**UNIDAD Nº1**

# INTRODUCCIÓN A LA ESTADÍSTICA

**1.-** ESTADÍSTICA

* 1. Concepto. - Se entiende por estadística a los métodos científicos mediante los cuales podemos recolectar, organizar, analizar, resumir y presentar informes de datos numéricos relativos a un conjunto de individuos u observaciones las cuales nos permiten sacar conclusiones y tomar decisiones lógicas basadas en dichos análisis. Dichos métodos científicos consisten en la utilización de tablas y gráficos mediante los cuales se procesa la información.
	2. Población y Muestra. - Población o Universo se define como un conjunto de elementos que poseen algunas características determinadas y que constituyen un objeto de estudio, es decir que el total de observaciones susceptibles de ser registradas o empadronadas, se les denomina Población.

Muestra. - Es la recolección de observaciones provenientes de una población la cual posee una característica definida. Toda característica de una muestra se la conoce con el Nombre de **"**Estadígrafo**"** Ejemplo. - La media Aritmética de una muestra sea un estadígrafo.

* 1. Clasificación de la Estadística. - Existen dos clases o tipos de estadísticas las cuales son: Estadística Descriptiva y Estadística Inductiva.
1. Estadística Descriptiva. - Es la descripción de los datos muéstrales que son objeto de estudio mediante la utilización de ciertos tipos de indicadores. Los datos muéstrales se pueden recopilar de dos formas:
	1. Cuando se consideran a todos los elementos de la población se denomina "CENSO"
	2. Cuando se realiza una selección al azar de los elementos de la población se denomina "MUESTREO"
2. Estadística Inductiva. - Conocida también con el nombre de estadística "INFERENCIAL", está basada en una serie de teorías e hipótesis acerca del comportamiento de un fenómeno observado. En este tipo de estadística, muchas decisiones son tomadas en condiciones de incertidumbre por lo cual es preciso utilizar el término de probabilidades.

**1.4)** Aplicaciones. - La utilización de datos estadísticos es de vital importancia en todos los campos de la ciencia donde esta sea aplicada como ser:

Estadística en las Empresas. - Ayuda a tomar decisiones económicas y financieras para el futuro.

Estadística en la Medicina. - Ayuda a conocer la efectividad de los medicamentos y los efectos secundarios que producen.

Estadística en la Educación. - Ayuda a decidir sobre el empleo de los mejores métodos de enseñanza.

Estadística en la Agricultura. - Ayuda a conocer el efecto de diferentes tipos de semillas e insecticidas que van a ser utilizados.

Estadística en la Veterinaria. - Ayuda a producir ganado de mejor calidad y peso.

**2.-** Variables. - Son símbolos o signos que pueden tomar algún valor o infinitos valores en una muestra, dentro de un intervalo determinado. En estadística existen dos tipos de variables que son:

1. Variables Ordinales. - Son las cuales están sujetas a una ordenación cualitativa y no a una mediación cuantitativa; Ejemplo el grado de cultura de una persona (Muy culta, regularmente culta, poco culta etc.)
2. Variables Cardinales. - Son las que están sujetas a una medición cuantitativa y no a una ordenación cualitativa y pueden ser de dos tipos: Discretas y Continuas.
	1. Variables Discretas. - Son aquellas que pueden tomar algún valor específico dentro de un intervalo determinado, es decir que solamente pueden tomar valores de números enteros. También podemos decir que son el resultado de "CONTAR". Ejemplo: Números de estudiantes de Estadística I; 70,80,90 etc.
	2. Variables Continuas. - Son aquellas que pueden tomar cualquier valor dentro de un intervalo determinado, es decir, que toman valores de números fraccionados. También se puede decir que son el resultado de "MEDIR". Ejemplo: La estatura y el peso de los estudiantes de Estadística I ; 70,5 ; 83,6 etc.

**3.-**Investigación Estadística.- En toda investigación científica es importante definir el marco de referencia del estudio, y el grado de requerimiento de la información, para lo cual se establecen ciertos parámetros o etapas de la investigación.

**3.1** Etapas de la Investigación Estadística.- Existen cuatro etapas :

1. Recolección de datos mediante técnicas definidas como ser: Encuestas, Censos, Muestreos, etc, para la obtención de datos fidedignos.
2. Elaboración, ordenación, y reducción de los datos para ayudar a su estudio.
3. Análisis de los datos para la obtención de medidas estadísticas como ser: Medidas de Tendencia Central y Medidas de Dispersión.
4. Interpretación de los resultados obtenidos para sacar conclusiones y emitir recomendaciones y diagnósticos.

**4.-** Tabulación de Datos.- Los datos obtenidos en base a las diferentes técnicas como ser: Censo, Muestreo, etc. Se elaboran y se clasifican para obtener series estadísticas, tablas y distribuciones de frecuencias.

* 1. Representación Tabular.- Una vez los datos han sido recolectados y seleccionados es necesario presentarlos de tal forma que sean de fácil comprensión, para su posterior análisis e interpretación. Para ello se elaboran cuadros numéricos, tablas de frecuencia y se los representa en distintos tipos de gráficos como ser: Barras, Círculos, Histogramas, Polígonos de frecuencia, Pictogramas, Ojivas, etc.
	2. Distribución de Frecuencias.- Es una tabla donde se agrupan todos los valores posibles para una variable y se registran los números de valores observados correspondiente a cada clase. Los datos organizados en una tabla de distribución de frecuencia se llaman "Datos Agrupados", mientras que los datos sin organizar se denominan “Datos Sencillos”.
		1. Elementos del Cuadro de Distribución de Frecuencia
1. Valor máximo y valor mínimo de un conjunto de datos .- Se trata de determinar el valor superior e inferior de un conjunto de observaciones para lo cual es conveniente ordenar los datos en forma ascendente o descendente para una fácil identificación de los valores máximos y mínimos.

**Ejemplo.-** Dadas las notas obtenidas por 15 estudiantes de Estadística I .

|  |  |  |  |
| --- | --- | --- | --- |
| 40 | 50 | 70 | Ordenación Ascendente: 10-15-20-25-30-35-40-45 |
| 30 | 65 | 15 | 50-65-70-75-85-90-95 |
| 85 | 10 | 90 | V. MAX = 95 |
| 75 | 45 | 25 | V. MIN = 10 |
| 95 | 20 | 35 |  |

1. Rango o recorrido de la variable (R).- Es la diferencia entre el valor Máximo y el valor Mínimo en una serie de datos.

R= V.MAX. - V.MIN. => 95-10 = 85 **R = 85**

1. Amplitud de intervalos (C).- Es el tamaño o medida que existe entre el limite inferior y el limite superior de cada número de intervalo. La amplitud de cada intervalo puede ser constante o variable. En una serie donde se precisa determinar intervalos, uniformes existen dos formas de cálculo:
2. De acuerdo a un número fijo de intervalos: R R = Rango o recorrido

C = ─── => ---

 K K = Número de intervalos fijos

1. De acuerdo al número de intervalos ideales(Regla de Sturges) R

C = ───────────────

1 + 3,3222 Log N donde N = Número de observaciones

R 85

C = ─────────────── => ────────────────── = 17.32  **C = 17**

1+(3.3222 LOG 15) 1+(3.3222 X 1.1761)

**4.-**Número de Intervalos (K).- Es la cantidad de clases o categorías que resulta del procesamiento de una distribución de frecuencia. Un cuadro de distribución de frecuencia convencional, no debe excederse de 20 intervalos de clases.

Cuando el número de intervalos es desconocido se busca la cantidad ideal de intervalos mediante la fórmula siguiente:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| K = | R--- | => | K = | 85---- = 5 | **K = 5** |
|  | C |  |  | 17 |  |

**Nota.-** El número de intervalos “siempre” deberá ser expresado en valores enteros.

**5º)** Rango o Recorrido Ideal (RI).- Muchas veces el recorrido existente, no es el más adecuado para una distribución por lo cual se tiene que encontrar un Recorrido Ideal para dicha distribución que actúe en función de la Amplitud y del Número de intervalos de la distribución. El proceso de cálculo se lo efectúa mediante la siguiente formula:

RI = C x NºI => 17 x 5 = 85 **RI = 85**

Nota. - El valor del **RI debe ser > = R**; caso contrario se deberá redondear el valor de “**C”** al número entero siguiente.

**6º)** Margen de Desplazamiento (MD). -En la mayoría de los casos la información no se encuentra centralizada, por lo que se corre el riesgo de cargar mayor información a uno de los extremos; para lo cual la información es centralizada mediante el margen de desplazamiento en base a la formula siguiente:

RI-R 85-85

MD = ----- => ----- = 0 **MD = 0**

2 2

**7º) Nuevo Valor Mínimo o** Limite Inicial de los Intervalos de Clases (Li).- Se obtiene disminuyendo al valor mínimo, el valor del margen de desplazamiento obtenido.

NVmin = V.MIN - MD => 10 - 0 = 10 **LI = 10**

**8º)Nuevo Valor Máximo o** Limite Final de Intervalos de Clases (LF).- Se obtiene aumentando al valor máximo, el valor del margen de desplazamiento obtenido.

NVmax= V.MAX + MD => 95 + 0 = 95 **LF= 95**

**9º)** Limites Reales de cada Intervalo de clase.-Se obtiene sumando al limite inicial, la amplitud de intervalos obtenido. Existen dos tipos de limites de cada intervalo que son: Limite Inferior (Y'i-1) y el Limite Superior (Y'i). El límite superior de un intervalo de clases es el límite inferior de la siguiente clase o intervalo.

 Li Ls

|  |
| --- |
| K │(Y'i-1)(Y'i) │──────┼──────────────┤ |
| 1 │ | 10 | 27 │ | => | Y'1-1 | + | C | = | Y'1 | => | 10 | + | 17 | = | 27 |
| 2 │ | 27 | 44 │ | => | Y'2-1 | + | C | = | Y'2 | => | 27 | + | 17 | = | 44 |
| 3 │ | 44 | 61 │ | => | Y'3-1 | + | C | = | Y'3 | => | 44 | + | 17 | = | 61 |
| 4 │ 5 │  | 6178  | 78 │95 │ | =>=> | Y'4-1Y'5-1 | ++ | CC | == | Y'4Y'5 | =>=> | 6178 | ++ | 1717 | == | 7895 |

**10º)** Marca de Clases o Punto Medio (Yi). - Es el valor central de cada intervalo de clases y se obtiene dividiendo entre dos la suma del valor del límite inferior más el valor del límite superior de cada intervalo de clases.

|  |  |  |  |
| --- | --- | --- | --- |
| Y'i-1 + Y'i |  | NºI │Y'i-1 Y'i │ Y'i │ |  |
| Yi= ───────────── | => | ────┼───────────┼─────┤ |
| 2 |  | 1 │ 10 27 │ 18.5│=> | Y1 | = | 10+27 |
|  |  | │ │ │ |  |  | 2 |
|  | 2 │ 27 44 │ 35.5│=> Y2 = 27+44│ │ │ 2 3 │ 44 61 │ 52.5│=> Y3 = 44+61│ │ │ 2 4 │ 61 48 │ 69.5│=> Y4 = 61+78│ │ │ 25 │ 78 95 │ 86.5│=> Y5 = 78+95 |

2

A todo este procedimiento realizado se lo conoce con el nombre de Representación Tabular. Los procedimientos adicionales a utilizarse posterior a esta Representación se los conoce como Distribución de Frecuencia.

# Ejercicio Nº2

En base a las notas obtenidas del ejercicio anterior se pide elaborar una nueva tabla de distribución de frecuencia para una serie de 8 intervalos de clases de Amplitud Constante.

**2)** R = V. MAX - V.MIN => 95-10 =85 R 85

**3)** C=----- => ------ = 10.62  11

NºI 8

R 85

**4)** K =---- => ---- =7.33  8 **(NO)** Por que sabemos que: **NºI=8**

C 11

**5)** RI =C X NºI => 11 X 8 = 88

RI - R 88-85

**6)** MD =------- => ------ = 1.5

2 2

**7)** LI = V.MIN - MD => 10-1.5 = 8.5

**8)** LF = V.MAX + MD => 95+1.5 = 96.5

REPRESENTACION TABULAR

┌────┬──────────────┬────────────┐

│NºI │(Y'i-1) (Y'i)│ Yi │ fi...etc.

│────┼──────────────┼────────────│

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ││ | 1 | ││ | 8.5 | 19.5 | ││ | 14 | C | │=11│ |
| │ | 2 | │ | 19.5 | 30.5 | │ | 25 |  | │ |
| │ |  | │ |  |  | │ |  | C | =11│ |
| │ | 3 | │ | 30.5 | 41.5 | │ | 36 |  | │ |
| │ |  | │ |  |  | │ |  | C | =11│ |
| │ | 4 | │ | 41.5 | 52.5 | │ | 47 |  | │ |
| │ |  | │ |  |  | │ |  | C | =11│ |
| │ | 5 | │ | 52.5 | 63.5 | │ | 58 |  | │ |
| │ |  | │ |  |  | │ |  | C | =11│ |
| │ | 6 | │ | 63.5 | 74.5 | │ | 69 |  | │ |
| │ |  | │ |  |  | │ |  | C | =11│ |
| │ | 7 | │ | 74.5 | 85.5 | │ | 80 |  | │ |
| │ |  | │ |  |  | │ |  | C | =11│ |

│ 8 │ 85.5 96.5 │ 91 │

Frecuencia Absoluta (fi).- Es el número de observaciones que se repiten dentro de cada intervalo de clase. Se obtiene mediante el recuento y clasificación de los datos en la que al final la suma total de la frecuencia encontrada, será igual al número total de observaciones estudiadas. ∑**fi= N =15**

# (…Continuamos con la tabla original:)

|  |
| --- |
| ─────────────────────────────────────────────────────────────NºI│(Y'i1)(Y'i)│ Yi│fi│ hi │h% │FiA│ HiA│H%A│FID│HID │HI%D |
| 1 | │ | 10 | 27│18.5│ 4│0.27│27 | │ 4 | │0.27│ 27│15 | │1.00│100 |
| 2 | │ | 27 | 44│35.5│ 3│0.20│20 | │ 7 | │0.47│ 47│11 | │0.73│ 73 |
| 3 | │ | 44 | 61│52.5│ 2│0.13│13 | │ 9 | │0.60│ 60│ 8 | │0.53│ 53 |
| 4 | │ | 61 | 78│69.5│ 3│0.20│20 | │12 | │0.80│ 80│ 6 | │0.40│ 406 |
| 5 | │ | 78 | 95│86.5│ 3│0.20│20  | │15  | │1.00│100│ 3 | │0.20│ 20 |

│**15**│**1.00**│**100**│

∑fi=> N => 15

Frecuencia Relativa (hi).- Es el cociente que resulta de dividir cada una de las frecuencias absolutas entre el número total de observaciones. Todas las frecuencias relativas deben ser menores que la unidad y la suma total de ellos, da como resultado el valor de la unidad.

h1 = 4 = 0.27

hi= fi => fi 15

∑fi N

h2 = 3 = 0.20 etc......

15

Frecuencia Relativa Porcentual (hi %).-Es una nueva forma de expresar la frecuencia relativa y a la vez toma las mismas características de esta. Su forma de cálculo es tan sencilla que simplemente resulta de multiplicar por 100 la frecuencia relativa obtenida. La suma del total de frecuencias relativas porcentuales, será igual al 100%

hi% = hi x 100

h1 % = 0.27 x 100 = 27

h2 % = 0.20 X 100 = 20 etc.....

Frecuencias Acumuladas.-Todas las frecuencias: Absolutas, Relativas, y Porcentuales se pueden expresar en valores acumulados, es decir que la suma de las distintas frecuencias hasta un valor determinado, se denomina frecuencia acumuladas y pueden ser de dos tipos; Frecuencia Acumuladas Ascendente y Frecuencia Acumuladas Descendente.

1. Frecuencia Acumuladas Ascendente.-Es la sumatoria acumulativa de los valores de las frecuencias a partir de los intervalos inferiores y dirigida hacia los intervalos superiores es decir que se debe comenzar a partir del 1º intervalo; hacia el ultimo intervalo de clase.
* En el caso de las frecuencias absolutas, el último será igual al número total de observaciones.
* En el caso de las frecuencias relativas, el último valor será igual a la unidad.
* En el caso de las frecuencias porcentuales, el último valor será igual al 100%

# Ejemplo

F1A = f1 = 4

F2A = f1 + f2 = 4 + 3 = 7 etc...... H1A = h1 = 0.27

H2A = h1 + h2 = 0.27 + 0.20 = 0.47 etc......

H1%A = h1% = 27

H2%A = h1% + h2% = 27 + 20 = 47 etc....

* Otra forma:

H1%A = H1A X 100 = 0.27 X 100 = 27

H2%A = H2A X 100 = 0.47 X 100 = 47

1. Frecuencias Acumuladas Descendente.- Es la diferencia acumulativa de las frecuencias a partir del numero total de observaciones (en frecuencias absolutas) y de la unidad (en frecuencias relativas)y dirigidos hacia los intervalos superiores; también se puede comenzar desde el último intervalo de clase, sumando acumulativamente hacia intervalos inferiores.
* En el caso de las frecuencias absolutas el primer valor será igual al total de observaciones.
* En el caso de las frecuencias relativas el valor del primer intervalo será igual a la unidad.

-En el caso de las frecuencias porcentuales el valor del primer intervalo será igual al 100%

# Ejemplo.-

F5D = F5 = 3

F5D = F5 + F4 = 3 + 3 = 6 etc.....

H5D = h5 = 0.20

H5D = h5 + h4 = 0.20 + 0.20 = 040 etc....

H5%D = h5% = 20

H5%D = h5% + h4% = 20 + 20 = 40 etc....

* De otra forma:

H5%D = H5D X 100 = 0.20 X 100 = 20

H4%D = H4D X 100 = 0.40 X 100 = 40

* + 1. Reglas Generales para Formar una Tabla de Distribución de Frecuencia. -

**1º** Ordenar los datos en forma ascendente o descendente e Identificar el valor máximo y mínimo de las observaciones.

**2º** Hallar el rango o recorrido.

**3º** Encontrar la amplitud de intervalos de clases.

**4º** Determinar el número de intervalos en caso de no conocerlo

**5º** Encontrar el rango ideal bajo la siguiente condición:

**RI > = R** En el caso que no se cumpla la mencionada condición, se deberá ajustar la amplitud de intervalos (C) al número entero inmediato superior.

85

Ej. **(**R=85; NºI=7; C= ----- = 12.14 13; RI = 13 x 7 = 91 > 85**)**

7

**6º** Hallar el margen de desplazamiento.

**7º** Hallar los límites iniciales y finales para la distribución

**8º** Hallar los límites superiores e inferiores para cada uno de los intervalos de clase.

**9º** Hacer el conteo de las observaciones que se encuentran dentro de cada intervalo y formar una tabla de distribución de frecuencias Absolutas, Relativas, y Porcentuales.

**5.-**Representación Gráfica

Coordenadas Rectangulares. - Las distribuciones de frecuencias se representan mediante diferentes tipos de gráficos en el plano de los ejes cartesianos. Los intervalos de clases se representan en el plano horizontal, eje de las Abscisas (Variable X) y las diferentes distribuciones de las frecuencias Absolutas se las representa en el plano vertical, ejes de las Ordenadas (Variable Y).

# Y

**ORDENADAS** (VERTICAL)

────────────>**X**

**ABSCISAS** (HORIZONTAL)

Tipos de Representación Gráficas. - Existen 3 tipos de Representación Gráfica que son: Histograma, Polígono de Frecuencia y Curva de Frecuencia.

1. Histograma. - Es una serie de rectángulos expresados en el primer cuadrante de los ejes de las coordenadas cartesianas, donde cada uno de los rectángulos, representa las Frecuencias Absolutas de un determinado intervalo de clase. El conjunto total de los rectángulos forma el total de observaciones estudiadas.
2. Polígono de Frecuencia. - Es la unión de los puntos centrales del histograma es decir las distintas marcas de clases, mediante líneas quebradas que se prolongan a una marca de clases inmediatamente superior e inferior, En la práctica para obtener rápidamente un polígono de frecuencia, se transfiere la información del histograma de un aérea de la curva quebrada hacia el polígono de frecuencia.
3. Curvas de Frecuencia. - Al igual que con el histograma, el polígono de frecuencia es utilizado para formar una curva de frecuencia. En la práctica se transfiere toda la información existente en el polígono de frecuencia haciendo pasar una curva suavizada, por los puntos medios o marcas de clase.

(y’i-1) (y’i)

10 27 44 61 78 95

4.5

4

3.5

3

2.5

2

1.5

1

0.5

0

**fi**

Tipos de Curvas de Frecuencias. -Las frecuencias Absolutas (Simples) se pueden clasificar de acuerdo a su punto máximo y de acuerdo a su estructura.

(y'í-1) (yí')

1. De acuerdo a su Punto Máximo se clasifican en Unimodal, Bimodal, y Multimodal.
	1. Curva Unimodal. - Se denomina a la curva de frecuencia que se registra en un solo vértice.
	2. Curva Bimodal. - Es cuando la curva de frecuencia se registra en dos vértices.
	3. Curva Multimodal. - Es cuando la curva de frecuencia se registra en más de dos vértices.
2. De acuerdo a su Estructura se clasifica en Simétrica Positiva, Simétrica Negativa, y Curva Simétrica.
	1. Asimétrica Positiva. - Es cuando la curva se desplaza hacia la derecha.
	2. Asimétrica Negativa. - Es cuando la curva se desplaza hacia la izquierda.
	3. Simétrica Perfecta. - Es cuando la curva se desplaza hacia ambos lados del punto máximo.

**-** Las Frecuencias Absolutas Acumuladas Ascendentes y Descendente, se las puede representar mediante diferentes tipos de curvas denominadas OJIVAS las cuales se clasifican en función a la marca de clases y en función de los limites reales.

1. En función a las marcas de clases se clasifican en Ojivas Suavizadas Ascendente y Ojivas Suavizadas Descendente.
	1. Ojivas Suavizadas Ascendentes. - La representación gráfica se la efectúa en base a la frecuencia absoluta acumulada ascendente en relación con la marca de clases.

**Fi**

# Yi

1 2 3 4 5

86.5

52.5 69.5

18.5 35.5

20

15

10

5

0

* 1. Ojivas Suavizadas Descendente. - La representación gráfica se la realiza en base a la frecuencia absoluta acumulada descendente, en relación con la marca de clases.

**Fi**

# Yi

1 2 3 4 5

86.5

52.5 69.5

18.5 35.5

20

15

10

5

0

1. En función a los limites reales de los intervalos de clase. - Se clasifican en Ojivas escalonadas ascendentes y Ojivas escalonadas descendente.
	1. Ojivas Escalonadas Ascendente. -La representación gráfica se la efectúa en base a las frecuencias absolutas acumuladas ascendente con relación a los limites reales de los intervalos de clase.

(**Y’i-1) (Yì**)

# Fi

* 1. Ojivas Escalonada Descendente. -La representación gráfica se la realiza en base a las frecuencias absolutas acumuladas descendente con relación a los limites reales de los intervalos de clase.

**Fi**

(**Y’i-1) (Yì**)

**-** Las Frecuencias Relativas Porcentuales(Simples). - Se las puede representar mediante diagramas de forma circular tomando como referencia que el total de la circunferencia es igual al 100% de la información.

Dicha representación gráfica se la conoce con el nombre de Ciclograma.

**0.13**

**0.20**

**0.20**

**hi**

**0.27**

**0.20**

**5.2** Otras Gráficas. - Cuando se trabaja con datos no tabulados, a veces es necesario la graficación con la finalidad de una mejor interpretación de la información. Para este caso se pueden utilizar diagramas de bases simples, barras compuestas, barras proporcionales etc.

# FIN …

Ejercicio 5.3. Los valores del pH sanguíneo de 32 individuos son los siguientes