudio para obtener más herramientas con las cuales calcular probabilidades de eventos relacionados con fenómenos naturales y procesos sociales.

Distribución de probabilidad normal

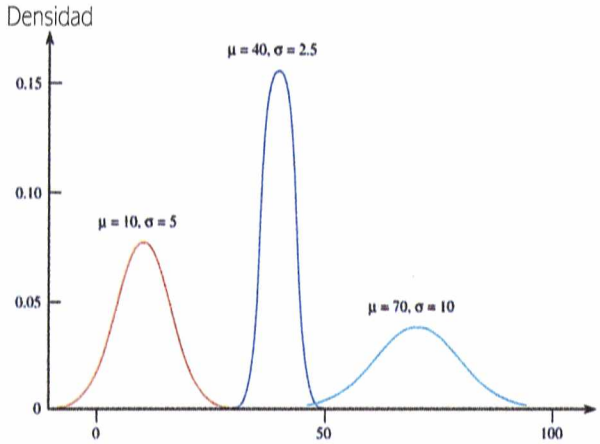
Con anterioridad definimos a las **distribuciones normales** como distribuciones de probabilidad continuas o de curvas normales que tienen forma de campana y son simétricas, como se muestra en la siguiente figura.



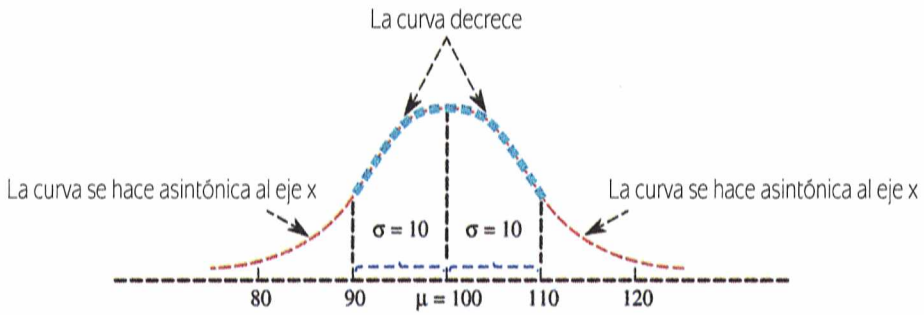
Los tipos de distribuciones normales se distinguen entre sí por los valores de su media μ y su desviación estándar σ. Mientras la media μ de una distribución normal describe dónde está centrada la curva correspondiente, la desviación estándar define la dimensión de la curva cuando se extiende alrededor de ese centro. Estas características son análogas o iguales a las de la media y la desviación estándar, ya que la medida μ es una medida de tendencia central y la desviación estándar σ una medida de dispersión. Al igual que todas las distribuciones de probabilidad continua, el área total bajo cualquier curva normal es igual a 1.

La siguiente figura muestra tres distribuciones normales en las que se puede observar cuánto es menor la desviación estándar o cuánto más alta y más estrecha se vuelve la curva correspondiente, ya que al haber menor variabilidad la mayoría de los valores de la correspondiente distribución normal se concentran a una distancia menor de su media. Recordemos que las áreas bajo la curva de una distribución de probabilidad continua representan probabilidades, entonces cuando la desviación estándar es pequeña, un área más grande se concentra cerca del centro

de la curva, y por lo tanto la posibilidad de observar un valor cercano de la media es mucho mayor, como lo puedes apreciar en la siguiente figura en la que μ y σ es para una curva normal.



El valor de la media es el número que en el eje de medición (eje x) se extiende debajo de la parte superior de la campana. El valor de la desviación estándar σ también puede determinarse a partir de la gráfica de una curva normal, tal y como se observa en la curva normal de la siguiente figura.



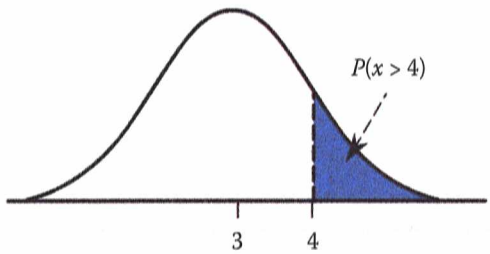
En la parte superior de la campana (por encima de $\mu = 100$) y hacia la derecha, la curva tiende a bajar hasta que está alrededor del valor 110. Después de ese punto, continúa decreciendo, de manera menos notable, haciéndose **asintótica** al eje x. Lo mismo sucede del lado izquierdo de la media $\mu = 100$, ya que la curva decrece hasta que llega al 90 y a partir de este punto el decrecimiento es menos notable haciéndose asintótica al eje x. La curva cambia entonces su intensidad de decrecimiento a una distancia de 10 a cada lado de $\mu = 100$. En general, σ es la máxima distancia a cada lado de μ, a partir de la cual una curva normal disminuye su decrecimiento. Así entonces, para este ejemplo $\sigma = 10$ es el valor de la desviación estándar para esta curva normal.

Si una distribución normal se utiliza como un modelo poblacional, su media y su desviación estándar deben especificarse. Por ejemplo, una distribución normal con media 3 y desviación estándar 1 podría usarse como modelo para la distribución

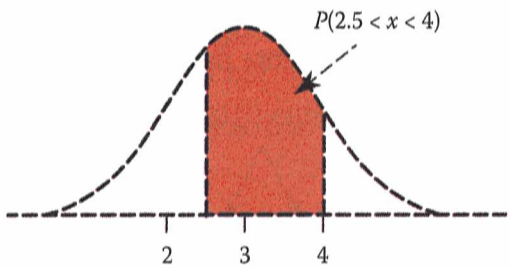
glosario

Asintótico: término utilizado en matemáticas para describir el comportamiento de la gráfica de una función que consiste en la aproximación de la curva de la gráfica de una función a una línea recta horizontal (por arriba o por abajo) pero sin llegar jamás a cortarla.

del peso de niños recién nacidos. Si este modelo es una descripción razonable de esta distribución de probabilidad podríamos usar áreas bajo la curva normal con $\mu = 3$ y $\sigma = 1$ para aproximar probabilidades de diversos eventos relacionados con el peso de un recién nacido. La probabilidad de que el peso de un recién nacido sea mayor a 4 kg —expresado simbólicamente como $P(x > 4)$— corresponde al área sombreada en la siguiente gráfica "Distribución normal para el peso de un recién nacido".



Por otra parte, la probabilidad $P(2.5 < x < 4)$ de que un recién nacido pese entre 2.5 y 4 kg corresponde al área sombreada en la siguiente gráfica.



Gestión del aprendizaje

Es posible realizar el cálculo de las probabilidades bajo el área de la curva de una distribución normal en una hoja de cálculo electrónica, de manera sencilla.

Basta con que sigas los pasos descritos en el Apéndice 2.

El cálculo directo de estas probabilidades (áreas bajo la curva normal) es complejo por lo que se utiliza una tabla de áreas para una distribución normal de referencia, llamada distribución normal estándar.

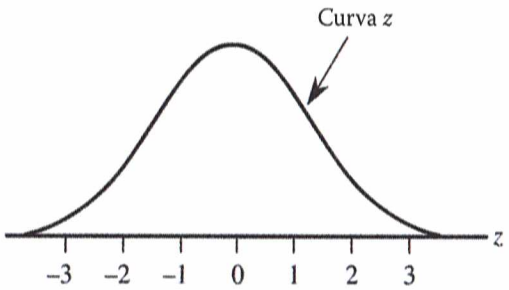
La distribución normal estándar es una distribución normal con $\mu = 0$ y $\sigma = 1$. La curva de densidad correspondiente también es denominada curva z o curva normal estándar. Se acostumbra utilizar la letra z como la variable para representar los posibles valores que puede tomar una distribución normal.

Algunas variables pueden describirse mediante la distribución normal estándar, que también se utiliza en los cálculos de probabilidad para otras distribuciones normales. Cuando estamos interesados en encontrar una probabilidad basada en otra curva normal, podemos traducir el problema a uno "equivalente", que suponga encontrar un área bajo la curva normal estándar. Puede utilizarse una tabla para la distribución normal estándar para encontrar el área deseada, pero para poder hacerlo, primero habría que entender la distribución normal estándar.

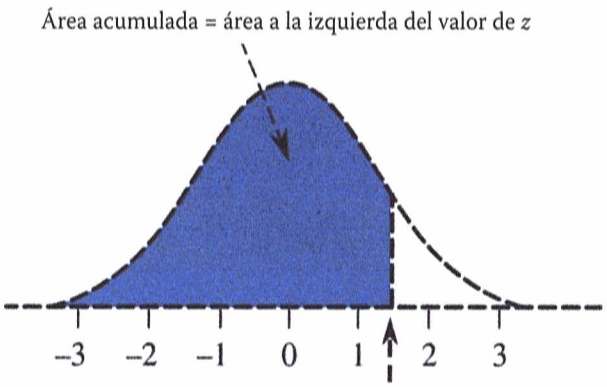
La distribución normal estándar

Para trabajar con la **distribución normal estándar** es necesario saber calcular probabilidades o áreas bajo una curva normal estándar por encima de los intervalos dados de posibles valores para z.

La curva normal estándar (z) que se muestra en la siguiente figura está centrada en $\mu = 0$ y tiene desviación estándar $\sigma = 1$; este último valor se emplea como medida del grado en que posibles valores de la variable z se extienden alrededor de la media (en este caso, 0). Por ejemplo, alrededor del 95% del área (probabilidad) se asocia con valores que están dentro de dos desviaciones estándar de la media (entre -2 y 2) y casi toda el área en la zona de los valores que están dentro de tres desviaciones estándar alrededor de la media (entre -3 y 3).



Ahora bien, para calcular la curva normal estándar se utiliza una tabla con los valores acumulados bajo el área de la curva z para valores z^* distintos de la variable z, que puedes consultar en el Apéndice 4 y del tipo mostrado en la siguiente figura:



Un valor particular del valor de z

En esta tabla, el valor más pequeño de z^* es -3.89, el siguiente es -3.88, y así sucesivamente en incrementos de 0.01 terminando con el área acumulada total a la izquierda de 3.89 de 1 o 100%.

Gestión del aprendizaje

Uso de la tabla normal estándar para el cálculo de probabilidades

Para cualquier número z* entre -3.89 y 0, redondeado a dos cifras decimales, la tabla del Apéndice 4 calcula los valores $P(z<z^*)$, cuya expresión representa el área de la curva acumulada a la izquierda de z^*.

Para calcular el valor de esta probabilidad con la tabla se deben llevar a cabo dos pasos:

1. Localizar el renglón correspondiente al dígito a cada lado del punto decimal del valor observado de z*. Por ejemplo si el valor observado es z*=–1.75, en la tabla se debe localizar el renglón correspondiente a z* de –1.7.
2. Ubicar la columna identificada con el segundo dígito a la derecha del punto decimal en z^*. Por ejemplo, con el valor observado de z*=–1.75 en la tabla se debe localizar la columna correspondiente al valor de .05.

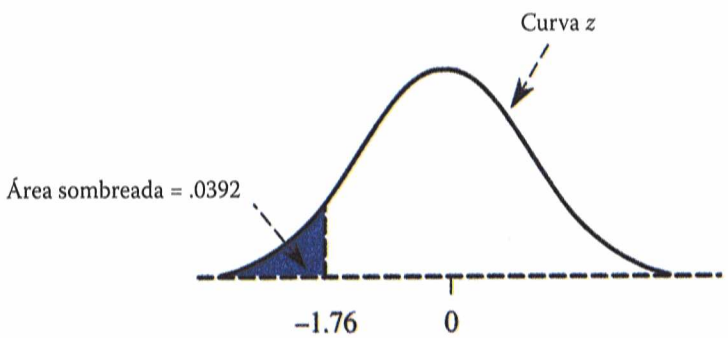
La probabilidad deseada P(z<z*) es el número en la intersección de este renglón con esta columna. Por ejemplo, para z*=–1.75 la probabilidad deseada es P(z<–1.75) = 0.0401.

En la siguiente sección de la tabla de la curva normal estándar se puede ubicar el área bajo la curva *z* a la izquierda de 1.42. Para hacerlo hay que buscar la fila con la etiqueta 1.40 y la columna etiquetada con 0.02 (la fila y la columna resaltadas en azul). El valor correspondiente a la intersección de esta fila y esta columna es 0.9222. Por lo tanto, el área de la curva z a la izquierda de 1.42 = 0.9222.

También se puede utilizar en la tabla el área a la derecha de un valor particular. Debido a que el área total bajo la curva de z es 1, se deduce que el área de la curva z a la derecha de 1.42 es 1-(el área de la curva z a la izquierda de 1.42) = 1-0.9222 = 0.0778.

Estas probabilidades se pueden interpretar en el sentido de que en una larga secuencia de observaciones, aproximadamente 92.22% de los valores observados z será menor que 1.42 y 7.78% será mayor de 1.42.

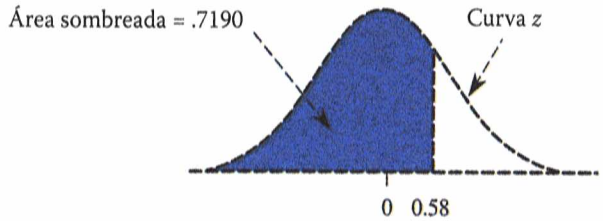
z*	0.00	0.01	0.02	0.03	0.04	0.05
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678



Trabajemos el siguiente ejemplo:

La probabilidad $P(z < -1.76)$ se encuentra en la intersección de la fila -1.7 y de la columna de 0.06. El resultado es $P(z < -1.76) = 0.0392$ como se muestra en la figura de la izquierda.

En otras palabras, en una secuencia larga de observaciones 3.9% de los valores observados z, aproximadamente, será menor que -1.76. Del mismo modo, el valor corresponde $P(z \leq 0.58)$ al valor 0.7190 identificado en la intersección de la fila 0.5 y la columna 0.08 de la tabla. El área bajo la curva correspondiente a esta probabilidad se muestra en la siguiente gráfica:



Ahora considérese $P(z < -4.12)$. Esta probabilidad no aparece en la tabla del Apéndice 4, pues no hay un renglón con el valor de -4.1. Sin embargo, debe ser menor que $P(z < -3.89)$, el valor más pequeño de z* en la tabla, ya que -4.1 está más lejos en la cola izquierda de la curva z. Como $P(z < -3.89) = .0000$ (esto es cero, redondeado a cuatro cifras decimales), se sigue que

$$P(z < -4.12) \approx \text{(se lee aproximadamente). } 0000.$$

De manera similar,

$$P(z < 4.12) > P(z < 3.89) = 1.0000$$

de donde concluimos que

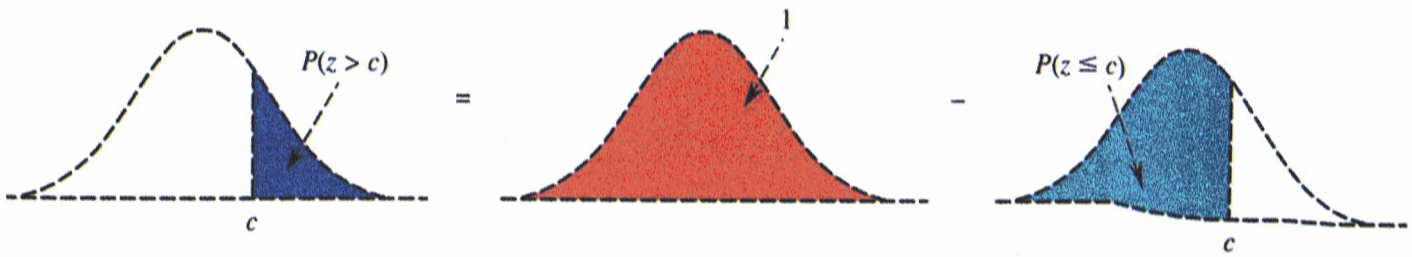
$$P(z < -4.12) \approx 1.0000$$

Podemos usar las áreas acumuladas tabuladas en la tabla del Apéndice 4 para calcular otras probabilidades que involucran a z. La probabilidad de que z sea mayor que un valor c es:

$$P(z > c) = \text{área bajo la curva z al lado derecho de c.}$$

O en forma equivalente $P(z > c) = 1 - P(z \leq c)$

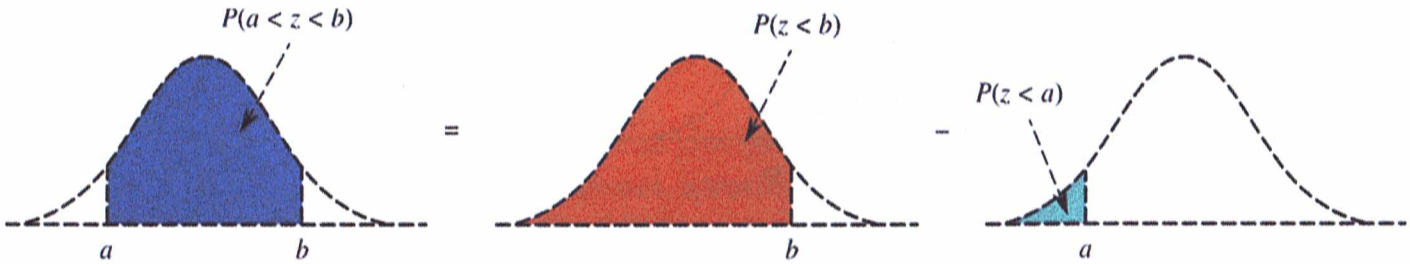
En otras palabras, el área a la derecha de un valor (un área de cola derecha) es 1 menos su correspondiente área acumulada, como se ilustra en la siguiente gráfica.



De igual manera, la probabilidad de que z caiga en un intervalo con límite inferior y límite superior b es

$P(a < z < b)$ = área bajo la curva z en el intervalo de a a b = $P(a < b) - P(z < a)$.

Esto es, $P(a < z < b)$ es la diferencia entre dos áreas acumuladas, como se muestra en la siguiente figura.

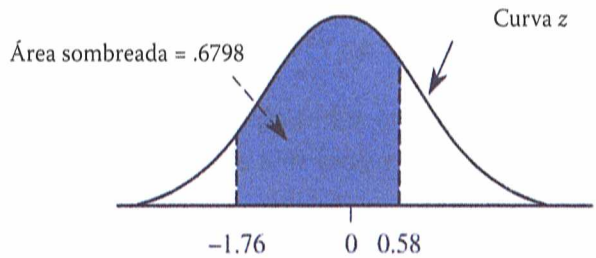


Vayamos con el siguiente ejemplo:

La probabilidad de que z esté entre -1.76 y 0.58 es

$$P(-1.76 < z < 0.58) = P(z < 0.58) - P(z < -1.76)$$
$$= 0.7190\text{-}0.0392$$
$$= 0.6798$$

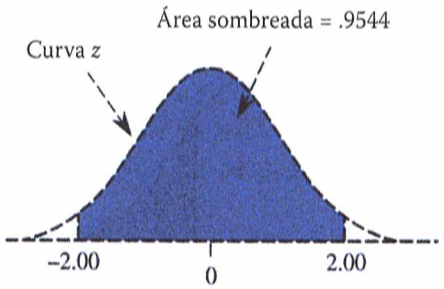
Como se muestra en la siguiente figura:



La probabilidad de que z esté entre 2 y -2 (dentro de dos desviaciones estándar de la media, ya que $\mu = 0$ y $\sigma = 1$) es

$$P(-2 < z < 2) = P(z < 2) - P(z < -2)$$
$$= 0.9544$$
$$= 0.95$$

como se muestra en la siguiente gráfica:

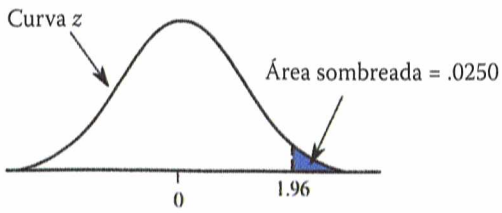


Esta última probabilidad es la base para una parte de la regla empírica que establece que cuando un histograma está bien aproximado por una curva normal, aproximadamente 95% de los valores están dentro de dos desviaciones estándar de la media.

La probabilidad de que el valor de z exceda 1.96 es

$$P(z > 1.96) = 1 - P(z \leq 1.96)$$
$$= 1\text{-}0.9750$$
$$= 0.0250$$

como se muestra en la siguiente figura:



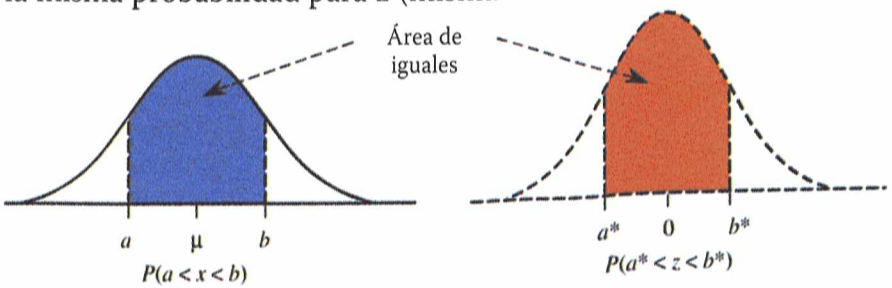
Es decir, 2.5% del área bajo la curva z se encuentra a la derecha de 1.96, en el extremo superior. De igual forma:

$$P(z > -1.28) = \text{área a la derecha de } -1.28$$
$$P(z > -1.28) = 1 - P(z \leq -1.28)$$
$$= 1\text{-}.1003$$
$$= .8997$$
$$= .90$$

Otras distribuciones normales

Las áreas de la curva z también pueden emplearse para el cálculo de probabilidades de cualquier distribución normal. Usaremos la letra z sólo para aquellas variables que tienen una distribución normal estándar y la letra x para cualquier variable cuya distribución se describe mediante una curva normal con media μ y desviación estándar σ. Si queremos calcular $P(a < z < b)$, la probabilidad de que la variable x se encuentre en un rango determinado. Esta probabilidad corresponde al área bajo la curva normal y por encima del intervalo de a a b.

Para obtener esta probabilidad se debe encontrar un problema equivalente, relacionado con la distribución normal estándar. Lo anterior significa que la determinación de un intervalo (a*, b*), tiene la misma probabilidad para z (misma área bajo la curva z), al igual que el intervalo (a, b) en nuestra distribución normal original, como se muestra en la siguiente figura en la que se representa la igualdad entre las áreas de una curva normal estándar y una curva normal y uno no estándar.



La notación asterisco (*) se utiliza para distinguir a y b (los valores de la distribución normal original con media μ y desviación estándar σ), de a* y b *, los valores correspondientes a la curva de z. Para encontrar a* y b*, simplemente se deben calcular las puntuaciones z para los puntos finales del intervalo para el que se desea calcular una probabilidad. Este proceso se llama normalización de los criterios de valoración. Por ejemplo, si la variable x tiene una distribución normal con media $\mu = 100$ y desviación estándar $\sigma = 5$, para encontrar

$$P(98 < z < 107)$$

lo primero que se debe hacer es convertir este problema en un problema equivalente a la distribución normal estándar. Para ello, comenzaremos calculando la puntuación z en el punto final inferior a = 98, para lo cual a partir de dicho valor se le resta su media y posteriormente se le divide entre su desviación estándar, convirtiendo el punto final inferior a = 98 en el puntaje estandarizado dado por

$$a^* = \frac{98-100}{5} = -\frac{2}{5} = -0.40$$

y convirtiendo el punto final b=107, para transformarse en

$$b^* = \frac{107-100}{5} = \frac{7}{5} = 1.40$$

Entonces

$$P(98 < z < 107) = P(-0.40 < z < 1.40).$$

La probabilidad puede ahora evaluarse utilizando la tabla del Apéndice 4.



La puntuación z es la puntuación típica o resultado estándar individual más utilizada. Gracias a la puntuación z podemos comparar los resultados de un individuo (o de varios) en distintas pruebas en las que las distribuciones de los resultados de los otros miembros de la población sean distintas.

La puntuación z correspondiente a un valor particular es

$$z = \frac{\text{valor} - \text{media}}{\text{desviación estándar}}$$

La puntuación z nos dice que tantas desviaciones estándar está el valor de la media. Es positivo o negativo a si el valor se encuentra por encima o por debajo de la media.

Encontrando probabilidades de una distribución normal

A fin de calcular las probabilidades para cualquier distribución normal se deben estandarizar los valores correspondientes al punto inferior y superior de la distribución normal original y luego usar la tabla de áreas de la curva z. De manera específica, si x es una variable cuyo comportamiento está descrito por una distribución normal con media y desviación estándar, entonces

$$P(x<b)=P(z<b^*)$$
$$P(x<a)=P(z<a^*)$$
$$P(a<x<b)=P(a^*<z<b^*)$$

donde z es una variable cuya distribución es normal estándar y

$$a^*=\frac{a-\mu}{\sigma} \quad b^*=\frac{b-\mu}{\sigma}$$

Una vez más, ejemplifiquemos.

En el estudio *Los parámetros de crecimiento fetal y peso al nacer: su relación con la composición corporal neonatal* se sugiere que una distribución normal con media $\mu = 3500$ gramos y desviación estándar $\sigma = 600$ gramos representa un modelo razonable para la distribución de probabilidad de variable numérica continua x = peso al nacer de un bebé recién nacido seleccionado al azar. ¿Qué proporción de pesos de recién nacidos está entre 2900 y 4700 gramos?

Para responder esta pregunta, debemos encontrar

$$P(2900<x<4700)$$

En primer lugar, se transforman los extremos $a = 2900$ y $b = 4700$ del intervalo correspondiente a la distribución de la variable x, a los extremos a^* y b^* del intervalo equivalente a la distribución normal estándar:

$$a^*=\frac{a-\mu}{\sigma}=\frac{2900-3500}{600}=-1.00$$

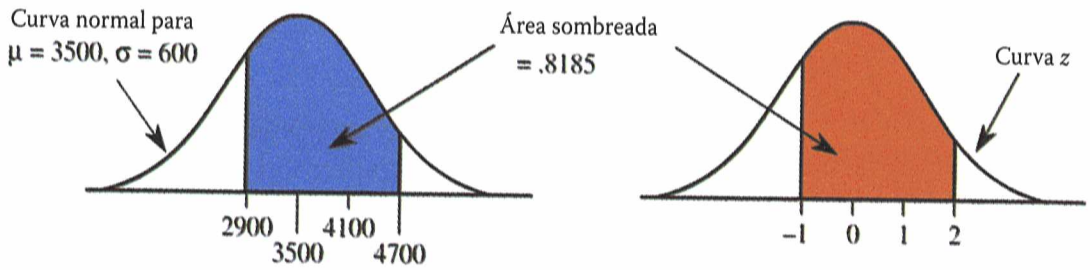
$$b^*=\frac{b-\mu}{\sigma}=\frac{4700-3500}{600}=2.00$$

Entonces

$$P(2900<x<4700)=P(-1.00<z<2.00)$$

= (área de la curva z a la izquierda de 2.00)
– (área de la curva z a la izquierda de)
= .9772 – 0.1587
= 0.8185

Las probabilidades de x y z se muestran en la siguiente gráfica.



Este resultado significa que si se midieran los pesos al nacer para muchos bebés de esta población, el peso del 82% de ellos se situaría entre 2900 y 4700 gramos. Ahora bien, ¿cuál es la probabilidad de que un niño elegido al azar tenga un peso al nacer mayor de 4500? Para calcular esta probabilidad debemos evaluar $P(x > 4500)$.

Así que primero calculamos

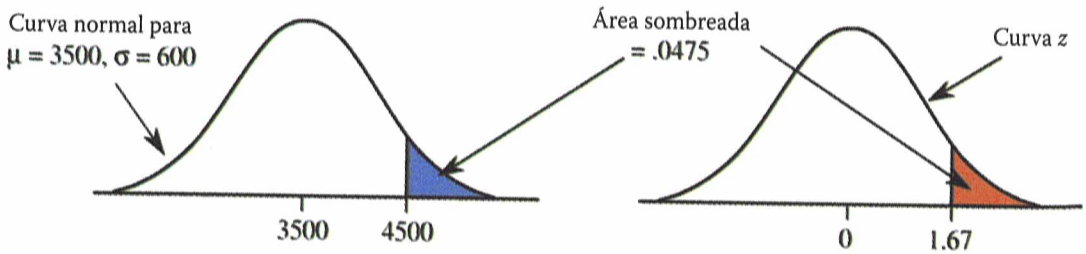
$$a^* = \frac{a-\mu}{\sigma} = \frac{4500-3500}{600} = 1.67$$

Entonces

$$P(x > 4500) = P(z > 1.67)$$

= área de la curva z a la derecha de 1.67
= 1 – (área de la curva z a la izquierda de 1.67)
= 1 – 0.9525
= 0.0475

Las probabilidades de x y z se muestran en la siguiente gráfica.



Distribución de probabilidad binomial

Si quisiéramos registrar el sexo de cada uno de los próximos cinco niños recién nacidos en un hospital comunitario, podríamos cuestionarnos: ¿Cuál es la probabilidad de que uno de los cinco sea mujer?, ¿cuál de que entre dos y cuatro sean mujeres? Antes de responder las dos preguntas centremos nuestra atención en la naturaleza del experimento. Si lo analizamos, podemos percatarnos que éste posee las siguientes características:

1. El experimento consiste de $n = 5$ intentos (ensayos) idénticos, pues cada intento consiste en registrar el sexo de un recién nacido.
2. Cada intento da lugar exactamente a dos resultados, ya que cada recién nacido sólo puede ser registrado como hombre o mujer.
3. Los cinco intentos son independientes pues el resultado obtenido de cada registro es independiente de los resultados obtenidos de los otros registros.
4. La probabilidad de que al registrar el sexo de un recién nacido sea niña no varía de un registro a otro. Esta probabilidad a la cual denotamos como p representa la probabilidad de que en cualquier ensayo particular se obtenga un éxito (en este caso que el sexo del recién nacido sea niña).

Por las características descritas el experimento es de carácter **binomial**, ya analizado en la unidad 1 es fórmula para calcular las probabilidades de sus valores asociados es.

Fórmula de la distribución binomial

Sea
n = número de ensayos independientes en un experimento binomial.
p = la probabilidad constante de que en cualquier ensayo particular se obtenga un éxito.
Entonces

$$P(x) = P(x \text{ éxitos en } n \text{ ensayos}) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$x = 0, 1, \ldots, n$$

La expresión $\binom{n}{x}$ algunas veces se utiliza en lugar de $\frac{n!}{x!(n-x)!}$, cuya expresión representa el número de maneras de elegir x elementos de un conjunto de n elementos.

El símbolo de $n!$ (Léase " n factorial ") se define como

$$n! = n(n-1)(n-2)\ldots(2)(1)$$

y $0! = 1$

La fórmula de la probabilidad binomial puede entonces expresarse como

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$x = 0, 1, \ldots, n$$



La fórmula de la distribución binomial se puede deducir a partir de la aplicación de conceptos de probabilidad básica. Si tienes interés en conocer como se deduce esta fórmula puedes consultar el Apéndice 5.

Por lo que si suponemos que la probabilidad de éxito es $p = 0.4$ y queremos calcular la probabilidad de que uno de los próximos cinco niños recién nacidos sea mujer, basta con sustituir en la fórmula los valores $n = 5$ y $x = 1$ para obtener:

$$P(1) = P(x = 1)$$
$$= \binom{5}{1}(.4)^1(.6)^4$$
$$(.6)^4 = .26$$

Y si ahora queremos calcular la probabilidad de que entre dos y cuatro recién nacidos sean mujeres obtenemos

$$P(2 \leq x \leq 4) = P(x = 2 \text{ o } x = 3 \text{ o } x = 4)$$

Dado que estos resultados son disjuntos, esto es igual a

$$P(x = 2 \text{ o } x = 3 \text{ o } x = 4) = P(x = 2) + P(x = 3) + P(x = 4)$$
$$= P(2) + P(3) + P(4)$$
$$= \binom{5}{2}(.4)^2(.6)^3 + \binom{5}{3}(.4)^3(.6)^2 + \binom{5}{4}(.4)^4(.6)^1$$
$$= \frac{5!}{2!3!}(.4)^2(.6)^3 + \frac{5!}{3!2!}(.4)^3(.6)^2 + \frac{5!}{4!1!}(.4)^4(.6)^1$$
$$= (10)(.4)^2(.6)^3 + (10)(.4)^3(.6)^2 + (5)(.4)^4(.6)^1$$
$$= .3456 + .2304 + .0768$$
$$= .65$$

Para saber más

Dos eventos son disjuntos si no suceden al mismo tiempo (en otras palabras, si no tienen resultados en común). Por ejemplo, si consideramos los eventos llueve (L) y no llueve (NL), es claro que estos eventos son disjuntos, ya que no pueden suceder los dos al mismo tiempo. El cálculo de la probabilidad para dos eventos disjuntos se realiza de la siguiente manera: supongamos que queremos calcular la probabilidad del evento de que llueva o no llueva al cual podemos representar como (L o NL). La probabilidad de este evento a la cual denotamos como $P(L \text{ o } NL)$ se calcula como resultado de la suma de las probabilidades $P(L) + P(NL)$. Este resultado es generalizable para más de dos eventos, pues si $A1, A2, \ldots, An$ son n eventos disjuntos, entonces $P(A1 \text{ o } A2 \text{ o} \ldots 0\ An) = P(A1) + P(A2) + \ldots + P(An)$.

Ejemplo

Un botánico realiza la fertilización cruzada de especies afines de plantas, con flores blancas y azules, produciendo vástagos de los cuales 20% tiene flores blancas. Este experimento es una aplicación de la teoría de Mendel sobre caracteres hereditarios. El ensayo consiste en cruzar una planta de flores azules con una de flores blancas; un éxito es obtener de esta cruza un vástago de flores blancas y $p = .20$. De acuerdo con la teoría de Mendel, los ensayos son independientes; en consecuencia el experimento es binomial con $n = 6$ y $p = 0.2$. Si deseamos determinar la probabilidad de que a lo más haya dos flores blancas entre sus vástagos, se define la variable.

x = el número de vástagos con flores blancas

Entonces las probabilidades para los posibles valores de x pueden calcularse a partir de la fórmula de la probabilidad binomial

$$P(x) = \binom{6}{x}(.2)^x(.8)^{6-x}$$

$$x = 0, 1, 2, 3, 4, 5, 6.$$

La probabilidad de que haya a lo más dos flores blancas entre sus vástagos es

$$P(x \leq 2) = P(x = 0 \text{ o } x = 1 \text{ o } x = 2)$$

$$= P(x = 0) + P(x = 1) + P(x = 2)$$

$$\binom{6}{0}(.2)^0(.8)^6 + \binom{6}{1}(.2)^1(.8)^5 + \binom{6}{2}(.2)^2(.8)^4$$

$$= \frac{6!}{0!6!}(.2)^0(.8)^6 + \frac{6!}{1!5!}(.2)^1(.8)^5 + \frac{6!}{2!4!}(.2)^2(.8)^4$$

$$= (.2)^0(.8)^6 + (6)(.2)^1(.8)^5 + (15)(.2)^2(.8)^4$$

$$= .2621 + .3932 + .2458$$

$$= .90$$

Distribución de probabilidad de Poisson

En un hospital de cierta comunidad llegan tres pacientes por hora al servicio de emergencias. ¿Cuál es la probabilidad de que en una hora lleguen exactamente dos?, ¿cuál es la probabilidad de que en una hora lleguen al menos dos pacientes? Antes de poder responder tales preguntas es necesario establecer las siguientes condiciones sobre el evento:

1. El número promedio de pacientes que acude al hospital por unidad de tiempo (para este evento la unidad de tiempo será considerada una hora) es constante.
2. La probabilidad de que más de un paciente acuda en cualquier intervalo de tiempo menor es casi cero. Esto significa que la probabilidad de que dos o más

pacientes lleguen a la clínica durante un intervalo de tiempo, por ejemplo, de un segundo es muy pequeña o casi cero.

3. El número de pacientes que acude al hospital en intervalos ajenos de tiempo son independientes unos de otros; es decir, que el número de pacientes que acudió a la clínica en un intervalo concreto de tiempo no influye sobre el número de pacientes que acudirá durante un próximo intervalo de tiempo.

glosario

Experimento de Poisson: experimento en el que se analiza el número de veces que ocurre un evento por unidad de tiempo o unidad de espacio.

Si este evento cumple con estas hipótesis entonces es posible suponer que se trata de un **experimento de Poisson**, ya estudiado, al igual con su distribución de probabilidad, en la unidad anterior.

Si expresamos con x la variable que representa el número de sucesos resultantes de este experimento, entonces ésta representa una variable discreta que puede tomar valores enteros (0, 1, 2, 3, 4, 5). La fórmula que nos permitirá calcular las probabilidades de los valores asociados con x se da a continuación:

Fórmula de la distribución de Poisson

Sea

λ = número medio de observaciones por unidad de tiempo
e = base de los logaritmos neperianos. Su valor aproximado es de 2.71828

$$P(x)=\frac{\lambda^x e^{-\lambda}}{x!}$$
$$x=0,1,2,\ldots$$

Un ejemplo nos auxiliará a una mejor comprensión del experimento.

Ahora podemos calcular la probabilidad de que en una hora lleguen exactamente dos pacientes al tomar $x = 2$ y $\lambda = 3$ (número promedio de pacientes que arriban al servicio de emergencias por unidad de tiempo) y sustituir en la fórmula de la distribución Poisson

$$p(2)=\frac{(3)^2 e^{-3}}{2!}$$
$$=\frac{(9)(0.4979)}{2}$$
$$=.22$$

Asimismo, la probabilidad de que en una hora lleguen al menos dos pacientes es

$$P(x\geq 2)=P(x=2 \text{ o } x=3 \text{ o } x=4 \text{ o } \ldots\ldots)$$

Dado que este resultado es difícil de evaluar pues considera una infinidad de eventos (tantos como valores puede tomar la x). Entonces recurrimos al cálculo de esta

probabilidad a partir de su complemento, de donde resulta que

$$\begin{aligned} P(x \geq 2) &= 1 - P(x < 2) \\ &= 1 - P(x = 0 \text{ o } x = 1) \\ &= 1 - P(x = 0) - P(x = 1) \\ &= 1 - P(0) - P(1) \\ &= 1 - \frac{(3)^0 e^{-3}}{0!} - \frac{(3)^1 e^{-3}}{1!} \\ &= 1 - e^{-3} - 3e^{-3} \\ &= .80 \end{aligned}$$



Un evento A es complementario a un evento B si A está compuesto por los eventos que no están en el evento B. Si A es un evento complementario de B en términos probabilísticos se tiene que $P(A) = 1\text{-}P(B)$. Por ejemplo, si consideramos al fenómeno natural temperatura ambiente, entonces el evento frio (F) es complementario del evento templado (T) o caluroso (C). De donde en términos probabilísticos $P(F) = 1\text{-}P(T \text{ o } C)$.

Debe tomarse en cuenta que si una distribución de probabilidad se especifica mediante una fórmula, ésta permite el cálculo de distintas probabilidades sin tener que recurrir al uso de una tabla o un histograma de frecuencias.

¡Ahora te toca a ti!

Ya aprendiste a utilizar algunos modelos de distribuciones de probabilidad teórica para estimar las probabilidades de ocurrencia de cierto tipo de eventos. Ahora es momento de aplicarlos en el cálculo de probabilidades de eventos en contextos más específicos relacionados con fenómenos naturales y procesos sociales.

Para hacerlo realiza la siguiente actividad en la cual retomarás la investigación del impacto migratorio en la comunidad de El Encino.

Si decidiste llevar a cabo la actividad alternativa te sugiero que identifiques variables que sean de interés y que además por sus características puedan modelarse con algunos de los distintos tipos de distribuciones de probabilidad estudiados: distribución de probabilidad normal estándar, binomial y de Poisson.



A partir de los modelos de distribución de probabilidad normal, binomial y de Poisson realizarás el cálculo de probabilidades de eventos relacionados con fenómenos naturales y procesos sociales, que por sus características pueden llegar a tener impacto en el entorno bajo estudio. Podrás identificar cuáles de estos eventos tienen mayor probabilidad de ocurrir.

Para el desarrollo de esta actividad es recomendable que tengas acceso a una computadora con procesador de texto y hoja de cálculo electrónica.

Gestión del aprendizaje

Es posible realizar el cálculo de probabilidades de las distribuciones de probabilidad binomial y de Poisson en una hoja de cálculo electrónica de manera sencilla. Basta con que sigas los pasos descritos en el Apéndice 2.

El desarrollo de la actividad puede tomarte 10 horas, aproximadamente y los pasos para realizarla son los siguientes:

1. Lee y analiza el seguimiento del caso planteado al inicio de la unidad.

   El investigador de El Encino contacta a un investigador de El Maguey, población aledaña que por sus características geográficas guarda muchas similitudes con El Encino y en la cual el investigador contactado realizó un estudio del impacto del flujo migratorio.

   Después de hablar sobre su estudio, el investigador solicita información sobre las variables pertinentes y sus correspondientes distribuciones teóricas que haya obtenido para El Maguey. Dadas las similitudes entre las dos comunidades considera que los modelos probabilísticos pueden ser de utilidad para el cálculo de probabilidades de eventos relacionados con estas variables en El Encino. Las variables consideradas y sus correspondientes distribuciones de probabilidad teóricas se muestran en la siguiente tabla.

Variable	Tipo de distribución teórica
x_1 = Ingreso diario de la población económicamente activa de origen foráneo.	normal con media $\mu = 60$ y desviación estándar $\sigma = 12$
x_2 = Ingreso diario de la población económicamente activa oriunda.	normal con media $\mu = 50$ y desviación estándar $\sigma = 5$
x_3 = Número de trabajadores de origen foráneo que pueden ser encontrados en una muestra de 100 trabajadores que se dedican a labores que requieren para su desempeño un alto nivel de especialización.	Binomial con $n = 100$ y $p = .6$
x_4 = Número de personas que inmigran a la comunidad de El encino por semana.	Poisson con $\lambda = 12$
x_5 = Número de personas que emigran de la comunidad de El encino.	Poisson con $\lambda = 12$

2. Utiliza estos cinco modelos teóricos para calcular las siguientes probabilidades:

   Si la percepción salarial diaria que un trabajador debe tener para ser catalogado dentro del grupo de nivel de ingreso medio a alto de la comunidad es de al menos $65 diarios. Calcula la probabilidad de los siguientes eventos:

   a) $P(x_1 \geq 65)$

   b) $P(x_2 \geq 65)$

   La interpretación que se puede dar a ambas probabilidades es la siguiente: si dos individuos son seleccionados al azar entre la población económicamente activa de El Encino (uno de origen foráneo y otro oriundo de la comunidad), calcula la probabilidad de que ambos tengan de un sueldo al menos de $65 diarios.

   Con base en los resultados obtenidos, ¿tendrás evidencia sufiente para afirmar que las condiciones salariales de los individuos de la población económicamente activa de

origen foráneo son mejores que las de los originarios de la comunidad? De ser así, ¿qué medidas propondrías para equilibrar esta desigualdad en el ingreso?

3. Calcula la probabilidad del siguiente evento:

$$P(x_3 > 50)$$

La probable interpretación a esta probabilidad es la siguiente: si 100 individuos de la población económicamente activa son seleccionados al azar, ¿cuál es la probabilidad de que más de la mitad de ellos sean de origen foráneo?

Con base en el resultado obtenido, ¿ hay evidencia para argumentar que dentro de la población económicamente activa hay un mayor número de empleados de origen foráneo dedicados a labores que requieren una alta calificación para su desempeño que empleados oriundos de la localidad con esta característica? De ser así, ¿qué medida propondrías para equilibrar esta desigualdad?

4. Calcula la probabilidad de los siguientes eventos:

a) $P(x_4 = 7)$

b) $P(x_5 = 7)$

La interpretación posible a ambas probabilidades es la siguiente: ¿cuál es la probabilidad de que en el transcurso de una semana exactamente siete individuos hayan inmigrado hacia la comunidad?, ¿cuál es la de que en el transcurso de una semana exactamente siete individuos hayan emigrado de la comunidad?

Con base en los resultados obtenidos responde si consideras que tales probabilidades evidencian que la comunidad se caracteriza por ser una zona receptora más que expulsora de población.

De la selección de un modelo de probabilidad teórico adecuado para una variable depende la validez de los cálculos probabilísticos que se obtengan. Si decidiste realizar una actividad alternativa es importante que al momento de seleccionar variables verifiques sus características para definir si pueden modelarse mediante alguna de las distribuciones teóricas: normal, binomial y de Poisson.

Reporta los resultados obtenidos en un reporte escrito con una extensión aproximada de una cuartilla. Al inicio del reporte deberás describir las fórmulas que utilizaste en el cálculo de probabilidades. En el cuerpo del reporte calcula las probabilidades consideradas, indicando cada uno de los pasos que realizaste para obtener los resultados y para concluir, describe brevemente cómo aplicaste el criterio de selección de distribuciones teóricas. Da respuesta a cada una de las preguntas realizadas durante el desarrollo de las actividades; expón los argumentos que te llevaron a contestarlas

Al terminar la elaboración del reporte, léelo y corrígelo para después evaluarlo por medio de la siguiente tabla. Marca con una cruz (✗) si lograste (L) o no (NL) desarrollar cada uno de los aspectos considerados. En cada caso indica el nivel desempeño que

tuviste en el desarrollo de cada uno de estos aspectos. De ser necesario indica que debes mejorar.

Escala de evaluación del nivel de desempeño: 0 Nulo 1 Deficiente 2 Aceptable 3 Satisfactorio				
Aspectos a evaluar	**L**	**NL**	**Escala de evaluación**	**¿Qué debo mejorar?**
Describí las fórmulas y los procedimientos necesarios para el cálculo de probabilidades de eventos relacionados con los fenómenos considerados en el contexto tratado.				
Realicé los cálculos de las probabilidades de los eventos considerados.				
Respondí las preguntas formuladas y argumenté los resultados obtenidos.				

La naturaleza de cierto tipo de mediciones continuas, como la estatura y el peso de un individuo, determina que la distribución de frecuencias se comporte de manera simétrica alrededor de su media, lo que propicia que las mediciones asociadas a este tipo de características puedan modelarse adecuadamente por medio de una distribución de probabilidad normal.

Ciertas situaciones experimentales o eventos como el lanzamiento sucesivo de una moneda un cierto número de veces donde en cada lanzamiento sólo se obtienen dos resultados posibles (cara o cruz) o bien el número de personas enfermas en una muestra representativa, donde la condición de salud de cada persona seleccionada al azar tiene sólo dos posibilidades (enfermo o sano), son sólo algunos ejemplos de eventos o experimentos que generan variables que —bajo el cumplimiento de ciertas hipótesis— pueden expresarse mediante una distribución de probabilidad binomial.

Eventos como el número de inmigrantes que llegan a una cierta región en un determinado tiempo o el número de huracanes ocurridos durante el verano en una zona del mar Caribe son situaciones que, bajo el cumplimiento de ciertos supuestos, generan variables, que pueden modelarse mediante una distribución de probabilidad de Poisson.

A partir de lo expuesto se entiende porque estos tres tipos de distribuciones de probabilidad teórica se utilizan como modelos teóricos para el cálculo de probabilidades de eventos asociados con sus mediciones.

No olvides verificar tus respuestas en el Apéndice 1.

El modelo de regresión y el de correlación lineal como medidas para describir la asociación entre variables

El tamaño inusualmente grande del cerebro de niños entre 2 y 5 años puede ser un indicador para diagnosticar autismo. Algunos médicos investigadores suponen que la circunferencia de la cabeza de niños entre 6 y 14 meses podría ser una forma para estimar el tamaño de la materia gris cerebral que podrá llegar a tener un niño entre los 2 y los 5 años. Los datos sobre la circunferencia cefálica (medida entre los 6 y los 14 meses) y la materia gris cerebral (medida de los 2 a los 5 años) de 18 niños varones con autismo son usados para explorar la relación entre estas dos variables.



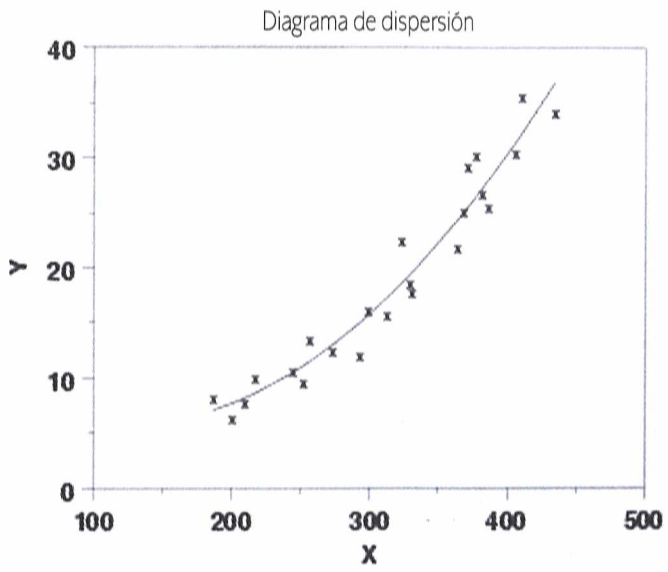
Estás trabajando para correlacionar las variables involucradas entre fenómenos naturales y procesos sociales de tu contexto mediante el uso de las TIC para obtener información y procesarla a través de una hoja de cálculo.

Los médicos investigadores se plantean preguntas como las siguientes: ¿existe una relación entre la circunferencia de la cabeza a la edad de 6 a 14 meses y la materia gris cerebral medida a la edad de 2 a 5 años? Si así es, ¿servirá la medida de la circunferencia de la cabeza tomada a una edad temprana para predecir la medición de la materia gris, permitiendo detectar el autismo a edades más jóvenes?, ¿qué tan exactas son estas predicciones de la materia gris?

Podrás responder las preguntas anteriores con el estudio de los métodos que complementan las gráficas de dispersión para describir las relaciones entre dos variables numéricas y que también permiten evaluar la fortaleza de esta relación. Comencemos por la correlación.

Correlación

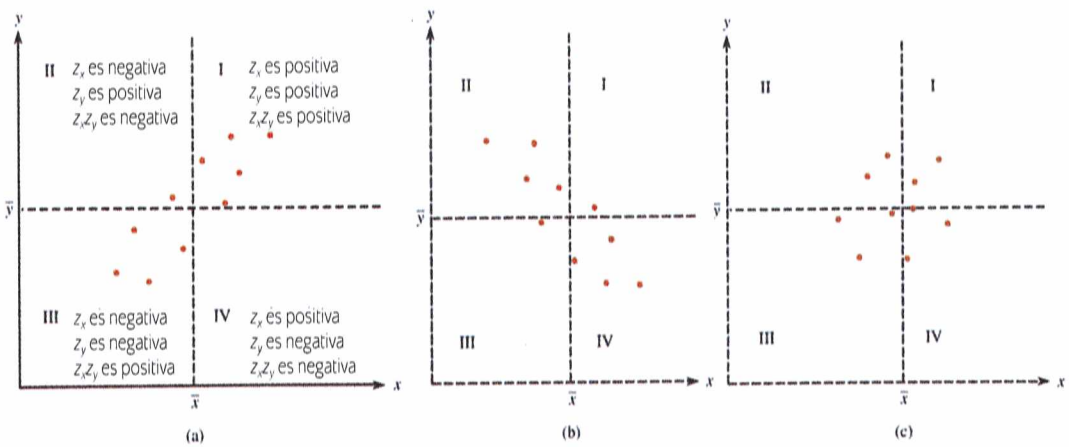
Como recordarás por lo estudiado en la unidad anterior, las gráficas de dispersión de dos variables son instrumentos con los que se obtiene una imagen de la tendencia con la que los valores de las observaciones pareadas (x, y) se relacionan. Sin embargo, una imagen no es suficiente y para argumentar de forma precisa y sacar conclusiones, hay que usar el **coeficiente de correlación muestral**, cuyo valor nos da una evaluación numérica de la tendencia y de la fuerza de la relación entre los valores x y y en un conjunto de datos con observaciones pareadas de la forma (x, y). El coeficiente de correlación muestral más utilizado es el de Pearson.



Coeficiente de correlación muestral de Pearson

El coeficiente muestral de Pearson mide la fuerza de una relación lineal entre dos variables numéricas mediante el uso de las puntuaciones z. Lo hace al reemplazar cada valor de x por su correspondiente puntuación z y z_x (obtenido de restar $\bar{x}$ y x y dividir este resultado por S_x, la desviación estándar de los valores de x); de manera similar, se reemplaza cada valor y por su puntaje z_y. Se debe tomar en cuenta que los valores de x mayores que generan puntuaciones z_x positivas y los menores un efecto negativo en las puntuaciones z_x. También los valores y más grandes que $\bar{y}$ generan puntuaciones z_y positivas y los más pequeños dan como resultado puntuaciones z_y negativas. El coeficiente de correlación muestral de Pearson se basa en la suma de los productos de z_x y z_y para cada observación en el conjunto de observaciones pareadas. Su notación algebraica es $\sum z_x z_y$..

Veamos algunas gráficas de dispersión para comprender con mayor profundidad el coeficiente descrito. La gráfica (a) presenta una marcada tendencia creciente y por lo tanto una fuerte relación positiva. Si se traza una línea vertical que pase por $\bar{x}$ y una horizontal a través de la media $\bar{y}$ se divide el primer cuadrante del plano cartesiano en cuatro regiones. En la región I, los valores graficados de x y y superan sus valores medios, por lo que la puntuación z_x de x y el puntaje z_y de y son números positivos. De ello resulta que $z_x z_y$ son positivos. El producto de las puntuaciones z también es positivo para cualquier punto de la región III ya que tanto la puntuación z_x como de z_y es negativa en la Región III y la multiplicación de dos números negativos da uno positivo. En cada una de las otras dos regiones, una puntuación z es positiva y la otra negativa, por lo que $z_x z_y$ es negativo. Pero como en este caso, por lo general, los puntos caen en las regiones I y III, los productos de las puntuaciones z tienden a ser positivos. Por lo tanto, la suma de los productos será un número positivo relativamente grande. Aplicando un razonamiento similar para los datos mostrados en la gráfica (b), donde se exhibe una fuerte relación negativa, implica que $\sum z_x z_y$. será un número negativo relativamente grande (en magnitud). Cuando no hay una relación fuerte, como se muestra en la gráfica (c), los productos positivos y negativos tienden a contrarrestarse el uno con el otro, generando un valor de $\sum z_x z_y$. que es cercana a cero. En resumen, $\sum z_x z_y$. parece ser una medida razonable para encontrar el grado de asociación entre x y y, el cual puede ser un número positivo



grande, un número negativo grande, o un número cercano a 0, dependiendo de si la relación lineal positiva o negativa es fuerte, o no existe una fuerte relación lineal.

El **coeficiente de correlación muestral de Pearson**, al que denotaremos mediante r, se obtiene dividiendo $\sum z_x z_y$. por (n - 1).

El coeficiente de correlación muestral de Pearson está dado por

$$\frac{\sum z_x z_y}{n-1}$$

Aunque hay varios coeficientes de correlación, el de Pearson es el más utilizado, por lo que el nombre de "Pearson" se omite con frecuencia y simplemente se denominada como el **coeficiente de correlación**.

Gestión del aprendizaje

Es posible realizar el cálculo del coeficiente de correlación muestral en una hoja de cálculo electrónica de manera sencilla.

Basta seguir los pases descritos en el Apéndice 2.

Calcular el coeficiente de correlación es bastante laborioso pero hoy la mayoría de los paquetes estadísticos y las calculadoras científicas pueden calcular r una vez introducidos los valores x y y.

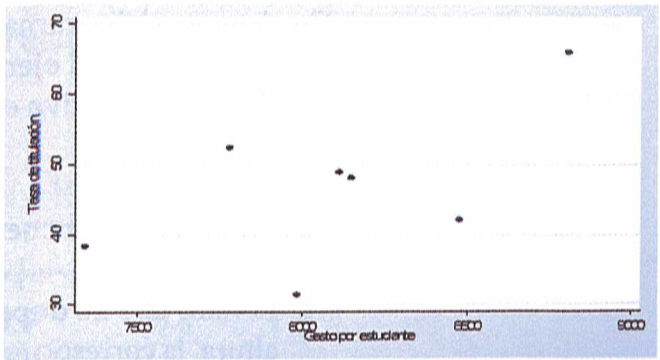
Para adentrarnos más en el estudio del coeficiente de correlación veamos un ejemplo:

Ejemplo

Una preocupación permanente de las universidades públicas en países como México es la obtención de recursos para su financiamiento. En cierto estado del país existen siete universidades públicas con una matrícula registrada en conjunto de entre 10,000 y 20,000 estudiantes. Las tasas de titulación, así como el costo en pesos invertido por estudiante durante el último año en cada una de las siete universidades, se registraron de la siguiente manera:

Universidad	Tasa de titulación (%)	Gasto por estudiante ($)
1	66.1	8,810
2	52.4	7,780
3	48.9	8,112
4	48.1	8,149
5	42.0	8,477
6	38.3	7,342
7	31.3	7,984

La siguiente es la gráfica de dispersión de estos datos.



Si x representa el gasto por estudiante e y denota la tasa de graduación por año. Es fácil comprobar que

$$\bar{x} = 8093.43 \quad S_x = 472.39 \quad \bar{y} = 46.73 \quad S_y = 11.15$$

Para calcular el coeficiente de correlación se empieza por el cálculo de las puntuaciones z para cada par (x, y) en el conjunto de datos. Por ejemplo, la primera observación es (8810, 66.1). Los correspondientes valores z son

$$z_x = \frac{8810 - 8093.43}{472.39} = 1.52 \quad z_y = \frac{61.1 - 46.73}{11.15} = 1.74$$

La siguiente tabla muestra las puntuaciones z, el producto $z_x z_y$ para cada observación, así como la suma total de los productos $z_x z_y \left(\sum z_x z_y\right)$:

y	x	zx	zy	zx zy
66.1	8,810	1.52	1.74	2.64
52.4	7,780	-0.66	0.51	-0.34
48.9	8,112	0.04	0.20	0.01
48.1	8,149	0.12	0.13	0.01
42.0	8,477	0.81	-0.42	-0.34
38.3	7,342	-1.59	-0.75	1.20
31.3	7,984	-0.23	-1.38	0.32
				$\sum z_x z_y = 3.52$

Por lo que ahora con la información anterior es posible calcular el valor del coeficiente de correlación r:

$$\sum z_x z_y = 3.52$$

Entonces con n = 7

$$r = \frac{\sum z_x z_y}{n-1} = \frac{3.52}{6} = .587$$

Con base en el diagrama de dispersión y las propiedades del coeficiente de correlación ilustradas en el ejemplo anterior, es posible concluir que existe una moderada relación lineal positiva entre el gasto por estudiante y la tasa de titulación en estas siete universidades.

Las propiedades del coeficiente de correlación (r) son las siguientes:

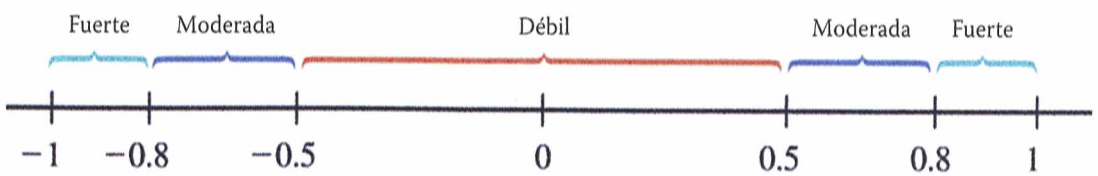
- El valor de r no depende de la unidad de medida de cada variable; así, si x es la altura, la correspondiente puntuación z es la misma; es decir, que si la altura se

expresa en centímetros, metros o kilómetros el valor del coeficiente de correlación no se afecta.

- El valor de r no depende de cuál de las dos variables se considera x. Por lo tanto, si en el ejemplo anterior tuviéramos que x = tasa de graduación e y = gastos por estudiante, se habría obtenido el mismo valor, r =.587.
- El valor de r está entre -1 y 1. Un valor cercano al límite superior 1, indica una fuerte relación positiva. En términos de la relación entre la variable x y la variable y se traduce en que al aumentar el valor de la variable x existe una fuerte tendencia lineal a que aumente el valor de la variable y. Mientras un valor de r cercano al límite inferior, -1, sugiere una fuerte relación negativa, lo que indica que al aumentar el valor de la variable x existe una fuerte tendencia lineal a que disminuya el valor de la variable y.
- Un coeficiente de correlación de r =1 sólo se produce cuando todos los puntos en una gráfica de dispersión se encuentran exactamente en una línea recta que tiene pendiente ascendente. Del mismo modo, r =-1 sólo se da cuando todos los puntos están exactamente en una línea descendente. Únicamente cuando hay una relación lineal perfecta entre x e y en la muestra, r toma alguno de sus dos valores extremos posibles. Encontrar estos niveles de relación entre dos variables es muy raro.
- El valor de r es una medida del grado en que x e y están relacionadas linealmente; es decir, el grado en que los puntos de una dispersión está cerca de una línea recta. Un valor cercano a 0 no descarta una relación fuerte entre x e y, pues podría haber una fuerte relación que no es lineal entre x e y. De hecho la imposibilidad de detectar relaciones no lineales entre dos variables constituye un seria desventaja de esta medida de asociación.

La siguiente figura muestra los tres niveles de asociación en que se clasifican los posibles valores que puede tomar el coeficiente de correlación:

- Correlación débil si el coeficiente de correlación toma valores entre -0.5 y 0.5.
- Correlación moderada si sus valores están entre 0.5 y 0.8 o entre -0.5 y -0.8.
- Correlación fuerte si los valores del coeficiente se encuentran entre 0.8 y 1 o entre -0.8 y -1.



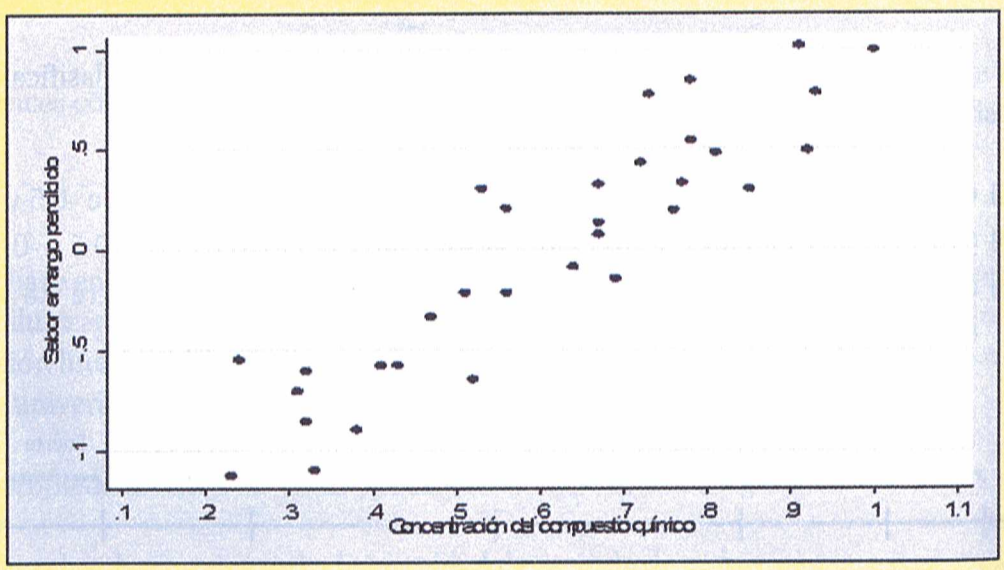
Estudiemos el siguiente ejemplo:

Ejemplo

Un laboratorio encargado de surtir medicamentos a un hospital realiza un estudio para determinar si existe una relación entre el sabor amargo de un jarabe pediátrico contra la tos y la concentración de un compuesto químico que se utiliza en su elaboración. A continuación se proporcionan los datos de x = concentración del compuesto químico y de y = el sabor amargo percibido, determinados por un panel de catadores del laboratorio.

x	y	x	y	x	y
0.72	0.43	0.76	0.19	0.52	-0.65
0.81	0.48	0.67	0.07	0.69	-0.15
0.92	0.49	0.56	-0.22	0.91	1.01
1.00	0.99	0.38	-0.90	0.64	-0.09
0.67	0.32	0.78	0.84	0.23	-1.13
0.53	0.30	0.67	0.13	0.78	0.54
0.51	-0.22	0.85	0.30	0.33	-1.10
0.56	0.20	0.41	-0.58	0.43	-0.58
0.77	0.33	0.93	0.78	0.32	-0.86
0.47	-0.34	0.31	-0.71	0.24	-0.55
0.73	0.77	0.32	-0.61		

El coeficiente de correlación $r = 0.916$ indica una fuerte relación positiva entre la concentración del compuesto químico y el sabor amargo percibido. Esto indica que conforme se aumenta la concentración del compuesto químico, la percepción del sabor amargo del jarabe tiende a aumentar. Se da la siguiente gráfica de dispersión para este conjunto de datos:



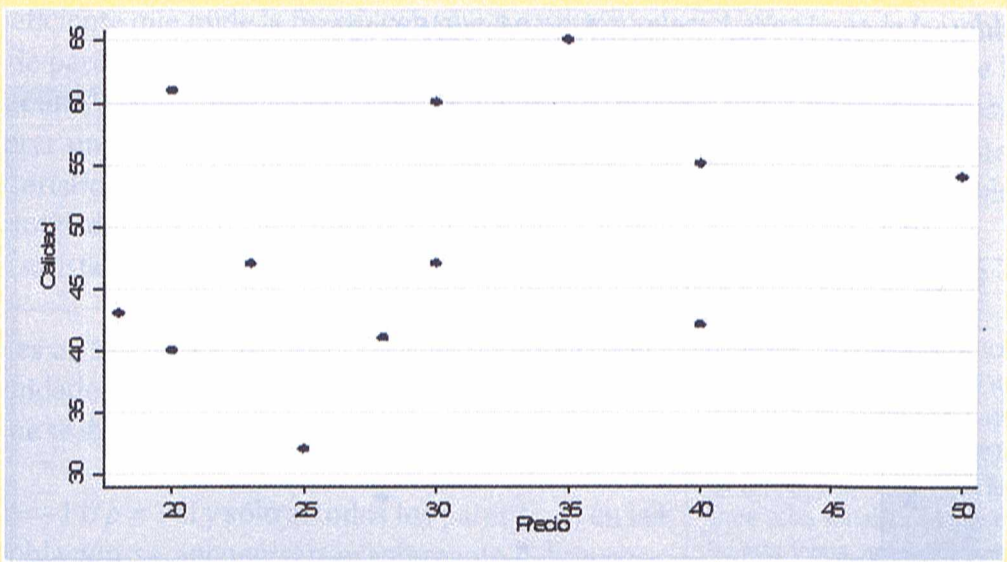
Trabajemos un ejemplo más.

Ejemplo

¿Los clavos para cirugía ortopédica más caros son más seguros que los más baratos? Los datos mostrados proporcionan la variable x = precio y la de y = calificación de calidad para 12 marcas diferentes de clavos ortopédicos para cirugía. La valoración de la calidad es un número entre 0 (la peor calificación posible) y 100 y está basada en su funcionalidad, el impacto y la resistencia después de implantados.

Precio	Calidad
35	65
20	61
30	60
40	55
50	54
23	47
30	47
18	43
40	42
28	41
20	40
25	32

La siguiente es la gráfica de dispersión para este conjunto de datos.

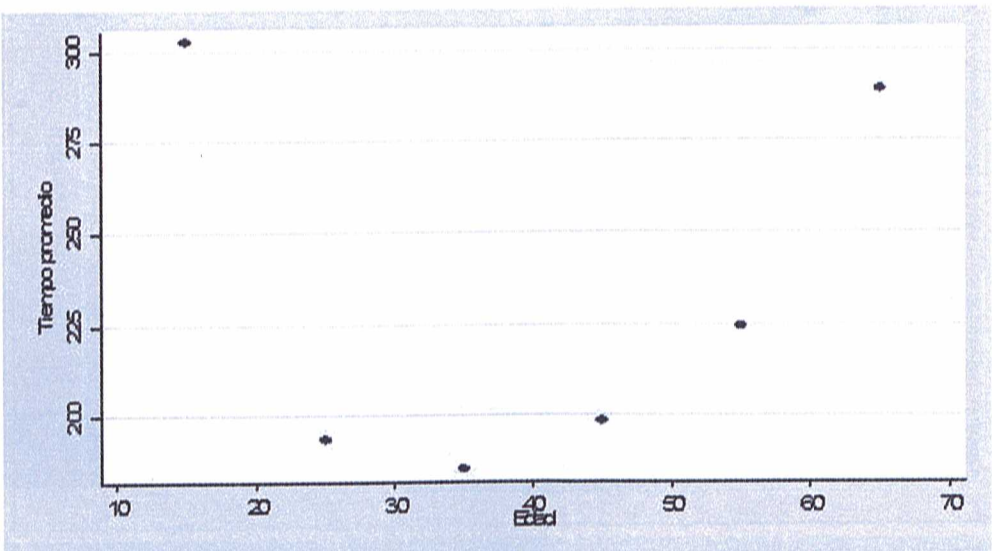


El análisis del diagrama de dispersión permite observar la débil relación positiva entre el precio y nivel de calidad. El coeficiente de correlación, obtenido por medio de una hoja electrónica de cálculo, es de 0.303. Un coeficiente de correlación de $r = 0.303$ confirma que existe una tendencia a que las mayores clasificaciones de calidad estén asociadas a un mayor precio de los clavos. Sin embargo, como el valor del coeficiente de correlación calculado cae en el rango de valores comprendido entre -0.5 y 0.5, se puede concluir que la relación lineal entre estas dos variables es débil. Resultado que concuerda totalmente con lo representado en el diagrama de dispersión.

Un grupo de médicos del deporte, contratado para llevar el seguimiento médico de los participantes en una carrera de 20 km, está interesado en medir la relación que pueden tener el grupo de edad de las mujeres participantes, con el tiempo en minutos que tardan en cruzar la meta. Para ello se toma la medición del tiempo promedio que le tomó a cada grupo de edad llegar a la meta. Se muestran en seguida las mediciones de tiempo promedio en minutos y grupo de edad para las mujeres participantes.

Grupo de edad	Edad representativa	Tiempo promedio
10–19	15	302.38
20–29	25	193.63
30–39	35	185.46
40–49	45	198.49
50–59	55	224.30
60–69	65	288.71

La siguiente gráfica muestra la dispersión del tiempo promedio de recorrido (en minutos) *versus* la edad representativa de cada grupo.



El coeficiente de correlación de Pearson entre la edad representativa y el tiempo promedio de recorrido fue $r = -0.038$. Cantidad que bajo los parámetros establecidos para el coeficiente de correlación hace patente la débil asociación lineal entre las variables.

Este ejemplo muestra que el coeficiente de correlación r es poco eficiente para detectar la relación entre dos variables cuando ésta no se comporta de manera lineal. Para el caso anterior, el valor calculado de r es cercano a cero, lo que indica la presencia de una relación lineal muy débil entre las variables. Sin embargo, al inspeccionar la gráfica de dispersión correspondiente, es fácil observar que existe evidencia de una fuerte relación no lineal entre la edad representativa de cada grupo y el tiempo promedio de recorrido; relación que no pudo detectar el coeficiente de correlación calculado. Éste es un punto importante a destacar, ya que nos lleva a concluir que no basta con el hecho de obtener un valor de pequeño r, para concluir la ausencia de relación entre dos variables. Por lo tanto, siempre es importante estudiar la gráfica de dispersión de los datos antes de concluir, con base en un coeficiente de correlación cuyo valor es cercano a cero, que no existe relación entre dos variables.

El coeficiente de correlación de la población

El coeficiente de correlación muestral r —que mide la fuerza de la relación de los valores x y los y en una muestra de parejas ordenadas que están linealmente relacionadas entre sí— no es una medida análoga de la fuerza con que x e y están relacionadas en toda la población de parejas ordenadas de la cual se obtuvo la muestra. Al coeficiente que mide la fuerza con que x e y están relacionadas en toda la población de parejas ordenadas se le conoce como el **coeficiente de correlación de la población** y se define como ρ (nótese de nuevo el uso de una letra griega para nombrar una característica de la población y una letra romana para designar una característica de la muestra). Aunque nunca tendrás que calcular ρ de una población total de pares ordenados, es importante saber que ρ satisface propiedades paralelas a las de r:

1. ρ es un número entre -1 y 1, que no depende de la unidades de medición de x e y, ni de cuál variable que se denomine como x y cuál se denomine como y.
2. $\rho = -1$ o $\rho = 1$ si y sólo si todos los pares (x, y) en la población se encuentran exactamente sobre una línea recta, por lo que mide el grado en que existe una relación lineal entre la población.

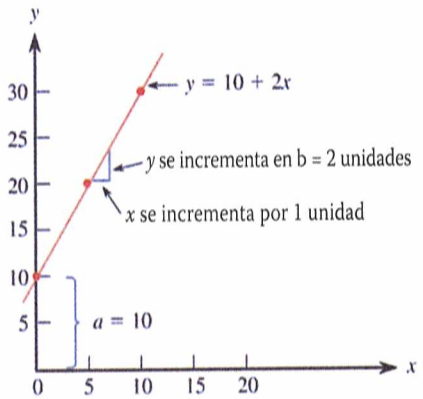
Gestión del aprendizaje

Es posible encontrar un coeficiente de correlación elevado entre dos variables que no tienen relación alguna, es decir, variables que no presentan relación justificada a través de alguna teoría específica (biología, astronomía, economía, entre otras). Cuando sucede esto, se dice que la correlación estadística existente entre estas variables es una correlación espúrea o sin sentido. De esta forma, es posible hablar de correlación espúrea entre variables relacionadas a la economía, a la biología, a la astronomía, entre otras.

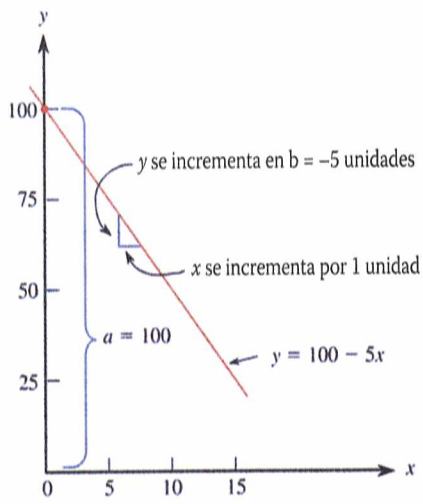
Regresión lineal: ajuste de una recta a los datos bivariados

El objetivo del análisis de regresión es el uso de la información sobre una variable x, con la finalidad de obtener alguna conclusión acerca de una segunda variable, y. Por ejemplo, podríamos tratar de predecir y = la proporción de enfermos de un mal respiratorio, cuando la temperatura ambiente promedio se encuentra en x =2°C. Las dos variables en un análisis de regresión desempeñan funciones diferentes: se denomina a la **variable y dependiente** o respuesta; y a la variable x **variable independiente, predictora o explicativa**. Cuando se observa que una gráfica de dispersión presenta un patrón lineal tiene sentido analizar la relación entre las variables mediante la búsqueda de una línea que esté lo más cerca posible de los puntos descritos en la gráfica de dispersión.

Antes estudiar cómo se hace esto, haremos un breve repaso sobre las características elementales de las lineales y las relaciones lineales.



(a)



(b)

Características de las lineales y las relaciones lineales

La ecuación de una recta es $y = a + bx$. Una línea en particular se determina eligiendo los valores de a y b. Por ejemplo, una recta es $y = 10 + 2x$; y otra es $y = 100 - 5x$. Si optamos por dar algunos valores a la variable x y calcular $y = a + bx$ para cada uno de estos valores, las parejas ordenadas resultantes (x, y), pertenecen a una línea recta.

En la ecuación de una línea recta

$$y = a + bx$$

A la cantidad b, se le llama la pendiente la línea recta. Representa la cantidad en que se incrementa la variable y cuando x se incrementa en 1 unidad. El valor a se conoce como la intersección de la línea recta con el eje y y representa la altura que toma la línea por encima del valor $x = 0$.

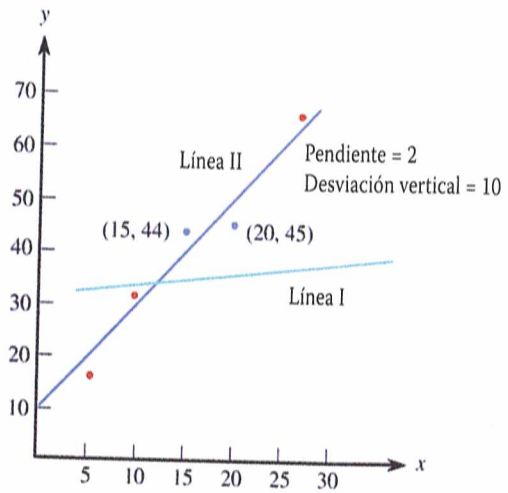
La recta $y = 10 + 2x$ tiene pendiente $b = 2$ por lo que cada incremento de 1 unidad en x está asociado con un aumento de 2 unidades en y. Cuando $x = 0$, $y = 10$, por lo que la altura a la que esta línea cruza el eje vertical (cuando $x = 0$) es 10. Esto se ilustra en la gráfica (a). La pendiente de la línea $y = 100 - 5x$ es igual a -5, así que y decrece en 5 cuando x se incrementa en 1. La altura de la línea cuando $x = 0$, se encuentra en 100. La línea de resultante se representa en la gráfica (b).

Es fácil trazar la línea que corresponde a la ecuación de una línea recta en particular. Seleccionamos dos valores de x, cualesquiera, y los sustituimos

en la ecuación para obtener los valores correspondientes de la variable y. Si representamos los valores resultantes como parejas de puntos (x, y), la línea deseada es la que pasa a través de estas dos parejas de puntos. Para la ecuación $y = 10 + 2x$, si sustituimos los valores $x = 5$ y $x = 10$ obtenemos los valores para la variable $y = 20$ e $y = 30$, respectivamente. Así, las parejas ordenadas resultantes son (5, 20) y (10, 30), por lo que la línea recta representada por esta ecuación es la que pasa por estos dos puntos.

El ajuste de una línea recta: principio de mínimos cuadrados

En la figura siguiente se muestra una gráfica de dispersión con dos líneas superpuestas entre la nube de puntos. La línea II se ajusta mejor a los datos que la línea I. Para medir el grado en que una línea particular proporciona un buen ajuste a los datos, nos centramos en las desviaciones verticales de la línea. Por ejemplo, la línea II tiene como ecuación $y = 10 + 2x$ y los puntos tercero y cuarto a la izquierda en el diagrama de dispersión son (15, 44) y (20, 45). Para estos dos puntos, las desviaciones verticales de esta línea



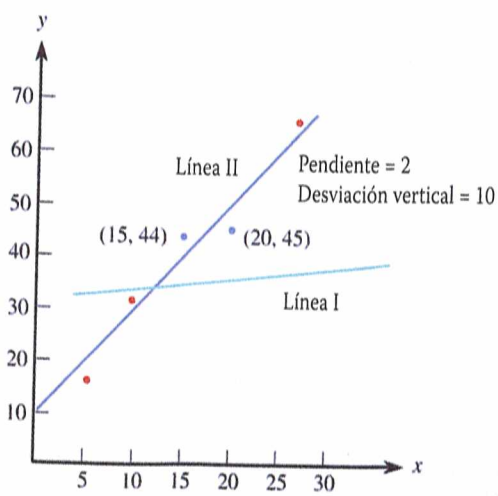
son

$$\begin{aligned} \textit{Tercera desviación} &= y_3 - \textit{altura de la línea por encima de } x_3 \\ &= 44 - [10 + 2(15)] \\ &= 4 \end{aligned}$$

y

$$\begin{aligned} \textit{Cuarta desviación} &= 45 - [10 + 2(20)] \\ &= -5 \end{aligned}$$

Una desviación vertical positiva corresponde a un punto situado por encima de la línea elegida; se produce una desviación vertical negativa si el punto correspondiente está situado por debajo de dicha línea. Una línea en particular se dice que es un buen ajuste de los datos, si las desviaciones de la línea son de pequeña magnitud. La línea I de la gráfica que se muestra a continuación se adapta mal, porque todas las desviaciones de la línea son más grandes que las de la línea II.



Para evaluar el ajuste global de una línea (bondad de ajuste) en un conjunto de n mediciones, necesitamos una manera de combinar las n desviaciones en una sola medida de ajuste. El método estándar consiste en tomar los cuadrados de las desvia-

ciones (para obtener un número no negativo) y luego considerar la suma de los cuadrados de estas desviaciones.

La medida más utilizada de la bondad del ajuste de una línea a un conjunto de datos bivariados

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

es la suma de los cuadrados de las desviaciones alrededor de la línea

$$\sum [y - (a + bx)]^2 = [y_1 - (a + bx_1)]^2 + [y_2 - (a + bx_2)]^2 + \cdots + [y_n - (a + bx_n)]^2$$

La línea de mínimos cuadrados, también llamada la línea de regresión muestral, es la que minimiza la suma de los cuadrados de las desviaciones.

La ecuación de la recta de mínimos cuadrados se puede obtener sin necesidad de calcular las desviaciones de cualquier línea en particular. En el siguiente cuadro están las fórmulas relativamente simples para la pendiente y la intersección de la línea de mínimos cuadrados.

La pendiente de la línea de mínimos cuadrados es

$$= \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

y la intersección es

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

Escribimos la ecuación de mínimos cuadrados como la línea

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

donde la ^ arriba de y (que se lee y gorro) indica que $\hat{y}$ es la predicción de y que resulta de sustituir un valor particular de x en la ecuación.

Gestión del aprendizaje

Es posible realizar el cálculo de los coeficientes de una línea de regresión en una hoja de cálculo, de maneara sencilla. Basta con que sigas los pasos que para realizarlo están descritos en el Apéndice 2.

Varios paquetes estadísticos y calculadoras pueden ser útiles para calcular la pendiente y la intersección de la línea de mínimos cuadrados. Si la pendiente y la intersección se van a calcular a mano, la siguiente fórmula de cálculo sirve para reducir el tiempo para realizar los cálculos.

Fórmula para calcular la pendiente de la línea de mínimos cuadrados

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

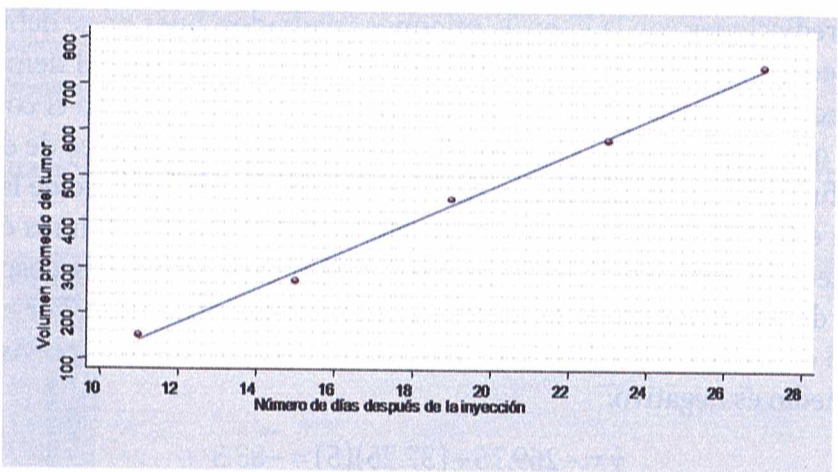
Vayamos al ejemplo:

La granada, fruta originaria de Persia, se ha utilizado en la medicina popular de muchas culturas para el tratamiento de diversas dolencias. Los investigadores están ahora estudiando las propiedades antioxidantes de la granada para estimar su efecto benéfico en el tratamiento del cáncer. El tema de investigación de uno de los estudios es si el extracto de fruta de la granada (EFG) fue eficaz en la ralentización o el hacer más lento el crecimiento de los tumores del cáncer de próstata. En este estudio, 24 ratones fueron inyectados con células cancerígenas. Los ratones se asignaron aleatoriamente a uno de tres grupos de tratamiento. Un grupo de ocho ratones tomó agua potable normal, el segundo grupo de ocho ratones bebió agua potable complementada con 0.1% de EFG y el tercer grupo recibió agua potable complementada con 0.2% de EFG. El volumen tumoral promedio de los ratones en cada grupo se registró en varios puntos en el tiempo.

Los datos para las variables y = el volumen promedio de tumor (en mm^3) y x = número de días después de la inyección de las células cancerosas en los ratones que recibieron el agua potable normal se presentan a continuación:

x	11	15	19	23	27
y	150	270	450	580	740

La siguiente gráfica de dispersión de estos datos muestra que la relación entre el número de días después de la inyección de las células cancerosas y el volumen tumoral promedio podría, razonablemente, resumirse mediante una línea recta.



Las cantidades necesarias para calcular la ecuación de mínimos cuadrados correspondiente a este conjunto de datos son:

$$\sum x = 95 \quad \sum x^2 = 1965 \quad \sum xy = 47570$$
$$\sum y = 2190 \quad \sum y^2 = 1181900$$

A partir de estas ecuaciones calculamos

$$\bar{x} = 19 \quad \bar{y} = 438$$

$$b = \frac{\sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} = \frac{47570 - \frac{(95)(2190)}{5}}{1965 - \frac{(95)^2}{5}} = \frac{5960}{160} = 37.25$$

y

$$a = \bar{y} - b\bar{x} = 438 - (37.25)(19) = -269.75$$

La línea de mínimos cuadrados es entonces

$$\hat{y} = -269.75 + 37.25x$$

Si queremos predecir el volumen promedio del tumor 20 días después de la inyección de las células cancerosas, usamos el valor de y del punto en la línea de mínimos cuadrados por encima de $x = 20$:

$$\hat{y} = -269.75 + (37.25)(20) = 475.25$$

El volumen promedio del tumor predicho para otro número de días después de la inyección de células cancerosas se calcula de manera similar. Debe tenerse cuidado al hacer predicciones con la línea de mínimos cuadrados, pues no se debe emplear para predecir el volumen promedio de un tumor para instantes del tiempo demasiado alejados del rango de 11 a 27 días (el rango de valores de x en el conjunto de datos), ya que no sabemos si el patrón lineal observado en la gráfica de dispersión continúa fuera de este rango. Esto a veces se conoce como el peligro de la extrapolación. En este ejemplo podemos ver que el uso de la línea de mínimos cuadrados para predecir el volumen tumoral promedio de menos de 10 días después de la inyección de las células cancerosas, lleva a predicciones sin sentido. Por ejemplo, si el número de días después de la inyección es de cinco, el volumen previsto del tumor promedio es negativo:

$$\hat{y} = -269.75 + (37.25)(5) = -83.5$$

Debido a que es imposible que el volumen del tumor promedio sea negativo, esto es una clara indicación de que el patrón observado para los valores de x en el rango de 11 a 27 no se continúa fuera de este rango. Sin embargo, la línea de mínimos cuadrados constituye una herramienta útil para hacer predicciones para valores de x dentro del rango de días 11 a 27.

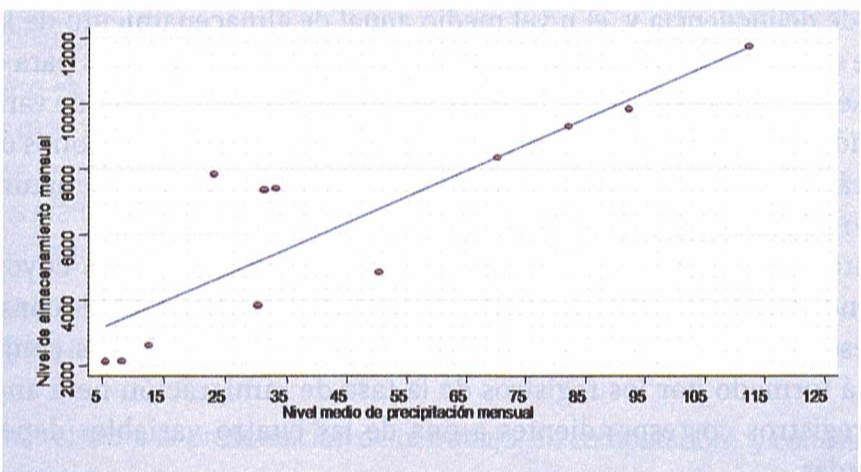
Veamos un ejemplo más.



El nivel de la presa de Valle de Bravo fluctúa durante el año dependiendo de la precipitación pluvial. Por lo general, su nivel aumenta a partir de septiembre, con la acumulación de escurrimientos de lluvia de las laderas; llega a su nivel más alto a finales de octubre y empieza a bajar en marzo, alcanzando su nivel más bajo en junio, cuando reinicia la época de lluvia. El lago se encuentra, entonces, en niveles bajos durante el verano. Los datos mostrados a continuación presentan el nivel medio de precipitación pluvial mensual (medida en mm^3) así como los correspondientes niveles de almacenamiento mensual de agua, medidos en hm^3 (1 hm^3 = 1000000 m^3).

Mes del año	Nivel medio de precipitación pluvial mensual	Niveles de almacenamiento mensual
Enero	70.3	8,329
Febrero	23.2	7,832
Marzo	33.2	7,392
Abril	50.4	4,835
Mayo	30.4	3,832
Junio	5.6	2,111
Julio	8.3	2,132
Agosto	12.7	2,604
Septiembre	31.4	7,342
Octubre	112.33	11,746
Noviembre	92.3	9,828
Diciembre	82.2	9,301

A continuación observamos la gráfica de dispersión junto con su línea de regresión ajustada.



El estudio de la gráfica posibilita observar una tendencia creciente en el comportamiento de los niveles de almacenamiento de la presa y los niveles promedio mensuales de la precipitación pluvial. Esta evidencia visual se confirma al calcular el coeficiente de correlación muestral, que para estos dados es de . Cantidad que confirma que existe una fuerte relación lineal entre las variables.

La línea de mínimos cuadrados calculada para estos datos, por medio de una hoja electrónica de cálculo, queda representada por la ecuación

$$\hat{y} = -2762.95 + 79.8x$$

La pendiente de esta línea recta es positiva. Lo anterior revela que existe una relación lineal creciente entre el nivel medio de precipitación pluvial mensual y el nivel de almacenamiento mensual de la presa.

¡Ahora te toca a ti!

La aplicación del coeficiente de correlación proporciona información adicional que, junto con la evidencia visual, permite aceptar o desechar la relación causa-efecto planteada entre dos variables. Con su estudio tienes herramientas complementarias a la gráfica de dispersión para aceptar o refutar las hipótesis planteadas. Así como con medidas para tener indicios sobre la posible relación entre dichas variables. Estás listo ya para aplicar lo aprendido. En la siguiente actividad identificarás variables relacionadas con el fenómeno migratorio para responder cuál es el impacto que tiene en la comunidad de El Encino.

El criterio para la selección de las variables está basado en la identificación de variables cuyo comportamiento pudiera afectarse gracias a la intensidad de la inmigración de personas hacia la comunidad de El Encino (intensidad que se mide a partir de la tasa de inmigración neta anual).

El producto interno bruto (PIB) de la comunidad, el índice de pobreza humana, el índice de delincuencia y el nivel medio anual de almacenamiento de la represa que surte de agua a la comunidad serán las variables dependientes para este análisis. Tal selección se explica ya que la naturaleza de las magnitudes de variación en la cantidad de población inmigrante podrían afectarlas, permitiéndonos utilizarlas como parámetros para cuantificar los impactos económico, social y natural que el fenómeno migratorio producen en El Encino.

Cuentas con cuatro conjuntos de 50 observaciones pareadas cuyos valores corresponden a los registros de los últimos 50 años de las variables consideradas (consúltese la base de datos del Apéndice 3). Cada uno de estos conjuntos de datos está formado por los registros de la tasa de inmigración neta anual junto con los registros correspondientes a una de las cuatro variables dependientes consideradas.

Como se ha mencionado en las actividades anteriores, la propuesta anterior puede ser tu guía para el desarrollo de una actividad alterna. Si optas por realizar ésta será necesario que plantees las hipótesis, identifiques las variables dependientes e independientes así como las fuentes de información de las cuales obtendrás las mediciones. Tales acciones te permitirán calcular las correlaciones para solucionar el problema de estudio planteado.



Con base en uno o más conjuntos de observaciones pareadas —correspondientes a las variables por cuya naturaleza planteaste la hipótesis para probar el nivel de asociación— realiza los diagramas de dispersión necesarios y establece si hay o no relación entre las variables estudiadas.

Realizados los diagramas, aplica el cálculo del coeficiente de correlación muestral y el ajuste de una línea recta por el método de mínimos cuadrados al conjunto o conjuntos de observaciones pareadas. A partir del análisis de los resultados, formula conclusiones con base en la relación de dependencia identificada entre estas variables.

Es aconsejable que te auxilies con computadora con procesador de textos, hoja de cálculo electrónica y programa elaborar presentaciones. Para el desarrollo de la actividad que requerirás 10 horas de trabajo.

Con los datos considerados para esta actividad (véase el Apéndice 3) desarrolla lo que se te pide a continuación:

1. Para cada uno de los cuatro conjuntos de observaciones pareadas elabora una gráfica de dispersión donde identifiques la tasa de inmigración neta anual con el eje *x*, y una de las cuatro variables dependientes consideradas con el eje *y*.
2. Analiza las gráficas de dispersión elaboradas y estudia si hay alguna tendencia que pueda llevar a suponer que las variables consideradas tienen algún grado de asociación. De existir está asociación, índica en qué sentido se presenta.
3. Calcula el coeficiente de correlación y la línea de regresión de mínimos cuadrados para cada conjunto de datos.
4. Con base en los valores obtenidos para el coeficiente de correlación responde: ¿estos valores indican algún tipo de asociación entre las variables?, ¿el tipo de asociación detectada por los coeficientes de correlación coincide con la tendencia observada en sus respectivas gráficas de dispersión?, ¿qué conclusiones obtienes acerca del impacto del fenómeno migratorio en la comunidad de El Encino?
5. Utiliza la línea de regresión de mínimos cuadrados para pronosticar el producto interno bruto anual para dos valores distintos de la tasa de inmigración neta anual.

Para terminar elabora la presentación de tu trabajo, en computadora, y con las siguientes características:

Como inicio elabora una diapositiva en la que describas las fórmulas para el cálculo del coeficiente de correlación y la recta de mínimos cuadrados. En otra diapositiva expón la hipótesis y los argumentos que te permiten sustentar la posible naturaleza de la relación entre tasa de migración anual y el producto interno bruto.

Desarrollo. Elabora las diapositivas necesarias describir los pasos que realizaste para obtener los cálculos del coeficiente de correlación así como de los pronósticos del producto interno bruto para las dos tasas de marginación consideradas. Documenta la gráfica de dispersión que realizaste durante el desarrollo de la actividad.

Cierre. Expresa las conclusiones obtenidas a partir del análisis de la gráfica de dispersión y del coeficiente de correlación. Responde si la gráfica de dispersión y el coeficiente de correlación muestran evidencia de exista algún tipo de relación de causa-efecto entre la tasa de migración y el producto interno bruto. Para terminar concluye si el fenómeno migratorio tiene o no un efecto económico en la comunidad de El Encino.

En la siguiente tabla marca con una cruz (✗) cuál fue tu nivel de cumplimiento de la actividad y de ser necesario señala que aspecto habrías de trabajar para lograr un mejor desempeño.

Actividad a evaluar	Nivel de cumplimiento en caso de haber realizado la actividad			En caso de ser necesario indica que aspectos de la actividad debes mejorar
	Bueno	Regular	Malo	
Realicé la gráfica de dispersión de las variables consideradas.				
Calculé el valor del coeficiente de correlación y ajuste la línea recta de mínimos cuadrados.				
Calculé los valores pronosticados para la variable y correspondientes a los dos valores de la variable explicativa *x*.				
Interpreté los resultados obtenidos para obtener conclusiones acerca de la situación problemática planteada.				

La aplicación del modelo de regresión y el coeficiente de correlación en el estudio de fenómenos o eventos sobre el entorno permite identificar las causas o factores que originan su ocurrencia. Su identificación hace posible una mejor comprensión de su naturaleza; su análisis, la obtención de beneficios como resultado de su ocurrencia así

como la implementación de medidas preventivas que mitiguen posibles consecuencias negativas de su realización. Ejemplos de la aplicación del modelo de regresión y el coeficiente de correlación hay muchos. Mediante tales modelos, un economista puede determinar las causas que originan la inflación y así aplicar las medidas económicas necesarias para tenerla bajo control y disminuir el impacto de sus efectos negativos. Un agrónomo puede conocer los factores climáticos que inciden en la producción de un cierto tipo de grano e identificar cuál es la mejor época del año para realizar su cultivo, maximizando la producción. Es en este tipo de situaciones en las que toma sentido la aplicación del modelo de regresión y el coeficiente de correlación lineal, pues por sus características sirven como medidas objetivas que permiten identificar las causas relacionadas con la ocurrencia de eventos o fenómenos, para tomar medidas más acertadas ymaximizar sus beneficios o bien disminuir o eliminar sus efectos negativos.

CIERRE

Has concluido el estudio de esta unidad y es momento para llevar a cabo el recuento de los saberes aprendidos. Al igual que en la unidad 1, los fenómenos naturales y los procesos sociales fueron los ejes temáticos del estudio de los contenidos de la unidad 2. En ella te acercaste a las medidas de tendencia central y dispersión (medidas resumen que pueden servir como características que describen la naturaleza de una población o muestra), sirviendo además como parámetros de comparación entre dos poblaciones o muestras. Profundizaste el estudio de las distribuciones de probabilidad teórica normal, binomial y de Poisson, al introducir un conjunto de herramientas y fórmulas útiles para el cálculo de las probabilidades asociadas con eventos cuya naturaleza de ocurrencia hace posible la aplicación de modelos teóricos para la descripción de un fenómeno natural o un proceso social. Finalmente, estudiaste el coeficiente de correlación muestral y el ajuste de una línea de regresión por mínimos cuadrados (conceptos complementarios a la gráfica de dispersión) herramientas útiles para establecer la posible relación de dependencia entre dos variables.

Ya cuentas con elementos suficientes para solucionar la problemática planteada al inicio. Comienza con el análisis comparativo para medir el impacto que tiene para los residentes de origen foráneo y los oriundos de la comunidad el fenómeno migratorio. Para ello, haz el análisis comparativo de las medidas de tendencia central y dispersión del nivel de ingreso y edad para cada uno de estos dos grupos de población (véase la solución en el Apéndice 1). Los cálculos correspondientes a estas medidas resumen para cada uno de estos grupos son presentados en las siguientes tablas:

Residentes Oriundos				
Variable	Media	Mediana	Desviación Estándar	Rango
Nivel de ingreso	48.96	49.00	4.96	26
Edad	40.0	40.0	10.03	54

Residentes de Origen Foráneo				
Variable	Media	Mediana	Desviación Estándar	Rango
Nivel de ingreso	60.91	61.50	14.19	65
Edad	35.3	35.0	8.68	44

Después de estudiar las tablas, elabora un análisis comparativo de estas cantidades, podemos observar que tanto la media como la mediana del ingreso para los residentes de origen foráneo muestra ser considerablemente mayor que para los residentes oriundos. Por otra parte, las medidas de dispersión muestran una mayor variabilidad en el ingreso para los residentes de origen foráneo que para los oriundos. Lo que puede hacer que nos planteemos la hipótesis en el sentido de que esta mayor variabilidad se explica mediante una mayor heterogeneidad entre los niveles de especialización y capacitación de los residentes de origen foráneo, lo que les permite incursionar a una mayor diversidad de áreas del sector productivo y por lo tanto acceder a una mayor diversidad de escalas salariales.

A partir del análisis comparativo correspondiente a la edad, se puede advertir que el promedio de edad para los residentes de origen foráneo es menor que para los residentes oriundos; sin embargo la variabilidad entre las edades es menor para los residentes oriundos que para los foráneos. Dicho cambio en la variación muestral se explica, por la naturaleza de la estructura por edades de los residentes de origen foráneo, cuyo rango de edades varía entre los 18 y 62 años mientras que la estructura por edades para los residentes oriundos varía entre los 18 y 72 años.

Ahora centremos nuestro interés en dar interpretación a los resultados de las probabilidades $P(x_1 \geq 65) = .3372$ y $P(x_2 \geq 65) = .0014$. Estos valores indican que es más probable encontrar individuos con niveles salariales mayores a $65 diarios (considerados de medios a buenos) entre residentes de origen foráneo que entre residentes oriundos de la comunidad. Por otra parte, el valor calculado de la probabilidad $P(x_3 > 50) = .9729$ muestra que es más probable que la mitad de la mano de obra especializada de trabajadores altamente calificados sea de origen foráneo.

Si complementamos estos resultados con los obtenidos en el análisis comparativo, entonces tenemos la evidencia necesaria para afirmar o negar la marcada desigualdad en el ingreso entre los trabajadores de origen foráneo y los oriundos en la comunidad de El Encino.

La desigualdad puede explicarse por el mayor nivel de especialización de los trabajadores de origen foráneo. Para disminuir el desequilibrio habría que capacitar a los trabajadores oriundos; especializarlos con el fin de mejorar las condiciones salariales. La medida se sustenta en el hecho de que la intensidad de la inmigración hacia la comunidad de El Encino es mayor que la de la migración. Los cálculos probabilísticos, $P(x_4 = 7) = .0437$ y $P(x_5 = 7) = .0034$, revelan que es más probable que ingresen siete personas a la comunidad, a que este mismo número decida salir de ella.

Queda claro entonces que hay indicios para suponer que el fenómeno migratorio ha tenido un impacto económico benéfico en la comunidad de El Encino. Sin embargo, no está demás contar con otra medida objetiva que nos permita corroborarlo. Para ello proseguimos este análisis con el cálculo del coeficiente de correlación para los datos correspondientes al conjunto de 50 observaciones pareadas anuales (véase el Apéndice 2) correspondientes a los registros comunitarios de los últimos 50 años del producto interno bruto anual así como de la tasa neta de inmigración. Con el cálculo del coeficiente de correlación muestral se obtuvo un valor calculado de $r = .75$ (para solución véase el Apéndice 1), lo que evidencia la tendencia positiva creciente entre la tasa neta de inmigración anual y el producto interno bruto.

El cálculo de los coeficientes de correlación para cada uno de los tres conjuntos de datos pareados conformados por la tasa neta de inmigración —y los correspondientes valores observados del índice de pobreza humana, el índice de delincuencia así como el nivel medio anual de almacenamiento de agua de la presa El Encino— se obtuvieron los valores -.76, -.02 y -.09 respectivamente (para su solución véase el Apéndice 1).

A partir de estos valores se concluye que el fenómeno migratorio favoreció el abatimiento de los niveles de pobreza, pues el coeficiente de correlación de -.76 muestra que en el periodo considerado, el aumento en la inmigración hacia la comunidad trajo consigo una disminución de los niveles de pobreza.

Por último, los valores -.02 y -,09 correspondientes a los indicadores delictivos y de nivel de abastecimiento de agua, muestran una asociación muy débil con la tasa neta anual de inmigración lo que nos hace suponer que la influencia de este factor no es determinante en el comportamiento de estas variables. Ambos coeficientes permiten conjeturar que los valores de estas variables tienden a a disminuir ante la presencia de incrementos en la tasa neta anual de inmigración.

A partir de los resultados expuestos concluimos que la inmigración ha sido más benéfica que perjudicial para el desarrollo de la comunidad de El Encino pues,

según el análisis, no se presentó un aumento en los índice de pobreza humana y delincuencia. Además, el flujo migratorio no impacta sobre el abasto de agua, problema preocupante para la comunidad.

Los saberes desarrollados en esta unidad y su aplicación en el fenómeno migratorio de El Encino pueden ser una guía para el análisis y la interpretación de cualquier otro problema. Si decidiste trabajar uno de tu contexto local o regional ya cuentas con las herramientas y los ejemplos para solucionarlo.

Identificación de diferentes escenarios donde aplicar lo aprendido

La aplicación de los métodos y procedimientos estudiados en esta unidad va más allá de la naturaleza de los ejemplos planteados. Las medidas de tendencia central y dispersión, por ejemplo, son herramientas de comparación del volumen de ventas entre dos almacenes que están en competencia, o bien pueden usarse como medidas para comparar la eficiencia en el tiempo de atención a clientes entre dos sucursales bancarias.

El cálculo de probabilidades resulta de gran utilidad para el pronóstico de las ventas en una tienda de autoservicio o bien para pronosticar el número de personas que asiste a un centro comercial en cierta hora del día.

El coeficiente de correlación y el modelo de regresión son, entonces, herramientas que permiten verificar si una estrategia de inversión en publicidad deriva en un aumento en las ventas de cierto producto.

Es esta diversidad de aplicaciones y usos lo que los ha convertido en herramientas de análisis de información y fundamento para la toma de decisiones. Aprovecha su conocimiento.

Evalúa tu aprendizaje

¡Felicidades, terminaste el estudio de la segunda unidad! Has un recuento de lo aprendido evaluando tus logros.

I. Relaciona las dos columnas de tal forma que, delante de cada enunciado de la columna A anotes el número de la columna B que complete correctamente el enunciado.

Columna A		Columna B
Se característica por no ser afectada por valores extremos.	( )	1. Media, mediana y moda.
Es una característica del coeficiente de correlación muestral.	( )	2. Cantidad que sirve para cuantificar que tanto las observaciones de una muestra difieren unas de otras.
Característica de la línea de regresión.	( )	3. Moda.
Definición de medida de tendencia central.	( )	4. Tiene la desventaja de que su valor es afectado por la presencia de valores extremos.
Desventaja del uso de la media muestral como medida de tendencia central.	( )	5. Tiene la desventaja de no poder detectar relaciones de asociación no lineales entre dos variables.
Son ejemplos de medidas de tendencia central.	( )	6. Rango, varianza y desviación estándar.
Desventaja del uso del rango como medida de dispersión.	( )	7. Se define como el número que describe aproximadamente donde están localizados o centrados los datos a lo largo de la recta numérica.
Son ejemplos de medidas de dispersión.	( )	8. Proporción muestral.
Definición de medida de dispersión.	( )	9. Tiene la desventaja de que su suma anula contribuciones con magnitudes absolutas guales pero con distinto signo.
Es un ejemplo de medida de tendencia central para el caso de variables categóricas.	( )	10. Su valor siempre varía entre -1 y 1.
Desventaja del uso de la desviación media como medida de dispersión.	( )	11. Es una característica de la mediana como medida de tendencia central.
Es la medida de tendencia central que se define como el dato que más se repite en una muestra.	( )	12. Tiene la característica de ser una línea recta que está lo más cerca posible de los puntos descritos en una gráfica de dispersión.
Es una desventaja del uso del coeficiente de correlación muestral.	( )	13. Tiene la desventaja de no considerar para su cálculo todos los datos que pueden llegar a ocultar información sobre la verdadera naturaleza de variación de un conjunto de datos.

II Aplica las medidas de tendencia central y dispersión, del cálculo de probabilidades de eventos relacionados con las distribuciones de probabilidad normal, binomial, de Poisson y del cálculo e interpretación del coeficiente de correlación para responder nuevas interrogantes sobre la comunidad de El Encino y la migración.

Toma en cuenta los siguientes elementos:

1. Un conjunto de 100 nuevas mediciones para cada una de las variables nivel de ingreso y edad para residentes oriundos y de origen foráneo pertenecientes a la población económicamente activa de la comunidad de El Encino. Los datos correspondientes los puedes obtener en la base de datos del Apéndice 3.
2. Las cinco variables consideradas durante el desarrollo de la actividad 2, pero con valores distintos para sus parámetros. Las variables así como sus parámetros se muestran en la siguiente tabla.

Variable	Tipo de distribución teórica
x_1 = Ingreso diario de la población económicamente activa de origen foráneo.	normal con media $\mu = 52$ y desviación estándar $\sigma = 6$
x_2 = Ingreso diario de la población económicamente activa oriunda.	normal con media $\mu = 58$ y desviación estándar $\sigma = 4$
x_3= Número de trabajadores de origen foráneo que pueden ser encontrados en una muestra de 100 trabajadores que se dedican a labores que requieren para su desempeño un alto nivel de especialización.	Binomial con $n = 100$ y $p = .2$
x_4 = Número de personas que inmigran a la comunidad de El Encino por semana.	Poisson con $\lambda = 10$
x_5 = Número de personas que emigran de la comunidad de El Encino.	Poisson con $\lambda = 5$

3. Cuatro conjuntos de 50 nuevas observaciones pareadas anuales correspondientes a los registros comunitarios de los últimos 50 años, conteniendo cada uno de estos conjuntos de datos observaciones pareadas, correspondientes a la tasa de inmigración neta anual junto al registro del valor correspondiente a una de las cuatro variables dependientes consideradas. Los conjuntos de datos referidos los puedes obtener en el Apéndice 3.

No olvides verificar tus respuestas en el Apéndice 1.

I. Lee con atención la siguiente situación y escoge en cada reactivo la opción correcta.

> Eres habitante de una ciudad costera cuya población es de 1 millón de habitantes, aproximadamente. La ciudad está enclavada en un corredor turístico ubicado en la costa del mar Caribe. Por su ubicación y desarrollo económico es un polo de atracción para habitantes de otras regiones y del extranjero. Además sus condiciones climáticas posibilitan condiciones laborales y salariales que permiten una calidad de vida aceptable.

A partir de la situación, identifica cuáles de los siguientes fenómenos pueden influir el entorno geográfico de la ciudad:

1. Turismo
2. Pesca
3. Huracanes
4. Emigración
5. Inmigración
6. Tormentas Tropicales
7. Tsunamis
8. Marginación
9. Desarrollo Económico
10. Comercio
11. Inundaciones
12. Epidemias

Con base en esta lista de fenómenos responde las preguntas que se formulan a continuación:

1. ¿Cuáles de los siguientes fenómenos descritos pueden ser clasificados como fenómenos naturales?

   a) 1, 2, 5, 8 y 10
   b) 3, 6, 7, 11 y 12
   c) 1, 2, 8, 9 y 10
   d) 4, 5, 8, 11 y 12

2. ¿Cuáles de los siguientes fenómenos, que son característicos de la región descrita la situación anterior, pueden clasificarse como procesos sociales?

   a) 1, 2, 4, 5, 8, 9 y 10
   b) 3, 6, 7, 8, 10, 11 y 12
   c) 2, 3, 4, 8, 10, 11 y 12
   d) 1, 4, 5, 6, 7, 9 y 11

Los siguientes son eventos que por su naturaleza ocurren o pudieran ocurrir en el entorno geográfico de la ciudad descrita:

1. Día y noche
2. Niveles de precipitación pluvial durante el verano
3. Estaciones del año
4. Niveles de ocupación hotelera durante un año
5. Pérdidas económicas anuales
6. Número de defunciones anuales
7. Número de nacimientos anuales
8. Generación de empleos anual
9. Percepción mensual de una trabajador asalariado

Con base en la lista de eventos responde las preguntas que se formulan a continuación:

3. ¿Cuáles de los siguientes eventos que por su naturaleza ocurren o pudieran llegar a ocurrir en el entorno geográfico descrito pueden ser clasificados como eventos determinísticos?
   a) 2 , 4 y 6
   b) 4, 7 y 9
   c) 3, 6 y 7
   d) 1, 3 y 9

4. ¿Cuáles de los siguientes eventos que por su naturaleza ocurren o pudieran llegar a ocurrir en el entorno geográfico descrito pueden ser clasificados como eventos aleatorios?
   a) 1, 2, 3, 4, 5 y 6
   b) 3, 5, 6, 7, 8 y 9
   c) 2, 4, 5, 6, 7 y 8
   d) 1, 3, 6, 7, 8 y 9

II. Un investigador está interesado en realizar un estudio que le permita identificar las características ambientales que propician la propagación de parásitos que provocan enfermedades gastrointestinales. Así mismo, pretende recopilar información del entorno social con la finalidad de identificar los hábitos y las conductas sociales que propician la diseminación de estos padecimientos. También busca identificar las características de la población que resulta más afectada por la prevalencia de estas enfermedades. El investigador cuenta con un protocolo de investigación cuyos elementos metodológicos principales puedes identificar a partir de resolver acertadamente los siguientes reactivos:

5. El investigador en su protocolo de investigación identificó cuatro variables de interés:
   a) Ingreso per cápita
   b) Nivel socioeconómico
   c) Tipo de padecimiento gastrointestinal
   d) Número de pacientes con algún tipo padecimiento gastrointestinal

Utilizando el siguiente criterio de clasificación de variables:

1. Variable numérica continua
2. Variable numérica discreta
3. Variable categórica ordinal
4. Variable categórica nominal

Selecciona la clasificación correcta de entre los cuatro posibles criterios de clasificación mostrados a continuación.

a) 1a; 2b; 3c; 4d
b) 1a; 2d; 3b; 4c
c) 1c; 2d; 3a; 4c
d) 1b; 2c; 3a; 4d

6. Para obtener la información el investigador debe decidir aplicar un muestreo o un censo. No puede realizarlo sin conocer los conceptos propios de estas metodologías. Identifica cuáles de los siguientes son propios de ellas relacionando el con su definición.

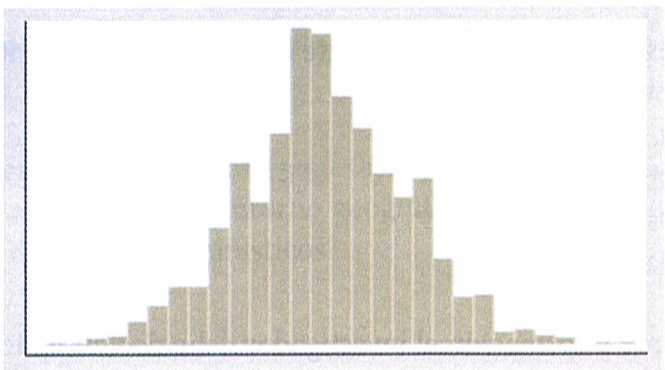
   Selecciona entre las cuatro opciones de respuesta desarrolladas al final la que consideres es la correcta.

1. Población	a) Recuento de una población con la finalidad de conocer cuestiones inherentes a ella.
2. Método de muestreo	b) Subconjunto extraído de la población mediante la aplicación de algún método de muestreo, que permite obtener conclusiones válidas que se pueden generalizar hacia toda la población.
3. Censo	c) Conjunto de todos los elementos o individuos que son objeto de estudio.
4. Muestra	d) Técnicas basadas en la recolección de una fracción de la información de la población objeto de estudio.
5. Muestra Representativa	e) Es una muestra seleccionada mediante un método que asegura que cada muestra posible del tamaño deseado tiene la misma oportunidad de ser elegida.

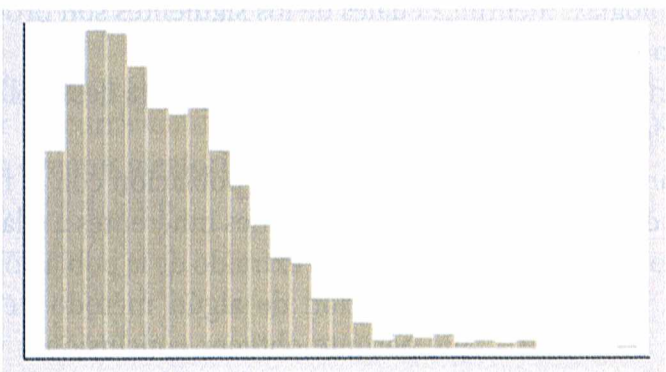
a) 1a; 2b; 3c; 4d; 5e
b) 1c; 2e; 3a; 4d; 5b
c) 1c; 2d; 3a; 4b; 5e
d) 1e; 2d; 3c; 4b; 5a

7. Con base en la información obtenida a partir de las mediciones de la variable ingreso per cápita, el investigador realiza un histograma de frecuencia. La forma del histograma le permite concluir que el modelo de la distribución de probabilidad normal es la opción de modelo teórico adecuado para modelar esta variable continua. Selecciona entre las siguientes cuatro gráficas el histograma que más de asemeja a una distribución normal y llevó al investigador a seleccionar el modelo teórico.

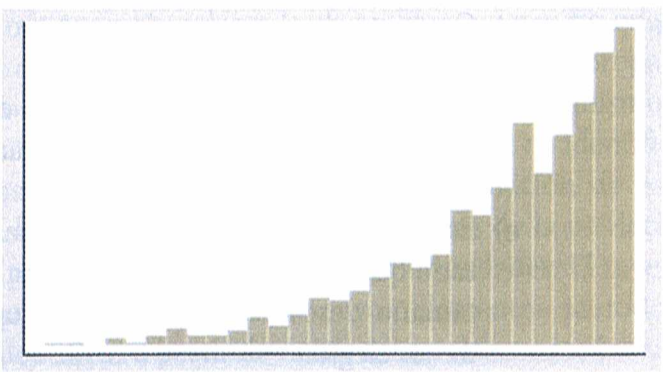
1



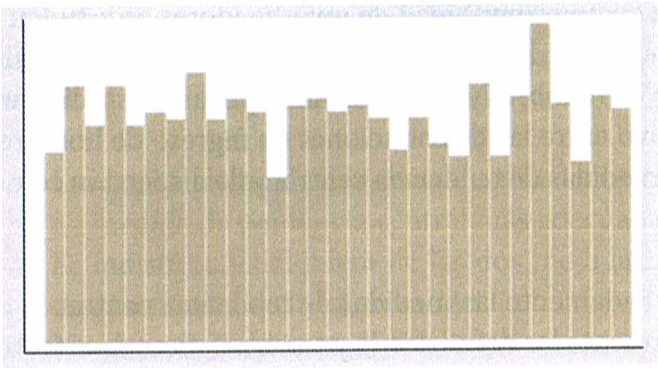
2.



3.



4.



Utiliza un buscador para identificar las fotos que pueden ser de utilidad.

a) 4
b) 2
c) 1
d) 3

8. Para investigar la prevalencia de los padecimientos gastrointestinales un investigador médico decide seleccionar una muestra aleatoria de tamaño $n = 100$ a partir de la cual toma una medición de la variable número de pacientes con algún tipo de padecimiento gastrointestinal. Considera que hay suficientes elementos para considerar a este experimento como uno de tipo binomial. Identifica los supuestos que asumió el investigador para ser considerado experimento binomial.

1. El experimento consta de 100 intentos idénticos.
2. El número promedio de veces en que ocurre un evento por unidad de tiempo o de espacio es constante.
3. Los 100 intentos son independientes.
4. La probabilidad de más de un suceso en una unidad de tiempo o espacio es muy pequeña.
5. Cada intento da lugar exactamente a dos resultados, éxito y fracaso. Tales resultados pueden ser medidos por la variable x=0 correspondiente al fracaso y x=1 correspondiente al éxito.
6. La probabilidad "p" de un éxito permanece constante de un intento a otro.

Selecciona la respuesta correcta de entre las cuatro opciones que se te presentan a continuación:

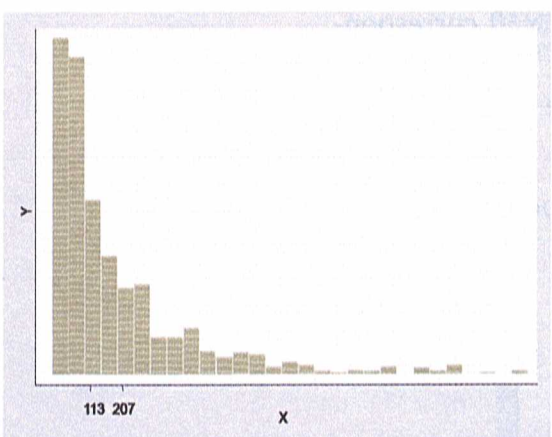
a) 1; 3; 5; 6
b) 2; 4; 5; 6
c) 1; 2; 3; 4
d) 1; 3; 4; 6

III. Hay interés por parte de una comunidad de investigadores de estudiar la equidad y género. Su objetivo es determinar si existen diferencias salariales entre los niveles de ingreso de hombres y mujeres. Para apoyar el estudio, realizarán un análisis comparativo entre las distribuciones de ingreso de hombres y mujeres; y utilizarán las medidas de tendencia central y de dispersión como herramientas.

9. Para aplicar adecuadamente las medidas de tendencia central, la comunidad debe tener presente las definiciones y características de estas medidas. Vincula los conceptos con sus definiciones y características relacionando las columnas. Selecciona entre las cuatro opciones la respuesta correcta.

1. Desventaja de la media muestral	a) Describe aproximadamente donde están localizados o centrados los datos a lo lago la recta numérica.
2. Mediana muestral	b) Su valor puede verse muy afectado por la presencia de valores extremos o atípicos.
3. Medida de tendencia central	c) Se define como la cantidad que divide al conjunto de datos en dos partes iguales.
4. Media muestral	d) Se define como el resultado del cociente de la suma de las observaciones dividida entre el número de observaciones.
5. Es una característica de la mediana muestral	e) Su valor no se ve afectado por la presencia de valores extremos.

a) 1a; 2b; 3c; 4d; 5e
b) 1c; 2e; 3a; 4d; 5b
c) 1c; 2d; 3a; 4b; 5e
d) 1b; 2c; 3a; 4d; 5e



10. Para decidir qué medida de tendencia central utilizar en el estudio de la distribución del ingreso de las mujeres, la comunidad calculó el valor de la media y la mediana a partir de los valores obtenidos en una muestra representativa. Los valores obtenidos para la media y la mediana fueron de $207.00 y $113.00, respectivamente. Junto a estos valores graficó sus respectivo histograma de frecuencias el cual se muestra a tu izquierda.

Con base en el análisis de la gráfica selecciona una respuesta sobre la medida de tendencia central más adecuada a utilizar como medida representativa de la información disponible.

a) La medida de tendencia central más adecuada es la media muestral.
b) La medida de tendencia central más adecuada es la mediana muestral.
c) Es indistinto, cualquiera de las dos medidas puede ser tomada como representativa de la información.
d) Ninguna de las dos medidas.

11. Para aplicar adecuadamente las medidas de dispersión la comunidad elabora un cuadro con los conceptos y las definiciones básicas. Identifica estas definiciones y características relacionando las siguientes columnas, vinculando los conceptos con sus definiciones respectivas. Selecciona entre las cuatro opciones de respuesta desarrolladas al final la que consideres es la correcta.

1. Desventaja del rango muestral como medida de dispersión	a. Medida que permite cuantificar que tanto las observaciones difieren unas de otras.
2. Desviación estándar muestral	b. Se define como el resultado de la diferencia entre la observación más grande y la observación más pequeña de un conjunto de datos.
3. Medida de dispersión	c. Para una muestra de tamaño n se define como la suma de los cuadrados de las desviaciones de la media divididas entre n-1.
4. Varianza muestral	d. Se obtiene como resultado de la raíz cuadrada de la varianza..
5. Rango muestral	e. Tiene la desventaja de no considerar en su cálculo la contribución de cada observación a la variabilidad.

a) 1e; 2d; 3a; 4c; 5b
b) 1c; 2e; 3a; 4d; 5b
c) 1c; 2d; 3a; 4b; 5e
d) 1b; 2c; 3a; 4d; 5e

12. En una región de difícil acceso, la comunidad de investigación se vio forzado a tomar dos muestras aleatorias de tamaño n=10 del ingreso diario en pesos de mujeres y hombres, con la finalidad de realizar un comparativo de la variabilidad en el ingreso de hombres y mujeres de este distrito. Los valores obtenidos del ingreso para ambas muestras aleatorias son presentados en las siguientes dos tablas:

Ingreso diario en pesos de 10 mujeres seleccionadas al azar				
$55.00	$38.00	$41.00	$54.00	$48.00
$48.00	$48.00	$52.00	$63.00	$57.00

Ingreso diario en pesos de 10 hombres seleccionadas al azar				
$47.00	$75.00	$35.00	$78.00	$46.00
$63.00	$85.00	$73.00	$59.00	$40.00

Para cada conjunto de datos calcula la desviación estándar y selecciona la respuesta que da una adecuada interpretación a los resultados obtenidos. Recuerda que para el cálculo de la desviación estándar puedes utilizar una hoja de cálculo.

a) A partir de los valores calculados para la desviación estándar de ambas muestras se puede concluir que existe evidencia que nos hace suponer que entre los hombres existe mayor variabilidad en sus ingresos diarios que entre las mujeres.
b) A partir de los valores calculados para la desviación estándar de ambas muestras se puede concluir que existe evidencia que nos hace suponer que entre las mujeres existe mayor variabilidad de ingresos diarios que entre los hombres.
c) A partir de los valores calculados para la desviación estándar de ambas muestras no se puede obtener algún tipo de conclusión.
d) A partir de los valores calculados para la desviación estándar de ambas muestras se puede concluir que existe evidencia que nos hace suponer que la variabilidad en las percepciones es la misma para hombres y mujeres.

IV. Un grupo interdisciplinario de investigación ha realizado un estudio sobre el impacto delos fenómenos meteorológicos sobre una región de alto desarrollo turístico ubicada en la costa del mar de Cortés. Durante su investigación, el equipo identificó un conjunto de variables cuyas características le permitieron asignar a cada una, una distribución de probabilidad teórica. Con los modelos probabilísticos determinados este grupo de investigación tuvo la posibilidad de estimar probabilidades de ocurrencia de eventos relacionados con fenómenos naturales y procesos sociales.

Determina los eventos y los modelos probabilísticos que fueron utilizados en la estimación de sus probabilidades de ocurrencia.

13. La variable x_1 = precipitación pluvial fue identificada como una variable cuya distribución de frecuencia se puede modelar de forma adecuada con

una distribución de probabilidad normal con media $\mu = 121.9$ *millones de* m^3 y desviación estándar $\sigma = 50.3$ *millones de* m^3. Utilizando esta distribución de probabilidad, ¿cuál es el valor de la probabilidad $P(71.6 \leq x_1 \leq 172.2)$?

a) .1587
b) .3174
c) .6826
d) .5000

14. Para poder medir el impacto social que una tormenta tropical dejo a su paso por la región, el grupo de investigadores seleccionó una muestra de 50 habitantes de la región a partir de la cual midió la variable x_2=número de habitantes damnificados. El modelo probabilístico adecuado para modelar la distribución de frecuencia de esta variable es una distribución binomial con $n = 50$ $p = .3$. Aplicando esta distribución de probabilidad, ¿cuál es el valor de la probabilidad $P(25 \leq x_2)$? Recuerda que puedes utilizar la hoja de cálculo para realizar este cálculo.

a) .9976
b) .0032
c) .9968
d) .0024

15. Para poder medir el impacto que los huracanes tienen sobre la región. El equipo de investigadores recabó la información necesaria para suponer que la variable x_3= número de huracanes por año, puede moldearse de manera adecuada por medio de una distribución de probabilidad de Poisson con $\lambda = 15$.

Identifica los supuestos o hipótesis que el grupo de investigadores tomó como base seleccionando de la siguiente lista aquellos que hacen que el evento que da origen a las observaciones de esta variable pueda ser considerada como un experimento de Poisson.

1. El número promedio de huracanes que ocurren en la región por unidad de tiempo (para este evento la unidad de tiempo será considerada un año) es constante.
2. La variable x_3 es una variable numérica discreta
3. La probabilidad de que más de un huracán ocurra en la región en cualquier intervalo de tiempo breve es casi cero. Esto significa que la probabilidad de que dos o más huracanes lleguen a ocurrir en la región durante un intervalo de tiempo breve por ejemplo de un día es muy pequeña o casi cero.

4. La variable x_3 toma un número finito de valores
5. El número de huracanes que suceden en la región en intervalos ajenos de tiempo son independientes unos de otros. Es decir que el número de huracanes que ocurrió en un intervalo concreto de tiempo no influye sobre el número de huracanes que serán observados en la región durante un próximo intervalo de tiempo.

Selecciona la respuesta correcta de entre las cuatro opciones de respuesta que se te muestran a continuación:

a) 1; 3; 5
b) 2; 4; 5
c) 1; 2; 3
d) 1; 3; 4

16. Utilizando esta distribución de probabilidad de Poisson con $\lambda = 15$, ¿cuál es el valor de la probabilidad $P(12 \leq x_3 \leq 14)$? Puedes utilizar una hoja de cálculo para realizar este cálculo.

a) .0829
b) .1785
c) .2809
d) .0024

V. Un grupo de investigación pretende identificar la relación entre las actividades del ser humano y el cambio climático, a partir del vínculo entre fenómenos como la sequía, aridez de una región o la degradación de recursos naturales y su efecto socioeconómico.

Por la naturaleza de la investigación planteada, el grupo de investigación deberá recurrir a la búsqueda de interacciones de variables que sean de utilidad para describir los fenómenos y procesos estudiados, por lo que tendrán que recurrir a las gráficas de dispersión, coeficiente de correlación y líneas de regresión lineal, como herramientas que les permitan indagar la relación entre variables.

Resuelve las siguientes preguntas para sabe qué es lo que el grupo de investigación debe saber acerca de los conceptos y aplicaciones de la gráfica de dispersión, coeficiente de correlación y línea de regresión lineal para adecuarlos a su problema de investigación.

17. Para la aplicación correcta de la gráfica de dispersión, el coeficiente de correlación y la línea de regresión es importante que el investigador conozca las definiciones y características de éstas. Identifícalas relacionando las siguientes columnas.

Vincula los conceptos con sus definiciones.
Selecciona entre las cuatro opciones de respuesta desarrolladas al final la que consideres es la correcta.

1. Gráfica de dispersión
2. Coeficiente de correlación muestral
3. Desventaja del coeficiente de correlación como medida de asociación
4. Línea de regresión
5. Tendencia crecientes
6. Tendencia decreciente

a. Tiene la característica de ser una línea recta que está lo más cerca posible de los puntos descritos en una gráfica de dispersión..
b. Tendencia que se caracteriza por observar incrementos en la variable Y como resultado de incrementos en la variable X.
c. Coeficiente que sirve para medir la posible relación lineal entre dos variables.
d. Tendencia que se caracteriza por observar decrementos en la variable Y como resultado de incrementos en la variable X.
e. Técnica que utiliza un conjunto de registros apareados de la forma (X,Y), con X como variable independiente e Y como variable dependiente, a partir de los que se elabora una gráfica conformada por parejas ordenadas en el plano cartesiano con base en la identificación del valor de la variable X respecto al eje horizontal mientras que el valor de la variable Y se identifica con el eje vertical.
f. Tiene la desventaja de no poder detectar relaciones de asociación no lineales entre dos variables.

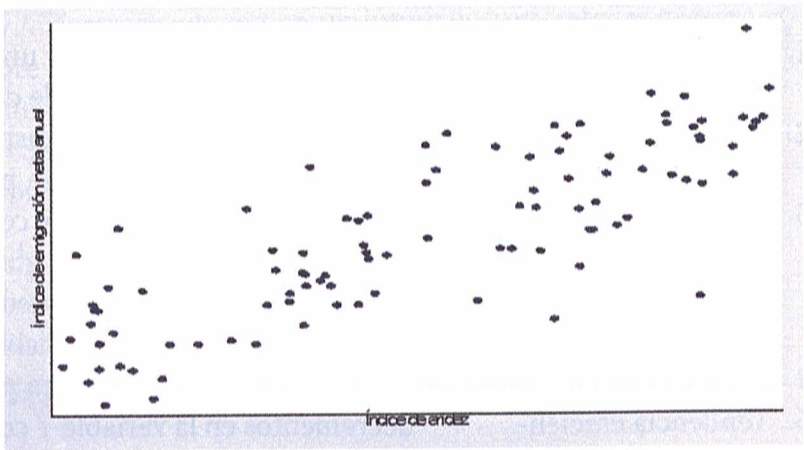
a) 1a; 2b; 3c; 4d; 5e; 6f
b) 1e; 2c; 3f; 4a; 5b; 6d
c) 1e; 2c; 3a; 4b; 5d; 6f
d) 1b; 2c; 3a; 4d; 5e; 6f

18. Durante el desarrollo de la investigación el equipo determinó la presencia de una tendencia creciente entre las variables índice de aridez e índice de emigración neta anual. Detectó la presencia de esta tendencia a partir de la construcción de una gráfica de dispersión junto con el cálculo del coeficiente de correlación muestral asociado.

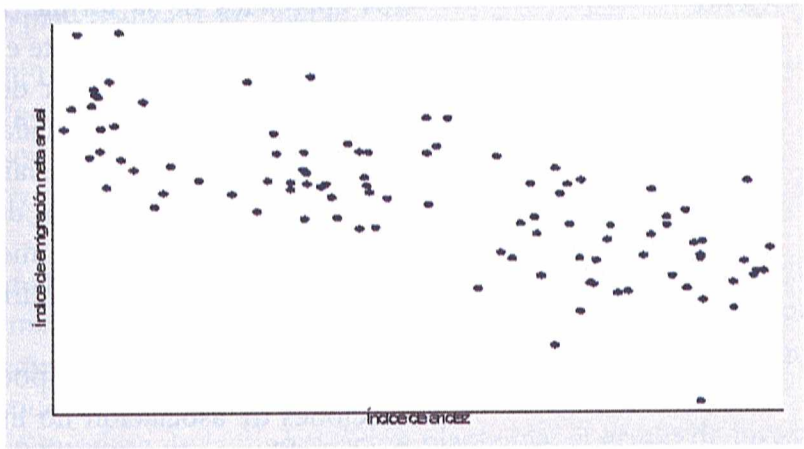
A partir de las gráficas de dispersión y de los coeficientes de correlación calculados selecciona de entre las cuatro opciones que se te presentan, la respuesta correcta que combine la gráfica de dispersión y el coeficiente de

correlación que permitió detectar esta tendencia creciente entre las variables.

a.



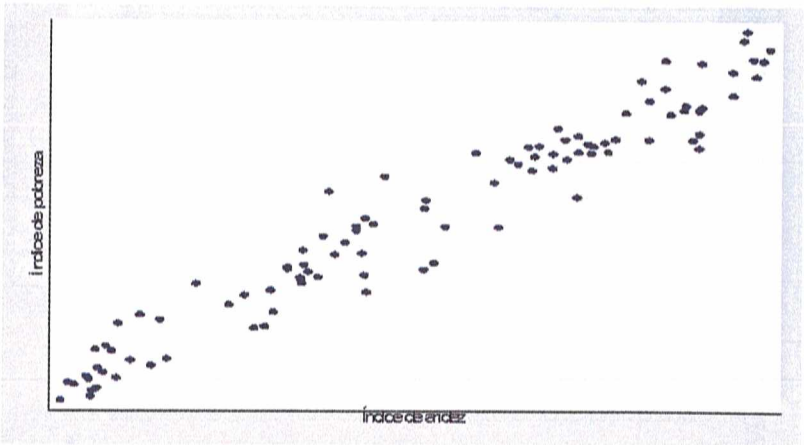
b.



1. r = -0.76
2. r = 0.82
   a) 1; b
   b) 2; b
   c) 1; a
   d) 2; a

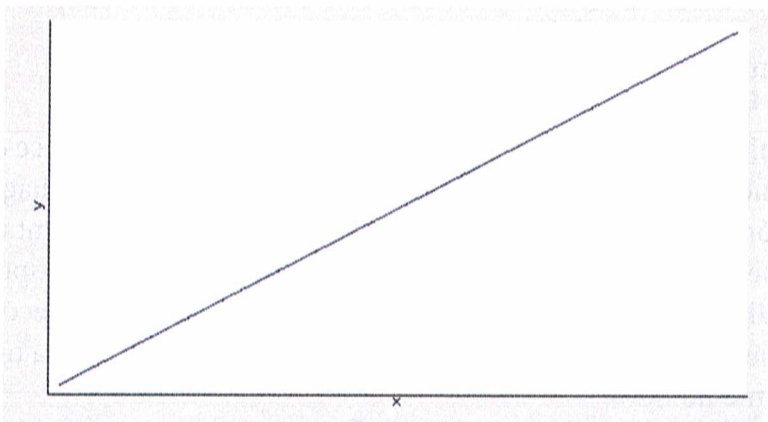
19. Para ajustar una línea de regresión que permita predecir el índice de pobreza como función del índice de aridez, con los datos recabados para ambas variables, el equipo de investigación realizó un diagrama de dispersión con el fin de definir qué tipo de línea de regresión es la más adecuada para representar la tendencia que se observa en el diagrama de dispersión.

La gráfica del diagrama de dispersión elaborada por el equipo de investigación se muestra a continuación.

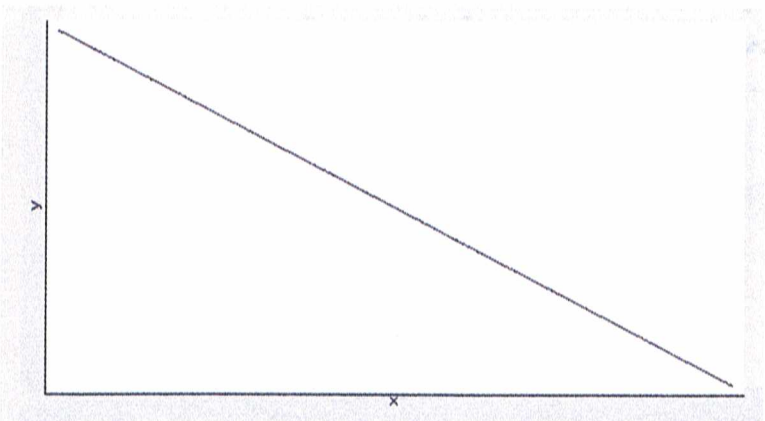


Con base en la gráfica de dispersión determina cuál de las siguientes líneas es la más adecuada para representar la línea de regresión de los datos mostrados en la gráfica de dispersión.

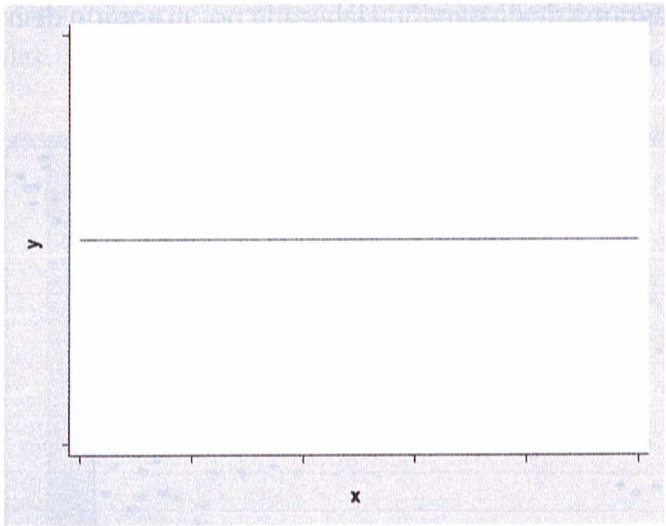
1.



2.



3.

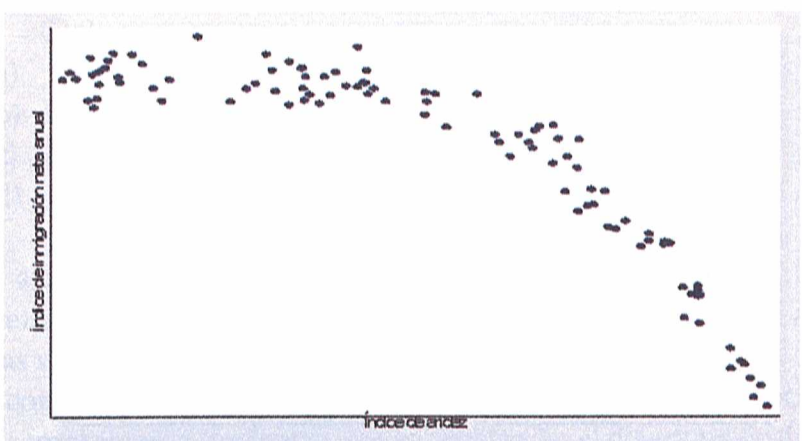


Selecciona la respuesta correcta de entre las cuatro opciones de respuesta que se te muestran a continuación:

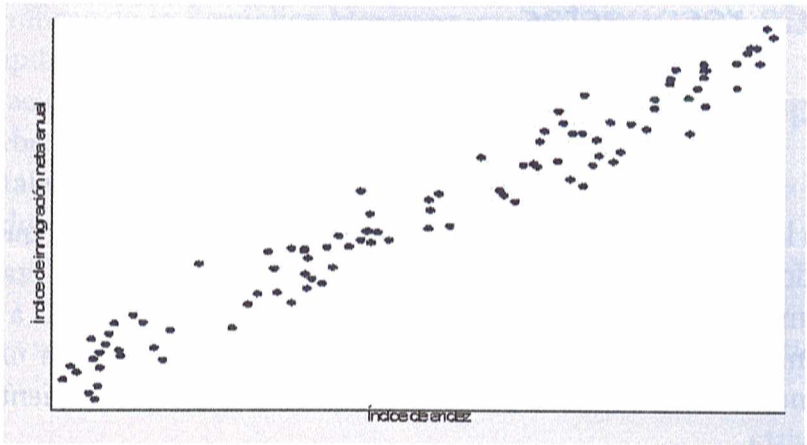
a) 1.
b) 3.
c) 2.
d) Ninguna de las anteriores.

20. Al estudiar la relación entre las variables índice de aridez e índice de inmigración neta anual, el grupo de investigación se percató que el diagrama de dispersión correspondiente a las observaciones pareadas de ambas variables presentó una tendencia decreciente no lineal. Identifica el tipo de gráfica de dispersión que el grupo de investigación observó, seleccionando cuál de las cuatro gráficas de dispersión mostradas muestra una tendencia decreciente no lineal.

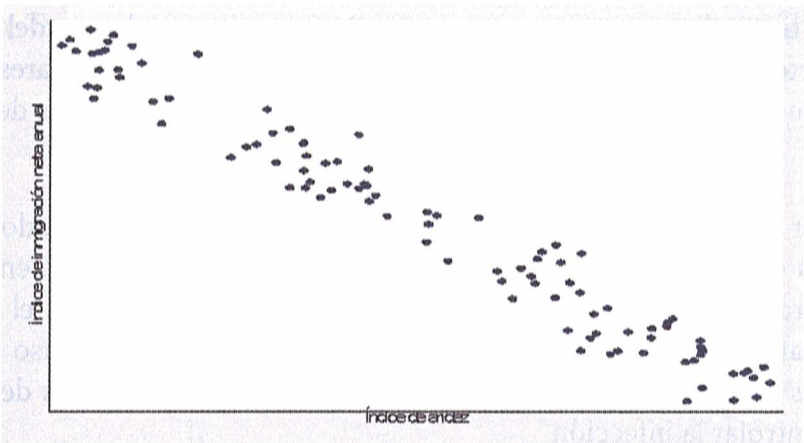
1.



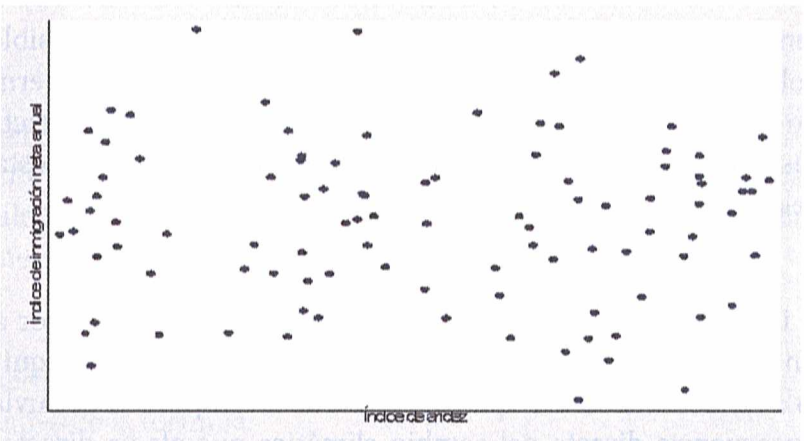
2.



3.



4.



a) 3
b) 2
c) 1
d) 4

Clave de respuestas

¿Con qué saberes cuento?

1. b)

   En la elección acertada de esta respuesta debiste considerar que el texto citado en la pregunta da un panorama general de la naturaleza del dengue como enfermedad de transmisión, sin la necesidad recurrir a un lenguaje demasiado técnico lo que permite ser leído y entendido por un público en general, convirtiéndolo así en un artículo de divulgación científica de una revista.

2. g)

   El texto claramente se enfoca en relacionar la prevalencia del dengue con factores de carácter sanitario de entornos humanos particulares. Es por ello que la respuesta g) es la que mejor caracteriza las condiciones descritas.

3. a)

   Por los condicionantes de transmisión del dengue considerados en el texto los cuales involucran factores asociados con entorno ambiental propicio para la reproducción del mosquito transmisor del dengue, el impedir encharcamientos, reservas inadecuadas de agua así como el uso de insecticidas se pueden considerar como las formas más adecuadas de erradicar o controlar la infección.

4. b)

   Con base en la observación directa de la enfermedad es posible plantear el problema así como dar un diagnóstico general del mismo, permitiendo con ello generar hipótesis así como realizar la comprobación directa de las mismas obteniendo con ello resultados. Es por ello que la respuesta b) es la que mejor caracteriza la situación descrita en el texto.

5. d)

   La influencia directa del cambio climático se caracteriza por alterar el orden normal de la frecuencia e intensidad de las lluvias, de aquí la pertinencia de considerar una amplia duración de la temporada de lluvias como una consecuencia directa del cambio climática que afecta directamente en la proliferación del dengue.

6. c)

   Los condicionantes ambientales y sociales que caracterizan la transmisión de la enfermedad del dengue, hacen necesario la identificación de las zonas

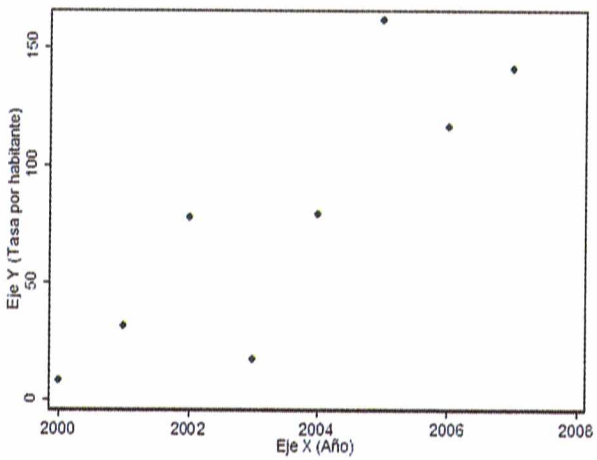
geográficas de la República Mexicana que presentan un mayor índice de precipitación pluvial, así como la difusión de las medidas preventivas y la aplicación de tratamientos adecuados para su tratamiento. Es por ello que se debe considerar importante la participación de dependencias gubernamentales como La Secretaría de Salud, COANAGUA así como la Secretaría de Educación Pública.

7. d)

Debes recordar que el 2.5% se define como 2.5/100 = 0.25. Cantidad que al ser multiplicada por 500,000 nos da la cifra correcta.

8.

Para encontrar la respuesta correcta, debes recordar que las parejas ordenadas que vas a graficar en el plano cartesiano deben considerar el arreglo (Año, Tasa por 100 mil habitantes).



9. d)

Recuerda que para encontrar un valor promedio basta con que sumes el total de casos de dengue para cada uno de los 8 años, con lo cual obtienes un valor de 45270 casos. Para posteriormente dividir este valor entre 8 y así obtener 5658.75 que es el valor de la respuesta correcta.

10. b)

Recuerda que para la obtención del resultado correcto debes aplicar adecuadamente la fórmula

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

Expresión que corresponde a la ecuación de una recta que pasa por los puntos (x_1, y_1) y (x_2, y_2) para $x_1 \neq x_2$.

Unidad 1

Actividad 1

1. En la siguiente tabla se muestra un listado de fenómenos cuyas características hacen que puedan ser ubicados en el entorno geográfico de la comunidad de Santiago. A continuación se te pide que en la columna de la derecha clasifiques cada uno de éstos como fenómeno natural o proceso social.

Fenómeno	Clasificación
Huracanes	Fenómeno natural
Inseguridad	Proceso social
Lluvias	Fenómeno natural
Escolaridad en la comunidad	Proceso social
Obesidad en la comunidad	Proceso social
Inundaciones	Fenómeno natural
Dengue en la comunidad	Fenómeno natural
Temperatura	Fenómeno natural
Pobreza y marginación	Proceso social
Sequía	Fenómeno natural
Crecimiento demográfico	Proceso social
Epidemias	Fenómeno natural
Migración	Proceso social
Inmigración	Proceso social
Enfermedades en la comunidad	Fenómeno Natural
Natalidad en la comunidad	Fenómeno natural
Servicios de salud en la comunidad	Proceso social

2. A partir de la tabla de clasificación de eventos selecciona 4 fenómenos naturales y 4 procesos sociales que por sus características consideres que pueden incidir en el contexto referido.

Fenómeno Natural	Proceso Social
Temperatura	Servicios de salud en la comunidad
Lluvias	Escolaridad en la comunidad
Huracanes	Crecimiento demográfico
Enfermedades en la comunidad	Pobreza y marginación

Actividad 2

1. Clasifica los 4 fenómenos naturales y los 4 procesos sociales considerados en la actividad 1 en determinísticos y aleatorios. En la siguiente tabla anota tus respuestas.

Fenómeno	Clasificación
Temperatura	Aleatorio
Lluvias	Aleatorio
Huracanes	Aleatorio
Enfermedades de la comunidad	Aleatorio
Servicios de salud en la comunidad	Determinístico
Escolaridad en la comunidad	Determinístico
Crecimiento demográfico	Aleatorio
Pobreza y marginación	Aleatorio

2. Supón que con la intención de buscar medidas encaminadas a mejorar las condiciones de salud de los habitantes de la comunidad de Santiago, estás interesado en conocer los principales padecimientos que se pueden constituir en problemas de salud pública para la comunidad. Para realizar esta medición solicitas al director de la clínica de esta localidad los datos correspondientes al tipo y número de casos de los padecimientos más comunes dentro de la comunidad. En la tabla siguiente se muestra el registro de ocurrencia anual para los distintos tipos de padecimientos presentes dentro de la comunidad.

Tipo de enfermedad	Número "n" de casos registrados
Enfermedades Respiratorias	245
Dengue	675
Infecciones intestinales	300
Cáncer	50
Diabetes	270
Número Total de Casos "N"	1540

Con base en estos registros calcula la frecuencia relativa asociada.
La frecuencia relativa asociada se presenta en la siguiente tabla:

Tipo de Enfermedad	Frecuencia Relativa
Enfermedades Respiratorias	0.16
Dengue	0.44
Infecciones Intestinales	0.19
Cáncer	0.03
Diabetes	0.18
Total	1.00

3. Como puedes observar a partir de la tabla anterior el dengue es el padecimiento que más prevalencia tuvo en la comunidad (.44 o 44%) a lo largo de un año. Por lo que ahora estás interesado en medir el impacto dentro de la comunidad de los 4 diferentes tipos dengue (serotipo 1, 2, 3 y 4). En la tabla siguiente se muestra el registro anual del número de casos.

Tipos de dengue Serotipo	Número "n" de casos registrados
1	235
2	145
3	200
4	95
Número Total de Casos "N"	675

Con base en estos registros calcula la frecuencia relativa asociada.
La frecuencia relativa asociada se presenta en la siguiente tabla:

Tipo de dengue Serotipo	Frecuencia Relativa
1	0.35
2	0.21
3	0.30
4	0.14
Total	1.00

Actividad 3

1. Con base en las características del entorno en el cual se encuentra ubicada la comunidad de Santiago, en el siguiente cuadro se presenta un conjunto de variables relacionadas con la naturaleza del entorno de esta comunidad.

Variables identificadas con eventos aleatorios relacionados con el entorno de la comunidad de Santiago
Sexo de los habitantes de Santiago
Escolaridad de los habitantes mayores de 18 años
Temperatura media anual en la comunidad
Tipo de enfermedades más frecuentes
Porcentaje anual de habitantes que se contagiaron de dengue durante el verano
Porcentaje de habitantes que se contagiaron de alguna enfermedad respiratoria durante el invierno
Número de altas semanales en la clínica de la comunidad
Número diario de camas disponibles en la clínica de la comunidad
Porcentaje de familias cuyo sustento económico es obtenido en actividades derivadas de la comercialización de productos del mar
Número anual de huracanes y tormentas tropicales que han impactado la comunidad
Porcentaje anual de la población que ha migrado hacia la comunidad
Porcentaje anual de la población que ha emigrado fuera de la comunidad
Índice de analfabetismo en la comunidad
Número diario de personas detectadas con diabetes en el hospital de la comunidad.
Presión sanguínea sistólica registrada en las personas que acuden a la clínica de la comunidad.
Tipo de accidente por el que ingresan los pacientes a la sala de emergencia en la clínica de la comunidad

A partir de las variables registradas en este cuadro, realiza una clasificación de éstas en cuantitativas (continuas o discretas) y cualitativas (ordinales y nominales). Utiliza para ello la tabla que se muestra a continuación.

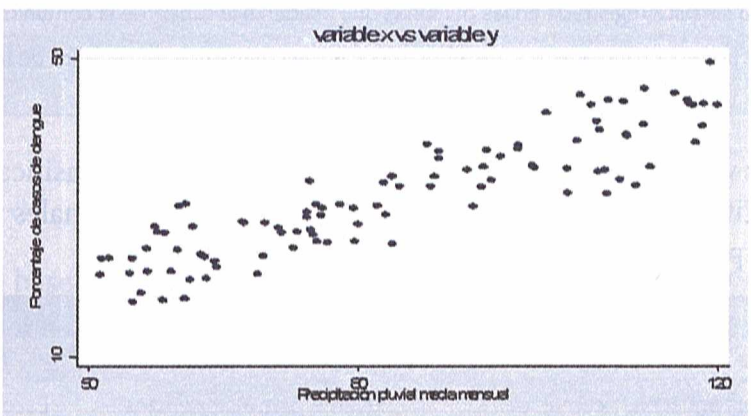
Variables Cuantitativas		Variables Cualitativa	
Continuas	**Discretas**	**Ordinales**	**Nominales**
Temperatura media anual en la comunidad Porcentaje anual de habitantes que se contagiaron de dengue durante el verano Porcentaje de habitantes que se contagiaron de alguna enfermedad respiratoria durante el invierno Porcentaje de familias cuyo sustento económico es obtenido en actividades derivadas de la comercialización de productos del mar Porcentaje anual de la población que ha migrado hacia la comunidad Porcentaje anual de la población que ha emigrado fuera de la comunidad Índice de analfabetismo en la comunidad Presión sanguínea sistólica registrada en las personas que acuden a la clínica de la comunidad	Número de altas semanales en la clínica de la comunidad Número diario de camas disponibles en la clínica de la comunidad Número anual de huracanes y tormentas tropicales que han impactado a la comunidad Número diario de personas detectadas con diabetes en el hospital de la comunidad.	Escolaridad de los habitantes mayores de 18 años	Sexo de los habitantes de Santiago Tipo de enfermedades más frecuentes Tipo de accidente por el que ingresan los pacientes a la sala de emergencia en la clínica de la comunidad

2. De las variables que clasificaste en la tabla anterior identifica aquellas que puedan tener una relación de causa-efecto. Utiliza la siguiente tabla para registrar cada variable independiente identificada con su correspondiente variable dependiente.

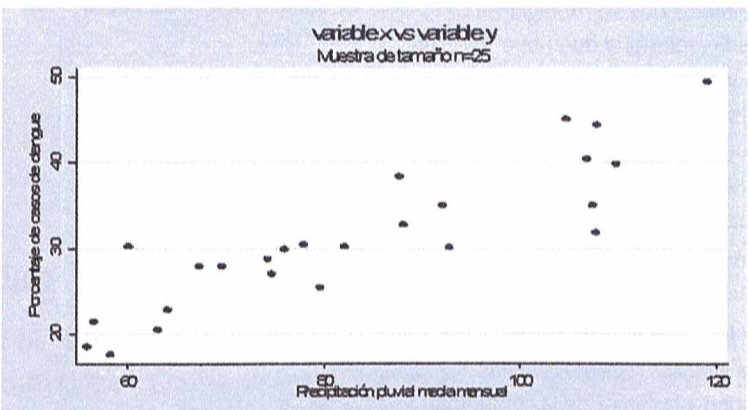
Variables Independientes	Variable Dependiente
Número anual de huracanes y tormentas tropicales que han impactado a la comunidad	Porcentaje anual de habitantes que se contagiaron de dengue durante el verano.
Temperatura media anual en la comunidad	Porcentaje de habitantes que se contagiaron de alguna enfermedad respiratoria durante el invierno.
Sexo de los habitantes de Santiago	Tipo de enfermedades más frecuentes

Actividad 4

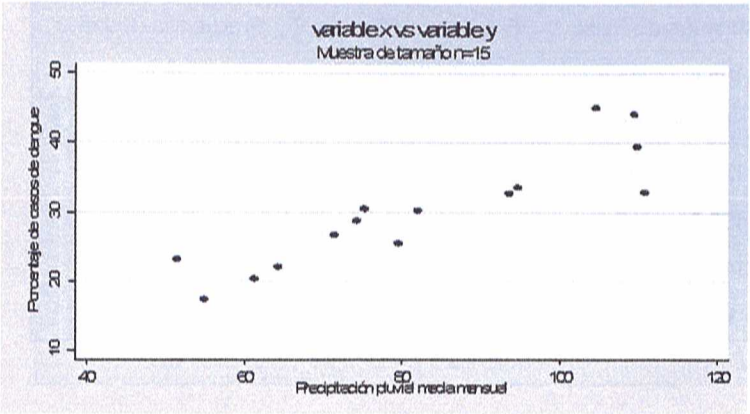
2. Realiza tres gráficas de dispersión, uno correspondiente a la totalidad de las observaciones pareadas y las otros dos correspondientes a las dos muestras de observaciones pareadas que obtuviste.
   a) Gráfica de dispersión de la totalidad de los datos



   b) Gráfica de dispersión de la muestra de n=25 datos



c) Gráfica de dispersión de la muestra de n=15 datos



3. Las tres gráficas muestran una tendencia creciente lo que nos indica que existe asociación entre las variables, por lo que a mayor cantidad de precipitación pluvial promedio mensual aumenta el porcentaje de casos de esta enfermedad entre los habitantes. Se puede concluir que las dos muestras aleatorias son muestras representativas

Actividad 5

1. Para cada una de las variables descritas identifica a cada una de estas en continuas o discretas.

Variable	Tipo
Presión sanguínea sistólica	Continua
Número diario de camas disponibles en la clínica de la comunidad	Discreta
Número de altas médicas semanales en la clínica de la comunidad	Discreta

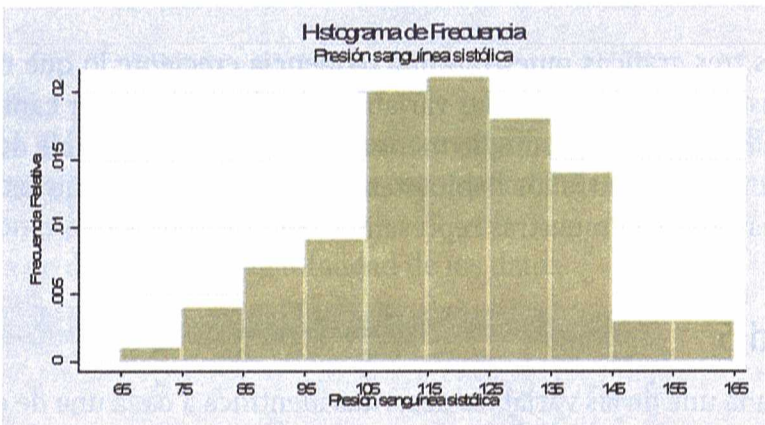
2. Por ser continua, se tiene la posibilidad de que la variable "Presión sanguínea sistólica" sea modelada como una distribución de probabilidad normal. Para corroborar esto a continuación se presenta una tabla de su distribución de frecuencias así como su correspondiente histograma.

Distribución de frecuencias para la variable presión sanguínea sistólica		
Intervalos de Clase	Frecuencia Absoluta	Frecuencia Relativa
40 a <75.1	1	0.01
75.1 a <85.2	4	0.04
85.2 a <95.3	7	0.07
95.3 a <105.4	9	0.09

(Continúa...)

(Continuación...)

Distribución de frecuencias para la variable presión sanguínea sistólica		
Intervalos de Clase	**Frecuencia Absoluta**	**Frecuencia Relativa**
105.4 a <115.5	20	0.20
115.5 a <125.6	21	0.21
125.6 a <135.7	18	0.18
135.7 a <145.8	14	0.14
145.8 a <155.9	3	0.03
155.9 a <126	3	0.03



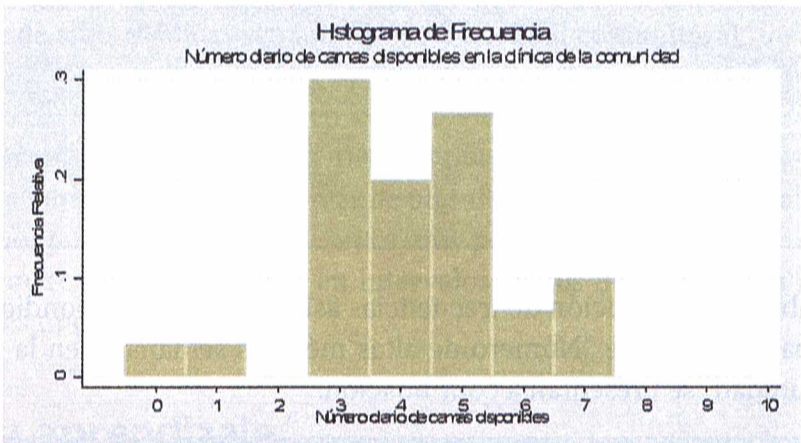
Al realizar una inspección visual del histograma se realizan las siguientes dos observaciones:

a) Se observa que este tiene una forma simétrica. Por lo tanto, es posible modelar esta variable con una distribución de probabilidad normal.
b) Se observa que los registros más frecuentes de presión sanguínea sistólica rondan el valor 120.

La tabla de distribución de frecuencias así como su correspondiente histograma para la variable "Número diario de camas disponibles en la clínica de la comunidad" se presentan a continuación:

Distribución de frecuencias para la variable número diario de camas disponibles en la clínica de la comunidad		
Número de camas	**Frecuencia Absoluta**	**Frecuencia Relativa**
0	1	0.03
1	1	0.03

2	0	0.00
3	9	0.30
4	6	0.20
5	8	0.27
6	2	0.07
7	3	0.10
8	0	0.00
9	0	0.00
10	0	0.00



Al realizar una inspección visual del histograma se observa que el número de camas disponibles para la atención de enfermos en el hospital de la comunidad durante el periodo observado la mayoría de las veces estuvo entre las 3 y las 5 por día. Sólo en una ocasión no se tuvo disponibilidad de camas.

Si sobre esta variable se establecen los siguientes supuestos:

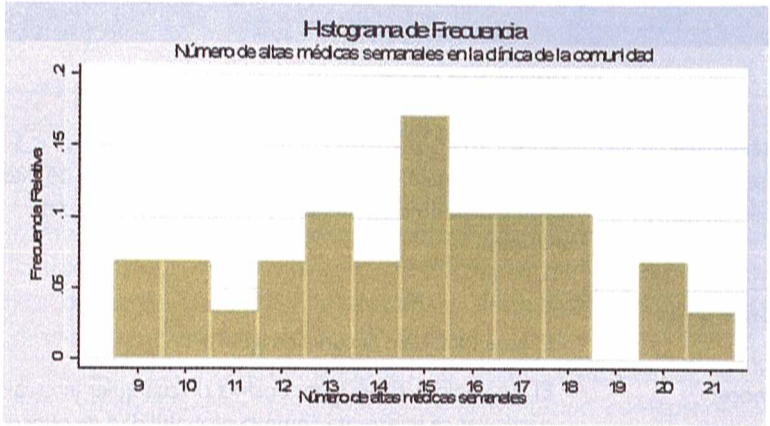
1. El experimento consiste de n=10 intentos idénticos (en este caso cada intento representa observar cada cama del hospital)
2. Cada intento da lugar a exactamente dos resultados, llamados éxito y fracaso. En este caso el éxito se puede considerer como el hecho de observer una cama disponible.
3. Los n=10 intentos son independientes
4. La probabilidad "p" de un éxito permanece constante de un intento a otro.

Si estos supuestos son asumidos, entonces podemos modelar esta variable con una distribución de probabilidad binomial.

Distribución de frecuencias para la variable número de altas médicas semanales en la clínica de la comunidad		
Número de camas	Frecuencia Absoluta	Frecuencia Relativa
9	3	0.07
10	2	0.05
11	1	0.02
12	2	0.05
13	3	0.07
14	14	0.33
15	5	0.12
16	3	0.07
17	3	0.07
18	3	0.07
19	0	0
20	2	0.05
21	1	0.02

La tabla de distribución de frecuencias así como su correspondiente histograma para la variable "Número de altas médicas semanales en la clínica de la comunidad" se presentan a continuación:

Distribución de frecuencias para la variable número de altas médicas semanales en la clínica de la comunidad		
Número de camas	Frecuencia Absoluta	Frecuencia Relativa
9	3	0.10
10	2	0.07
11	1	0.03
12	2	0.07
13	3	0.10
14	2	0.07
15	5	0.17
16	3	0.10
17	3	0.10
18	3	0.10
19	0	0
20	2	0.07
21	1	0.03



Al realizar una inspección visual del histograma se observa que el número más frecuente de altas médicas semanales en el hospital comunitario fue de 15.

Si sobre esta variable se establecen los siguientes supuestos:

1. El número promedio de altas médicas por semana es constante.
2. La probabilidad de que haya más de una alta médica en un interval corto de tiempo (por ejemplo un minute) es muy pequeña.
3. El número de altas médicas en intervalos ajenos de tiempo son independientes unos de otros.

Evalúa tu aprendizaje

I.

Columna A			
Son ejemplos de fenómenos naturales	(14)	Definición de Frecuencia Relativa	(3)
Son ejemplos de procesos sociales	(17)	Utilidad de un Diagrama de Dispersión	(25)
Son ejemplos de eventos determinísticos	(30)	Es una característica de las variables dependientes	(20)
Son ejemplos de eventos aleatorios	(26)	Son tipos de variables cualitativas	(7)
Definición de fenómeno natural	(9)	Son tipos de variables cuantitativas	(23)
Definición de proceso social	(27)	Característica de un muestreo aleatorio simple	(11)
Definición de censo	(2)	Característica de un muestreo aleatorio estratificado	(24)
Definición de muestra	(28)	Característica de un muestreo por conveniencia	(33)
Definición de muestra representativa	(34)	Definición de población	(5)
Definición de evento determinístico	(36)	Son característica de una curva de densidad	(22)
Definición de evento aleatorio	(35)	Son las características de un experimento binomial	(1)
Definición de Estadística	(31)	Son las características de un experimento Poisson	(16)
Es una característica de los eventos aleatorios	(6)	Característica de una curva que sigue una distribución teórica normal	(15)
Es una característica de los eventos determinísticos	(13)	Son algunos tipos de distribuciones de probabilidad teóricas	(4)
Ventajas de trabajar con una muestra	(18)	Es un ejemplo de una distribución de probabilidad continua	(10)
Ventaja de trabajar con un censo	(21)	Son ejemplos de distribuciones de probabilidad discreta	(19)
Son ejemplos de variables cualitativas	(32)	Es un elemento que sirve para elaborar un histograma	(8)
Son ejemplos de variables cuantitativas	(12)		
Definición de Histograma	(29)		

Columna B

1.
El experimento consiste de n intentos idénticos
- Cada intento da lugar a exactamente a dos resultados, llamados éxito y fracaso (estos resultados pueden ser medidos por la variable x=0 correspondiente al fracaso y x=1 correspondiente al éxito).
- Los n intentos son independientes
- La probabilidad "p" de un éxito permanece constante de un intento a otro.

2. Recuento de una población con la finalidad de conocer cuestiones inherentes a ella

3. Se define como el cociente entre la frecuencia absoluta del evento y el número total de eventos observados dentro de un periodo específico de tiempo o un espacio geográfico determinado

4. Distribuciones de probabilidad normal, binomial y de Possion

5. Conjunto de todos los elementos o individuos que son objeto del estudio

6. Incertidumbre

7. Variables categóricas nominales y ordinales

8. Intervalos de clase

9. Todo evento en la naturaleza que sucede por si solo sin ser propiciado por la intervención directa del hombre

10. Distribución de probabilidad normal

11. Es una muestra seleccionada mediante un método que asegura que cada muestra posible del tamaño deseado tiene la misma oportunidad de ser elegida

12. Porcentaje de población desocupada, peso, estatura.

13. Se puede predecir su ocurrencia con exactitud

14. Terremotos, erupciones volcánicas, sequías

15. Tiene forma de campana y es simétrica

16.
- El número promedio de veces que ocurre un evento por unidad de tiempo o de espacio es constante.
- La probabilidad de más de un suceso en una unidad de tiempo o espacio es muy pequeña.
- El número de acontecimientos en intervalos ajenos de tiempo o de espacio son independientes unos de otros

17. Analfabetismo, migración, marginación

18. Tiene la ventaja de requerir para su realización de una menor inversión de recursos humanos, financieros y materiales.

19. Distribución binomial y de Poisson.

20. Relación de causa efecto.

21. Tiene la ventaja de poder medir la característica de interés a cada individuo o elemento de la población objeto de estudio.

22.
- El área total bajo la curva es igual a 1
- El área bajo la curva y por encima de cualquier intervalo particular se interpreta como la probabilidad de observar un valor en el intervalo correspondiente, cuando una persona u objeto es seleccionado al azar de la población.

23. Variables continuas y variables discretas

24. Divide la población objeto de estudio en subgrupos, y selecciona muestras aleatorias simples hacia dentro de cada subgrupo

25. Es un diagrama que sirve para determinar visualmente la posible relación de causa efecto entre dos variables.

26. Terremotos, inundaciones, migración epidemias.

27. Proceso que genera un cambio en cualquier ámbito de la sociedad, se caracteriza por la intervención directa del hombre, pues a diferencia de la naturaleza éste siempre tiene la posibilidad de interferir directamente en su contexto social para modificarlo

28. Subconjunto extraído de la población mediante la aplicación de algún método de muestreo

29. Representación gráfica de una distribución de frecuencias.

30. El fenómeno del día y la noche, el salario diario de un trabajador, la duración de una jornada laboral

31. Ciencia referente a la recolección, análisis e interpretación de datos, que busca explicar condiciones regulares en fenómenos de tipo aleatorio.

32. Nivel Socioeconómico, sexo, religión, filiación política.

33. Técnica de muestreo que recurre a un grupo de fácil acceso o conveniente para formar una muestra.

34. Muestra caracterizada por incluir representantes de toda la población objeto de estudio.

35. Es aquel evento cuya naturaleza de ocurrencia es incierta y no se puede predecir con exactitud.

36. Es aquel evento cuya naturaleza de ocurrencia se puede predecir con exactitud.

II. Lee con atención el siguiente texto:

La sequía, pobreza y destrucción de los recursos naturales

La desertificación es un proceso complejo de los ecosistemas de las tierras secas, se le define como la degradación de la tierra en zonas áridas, semiáridas y sub húmedas secas, resultante de factores diversos como variaciones climáticas y las actividades humanas, está presente en todos los continentes (excepto en la Antártida) y afecta a millones de personas. La desertificación se produce como resultado del desequilibrio, provocado muchas veces por el ser humano, ocurre a largo plazo a causa de la demanda de recursos de los ecosistemas, de los que se obtienen beneficios como: cosechas, forrajes, madera, alimentación combustible. La combinación de los cambios climáticos y las acciones del hombre (agricultura, ganadería, pesca, silvicultura y crecimiento demográfico) suponen una gran amenaza para las tierras secas y a su vez, afecta a las sociedades humanas y a todos los seres vivos del planeta.

A partir de la lectura que hiciste de este texto realiza lo que se te pide a continuación:

III. En la siguiente tabla se muestra un registro de un conjunto de fenómenos que pueden ser observados en el contexto del texto que acabas de leer.

Fenómeno	
Desertificación	Incendios
Migración	Erosión del suelo
Sequia	Calentamiento global
Pobreza	Desigualdad social
Marginación	Precipitación pluvial
Inundaciones	Desnutrición
Deforestación	Estaciones del año
Contaminación ambiental	

Realiza un registro en la siguiente tabla de este conjunto de eventos en fenómenos naturales y procesos sociales.

Fenómenos Naturales	Procesos Sociales
Desertificación	Migración
Sequía	Pobreza
Inundaciones	Marginación
Incendios	Contaminación ambiental
Erosión del suelo	Desigualdad social
Calentamiento global	Desnutrición
Precipitación pluvial	
Estaciones del año	

IV. Realiza un registro en la siguiente tabla de este conjunto de fenómenos en eventos aleatorios y determinísticos.

Eventos Aleatorios	Eventos Determinísticos
Desertificación Migración Sequia Pobreza Marginación Inundaciones Deforestación Contaminación ambiental Incendios Erosión del suelo Calentamiento global Desigualdad social Precipitación pluvial Desnutrición	Estaciones del año

V. El siguiente cuadro muestra un conjunto de variables que pueden ser medidas a partir del contexto de la sequía, pobreza y destrucción de los recursos naturales. Estas variables que se muestran en el siguiente cuadro han sido clasificadas en variables de acidez o sequía, sociales, económicas, demográficas y de presencia del Estado.

Variables de acidez o sequía	Variables Sociales	Económicas	Demográficas	Presencia del Estado
Precipitación pluvial Ocurrencia de la precipitación pluvial Evaporación potencial Índice de aridez	Tasa de migración neta Porcentaje de mujeres jefas del hogar Densidad demográfica Tasa media de crecimiento anual de la población. Estructura de la población por edad y percepción económica Incidencia de pobreza Acceso a la educación Acceso a servicios de salud Enfermedades de mayor incidencia Brecha de pobreza Tasa de migración neta Intensidad de la pobreza Índice de desarrollo humano Mortalidad infantil Escolaridad Porcentaje de personas o familias con energía eléctrica Porcentaje de personas o familias con agua potable Porcentaje de personas con drenaje	Destino de los recursos para la recuperación de tierras Ingreso agrícola de las familia/Ingreso total por familia Producción para la subsistencia/Producción total Expansión de la frontera agropecuaria Tamaño de las propiedades rurales Porcentaje de la población que usa leña y carbón como combustible Tasa de endeudamiento Extracción de productos forestales Estructura de la propiedad Población económicamente activa por género.	Población Densidad (por km2) Tasa de crecimiento de la población Tasa media anual de migración neta Población mayor de 75 años	Número de políticas destinadas al área Coherencia y articulación en las políticas de estado de aplicación a la zona Tipo de política destinada (asistencias, promocional) Fase de proyecto en que participa la población(en todo el proceso, en la ejecución, ausencia de participación) Aportes efectivos de capital Áreas atendidas sobre áreas no atendidas

En el cuadro siguiente realiza una clasificación de estas variables en cuantitativas (continuas y discretas) y cualitativas (ordinales y nominales)

Variables Cuantitativas		Variables Cualitativa	
Continuas	**Discretas**	**Ordinales**	**Nominales**
Precipitación pluvial Evaporación potencial Índice de aridez Tasa de migración neta Porcentaje de mujeres jefas del hogar Tasa media de crecimiento anual de la población. Estructura de la población por edad y percepción económica Incidencia de pobreza Acceso a la educación Acceso a servicios de salud Brecha de pobreza Tasa de migración neta Intensidad de la pobreza Índice de desarrollo humano Mortalidad infantil Porcentaje de personas o familias con energía eléctrica Porcentaje de personas o familias con agua potable Porcentaje de personas con drenaje Ingreso agrícola de las familia/ Ingreso total por familia Producción para la subsistencia/Producción total Expansión de la frontera agropecuaria Tamaño de las propiedades rurales Porcentaje de la población que usa leña y carbón como combustible Tasa de endeudamiento Población económicamente activa por género. Tasa de crecimiento de la población Tasa media anual de migración neta Aportes efectivos de capital	Ocurrencia de la precipitación pluvial Densidad demográfica Enfermedades de mayor incidencia Escolaridad Población Densidad(por km2) Población mayor de 75 años Número de políticas destinadas al área	Fase de proyecto en que participa la población(en todo el proceso, en la ejecución, ausencia de participación)	Destino de los recursos para la recuperación de tierras Estructura de la propiedad Extracción de productos forestales Coherencia y articulación en las políticas de estado de aplicación a la zona Tipo de política destinada (asistencias, promocional) Áreas atendidas sobre áreas no atendidas

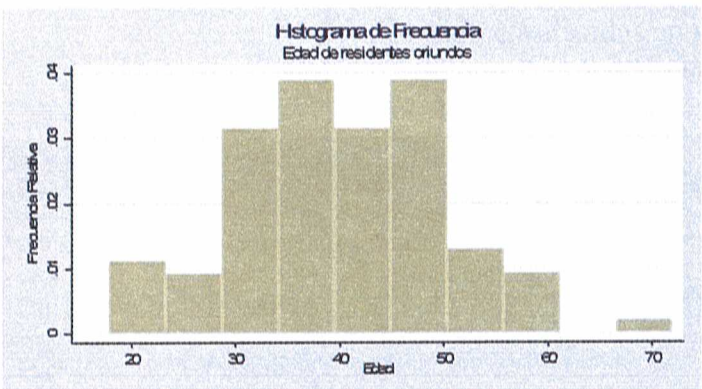
Unidad 2

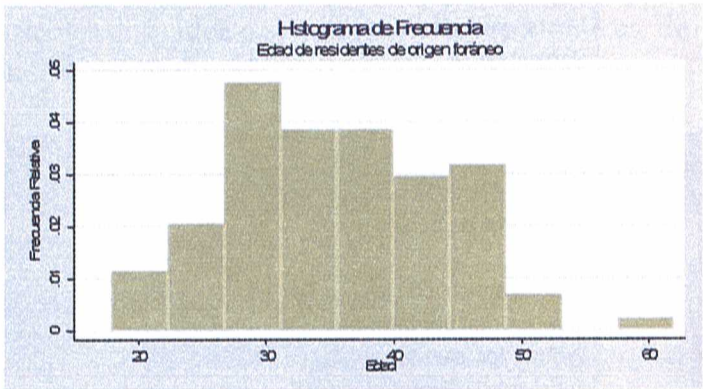
Actividad 1

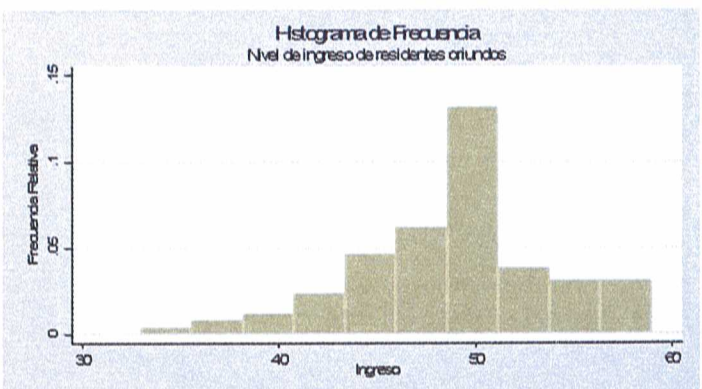
3. Los cálculos correspondientes a las medidas de tendencia central y de dispersión para las variables consideradas en esta actividad (las cuales fueron calculadas utilizando la hoja de cálculo de Microsoft Excel) así como sus respectivos histogramas de frecuencia, se presentan a continuación.

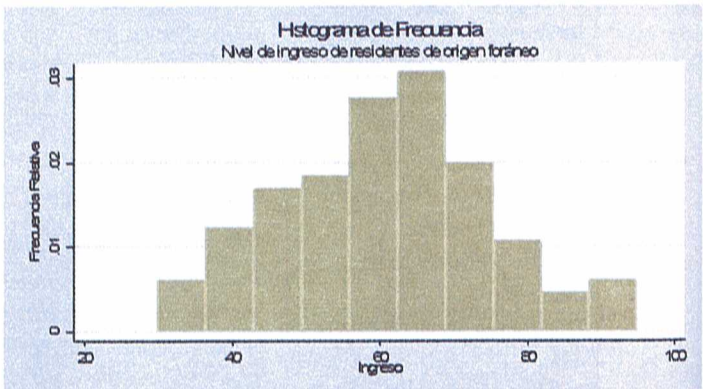
Residentes Oriundos				
Variable	**Media**	**Mediana**	**Desviación Estándar**	**Rango**
Nivel de Ingreso	48.96	49.00	4.96	26
Edad	40.0	40.0	10.03	54

Residentes de Origen Foráneo				
Variable	**Media**	**Mediana**	**Desviación Estándar**	**Rango**
Nivel de Ingreso	60.91	61.50	14.19	65
Edad	35.3	35.0	8.68	44









4. Al realizar un análisis comparativo de los valores calculados de las medidas de tendencia central y de dispersión así como de los correspondientes histogramas de frecuencia de las variables nivel de ingreso y estructura por edad entre residentes oriundos y de origen foráneo, se pueden realizar las siguientes observaciones:

   a) Se puede observar que tanto la media como la mediana del ingreso para los residentes de origen foráneo muestra ser considerablemente mayor que para los residentes oriundos.
   b) Las medidas de dispersión así como los respectivos histogramas muestran una mayor variabilidad en el ingreso para los residentes de origen foráneo que para los oriundos.
   c) La comparación entre ambos grupos de las medidas de tendencia central para la variable edad mostró que tanto el promedio como la mediana de la edad para los residentes de origen foráneo es menor que para los residentes oriundos.
   d) Al comparar la variabilidad de la variable edad entre ambos grupos a partir de las medidas de dispersión y de los histogramas correspondientes se encontró que ésta es menor entre los residentes oriundos que entre los residentes foráneos.

5. A partir de estas observaciones en posible concluir que el flujo migratorio hacia la comunidad de El Encino está caracterizado por aportar un mayor dinamismo a la economía de la comunidad. Los niveles salariales de los habitantes de origen foráneo, permiten tener mejores niveles de consumo en la comunidad; propiciando con ello una mayor producción de bienes y servicios lo que se traduce en un impacto directo en la generación de empleos. La estructura por edad cuyo promedio de edad es menor que la de los habitantes oriundos, propicia la existencia de una planta laboral cuya fuerza laboral contribuirá por un periodo mayor de tiempo al desarrollo económico de la comunidad.

Actividad 2

2. Calcula la probabilidad de los siguientes eventos

   a) $P(x_1 \geq 65) = 1 - P(x_1 < 65) = 1 - .6828 = .3372$

   b) $P(x_2 \geq 65) = 1 - P(x_2 < 65) = 1 - .9986 = .0014$

   Dado que la probabilidad de que un trabajador de origen foráneo perciba $65 o más es mayor a la probabilidad de que un trabajador oriundo lo haga, entonces existe evidencia que hace suponer que las condiciones salariales de

los individuos de la población económicamente activa que son de origen foráneo son mejores a los originarios de la comunidad . Por lo tanto se deberían promover cursos de capacitación que permitan a los habitantes oriundos de la comunidad capacitarse para acceder a mejores puestos laborales y mejores salarios.

3. Calcula la probabilidad del siguiente evento

$$P(x_3 > 50) = 1 - P(x_3 \leq 50) = 1 - .0271 = .9729$$

El resultado de esta probabilidad nos indica que de cada 100 trabajadores que se dedican a labores que requieren una alta calificación para su desempeño es muy probable que 50 o más sean de origen foráneo. Por lo tanto de deberían promover los cursos de especialización de los trabajadores oriundos con la finalidad de obtener mayores niveles de equidad laboral en la comunidad.

4. Calcula la probabilidad de los siguientes eventos

   a) $P(x_4 = 7) = .0437$

   b) $P(x_5 = 7) = .0034$

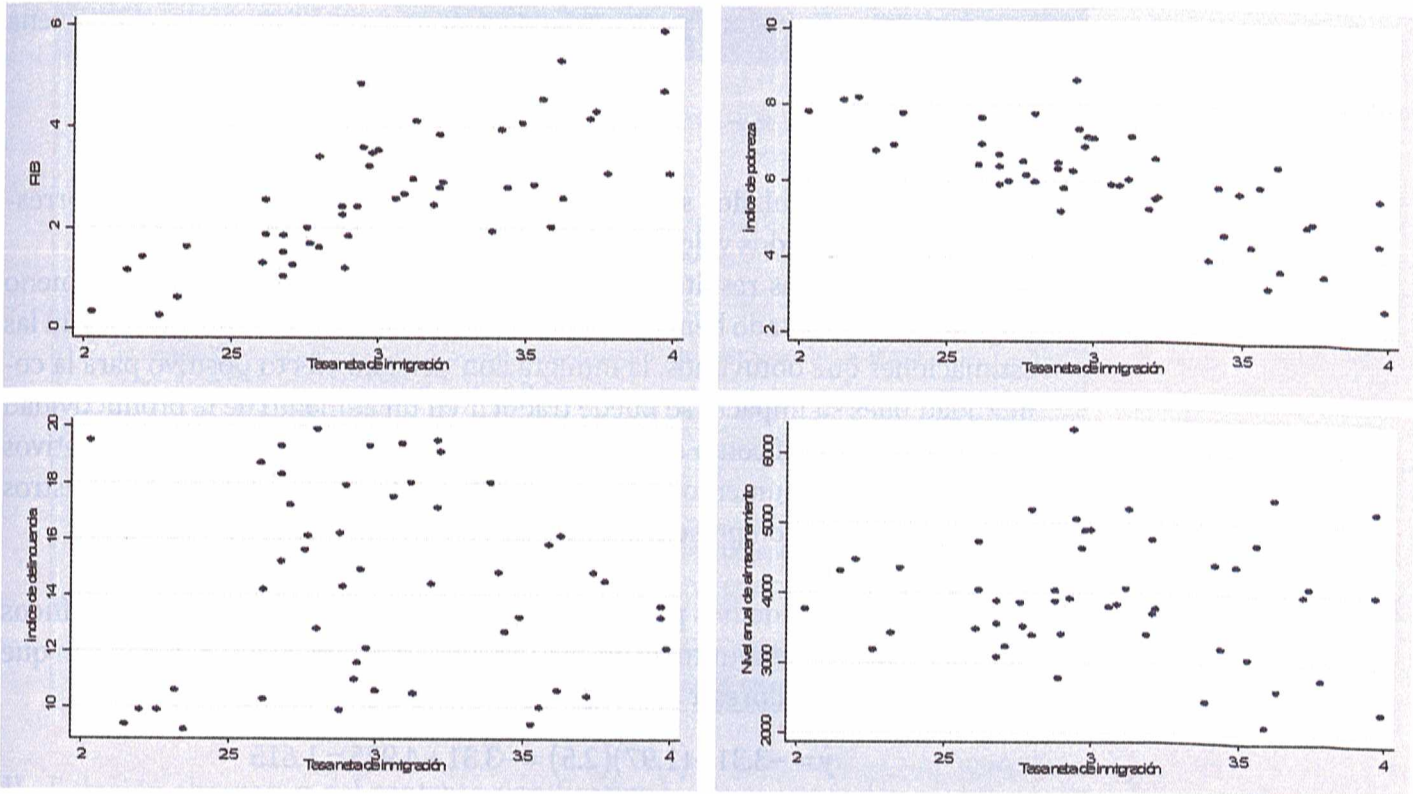
El resultado de estas probabilidades nos indiaca que es más probable que inmigren hacia la comunidad 7 individuos a que emigren fuera de la comunidad 7 individuos. Esta diferencia en las probabilidades marca la naturaleza de El Encino como una comunidad receptora más que expulsora de población.

Actividad 3

Los coeficientes de correlación muestral para la variable tasa de inmigración neta anual contra cada una de las cuatro variables dependientes se muestra en la siguiente tabla:

Tasa neta de inmigración vs PIB	Tasa neta de inmigración vs Índice de pobreza	Tasa neta de inmigración vs Índice de delincuencia	Tasa neta de inmigración vs Nivel anual de almacenamiento
0.75	-0.76	-0.02	-0.09

1. Las gráficas de dispersión correspondientes a cada uno de estos coeficientes de correlación se muestran a continuación:



$$\hat{y}0 - 3.31 + 1.97x$$

2 y 3. A partir de las gráficas de dispersión y de los correspondientes coeficientes de correlación se identifican las siguientes tendencias en los datos:

a) Se identifica la existencia de una tendencia lineal creciente (positiva) entre la variables tasa neta de inmigración anual y PIB, relación que coincide con la tendencia marcada por su correspondiente coeficiente de correlación $r = .75$.
b) Se identifica la existencia de una tendencia lineal decreciente(negativa) entre la variables tasa neta de inmigración anual e índice de pobreza, relación que coincide con la tendencia marcada por su correspondiente coeficiente de correlación $r = -.76$.
c) No se identifica la existencia de algún tipo tendencia entre la variables tasa neta de inmigración anual e índice de delincuencia, situación que coincide con la débil relación de asociación detectada por su correspondiente coeficiente de correlación $r = -.02$.
d) No se identifica la existencia de algún tipo tendencia entre la variables tasa neta de inmigración anual y nivel anual de abastecimiento, situación que coincide con la débil relación de asociación detectada por su correspondiente coeficiente de correlación $r = -.09$.

La línea de regresión ajustada para la tasa de inmigración neta anual y el PIB tiene la siguiente ecuación:

$$\hat{y}0-3.31+1.97x$$

4. Cuyos valores calculados son los valores pronosticados del PIB anual correspondientes a estos dos valores de la tasa de inmigración neta anual.

   Con base en los resultados obtenidos se puede concluir que el fenómeno migratorio ha resultado benéfico para la comunidad de El Encino, pues según las estimaciones que obtuvimos, la inmigración tiene un efecto positivo para la comunidad pues su impacto se puede traducir en un aumento de la productividad en la comunidad así como en una disminución en la pobreza. Efectos negativos como podrían ser aumento en la criminalidad o disminución en los suministros de agua no estuvieron presentes como efectos directos de la inmigración.

5. Por lo que si pretendemos pronosticar el PIB anual para dos valores específicos que no están en el conjunto de observaciones como son 2.5 y 3.6. Basta con que sustituyamos los valores $x = 2.5$ y $x = 3.6$ de la siguiente manera

$$\hat{y} = -3.31+(1.97)(2.5) = -3.31+4.925 = 1.615$$

$$\hat{y} = -3.31+1.97x = -3.31+(1.97)(3.6) = -3.31+7.092 = 3.782$$

Evalúa tu aprendizaje

I. Relaciona las dos columnas de tal forma que, delante de cada enunciado de la columna A escribas el número de la columna B que complete correctamente el enunciado.

Columna A	
Se característica por no ser afectada por valores extremos	(11)
Es una característica del coeficiente de correlación muestral	(10)
Característica de la línea de regresión	(12)
Definición de medida de tendencia central	(7)
Desventaja del uso de la media muestral como medida de tendencia central	(4)
Son ejemplos de medidas de tendencia central	(1)
Desventaja del uso del rango como medida de dispersión	(13)
Son ejemplos de medidas de dispersión	(6)
Definición de medida de dispersión	(2)
Es un ejemplo de medida de tendencia central para el caso de variables categóricas	(8)
Desventaja del uso de la desviación media como medida de dispersión	(9)
Es la medida de tendencia central que se define como el dato que más se repite en una muestra	(3)
Es una desventaja del uso del coeficiente de correlación muestral	(5)

Columna B
1. Media, mediana y moda
2. Cantidad que sirve para cuantificar que tanto las observaciones de una muestra difieren unas de otras.
3. Moda
4. Tiene la desventaja de que su valor es afectado por la presencia de valores extremos.
5. Tiene la desventaja de no poder detectar relaciones de asociación no lineales entre dos variables.
6. Rango, varianza y desviación estándar.
7. Se define como el número que describe aproximadamente donde están localizados o centrados los datos a lo largo de la recta numérica.
8. Proporción muestral.
9. Tiene la desventaja de que su suma anula contribuciones con magnitudes absolutas iguales pero con distinto signo.
10. Su valor siempre varía entre -1 y 1.
11. Es una característica de la mediana como medida de tendencia central.
12. Tiene la característica de ser una línea recta que está lo más cerca posible de los puntos descritos en una gráfica de dispersión.
13. Tiene la desventaja de no considerar para su cálculo todos los datos lo que puede llegar a ocultar información sobre la verdadera naturaleza de variación de un conjunto de datos.

II. Solución alternativa del contexto significativo.
Los cálculos correspondientes a las medidas de tendencia central y de dispersión para residentes oriundos y de origen foráneo son presentados en las siguientes tablas:

Residentes Oriundos				
Variable	**Media**	**Mediana**	**Desviación Estándar**	**Rango**
Nivel de Ingreso	57.79	58	7.50	34
Edad	28.9	28	5.69	28

Residentes de Origen Foráneo				
Variable	**Media**	**Mediana**	**Desviación Estándar**	**Rango**
Nivel de Ingreso	51.86	52	6.07	26
Edad	24.62	25	2.16	9

Al realizar una inspección de estas tablas y realizar un análisis comparativo de estas cantidades, podemos observar que tanto la media como la mediana del ingreso para los residentes oriundos es mayor que para los residentes de origen foráneo. Así mismo, las medidas de dispersión muestran una mayor variabilidad en el ingreso para los residentes oriundos que para los residentes de origen foráneo. Lo que

puede hacer que nos planteemos la hipótesis en el sentido de que esta mayor variabilidad puede ser explicada por una mayor heterogeneidad entre los niveles de especialización y capacitación entre los residentes oriundos, lo que les permite incursionar a una mayor diversidad de áreas del sector productivo y por lo tanto acceder a una mayor diversidad de escalas salariales.

A partir del análisis comparativo correspondiente a la edad, se puede advertir que el promedio de edad así como la variabilidad de ésta es menor para los residentes de origen foráneo; lo que muestra una estructura por edades menor entre los miembros de la población económicamente activa de origen foráneo que entre los oriundos de la comunidad. Dicha diferencia en las edades puede explicar en gran medida la diferencia entre los niveles salariales, ya que es posible pensar que por ser más jóvenes, los habitantes de origen foráneo tienen una instrucción menor lo que los hace tener menores posibilidades de acceder a mejores puestos laborales y por lo tanto a mejores condiciones salariales.

Ahora centremos nuestro interésen dar interpretación a los resultados de las probabilidades $P(x_1 \geq 65) = .0151$ y $P(x_2 \geq 65) = .0846$ .Estos valores indican que es más probable (aunque no demasiado) encontrar individuos con niveles salariales mayores a $65 diarios (niveles salariales considerados de medios a buenos) entre residentes oriundos que entre residentes de origen foraneo de la comunidad. Por otra parte, el valor calculado de la probabilidad $P(x_3 \geq 50) = 5.17419 \times 10^{-12}$ muestra que la probabilidad de encontrar más de 50 trabajadores de origen foráneo por cada 100 trabajadores altamente calificados es un evento extremadamente raro. Si complementamos estos resultados con los obtenidos del análisis comparativo, entonces a partir de los datos muestrales y de las probabilidades calculadas obtenemos la evidencia necesaria para afirmar que existe una marcada desigualdad en el ingreso así como en los niveles de especialización entre los trabajadores de origen foráneo y los trabajadores oriundos de la comunidad. Por lo que una buena medida para disminuir este desequilibrio, sería generar programas de capacitación para los trabajadores de origen foráneo, con la finalidad de que estos puedan especializarse para acceder a mejores condiciones salariales. La urgencia de tomar esta medida viene sustentada por el hecho de la intensidad de la inmigración hacia la comunidad de El Encino es mayor que la de la migración, pues como lo indican las siguientes cálculos probabilísticos, $P(x_4 = 7) = 0.0710$ y $P(x_5 = 7) = .0181$, es más probable que ingresen 7 personas a la comunidad a que este mismo número que decida salir de ella.

Con base en estas mediciones hay elementos para suponer que el fenómeno migratorio no está teniendo un efecto económico benéfico en la comunidad de El Encino. Para corroborar esto, realizamos el cálculo del coeficiente de correlación para los datos correspondientes al conjunto de 50 observaciones pareadas anuales (ver apéndice 2) correspondientes a los registros comunitarios de los últimos 50 años del producto interno bruto anual, así como de la tasa neta de inmigración. Al

realizar el cálculo del coeficiente de correlación muestral se obtuvo un valor calculado de $r = -.02$, lo que nos otorga evidencia de la existencia de una relación débil entre la tasa neta de inmigración anual y el producto interno bruto, corroborando con ello, que el fenómeno migratorio no ha tenido un efecto económico benéfico para la comunidad.

El cálculo de los coeficientes de correlación para cada uno de los tres conjuntos de datos pareados conformados por la tasa neta de inmigración y los correspondientes valores observados de las variables, índice de pobreza humana, índice de delincuencia así como nivel medio anual de almacenamiento de agua de la presa El Encino se obtuvieron los valores, .87, .92 y .75 respectivamente. A partir de estos valores se puede concluir que el fenómeno migratorio tuvo un efecto benéfico en el abatimiento de los niveles de pobreza, pues el coeficiente de correlación de .87 muestra que en el periodo considerado el impacto de la inmigración generó una tendencia creciente en el indicador de los niveles de pobreza.

Por último, los valores .92 y .75 correspondientes a los indicadores delictivos y de nivel de abastecimiento de agua respectivamente, muestran una asociación fuerte con la tasa neta anual de inmigración, lo que hace suponer que este **factor** es determinante en el comportamiento de estas variables, pues ambos coeficientes muestran que los valores de estas variables presentan una tendencia creciente ante la presencia de incrementos en la tasa neta anual de inmigración.

A partir de los resultados expuestos, concluimos que el fenómeno migratorio, más particularmente la inmigración no se ha traducido en beneficios para el desarrollo de la comunidad de El Encino, pues hay evidencia que nos permite vincular este fenómeno con el aumento en los índices de pobreza humana y de delincuencia, así como con el abastecimiento de agua en la comunidad.

¿Ya estoy preparado?

1. a) 1, 2, 5, 8 y 10

Respuesta Incorrecta
Los fenómenos considerados en la respuesta surgen como consecuencia de la intervención directa del hombre; por lo que no pueden ser considerados como fenómenos naturales. Recuerda que la definición de fenómeno natural hace énfasis en el hecho de que para que un fenómeno sea considerado natural éste no debió ser propiciado por la intervención directa del hombre. Para aclarar tus dudas te recomiendo repases la sección 1.1 de la Unidad 1.

b) 3, 6, 7, 11 y 12

Respuesta correcta
Los fenómenos considerados en esta respuesta no surgen como consecuencia de la intervención directa del hombre; por lo que pueden ser considerados como fenómenos naturales. Recuerda que la definición de fenómeno natural hace énfasis en esto. Si todavía tienes alguna duda, te recomiendo repases la sección 1.1 de la Unidad 1.

c) 1, 2, 8, 9 y 10

Respuesta incorrecta
La marginación, desarrollo económico y comercio son fenómenos propiciados por la intervención directa del hombre, por lo que no se les puede considerar como fenómenos naturales. Recuerda que la definición de fenómeno natural hace énfasis en el hecho de que para que un fenómeno sea considerado natural éste no debió ser propiciado por la intervención directa del hombre. Para aclarar tus dudas te recomiendo repases la sección 1.1 de la Unidad 1.

d) 4, 5, 8, 11 y 12

Respuesta incorrecta
Los emigración, inmigración y marginación, son fenómenos propiciados por la intervención directa del hombre, por lo que no se les puede considerar como fenómenos naturales. Recuerda que la definición de fenómeno natural hace énfasis en el hecho de que para que un fenómeno sea considerado natural éste no debió ser propiciado por la intervención directa del hombre. Para aclarar tus dudas te recomiendo repases la sección 1.1 de la Unidad 1.

2. a) 1, 2, 4, 5, 8, 9 y 10

Respuesta correcta
Los fenómenos considerados en esta respuesta surgen como consecuencia de la intervención directa del hombre; por lo que pueden ser considerados como procesos sociales. Recuerda que la definición de proceso social hace énfasis en este aspecto. Si todavía tienes alguna duda, te recomiendo repases la sección 1.1 de la Unidad 1.

b) 3, 6, 7, 8, 10, 11 y 12

Respuesta incorrecta
Los huracanes, tormentas tropicales, tsunamis, inundaciones y epidemias no son propiciados por la intervención directa del hombre, por lo que no se les puede considerar como procesos sociales. Recuerda que la definición de proceso social hace énfasis en el hecho de que su ocurrencia se da como resultado de la intervención directa del hombre. Para aclarar tus dudas te recomiendo repases la sección 1.1 de la Unidad 1.

c) 2, 3, 4, 8, 10, 11 y 12

Respuesta incorrecta
Los huracanes, inundaciones y epidemias no son propiciados por la intervención directa del hombre, por lo que no se les puede considerar como procesos sociales. Recuerda que la definición de proceso social hace énfasis en el hecho de que su ocurrencia se da como resultado de la intervención directa del hombre. Para aclarar tus dudas te recomiendo repases la sección 1.1 de la Unidad 1.

d) 1, 4, 5, 6, 7, 9 y 11

Respuesta incorrecta
Las tormentas tropicales, tsunamis e inundaciones no son propiciados por la intervención directa del hombre, por lo que no se les puede considerar como procesos sociales. Recuerda que la definición de proceso social hace énfasis en el hecho de que su ocurrencia se da como resultado de la intervención directa del hombre. Para aclarar tus dudas te recomiendo repases la sección 1.1 de la Unidad 1.

3. a) 2, 4 y 6

Respuesta incorrecta
La ocurrencia de los eventos considerados en esta respuesta no puede ser pronosticada con exactitud; por está razón no pueden ser considerados

como eventos deterministicos. Recuerda que por su naturaleza un evento determinístico puede ser pronosticado con exactitud. Para que puedas aclarar tus dudas sobre el tema te recomiendo des un repaso a la sección 1.2 de la Unidad 1.

b) 4, 7 y 9

Respuesta incorrecta.
Los niveles de ocupación hotelera anual así como el número de nacimientos anuales no puede ser pronosticada con exactitud; por está razón no pueden ser considerados como eventos determinísticos. Recuerda que por su naturaleza un evento determinístico puede ser pronosticado con exactitud. Para que puedas aclarar tus dudas sobre el tema te recomiendo des un repaso a la sección 1.2 de la Unidad 1.

c) 3, 6 y 7

Respuesta incorrecta
El número de defunciones y de nacimientos anuales no puede ser pronosticada con exactitud; por está razón no pueden ser considerados como eventos determinísticos. Recuerda que por su naturaleza un evento determinístico puede ser pronosticado con exactitud. Para que puedas aclarar tus dudas sobre el tema te recomiendo des un repaso a la sección 1.2 de la Unidad 1.

d) 1, 3 y 9

Respuesta correcta
La ocurrencia de los eventos considerados en esta respuesta si puede ser pronosticada con exactitud; por está razón estos eventos pueden ser considerados como eventos deterministicos. Recuerda que por su naturaleza un evento determinístico cumple esta característica. Si todavía tienes alguna duda, te recomiendo repases la sección 1.2 de la Unidad 1.

4. a) 1, 2, 3, 4, 5 y 6

Respuesta incorrecta.
El día y la noche así como las estaciones del año son eventos que por su naturaleza de ocurrencia (se tiene la certeza de que siempre serán observados así como del momento en que ocurrirán) pueden ser pronosticarlos con exactitud, por lo que no se pueden considerar como eventos aleatorio. Recuerda que un evento aleatorio está caracterizado por la incertidumbre asociada con su ocurrencia lo que hace que no se puede pronosticar con

exactitud. Para que puedas aclarar tus dudas sobre el tema te recomiendo des un repaso a la sección 1.2 de la Unidad 1.

b) 3, 5, 6, 7, 8 y 9

Respuesta incorrecta.
Las estaciones del año son eventos que por su naturaleza de ocurrencia (se tiene la certeza de que siempre serán observadas así como del momento en que ocurrirán) pueden ser pronosticarlos con exactitud; por otra parte las percepciones mensuales de un trabajador asalariado por lo regular son fijas, circunstancia que las hace poder ser pronosticadas con exactitud. Por estas razones ambos eventos no pueden ser considerados como eventos aleatorios. Recuerda que un evento aleatorio está caracterizado por la incertidumbre asociada con su ocurrencia lo que hace que no se puede pronosticar con exactitud. Para que puedas aclarar tus dudas sobre el tema te recomiendo des un repaso a la sección 1.2 de la Unidad 1.

c) 2, 4, 5, 6, 7 y 8

Respuesta correcta.
La naturaleza incierta en la ocurrencia de los eventos considerados en esta respuesta los hace difíciles de pronosticar con exactitud. Es por ello que estos eventos pueden ser considerados como aleatorios. Recuerda que un evento aleatorio está caracterizado por la incertidumbre asociada con su ocurrencia lo que hace que no se puede pronosticar con exactitud. Si tienes alguna duda al respecto de este tema, te sugiero revises la sección 1.2 de la Unidad 1.

d) 1, 3, 6, 7, 8 y 9

Respuesta incorrecta.
El día y la noche así como las estaciones del año son eventos que por su naturaleza de ocurrencia (se tiene la certeza de que siempre serán observados así como del momento en que ocurrirán) pueden ser pronosticarlos con exactitud; por otra parte las percepciones mensuales de un trabajador asalariado por lo regular son fijas, circunstancia que las hace poder ser pronosticadas con exactitud. Por estas razones estos eventos no pueden ser considerados como eventos aleatorios. Recuerda que un evento aleatorio está caracterizado por la incertidumbre asociada con su ocurrencia lo que hace que no se puede pronosticar con exactitud. Para que puedas aclarar tus dudas sobre el tema te recomiendo des un repaso a la sección 1.2 de la Unidad 1

5. a) 1a; 2b; 3c; 4d

Respuesta incorrecta.
Las variables *b*, *c* y *d* están mal clasificadas. La variable *b* es categórica ordinal, la variable *c* es categórica nominal y la variable *d* es numérica discreta. Para que aclares tus dudas respecto a los criterios de clasificación, es importante revises la sección 1.3 de la Unidad 1

b) 1a; 2d; 3b; 4c

Respuesta correcta.
El criterio de clasificación es el adecuado. Si aún tienes alguna duda sobre los criterios de clasificación, es importante revises la sección 1.3 de la Unidad 1 Para que aclares tus dudas respecto a los criterios de clasificación, es importante revises la sección 1.3 de la Unidad 1

c) 1c; 2d; 3a; 4c

Respuesta incorrecta.
Las variables *a* y *c* están mal clasificadas. La variable *a* es numérica continua y la variable *c* es categórica nominal.

d) 1b; 2c; 3a; 4d

Respuesta incorrecta.
Las variables *a*, *b*, *c* y *d* están mal clasificadas. La variable *a* es numérica continua, la variable *b* es categórica ordinal, la variable *c* es categórica nominal y la variable *d* es numérica discreta. Para que aclares tus dudas respecto a los criterios de clasificación, es importante revises la sección 1.3 de la Unidad 1

6. a) 1a; 2b; 3c; 4d; 5e

Respuesta incorrecta.
No tienes claras las definiciones. Es importante que revises la sección 1.4 de la Unidad 1 para que repases las definiciones planteadas.

b) 1c; 2e; 3a; 4d; 5b

Respuesta incorrecta.
No tienes claras las definiciones. Es importante que revises la sección 1.4 de la Unidad 1 para que repases las definiciones planteadas.

c) 1c; 2d; 3a; 4b; 5e

Respuesta correcta.
Identificaste adecuadamente los conceptos con sus definiciones respectivas.

d) 1e; 2d; 3c; 4b; 5a

Respuesta incorrecta.
No tienes claras las definiciones. Es importante que revises la sección 1.4 de la Unidad 1 para que repases las definiciones planteadas.

7. a) 4

Respuesta incorrecta.
Recuerda que un histograma para representar adecuadamente una distribución de probabilidad normal debe tener forma simétrica y de campana. Para que aclares tus dudas con respecto a este tema te recomiendo revises la sección 5 de la Unidad 1.

b) 2

Respuesta incorrecta.
Recuerda que un histograma para representar adecuadamente una distribución de probabilidad normal debe tener forma simétrica y de campana. Para que aclares tus dudas con respecto a este tema te recomiendo revises la sección 5 de la Unidad 1.

c) 1

Respuesta correcta.
Este es el histograma que por su forma simétrica y de campana puede representar mejor a una distribución de probabilidad normal.

d) 3

Respuesta incorrecta.
Recuerda que un histograma para representar adecuadamente una distribución de probabilidad normal debe tener forma simétrica y de campana. Para que aclares tus dudas con respecto a este tema te recomiendo revises la sección 5 de la Unidad 1.

8. a) 1; 3; 5; 6

Respuesta correcta.
Identificaste adecuadamente los supuestos de un experimento binomial.

b) 2; 4; 5; 6

Respuesta incorrecta
No seleccionaste adecuadamente los supuestos. Para aclarar tus dudas respecto del tema te recomiendo revises la sección 1.5 de la Unidad 1.

c) 1; 2; 3; 4

Respuesta incorrecta
No seleccionaste adecuadamente los supuestos. Para aclarar tus dudas respecto del tema te recomiendo revises la sección 1.5 de la Unidad 1.

d) 1; 3; 4; 6

Respuesta incorrecta
No seleccionaste adecuadamente los supuestos. Para aclarar tus dudas respecto del tema te recomiendo revises la sección 1.5 de la Unidad 1.

9. a) 1a; 2b; 3c; 4d; 5e

Respuesta incorrecta.
No tienes claras las definiciones. Es importante que revises la sección 2.1 de la Unidad 2 para que repases las definiciones planteadas.

b) 1c; 2e; 3a; 4d; 5b

Respuesta incorrecta.
No tienes claras las definiciones. Es importante que revises la sección 2.1 de la Unidad 2 para que repases las definiciones planteadas.

c) 1c; 2d; 3a; 4b; 5e

Respuesta incorrecta.
No tienes claras las definiciones. Es importante que revises la sección 2.1 de la Unidad 2 para que repases las definiciones planteadas.

d) 1b; 2c; 3a; 4d; 5e

Respuesta correcta.
Identificaste adecuadamente los conceptos con sus definiciones respectivas.

10. a) La medida de tendencia central más adecuada es la media muestral.

Respuesta incorrecta.
No te ha quedado claro la interpretación de la media y la mediana como medidas de tendencia central. Te sugiero repases la sección 2.1 de la Unidad 2 para que despejes todas tus dudas sobre el tema.

b) La medida de tendencia central más adecuada es la mediana muestral.

Respuesta correcta.
Al inspeccionar el histograma de frecuencia se puede observar que el valor de la mediana de $113 se encuentra más hacia el centro de la distribución de los datos que el de la media de $207 que está más alejado. Por esta razón es más adecuado utilizar la mediana como medida representativa de la información disponible.

c) Es indistinto, cualquiera de las dos medidas puede ser tomada como representativa de la información.

Respuesta incorrecta.
No te ha quedado claro la interpretación de la media y la mediana como medidas de tendencia central. Te sugiero repases la sección 2.1 de la Unidad 2 para que despejes todas tus dudas sobre el tema.

d) Ninguna de las dos medidas.

Respuesta incorrecta.
No te ha quedado claro la interpretación de la media y la mediana como medidas de tendencia central. Te sugiero repases la sección 2.1 de la Unidad 2 para que despejes todas tus dudas sobre el tema.

11. a) 1e; 2d; 3a; 4c; 5b

Respuesta correcta.
Identificaste adecuadamente los conceptos con sus definiciones respectivas.

b) 1c; 2e; 3a; 4d; 5b

Respuesta incorrecta.
No tienes claras las definiciones. Es importante que revises la sección 2.1 de la Unidad 2 para que repases las definiciones planteadas.

c) 1c; 2d; 3a; 4b; 5e

Respuesta incorrecta.
No tienes claras las definiciones. Es importante que revises la sección 2.1 de la Unidad 2 para que repases las definiciones planteadas.

d) 1b; 2c; 3a; 4d; 5e

Respuesta incorrecta.
No tienes claras las definiciones. Es importante que revises la sección 2.1 de la Unidad 2 para que repases las definiciones planteadas.

12. a) A partir de los valores calculados para la desviación estándar de ambas muestras se puede concluir que existe evidencia que nos hace suponer que entre los hombres existe mayor variabilidad en sus ingresos diarios que entre las mujeres.

Respuesta correcta.
El valor calculado para la desviación estándar del ingreso diario de mujeres y hombres fue de 7.6 y 17.5 pesos respectivamente. Estas cantidades nos llevan a concluir que existe evidencia para suponer que en la región considerada hay mayor variabilidad entre los ingresos diarios de los hombres que entre los ingresos diarios de las mujeres.

b) A partir de los valores calculados para la desviación estándar de ambas muestras se puede concluir que existe evidencia que nos hace suponer que entre las mujeres existe mayor variabilidad en sus ingresos diarias que entre los hombres.

Respuesta incorrecta.
El valor calculado para la desviación estándar del ingreso diario de mujeres y hombres fue de 7.6 y 17.5 pesos respectivamente. En caso de que hayas obtenido valores distintos debes revisar las fórmulas o tus procedimientos de cálculo. Si obtuviste estos valores entonces no te ha quedado claro como interpretar los resultados, por lo que te recomiendo revises la sección 2.1 de la Unidad 2 para que aclares todas las dudas que tengas con respecto a este tema.

c) A partir de los valores calculados para la desviación estándar de ambas muestras no se puede obtener algún tipo de conclusión.

Respuesta incorrecta.
El valor calculado para la desviación estándar del ingreso diario de mujeres y hombres fue de 7.6 y 17.5 pesos respectivamente. En caso de que hayas obtenido valores distintos debes revisar las fórmulas o tus procedimientos de cálculo. Si obtuviste estos valores entonces no te ha quedado claro como interpretar los resultados, por lo que te recomiendo revises la sección 2.1 de la Unidad 2 para que aclares todas las dudas que tengas con respecto a este tema.

d) A partir de los valores calculados para la desviación estándar de ambas muestras se puede concluir que existe evidencia que nos hace suponer quela variabilidad en las percepciones es la misma para hombres y mujeres.

Respuesta incorrecta.
El valor calculado para la desviación estándar del ingreso diario de mujeres y hombres fue de 7.6 y 17.5 pesos respectivamente. En caso de que hayas obtenido valores distintos debes revisar las fórmulas o tus procedimientos de cálculo. Si obtuviste estos valores entonces no te ha quedado claro como interpretar los resultados, por lo que te recomiendo revises la sección 2.1 de la Unidad 2 para que aclares todas las dudas que tengas con respecto a este tema.

13. a) .1587

Respuesta incorrecta.
No calculaste de forma adecuada la probabilidad de este evento. Es necesario que repases las técnicas para calcular las probabilidades de eventos asociados con una distribución normal que fueron descritas en la sección 2.2 de la Unidad 2.

b) .3174

Respuesta incorrecta.
No calculaste de forma adecuada la probabilidad de este evento. Es necesario que repases las técnicas para calcular las probabilidades de eventos asociados con una distribución normal que fueron descritas en la sección 2.2 de la Unidad 2.

c) .6826

Respuesta correcta.
Calculaste correctamente la probabilidad del evento considerado.

d) .5000

Respuesta incorrecta.
No calculaste de forma adecuada la probabilidad de este evento. Es necesario que repases las técnicas para calcular las probabilidades de eventos asociados con una distribución normal que fueron descritas en la sección 2.2 de la Unidad 2.

14. a) .9976

Respuesta incorrecta.
No calculaste de forma adecuada la probabilidad de este evento. Es necesario que repases las técnicas para calcular las probabilidades de eventos aso-

ciados con una distribución binomial que fueron descritas en la sección 2.2 de la Unidad 2.

b) .0032

Respuesta incorrecta.
No calculaste de forma adecuada la probabilidad de este evento. Es necesario que repases las técnicas para calcular las probabilidades de eventos asociados con una distribución binomial que fueron descritas en la sección 2.2 de la Unidad 2.

c) .9968

Respuesta incorrecta.
No calculaste de forma adecuada la probabilidad de este evento. Es necesario que repases las técnicas para calcular las probabilidades de eventos asociados con una distribución binomial que fueron descritas en la sección 2.2 de la Unidad 2.

d) .0024

Respuesta correcta.
Calculaste correctamente la probabilidad del evento considerado.

15. a) 1; 3; 5

Respuesta correcta.
Identificaste adecuadamente los supuestos de un experimento de Poisson.

b) 2; 4; 5

Respuesta incorrecta
No seleccionaste adecuadamente los supuestos. Para aclarar tus dudas respecto del tema te recomiendo revises la sección 2.2 de la Unidad 2.

c) 1; 2; 3

Respuesta incorrecta
No seleccionaste adecuadamente los supuestos. Para aclarar tus dudas respecto del tema te recomiendo revises la sección 2.2 de la Unidad 2.

d) 1; 3; 4

Respuesta incorrecta
No seleccionaste adecuadamente los supuestos. Para aclarar tus dudas respecto del tema te recomiendo revises la sección 2.2 de la Unidad 2.

16. a) .0829

Respuesta incorrecta.
No calculaste de forma adecuada la probabilidad de este evento. Es necesario que repases las técnicas para calcular las probabilidades de eventos asociados con una distribución de Poisson que fueron descritas en la sección 2.2 de la Unidad 2.

b) .1785

Respuesta incorrecta.
No calculaste de forma adecuada la probabilidad de este evento. Es necesario que repases las técnicas para calcular las probabilidades de eventos asociados con una distribución de Poisson que fueron descritas en la sección 2.2 de la Unidad 2.

c) .2809

Respuesta correcta.
Calculaste correctamente la probabilidad del evento considerado.

d) .0024

Respuesta incorrecta.
No calculaste de forma adecuada la probabilidad de este evento. Es necesario que repases las técnicas para calcular las probabilidades de eventos asociados con una distribución de Poisson que fueron descritas en la sección 2.2 de la Unidad 2.

17. a) 1a; 2b; 3c; 4d; 5e; 6f

Respuesta incorrecta.
No tienes claras las definiciones. Es importante que revises la sección 1.5 de la Unidad 1 y la sección 2.3 de la Unidad 2 para que repases las definiciones planteadas.

b) 1e; 2c; 3f; 4a; 5b; 6d

Respuesta correcta.
Identificaste adecuadamente los conceptos con sus definiciones respectivas.

c) 1e; 2c; 3a; 4b; 5d; 6f

Respuesta incorrecta.
No tienes claras las definiciones. Es importante que revises la sección 1.5 de la Unidad 1 y la sección 2.3 de la Unidad 2 para que repases las definiciones planteadas.

d) 1b; 2c; 3a; 4d; 5e;6f

Respuesta incorrecta.
No tienes claras las definiciones. Es importante que revises la sección 1.5 de la Unidad 1 y la sección 2.3 de la Unidad 2 para que repases las definiciones planteadas.

18. a) 1;b

Respuesta incorrecta.
Tienes problemas para interpretar adecuadamente la gráfica de dispersión y el coeficiente de correlación muestral. Para que aclares tus dudas con respecto a estos temas, te sugiero revises la sección 1.4 de la Unidad 1 así como la sección 2.3 de la Unidad 2.

b) 2;b

Respuesta incorrecta.
Tienes problemas para interpretar adecuadamente la gráfica de dispersión y el coeficiente de correlación muestral. Para que aclares tus dudas con respecto a estos temas, te sugiero revises la sección 1.4 de la Unidad 1 así como la sección 2.3 de la Unidad 2.

c) 1;a

Respuesta incorrecta.
Tienes problemas para interpretar adecuadamente la gráfica de dispersión y el coeficiente de correlación muestral. Para que aclares tus dudas con respecto a estos temas, te sugiero revises la sección 1.4 de la Unidad 1 así como la sección 2.3 de la Unidad 2.

d) 2;a

Respuesta correcta.
Sabes interpretar adecuadamente la gráfica de dispersión y el coeficiente de correlación muestral.

19. a) 1.

Respuesta correcta.
Seleccionaste la respuesta correcta. Sabes identificar la tendencia creciente y una gráfica de dispersión y en una línea.

b) 3.
Respuesta incorrecta.

No sabes distinguir la tendencia creciente en una gráfica de dispersión. Para aclarar tus dudas en relación con este tema te sugiero que revises la sección 1.4 de la Unidad 1 así como la sección 2.3 de la Unidad 2.

c) 2.

Respuesta incorrecta.
No sabes distinguir la tendencia creciente en una gráfica de dispersión. Para aclarar tus dudas en relación con este tema te sugiero que revises la sección 1.4 de la Unidad 1 así como la sección 2.3 de la Unidad 2.

d) Ninguna de las anteriores.

Respuesta incorrecta.
No sabes distinguir la tendencia creciente en una gráfica de dispersión. Para aclarar tus dudas en relación con este tema te sugiero que revises la sección 1.4 de la Unidad 1 así como la sección 2.3 de la Unidad 2.

20. a) 3

Respuesta incorrecta.
No sabes distinguir la tendencia decreciente no lineal en una gráfica de dispersión. Para aclarar tus dudas en relación con este tema te sugiero que revises la sección 2.3 de la Unidad 2.

b) 2

Respuesta incorrecta.
No sabes distinguir la tendencia decreciente no lineal en una gráfica de dispersión. Para aclarar tus dudas en relación con este tema te sugiero que revises la sección 2.3 de la Unidad 2.

c) 1

Respuesta correcta.
Seleccionaste la respuesta correcta. Sabes identificar la tendencia decreciente no lineal en una gráfica de dispersión.

d) 4

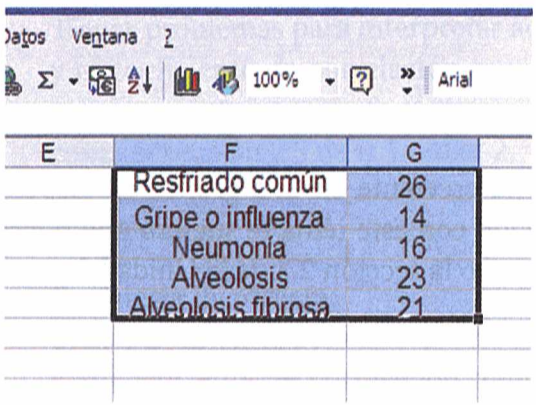
Respuesta incorrecta.
No sabes distinguir la tendencia decreciente no lineal en una gráfica de dispersión. Para aclarar tus dudas en relación con este tema te sugiero que revises la última sección de la Unidad 2.

Guía de uso de una hoja de cálculo para la aplicación de herramientas estadísticas

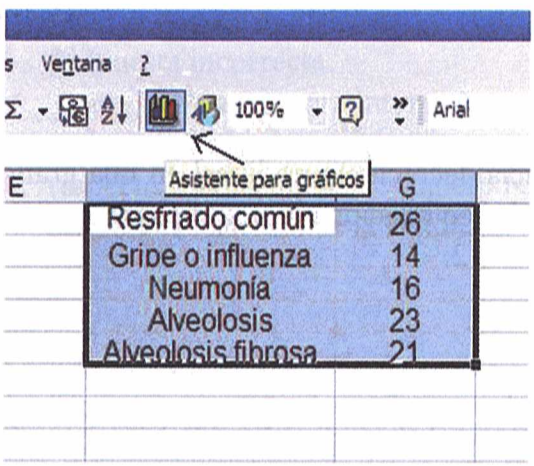
1. Gráficas con Excel

Para realizar gráficas en Excel debes seguir los siguientes pasos.

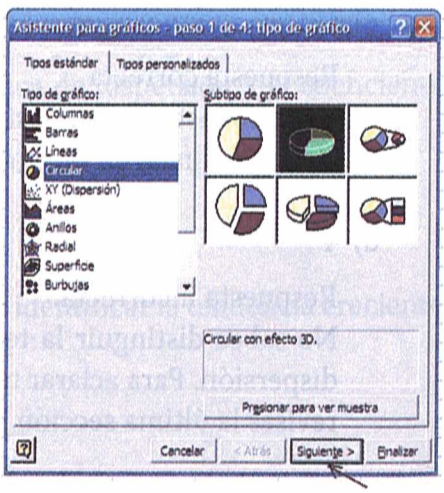
1. Selecciona el conjunto de datos que deseas graficar.



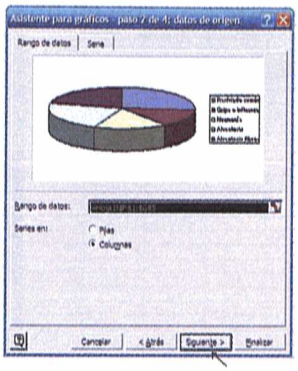
2. Elige la opción asistente para gráficos.



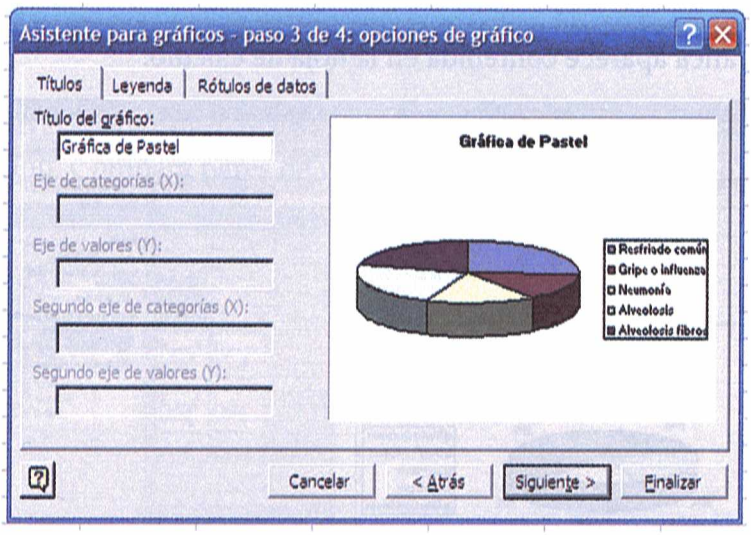
3. En el cuadro de diálogo escoge la opción circular y la opción circular con efecto 3D. Da un clic izquierdo en finalizar.



4. En el cuadro de diálogo selecciona la opción columnas y da un clic izquierdo en finalizar.



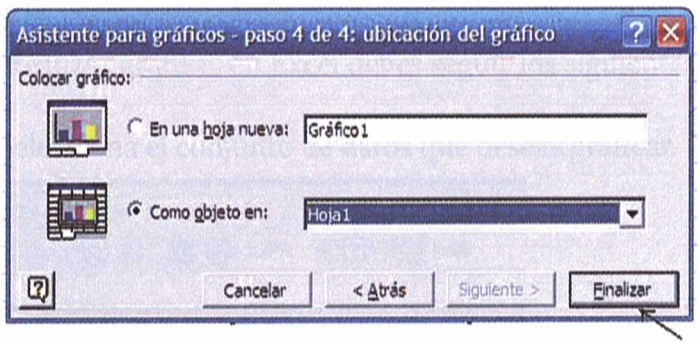
5. En el cuadro de diálogo escribe le leyenda Gráfica de Pastel.



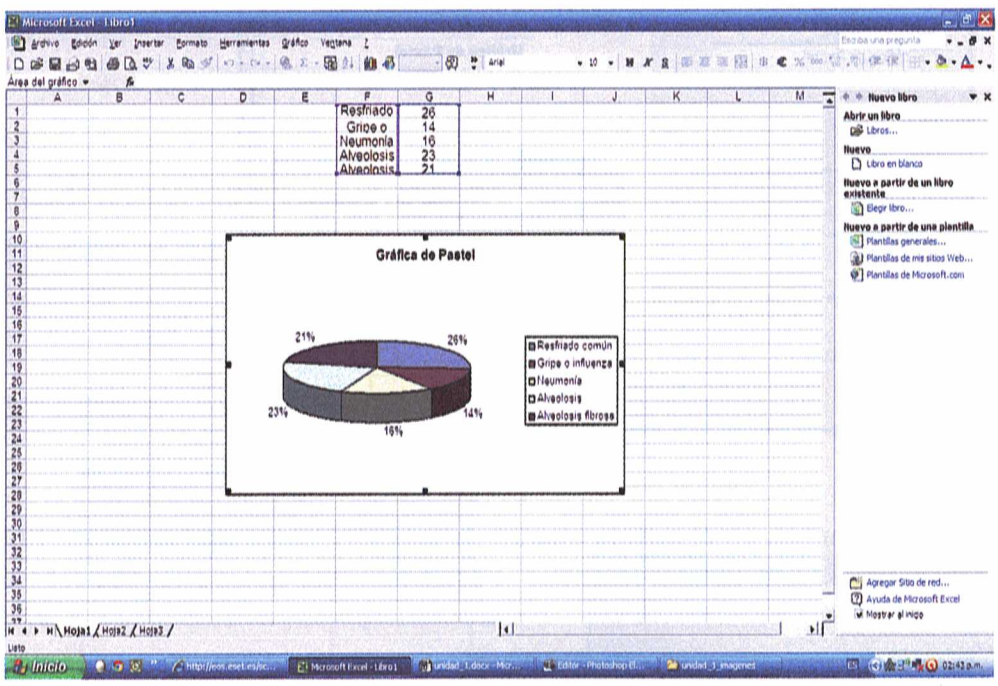
6. Posteriormente, en el mismo cuadro de diálogo selecciona la opción "Rótulos de datos" y selecciona "porcentaje". Posteriormente da un clic izquierdo en siguiente.



7. En el cuadro de diálogo selecciona la opción "Como objeto en" si prefieres que tu gráfico aparezca en tu hoja de cálculo. En caso contrario puedes elegir la opción "En una hoja nueva" si prefieres que tu gráfico esté ubicado en una hoja aparte. Posteriormente da clic en finalizar.



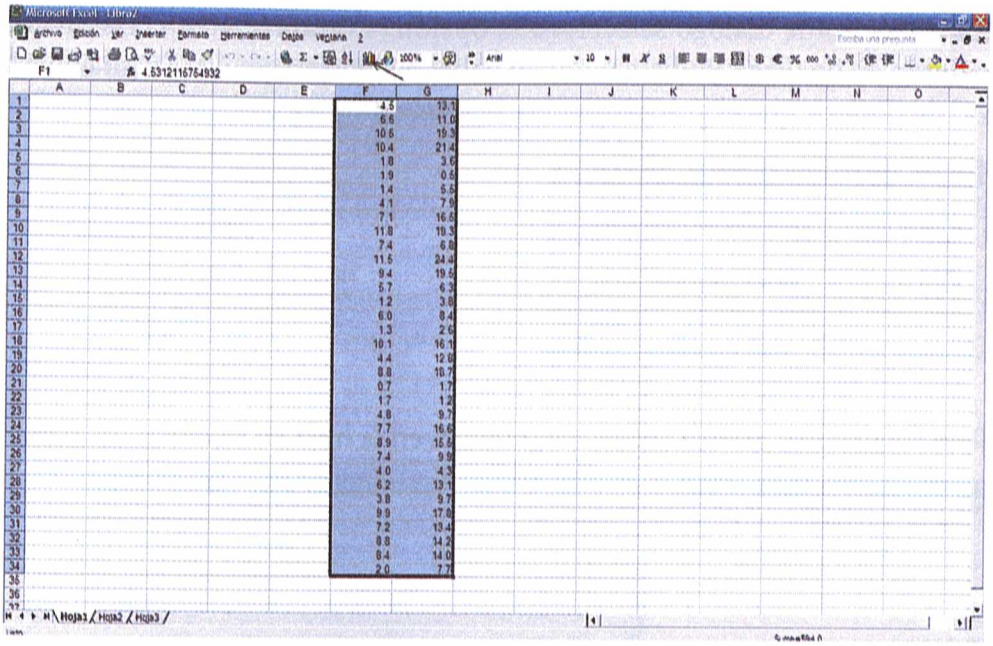
8. Para finalizar tu gráfica aparece contenida en la hoja de cálculo.



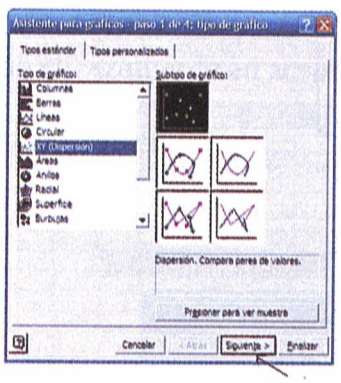
2. Gráfica de dispersión con Excel

Para realizar una gráfica de dispersión en Excel debes seguir los siguientes pasos.

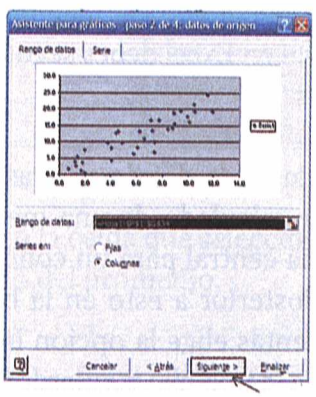
1. Selecciona el conjunto de datos que deseas graficar y da clic izquierdo en el asistente para gráficos



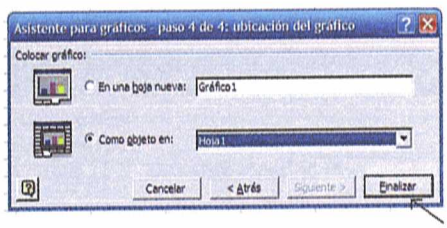
2. En el cuadro de diálogo escoge la opción "XY(Dispersión)" y la opción "Dispersión. Compara pares de valores. Da un clic izquierdo en siguiente.



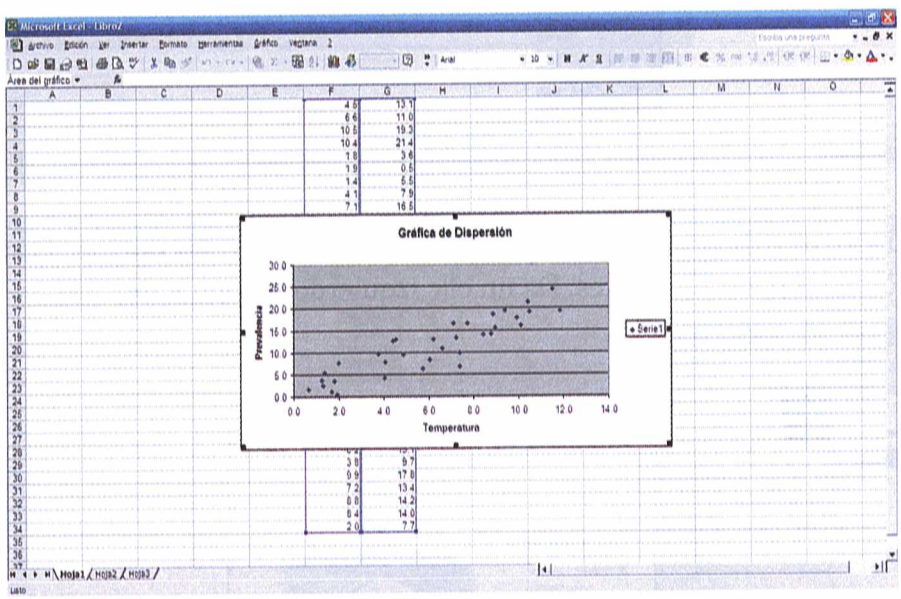
3. En el cuadro de diálogo selecciona la opción columnas. Da un clic izquierdo en siguiente.

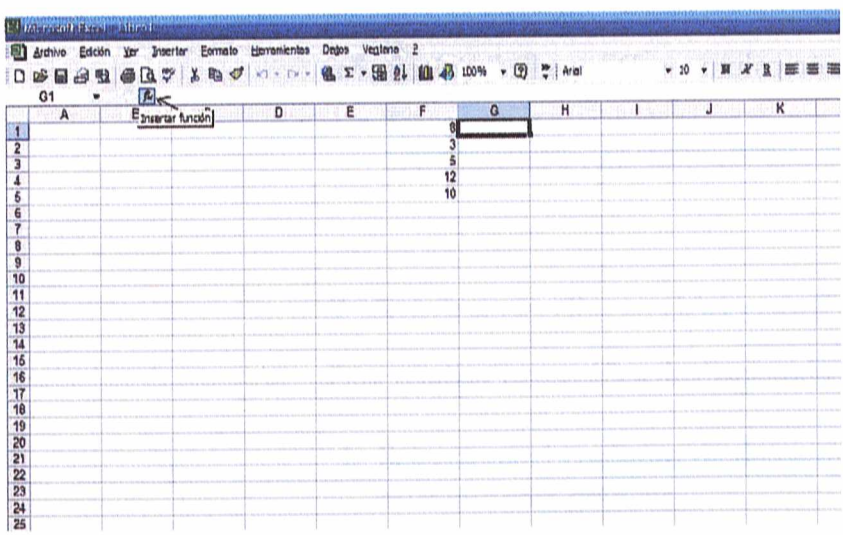


4. En el cuadro de diálogo selecciona la opción "Como objeto en". Da un clic izquierdo en finalizar. Pudiste también haber elegido la opción "En una hoja nueva" de preferir que el gráfico aparezca en una hoja a parte.



5. Para finalizar tu gráfica aparece contenida en la hoja de cálculo.



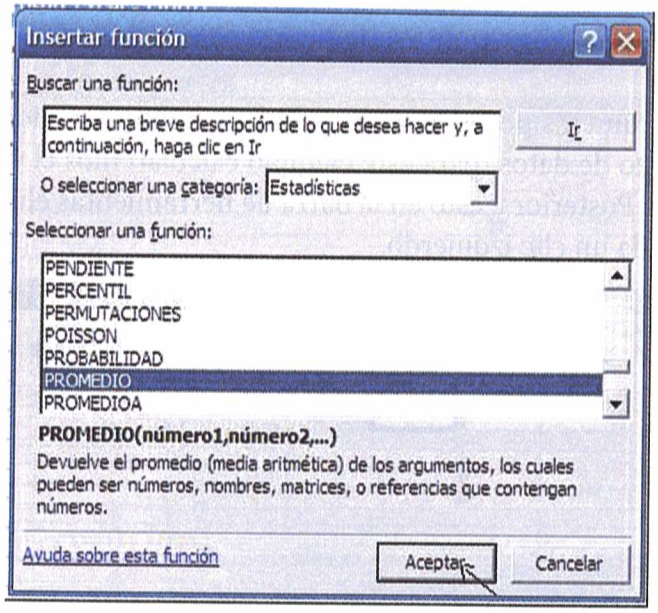


3. Cálculo de las medidas de tendencia central con Excel

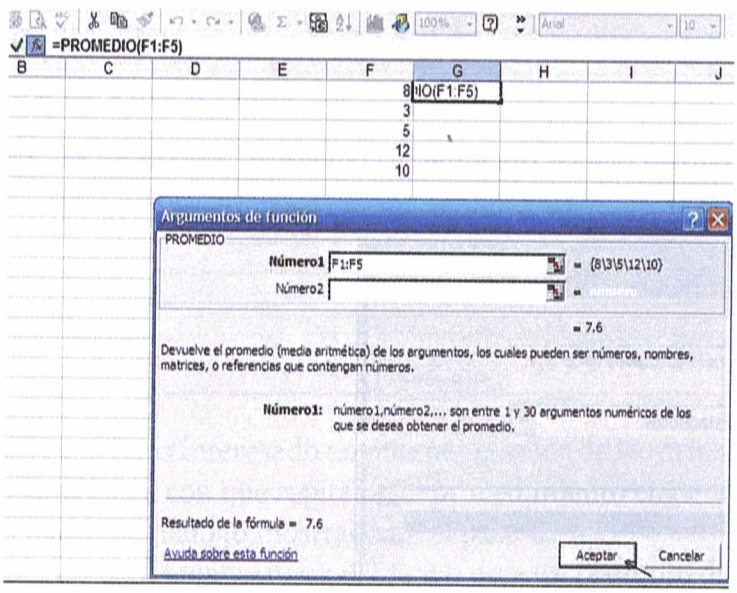
Para realizar el cálculo de las medidas de tendencia central en Excel debes seguir los siguientes pasos.

1. Selecciona la celda donde quieras poner el valor calculado de una medida de tendencia central para un conjunto de datos. Posterior a esto en la barra de herramientas elige la opción insertar función y da un clic izquierdo.

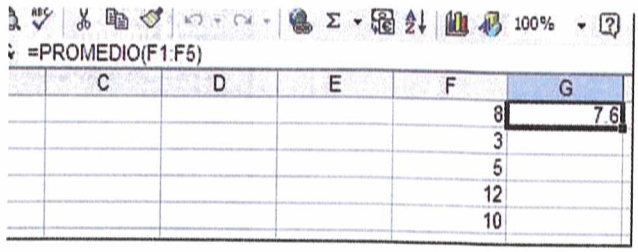
2. En el cuadro de diálogo escoge las opciones " Estadísticas " y "PROMEDIO". Da un clic izquierdo en aceptar.



3. En el cuadro de diálogo en el espacio indicado como " Número 1" selecciona las celdas que contienen los valores sobre los que quieres calcular el promedio. Da un clic izquierdo en aceptar.



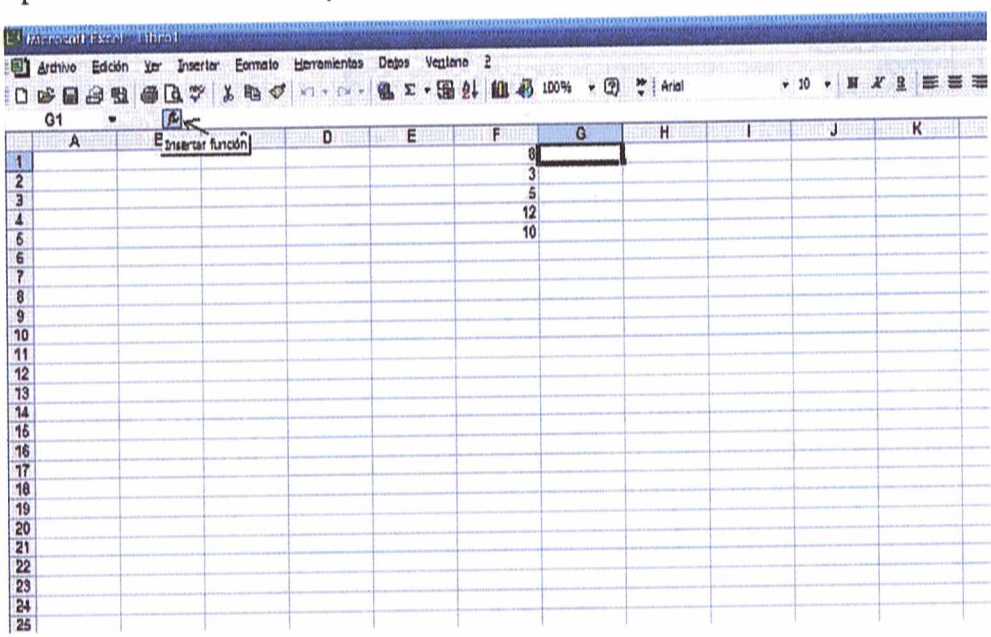
4. En la celda que seleccionaste aparece el valor calculado del promedio.



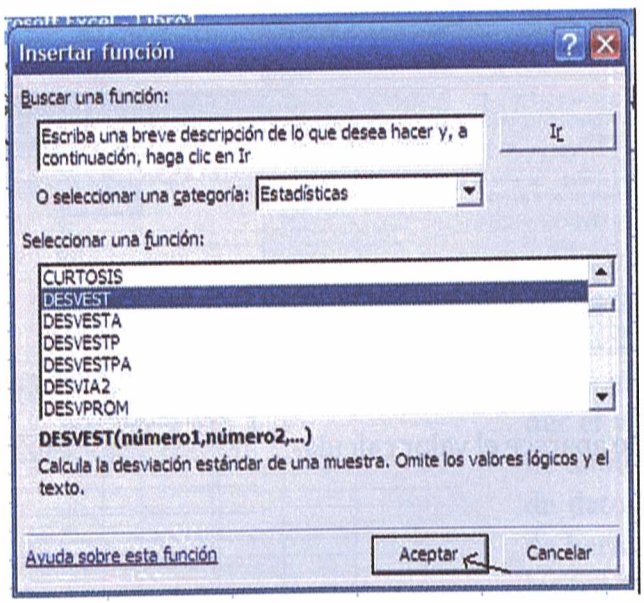
4. Cálculo de las medidas de dispersión con Excel

Para realizar el cálculo de las medidas de dispersión como desviación estándar o varianza de una muestra en Excel debes seguir los siguientes pasos.

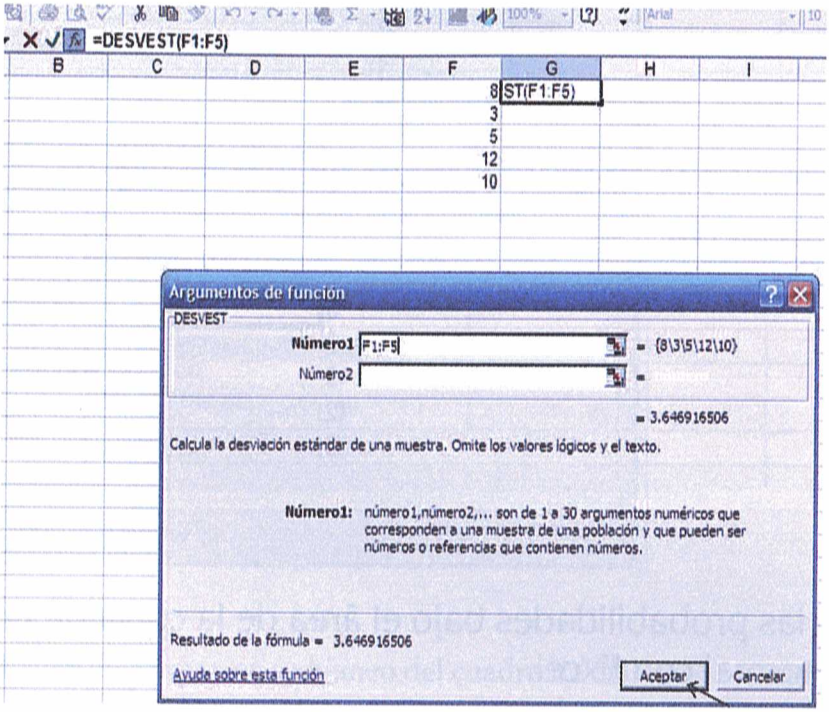
1. Selecciona la celda donde quieras poner el valor calculado de una medida de dispersión para un conjunto de datos (para este ejemplo calcularemos el valor de la desviación estándar). Posterior a esto en la barra de herramientas elige la opción insertar función y da un clic izquierdo.



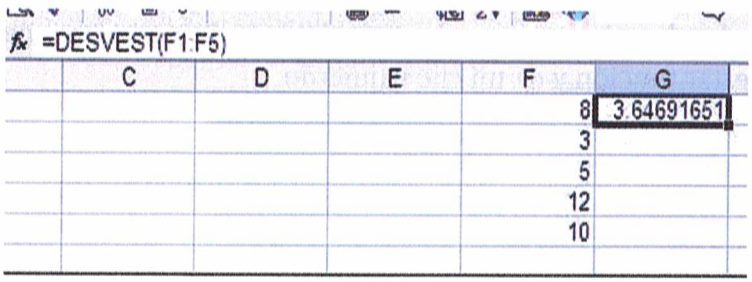
2. En el cuadro de diálogo escoge las opciones "Estadísticas" y "DESVEST". Da un clic izquierdo en aceptar.



3. En el cuadro de diálogo en el espacio indicado como " Número 1" selecciona las celdas que contienen los valores sobre los que quieres calcular el promedio. Da un clic izquierdo en aceptar.



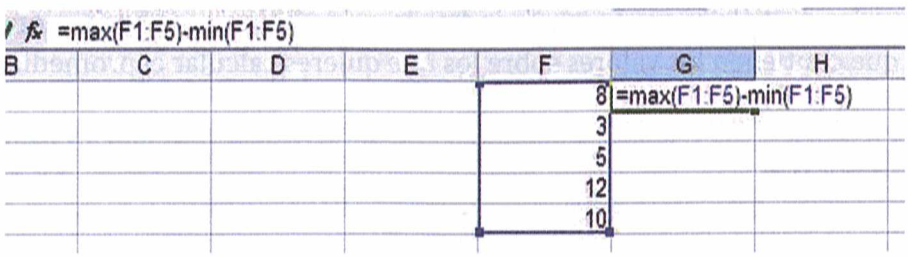
4. En la celda que seleccionaste aparece el valor calculado de la desviación estándar para una muestra.



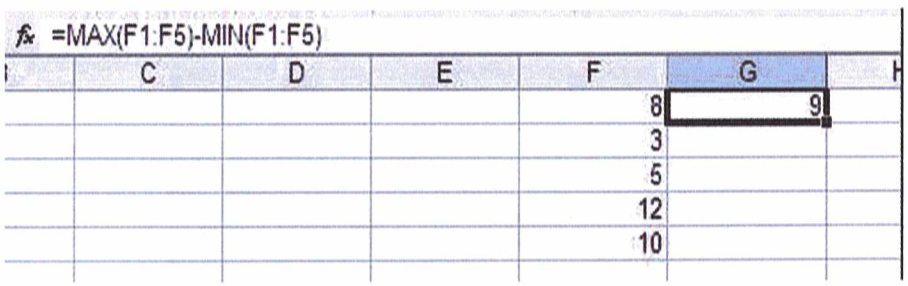
Si estuvieras interesado en obtener el valor de la varianza para este conjunto de datos basta con que repitas este procedimiento eligiendo la opción "VAR" en el cuadro de dialogo mostrado en el paso 2.

Para obtener el valor del rango para un conjunto de valores debes seguir los siguientes pasos.

1. Selecciona la celda donde quieras poner el valor calculado del rango para un conjunto de datos. Posterior a esto en la celda seleccionada deberás insertar la fórmula cuya operación indique la diferencia entre el valor máximo y el valor mínimo del conjunto de datos.



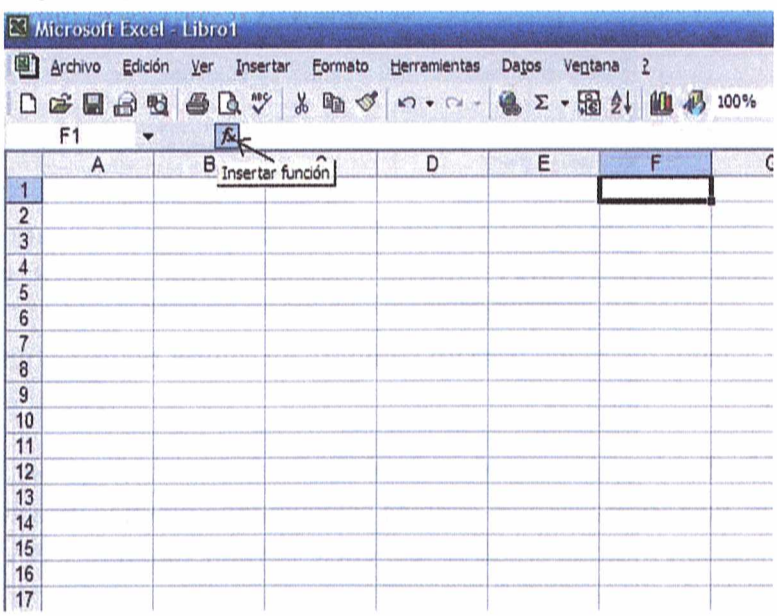
2. En la celda que seleccionaste aparecerá el valor calculado del rango.



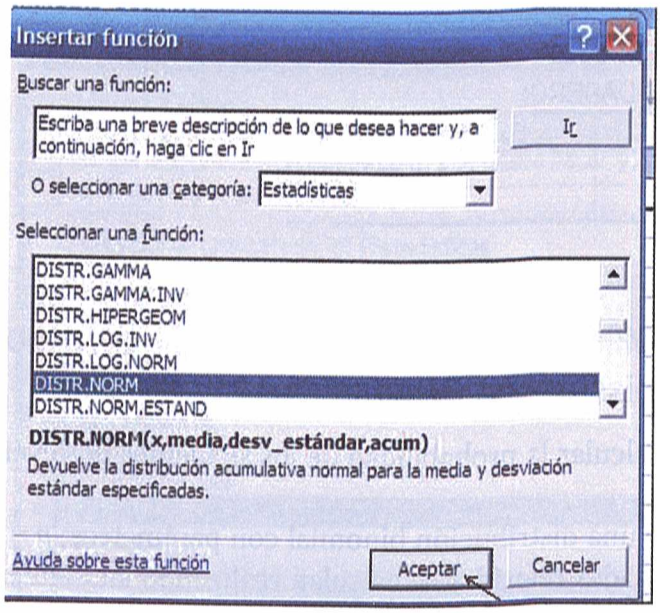
5. Cálculo de las probabilidades bajo el área de la curva de una distribución normal con Excel.

Supongamos que queremos calcular la probabilidad del evento $P(x \leq 1)$, donde la variable x tiene una distribución de probabilidad normal con $\mu = 0$ y $\sigma = 1$. Para ello debes realizar los siguientes pasos

1. Selecciona la celda donde quieres poner el valor calculado de esta probabilidad, elige la opción insertar función y da un clic izquierdo.



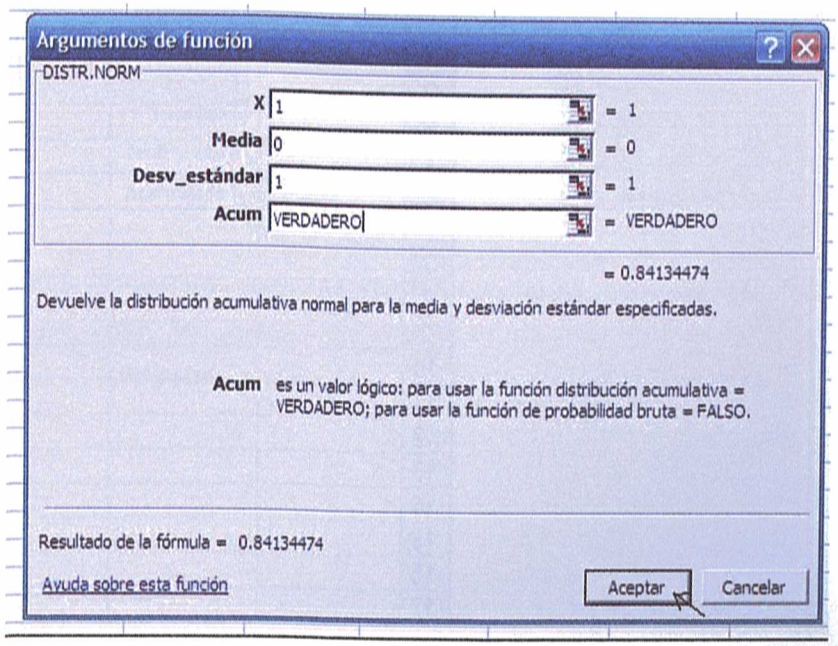
2. En el cuadro de diálogo escoge las opciones " Estadísticas " y "DIST.NORM". Da un clic izquierdo en aceptar.



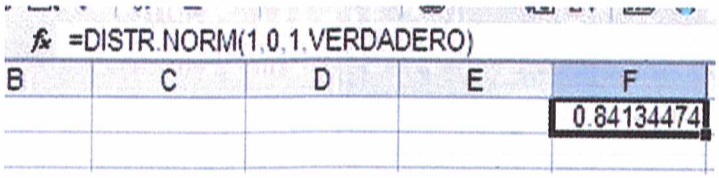
3. Rellena los espacios en blanco del cuadro de diálogo de la siguiente forma:

   a) $X = 1$; valor que corresponde al límite superior del intervalo cuya probabilidad se desea calcular
   b) Media = 0; valor que corresponde a la media de la distribución de probabilidad normal considerada.
   c) Desv_estándar = 1; valor que corresponde a la desviación estándar de la distribución de probabilidad normal considerada.
   d) acum. = VERDADERO; valor lógico necesario para calcular la probabilidad que se desea obtener.

   Da un clic izquierdo en aceptar.



4. El valor calculado de la probabilidad aparece en la celda que en un inicio seleccionaste. Ahora ya cuentas con el valor calculado para la probabilidad de este evento, cuyo valor es $P(x \leq 1) = .84134474$.

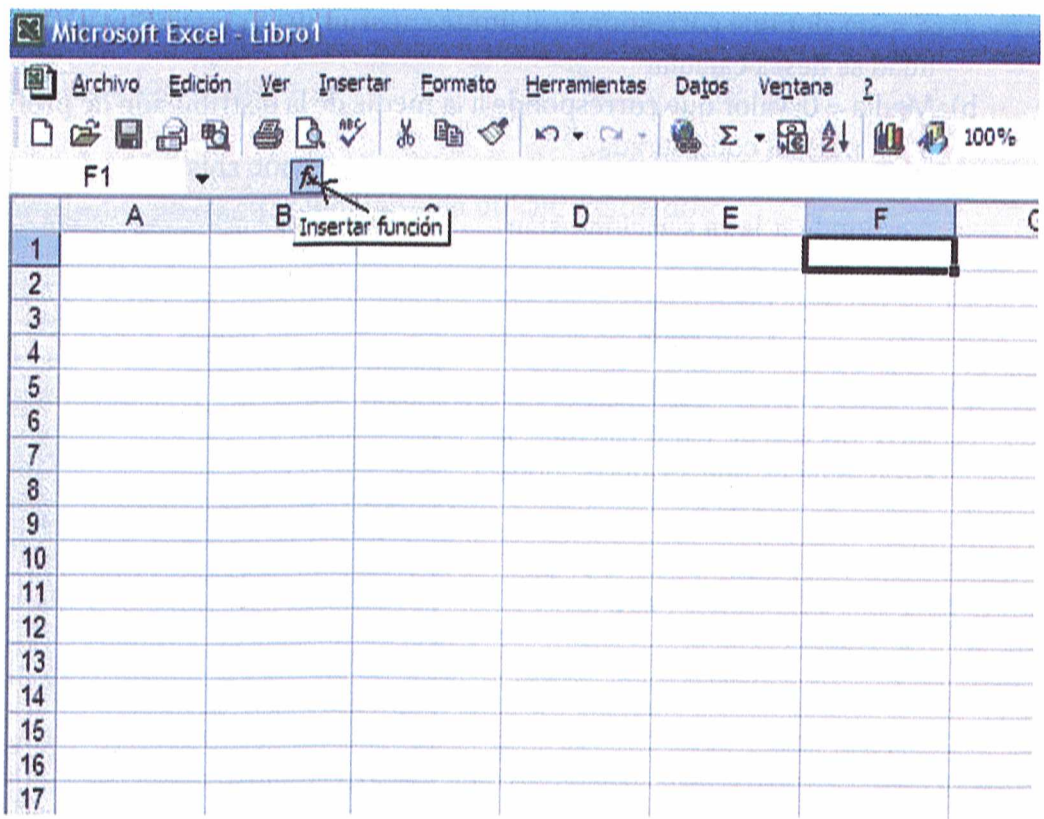


6. Cálculo de las probabilidades de las distribuciones de probabilidad binomial y de Poisson con Excel.

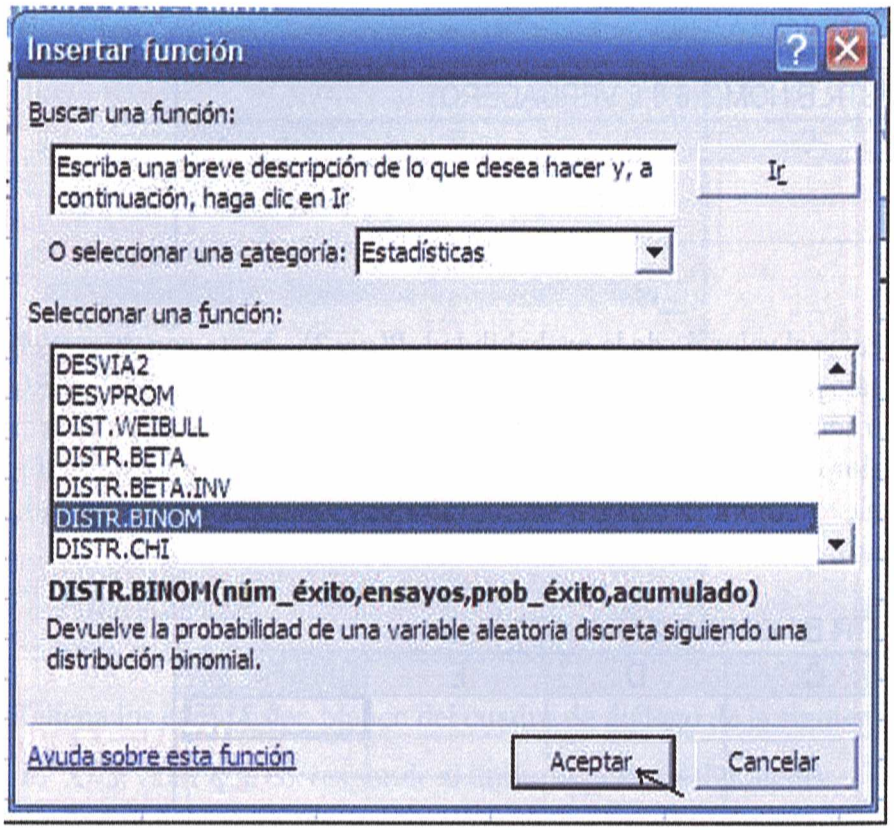
Supongamos que queremos calcular la probabilidad de los siguientes dos eventos $P(x \leq 3)$ y $P(x = 3)$.

Si suponemos que x tiene una distribución binomial con parámetros $p = .5$ y $n = 6$. Entonces las probabilidades referidas se calculan realizando los siguientes pasos.

1. Selecciona la celda donde quieres poner el valor calculado de esta probabilidad, elige la opción insertar función y da un clic izquierdo.

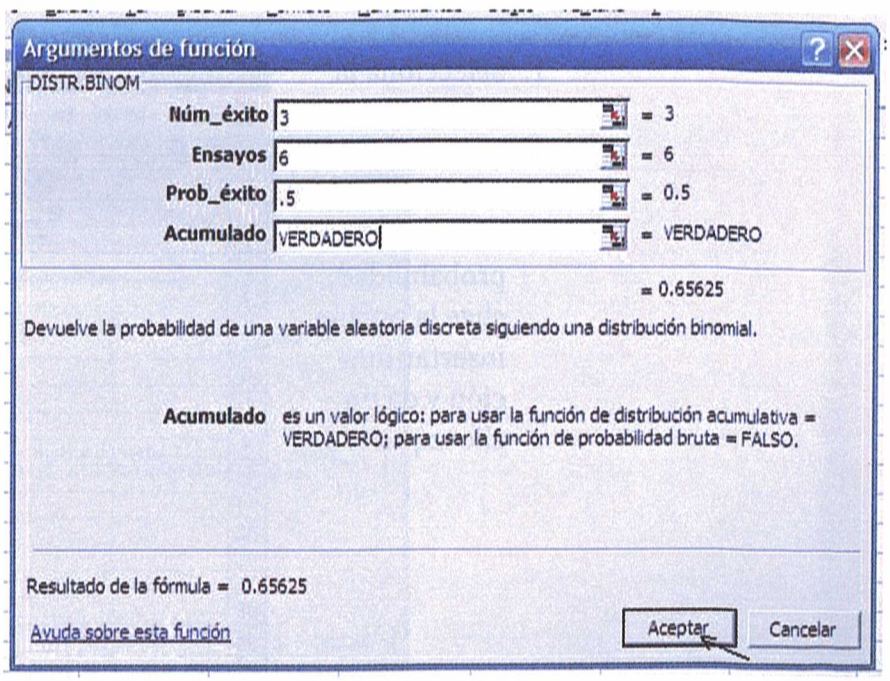


2. En el cuadro de diálogo escoge las opciones " Estadísticas " y "DIST.BINOM". Da un clic izquierdo en aceptar.



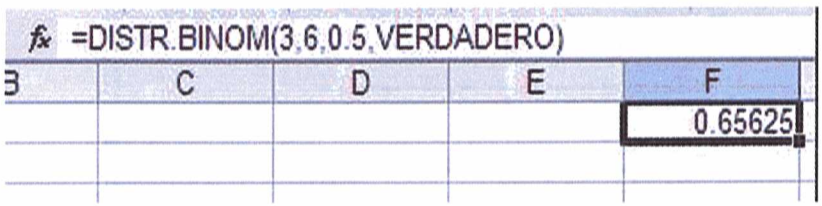
3. Rellena los espacios en blanco del cuadro de diálogo de la siguiente forma:

   a) Núm_éxito=3; valor que corresponde al límite superior del intervalo cuya probabilidad se desea calcular
   b) Ensayos = 6; valor que corresponde al parámetro n=6.
   c) Prob_éxito = .5; valor que corresponde al parámetro p=.5.
   d) acum. = VERDADERO; valor lógico necesario para calcular la probabilidad que se desea obtener.



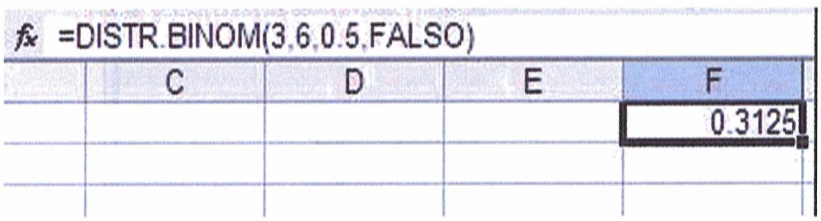
Da un clic izquierdo en aceptar.

4. El valor calculado de la probabilidad aparece en la celda que en un inicio seleccionaste. Ahora ya cuentas con el valor calculado para la probabilidad de este evento, cuyo valor es $P(x \leq 3) = .65625$.



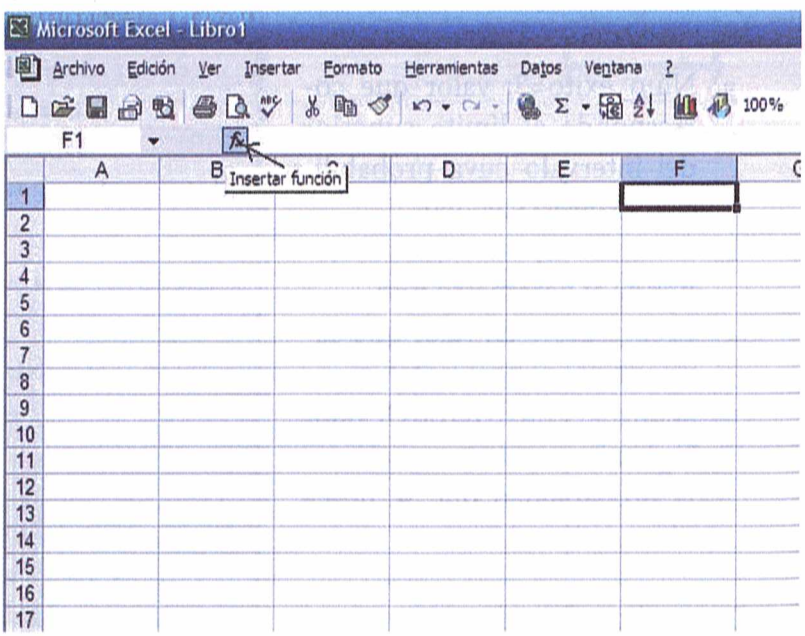
5. Para calcular el valor de de la probabilidad $P(x = 3)$, basta con que en el cuadro de diálogo que aparece en 3 cambies la opción acum. = VERDADERO por la opción acum. = FALSO. Da un clic izquierdo en aceptar.

   El valor calculado de la probabilidad aparece en la celda que en un inicio seleccionaste. Ahora ya cuentas con el valor calculado para la probabilidad de este evento, cuyo valor es $P(x = 3) = .3125$.

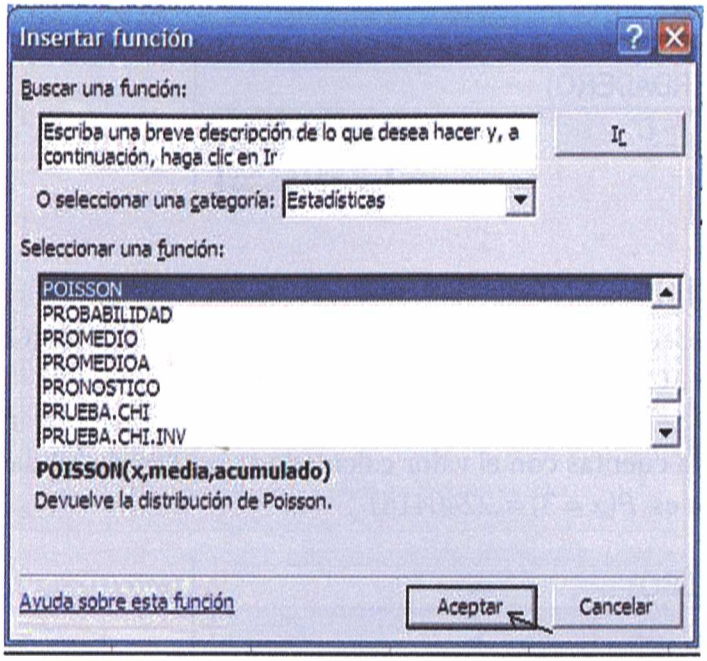


Si suponemos que x tiene una distribución de Poisson con parámetro $\lambda = 3$. Entonces las probabilidades referidas se calculan realizando los siguientes pasos.

1. Selecciona la celda donde quieres poner el valor calculado de esta probabilidad, elige la opción insertar función y da un clic izquierdo.



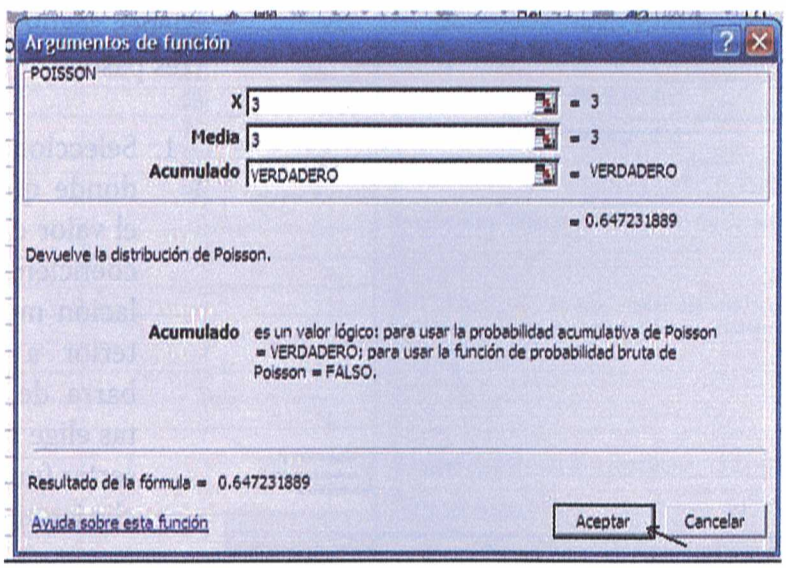
2. En el cuadro de diálogo escoge las opciones " Estadísticas " y "POISSON". Da un clic izquierdo en aceptar.



3. Rellena los espacios en blanco del cuadro de diálogo de la siguiente forma:

   a) X=3; valor que corresponde al límite superior del intervalo cuya probabilidad se desea calcular
   b) Media = 3; valor que corresponde al parámetro $\lambda = 3$.
   c) acum. = VERDADERO; valor lógico necesario para calcular la probabilidad que se desea obtener.

   Da un clic izquierdo en aceptar.

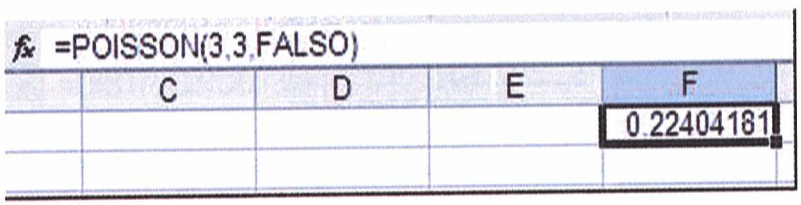


4. El valor calculado de la probabilidad aparece en la celda que en un inicio seleccionaste. Ahora ya cuentas con el valor calculado para la probabilidad de este evento, cuyo valor es $P(x \leq 3) = .64723189$.

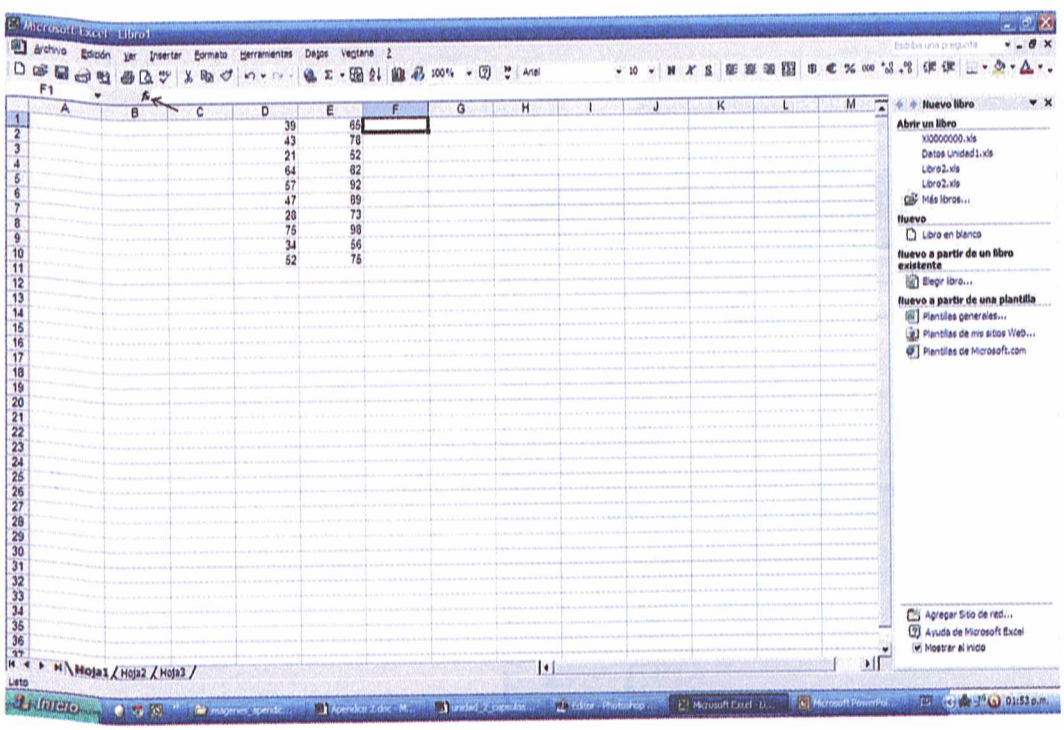


5. Para calcular el valor de de la probabilidad $P(x = 3)$, basta con que en el cuadro de diálogo que aparece en 3 cambies la opción acum. = VERDADERO por la opción acum. = FALSO. Da un clic izquierdo en aceptar.

   El valor calculado de la probabilidad aparece en la celda que en un inicio seleccionaste. Ahora ya cuentas con el valor calculado para la probabilidad de este evento, cuyo valor es $P(x = 3) = .22404181$.



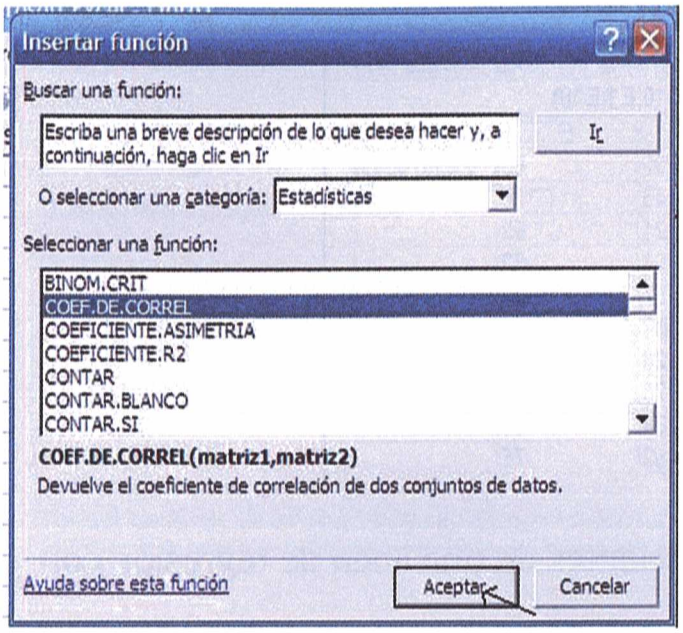
7. Cálculo del coeficiente de correlación muestral con Excel.



Para realizar el cálculo del coeficiente de correlación muestral en Excel debes seguir los siguientes pasos.

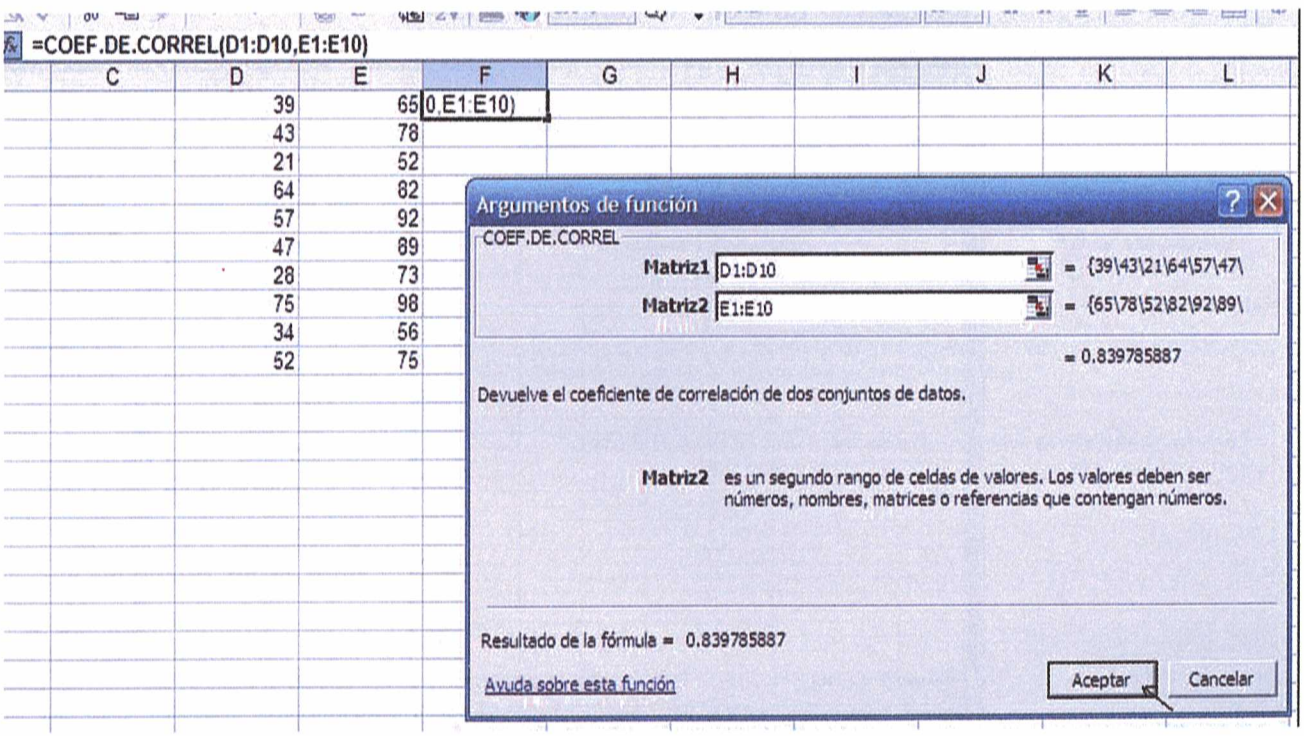
1. Selecciona la celda donde quieras poner el valor calculado del coeficiente de correlación muestral. Posterior a esto en la barra de herramientas elige la opción insertar función y da un clic izquierdo.

2. En el cuadro de diálogo escoge las opciones " Estadísticas " y "COEF. DE.CORREL". Da un clic izquierdo en aceptar.

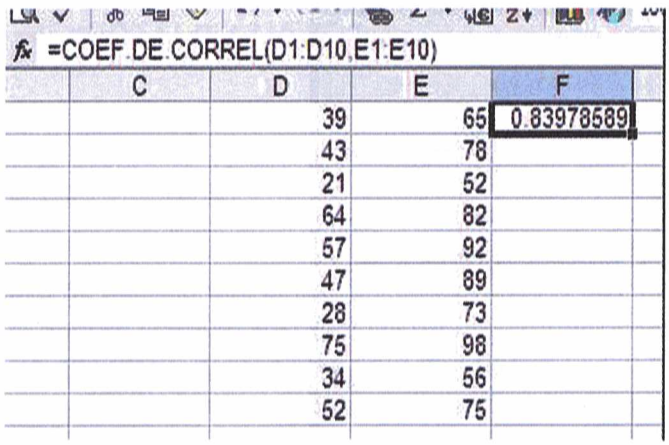


3. Rellena los espacios en blanco del cuadro de diálogo con las celdas correspondientes a las dos columnas que representan las observaciones pareadas sobre las que vas a calcular el coeficiente de correlación.

   Da un clic izquierdo en aceptar.



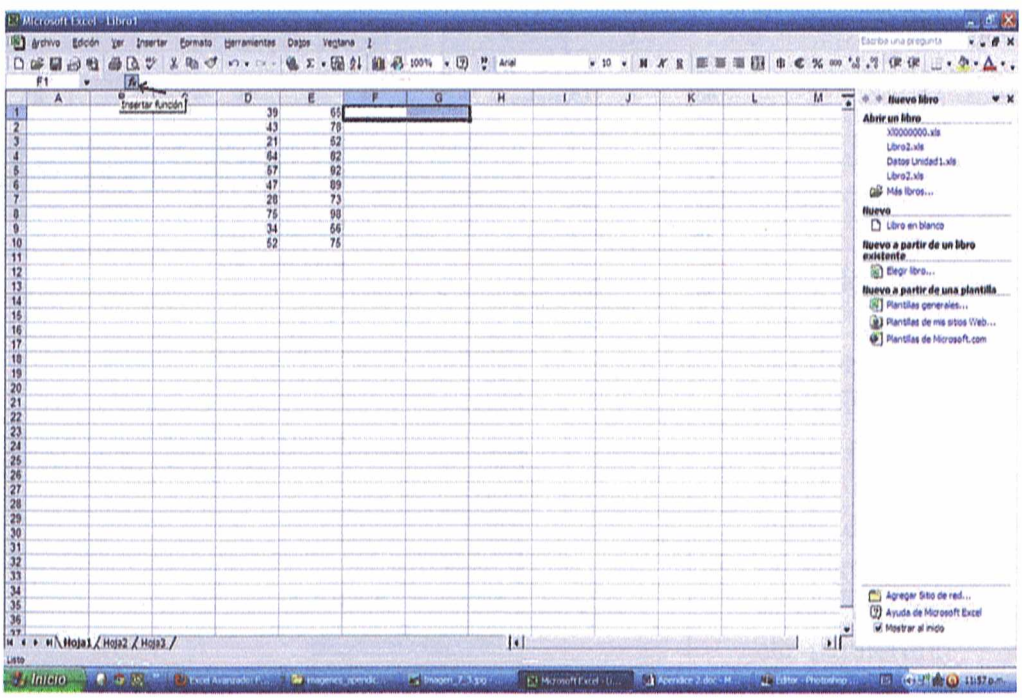
4. El valor calculado del coeficiente de correlación muestral aparece en la celda que en un inicio seleccionaste. Ahora ya cuentas con el valor calculado para este coeficiente de correlación muestral, cuyo valor es $r = .83978589$.



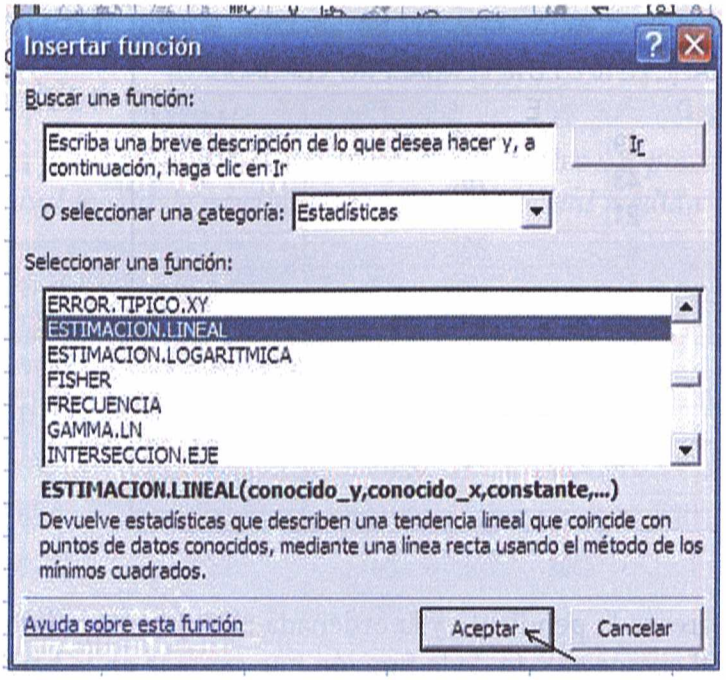
8. Cálculo de los coeficientes de una línea de regresión con Excel.

Para realizar el cálculo de los coeficientes de una línea de regresión en Excel debes seguir los siguientes pasos.

1. Selecciona dos celdas donde quieras poner los coeficientes correspondientes a la línea de regresión. Posterior a esto en la barra de herramientas elige la opción insertar función y da un clic izquierdo.



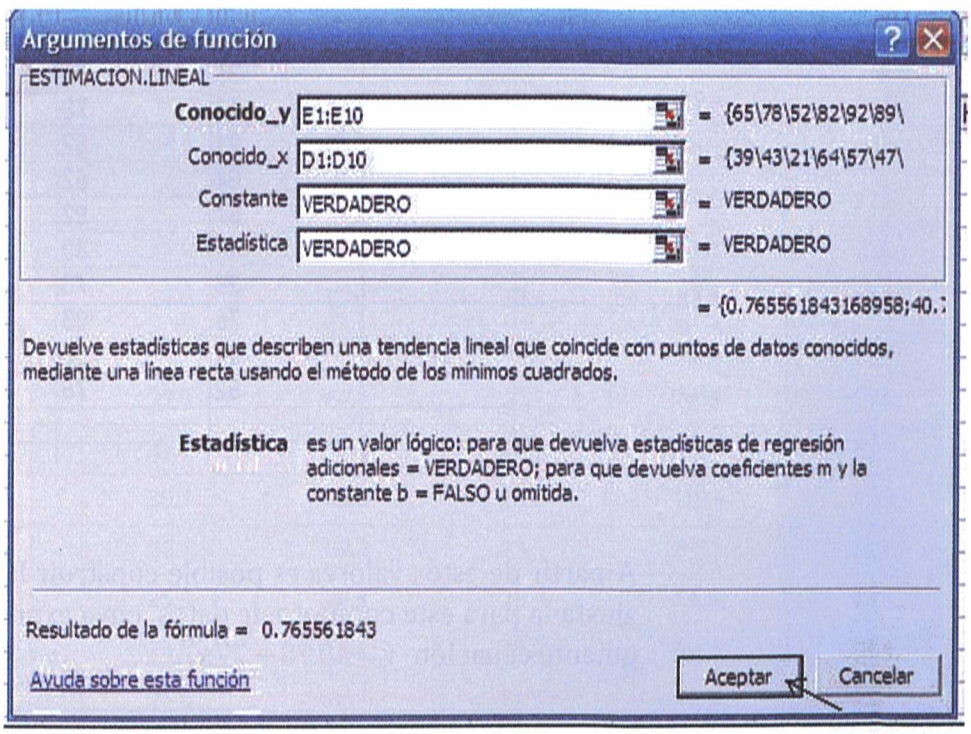
2. En el cuadro de diálogo escoge las opciones " Estadísticas " y "ESTIMACION. LINEAL". Da un clic izquierdo en aceptar.



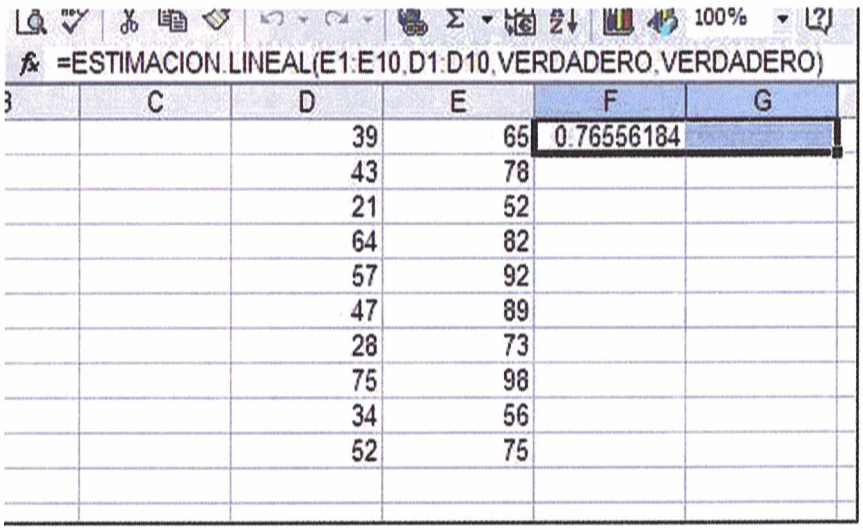
3. Rellena los espacios en blanco del cuadro de diálogo de la siguiente forma:

a) Conocido_y=E1:E10; valores correspondientes a la variable dependiente y
b) Conocido_x = D1:D10; valores correspondientes a la variable independiente x.
c) Constante = VERDADERO; valor lógico necesario para obtener el valor de la ordenada al origen en la ecuación de la línea de regresión.
d) Estadística = VERDADERO; valor lógico necesario para obtener el valor de la pendiente en la ecuación de la línea de regresión.

Da un clic izquierdo en aceptar



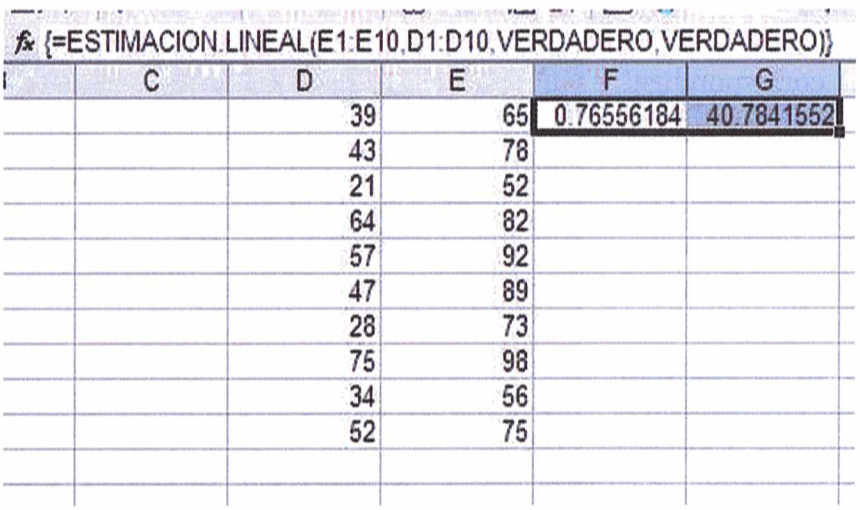
4. El valor obtenido en la primera de las dos celdas seleccionadas, es el correspondiente a la pendiente de la línea de regresión



=ESTIMACION.LINEAL(E1:E10,D1:D10,VERDADERO,VERDADERO)

C	D	E	F	G
	39	65	0.76556184	
	43	78		
	21	52		
	64	82		
	57	92		
	47	89		
	28	73		
	75	98		
	34	56		
	52	75		

5. Si deseamos que aparezca la pendiente y la ordenada al origen de la línea de regresión, situamos el cursor al lado de la función que aparece en la barra de fórmulas y pulsamos ctrl.+Shift+Enter (a la vez). Una vez hecho esto aparecen la pendiente y la ordenada al origen de la recta de regresión en las celdas que seleccionamos en un principio.



{=ESTIMACION.LINEAL(E1:E10,D1:D10,VERDADERO,VERDADERO)}

C	D	E	F	G
	39	65	0.76556184	40.7841552
	43	78		
	21	52		
	64	82		
	57	92		
	47	89		
	28	73		
	75	98		
	34	56		
	52	75		

A partir de estos valores es posible construir la ecuación de la línea de regresión ajustada para este conjunto de datos, cuya expresión queda determinada por la siguiente ecuación $y = 40.78 + .77x$.

Bases de datos

Unidad 1

Actividad 4

Datos pareados correspondientes a las variables x = *porcentaje de casos de dengue* y = *nivel medio de precipitación pluvial mensual medida en hectómetros cúbicos.*

Observaciones pareadas porcentaje de casos de dengue vs precipitación pluvial media mensual							
X	y	X	Y	x	Y	x	Y
51.3	21.0	93.9	35.4	110.9	32.9	104.4	38.9
117.5	38.7	89.0	37.5	82.8	33.3	106.7	34.8
119.1	49.4	111.8	45.9	59.2	21.4	116.6	44.4
60.8	30.5	106.8	40.4	119.9	43.7	63.1	20.5
64.3	22.0	60.1	30.2	83.8	34.1	77.9	30.4
95.8	36.8	105.9	43.7	93.7	32.7	55.9	18.5
55.0	17.3	67.1	28.1	111.7	41.1	109.8	39.8
62.9	23.4	117.2	43.7	112.5	35.4	104.7	45.1
89.0	36.5	75.0	26.3	99.6	35.3	80.0	27.7
88.1	32.8	54.6	21.2	82.1	30.2	69.6	27.9
103.2	35.2	107.7	31.8	61.3	20.3	56.6	21.4
115.2	45.3	83.8	25.1	68.8	21.1	100.9	42.7
74.6	33.5	109.5	44.2	116.8	43.9	99.4	35.6
72.8	24.6	73.2	26.7	83.7	34.1	109.9	39.5
118.3	40.9	97.8	37.8	75.4	25.4	84.6	32.8
94.8	33.6	92.1	35.0	59.9	24.3	71.1	27.2
76.0	29.9	94.3	37.6	97.8	38.3	83.1	29.0
75.3	30.4	57.7	26.7	109.1	33.7	88.5	34.1
58.3	17.6	67.3	27.9	57.4	27.4	61.6	27.4
107.8	44.4	60.7	17.8	56.5	24.5	107.4	35.0
52.3	23.2	92.8	30.1	79.5	29.9	118.4	43.9
87.7	38.4	74.3	28.7	69.4	23.5	64.1	22.8
103.3	31.9	76.6	25.3	74.3	29.4	54.9	23.1
51.5	23.1	79.6	25.4	75.9	28.9	58.5	26.6
74.7	27.0	71.5	26.6	106.5	41.5	62.5	23.7

Muestra de tamaño n = 25 de observaciones pareadas porcentaje de casos de dengue vs precipitación pluvial media mensual			
x	Y	X	Y
30.1	92.8	27	74.7
20.5	63.1	40.4	106.8
27.9	67.3	45.1	104.7
44.4	107.8	38.4	87.7
18.5	55.9	35	92.1
49.4	119.1	39.8	109.8
21.4	56.6	35	107.4
29.9	76	31.8	107.7
17.6	58.3	27.9	69.6
22.8	64.1	28.7	74.3
30.4	77.9	32.8	88.1
30.2	82.1	30.2	60.1
25.4	79.6		

Muestra de tamaño n=15 de observaciones pareadas porcentaje de casos de dengue vs precipitación pluvial media mensual					
x	y	X	y	x	y
28.7	74.3	23.1	51.5	30.4	75.3
26.6	71.5	17.3	55	39.5	109.9
22	64.3	25.4	79.6	44.2	109.5
45.1	104.7	30.2	82.1	32.9	110.9
32.7	93.7	33.6	94.8	20.3	61.3

Unidad 2

Actividad 1

Edad de residentes oriundos				
39	37	35	43	43
46	32	38	47	41
48	45	35	47	46
20	45	27	39	46
36	35	30	54	40
33	42	37	35	72
30	58	38	31	34
33	30	30	51	37
46	29	42	43	30
43	34	36	49	59
51	48	39	21	57
28	45	36	54	34
42	26	54	33	52
51	47	47	39	44
56	24	33	42	37
40	33	50	43	43
39	37	36	42	50
28	45	45	23	43
18	49	48	18	31
20	61	46	38	43

Nivel de ingreso para residentes oriundos				
38	47	49	49	53
50	48	49	51	47
48	44	48	39	33
43	49	56	44	58
45	45	49	59	56
43	51	48	51	47
47	48	50	57	50
44	50	42	51	54
47	46	49	56	57
49	41	53	47	53
52	46	57	49	49
49	47	57	59	45
52	52	49	53	52
48	52	51	43	47
50	50	54	48	46
50	49	51	51	50
40	58	40	49	54
45	53	36	44	50
44	50	49	51	50
41	55	49	55	47

Edad de residentes de origen foráneo				
34	51	32	37	49
35	22	44	26	26
25	30	40	44	35
19	35	37	30	31
27	39	36	30	31
46	37	27	39	29
29	24	38	40	62
42	33	37	42	41
18	37	28	48	41
31	18	25	36	36
33	28	32	36	32
29	31	24	27	45
43	47	46	43	34
32	53	33	42	47
36	48	48	43	47
45	33	23	23	29
38	45	32	38	25
41	39	27	19	36
34	28	29	48	29
34	48	45	32	31

Nivel de ingreso para residentes de origen foráneo				
60	88	65	62	76
59	60	64	63	86
47	59	66	71	44
58	69	60	61	62
55	46	63	63	89
59	33	43	60	40
37	74	30	77	76
64	78	64	30	53
59	53	89	70	54
74	65	95	66	71
59	56	41	46	53
47	66	75	66	75
70	92	78	66	49
43	73	69	87	39
54	61	52	59	42
50	48	70	49	69
70	38	31	68	57
46	63	75	50	50
61	53	46	47	68
62	77	66	76	73

Actividad 3

Tasa neta de inmigración vs PIB			
X	Y	x	y
2.77	1.679	3.49	4.081
3.99	3.084	2.80	3.412
3.97	5.925	2.62	2.553
3.13	4.123	2.71	1.259
3.56	4.552	2.90	1.834
2.32	0.615	3.42	3.945
2.35	1.615	2.88	2.258
3.39	1.932	3.53	2.845
3.59	2.017	3.19	2.452
3.21	3.857	2.03	0.334
3.63	2.596	2.62	1.862
2.68	1.846	3.09	2.669
2.88	2.413	2.94	4.864
3.06	2.574	3.62	5.325
3.78	3.087	2.89	1.198
2.80	1.610	2.76	1.995
3.21	2.793	2.26	0.258
2.93	2.421	3.22	2.899
2.95	3.596	2.68	1.037
2.68	1.507	3.74	4.311
2.20	1.415	2.97	3.218
2.15	1.156	3.12	2.967
3.00	3.544	3.44	2.798
2.98	3.484	2.61	1.301
3.72	4.165	3.97	4.724

Tasa neta de inmigración vs Índice de pobreza			
X	Y	x	y
2.77	6.163	3.49	5.709
3.99	2.695	2.80	7.775
3.97	5.605	2.62	7.669
3.13	7.169	2.71	5.993
3.56	5.887	2.90	5.824
2.32	6.930	3.42	5.853
2.35	7.777	2.88	6.331
3.39	3.939	3.53	4.312
3.59	3.221	3.19	5.266
3.21	6.588	2.03	7.807
3.63	3.673	2.62	6.980
2.68	6.702	3.09	5.880
2.88	6.487	2.94	8.665
3.06	5.910	3.62	6.435
3.78	3.535	2.89	5.199
2.80	5.988	2.76	6.527
3.21	5.531	2.26	6.786
2.93	6.269	3.22	5.579
2.95	7.371	2.68	5.904
2.68	6.382	3.74	4.925
2.20	8.189	2.97	6.910
2.15	8.117	3.12	6.047
3.00	7.129	3.44	4.611
2.98	7.158	2.61	6.430
3.72	4.844	3.97	4.411

Tasa neta de inmigración vs Índice de delincuencia			
X	Y	x	Y
2.77	16.1	3.49	13.2
3.99	12.1	2.80	12.8
3.97	13.2	2.62	10.3
3.13	10.5	2.71	17.2
3.56	10.0	2.90	17.9
2.32	10.6	3.42	14.8
2.35	9.2	2.88	9.9
3.39	18.0	3.53	9.4
3.59	15.8	3.19	14.4
3.21	19.5	2.03	19.5
3.63	16.1	2.62	14.2
2.68	19.3	3.09	19.4
2.88	16.2	2.94	11.6
3.06	17.5	3.62	10.6
3.78	14.5	2.89	14.3
2.80	19.9	2.76	15.6
3.21	17.1	2.26	9.9
2.93	11.0	3.22	19.1
2.95	14.9	2.68	18.3
2.68	15.2	3.74	14.8
2.20	9.9	2.97	12.1
2.15	9.4	3.12	18.0
3.00	10.6	3.44	12.7
2.98	19.3	2.61	18.7
3.72	10.4	3.97	13.6

Tasa neta de inmigración vs Nivel anual de almacenamiento			
X	Y	x	y
2.77	3521	3.49	4401
3.99	2330	2.80	5189
3.97	5208	2.62	4731
3.13	5198	2.71	3234
3.56	4717	2.90	3412
2.32	3431	3.42	4415
2.35	4350	2.88	3881
3.39	2455	3.53	3080
3.59	2112	3.19	3404
3.21	4765	2.03	3767
3.63	2623	2.62	4041
2.68	3885	3.09	3833
2.88	4037	2.94	6342
3.06	3804	3.62	5370
3.78	2779	2.89	2782
2.80	3395	2.76	3862
3.21	3705	2.26	3188
2.93	3924	3.22	3779
2.95	5060	2.68	3082
2.68	3557	3.74	4092
2.20	4476	2.97	4638
2.15	4316	3.12	4061
3.00	4907	3.44	3217
2.98	4894	2.61	3486
3.72	3980	3.97	4011

Unidad 2

Evalúa tu aprendizaje

II.

Edad de residentes oriundos				
22	30	22	36	28
27	22	30	25	31
22	31	34	36	25
23	30	31	37	26
42	37	28	30	39
44	24	35	25	31
30	25	30	16	42
31	27	29	33	26
26	29	35	25	28
25	36	26	37	23
23	27	36	20	29
36	27	21	32	38
31	26	36	25	22
31	29	40	22	24
30	27	28	21	38
24	17	27	39	26
27	28	29	24	29
30	28	27	27	31
34	33	25	31	20
20	34	29	26	24

Nivel de ingreso para residentes oriundos				
64	71	51	46	63
44	56	71	55	56
69	57	46	58	45
63	61	56	59	58
58	58	45	44	61
51	53	51	49	49
61	61	53	65	53
58	66	50	52	59
64	49	58	78	71
60	61	57	48	58
49	53	65	64	59
63	61	66	63	59
49	66	51	65	60
60	56	68	60	67
50	66	47	51	61
58	48	47	51	64
74	52	45	55	58
53	59	57	51	69
61	50	55	62	65
56	73	65	55	67

Edad de residentes de origen foráneo				
21	25	28	27	25
25	23	23	28	24
27	24	26	26	23
28	24	23	21	23
21	23	28	21	23
26	27	21	22	24
22	24	23	27	28
25	23	27	25	25
26	25	22	28	27
27	25	25	23	26
25	25	26	24	28
24	24	26	25	28
23	26	29	23	25
24	25	21	25	26
26	23	26	23	28
28	28	25	26	21
21	26	22	25	23
24	23	28	21	23
26	20	24	26	25
24	24	22	25	21

Nivel de ingreso para residentes de origen foráneo				
59	51	44	60	53
52	37	50	61	46
57	51	58	55	44
46	41	52	63	51
53	57	48	61	39
48	46	58	56	56
63	48	61	54	49
55	55	47	48	53
59	46	49	51	57
47	51	50	39	49

Nivel de ingreso para residentes de origen foráneo				
48	54	54	41	53
50	51	55	60	43
59	51	43	50	50
57	45	61	55	51
54	53	56	54	46
59	46	62	57	43
48	63	42	59	55
47	63	56	55	57
51	47	50	42	46
54	55	53	45	53

Tasa neta de inmigración vs PIB			
X	y	x	y
2.77	3.132	3.49	4.106
3.99	2.100	2.80	4.804
3.97	4.976	2.62	4.322
3.13	4.857	2.71	2.837
3.56	4.431	2.90	3.040
2.32	2.983	3.42	4.110
2.35	3.907	2.88	3.506
3.39	2.147	3.53	2.790
3.59	1.830	3.19	3.070
3.21	4.434	2.03	3.282
3.63	2.346	2.62	3.632
2.68	3.485	3.09	3.485
2.88	3.661	2.94	5.976
3.06	3.453	3.62	5.091
3.78	2.522	2.89	2.409
2.80	3.010	2.76	3.472
3.21	3.373	2.26	2.733
2.93	3.556	3.22	3.450
2.95	4.694	2.68	2.681
2.68	3.156	3.74	3.829
2.20	4.013	2.97	4.275
2.15	3.847	3.12	3.718
3.00	4.548	3.44	2.915
2.98	4.532	2.61	3.077
3.72	3.716	3.97	3.779

Tasa neta de inmigración vs Índice de pobreza			
X	y	x	y
2.77	4.318	3.49	7.506
3.99	7.064	2.80	6.086
3.97	9.886	2.62	5.020
3.13	7.158	2.71	3.831
3.56	8.058	2.90	4.610
2.32	2.751	3.42	7.293
2.35	3.793	2.88	5.011
3.39	5.253	3.53	6.314
3.59	5.559	3.19	5.551
3.21	6.979	2.03	2.152
3.63	6.172	2.62	4.328
2.68	4.383	3.09	5.659
2.88	5.166	2.94	7.692
3.06	5.530	3.62	8.892
3.78	6.836	2.89	3.971
2.80	4.279	2.76	4.622
3.21	5.914	2.26	2.336
2.93	5.236	3.22	6.035
2.95	6.430	2.68	3.572
2.68	4.040	3.74	8.016
2.20	3.425	2.97	6.075
2.15	3.114	3.12	5.993
3.00	6.431	3.44	6.172
2.98	6.346	2.61	3.764
3.72	7.851	3.97	8.683

Tasa neta de inmigración vs Índice de delincuencia			
X	y	x	y
2.77	10.5	3.49	14.3
3.99	14.4	2.80	12.3
3.97	17.2	2.62	11.1
3.13	13.7	2.71	10.0
3.56	15.0	2.90	10.9
2.32	8.5	3.42	14.1
2.35	9.6	2.88	11.3
3.39	12.0	3.53	13.2
3.59	12.5	3.19	12.1
3.21	13.6	2.03	7.7
3.63	13.1	2.62	10.4
2.68	10.5	3.09	12.1
2.88	11.5	2.94	14.0
3.06	12.0	3.62	15.8
3.78	13.9	2.89	10.3
2.80	10.5	2.76	10.8
3.21	12.5	2.26	8.1
2.93	11.6	3.22	12.6
2.95	12.8	2.68	9.7
2.68	10.1	3.74	15.1
2.20	9.1	2.97	12.4
2.15	8.8	3.12	12.5
3.00	12.8	3.44	13.0
2.98	12.7	2.61	9.8
3.72	14.9	3.97	16.0

Tasa neta de inmigración vs Nivel anual de almacenamiento			
X	y	x	y
2.77	3541	3.49	5370
3.99	4373	2.80	4701
3.97	7214	2.62	3842
3.13	5412	2.71	2548
3.56	5841	2.90	3123
2.32	1904	3.42	5234
2.35	2904	2.88	3547
3.39	3221	3.53	4134
3.59	3306	3.19	3741
3.21	5146	2.03	1623
3.63	3885	2.62	3151
2.68	3135	3.09	3958
2.88	3702	2.94	6153
3.06	3863	3.62	6614
3.78	4376	2.89	2487
2.80	2899	2.76	3284
3.21	4082	2.26	1547
2.93	3710	3.22	4188
2.95	4885	2.68	2326
2.68	2796	3.74	5600
2.20	2704	2.97	4507
2.15	2445	3.12	4256
3.00	4833	3.44	4087
2.98	4773	2.61	2590
3.72	5454	3.97	6013

Tabla de distribución normal estándar

Tabla de probabilidades acumuladas de una distribución normal estándar										
Z*	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000
-3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0046	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2110	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

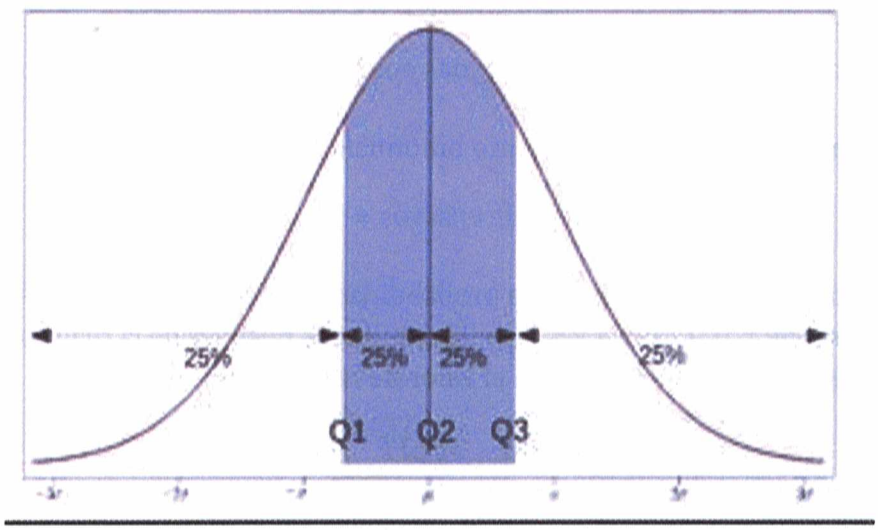
Tabla de probabilidades acumuladas de una distribución normal estándar										
Z*	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	1.0000

Formas estandarizadas para dividir los intervalos de clase

Existen medidas conocidas como cuantiles las cuales agrupan la distribución de frecuencias en partes iguales; entendidas estas como intervalos de clase que comprenden la misma proporción de valores. Los más usados son:

- Los Cuartíles, que dividen a la distribución en cuatro partes (corresponden a los cuantiles 0.25, 0.50 y 0.75)
- Los Quíntiles, que dividen a la distribución en cinco partes (0.20, 0.40, 0.60 y 0.80),
- Los Percentiles, que dividen a la distribución en 100 partes

Como ejemplo de la manera como se divide una distribución de frecuencias en cuantiles observe la siguiente gráfica



Deducción de la fórmula de una distribución binomial

Si estás interesado en conocer la forma en como se deduce la fórmula de la distribución binomial te invito a que leas con atención el siguiente ejemplo:

Supongamos que decidimos registrar el género de cada uno de los próximos 5 niños recién nacidos en un hospital comunitario. ¿Cuál es la probabilidad de que uno de los cinco sea mujer? ¿Cuál es la probabilidad de que dos de los cinco sean mujeres? Antes de contestar estas preguntas fijémonos en de la naturaleza del experimento. Si analizamos la naturaleza de este experimento nos podemos percatar que este posee las siguientes características:

1. El experimento consiste de $n = 5$ intentos (ensayos) idénticos, pues cada intento consiste en registrar el género de un recién nacido.
2. Cada intento da lugar a exactamente a dos resultados, ya que cada recién nacido sólo puede ser registrado como hombre o mujer.
3. Los 5 intentos son independientes pues el resultado obtenido de cada registro es independiente de los resultados obtenidos de los otros registros.
4. La probabilidad de que al registrar el género de un recién nacido sea niña no varia de un registro a otro.

Por las características antes descritas nos encontramos ante la presencia de un experimento binomial. El experimento binomial junto con su distribución de probabilidad fue tratado en la unidad 1. Por lo que ahora solo resta exponer la fórmula que nos permitirá calcular las probabilidades de sus valores asociados.

Si consideramos que el éxito (E) asociado con este experimento es el nacimiento de una niña y que el fracaso (F) está asociado con el nacimiento de un niño, si además suponemos que 40% de todos los nacimientos son niñas. Definimos a la variable

x = número de niñas registradas en los 5 nacimientos observados.

Este experimento es un experimento binomial con

número de ensayos = 5 y $P(E) = 4$

La distribución binomial nos da la probabilidad asociada con cada uno de los posibles valores de x 0, 1, 2, 3, 4 y 5. Hay 32 resultados posibles de la realización de este experimento, y de ellos 5 generan el valor de la variable $x = 1$:

EFFFF FEFFF FFEFF FFFFE FFFEF

Por independencia de eventos, el primero de los resultados posibles tiene una probabilidad de

$$\begin{aligned} P(EFFFF) &= P(E)P(F)P(F)P(F)P(F) \\ &= (.4)(.6)(.6)(.6)(.6) \\ &= (.4)(.6)^4 \\ &= .05184 \end{aligned}$$

Como la probabilidad calculada es la misma para cualquier resultado con un solo éxito ($x = 1$). No importa en qué lugar de la secuencia el éxito se produzca. Así

$$
\begin{aligned}
P(1) &= P(x=1) \\
&= P(EFFF \text{ o } FEFFF \text{ o } FFEFF \text{ o } FFFFE \text{ o } FFFEF) \\
&= .05184 + .05184 + .05184 + .05184 + .05184 \\
&= 5(.05184) \\
&= .25920
\end{aligned}
$$

Similarmente, hay 10 resultados para los cuales x =2, ya que hay 10 maneras de seleccionar dos resultados exitosos entre los cinco ensayos: *EEFFF*, *EFEFF*,. . .,y *FFFEE*. Por independencia de eventos, la probabilidad de cada uno de estos resultados se obtiene al multiplicar entre sí dos veces (. 4) por tres veces (6.). Por ejemplo,

$$
\begin{aligned}
P(EEFFF) &= P(E)P(E)P(F)P(F)P(F) \\
&= (.4)(.4)(.6)(.6)(.6) \\
&= (.4)^2(.6)^3 \\
&= .03456
\end{aligned}
$$

y por lo tanto

$$
\begin{aligned}
P(2) &= P(x=2) \\
&= P(EEFFF) + \cdots + P(FFFEE) \\
&= 10(.4)^2(.6)^3 \\
&= .34560
\end{aligned}
$$

La forma general de la distribución es

$$
\begin{aligned}
P(x) &= P(x) \\
&= \begin{pmatrix} \text{número de resultados} \\ \text{con } x \text{ éxitos} \end{pmatrix} \begin{pmatrix} \text{Probabilidad de cualquier resultado} \\ \text{particular con } x \text{ éxitos} \end{pmatrix} \\
&= (\text{número de resulados con } x \text{ éxitos})(.4)^x(.6)^{3-x}
\end{aligned}
$$

Si n es el número de ensayos en el experimento. Entonces el número resultados con x éxitos, es el número de formas de elegir x éxitos de entre los n ensayos. Una simple expresión de esta cantidad es

$$\text{número de resultados con } x \text{ éxitos} = \frac{n!}{x!(n-x)!}$$

donde, para cualquier número entero positivo n, el símbolo de $n!$ (Léase "n factorial") es definido por

$$n! = n(n-1)(n-2)\ldots(2)(1)$$

y $0! = 1$

Fórmula de la distribución binomial

Sea
n = número de ensayos independientes en un experimento binomial
p = la probabilidad constante de que en cualquier ensayo particular se obtenga un éxito

Entonces

$$P(x) = P(x \text{ éxitos en } n \text{ ensayos}) = \frac{n!}{x!(n-x)} p^x (1-p)^{n-x}$$
$$x = 0, 1, \ldots, n$$

La expresión $\binom{n}{x}$ algunas veces es utilizada en lugar de

$$\frac{n!}{x!(n-x)},$$

cuya expresión representa el número de maneras de elegir x elementos de un conjunto de n elementos.

La fórmula de la probabilidad binomial puede entonces ser escrita como

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$x = 0, 1, \ldots, n$$

La consulta en fuentes de información por Internet

La información es un punto nodal para la sociedad de hoy. Diferenciarla, manejarla y utilizarla son acciones básicas para nosotros los miembros de la sociedad del siglo XXI y por ello hay que acercarse a ella. Saber qué hacer es el primer paso.

La información se define como el conjunto de datos sobre algún fenómeno determinado; se obtiene de diversas formas, como la observación o la búsqueda intencionada. En el primer caso es natural pero en el segundo no. Para aprender se utilizan las dos pero para estudiar se usa principalmente la segunda.

La información se obtiene de fuentes primarias y secundarias, escritas, orales y visuales, mediante medios impresos, electrónicos y personales. El conjunto de datos por obtener es tan amplio que después de obtenidos se deben analizar, pues no todo lo percibido o encontrado es certero y confiable y tampoco responde de manera puntual al objeto de estudio.

En estos días es común el acceso a la información a través de Internet o red global de información a la que se llega y se mantiene por medio de computadoras. Son millones y millones de datos, documentos, imágenes, fotografías lo que se almacena y a lo que uno tiene acceso. Por eso, diferenciar entre una buena información y la información basura es difícil. Los siguientes son algunos consejos o recomendaciones para guiar tu búsqueda.

1. Para distinguir el valor de la información para ti debes planear el objetivo antes de comenzar a buscar. Los siguientes criterios de búsqueda pueden ayudarte: ¿qué voy a buscar?, ¿qué quiero saber de lo que voy a buscar?, ¿para qué lo estoy buscando?
2. Es muy importante que no busques saber TODO de un tema. Entre más específica sea tu búsqueda, mayor oportunidad tienes de encontrar rápida y fácilmente la información. Puedes caer en dos errores:
   a) Especificar demasiado las cosas.
   b) Dejar sin especificar las cosas.
3. Define qué sabes. Para comenzar a investigar hay que partir de tus conocimientos previos. Lo que ya conoces te servirá para realizar tu investigación y para diferenciar datos correctos de los incorrectos, los útiles de los inútiles.
   a) Asegúrate que la información que tú conoces previamente es correcta.
   b) Asegúrate que la información que es actual.
   c) Recuerda que, aunque no sepas del tema, sí sabes cómo comenzar a buscarlo.
4. Decide dónde y cómo vas a buscar.
5. Pregúntate: ¿qué palabras voy a utilizar?, ¿qué criterios de búsqueda? Tienes que enlistar las palabras clave para tu búsqueda. Conforme avances, agrega más palabras clave.
6. Planea la búsqueda de acuerdo a tu nivel de conocimientos: vas a investigar algo muy básico o más avanzado. Los mejores lugares para comenzar a infor-

marte son diccionarios, enciclopedias, las lecturas sugeridas en los libros de texto, las páginas de Internet "oficiales" (aquellas del gobierno, de las organizaciones importantes (como la ONU, la UNICEF), páginas de universidades de prestigio (como la UNAM, el IPN) Estas páginas "oficiales" tienen CONTROL sobre sus contenidos por lo que la información encontrada, aunque puede ser subjetiva (que depende de un punto de vista), es la "oficialmente correcta".

Es muy importante que pongas MUCHA ATENCIÓN en tus primeras lecturas. Debes encontrar información correcta. Para ello es necesario que compares los datos obtenidos entre sí.

7. Busca y consulta la información utilizando un buscador (el que te va a encontrar dónde, de todo el Internet, está tu tema).

   Algunos buscadores son:
   - mx.yahoo.com
   - www.google.com.mx
   - mx.altavista.com

   Si quieres noticias probablemente las encuentres en:
   - www.bbc.co.uk/mundo/index.shtml
   - mx.reuters.com
   - mx.news.yahoo.com

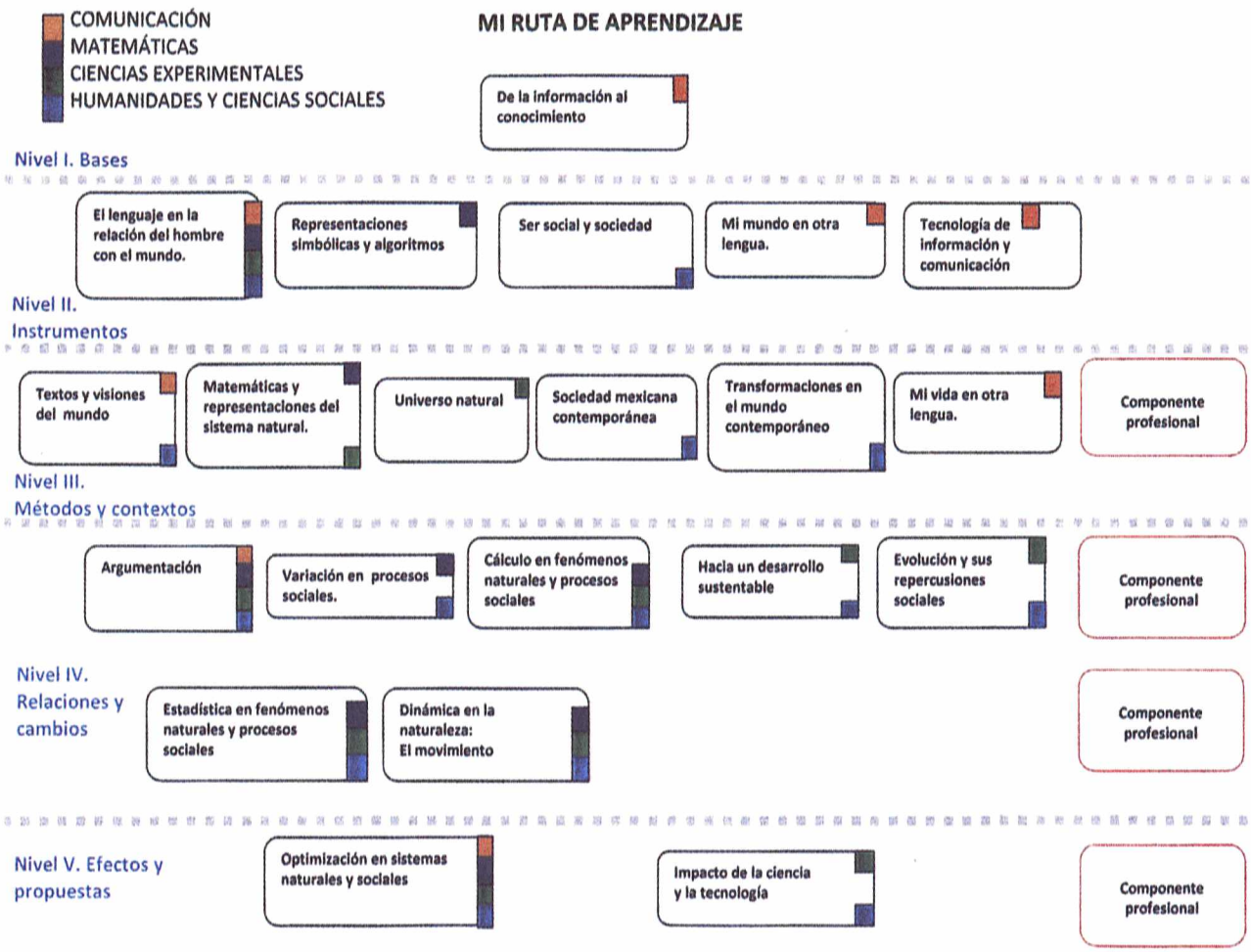
   Si buscas libros los puedes encontrar (además de en una librería) en:
   - books.google.es
   - www.booksfactory.com/indice.html
   - www.ucm.es/BUCM/atencion/25403.php

   Si lo que deseas son diccionarios:
   - rae.es/rae.html
   - www.diccionarios.com
   - www.elmundo.es/diccionarios

   ¿Qué opciones del buscador me conviene utilizar?

   Los buscadores presentan algunas opciones tales como:
   - Opciones de Búsqueda: Incluye "buscar videos", "buscar imágenes", "buscar noticias", "búsqueda en español", "búsqueda en México" etc. Lo que hacen es especificar tu búsqueda.
   - Dentro de "búsqueda avanzada" podrás elegir cómo preferirías que te ayudara a buscar. Utilizando las opciones de: "buscar con las palabras" y "que no contenga las palabras" puedes hacer tu búsqueda aún más pequeña y te será más fácil encontrar lo que quieres.
8. Una vez obtenida la información: analiza. Los puntos más importantes ahora son: ¿es lo que necesito?, ¿qué tan bueno es el contenido?, ¿qué tan confiable es el autor?, ¿cuáles son algunos lugares de donde viene la información?

Rodrigo Zepeda Tello. "Guía básica para el manejo de Internet", en Liliana Almeida *et al.* (2011). *Ciencia Contemporánea ¿Para qué?* México: Edere/Esfinge, pp. 142-148.


COMUNICACIÓN
MATEMÁTICAS
CIENCIAS EXPERIMENTALES
HUMANIDADES Y CIENCIAS SOCIALES
MI RUTA DE APRENDIZAJE
De la información al conocimiento
Nivel I. Bases
El lenguaje en la relación del hombre con el mundo.
Representaciones simbólicas y algoritmos
Ser social y sociedad
Mi mundo en otra lengua.
Tecnología de información y comunicación
Nivel II. Instrumentos
Textos y visiones del mundo
Matemáticas y representaciones del sistema natural.
Universo natural
Sociedad mexicana contemporánea
Transformaciones en el mundo contemporáneo
Mi vida en otra lengua.
Componente profesional
Nivel III. Métodos y contextos
Argumentación
Variación en procesos sociales.
Cálculo en fenómenos naturales y procesos sociales
Hacia un desarrollo sustentable
Evolución y sus repercusiones sociales
Componente profesional
Nivel IV. Relaciones y cambios
Estadística en fenómenos naturales y procesos sociales
Dinámica en la naturaleza: El movimiento
Componente profesional
Nivel V. Efectos y propuestas
Optimización en sistemas naturales y sociales
Impacto de la ciencia y la tecnología
Componente profesional

Cáceres Hernández, J. (2007). *Conceptos básicos de estadística para ciencias sociales.* España: Delta/Publicaciones universitarias.

Cobo, E., P. Muñoz y J. A. González (2007). *Bioestadística para no estadísticos.* España: Elsevier Masson.

Meyer Paul, L. (1992). *Probabilidad y aplicaciones estadísticas.* México: Fondo Educativo Interamericano.

Morales C. y S. Parada (2005). *Pobreza, desertificación y degradación de los recursos naturales. Comisión Económica para América Latina y el Caribe.* Cesar Morales y Soledad Parada Editores. Naciones Unidas.

Disponible en: http://www.eclac.org/cgi-bin/getProd.asp?xml=/publicaciones/xml/8/24268/P24268.xml&xsl=/ddpe/tpl/p9f.xsl&base=/ddpe/tpl/top-bottom.xsl

Peck, R. y J. L. Devore (2012). *Statistics: The Exploration & Analysis of Data.* Boston: Brooks/Cole/Cengage Learning.

Weimer, R. C. (1996). *Estadística.* México: CECSA.

Sánchez, E. (2008). *Probabilidad y estadística I.* México: Publicaciones Cultural.

Spiegel, M. R. (2003). *Estadística* México: McGraw-Hill (Serie Schaum).

Estadística en Fenómenos Naturales y Procesos Sociales



ISBN: 978-607-8229-42-0



Secretaría de Educación Pública
Subsecretaría de Educación Media Superior
Dirección General del Bachillerato