|  |  |
| --- | --- |
| **Título del guion** | Estadística |
| **Código de guion** | MA\_09\_12\_CO |
| **Descripción** | En un mundo globalizado como el de hoy, se maneja mucha información; alguna de esta información puede ser recogida por medio de variables y luego se puede organizar, representar y analizar, pero, ¿sabes cómo se puede representar y analizar? Te invitamos a que conozcas cómo se puede representar y analizar la información, utilizando algunas herramientas estadísticas. |

[SECCIÓN 1] **1 Análisis de variables**

La **estadística** es la rama de las matemáticas que se encarga de recoger, organizar, clasificar, presentar y analizar datos, numéricos o no numéricos, con el fin de ayudar a resolver problemas, tomar de decisiones, presentar la información, entre otras finalidades.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_09\_12\_IMG01 |
| **Descripción** | Gráficos utilizados en la estadística |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | http://thumb101.shutterstock.com/display_pic_with_logo/499414/171204833/stock-photo-business-man-drawing-different-graphs-and-charts-171204833.jpg  <http://thumb101.shutterstock.com/display_pic_with_logo/499414/171204833/stock-photo-business-man-drawing-different-graphs-and-charts-171204833.jpg> |
| **Pie de imagen** | La estadística utiliza diferentes formas para presentar la información. |

Las **variables** en estadística se definen como las características que pueden tener los individuos de una población determinada, como por ejemplo la edad, el sexo, la altura, la nacionalidad, sus salarios, entre otros.

|  |  |
| --- | --- |
| **Recuerda** | |
| **Contenido** | Existen dos clases de variables:  **Variables cualitativas:** son aquellas variables que no se pueden medir numéricamente, ya que representan atributos o rasgos, por ejemplo: nombre, religión, deporte favorito o género, entre otros.  **Variables cuantitativas:** son aquellas variables que toman valores numéricos, por ejemplo: edad, peso, altura, precio, entre otras. |

Al conjunto de datos que se obtiene luego de recoger la información se le puede realizar un **análisis de variable**, utilizando diferentes herramientas como: distribución de frecuencias, gráficos, histogramas, medidas de posición, medidas de dispersión, entre otras. La escogencia de las herramientas dependerá de la clase de variable que se quiere analizar (cuantitativa o cualitativa).

El análisis debe describir el comportamiento de la variable; con dicho comportamiento es posible elaborar conclusiones, las cuales podrán ser utilizadas para responder algunas preguntas, resolver algún problema o ayudar a tomar una decisión; a continuación, se presentan algunas herramientas que permiten realizar el análisis de **variables cualitativas**.

|  |  |
| --- | --- |
| **Profundiza (recurso de exposición)** | |
| **Código** | MA\_09\_12\_REC10 |
| **Título** | El estudio estadístico. Población y muestra |
| **Descripción** | Interactivo que sirve para introducir los conceptos básicos de la estadística como población, muestra y variables estadísticas |

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC20 |
| **Título** | Repasa conceptos básicos de estadística |
| **Descripción** | Actividad que permite profundizar los conceptos básicos sobre la estadística |

[SECCIÓN 2] **1.1 Análisis de una variable cualitativa**

Las **variables cualitativas,** también llamadas **variables categóricas,** son aquellas que se expresan solo en forma de atributo; son ejemplos: sexo, ciudad donde vive, tipo de sangre, color de ojos, entre otros. Las variables cualitativas se pueden dividir en dos grupos:

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Clasificación de variables cualitativas** |
|  | Las variables cualitativas en estadística se clasifican en:  **Variables cualitativas nominales:** son las que no permiten definir un orden jerárquico entre las posibles respuestas, por ejemplo, para la variable color de cabello, las posibles respuestas serían: negro, rubio, castaño, entre otras.  **Variables cualitativas ordinales:** son las variables que sí permiten definir un orden específico que no es numérico, por ejemplo, para usted cómo ha sido la gestión del nuevo alcalde: excelente, buena, regular, mala. |

Otra clasificación de las variables cualitativas se realiza dependiendo del número de datos que puede tener la variable, como se muestra a continuación:

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Clasificación de las variables cualitativas de acuerdo con el número de datos** |
|  | Según el número de posibles respuestas, la variable estadística se puede clasificar en:  **Las variables dicotómicas** son aquellas que tienen solo dos respuestas posibles, como por ejemplo: está de acuerdo con el alza en el trasporte; en este caso las posibles respuestas pueden ser sí o no.  **Las variables politómicas** son aquellas que poseen más de dos respuestas posibles, por ejemplo: cuál es el color de sus ojos; para este hecho las posibles respuestas son más de dos. |

Algunas herramientas que se pueden utilizar para realizar el **análisis de una variable cualitativa** son:

* Las tablas de frecuencia
* Los diagramas de barras
* Los diagramas de líneas
* Las gráficas de sectores
* La moda

Para mostrar la aplicación de estas herramientas se presenta el siguiente ejemplo:

Se realizó una encuesta a 450 estudiantes de un colegio, donde una de las preguntas era: ¿cuál es el color de sus ojos?; los resultados fueron los siguientes:

color castaño 190, color ámbar 140, color avellana 105, color verde 10, color azul 5.

**La tabla de frecuencias:** también conocida como tabla de distribución de frecuencias, es una tabla en la que se ordenan los datos de cada variable y se asigna a cada dato su frecuencia correspondiente.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Clases de frecuencia** |
|  | Las frecuencias se clasifican en:  **Frecuencia absoluta (*fi*):** se define como el número de veces que aparece un determinado dato en un estudio estadístico.  **Frecuencia relativa:** es el cociente entre la frecuencia absoluta de un dato determinado y el número total de datos de la muestra. La frecuencia relativa se puede expresar como porcentaje multiplicando este cociente por 100.  **Frecuencia acumulada (*Fi*):** es la suma de las frecuencias de los datos anteriores; esta puede ser frecuencia absoluta acumulada o frecuencia relativa acumulada, según sea la clase de frecuencias que se suman. La interpretación de la frecuencia acumulada solo tiene sentido en variables cuantitativas o en variables cualitativas ordinales. |

Del ejemplo anterior, se presenta la siguiente tabla de frecuencias:

|  |  |  |
| --- | --- | --- |
| **Color ojos** | **Frecuencia absoluta** | **Frecuencia relativa** |
| Castaño | 190 | 42,2 % |
| Ámbar | 140 | 31,1 % |
| Avellana | 105 | 23,3 % |
| Verde | 10 | 2,3 % |
| Azul | 5 | 1,1 % |
| **Total** | 450 | 100 % |

Se debe resaltar que en la tabla de frecuencias no se calcula la frecuencia acumulada, puesto que la variable que se presenta aquí es cualitativa nominal. Por lo tanto, el cálculo de la frecuencia acumulada no tiene ningún sentido.

**El diagrama de barras:** describe cada uno de los datos de acuerdo con su frecuencia, utilizando barras representadas en un diagrama cartesiano.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_09\_12\_IMG02 |
| **Descripción** | Diagrama de barras |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | El diagrama de barras da cuenta del número de estudiantes que posee determinado color de ojos. Este diagrama hace uso de la frecuencia absoluta. |

* **Gráfica de sectores o torta:** describe cada uno de los datos de acuerdo con su frecuencia, que se representa mediante la fragmentación de un círculo; casi siempre se presenta en porcentajes.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_09\_12\_IMG04 |
| **Descripción** | Gráfica de torta |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | El diagrama de torta representa el porcentaje de estudiantes que hay en el colegio de cada color de ojos. |

**La moda:** es una medida de tendencia central que se define como el dato que tiene mayor frecuencia absoluta.

En el ejemplo, el color castaño es la moda, puesto que su frecuencia absoluta es 190.

Algunas de las conclusiones que se pueden extraer gracias a la descripción del comportamiento de la variable, que nos brindan las diferentes herramientas que se utilizaron para su análisis, son:

* En el colegio la mayoría de estudiantes tiene color de ojos castaño.
* Solo el 1,1 % de los estudiantes del colegio tiene los ojos de color azul.
* 105 estudiantes tienen el color de ojos avellana, número que equivale al 23,3 % de la totalidad de estudiantes del colegio.

Pero se pueden construir muchas más conclusiones, gracias a las herramientas de análisis que se utilizaron (la tabla de frecuencia, diagrama de barras, diagrama de líneas, gráficas de sectores y la moda).

En el ejemplo anterior se mostró cómo se realiza el **análisis de una variable cualitativa**, de una manera sencilla; hay que aclarar que algunas de las herramientas utilizadas como las tablas de frecuencia y la moda se desarrollarán de manera más profunda más adelante.

A continuación, se describirá el análisis que se puede realizar de dos variables cualitativas simultáneamente.

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC40 |
| **Título** | Resuelve situaciones analizando la información |
| **Descripción** | Actividad que permite reconocer el tipo de variable estadística |

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC60 |
| **Título** | Variables cuantitativas: discreta y continuas |
| **Descripción** | Ejercicios que permiten diferenciar variables cuantitativas discretas y continuas |

[SECCIÓN 2]**1.2 Análisis de dos variable cualitativas simultáneas**

El **análisis de dos variables cualitativas simultáneas** se realiza cuando estasvariables están relacionadas entre sí, lo cual se cumple cuando ciertos valores de una de las variables se asocian con algunos valores de la otra variable, como por ejemplo:

* El consumo de harinas con el sobrepeso.
* La adicción al cigarrillo con el cáncer.

El análisis de dos variables intenta describir la relación existente entre una variable respecto a otra; para hacer este análisis es necesario establecer si existe una relación de dependencia entre las dos variables o no existe, lo cual se puede llevar a cabo de una manera intuitiva observando los datos directamente y determinando si existe dicha relación o no, sin embargo, es más fácil desarrollar este análisis por medio de una tabla cruzada.

[SECCIÓN 3]**1.2.1 La tabla cruzada o de contingencia**

La **tabla de contingencia** es una herramienta que permite estudiar la información obtenida de dos variables dentro de una población determinada, donde los valores de la primera variable se representan en *n* clases (filas) y los valores de la segunda variable en *m* clases (columnas); dichos valores son clasificados y ordenados en una tabla de doble entrada de dimensión *n ×* *m*, donde cada casilla de la tabla es una de las posibles categorías; para su construcción, por lo general, se utilizan las frecuencias absolutas de ambas variables, esto permite realizar un cruce de sus categorías con el fin de determinar si existe algún tipo de relación de dependencia entre las dos variables; observa el siguiente ejemplo donde se crea una tabla cruzada:

* Se realizó una encuesta a 200 personas del barrio Casa Blanca; las preguntas fueron las siguientes:

1. Consume pan todos los días.
2. Tiene sobrepeso.

Los resultados obtenidos se organizaron en la siguiente tabla de contingencia:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Consume pan todos los días** | |  |
| **Tiene sobrepeso** | **Sí** | **No** | **Total filas** |
| **Sí** | 142 | 15 | 157 |
| **No** | 20 | 23 | 43 |
| **Total columnas** | 162 | 38 | 200 |

Las **tablas de contingencia** permiten efectuar una observación detallada y determinar si existe una relación de dependencia entre las dos variables. De la tabla de contingencia se puede decir:

1. De las 162 personas que consumen pan todos los días 142 tienen sobrepeso y 20 no tienen sobrepeso.
2. De las 38 personas que no consumen pan todos los días 15 tienen sobrepeso y 23 no tienen sobrepeso.

A continuación, se mostrará otra herramienta que es utilizada para el análisis de dos variables cuantitativas, denominada tabla marginal.

[SECCIÓN 3]**1.2.2 Las tablas marginales**

La **tabla marginal** es aquella representación que permite centrar el análisis en una sola de las variables sin importar el comportamiento de la otra; los datos de las tablas marginales están plasmados en la tabla de contingencia en total columnas y total filas. Para su construcción, por lo general, se utiliza la frecuencia absoluta de cada una de las variables, las cuales permitirán describir su comportamiento individualmente. Observa el siguiente ejemplo:

De acuerdo con la tabla de contingencia:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Consume pan todos los días** | |  |
| **Tiene sobrepeso** | **Sí** | **No** | **Total filas** |
| **Sí** | 142 | 15 | **157** |
| **No** | 20 | 23 | **43** |
| **Total columnas** | **162** | **38** | **200** |

Se presentan las siguientes dos tablas marginales:

Distribución marginal de la variable de sobrepeso:

|  |  |
| --- | --- |
| **Tiene sobrepeso** | ***fi*** |
| sí | 157 |
| no | 43 |

Distribución marginal de la variable de consumo de pan a diario:

|  |  |
| --- | --- |
| **Consume pan todos los días** | ***fi*** |
| sí | 162 |
| no | 38 |

En resumen, de las tablas de distribución marginal se puede decir que:

1. De la muestra total 157 personas tienen sobrepeso y 43 no lo tienen.
2. De la muestra total 162 personas consumen pan todos los días y 38 personas no consumen pan todos los días.

Estos datos permitirán extraer diferentes conclusiones como:

* La probabilidad de encontrar personas con sobrepeso en Casa Blanca es más alta que la probabilidad de encontrar personas que no tengan sobrepeso.

En la siguiente sección se presentan las variables cuantitativas y el análisis a estas variables.

|  |  |
| --- | --- |
| **Profundiza (recurso de exposición)** | |
| **Código** | MA\_09\_12\_REC30 |
| **Título** | Caracterización de variables |
| **Descripción** | Interactivo que explica la caracterización de las variables cualitativas y cuantitativas. Se describen las tablas cruzadas, las tablas marginales y las tablas de frecuencia |

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC50 |
| **Título** | Estudia las tabla de contingencia |
| **Descripción** | Práctica que permite determinar datos de una tabla de contingencia |

[SECCIÓN 2]**1.3 La caracterización de variables cuantitativas**

La **caracterización de variables cuantitativas** se refiere a la forma como se presentan los datos de las variables cualitativas; dependiendo del número de datos que se recogen, se pueden definir dos formas:

**Los datos no agrupados:** son aquellos que no requieren ser agrupados para poderlos interpretar, por lo tanto, los datos se presentan como fueron recogidos, lo cual aplica cuando los datos recogidos en la muestra no son muchos.

Por ejemplo: las edades de 20 estudiantes de grado 9 de un colegio son estas:

14, 13, 12, 15, 16, 14, 14, 16, 12, 13, 15, 14, 17, 14, 13, 15, 12, 16, 17, 14.

De estos datos se puede concluir que:

* Las edades de los estudiantes están entre los 12 años y los 17 años.
* La moda es 14 años.

**Los datos agrupados**: son aquellos que pueden ser divididos en clases o intervalos, debido a que los datos recogidos en la muestra son muchos y no se puede realizar un análisis sino se clasifican y se ordenan previamente.

Las herramientas para analizar y representar los datos de forma agrupada son: las tablas de distribución de frecuencia, los histogramas, polígonos de frecuencia, la ojiva, entre otras.

[SECCIÓN 2] **1.4 Consolidación**

Actividades para consolidar lo que has aprendido en esta sección.

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC70 |
| **Título** | Refuerza tu aprendizaje: El análisis de variables |
| **Descripción** | Actividades sobre El análisis de variables |

[SECCIÓN 1] **2 Las tablas de distribución de frecuencia**

Una **tabla de distribución de frecuencia** es una herramienta que permite organizar los datos en columnas, las cuales representan los diferentes valores que se han recogido en la muestra y las frecuencias de cada uno de ellos. Los elementos que hacen parte de la tabla de frecuencia son: variable, datos, frecuencia absoluta, frecuencia absoluta acumulada, frecuencia relativa y frecuencia relativa acumulada. A continuación se desarrollan los conceptos de las frecuencias.

[SECCIÓN 2] **2.1 Las frecuencias absoluta, relativa y acumulada**

Las frecuencias son los valores numéricos sobre los cuales se realiza el estudio estadístico de cada uno de los datos. Estas frecuencias son:

La **frecuencia absoluta (*fi*)** es el número de veces que aparece el mismo dato en un estudio estadístico; la suma de todas las frecuencias absolutas es igual al total de datos recogidos.

La **frecuencia absoluta acumulada (*Fi*)** es la suma de las frecuencias absolutas de los datos o intervalos menores o iguales al dato que se quiere calcular la frecuencia absoluta acumulada.

La **frecuencia relativa (*ni*)** es el cociente entre la frecuencia absoluta y el número total de datos; se puede presentar de forma decimal, en fracción o en porcentaje. Una de sus propiedades es que la suma de todas las frecuencias relativas de forma decimal es igual a 1 y la suma de todas las frecuencias relativas en porcentaje es igual a 100 %.

La **frecuencia relativa acumulada (*Ni*)** es el cociente entre la frecuencia absoluta acumuladay el número total de datos. Este valor se puede expresar en forma de porcentaje multiplicando la forma decimal por 100.

Veamos el siguiente ejemplo: se lanza un dado 25 veces; los resultados obtenidos son los siguientes:

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

La siguiente tabla relaciona las frecuencias con los datos obtenidos en el experimento.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***x*** | ***fi*** | ***Fi*** | ***ni*** | | ***Ni*** | |
| **Decimal** | **Porcentaje**  **%** | **Decimal** | **Porcentaje**  **%** |
| 1 | 5 | 5 | 0,2 | 20 | 0,2 | 20 |
| 2 | 4 | 9 | 0,16 | 16 | 0,36 | 36 |
| 3 | 3 | 12 | 0,12 | 12 | 0,48 | 48 |
| 4 | 3 | 15 | 0,12 | 12 | 0,6 | 60 |
| 5 | 4 | 19 | 0,16 | 16 | 0,76 | 76 |
| 6 | 6 | 25 | 0,24 | 24 | 1 | 100 |
| **Total** | 25 |  | 1 | 100 |  |  |

Es decir que para calcular la frecuencia absoluta acumulada de obtener 4 con los dados, se adicionan las frecuencias absolutas anteriores: 5 + 4 + 3 + 3 = 15. Esta frecuencia acumulada indica que en 15 ocasiones diferentes se obtuvo un número menor o igual a 4 al lanzar un dado.

Mientras que la frecuencia relativa acumulada de obtener 3 con los dados se obtiene al adicionar las frecuencias relativas anteriores, así: 20 % + 16 % + 12 % = 48 %, es decir, el 48 % de los lanzamientos obtuvieron resultados menores o iguales a 3.

|  |  |
| --- | --- |
| **Profundiza (recurso de exposición)** | |
| **Código** | MA\_09\_12\_REC80 |
| **Título** | La frecuencia de los datos. Los gráficos |
| **Descripción** | Interactivo que permite establecer las diferencias entre frecuencia relativa y absoluta, así como conocer algunos gráficos estadísticos |

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC90 |
| **Título** | Analiza información en tablas |
| **Descripción** | Ejercicios que permiten identificar los elementos de una tabla de frecuencia |

[SECCIÓN 2] **2.2 Las tablas de frecuencia para variables discretas y variables continuas**

Para elaborar una tabla de frecuencias es indispensable tener en cuenta el tipo de variable cuantitativa que se maneja; existen dos clases: la variable cuantitativa discreta y la variable cuantitativa continua.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Las variables cuantitativas discretas** |
|  | Son las variables que toman datos dentro de un conjunto dado, además de no tener valores intermedios entre dos valores consecutivos. Un ejemplo de variable discreta puede ser número de hijos, en el que las respuestas que se pueden obtener serán: 0 hijos, 1 hijo, 2 hijos, 4 hijos, 5 hijos, pero nunca se podrá encontrar una respuesta como 2,5 hijos. |

En seguida se presenta un ejemplo de una tabla de distribución de frecuencias para variables discretas:

En una encuesta a 40 personas del barrio el Tunal, cuyas edades estaban entre los 18 y 65 años, una de las preguntas era: ¿cuántos hijos tiene? Las respuestas fueron las siguientes:

1, 2, 4, 0, 5, 2, 4, 5, 6, 1, 2, 5, 6, 2, 0, 2, 0, 3, 0, 4, 3, 2, 4, 5, 6, 7, 6, 4, 3, 2, 4, 5, 1, 0, 0, 0, 1, 2, 3, 0.

La tabla de frecuencias presenta la siguiente estructura:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***x*** | ***fi*** | ***Fi*** | ***ni*** | | ***Ni*** | |
| **Decimal** | **Porcentaje**  **%** | **Decimal** | **Porcentaje**  **%** |
| 0 | 8 | 8 | 0,2 | 20 | 0,2 | 20 |
| 1 | 4 | 12 | 0,1 | 10 | 0,3 | 30 |
| 2 | 8 | 20 | 0,2 | 20 | 0,5 | 50 |
| 3 | 4 | 24 | 0,1 | 10 | 0,6 | 60 |
| 4 | 6 | 30 | 0,15 | 15 | 0,75 | 75 |
| 5 | 5 | 35 | 0,125 | 12,5 | 0,875 | 87,5 |
| 6 | 4 | 39 | 0,1 | 10 | 0,975 | 97,5 |
| 7 | 1 | 40 | 0,025 | 2,5 | 1 | 100 |
| **Total** | 40 |  | 1 | 100 |  |  |

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Las variables cuantitativas continuas** |
|  | Las **variables continuas** son aquellas que toman valores dentro de un conjunto definido, donde siempre se puede encontrar un elemento entre dos elementos dados. Un ejemplo de variable continua es la estatura de las personas, puesto que se obtienen respuestas como: 1,70 m, 1,71 m, 1,712 m, 1,82 m, 1,83 m, 1,822 m; como puedes observar, siempre se podrá hallar un elemento intermedio entre dos elementos dados. |

La estructura de la tabla de frecuencias para variables continuas es similar a las tablas de frecuencias de variables discretas, sin embargo, es preciso elaborar primero los intervalos en los que se desea particionar los datos.

**Agrupación en intervalos de variables continuas**

Para determinar los intervalos que particionan al conjunto de datos de la muestra se deben seguir estos pasos:

1. **Ordenar los valores de menor a mayor o de mayor a menor.**

Por ejemplo:

En la siguiente tabla se recoge la estatura en metros de 100 personas que trabajan en una empresa:

1,65; 1,55; 1,70; 1,53; 1,81; 1,78; 1,68; 1,56; 1,89; 1,90;

1,67; 1,77; 1,85; 1,65; 1,72; 1,73; 1,89; 1,50; 1,69; 1,54;

1,78; 1,65; 1,77; 1,69; 1,81; 1,83; 1,90; 1,77; 1,89; 1,51;

1,66; 1,95; 1,49; 1,51; 1,98; 1,77; 1,48; 1,72; 1,55; 1,89;

1,44; 1,67; 1,88; 1,66; 1,77; 1,68; 1,67; 1,92; 1,67; 1,47;

1,45; 1,56; 1,78; 1,54; 1,87; 1,67; 1,56; 1,45; 1,87; 1,78;

1,58; 1,45; 1,57; 1,88; 1,98; 1,67; 1,45; 1,70; 1,74; 1,77;

1,78; 1,67; 1,54; 1,77; 1,54; 1,65; 1,67; 1,65; 1,68; 1,56;

1,67; 1,56; 1,42; 1,44; 1,45; 1,45; 1,54; 1,64; 1,65; 1,65;

1,76; 1,85; 1,55; 1,74; 1,86; 1,49; 1,84; 1,44; 1,51; 1,79.

De acuerdo con el paso 1, estos datos se ordenan de menor a mayor y se obtiene:

1,42; 1,44; 1,44; 1,44; 1,45; 1,45; 1,45; 1,45; 1,45; 1,45;

1,47; 1,48; 1,49; 1,49; 1,50; 1,51; 1,51; 1,51; 1,53; 1,54;

1,54; 1,54; 1,54; 1,54; 1,55; 1,55; 1,55; 1,56; 1,56; 1,56;

1,56; 1,56; 1,57; 1,58; 1,64; 1,65; 1,65; 1,65; 1,65; 1,65;

1,65; 1,65; 1,66; 1,66; 1,67; 1,67; 1,67; 1,67; 1,67; 1,67;

1,67; 1,67; 1,67; 1,68; 1,68; 1,68; 1,69; 1,69; 1,70; 1,70;

1,72; 1,72; 1,73; 1,74; 1,74; 1,76; 1,77; 1,77; 1,77; 1,77;

1,77; 1,77; 1,77; 1,78; 1,78; 1,78; 1,78; 1,78; 1,79; 1,81;

1,81; 1,83; 1,84; 1,85; 1,85; 1,86; 1,87; 1,87; 1,88; 1,88;

1,89; 1,89; 1,89; 1,89; 1,90; 1,90; 1,92; 1,95; 1,98; 1,98.

1. **Determinar el número de intervalos (*k*):** existen varios métodos para determinar el número de intervalos en que se desea particionar la muestra; los más utilizados son:

* **La regla de Sturges:** permite conocer el número de intervalos aplicando la fórmula:

*k* = 1 + 3,22 · log10*N*

donde *k* es el número de intervalos y *N* es el tamaño de la muestra. Además, el valor de *k* se debe redondear al número entero más cercano. Esta regla se aplica cuando *N* es un número bastante grande de la muestra.

* **La regla de la raíz cuadrada:** cuando el número de elementos de la muestra no es muy grande, se aplica la fórmula:

*k =* √*N*

donde *k* es el número de intervalos, que se aproxima al número entero más cercano, y *N* es el tamaño de la muestra.

* **Por conveniencia:** el estadista puede considerar el número de intervalos que requiere para su estudio, cuando lo ha definido con alguna anticipación y empleando algún criterio para hacer el análisis.

En el ejemplo de la estatura, como el número de datos recogidos es relativamente pequeño, los intervalos se definen utilizando la regla de la raíz:

*k = √N, k = √*100 *=* 10*, k =* 10, luego el conjunto de datos se dividirá en 10 intervalos.

1. **Determinar el rango de los datos *R*:** el rango de los datos es la diferencia entre el dato mayor y el dato menor de la muestra. Es decir:

*R* = *x*mayor*– x*menor

En el ejemplo, se tiene:

*x*mayor = 1,98

*x*menor = 1,42

Por lo tanto: *R* = *x*mayor*– x*menor ⟶ *R* = 1,98 – 1,42 = 0,56. Es decir que el rango es *R* = 0,56.

1. **Determinar la amplitud del intervalo:** se define como el tamaño de cada intervalo y se calcula por medio de la fórmula:

*<<MA\_09\_12\_01.gif>>*

donde *a* es la amplitud del intervalo, *R* es el rango de los datos y *k* es el número de intervalos.

En el ejemplo de la estatura de los empleados de la empresa se tiene:

*<<MA\_09\_12\_02.gif>>*

Por consiguiente, la amplitud de cada intervalo es *a* = 0,056.

1. **Hallar los extremos de cada intervalo:** los extremos de los intervalos, los cuales serán: *l*0 = *x*menor, *l*1 = *l*0 + *a*, *l*2 = *l*1 + *a*, se determinan, de modo general:

*l*i = *l*i – 1 + *a*

El primer intervalo [*l*0, *l*1] incluye los extremos, por ende, es cerrado; los demás intervalos incluyen el extremo inferior, pero no incluyen el extremo superior, en consecuencia, son semiabiertos a izquierda y se representan como (*l*i, *l*i – 1].

Continuando con el ejemplo, el primer intervalo se obtiene así:

*l*0 = *x*menor = 1,42; *l*1 = *l*0 + *a ⟶ l*1 = 1,42 + 0,056 = 1,476. Por ello, el primer intervalo de clase es [1,42; 1,476].

El segundo intervalo se obtiene: *l*2 = *l*1 + *a* ⟶ *l*2 = 1,476+ 0,056 = 1,532. Así, el segundo intervalo es (1,476, 1,532].

Si se continúa el mismo proceso, se obtienen los intervalos:

[1,42; 1,476], (1,476, 1,532], (1,532, 1,588], (1,588, 1,644], (1,644, 1,7], (1,7, 1,756], (1,756, 1,812], (1,812, 1,868], (1,868, 1,924], (1,924, 1,98]

1. Identificar la marca de clase de cada intervalo.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La marca de clase de un intervalo** |
|  | Es el número que representa a todos los elementos que están en el intervalo; se calcula adicionando los extremos de un intervalo y dividiéndolos entre dos, así  <<MA\_09\_12\_GIF54>>  donde *mi*es la marca de clase del intervalo *i*, *li* es el límite superior y *l*i – 1 es el límite inferior del intervalo. |

Por ejemplo, para calcular la marca de clase del tercer intervalo, se identifican

*l*3 = 1,588; *l*3 – 1 = *l*2 = 1,532, que al reemplazarlos en la fórmula, se obtiene:

<<MA\_09\_12\_GIF55>>

Por lo tanto, la marca de clase del tercer intervalo es 1,56.

1. Organizar los intervalos en la tabla de frecuencias con sus respectivas frecuencias.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_09\_12\_IMG05 |
| **Descripción** | Tabla de frecuencias |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | En la tabla de frecuencias se organiza la información en intervalos, se identifica la marca de clase de cada intervalo y se calculan sus frecuencias. |

Como se observa, la tabla de frecuencias es una herramienta muy útil para el análisis de variables tanto cuantitativas como cualitativas. En la siguiente sección se identifican algunos valores numéricos que son relevantes para interpretar un conjunto de datos estadísticos.

[SECCIÓN 2] **2.3 Consolidación**

Actividades para consolidar lo que has aprendido en esta sección.

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC100 |
| **Título** | Refuerza tu aprendizaje: Tablas de distribución de frecuencias |
| **Descripción** | Actividades sobre Las tablas de frecuencia |

[SECCIÓN 1] **3 Las medidas de posición**

Las **medidas de posición** son aquellasque permiten reconocer las características que posee una serie de datos ordenados que fueron recogidos por medio de una variable cuantitativa. Las medidas de posición se dividen en dos clases: **las medidas de posición central** y **las medidas de posición no central**.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Las medidas de posición central** |
|  | Las medidas de posición o tendencia central son los valores numéricos que representan al conjunto de datos; estas medidas se tienden a localizar en la parte media del conjunto de datos ordenados de una variable cuantitativa.  Las medidas de tendencia central más utilizadas en el análisis estadístico son: la media aritmética, la mediana y la moda. |

[SECCIÓN 2] **3.1 La media aritmética**

La media aritmética también recibe el nombre de promedio o esperanza matemática, y se representa como

<<MA\_09\_12\_GIF53>>

Y se define a continuación:

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La media aritmética** |
|  | La media aritmética es el dato que no necesariamente está en la conjunto de datos, pero sí es el que representa a todo el conjunto de datos, puesto que se considera que los demás datos son muy cercanos a la media. |

**En un conjunto de datos no agrupados**, la media aritmética se calcula mediante la fórmula:

*<<MA\_09\_12\_03.gif>>*

donde *x*1, *x*2,…, *x*n son elementos del conjunto de datos y *n* es el número de datos del conjunto.

Por ejemplo, el promedio de peso de una familia que tiene 5 integrantes, si sus pesos son los siguientes:

padre: 85 kg, madre: 57 kg, hermano mayor: 68 kg, hermano del medio 56 kg, hermano menor; 46 kg, se calcula así aplicando la fórmula:

<<MA\_09\_12\_05.gif>>

Se obtiene que el promedio de peso de la familia es 62,4 kilogramos.

**En un conjunto de datos agrupados,** la media aritmética se calcula utilizando la fórmula:

<<MA\_09\_12\_04.gif>>

donde *mi* es la marca de clase del intervalo *i*, *fi* es la frecuencia absoluta del intervalo *i*, *k* es el número de intervalos y *n* es el número de datos de la muestra.

La siguiente tabla de frecuencias corresponde a la estatura de los 100 empleados que trabajan en una fábrica:

|  |  |  |  |
| --- | --- | --- | --- |
| **Clase** | **Intervalos** | **Marca de clase del intervalo** | **Frecuencia absoluta** |
| ***K*** | ***li*, *li* - 1** | ***mi*** | ***fi*** |
| **1** | [1,42, 1,476] | 1,448 | 11 |
| **2** | (1,476, 1,532] | 1,504 | 8 |
| **3** | (1,532, 1,588] | 1,56 | 15 |
| **4** | (1,588, 1,644] | 1,616 | 1 |
| **5** | (1,644, 1,70] | 1,672 | 25 |
| **6** | (1,70, 1,756] | 1,728 | 5 |
| **7** | (1,756, 1,812] | 1,784 | 16 |
| **8** | (1,812, 1,868] | 1,84 | 5 |
| **9** | (1,868, 1,924] | 1,896 | 11 |
| **10** | (1,924, 1,98] | 1,952 | 3 |
| **Total** |  |  | **100** |

Con base en ella, la media aritmética se calcula con la fórmula así:

*<<MA\_09\_12\_06.gif>>*

Por lo tanto, el promedio de estatura de los trabajadores esaproximadamente 1,67 m.

[SECCIÓN 2] **3.2 La mediana**

Una medida de tendencia central es la mediana, que se define a continuación:

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La mediana** |
|  | La mediana es el dato que divide al conjunto de datos en dos, de manera que haya el mismo número de datos mayores que la mediana como menores que la mediana. La mediana se representa como  <<MA\_09\_12\_56.gif>> |

Para calcular la mediana se debe realizar el siguiente proceso, **si los datos no son agrupados**:

1. Ordenar los datos de menor a mayor.
2. Si el número de datos *n* es un número impar, el dato que corresponde a la mediana está en la mitad del conjunto ordenado de datos, es decir:

<<MA\_09\_12\_57.gif>>

Si el número de datos *n* es par, la mediana es la media aritmética de los dos datos de la mitad. Así:

<<MA\_09\_12\_58.gif>>

En el ejemplo del peso de los integrantes de la familia se tienen los siguientes datos

85, 57, 68, 56, 46

Se ordenan los datos de menor a mayor y se obtiene

46, 56, 57, 68, 85.

Como *n* = 5 y es un número impar, la mediana se calcula:

<<MA\_09\_12\_59.gif>>

En consecuencia, la mediana es el tercer dato del conjunto de datos, es decir:

<<MA\_09\_12\_60.gif>>

**Si los datos son agrupados**, la mediana se calcula de la siguiente manera:

1. El intervalo donde está la mediana, el cual es aquel donde la frecuencia acumulada llega a la mitad del total de datos.
2. Del intervalo que contiene a la mediana se determina el límite inferior *li –* 1, la amplitud *ai* y la frecuencia absoluta *fi*. Además, la frecuencia acumulada anterior al intervalo donde se encuentra la mediana *Fi –* 1.
3. La mediana se calcula a través de la siguiente expresión:

*<<MA\_09\_12\_08.gif>>*

Por ejemplo, en la siguiente tabla se encuentran los datos recogidos en un estudio estadístico, determine su mediana.

|  |  |  |
| --- | --- | --- |
| **Intervalo** | ***fi*** | ***Fi*** |
| [0, 5] | 15 | 15 |
| (5, 10] | 20 | 35 |
| (10, 15] | 25 | 60 |
| (15, 20] | 30 | 90 |
| (20, 25] | 12 | 102 |
| (25, 30] | 10 | 112 |
| (30, 35] | 5 | 117 |

1. Se determina el intervalo donde está la mediana, es decir, aquel donde la frecuencia acumulada llega a la mitad:

<<MA\_09\_12\_09.gif>>

Es decir que la mitad de la frecuencia acumulada es 58,5 y se encuentra en el intervalo (10,15].

1. Se identifican los datos que hacen referencia a este intervalo: *li –* 1=10, *ai* =5, *Fi –* 1 = 35*, fi* = 25*.*
2. Se sustituyen estos datos en la fórmula:

*<<MA\_09\_12\_10.gif>>*

Por lo tanto, la mediana es 14,7.

[SECCIÓN 2] **3.3 La moda**

La **moda** es la única medida de tendencia central que se calcula tanto para variables cuantitativas como cualitativas y se define a continuación:

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La moda** |
|  | Es el dato que tiene mayor frecuencia absoluta del conjunto de datos. La moda se representa como:  <<MA\_09\_12\_61.gif>>  Si existen dos datos que tienen la mayor frecuencia absoluta se indica que la distribución es bimodal. La distribución es multimodal si tiene más de dos modas. |

Para hallar la moda cuando **los datos son agrupados,** se debe desarrollar el siguiente proceso:

1. Precisar el intervalo con mayor frecuencia absoluta; este intervalo se define como clase modal.
2. Determinar de la clase modal, los siguientes datos: amplitud *ai*,límite inferior *li* – 1, frecuencia absoluta *fi*, así como la frecuencia absoluta del intervalo inmediatamente menor a la frecuencia absoluta *fi* – 1, y la frecuencia absoluta del intervalo inmediatamente mayor a la clase modal *fi* + 1.
3. Se reemplazan los valores anteriores en la fórmula para determinar la moda:

<<MA\_09\_12\_11.gif>>

Por ejemplo, en la siguiente tabla se encuentran las calificaciones obtenidas por los 500 estudiantes en el segundo período con su frecuencia absoluta.

|  |  |
| --- | --- |
| **Notas** | **Frecuencia absoluta** |
| [0, 2] | 50 |
| (2, 4] | 125 |
| (4, 6] | 100 |
| (6, 8] | 150 |
| (8, 10] | 75 |
| Total | 500 |

Para determinar la moda se ejecuta el siguiente proceso:

1. Se identifica la clase modal, en este caso es el intervalo: (6, 8].
2. Se determinan los valores de la clase modal: *ai*= 2, *li* – 1 = 6, *fi* = 150*, fi* – 1*=* 100, *fi* + 1*=* 75.
3. Se reemplazan en la fórmula:

<<MA\_09\_12\_12.gif>>

Es decir que la moda es 6,8.

|  |  |
| --- | --- |
| **Profundiza (recurso de exposición)** | |
| **Código** | MA\_09\_12\_REC120 |
| **Título** | Encuentra las medias de posición |
| **Descripción** | Actividad para resolver situaciones aplicando las medidas de centralización |

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC110 |
| **Título** | Actividad para calcular la media, la mediana y la moda |
| **Descripción** | Calcula las medidas de centralización |

[SECCIÓN 2] **3.4 Los valores no centrales**

**Medidas de posición no central** o **valores no centrales** son las medidas que permiten reconocer otros valores característicos del conjunto de datos ordenados, que no son los que se encuentran en la parte media de la distribución.

|  |  |
| --- | --- |
| **Profundiza (recurso de exposición)** | |
| **Código** | MA\_09\_12\_REC130 |
| **Título** | Las medidas de posición no central |
| **Descripción** | Interactivo que muestra cómo calcular medidas de posición no central como cuartiles, deciles y percentiles |

[SECCIÓN 2] **3.4.1 Los cuartiles, los deciles y los percentiles**

Las medidas de posición no central son los cuartiles, los deciles y los percentiles.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Los cuartiles** |
|  | Los **cuartiles** son los tres valores que particionan al conjunto ordenado de datos en cuatro partes iguales, de manera que en cada tramo concentra el 25% de los datos de la muestra. |

**Cuando los datos no son agrupados** un método para encontrar los 3 valores de los cuartiles es el siguiente:

1. Ordenar los datos de menor a mayor.
2. Buscar el lugar de cada uno de los cuartiles utilizando la siguiente fórmula:

<<MA\_09\_12\_13.gif>>

donde *Qk* es la posición del cuartil, *k* es el cuartil que se quiere calcular y *n* es el número total datos de la distribución.

Si, por ejemplo, estas son las edades de 12 personas:

24, 25, 39, 34, 56, 45, 19, 34, 23, 20, 34, 36.

Para calcular los cuartiles se realiza el siguiente proceso:

1. Se ordenan los datos de menor a mayor:

19, 20, 23, 24, 25, 34, 34, 34, 36, 39, 45, 56.

1. Se encuentra cada uno de los tres cuartiles aplicando la fórmula:

<<MA\_09\_12\_15.gif>>

De la fórmula se deduce que el cuartil uno se halla en la posición 3 que equivale a 23, el segundo cuartil se encuentra en la posición 6 que equivale a 34 y el cuartil tres está en la posición 9 que equivale a 36.

Se puede decir que el 25 % de las personas encuestadas tiene menos de 23 años, el 50 % de las personas tiene entre 19 y 34 años, el 75 % de los encuestados tiene entre 19 y 36 años.

|  |  |
| --- | --- |
| **Profundiza (recurso de exposición)** | |
| **Código** | MA\_09\_12\_REC140 |
| **Título** | Los cuartiles y el diagrama de caja y bigotes |
| **Descripción** | Interactivo que permite a los estudiantes conocer e interpretar el concepto de cuartil, diagrama de caja y bigotes y conocer sus aplicaciones |

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Los deciles** |
|  | Los **deciles** se definen como los nueve valores que distribuyen al conjunto de datos ordenados de menor a mayor, en tramos iguales, es decir que cada tramo concentra el 10 % de los datos de la muestra. |

Cuando los datos no están agrupados un método para encontrar los 9 valores de los deciles es el siguiente:

1. Ordenar los datos de menor a mayor.
2. Buscar el lugar de cada uno de los deciles utilizando la fórmula:

<<MA\_09\_12\_25.gif>>

donde *Dk* es la posición del *k* decil en el conjunto ordenado de datos y *n* es el total de datos de la muestra.

A continuación se presenta el número de automóviles de una conocida marca que se vendieron en Colombia cada día, durante los últimos 55 días.

25, 59, 26, 68, 27, 57, 56, 44, 53, 43, 33, 57, 34, 61, 67, 30, 35, 71, 43, 35,

24, 69, 37, 22, 66, 38, 55, 35, 41, 62, 31, 49, 38, 61, 43, 36, 21, 29, 48, 45,

36, 42, 29, 64, 32, 18, 33, 64, 55, 46, 65, 37, 45, 47, 55.

Para encontrar el segundo, el cuarto y el séptimo *k* decil, primero se deben ordenar los datos de menor a mayor:

18, 21, 22, 24, 25, 26, 27, 29, 29, 30, 31, 32, 33, 33, 34, 35, 35, 35, 36, 36,

37, 37, 38, 38, 41, 42, 43, 43, 43, 44, 45, 45, 46, 47, 48, 49, 53, 55, 55, 55,

56, 57, 57, 59, 61, 61, 62, 64, 64, 65, 66, 67, 68, 69, 71.

Se aplica la fórmula para conocer la posición de cada uno de los deciles:

<<MA\_09\_12\_26.gif>>

Se observa que el segundo decil está en la posición 11, que corresponde a 31 en el conjunto ordenado de datos, es decir que el 20 % de los días se vendieron menos de 31 vehículos. Asimismo, el cuarto decil está en la posición 22, que indica que 37 automóviles o menos fueron vendidos en el 40% de los días.

Respecto al séptimo decil, se observa que *Dk* = 38,5; se utilizan entonces los datos más cercanos a 38,5, es decir, el dato 38 y el dato 39 que son ambos 55, se calcula su diferencia y se multiplica por la parte decimal de 38,5, es decir, 0,5. Luego se le adiciona el valor del dato 38, esto es, 55, así:

(55 – 55) · 0,5 + 55 = 0 · 0,5 + 55 = 0 + 55 = 55

Como los datos 38 y 39 tienen el mismo valor, se deduce que el 70 % de los días se vendieron 55 vehículos o menos.

En general, el conjunto de datos se puede dividir en 100 partes mediante los percentiles.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Los percentiles** |
|  | Los **percentiles** se definen como los noventa y nueve valores que distribuyen al conjunto de datos ordenados de menor a mayor, en tramos iguales, es decir que cada tramo concentra el 1 % de los datos de la muestra. |

Para calcular un percentil se debe llevar a cabo el siguiente proceso:

1. Ordenar los datos de menor a mayor.
2. Determinar el lugar que ocupa en el conjunto ordenado de datos cada uno de los percentiles mediante la fórmula:

<<MA\_09\_12\_30.gif>>

donde *Pk* es la posición del *k* percentil y *n* es el total de datos de la muestra.

Por ejemplo, la producción diaria de una fábrica de zapatos, durante el último mes, fue la siguiente:

722, 850, 841, 890, 777, 731, 657, 754, 854, 745, 658, 662, 631, 811, 781,

881, 654, 647, 820, 690, 781, 783, 846, 806, 848, 736, 819, 672, 875, 602.

Determinar cuál fue la producción de pares de zapatos en el 32 % de los días.

Para saberlo se debe hallar el percentil 32, por lo tanto, se ordena el conjunto de datos de menor a mayor:

602, 631, 647, 654, 657, 658, 662, 672, 690, 722, 731, 736, 745, 754, 777

781, 781, 783, 806, 811, 819, 820, 841, 846, 848, 850, 854, 875, 881, 890.

Se aplica la fórmula:

<<MA\_09\_12\_31.gif>>

Como *Pk* = 9,6, se toman el noveno y décimo valores y se resuelve en la fórmula:

690 + (722 – 690) · 0,6 = 690 + (32) · 0,6 = 690 + 19,2 = 709,2

Por consiguiente, el 32 % de los días se produjeron 709 o menos pares de zapatos aproximadamente.

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC150 |
| **Título** | ¿Qué sabes de la estadística? |
| **Descripción** | Actividad que permite repasar los conceptos básicos de las medidas de posición |

[SECCIÓN 2] **3.6 Consolidación**

Actividades para consolidar lo que has aprendido en esta sección.

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC160 |
| **Título** | Refuerza tu aprendizaje: Las medidas de posición |
| **Descripción** | Actividad que permite reforzar el conocimiento sobre Las medidas de posición |

[SECCIÓN 1] **4 Las medidas de dispersión**

Las **medidas de dispersión** son las encargadas de estudiar la distribución de los datos, analizando si dichos valores se encuentran en el centro de la distribución o muy dispersos en esta. Las medidas de dispersión más utilizadas son **el rango, la varianza, la desviación típica y el coeficiente de variación.**

|  |  |
| --- | --- |
| **Profundiza (recurso de exposición)** | |
| **Código** | MA\_09\_12\_REC170 |
| **Título** | Las medidas de dispersión |
| **Descripción** | Interactivo mediante el cual se definen las medidas de dispersión como la desviación estándar, la varianza y el coeficiente de variación |

[SECCIÓN 1] **4.1 El rango**

Una de las medidas de dispersión es el rango, que se define como:

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **El rango** |
|  | El **rango *R*** es la distancia entre el dato menor y el dato mayor del conjunto de datos;se calcula mediante su diferencia, así:  *R = xmáximo – xmínimo*  donde *x*máximoes el dato mayor y *x*mínimoes el dato menor de la distribución. |

El peso en kilos de cada uno de los 15 estudiantes hombres de noveno grado es:

50, 70, 60, 75, 65, 78, 89, 60, 64, 76, 78, 78, 67, 65, 67.

Para calcular el rango de datos se identifica el peso menor y el peso mayor de los estudiantes: *x*máximo= 89 *y x*mínimo= 50.

Se sustraen estos valores, así:

*R* =89 – 50 = 39

El rango de la distribución es 39.

[SECCIÓN 2] **4.2 La desviación media**

Para poder definir lo que es la **desviación media** se debe primero puntualizar qué es la desviación respecto a la media.

La **desviación respecto a la media (*DM*)** se puede definir como la distancia que hay entre cada dato y la media aritmética, y se calcula con la fórmula:

<<MA\_09\_12\_62.gif>>

donde *Di* es la desviación de cada dato respecto a la media, *xi*es cualquiera de los datos y la media aritmética es

<<MA\_09\_12\_53.gif>>

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La** **desviación media** |
|  | **La desviación media** es el promedio de las deviaciones respecto a la media de cada uno de los datos.  En los datos no agrupados se calcula con la ayuda de la fórmula:  <<MA\_09\_12\_35.gif>>  donde *x*1, *x*2, *x*3,…, *xn* son los valores de los datos y *n* es el número de datos. |

En datos agrupados se calcula mediante la fórmula:

*<<MA\_09\_12\_36.gif>>*

donde *m1*, *m2*, *m3*,…, *mn* son las marcas de clase de cada intervalo, *f1*, *f2*, *f*3,…, *f*n son las respectivas frecuencias absolutas, *n* es el número de intervalos y *N* es el número de datos de la muestra.

Para datos no agrupados, se presenta el siguiente ejemplo:

Encuentra la desviación media de los siguientes datos que representan la altura en metros de 5 estudiantes de noveno grado:

1,70; 1,60; 1,50, 1,56: 1,69

Se calcula la media aritmética del conjunto de datos:

<<MA\_09\_12\_63.gif>>

Se identifican los datos de la situación *n* = 5, *x*1 = 1,50; *x*2 = 1,56; *x*3 = 1,60; *x*4 = 1,69; *x*5 = 1,70.

Se reemplazan en la fórmula:

*<<MA\_09\_12\_37.gif>>*

Se obtiene que la deviación media *DM* = 0,068.

[SECCIÓN 2] **4.3 La varianza y la desviación típica**

La varianza es una medida de dispersión que se define como:

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La** **varianza** |
|  | La **varianza** que se representa comoσ2, es la media aritmética de los cuadrados de las desviaciones respecto a la media.  En los datos no agrupados se define como:  *<<MA\_09\_12\_38.gif>>*  donde *x1*, *x2*, *x3*,…, *xn* son los valores de los datos, *n* es el número de datos. |

Cuando los datos están agrupados, la varianza se calcula por medio de la fórmula:

<<MA\_09\_12\_39.gif>>

donde *m*1, *m*2,…, *mn* son las marcas de clase de cada intervalo, *n* es el número de intervalos, *N* es el número de datos de la muestra y *f*1, *f*2,…, *fn* es la frecuencia absoluta de cada intervalo.

El siguiente es un ejemplo para datos agrupados:

Calcule la varianza de las edades de 2500 encuestados, que se encuentran consignadas en la siguiente tabla de frecuencias:

|  |  |  |
| --- | --- | --- |
| ***li*, *li* – 1** | ***fi*** | ***Fi*** |
| [18-23] | 520 | 520 |
| (23-28] | 620 | 1140 |
| (28-33] | 410 | 1550 |
| (33-38] | 325 | 1875 |
| (38-43] | 300 | 2175 |
| (43-48] | 325 | 2500 |
| **Total** | 2500 |  |

Para calcular la media aritmética con base en los datos anteriores, se determina la marca de clase de cada intervalo *mi*, y el producto de la frecuencia absoluta de cada intervalo con su respectiva marca de clase *mi* · *fi*, como se muestra en la siguiente tabla.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***li , li –* 1** | ***mi*** | ***fi*** | ***mi*· *fi*** | ***Fi*** |
| [18-23] | 20,5 | 520 | 10 660 | 520 |
| (23-28] | 25,5 | 620 | 15 810 | 1140 |
| (28-33] | 30,5 | 410 | 12 505 | 1550 |
| (33-38] | 35,5 | 325 | 11 537,5 | 1875 |
| (38-43] | 40,5 | 300 | 12 150 | 2175 |
| (43-48] | 45,5 | 325 | 14 787,5 | 2500 |
| **Total** |  | 2500 | 77 450 |  |

Al dividir el total de la columna *mi* · *fi*  entre el total de la muestra *N* = 2500 se obtiene que la media aritmética de los datos es:

<<MA\_09\_12\_64.gif>>

En algunos casos una tabla permite calcular con mayor facilidad la varianza como se muestra a continuación.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***li , li –* 1** | ***mi*** | ***fi*** | ***mi – 30,98*** | **(*mi* – 30,98)2** | **(*mi* – 30,98)2 · *fi*** |
| [18-23] | 20,5 | 520 | -10,48 | 109,8304 | 57 111,808 |
| (23-28] | 25,5 | 620 | -5,48 | 30,0304 | 18 618,848 |
| (28-33] | 30,5 | 410 | -0,48 | 0,2304 | 94,464 |
| (33-38] | 35,5 | 325 | 4,52 | 20,4304 | 6639,88 |
| (38-43] | 40,5 | 300 | 9,52 | 90,6304 | 27 189,12 |
| (43-48] | 45,5 | 325 | 14,52 | 210,8304 | 68 519,88 |
| **Total** |  | 2500 |  |  | **178 174** |

Esta tabla permite calcular la varianza, mediante el cociente:

<<MA\_09\_12\_41.gif>>

Siendo 𝜎2 = 71,2692 la varianza requerida.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La** **desviación típica** |
|  | La **desviación típica o desviación estándar (σ)** se define como la raíz cuadrada de la **varianza.**  Cuando los datos no son agrupados se calcula mediante la fórmula:  <<MA\_09\_12\_42.gif>>  donde *x*1, *x*2,…, *xn*son los valores de los datos, *N* el número de datos de la muestra, y la media aritmética de los datos es:  <<MA\_09\_12\_53.gif>> |

Cuando los datos están agrupados en intervalos y presentados en una tabla de frecuencias, la desviación estándar se define como:

*<<MA\_09\_12\_43.gif>>*

donde *m*1, *m*2,…, *mn*son la marca de clase de cada intervalo, *N* el número de datos de la muestra, *n* el número de intervalos y la media aritmética de los datos es:

<<MA\_09\_12\_53.gif>>

En el ejemplo sobre las edades de los 2500 encuestados, y cuya varianza es 𝜎2 = 71,2692, la desviación estándar se calcula aplicando la raíz cuadrada a la varianza, así:

*<<MA\_09\_12\_44.gif>>*

donde la desviación típica será σ = 8,4421.

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC180 |
| **Título** | Determina medidas estadísticas |
| **Descripción** | Ejercicios que permiten la práctica sobre las medidas estadísticas dentro de un conjunto de datos |

SECCIÓN 2] **4.4 El coeficiente de variación**

El coeficiente de variación expresa la relación entre la media y la desviación estándar de la siguiente manera:

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Coeficiente de variación (*CV*)** |
|  | **El coeficiente de varianza** o de variación **(*CV*)** se define como el cociente entre la **desviación típica** de una muestra y el valor absoluto de su **media aritmética.**  Se calcula a través de la fórmula:  <<MA\_09\_12\_45.gif>>  También se puede expresar como porcentaje mediante la fórmula:  <<MA\_09\_12\_65.gif>> |

Por ejemplo, en una muestra de productos lácteos de una tienda se encontró que, en promedio, estos productos se vencen en 30,9 días y su desviación típica es de *σ =* 1,5. El coeficiente de varianza es:

Se sustituyen los valores en la fórmula:

<<MA\_09\_12\_46.gif>>

Se obtiene que el coeficiente de varianza es *CV* = 4,8 %, esto quiere decir que los datos se desvían de la media aritmética en un 4,8 %.

En la siguiente sección el trabajo consistirá en resolver situaciones problema que involucran el trabajo de análisis estadístico.

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC200 |
| **Título** | Interpreta el concepto de coeficiente de variación |
| **Descripción** | Actividad creada para calcular el coeficiente de variación |

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC190 |
| **Título** | Practica las medidas de posición y dispersión |
| **Descripción** | Actividad que permite identificar y practicar las medidas de posición y de dispersión |

[SECCIÓN 2] **4.5 Consolidación**

Actividades para consolidar lo que has aprendido en esta sección.

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC210 |
| **Título** | Refuerza tu aprendizaje: Medidas de dispersión |
| **Descripción** | Actividad que refuerza las medidas de dispersión |

[SECCIÓN 1] **5 La estadística bidimensional**

**La estadística bidimensional** estudia las relaciones entre dos variables en la misma población; para esto se puede considerar la variable *X* cuyos datos son *x*1, *x*2,…, *xn* y la variable *Y* cuyos datos son *y*1, *y*2,…, *yn*. Las parejas ordenadas (*x*1, *y*1), (*x*2, *y*2),…, (*xn*, *yn*) expresan la relación entre los datos de la variable *X* con los datos de la variable *Y*. Estos datos se pueden presentar en una tabla de contingencia.

La relación entre estas dos variables se mide a través de la covarianza que se define a continuación:

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La covarianza** |
|  | **La covarianza** de una variable bidimensional que se representa como *𝜎xy*, es la diferencia entre el promedio del producto de cada uno de los datos relacionados entre sí con el producto de sus medias. Es decir:  <<MA\_09\_12\_66.gif>>  donde *x*1, *x*2,…, *xn* son los datos de la variable *X*, *y*1, *y*2,…, *yn* son los datos de la variable *Y*, *n* es el número de datos de la muestra y las respectivas medias de la variable *X* y de la variable *Y* son:  <<MA\_09\_12\_67.gif>> |

Asimismo, existe un valor numérico que permite establecer cómo se relacionan las variables que se denomina coeficiente de variación.

**La correlación:** es el valor numérico que permite identificar cómo se relacionan las variables en un análisis bidimensional.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **El coeficiente de correlación de Pearson** |
|  | El coeficiente de correlación de Pearson que se nota como *r*, se calcula mediante la fórmula:  <<MA\_09\_12\_68.gif>>  donde *𝜎XY* es la covarianza de la variable (*X*, *Y*), 𝜎x es la desviación estándar respecto a la variable *X* y 𝜎Y es la desviación estándar respecto a la variable *Y.* |

El signo de *r* está determinado por la covarianza, puesto que las desviaciones típicas de una sola variable son siempre positivas. Este signo decide el comportamiento de la correlación de esta manera:

* Si *r* > 0, la correlación es directa.
* Si *r* < 0, la correlación es inversa.
* Si *r* = 0, no hay correlación entre las variables.

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC220 |
| **Título** | Practica la estadística bidimensional |
| **Descripción** | Ejercicios que ayudan a repasar los conceptos de correlación y covarianza entre dos variables estadísticas |

[SECCIÓN 2] **5.1 Consolidación**

Actividades para consolidar lo que has aprendido en esta sección.

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC230 |
| **Título** | Refuerza tu aprendizaje: La estadística binomial |
| **Descripción** | Actividades sobre La estadística binomial |

[SECCIÓN 1] **6 Competencias**

Pon a prueba tus capacidades y aplica lo aprendido con estos recursos.

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_11\_REC240 |
| **Título** | Competencias: aplica las medidas estadísticas |
| **Descripción** |  |

[SECCIÓN 1] **Fin de unidad**

|  |  |
| --- | --- |
| **Mapa conceptual** | |
| **Código** | MA\_09\_11\_REC250 |
| **Título** | Mapa conceptual |
| **Descripción** | Mapa conceptual sobre el tema La estadística |

|  |  |
| --- | --- |
| **Practica (recurso de ejercitación)** | |
| **Código** | MA\_09\_12\_REC260 |
| **Título** | Evaluación |
| **Descripción** | Evalúa tus conocimientos sobre el tema La estadística |

|  |  |  |
| --- | --- | --- |
| **Webs de referencia** | | |
| **Código** |  | |
| **Web 01** | El análisis de los datos en estadística. | <http://liceu.uab.cat/~joaquim/phonetics/fon_met_exper/anal_datos.html> |
| **Web 02** | La distribución de frecuencias en estadística. | <http://www.ecured.cu/Tablas_de_frecuencias> |
| **Web 03** | Las medidas de posicionamiento en estadística. | [http://www.analyzemath.com/spanish/Graphing/GraphLoghttp://www.tuveras.com/estadistica/estadistica02.htm](http://www.analyzemath.com/spanish/Graphing/GraphLogarithmicFunction.html) |