Juan Alberto Cervantes Pasqualli

Probabilidad y estadística

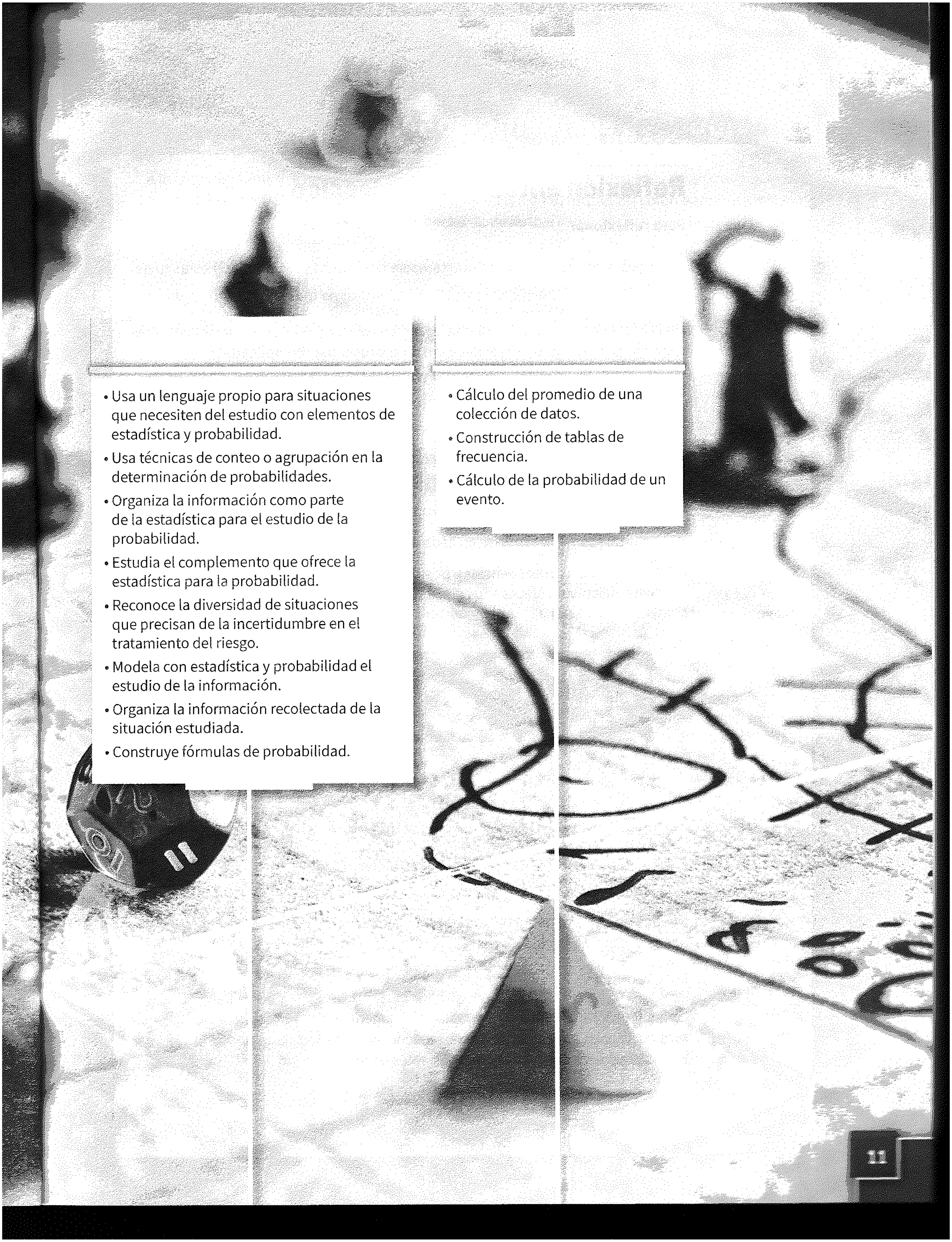
Primer parcial

Eje: Del manejo de la información al pensamiento estocástico

- Riesgo, inferencia y aleatoriedad: elementos de la estadística y la probabilidad.

- Conceptos básicos de estadística y probabilidad.
- Recolección de datos y su clasificación en clases.
- Uso del conteo y la probabilidad para eventos.
- Concepto de riesgo en situaciones contextuales.
- Contextualización de los elementos de probabilidad condicional e interpretación intuitiva del teorema de Bayes (probabilidad subjetiva).

- Nociones y conceptos básicos de estadística y probabilidad.
- Enfoques de probabilidad. ¿Qué significa cada enfoque de probabilidad?, ¿qué significan las medidas de tendencia central?, ¿para qué obtener estos valores?
- Técnicas de conteo y agrupación en clases para la determinación de probabilidades.
- ¿Qué es el riesgo?, ¿qué papel juega la probabilidad y estadística en el estudio del riesgo?
- Usos de la estadística y probabilidad en situaciones dadas.
- Análisis de la información.
- Nociones de incertidumbre, azar y aleatoriedad.
- Tipos de eventos en el estudio de la probabilidad.



- Usa un lenguaje propio para situaciones que necesiten del estudio con elementos de estadística y probabilidad.
- Usa técnicas de conteo o agrupación en la determinación de probabilidades.
- Organiza la información como parte de la estadística para el estudio de la probabilidad.
- Estudia el complemento que ofrece la estadística para la probabilidad.
- Reconoce la diversidad de situaciones que precisan de la incertidumbre en el tratamiento del riesgo.
- Modela con estadística y probabilidad el estudio de la información.
- Organiza la información recolectada de la situación estudiada.
- Construye fórmulas de probabilidad.

- Cálculo del promedio de una colección de datos.
- Construcción de tablas de frecuencia.
- Cálculo de la probabilidad de un evento.

Reflexión antes de una elección

Para reflexionar

¿Recuerdas situaciones en las que tomaste decisiones precipitadas y estas tuvieron consecuencias desagradables?

¿Sabías que hay estrategias que te permiten pensar y revisar las posibles consecuencias para tomar la mejor decisión ante situaciones complicadas?

Paso a paso:

1. Identifica una situación sobre la cual debes tomar una decisión.
2. Elabora una tabla con la información: detonador (situación), disyuntiva, pros, contras, y evaluación general y decisión.
3. Valora las consecuencias de cada opción a corto, mediano y largo plazo.
4. Analiza: ¿cómo te sentirías en cada una de las decisiones?
5. Considera las consecuencias y posibles resultados: ¿te afecta de manera positiva o negativa? ¿Afecta a tus amistades o a tu familia?
6. Con base en el análisis, toma la decisión que consideres más adecuada.

Detonador:			
Disyuntiva (opciones)	Pros	Contras	Evaluación general y decisión

Puedes utilizar esta tabla para analizar consecuencias en diversos temas, ya sean académicos o de la vida diaria.

Para terminar...

La habilidad para identificar y evaluar críticamente las repercusiones de tus posibles decisiones es el análisis de consecuencias. Al desarrollarla podrás considerar los resultados de diversas alternativas antes de tomar alguna decisión.

Proyecto de vida

A lo largo del semestre desarrollaremos tu Proyecto de vida, con el fin de clarificar lo que deseas para ti y que puedas tomar decisiones que marquen la dirección de tu futuro, así como hacer una reflexión sobre las implicaciones que tiene en tu vida el hecho de llevarlo a cabo.

Organiza la información de tu Proyecto de vida.

1. Copia y completa en tu cuaderno el siguiente cuadro.
2. Escribe en la primera columna las metas más importantes que deseas lograr. Agrega las filas que sean necesarias.

Meta	Ideas importantes a rescatar	Acciones que debo efectuar	Beneficios que representa para mí	Beneficios que representa para mi entorno

3. Una vez que organizaste la información en la tabla, léela con atención para que tengas un panorama general de lo que implica el planteamiento de tu Proyecto de vida.
4. Escribe por qué consideras que es importante elaborar un Proyecto de vida.
5. Elabora un organizador gráfico en el que incorpores frases que te motiven a trabajar día con día en tu Proyecto de vida, incluye tus metas para que las tengas presentes todo el tiempo. Coloca este organizador en un lugar visible en tu casa. Utiliza el siguiente esquema para hacer un esbozo y ve enriqueciéndolo paulatinamente.

Proyecto de vida de: ____________________

	Metas a corto plazo	Metas a mediano plazo	Metas a largo plazo
Personal			
Familiar			
Escolar			
Laboral			
Social			

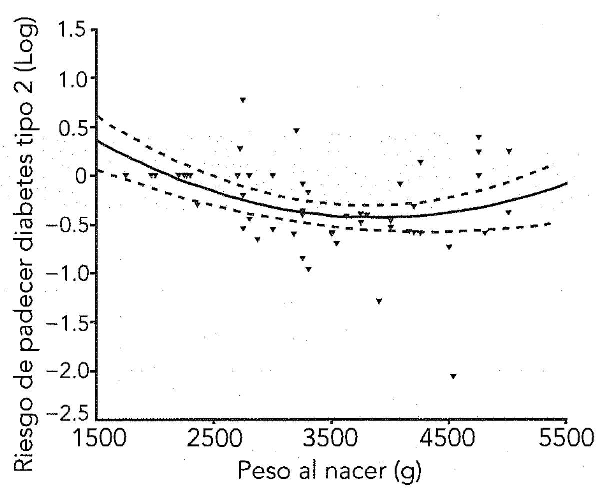
Conceptos básicos de estadística y probabilidad

Nociones de estadística y probabilidad

La estadística en la vida diaria

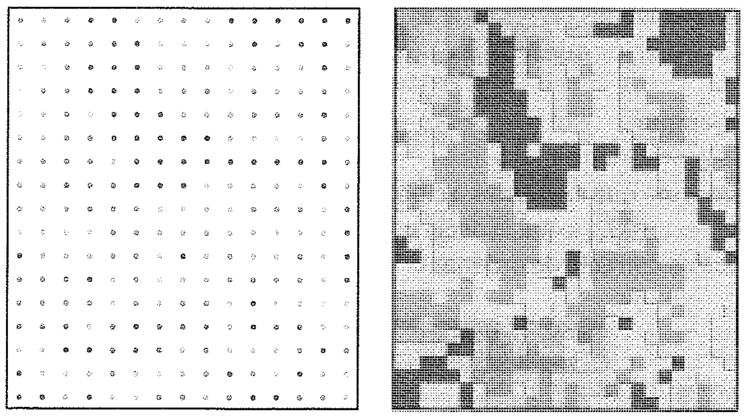
En una playa, es más probable que mueras por la caída de un coco en la cabeza que por un ataque de tiburón. Los hombres son, en promedio, más altos que las mujeres. En 2016, las personas más buscadas en Google fueron Michael Phelps, Hillary Clinton y Donald Trump. En el mundo se produce comida para alimentar once mil millones de personas, sin embargo, somos siete mil millones y hay gente muriendo de hambre.

En enero de 2017 en México, de cada 1000 personas mayores de 15 años y con la capacidad de trabajar, 36 se encontraban desempleadas. En los hogares mexicanos tres de cada diez personas mayores de 12 años se han sentido deprimidos en algún momento de sus vidas. Las aspirinas disminuyen la probabilidad de que padezcas enfermedades cardiovasculares. Absolutamente todas estas afirmaciones provienen de la exploración y algunas de ellas del análisis de información contenida en bases de datos. La estadística es una rama de las matemáticas que recopila, ordena, explora y analiza información con el objetivo de hacer inferencias, por lo general, útiles.



La estadística es una herramienta utilizada en todas las ciencias y en muchos otros ámbitos; la versatilidad de los campos en los que se utiliza se debe a la capacidad que tienen de sintetizar y analizar bases de datos enormes, e inclusive varias bases de datos en un solo análisis. Por ejemplo, en una investigación publicada en 2007 sobre la relación que hay entre el peso al nacer y el riesgo de padecer diabetes tipo 2, un grupo de investigadores en Alemania analizó 14 bases de datos que acumulaban los registros de 132 180 personas. Ellos encontraron que los recién nacidos con bajo y alto peso al nacer tienen mayor riesgo que los que tienen un peso promedio. (bkmrt.com/nlVDFX)

La estadística es una herramienta sumamente poderosa debido a las decisiones que se toman con base en las inferencias que se pueden hacer sobre los resultados. Por ejemplo, las compañías mineras utilizan la estadística para decidir dónde cavar y de esta manera ahorrarse millones de pesos; si no la utilizaran perderían mucho dinero cavando en lugares donde no sacarían metal o muy poco. Un ejemplo de lo que hacen estas compañías lo puedes observar en la siguiente figura. Primero, toman muestras de suelo profundo y las analizan para ver las cantidades de metal que contienen, en el cuadro de la izquierda cada punto representa una muestra. Con esta información utilizan la estadística e infieren la distribución del metal en toda el área que muestrearon (cuadro de la derecha) y entonces deciden dónde cavar minimizando el costo de excavación y maximizando la ganancia.



Otros ejemplos de su aplicación son los siguientes: la usan los médicos para considerar tu predisposición genética a ciertos males o enfermedades; las farmacéuticas para probar que un medicamento es más efectivo que perjudicial y poder venderlo; los biólogos para descubrir los patrones que se esconden en las relaciones entre especies en los ecosistemas; los químicos para probar que los resultados de sus experimentos son confiables y no son resultados que pudieron haber obtenido por puro azar; los epidemiólogos para predecir y prevenir epidemias; los políticos para conocer las preferencias electorales; los casinos para detectar a los estafadores; cualquier compañía para monitorear las demandas de sus clientes y ganar más dinero; las oficinas de tránsito para coordinar los semáforos y agilizar el flujo vehicular; la utiliza Facebook para monitorear tus likes y bombardearte con publicidad compatible con tus gustos, también para analizar tus contactos y sugerirte "amigos".

Quizás te preguntes si usas la estadística en tu vida diaria. Bueno, pues cada vez que te preguntas cosas como: ¿cuánto dinero gasto en promedio al día?, ¿existe alguna relación entre el tiempo que dedico a jugar videojuegos y mis calificaciones?, te estás haciendo preguntas estadísticas. Y si no eres de los que se hacen estas preguntas, me temo que eres una persona poco curiosa, pero que seguramente vive de las decisiones que se toman a partir del uso de la estadística. Por ejemplo, el sistema educativo en teoría está basado en las necesidades del país y es modificado a partir de las deficiencias académicas que se detectan en toda la República.

La estadística se divide en dos áreas, la estadística descriptiva y la inferencial. La primera aborda lo relacionado con recopilar y ordenar información en bases de datos, así como el cálculo de valores sencillos que describen de manera muy general la información en esas bases. Parte de la estadística descriptiva consiste en la elaboración de una gran variedad de gráficos que muestren un resumen de toda la información recabada. Por otro lado, la estadística inferencial se dedica a estudiar las relaciones entre las variables de una base de datos y a la construcción de modelos que permitan predecir el comportamiento de alguna de las variables estudiadas para poder hacer deducciones útiles y tomar decisiones.

En otras palabras, la estadística descriptiva responde preguntas como: ¿cuál es la altura promedio de mis compañeros de clase?, ¿cuál es la fecha más común de cumpleaños en mi escuela?, ¿qué tan variable es el peso de los profesores en mi escuela?, ¿cuántas mascotas tienen en promedio las casas de mi colonia?, ¿cuál es la mascota más común?, ¿cuánta agua en promedio toman los miembros de mi familia? La estadística inferencial, por su parte, responde preguntas como: ¿qué relación hay entre la altura y el peso de mis compañeros de clase?, ¿los hombres son más altos que las mujeres?, ¿existe una relación entre la cantidad de agua que ingiero con el número de dolores de cabeza que tengo al mes?, ¿cuál sería mi calificación en un examen de opción múltiple si respondiera totalmente al azar?

Antes de empezar de lleno con conceptos estadísticos es importante que sepas qué es una base de datos y las partes básicas que la conforman. Esto es importante, pues si no sabes acomodar la información que recabas, no podrás analizarla adecuadamente.

Ejemplo

Imagina que encuestas a cinco compañeros de tu clase y les preguntas: ¿cuántas mascotas tienen en su casa?, ¿a qué especie pertenecen?, ¿cuál es su sexo?, ¿cuántos años tiene cada mascota? La forma correcta de ordenar la información recabada por esa encuesta sería:

Compañero	Mascota	Especie	Sexo	Edad
Daniel	Brutus	perro	M	11
Daniel	Caníbal	hurón	M	2
Daniel	Zeus	perro	M	4
Daniel	Tuna	cuyo	F	1
Ana	Kiki	gato	F	11
Emanuel	Baco	gato	M	7
Emanuel	Tobías	tortuga	M	12
Joab	Chomak	perro	F	5
Joab	Luna	perro	F	12
Enrique	Ra	perro	F	8
Enrique	Chela	perro	F	5
Enrique	Darwin	cuyo	M	3

Esta tabla es una base de datos de cinco compañeros encuestados, Daniel, Ana, Emanuel, Joab y Enrique, en casa de Daniel tienen cuatro mascotas, en la de Ana sólo una, en la de Emanuel dos, en la de Joab dos y en la de Enrique tres. En una base de datos cada fila es un dato y cada columna una variable, por lo tanto, esta base de datos tiene doce datos y cinco variables. Más adelante veremos los diferentes tipos de variables y lo que puedes hacer con ellas.

Conceptos básicos de estadística

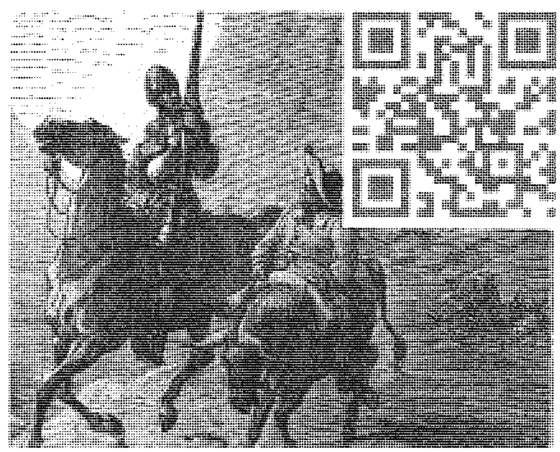
Es común que el estudiante tenga la impresión de que la estadística es una materia difícil o confusa. Esto se debe principalmente a que existe una gran diversidad de formatos en que se enseña y a que los conceptos estadísticos giran alrededor de la pregunta que se quiera responder. Quizás esto no parezca tener mucho sentido, así que te pondré un ejemplo con la base de datos de la página anterior.

Si la pregunta de esa investigación fuera solamente ¿cuántas mascotas viven en las casas de mis compañeros?, la base de datos sería como la de la derecha. Sin embargo, la pregunta original no sólo es para conocer el número de mascotas sino también su especie, edad y sexo, por lo que la base de datos debe mostrar también toda esa información. Si te fijas, la variable "Mascotas" se expresa de maneras muy diferentes en ambas bases. En la primera base se trata de una variable con los nombres de las mascotas y en la de la derecha como el número de mascotas. El número de datos también cambió, en la primera hay doce y en esta sólo cinco. Todos esos cambios están relacionados con la pregunta que antecede a la investigación y lo mismo sucede con los términos estadísticos que veremos a continuación.

Compañero	Mascotas
Daniel	4
Ana	1
Emanuel	2
Joab	2
Enrique	3

Ejemplo

Cada uno de los siguientes conceptos estará ejemplificado con el siguiente caso de investigación: un estudiante muy curioso de cuyo nombre no quiero acordarme quiere comprobar que en su colonia los hombres adultos son más altos que las mujeres adultas. Para esto, mide la estatura (en metros) de 50 hombres y 50 mujeres mayores de 18 años en su colonia. Los resultados de su trabajo lo puedes ver y descargar escaneando con tu celular el código QR que aparece a la derecha.



Población

La población es el universo de datos que se estudian. El número total de datos que conforman una población a menudo se denota con la letra ene mayúscula (N). Estos datos pueden ser personas, lugares, cosas, épocas, animales, drogas, bancas, lápices, planetas, patinetas, actores, videojuegos o muchas cosas más. Toda la estadística se trata de estimar y estudiar las propiedades numéricas de una población.

En el último ejemplo (el de Alonso Quijano), la población sería toda la gente adulta (mayor a 18 años) de su colonia, pues su pregunta se enfoca específicamente en la diferencia de alturas entre sexos en su colonia. Si la pregunta de Alonso fuera ¿los hombres son más altos que las mujeres? Su población sería toda la gente de todas las edades, de todas las razas y de todos los países.

Muestra

La muestra es una fracción de la población. Casi nunca en una investigación donde se ocupa la estadística se mide o se registra a todo el universo de datos que existen, principalmente porque es imposible, o muy costoso, o innecesario. La característica más importante que debe tener una muestra, es que sea tomada aleatoriamente, de no hacerlo así corres el riesgo de estar obteniendo resultados espurios.

Siguiendo con el ejemplo de Alonso Quijano, su muestra es la gente a la que midió, es decir, 100 individuos, pues de toda la gente adulta que vive en su colonia (población) él midió únicamente a 50 hombres y 50 mujeres. Como Alonso es muy inteligente, eligió aleatoriamente a las personas que mediría. Puso papelitos con los nombres de las personas adultas que viven en su colonia en una tómbola y sacó 50 de hombres y 50 de mujeres.

¿Por qué es importante que el muestreo sea aleatorio? Imagínate que en su colonia viven personas de raza blanca (altas) y personas de raza morena (chaparras) y que Alonso Quijano prefiere la compañía de gente de raza morena. Si él midiera sólo a personas con las que se junta, mediría en su mayoría a gente de raza morena y sus datos no estarían reflejando la realidad de la altura que tienen las personas adultas en su colonia.

Tamaño de la muestra

Se le dice así al número de datos o registros en una muestra de una población, y por lo general se denota con una ene minúscula (n). Una característica muy importante que debe tener el tamaño de muestra es su representatividad. Es decir, el tamaño de muestra debe ser lo suficientemente grande como para asegurar que los datos servirán para responder la pregunta de la investigación. Existen procesos estadísticos que sirven para definir el tamaño de muestra ideal (que no sea ni muy grande ni muy pequeño) para determinada investigación, sin embargo, para fines didácticos a nivel preparatoria es recomendable un tamaño de muestra de 100 datos o registros y un tamaño mínimo de 50.

En el ejemplo, el tamaño de muestra que el ilustre caballero Alonso Quijano decidió utilizar fue de 100 individuos, 50 hombres y 50 mujeres. ¿Por qué es importante la representatividad en el tamaño de muestra? Imagínate que Alonso hubiera medido aleatoriamente sólo a tres hombres y a tres mujeres, y que por casualidad midiera a las tres mujeres más altas y a los tres hombres más chaparros. Sus resultados estarían severamente sesgados por su tamaño de muestra tan pequeño y no podría responder acertadamente su pregunta.

Actividad de aprendizaje 1

Efectúa lo que se te pide.

1. Debes saber que actualmente es meramente didáctico hacer estadística en un cuaderno; para fines aplicados, la estadística se hace en computadoras. Practica el uso de las tecnologías capturando tu base de datos en Excel y conserva el archivo, pues lo necesitarás más adelante. Escribe en tu cuaderno cuatro preguntas que atiende la estadística descriptiva y cuatro preguntas que se respondan utilizando estadística inferencial. Coméntalas con dos de tus compañeros y discutan si todas son correctas.

2. Investiga de qué manera la estadística está relacionada con tu pasatiempo favorito. Escribe una carta de una cuartilla donde le expliques a un ser querido la importancia de la estadística en tu pasatiempo favorito.

   a. Formen equipos de tres personas y compartan entre ustedes el resultado de su investigación. Elijan uno de los tres pasatiempos y generen una base de datos de al menos cuatro variables que podría llenarse utilizando la información del pasatiempo que eligieron.

   b. Llenen la base de datos hasta 30 registros (filas), detecten las complicaciones que tuvieron para llenar la base.

   c. Expongan al grupo sus resultados, incluyendo la base de datos, lo que aprendieron de haber colectado la información, las complicaciones que tuvieron y por último pregúntenles a sus compañeros si ellos consideran que hubo algo relevante que ustedes pudieron haber omitido.

3. ¿Puedes pensar en algo relacionado con la existencia humana que no pueda ser registrable en una base de datos? ¿Será que es imposible no ser parte de una estadística? Comenta tus reflexiones con alguien que piense diferente.

4. Sobre tu animal favorito escribe una pregunta que atienda la estadística descriptiva y otra que atienda la estadística inferencial. Genera una base de datos de 15 registros (filas) con las variables que crees que necesites para contestar tus preguntas. Anota las dificultades que tuviste para pensar en las preguntas y las variables que responderían tus preguntas. Por último, intercambia tu trabajo con otros dos compañeros, revisa y comenta su trabajo considerando los siguientes puntos:

   a. ¿Sus preguntas están correctamente planteadas (una descriptiva y la otra inferencial)?

   b. ¿Consideras que las variables que seleccionó tu compañero serían suficientes para responder sus preguntas?

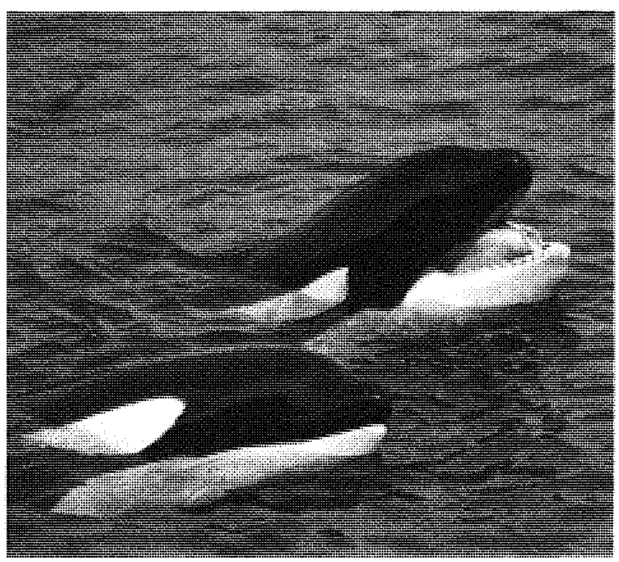
   c. ¿Coincidieron tus dificultades con las de tus compañeros?

5. Plantea una investigación donde tu población sean todas las orcas del planeta.

6. Plantea una investigación donde la población sea sólo una fracción de las orcas en el planeta.

7. Imagina que quieres saber el tamaño de las palomas que viven en la catedral de la Ciudad de México, ¿cuál sería tu población y cuál tu muestra?

8. Imagina que quieres saber el tamaño de las palomas que viven en todas las catedrales de toda la República Mexicana, ¿cuál sería tu población y cuál tu muestra?



Enfoques de probabilidad. ¿Qué significa cada enfoque de probabilidad?

Una probabilidad se define como la posibilidad de que suceda un determinado evento. Una probabilidad se expresa como un número que va de cero a uno, donde "cero" significa que no hay forma de que suceda el evento y "uno" que tienes toda la certeza de que sucederá. Sin embargo, es común que en algunos textos se exprese como un porcentaje o fracción equivalente. Por ejemplo, una probabilidad de 0.2 la puedes encontrar como 20% o 1/5.

Probabilidad clásica

Es cuando la probabilidad de un evento es determinada por la proporción que ocupa ese evento del total de resultados posibles. El ejemplo más sencillo de esto es el de un volado, en él hay dos posibles resultados (águila o sol), por lo tanto, la probabilidad de que caiga águila será una de dos, es decir 1/2, o sea que la probabilidad de que caiga águila en un volado es de 0.5.

La expresión matemática que se utiliza para calcular una probabilidad es:

$$P(A)=\frac{n}{N},$$

donde *P* es la probabilidad, *A* el evento del que se quiere conocer su probabilidad, *n* el número de resultados favorables y *N* el total de posibles resultados.

Ejemplo

Ahora imagina que apuestas 15 pesos a que el resultado de tirar un dado de seis caras será un número par. Ahora tienes que tres de las caras del dado son resultados favorables para ti (2, 4 y 6), por lo tanto, la probabilidad de que ganes es tres de cada seis, aplicando la fórmula tienes:

$$P=\frac{n}{N}=\frac{3}{6}=0.5.$$

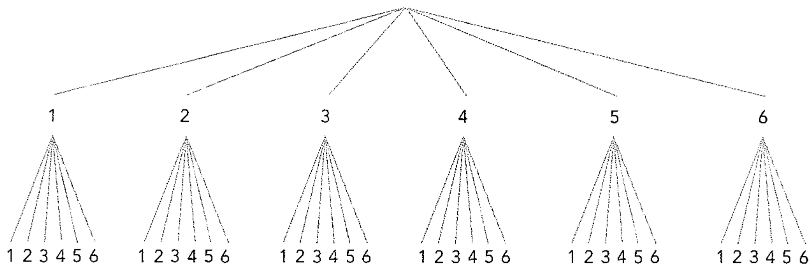
Ejemplo

Dado 1

		1	2	3	4	5	6
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
Dado 2	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Darwin y Wallace están jugando un juego de azar. El juego se trata de lanzar dos dados de seis caras y los jugadores apuestan al total que sumen ambos dados. Darwin apuesta a que la suma será de nueve y Wallace apuesta a que será de siete, ¿quién tiene una mayor probabilidad de ganar?

Hay 36 posibles maneras en las que puede resultar la tirada de dos dados, sin embargo, la apuesta es sobre la suma de ellos, por lo tanto, habría que sumar cada uno de los 36 posibles resultados. En el cuadro de la página anterior puedes observar las 36 posibles sumas, de las cuales cuatro suman nueve. Usando la fórmula tenemos que: $P = \frac{4}{36} = 0.\overline{1}$.



Por otra parte, hay seis de 36 formas en las que los dados sumarían siete, por lo que tendríamos:

$$P = \frac{6}{36} = 0.1\overline{6}.$$

Darwin tiene una probabilidad de ganar la apuesta de 0.11 mientras que la de Wallace es de 0.16. Entonces, Wallace tiene mayor probabilidad de ganar en este juego.

Probabilidad empírica

No en todos los eventos se pueden conocer todos los posibles resultados. Sin embargo, puedes ocupar el registro histórico de los resultados que ya sucedieron para calcular la probabilidad de que ocurra cualquiera de ellos. Un ejemplo sencillo sería calcular la probabilidad de que tu profesor(a) de probabilidad y estadística se presente en clase, si de las últimas 23 clases sólo faltó una vez, la probabilidad de que se presente en clase sería de 22/23, o sea 0.95.

Ejemplo

Los últimos 30 días de clase, un profesor de derecho se ha puesto 16 días una corbata azul, 10 días una corbata gris y cuatro días una corbata rosa. ¿Cuál es la probabilidad de que el siguiente día traiga puesta una corbata gris? Si de los últimos 30 días, diez se ha puesto una corbata gris, entonces la probabilidad es 10 de 30, es decir 10/30 = 0.333. Existe una probabilidad de 0.333 de que el profesor se ponga una corbata gris.

Probabilidad subjetiva

Esta probabilidad no es calculable como las dos anteriores, si no que depende de la opinión de un experto en el tema. Son, a veces, probabilidades muy controversiales, pues pueden variar mucho de un experto a otro.

Ejemplo

La probabilidad de que en 2020 las elecciones presidenciales en Estados Unidos las gane Donald Trump no es calculable. Sin embargo, politólogos expertos podrían opinar y dar a conocer su pronóstico, que sería una probabilidad totalmente subjetiva.

Actividad de aprendizaje 2

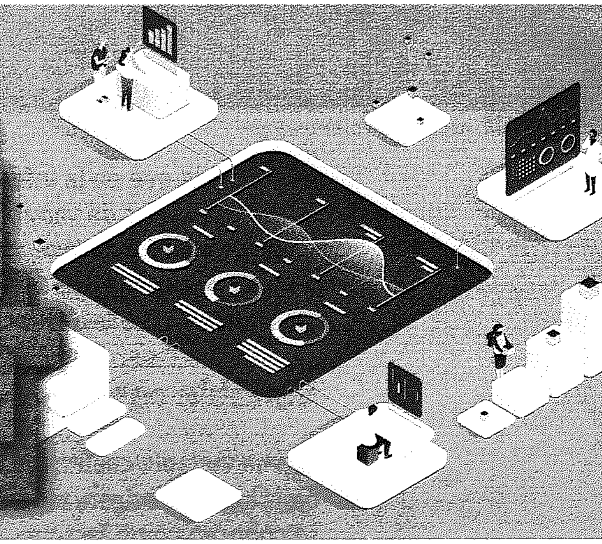
Efectúa lo que se te pide.

1. Investiga de qué manera la estadística está relacionada con tu pasatiempo favorito. Escribe una carta de una cuartilla donde le expliques a un ser querido la importancia de la estadística en tu pasatiempo favorito.
2. Formen equipos de tres personas y compartan entre ustedes el resultado de su investigación. Elijan uno de los tres pasatiempos y generen una base de datos de al menos cuatro variables que podría llenarse utilizando la información del pasatiempo que eligieron.
3. Llenen la base de datos hasta 30 registros (filas), detecten las complicaciones que tuvieron para llenar la base. Expongan al grupo sus resultados, incluyendo la base de datos, lo que aprendieron de haber colectado la información, las complicaciones que tuvieron y por último pregúntenles a sus compañeros si ellos consideran que hubo algo relevante que ustedes pudieron haber omitido.
4. ¿Puedes pensar en algo relacionado con la existencia humana que no pueda ser registrable en una base de datos? ¿Será que es imposible no ser parte de una estadística? Comenta tus reflexiones con alguien que piense diferente.

Actividad de aprendizaje 3

Efectúa lo que se te pide.

1. Entrevista por lo menos a 7 compañeros de tu salón y pregúntales lo siguiente:
   a. ¿Cuántas personas viven en su casa?
   b. ¿Cuál es el sexo de cada uno?
   c. ¿Cuántos años tiene cada uno?
   d. Máximo grado de estudio que tienen.
2. Busca en distintas fuentes de información o en la web sobre el concepto de estadística, estadística descriptiva y estadística inferencial. Registra la información en tu cuaderno. Comenta en equipo la información que obtuviste analizando los elementos que integran el concepto de estadística, destaca la diferencia entre estadística descriptiva e inferencial.
3. Efectúa en equipo una investigación de campo, entrevistando cada uno de manera aleatoria, a por lo menos 10 vecinos de su casa. Recaba la información y regístrala en una base de datos:
   a. Productos que más consumen en una semana, número de veces que consumen cada producto en un día, costo de cada producto. ¿Cuál sería una pregunta de estadística descriptiva que podrías responder con la información recabada? ¿Cuál sería una pregunta de estadística inferencial que podrías responder con la misma información?
   b. El número y tipo de lámparas que tiene su casa, potencia en watts de la lámpara y el costo del recibo de luz por bimestre. A partir de los datos comenta con tu equipo los beneficios y perjuicios para las familias encuestadas acerca de la información obtenida.
4. Identifica las diferentes actividades deportivas que hay en tu escuela. ¿Qué información de esas actividades se puede registrar en una base de datos?



Recolección de datos y su clasificación en clases

¿Qué es un dato?

Un dato se traduce como un registro o toda una fila en una base de datos. Puede resultar poquito confuso, pues en México usamos la palabra dato en el lenguaje cotidiano para referirnos a un número o palabra. Sin embargo, en estadística un dato no es sólo un número o una palabra, es toda la información que hayas recabado sobre un elemento de tu población. Si le mediste 33 cosas a las orcas de tu investigación en la página 19, no tienes 33 datos por orca, sino que cada orca es un dato. Al concepto de dato también se le puede encontrar en algunos libros y páginas de internet como "unidad muestral", "unidad de muestra", "registro", "unidad básica", "individuo" y "elemento".

Una variable es un valor numérico o una característica que puede variar de un dato a otro. En términos prácticos es una característica o algo que le mides a cada elemento de tu muestra. Es importante que recuerdes que cada variable se convertirá en una columna cuando generes tu base de datos. Existen distintos tipos de variables y, dependiendo de la fuente de información, puedes encontrarlas clasificadas de manera diferente, por esta razón algunas personas batallan para aprender y entender los tipos de variables y cómo se clasifican.

Para fines prácticos y considerando el temario del curso, veremos tres tipos de variables –categóricas, discretas y continuas –, que se clasifican en cualitativas y cuantitativas.

Las variables **cualitativas** son aquellas cuyos valores son adjetivos, tales como: colores, géneros, nacionalidades... Mientras que las **cuantitativas** son variables cuyos valores son numéricos y que además poseen un sentido aritmético, como por ejemplo conteos, temperaturas, pesos, dinero, distancias, áreas, etcétera. ¿Qué significa que los valores numéricos deban poseer sentido aritmético? Existen números que no lo tienen, es decir, que su suma, resta, multiplicación o división no significan nada, como por ejemplo números de tarjetas de crédito, cuentas de banco, matrículas y códigos.

Variables

Cualitativas

- Categóricas

Cuantitativas

- Discretas
- Continuas

Es probable que en la información que consultes en internet o en otras fuentes de información, encuentres tipos de variables que no están mencionadas aquí como: variable aleatoria, ordinal, nominal, binaria, independiente, dependiente, entre otras. Dichas variables no forman parte del temario del curso y en este momento en el que apenas estás descubriendo la estadística, sería más confuso que revelador ahondar en ellas.

Variable categórica

Es una variable cualitativa cuyo rango de valores son categorías, algunos ejemplos son color de ojos, lugar de origen, tamaño, raza, tipo de pelo. En estadística, las variables categóricas también se conocen como factores. Por ejemplo, la variable categórica raza podría tener cuatro posibles valores: blanca, africana, árabe y asiática; y si hablamos de esta variable como un factor, se dice que es un factor con cuatro niveles.

Ejemplo

La base de datos realizada por Alonso Quijano en la página 17 tiene una variable categórica que es el sexo y tiene dos categorías: femenino y masculino. Otra manera de decir esto es que en la investigación de Alonso la variable sexo es un factor con dos niveles.

Es posible que a veces los valores que puede tener una variable categórica sean números. Por ejemplo, imagina que tu salón quiere saber cuánta basura plástica produce y cuánta de esta basura podría separarse y reciclarse. Para esto, revisan durante una semana diez botes de basura de tu escuela elegidos al azar y registran cada plástico en una base de datos que podría verse como la siguiente:

Producto	Valor ($)	Peso (g)	Tipo de plástico (1-7)	Reciclable (Si/No)

Al final de la semana tendrás cientos de datos (filas) de todos los plásticos que encontraron. La variable "Tipo de plástico" se refiere a la clasificación propuesta por la Asociación Americana de la Industria, en la que ellos plantean siete categorías que puedes ver en la figura de los triángulos. A pesar de tratarse de una variable numérica, los números no tienen sentido aritmético, en otras palabras, no obtendrías información relevante si decidieras sumar todos los valores de esa variable. Por otra parte es una variable que permite clasificar todos tus datos (filas) en siete categorías, por esta razón se trata de una variable categórica.

No siempre una variable categórica es útil para fines estadísticos, por ejemplo, si hicieras una base de datos de los padres de familia de tu clase y de sus números de tarjetas de crédito, ambas variables (nombre del padre y número de tarjeta) serían variables categóricas. Sin embargo los nombres de los padres serían únicos así como sus números de tarjetas de crédito. Una variable categórica que tiene tantas posibles categorías como número de registros, es totalmente inútil para fines estadísticos. Reflexiona por qué.

Variable discreta

Es una variable cuantitativa cuyos posibles valores son contables enteros. Ejemplos típicos de variables discretas son el número de miembros familiares, edad, número de bancas, número de parásitos. Es relativamente fácil identificar una variable discreta, pues si no tiene sentido que los valores numéricos tengan decimales seguramente se trata de una variable discreta. Por ejemplo, no tiene sentido que una familia tenga 3.7 miembros familiares, o tiene tres o tiene cuatro pero no es posible que tenga 3.7 personas.

En la base de datos de Alonso hay una variable discreta que es la edad, pues en este estudio Alonso consideró la edad como los años cumplidos.

Variable continua

Una variable cuantitativa es continua cuando los valores que puede tomar son incontables o pueden tener un valor infinito de decimales como se muestra en el cuadro de la derecha. Ejemplos de este tipo de variables son temperatura, altura, distancia, peso. Prácticamente cualquier medición con decimales es una variable continua. En el caso de estudio de Alonso Quijano, su variable de interés que es la estatura de las personas, es una variable continua, pues los valores que puede haber entre una estatura y otra tienen diferentes decimales.

Altura
1.72
1.723
1.7239
1.72394
1.723944
1.7239441

Importante. Como probablemente algunos de ustedes ya se habrán dado cuenta, a veces un mismo atributo se puede registrar como variable continua, discreta o categórica según sea la investigación. Tomemos la variable "edad" como ejemplo, si se mide por los años cumplidos se trata de números enteros, y por lo tanto de una variable cuantitativa discreta; si se consideran fracciones de año, es decir que se considere que alguien pueda tener 16.4 años, se trata entonces de una variable cuantitativa continua; y por último, se puede registrar la edad como una variable cualitativa categórica si se usan categorías como: infante, niño, adolescente, joven, adulto y anciano. Recuerda que en estadística todo depende de la pregunta que quieres contestar.

Agrupación en clases para la determinación de probabilidades

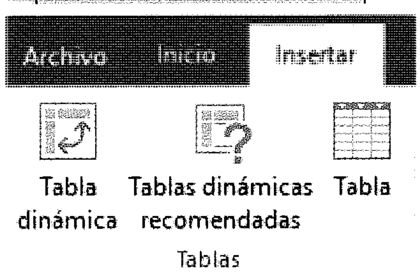
Una herramienta que te permite explorar inicialmente toda la información contenida en una variable es la frecuencia de cada elemento en esa variable. La **frecuencia** es el número de veces que algo se repite. Por lo general, las frecuencias se analizan en unos cuadros llamados "Tablas de frecuencias". La versión más sencilla de estas tablas se compone de dos columnas, la primera contiene los diferentes valores que la variable puede tomar, y la segunda columna el número de veces que cada posible valor se repite (frecuencia).

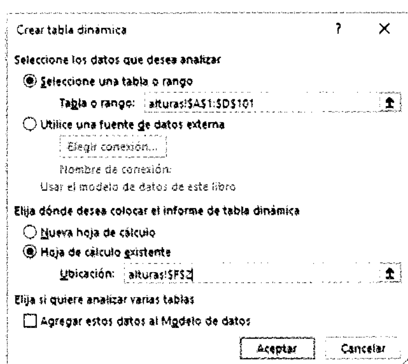
No. de refrescos	Frecuencia
5	5
6	7
7	10
8	4
9	3
10	3

Imagina que Zutanito en una fiesta familiar quiere saber la cantidad de refrescos de 600 ml que sus parientes se toman en un mes. Encuesta a sus 32 familiares y obtiene los siguientes datos: 5, 5, 9, 8, 7, 7, 7, 6, 6, 6, 6, 8, 9, 10, 9, 5, 5, 6, 6, 6, 7, 7, 8, 10, 7, 7, 8, 7, 7, 7,10, 5. Esos 32 datos forman una variable, en este caso se trata de una variable discreta. La tabla de frecuencias que resumiría esta información se vería como la de la página previa. Además de resumir información, las tablas de frecuencias son útiles para calcular la probabilidad que tiene cada valor de aparecer en la variable. Por ejemplo, la probabilidad de que un familiar de Zutanito, tomado al azar, tome ocho refrescos al mes es de: 4/32 = 0.125. La misma probabilidad se puede calcular para cada número de refrescos.

	A	B	C	D	E
77	76	M	1.85	34	
78	77	M	1.78	32	
79	78	M	1.85	24	
80	79	M	1.81	28	
81	80	M	1.75	32	
82	81	M	1.67	31	
83	82	M	1.68	36	
84	83	M	1.88	31	
85	84	M	1.83	26	
86	85	M	1.77	36	
87	86	M	1.72	28	
88	87	M	1.8	23	
89	88	M	1.75	31	
90	89	M	1.76	27	
91	90	M	1.71	30	
92	91	M	1.64	24	
93	92	M	1.76	20	
94	93	M	1.79	37	
95	94	M	1.79	40	
96	95	M	1.87	37	
97	96	M	1.77	27	
98	97	M	1.68	25	
99	98	M	1.76	20	
100	99	M	1.72	25	
101	100	M	1.71	23	

Cualquiera de los tres tipos de variables que aprendiste en el tema anterior (categórica, continua y discreta) se puede ordenar en una tabla de frecuencias, sin embargo, la forma de elaborar la tabla varía según las características de la variable que se quiere ordenar. En el caso de variables categóricas y variables discretas con pocos posibles valores (diez o menos), debes identificar los diferentes valores de la variable y contar cuántos hay de cada uno. Un ejemplo de una variable cuantitativa discreta es el de los refrescos, esa variable sólo tiene seis valores posibles (del cinco al diez) y se cuenta cuántos hay de cada uno (frecuencia). Un ejemplo de una variable categórica es el sexo en el estudio de Alonso Quijano, dicha variable puede tener dos valores (femenino y masculino) y hay 50 de cada uno.



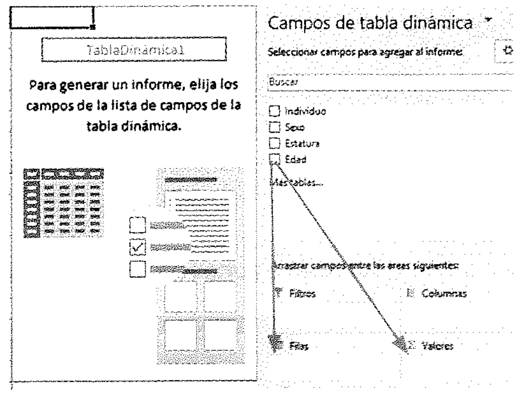


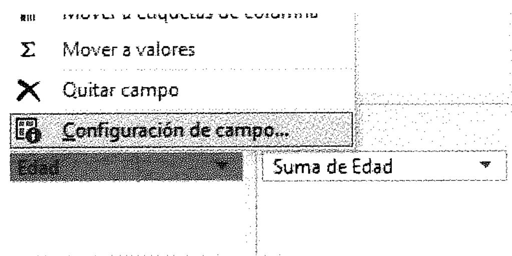
En Excel existe una herramienta muy útil que puedes utilizar para explorar bases de datos, entre sus muchas aplicaciones está la de crear tablas de frecuencias, esta herramienta se llama "Tablas dinámicas". A continuación, utilizando esta herramienta crearás en Excel una tabla de frecuencias de la variable discreta "Edad" que está en la base de datos de Alonso Quijano.

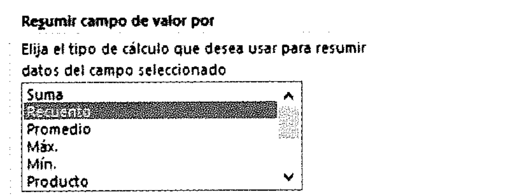
1. Una vez abierta la base de datos en Excel, selecciona de la celda A1 a la celda D101, como se muestra en la imagen de la izquierda.
2. Busca en el menú "Insertar" la herramienta "Tabla dinámica" y selecciónala.
3. Aparecerá un cuadro titulado "Crear tabla dinámica". Si seleccionaste correctamente la base de datos, el espacio de "Tabla o rango" aparece con algo escrito muy similar a lo que se ve en la imagen. A continuación debes elegir en dónde quieres colocar tu tabla dinámica, en mi caso siempre opto por crear la tabla en la "Hoja de cálculo

existente" donde está la base, luego sólo escribe o selecciona la "Ubicación", en este caso particular especifica que aparezca en la celda F2. Da clic en "Aceptar".

4. De momento la tabla dinámica sólo será un rectángulo en blanco y aparecerá una "Lista de campos" asociada a la tabla como se aprecia en la figura de la derecha (la presentación puede variar dependiendo de la versión de Office y otros factores).
5. Para crear la tabla de frecuencias debes arrastrar la variable de interés (que en este caso es "Edad" hacia las áreas "Etiquetas de fila" y "Valores", como muestran las flechas azules en la figura.
6. Deberá salir la tabla solicitada, sin embargo, es probable que Excel automáticamente considere la suma de la variable "Edad" por tratarse de una variable cuantitativa y no el conteo de ella (frecuencia). Puedes cambiar esto dando clic en "Suma de Edad", luego en "Configuración de campo de valor...", y entonces aparecerá un cuadro donde debes cambiar de "Suma" a "Cuenta" como muestran las figuras de la derecha.
7. ¡Listo! Ahora ya sabes hacer tablas de frecuencias en Excel.







Rango

El rango es el par de números entre los que se encuentran todas las observaciones de una variable cuantitativa y la diferencia entre ellos. En otras palabras, el valor máximo y el mínimo así como su diferencia. En el caso de los refrescos el valor mínimo es cinco, el máximo diez y la diferencia también es de cinco (10 – 5 = 5). En el caso de la edad en la investigación de Alonso, la persona más joven tuvo 18 años y el más grande 41, la diferencia entre ambos es de 23 (41 – 18 = 23). En algunas fuentes de información al mismo concepto lo llaman "intervalo".

Cuando una variable cuantitativa discreta tiene un **rango** muy amplio (8 o más), las tablas de frecuencias como las que hemos visto hasta ahora no resultan muy útiles, pues son demasiadas filas (como el caso de la edad en la base de datos de Alonso). Algo similar pasa con las variables cuantitativas continuas, pues cuando determinas que una variable es continua asumes que puede haber tantos valores como decimales, es decir un número infinito de valores entre el rango. La forma en la que se soluciona esto es dividiendo el rango en segmentos del mismo tamaño y contando cuántos valores (frecuencia) están dentro de cada segmento. A este tipo de tablas se les conoce como tablas de frecuencias para datos agrupados.

Ejemplo

Abajo hay dos ejemplos de tablas de frecuencias para datos agrupados utilizando las variables edad y estatura de la base de datos de Alonso. En estas tablas puedes observar, por ejemplo, que hubo 43 personas que midieron entre 1.71 y 1.80 metros, y que 27 personas medidas tuvieron entre 26 y 29 años.

A continuación verás algunos conceptos relacionados con este tipo de tablas.

Variable continua	
Estatura	**Frecuencia**
[1.41 - 1.51)	1
[1.51 - 1.61)	10
[1.61 - 1.71)	37
[1.71 - 1.81)	43
[1.81 - 1.91]	9

Variable discreta	
Edad	**Frecuencia**
[18 - 22)	7
[22 - 26)	13
[26 - 30)	27
[30 - 34)	25
[34 - 38)	19
[38 - 42]	9

Intervalo o clase

Es cada uno de los segmentos en los que divides el rango. Es importante que consideres que en algunas fuentes de información le llaman "intervalo" a cada segmento y en otras le llaman "clase". En los ejemplos de arriba, la tabla de frecuencias de la estatura tiene cinco intervalos, mientras que la tabla de edad tiene seis. Todos los datos deben caer dentro de algún intervalo y sólo en un intervalo.

Amplitud de intervalo o clase

Se le llama **amplitud de intervalo** o **amplitud de clase** al tamaño de cada intervalo. La amplitud la determinas dividiendo el rango entre el número de intervalos que deseas. Dependiendo del caso, redondeas el resultado de la división a números enteros, décimas o centésimas. En el caso de la edad en la base de Alonso, recuerda que el rango es de 23 (la persona más joven tuvo 18 años y la más grande 41), si deseas tener seis intervalos, divides 23 (rango) entre 6 (número de intervalos) y obtienes 3.83 que se redondea a 4. Empezando por el valor más bajo de la edad (18), le vas sumando la amplitud para definir los límites de cada intervalo.

En ocasiones, dependiendo de la pregunta que quieres responder, puedes fijar la amplitud del intervalo arbitrariamente. Recuerda que las tablas de frecuencias son una herramienta para ordenar y explorar la información. En el ejemplo de la tabla de frecuencias de la estatura, el intervalo está arbitrariamente ajustado a diez centímetros y a empezar en 1.41, de esta manera la tabla queda más intuitiva de asimilar pues nuestros cerebros están acostumbrados a contar de diez en diez.

Probablemente te estés preguntando dos cosas: a) ¿dónde va un valor que coincide exactamente con el límite de un intervalo? y b) ¿qué significan los paréntesis y los corchetes en los intervalos de las tablas de frecuencias? Las respuestas a estas dos preguntas están relacionadas. El corchete significa que ese límite de intervalo está "cerrado" y el paréntesis que está "abierto", un valor que coincide exactamente con un límite de intervalo va en aquél que esté cerrado. En el ejemplo de la edad, las personas de 22 años se cuentan en el segundo intervalo.

Marca de clase

La marca de clase es el valor intermedio que tiene cada intervalo. Se calcula promediando los límites de cada intervalo. Se puede ver el ejemplo con la variable edad en la tabla de frecuencias de la derecha.

Variable discreta		
Edad	Frecuencia	Marca de Clase
[18 - 22)	7	20
[22 - 26)	13	24
[26 - 30)	27	28
[30 - 34)	25	32
[34 - 38)	19	36
[38 - 42]	9	40

Actividad de aprendizaje 4

Para llevar a cabo esta actividad, formen equipos de cuatro personas y hagan lo que se pide.

1. Cada integrante encuestará a mínimo 15 estudiantes de su preparatoria para conseguir la siguiente información de cada uno (recuerda llevar una cinta métrica o alguna otra herramienta para medir longitudes):

   - Semestre
   - Edad
   - Altura
   - Longitud de su dedo meñique
   - ¿Usa anteojos?
   - ¿Es soltero(a)?
   - Género musical favorito
   - Cuántos hermanos tiene
   - ¿Tiene Facebook?
   - ¿Vegetariano u omnívoro?
   - Cuántas mascotas tiene en su casa
   - Salón
   - Sexo
   - Número de dedos en la mano izquierda
   - Longitud de su dedo índice
   - Número de lista en su salón
   - Horas de ejercicio a la semana
   - Tiempo que hace de la escuela a su casa
   - Número de libros que lee al año
   - ¿Evita el unicel?
   - Nacionalidad
   - Cuántas tortillas come al día

   Recuerda ser respetuoso con tus compañeros, se trata de información personal y las personas protegemos ese tipo de información, si quieres que te compartan información debes respetar a quien la tiene.

2. Entre los cuatro integrantes del equipo debieron reunir mínimo 60 estudiantes, capturen toda su información en una base de datos en Excel y discutan sobre el tipo de cada una de las variables registradas (categórica, continua o discreta).

3. Responde las preguntas:

   a. ¿Qué complicaciones tuvieron al momento de aplicar las encuestas?, ¿los cuatro integrantes aplicaron la encuesta exactamente igual? Si no, ¿cuáles fueron las discrepancias?

b. ¿Todas las variables presentaron diversidad en sus valores?, ¿cuáles no?

c. ¿Consideran que alguna de las variables numéricas no tiene sentido aritmético? ¿Cuáles? y ¿por qué?

d. ¿Hubo variables que, al momento de aplicar las encuestas, algunos registraron como discretas y otros como continuas?

e. ¿Cuál fue la variable categórica con menos categorías?

f. Sin contar variables que hayan tenido tantas categorías como estudiantes encuestados, ¿cuál fue la variable categórica con más categorías?

g. Supón que esta recopilación de la información es para un estudio de tu preparatoria. ¿Cuál es la población? ¿Cuál es la muestra? ¿Consideras que la muestra es representativa? ¿Por qué?

4. Utiliza Excel y la herramienta de tablas dinámicas para hacer la tabla de frecuencias con los datos de los refrescos de la página 25.

5. Genera una tabla de frecuencias para datos agrupados con la variable edad de la base de Alonso, pero con siete intervalos en vez de seis. Determina la amplitud de intervalo con los cálculos en el texto y no de manera arbitraria. ¿Qué problemas tuviste? Y ¿por qué? Investiga la solución para evitar este problema. Pista: tiene que ver con la forma de redondear.

6. Calcula las marcas de clase para la tabla de frecuencias del ejemplo de la variable estatura en la página 28.

7. Imagina que tienes la variable: 19.2, 19.2, 18.3, 16.3, 19.2, 19.8, 20.3, 18.3, 18.4, 19.4, 20.4, 16.6, 18.1, 19.2, 18.9, 21.6, 20, 17.8, 20.2, 19.5, 19.3, 18.5 y 20.5.

a. Completa la tabla anotando la frecuencia y calculando las marcas de clase.

b. Determina el rango de la variable.

c. ¿Cuántos intervalos hay?

d. ¿De qué amplitud es cada intervalo?

Variable F	Frecuencia	Marca de clase
[16.3 – 17.7)		
[17.7 – 19.1)		
[19.1 – 20.5)		
[20.5 – 21.9]		

Actividad de aprendizaje 5

Efectúa lo que se te pide.

1. Observa las bancas de tu salón de clase y clasifícalas con base en dos características que tú decidas, registra la información en tu cuaderno en una base de datos y anota también los criterios que utilizaste para clasificarlas. Comenta con algún compañero(a) si puedes clasificar media banca, o un cuarto de banca, y si todo esto tiene algún vínculo con la estadística.
2. En un potrero hay 500 vacas, el vaquero observa que algunas están enfermas de un agente desconocido, le comenta al dueño y este decide verificar la observación. Para esto manda traer 30 vacas del potrero eligiéndolas al azar para ver si están o no enfermas. Identifica en este problema lo siguiente: la población, la muestra y su tamaño, dato y el tipo de variable que se investiga.

3. Un canal de televisión desea conocer las edades de los televidentes en el área oeste de Puebla que ven la telenovela "El amor de Timotea", para esto lleva a cabo un estudio y se seleccionan al azar 350 adultos de familias de 5 municipios de esta área. ¿Cuál es un dato de este estudio? ¿Cuál es la población? ¿Cuál es la muestra?
4. Un botánico estudia la forma, el tamaño, el color, el número de pétalos y el momento en que la flor abre (antesis) y cree que las tablas de frecuencia podrían ayudarle a descubrir algún patrón. Descarga la base de datos de las flores escaneando el código QR de la derecha.
5. Determina el tipo (categórica, discreta o continua) de cada una de las variables en esa base de datos.
6. Usando tablas dinámicas construye una tabla de frecuencias para las variables categóricas, guarda el archivo de tus tablas de frecuencias pues las usarás más adelante. Si tu profesor lo considera necesario copia las tablas de frecuencias en tu cuaderno.
7. Observa la variable "Número de pétalos" ¿cuántos posibles valores tiene?, ¿cuál es el rango de esa variable?, ¿qué tipo de tabla de frecuencias consideras apropiada para esa variable, una sencilla o de datos agrupados? Coméntalo con tus compañeros y haz la tabla que hayas decidido que es mejor.
8. ¿Cuál es el rango de la única variable cuantitativa continua? Construye una tabla de frecuencias de datos agrupados con esta variable. No olvides calcular las marcas de clase.
   a. ¿Cuántos intervalos decidiste poner? Y ¿por qué?
   b. ¿De qué amplitud quedaron los intervalos?
   c. ¿Existe alguna relación entre la amplitud de cada intervalo y la diferencia de una marca de clase a otra?

Compara tus tablas de frecuencias con las de tus compañeros. ¿En qué se parecen y en qué son diferentes?

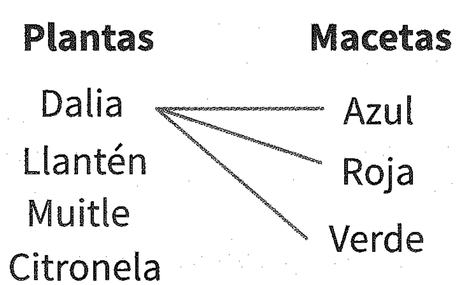
Uso del conteo y la probabilidad para eventos

Técnicas de conteo

Como ya vimos, una probabilidad se calcula a partir del número de posibles resultados de un evento, esos posibles resultados hay que saber contarlos. Contar puede parecerte algo muy fácil, pues estás acostumbrado a contar cosas sencillas y ordenadas, pero imagina que quisieras contar cuántas palabras se pueden hacer con cinco letras del abecedario; o de cuántas formas puedes combinar los rasgos faciales de un avatar en alguna página de internet. Es ahí cuando contar ya no es tan fácil. Para esto, se han desarrollado técnicas y estrategias de conteo. El principio más utilizado para contar es el **principio multiplicativo**.

El principio multiplicativo, al igual que en la multiplicación aritmética, es una abreviación de una suma aritmética ($5 \cdot 2$ es una forma abreviada de sumar cinco veces el dos o dos veces el cinco), en el conteo de posibles resultados también puedes abreviar sumas de conteos, a estas abreviaturas se les conoce como principio multiplicativo.

Ejemplo



Supón que en un pequeño vivero tienen tres colores de macetas (azul, rojo y verde) y cuatro especies de plantas (dalia, llantén, muitle y citronela). Si cualquier planta puede ir en cualquier maceta, ¿de cuántas formas diferentes puede el vivero vender una planta con una maceta? Observa el diagrama de la izquierda, en él puedes ver que hay tres opciones de maceta para la dalia. De la misa manera habrá tres opciones de maceta para cada planta, por lo que el número de formas diferentes en el que pueden vender una planta con una maceta será:

$$3 + 3 + 3 + 3 = 3 \cdot 4,$$

es decir, 12 opciones.

Ahora imagina que a cada maceta se le puede poner dos tipos de adorno, ¿cuántas opciones hay ahora? Bueno, pues sabes que para cada tipo de adorno hay 12 opciones de color de maceta con especie de planta, si para cada tipo de adorno tienes 12 opciones y son dos tipos de adorno, entonces las opciones son $12 + 12$ o lo que es lo mismo $12 \cdot 2$, es decir, 24 opciones en total.

El principio multiplicativo establece que si un evento está compuesto por varias partes y cada parte tiene varias formas de suceder, entonces multiplicas el número de formas en las que puede suceder cada parte. En el ejemplo de las plantas con las macetas y los adornos se trata de un evento de tres partes, el tipo de adorno tiene dos posibles formas de hacerse, el color de la maceta tres posibles formas y la especie de planta cuatro formas, por lo que el número de opciones en las que el evento se puede efectuar es de $2 \cdot 3 \cdot 4$, lo que es igual a 24.

Técnica de las casillas

Es una técnica práctica para aplicar el principio multiplicativo. Consiste en crear tantas casillas como partes compongan determinado evento y en cada casilla anotar las posibles formas en las que se puede llevar a cabo esa parte, y al final multiplicar todos los dígitos de las casillas. Puede parecer una técnica muy simple pero en ocasiones es muy útil para entender y resolver ciertos problemas.

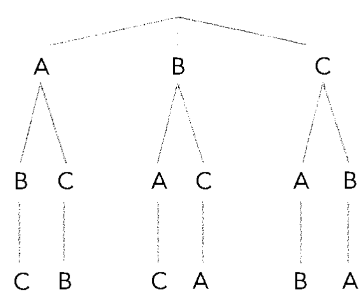
Para el caso de las plantas, las macetas y los adornos, al aplicar la técnica de las casillas se dibuja una casilla para cada parte del evento y luego se multiplican los dígitos de cada casilla.

2	3	4	= 24 posibilidades
Adorno	Maceta	Especie	

Permutaciones

En teoría del conteo, las permutaciones son todas las posibles formas de ordenar un conjunto o subconjunto de n elementos.

Por ejemplo ¿De cuántas maneras puedes ordenar las letras A, B y C? En este ejemplo, se trata de un conjunto de tres elementos (A, B y C). El diagrama de árbol que resultaría para intentar contestar esta pregunta sería como el que se ve a la derecha. Si lees el diagrama de arriba hacia abajo, en la primera posición puedes poner cualquiera de las tres letras, pero una vez que se ocupa la primera posición para la segunda sólo quedan dos letras disponibles, y una vez que eliges la segunda letra sólo queda una letra disponible para la tercera posición. Dicho de otra forma, por cada letra que elijas al inicio, tendrás dos opciones para elegir después y sólo una alternativa para la última posición. Si consideras que tienes tres posibles letras para comenzar, utilizando la técnica de las casillas, tienes que el número de maneras en que puedes ordenar este set de tres elementos, es:



3	2	1	= 6 maneras
1ª letra	2ª letra	3ª letra	

Al momento de multiplicar los números en las casillas, estás aplicando el principio multiplicativo.

Ejemplo

Si en un salón hay diez bancas y diez estudiantes, ¿de cuántas formas diferentes se pueden sentar los diez estudiantes? Siguiendo la lógica del ejercicio anterior, para la primera banca hay diez estudiantes disponibles, una vez que se ocupa la primera banca quedan nueve estudiantes para la segunda, ocho para la tercera y así sucesivamente hasta que queda sólo un estudiante para el último lugar disponible. Utilizando la técnica de las casillas el ejercicio se resolvería de la forma:

10	9	8	7	6	5	4	3	2	1	= 3 628 800 maneras

Usando el principio multiplicativo a través de la técnica de las casillas, te puedes dar cuenta de que la regla para calcular el número de permutaciones de n elementos es:

$$n(n-1)(n-2)(n-3)\ldots(n-n+1).$$

En el ejemplo de los estudiantes habría que sustituir n por 10, lo que resultaría en:

$$10(10-1)(10-2)(10-3)\ldots(10-10+1) = 3\,628\,800,$$

en otras palabras, el número de permutaciones se calcula multiplicando todos los números de n hasta 1. A esta multiplicación se le conoce como **factorial** y se denota con el valor de n seguida de un signo de exclamación de cierre. En el ejemplo, el número de formas en las que se pueden sentar los diez estudiantes es igual a "diez factorial" y se escribe:

$$10! = 3\,628\,800.$$

En ocasiones hay ciertas restricciones o reglas acerca de algunas ordenaciones. En esos casos hay que determinar cómo las restricciones disminuyen las posibilidades en cada casilla. Por ejemplo, supón que en el caso de los diez estudiantes y las diez bancas, fueran siete mujeres y tres hombres.

¿Cuántas permutaciones hay si las primeras dos bancas deben ser ocupadas por mujeres? En este escenario, no son diez posibles personas las que pueden ocupar el primer lugar, pues sólo hay siete mujeres, entonces el número en la primera casilla debe ser siete, pues hay siete posibles mujeres a ocuparla; según la pregunta, la segunda casilla debe ocuparse también por una mujer; como una mujer ya se sentó en la primera banca sólo quedan seis mujeres disponibles, entonces en la segunda casilla va un seis. El resto de las casillas no tiene restricción alguna, por lo que de la tercera banca en adelante puede sentarse cualquiera de los ocho estudiantes disponibles. Las casillas quedarían como sigue:

7	6	8	7	6	5	4	3	2	1	= 1 693 440 maneras

Ejemplo

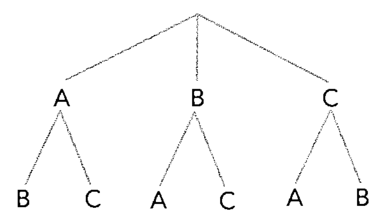
¿Cuántas permutaciones hay con las primeras cuatro letras del alfabeto (A, B, C y D) y con la tercera posición ocupada por la letra B? Para la primera posición sólo hay tres letras disponibles, pues la B debe ir en la tercera posición; para la segunda posición sólo quedan dos letras disponibles, en la tercera sólo puede ir la B y en la cuarta sólo quedaría una posible opción. Las casillas quedarían como sigue:

3	2	1	1	= 6 permutaciones

Combinaciones

En teoría del conteo, las combinaciones son todas las posibles formas en que puedes seleccionar un subconjunto de un conjunto de elementos.

Por ejemplo ¿de cuántas formas puedes seleccionar dos letras de las tres primeras letras del abecedario? En el diagrama de árbol de la derecha puedes ver que hay seis maneras en las que puedes ordenar dos de las tres primeras letras del abecedario que serían: AB, AC, BA, BC, CA y CB. Sin embargo, la pregunta es acerca de las formas en que se pueden seleccionar dos de tres letras y no sobre el número de formas en que pueden **ordenarse**. ¿Qué quiere decir esto?, quiere decir que cuando **no importa el orden**, AB y BA son exactamente lo mismo, igual que AC con CA y BC con CB. Viéndolo de este modo, sólo hay tres formas de seleccionar dos de las tres primeras letras del abecedario. En palabras matemáticamente educadas la pregunta de este ejemplo es: ¿cuántas combinaciones de dos letras hay, considerando las tres primeras letras del abecedario? Y la respuesta es tres.



Veamos un ejemplo con más combinaciones. ¿Cuántas combinaciones puedes tener tomando tres de las cinco vocales del alfabeto? Primero calculemos cuántas ordenaciones sin repetición puedes formar con tres de las cinco vocales:

5	4	3	= 60 ordenaciones

Pero en esas 60 ordenaciones sabemos que se repiten grupos de letras en diferente orden, por ejemplo, la ordenación AEI repite las letras que la ordenación AIE. Entonces surge la pregunta: ¿de cuántas formas se puede ordenar un grupo de tres letras? Esa pregunta ya la sabes responder con la técnica de las casillas y el uso de factorial:

3	2	1	= 3! = 6 maneras

O sea que cada grupo de tres letras se puede repetir seis veces con un orden diferente, por lo tanto, si divido el total de ordenaciones de tres letras (60) entre el número de maneras en que puede repetirse un grupo de tres letras con diferente orden (6) tendré el total de combinaciones.

En este ejemplo:

$$60/\,6 = 10 \text{ combinaciones.}$$

Observa el cuadro siguiente en donde se muestran las 60 ordenaciones clasificadas por las diez combinaciones posibles.

Diez combinaciones

Seis maneras en que puedes ordenar cada combinación

A E I	A E O	A E U	A I O	A I U	A O U	E I O	E I U	E O U	I O U
A I E	A O E	A U E	A O I	A U I	A U O	E O I	E U I	E U O	I U O
E A I	E A O	E A U	I A O	I A U	O A U	I E O	I E U	O E U	O I U
E I A	E O A	E U A	I O A	I U A	O U A	I O E	I U E	O U E	O U I
I A E	O A E	U A E	O A I	U A I	U A O	O E I	U E I	U E O	U I O
I E A	O E A	U E A	O I A	U I A	U O A	O I E	U I E	U O E	U O I

Repasa con atención el ejemplo de las vocales hasta que estés seguro de que puedes explicarlo.

Ahora sabes que para calcular el número de combinaciones sólo debes dividir la cantidad de ordenaciones sin repetición entre el número de maneras en que se puede ordenar cada conjunto de elementos seleccionados (factorial del tamaño de la selección). Y ese es el origen de la expresión matemática para calcular el número de combinaciones, la expresión es:

$$C(n,k)=\frac{n!}{(n-k)!k!}.$$

Donde C es el número de combinaciones, n el universo de posibles elementos y k el tamaño del conjunto de los elementos seleccionados.

Si aplicas la fórmula para el ejemplo de las vocales quedaría así:

$$C(5,3)=\frac{5!}{(5-3)!3!}=\frac{120}{2!\cdot 3!}=\frac{120}{2\cdot 6}=\frac{120}{12}=10$$

Hay otras notaciones de combinatoria que puedes hallar en los libros o en internet, inclusive en algunas notaciones se usa la letra r en vez de la k, sin embargo, tienen el mismo significado, algunas de estas notaciones son:

$$C(n,k)={}^{n}C_{k}={}_{n}C_{k}=\binom{n}{k}=\frac{n!}{k!(n-k)!}$$

Es importante que recuerdes que el uso de la palabra "combinación" en teoría del conteo es diferente al que se le da en el lenguaje cotidiano. Por ejemplo, en el lenguaje cotidiano tú podrías preguntar ¿de cuántas maneras puedo combinar una playera con un pantalón si tengo tres playeras y cuatro pantalones disponibles? Sin embargo, este no es un problema de combinatoria, sino de posibles eventos utilizando el principio multiplicativo, pues si tienes tres posibilidades para el evento "playera" y cuatro posibilidades para el "evento" pantalón, multiplicas las posibilidades de un evento por las posibilidades del otro y en total tienes 12 posibilidades.

Ejemplo

Una baraja inglesa tiene 13 cartas distintas de cada uno de los cuatro palos (corazones, diamantes, tréboles y espadas). ¿De cuántas formas se pueden elegir cinco cartas de una baraja inglesa de tal manera que sean dos reyes y tres reinas?

Primero calculemos de cuántas maneras pueden combinarse dos de los cuatro reyes en la baraja:

$$C(4,2)=\frac{4!}{(4-2)!2!}=\frac{24}{2!\cdot 2!}=\frac{24}{2\cdot 2}=\frac{24}{4}=6\,.$$

Luego calculamos de cuántas maneras pueden combinarse tres de las cuatro reinas en la baraja:

$$C(4,3)=\frac{4!}{(4-3)!3!}=\frac{24}{1!\cdot 3!}=\frac{24}{1\cdot 6}=\frac{24}{6}=4\,.$$

Luego razonemos: por cada combinación de dos reyes hay cuatro posibles combinaciones de tres reinas, entonces multiplico el número de combinaciones de un evento por el número de combinaciones del otro (principio multiplicativo) y tenemos que:

6	4	= 24 formas

Actividad de aprendizaje 6

Resuelve las siguientes cuestiones.

1. En la comida corrida de una cocina económica ofrecen tres opciones de entrada: sopa de pasta, arroz o crema de champiñones; tienen tres opciones de plato fuerte: pollo con mole, tortitas de atún o chuletas con crema; y de postre te dan a elegir entre gelatina y arroz con leche. Además, puedes pedir de beber agua fresca o refresco. ¿De cuántas maneras diferentes podría un cliente ordenar su comida corrida? Utiliza la técnica de las casillas.
2. ¿De cuántas formas diferentes pueden caer tres dados de seis caras cada uno? Llena las casillas y resuelve el ejercicio.

			= ______ posibilidades
Dado 1	Dado 2	Dado 3	

3. ¿De cuántas formas se puede contestar un examen de opción múltiple de cinco preguntas y con cuatro posibles respuestas cada pregunta? Llena las casillas y resuelve el ejercicio.

					= ______ posibilidades
Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4	Pregunta 5	

4. Una ensalada se va a preparar con lechuga, jitomate, pepino y nuez. Hay tres posibles lechugas (romana, iceberg o italiana), dos opciones de jitomate (jitomate cereza o jitomate bola), una opción de pepino y cuatro opciones de nuez (macadamia, de castilla, pacanas o de cedro). ¿De cuántas maneras se puede preparar la ensalada si sólo puedes elegir una opción de cada tipo de ingrediente?

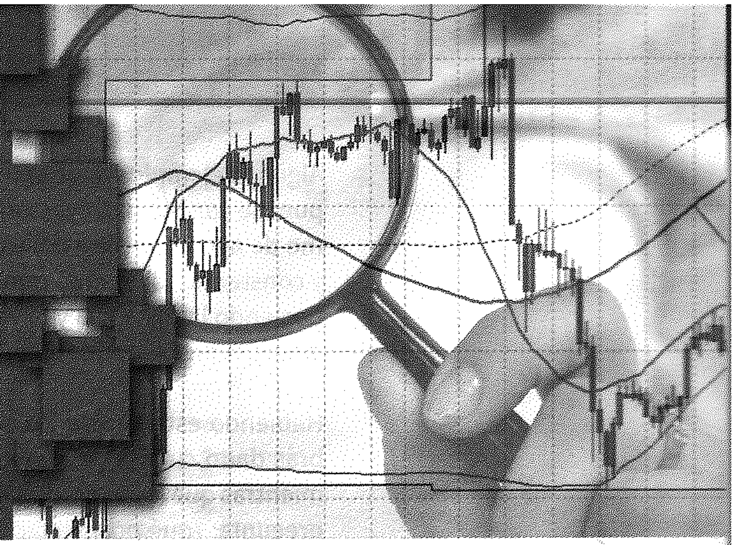
5. Supón que para preparar una torta tienes 16 ingredientes disponibles pero sólo usarás siete. ¿De cuántas formas puedes preparar tu torta?, o dicho de otra forma, ¿cuántas combinaciones de siete ingredientes puedes hacer con 16 ingredientes disponibles?
6. ¿Cuántas combinaciones de cuatro cartas puedes hacer con una baraja inglesa?
7. En un festival al que asistieron 783 personas se rifarán diez copias de un libro. ¿Cuántas combinaciones de personas pueden ganarse los libros?
8. ¿De cuántas maneras puedes combinar seis letras del abecedario?
9. Imagina que vas a comprar tres chocolates y hay ocho sabores. ¿De cuántas maneras puedes combinar tres sabores?
10. ¿Cuántas combinaciones de siete cartas puedes hacer con una baraja inglesa considerando que tres cartas deban ser reyes, dos sean ases y las otras dos sean cuatros?
11. En un grupo de danza hay ocho hombres y cinco mujeres, para la presentación de un baile necesitan tres parejas. ¿Cuántas combinaciones de tres hombres y tres mujeres hay?

Actividad de aprendizaje 7

Resuelve las siguientes cuestiones.

1. Considera una baraja inglesa con 52 cartas y responde:
   a. ¿De cuántas formas diferentes puedes ordenar en una fila las 52 cartas?
   b. ¿Cuántas ordenaciones de seis cartas que empiecen y terminen con rey se pueden formar?
   c. ¿Cuántas ordenaciones de siete cartas se pueden formar con los cuatro reyes alternándose con cualquier otra carta?
      Ejemplo: K Q K 9 K 4 K.
   d. ¿Cuántas combinaciones se pueden hacer de siete cartas?
   e. ¿Cuántas combinaciones se pueden hacer de siete cartas con cuatro reyes y tres reinas?
2. En una competencia de natación hay ocho participantes, ¿de cuántas formas pueden quedar los tres primeros lugares?
3. En un juego de dominó hay 28 fichas diferentes, cuando juegan cuatro personas, cada jugador toma siete fichas de manera que ninguna sobra. ¿De cuántas maneras se puede componer el grupo de siete fichas que toma cada jugador?
4. Para llenar un álbum de 350 estampas hay que comprar sobres que traen cinco estampas. Suponiendo que no pueden salir estampas repetidas en un sobre, ¿cuántas combinaciones de estampas pueden salir en un sobre?

Concepto de riesgo en situaciones contextuales



¿Qué es el riesgo?, ¿qué papel juega la probabilidad y la estadística en el estudio del riesgo?

Nociones de incertidumbre, azar y aleatoriedad

Todos los eventos, es decir, todas las cosas que pasan, pueden catalogarse en dos tipos: **eventos aleatorios y deterministas**. La diferencia es sencilla, los eventos aleatorios, también conocidos como eventos al azar, son todos aquellos en los que no puedes determinar el resultado antes de que sucedan sin importar cuántas veces se haya repetido, es decir, no puedes predecir con absoluta precisión lo que sucederá. Por otra parte, los deterministas o determinísticos, son todos los eventos en los que sí puedes predecir el resultado con absoluta seguridad, aún antes de que se hayan llevado a cabo.

El **azar** es una casualidad que no es predecible, es un grupo de circunstancias impredecibles sin un propósito particular que hacen que suceda algo. Por ejemplo, cuando tiras un dado de seis caras, el azar determina la cara que caerá hacia arriba. Obviamente hay circunstancias que determinan la cara que cayó hacia arriba: la fuerza con la que tiraste el dado, la posición inicial con la que lo tiraste, la fuerza del viento en ese espacio, la superficie en la que cae el dado, la forma en la que fue lanzado, el material del dado, en fin, muchas cosas, pero entre todas hacen imposible determinar qué cara caerá.

Ejemplos de eventos aleatorios son el lanzamiento de una moneda o un dado, pues no sabes qué cara de la moneda o el dado resultará, no importa cuántas veces lo hayas lanzado antes. La mayoría de las cosas que suceden a nuestro alrededor son eventos aleatorios, un choque automovilístico, un dolor en el cuerpo, olvidar las llaves de tu casa, perder o encontrar dinero. Para ninguna de estas cosas tienes la certeza de qué te sucederá o no te sucederá en algún día o momento. Inclusive al untarte una pomada para un dolor muscular, no tienes la certeza de que funcionará, siempre existe la posibilidad de que no funcione. Comer picante en exceso puede terminar en cáncer de estómago, pero no es una verdad absoluta, pues cada quién tiene diferentes niveles de tolerancia a la capsaicina.

Algunos ejemplos de eventos deterministas son la posición de la Tierra respecto al Sol, pues se conoce su trayectoria y velocidad (30 km/s); el volumen que ocupará un kilogramo de agua que pasa del estado líquido al sólido, pues sus propiedades son conocidas y constantes; el tiempo que tardaría un objeto en llegar de un punto A a un punto B conociendo la distancia y la velocidad; la reacción al mezclar agua con nitrógeno líquido, pues sus propiedades químicas están descritas y no cambian.

Habiendo establecido las diferencias entre eventos aleatorios y deterministas, te habrás dado cuenta de que los eventos deterministas tienen sólo un posible resultado, mientras que los eventos aleatorios tienen varios posibles resultados. Entonces surge la pregunta: ¿qué posibilidades tienen cada uno de los posibles resultados en un evento aleatorio?, es decir, ¿qué posibilidades hay de que en un volado caiga águila, o de que choques el día de hoy, o de que salga un cuatro en un dado de seis caras, o de que te duela la cabeza un determinado día, o de que tendrás cáncer de estómago si consumes 5 habaneros al día? A esas posibilidades se les conoce como probabilidades. Cuando hablas de probabilidades, forzosamente estás hablando de eventos aleatorios, pues es el tipo de evento en el que tienes varios posibles resultados. La ciencia, en su mayoría, estudia eventos aleatorios y las probabilidades de que suceda una u otra cosa. En esta materia de estadística estudiarás, casi en su totalidad, eventos aleatorios.

La **certidumbre** es el nivel de seguridad que tienes de saber el resultado de un evento y es, prácticamente, una probabilidad. Por ejemplo, en un lanzamiento de un dado la certidumbre que tienes de que caiga un tres, es de $1/6 = 0.16$, pues sólo una cara de las seis posibles es tres. En estadística el **riesgo** es la certidumbre de que, en un evento, el resultado involucre una pérdida de algún tipo. En otras palabras, el riesgo es la probabilidad de que el resultado de un evento no sea favorable.

Espacio muestral

Se le llama espacio muestral al conjunto de posibles resultados en un evento aleatorio. Es importante, pues para poder calcular el riesgo (la probabilidad de que un resultado no sea favorable) es necesario saber el total de posibles resultados. A pesar de no poder predecir el resultado de un evento aleatorio, en muchas ocasiones tienes una idea de todos los posibles escenarios que pudieran resultar. Por ejemplo, en un volado sabes que todos los posibles escenarios son dos: águila o sol, y ese es el espacio muestral. Hay muchas estrategias para conocer el espacio muestral de un evento aleatorio, algunas de ellas son la técnica de las casillas y las combinaciones que viste en temas anteriores.

Ejemplo

Determina el espacio muestral de lanzar cuatro volados consecutivos. Si usas la técnica de las casillas obtendrías que:

2	2	2	2	= 16 posibilidades

El espacio muestral de lanzar cuatro volados consecutivos es de 16 posibles resultados.

Actividad de aprendizaje 8

Efectúa lo que se te pide.

1. Determina si los siguientes eventos son aleatorios o deterministas:
   a. Que te tropieces o no te tropieces mañana.
   b. Que te ganes o no te ganes la lotería.
   c. Que el día que sigue del viernes sea sábado.
   d. Que la aceleración de un objeto lanzado desde un edificio de 139 pisos hacia el suelo sea constante.
   e. Que al saltar regreses al piso.
   f. Que tires un tres de un dado de seis caras.
   g. Que comas el día de hoy.
   h. Que encuentres una nube con forma de nave espacial.
   i. Que la lluvia moje.
   j. Que al tirar un dado caiga alguna de las seis caras que tiene.
   k. Que tu novia termine contigo.
   l. Que mueras algún día.
   m. Que te mueras hoy.
   n. Que seas feliz.
   o. Longitud máxima a la que un resorte se pueda comprimir.
   p. Que veas un eclipse.
   q. Que se extinga una especie animal.
   r. Que surja una especie de planta.
   s. Que una persona sobreviva al aventarse de un edificio de 40 pisos.
   t. Que este año tiemble en algún lugar del planeta.
   u. Que en algún lugar del mundo alguien nazca el día de mañana.
   v. Que la vida continúe.
   w. Que mañana salga el sol.
2. Determina el espacio muestral de tirar un dado de seis caras.
3. ¿Cuál es el espacio muestral de tirar un dado de seis caras dos veces consecutivas?
4. Determina el espacio muestral de crear un número de dos dígitos si tienes disponibles los números 3, 5, 6 y 7 (se pueden repetir).

5. Paco tiene tres pares de zapatos, cinco pantalones, y dos camisas que le gustan. ¿Cuál es el espacio muestral de las diferentes formas en que Paco se puede vestir hoy?

Análisis de situaciones de riesgo mediante la estadística y probabilidad

Piensa que la probabilidad de riesgo es como la probabilidad de perder una apuesta. En un evento aleatorio, una vez que puedes determinar los posibles resultados (espacio muestral), la probabilidad de cada uno de ellos, y la proporción de resultados favorables/desfavorables, entonces puedes determinar la probabilidad de riesgo.

Ejemplo

Imagina que Fernando y Natalia juegan apuestas tirando dos dados de seis caras. Fernando apuesta a que el resultado de tirar los dos dados será 12 y Natalia apuesta a que el resultado será ocho. Sabemos, por el cuadro de la página 20, que al tirar dos dados hay 36 posibles resultados, uno de esos 36 resultados es 12 y cinco de los 36 son ocho. Es decir, la probabilidad de que Fernando gane es de:

$$P(\text{Fernando gane}) = 1/36 = 0.027,$$

mientras que la de Natalia es:

$$P(\text{Natalia gane}) = 5/36 = 0.138.$$

Hay otra manera de verlo, que es calculando la probabilidad de perder. Si uno de 36 posibles resultados es doce, significa que 35 resultados no dan doce, por lo tanto la probabilidad de que Fernando pierda es de:

$$P(\text{Fernando pierda}) = 35/36 = 0.972.$$

Siguiendo la misma lógica tenemos que la probabilidad de que Natalia pierda es:

$$P(\text{Natalia pierda}) = 31/36 = 0.861.$$

La apuesta de Fernando es de **mayor riesgo** pues su probabilidad de perder es mayor que la de Natalia. Ahora imagina que se le agrega dinero a la apuesta, por lo general en una apuesta de un volado o de una tirada de dados, la cantidad de dinero que se apuesta es la misma, pero para este ejemplo imagina que Fernando apuesta cinco pesos y Natalia apuesta diez ¿qué apuesta es de mayor riesgo ahora? En estos casos debes multiplicar la probabilidad de perder por la pérdida (que en este caso se trata de dinero).

Entonces tenemos que el riesgo de la apuesta de Fernando es:

$$\text{Riesgo de Fernando} = 0.972 \cdot 5 = 4.86.$$

Y el riesgo de la apuesta de Natalia sería:

$$\text{Riesgo de Natalia} = 0.861 \cdot 10 = 8.61.$$

Ahora la apuesta de Natalia es la que es de mayor riesgo. ¿por qué? Piénsalo de la siguiente manera: a pesar de que es más probable que salga ocho que doce, lo que Fernando pierde en dos apuestas perdidas (diez pesos) Natalia lo pierde en una, por lo tanto con la apuesta de Natalia se pierde más. Esto es sólo un ejemplo, en la vida real por lo general las apuestas pagan más si son más riesgosas precisamente para intentar compensar el riesgo.

Ejemplo

En un estudio probaron la eficiencia de un medicamento, a 549 personas enfermas de gripa las trataron con un placebo, y a 742 con medicamento. Se curaron 391 personas de las tratadas con el placebo y 409 con el medicamento. De aquí podrías observar que la probabilidad de curarte con el medicamento es de:

$$P(\text{curarte con placebo}) = 391/549 = 0.71,$$

mientras que con el medicamento es de:

$$P(\text{curarte con medicamento}) = 409/742 = 0.55,$$

es decir, la probabilidad de curarte es más alta con el placebo. Esto significa que el medicamento no sólo no es efectivo sino que impide que la gente se cure. En este caso, el riesgo sería la probabilidad de no curarte, la cuál es mayor tomando el medicamento ($P = 333/742 = 0.45$) que tomando el placebo ($P = 158/549 = 0.29$).

El ejemplo de los dos dados es un caso de estadística clásica, en el que conoces los posibles resultados antes de tirar los dados. El ejemplo de la medicina es un caso de estadística empírica, en el que calculas probabilidades según el registro histórico de varios eventos. Cuando se utiliza la probabilidad subjetiva el riesgo arrastra esa subjetividad y depende del criterio del experto.

La medición del riesgo es utilizada en muchos campos del quehacer humano. Se utiliza en economía para saber que segura/riesgosa es una inversión; en medicina para saber que tan riesgosa es una operación considerando la condición del paciente; en la compañías de seguros para saber qué tan riesgoso es venderle un seguro a una persona dependiendo de su edad, educación, trabajo...; en arquitectura para saber que tan segura/riesgosa es una casa o un edificio después de un sismo; en biología para determinar los riesgos cuando se lleva acabo cualquier proyecto que tenga un impacto en el medio ambiente. Aunque para calcular estos riesgos especializados se usan fórmulas y métodos más complicados.

Actividad de aprendizaje 9

Efectúa lo que se te pide.

1. Imagina que se tirará un dado de seis caras y hay dos opciones para apostar:

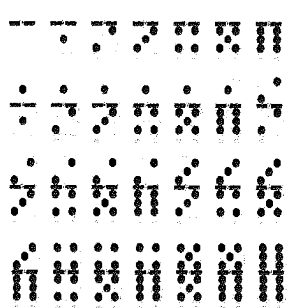
   el dado caerá en seis y

   el dado caerá en cualquier número impar ¿Qué opción elegirías? y ¿por qué?

2. Intenta determinar qué posibilidad tienes de perder cuando:
   a. Lanzas un volado.
   b. Lanzas un dado de seis caras y una cara es la ganadora.
   c. Participas en una rifa de 120 boletos.
   d. Compites en una carrera de ocho concursantes.

3. Determina el espacio muestral de tirar tres volados consecutivos, luego responde ¿cuál es el riesgo que corres de perder? si apuestas a que:
   a. Caerán tres águilas en tres volados.
   b. Caerán al menos dos águilas en tres volados.
   c. Tirarás exactamente dos águilas en tres volados (con tres pierdes).

4. Considera las 52 cartas de una baraja inglesa (página 37) y responde:
   a. ¿Cuál es la probabilidad de sacar una carta y que sea de corazones?
   b. ¿Cuál es la probabilidad de sacar una carta negra?
   c. ¿Cuál es la probabilidad de sacar una carta que sea un cinco o un siete?
   d. ¿Cuál es la probabilidad de sacar dos cartas y que sean un par de ochos?
   e. ¿Cuál es la probabilidad de sacar una corrida del as al cuatro de cualquier palo?



5. Considera las 28 fichas de un dominó tradicional. Calcula:
   a. La probabilidad de que al tomar, cara abajo, una ficha al azar, ésta tenga un cuatro.
   b. La probabilidad de tomar dos fichas y que ambas tengan un seis.
   c. La probabilidad de que al tomar siete fichas te salgan las siete mulas (una mula es una ficha que tiene el mismo número en los dos lados).

6. Para una fiesta infantil se compraron 23 aguas de limón, 10 de jamaica y siete de horchata. Si las aguas se repartieran aleatoriamente y suponiendo que no te gusta el agua de horchata, ¿cuál es el riesgo de que te toque un agua de horchata?

7. Si en un salón de clases de 33 estudiantes, 29 saben quién es Germán Garmendia, ¿cuál es la probabilidad de que un estudiante nuevo no sepa quién es dicho personaje?

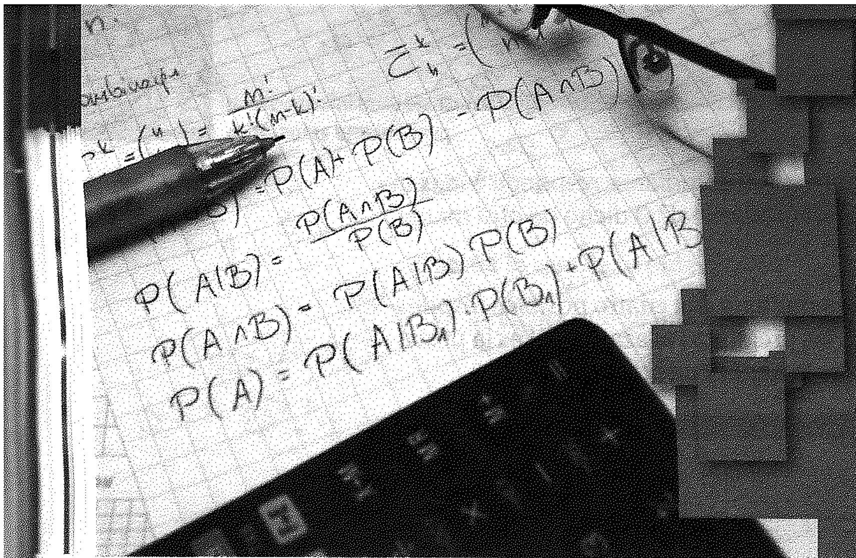
8. En un examen de 15 preguntas de falso y verdadero, si un estudiante responde todas las pregunta al azar ¿cuál es la probabilidad de responder todas mal? O dicho de otra manera ¿qué riesgo corre de responder todas las mal?

9. En un pequeño parcial de cinco preguntas de opción múltiple con cuatro posibles respuestas, un estudiante, que no estudió, responde todas las preguntas al azar, ¿cuál es la probabilidad de que el estudiante tenga tres preguntas contestadas correctamente?

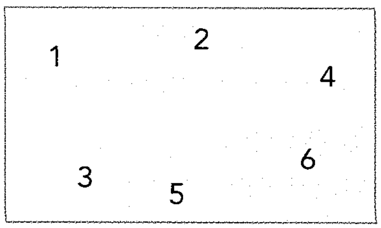
Actividad de aprendizaje 10

Efectúa lo que se te pide.

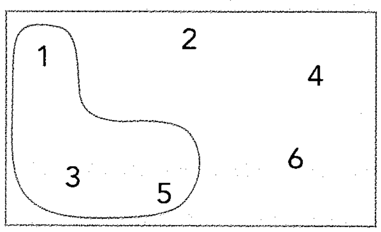
1. En equipo lleva a cabo las actividades y apunta en tu cuaderno los resultados:
   a. Lanza dos volados, registra todas las formas posibles en que pueden caer las monedas.
   b. Lanza 10 volados, registra el número de veces que cae águila y sol.
   c. Si lanzas dos monedas al mismo tiempo, ¿qué posibilidad existe de:
      - que caigan dos águilas?
      - que caigan dos soles?
      - que caigan al mismo tiempo un águila y un sol?
2. Resuelve los siguientes problemas y argumenta las respuestas de cada cuestionamiento:
   a. Un grupo de preparatoria está conformado por 20 alumnos y tres de ellos poseen una beca. Se decide otorgar una beca más. ¿Cuál es la probabilidad de que un alumno ya becado salga sorteado nuevamente?
   b. Si 12 de 15 competidores cumplieron con la prueba de tiempo en los 100 metros planos, ¿cuál es el riesgo de que un competidor tomado al azar no cumpla con el tiempo de dicha prueba?
   c. Registra en tu cuaderno las últimas 20 veces que llegaste tarde a tu casa. ¿Cuántas veces te regañaron o te llamaron la atención? Qué riesgo tienes de que la próxima vez que llegues tarde a casa te regañen.
3. Imagina dos fábricas de tornillos A y B. La fábrica A produce 1 tornillo defectuoso por cada 10 tornillos y vende cada tornillo a 26 centavos, mientras que la fábrica B produce 1 tornillo defectuoso cada 4 tornillos producidos y los vende a 19 centavos el tornillo. ¿En qué fábrica hay menor riesgo de comprar un tornillo defectuoso? ¿en qué fábrica conviene comprar 1000 tornillos?
4. Un empresario tiene un millón de pesos y le gustaría invertirlos en negocios locales. En este momento tiene tres opciones: una tortillería, un café y una papelería. En los últimos cinco años los dueños de la tortillería han invertido 42 000 pesos y han tenido ganancias de 63 000 pesos; los del café han invertido 68 000 pesos y han tenido ganancias de 75 000; y los de la papelería han invertido 95 000 pesos y han tenido ganancias de 90 000. ¿Cuál es el negocio con menor riesgo para invertir?



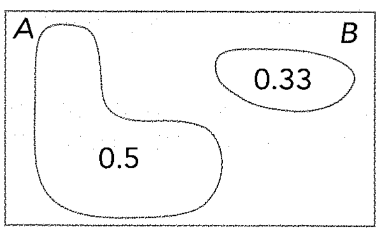
Contextualización de los elementos de probabilidad condicion e interpretación intuitiva del teorema de Bayes



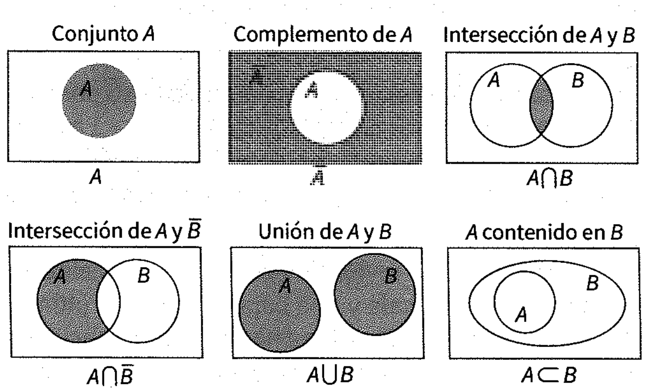
El espacio muestral es un conjunto, por esta razón en probabilidad se usa la notación y las herramientas gráficas (diagramas de Euler y Venn) de la teoría de conjuntos. Se puede representar al conjunto de posibles resultados (espacio muestral) como un área. Por ejemplo, pensemos en el lanzamiento de un dado de seis caras, el espacio muestral se podría representar como en el diagrama de la izquierda, donde el área del rectángulo representa al espacio muestral.



Ahora imagina que quieres representar en esa área la probabilidad de que el dado caiga en un número impar. Independientemente del cálculo y la magnitud de la probabilidad, puedes representarla con una figura dentro del espacio muestral, como se observa en el diagrama de la izquierda.



A continuación, llamaremos a ese evento el evento A, cuya probabilidad es de 3/6 = 0.5, agregaremos la probabilidad de tirar un dos o un cuatro y llamaremos a ese evento el evento B, además borraremos los números que representan las seis caras del dado. El diagrama queda como el de la izquierda.



Ahora debes entender que estos diagramas se utilizan como una herramienta para representar y entender los axiomas y teoremas de la probabilidad. De aquí en delante debes recordar que el tamaño y forma de las figuras dentro de esos diagramas no representan la magnitud de la probabilidad, sólo debes usarlos como instrumentos para visualizar la interacción y distribución de las probabilidades.

En la imagen se presenta la notación y diagramas en teoría de conjuntos, el área azul denota el conjunto que nos interesa.

Axiomas de probabilidad

Los axiomas son verdades intuitivas que no necesitan una demostración, sobre ellos se construyen los teoremas y las teorías matemáticas. A continuación veremos los tres axiomas de la probabilidad.

Primer axioma

$$P(A) \geq 0$$

La probabilidad de un evento cualquiera ($P(A)$) es mayor o igual que cero, es decir, no existen las probabilidades negativas.

Segundo axioma

$$P(E) = 1$$

La probabilidad del espacio muestral ($P(E)$) es de uno, es decir, es totalmente seguro que algo va a pasar. En un diagrama de Venn, el área del rectángulo representa al espacio muestral, o sea que el área del rectángulo es igual a uno.

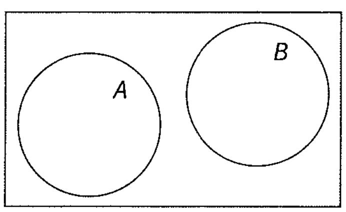
Tercer axioma

$$\text{Si } A \cap B \cap C \cap \ldots = \varnothing \text{ entonces } P(A \cup B \cup C \cup \ldots) = P(A) + P(B) + P(C) + \ldots$$

Si dos o más eventos no pueden suceder al mismo tiempo, es decir, la intersección de sus conjuntos está vacía, entonces la probabilidad del evento que representa la unión de ellos será la suma de sus probabilidades.

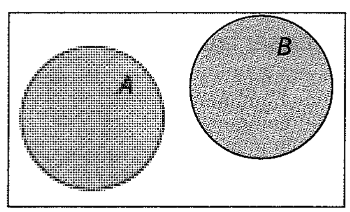
Por ejemplo, en el siguiente diagrama, la intersección de A y B está vacía ($A \cap B = \varnothing$), entonces la probabilidad de que ocurra A o B (la unión de A y B) es la suma de $P(A)$ y $P(B)$.

Si



$A \cap B = \varnothing$

entonces

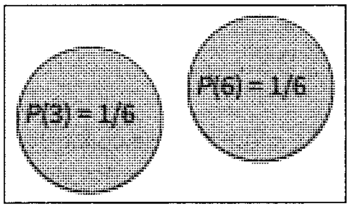


$P(A \cup B) = P(A) + P(B)$

Este axioma aplica para n cantidad de eventos sin intersección.

Ejemplo

En un dado de seis caras, ¿cuál es la probabilidad de que caiga tres o seis? En el lanzamiento de un dado no puede caer un tres y un seis en el mismo lanzamiento, por lo tanto, la intersección de que



$P(3 \bigcup 6) = P(3) + P(6)$

caiga tres y seis está vacía. Entonces, la probabilidad de que caiga tres o seis será la suma de las dos probabilidades; sustituyendo en la fórmula del axioma, tenemos:

Si caiga 3 $\bigcap$ caiga 6 = $\varnothing$ entonces P(caiga 3 $\bigcup$ caiga 6) = P(caiga 3) + P(caiga 6).

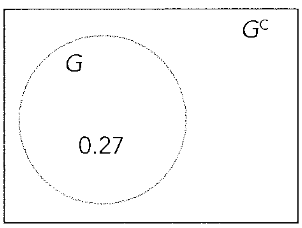
Es decir:

$$P(\text{caiga 3 U caiga 6}) = 1/6 + 1/6 = 2/6 = 0.333.$$

Teoremas básicos

Un teorema es una verdad construida a partir de axiomas, a continuación conocerás los teoremas básicos de la probabilidad explicados a través de axiomas y diagramas de Venn. Los teoremas aquí presentados son sólo algunos y los nombres de “primer teorema”, “segundo teorema”, sólo marcan el inicio de la explicación de cada uno.

Primer teorema



Imagina un juego donde puedes ganar (G), empatar (D) o perder (L) y que la probabilidad de ganar es de 0.27. ¿Cuál es la probabilidad de no ganar? Si sólo hay tres posibles resultados, las probabilidades de perder o empatar representan al conjunto de todo lo que no es ganar. Recuerda que en teoría de conjuntos todo lo que no es parte de un conjunto es su complemento. Dicho esto, el problema se puede representar como el diagrama de la izquierda, donde G es ganar y su complemento (G^c) es todo lo que no es ganar (empatar o perder). Sabemos por el segundo y tercer axioma, que la probabilidad del espacio muestral (todo el rectángulo) es igual a uno y que la unión de dos eventos sin intersección es igual a la suma de sus probabilidades. Si miras el diagrama, notarás que G y G^c no tienen intersección y forman todo el espacio muestral, entonces la suma de $P(G)$ y $P(G^c)$ debe ser igual a uno:

$$P(G) + P(G^c) = 1.$$

En este ejemplo queremos saber la probabilidad de no ganar (G^c), así que haciendo uso de nuestras habilidades algebraicas despejamos $P(G^c)$:

$$P(G^c) = 1 - P(G),$$

si sabemos que la probabilidad de ganar es de 0.27, sustituimos $P(G)$ por 0.27:

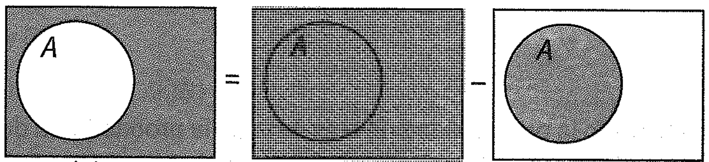
$$P(G^c) = 1 - 0.27 = 0.73.$$

En otras palabras, si la probabilidad de ganar es 0.27 y el no ganar no puede ocurrir al mismo tiempo que ganar, entonces la probabilidad de ganar debe ser la diferencia entre 1 y 0.27.

La expresión algebraica de este teorema es:

Teorema 1. $P(\bar{A}) = 1 - P(A)$.

La probabilidad del complemento de un evento A es igual a uno menos la probabilidad de ese evento. Si representamos el teorema con diagramas de Venn, se vería así:



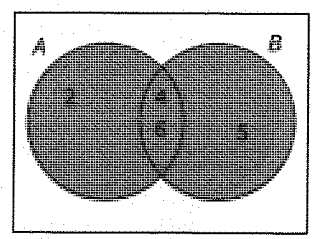
$P(A)$ 1 $P(A)$

Ejemplo

La probabilidad de un evento A es de 0.11, ¿cuál es la probabilidad de su complemento?

$$P(\bar{A}) = 1 - 0.11 = 0.89.$$

Segundo teorema



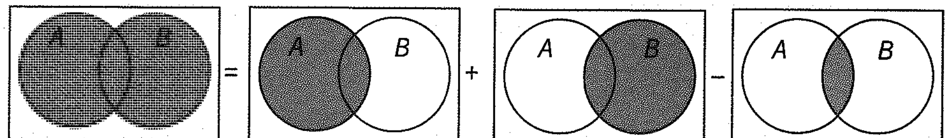
$A \cup B$

Imagina que en el lanzamiento de un dado, el evento A es que caiga un número par y el evento B que caiga cuatro, cinco o seis. ¿Cuál sería la probabilidad de que suceda A o B?, es decir, ¿cuál es la probabilidad de $A \cup B$? La representación gráfica de la unión de A y B la puedes ver en la imagen de la derecha. Sabemos que la probabilidad del evento A es de 0.5, pues la mitad de las caras del dado es par y la probabilidad del evento B también es 0.5, pues su conjunto también representa a la mitad de los posibles resultados. Pero si sumamos $P(A)$ más $P(B)$, estaríamos contando dos veces la probabilidad de tirar cuatro o seis pues ambos resultados se encuentran en los dos eventos, es decir, estaríamos contando dos veces la intersección de A y B. Por lo tanto, a la suma de $P(A)$ más $P(B)$ tendríamos que restarle una vez la intersección de A y B. Si expresamos todo esto en una ecuación, tenemos el teorema:

Teorema 2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Conocido como la ley aditiva de probabilidad, dice que la probabilidad del evento que representa la unión de dos eventos no excluyentes (con intersección) es igual a la suma de las probabilidades de cada evento menos la probabilidad de su intersección.

La representación de este teorema con diagramas sería:



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

Continuando con el caso de los eventos antes planteados, la intersección de A y B es de 2/6 (0.333), por lo tanto, tendríamos:

$$P(A \cup B) = 0.5 + 0.5 - 0.333 = 0.666.$$

Lo cual coincide con el cálculo clásico, pues la probabilidad de tirar dos, cuatro, cinco o seis con un dado de seis caras es de 4/6 = 0.666.

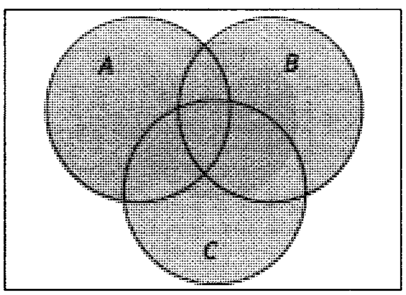
Ejemplo

En un experimento aleatorio la probabilidad de H es de 0.12, la probabilidad de K de 0.75 y la probabilidad de su intersección es de 0.09. ¿Cuál es la probabilidad de que suceda H o K? Aplicando el teorema tenemos que:

$$P(H \cup K) = P(H) + P(K) - P(H \cap K),$$

sustituyendo:

$$P(H \cup K) = 0.12 + 0.75 - 0.09 = 0.78.$$



$A \cup B \cup C$

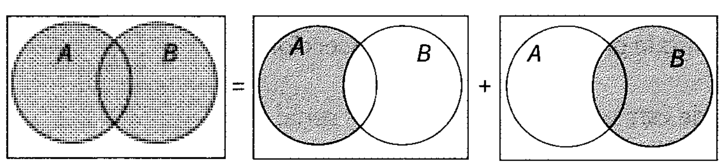
El mismo razonamiento para entender el teorema que acabas de aprender lo puedes aplicar a la unión de más de dos eventos. Por ejemplo, en la unión de tres eventos como el que se muestra en la imagen de la izquierda. Si sumas A más B más C y luego restas las intersecciones de cada par de conjuntos, habrás restado tres veces la intersección de los tres conjuntos (el triángulo del centro), por lo que es necesario sumar esa intersección una vez. La expresión para la unión de tres eventos no excluyentes quedaría:

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

Otra forma que tiene el teorema de la ley aditiva de probabilidad es:

Teorema 3. $P(A \cup B) = P(A \cap \bar{B}) + P(B)$.

Si lo proyectamos con diagramas, sería:

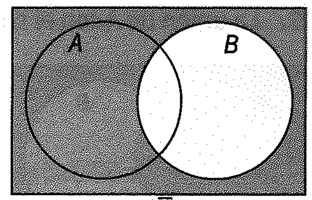


$P(A \cup B)$ $P(A \cap \bar{B})$ $P(B)$

Ejemplos

Si la $P(A) = 0.7$, la $P(B) = 0.4$ y la $P(A \cap B) = 0.2$, determinaremos la probabilidad de algunos eventos.

1. El complemento de B:



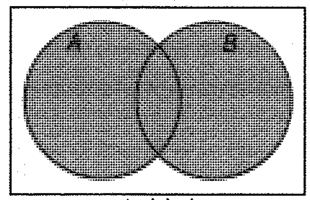
$P(\overline{B})$

$P(\overline{B}) = 1 - P(B)$

$P(\overline{B}) = 1 - 0.4$

$P(\overline{B}) = 0.6$

2. La unión de A y B:



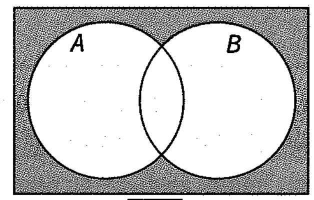
$P(A \cup B)$

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

$P(A \cup B) = 0.7 + 0.4 - 0.2$

$P(A \cup B) = 0.9$

3. El complemento de la unión de A y B:



$P(\overline{A \cup B})$

Usando el teorema 1 sabemos que:

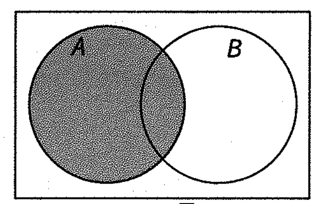
$P(\overline{A \cup B}) = 1 - P(A \cup B)$,

por lo tanto:

$P(\overline{A \cup B}) = 1 - 0.9$

$P(\overline{A \cup B}) = 0.1$

4. La intersección de A con el complemento de B:



$P(A \cap \overline{B})$

Por el teorema 3 sabemos que:

$P(A \cap \overline{B}) = P(A \cup B) - P(B)$,

por lo tanto:

$P(A \cap \overline{B}) = 0.9 - 0.4$

$P(A \cap \overline{B}) = 0.5$

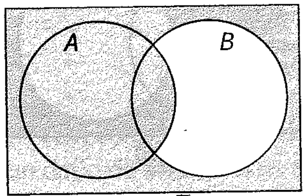
Si ves el diagrama, puedes darte cuenta de que también resuelves el ejercicio si a la $P(A)$ le restas la intersección de $P(A)$ con $P(B)$:

$P(A \cap \overline{B}) = P(A) - P(A \cap B)$

$P(A \cap \overline{B}) = 0.7 - 0.2$

$P(A \cap \overline{B}) = 0.5$

5. La unión de A con el complemento de B:



$P(A\cup\overline{B})$

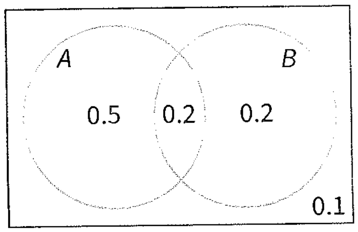
Por el teorema 1 sabemos que:

$$P(A\cup\overline{B}) = P(A) + P(\overline{B}) - P(A\cap\overline{B}),$$

por lo tanto:

$$P(A\cup\overline{B}) = 0.7 + 0.6 - 0.5 = 0.8$$

Una estrategia visualmente útil cuando se trabaja con probabilidades y diagramas de Venn es anotar las probabilidades de cada conjunto delimitado, es decir, de los conjuntos sin intersecciones. En este ejemplo tenemos cuatro conjuntos delimitados dentro del espacio muestral que son: $P(A\cap\overline{B})$, $P(\overline{A}\cap B)$, $P(A\cap B)$ y $P(\overline{A\cup B})$, y además hemos calculado las probabilidades de casi todos ellos. En la imagen de la izquierda verás estos cuatro conjuntos con su respectiva probabilidad.



La suma de todos los conjuntos debe ser igual a la probabilidad del espacio muestral (uno). Puedes corroborar todos los cálculos previos sumando y restando a partir de este diagrama.

Actividad de aprendizaje 11

Resuelve los ejercicios.

1. Resuelve en tu cuaderno el siguiente ejercicio. Si la probabilidad de un evento F es de 0.29, la probabilidad de un evento D es 0.41 y la probabilidad de la intersección de ambos eventos es de 0.17, haz los diagramas de Venn correspondientes y determina la probabilidad de:

   a. El complemento de F.

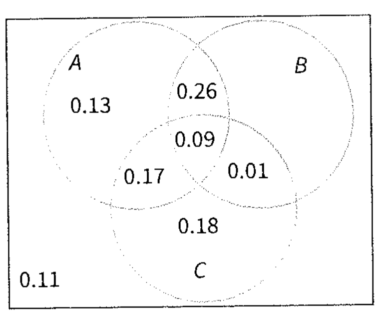
   b. La unión de F y D.

   c. El complemento de la unión de F y D.

   d. La intersección de D con el complemento de F.

   e. La unión de F con el complemento de D.

   f. La unión de D con el complemento de F.



2. Utiliza el diagrama de la izquierda para determinar:

   a. $P(A)$

   b. $P(B)$

   c. $P(B\cup C)$

   d. $P(A\cap C)$

   e. $P(\overline{A\cap B\cap C})$

Eventos condicionales en el estudio de la probabilidad

La relevancia de este tipo de eventos radica en su utilidad para calcular probabilidades de eventos que dependen de otros, básicamente incorpora al cálculo información (probabilidades) de otros eventos que afectan tu evento de interés.

Primero debes aprender qué es una probabilidad condicional. Imagina que después de haber encuestado a 119 personas de tu ciudad acerca de si juegan videojuegos, obtienes la tabla de frecuencias de abajo.

	Juega	No juega
Hombre	39	12
Mujer	17	51

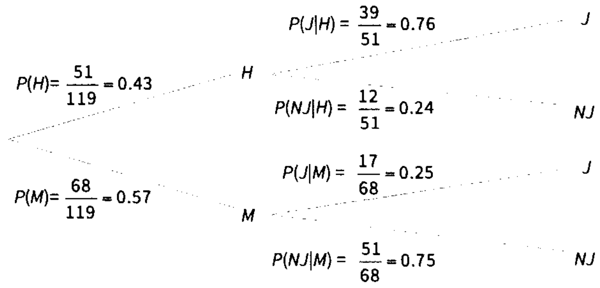
Ahora piensa en las siguientes preguntas, si tomaras a una persona al azar de esta ciudad, ¿cuál es la probabilidad de que sea mujer y juegue videojuegos?: si sabes que una persona es mujer, ¿la de que juegue videojuegos?, y si sabes que una persona juega videojuegos, ¿la de que sea mujer?

Si haces uso de tus conocimientos probabilísticos clásicos, sabes que la probabilidad de un evento es la proporción que ocupa ese evento del espacio muestral. Entonces, la probabilidad de que tomes a una persona al azar y esta sea mujer y juegue videojuegos es de 17/119 = 0.14, pues hubo 17 mujeres jugadoras de un total de 119 personas. De esas 119 personas 68 fueron mujeres y de ellas, 17 fueron jugadoras, por lo tanto, si sabes que una persona es mujer, la probabilidad de que sea jugadora es de 17/68 = 0.25. Por último, de los 119 encuestados 56 fueron jugadores y de ellos, 17 fueron mujeres, por lo tanto, si sabes que una persona juega videojuegos la probabilidad de que sea mujer es de 17/56 = 0.30.

Las últimas dos probabilidades se dice que son probabilidades condicionadas, pues su valor está condicionado por un evento anterior. En el primer caso, la probabilidad de ser jugadora está condicionada por el hecho de que se trate de una mujer; y en el segundo caso la probabilidad de ser mujer está condicionada por ser jugador.

La notación para la probabilidad de un evento condicionado es $P(A|B)$, esto es la probabilidad de que suceda A sabiendo que ha sucedido B. Dependiendo de la fuente bibliográfica, ésta probabilidad se puede leer de diferentes formas pero todas significan lo mismo. Algunos ejemplos son: la probabilidad de A dado B; si ha sucedido B, la probabilidad de que pase A; la probabilidad de que si pasó B suceda A; probabilidad de A sabiendo que ha ocurrido B.

Si representamos el ejercicio con un diagrama de árbol, podemos ver cuatro probabilidades condicionadas (H-hombre, M-mujer; J-jugador; NJ-No jugador).



Por definición, la **probabilidad condicional** de un evento A sabiendo que ha ocurrido B, es igual a:

$$P\left(A\middle|B\right)=\frac{P\left(A\cap B\right)}{P\left(B\right)}.$$

Ejemplo

Calcula la probabilidad condicionada de que una persona sea jugadora sabiendo que es hombre. Utilizando la ecuación, sería como sigue:

$$P\left(J\middle|H\right)=\frac{P\left(J\cap H\right)}{P\left(H\right)},$$

donde $P(J\cap H)$ es la probabilidad de que una persona sea hombre y jugador de videojuegos, y $P(H)$ la probabilidad de ser hombre. Ambas probabilidades ya las conocemos, si sustituimos tenemos:

$$P\left(J\middle|H\right)=\frac{\frac{39}{119}}{\frac{51}{119}}=\frac{0.33}{0.43}=0.76.$$

Lo cual coincide con la probabilidad que aparece en el diagrama de árbol.

Ejemplo

Ahora supón que no tienes las frecuencias de las personas encuestadas y que sólo tienes las probabilidades del diagrama de árbol. Retomemos ahora las tres preguntas del tercer párrafo de la página anterior. Si tomaras a una persona al azar de esta ciudad:

¿Cuál es la probabilidad de que sea mujer y juegue videojuegos?, en otras palabras, ¿cuál es $P(J\cap M)$?

Si utilizamos nuestros conocimientos algebraicos, podemos despejar $P(A\cap B)$ de la ecuación de probabilidad condicional y tendríamos:

$$P(A\cap B) = P(B)\cdot P(A \mid B).$$

A esta operación se le conoce como la ley multiplicativa de probabilidad y establece que la probabilidad de la intersección de dos eventos A y B es igual a multiplicar la probabilidad de uno por la probabilidad condicionada del otro.

Recuerda que:

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

Por lo anterior:

$$P(B)\cdot P(A \mid B) = P(A)\cdot P(B \mid A).$$

Aplicando la ley multiplicativa, tenemos que:

$$P(J \cap M) = P(M)\cdot P(J \mid M),$$

tanto $P(M)$ como $P(J|M)$ se encuentran en el diagrama de árbol, si sustituimos:

$$P(M \cap J) = 0.57\cdot 0.25 = 0.14.$$

Verás que el resultado es el mismo que la probabilidad que calculamos usando nuestros conocimientos probabilísticos clásicos.

Ejemplo

¿Cuál es la probabilidad de que si tomas a una mujer, esta juegue videojuegos? Esta probabilidad ya está escrita en el diagrama, corresponde a la $P(J|M)$, que es de 0.25.

Ejemplo

¿Cuál es la probabilidad de que si tomas a una persona que juega videojuegos, sea mujer? En otras palabras, ¿cuál es el valor de $P(M|J)$? Si ves el diagrama de árbol esta probabilidad condicionada no está ahí. Es en este tipo de situaciones donde se usa el teorema de Bayes; aplicándolo, la forma de calcular $P(M|J)$ sería:

$$P\left(M \middle| J\right) = \frac{P\left(M\right)\cdot P\left(J \middle| M\right)}{P\left(M\right)\cdot P\left(J \middle| M\right) + P\left(H\right)\cdot P\left(J \middle| H\right)}$$

todas esas probabilidades sí las tenemos en nuestro diagrama de árbol, sustituyendo:

$$P(M \mid J) = \frac{0.57\cdot 0.25}{0.57\cdot 0.25 + 0.43\cdot 0.76} = \frac{0.14}{0.14 + 033} = \frac{0.14}{0.47} = 0.30.$$

La $P(M|J)$, es decir, la probabilidad de que una persona sea mujer sabiendo que esa persona juega videojuegos es de 0.30 y coincide con la probabilidad calculada cuando utilizamos la tabla de frecuencias.

Si lo meditas con calma, el teorema de Bayes sigue el mismo razonamiento que usamos en la probabilidad clásica, sólo que el teorema usa probabilidades en vez de frecuencias. Cuando conocíamos el número de personas encuestadas, para calcular la probabilidad de que una persona fuera mujer sabiendo que era una persona jugadora, calculamos la proporción de mujeres que juegan del total de personas jugadoras. Con el teorema de Bayes estamos haciendo lo mismo, estamos calculando la proporción de la probabilidad de una mujer jugadora del total de la probabilidad de que una persona sea jugadora. Analiza y compara los siguientes dos diagramas:

Probabilidad clásica

$H \cap J$ 39	$H \cap NJ$ 12
$M \cap J$ 17	$M \cap NJ$ 51

$$P(M|J) = \frac{17}{17+39}$$

$$P(M|J) = \frac{17}{56} = 0.30$$

Mujeres jugadoras

Teorema de Bayes

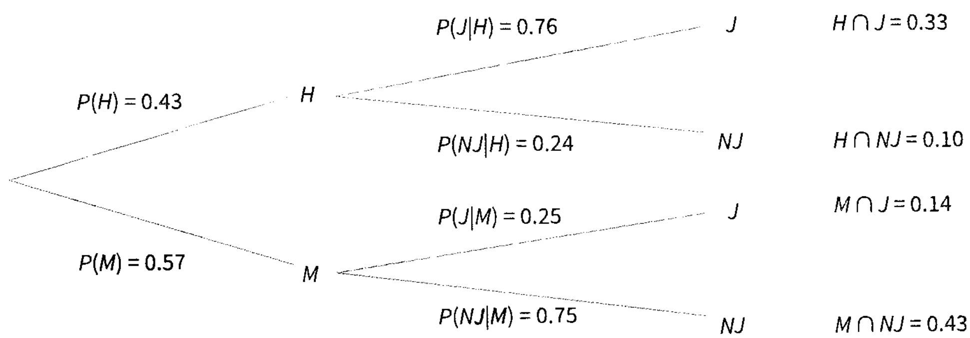
$H \cap J$ 0.33	$H \cap NJ$ 0.10
$M \cap J$ 0.14	$M \cap NJ$ 0.43

$$P(M|J) = \frac{0.14}{0.14+0.33}$$

$$P(M|J) = \frac{0.14}{0.47} = 0.30$$

Total de jugadores

Las probabilidades del diagrama de la derecha (Teorema de Bayes) las puedes calcular con la ley multiplicativa de probabilidad y las probabilidades del diagrama de árbol. Otra forma de visualizar el teorema de Bayes es con el siguiente diagrama de árbol de probabilidades:



Y ahora pregúntate: ¿cuál es el total de la probabilidad de que una persona juegue videojuegos? Pues debe ser la suma de la probabilidad de los hombres jugadores, más la probabilidad de las mujeres jugadoras:

$$0.14 + 0.33 = 0.47,$$

y de las personas jugadoras, ¿cuál es la proporción de las mujeres que juegan?

$$0.14 / 0.47 = 0.30.$$

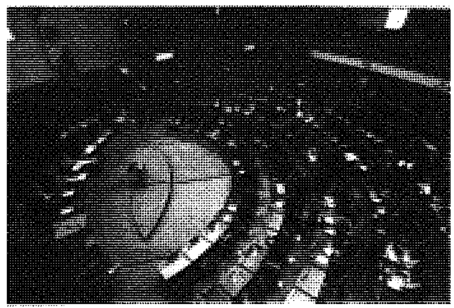
Este ejemplo es muy sencillo pues cada evento (sexo y jugar videojuegos) sólo tiene dos posibles resultados (hombre, mujer; juega, no juega), sin embargo, el teorema de Bayes está formulado para eventos con muchos posibles resultados.

Teorema de Bayes

$$P\left(B_j \middle| A\right)=\frac{P\left(B_j\right)P\left(A \middle| B_j\right)}{\sum_{i=1}^{k} P\left(B_j\right)P\left(A \middle| B_j\right)}$$

Donde B_j y B_i representan posibles resultados de B dado A, y k representa el número de posibles resultados de A.

Ejemplo



En la LXIV legislatura, 52% de los diputados son de MORENA, 16% del PAN, 9% del PRI y el resto pertenece a otros partidos. Imagina que se promueve una iniciativa de ley para eliminar el fuero constitucional a diputados y senadores. En MORENA, 88% de los diputados se declara en contra, en el PAN, 70% se declara en contra, en el PRI, 44% se declara en contra y de los otros partidos 91% se declara a favor.

¿Cuál es la probabilidad de que un diputado que esté a favor, sea del PRI? En palabras probabilísticas, ¿cuál es la probabilidad de que un diputado sea del PRI sabiendo que está a favor?, es decir, $P(\text{PRI}|F)$. Usando el teorema de Bayes, tendríamos:

$$P\left(\text{PRI} \middle| F\right)=\frac{P\left(\text{PRI}\right)\cdot P\left(F \middle| \text{PRI}\right)}{P\left(\text{PRI}\right)\cdot P\left(F \middle| \text{PRI}\right)+P\left(\text{MOR}\right)\cdot P\left(F \middle| \text{MOR}\right)+P\left(\text{PAN}\right)\cdot P\left(F \middle| \text{PAN}\right)+P\left(\text{Otros}\right)\cdot P\left(F \middle| \text{Otros}\right)}.$$

Si 9% de la cámara de diputados pertenece al PRI, entonces la $P(\text{PRI})$, es decir, la probabilidad de que un diputado sea del PRI es de 0.09. Si aplicamos el mismo razonamiento para el resto de los partidos, tenemos que la $P(\text{MORENA}) = 0.52$ y la $\text{P}(\text{PAN}) = 0.16$. Si entre MORENA, el PAN y el PRI suman 77% de la cámara, entonces los otros partidos deben sumar el 23% restante, por lo tanto, $P(\text{Otros}) = 0.23$.

Ahora vamos con las probabilidades condicionadas. ¿Cómo determinar la $P(F|\text{PRI})$?, es decir, la probabilidad de que un diputado del PRI esté a favor. Si 44% de los diputados del PRI está en contra, entonces el 56% restante debe está a favor, por lo tanto, $P(F|\text{PRI}) = 0.56$. Siguiendo ese razonamiento puedes determinar que $P(F|\text{MORENA}) = 0.12$, $P(F|\text{PAN}) = 0.3$ y $P(F|\text{Otros}) = 0.91$. Ahora sustituimos todos los valores en la ecuación:

$$P\left(\text{PRI} \middle| F\right)=\frac{0.09\cdot 0.56}{0.09\cdot 0.56+0.52\cdot 0.12+0.16\cdot 0.3+0.23\cdot 0.91}.$$

Si resolvemos, tenemos:

$$P(\text{PRI}|F)=\frac{0.0504}{0.0504+0.0624+0.048+0.2093}=\frac{0.0504}{0.3701}=0.13\,.$$

Actividad de aprendizaje 12

Efectúa lo que se te pide.

1. Utiliza el teorema de Bayes para calcular las siguientes probabilidades condicionadas utilizando la información del diagrama de árbol de la relación de mujeres y hombres que juegan o no videojuegos:
   a. $P(M|NF)$
   b. $P(H|F)$
   c. $P(H|NF)$
2. Elabora el diagrama de árbol que correspondería a la información acerca de la cámara de diputados y su postura a favor o en contra de la iniciativa de ley para eliminar el fuero constitucional a diputados y senadores. Anota en cada rama del árbol su propia probabilidad.
3. Utiliza tus conocimientos, la información en tu diagrama de árbol del ejercicio anterior y el teorema de Bayes para responder las preguntas:
   a. ¿Cuál es la probabilidad de que un diputado que esté a favor pertenezca al PAN?
   b. ¿Cuál es la probabilidad de que un diputado que esté en contra pertenezca a MORENA?
   c. ¿Cuál es la probabilidad de que un diputado que esté en contra no pertenezca ni a MORENA ni al PAN ni al PRI?
   d. ¿Cuál es la probabilidad de que un diputado que está en contra pertenezca al PRI o al PAN?

Actividad de aprendizaje 13

Efectúa lo que se te pide.

1. En equipos lleven al salón dos dados de seis caras y diferente color, tres canicas (una verde, una azul y una roja) y lleven a cabo las actividades:
   a. Lancen cien veces los dos dados y registren la suma de cada lanzamiento, luego calculen la probabilidad de que:
      - La suma de las caras sea 13.
      - La suma de las caras sea cinco.
      - La suma de las caras sea menor que trece.
      - La suma de las caras sea menor o igual a seis.

- La suma de las caras sea mayor o igual a seis.
- Encuentra la intersección y la unión de la probabilidad obtenida en los dos incisos anteriores.

b. Introduzcan las tres canicas en una bolsa, saquen una canica 60 veces y registren el color de cada evento. Con los resultados registrados obtengan la probabilidad de que al extraer una canica sea azul.

2. Anota en tu cuaderno la información que obtuviste de estas actividades y coméntalo con el grupo.
3. Resuelvan los problemas en equipo:

a. Si A y B son eventos mutuamente excluyentes (sin intersección), $P(A) = 0.29$ y $P(B) = 0.43$, haz los diagramas correspondientes y determina:

- $P(\bar{A})$
- $P(A \cup B)$
- $P(A \cap B)$
- $P(\bar{A} \cap \bar{B})$

b. En una empresa fabricante de circuitos para computadoras, los trabajadores del grupo A producen 68% del total de la fabricación, mientras que el grupo B produce 32%. De acuerdo con el departamento de control de calidad, 2% de la producción del grupo A es defectuosa, mientras que del grupo B el 5%. Calcula la probabilidad de que al revisar al azar un circuito, éste sea defectuoso y provenga de la producción del grupo B.

4. Participa con respeto y tolerancia en un foro de discusión acerca del uso de los axiomas y teoremas de la probabilidad para la toma de decisiones; concluye acerca de la importancia de utilizarlos en situaciones reales, por ejemplo: la Lotería Nacional, compañías de seguros, controles de calidad. Anota tus conclusiones en tu cuaderno de apuntes.

Diagonales

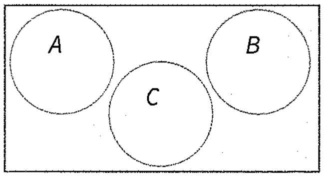
Las matemáticas relacionan conceptos y aspectos prácticamente ajenos de una forma sorprendente, por ejemplo, ¿qué tienen en común el triángulo de Pascal, los vértices de un polígono y la selección de un comité de r miembros entre n disponibles?

La fórmula de combinatoria nCr, se aplica para todos los casos descritos antes, investiga y encuentra la relación entre ellos.

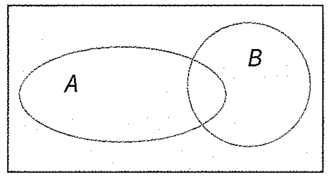
Entrena tus conocimientos

Responde los ejercicios usando lo aprendido en este parcial.

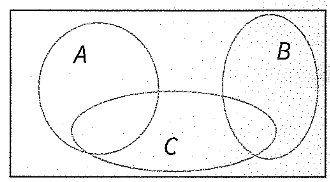
1. En los diagramas colorea el área correspondiente al conjunto señalado.



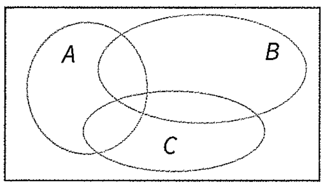
$B \cup C$



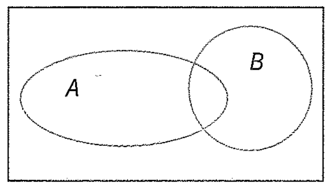
$\overline{A}$



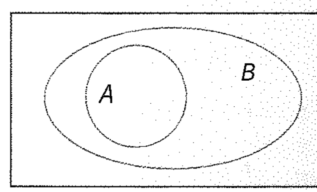
$\overline{A \cup C}$



$A \cap B \cap C$



$\overline{A \cap B}$



$A \cap B$

2. Si la probabilidad de un evento A es de 0.62, la probabilidad de un evento C es 0.35 y la probabilidad de la intersección de ambos eventos es de 0.21, haz los diagramas de Venn correspondientes y determina la probabilidad de:

   a. El complemento de A.
   b. La unión de A y C.
   c. El complemento de la unión de A y C.

3. Supón que la probabilidad de que te encuentres en la calle a una persona rica es de 0.0001. Por otro lado, considera que la probabilidad de que una persona rica traiga 5 000 pesos en efectivo es de 0.9 y que la probabilidad de que una persona que no es rica traiga 5 000 pesos en efectivo es de 0.17.

   a. Dibuja el diagrama de árbol que corresponde a las probabilidades de los eventos en el ejercicio. Anota en cada rama su respectiva probabilidad.
   b. Utiliza el teorema de Bayes para determinar la probabilidad de que sabiendo que una persona trae 5 000 pesos en efectivo, ésta sea rica.
   c. Usando el teorema de Bayes calcula la probabilidad de que sabiendo que alguien no trae 5 000 pesos esa persona no sea rica.

4. Tu papá necesita comprar neumáticos nuevos para su auto, y existen 0.17, 0.22, 0.03, 0.29, 0.21 y 0.08 de probabilidad de que compre neumáticos de las marcas: Uniroyal, Goodyear, Michelin, General, Goodrich o Armstrong. Determina las probabilidades de que compre:

   a. Neumáticos Goodyear o Goodrich.
   b. Neumáticos Uniroyal, General o Goodrich.
   c. Neumáticos Michelin o Armstrong.
   d. Neumáticos Goodyear, General o Armstrong.

Suma anécdota

La toma de decisiones y el análisis de riesgos están presentes en todos los aspectos de nuestras vidas. Todo el tiempo se nos presentan opciones de las cuales, según nuestros intereses, podemos adoptar: lo más fácil, lo más redituable, lo más bonito, lo más placentero, etcétera.

Decisiones acertadas impactan de tal forma que mejoran nuestro entorno y nuestras relaciones de forma considerable, pero también, asumir un riesgo grande y tomar malas decisiones, puede arruinarnos y perjudicar a quienes nos rodean.

Girolamo Cardano fue un científico italiano que destacó en diversidad de disciplinas, hizo grandes contribuciones a las ciencias y tuvo una prolífica carrera como escritor de obras científicas y filosóficas. *Publicó el primer tratado sistemático de probabilidades*, además de un método de resolución de ecuaciones cúbicas y cuárticas, introduciendo los números complejos. Fue el médico más famoso de su tiempo, por las curaciones casi milagrosas que efectuaba y sus publicaciones al respecto, siendo consultado, incluso por los mejores médicos de la época.

Sin embargo, tenía una afición al juego que le acarreó muchos problemas desde la adolescencia y hasta poco tiempo antes de su muerte. Su vida se vio envuelta en escándalos y disputas que le impidieron conseguir empleo, mantenerlo o ejercer su profesión, pues se vio varias veces confrontado con autoridades religiosas, civiles y educativas. A pesar de ser un brillante médico y gran matemático, su carácter y ludopatía lo llevaron a tomar decisiones arriesgadas.

Cardano era consciente de sus debilidades. Escribió de sí mismo: "He recibido de la naturaleza un espíritu filosófico e inclinado a la ciencia. Soy ingenioso, amable, elegante, voluptuoso, alegre, piadoso, amigo de la verdad, apasionado por la meditación, dotado de talento inventivo y lleno de doctrina. Me entusiasman los conocimientos médicos y adoro lo maravilloso. Astuto, investigador y satírico, cultivo las artes ocultas. Sobrio, laborioso, aplicado, detractor de la religión, vengativo, envidioso, triste, pérfido y mago, sufro mil contradicciones. Lascivo, misántropo, dotado de facultades adivinatorias, celoso, calumniador e inconstante, contemplo el contraste entre mi naturaleza y mis costumbres".

Por todo esto, Cardano atrajo sobre su vida y obra, simultáneamente, la admiración y el rechazo de los mayores pensadores, matemáticos, científicos, juristas y teólogos de su tiempo. Su espíritu científico lo llevó muy lejos, pero su arrogancia y la tendencia a arriesgarlo todo, lo pusieron muchas veces en situaciones deplorables. De Cardano podemos vislumbrar que las decisiones siempre tienen consecuencias.

- **Elabora un cuadro comparativo de las decisiones que debes tomar en los años venideros y las repercusiones que podrían tener en tu vida, así como la recompensa inmediata que ofrecen. Sopesa las ventajas y desventajas de las opciones disponibles y determina la motivación que te lleva, en primera instancia, a preferir una de las opciones disponibles.**

Proyecto integrador

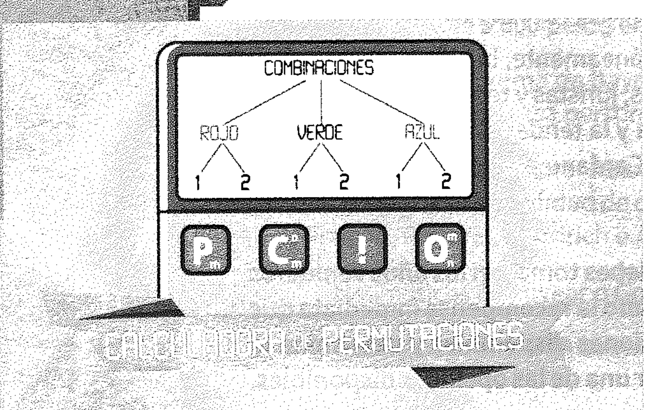
Tratamiento de la información.

El hecho de aleatoriedad, muchas veces es opuesto a nuestra percepción, de tal modo que nos vemos obligados a encontrar patrones ante una secuencia de datos. Aunque el hecho de que un evento sea aleatorio, impide que haya una regla que nos ofrezca los datos arrojados, podemos establecer conclusiones de acuerdo con la naturaleza del evento.

Considera lo anterior para desarrollar la actividad.

En parejas, contesten las preguntas y realicen lo que se pide.

1. ¿Cómo creen que se distribuyan los resultados al lanzar 20 veces una moneda?
2. ¿Serían capaces de describir 20 resultados de lanzar una moneda de forma que otras personas piensen que han lanzado la moneda realmente?
3. Inventen una secuencia de 20 posibles resultados al lanzar una moneda equilibrada, de modo que la secuencia pueda pasar como aleatoria para otra persona.
4. ¿Creen que al mostrar esos resultados, una persona piense que realmente lanzaron la moneda 20 veces?
5. Comparen su resultado con 20 lanzamientos reales de una moneda.
6. Presenten uno de los resultados a sus compañeros del grupo para que indiquen si fue el inventado o el generado por el lanzamiento real de la moneda.
7. ¿Cómo se puede distinguir una secuencia realmente aleatoria de otra inventada?
8. Discutan los resultados en grupo y lleguen a una conclusión con el docente.



Calculadora de permutaciones

Muchas situaciones se resuelven de forma más rápida utilizando las herramientas diseñadas para determinadas tareas. Sin embargo, al conocer el funcionamiento de la herramienta, se podrá emplear de forma más efectiva.

Así, te ofrecemos una calculadora de permutaciones para facilitar los resultados en tus operaciones.

Hacia la prueba Planea

Resuelve los problemas que se plantean.

1. En el colegio se realizó una campaña para medir el grado de obesidad de la población infantil. En la escuela hay 1 430 estudiantes, pero sólo 8 de cada 10 aceptaron pesarse, ¿cuántos alumnos faltaron por pesar?

   a. 143　　b. 286　　c. 715　　d. 1 144

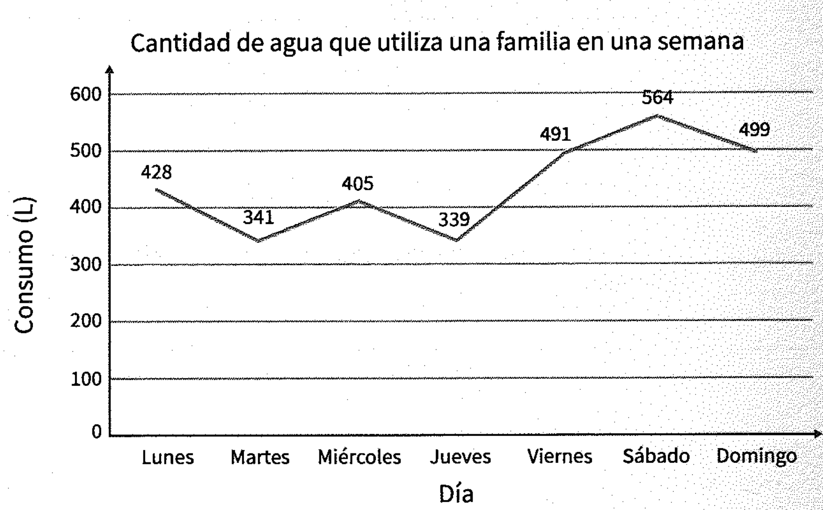
2. Los estudiantes, en un curso de matemáticas, anotaron en el pizarrón el número de hermanos que tienen. Al final, el profesor solicitó que obtuvieran el promedio. Las cantidades anotadas fueron: 3, 5, 0, 5, 2, 1, 1, 2, 3, 0, 1. ¿Cuál es el promedio?

   a. 2　　b. 1.9　　c. 1.5　　d. 1

3. Una urna contiene 5 bolas rojas, 6 verdes y 4 blancas. Si se saca una sola de ellas, ¿cuál es la probabilidad de que sea verde?

   a. $\frac{6}{1}$　　b. $\frac{5}{2}$　　c. $\frac{2}{5}$　　d. $\frac{1}{6}$

4. Una familia consumió 3 067 L de agua en la semana. La gráfica muestra la cantidad de agua en litros que la familia utilizó cada día de la semana.



   ¿Qué día de la semana se consumió aproximadamente el promedio semanal de agua?

   a. Sábado　　b. Jueves

   c. Martes　　d. Lunes

5. ¿Cuál es la probabilidad de que al lanzar un dado se obtenga un número impar menor que 5?

   a. $\frac{6}{2}$　　b. $\frac{1}{6}$　　c. $\frac{1}{3}$　　d. $\frac{3}{6}$

Evalúa tus evidencias

Productos	Criterios	Sí	No
Cálculo del promedio de una colección de datos.	Se presenta un gráfico que describe la situación.		
	El cálculo efectuado es correcto y puntual.		
	Al cálculo le acompaña una conclusión puntual.		
Construcción de tablas de frecuencia.	La información presentada es verídica y demuestra una investigación profunda.		
	Se presenta en una tabla la información investigada.		
	La gráfica presentada está en correspondencia con la información de la tabla.		
Cálculo de la probabilidad de un evento.	Los datos presentados son correctos.		
	Se presenta una representación no numérica para respaldar el cálculo.		
	Se da una conclusión significativa.		

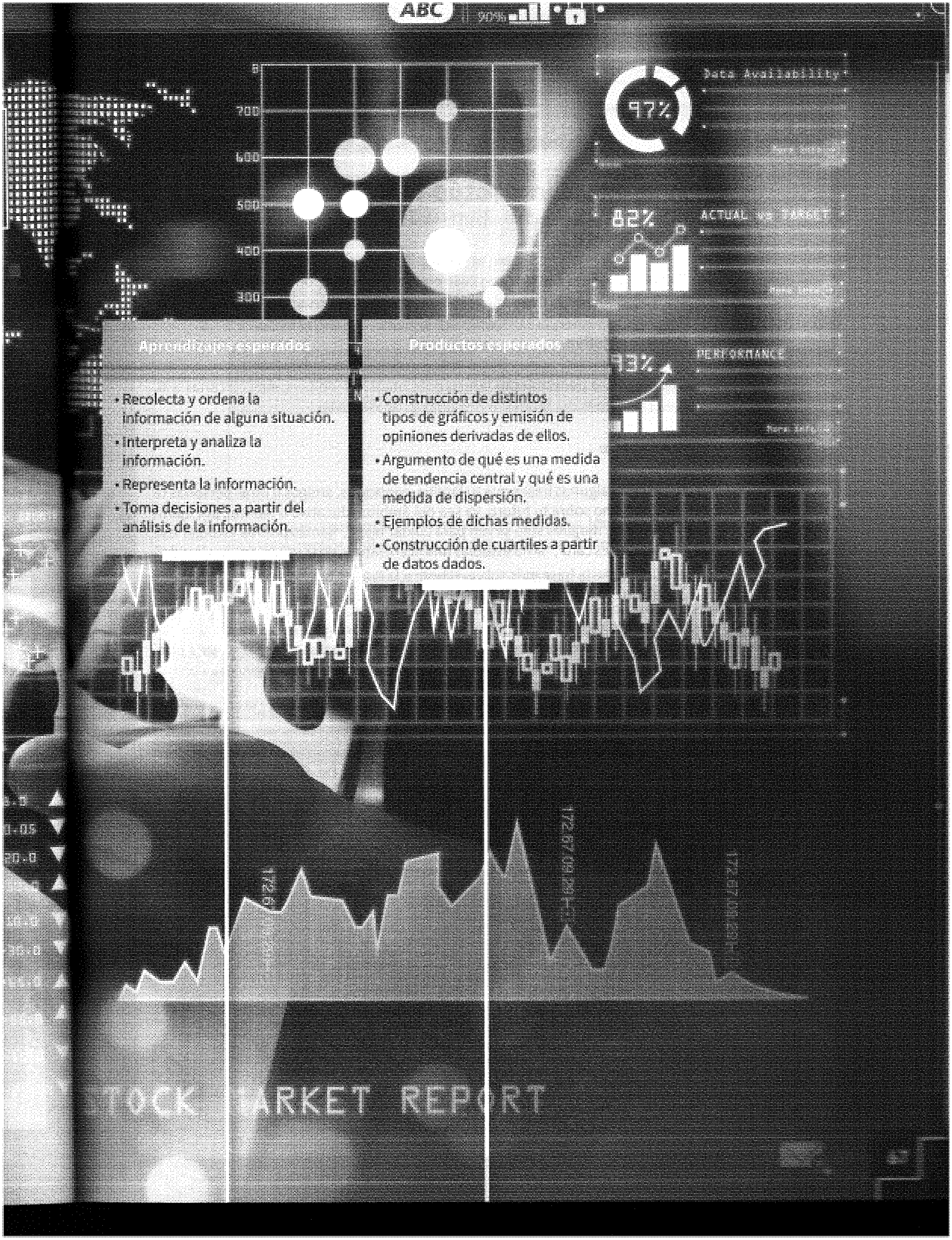
Rúbrica

Aprendizajes esperados	Básico	Autónomo	Estratégico
Usa un lenguaje propio para situaciones que necesiten del estudio con elementos de estadística y probabilidad.	Conoce los conceptos básicos de probabilidad y estadística.	Uso del lenguaje apropiado para la presentación de situaciones asociadas con la probabilidad y la estadística.	Mediante el uso del lenguaje de la probabilidad y la estadística resuelve situaciones de la vida cotidiana.
Usa técnicas de conteo o agrupación en la determinación de probabilidades.	Conoce alguna de las técnicas de conteo o agrupación.	Determina probabilidades haciendo uso de las técnicas de agrupación y de conteo.	Resuelve situaciones de la vida cotidiana mediante el uso de técnicas de conteo o agrupación para la determinación de sus probabilidades.
Organiza la información como parte de la estadística para el estudio de la probabilidad.	Reconoce la información que busca analizar.	Resuelve una situación dada mediante la organización de información recolectada.	Analiza situaciones de la vida cotidiana mediante la organización de la información recolectada.
Estudia el complemento que ofrece la estadística para la probabilidad.	Entiende las definiciones de probabilidad y de estadística.	Comprende la relación entre la probabilidad y la estadística.	Resuelve situaciones de la vida cotidiana mediante el análisis de la relación entre la probabilidad y la estadística.
Reconoce la diversidad de situaciones que precisan de la incertidumbre en el tratamiento del riesgo.	Conoce el concepto de incertidumbre y de riesgo.	Entiende situaciones en donde se involucra el concepto de riesgo.	Resuelve situaciones que precisan de la incertidumbre y el tratamiento del riesgo.
Modela con estadística y probabilidad el estudio de la información.	Reconoce la información que busca analizar.	Modela una situación dada mediante el uso de la probabilidad y la estadística.	Resuelve situaciones de la vida cotidiana mediante modelos relacionados con la probabilidad y la estadística.
Organiza la información recolectada de la situación estudiada.	Reconoce la información de la situación que busca analizar.	Analiza una situación dada mediante la organización de información recolectada.	Resuelve situaciones de la vida cotidiana mediante la organización de la información recolectada.
Construye fórmulas de probabilidad.	Conoce algunas fórmulas probabilísticas.	Hace uso de fórmulas de probabilidad.	Resuelve ejercicios de la vida cotidiana mediante el uso de fórmulas de probabilidad.

Segundo parcial

Eje: Del manejo de la información al pensamiento estocástico

Componentes	Contenidos centrales	Contenidos específicos
• Riesgo, inferencia y aleatoriedad: elementos de la estadística y la probabilidad.	• Manejo de la información en situaciones de la vida cotidiana.	• Estudio de la información. ¿Qué papel juegan las medidas de tendencia central?, ¿cómo representar la información en un gráfico estadístico?, ¿cómo estudiar un gráfico estadístico?, ¿qué papel juega la probabilidad en el manejo de la información? • Cálculo de las medidas de tendencia central y su representatividad en términos de la variabilidad y contexto situacional. • Construcción de gráficos estadísticos en la representación de la información. • Análisis de tipos de gráficos estadísticos.



Aprendizajes esperados

- Recolecta y ordena la información de alguna situación.
- Interpreta y analiza la información.
- Representa la información.
- Toma decisiones a partir del análisis de la información.

Productos esperados

- Construcción de distintos tipos de gráficos y emisión de opiniones derivadas de ellos.
- Argumento de qué es una medida de tendencia central y qué es una medida de dispersión.
- Ejemplos de dichas medidas.
- Construcción de cuartiles a partir de datos dados.

Analizo, elijo y actúo

Para reflexionar

En clase, ¿tiendes a actuar con base en comentarios de otras personas?, ¿conoces a alguien que se deje influenciar fácilmente por otros?

Nuestro objetivo

Desarrollar tu pensamiento crítico ante mensajes de personas cercanas a ti, para que tomes decisiones informadas y responsables.

Paso a paso:

1. Escribe algunas frases que tu familia, conocidos, amigos u otras personas te han dicho sobre tu futuro, ya sea del ámbito educativo o laboral; por ejemplo: "las mujeres sólo sirven para cuidar hijos", "jamás podrás estudiar eso porque es caro", "si no estudias una licenciatura, no la vas a hacer en la vida".
2. Comparte la frase más sobresaliente o la que no haya sido mencionada por algún compañero.
3. Cuando terminen de participar todos, reflexionen sobre las frases: siempre escucharán comentarios a su alrededor sobre sus elecciones y éstos pueden ser buenos o malos; sin embargo, lo importante es elegir, vivir con esa elección y asumir sus consecuencias.
4. En casa, piensa en una decisión que debas tomar. Escribe cuáles son las posibles alternativas, por ejemplo: trabajar y dejar de estudiar, estudiar hasta terminar la carrera, trabajar y estudiar. Una vez identificadas las alternativas, escribe los pros y contras de cada una. De esta forma tendrás más argumentos antes de elegir.

Para terminar...

Recuerda que cada decisión tiene consecuencias. Antes de actuar, reflexiona y analiza las alternativas.

Proyecto de vida

A lo largo del semestre desarrollaremos tu Proyecto de vida, con el propósito de aclarar qué deseas para ti; que puedas tomar decisiones que marquen la dirección de tu futuro, así como reflexionar acerca de las implicaciones que tendrá en tu vida el hecho de llevarlas a cabo.

Organiza la información de tu Proyecto de vida.

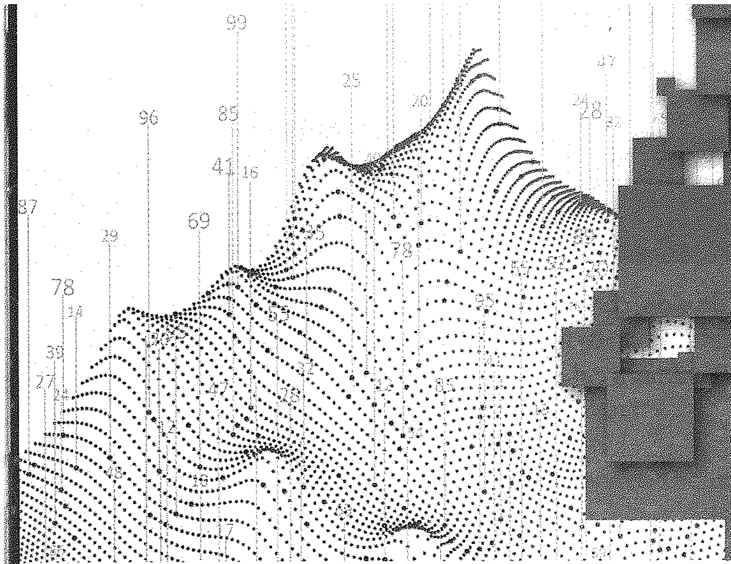
1. Copia y completa en tu cuaderno el cuadro.
2. Escribe en la primera columna tus metas más importantes. Agrega las filas que sean necesarias.

Meta	Ideas importantes a rescatar	Acciones que debo llevar a cabo	Beneficios que representa para mí	Beneficios que representa para mi entorno

3. Una vez que organizaste la información en el cuadro, léela con atención para tener un panorama de lo que implica el planteamiento de tu Proyecto de vida.
4. Escribe por qué es importante elaborar tu Proyecto de vida.
5. Diseña un organizador gráfico en el cual incorpores frases que te motiven a trabajar día con día en tu Proyecto de vida, incluye tus metas para que las tengas presentes todo el tiempo. Coloca este organizador en un lugar visible en tu casa. Utiliza el siguiente esquema para hacer un esbozo y enriquécelo paulatinamente.

Proyecto de vida de: ____________________

	Metas a corto plazo	Metas a mediano plazo	Metas a largo plazo
Personal			
Familiar			
Escolar			
Laboral			
Social			



Manejo de la información en situaciones de la vida cotidiana

Medidas de tendencia central

Así como puedes describir a las personas por su altura, peso, color de ojos, tipo de cabello y algunos otros adjetivos, a las variables cuantitativas (discretas y continuas) las puedes describir mediante los estadísticos descriptivos. Algunos de los más básicos ya los conociste (rango, marca de clase), otros estás por conocerlos (media, mediana, moda) y algunos más los conocerás en el siguiente parcial (desviación estándar y varianza). Estas características que tienen las variables cuantitativas, además de describirlas, te permiten hacer inferencias sobre ellas.

Las medidas de tendencia central son valores que resumen una variable y que de manera general intentan darte una idea de dónde está la mitad de los datos de una variable. En teoría, las medidas de tendencia central son un valor cercano al centro del rango, y alrededor de ese valor central está repartida toda la variable. Las medidas de tendencia central más utilizadas son la media y la mediana. El centro de una variable cuantitativa podría parecernos muy evidente, por ejemplo, si tu variable es un 1 y un 99 es evidente que la mitad está en el 50 ¿no? Pero, ¿qué pasaría si tienes la siguiente: 1, 99, 99, 99, 99, 99, 99, 99 y 99? O ¿cuál sería el centro de una variable como esta: 1, 10, 10, 10, 10, 10, 99, 99, 99, 99 y 99?

A continuación aprenderás las medidas de tendencia central más comunes, cómo calcularlas y cuando usarlas.

Media aritmética

Hay muchos tipos de medias que se utilizan en la estadística; la media aritmética, la ponderada y la geométrica son las más conocidas. Sin embargo, de estas tres la más utilizada es la media aritmética que seguramente conoces como "promedio". La expresión matemática de la media aritmética es la siguiente:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} a_i,$$

donde $\overline{X}$ significa media aritmética, n el número de valores en la variable a promediar, a_i es el i-ésimo valor de la variable y Σ se refiere a la sumatoria desde el primer valor (i = 1) hasta el n-ésimo valor de a.

En otras palabras, sumas todos los valores de la variable y divides esa sumatoria entre el número de valores que sumaste. Es importante que aunque conozcas los pasos mecánicos para calcular un promedio, aprendas a leer las expresiones matemáticas, pues probablemente más adelante en tu formación académica te enfrentarás con expresiones más difíciles de leer. Es un buen ejercicio habituarte a leer expresiones sencillas como esta para que luego batalles menos con las difíciles.

Ejemplo

En una preparatoria hay cinco salones de sexto semestre con la siguiente cantidad de alumnos en cada salón: 29, 37, 32, 35 y 33. ¿Cuál es la media de esa variable?

A continuación verás paso a paso la sustitución de los valores en la expresión matemática:

$$\bar{X} = \frac{1}{5}\sum_{i=1}^{5} a_i$$

$$\bar{X} = \frac{1}{5}(29+37+32+35+33)$$

$$\bar{X} = \frac{166}{5} = 33.2.$$

La media aritmética es 33.2 estudiantes.

Ejemplo

Alexander es un estudiante universitario que desea saber cuánto dinero gasta en transporte y aperitivos al día (entre semana) para poder administrar su beca exitosa y responsablemente. A continuación verás la cantidad de pesos que ha gastado las últimas cuatro semanas: 40, 47, 54, 40, 40, 57, 40, 40, 47, 64, 40, 47, 54, 65, 40, 47, 40, 54, 65 y 54.

Para calcular el promedio de su gasto diario debes sumar todos los valores y dividir la sumatoria entre el número de valores:

$$\frac{40+47+54+\ldots+54+65+54}{20} = \frac{975}{20} = 48.75.$$

Alexander necesita aproximadamente 50 pesos diarios (de lunes a viernes) para transporte y aperitivos, es decir, con 1 000 pesos puede transportarse y no morir de hambre por un mes.

Ahora sigue las siguientes instrucciones para aprender a usar la función en Excel que sirve para calcular el promedio de una variable. Usarás como ejemplo la variable "tamaño" de la base de datos de las flores que descargaste en la página 31.

1. En la hoja de cálculo donde está la base de datos, elige una celda en la que calcularás el promedio de la variable.
2. En la celda elegida escribe la siguiente fórmula: =PROMEDIO(C2:C49).

3. Dentro de los paréntesis debes poner la celda donde está el primer valor de tu variable, que en este caso es la celda C2, seguida de dos puntos (:) y luego la celda donde está el último valor de tu variable, que en este caso es C49.

4. Si tu Excel está en inglés la función en vez de "PROMEDIO" es "MEAN".

5. El resultado debe ser 63.37 cm, corrobóralo.

La mediana es también conocida como el valor medio de una lista de valores. Para calcularla, debes ordenar de manera ascendente o descendente todos los valores y tomar el valor que esté exactamente a la mitad de la lista. Si la variable tiene un número impar de valores, la mediana será el valor que divida en dos partes iguales a la variable (iguales en la cantidad de valores), pero si tiene un número par de valores, se toman los que estén más cercanos a la mitad y los promedias. En otras palabras, si tienes cinco valores, una vez ordenados de menor a mayor, el tercer valor será la mediana; si tienes 25 valores el treceavo es la mediana y si tienes 26 valores agarras el treceavo y el catorceavo (los más cercanos a la mitad) y los promedias (los sumas y los divides entre dos).

Ejemplo

Si ordenas de menor a mayor las 20 cantidades de dinero que gastó Alexander diariamente (ejemplo de la página anterior) obtienes lo siguiente: 40, 40, 40, 40, 40, 40, 40, 40, 47, 47, 47, 47, 54, 54, 54, 54, 57, 64, 65 y 65. Como se trata de 20 valores, los más cercanos a la mitad son el décimo y el onceavo, ambos son 47, si promedias estos dos números obtienes 47, así que la mediana de esta variable es 47. En este ejemplo, la media aritmética y la mediana resultan muy similares, 48.75 y 47, respectivamente.

Ahora determina en Excel la mediana de la variable "tamaño" de la base de datos de las flores. Utiliza las mismas instrucciones de la página anterior sobre cómo calcular el promedio en Excel pero cambia la función "PROMEDIO" por la función "MEDIANA", si tu Excel está en inglés la función es "MEDIAN". Compara la media y la mediana de esa variable. ¿Crees que la diferencia es importante? Coméntalo con tus compañeros.

¿Cuándo usar la media y cuándo la mediana?

Casi siempre utilizarás la media como medida de tendencia central. Sin embargo, en ocasiones las variables están fuertemente inclinadas hacia uno de los límites del rango, en estos casos, si utilizas la media probablemente estés sobreestimando los pocos valores altos que jalan la media hacia ellos. En estos casos la mediana es un mejor indicador del centro de la variable.

Ejemplo

La siguiente variable es el número de helados que diez niños comen a la semana: 1, 1, 0, 2, 1, 12, 1,1, 0 y 1. Si calculas ambas medidas obtendrás que la media de esa variable es dos helados, mientras que la mediana es un helado. Usar la media como medida de tendencia central estaría dando la falsa impresión de que esos diez niños comen dos helados a la semana cada uno. La realidad es que la mayoría come entre cero y dos helados: sólo un niño, Paquito, que tiene varicela y lo estuvieron sobrealimentando con helado, comió 12 helados; él "infla" el promedio hacia dos helados por niño. En este caso la mediana es un mejor indicador del centro de la variable.

¿Puedes pensar en alguna variable que tenga muchos valores bajos y muy pocos valores altos?

Moda

Es la menos usada de las medidas de tendencia central y es la más sencilla de asimilar. La moda es el valor más frecuente en una variable, es decir, el más repetido. Si hay un empate entre los valores más repetidos, todos los valores empatados son la moda, por lo tanto una variable puede tener más de una moda.

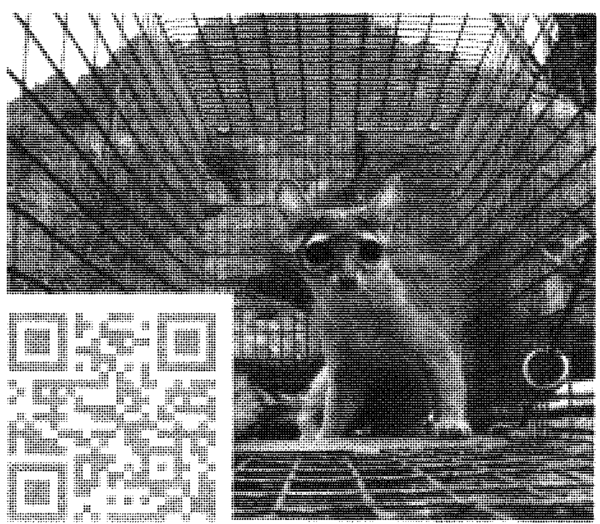
Ejemplo

Retomando los valores de los gastos diarios en transporte y aperitivos escolares de Alexander, que son: 40, 40, 40, 40, 40, 40, 40, 40, 47, 47, 47, 47, 54, 54, 54, 54, 57, 64, 65 y 65, el valor más frecuente es el 40, por lo tanto la moda es 40. En este caso, la diferencia de esta medida de tendencia central sí cambia bastante con respecto a la media o mediana. Si Alexander tomara la moda como indicador de lo que gasta al mes, creería que con 800 pesos sería suficiente, cuando la realidad es que 800 pesos no le alcanzarían.

Actividad de aprendizaje 1

En cada inciso, resuelve los problemas planteados.

1. Calcula el promedio de la cantidad de dinero que gastas diariamente en transporte y aperitivos escolares.
2. Utiliza Excel para calcular el promedio y la mediana de la variable "edad" de la base de datos que construiste a partir de la encuesta de la página 29.
3. Un biólogo está estudiando una población de cacomixtles (*Bassariscus astutus*) en el estado de Puebla. Hasta el momento lleva registrados 37 cacomixtles, a los cuales ha medido, sexado, pesado, despulgado y les ha tomado la temperatura corporal. La base de datos la puedes descargar escaneando el código QR de la derecha. Responde y efectúa lo siguiente:



   a. ¿Cuál es la población estadística?
   b. ¿Cuál es la muestra?
   c. ¿De qué tipo son las variables que han registrado?

d. Elige una variable categórica de la base de datos, construye una tabla de frecuencias de esa variable y elige un gráfico para representarla.

e. Calcula la media aritmética, mediana y moda de las cuatro variables cuantitativas que aparecen en la base.

f. Comenta tus dudas e inquietudes con tus compañeros de clase.

4. Considera que los once jugadores titulares de un equipo de fútbol tienen las estaturas que se indican: 1.68 m, 1.83 m, 1.75 m, 1.72 m, 1.65 m, 1.78 m, 1.74 m, 1.68 m, 1.79 m, 1.67 m y 1.81 m. Determina en tu cuaderno:

   a. La estatura promedio de los jugadores.

   b. Agrupa los jugadores de acuerdo con su estatura.

   c. Si ordenas las estaturas de manera ascendente, qué estatura es la que está a la mitad de todas.

5. Registra en una base de datos las calificaciones que hayas obtenido en los semestres cursados. Luego calcula las medidas de tendencia central (media, mediana y moda).

6. Los siguientes valores indican el saldo de 15 cajeros automáticos expresado en miles de pesos: 10, 12, 16, 8, 4, 12, 3, 14, 11, 7, 8, 12, 15, 16 y 4. Determina el valor de las medidas de tendencia central.

Actividad de aprendizaje 2

En cada inciso, resuelve los problemas planteados.

1. Retoma la base de datos que hiciste con tus compañeros encuestando 60 estudiantes (página 29). Elige tres variables cuantitativas y:

   a. Haz una tabla de frecuencias para cada variable.

   b. Calcula las tres medidas de tendencia central para cada variable.

2. Dos empresas de trenes tienen las siguientes cantidades de gastos e ingresos de los últimos doce meses, en miles de pesos.

Empresa	Gastos	Ingresos
A	7, 10, 14, 12, 11, 9, 9, 7, 10, 11, 9, 7	28, 48, 37, 34, 34, 22, 27, 26, 34, 20, 31, 13
B	8, 7, 6, 14, 9, 10, 13, 9, 7, 4, 9, 5	38, 34, 21, 30, 38, 30, 28, 32, 26, 27, 31, 29

Responde lo siguiente:

   a. ¿Cuántas variables hay en este ejercicio?

   b. ¿Consideraste la variable "mes"?

   c. El cuadro, ¿es una base de datos? ¿por qué?

d. Construye en Excel la base de datos que representaría correctamente esta información (Pista: debe tener cuatro columnas y 24 filas).

e. Utiliza las funciones PROMEDIO y MEDIANA de Excel para completar el siguiente cuadro:

	Gastos			Ingresos		
	Media	Mediana	Moda	Media	Mediana	Moda
Empresa A						
Empresa B						
Ambas						

f. ¿Qué medida de tendencia central consideras que representa mejor los gastos en ingresos mensuales de estas dos compañías?

g. ¿Qué relación hay entre las medidas de tendencia central de las compañías con respecto a las medidas de tendencia central de las dos juntas?

h. Guarda la base de datos, pues la necesitarás más adelante.

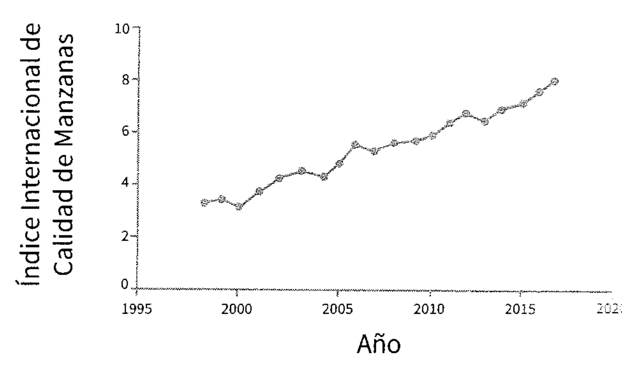
Construcción de gráficos estadísticos en la representación de la información

Se entiende por representación gráfica al acto de transformar la información contenida en tus bases de datos, en una forma visualmente entendible.

Imagina que eres un importante productor de manzanas y que tienes veinte años produciendo y mejorando su calidad nutricional. Tres veces al año tomas cien manzanas al azar de cada uno de los cuatro huertos que tienes y mides su calidad nutricional.

Si haces cuentas, en total mides la calidad nutricional de mil 200 manzanas al año. Además, has hecho esto durante los últimos veinte años; es decir, has valorado 24 mil manzanas.

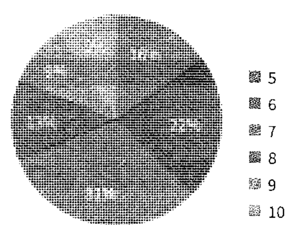
Ahora, imagínate que quieres convencer a tus compradores de que la calidad de tus manzanas ha mejorado en los últimos veinte años: ¿qué haces?, ¿les muestras una base de datos de 24 mil registros? o ¿les muestras un gráfico resumido, como el de la derecha?

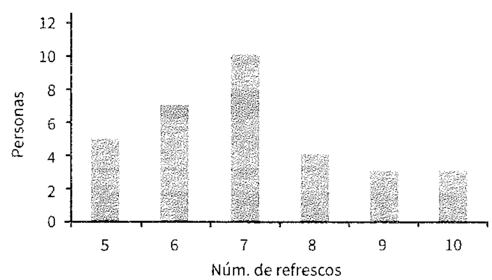


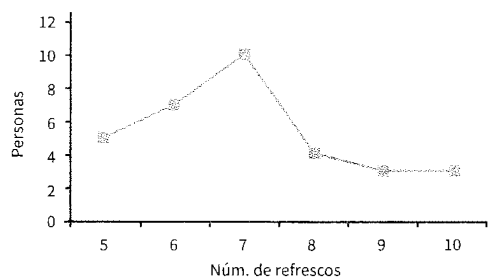
La principal cualidad de un gráfico es su poder de presentar varios datos de una forma resumida; inclusive para bases de datos pequeñas, son una excelente forma de visualizar datos y estadísticos para que la mayoría de las personas la comprendan.

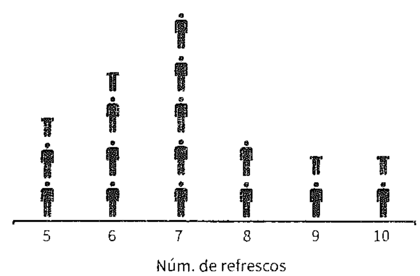
Diferentes tipos de gráficos se utilizan, dependiendo de la información que se quiera visualizar, de cuántas variables quieres mostrar simultáneamente, y del tipo de variables que vas a presentar (categórica, discreta o continua). En su mayoría, cuando uses un gráfico en estadística descriptiva, necesitarás una tabla de frecuencias.

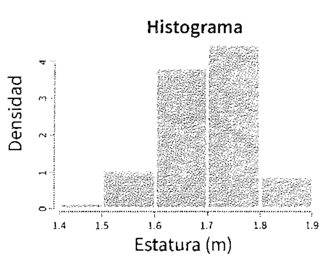
La elección de un tipo de gráfico depende de la pregunta y del mensaje que se quiere transmitir. La estadística y sus mecanismos son como cualquier otra herramienta, tienen un uso y un propósito, pero si el usuario la desconoce, la usará de manera equivocada. Piensa en la estadística como un molino: si le pones carne, tendrás carne molida; pero si le pones un tronco, la descompondrás. La estadística no funciona por sí sola, necesita un usuario que sepa lo que hace; por ejemplo, si tomas la tabla de frecuencias de los refrescos de la página 25, y dando unos cuantos teclazos en la computadora, obtienes cualquiera de los gráficos que se muestran:

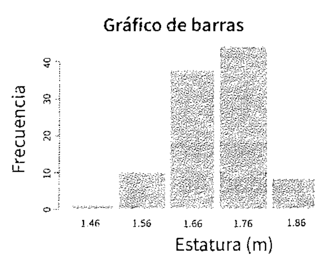












Sin embargo, no todos esos gráficos están bien utilizados, y otros son menos eficientes en transmitir determinado mensaje.

Observa y compara los dos gráficos de la izquierda, ¿en qué se parecen? ¿en qué se diferencian?

El histograma es un tipo de gráfico que resume cómo los datos de una variable cuantitativa están distribuidos a lo largo del rango. Se compone por rectángulos adyacentes cuya base se encuentra en el eje horizontal.

Los histogramas se construyen a partir de tablas de frecuencias de datos agrupados; es decir, el número de intervalos y su amplitud tienen un uso para construir el histograma, aunque en realidad es poco común trazarlos a mano, actualmente se hacen en algún programa estadístico.

El ancho de cada rectángulo que forma el histograma está dado por la amplitud del intervalo, y la altura es tal que el área de cada rectángulo es proporcional a la fracción de las observaciones que se encuentran en ese intervalo.

Lee el párrafo anterior las veces que sea necesario, hasta que lo asimiles.

Las unidades en el eje vertical son porcentaje por unidad del eje horizontal, de manera que si calcularas el área de cada rectángulo y la sumaras, tendrías un valor de uno o cien por ciento.

Por todo lo anterior, no es posible construir un histograma con una tabla de frecuencias de una variable categórica.

Nota que el histograma de la página anterior tiene cinco rectángulos; es decir, consta de cinco intervalos de clase y la amplitud de cada intervalo es 0.1.

La base, altura y área de cada rectángulo del histograma de la página anterior se pueden observar en la tabla de la derecha.

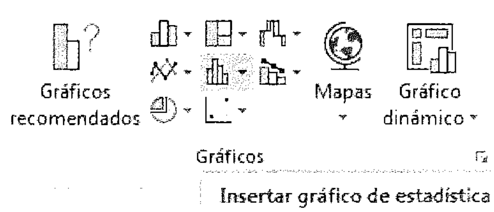
Analiza los datos y relaciona la representación gráfica con los datos de la tabla.

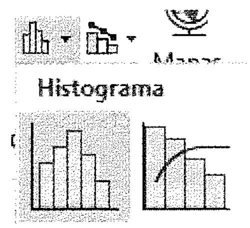
Base	Altura	Área
0.1	0.1	0.01
0.1	1	0.1
0.1	3.7	0.37
0.1	4.3	0.43
0.1	0.9	0.09
	SUMA	1

Ejemplo

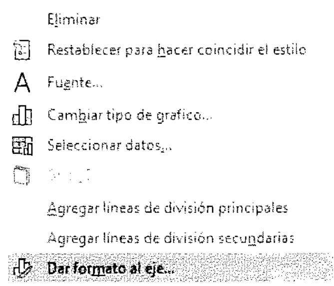
Utiliza la base de datos de los cacomixtles de la página 73, y sigue las instrucciones para que aprendas a generar histogramas en Excel.

1. Selecciona la variable longitud, seleccionando de la celda D1 a la celda D38.
2. En el menú "Insertar" encuentra el menú "Gráficos" y entra a la opción "Insertar gráfico de estadística" como el que se muestra a la derecha.



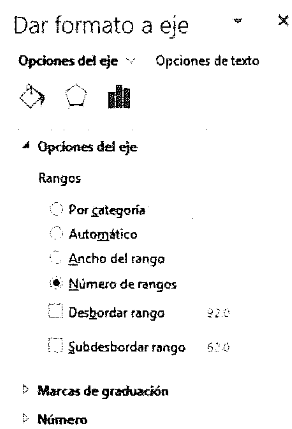


3. Ahora selecciona la opción “histograma” como se muestra en la imagen de la izquierda.



4. Automáticamente Excel hará el histograma con cuatro intervalos (con cuatro rectángulos).

Para modificar esto da clic derecho en el eje X, y elige la última opción del menú: “Dar formato al eje...”



5. En el menú que aparecerá a la derecha de la pantalla, elige “Opciones del eje”, dentro del submenú con el mismo nombre (Opciones del eje) cambia la selección de “Automático” a “Número de rangos” y cambia el cuatro por el número de intervalos que quieras (en este ejemplo usamos seis).

6. Explora las diferentes opciones para editar el histograma.

Gráfica de barras

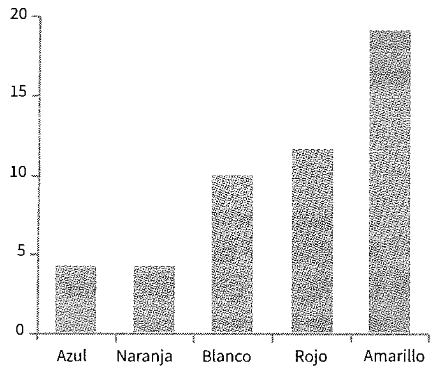
Un gráfico de barras presenta datos agrupados con barras rectangulares cuya altura es proporcional a los valores que representan. Las barras pueden ser verticales u horizontales. Los gráficos de barras son excelentes para datos cualitativos (variables categóricas). A pesar de que también se pueden utilizar para graficar variables cuantitativas (usando tablas de frecuencias), existen mejores opciones de gráficos para ese tipo de variables.

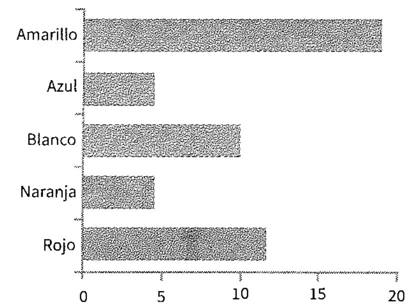
Los gráficos de barras son muy buenos para mostrar diferencias entre categorías o intervalos, porque puedes acomodar las barras de menor a mayor tamaño o viceversa.

Ejemplo

Los siguientes gráficos de barras están construidos a partir de los colores de las flores que el botánico de la página 31 registró.

En el gráfico de la izquierda, las barras son verticales y están acomodadas de menor a mayor, mientras que en el de la derecha son barras horizontales y las categorías están acomodadas en orden alfabético.

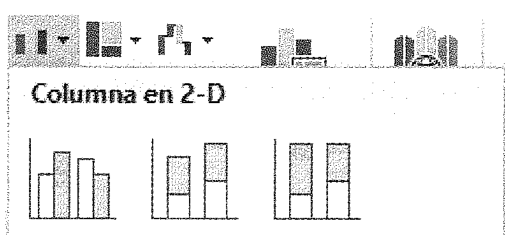




Sigue las instrucciones para que aprendas a elaborar gráficos de barras en Excel. Utiliza la tabla de frecuencias de la variable "forma" que hiciste en los ejercicios de la página 31:

1. Selecciona tu tabla de frecuencias en Excel como se muestra en la figura de la derecha.
2. En el menú "Insertar" encontrarás una gran variedad de gráficos disponibles. Los que dicen "Columna" y "Barra" se conocen como "Gráficos de barras". Selecciona la primera opción de "2-D".
3. ¡Listo! Ya tienes un gráfico de barras, mejora la presentación quitándole las líneas y las etiquetas innecesarias.
4. Explora las opciones para modificar el gráfico. Genera otras tres versiones del mismo cambiando la orientación y orden de las barras.

	A	B
1	Forma	Frecuencia
2	Circular	4
3	Ovalada	19
4	Triangular	10
5	Tubular	15



Diferencias entre histogramas y gráficos de barras:

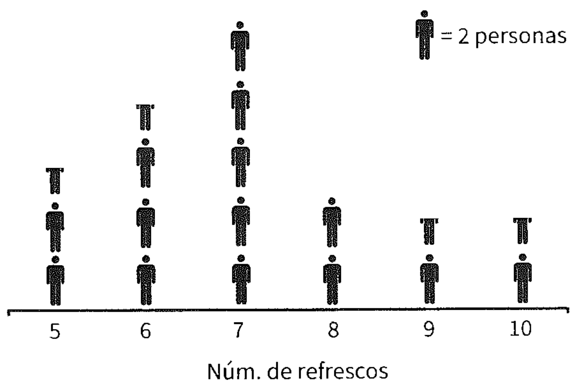
A pesar de que los histogramas y los gráficos de barras en ocasiones pueden parecer similares, son gráficos que tienen diferentes objetivos.

En palabras más concretas, el histograma sirve para ver la distribución de los datos de una variable a lo largo del rango, mientras que el gráfico de barras muestra cuál de tus categorías o intervalos es más grande y cuál más pequeño.

En los histogramas no puedes cambiar el orden de los rectángulos, pues el eje horizontal va del límite inferior del rango al superior, mientras que en el gráfico de barras puedes mover el orden de las barras a tu antojo.

Los histogramas no tienen espacios entre las barras, a menos que algún intervalo tenga una frecuencia de cero (casi no sucede), en cambio, los gráficos de barras por lo general tienen un espacio entre cada barra, pues cada una es independiente de la otra.

Por último, el eje vertical de un histograma no dice cuántos valores hay en cada rectángulo, mientras que en un gráfico de barras, el eje paralelo a éstas sí tiene esa función.



Un pictograma es una manera de representar tablas de frecuencias con figuras que indican las unidades de las frecuencias.

Es como un gráfico de barras, sólo que en vez de usar barras, utilizas una figura que apilas hasta llegar al tamaño de la barra. La figura o dibujo representa una cantidad que debe especificarse en la gráfica; si la frecuencia de una categoría no es múltiplo de la cantidad que la figura representa, se grafica sólo la parte proporcional.

En el ejemplo de los refrescos, cada silueta representa dos personas; si te fijas, la pila de los seis refrescos tiene 3.5 siluetas, lo cual se traduce como siete personas. Este tipo de gráfico es utilizado cuando el mensaje va dirigido a un público muy diverso, pues transmite información concreta de una manera muy sencilla.

Este tipo de gráfico se elabora con software relacionado al diseño gráfico, y no con estadístico. Básicamente representa lo mismo que un gráfico de barras pero con ilustraciones sustituyendo las barras.

Un gráfico de línea es, por lo general, utilizado para mostrar la relación entre dos variables cuantitativas. Es muy común que se utilice para mostrar el comportamiento de una variable en el tiempo, como el ejemplo de las manzanas en la página 75. Se utiliza una línea para transmitir la idea de que los puntos unidos por la línea llevan una secuencia ordenada.

Cuando la relación entre las dos variables no cumple con esta condición, el gráfico de línea no es el adecuado. Por esta razón, el gráfico de línea de los refrescos de la página 76 está mal empleado; es decir, no existe una secuencia entre las personas que se toman cinco refrescos a la semana y los que se toman seis, en este caso, el número de refrescos es una categoría y se pueden poner en el orden que sea.

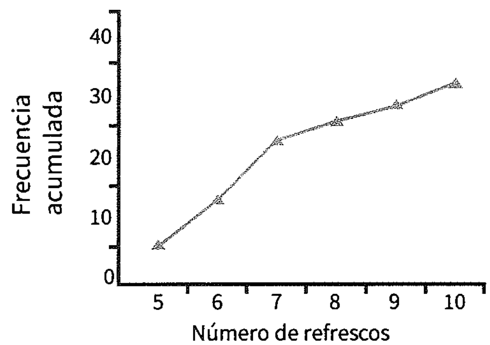
Existen varios tipos de gráficos de línea, el que se utiliza para graficar tablas de frecuencias es también llamado gráfico de ojiva, el cual muestra la frecuencia acumulada de cada categoría.

La frecuencia acumulada es la frecuencia de cierta categoría, más las frecuencias de las categorías anteriores; por esta razón nunca decrece y su último valor es el total de datos que tiene la variable.

Ejemplo

Retomando la tabla del número de refrescos por semana, el cálculo de la frecuencia acumulada y el gráfico quedarían como se muestra a continuación.

Núm. de refrescos	Frecuencia	Frecuencia acumulada
5	5	5
6	7	12
7	10	22
8	4	26
9	3	29
10	3	32



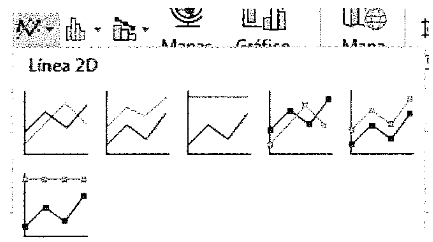
Sigue las instrucciones para elaborar un gráfico de línea en Excel.

1. El cuadro de la derecha muestra las ventas de una papelería de las últimas once semanas, cópialo en una hoja de Excel.

	A	B
1	Semana	Ventas
2	1	6270
3	2	5069
4	3	4822
5	4	5049
6	5	6133
7	6	5257
8	7	5798
9	8	6641
10	9	5104
11	10	4680
12	11	6538

2. Selecciona únicamente la columna "Ventas" en el menú "Insertar"; en la sección de gráficos selecciona "Insertar gráfico de líneas o áreas", como se muestra en la imagen.

3. En el apartado "2D" selecciona el gráfico "Línea" o "Línea con marcadores", dependiendo si quieres el gráfico sólo con la línea o con los puntos de cada dato.



4. Debe aparecerte un gráfico de línea con la variable ventas. Explora las opciones en Excel para editar y modificar el gráfico.

Gráfico de dispersión

El gráfico de dispersión permite explorar la relación entre dos variables cuantitativas.

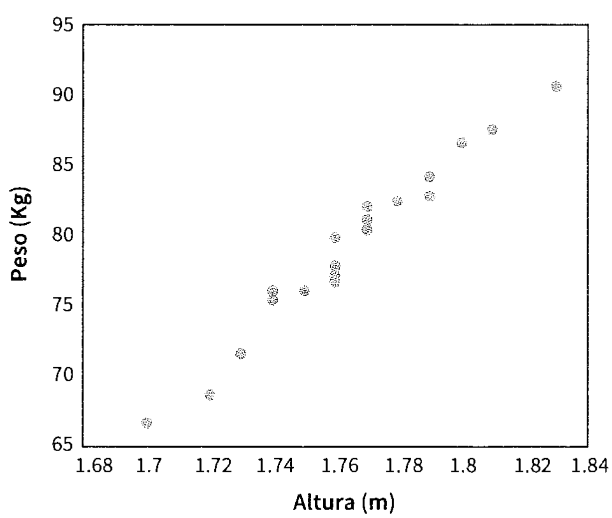
Se trata de un gráfico que se construye en un plano cartesiano, por lo general de dos ejes, donde cada eje representa una variable. De manera que cada dato queda graficado como un punto en el plano, al final tienes tantos puntos como datos.

Ejemplo

Las tablas muestran la altura en metros y el peso en kilogramos de veinte individuos tomados al azar de un pueblito en Chihuahua.

Individuo	1	2	3	4	5	6	7	8	9	10
Altura	1.7	1.72	1.73	1.74	1.74	1.74	1.75	1.76	1.76	1.76
Peso	66.7	68.7	71.7	75.5	76.1	76.1	76.1	76.8	77.2	77.9

Individuo	11	12	13	14	15	16	17	18	19	20
Altura	1.76	1.77	1.77	1.77	1.78	1.79	1.79	1.8	1.81	1.83
Peso	79.9	80.4	81.1	82.1	82.4	82.8	84.2	86.6	87.5	90.6

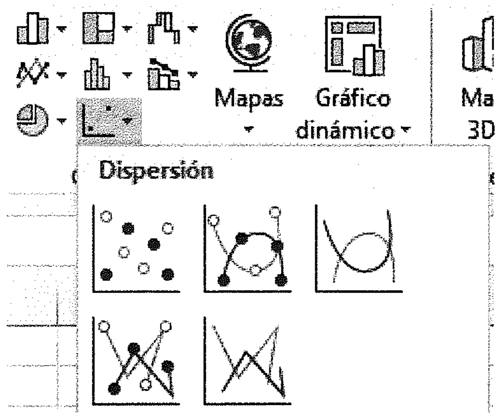


Con un gráfico de dispersión puedes explorar la relación entre las dos variables. Observa el gráfico de la izquierda. Muestra que, aparentemente, entre más alta es una persona, más pesada será. Ese tipo de relaciones son las que puedes explorar en un gráfico de dispersión.

Se llama gráfico de dispersión porque se observa la disgregación de los puntos en el gráfico, y esa disgregación te puede dar una idea sobre las variables: si están relacionadas y, en caso de estarlo, cómo es esa relación.

El gráfico de línea tiene un objetivo muy parecido: ver la relación entre dos variables cuantitativas. Recuerda que el gráfico de línea sólo se usa para transmitir la idea de que los puntos unidos por la línea llevan una secuencia ordenada. No tendría sentido dibujar una línea que uniera todos los puntos del gráfico de dispersión.

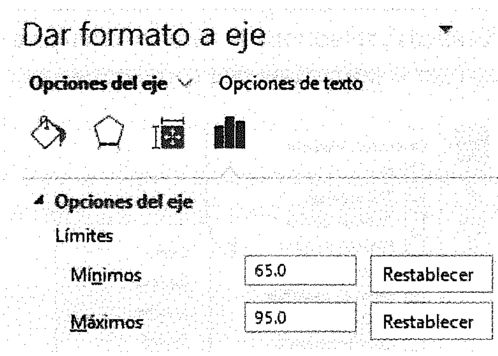
Como podrás darte cuenta, tanto el gráfico de línea como el de dispersión los puedes hacer fácilmente a mano, pues se trata sólo de un sistema de coordenadas en el que una variable va sobre el eje X, y la otra en el eje Y.



Sin embargo, es importante que sepas usar las herramientas digitales más elementales para construir estos gráficos. Para elaborar este gráfico en Excel debes seleccionar las dos variables que quieres relacionar.

Luego, en el menú "Insertar", en el área "Gráficos", seleccionar el tipo de gráfico "Dispersión".

En ocasiones, Excel construye el gráfico usando el cero como límite inferior de alguna de las dos variables. Esto provoca que los puntos se "amontonen" y quede un vacío en el gráfico. Para modificar los límites de los ejes da clic derecho en el eje que quieras modificar, luego selecciona "Dar formato al eje".



Por último, en las opciones cambia el cero en "Límites mínimos" por un valor cercano al valor mínimo de la variable. En el ejemplo de la imagen cambiamos el cero por 65, pues el valor mínimo de la variable "Peso" es 66.7.

Gráfica de pastel

Es un gráfico circular dividido en "rebanadas", donde cada rebanada representa una categoría, y el tamaño de la rebanada es proporcional a la frecuencia que tiene esa categoría.

Se recomienda que se use para representar diez o menos categorías, pues más de diez resulta poco ilustrativo, si tienes más de diez categorías es mejor que uses gráficos de barras.

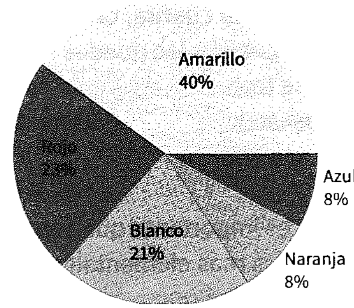
El gráfico de pastel por lo general se usa para mostrar porcentajes o la predominancia de ciertas variables, y es común que en el gráfico se escriba el porcentaje que representa cada rebanada. También, es común que las categorías se ordenen de mayor a menor, de manera que las rebanadas vayan reduciéndose y no estén intercaladas las grandes con las pequeñas.

Quien ve el gráfico puede ver claramente qué categoría domina sobre las demás; sin embargo, puede ser complicado distinguir las categorías que ocupan un porcentaje pequeño. Aunque esto se soluciona etiquetando las rebanadas, el gráfico puede verse mal si tiene exceso de estas etiquetas.

Ejemplo

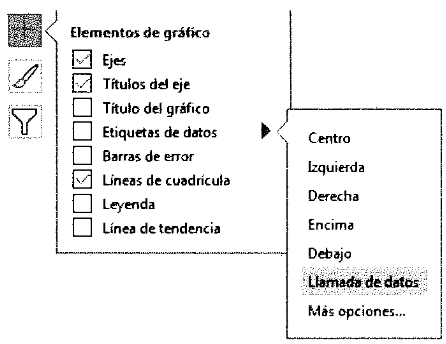
A continuación, puedes ver una tabla de frecuencias de la variable "color" en la base de datos de las flores, y su respectivo gráfico de pastel.

Color	Frecuencia
Amarillo	19
Rojo	11
Blanco	10
Azul	4
Naranja	4



El gráfico de barras cumple con un objetivo muy parecido. La principal diferencia es que en el gráfico de pastel puedes ver cuantitativamente la proporción de cada categoría y así concluir declaraciones como: de la comunidad de flores estudiada, cuarenta de cada cien son amarillas.

Para construir este gráfico en Excel, selecciona la tabla y luego en el menú "Insertar" submenú "Gráficos", selecciona el "Gráfico circular o de anillos". Hay diferentes maneras en las que puedes agregar el porcentaje que corresponde a cada rebanada, a continuación una de las formas:



1. Selecciona el gráfico y da clic en el signo de "más" que aparece en la esquina superior derecha.
2. Da clic en el triángulo que aparece al poner el cursor sobre la opción "Etiquetas de datos"
3. Por último, elige la opción "Llamada de datos".

Explora las múltiples formas en las que puedes editar el gráfico en Excel.

Actividad de aprendizaje 3

En cada inciso, resuelve los problemas planteados.

1. Haz un histograma de siete intervalos con la variable "Estatura" de la base de datos de Alonso Quijano (página 17).
2. Genera un gráfico de barras utilizando la variable "Forma" de la base de datos de las flores (página 31), acomoda las barras desde la más pequeña hasta la más grande (recuerda que necesitarás una tabla de frecuencias de la variable).
3. Crea un pictograma con la tabla de frecuencias que obtuviste de la variable "color", de la base de datos de las flores que descargaste de la página 31.
4. Elabora un gráfico de ojivas utilizando la variable "especie" de la base de datos que obtuviste de la página 16. ¿Es necesario ordenar de alguna manera en particular las categorías en esa variable?

5. Las tablas de abajo muestran el ancho, largo y área de 25 potreros en Veracruz. Traza tres gráficos de dispersión usando la tabla:

Ancho	58	61	62	64	65	67	68	69	73	75	77	82	83
Largo	180	162	151	142	141	141	139	138	135	133	133	132	132
Área	10 440	9 882	9 362	9 088	9 165	9 447	9 452	9 522	9 855	9 975	10 241	10 824	10 956

Ancho	86	86	88	88	88	89	91	92	94	96	100	122
Largo	131	131	130	130	128	127	124	123	122	114	106	99
Área	11 266	11 266	11 440	11 440	11 264	11 303	11 284	11 316	11 468	10 944	10 600	12 078

6. Haz un gráfico de pastel utilizando la variable "Antesis" de la base de datos de las flores. Necesitarás una tabla de frecuencias, recuerda acomodar las categorías de la más a la menos frecuente.

Actividad de aprendizaje 4

En cada inciso, resuelve los problemas planteados.

1. Completa el siguiente cuadro poniendo un cero donde el gráfico no sirva para esa variable, un cinco para cuando sirve pero hay mejores opciones, y un diez para cuando es la mejor opción o una de las mejores.

	Variable		
	Categórica	Discreta	Continua
Histograma			
Gráfico de barras			
Pictograma			
Gráfico de línea			
Gráfico de pastel			

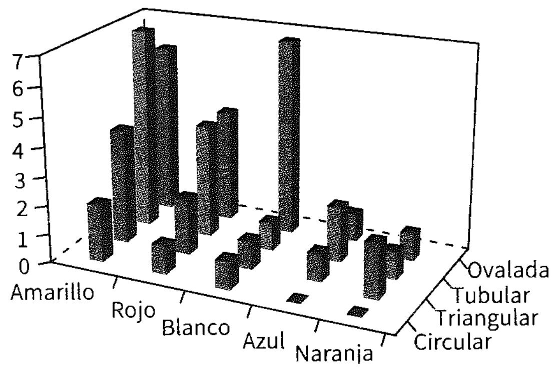
2. Retoma la base de datos de la encuesta de la página 29 y elige cinco variables al azar, asegúrate de tener mínimo dos variables cuantitativas.

   a. Para cada variable, haz la tabla de frecuencias correspondiente (sencilla o de datos agrupados) según sea necesario. Si la variable tiene sólo una categoría, no tendrá sentido hacer una tabla de frecuencias, si sucede esto con alguna variable, cámbiala por una que tenga más de una categoría.

b. Para cada tabla generada, elige y elabora el tipo de gráfico que mejor represente la información. Justifica tu elección.

c. Considerando que tus compañeros no necesariamente tienen la misma base de datos que tú, pero que la muestra viene de la misma población ¿crees que sus gráficos deberían coincidir con los tuyos?

d. Busca en el grupo alguien que haya graficado la misma variable y compara. Responde en tu cuaderno las siguientes preguntas: ¿en qué se parecen?, ¿en qué no se parecen? y ¿por qué?

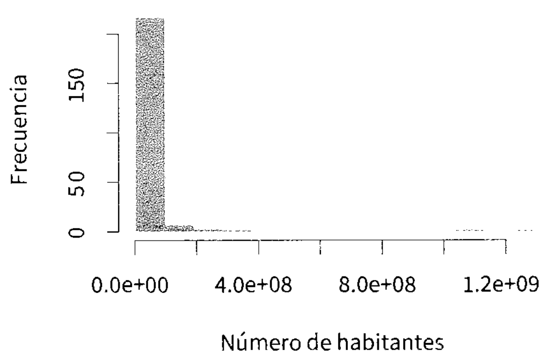


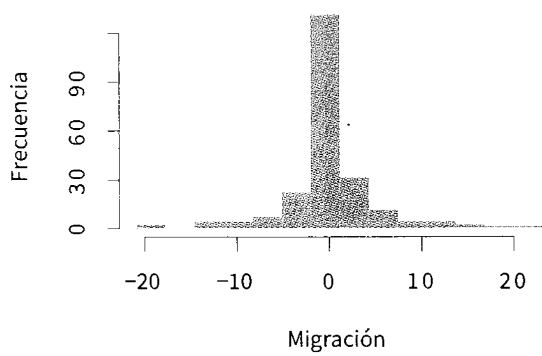
3. Observa la gráfica de la izquierda. Se trata de un gráfico de barras en 3D producto de una tabla de frecuencias combinando dos variables categóricas (color y forma) de la base de datos de las flores. Usa la herramienta tablas dinámicas y la base de datos de las flores para construir la tabla que dio origen al gráfico de arriba, y busca en Excel la forma de recrearlo.

Análisis de tipos de gráficos estadísticos

De los gráficos vistos antes, vale la pena ahondar en los histogramas y los gráficos de dispersión.

Como ya leíste, ambos gráficos son útiles para explorar variables y la relación entre ellas. Observa algunos detalles en los siguientes histogramas.





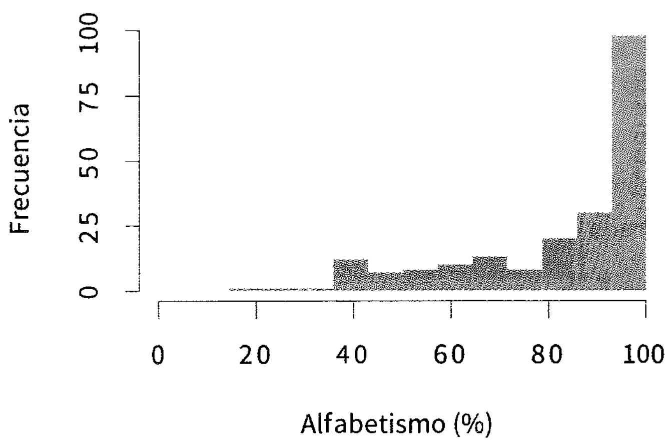
Estos histogramas están hechos con una base de datos que tiene información de 227 países, entre 1995 y 2006.

El primer histograma muestra la distribución de la variable "población": el número de personas que viven en cada país; el segundo, de la variable "migración": el número de personas que se va o llega al país, por cada mil habitantes.

El tercero, que se observa a la derecha, muestra la distribución de la variable "alfabetismo", que expone el porcentaje de alfabetización en el país.

¿Qué puedes concluir al verlos?, ¿para qué sirven? Los histogramas muestran hacia qué valores se encuentran la mayoría de los datos.

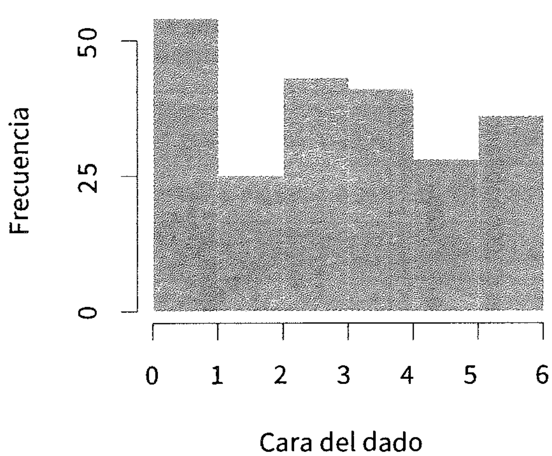
Por ejemplo, el primer histograma tiene muchos valores bajos y pocos valores altos, por eso se dice que la distribución de esa variable está sesgada a la izquierda. Puedes concluir que la mayoría de los países tienen menos de 200 millones de habitantes, y muy pocos (sólo dos) tienen más de 400 millones de habitantes.



En el segundo histograma puedes ver que la mayoría de los datos está cercana al centro de la variable, con pocos valores por debajo del centro, y pocos por encima del centro; esta distribución es simétrica y tiende a los valores centrales. Puedes concluir que la gran mayoría de los países tiene poca migración e inmigración, y pocos son los países que tienen migraciones e inmigraciones muy altas (de veinte habitantes por cada mil).

En el tercer histograma, puedes ver la contraparte del primero, la mayoría de los datos son valores altos y pocos son valores bajos, se dice que la variable está sesgada a la derecha. Puedes concluir que la gran mayoría de los países tiene altos índices de alfabetización, y pocos tienen bajos índices de alfabetización.

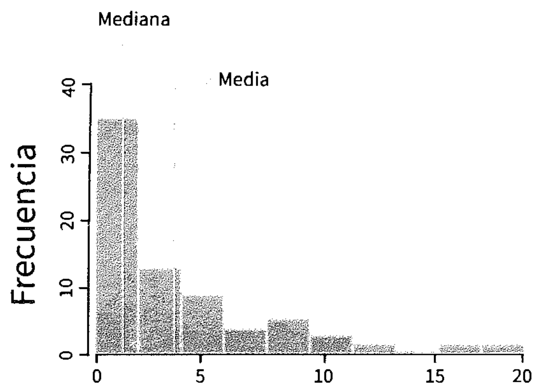
Hay ocasiones en las cuales una variable no tiende hacia ningún valor; por ejemplo, imagina que tiras un dado de seis caras 227 veces, el histograma que obtendrías de esa variable sería uno parecido al de la derecha; a pesar de que el uno fue el más frecuente, la variable no tiende a valores bajos, ni a altos, ni a los centrales.



La distribución de esa variable es simétrica y uniforme. Puedes concluir que todas las caras aparecieron más o menos con la misma frecuencia.

La distribución de las variables determina la medida de tendencia central más adecuada para cada tipo de variable.

El histograma de abajo muestra por qué a veces es necesario usar la mediana como medida de tendencia central. En una distribución sesgada, como en el gráfico, la media es "jalada" por los pocos valores altos de la variable, pero la mediana se mantiene en el "centro" de la variable.

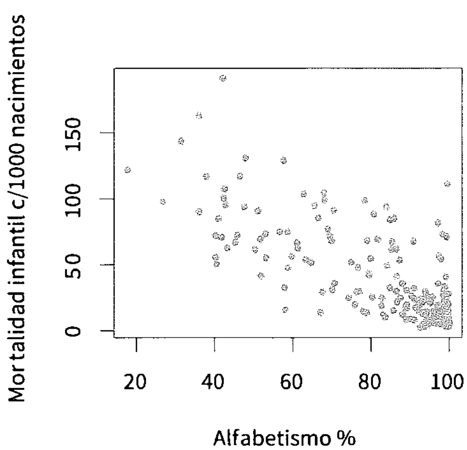


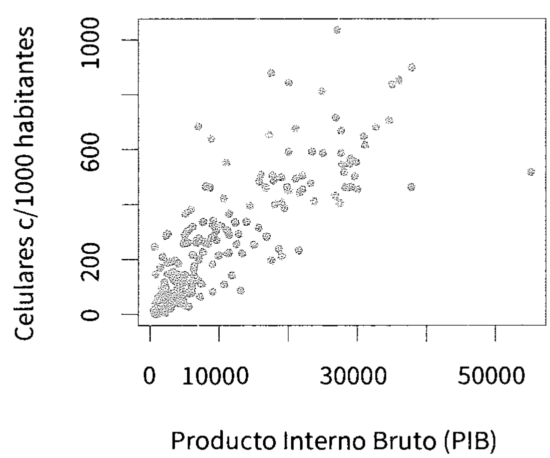
En la variable "población" usada para el histograma de la página 86, la media aritmética es de 28 740 284 habitantes, y la mediana es 4 786 994 habitantes (seis veces más pequeña).

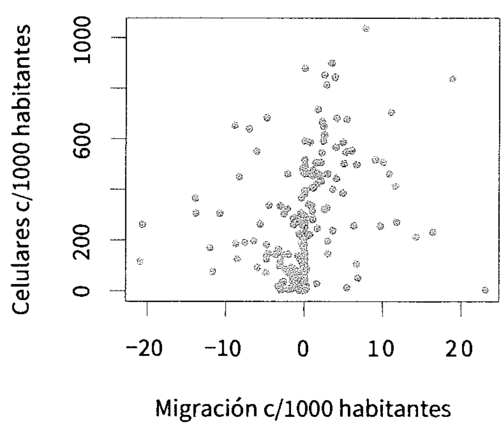
De los 227 países que forman la base de datos, sólo 38 tienen más habitantes que la media y 189 menos que la media. En cambio, 113 tienen más habitantes que la mediana y 113 menos que la mediana. En estos casos, la mediana es más adecuada en determinar el centro de la variable.

Ejemplo

Observa los siguientes gráficos de dispersión de la base de datos de los 227 países:







En el primer gráfico se muestra la relación entre las variables "alfabetismo" y "mortalidad infantil", en el segundo la relación entre las variables "Producto Interno Bruto (PIB)" y "celulares por cada mil habitantes", y en el último la relación entre las variables "migración" y "celulares por cada mil habitantes".

¿Qué puedes concluir de estos gráficos de dispersión?

Primero, te debe quedar claro que cada punto en el gráfico es uno de los 227 países que conforman la base de datos. En el primer gráfico se ve cómo, a medida que el valor de la variable en el eje X aumenta, la variable del eje Y disminuye, se dice entonces que la relación entre estas dos variables es negativa. Específicamente en este caso, significa que los países más alfabetizados tienen menos mortalidad infantil, y los menos alfabetizados, mayor mortalidad infantil.

En el segundo gráfico es lo opuesto, se observa cómo, a medida que los valores en el eje X aumentan, también lo hacen los valores en el eje Y, se dice que la relación entre estas dos variables es positiva. En este caso específico, significa que los países con mayor PIB tienen más celulares por cada mil habitantes.

Por último, en el tercer gráfico no se ve una aparente relación entre las dos variables graficadas, puedes pensar que las variables no tienen ninguna relación. En el ejemplo parece que la migración no está relacionada con la cantidad de celulares por cada mil habitantes.

En el próximo parcial aprenderás a medir la relación entre las variables, pues el ojo humano puede equivocarse cuando cree ver una relación positiva o negativa entre dos variables o cuando cree no verla.

Muertes dolosas en México

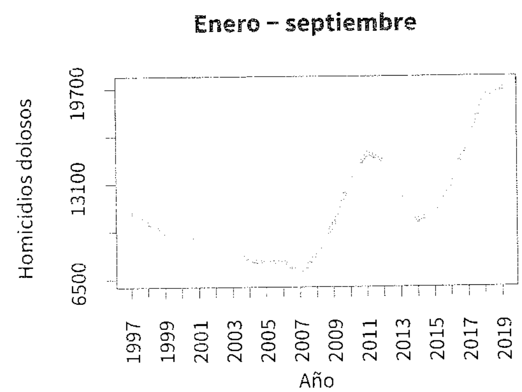
En abril de 2019 se publicó la noticia de que "éste fue el primer trimestre más sangriento en la historia moderna de México" en varios medios de comunicación y en meses subsecuentes se ha estado comentando lo mismo para todo el año.

En el código QR que se muestra a la derecha puedes entrar a la página bkmrt.com/nKeAca, donde puedes revisar las publicaciones que contienen información por entidad federativa sobre los diferentes delitos que ocurren en el país desde 1997.

En la siguiente tabla puedes ver un fragmento de esa base de datos, que contiene la información sobre los homicidios dolosos en el país de enero a septiembre de cada año, y la diferencia en homicidios con respecto al año anterior, desde 1997 al 2019.

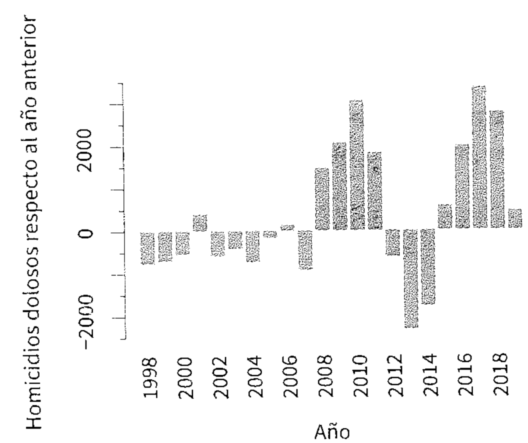
Año	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Dolosos	11 169	10 406	9 703	9 555	8 956	8 534	7 797	7 634	6 835	8 297	10 357	8 297
Diferencia		−763	−703	−546	398	−599	−422	−737	−163	135	−934	1 462

Año	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Dolosos	10 357	13 411	15 237	14 616	12 293	10 516	11 093	13 076	16 425	19 191	19 648
Diferencia	2 060	3 054	1 826	−621	−2 323	−1 777	577	1 983	3 349	2 766	457



Si ves la tabla, puedes darte una idea de que el número de homicidios dolosos ha aumentado desde 1997, pero los gráficos nos dan una idea más clara de cómo ha sido ese aumento, puedes usar un gráfico de línea como el que se muestra a la izquierda.

Efectivamente, puedes ver que ningún año antes del 2019 ha tenido más homicidios dolosos, al menos hasta lo que va del 2019 (hasta septiembre).



Ahora utilicemos un gráfico de barras para visualizar la diferencia de homicidios con respecto al año anterior. En este gráfico puedes ver que, a pesar de que en 2019 han muerto más personas que en los últimos 22 años, el aumento con respecto a los cuatro años anteriores es menor.

La estadística, como parte de las matemáticas, no sólo es importante para que la utilices como profesionista en tu día a día. Es importante que sepas estadística para que formes tu propio criterio en situaciones en que la información se transmite con intereses.

En este ejemplo, los que quieran desprestigiar la administración presidencial actual, dirán que lo que ha hecho, no ha servido para nada, pues habrá más muertos; descartando que el presidente recibió una administración con un alza de homicidios dolosos altísima, y que entre 2017 y 2018 también se rompió el mismo récord de homicidios desde 1997.

Los que quieran favorecer a la actual administración presidencial, dirán que el aumento del número de homicidios es cada mes menor, y que es el crecimiento más pequeño en los últimos diez años (pues entre 2012 a 2014 prácticamente se podría considerar como disminución y estabilización), ignorando que es posible un cambio radical, de un crecimiento a un decrecimiento, como el que ocurrió entre 2011 y 2012.

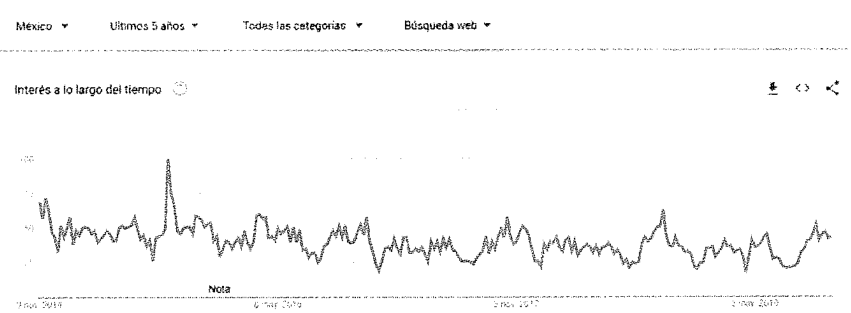
¿Qué conclusiones puedes sacar de esos dos gráficos? ¿Es válido decir que la actual administración es responsable del año más sangriento, de los últimos 23 años en México? ¿Es válido decir que las actuales medidas son las más eficientes? ¿En estos dos gráficos está la información suficiente para concluir algo inapelable?

Actividad de aprendizaje 5

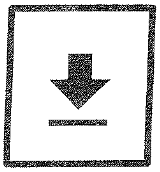
En cada inciso, efectúa lo que se pide.

1. Entrevista a cuarenta personas sobre el color de su fruta favorita y, previamente, a cada persona entrevistada solicítale su permiso para medir la circunferencia de su cráneo (hasta milímetros), también mide su altura y captura las tres variables en Excel.
   a. Utiliza tablas dinámicas en Excel para construir una tabla de frecuencias con el color de frutas, y construye dos gráficos que mejor representen la variable.
   b. Explora la distribución de las variables: "circunferencia del cráneo" y "altura" creando un histograma para cada una. ¿Qué distribución tiene cada una (sesgada a la izquierda, a la derecha, simétrica uniforme o simétrica con tendencia al centro)?
   c. Explora la relación entre las dos variables: "circunferencia del cráneo" y "altura", creando un gráfico de dispersión entre ellas. ¿Tienen alguna relación? Si la tienen, ¿es negativa o positiva?, ¿qué puedes concluir al ver el gráfico?

2. Google tiene una herramienta llamada "Google Tendencias", que permite conocer la popularidad de búsqueda que ha tenido cualquier tema o palabra buscada en Google. Sigue las siguientes instrucciones:
   a. Ingresa a "Google tendencias" buscándola en Google, con el enlace bkmrt.com/bployF o por medio del código QR de la derecha.
   b. Una vez en "Google tendencias" busca un tema de tu interés: político, histórico, algún actor, youtuber, videojuego, atleta, deporte...
   c. Ajusta las opciones de búsqueda para México en los últimos cinco años, como se muestra en la imagen:



   d. Dependiendo del tema elegido, el gráfico será diferente, con un pico de popularidad de búsqueda con un valor de cien. Si pones el cursor sobre el pico, puedes ver la fecha en la que ocurrió ese pico de popularidad, en este ejemplo, el pico se encuentra en febrero de 2019. Busca en internet el tema con la fecha y el pico de popularidad ¿puedes determinar a qué se debió ese pico de popularidad?



e. En dicha página, descarga la base de datos dando clic en el ícono de descarga en la esquina superior derecha del gráfico (como el que se muestra aquí).

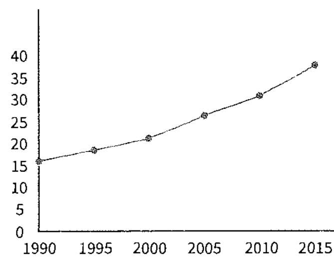
f. Se descargará un archivo .csv que puedes abrir en Excel. Una vez hayas abierto la base de datos en Excel, haz una copia del gráfico utilizando las herramientas para hacer gráficos en Excel.

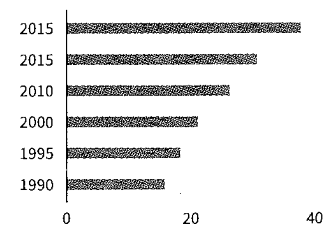
g. Debajo del gráfico de línea en "Google Tendencias", verás un mapa de México con la popularidad del tema que buscaste en cada estado de la república. Descarga también esa base de datos y construye en Excel un gráfico de barras, donde cada barra represente a un estado.

3. En México, el organismo encargado de almacenar, manejar y estudiar la información del país se llama Instituto Nacional de Estadística y Geografía (INEGI). A continuación, hay dos tablas con información sobre el índice de envejecimiento (número de personas de la tercera edad por cada cien niños y jóvenes) y sobre el porcentaje de mexicanos que hablan alguna lengua indígena. Para cada tabla, selecciona el gráfico que mejor represente la información a lo largo del tiempo, y responde las preguntas:

Índice de envejecimiento

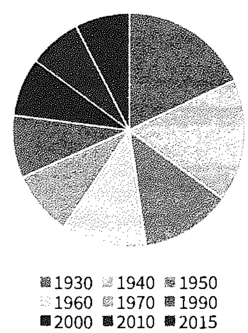
Periodo	Personas
1990	16
1995	18.5
2000	21.3
2005	26.4
2010	30.9
2015	38

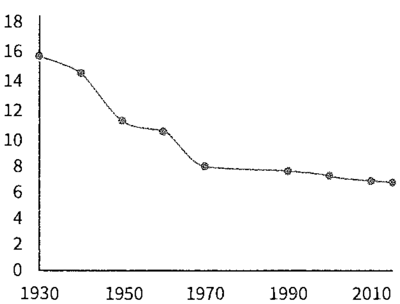




Hablantes de lenguas indigenas

Periodo	Personas
1930	16
1940	14.8
1950	11.2
1960	10.4
1970	7.8
1990	7.5
2000	7.1
2010	67
2015	6.6





a. ¿Cuáles de los siguientes enunciados son correctos, según las tablas y gráficas?

- En la actualidad, hay más personas de la tercera edad que antes.
- En 2015 había más del doble de personas de la tercera edad, por cada cien niños y jóvenes, que en 1990.

- En el 2000 había 21.3 personas de la tercera edad por cada cien niños y jóvenes en México.
- Según el INEGI, en 2015 había 32 863 142 niños y jóvenes en México; si había 38 adultos mayores por cada cien niños y jóvenes, entonces había aproximadamente 864 819 adultos mayores.
- Según el INEGI, en 2015 había 32 863 142 niños y jóvenes en México; si había 38 adultos mayores por cada cien niños y jóvenes, entonces había aproximadamente 12 487 994 adultos mayores.
- En 1930, casi una cuarta parte de la población en México, hablaba alguna lengua indígena.
- Los mexicanos que hablan lenguas indígenas están aumentando en el país.
- En los últimos cincuenta años ha decrecido el porcentaje de mexicanos que hablan alguna lengua indígena.
- El periodo entre 1960 y 1970 fue cuando más decreció el porcentaje de mexicanos que hablaban alguna lengua indígena.
- Desde 1930 no ha crecido el porcentaje de mexicanos que hablan alguna lengua indígena.

b. Marca la opción correcta:

- El índice de envejecimiento ha crecido porque ancianos de otros países migraron a México.
- El índice de envejecimiento ha crecido porque ahora las personas tienen menos hijos.
- El índice de envejecimiento ha crecido porque ha aumentado la mortalidad infantil/juvenil por enfermedades y narcotráfico.
- No puedes saber la causa con la información en la tabla.

c. Para más información del INEGI sobre la población mexicana, entra al enlace bkmrt.com/WwKxhS, escaneando el código QR.

Actividad de aprendizaje 6

En cada inciso, efectúa lo que se pide.

1. Del código QR que aparece abajo, descarga una base de datos con todos los partidos de futbol soccer que México ha jugado desde 1923. Utiliza tablas dinámicas para resolver los siguientes problemas y preguntas:

   a. ¿Cuántos partidos han sido ganados, perdidos y empatados? Construye un gráfico de pastel con esos tres valores y marca si los siguientes enunciados son falsos o verdaderos:

   - Más de la mitad de los partidos han sido ganados.
   - Más de una cuarta parte de los juegos han sido empates.

h. Explora la distribución de la variable “goles anotados” por México, haciendo un histograma donde tengas un rectángulo por número de goles (un rectángulo para el cero, uno para un gol, otro para dos goles y así hasta diez goles). ¿Cómo es su distribución: sesgada hacia la izquierda, sesgada a la derecha, simétrica uniforme o simétrica con tendencia al centro?

2. El ciclo de vida de las mariposas empieza en un huevo del cual sale una oruga que crece para luego encapsularse en una pupa, de la cual saldrá una mariposa adulta que pondrá huevos. Una bióloga quiere saber qué variables afectan la duración de la fase de pupa, desde que se convierte en pupa hasta que sale la mariposa. Tiene información de cien orugas de la misma especie, desde que puparon, hasta que salió la mariposa. De cada una registró: los días que duró la pupa, el tamaño de la oruga en milímetros antes de pupar, y la temperatura promedio a la que cada pupa estuvo durante su desarrollo. Puedes descargar la base de datos escaneando el código QR de la izquierda para las siguientes actividades:

a. Haz un histograma de cada una de las tres variables y completa el cuadro:

Variable	¿Qué distribución tiene?	¿Qué puedes concluir de su distribución?
Tamaño (mm)		
Temperatura (°C)		
Días		

b. Al ver la distribución de las variables, ¿puedes saber qué variable (tamaño de la pupa o temperatura) tiene efecto en la duración de la fase de pupa (días)?

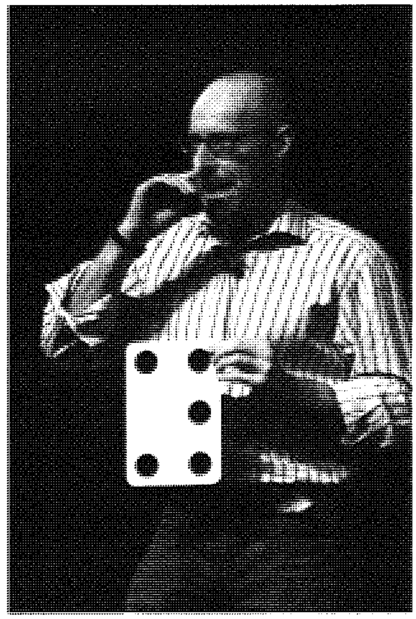
c. Haz los diagramas de dispersión (días/tamaño, días/temperatura) y completa el cuadro.

Variables relacionadas	Relación (positiva, negativa o sin relación)	¿Qué concluyes de la relación entre estas dos variables?
Días vs Tamaño		
Días vs Temperatura		

d. ¿Qué variable tiene efecto en la duración de la pupa?

e. Haz un gráfico de dispersión entre el tamaño y la temperatura. ¿Qué significado tiene la relación entre estas dos variables? ¿qué concluyes de esa relación?

Suma anécdota



Históricamente, la probabilidad nació con los juegos de azar, pues en la mayoría de los casos había una apuesta involucrada que motivaba a ganar o, en el peor de los casos, se buscaba perder lo menos posible, tanto en la partida simple, como en el transcurso de varias rondas.

Al tratar de averiguar una estrategia para ganar en más partidas o detectar si existía algún truco puesto por el rival, comenzaron las prácticas del conteo de cartas, de estudiar los lanzamientos de los dados, de analizar el lenguaje verbal del contrincante, en fin, de minimizar el factor suerte.

Contrario a lo que puede pensarse, la suerte tiene que ver más con la actitud y el pensamiento de la persona que con los eventos aleatorios que le suceden cotidianamente. Es decir, una persona con suerte es aquella que ha trabajado para tenerla.

Richard Wisemann es un investigador británico que, como la probabilidad, tuvo sus inicios en el estudio de patrones y trucos. Concretamente, aprendió el oficio de mago, siendo miembro del Círculo Mágico de Londres. Dejó el oficio luego de darse cuenta que no era rentable y estudió psicología y luego el doctorado.

Por sus orígenes, Wisemann ha estudiado los principios de la buena y la mala suerte. Demostró que la buena y mala suerte son resultado de hábitos medibles. Por ejemplo, personas suertudas, esperando buena suerte, pueden esforzarse más en alcanzar el resultado que desean, lo que se traduce en más éxito, lo que refuerza su creencia de tener buena suerte.

Las personas suertudas son extrovertidas y observadoras, por tanto, tienen más encuentros azarosos que las personas con mala suerte, lo que les da más oportunidad de tener una situación favorable. Más aún, las personas suertudas tienden a ver más el lado bueno de los encuentros "malos".

Contesta las preguntas.

1. De acuerdo con la lectura, ¿qué es la suerte?
2. ¿Compartes la visión de Wisemann sobre las personas suertudas?
3. ¿Cómo lograría una persona sin suerte convertirse en alguien suertudo?
4. ¿De qué depende tomar una situación como buena o mala?

Proyecto integrador

Tratamiento de la información.

El objetivo de este proyecto es elaborar un perfil de tus compañeros, identificando el alumno típico. Para ello se recogerán datos sobre sus características físicas.

Efectúa, con el grupo y tu docente lo que se pide. Contesta las preguntas.

1. Preparen una lista de las características que desean incluir.
2. Planifiquen cómo se obtendrán los datos de las características establecidas.
3. ¿Cuáles variables se obtendrán?
4. Para las variables, ¿cuáles datos se obtienen por simple observación?, ¿cuáles requieren de una medición?, ¿cuáles deben ser recabados mediante una encuesta?
5. Reúnan los materiales necesarios para tomar los datos de todos los compañeros del grupo.
6. ¿Qué tipo de variable es el sexo?, ¿qué identificador le asignaron?, ¿tiene sentido calcular la media de esta variable?, ¿tiene sentido calcular su moda?
7. ¿Cómo se distribuye el sexo de los alumnos del grupo?
8. ¿Cómo es el alumno típico de tu salón?, ¿es hombre o mujer?
9. Preparen una tabla de frecuencias y describan al chico y a la chica promedio de su salón, ¿coincide visualmente con los compañeros del salón?, ¿en qué se diferencia?, ¿el chico promedio también lo es para todos los estudiantes de la escuela?, ¿por qué?

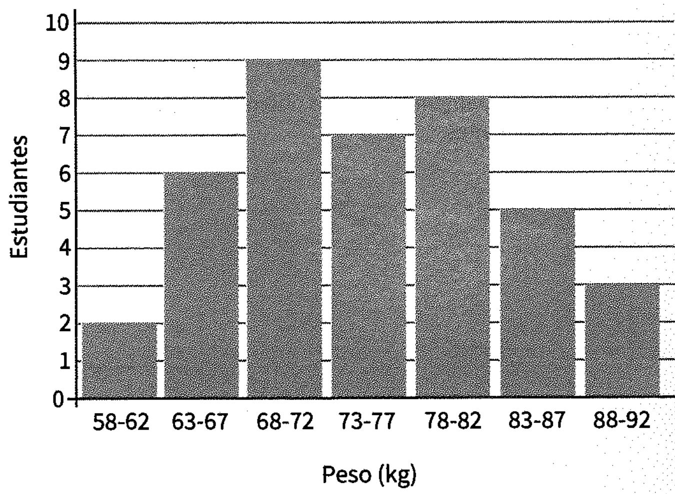
Media de salarios

Las medidas de tendencia central son muy útiles para conocer cómo es una muestra, sin embargo, hay datos que alteran mucho el resultado de estas medidas, por ejemplo, casos atípicos hacen que la media se "jale" hacia un extremo, haciendo que pierda la "generalidad" de los datos de la muestra y se infieran resultados distintos a los reales.

Hacia la prueba Planea

Resuelve los problemas que se plantean.

1. La gráfica muestra el peso en kilogramos de 40 estudiantes de sexto semestre de bachillerato.



Al seleccionar al azar a un estudiante, ¿en cuál intervalo es más probable que se ubique su peso?

a. 58 - 62 b. 68 - 72 c. 78 - 82 d. 88 - 92

2. Una cadena de restaurantes hará un sorteo, para lo cual se entrega un boleto a cada uno de sus clientes. En la ciudad de Puebla se repartieron $\frac{1}{16}$ del total de los boletos, en Guadalajara, $\frac{1}{4}$, en Monterrey, $\frac{5}{16}$ y en la Ciudad de México, $\frac{3}{8}$. ¿En qué ciudad es más probable que viva el ganador?

a. Puebla b. Guadalajara c. Monterrey d. Ciudad de México

3. Para comprar una computadora, seis amigos deben aportar en promedio $900.00, Los primeros cinco colaboran con $840, $1 090, $720, $900 y $920, ¿cuánto debería aportar el sexto amigo para poder comprarla?

a. $930 b. $900 c. $895 d. $894

4. Se realizó una encuesta en un edificio para determinar cuántas personas viven en cada uno de sus departamentos. Las respuestas obtenidas fueron: 6, 5, 3, 2, 6, 5, 1, 2, 0, 3, 1, 6, 5, 8, 3, 2, 5, 3, 6, 5. ¿Cuál es el valor de la mediana del número de personas que habitan en los departamentos?

a. 4 b. 2 c. 3 d. 5

Evalúa tus evidencias

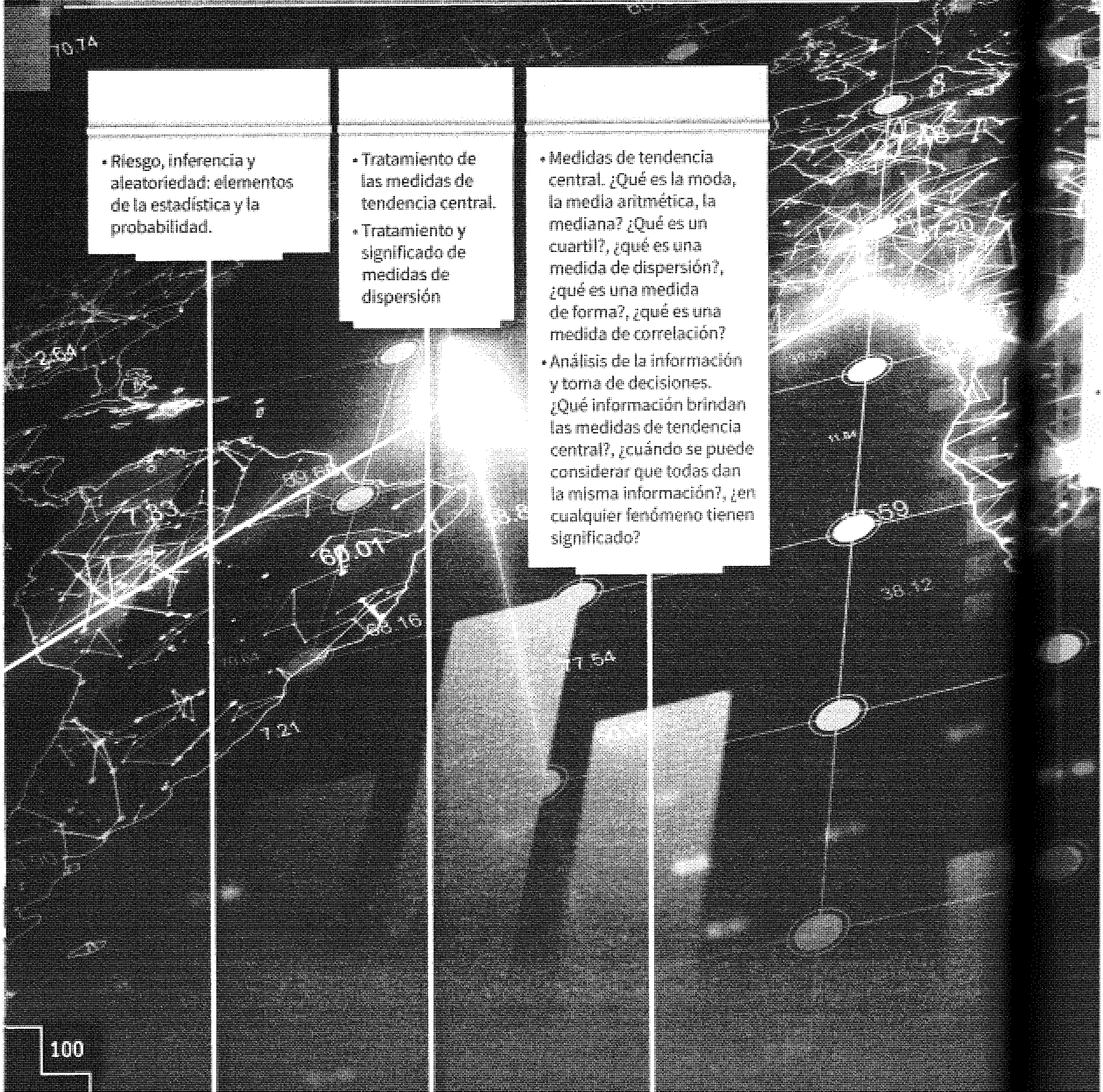
Productos	Criterios	Sí	No
Construcción de distintos tipos de gráficos y emisión de opiniones derivadas de ellos.	Coloca correctamente los datos obtenidos en los ejes correspondientes.		
	Utiliza la escala adecuada para la construcción de su gráfica.		
	Interpreta correctamente el comportamiento de los datos, aún variando la escala e intercambiando los datos de los ejes.		

Rúbrica

Aprendizajes esperados	Básico	Autónomo	Estratégico
Recolecta y ordena la información de alguna situación.	Conoce los distintos métodos de obtención de información.	Reconoce cada variable y las asocia correctamente con los datos obtenidos.	Identifica el tipo de variable y cómo recabar y organizar los datos de acuerdo con cada una de ellas.
Interpreta y analiza la información.	Distingue entre datos y comprende los cambios obtenidos en las mediciones.	Identifica las variaciones en las gráficas y los histogramas, y relaciona las causas posibles.	Relaciona correctamente la influencia entre variables, cuando existe y puede hacer una predicción con base en esas relaciones.
Representa la información.	Conoce los tipos de gráficas para representar datos.	Codifica y decodifica la información presentada en gráficas y tablas.	Identifica la gráfica más adecuada de acuerdo con el tipo de información que posee.
Toma decisiones a partir del análisis de la información.	Interpreta correctamente las medidas de tendencia central en las gráficas.	Infiere, a partir de las medidas de tendencia central, el comportamiento de la muestra y la población.	Observa la distribución de frecuencias y analiza las medidas de tendencia central para emitir una conclusión correcta de los datos.

Tercer parcial

Eje: Del manejo de la información al pensamiento estocástico



- Riesgo, inferencia y aleatoriedad: elementos de la estadística y la probabilidad.

- Tratamiento de las medidas de tendencia central.
- Tratamiento y significado de medidas de dispersión

- Medidas de tendencia central. ¿Qué es la moda, la media aritmética, la mediana? ¿Qué es un cuartil?, ¿qué es una medida de dispersión?, ¿qué es una medida de forma?, ¿qué es una medida de correlación?
- Análisis de la información y toma de decisiones. ¿Qué información brindan las medidas de tendencia central?, ¿cuándo se puede considerar que todas dan la misma información?, ¿en cualquier fenómeno tienen significado?

- Calcula las medidas de tendencia central, medidas de dispersión, medidas de forma y medidas de correlación.
- Interpreta las medidas de tendencia central desde el análisis del gráfico estadístico, así como su variabilidad y representación de la situación contextual.
- Toma decisiones a partir de las medidas de tendencia central y su representación con respecto a un conjunto de datos.

- Argumento de qué es una medida de tendencia central y qué es una medida de dispersión.
- Ejemplos de dichas medidas.
- Construcción de cuartiles a partir de datos dados.

Elijo, decido y quiero

Para reflexionar

¿Cuántas veces realizas actividades sin motivación porque consideras que las haces por obligación?

¿Cómo puedes responsabilizarte por las decisiones que tomas y las acciones que realizas?

Nuestro objetivo

Reconocer la responsabilidad de tomar nuestras decisiones.

Paso a paso:

En muchas ocasiones de nuestra vida cotidiana sentimos que estamos atrapados frente a acciones que consideramos obligatorias. Cuando las vemos de esta forma, las sentimos como una imposición y muchas veces no estamos contentos al hacer esas actividades.

1. Escribe rápidamente diez frases en las que tengas que iniciar la oración con: "Tengo que...".
2. Analiza tus frases anteriores y selecciona al menos cinco de ellas, que podrías iniciar de la siguiente forma: Elijo, decido y quiero...".
3. ¿Qué sientes cuando debes hacer algo por obligación?, ¿qué sientes cuando haces un proyecto o actividad por convicción propia?, ¿el resultado es el mismo si haces una actividad por obligación o por convicción?

Para terminar...

En todo momento estamos tomando decisiones, tú eres artífice de tu vida y de tu destino. Es mucho más fácil quejarse, que tomar las riendas y enfrentar las consecuencias de nuestra vida.

Proyecto de vida

A lo largo del semestre desarrollaremos tu Proyecto de vida, con el propósito de aclarar qué deseas para ti; que puedas tomar decisiones que marquen la dirección de tu futuro, así como reflexionar acerca de las implicaciones que tendrá en tu vida el hecho de llevarlas a cabo.

Organiza la información de tu Proyecto de vida.

1. Copia y completa en tu cuaderno el cuadro.
2. Escribe en la primera columna tus metas más importantes. Agrega las filas que sean necesarias.

Meta	Ideas importantes a rescatar	Acciones que debo llevar a cabo	Beneficios que representa para mí	Beneficios que representa para mi entorno

3. Una vez que organizaste la información en el cuadro, léela con atención para tener un panorama de lo que implica el planteamiento de tu Proyecto de vida.
4. Escribe por qué es importante elaborar tu Proyecto de vida.
5. Diseña un organizador gráfico en el cual incorpores frases que te motiven a trabajar día con día en tu Proyecto de vida, incluye tus metas para que las tengas presentes todo el tiempo. Coloca este organizador en un lugar visible en tu casa. Utiliza el siguiente esquema para hacer un esbozo y enriquécelo paulatinamente.

Proyecto de vida de: ____________________

	Metas a corto plazo	Metas a mediano plazo	Metas a largo plazo
Personal			
Familiar			
Escolar			
Laboral			
Social			

Tratamiento y significado de las medidas de dispersión

Los gráficos son una herramienta importante para visualizar variables, en el parcial anterior aprendiste a graficar la distribución de las variables usando histogramas. Sin embargo, muchas veces la diferencia entre un tipo de distribución y otra es difícil de percibir a simple vista. Por esta razón es importante que aprendas a describir numéricamente la distribución de las variables.

En este parcial aprenderás diferentes estadísticos descriptivos (valores que describen a las variables) relacionados con la distribución de variables, tales como: percentiles, desviación estándar, varianza, curtosis y asimetría; así como dos estadísticos relacionados con medir la relación entre dos variables: la covarianza y un coeficiente de correlación.

¿Qué es un percentil?

Los percentiles son estadísticos descriptivos que te dicen el dato por debajo del cuál se encuentra determinado porcentaje de los datos.

Ejemplo

Retomemos la variable de la página 26: de los refrescos que 32 familiares de Zutanito se toman en un mes. Los 32 valores ordenados, de menor a mayor, son los siguientes:

5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 10, 10 y 10.

El percentil 50 de esta variable sería el valor por debajo del cuál está el 50% de los datos (la mitad). Si la variable tiene 32 datos, el percentil 50 será que esté entre el dato 16 y 17; en este ejemplo, siete.

Si te das cuenta, básicamente se trata de calcular un porcentaje. Multiplicas el número de datos por el porcentaje que quieres conocer; si el resultado de esa multiplicación es un número decimal, el percentil buscado será el siguiente número entero; si el resultado es un número entero entonces promedias ese valor con el siguiente.

Veamos el ejemplo del percentil 50:

$$P_{50} = 32 \text{ datos} \cdot 0.50 = \text{posición } 16.$$

Como se trata de un número entero, el percentil 50 será el promedio del dato 16 con el dato 17:

$$P_{50} = \frac{7+7}{2} = 7.$$

Ejemplo

El percentil 86 sería el valor por debajo del cuál está el 86% de los datos:

$$P_{86} = 32 \cdot 0.86 = 27.52.$$

Como no es un valor entero, se considera la siguiente posición entera, es decir, la posición 28.

Dicha posición en la variable la ocupa el número nueve, por lo tanto, $P_{86} = 9$.

Cuartiles

Los percentiles más utilizados en la estadística descriptiva son los cuartiles. Son los percentiles que dividen a una variable en cuatro partes iguales.

En otras palabras, los cuartiles son los percentiles 25, 50 y 75.

Los cuartiles se representan con la letra Q y son tres: el primer cuartil ($Q_1 = P_{25}$), el segundo ($Q_2 = P_{50}$) y el tercero ($Q_3 = P_{75}$). Los cuartiles se calculan igual que el percentil al que corresponden.

Ejemplo

En el siguiente cuadro puedes ver el cálculo de los cuartiles de la siguiente variable de 16 datos:

0.3, 0.9, 2.7, 2.9, 3.6, 4.1, 5.0, 5.5, 7.0, 7.1, 7.2, 7.5, 7.9, 8.7, 9.0 y 9.5.

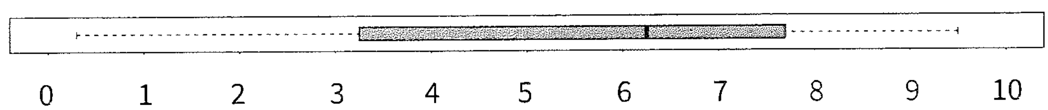
<table>
<tr><th>Cuartil</th><th>Percentil</th><th>Posición</th><th colspan="2">Cálculo</th></tr>
<tr><td>Q_1</td><td>P_{25}</td><td>$16 \cdot 0.25 =$ posición 4</td><td rowspan="3">Al ser número entero se toma el dato de esa posición y se promedia con el de la siguiente posición.</td><td>$\frac{2.9+3.6}{2} = 3.25$</td></tr>
<tr><td>Q_2</td><td>P_{50}</td><td>$16 \cdot 0.5 =$ posición 8</td><td>$\frac{5.5+7.0}{2} = 6.25$</td></tr>
<tr><td>Q_3</td><td>P_{75}</td><td>$16 \cdot 0.75 =$ posición 12</td><td>$\frac{7.5+7.9}{2} = 7.7$</td></tr>
</table>

Si recuerdas las medidas de tendencia central, ya te habrás dado cuenta que el segundo cuartil y el percentil 50 son lo mismo que la mediana.

Boxplots

Los gráficos utilizados para mostrar los cuartiles se conocen como gráficos de caja y bigote o *boxplots*. En ellos se coloca un rectángulo que abarca desde el primer cuartil hasta el tercero, y se marca el segundo cuartil (mediana) con una línea. A cada extremo de la caja se le dibuja un "bigote" que abarca el valor mínimo por un lado y máximo por el otro.

El *boxplot* que se muestra a continuación presenta la variable del ejemplo anterior. Como puedes ver, la caja comienza en el primer cuartil (3.25) y termina en el tercer cuartil (7.7); la línea negra dentro del rectángulo marca la mediana (6.25) y los extremos de los bigotes van del valor más bajo (0.3) al más alto (9.5).



Si te fijas, la variable queda partida en cuatro partes, en cada una de esas partes cae el 25% de los datos, es decir, en cada sección caen cuatro de los 16 datos de la variable.

En este gráfico, el 50% de los datos se encuentran en el rango del rectángulo. Puedes concluir que el 50% de los datos está entre los valores 3.25 y 7.7.

El *boxplot* se puede hacer vertical u horizontal, representan lo mismo. Los *boxplots* son otro tipo de gráfico, además del histograma, que se utiliza para explorar la distribución de las variables.

Con el objetivo de evidenciar datos "atípicos" (datos que se alejan mucho de la mediana), a veces la extensión de los bigotes no se grafica utilizando los valores mínimo y máximo de la variable, en vez de esos, se utiliza 1.5 veces el rango intercuartil (la diferencia que hay entre Q_1 y Q_3), a partir de Q_1 para el valor mínimo y de Q_3 para el valor máximo.

Siguiendo esta regla, los valores que queden más allá de ese límite se grafican como puntos aislados fuera de los bigotes.

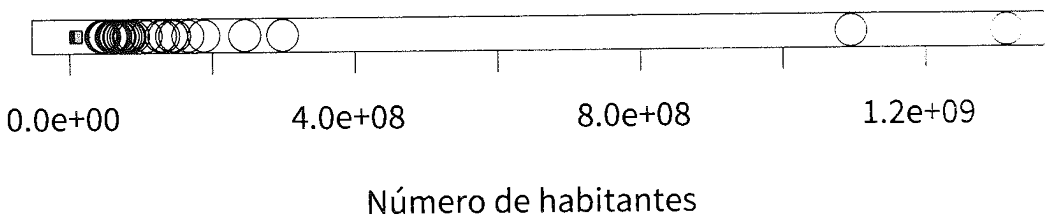
Ejemplo

Cuartil	Valor
Q_1	437 624
Q_2	4 786 994
Q_3	17 497 773

Considera la variable "población" de la base de datos de los 227 países que viste en la página 86.

Como recordarás, la mayoría de los países tienen pocos habitantes y muy pocos países tienen muchos habitantes. En esta variable, los cuartiles son los que se muestran en el cuadro de la izquierda.

El *boxplot* de la variable “población” de se vería como la siguiente imagen:



El rango intercuartil sería: $Q_3 - Q_1$ = 17 497 773 – 437 624 = 17 060 149.

1.5 veces el rango intercuartil sería: 17 060 149 · 1.5 = 25 590 223.5.

El bigote inferior quedaría entonces en el valor:

$$Q_1 - 1.5 \text{ rango intercuartil} = 437\,624 - 25\,590\,223.5 = -25\,152\,599.5,$$

sin embargo, el valor más pequeño de los 227 datos es 7 026, por lo que el bigote inferior está graficado en 7 026. El bigote superior debería quedar en el valor:

$$Q_3 + 1.5 \text{ rango intercuartil} = 17\,497\,773 + 25\,590\,223.5 = 43\,087\,996.5.$$

Como no hay un país que tenga esa cantidad específica de habitantes, el dato más cercano hacia la mediana es 41 236 378, por lo que es en ese valor donde está graficado el bigote superior, los países con una población mayor a 41 236 378 habitantes, están graficados como puntos aislados

Como podrás darte cuenta, los valores están tan cercanos entre sí, que en el gráfico no se distinguen el bigote inferior, el primer y el segundo cuartil.

Actividad de aprendizaje 1

En cada inciso, realiza lo que se te pide.

1. Utiliza la base de datos de los cacomixtles que descargaste en la página 73 para calcular los percentiles señalados y llenar la tabla:

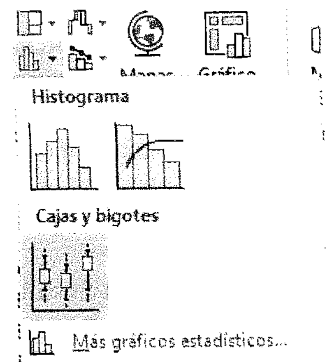
Percentil	Longitud	Peso	Temperatura	Pulgas
P_5				
P_{10}				
P_{30}				

percentil	longitud	peso	temperatura	pulgas
P_{55}				
P_{70}				
P_{90}				
P_{95}				

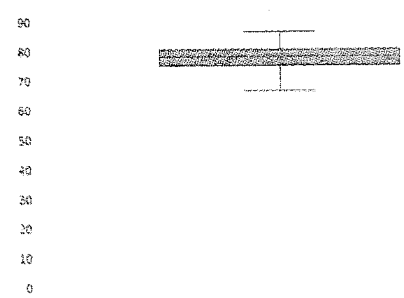
2. Sigue las siguientes instrucciones para hacer el *boxplot* de la variable "longitud" de la base de datos de los cacomixtles utilizando Excel:

   a. Selecciona la variable longitud.

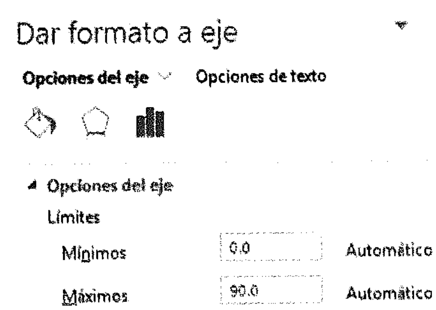
   b. En el menú "Insertar", dentro del submenú "Gráfico", selecciona la opción "Insertar gráfico de estadística" y luego la opción "Cajas y bigotes" como se muestra en la imagen.



   c. Es probable que el gráfico salga con el eje Y empezando desde el cero, como se muestra en la imagen.



   d. Para ajustar esto, da clic derecho sobre el eje y selecciona la opción "Dar formato al eje". Se abrirá un menú a la derecha de la pantalla y deberás cambiar el cero en "Mínimos" por el valor mínimo que quieres que tome el eje Y. Para el caso de la longitud, 60 es un buen valor pues el valor más chico es de 66.08.



3. Haz un *boxplot* para cada una de las otras variables de la misma base de datos ("peso", "temperatura" y "pulgas").

Actividad de aprendizaje 2

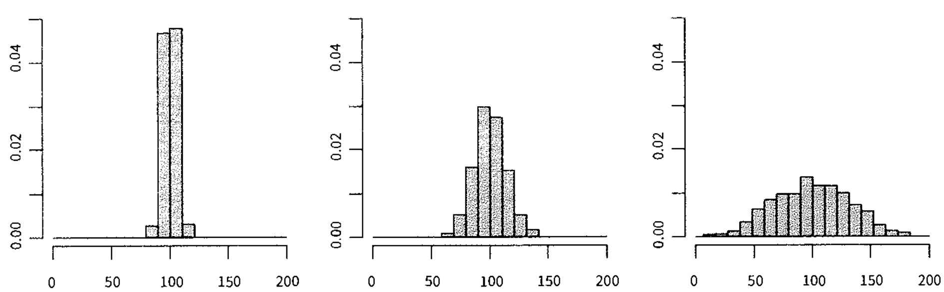
Utiliza la base de datos de las orugas de la página 94 para completar el cuadro.

			Días
Q_1			
Q_2			
Q_3			
Rango intercuartil			
1.5 rango intercuartil			
$Q_1 - 1.5$ rango intercuartil			
Valor mínimo de la variable			
$Q_3 + 1.5$ rango intercuartil			
Valor máximo de la variable			

¿Qué es una medida de dispersión?

Son estadísticos que te dicen qué tan "desparramada" está una variable alrededor de la media. Las medidas de dispersión te dan una idea de la distribución de los datos a lo largo del rango, y te permiten hacer inferencias dependiendo de si la mayoría de los datos están cerca de la media o no.

A continuación una representación gráfica de cómo puede variar la dispersión en torno a la media:



En los tres histogramas estás viendo una variable que tiene la misma cantidad de datos (1 000 valores), la misma media (100), pero distinta dispersión.

El histograma de la izquierda tiene poca dispersión, es decir, todos los valores están cerca de la media; el del centro tiene un poco más de dispersión, observa que hay valores más alejados de la media; y el de la derecha tiene mucha dispersión, pues hay valores muy lejanos a la media.

Al igual que las medidas de tendencia central, existen muchas medidas de dispersión aunque son dos las más famosas y las que casi siempre se usan: la varianza y la desviación estándar. En realidad estas dos medidas están muy relacionadas, pues la desviación estándar es la raíz cuadrada de la varianza, por lo que si sabes calcular la varianza, sólo tendrás que calcular su raíz cuadrada para conocer la desviación estándar.

Algo que debes saber y siempre recordar sobre estos estadísticos descriptivos es que conceptualmente todos ellos (media, mediana, moda, desviación estándar y varianza) tienen dos versiones, la muestral y la poblacional. En otras palabras, si calculas la media de una muestra obtienes la media muestral y si calcularas la media de la población obtendrías la media poblacional.

Sin embargo, nunca o casi nunca se mide a toda la población, así que casi nunca calcularás los estadísticos poblacionales. La mayoría de las veces registrar toda la población es muy demandante de tiempo o de dinero, y además porque si tu muestra es representativa, el estadístico muestral será un aproximado bastante confiable del poblacional.

Varianza

La varianza es un promedio de las distancias de cada dato a la media (lee el enunciado anterior hasta que lo asimiles), sólo que cada distancia está elevada al cuadrado y se representa con la letra griega sigma minúscula elevada al cuadrado (σ^2). Su expresión matemática es la siguiente:

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

Donde x_i es el i–ésimo valor en la variable, $\overline{x}$ representa a la media de la variable, n es el número de valores y Σ a la sumatoria desde el primer valor hasta n.

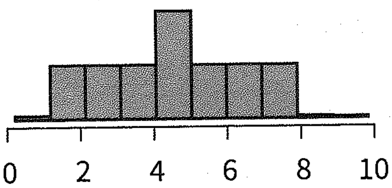
En otras palabras, a cada valor le calculas su diferencia con la media, la elevas al cuadrado, sumas todas las diferencias elevadas al cuadrado y las divides entre el número de valores.

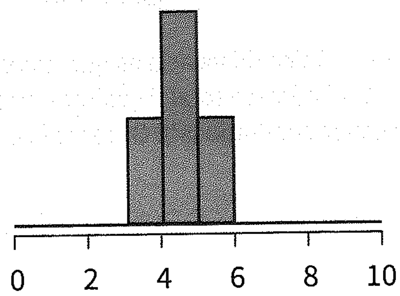
Ejemplo

Considera las siguientes dos variables que tienen el mismo promedio pero diferente dispersión. Por fines didácticos, cada variable tiene sólo ocho datos:

									Media
V_1	2	3	4	5	5	6	7	8	5
V_2	4	4	5	5	5	5	6	6	5

¿Puedes determinar qué histograma corresponde a cada variable?





Es evidente que una variable está más dispersa alrededor del cinco (la media) que la otra, pero ¿cómo medirlo?

El primer paso es medir la distancia de cada valor a la media, es decir, calcular la diferencia que hay entre cada valor y su media:

V_1	2 − 5 = −3	3 − 5 = −2	4 − 5 = −1	5 − 5 = 0	5 − 5 = 0	6 − 5 = 1	7 − 5 = 2	8 − 5 = 3
V_2	4 − 5 = −1	4 − 5 = −1	5 − 5 = 0	5 − 5 = 0	5 − 5 = 0	5 − 5 = 0	6 − 5 = 1	6 − 5 = 1

Puedes ver que V_1 tiene valores de hasta tres unidades por debajo y por arriba del promedio, mientras que en V_2, los valores más alejados del promedio están a sólo una unidad.

Ahora que tienes las diferencias, las elevas al cuadrado y las promedias.

									Media
V_1	9	4	1	0	0	1	4	9	3.5
V_2	1	1	0	0	0	0	1	1	0.5

Y ahí tienes la varianza.

En este ejemplo, V_1 tiene una varianza poblacional de 3.5 y V_2 de 0.5, es decir, los valores en V_1 están más desparramados alrededor de la media que los valores en V_2. Si sustituyes en la expresión matemática para V_1 se vería así:

$$\frac{(2-5)^2+(3-5)^2+(4-5)^2+(5-5)^2+(5-5)^2+(6-5)^2+(7-5)^2+(8-5)^2}{8}=\frac{28}{8}=3.5.$$

A pesar de tener un cálculo de la dispersión de la variable alrededor de la media, la magnitud de este valor a veces carece de interpretación si no hay otro con cuál compararlo.

Es decir, si sólo hubieras tenido la varianza de V_1, no podrías interpretar gran cosa, sin embargo, como tienes la varianza de V_2, entonces puedes decir que V_1 tiene mayor dispersión que V_2, lo que sea que tus variables signifiquen.

Algo que debes saber es que el cálculo y la expresión matemática que acabas de ver es la fórmula para la varianza poblacional, sin embargo, casi siempre la que usarás es la de la varianza muestral.

En el cálculo, la única diferencia es que en vez de dividir entre n (el número total de valores), divides entre n–1 (el número total de valores menos uno), además la varianza muestral se representa por la letra ese al cuadrado (s^2). Observa la expresión matemática:

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

Desviación estándar

La desviación estándar es la raíz cuadrada de la varianza, la desviación estándar muestral se representa por una ese (s) y la poblacional por la letra sigma (σ).

En el cálculo de la varianza, cada distancia a la media se eleva al cuadrado para evitar un promedio de cero. Sin embargo, esto hace que las unidades (pesos, metros, centímetros, etcétera) de la varianza sean unidades cuadráticas respecto a la variable, de este modo, al sacarle la raíz cuadrada obtienes una medida de dispersión en las mismas unidades que la variable. Esa es su principal propiedad.

Ejemplo

La desviación estándar poblacional de V_1 y V_2 del ejemplo anterior es la raíz cuadrada de 3.5 y 0.5 respectivamente, es decir, 1.87 para V_1 y 0.71 para V_2.

Actividad de aprendizaje 3

En cada inciso, realiza lo que se te pide.

1. Calcula la varianza muestral de la cantidad de dinero al día que gasta Miguel en transporte y aperitivos escolares. Los gastos de las últimas cuatro semanas de lunes a viernes fueron:

   15, 47, 54, 5, 40, 120, 40, 31, 47, 64, 40, 64, 54, 65, 20, 47, 18, 54, 65 y 85.

2. Así como en Excel puedes calcular la media y mediana usando funciones, investiga cuál es y cómo se usa la función en Excel para calcular la varianza muestral y la función para la varianza poblacional. Compara la varianza muestral que calculaste en el ejercicio anterior con la calculada en Excel. ¿Tuviste algún problema?, ¿cuál?
3. En un mismo gráfico dibuja el *boxplot* para V_1 y el *boxplot* para V_2 utilizadas en el ejemplo de la página anterior. ¿Qué relación te parece que hay entre la varianza/desviación estándar de cada variable y los *boxplots*?
4. Utiliza Excel para calcular la varianza y desviación estándar de las variables en la base de datos de los cacomixtles que descargaste en la página 73 y completa el cuadro de la siguiente página.

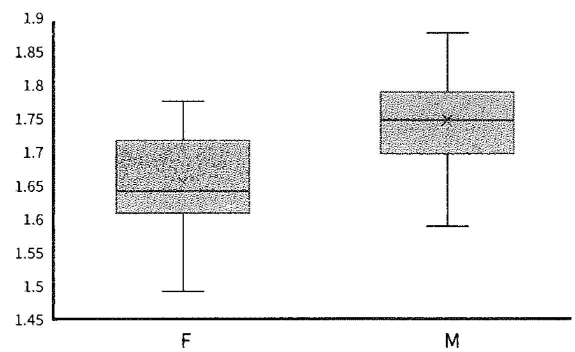
Medida de dispersión	Longitud	Peso	Temperatura	Pulgas
Varianza				
Desviación estándar				

Actividad de aprendizaje 4

En cada inciso, realiza lo que se te pide.

1. Sigue las instrucciones de la página 108 para hacer un *boxplot* en Excel, pero en vez de seleccionar sólo una variable, selecciona simultáneamente las variables "estatura" y "sexo" de la base de datos de las alturas de Alonso Quijano de la página 17.

   Como resultado debes tener dos *boxplots* en un mismo gráfico como se muestra en la imagen de la derecha.



2. Utilizando Excel y la misma base de datos, completa el siguiente cuadro y responde los incisos (Sugerencia: para facilitarte el trabajo puedes utilizar filtros o tablas dinámicas).

Estadístico descriptivo	Todas las estaturas	Estaturas de mujeres	Estaturas de hombres
Media aritmética			
Varianza			
Desviación estándar			

   a. ¿Qué tan similar o diferente dirías que es la dispersión de las estaturas entre hombres y mujeres?

   b. ¿Por qué la varianza y la desviación estándar de todas las alturas juntas es mayor que la de los dos sexos por separado?

   c. ¿Qué relación encuentras entre los *boxplots* que hiciste en el primer inciso con la información del cuadro?

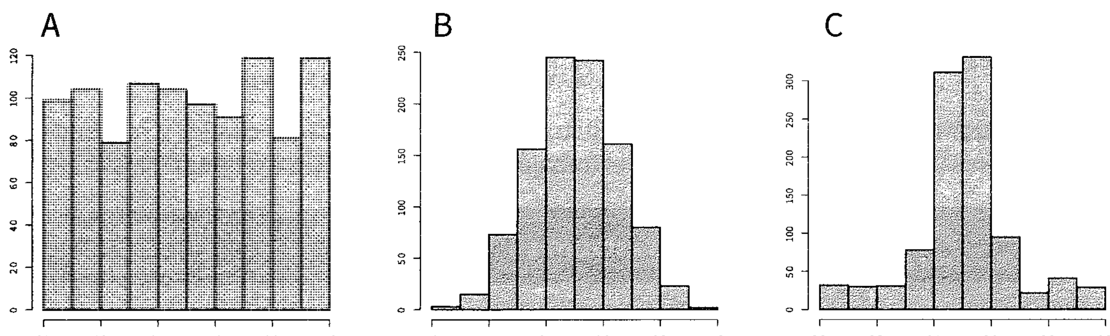
3. Repite los dos ejercicios anteriores, pero ahora con las variables “edad” y “sexo”.

Estadístico descriptivo	Todas las estaturas	Estaturas de mujeres	Estaturas de hombres
Media aritmética			
Varianza			
Desviación estándar			

4. ¿Qué diferencias encuentras entre las comparaciones de estaturas y edades entre sexos? ¿por qué la varianza y desviación estándar de todas las edades juntas no es mayor que la de los dos sexos por separado?

¿Qué es una medida de forma?

Observa los siguientes histogramas:



Los histogramas tienden a tener ciertos patrones de formas. A la forma de cerro que a veces se ve en los histogramas se le conoce como campana y a los lados de la campana se les llama colas, una campana por lo general tiene dos colas: la de la derecha y la de la izquierda.

La forma de campana es muy famosa, pues aparece en muchos tipos de distribuciones. La **curtosis** es una medida de apuntalamiento de la distribución de una variable, en otras palabras nos dice qué tan puntiaguda es la distribución. Existen muchas fórmulas para calcular la curtosis de una variable, una de ellas es la siguiente:

$$k(x)=\left\{\frac{n(n+1)}{(n-1)(n-2)(n-3)}\cdot\sum\left(\frac{x_i-\overline{x}}{s}\right)^4\right\}-\frac{3(n-1)^2}{(n-2)(n-3)}.$$

Donde n es el número de datos en la variable, x_i cada dato, $\overline{x}$ la media y s la desviación estándar. Usando esta ecuación obtendrás un valor de curtosis k.

Cuando k es menor que cero, se dice que la distribución es **platicúrtica**, esto es chata o aplanada. Lo que significa que los valores están homogéneamente distribuidos a lo largo del rango e implica que el centro de la variable no sobresale de manera evidente. Por ejemplo, la distribución en el histograma A es platicúrtica, su curtosis es $k = -1.19$.

Cuando k es cero, se dice que la distribución es **mesocúrtica**, es decir, tiene una campana considerada "normal"; ni puntiaguda ni chata. Esto significa que el centro de la variable sobresale, pero no abruptamente. Los valores de k iguales a cero sólo existen en distribuciones teóricas.

Para términos prácticos de este curso puedes asumir (únicamente con esta ecuación) que una variable tiene distribución mesocúrtica si el valor de k está entre -0.27 y 0.33. Por ejemplo, la distribución en el histograma B es mesocúrtica, su coeficiente de curtosis es $k = -0.18$.

Por último, si k es mayor que cero se dice que la distribución de la variable es **leptocúrtica**, esto es muy puntiaguda, lo que significa que el centro de la variable resalta de manera abrupta porque hay un valor o valores muy cercanos entre sí que se repiten muchas veces en comparación con el resto de los valores; un ejemplo de una distribución leptocúrtica sería la del histograma C, su k es 1.77.

Calculemos la curtosis paso a paso en V_1 del ejemplo para calcular la varianza: 2, 3, 4, 5, 5, 6, 7 y 8. Son ocho datos ($n = 8$), y sabemos que el promedio es $\bar{x} = 5$ y la desviación estándar muestral (recuerda la diferencia entre la desviación estándar poblacional y muestral) es $s = 2$.

Si sustituimos en la ecuación tenemos que:

$$
\begin{aligned}
k(V_1) &= \left\{ \frac{8(8+1)}{(8-1)(8-2)(8-3)} \cdot \sum \left(\frac{x_i - 5}{2} \right)^4 \right\} - \frac{3(8-1)^2}{(8-2)(8-3)} \\
&= \left\{ \frac{8(9)}{(7)(6)(5)} \cdot \sum \left(\frac{x_i - 5}{2} \right)^4 \right\} - \frac{3(7)^2}{(6)(5)} = \left\{ \frac{72}{210} \cdot \sum \left(\frac{x_i - 5}{2} \right)^4 \right\} - \frac{147}{30} \\
&= \left\{ 0.3428 \cdot \sum \left(\frac{x_i - 5}{2} \right)^4 \right\} - 4.9.
\end{aligned}
$$

Resolvamos ahora la parte de la sumatoria $\sum \left(\frac{x_i - 5}{2} \right)^4$:

$$
\begin{aligned}
\sum \left(\frac{x_i - 5}{2} \right)^4 &= \left(\frac{2-5}{2} \right)^4 + \left(\frac{3-5}{2} \right)^4 + \left(\frac{4-5}{2} \right)^4 + \left(\frac{5-5}{2} \right)^4 + \left(\frac{5-5}{2} \right)^4 + \left(\frac{6-5}{2} \right)^4 + \left(\frac{7-5}{2} \right)^4 + \left(\frac{8-5}{2} \right)^4 \\
&= \left(\frac{-3}{2} \right)^4 + \left(\frac{-2}{2} \right)^4 + \left(\frac{-1}{2} \right)^4 + \left(\frac{0}{2} \right)^4 + \left(\frac{0}{2} \right)^4 + \left(\frac{1}{2} \right)^4 + \left(\frac{2}{2} \right)^4 + \left(\frac{3}{2} \right)^4 \\
&= (-1.5)^4 + (-1)^4 + (-0.5)^4 + (0)^4 + (0)^4 + (0.5)^4 + (1)^4 + (1.5)^4 \\
&= 5.0625 + 1 + 0.0625 + 0 + 0 + 0.0625 + 1 + 5.0625 = 12.25.
\end{aligned}
$$

Sustituimos en la ecuación original:

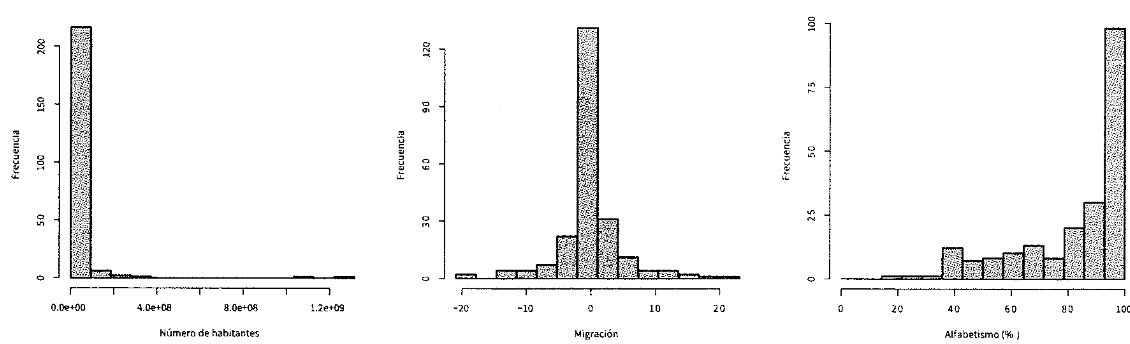
$$k(V_1) = \{0.3428 \cdot 12.25\} - 4.9 = 4.1993 - 4.9 = -0.7007.$$

Esto significa que la distribución de la variable tiene curtosis platicúrtica, o sea que la mayoría de los valores están distribuidos homogéneamente a lo largo del rango, el histograma no mostraría un apuntalamiento pronunciado del centro de la variable.

¿Qué es una medida de sesgo?

Otra característica que tiene una distribución es su simetría: una distribución puede ser simétrica o asimétrica. Cuando una distribución es simétrica, significa que hay tantos valores menores que la media como los hay mayores que ella; que la distribución sea asimétrica significa que hay más valores mayores que la media de los que hay menores que ella o viceversa (más menores que mayores).

Como ya vimos, usando histogramas puedes darte una idea de si una variable es simétrica o asimétrica, los siguientes histogramas muestran los tres casos más comunes de simetría (centro) y asimetría (izquierda y derecha), son una copia de los vistos en las páginas 86 y 87 utilizando las variables de la base de datos de los países del mundo.



Muchas veces no es fácil deducir a simple vista si una distribución es simétrica o asimétrica. Se han creado varias fórmulas para calcular si una distribución es simétrica o hacia qué lado es la asimetría. La más popular es la fórmula para calcular el coeficiente de simetría de Fisher:

$$g_1(x) = \frac{\sum (x_i - \bar{x})^3}{ns^3}.$$

Donde n es el número de datos en la variable, $\bar{x}$ el promedio y s la desviación estándar.

Si g_1 es mayor que cero, se dice que la distribución tiene asimetría positiva, lo cual implica que la mayoría de los valores son menores que la media, en un histograma se ve que la campana está cargada a la izquierda. El histograma de la izquierda muestra asimetría positiva (g_1 = 9.07). Esto significa que hay muchos países con poblaciones pequeñas (menor al promedio) y pocos países con una población grande (mayor a la del promedio).

Cuando g_1 es igual a cero, la distribución es simétrica; lo que implica que la mitad de los valores son mayores que la media y la otra mitad menores. Al igual que la medida de curtosis, sólo distribuciones teóricas tienen un coeficiente g_1 exactamente igual a cero. Por lo general, si obtienes un valor g_1 entre −0.15 y 0.15, puedes asumir que la distribución es simétrica.

El histograma del centro muestra un ejemplo de distribución simétrica (g_1 = 0.12). Esto significa que los valores de migración están distribuidos equitativamente alrededor del promedio, si lo piensas un poco descubrirás que esto tiene mucho sentido.

Por último, si el valor g_1 es menor que cero, se dice que tiene asimetría negativa, lo que significa que la mayoría de los valores son mayores que la media. Cuando haces un histograma con asimetría negativa, se puede apreciar que la campana está cargada a la derecha.

El histograma de la derecha muestra asimetría negativa (g_1 = −1.2). Esto significa que hay muchos países con niveles altos de alfabetismo (mayores que el promedio) y pocos países con niveles bajos de alfabetismo (menores que el promedio).

Ahora veamos un ejemplo paso a paso usando V_1 de la página 110. Al sustituir en la fórmula la media, el número de datos y la desviación estándar, tenemos que:

$$g_1(V_1) = \frac{\sum (x_i - 5)^3}{8 \cdot 2^3}.$$

Luego desarrollemos la sumatoria (Σ):

$$\begin{aligned} g_1(V_1) &= \frac{(2-5)^3+(3-5)^3+(4-5)^3+(5-5)^3+(5-5)^3+(6-5)^3+(7-5)^3+(8-5)^3}{8 \cdot 2^3} \\ &= \frac{(-3)^3+(-2)^3+(-1)^3+(0)^3+(0)^3+(1)^3+(2)^3+(3)^3}{8 \cdot 2^3} \\ &= \frac{-27-8-1+0+0+1+2+27}{8 \cdot 2^3} = \frac{0}{64} = 0. \end{aligned}$$

Esto significa que la distribución de V_1 es simétrica, es decir, que hay tantos valores mayores que la media como menores a esta.

Actividad de aprendizaje 5

En cada inciso, realiza lo que se pide.

1. Crea un histograma para cada variable cuantitativa de la base de datos de los cacomixtles (longitud, peso, temperatura y pulgas) que descargaste en la página 73, y a ojo de buen cubero determina la curtosis (platicúrtica, mesocúrtica o leptocúrtica) y simetría de cada variable (asimétrica positiva, asimétrica negativa o simétrica).

   Anota tus predicciones en el cuadro de la siguiente página.

Estadístico	Longitud	Peso	Temperatura	Pulgas
Curtosis				
Simetría				

2. Investiga cuáles son las funciones en Excel para calcular la curtosis y la simetría de una variable. Utiliza ambas funciones para completar el siguiente cuadro. Además de colocar el valor de cada estadístico, anota si se trata de una distribución platicúrtica, mesocúrtica o leptocúrtica; y si es asimétrica positiva, asimétrica negativa o simétrica. Compara los resultados con tus predicciones.

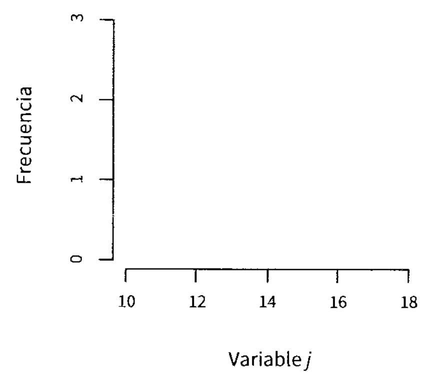
Estadístico	Longitud	Peso	Temperatura	Pulgas
Curtosis				
Simetría				

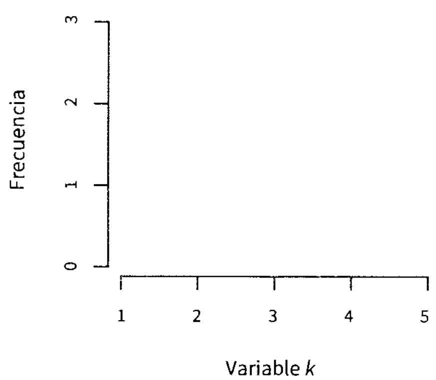
Actividad de aprendizaje 6

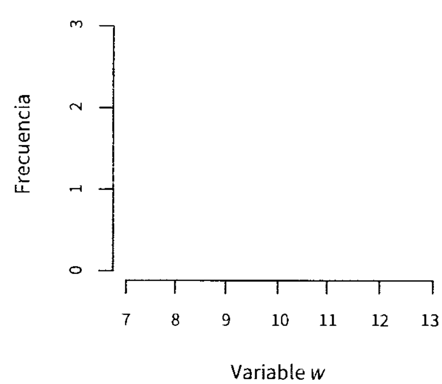
En cada inciso, realiza lo que se pide.

1. Crea tres histogramas utilizando las variables *j*, *k* y *w* de la pequeña base de datos que se muestra a continuación.

j	13	18	12	14	11	12	15
k	4	3	5	2	4	3	3
w	10	9	13	12	8	13	11







2. Calcula la media, mediana, curtosis y simetría de cada variable y completa el cuadro de la derecha.

Variable	*j*	*k*	*w*
Media			
Mediana			
Curtosis			
Asimetría			

3. Marca en cada histograma con color azul la media y color rojo la mediana.

4. Determina si los siguientes enunciados son falsos o verdaderos:

   a. Si la media de una variable es mayor que diez, es una media alta.

   b. Si una distribución tiene asimetría positiva, la media será mayor que la mediana.

   c. Si una distribución tiene asimetría negativa, la mayor cantidad de datos será mayor que la mediana.

   d. La distribución de la variable *k* se puede considerar mesocúrtica.

   e. La distribución de la variable *w* se considera leptocúrtica, pues los valores están homogéneamente distribuidos a lo largo del rango.

   f. Si una distribución tiene asimetría positiva, la mediana será mayor que la media.

   g. La distribución de la variable *j* se considera mesocúrtica porque el índice es mayor que cero pero por poco.

¿Qué es una medida de correlación?

Covarianza

La covarianza es un estadístico que dice cómo varían en conjunto dos variables, sirve para describir la relación entre dos variables cuantitativas. Esta relación puede ser positiva, negativa o neutra.

Si la relación es positiva, significa que conforme una variable aumenta, la otra también; si la relación es negativa, significa que a medida que aumenta una variable, la otra disminuye; si la relación es neutra, significa que las variables no tienen ninguna relación.

Hay dos fórmulas que puedes usar para calcular la covarianza, ambas dan exactamente lo mismo.

$$\text{cov}(x,y)=\frac{1}{n}\sum_{i=1}^{n}(xi-\bar{x})(yi-\bar{y}) \text{ o } \text{cov}(x,y)=\frac{1}{n}\sum_{i=1}^{n}(x_i\cdot y_i)-\bar{x}\cdot\bar{y},$$

Donde *x* es una variable, y es la otra variable y *n* el número de datos. Además, $\bar{x}$ y $\bar{y}$ son los promedios respectivos de ambas variables.

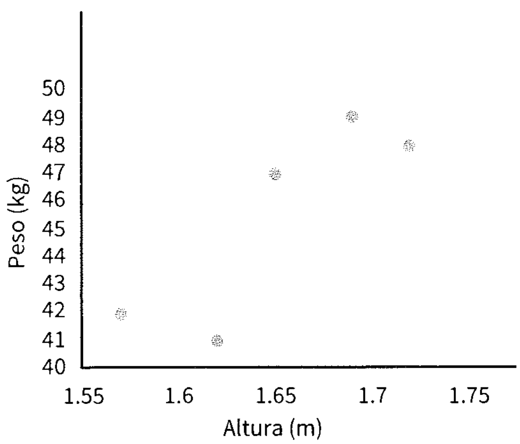
La relación entre las variables será positiva o negativa si la covarianza es positiva o negativa, respectivamente. La relación será neutra si la covarianza es igual a cero.

Altura (m)	Peso (kg)
1.62	41
1.72	48
1.65	47
1.69	49
1.57	42

Veamos un ejemplo de la fórmula paso a paso con la pequeña base de datos de la izquierda. En ella se muestra la altura y peso de cinco mujeres tomadas al azar.

El promedio de la altura es 1.65 m y del peso 45.4 kg. Si utilizamos la segunda fórmula, al sustituir, tenemos que:

$$\begin{aligned}\operatorname{cov}(x,y) &= \tfrac{1}{5}(1.62\cdot 41+1.72\cdot 48+1.65\cdot 47+1.69\cdot 49+1.57\cdot 42)-1.65\cdot 45.4\\ &= \tfrac{1}{5}(66.42+82.56+77.55+82.81+65.94)-74.91\\ &= \tfrac{1}{5}(375.28)-74.91=75.06-74.91=0.15.\end{aligned}$$

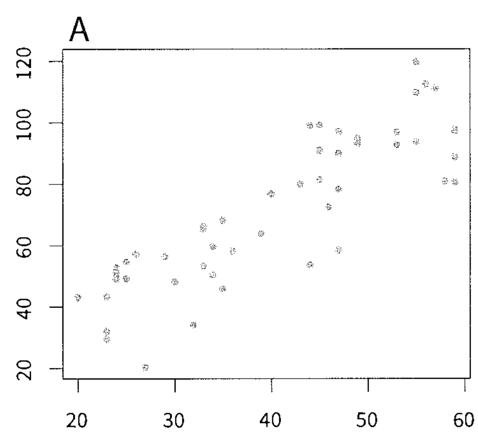


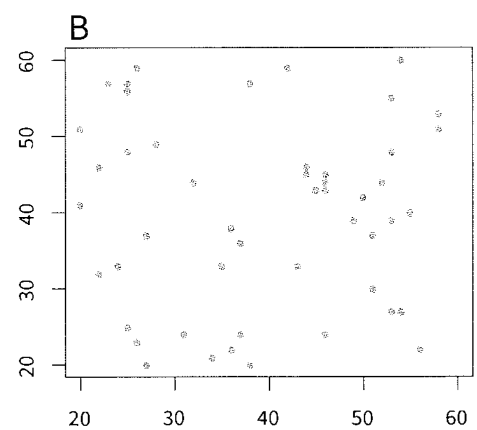
Como la covarianza es positiva, significa que las variables tienen una relación positiva, es decir, a medida que una variable aumenta, la otra también. Específicamente en este ejemplo, significa que a medida que una persona es más alta, también es más pesada.

Si te fijas en la pequeña base de datos, puedes ver que los valores de altura más bajos se relacionan con los valores de peso más bajos y los de altura elevada, con los pesos más altos. Usando un gráfico de dispersión como el de la izquierda, resulta evidente esta relación.

Coeficiente de correlación de Pearson

Observa los siguientes gráficos de dispersión.





Los gráficos representan variables inventadas.

En el gráfico A, la covarianza entre las variables es de 260.69 y en el gráfico B es de 1.98, esto indica que en ambos casos, la relación entre las variables es positiva. Lo cual significa que conforme una variable aumenta, la otra también lo hace. Sin embargo, es evidente que en el caso B la relación no es tan fuerte como en el caso A.

Para medir la "fuerza" o la "intensidad" de la relación que hay entre dos variables, existen los **coeficientes de correlación**. Uno de los más populares es el de Pearson, cuya fórmula es:

$$r(x,y)=\frac{\operatorname{cov}(x,y)}{s_x s_y}.$$

Donde cov(x, y) es la covarianza de las variables y s la desviación estándar de cada variable. El valor del coeficiente estará siempre entre −1 y 1.

Al igual que la covarianza, si el coeficiente es positivo, significa que la relación entre las variables es positiva; si es negativo, quiere decir que la relación es negativa.

Lo que distingue al coeficiente de correlación de Pearson respecto a la covarianza es que entre más cercano esté a −1 y 1, más "fuerte" será la relación de las variables.

Por ejemplo, el coeficiente de correlación entre las variables del gráfico A es de 0.86, mientras que el de las variables en el gráfico B es de 0.01. Como regla de dedo, se considera que dos variables están correlacionadas de manera positiva si su coeficiente de correlación es mayor que 0.6, o de manera negativa si es menor que −0.6.

Veamos un ejemplo paso a paso usando las variables de altura y peso de la página anterior.

Ejemplo

Considerando que la desviación estándar de la altura es 0.05 y del peso, 3.26 (usando la desviación estándar poblacional), tenemos que:

$$r(x,y)=\frac{0.14}{0.05\cdot 3.26}=\frac{0.14}{0.16}=0.87.$$

Esto significa que no sólo la relación entre altura y peso es positiva, sino que existe una fuerte correlación, pues 0.87 está muy cerca del 1.

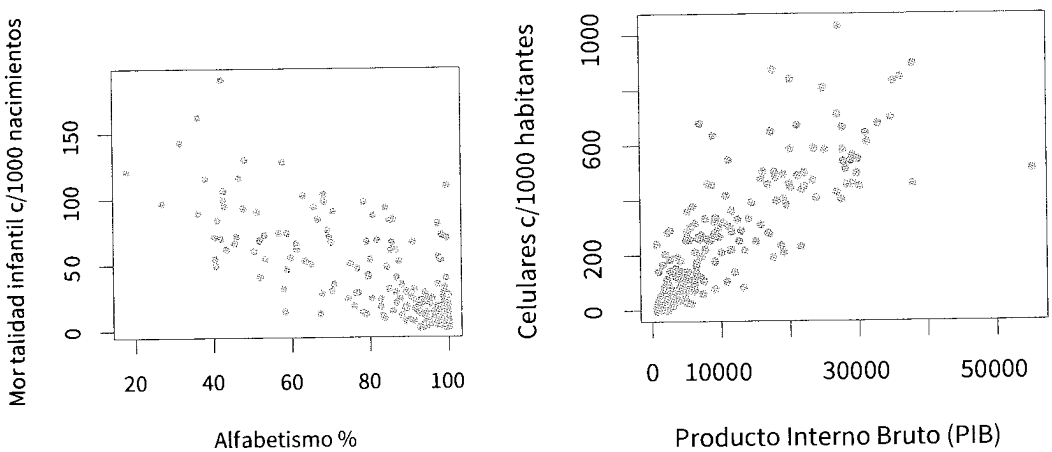
En un gráfico de dispersión, un coeficiente de correlación de 1, se vería como un montón de puntos perfectamente alineados con una pendiente positiva (como una rampa que sube); mientras que un coeficiente de correlación de −1 se vería como una línea de puntos con pendiente negativa (como una rampa que va hacia abajo).

En estadística inferencial, los coeficientes de correlación se usan para inferir causalidad entre una variable y otra. Sin embargo, la inferencia requiere de una investigación bibliográfica y justificación teórica. Ejemplos de correlaciones fuertes y absurdas como: los divorcios en el estado de Maine (USA) y el consumo de margarina, tienen una correlación de 0.99; o el número de doctorados en ingeniería y el consumo de queso mozzarella, tienen una correlación positiva de 0.95; y otros ejemplos los puedes ver en la página: bkmrt.com/MVbqz8.

Observa que los gráficos en el enlace de las correlaciones absurdas no son de dispersión sino dos gráficos de línea sobrepuestos, es una forma poco convencional de mostrar una correlación, en este caso se hizo así porque las variables están graficadas en el tiempo.

Ejemplo

Retomemos los gráficos de dispersión del segundo parcial sobre la relación entre las variables alfabetismo y mortalidad infantil, y entre las variables producto interno bruto y la cantidad de celulares por cada mil habitantes.



El coeficiente de correlación de Pearson entre las variables alfabetismo y mortalidad infantil es −0.76, es decir que su relación es negativa y fuerte (pues el valor es menor a −0.6).

Esto significa que los países que tienen mayor alfabetismo tienen menor mortalidad infantil, aunque el coeficiente por sí sólo no nos permite inferir que saber leer causa que menos niños se mueran, o viceversa, que mueran pocos niños cause que la gente sepa leer.

Se requiere una investigación que permita inferir esa causalidad. En este caso, lo más probable es que existan otras variables que afectan simultáneamente a estas dos variables, como por ejemplo programas educativos.

En el segundo caso, el coeficiente es de 0.83, esto es que la relación entre el PIB y la cantidad de celulares que tienen la gente es positiva y fuerte (el valor es mayor a 0.6). Esto significa que los países con mayor riqueza tienen más celulares por cada mil habitantes.

Es probable que en este caso sí exista causalidad entre las dos variables: suena consecuente que un país con recursos cause que la gente tenga mayor poder adquisitivo y éste se refleje en la cantidad de celulares. Sin embargo, el coeficiente por sí solo no permite hacer la inferencia, se requiere de un planteamiento lógico producto de una investigación.

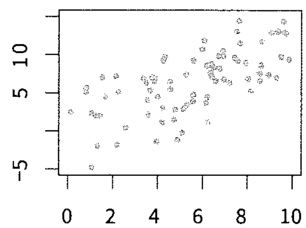
Es muy importante recordar que la estadística es una herramienta de análisis, no piensa por sí sola.

Actividad de aprendizaje 7

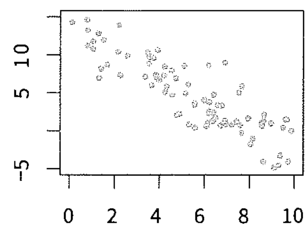
En cada inciso, realiza lo que se pide.

1. Relaciona los siguientes coeficientes de correlación con su respectivo gráfico de dispersión: 1, 0.65, −0.87, 1, −0.30, −1, 0.94, 0, 0.

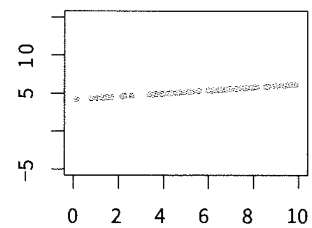
r = ____________



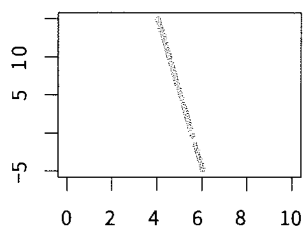
r = ____________



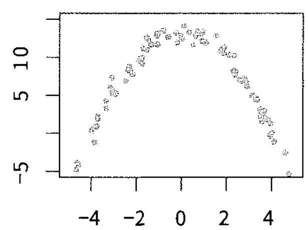
r = ____________



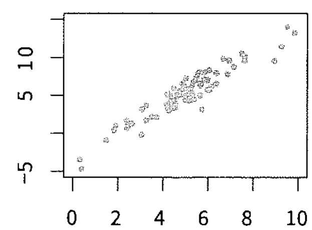
r = ____________



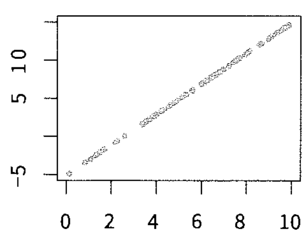
r = ____________



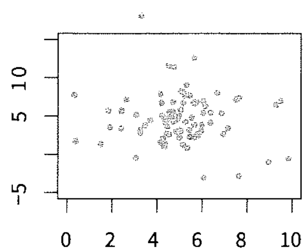
r = ____________



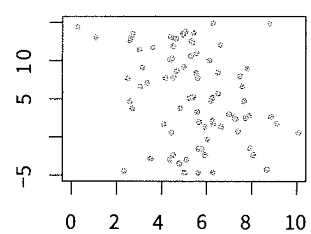
r = ____________



r = ____________



r = ____________



2. Investiga en Excel cómo calcular la covarianza y el coeficiente de correlación de Pearson.

3. Utiliza las funciones en Excel para calcular la covarianza y el coeficiente de correlación entre las diferentes variables de la base de datos de las mariposas de la página 94.

   Completa el cuadro de la siguiente página y contesta las preguntas que se presentan.

Estadístico	Días y Tamaño	Tamaño y Temperatura	Días y Temperatura
Covarianza			
Coeficiente de correlación			
Conclusión			

a. ¿Qué variable está estrechamente relacionada con la duración de la fase de pupa de las mariposas?

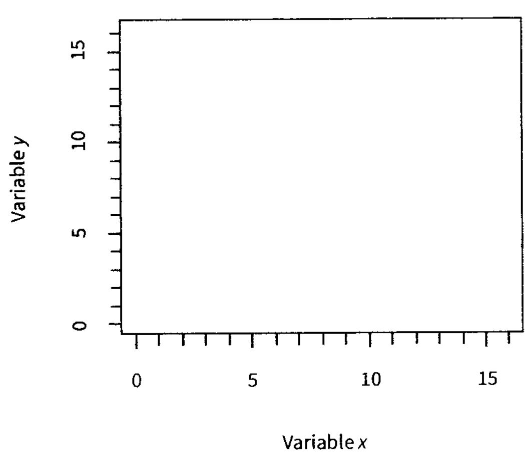
b. ¿Qué correlación es totalmente irrelevante para la pregunta del estudio?

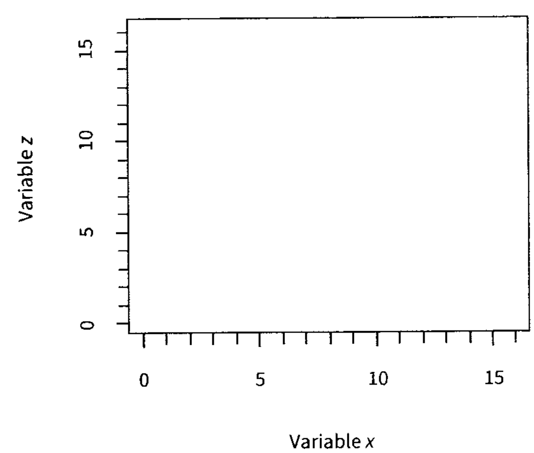
Actividad de aprendizaje 8

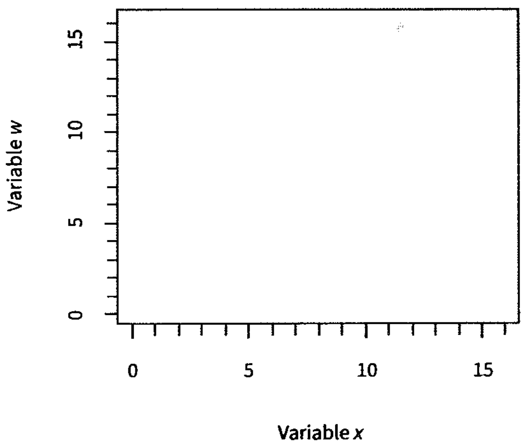
En cada inciso, realiza lo que se pide.

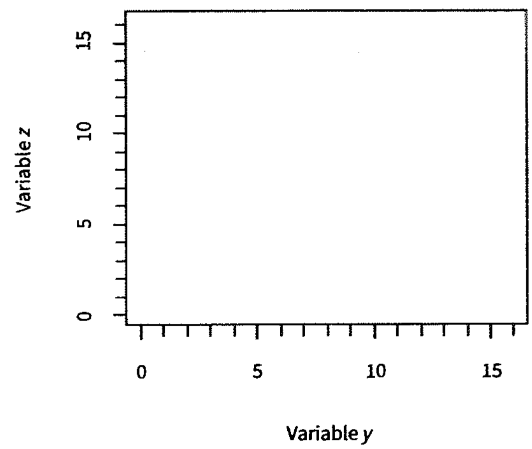
x	y	z	w
6	8	4	9
10	7	7	17
16	16	5	12
8	12	11	17
4	1	12	14
2	3	10	12
12	10	10	9
14	9	7	14

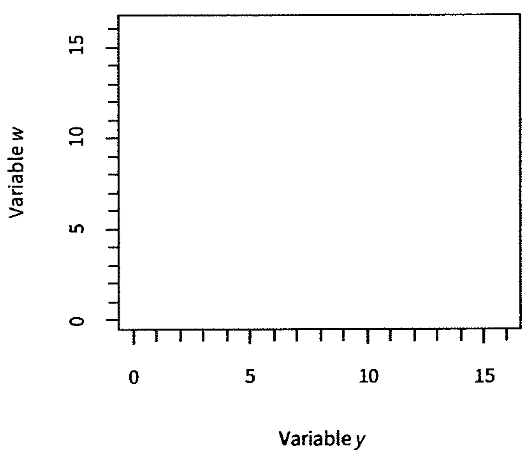
1. Completa los gráficos de dispersión de abajo y de la siguiente página utilizando la pequeña base de datos que se muestra a la izquierda.
2. A ojo de buen cubero, anota en cada gráfico si te parece que la relación entre las variables es positiva, negativa o neutra.
3. Calcula el coeficiente de correlación de Pearson para cada combinación de variables y anótalo en su respectivo gráfico de dispersión.

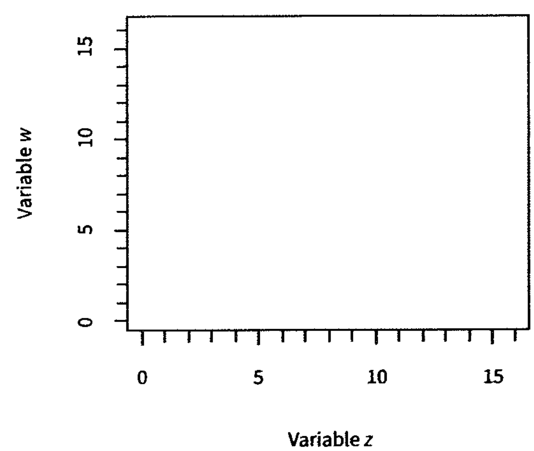












4. ¿Entre qué variables no existe ninguna relación?
5. ¿Entre qué variables a medida que disminuye una de las dos la otra también lo hace?
6. ¿Qué par de variables tiene la correlación negativa más fuerte?
7. ¿Entre qué variables a medida que una variable disminuye la otra aumenta?
8. ¿Qué par de variables tiene la correlación positiva más fuerte?

Análisis de la información y toma de decisiones

La estadística es una herramienta que se utiliza para resumir información, analizarla y tomar decisiones. A pesar de que la estadística que has visto hasta ahora es la más básica, es suficiente para llegar a algunas conclusiones.

En este último apartado, utilizaremos las distintas herramientas estadísticas vistas a lo largo del curso aplicadas en un ejemplo muy sencillo.

Ejemplo

Veintidós alumnos del mismo grupo desean saber qué pueden hacer para mejorar sus calificaciones.

Una parte del grupo cree que las calificaciones mejorarían si los estudiantes dedicaran menos tiempo a los videojuegos, redes sociales e internet; el resto del grupo cree que las calificaciones mejorarían si los estudiantes dedicaran más tiempo al estudio en sus casas.

Para resolver esta duda, los estudiantes crearon la base de datos de abajo.

Promedio	Horas de estudio	Horas de distracción
9.2	1.9	1.0
7.2	0.6	2.1
8.9	1.9	1.5
9.9	2.9	2.6
9.0	2.2	1.7
5.8	0	2.8
8.5	2.1	3.0
7.5	0.5	0.6
6.3	0.3	0.4
9.6	2.3	2.5
9.1	2.1	0.4

Promedio	Horas de estudio	Horas de distracción
9.5	3.4	1.1
7.0	0.6	2.1
7.3	0.8	2.2
5.0	0.0	4.0
9.7	2.9	2.4
7.3	1.5	3.9
7.6	0.8	2.9
7.5	0.6	2.6
6.3	0.4	2.6
8.5	1.4	2.4
8.1	1.6	0.1

Donde la variable "horas de estudio" es el promedio diario de horas que cada estudiante dedicó al estudio en casa las últimas dos semanas; la variable "horas de distracción" se midió igual, pero con horas dedicadas a videojuegos, redes sociales e internet.

En el cuadro de abajo puedes ver los estadísticos descriptivos básicos de cada variable. En ellos hay un resumen de la información que puede darte una idea de aspectos generales.

	Promedio	Horas de estudio	Horas de distracción
Media	7.95	1.40	2.04
Mediana	7.85	1.45	2.30
Varianza	1.78	0.96	1.12
Desviación estándar	1.33	0.98	1.06

Hacia PLANEA

Subraya la respuesta correcta.

La media de alturas de un grupo donde 6 personas miden 1.65; 4 miden 1.68; 5 miden 1.72; 3 miden 1.75, 3 miden 1.54 y 2 miden 1.58 es:

a. 1.66
b. 1.68
c. 1.70
d. 1.72

	Promedio	Horas de estudio	Horas de distracción
Curtosis	−0.62	−0.95	−0.53
Simetría	−0.37	0.31	−0.18
Rango	5.0 − 9.9	0.0 − 3.4	0.1 − 4.0

Del cuadro notamos que el promedio del grupo es de 7.95; la distribución de las calificaciones se asimétrica negativa, lo que significa que a pesar de que los promedios van de 5 a 9.9, la mayoría está más cerca del 9.9 que del 5.

Con la curtosis negativa podemos concluir que la distribución es platicúrtica, esto es, que los promedios de los alumnos están distribuidos homogéneamente a lo largo del rango. El mismo tipo de observaciones las puedes hacer con los estadísticos de las otras dos variables.

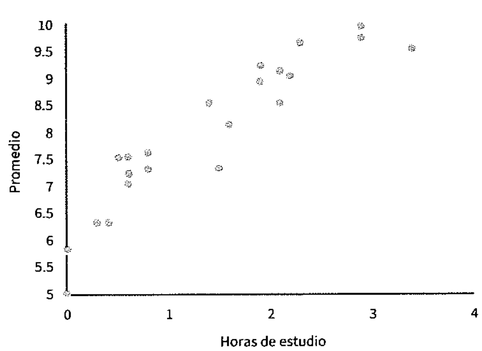
A estas alturas, la varianza y desviación estándar de los promedios no dicen mucho por sí solas y no son comparables con la varianza y desviación estándar de las otras dos variables, pues están en distintas unidades (calificaciones vs horas).

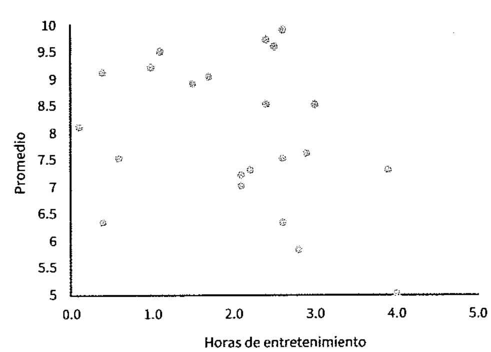
Sin embargo, las medidas de dispersión entre las variables que están en horas sí son comparables. En realidad, la diferencia es mínima (0.08 horas en la desviación estándar), pero aun así se observa que la dispersión de los datos alrededor de la media es mayor en las horas de entretenimiento que en las horas de estudio.

Esto significa que hay más estudiantes cerca de dedicar 1.4 horas de estudio diario, que estudiantes cerca de dedicar 2.04 horas a entretenimiento diario.

Como podrás darte cuenta, los estadísticos descriptivos no están contestando la pregunta original de los estudiantes (¿cómo mejorar el promedio? ¿estudiando más o jugando menos?).

Para atender la pregunta, será necesario correlacionar la variable de interés (promedio) con las variables que los estudiantes consideran que afectan a la variable de interés (horas de estudio y horas de entretenimiento). Observa los gráficos de dispersión entre las variables relacionadas.





Hacia PLANEA

4 **Subraya la respuesta correcta.**

La mediana de los pesos de un grupo donde 3 personas pesan 65 kg; 8 pesan 68 kg; 5 pesan 72 kg; 3 pesan 80 kg, 2 pesan 54 kg y 4 pesan 90 kg es:

a. 66
b. 68
c. 70
d. 72

El coeficiente de correlación entre el promedio y las horas de estudio es de 0.93 (positiva y fuerte), significa que a medida que se incrementan las horas de estudio el promedio también es mayor.

Con respecto a la correlación entre e el promedio y las horas de entretenimiento, el coeficiente de es de -0.31 (negativa y débil), lo que significa que a medida que aumenta la cantidad de horas dedicadas al entretenimiento los promedios son menores, sin embargo, la correlación no es tan fuerte como la de las horas de estudio con el promedio.

En este caso totalmente inventado, se puede concluir que, para mejorar sus calificaciones, sería más efectivo que los estudiantes dedicaran más horas al estudio en casa, que disminuyendo las horas de distracción.

Esta conclusión se puede decir de distintas formas, por ejemplo, que no es necesario que un estudiante deje de jugar videojuegos para mejorar sus calificaciones, bastaría con que aprenda a administrar su tiempo y dedique más horas al estudio.

Es importante hacer énfasis en que esta conclusión es una generalización del patrón a gran escala. Es probable que para algunos estudiantes, en particular, fuera necesario disminuir las horas de entretenimiento además de aumentar las de estudio.

Por último, nota que es posible correlacionar las horas de entretenimiento con las horas de estudio, sin embargo, la pregunta que motivó a este análisis no requiere que se haga dicha correlación.

Si se quisiera saber si los alumnos que más estudian son lo que menos dedican horas a entretenerse con otras actividades, entonces sí sería necesaria esa correlación.

Eventos deterministas y aleatorios

Es verdad que hay variables que repercuten entre sí, pero no siempre los fenómenos obedecen a causas bien identificadas, sino que ocurren por azar.

En este vídeo verás la diferencia entre eventos deterministas y aleatorios. Esto te ayudará a comprender qué sucesos puedes controlar y prever y cuáles simplemente ocurren.

Actividad de aprendizaje 9

En cada inciso, realiza lo que se pide.

1. Retoma la base de datos que hiciste en equipo en la página 29. Selecciona cinco variables de la base de datos que sean cuantitativas y realiza las siguientes actividades.

   a. Construye un histograma para cada una de las cinco variables.

   b. Anota en el siguiente cuadro las variables elegidas y describe cada una utilizando los estadísticos descriptivos señalados:

Estadístico	V_1	V_2	V_3	V_4	V_5
Variable elegida					
Media					
Mediana					
Varianza					
Desviación estándar					
Curtosis					
Simetría					
Rango					
Q_1					
Q_2					
Q_3					

   c. Marca en cada histograma: la media, mediana, Q_1 y Q_3.

2. En la base de datos que encontrarás escaneando el código QR de la derecha, hay doce variables con información obtenida del INEGI de los 32 estados de la República Mexicana, las variables son:

   I. Estado: nombre del estado.

   II. Rezago social: índice indicador de carencias sociales (educación, salud, servicios básicos, calidad y espacios en la vivienda).

   III. Exportaciones: millones de dólares anuales exportados.

   IV. Población: cantidad de habitantes.

Hacia PLANEA

Subraya la respuesta correcta.

El promedio de un estudiante es 8.25. Cursó ocho materias, en dos tuvo 7, en dos sacó 8 y en las otras cuatro tuvo la misma calificación, ¿cuánto sacó en las cuatro materias restantes?

a. 6

b. 7

c. 8

d. 9

Suma anécdota

En sus inicios, la estadística no estaba definida como un campo disciplinario de las matemáticas, sino que era un conjunto de técnicas y procedimientos matemáticos que permitían el estudio de fenómenos naturales y el tratamiento de datos.

El estudio de Darwin generó un ambiente propicio para hablar sobre la herencia genética y naturalmente, se buscó la forma de estudiar los rasgos adquiridos por los descendientes de las especies, incluida la humana.

Karl Pearson fue un prominente científico, matemático y pensador británico, que estableció la disciplina de la estadística matemática. Desarrolló una intensa investigación sobre la aplicación de los métodos estadísticos a la biología y fue el fundador de la bioestadística.

Pearson conoció a Francis Galton, primo de Charles Darwin, y llegó a ser su discípulo. Al morir Galton, Pearson se aventuró a producir su biografía, donde exaltaba su vida, su trabajo y su herencia, seguro de que Galton, en vez de Charles Darwin, sería recordado como el nieto más prodigioso de Erasmus Darwin. Galton también se interesó en aspectos de la evolución, pero trabajó más desde las matemáticas.

En 1911, Pearson fundó el primer departamento de estadística en la Universidad de Londres, donde fue profesor e inauguró el laboratorio biométrico. Reformuló la "ley de herencia ancestral" de Galton, según la cual, al cabo de algunas generaciones sometidas a la selección, una población se multiplicaría de acuerdo con el carácter elegido. Pearson sostenía que el medio ambiente tenía poco que ver con el desarrollo de cualidades mentales y emocionales.

La idea que Pearson desarrolló ofreció un sustento científico al discurso sobre la superioridad de razas y sirvió como aliciente para la búsqueda de la creación de una raza dominante, a esto se le conoce como eugenesia. Finalmente, con estudios posteriores, se llegó a la conclusión de que en el desarrollo de las personas influyen, además de la herencia, factores ambientales y sociales que hacen a las personas distintas entre sí, aunque todavía se utiliza a la estadística, en particular, como un pretexto para sustentar ideologías peligrosas o para manipular la opinión pública.

Reflexiona sobre lo leído y contesta las preguntas con tus compañeros y el docente.

1. ¿En qué medida un argumento científico se vuelve peligroso para un sector poblacional?
2. ¿Los resultados estadísticos siempre son contundentes?
3. ¿Qué significa que 8 de cada 10 gatos prefieran una marca de alimento?
4. ¿Qué se necesita para que un resultado estadístico sea confiable?
5. ¿Cómo puedes evitar caer en manipulaciones con los datos que se te presentan?

Proyecto integrador

Estudio de la población.

En este proyecto, estudiarás los datos obtenidos en el proyecto del parcial anterior (p. 96), realizando las gráficas correspondientes para obtener conclusiones de la población.

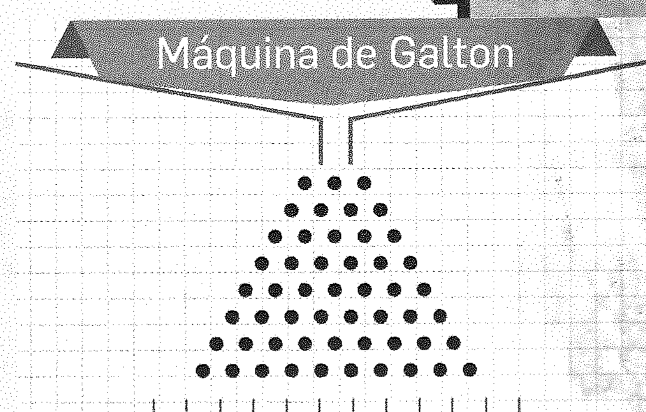
Realiza, con el grupo y tu docente lo que se indica y contesta las preguntas.

1. Elijan cinco variables de las obtenidas en el proyecto del segundo parcial, en caso de que falten variables, decidan cuáles incluir.
2. Describan la población (los alumnos del grupo) con los datos recabados.
3. Extraigan las medidas de tendencia central y las medidas de dispersión.
4. Grafiquen los datos obtenidos y estúdienlos, con las medidas de tendencia central y de dispersión, describan visualmente la distribución de los datos obtenidos.
5. Describan analíticamente la gráfica, anoten las medidas de sesgo y forma, digan cómo es la dispersión de los datos y cómo son las medidas de tendencia central con respecto a los datos obtenidos.
6. Establezcan cuáles variables de las que indagaron tienen alguna correlación y si esta es fuerte o débil, comenten por qué piensan que se relacionan y cómo repercuten entre sí.
7. Describan el grupo con base en los resultados realizados y discutan al respecto.

Máquina de Galton

Las distribuciones estadísticas presentan ciertas regularidades y pueden ser modeladas matemáticamente, lo que hace posible su estudio.

Una forma experimental de visualizar una distribución aleatoria es con la máquina de Galton, en ella se observa cómo la caída de cuerpos se distribuyen de forma particular.



V. Actividad económica: cantidad de habitantes mayores de 15 años económicamente activos.

VI. Internet: porcentaje de casas con internet.

VII. Piso de tierra: cantidad de hogares con piso de tierra

VIII. Analfabetismo: número de habitantes analfabetas mayores de 15 años.

IX. Corrupción: número de experiencias de corrupción por cada 100 000 habitantes en la realización de trámites, pagos, solicitudes de servicios u otros realizados personalmente ante servidores públicos por usuarios de 18 años y más.

X. Tuberculosis: casos de tuberculosis cada 100 000 habitantes.

XI. Jueces: número de jueces por cada 100 000 habitantes.

XII. Confianza en policía: porcentaje de personas que confían en la policía.

a. Elige cinco variables cuantitativas y completa el siguiente cuadro:

Estadístico	V_1	V_2	V_3	V_4	V_5
Variable elegida					
Media					
Mediana					
Varianza					
Desviación estándar					
Curtosis					
Simetría					
Rango					
Q_1					
Q_2					
Q_3					

b. Elige cinco combinaciones de variables cuantitativas que crees que pudieran estar relacionadas. Escribe en tu cuaderno cada par de variables y la razón o razones por las cuáles consideras que esas variables están correlacionadas.

c. Elabora los gráficos de dispersión de las cinco combinaciones y anota en cada uno tu predicción de cómo será la relación entre las dos variables (positiva, negativa o neutra).

d. Calcula el coeficiente de correlación de Pearson para cada una de las cinco combinaciones que elegiste y anótalo en cada gráfico de dispersión.

e. Anota tus conclusiones, ¿acertaste en todas las predicciones? En las que no acertaste ¿puedes imaginar a qué se debió?

f. Si quieres consultar más información de los estados visita la página: bkmrt.com/qMINJ7.

Actividad de aprendizaje 10

1. Piensa en una variable cuantitativa de tu interés e investiga sobre ella y busca al menos tres variables cuantitativas que, según tu investigación, consideres que están relacionadas con la variable que pensaste primero.

2. Investiga cifras reales que incluyan al menos las cuatro variables y construye la base de datos correspondiente con al menos diez datos (diez renglones).

3. Construye, con los métodos propuestos, los gráficos de dispersión que relacionan a la variable de tu interés con las otras tres.

4. A cada uno de los gráficos de dispersión, anótales tu predicción de cómo es la relación entre cada par de variables (positiva, negativa o neutra).

5. Calcula los coeficientes de correlación de las tres combinaciones de variables y anota el resultado en cada gráfico.

6. Responde las siguientes preguntas:

   a. ¿Los coeficientes de correlación coincidieron con la investigación previa?

   b. Con los coeficientes que sí coincidieron con la predicción basada en la investigación, concluye respondiendo:

      - ¿Cuáles son las implicaciones de tus resultados? Es decir, ¿qué significa esta correlación?
      - ¿Consideras que la información es útil para tomar alguna decisión?

   c. Con los coeficientes que no coincidieron con la predicción basada en la investigación, concluye respondiendo:

      - ¿Por qué crees que no resultó así?
      - ¿La información de la investigación está mal tomada o se tiene una percepción errónea de la realidad?
      - ¿Consideras que la información es útil para tomar alguna decisión?

Suma anécdota

En sus inicios, la estadística no estaba definida como un campo disciplinario de las matemáticas, sino que era un conjunto de técnicas y procedimientos matemáticos que permitían el estudio de fenómenos naturales y el tratamiento de datos.

El estudio de Darwin generó un ambiente propicio para hablar sobre la herencia genética y naturalmente, se buscó la forma de estudiar los rasgos adquiridos por los descendientes de las especies, incluida la humana.

Karl Pearson fue un prominente científico, matemático y pensador británico, que estableció la disciplina de la estadística matemática. Desarrolló una intensa investigación sobre la aplicación de los métodos estadísticos a la biología y fue el fundador de la bioestadística.

Pearson conoció a Francis Galton, primo de Charles Darwin, y llegó a ser su discípulo. Al morir Galton, Pearson se aventuró a producir su biografía, donde exaltaba su vida, su trabajo y su herencia, seguro de que Galton, en vez de Charles Darwin, sería recordado como el nieto más prodigioso de Erasmus Darwin. Galton también se interesó en aspectos de la evolución, pero trabajó más desde las matemáticas.

En 1911, Pearson fundó el primer departamento de estadística en la Universidad de Londres, donde fue profesor e inauguró el laboratorio biométrico. Reformuló la "ley de herencia ancestral" de Galton, según la cual, al cabo de algunas generaciones sometidas a la selección, una población se multiplicaría de acuerdo con el carácter elegido. Pearson sostenía que el medio ambiente tenía poco que ver con el desarrollo de cualidades mentales y emocionales.

La idea que Pearson desarrolló ofreció un sustento científico al discurso sobre la superioridad de razas y sirvió como aliciente para la búsqueda de la creación de una raza dominante, a esto se le conoce como eugenesia. Finalmente, con estudios posteriores, se llegó a la conclusión de que en el desarrollo de las personas influyen, además de la herencia, factores ambientales y sociales que hacen a las personas distintas entre sí, aunque todavía se utiliza a la estadística, en particular, como un pretexto para sustentar ideologías peligrosas o para manipular la opinión pública.

Reflexiona sobre lo leído y **contesta las preguntas con tus compañeros y el docente.**

1. ¿En qué medida un argumento científico se vuelve peligroso para un sector poblacional?
2. ¿Los resultados estadísticos siempre son contundentes?
3. ¿Qué significa que 8 de cada 10 gatos prefieran una marca de alimento?
4. ¿Qué se necesita para que un resultado estadístico sea confiable?
5. ¿Cómo puedes evitar caer en manipulaciones con los datos que se te presentan?

Proyecto integrador

Estudio de la población.

En este proyecto, estudiarás los datos obtenidos en el proyecto del parcial anterior (p. 96), realizando las gráficas correspondientes para obtener conclusiones de la población.

Realiza, con el grupo y tu docente lo que se indica y contesta las preguntas.

1. Elijan cinco variables de las obtenidas en el proyecto del segundo parcial, en caso de que falten variables, decidan cuáles incluir.
2. Describan la población (los alumnos del grupo) con los datos recabados.
3. Extraigan las medidas de tendencia central y las medidas de dispersión.
4. Grafiquen los datos obtenidos y estúdienlos, con las medidas de tendencia central y de dispersión, describan visualmente la distribución de los datos obtenidos.
5. Describan analíticamente la gráfica, anoten las medidas de sesgo y forma, digan cómo es la dispersión de los datos y cómo son las medidas de tendencia central con respecto a los datos obtenidos.
6. Establezcan cuáles variables de las que indagaron tienen alguna correlación y si esta es fuerte o débil, comenten por qué piensan que se relacionan y cómo repercuten entre sí.
7. Describan el grupo con base en los resultados realizados y discutan al respecto.

Máquina de Galton

Las distribuciones estadísticas presentan ciertas regularidades y pueden ser modeladas matemáticamente, lo que hace posible su estudio.

Una forma experimental de visualizar una distribución aleatoria es con la máquina de Galton, en ella se observa cómo la caída de cuerpos se distribuyen de forma particular.