|  |  |
| --- | --- |
| Título del guion | **La estadística** |
| Código del guion | MA\_08\_11\_CO |
| Descripción | A través de la estadística podemos estudiar y analizar el comportamiento de una población o una serie de eventos haciendo una recolección y análisis de los datos. En este tema vamos a ver cómo podemos aplicar la estadística en el estudio de diferentes situaciones. |

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_10\_CO\_IMG01 |
| **Descripción** | C:\Users\EdgarJosué\Documents\GitHub\RecursosGenerales\graficos\iconosGuiones\Matematicas\MTP_09_11.png |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | Icono de guion: MTP\_09\_11 |
| **Pie de imagen** |  |
|  |  |

[SECCIÓN 1] **1 La estadística y sus elementos**

La estadística en una disciplina de gran importancia debido a sus múltiples aplicaciones en los diferentes ámbitos del mundo, ya sea en la ciencia, la economía, las ciencias sociales, la industria y cualquier situación en la que nos debamos enfrentar a la recolección y análisis de datos.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG01 |
| **Descripción** | Businesspeople, man and woman, talking, discussing in meeting room. With chart and graph statistics background. Diverse, multi-ethnic, flat design.  - stock vector |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | 301469477 |
| **Pie de imagen** | A través de la estadística las empresas toman decisiones sobre el rumbo de sus compañías. |

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La estadística** |
| **Contenido** | Es la disciplina que se encarga organización y análisis de la información a través de la recolección de datos para hacer inferencias o predicciones sobre el comportamiento de algún suceso o evento. |

[SECCIÓN 2] **1.1 La población y la muestra**

Las estadísticas no tienen sentido si no se consideran o se relacionan dentro del contexto con que se trabajan. Por lo tanto es necesario entender los conceptos de **población** y de **muestra**.

**Población:** Lo definimos como el conjunto total de individuos, objetos o medidas que poseen características comunes observables en un lugar y en un momento determinado.

**Muestra:** La definimos como un subconjunto representativo de la población que conserva las mismas características.

Es decir que, la población es el objeto de estudio y la muestra son los elementos que se seleccionan de forma aleatoria para realizar dicho estudio, esto se debe a que la estadística maneja grandes cantidades de datos y realizar un estudio sobre toda una población no es conveniente por tiempos costos y recursos.

Por ejemplo si se realiza un estudio sobre sobre las preferencias musicales de los estudiantes de grado octavo de una institución educativa, no se entrevistan a todos los estudiantes, se puede seleccionar un curso en particular que sea representativo de todos los estudiantes de grado octavo.

En este caso la **población** son todos los estudiantes de grado octavo y la **muestra** es el curso seleccionado de grado octavo

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG02 |
| **Descripción** | http://thumb9.shutterstock.com/display_pic_with_logo/563656/234415222/stock-vector-vector-flat-illustration-of-society-members-with-men-and-women-population-social-network-concept-234415222.jpg |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | La muestra siempre debe ser representativa y caracterizar a toda la población que se estudia. |

[SECCIÓN 2] **1.2 Las variables y su clasificación**

El proceso de recolección de datos en los estudios estadísticos se puede hacer de diversas maneras, a través de la observación en la repetición de un suceso, a través de la entrevista, o través de la investigación en acción, entre otros, y los datos que se obtienen pueden ser **cualitativos** estos se refieren a cualidades o características del objeto de estudio, por ejemplo el color de ojos de una persona, la calidad del servicio prestado en un punto de atención y es **cuantitativo** cuando el dato recolectado es medible mediante un número, por ejemplo la edad de un persona, el tiempo de reacción de un grupo de emergencias, la cantidad de llamadas que recibe la línea de emergencias en un día.

Las **variables cualitativas** pueden ser **nominales** si describen una característica como el color de preferencia en la elección de un auto u **ordinales** si clasifican el dato por ejemplo el nivel de satisfacción en la prestación de un servicio por parte de una compañía.

Mientras que las **variables cuantitativas** pueden ser **continúas** si el dato recolectado es un número real, por ejemplo el tiempo que marca el cronometro en una carrera de fórmula 1 y **discreta** si el dato recolectado es un número entero, por ejemplo la edad de una persona.

Observa el siguiente ejemplo para determinar los conceptos vistos hasta el momento.

Un grupo ambientalista ha decidió realizar un estudio acerca de la vegetación que existe en un pequeño bosque cerca de la ciudad, para ello deciden determinar qué clase de plantas se encuentran, y cuál es su edad y tamaño, para ello delimitan 2 hectáreas del bosque donde se encuentra gran variedad de plantas.

En esta situación ¿cuál es la población y la muestra? ¿Cuáles son las variables del estudio?

**Población:** El bosque y las plantas que se encuentran en él.

**Muestra**: Las dos hectáreas que se han delimitado del bosque.

**Variable cualitativa nominal:** los tipos de plantas que se encuentran en el bosque.

**Variable cuantitativa discreta:** Edad de cada planta, cantidad.

**Variable cuantitativa continua**: Tamaño de cada planta

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG03 |
| **Descripción** |  |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | Clasificación de las variables estadísticas. |

[SECCIÓN 2] **1.3 Las tablas de frecuencias.**

Para poder organizar y analizar la información de los datos que se recolectan es necesario acudir a tablas que permitan observar de una forma resumida los datos recolectados.

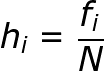
Si el dato que se analiza corresponde a una sola variable usaremos una **tabla de frecuencias** que se compone de filas y columnas.

En cada columna vamos a encontrar:

**Variable**: Corresponde al dato que se ha recolectado se representa como ***Xi***.

**Frecuencia absoluta:** Corresponde a la cantidad de veces que se repite un dato se representa como ***fi***.

**Frecuencia relativa**, es la **razón** entre la frecuencia absoluta ***fi*** y el total de datos **N**,se representa mediante ***hi*** y se calcula como:



La frecuencia relativa también la podemos expresar como un porcentaje multiplicando cada ***hi*** por 100:

**Frecuencia absoluta acumulada:** Corresponde a la suma de las frecuencias absolutas se representa como **Fi**.

**Fi = *f1 + f2 + f3 + . . .+ fi***

**Frecuencia relativa acumulada**: Corresponde a la suma de las frecuencias relativas se representa como **Hi**.

**Hi = *h1 + h2 + h3 + . . .+ hi***

La frecuencia relativa acumulada también la podemos expresar como un porcentaje multiplicando cada ***Hi*** por 100:

Observemos el siguiente ejemplo:

Ante el inminente cambio climático, el instituto meteorológico de una ciudad ha registrado la temperatura media de su ciudad durante los 30 primeros días del año como se observa en la siguiente tabla.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Día | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Temperatura °C | 22 | 25 | 22 | 26 | 24 | 23 | 22 | 25 | 27 | 26 | 25 | 24 | 22 | 24 | 25 |
| Día | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| Temperatura °C | 24 | 27 | 26 | 22 | 25 | 25 | 23 | 23 | 23 | 24 | 26 | 22 | 24 | 25 | 25 |

Para resumir la información y analizar los datos recolectados usaremos la tabla de frecuencias. En este caso **N = 30** datos

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Temperatura | ***fi*** | ***hi*** | % | **Fi** | **Hi** | % |
| 22 °C | 6 | 6/30 = 0.2 | 20% | 6 | 6/30 = 0.2 | 20% |
| 23 °C | 4 | 4/30 = 0,13 | 13% | 10 | 10/30 = 0,33 | 33% |
| 24 °C | 6 | 6/30 = 0,2 | 20% | 16 | 16/30 = 0,53 | 53% |
| 25 °C | 8 | 8/30 = 0,27 | 27% | 24 | 24/30 = 0,8 | 80% |
| 26 °C | 4 | 4/30 = 0,13 | 13% | 28 | 28/30 = 0,93 | 93% |
| 27 °C | 2 | 2/30 = 0,07 | 7% | 30 | 30/30 = 1 | 100% |

Como observas la tabla de frecuencias resume la información y nos permite tener un panorama acerca del comportamiento de la temperatura de la ciudad en la que se recolectaron los datos. Por ejemplo la temperatura que más se repitió fue la de 25 °C y la que menos ocurrió fue la de 27 °C.

Por otra parte sí el análisis que se pretende realizar requiere la combinación de dos variables usaremos una tabla de **doble entrada** o **tabla de contingencia**

Por ejemplo: la siguiente tabla muestra los datos recogidos a un grupo de 30 estudiantes de grado once en una institución educativa acerca de la clase de carrera universitaria que les gustaría estudiar.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Genero | Carrera | Genero | Carrera | Genero | Carrera |
| M | Ingeniería | F | Medicina | F | Medicina |
| M | Medicina | M | Ingeniería | F | Ingeniería |
| F | Ingeniería | F | Ingeniería | M | Derecho |
| F | Derecho | F | Medicina | M | Ingeniería |
| F | Derecho | M | Derecho | M | Medicina |
| M | Medicina | M | Medicina | F | Derecho |
| M | Ingeniería | F | Ingeniería | M | Derecho |
| F | Medicina | M | Ingeniería | F | Medicina |
| F | Medicina | M | Ingeniería | F | Ingeniería |
| M | Derecho | F | Medicina | F | Medicina |

Para resumir la información y analizar la preferencia de carrera por género construiremos la **tabla de doble entrada**, en la primera columna ubicaos el género y en la primera fila ubicamos las carreras.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Ingeniería | Medicina | Derecho | Total |
| M | 6 | 4 | 4 | 14 |
| F | 5 | 8 | 3 | 16 |
| Total | 11 | 12 | 7 | 30 |

A partir de esta tabla podemos hacer varias observaciones, por ejemplo la carrera de mayor preferencia para los hombres es la Ingeniería y para las mujeres la Medicina, la carrera que menos les gusta es el derecho.

[SECCIÓN 2] **1.4 Consolidación**

Actividades para consolidar lo que has aprendido en esta sección.

[SECCIÓN 1] **2 La representación gráfica de la información estadística. El uso de Excel**

Hasta el momento hemos revisado como organizar y resumir los datos que se recolectan en un estudio estadístico mediante las tablas de frecuencias y de doble entrada, sin embargo al momento de presentar los resultados se debe hacer mediante un medio que sea fácil, en este caso las **gráficas estadísticas.**

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Gráficas Estadísticas** |
| **Contenido** | Las gráficas son la herramienta a través de la cual se hace más fácil presentar los resultados de los estudios estadísticos. |

[SECCIÓN 2] **2.1 El uso de Excel**

Actualmente la tecnología ha revolucionado todo el mundo y el manejo de la información y Excel es una de las herramientas que más se usa para representar la información. A continuación veremos cómo usar este programa para realizar gráficos estadísticos.

Para graficar datos en Excel primero debemos insertar los datos que deseamos en la hoja de cálculo. Observemos el siguiente ejemplo con las preferencias deportivas de un grupo de estudiantes que muestra la siguiente tabla.

|  |  |
| --- | --- |
| Deporte | Estudiantes  Que lo prefieren |
| Futbol | 15 |
| Baloncesto | 10 |
| Béisbol | 8 |
| Voleibol | 12 |

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG04 |
| **Descripción** |  |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | Los datos se insertan en la hoja de cálculo para luego ser graficados. |

Ahora en el menú de inicio escogemos la opción insertar y vamos a la sección de gráficos. Allí tendremos varias opciones para realizar el grafico.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG05 |
| **Descripción** | http://profesores.aulaplaneta.com/DNNPlayerPackages/Package13436/InfoGuion/cuadernoestudio/images_xml/MT_09_11_img6_small.jpg |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | Con la opción insertar, puedes escoger el grafico que necesites dependiendo de la información que pretendas presentar. |

[SECCIÓN 3] **2.1.1 El diagrama de barras**

Para elaborar un diagrama de barras se debe partir de los datos que se resumen en una tabla de frecuencias, cada variable se ubica en un eje horizontal y sus respectivas frecuencias en un eje vertical, luego se construye un rectángulo sobre cada variable en el que la altura de cada uno es el valor de la frecuencia.

Tomemos como ejemplo el caso de las temperaturas, la siguiente tabla nos resume las frecuencias.

|  |  |
| --- | --- |
| Temperatura | ***fi*** |
| 22 °C | 6 |
| 23 °C | 4 |
| 24 °C | 6 |
| 25 °C | 8 |
| 26 °C | 4 |
| 27 °C | 2 |

El respectivo diagrama de barras se representa a continuación.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG06 |
| **Descripción** |  |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | Gráfica del registro de la temperatura media en una ciudad durante los primeros 30 días del año. |

Como observas el diagrama permite al lector tener una visión del comportamiento de los datos que se han resumido en la tabla de frecuencias.

Si los datos que se desean presentar corresponden a los resumidos en una tabla de **doble entrada** el diagrama de barras se puede presentar de dos formas:

La primera usa una misma barra para representar la información de los datos de una variable con respecto a la otra. Y en la segunda se usas barras diferentes para representar los datos de una variable con respeto a la otra.

Observa las gráficas para el caso las carreras universitarias de un grupo de 30 estudiantes.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Ingeniería | Medicina | Derecho |
| M | 6 | 4 | 4 |
| F | 5 | 8 | 3 |

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG07 |
| **Descripción** |  |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | Gráfica de las preferencias profesionales en un grupo de 30 estudiantes discriminado por género. |

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **El diagrama de barras** |
| **Contenido** | El diagrama de barras se emplea para representar gráficamente las frecuencias de las variables cuantitativas y cualitativas. |

[SECCIÓN 3] **2.1.2 El diagrama circular**

El **diagrama circular** es una representación que se emplea para representar no solo la frecuencia de un dato sino su proporción con respecto al total de datos recolectados, y para esto el círculo es un buen elemento de representación, cada frecuencia de cada dato ocupa una porción del mismo.

Se debe tener en cuenta que para representar los datos mediante el diagrama circular se usan los datos de la frecuencia relativa en forma porcentual.

Tomemos el ejemplo de las temperaturas, la siguiente tabla muestra los datos recolectados con su frecuencia absoluta y su frecuencia relativa.

|  |  |  |  |
| --- | --- | --- | --- |
| Temperatura | ***fi*** | ***hi*** | % |
| 22 °C | 6 | 6/30 = 0.2 | 20% |
| 23 °C | 4 | 4/30 = 0,13 | 13% |
| 24 °C | 6 | 6/30 = 0,2 | 20% |
| 25 °C | 8 | 8/30 = 0,27 | 27% |
| 26 °C | 4 | 4/30 = 0,13 | 13% |
| 27 °C | 2 | 2/30 = 0,07 | 7% |

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG08 |
| **Descripción** |  |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | Diagrama de las temperaturas registradas durante un mes en una ciudad. |

Para determinar qué sector circular le corresponde a cada frecuencia multiplicamos cada frecuencia relativa ***hi*** por la razón **360°** Así por ejemplo el sector que corresponde a a la temperatura de 22° con una frecuencia de 6 repeticiones y equivale al 20% se halla del siguiente modo.

α = 0,2∙360° = 72°

Así 72° equivalen al 20%

[SECCIÓN 3] **2.1.3 El histograma de frecuencias**

EL histograma de frecuencias es un diagrama de barras que se emplea cuando la variable es cuantitativa continua, o cuando el número de datos es bastante grande que se agrupan los datos en **intervalos de igual** longitud.

Para representar los datos utilizaremos el **histograma**, en el que señalaremos los valores de los intervalos mediante áreas de rectángulos que tienen por base la amplitud del intervalo y por altura la frecuencia absoluta.

Si unimos los puntos medios de los extremos superiores de las barras mediante una línea, obtenemos otra representación gráfica que llamaremos **polígono de frecuencias**.

El siguiente ejemplo muestra el resumen de los datos obtenidos con un velocímetro que determino la velocidad en *km/h* de 100 autos en una autopista durante una hora.

|  |  |
| --- | --- |
| Velocidad  *Xi* | Numero de autos  *fi* |
| [50, 60) | 15 |
| [60, 70) | 20 |
| [70, 80) | 40 |
| [80, 90) | 10 |
| [90, 100) | 15 |

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG09 |
| **Descripción** |  |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | En el histograma de frecuencias por tratarse de una variable continua, las barras no pueden estar separadas. |

[SECCIÓN 2] **2.2 Diagrama de tallos y hojas**

El diagrama de **tallos y hojas** es una representación que organiza y clasifica datos cuantitativos de una misma variable, y a la vez permite obtener una distribución de las frecuencias que presenta la variable.

Para representar el diagrama tomamos el dato y lo dividimos en dos partes en el que uno o algunos de sus primeros dígitos representaran el tallo y los demás dígitos representaran la hoja. Así por ejemplo si tenemos las edades de tres personas que son 17, 15 y 14 años respectivamente diremos que el tallo son las decenas, en este caso es el 1 y las hojas son las unidades, para este tallo son 7, 5 y 4. Su representación queda del siguiente modo.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG10 |
| **Descripción** |  |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | Representación de un diagrama de tallos y hojas. |

Ahora observemos el diagrama con un conjunto de datos más amplio.

La siguiente tabla muestra las notas de un examen para 20 estudiantes en una escala del 1.0 al 5.0.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Notas de un examen | | | | |
| 4.5 | 3.0 | 3.0 | 3.4 | 2.6 |
| 5.0 | 1.2 | 4.2 | 3.5 | 2.4 |
| 4.8 | 2.2 | 2.8 | 1.6 | 4.0 |
| 3.9 | 4.1 | 5.0 | 2.7 | 3.8 |

A continuación elaboraremos el diagrama tomando como tallos la parte entera de cada nota y como hojas la parte decimal. Así obtenemos la siguiente representación.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG11 |
| **Descripción** |  |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | Representación de un diagrama de tallos y hojas. |

Como observas el diagrama nos permite ver en qué rango de valores se encuentra la mayor frecuencia de los datos, que para este caso es para las notas entre 3.0 y 3.9, además nos permite estudiar si existe dispersión entre los datos o si se concentran o tienen hacia un valor especifico.

[SECCIÓN 2] **2.3 Diagrama de caja y bigotes**

Un Diagrama de **caja y bigotes** es un gráfico, que usa los **cuartiles** (medida que divide un conjunto de datos en 4 partes), mediante el cual se visualiza un conjunto de datos. Se compone de un rectángulo que denominamos **la caja**, y dos brazos, que denominamos **los bigotes**.

Este diagrama nos suministra información acerca de los valores mínimo y máximo, los cuartiles Q1, Q2 o mediana y Q3, y sobre la existencia de valores atípicos y la simetría de la distribución.

Para construir el diagrama primero se debe hallar el valor de **la mediana** para luegoencontrar los 2 cuartiles restantes y el rango intercuartilico.

Revisemos el siguiente ejemplo**:**

El siguiente conjunto de datos muestra la cantidad de equipos móviles que un vendedor en periodo de prueba vendió durante 15 días.

12, 8, 25, 11, 11, 23, 17, 22, 18, 16, 16, 15, 12, 19, 22.

Reorganizamos los datos y tenemos:

8, 11, 11, **12**, 12, 15, 16, **16**, 17, 18, 19, **22**, 22, 23, 25.

El cuartil dos que corresponde a la mediana que es el dato de la mitad es:

Me = Q2 = 16,

Los cuartiles Q1 y Q3 son:

Q1 = 12 y Q3 = 22

Ahora hallamos el rango intercuartilico (RIC).

Q3 − Q1 = 22 − 12= 10

Ahora identificamos el dato mínimo xmin y el dato máximo xmax.

xmin = 8 y xmax =25.

Con todos estos datos procedemos a elaborar nuestro diagrama:

El bigote de la izquierda comprende a los datos que están entre el dato mínimo y el primer cuartil, el bigote de la derecha comprende los datos que están entre el tercer cuartil y el dato máximo. Mientras que la caja contempla a los datos ubicados entre el primer y el tercer cuartil.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG11 |
| **Descripción** |  |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | Diagrama de caja y bigotes. |

[SECCIÓN 2] **2.4 Consolidación**

Actividades para consolidar lo que has aprendido en esta sección.

[SECCIÓN 1] **3 Las medidas estadísticas**

Ya que hemos revisado como se recolectan, organizan y representan los datos estadísticos, debemos estudiar qué elementos nos ofrece la estadística para analizar los datos y para ello las medidas estadísticas son una herramienta que permite realizar esta labor.

[SECCIÓN 2] **3.1 Las medidas de centralización o de posición**

Las **medidas de centralización**, mejor conocidas como **medidas de tendencia central** son un instrumento que nos permiten analizar el comportamiento de un conjunto de datos a través de la representación de un solo número, como su nombre lo indican las medidas de tendencia central indican hacia qué valor o valores tiende el conjunto de datos. Entre ellas encontramos las siguientes:

[SECCIÓN 3] **3.1.1 la moda**

La moda indica el dato que se presenta con mayor frecuencia en un conjunto de datos, se representa mediante la expresión (**Mod**).

La moda tiene la particularidad de que es una medida que se puede aplicar para datos cuantitativos y cualitativos, esta situación no se presenta con las otras medidas que solo aplica a datos cuantitativos.

Observa estos ejemplos:

Ejemplo 1.

El siguiente conjunto de datos muestra el número de libros que lee un grupo 15 de estudiantes al año S = {3, 5, 2, **8**, 7, 7, 3, 9, 12, **8**, 4, 3 ,7, **8**, **8**}.

Como observas el dato que presenta mayor frecuencia o que más se repite es el 8, en este caso decimos que la moda es el 8 y la escribimos como:

Mod = 8.

En este caso también podemos decir que el conjunto de datos es **unimodal.**

Ejemplo 2.

En una ciudad se tomó una pequeña muestra de 10 autos para determinar el color que más les gusta para sus autos. Y se obtuvieron los siguientes resultados.

S = {**Gris plateado**, Negro, Negro, **Rojo**, **Gris plateado**, **Rojo**, Blanco, Blanco, **Rojo**, **Gris plateado**}

En este caso te puedes dar cuenta que existen dos datos que se repiten por tanto el conjunto tiene dos modas.

Mod = Gris plateado y Mod = rojo.

En este caso decimos que el conjunto de datos es multimodal.

Si en el conjunto de datos no existe ningún dato que se repita, decimos que en el conjunto no existe moda.

[SECCIÓN 3] **3.1.2 la media aritmética**

La **media aritmética** o **promedio** es una medida de tendencia central que solo se aplica en datos cuantitativos, su uso da una aproximación numérica acerca del comportamiento de un conjunto de datos. Un ejemplo claro es la forma en que usualmente obtienes la nota definitiva de una asignatura, ya que sumas las notas parciales y luego divides por el total de notas. En este caso lo que estas halando es la **media aritmética.**

La media aritmética se representa como:

https://latex.codecogs.com/png.latex?%5Cdpi%7B200%7D%20%5Cfn_jvn%20%5Clarge%20%5Coverline%7BX%7D

Y se definen como:

https://latex.codecogs.com/png.latex?%5Cfn_jvn%20%5Clarge%20%5Coverline%7BX%7D%3D%5Cfrac%7B%5Csum_%7Bi%3D1%7D%5E%7Bn%7Dx_%7Bi%7D%7D%7BN%7D%3D%5Cfrac%7Bx_%7B1%7D&plus;x_%7B2%7D&plus;x_%7B3%7D&plus;...x_%7Bn%7D%7D%7BN%7D

El símbolo **∑** es la **letra del alfabeto griego sigma** y significa sumatoria.

Veamos un ejemplo:

El siguiente conjunto de datos representa la edad de un grupo de 12 niños que pertenecen a una escuela de futbol 8, 5, 7, 7, 6, 5, 7, 5, 7, 5, 6, 5. Ahora determinemos la edad promedio del grupo de niños, es decir hallemos la media aritmética sumando todos los datos y dividiendo por el total de datos *N* = 12

https://latex.codecogs.com/png.latex?%5Cfn_jvn%20%5Clarge%20%5Coverline%7BX%7D%3D%5Cfrac%7B8&plus;5&plus;7&plus;7&plus;6&plus;5&plus;7&plus;5&plus;7&plus;5&plus;6&plus;5%7D%7B12%7D

https://latex.codecogs.com/png.latex?%5Cfn_jvn%20%5Clarge%20%5Coverline%7BX%7D%3D%5Cfrac%7B73%7D%7B12%7D%3D6.08

Por tanto podemos decir que la edad promedio de los niños esta alrededor de los 6 años.

Si los datos se encuentran agrupados en una tabla de frecuencias, la media la calculamos a partir de la suma del producto de cada dato pos su frecuencia y luego dividiendo por el total de datos. La fórmula queda expresada del siguiente modo:

C:\Users\MIGUEL MUÑOZ\Pictures\promedio.png

Por ejemplo la siguiente tabla resume los datos recolectados acerca del número de comparendos que se registraron durante 20 días en la ciudad de Bogotá por superar el límite de velocidad permitido.

|  |  |
| --- | --- |
| Comparendos por exceso de velocidad | |
| *x*i | *f*i |
| 123 | 7 |
| 345 | 4 |
| 234 | 8 |
| 128 | 11 |

Por tanto el promedio aritmético lo calculamos así.

https://latex.codecogs.com/png.latex?%5Cfn_jvn%20%5Clarge%20%5Coverline%7BX%7D%3D%5Cfrac%7B123%5Ccdot%207&plus;345%5Ccdot%204&plus;234%5Ccdot8&plus;128%5Ccdot11%7D%7B30%7D

https://latex.codecogs.com/png.latex?%5Cfn_jvn%20%5Clarge%20%5Coverline%7BX%7D%3D%5Cfrac%7B861&plus;1380&plus;1872&plus;1408%7D%7B30%7D

https://latex.codecogs.com/png.latex?%5Cfn_jvn%20%5Clarge%20%5Coverline%7BX%7D%3D%5Cfrac%7B5521%7D%7B30%7D%3D184

Lo cual implica que diariamente por incumplir esta norma se aplican en promedio 180 comparendos o multas de tránsito.

[SECCIÓN 3] **3.1.3 la mediana**

Para un conjunto de datos cuantitativos, la **mediana** es el dato que divide al conjunto en dos partes iguales luego de ser ordenados de menor a mayor, se representa como **Me**. También decimos que es el dato que ocupa el lugar de la mitad. Para el caso de la mediana encontramos dos casos:

**Si el total de datos es un número impar:**

Se organizan los datos de menor a mayor y se toma el dato de la posición **x(n+1)/2.**

Por ejemplo.

A continuación se presenta la estatura en metros de 7 personas:

1.67m, 1.58m, 1.72m 1.62m, 1.75m, 1.60m y 1.70m.

Reorganizando los datos tenemos:

1.58m, 1.60m, 1.62m, **1.67m**, 1.70m, 1.72m, 1.75m.

Como el total de datos es 7, el dato (n+1)/2 es el dato número 4 que corresponde **1.67m** por tanto.

**Me = 1.67m**

**Si el total de datos es un número par:**

En este caso se deben tomar los dos datos centrales luego de ser organizados de menor a mayor y se calcula el promedio entre estos dos datos, el cual corresponderá al mediana.

Por ejemplo si al grupo de las edades añadimos un dato más tenemos:

1,70m, 1.67m, 1.58m, 1.72m 1.62m, 1.75m, 1.60m y 1.70m.

Organizados de menor a mayor quedan así.

1.58m, 1.60m, 1.62m, **1.67m**, **1.70m**, 1.70m, 1.72m, 1.75m.

Como los datos centrales son **1,67** y **1.70** hallamos el promedio entre ellos así que

https://latex.codecogs.com/png.latex?%5Cfn_jvn%20%5Clarge%20%5Coverline%7BX%7D%3D%5Cfrac%7B1.67&plus;1.70%7D%7B2%7D%3D%5Cfrac%7B3.37%7D%7B2%7D%3D1.685

Por tanto la mediana es **Me = 1.685**

[SECCIÓN 3] **3.1.4 los Cuartiles**

Los cuartiles son tres medidas de posición que dividen un conjunto de datos ordenados en cuatro partes iguales. Se simbolizan con *Q*1, *Q*2 y *Q*3, y determinan los valores correspondientes al 25 %, al 50 % y al 75 % de los datos. **El segundo cuartil** coincide con la **mediana**.

Por ejemplo, reunimos y ordenamos los valores del experimento de tirar un dado 13 veces: 1, 1, 1, 2, 3, 3, 4, 4, 5, 5, 5, 5, 6.

* El segundo cuartil *Q*2 o mediana es el 4.
* El primer cuartil es la mediana de las seis primeras observaciones, las situadas a la izquierda de la mediana: *Q*1 = (1+2)/2 = 1,5.
* El tercer cuartil es la mediana de las observaciones a la derecha de la mediana:*Q*3 = (5 + 5)/2 = 5.

Por tanto los datos quedan divididos del siguiente modo:

1, 1, 1, **1.5**, 2, 3, 3 , **4**, 4, 5, 5, **5**, 5, 5, 6.

Cada cuartil permite analizar como es el comportamiento de los datos en cada porcentaje.

[SECCIÓN 2] **3.2 Las medidas estadísticas de desviación o de dispersión**

Hasta ahora hemos estudiado cómo se comportan los datos a través de su tendencia a un valor, pero estas medidas no son suficientes para hacer un buen análisis, debido a que las medidas de tendencia central no son datos representativos para el comportamiento de los datos en los extremos o para determinar qué tan alejados o próximos se encuentran los datos entre sí.

Por tanto las **medidas de dispersión** o de **desviación** miden cómo se encuentran agrupados los datos en torno a la media aritmética: si la desviación es muy grande, el valor central no es representativo del total de datos. Las medidas de dispersión más utilizadas son:

* El rango.
* La varianza.
* La desviación típica.
* El coeficiente de correlación.

Dos valores pueden tener la misma media y ser muy diferentes, si los valores de las variables se distribuyen de forma distinta. Fijémonos en la siguiente imagen:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_08\_11\_IMG12 |
| **Descripción** | http://profesores.aulaplaneta.com/DNNPlayerPackages/Package13436/InfoGuion/cuadernoestudio/images_xml/MT_09_11_img7_small.jpg |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | Dos distribuciones de datos que muestran diferentes dispersiones. |

La **media aritmética** de los valores de ambas tablas es la misma: 5,7. Pero en la primer grafica se observa una mayor dispersión de los datos mientras que en la segunda se encuentran concentrados entre sí.

[SECCIÓN 3] **3.2.1 El rango**

El **rango** o **recorrido** de un conjunto de datos es la **diferencia** entre el **mayor** y el **menor valor** de la variable. Se simboliza mediante la letra ***R***. Veamos cómo se calcula el rango con el siguiente ejemplo:

Para hacer seguimiento al control de peso de un grupo de estudiantes un médico ha recolectado el peso de 25 estudiantes de una clase: 67, 55, 58, 59, 72, 62, 65, 70, 61, 74, 60, 64, 63, 49, 57, 62, 68, 48, 62, 58, 59, 56, 55, 57, 62.

Para calcular el rango debemos seguir estos pasos:

1. Buscamos de la serie de datos los siguientes valores:
   * El valor mayor: 74.
   * El valor menor: 48.
2. Restamos al valor mayor el valor menor: 74 − 48 = 26**.**

Así, el rango de la variable estudiada es: *R* = 26 kg.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Interpretación del rango** |
| **Contenido** | Cuanto mayor es el valor que adquiere el rango implica que el conjunto de datos va a presentar una mayor dispersión. |

[SECCIÓN 3] **3.2.1 La desviación media**

Esta medida indica que tan alejado o que tan cerca se encuentra un dato con respecto a la **media aritmética.**

La **desviación respecto a la media** se halla calculando **diferencia** entre cada **dato** y la **media aritmética,** cada desviación queda representada por:

https://latex.codecogs.com/png.latex?%5Cfn_jvn%20%5Clarge%20D_%7Bi%7D%3Dx_%7Bi%7D-%5Coverline%7Bx%7D

Por tanto La **desviación media** es la **media aritmética** de los **valores absolutos de las desviaciones respecto a la media** yla representamos como:

C:\Users\MIGUEL MUÑOZ\Pictures\desviacionmedia.png

Por ejemplo calculemos la desviación para las notas parciales de un estudiante en un curso de matemáticas si sus notas son: 4.5, 3.0, 2.0 5.0 y 4.0

Primero hallamos la media aritmética

https://latex.codecogs.com/png.latex?%5Cfn_jvn%20%5Clarge%20%7B%5Coverline%7Bx%7D%7D%3D%5Cfrac%7B4.5&plus;3.0&plus;2.0&plus;5.0&plus;4.0%7D%7B5%7D%3D%5Cfrac%7B185%7D%7B5%7D%3D3.7

Ahora hallamos cada desviación

D1 = 4.5 − 3.7 = 0.8

D2 = 3.0 − 3.7 = −0.7

D3 = 2.0 − 3.7 = −1.7

D4 = 5.0 − 3.7 = 1.3

D5 = 4.0 − 3.7 = 0.3

Con cada desviación, procedemos a calcular la desviación media.

C:\Users\MIGUEL MUÑOZ\Pictures\desvi1.png

Por tanto la desviación media de los datos respecto a la media aritmética es de 0.96 puntos.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Interpretación de la desviación media** |
| **Contenido** | Cuanto mayor es la desviación media, más grande es la dispersión de los valores de la muestra. |

[SECCIÓN 3] **3.2.1 La varianza**

La varianza de un conjunto de datos es la media aritmética de los cuadrados de las desviaciones de esos datos respecto de la media aritmética. Se simboliza con la letra griega sigma minúscula, elevada al cuadrado:***σ*2**

C:\Users\MIGUEL MUÑOZ\Pictures\varianza.png

Revisemos la varianza para el ejemplo de las notas del estudiante 4.5, 3.0, 2.0 5.0 y 4.0.

Sus desviaciones medias son respectivamente:

D1 = 4.5 − 3.7 = 0.8

D2 = 3.0 − 3.7 = −0.7

D3 = 2.0 − 3.7 = −1.7

D4 = 5.0 − 3.7 = 1.3

D5 = 4.0 − 3.7 = 0.3

Por tanto su varianza es:

https://latex.codecogs.com/png.latex?%5Cfn_jvn%20%5Csigma%20%5E%7B2%7D%3D%5Cfrac%7B%280.8%29%5E2&plus;%28-0.7%29%5E2&plus;%28-1.7%29%5E2&plus;%281.3%29%5E2&plus;%280.3%29%5E2%7D%7B5%7D

https://latex.codecogs.com/png.latex?%5Cfn_jvn%20%5Csigma%20%5E%7B2%7D%3D%5Cfrac%7B0.64&plus;0.49&plus;2.89&plus;1.69&plus;0.9%7D%7B5%7D

https://latex.codecogs.com/png.latex?%5Cfn_jvn%20%5Csigma%20%5E%7B2%7D%3D%5Cfrac%7B6.61%7D%7B5%7D%3D1.33

Así el valor de la varianza es de **1.33**

La varianza tiene el inconveniente de que se expresa en unidades al cuadrado. Por ejemplo, si las observaciones se han realizado en centímetros, la varianza se expresará en centímetros cuadrados.

[SECCIÓN 3] **3.2.1 La desviación típica**

La desviación típica es la raíz cuadrada de la varianza *σ*2 Se simboliza con *σ*. A diferencia de la varianza, la desviación típica se expresa en las mismas unidades que las usadas al hacer las observaciones.

https://latex.codecogs.com/png.latex?%5Cdpi%7B120%7D%20%5Cfn_jvn%20%5Clarge%20%5Csigma%3D%5Csqrt%5Csigma%20%5E%7B2%7D

Por ejemplo, si tenemos una varianza *σ*2 = 1.33, la desviación típica será:

https://latex.codecogs.com/png.latex?%5Cdpi%7B120%7D%20%5Cfn_jvn%20%5Clarge%20%5Csigma%3D%5Csqrt%5Csigma%20%5E%7B2%7D%3D%5Csqrt%7B1.33%7D%5Capprox%201.15

Una distribución de resultados se considera normal si un 68 % del total está comprendido entre los números:

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package13436/InfoGuion/cuadernoestudio/images_xml/MT_09_11_formula13_resized.gif

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Aplicaciones de la desviación típica** |
| **Contenido** | La desviación típica sirve para medir el grado de dispersión. Cuando dos distribuciones tienen la misma media, la desviación típica nos indica lo alejados que están los valores respecto a este valor. |

[SECCIÓN 3] **3.2.1 El coeficiente de variación**

El coeficiente de variación es el cociente de la desviación típica entre la media. Se simboliza con ***Cv***. Cuando comparamos dos distribuciones, la que tiene mayor coeficiente de variación es también la que tiene mayor dispersión.

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package13436/InfoGuion/cuadernoestudio/images_xml/MT_09_11_formula14_resized.gif

Utilizamos el coeficiente de variación cuando tenemos dos distribuciones heterogéneas con distinta media y queremos compararlas para saber la dispersión de sus valores. Se expresa como un porcentaje (%).

Por ejemplo, vamos a calcular el coeficiente de variación, conociendo la desviación típica y la media del siguiente conjunto de datos:

4.5, 3.0, 2.0 5.0 y 4.0

De donde:

https://latex.codecogs.com/png.latex?%5Cdpi%7B120%7D%20%5Cfn_jvn%20%5Clarge%20%5Coverline%7Bx%7D%3D3.7%3B%20%5Csigma%3D%201.15

Luego su coeficiente de variación es:

https://latex.codecogs.com/png.latex?%5Cdpi%7B120%7D%20%5Cfn_jvn%20%5Clarge%20C_%7Bv%7D%3D%5Cfrac%7B1.15%7D%7B3.7%7D%3D0.31

Es decir que la variación de los datos es de aproximadamente un 31% con respecto a la media.

[SECCIÓN 2] **3.3 Consolidación**

Actividades para consolidar lo que has aprendido en esta sección.

[SECCIÓN 1] **4.** **Ejercitación y competencias**

|  |  |
| --- | --- |
| **Mapa conceptual** | |
| **Código** | MA\_08\_09\_REC00 |
| **Título** | Mapa conceptual |
| **Descripción** |  |

|  |  |
| --- | --- |
| **Evaluación: recurso nuevo** | |
| **Código** | MA\_08\_09\_REC00 |
| **Título** |  |
| **Descripción** |  |

|  |  |  |
| --- | --- | --- |
| **Webs de referencia** | | |
| **Código** | MA\_08\_09\_REC00 | |
| **Web 01** | *Aula fácil* | *http://www.aulafacil.com/cursos/l11213/ciencia/estadisticas/estadisticas/introduccion-a-la-estadistica-descriptiva* |
| **Web 02** | *Estadística descriptiva* | *http://www.ditutor.com/estadistica/estadistica\_descriptiva.html* |
| **Web 03** | *Medidas de dispersión* | *http://www.vitutor.net/2/11/medidas\_dispersion.html* |