|  |  |
| --- | --- |
| **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, genero, entre otras.  **Variables cuantitativas:** son aquellas variables que toman valores numéricos, ejemplo: edad, peso, altura, precio entre otras. |

Al conjunto de datos que se obtiene luego de recoger la información se puede realizar 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 decisión, a continuación, se presentan algunas herramientas que permiten realizar el análisis de **variables cualitativas**.

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

Recuerda que 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 estadísticas 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 mas de dos respuestas posibles, por ejemplo: cuál es el color de sus ojos, las posibles respuestas son más de dos. |

Algunas herramientas que se puede 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.

**Tabla de frecuencias:** también conocida como 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 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**

* **Diagrama de barras:** describir cada uno de los datos de acuerdo a su frecuencia utilizando barras.

|  |  |
| --- | --- |
| **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** | *Diagrama de barras de:* Color castaño 190, color ámbar 140, color avellano 105, color verde 10, color azul 5. |

* **Diagrama de polígono:** describir 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** | *Diagrama de polígono:* Color castaño 190, color ámbar 140, color avellano 105, color verde 10, color azul 5. |

* **Grafica de sectores o torta:** describir cada uno de los datos de acuerdo a su frecuencia utilizando el arrea correspondiente en una torta, 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** | *Diagrama de torta:* Color castaño 190, color ámbar 140, color avellano 105, color verde 10, color azul 5. |

**La moda:** es el dato que más se repite, en este caso es color castaño con

Color castaño con 190.

Estas son 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:

* Que en el colegio la mayoría de estudiantes tiene el color de ojos castaña.
* Que solo el 1,1 % de los estudiantes del colegio tiene los ojos de color azul.
* Que 105 estudiantes tiene el color de ojos avellana y que este 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 realizaría 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**

Cuando se pretenden realizar un **Análisis de dos variables cualitativas simultáneas,** dichas variables deberán estar relacionadas entre sí, esto se cumple cuando ciertos valores de una de las variables se asocian con ciertos valores de la otra variable, ejemplo: ¿si consumo pan tiene sobrepeso? ¿si fuma tiene cáncer? ¿el consumo de licor con su estado de salud?

Cuando se quiere realiza el análisis de dos variables cualitativas simultáneas se busca describir como el comportamiento de una de las variables se relaciona con el comportamiento de la otra variable, con el objetivo de buscar información que permita generar conclusiones.

Para realizar dicho 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, pero sería más fácil desarrollar este análisis utilizando una tabla cruzada y una tabla de contingencia.

[SECCIÓN 3]**1.2.1 Tabla cruzada o tabal de contingencia**

La **tabla de contingencia** es una herramienta que permite estudiar la información recogida con dos variables dentro de una población determinada, donde los valores de la primera variable se podrán dividir en *n* clases (filas) y los valores de la segunda variable en *m* clases (columnas), dichos valores podrán ser clasificados y ordenados en una tabla de doble entrada de dimensión *n* x *m*, que serán las posibles categorías que existan, por lo general para su construcción se utiliza las frecuencias absoluta de las dos variables, esto permite describir el comportamiento de las dos variables gracias a la combinación y crucé de sus categorías, con el fin de determinando si existe algún tipo de relación de dependencia entre las dos variables, observa el siguiente ejemplo donde se crea una tabla de 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 fueron los siguientes:

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

Como se pude observar las **tablas de contingencia** permite realizar una observación más detallada y determinar si existe una relación de dependencia entre las dos variables, como se puede ver en la tabla si hay una relación de dependencia entre consumir pan todos los días y tener sobre peso, de la tabla de contingencia se puede decir:

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

Estos datos permitirán realizar diferentes conclusiones como:

* Es más probable tener sobre peso si se consume pan todos los días.
* Es menos probable tener sobre peso si no se consume pan todo los días.

Entre otras conclusiones, como puedes ver la tabla de contingencia es una poderosa herramienta cuando se pretende realizar el análisis de dos variables, a continuación, se mostrará otra herramienta que es utilizada para el análisis de dos variables cuantitativas, la cual se denomina tablas marginales.

[SECCIÓN 3]**1.2.2 Tabla marginales**

La **tabla marginal** es aquella 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, pero estos datos se pueden observar mejor en tablas apartes, por eso se construyen dos **tablas marginales**, una para cada variable, por lo general para su construcción se utiliza la frecuencia absoluta de cada una de las variables, las cuales permitirán describir el comportamiento de cada una de las variables, observa el siguiente ejemplo:

* 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 fueron los siguientes:

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

Distribuciones marginales:

|  |  |
| --- | --- |
| **Tiene sobre peso** | ***fi*** |
| si | 157 |
| no | 43 |

|  |  |
| --- | --- |
| **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 sobre peso.

Como puedes observar las tablas de contingencias y las tablas marginales son herramientas que brindan al estadista la capacidad de organizar, clasificar y analizar la información recogida utilizando dos variables de tipo cualitativo, para arrojar algún o algunas conclusiones.

En la siguiente sección el trabajo se centrar en las variables cuantitativas, las cuales hasta el momento no han sido trabajadas.

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

Cuando se habla de la **caracterización de variables cuantitativas** en realidad se hace alusión a la forma como se presentan los datos, dependiendo de la cantidad de datos que se recogen se pueden definir dos formas:

**Datos no agrupados:** son los datos que se presenta tal y como fueron recogidos, esto se hace cuando los datos recogidos en la muestra no son muchos, y no hay necesidad de realizar una clasificación de dichos datos ya que se pueden sacar conclusiones realizando una observación directa.

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

**Datos agrupados**: son aquellos datos que pueden ser divididos en clases, rangos, grupos 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 presentar estos datos de forma agrupado pueden ser: las tablas de distribución de frecuencia, los histogramas, polígonos de frecuencia, la ojiva, entre otras herramientas.

|  |  |
| --- | --- |
| **Recuerda** | |
| Contenido | **Variables cuantitativas puede ser:**  **Variables cuantitativas discretas:** son aquellas que toma valores dentro de un conjunto dado en el cual no hay valores intermedios entre dos valores específicos: por ejemplo las edades de 3 personas 30, 31 y 32.    **Variables cuantitativas continuas:** son aquellas que pueden tomar valores infinitos entre dos dato cualquiera, por ejemplo la altura de 3 personas dada en metros: 1,70 1,72 1,733. |

En la siguiente sección el trabajo se centrará en las tablas de distribución de frecuencias, ya que es una de las herramientas más utilizadas para poder presentar los datos de forma agrupada.

[SECCIÓN 1] **2 Tablas de distribución de frecuencia**

Las **tablas 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 veces que se presenta cada uno de ellos, los elementos que hacen parte de la tabla de frecuencia son:

* Variable
* Datos
* Frecuencia absoluta
* Frecuencia absoluta acumulada
* Frecuencia relativa
* Frecuencia relativa acumulada

|  |  |
| --- | --- |
| **Recuerda** | |
| **Contenido** | **Variable (x):** es la característica que se pretende medir, por ejemplo:   * la edad de un grupo de personas. * la estatura de un grupo de personas. * Los números que salen cuando se lanza un dado.     **Datos (n):** son los valores observados en la variable, por ejemplo:   * En el grado novena hay 30 estudiantes sus edades se presentan a continuación   12,13,14,15,16,12,14,14,15,16,17,12,13,14,15,16,14,16,17,14,15,13,12,12,14,15,13,14,14,14  Cada una de las edades representa un dato es decir que se tiene 35 datos.   * 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  Cada uno de los resultados obtenidos en el lanzamiento es un dato, en total encontramos 25 datos |

Los conceptos de las frecuencias como tal serán desarrollados a continuación.

[SECCIÓN 2] **2.1 frecuencia absoluta, relativa y acumulada**

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

Observa los siguientes ejemplos:

* En el grado novena hay 30 estudiantes sus edades se presentan a continuación

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

Esta sería la representación de las frecuencias absolutas:

|  |  |
| --- | --- |
| **x** | **fi** |
| 12 | 5 |
| 13 | 4 |
| 14 | 10 |
| 15 | 5 |
| 16 | 4 |
| 17 | 2 |
| **Total** | **30** |

* 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

Cada uno de los resultados obtenidos en el lanzamiento es un dato, en total encontramos 25 datos.

Esta sería la representación de las frecuencias absolutas:

|  |  |
| --- | --- |
| **x** | **fi** |
| 1 | 5 |
| 2 | 4 |
| 3 | 3 |
| 4 | 3 |
| 5 | 4 |
| 6 | 6 |
| **Total** | **25** |

La **frecuencia absoluta acumulada (Fi)** es la suma de las **frecuencias absolutas,** dada su ubicación desde las frecuencias absolutas de la parte superior hasta las frecuencias absolutas de la parte inferior,observa los siguientes ejemplos:

* En el grado novena hay 30 estudiantes sus edades se presentan a continuación

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

|  |  |  |
| --- | --- | --- |
| **X** | **fi** | **Fi** |
| 12 | 5 | 5 |
| 13 | 4 | (5+4)=9 |
| 14 | 10 | (9+10)=19 |
| 15 | 5 | (19+5)=24 |
| 16 | 4 | (24+4)=28 |
| 17 | 2 | (28+2)=30 |
| **Total** | **30** |  |

* 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

Cada uno de los resultados obtenidos en el lanzamiento es un dato, en total encontramos 25 datos

|  |  |  |
| --- | --- | --- |
| **X** | **fi** | **Fi** |
| 1 | 5 | 5 |
| 2 | 4 | (5+4)=9 |
| 3 | 3 | (9+3)=12 |
| 4 | 3 | (12+3)=15 |
| 5 | 4 | (15+4)=19 |
| 6 | 6 | (19+6)=25 |
| **Total** | **25** |  |

La **frecuencia relativa (ni)** es el cociente entre la **frecuencia absoluta** y el número total de datos, se puede presentar de forma decimal o de forma tanto por ciento, 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 forma tanto por ciento es igual a 100%, observa los siguientes ejemplos:

* En el grado novena hay 30 estudiantes sus edades se presentan a continuación

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

|  |  |  |
| --- | --- | --- |
| **X** | **fi** | **ni** |
| 12 | 5 | 0,16 o 16% |
| 13 | 4 | 0,13 o 13% |
| 14 | 10 | 0,33 o 33% |
| 15 | 5 | 0,16 o 16% |
| 16 | 4 | 0,13 o 13 % |
| 17 | 2 | 0,06 o 6% |
| **Total** | **30** | **1 o 100%** |

* 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

Cada uno de los resultados obtenidos en el lanzamiento es un dato, en total encontramos 25 datos

|  |  |  |
| --- | --- | --- |
| **X** | **fi** | **ni** |
| 1 | 5 | 0,2 o 20% |
| 2 | 4 | 0,16 o 16% |
| 3 | 3 | 0,12 o 12% |
| 4 | 3 | 0,12 o 12% |
| 5 | 4 | 0,16 o 16% |
| 6 | 6 | 0,24 o 24% |
| **Total** | **25** | **1 o 100%** |

La **frecuencia relativa acumulada (Ni)** es el cociente entre la **frecuencia absoluta acumulada** y el número total de datos, se puede presentar de forma decimal o de forma tanto por ciento, observa los siguientes ejemplos:

* En el grado novena hay 30 estudiantes sus edades se presentan a continuación

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

|  |  |  |  |
| --- | --- | --- | --- |
| **X** | **fi** | **Fi** | **Ni** |
| 12 | 5 | 5 | 0,16 o 16% |
| 13 | 4 | (5+4)=9 | 0,3 o 30% |
| 14 | 10 | (9+10)=19 | 0,63 o 63% |
| 15 | 5 | (19+5)=24 | 0,8 o 80% |
| 16 | 4 | (24+4)=28 | 0,93 o 93% |
| 17 | 2 | (28+2)=30 | 1 o 100% |
| **Total** | **30** |  |  |

* 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

Cada uno de los resultados obtenidos en el lanzamiento es un dato, en total encontramos 25 datos

|  |  |  |  |
| --- | --- | --- | --- |
| **X** | **fi** | **Fi** | **Ni** |
| 1 | 5 | 5 | 0,2 o 20% |
| 2 | 4 | (5+4)=9 | 0,36 o 36% |
| 3 | 3 | (9+3)=12 | 0,48 o 48% |
| 4 | 3 | (12+3)=15 | 0,6 o 60% |
| 5 | 4 | (15+4)=19 | 0,76 o 76% |
| 6 | 6 | (19+6)=25 | 1 o 100% |
| **Total** | **25** |  |  |

Estas son las diferentes frecuencias que se pueden establecer en torno a los variables cuantitativas, ahora se deberán plantear todas las frecuencias en una sola tabla, la forma de dicha tabla dependerá de la clase de variable sea, variable discreta o variable continua.

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

Recuerda que las **variables discretas** son aquellas que solo pueden tomar valores dentro de un conjunto definido en el cual sus elementos tendrán un orden, y no existirá un elemento entre otros dos elementos consecutivos, se puede decir que estas variables surge de un conteo, 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.

Las **tablas de frecuencia** para las **variables discretas** podrán tener la siguiente estructura:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Frecuencia absoluta** | **Frecuencia relativa** | **Porcentaje**  **ni . 100%** | **Frecuencia absoluta acumulada** | **Frecuencia relativa acumulada** |
| **X** | **fi** | **ni** | **%** | **Fi** | **Ni** |

Observa el siguiente ejemplo:

* Se realizó una encuesta a 40 personas del barrio 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 frecuencia** quedaría de la siguiente manera:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **X** | **fi** | **ni** | **%** | **Fi** | **Ni** |
| 0 | 8 | 0,2 | 20% | 8 | 0,2 |
| 1 | 4 | 0,1 | 10% | 12 | 0,3 |
| 2 | 8 | 0,2 | 20% | 20 | 0,5 |
| 3 | 4 | 0,1 | 10% | 24 | 0,6 |
| 4 | 6 | 0,15 | 15% | 30 | 0,75 |
| 5 | 5 | 0,125 | 12,5% | 35 | 0,875 |
| 6 | 4 | 0,1 | 10% | 39 | 0,975 |
| 7 | 1 | 0,025 | 2,5% | 40 | 1 |
| **total** | **40** | **1** | **100%** |  |  |

Esta puede ser una de las estructuras que puede tener una tabla de frecuencia completa para variables discretas, pero ¿tendrá la misma estructura la tabla de frecuencia cuando la variable sea continua?, la respuesta se presentara a continuación.

|  |  |
| --- | --- |
| **Recuerda** | |
| **Contenido** | *Recuerda que las* ***variables continuas*** *son aquellas que podrán tomar valores dentro de un conjunto definido, donde siempre se podrá encontrar un elemento entre dos elementos dados, se puede decir que estas variables surge de una medición, por ejemplo la estatura de las personas, se obtendrá respuestas como: 1.70 mt, 1.71 mt, 1.712 mt, 1.82mt, 1,83 mt, 1.822mt como puedes observar siempre se podrá encontrar un elemento intermedio entre dos elementos dados.* |

Las **tablas de frecuencia** para las **variables continuas** en algunas ocasiones tendrá la misma estructura que tiene la tabla de frecuencia de **variables discretas**, esto se presenta cuando los datos recogidos con la variable continua se puedan enumerar fácilmente, pero en otras ocasiones los valores recogidos con la **variable continua** tiene muchos valores, y resulta muy dispendioso crear una categoría por cada valar que la variable pueda tomar, dichos valores están comprendidos entre un valor inicial y un valor final, por ese motivo se debe crear los intervalos y realizar las diferentes distribuciones.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | La distribución de **variables continuas** que tiene muchos valores distintos |
|  | *Se puede crear de la siguiente manera:*   1. *Organizar los valores de menor a mayor o de mayor a menor.* 2. *Determinar la cantidad de intervalos k , utilizando la regla de sturgers: k = 1+3,322.logN, cuando los datos recogidos en realidad son muy grandes o la regla de la raíz k = √N, cuando los datos no son tantos, donde N es la cantidad de datos.* 3. *Se determina el rango de los datos de la siguiente manera:*   *R= xmayor – xmenor*   1. *Se debe determinar la amplitud del intervalo que se definirá como:*   *<<MA\_09\_12\_01.gif>>*  *Redondeando el número que dé como resultado para facilitar los cálculos.*   1. *Se determinar los extremos de los intervalos los cuales serán: l0 = x menor, l1 = l0 + a, l2 = l1 + a, de modo generar li = li-1 +a, para ubicarlos en la tabla, los intervalos serán semiabiertos a izquierda, excepto el primero que será cerrado.*      1. *Crear la tabla de distribuciones que podrá tener la siguiente estructura:*  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **Clase** | **Intervalos** | **Frecuencia absoluta** | **Frecuencia relativa** | **Porcentaje**  **ni . 100%** | **Frecuencia absoluta acumulada** | **Frecuencia relativa acumulada** | | **k** | **(li , li-1 ]** | **fi** | **ni** | **%** | **Fi** | **Ni** | |

Observa el siguiente ejemplo:

En la siguiente tabla se recoge la estatura de 100 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 |
| 1,78 | 1,55 | 1,45 | 1,77 | 1,43 | 1,55 | 1,78 | 1,54 | 1,76 | 1,54 |

1. Ordenar los valores en este caso de mayor a menor

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1,98 | 1,88 | 1,78 | 1,77 | 1,69 | 1,67 | 1,65 | 1,56 | 1,54 | 1,45 |
| 1,98 | 1,88 | 1,78 | 1,77 | 1,69 | 1,67 | 1,65 | 1,55 | 1,54 | 1,45 |
| 1,95 | 1,87 | 1,78 | 1,77 | 1,68 | 1,67 | 1,65 | 1,55 | 1,53 | 1,45 |
| 1,92 | 1,87 | 1,78 | 1,76 | 1,68 | 1,67 | 1,64 | 1,55 | 1,51 | 1,45 |
| 1,9 | 1,85 | 1,78 | 1,74 | 1,68 | 1,66 | 1,58 | 1,55 | 1,51 | 1,45 |
| 1,9 | 1,83 | 1,77 | 1,73 | 1,67 | 1,66 | 1,57 | 1,54 | 1,50 | 1,45 |
| 1,89 | 1,81 | 1,77 | 1,72 | 1,67 | 1,65 | 1,56 | 1,54 | 1,49 | 1,44 |
| 1,89 | 1,81 | 1,77 | 1,72 | 1,67 | 1,65 | 1,56 | 1,54 | 1,48 | 1,44 |
| 1,89 | 1,78 | 1,77 | 1,70 | 1,67 | 1,65 | 1,56 | 1,54 | 1,47 | 1,43 |
| 1,89 | 1,78 | 1,77 | 1,70 | 1,67 | 1,65 | 1,56 | 1,54 | 1,45 | 1,42 |

1. Como los datos recogidos no son tantos se definirán los intervalos utilizando la regla de la raíz:

*k = √N, k = √100 = 10,* se dividirá en 10 intervalos, *k = 10*

1. Se determina el rango:

*R= xmayor – x*menor,  *R = 1,98 – 1,42 = 0,56* el rango es de *R = 0,56.*

1. Se encuentra la amplitud que tendrá el intervalo:

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

1. Se definen los 10 intervalos:

[1,42 – 1,47], (1,47 – 1,53], (1,53-1,58],(1,58-1,64], (1,64-1,70], (1,70- 1,75], (1,75-1,81], (1,81-1,86], (1,86-1,92], (1,92-1,98]

1. Se crea la tabla de frecuencias:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Clase** | **Intervalos** | **Frecuencia absoluta** | **Frecuencia relativa** | **Porcentaje**  **ni . 100%** | **Frecuencia absoluta acumulada** | **Frecuencia relativa acumulada** |
| **K** | **li , li-1** | **fi** | **ni** | **%** | **Fi** | **Ni** |
| **1** | [1,42 – 1,47] | 12 | 0,12 | 12% | 12 | 0,12 |
| **2** | (1,47 – 1,53] | 6 | 0,06 | 6% | 18 | 0,18 |
| **3** | (1,53-1,58] | 18 | 0,18 | 18% | 36 | 0,36 |
| **4** | (1,58-1,64] | 1 | 0,01 | 1% | 37 | 0,37 |
| **5** | (1,64-1,70] | 25 | 0,25 | 25% | 62 | 0,62 |
| **6** | (1,70- 1,75] | 4 | 0,04 | 4% | 66 | 0,66 |
| **7** | (1,75-1,81] | 18 | 0,18 | 18% | 84 | 0,84 |
| **8** | (1,81-1,86] | 2 | 0,02 | 2% | 86 | 0,86 |
| **9** | (1,86-1,92] | 11 | 0,11 | 11% | 97 | 0,97 |
| **10** | (1,92-1,98] | 3 | 0,03 | 3% | 100 | 1 |
| **Total** |  | **100** | **1** | **100%** |  |  |
|  |  |  |  |  |  |  |

Como puedes observas esta es una de las formas que existe para crear una tabla de frecuencias para una variable continua, existen otros formas las cuales podrás consultar en libros de estadística o en la internet, en la siguiente sección el trabajo girara entorno a las medidas de posición, las cuales pueden ser vistas como herramientas que son utilizadas para analizar las características que pueden tener los datos recogidos con las variables.

[SECCIÓN 1] **3 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 pueden dividir en dos clases:

**Medidas de posición central:** son las medidas que tienden a localizar la parte media del conjunto de datos ordenados de una variable cuantitativa, algunas de ellas serán desarrolladas a continuación.

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

La **media aritmética (X)** se encarga de mostrar el promedio de un conjunto de datos, la forma general para ser calculada es: sumando los valores de los datos y se divide por el número de datos recogidos.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Formula de la media aritmética** |
|  | * *Cundo los datos no son agrupados:*   *(x) es la media aritmética*  *x1, x2,……xn son los datos recogidos.*  *N número de datos recogidos*    *<<MA\_09\_12\_03.gif>>*   * *Cuando los datos son agrupados:*   *(x) es la media aritmética*  *x1, x2,……xn son los puntos medios de los intervalos.*  *f1, f2………fn son las frecuencias absolutas de cada intervalo*  *N número de datos recogidos*  *<<MA\_09\_12\_04.gif>>* |

Observa los siguientes Ejemplos:

* Cuál será 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.

Utilizando la formula:

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

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

* cuál será el promedio de la estatura de los 100 empleados que trabajan en la fábrica si se tiene la siguiente tabla de frecuencias:

|  |  |  |
| --- | --- | --- |
| **Clase** | **Intervalos** | **Frecuencia absoluta** |
| **K** | **li , li-1** | **fi** |
| **1** | [1,42 – 1,47] | 12 |
| **2** | (1,47 – 1,53] | 6 |
| **3** | (1,53-1,58] | 18 |
| **4** | (1,58-1,64] | 1 |
| **5** | (1,64-1,70] | 25 |
| **6** | (1,70- 1,75] | 4 |
| **7** | (1,75-1,81] | 18 |
| **8** | (1,81-1,86] | 2 |
| **9** | (1,86-1,92] | 11 |
| **10** | (1,92-1,98] | 3 |
| **Total** |  | **100** |

Cuando los datos son agrupados o definidos en intervalo, se debe calcular el punto medio del intervalo y multiplicarlo por su respectiva frecuencia, posteriormente estos resultados se suman y se dividen por la cantidad de datos:

|  |  |  |  |
| --- | --- | --- | --- |
| **Intervalos** | **Frecuencia absoluta** | **Punto medio intervalo** | **Promedio x frecuencia** |
| **li , li-1** | **fi** | **Pi** | **Pi .fi** |
| [1,42 – 1,47] | 12 | 1,445 | 17,34 |
| (1,47 – 1,53] | 6 | 1,50 | 9 |
| (1,53-1,58] | 18 | 1,555 | 27,99 |
| (1,58-1,64] | 1 | 1,61 | 1,61 |
| (1,64-1,70] | 25 | 1,67 | 41,75 |
| (1,70- 1,75] | 4 | 1,725 | 6,9 |
| (1,75-1,81] | 18 | 1,78 | 32,04 |
| (1,81-1,86] | 2 | 1,84 | 3,68 |
| (1,86-1,92] | 11 | 1,89 | 20,79 |
| (1,92-1,98] | 3 | 1,95 | 5,85 |
| **Total** | **100** |  | **166,9** |

Utilizando la fórmula de la media aritmética se obtiene:

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

El promedio de estatura de los trabajadores es *1,67 mt*

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **propiedades y características de la media aritmética** |
|  | * *si a todos los valores de la variable se le suma el mismo valor, el valor de la media aritmética queda aumentada en el mismo valor.* * *si a todos los valores de la variable se le multiplica el mismo valor, el valor de la media aritmética quedara multiplicado por dicho valor.* * *Solo se puede utilizar con variables cuantitativas.* * *La media aritmética es muy sensible a los datos extremos.* |

[SECCIÓN 2] **3.2 La mediana**

La **mediana (Me)** se encarga de mostrar el valor centrar de un conjunto de datos, el método como se puede calcular depende si los datos están agrupados o no.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Métodos para encontrar la mediana** |
|  | * *Cuando los datos no son agrupados un método para encontrar la mediana es:*  1. *Ordenar los datos de menor a mayor.* 2. *Si la serie tiene una cantidad impar de datos la mediana será el dato que se encuentra en el centro de los datos, pero si la serie de datos tiene una cantidad par de datos la mediana será el promedio entre los dos datos centrales.*  * *Cuando los datos son agrupados un método para encontrar la mediana es aplicar una formula, donde se deben definir:*  1. *El intervalo donde se va a encontrar la mediana, el cual es aquel donde la frecuencia acumulada llega a la mitad y se define como:*     *<<MA\_09\_12\_07.gif>>*   1. *El límite inferior donde se encuentra la mediana se define como:*   *Li*   1. *la amplitud de la clase se define como:*   *ai*   1. *la frecuencia absoluta acumulada anterior al intervalo donde se encuentra la mediana:*   *Fi-1*   1. *la frecuencia absoluta del intervalo donde se encuentra la mediana.*   *fi*  *la fórmula es la siguiente:*  *<<MA\_09\_12\_08.gif>>* |

Observa los siguientes ejemplos:

* encontrar la mediana de las edades de 30 estudiantes que hacen parte del grado noveno:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 16 | 16 | 15 | 15 | 15 | 14 | 14 | 14 | 13 | 12 |
| 16 | 16 | 15 | 15 | 14 | 14 | 14 | 14 | 13 | 12 |
| 16 | 16 | 15 | 15 | 14 | 14 | 14 | 13 | 13 | 12 |

Como lo datos no están agrupados y la cantidad de datos es par se deben organizar de mayor a menos o de menor a mayor, ubicar los dos datos medio, sumarlos y dividirlos entre dos:

16,16,16,16,16,16,15,15,15,15,15,15,15,14,14,14,14,14,14,14,14,14,14,13,13,13,13,12,12,12

Se suma 14+14 el resultado es 28 y se divide en 2, es decir que la mediana de las edades de los estudiantes de noveno es *Me = 14.*

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

|  |  |  |
| --- | --- | --- |
| **li , li-1** | **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. El intervalo donde se va a encontrar la mediana, es 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. El límite inferior donde se encuentra la mediana es:

*Li = 10*

1. la amplitud del intervalo es:

*ai = 5*

1. la frecuencia absoluta acumulada anterior al intervalo donde se encuentra la mediana es:

*Fi-1 =35*

1. la frecuencia absoluta del intervalo donde se encuentra la mediana.

*fi =25*

Remplazando en la formula:

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

Se obtiene que la mediana es 14,7

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **propiedades y características de la mediana** |
|  | * *para calculara no se necesitan todos los valores que ha podido tomar la variable.* * *El resultado de la mediana no se afecta por los resultados extremos.* |

[SECCIÓN 2] **3.3 La moda**

La **moda (Mo)** se encarga de mostrar el valor que más se repite en un conjunto de datos, el método de calcular depende si los datos están agrupados o no.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Métodos para encontrar la moda** |
|  | * Cuando los datos no son agrupados un método para encontrar la moda es:      1. Se debe observar la frecuencia absoluta de cada uno de los datos y escoger la mayor frecuencia absoluta, ese será el dato moda. 2. Si existen dos o más datos cuya frecuencia absoluta es la mayor y es la misma se dirá que la distribución es bimodal, es decir que tiene varias modas.        * Cuando los datos son agrupados un método para encontrar la moda es aplicar una formula, donde se deben:  1. Determinar el intervalo donde se encuentra la mayor cantidad de datos, es decir el intervalo que tenga la mayor frecuencia absoluta, que se definirá como clase modal. 2. Definir la amplitud del intervalo como:   ai   1. Definir el límite inferior de la clase modal como:   L1   1. Definir la frecuencia absoluta de la clase modal como:   fi   1. Definir la frecuencia absoluta inmediatamente menor a la clase modal como:   fi-1   1. Definir la frecuencia absoluta inmediatamente mayor la clase modal como:   fi+1  la formula es la siguiente:  *<<MA\_09\_12\_11.gif>>* |

Observa los siguientes ejemplos:

* Los siguientes datos corresponden a las 30 calificaciones obtenidas por un estudiante durante todo el año en la asignatura de matemáticas:

8,8,7,7,6,8,5,7,8,9,6,8,9,10,9,8,8,8,7,6,7,7,8,8,8,8,4,4,8,1

¿Cuál fue la nota que mas repitió?

|  |  |
| --- | --- |
| **notas** | **Frecuencia absoluta** |
| 1 | 1 |
| 2 | 0 |
| 3 | 0 |
| 4 | 2 |
| 5 | 1 |
| 6 | 3 |
| 7 | 6 |
| **8** | **13** |
| 9 | 3 |
| 10 | 1 |
| total | **30** |

Donde se puede observar que la nota que más se repite, es decir la moda es *Mo = 13.*

* 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 |
|  |  |

¿Cuál será la moda de esta distribución estadística?

1. Determinar la clase modal, que en este caso es el intervalo:

(6-8].

1. La amplitud de los intervalos es

ai = 2

1. El límite inferior de la clase modal es:

L1 = 6

1. La frecuencia absoluta de la clase modal es:

*fi = 150*

1. La frecuencia absoluta inmediatamente menor a la clase modal es:

*fi-1= 100*

1. Definir la frecuencia absoluta inmediatamente mayor la clase modal como:

*fi+1=75*

Ahora se remplaza en la formula:

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

Es decir que la moda es *Mo = 6,8*

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **propiedades y características de la moda** |
|  | * *para calculara no se necesitan todos los valores que ha podido tomar la variable.* * *El valor de la moda se ve afectado de acuerdo a la forma como se distribuyen los intervalos cuando la variable es continua.* * *No se ve afectada por valores extremos.* |

Ya se han analizado las medidas de posición central, ahora se deberán analizar las mediadas de posición no centrar, pero ¿Qué es una medida de posición no central?

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

* **Medidas de posición no centrar 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, para ello es necesario utilizar valores que dividir la muestra en partes iguales, los cuales serán desarrollados a continuación.

[SECCIÓN 2] **3.5 Los cuartiles, los deciles y Los percentiles**

Los **cuartiles** se definen como los tres valores que distribuyen al conjunto de datos ordenados de menor a mayor, en tramos iguales, es decir que cada tramo concentrara el 25% de los datos de la muestra.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Métodos para calcular los cuartiles** |
|  | * *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 : será la posición del cuartil que se quiere calcular*  *k: será el cuartil que se quiere calcular*  *n: total datos de la distribución.*     * *Cuando los datos son agrupados un método para encontrar los 3 valores de los cuartiles es el siguiente:*  1. *Encontrar en que intervalo se encuentra cada cuartil con la siguiente fórmula:*   *<<MA\_09\_12\_13.gif>>*  *Cuando k =1,2,3 ,será el cuartil que se quiere calcular y*  *n: total datos de la distribución.*   1. *Definir la frecuencia absoluta del intervalo donde se encuentra el cuartil como:*   *f1*   1. *Definir el límite inferior de la clase donde se encuentra el cuartil como:*   *L1*   1. *definir el numero de datos como:*   *n*   1. *Definir la frecuencia acumulada anterior al intervalo dodne se encuentra el cuartil como:*   *Fi-1*   1. *Definir la amplitud de los intervalos como:*   *ai*  *se deberá utilizar la siguiente fórmula:*  *<<MA\_09\_12\_15.gif>>* |

Observa los siguientes ejemplos:

* Estas son las edades de 12 personas:

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

Par calcular los cuartiles:

1. Se ordenan de menor a mayor:

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

1. Se encuentra el primero, el segundo y el tercer cuartil utilizando la formula:

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

Cuartil uno se encuentra en la posición 3 que equivales a 23

*<<MA\_09\_12\_16.gif>>*

Cuartil dos se encuentra en la posición 6 que equivales a 34

*<<MA\_09\_12\_17.gif>>*

Cuartil tres se encuentra en la posición 9 que equivales a 39

Se puede decir que:

Entre 19 y 23 corresponde al 25 % de los datos

Entre 19 y 34 corresponde al 50 % de los datos

Entre 19 y 39 corresponde al 75 % de los datos

Entre 19 y 56 corresponde al 100 % de los datos

* En la siguiente tabla se encuentra la distribución de las edades de 3530 personas que participaron en una encuesta.

|  |  |  |
| --- | --- | --- |
| **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 |
| (48-53] | 310 | 2810 |
| (53-58] | 290 | 3100 |
| (58-63] | 150 | 3250 |
| (63-68] | 120 | 3370 |
| (68-73] | 78 | 3448 |
| (73-78] | 50 | 3498 |
| (78-83] | 21 | 3519 |
| (83-88] | 11 | 3530 |
| total | 3530 |  |

Se deberán calcular los cuartiles:

1. Encontrar en que intervalo se encuentra cada cuartil:

Primer cuartil:

*<<MA\_09\_12\_19.gif>>*

882,5 está en el intervalo (23-28]

Segundo cuartil:

*<<MA\_09\_12\_20.gif>>*

1765 esa en el intervalo (33-38]

Tercer cuartil:

*<<MA\_09\_12\_21.gif>>*

2647,5 está en el intervalo (48-53]

1. Definir la frecuencia absoluta del intervalo donde se encuentra el cuartil:

Primer cuartil: f1 = 620

Segundo cuartil: f1 = 335

Tercer cuartil: f1 = 310

1. Definir el límite inferior de la clase donde se encuentra cada cuartil :

Primer cuartil: L1 = 23

Segundo cuartil: L1 = 33

Tercer cuartil: L1 = 48

1. definir el numero de datos:

n = 3530

1. Definir la frecuencia acumulada anterior al intervalo donde se encuentra el cuartil como:

Primer cuartil Fi-1 = 520

Segundo cuartil Fi-1 = 1550

Tercero cuartil Fi-1 = 2500

1. Definir la amplitud de los intervalos:

ai = 5

Teniendo todos los datos necesarios para remplazar en la formula se calcula cada curtil:

Cuartil uno:

*<<MA\_09\_12\_22.gif>>*

El resultado es 25,9

Curtil dos:

*<<MA\_09\_12\_23.gif>>*

El resultado es 36,2

Cuartil tres:

*<<MA\_09\_12\_24.gif>>*

El resultado es 50,3

Se puede concluir que:

El 25% de los datos corresponde a personas entre los 18 y 25, 9 años.

El 50 % de los datos corresponde a personas entre los 18 y 36,2 años.

El 75 % de los datos corresponde a personas entre los 18 y 50,3 años.

Como puedes ver los cuartiles permiten realizar un análisis de los datos recogidos en torno a su división en cuatro partes iguales 25%, 50%, 75% y 100% el cual puede aportar cierta información que puede ser utilizada para realizar algún análisis estadístico.

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

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Métodos para calcular los deciles** |
|  | * Cuando los datos no son 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 siguiente fórmula:   *<<MA\_09\_12\_25.gif>>*  donde:  Dk : será el valor del cuartil que se quiere calcular  k: será el cuartil que se quiere calcular  n: total datos de la distribución.     * Cuando los datos son agrupados un método para encontrar los 10 valores de los deciles es el siguiente:  1. Encontrar en que intervalo se encuentra cada decil con la siguiente fórmula:   *<<MA\_09\_12\_26.gif>>*  Cuando k =1,2,3 ,será el decil que se quiere calcular y  n: total datos de la distribución.   1. Definir la frecuencia absoluta del intervalo donde se encuentra el decil como:   f1   1. Definir el límite inferior de la clase donde se encuentra el decil como:   L1   1. definir el numero de datos como:   n   1. Definir la frecuencia acumulada anterior al intervalo donde se encuentra el decil como:   Fi-1   1. Definir la amplitud de los intervalos como:   ai  *se deberá utilizar la siguiente fórmula:*  *<<MA\_09\_12\_27.gif>>* |

Observa los siguientes ejemplos:

* Estas son las edades de 50 personas

18,34,45,67,89,34,23,18,45,67,56,45,34,56,76,23,24,34,23,25,45,34,23,18,34,

56,67,89,34,23,45,67,34,43,54,65,43,42,23,24,53,45,34,56,34,56,23,34,45,45

Encontrar los 9 deciles.

1. Ordenar de los datos de mayor a menor:

18,18,18,23,23,23,23,23,23,23,24,24,25,34,34,34,34,34,34,34,34,34,34,34,42,43,43,45,45,45,45,45,45,45,45,53,54,56,56,56,56,56,56,65,67,67,67,67,76,89

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

* En la siguiente tabla se encuentra la distribución de las edades de 3530 personas que participaron en una encuesta.

|  |  |  |
| --- | --- | --- |
| **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 |
| (48-53] | 310 | 2810 |
| (53-58] | 290 | 3100 |
| (58-63] | 150 | 3250 |
| (63-68] | 120 | 3370 |
| (68-73] | 78 | 3448 |
| (73-78] | 50 | 3498 |
| (78-83] | 21 | 3519 |
| (83-88] | 11 | 3530 |
| total | 3530 |  |

Se deberán calcular los deciles:

1. Encontrar en que intervalo se encuentra cada decil utilizando la formula:

*<<MA\_09\_12\_28.gif>>*

Donde *k* es el número de decil y *n* el numero de datos recogidos

* Primer decil: 353 está en el intervalo [18-23]
* Segundo decil : 706 está en el intervalo (23-28]
* Tercer decil: 1056 está en el intervalo (28-33]
* cuarto decil: 1412 está en el intervalo [28-33]
* quinto decil : 1765 está en el intervalo (33-38]
* sexto decil: 2118 está en el intervalo (38-43]
* séptimo decil: 2471 está en el intervalo [43-48]
* octavo decil : 2824 está en el intervalo (53-58]
* noveno decil: 3177 está en el intervalo (58-63]

1. Definir la frecuencia absoluta del intervalo donde se encuentra cada decil:

* Primer decil: f1 = 520
* Segundo decil : f1 = 620
* Tercer decil: f1 = 410
* cuarto decil: f1 = 410
* quinto decil : f1 = 325
* sexto decil: f1 = 300
* séptimo decil: f1 = 325
* octavo decil : f1 = 290
* noveno decil: f1 = 150

1. Definir el límite inferior de la clase donde se encuentra cada decil :

* Primer decil: L1 = 18
* Segundo decil : L1 = 23
* Tercer decil: L1 = 28
* cuarto decil: L1 = 28
* quinto decil : L1 = 33
* sexto decil: L1 = 38
* séptimo decil: L1 = 43
* octavo decil : L1 = 53
* noveno decil: L1 = 58

1. definir el numero de datos:

n = 3530

1. Definir la frecuencia acumulada anterior al intervalo donde se encuentra el decil:

* Primer decil: Fi-1 = 0
* Segundo decil : Fi-1 =520
* Tercer decil: Fi-1 =1140
* cuarto decil: Fi-1 =1140
* quinto decil : Fi-1 =1550
* sexto decil: Fi-1 =1875
* séptimo decil: Fi-1 =2175
* octavo decil : Fi-1 =2810
* noveno decil: Fi-1 =3100

1. Definir la amplitud de los intervalos:

ai = 5

Teniendo todos los datos necesarios para remplazar en la formula se calcula cada decil:

*<<MA\_09\_12\_29.gif>>*

* El valor del primer decil es: 21,3
* El valor del segundo decil es: 24,5
* El valor del tercero decil es:27,7
* El valor del cuarto decil es:31,3
* El valor del quinto decil es:33,6
* El valor del sexto decil es:42
* El valor del séptimo decil es:47,5
* El valor del octavo decil es:53,3
* El valor del noveno decil es:60,5

Con estos datos de los deciles es posible realizar análisis estadísticos.

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 concentrara el 1% de los datos de la muestra.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Métodos para calcular los percentiles** |
|  | * Cuando los datos no son agrupados un método para encontrar los 99 valores de los percentiles es el siguiente:      1. Ordenar los datos de menor a mayor. 2. Buscar el lugar de cada uno de los deciles utilizando la siguiente fórmula:   *<<MA\_09\_12\_30.gif>>*  donde:  Pk : será la posición donde se encuentra el percentil que se quiere calcular  k: será el percentil que se quiere calcular  n: total datos de la distribución.     * Cuando los datos son agrupados un método para encontrar los 99 valores de los percentiles es el siguiente:  1. Encontrar en que intervalo se encuentra cada percentil con la siguiente fórmula:   *<<MA\_09\_12\_31.gif>>*  Cuando k =1,2,3,….99 será el percentil que se quiere calcular y  n: total datos de la distribución.   1. Definir la frecuencia absoluta del intervalo donde se encuentra el percentil como:   f1   1. Definir el límite inferior de la clase donde se encuentra el percentil como:   L1   1. definir el numero de datos como:   n   1. Definir la frecuencia acumulada anterior al intervalo percentil se encuentra el percentil como:   Fi-1   1. Definir la amplitud de los intervalos como:   ai  *se deberá utilizar la siguiente fórmula:*  *<<MA\_09\_12\_32.gif>>* |

Observa el siguiente ejemplo:

* En la siguiente tabla se encuentra la distribución de las edades de 3530 personas que participaron en una encuesta.

|  |  |  |
| --- | --- | --- |
| **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 |
| (48-53] | 310 | 2810 |
| (53-58] | 290 | 3100 |
| (58-63] | 150 | 3250 |
| (63-68] | 120 | 3370 |
| (68-73] | 78 | 3448 |
| (73-78] | 50 | 3498 |
| (78-83] | 21 | 3519 |
| (83-88] | 11 | 3530 |
| total | 3530 |  |

Se deberán calcular los percentiles 10 y 75:

1. Encontrar en que intervalo se encuentra cada percentil utilizando la fórmula:

*<<MA\_09\_12\_33.gif>>*

Donde *k* es el número de percentil y *n* el número de datos recogidos

* Percentil diez: 353 está en el intervalo [18-23]
* Percentil setenta y cinco: 2647,5 está en el intervalo (48-53]

1. Definir la frecuencia absoluta del intervalo donde se encuentra cada percentil:

* Percentil diez: *f1 = 520*
* Percentil setenta y cinco: *f1 = 310*

1. Definir el límite inferior de la clase donde se encuentra cada percentil:

* Percentil diez: *L1 = 18*
* Percentil setenta y cinco: *L1 = 48*

1. definir el número de datos:

n = 3530

1. Definir la frecuencia acumulada anterior al intervalo donde se encuentra el percentil:

* Percentil diez: *Fi-1 = 0*
* Percentil setenta y cinco: *Fi-1 =2500*

1. Definir la amplitud de los intervalos:

*ai = 5*

Teniendo todos los datos necesarios para remplazar en la formula se calcula cada percentil: utilizando la fórmula:

*<<MA\_09\_12\_34.gif>>*

* El valor del percentil diez es: 21,3
* El valor del Percentil setenta y cinco es: 50,3

De esta misma forma se puede encontrar los 99 percentiles, los cuales podrán ser utilizados para realizar diferentes análisis estadísticos.

El trabajo en esta sección se centró en la medida de posición, tanto centrales como no centrales, en la siguiente sección el trabajo se centrará en las medidas de dispersión, ¿sabes que es una medida de dispersión en estadística?

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

Las **medidas de dispersión** son las encargadas de estudiar la distribución de los valores recogido, analizando si dichos valores se encuentran centrados, más o menos centrados, dispersos o más o menos dispersos, las más utilizadas son el **rango**, **la desviación media, la varianza, la desviación típica** y **el coeficiente de variación**.

[SECCIÓN 1] **4.1 El rango**

El **rango (R)** se define como la diferencia que se existe entre el dato mayo y el dato menor de la distribución.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Como calcular el rango** |
|  | 1. *Se debe ubicar el dato mayor y el dato menor de la muestra.* 2. *Utilizando la fórmula:*   *R = xmáximo –xmínimo*  Donde:  *xmáximo: es el valor más grande de la muestra.*  *xminimo: es el valor más pequeño de la muestra.* |

Observa el siguiente ejemplo:

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

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

¿Cuál será el rango del conjunto de datos?

1. El dato *xmáximo: 89 y el dato xminimo: 50*
2. *R = 89 – 50* = 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) |