**LECTURA PARA RESUMIR DATOS**

Generalmente cuando se tiene datos de variables aleatorias continuas se necesita resumir dichos datos para convertirla en información útil y poder interpretar.

En la tabla de distribución de frecuencias de una variable cuantitativa continua, cuando el número de datos es muy grande, es necesario establecer intervalos de clase. Esto implica definir el número de clases que se quiere considerar y sus correspondientes límites (inferior y superior).

**Para el establecimiento de los intervalos de clase se recomienda seguir el siguiente procedimiento:**

* 1. Definir el valor máximo y el valor mínimo
  2. Calcular el rango (R) de los datos: R = (máximo-mínimo) + 

Donde  toma un valor de acuerdo al número de decimales que tienen los datos.

Esto significa que si todos los datos recolectados son enteros, agregamos 1, si los valores están expresados en décimos, agregamos 0.1; si todos los datos están expresados en centésimos, SE AÑADE 0.01, etc.

* 1. Determinar el número de clases (K). El número de clases se establece de acuerdo a los objetivos y al manejo que se tenga sobre esa variable. Por ejemplo, en el campo de la salud la edad se agrupa con frecuencia en quinquenios (0 a 4, 5 a 9, 10 a 14, etc.).También puede seguirse un criterio estadístico, aplicando la regla de Sturges:

K= 1 + 3.322 (log10n)

Dónde:

K: es el número de clases

n: es el número total de datos

* 1. Determinar la amplitud de cada clase (A). Para lo cual: A = R/K, que se puede leer como el recorrido entre el número de clases.

Según Sturges la amplitud debe redondearse, de acuerdo al número de decimales que tienen los datos.

* 1. Determinar un exceso ( E ), este exceso debe cumplir la condición siguiente:

**E = KA-R0.**

Si el exceso es exactamente igual a CERO, el límite inferior de la primera clase es igual al valor mínimo determinado, y el límite superior de la última clase coincide con el valor máximo determinado.

Si el exceso es MAYOR QUE CERO, dividimos dicho exceso entre dos y luego restamos al mínimo y sumamos al máximo.

Si el exceso es MENOR QUE CERO, se aplica el siguiente criterio:

|  |  |
| --- | --- |
| **CRITERIO A** | **CRITERIO B** |
| **K: se mantiene constante** | **A : se mantiene constante** |
| **An = Aa +** | **Kn = Ka + 1** |
| **E 1N= K An -R** | **E 2N = Kn A -R** |

*Se toma el criterio de menor error*

* 1. Determinado el exceso de acuerdo a los criterios establecidos, se asume el criterio cuyo exceso es menor, se divide entre dos y se resta al mínimo y se suma al máximo. Dichos valores mínimo y máximo constituyen el límite inferior de la primera clase y el límite superior de la última clase.
  2. Teniendo en cuenta que cuando el exceso es igual a cero, no hay problema para determinar los límites de clase.
  3. Nos vamos a ocupar solamente para el caso que E>0 y cuando E< 0.

**CASO I: CUANDO EL EXCESO ES MAYOR QUE CERO**

* 1. Ejemplo: ¿En cuántas clases se puede agrupar a 350 obreros de acuerdo a su salario, si estos varían entre S/. 125.80 y S/. 342.75?
  2. Solución:
  3. De acuerdo a los datos del problema, se tiene el tamaño de la muestra n= 350, que es el número de obreros considerados para el estudio; como nos indican también los valores entre los que oscilan sus salarios, es fácil seguir los pasos para la construcción de la tabla que se nos pide:

Xi : es la variable en estudio, salario.

El Xi mínimo = 125.80 y el Xi máximo es 342.75.

1. El recorrido R = (342.75-125.80) + 0.01= 216.96, en este caso sumamos 0.01 por cuanto todos los datos tienen dos decimales.
2. Calculamos el número de intervalos *UTILIZANDO STURGES*, por lo tanto usamos la fórmula que sigue: K = 1 + 3.322 (log 350)= 9.45 9. *REDONDEAR*
3. Calculamos la amplitud mediante A = R/K = 216.96/9 = 24.10624.11, redondeado a dos decimales, por cuanto el número de decimales de los datos es dos.
4. Determinamos el exceso, E = KA-R =9\*24.11-216.96 = 0.03, el mismo que también debe ser redondeado a dos decimales, por la condición anterior.

Este exceso lo dividimos entre 2 y se tiene:

0.015 E1 0.02

E = 0.03

0.015 E2 0.01

1. Aproximamos a dos decimales (por condición que impone Sturges) o bien se aproxima el primer exceso o el segundo exceso; estos dos valores permiten encontrar el nuevo valor mínimo y el nuevo valor máximo.
2. Por tanto el nuevo valor mínimo es: Xi mínimo definitivo es: 125.80-0.02= 125.78, mientras que el valor máximo es: Xi máximo definitivo es: 342.75 + 0.01= 342.76.
3. Por consiguiente, ahora se tiene la seguridad de que el límite inferior de la primera clase es 125.78 y el límite superior de la última clase es de 342.76.
4. La tabla a construir es la siguiente:

**TABLA Nº 01: DISTRIBUCION DE LOS SALARIOS DE LOS OBREROS**

*# [suma 24.11]*

|  |  |  |
| --- | --- | --- |
| SALARIOS | Xi | CLASES REALES |
| 125.78 ; 149.88 | 137,83 | 125,775 149,885 |
| 149.89 ; 173.99 | 161,94 | 149,885 173,995 |
| 174.00 ; 198.10 | 186,05 | 173,995 |
| 198.11 ; 222.21 | 210.16 |  |
| 222.22 ; 246.32 | 234.27 |  |
| 246.33 ; 270.43 |  |  |
| 270.44 ; 294.54 |  |  |
| 294.55 ; 318.65 |  |  |
| 318.66 ; 342.76 |  |  |

Las marcas de clase que en la tabla se representa con Xi, se determinan dividiendo la suma de los límites en cada clase y dividiendo entre 2.

Además, las clases reales se determinan utilizando las fórmulas que siguen:

Xi – A/2: Límite real inferior de la clase.

Xi + A/2: Límite real superior de la clase.

Xi :

A=24.11

137.83-(A/2)

**CASO II: CUANDO EL EXCESO ES MENOR QUE CERO**

Ejemplo: Se tiene la las siguientes estaturas de 80 empleados de la Universidad Cayetano Heredia de Lima:

1.71, 1.58 , 1.69 , 1.70 , 1.55 , 1.70 , 1.65 , 1.62 ,1.78 ,1.60 , 1.62 ,1.62 , 1.55 ,1.64 , 1.66 , 1.63 , 1.68 , 1.74 ,1.57 , 1.73 , 1.63 , 1.80 , 1.76 , 1.70 , 1.70 , 1.70,1.72 , 1.65 ,1.72 , 1.72 , 1.70 , 1.79 , 1.60 , 1.68 , 1.72 , 1.57 , 1.57 , 1.60 , 1.66 , 1.74 , 1.61 , 1.51 , 1.63 , 1.81.75 , 1.62 , 1.66 , 1.58 , 1.68 , 1.60 , 1.67 , 1.68, 168, 1.70 , 1.71 , 1.65 , 1.68 , 1.65 , 1.57 , 1.65, 1.62 , 1.62 , 1.71 , 1.80 , 1.60 , 1.62 , 1.57 , 1.65 , 1.72 , 1.70 , 1.70 , 1.82 , 1.82 , 1.68 , 1.55 , 1.55 , 1.83 ,1.83 , 1.73 , 1.66.

**Objetivo:**

Resumir estos datos en una tabla de distribución de frecuencias.

Solución:

1. Se determina el mínimo y el máximo. En este caso:

*Xi  mínimo es 1.51*

*Xi  máximo es 1.88*.

1. Se determina el rango, en este caso *R = (1.88 – 1.51) + 0.01 = 0.38.*
2. Se calcula el número de clases: K = 1 + 3.322 log 80 = 7.31  7.
3. Se halla la AMPLITUD A = R/K = 0.38/7 = 0.054  0.05.
4. Se calcula el exceso: E = KA-R = 7(0.05) -0.38 = -0.03. (**NEGATIVO)**
5. Se deben aplicar los criterios considerados, es decir:

|  |  |
| --- | --- |
| CRITERIO A | CRITERIO B |
| K = 7 se mantiene constante | A= 0.05 se mantiene constante |
| An = Aa +0.01= 0.05 + 0.01 = 0.06 | Kn = Ka + 1 = 7 +1 = 8 |
| E 1N = K An –R = 7\*0.06-0.38 = 0.04 | **E 2N = Kn A-R = 8\*0.05-0.38= 0.02** |

**Se toma de los criterios, se toma el B :D**

Si evaluamos ambos criterios se observa que el menor exceso se encuentra en el criterio B cuyo E 2N = 0.02; por lo tanto este valor lo dividimos entre 2, para restar al mínimo y sumar al máximo, con lo cual determinamos el Xi mínimo y el Xi máximo definitivo.

0.01 restamos al mínimo inicial

E = 0.02

0.01 sumamos al máximo inicial

Esto significa que el mínimo y máximo definitivo están dados por:

Xi mínimo= 1.51-0.01 = 1.50 y el Xi máximo = 1.88 + 0.01 = 1.89

Con estos datos se construye la siguiente tabla de distribución de frecuencias:

**:0 QUE BUENA CLASE**

**TABLA Nº 02: DISTRIBUCION DE LOS EMPLEADOS SEGÚN ESTATURA, UNIVERSIDAD PERUANA CAYETANO HEREDIA**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| CLASES | fi | Xi | hi% | Fi | CLASES REALES |
| 1.50 ; 1.54 | 1 | 1.52 | 1.3 | 1 | 1.495 ; 1.545 |
| 1.55 ; 1.59 | 11 | 1.57 | 13.8 | 12 | 1.545 ; 1.595 |
| 1.60 ; .64 | 17 | 1.62 | 21.3 | 29 | 1.595 ; 1.645 |
| 1.65 ; 1.69 | 19 | 1.67 | 23.8 | 48 | 1.645 ; 1.695 |
| **1.70 ; .74** | **21** | **1.72** | **26.3** | **69** | **1.695 ; 1.745** |
| 1.75 ; 1.79 | 4 | 1.77 | 5.0 | 73 | 1.745 ; 1.795 |
| 1.80 ; 1.84 | 6 | 1.82 | 7.5 | 79 | 1.795 ; 1.845 |
| 1.85 ; 1.89 | 1 | 1.87 | 1.3 | 80 | 1.845 ; 1.895 |
| TOTAL | 80 |  | 100.0 |  |  |

Se puede ver que se obtiene exactamente el número de clases con el límite inferior de la primera clase (el Xi mínimo definitivo) y como límite superior de la k-ésima clase el Xi máximo definitivo.

A partir de esta tabla fácilmente se puede podemos interpretar, por ejemplo:

* Que existe un solo trabajador cuya estatura oscila entre 1.50 y 1.54.
* Del total de los 80 empleados de la Universidad Cayetano Heredia de Lima, 21 de ellos tienen estaturas que oscilan entre 1.70 y 1.74, lo cual representa el 26.3%.

Esta tabla nos muestra 6 columnas bien diferenciadas:

Clases: que son mutuamente excluyentes.

fi: Frecuencias absolutas simples, obtenidas mediante el conteo.

Xi: Marca de clase obtenida mediante el promedio de los dos extremos del intervalo.

hi%: frecuencias relativas porcentuales, obtenidas mediante hi% = (fi / n)\*100.

Fi: frecuencias absolutas acumuladas.

Clases reales obtenidas mediante las fórmulas:

Xi – A/2: Límite real inferior de la clase.

Xi + A/2: Límite real superior de la clase.

Con frecuencia se desea representar los datos a través de un solo valor.

Cuando la variable es cuantitativa se consideran dos aspectos al resumir los datos: ¿Cuál es el valor central, alrededor del cual se concentran los datos? (medidas de tendencia central o de posición) y ¿Cuán dispersos se encuentran los datos?

Cuando las variables son cualitativas se considera la comparación de frecuencias entre sí, con otro número o con el total (razones, proporciones y tasas).

**MEDIDAS DE RESUMEN DE UNA VARIABLE CUANTITATIVA**

Las medidas que más se utilizan son las medidas de tendencia central y las medidas de dispersión.

**MEDIDAS DE TENDENCIA CENTRAL**

Las medidas de tendencia central identifican un valor del conjunto de datos como representativo de todos. Estas medidas informan acerca del valor promedio de un conjunto de datos.

Algunas medidas de tendencia central son:

* La media aritmética o promedio aritmético.
* La media geométrica.
* La mediana.
* Los cuantiles.
* La moda.

**LA MEDIA ARITMÉTICA**

**MEDIA ARITMÉTICA POBLACIONAL**

Es la suma de todos los datos dividido por el número de datos (N).



Donde:

: es la media poblacional

Xi: Cada dato

N: Total de datos.

Ejemplo: Si se considera una población de 7 estudiantes: Las notas de estos estudiantes en el sistema vigesimal son las siguientes:

8, 8, 10, 12, 14, 16,16

= 

La media aritmética de las notas de los estudiantes es de 12.

La media aritmética o promedio aritmético de una muestra se representa con el símbolo 

Y se presentan dos casos fundamentalmente:

**CASO I: DATOS SIN AGRUPAR**

Se presenta cuando por la naturaleza de los datos no es necesario agrupar en clases

Ejemplo: Si de una población de 50 estudiantes de Ingeniería de Sistemas, solamente se toma en cuenta un tamaño de muestra de 10, para evaluar su peso, se utiliza la siguiente formula:

 = 

Pesos: Xi: 52; 60; 70; 65; 59; 61; 55; 68; 72; 53

 = = 

El peso promedio de los estudiantes de Ingeniería de Sistemas de la UNHEVAL es de 61.5 kilogramos, aproximadamente.

**CASO II: DATOS AGRUPADOS**

Cuando los datos se requieren agrupar en intervalos para un mejor manejo estadístico, se utiliza la siguiente formula:

 = 

Donde:

de la UNHEVAL. Huánuco 2008, distribuidos de acuerdo a la tabla que Xi: marca de clase de cada uno de los intervalos

fi: Frecuencias absolutas simples

n=: Total de datos considerados en la muestra

Ejemplo: Considérese la muestra de 316 estudiantes de la Facultad de Economía sigue.

**TABLA Nº 03: DISTRIBUCIÓN DE LA MUESTRA SEGÚN GRUPOS DE EDAD HUÁNUCO, 2004**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| EDAD | fi | hi% | Fi | Hi% | Xi | Xifi |
| 18 ; 25 | 41 | 13.0 | 41 | 13.0 | 21.5 | 881.5 |
| 26 ; 33 | 57 | 18.0 | 98 | 31.0 | 29.5 | 1681.5 |
| 34 ; 41 | 44 | 13.9 | 142 | 44.9 | 37.5 | 1650.0 |
| 42 ; 49 | 50 | 15.8 | 192 | 60.8 | 45.5 | 2275.0 |
| 50 ; 57 | 31 | 9.8 | 223 | 70.6 | 53.5 | 1658.5 |
| 58 ; 65 | 30 | 9.5 | 253 | 80.1 | 61.5 | 1845.0 |
| 66 ; 73 | 36 | 11.4 | 289 | 91.5 | 69.5 | 2502.0 |
| 74 ; 81 | 13 | 4.1 | 302 | 95.6 | 77.5 | 1007.5 |
| 82 ; 89 | 14 | 4.4 | 316 | 100.0 | 85.8 | 1197.0 |
| TOTAL | 316 | 100.0 |  |  |  | 14698.0 |

A partir de la tabla, se puede calcular la media aritmética para la edad de los estudiantes de la Facultad de Economía.

 = =

La edad promedio de los estudiantes de Economía es de aproximadamente 46.5 años.

Se denomina desviación de un dato, a la diferencia entre ese dato y la media aritmética. Es negativo si el dato es menor que la media aritmética, y positivo, si es mayor que la media aritmética.

**PROPIEDADES DE LA AMEDIA ARITMÉTICA**

1. Hay una sola media en un conjunto de datos.
2. La media es influenciada por los valores extremos.
3. La suma de las desviaciones de los datos con respecto a la media es cero.

**LA MEDIANA (Me**)

Es el valor que divide al conjunto de datos en dos partes iguales. Debajo de la mediana se encuentra el 50% de datos; y por encima de la mediana se encuentra el otro 50%.

**Para calcular la mediana**:

1. Primero se ordenan los datos en orden de magnitud (ascendente o descendente).
2. Luego se identifica la mediana:
   1. Si el número de datos es impar, la mediana es el valor medio.

Se calcula la posición de la mediana:

POSICION: (N+1)/2.

Ejemplo: Respecto de las notas de los 7 estudiantes.

Posición = (7+1)/2= 4

La posición 4 corresponde a: 8, 8, 10, **12**, 14, 16, 16.

La mediana de las notas o la nota mediana es 12.

* 1. Si el número de datos es par, la mediana es la semisuma de los valores centrales.

Ejemplo: Si las notas solo fueran:

8, 10, 12, 14, 16, 16

Posición= (6+1)/2 =3.5, este resultado se descompone en 3+0.5

La posición 3 corresponde a: 8, 10, **12,** 14, 16, 16.

Pero también, la posición 0.5 adicional corresponde a☹ 14-12)\*0.5 =1

Por lo tanto la mediana es igual a 12+1 =13.

Cuando los datos están agrupados la mediana se calcula de la siguiente manera:

Me= LRI +A ()

Dónde:

LRI: Límite real inferior de la clase mediana.

Fi-1: Frecuencia acumulada anterior a la clase mediana.

fi: Frecuencia absoluta simple correspondiente a la clase mediana.

n: Tamaño de la muestra.

n/2: Posición de la mediana.

A: Amplitud.

Ejemplo: Si se considera los datos de la tabla Nº 05 se puede calcular la mediana.

**TABLA Nº 04: DISTRIBUCIÓN DE LA MUESTRA SEGÚN GRUPOS DE EDAD: HUÁNUCO, 2004**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| EDAD | fi | hi% | Fi | Hi% | Xi | Xifi |
| 18 ; 25 | 41 | 13.0 | 41 | 13.0 | 21.5 | 881.5 |
| **26 ; 33** | **57** | **18.0** | **98** | **31.0** | **29.5** | **1681.5** |
| 34 ; 41 | 44 | 13.9 | 142 | 44.9 | 37.5 | 1650.0 |
| **42 ; 49** | **50** | **15.8** | **192** | **60.8** | **45.5** | **2275.0** |
| 50 ; 57 | 31 | 9.8 | 223 | 70.6 | 53.5 | 1658.5 |
| 58 ; 65 | 30 | 9.5 | 253 | 80.1 | 61.5 | 1845.0 |
| **66 ; 73** | **36** | **11.4** | **289** | **91.5** | **69.5** | **2502.0** |
| 74 ; 81 | 13 | 4.1 | 302 | 95.6 | 77.5 | 1007.5 |
| 82 ; 89 | 14 | 4.4 | 316 | 100.0 | 85.8 | 1197.0 |
| TOTAL | 316 | 100.0 |  |  |  | 14698.0 |

Procedimiento:

1. Se determina la clase mediana, para lo cual se divide el total de datos de la muestra entre 2, es decir 316/2 =158.
2. Determinamos la clase mediana, para lo cual en la cuarta columna de la tabla que antecede se busca la mínima frecuencia absoluta acumulada que supera a n/2, y vemos que el mínimo valor que cumple dicha condición es 192, por lo tanto en ese intervalo encontramos todos los datos que necesitamos para usar la formula.
3. El intervalo se encuentra remarcado con negrita, es decir :

Intervalo= 42-49. El límite real inferior de la clase mediana es 41.5, y se obtuvo de la siguiente manera: X4- A/2 = 45.5-8/2= 45.5-4= 41.5.

Fi-1 = 142, valor anterior al de la clase mediana.

fi: 50, frecuencia absoluta que corresponde a la clase mediana.

A=8

Si reemplazamos en la formula, se tiene que:

Me = 41.5 + 8( = 41.5+2.56 = 44.1

Por tanto, la mediana de la edad de la muestra en estudio es de 44.1 años.

**PROPIEDADES DE LA MEDIANA**

1. Hay una sola mediana para un conjunto de datos.
2. La mediana no está influenciada por los valores extremos.
3. Se puede calcular la mediana aun cuando las clase extremas tienen límites abiertos.

**LOS CUANTILES**

Son aquellos que dividen a un conjunto de datos en cuatro partes (cuarteles), diez partes (deciles) y cien partes iguales (percentiles) respectivamente.

Los cuarteles son aquellos valores (Q1, Q2, Q3) que dividen a un conjunto de datos en cuatro partes iguales. Cada cuartel contiene 25 por ciento de datos. Para calcular los cuarteles se utilizan fómulas semejantes a las que se utilizan para hallar la mediana, es decir:

Qk= LRI + A ()

Dónde:

LRI: Límite real inferior de la clase cuartílica.

A: Amplitud.

n/4: Permite determinar la clase cuartílica.

Fi-1: Frecuencia absoluta acumulada anterior a la clase cuartílica.

fi: Frecuencia absoluta simple correspondiente a la clase cuartílica.

De manera similar se puede generar las fórmulas tanto para los deciles cuanto para los percentiles, y las fórmulas para cada caso estarían dadas por:

DK = LRI + A ()

PK = LRI + A ()

A como aplicación solo se va a calcular:

Q3= LRI + A () =?

1. Determinamos el valor de 3(n/4) = 3(316/4)=273.
2. Buscamos el mínimo valor que supera a este valor y se tiene que el mínimo valor es: 289, quiere decir que la clase cuartilica está en el intervalo: 66-73, por tanto el límite real inferior correspondiente es: 65.5.
3. Fi-1= 253; fi=36.

Si reemplazamos valores en la fórmula se tiene:

Q3= 65.5 + 8(

El cuartil 3 de la edad es 69.94 años. Es decir, el 25% de las personas tiene edad menor a 69.94 años.

Como ejercicio, se sugiere que el lector determine: Q2, D5, P50, utilizando la misma tabla de las edades.

**LOS DECILES**

Son aquellos valores (D1, D2, ………D9) que dividen a un conjunto de datos en 10 partes iguales. Cada decil contiene 10 por ciento de datos. Para calcular los deciles se utilizan fórmulas semejantes que las que se utilizan para el cálculo de la mediana.

**LOS PRECENTILES**

Son aquellos valores (P1, P2,………P99) que dividen a un conjunto de datos en cien partes iguales. Cada percentil contiene 1 por ciento de los datos (1%). Para calcular los percentilles se utilizan también fórmulas semejantes a las del cálculo de la mediana.

Ambas fórmulas generalizadas se han representado anteriormente.

NOTA: Para el caso de la mediana y de los cuantiles primero se deben encontrar las frecuencias absolutas acumuladas.

**LA MODA: (Mo)**

Es el valor más frecuente en un conjunto de datos.

Ejemplo: En los datos correspondientes a las notas obtenidas por lo alumnos, la moda es 8 y 16.

Para el caso de datos agrupados, la moda se establece como la marca de clase, de aquella con mayor frecuencia absoluta simple.

Ejemplo: Si se considera la misma tabla que se ha usado para hallar la media, la mediana y la moda referida a la edad, se desprende que la clase con mayor frecuencia absoluta simple corresponde al intervalo 26 a 33, siendo su marca de clase 29.5. Esto quiere decir entonces que la moda de la edad es de 29.5 años.

Si se quiere utilizar formula esta sería la siguiente: Mo= LRI +A (,

Dónde:

Es la diferencia entre la clase modal y la premodal.

: Es la diferencia entre la clase modal y la postmodal.

Mo= 25.5 +8()=25.5 + 4= 29.5.

**PROPIEDADES DE LA MODA**

1. Puede haber más de una moda en un conjunto de datos.
2. No está influenciada por los valores extremos.
3. Se aplica a datos cualitativos y a datos cuantitativos.

**1.2.2 MEDIDAS DE DISPERSION**

Las medidas de dispersión sintetizan la variabilidad de los datos. Si existe poca variabilidad entre los datos, la dispersión es pequeña.

**Las principales medidas de dispersión son**:

1. La amplitud o rango.
2. El rango intercuartílico.
3. La varianza.
4. La desviación estándar.
5. El coeficiente de variación.

**EL RANGO**

Es la diferencia entre el valor máximo y el valor mínimo.

El rango es una medida apropiada para datos numericos cuando el propósito es enfatizar valores extremos.

**EL RANGO INTERCUARTÍLICO**

Es la diferencia entre el tercer y primer cuartel. Comprende el 50% de los datos.

RIQ = Q3-Q1

Se utiliza cuando hay mucha asimetría.

**LA VARIANZA POBLACIONAL: **

Es la suma de las desviaciones al cuadrado, entre el número de datos y la media poblacional.

= 

Dónde:

: Varianza poblacional.

(Xi-: Desviación de cada dato.

N: Total de datos de la población.

Cuando se calcula la varianza de una muestra, la suma de los cuadrados se divide entre el número de datos menos 1 (grados de libertad).

La fórmula que se utiliza en este caso está dada por:

, fórmula usada para datos sin agrupar.

Donde:

S2: varianza de la muestra.

(Xi-2 = Desviación de cada dato de la media muestral.

n-1: Grados de libertad.

Para el caso de datos agrupados la varianza de la muestra se calcula utilizando la siguiente fórmula:



Donde:

S2: Varianza de la muestra.

: Desviación de cada marca de clase respecto de la media muestral al cuadrado.

fi: Frecuencia absoluta simple de cada clase.

n-1: Grados de libertad.

A partir de la tabla que sigue se quiere calcular la varianza de la edad, cuya media aritmética es 46.5.

**TABLA Nº 05: DISTRIBUCIÓN DE LA MUESTRA SEGÚN GRUPOS DE EDAD: HUÁNUCO, 2004**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| EDAD | fi | hi% | Xi | (Xi-) |  | \*fi |
| 18 ; 25 | 41 | 13.0 | 21.5 | -25 | 625 | 25625 |
| 26 ; 33 | 57 | 18.0 | 29.5 | -17 | 289 | 16473 |
| 34 ; 41 | 44 | 13.9 | 37.5 | -9 | 81 | 3564 |
| 42 ; 49 | 50 | 15.8 | 45.5 | -1 | 1 | 50 |
| 50 ; 57 | 31 | 9.8 | 53.5 | 7 | 49 | 1519 |
| 58 ; 65 | 30 | 9.5 | 61.5 | 15 | 225 | 6750 |
| 66 ; 73 | 36 | 11.4 | 69.5 | 23 | 529 | 19044 |
| 74 ; 81 | 13 | 4.1 | 77.5 | 31 | 961 | 12493 |
| 82 ; 89 | 14 | 4.4 | 85.8 | 39 | 1521 | 12294 |
| TOTAL | 316 | 100.0 |  |  |  | 106812 |

Si reemplazamos el valor obtenido en la última columna de la tabla, se obtiene que:

S2=, que constituye la varianza muestral de los datos considerados.

**PROPIEDADES DE LA VARIANZA**

1. El valor de la varianza es mayor o igual que cero cualquiera sea su distribución.

**LA DESVIACIÓN ESTÁNDAR**

Es la raíz cuadrada de la varianza, sea poblacional o muestral.

**EL COEFICIENTE DE VARIACIÓN**

Es la expresión relativa de la desviación estándar respecto a la media aritmética.

CV =, cuando se conoce la media y desviación estándar poblacional

CV =, cuando se tiene la media y varianza muestral.

El coeficiente de variación es útil cuando la intención es comparar dos distribuciones numéricas medidas en escalas diferentes.