|  |  |
| --- | --- |
| **Titulo del guion** | Estadística |
| **Código de guion** | MA\_G09\_12\_CO |
| **Descripción** | En un mundo globalizado se maneja mucha información, alguna de esta información puede ser recogida por medio de variables, posteriormente se puede organizar representar y analizada, pero ¿sabes cómo se puede representar y análisis? te invitamos a que conozcas como 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, ejemplo: nombre, religión, deporte favorito, género, entre otras.  **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 puede le realiza un **análisis de variable**, utilizando diferentes herramientas como: distribución de frecuencias, gráficos, histogramas, medidas de posición, medidas de dispersión, entre otros, las escogencias de las herramientas dependerán 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 construir algunas 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**.

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

Las **variables cualitativas** o también llamadas **variables categóricas** son aquellas que se expresar solo en forma de atributo, 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 variables que no permiten definir un orden jerárquico entre las posibles respuestas, por ejemplo: la variable color de cabello, las posibles respuestas serian; negro, rubio, castaño, entre otras.  **Variables cualitativas ordinales:** son las variables que permiten definir un orden específico que no es numérico, por ejemplo, para usted como 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 al número de datos.** |
|  | Según la cantidad de posibles respuestas, la variable estadística se puede clasificar en:  **Las variables dicotómicas** son aquellas que tienen solo dos respuestas posibles, ejemplo: está de acuerdo con el alza en el trasporte, 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, 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 diagrama 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, 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 los datos de la muestra. La frecuencia relativa se puede expresar como porcentaje multiplicando este cociente por 100.  **Frecuencia acumulada (*F*i):** 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 a 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. |

* **El diagrama de polígono:** describe cada uno de los datos de acuerdo a su frecuencia utilizando una curva.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_09\_12\_IMG03 |
| **Descripción** | Diagrama de polígono |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | El diagrama de polígono da cuenta de la frecuencia absoluta de cada color de ojos que tienen los estudiantes del colegio del cual se toma la muestra. |

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

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_09\_12\_IMG04 |
| **Descripción** | Grafica 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 cabello. |

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

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

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

* En el colegio la mayoría de estudiantes tiene el color de ojos castaño.
* Sólo el 1,1 % de los estudiantes del colegio tiene los ojos de color azul.
* 105 estudiantes tienen el color de ojos avellana y número 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, graficas 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 serán desarrolladas de una manera más profunda más adelante.

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

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

El **análisis de dos variables cualitativas simultáneas,** se realiza cuando estasvariables están relacionadas entre sí, esto se cumple cuando ciertos valores de una de las variables se asocian con ciertos valores de la otra variable, 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 de una variable respecto a otra, para realizar este análisis es necesario establecer si existe o no existe una relación de dependencia entre las dos variables, esto se puede realizar de una manera intuitiva observando los datos directamente y determinando si existe dicha relación o no, sin embargo es más fácil desarrollar este análisis mediante una tabla cruzada.

[SECCIÓN 3]**1.2.1 La tabla cruzada o tabla 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* x *m*, donde cada casilla de la tabla es una de las posibles categorías, por lo general, para su construcción 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 sobre peso.

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

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

Las **tablas de contingencia** permiten realizar 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 sobre peso.
2. De las 38 personas que no consumen pan todos los días 15 tienen sobrepeso y 23 no tienen sobre peso.

A continuación, se mostrará otra herramienta que es utilizada para el análisis de dos variables cuantitativas, se denomina 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 variable, los datos de las tablas marginales se encuentran plasmados en la tabla de contingencia en total columnas y total filas, por lo general, para su construcción 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 sobre peso** | **Si** | **No** | **Total filas** |
| **Si** | 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*** |
| si | 157 |
| no | 43 |

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

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

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

1. De la muestra total 157 personas tiene sobre peso 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 realizar diferentes conclusiones como:

* La probabilidad de encontrar personas con sobre peso 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.

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

La **caracterización de variables cuantitativas** hace alusión a la forma como se presentan los datos de las variables cualitativas, dependiendo de la cantidad 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 tal y como fueron recogidos, esto se aplica cuando los datos recogidos en la muestra no son muchos.

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

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.
* La moda es 14 años

**Los datos agrupados**: son aquellos datos 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, si no 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 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. Los conceptos de las frecuencias se desarrollan a continuación.

[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 al que se requiere 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.

Por 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 | 1000% |
| **Total** | 25 |  | 1 | 100% |  |  |

Es decir que para calcular la frecuencia absoluta acumulada de obtener 4 con los dados, se suman 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 que 4 al lanzar un dado.

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

[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 no tienen valores intermedios entre dos valores consecutivos: por ejemplo una variable discreta puede ser número de hijos, 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. |

Un ejemplo de una tabla de distribución de frecuencias para variables discretas se presenta a continuación:

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, por ejemplo la estatura de las personas, es una variable continua, puesto que se obtienen respuestas como: 1,70 m, 1,71 m, 1,712 m, 1,82m, 1,83 m, 1,822 m como puedes observar siempre se podrá encontrar 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 indispensable elaborar primero los intervalos en los que se desea particionar los datos.

**Agrupación en intervalos de variables continúas:**

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

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

Por ejemplo:

En la siguiente tabla se recoge la estatura de 90 personas en metros 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, se ordenan estos datos de a 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 la cantidad de intervalos (*k*):** existen varios métodos para determinar el número de intervalos que se debe particionar la muestra, los más utilizados son:

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

*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 la cantidad de elementos de la muestra no es muy grande, se aplica la formula

*k =* √*N*

donde k es el número de intervalos, que se aproxima al número entero mas 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 la cantidad de los datos recogidos es relativamente pequeña, se definen los intervalos utilizando la regla de la raíz:

*k = √N, k = √100 = 10, k = 10,* se dividirá al conjunto de datos 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 que*

*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 mediante la formula

*<<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 que

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

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

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

*l*i = *l*i – 1 + *a*

El primer intervalo [*l*0, *l*1] incluye los extremos, por lo tanto es cerrado, los demás intervalos incluyen el extremo inferior, pero no incluyen el extremo superior por lo tanto 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 lo tanto 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 sigue 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 sumando 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 idéntica

*l*3 = 1,588; *l*3 – 1 = *l*2 = 1,532, que al remplazar 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 1] **3 Las medidas de posición**

Las **medidas de posición** son aquellasque permiten reconocer las características que poseen 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 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, que 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 es el dato 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 respectivos 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:

Para calcular el promedio, se aplica 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 mediante la formula

<<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.

Por ejemplo, La media aritmética de la estatura de los 100 empleados que trabajan en la fábrica, de acuerdo con la siguiente tabla de frecuencias:

|  |  |  |  |
| --- | --- | --- | --- |
| **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** |

Se calcula mediante la fórmula así:

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

Por lo tanto, el promedio de estatura de los trabajadores esaproximadamente de 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 tal manera que haya la misma cantidad 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 la cantidad 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 la cantidad de datos n es impar, 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>>

Por lo tanto 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 mediante 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 reemplazan estos datos en la formula

*<<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 desarrolla el siguiente proceso:

1. Determinar 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 remplazan 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 periodo 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 desarrolla 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 remplazan en la formula:

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

Es decir que la moda es 6,8

[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.

[SECCIÓN 2] **3.5 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 particionan la conjunto ordenado de datos en cuatro partes iguales, de tal 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.

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 formula se deduce que el cuartil uno se encuentra 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 tienen menos de 23 años, el 50 % de las personas tienen entre 19 y 34 años, el 75 % de los encuestados tienen menos de 75 % y entre 19 y 56 corresponde al 100 % de los datos.

|  |  |
| --- | --- |
| **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.

Por ejemplo el número

Observa los siguientes ejemplos:

El número de automóviles de una conocida marca que se venden en Colombia cada día durante los últimos 55 días se presenta a continuación

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 k 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 k 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 es un número decimal

1. Aplicar la fórmula para encontrar cada uno de los deciles:

Decil uno: posición 5 que es igual a 23

Decil dos: posición 10 que es igual a 23

Decil tres: posición 15 que es igual a 34

Decil cuatro: posición 20 que es igual a 34

Decil cinco: posición 25 que es igual a 42

Decil seis: posición 30 que es igual a 45

Decil siete: posición 35 que es igual a 45

Decil ocho: posición 40 que es igual a 56

Decil nueve: posición 45 que es igual a 67

|  |  |
| --- | --- |
| **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. |

Cuando los datos no son agrupados un método para alguno de los 99 valores de los percentiles es el siguiente:

Ordenar los datos de menor a mayor.

1. Determinar el lugar que ocupa en el conjunto ordenado de datos cada uno de los percentiles mediante la formula:

<<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

[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.**

[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*máximo*es el dato mayor y x*mínimoes el dato menor de la distribución. |

Por ejemplo, el peso en kilos de cada uno de los 15 estudiantes hombres de grado noveno 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 definir ¿qué es la desviación con respecto a la media?

La **desviación con respecto a la media (Di)** se puede definir como la diferencia en valor absoluto que existe entre cada uno de los valores de la muestra y la media aritmética de los datos

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La** **desviación con respecto a la media** |
|  | Se define como:  Di = |xi - (X) |  Donde:  xi: es cualquier valor de los datos  (X): es la media aritmética de los datos. |

La **desviación media (D(x))** se define como; la media aritmética que se establece entre los valores absolutos de las **desviaciones con respecto a la media.**

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La** **desviación media** |
|  | *Cuando los datos no son agrupados se define como:*  *<<MA\_09\_12\_35.gif>>*  *Donde:*  *x1, x2……xn  son los valores de los datos*  *(X): es la media aritmética de los valores de los datos.*  *n: es el número de datos.*  *Cuando los datos están agrupados en una tabla de frecuencia se define como:*  *<<MA\_09\_12\_36.gif>>*  *Donde:*  *x1, x2……xn  son los valores medios de los intervalos.*  *(X): es la media aritmética de los valores de los datos.*  *f1, f2…….fn son las frecuencias absolutas de cada intervalo*  *n: es el número de datos.* |

Observa el siguiente ejemplo:

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1,70 | 1,60 | 1,50 | 1,56 | 1,69 |

Donde:

(x) = 1,61

n = 5

x1 = 1,50 x2 = 1,56 x3 = 1,60 x4 = 1,69 x5 = 1,70

Se remplaza en la fórmula:

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

Se obtiene que la deviación media D(X) = 0,068

SECCIÓN 2] **4.3 La varianza**

La **varianza (σ2)** se define como la media aritmética del cuadrado de las **desviaciones con respecto a la media.**

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La** **varianza** |
|  | *Cuando los datos no son agrupados se define como:*  *<<MA\_09\_12\_38.gif>>*  *Donde:*  *x1, x2……xn  son los valores de los datos*  *(X): es la media aritmética de los valores de los datos.*  *n: es el número de datos.*  *Cuando los datos están agrupados en una tabla de frecuencia se define como:*  *<<MA\_09\_12\_39.gif>>*  *Donde:*  *x1, x2……xn  son los valores medios de los intervalos*  *(X): es la media aritmética de los valores de los datos.*  *f1, f2…….fn son las frecuencias absolutas de cada intervalo*  *n: es el número de datos.* |

Observa el siguiente ejemplo:

* Calcule la variancia de las edades de 2500 encuestados que se encuentran consignados en la siguiente tabla de frecuencia:

|  |  |  |
| --- | --- | --- |
| **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 |  |

* Se encuentran los valores medios de cada intervalo:

x1 = 20,5

x2 = 25,5

x3 = 30,5

x4 = 35,5

x5 = 40,5

x6 = 45,5

* Se encuentra la media aritmética de los datos multiplicando los valores medios de cada intervalo por su respectiva frecuencia, posteriormente se suman y el resultado se dividen por el número de datos:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **li , li-1** | **xi** | **fi** | **xi .fi** | **Fi** |
| [18-23] | 20,5 | 520 | 10660 | 520 |
| (23-28] | 25,5 | 620 | 15810 | 1140 |
| (28-33] | 30,5 | 410 | 12505 | 1550 |
| (33-38] | 35,5 | 325 | 11537,5 | 1875 |
| (38-43] | 40,5 | 300 | 12150 | 2175 |
| (43-48] | 45,5 | 325 | 14787,5 | 2500 |
| total |  | 2500 | 77450 |  |

La media aritmética de los datos es (X) = 30,9

Ahora se remplazan los datos en la fórmula:

*<<MA\_09\_12\_40.gif>>*

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

Se obtiene que la varianza es σ2 = 2,3

La siguiente medida de dispersión parte del resultado obtenido en la varianza.

SECCIÓN 2] **4.4 La desviación típica**

La **desviación típica (σ)** se define como la raíz cuadrada de la **varianza.**

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La**  **desviación típica** |
|  | *Cuando los datos no son agrupados se define como:*  *<<MA\_09\_12\_42.gif>>*  *Donde:*  *x1, x2……xn  son los valores de los datos*  *(X): es la media aritmética de los valores de los datos.*  *n: es el número de datos.*  *Cuando los datos están agrupados en una tabla de frecuencia se define como:*  *<<MA\_09\_12\_43.gif>>*  *Donde:*  *x1, x2……xn  son los valores medios de los intervalos*  *(X): es la media aritmética de los valores de los datos.*  *f1, f2…….fn son las frecuencias absolutas de cada intervalo*  *n: es el número de datos.* |

Ejemplo:

* Calcule la desviación típica de las edades de 2500 encuestados que se encuentran consignados en la siguiente tabla de frecuencia:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **li , li-1** | **xi** | **fi** | **xi .fi** | **Fi** |
| [18-23] | 20,5 | 520 | 10660 | 520 |
| (23-28] | 25,5 | 620 | 15810 | 1140 |
| (28-33] | 30,5 | 410 | 12505 | 1550 |
| (33-38] | 35,5 | 325 | 11537,5 | 1875 |
| (38-43] | 40,5 | 300 | 12150 | 2175 |
| (43-48] | 45,5 | 325 | 14787,5 | 2500 |
| total |  | 2500 | 77450 |  |

Se obtiene que la varianza de los datos es σ2 = 2,3 ahora se le saca la raíz cuadrada para obtener la desviación típica:

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

Donde la desviación típica será σ = 1,5

La siguiente medida de dispersión se basa en el resultado obtenido en la varianza típica.

SECCIÓN 2] **4.5 El coeficiente de varianza**

El **coeficiente de varianza** **(C.V)** se define como el cociente entre el resultado obtenido en la **desviación típica** de una muestra y su **media aritmética.**

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Coeficiente de varianza (C.V)** |
|  | *Se utiliza la siguiente formula, donde se suele expresar en forma de porcentaje:*  *<<MA\_09\_12\_45.gif>>* |

Observa el siguiente ejemplo:

* De una muestra se obtiene que su desviación típica es de *σ = 1,5* y su media aritmética es (X)= 30,9 se debe calcular su coeficiente de varianza.

Se remplazan los valores en la fórmula:

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

Se obtiene que el coeficiente de varianza que es C.V = 4,8%

Como puedes observar las medidas de dispersión se encargan de brindar información que permite definir qué tan dispersos se encuentran los datos de una muestra.

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

SECCIÓN 1] **5 resolución de problemas estadísticos**

Los problemas estadísticos surgen cuando se quiere analizar el comportamiento de una población por medio de variables, que puede ser cuantitativas o cualitativas, dicho análisis se puede realizar empleando:

* El análisis de variables.
* Tablas de distribución de frecuencia.
* Las medidas de posición.
* Las medidas de dispersión.

Observa los siguientes ejemplos:

* En una encuesta realizada a 2500 personas se realizaron las siguientes preguntas:

1. Sexo, los resultado fueron:

Mujer 1200, hombre 1300

1. Tiene cabellos, los resultados fueron:

Si 1100 mujeres, si 941 hombres, no 100 mujeres, no 359 hombres.

1. Rango de edad, los resultados fueron:

(18-23] fueron 320

(23-28] fueron 500

(28-33] fueron 380

(33-38] fueron 270

(38-43] fueron 220

(43-48] fueron 215

(48-53] fueron 204

(53-58] fueron 191

(58-63] fueron 200

1. Cuantos hijos tiene: los resultados fueron:

0 hijos 618, 1 hijo 820, 2 hijos 400, 3 hijos 320, 4hijos 120, 5 hijos 100, 6 hijos 50, 7 hijos 40, 8 hijos 20 personas,

|  |  |
| --- | --- |
| **Número de hijos** | **Frecuencia** |
| 0 | 618 |
| 1 | 820 |
| 2 | 400 |
| 3 | 320 |
| 4 | 120 |
| 5 | 100 |
| 6 | 50 |
| 7 | 40 |
| 8 | 32 |
| **total** | **2500** |

Resuelve:

1. Qué clase de variables se manejaron en la encuesta:

* La pregunta uno: es una variable cualitativa nominal.
* La pregunta dos: es una variable cualitativa nominal.
* La pregunta tres: variable cuantitativa discreta.
* La pegunta cuatro: variable cuantitativa discreta.

1. Crear una tabla cruzada para los datos recogidos en la pregunta uno y la pregunta dos.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Sexo** | |  |
| **Tiene cabello** | **mujer** | **Hombre** | **Total filas** |
| **Si** | 1100 | 941 | **2041** |
| **No** | 100 | 359 | **459** |
| **Total columnas** | **1200** | **1300** | **2500** |

1. Crear las tablas marginales para los resultados de la pregunta uno y la pregunta dos:

Pregunta uno:

|  |  |
| --- | --- |
| **Sexo** | ***fi*** |
| Mujer | 1200 |
| Hombre | 1300 |

Pregunta dos:

|  |  |
| --- | --- |
| **Tiene cabello** | ***fi*** |
| si | 2041 |
| no | 459 |

1. Crear la tabla de distribución de frecuencia para los datos recogidos con la pregunta numero tres.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Clase** | **Intervalos** | **Frecuencia absoluta** | **Frecuencia relativa** | **Porcentaje**  **ni . 100%** | **Frecuencia absoluta acumulada** | **Frecuencia relativa acumulada** |
| **X** | **(li , li-1 ]** | **fi** | **ni** | **%** | **Fi** | **Ni** |
| **1** | [18-23] | 320 | 0,128 | 12,8% | 320 | 0,128 |
| **2** | (23-28] | 500 | 0,2 | 20% | 820 | 0,328 |
| **3** | (28-33] | 380 | 0,152 | 15,2% | 1200 | 0,48 |
| **4** | (33-38] | 270 | 0,108 | 10,8% | 1470 | 0,588 |
| **5** | (38-43] | 220 | 0,088 | 8,8% | 1690 | 0,676 |
| **6** | (43-48] | 215 | 0,086 | 8,6% | 1905 | 0,762 |
| **7** | (48-53] | 204 | 0,0816 | 8,16% | 2109 | 0,8436 |
| **8** | (53-58] | 191 | 0,0764 | 7,64% | 2300 | 0,92 |
| **9** | (58-63] | 200 | 0,08 | 8% | 2500 | 1 |

1. Calcule las medidas de posición central, para los datos recogidos en la pregunta numero tres:

* Media aritmética (X):

|  |  |  |  |
| --- | --- | --- | --- |
| **Intervalos** | **Frecuencia absoluta** | **Punto medio intervalo** | **Promedio x frecuencia** |
| **li , li-1** | **fi** | **Pi** | **Pi .fi** |
| [18-23] | 320 | 20,5 | 6560 |
| (23-28] | 500 | 25,5 | 12750 |
| (28-33] | 380 | 30,5 | 11590 |
| (33-38] | 270 | 35,5 | 9585 |
| (38-43] | 220 | 40,5 | 8910 |
| (43-48] | 215 | 45,5 | 9782,5 |
| (48-53] | 204 | 50,5 | 10302 |
| (53-58] | 191 | 55,5 | 10600,5 |
| (58-63] | 200 | 60,5 | 12100 |
| **Total** | **2500** |  | **92180** |

Utilizando la fórmula para la media aritmética:

*<<MA\_09\_12\_47.gif>>*

Se obtiene que la media *(X) = 36,872*

* La mediana (Me)

Como los datos so agrupados:

La mitad de la frecuencia absoluta es 1250 y se encuentra en el intervalo (33-38], se define: *L1 = 33, a1 = 5, Fi-1 = 1200, fi = 270*

*<<MA\_09\_12\_48.gif>>*

Se obtiene que la mediana *(Me) =33,92*

* La moda (Mo)

Como los datos son agrupados:

El intervalo modal es (23-28] y se define:

*, a1 = 5, L1 = 23, fi = 500,* fi-1 = 320, fi+1 = 270

*<<MA\_09\_12\_49.gif>>*

Se obtiene que la moda *(Mo) = 25,19*

1. Calcule los cuartiles para los datos recogidos en la pregunta cuatro:

El primer cuartil:

*<<MA\_09\_12\_50.gif>>*

Se encuentra en la posición 625, 1 hijo

El segundo cuartil:

Se encuentra en la posición 1250, 2 hijos.

El tercer cuartil:

Se encuentra en la posición 1875, 3 hijos.

1. Calcular el rango y la desviación media, para los datos recogidos en la pregunta 4:

* El Rango es igual a:

*R = 9-0 = 8*

Rango R *= 8*

* La desviación media:

Se calcula media aritmética:

|  |  |  |
| --- | --- | --- |
| **xi** | **fi** | **fi. xi** |
| 0 | 618 | 0 |
| 1 | 820 | 820 |
| 2 | 400 | 800 |
| 3 | 320 | 960 |
| 4 | 120 | 480 |
| 5 | 100 | 500 |
| 6 | 50 | 300 |
| 7 | 40 | 280 |
| 8 | 32 | 256 |
| **total** | **2500** | **4396** |

*<<MA\_09\_12\_51.gif>>*

La media aritmética es: *(X) = 1,7*

Se remplaza en la formula:

*<<MA\_09\_12\_51.gif>>*

Se obtiene que la desviación media es D((x)) = 1,3

Como puedes observar estos son algunos de los problemas estadísticos que se pueden presentar, te invitamos a que investigues un poco más sobre los problemas que se pueden resolver utilizando como herramienta los conocimientos estadísticos adquiridos en este curso

|  |  |  |
| --- | --- | --- |
| **Webs de referencia** | | |
| **Código** |  | |
| **Web 01** | Es una página en la cual se desarrolla n estudio profundo sobre el análisis de los datos en estadística. | <http://liceu.uab.cat/~joaquim/phonetics/fon_met_exper/anal_datos.html> |
| **Web 02** | El tema centrar es la distribución de frecuencia en estadística. | <http://www.ecured.cu/Tablas_de_frecuencias> |
| **Web 03** | Se trabajan 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) |