

CONCEPTOS PREVIOS

• ESTADISTICA

Es una ciencia que nos proporciona un conjunto de métodos y procedimientos para la recolección, clasificación, organización, presentación, análisis e interpretación de datos en forma adecuada con el fin de tomar decisiones más efectivas.

• CLASES DE ESTADISTICA

ESTADISTICA DESCRIPTIVA

Es la parte de la estadística que trata de recopilar, clasificar, presentar y describir datos estadísticos.

ESTADISTICA INFERENCIAL

Es la parte de la estadística cuyo objetivo es investigar cómo deben ser utilizados los datos para producir resultados o probar algunas hipótesis. La estadística inferencial usa la teoría muestral que parte de una muestra para inferir conclusiones de una población.

• POBLACION

Es un conjunto de datos referentes a determinadas características de un grupo de individuos o elementos.

Ejemplo

Las edades de los alumnos de la Escuela Superior de la Policía Nacional del Perú.

• MUESTRA

Es un subconjunto de la población total tomado al azar. Es representativa de la población.

Ejemplo

Las edades de los alumnos de la Especialidad de Administración de la Escuela Superior de la Policía Nacional del Perú.

DISTRIBUCIONES DE FRECUENCIAS

Consiste en distribuir los datos de la muestra en clases o categorías e ir colocando el número de datos que caen en cada intervalo. // Tabla estadística de resumen en la que los datos se agrupan en clases cuando los datos se agrupan o condensan en tablas de distribución de frecuencias, el proceso de análisis e interpretación de los datos se vuelve mucho más manejable y significativo.

FRECUENCIA

Número de veces que un dato se presenta en un conjunto de observaciones. La frecuencia se clasifica en absoluta y relativa.

FRECUENCIA ABSOLUTA

Es el numero de datos que caen dentro de cada intervalo de clase. Se denota por **f_i** y debe cumplir con la siguiente propiedad :

FRECUENCIA RELATIVA

Cociente entre la frecuencia absoluta y el numero total de datos . // Cociente entre la frecuencia absoluta y el numero de observaciones .debe cumplir con las siguientes propiedades:

f_i Frecuencia Absoluta f_i

$$h_i = \frac{f_i}{n} = \frac{f_i}{N^\circ \text{ de Observaciones}}$$

n N° de Observaciones f_i

RANGO (R)

Es la longitud de un conjunto de observaciones, resulta de la diferencia entre el valor máximo y valor mínimo de un conjunto de datos.

$$R = V.\text{Max} - V.\text{Min}$$

N° DE INTERVALO DE CLASE (K)

Son grupos que resultan de particionar el rango. El numero de grupos a particionar o intervalos de clase (K) se determina por la Regla de Sturges.

$$K = 1 + 3.32 \text{ Log } n$$

(*) Redondeando el entero superior e inferior según convenga.

Donde :

K : N° de intervalos de clase

N : N° de observaciones

ANCHO DE CLASE (W)

Se refiere al tamaño que debe tener cada intervalo de clase. Es decir a su amplitud el ancho de clase (W) resulta de dividir al Rango (R) el N° de Intervalos de clase (K)

R Rango

$$W = \frac{R}{K} = \frac{\text{Rango}}{\text{N° de Intervalos de clase}}$$

K N° de Intervalos de clase

MARCA DE CLASE

Son los puntos medios de los intervalos de clase . Resulta de la suma del limite inferior del intervalo de clase (LI) mas el limite superior del intervalo de clase(LS) dividido entre 2

LI +LS Limite inferior del intervalo de clase + Limite superior del intervalo de clase

$$X_i = \frac{\text{LI} + \text{LS}}{2} =$$

• 2

COMPONENTES BASICOS DE UNA TABLA DE DISTRIBUCION DE FRECUENCIAS

Intervalos	fi
[14– 18 >	5
[18– 22 >	12
[22– 26 >	10
[26– 30 >	8

INTERPRETACION DEL INTERVALO ABIERTO E INTERVALO CERRADO

Intervalos	Interpretación del intervalo cerrado y abierto
[14– 18 >	Indica que se inicia en 14 hasta 17.999....
[18– 22 >	Indica que se inicia en 18 hasta 21.999.....
[22– 26 >	Indica que se inicia en 22 hasta 25.999.....
[26– 30 >	Indica que se inicia en 26 hasta 29.999.....

ELABORACION DE LA TABLA DE

DISTRIBUCION DE FRECUENCIAS

CASO 1

Paso 1: Observación

En un operativo en el distrito de Villa María del Triunfo se efectuó la detención de 50 consumidores de drogas observándose que sus edades fluctúan entre 12 años y 46 años

Paso 2 : Recopilación

Al momento de su detención se solicito sus edades de los 50 consumidores siendo los datos los siguientes

12	14	14.1	18	18
18	15	15	16	16
19	20	20	16	16
22	24	24	24	30
30	30	29	29	29
31	31	32	32	33
33	34	35	35	36
37	37	28	38	38
39	40	42	44	44
14	15	15	26	46

Paso 3 : Organización de los Datos

3.1 Determinar el Rango :

$$R= V.Max - V.Min$$

El rango es el Valor máximo menos el mínimo

$$El V.Max = 46 (de la tabla de datos)$$

$$El V.Min = 12 (de la tabla de datos)$$

Entonces :

$$R= 46 - 12$$

$$R= 34$$

3.2 Determinar el N° de Intervalos de Clase

$$K= 1 + 3.32 \text{ Log } n$$

(* Redondeando el entero superior e inferior según convenga

Donde

$n = N^\circ$ de observaciones

$$n= 50$$

Entonces :

$$K= 1 + 3.32 \text{ Log } n$$

$$K= 1 + 3.32 \text{ Log } (50)$$

$$K= 1 + 3.32 * 1.69$$

$$K= 1 + 5.64$$

$$K = 6.64$$

(* Redondeando el entero superior e inferior según convenga.

$$K = 7$$

3.3 Determinar el Ancho de Clase

R Rango

$$W = \text{-----} = \text{-----}$$

K N° de Intervalos de clase

Reemplazando el valor del rango (obtenido en 3.1) y el N° de Intervalos (obtenido en 3.2) en la formula 3.3 del ancho de clase

Entonces :

34

$$W = \frac{34}{7} = 4.857 = 5$$

7

3.4 Calculando el Limite Inferior y Superior de los Intervalos

1er Intervalo

Limite Inferior = Valor mínimo (del conjunto de datos)

Limite Inferior = 12

Limite Superior = Limite Inferior + Ancho de Clase

Limite Superior = 12 + 5

Limite Superior = 17

Entonces :

1er Intervalo : [12– 17 >

2do Intervalo

Limite Inferior = Limite superior anterior (1er intervalo)

Limite Inferior = 17

Limite Superior = Limite Inferior + Ancho de Clase

Limite Superior = 17 + 5

Limite Superior = 22

Entonces :

2do Intervalo : [17 – 22 >

3er Intervalo

Limite Inferior = Limite superior anterior (2do intervalo)

Limite Inferior = 22

Limite Superior =Limite Inferior +Ancho de Clase

Limite Superior =22 + 5

Limite Superior =27

Entonces :

3er Intervalo : [22 – 27 > y así sucesivamente hasta cubrir todos los datos.

Finalmente :

• **Tabla de Distribución de Frecuencias**

1er intervalo

Se contabiliza las observaciones de consumidores de drogas entre 12 años y 16.999999 ...años y se coloca en fi (frecuencia absoluta) del primer intervalo [12– 17 > en este caso se contabilizo 12 personas cuyas edades están entre 12 y 16.999.....

2do intervalo

Se contabiliza las observaciones de consumidores de drogas entre 17 años y 21.999999 ...años y se coloca en fi (frecuencia absoluta) del segundo intervalo [17– 22 > en este caso se contabilizo 6 personas cuyas edades están entre 17 años y 21.999.....

Y así sucesivamente hasta el 7mo intervalo.

Intervalos	fi
[12– 17 >	12
[17– 22 >	6
[22– 27 >	5
[27– 32 >	9
[32– 37 >	8
[37– 42 >	6
[42– 47 >	4

INTERPRETACION

Existe 12 Consumidores de drogas entre 12 y 16 años

Existe 6 Consumidores de drogas entre 17 y 21 años

Existe 5 Consumidores de drogas entre 22 y 26 años

Existe 9 Consumidores de drogas entre 27 y 31 años

Existe 8 Consumidores de drogas entre 32 y 36 años

Existe 6 Consumidores de drogas entre 37 y 41 años

Existe 14 Consumidores de drogas entre 42 y 46 años

*** Se puede concluir que el mayor consumo de drogas en el Distrito de Villa María del Triunfo se concentra en adolescentes personas menores de 18 años debido a que se ubican entre 12 y 16 años**

ELABORACION DE LA TABLA DE

DISTRIBUCION DE FRECUENCIAS

CASO 2

VARIABLE CUANTITATIVA CONTINUA (Decimal)

Paso 1: Observación

En el Centro de Lima se llevo a cabo un operativo para detener el robo a los transeúntes y comerciantes obteniéndose la captura de 84 delincuentes quienes se les encontró diferentes montos productos del robo.

Paso 2 : Recopilación

Al momento de su detención se recupero los siguientes montos

93	83	103	96	100	96	89
132	102	106	102	101	101	132
94	95	98	114	101	127	101
102	121	99	95	100	118	107
107	115	100	94	97	104	96
84	108	93	87	116	93	92
63	96	100	104	94	101	100
132	94	101	100	99	103	104
101	80	118	101	102	97	90
103	71	81	123	73	100	82
104	68.2	80	98	85	88	107
125	70.1	105	98	100	75	103

Paso 3 : Organización de los Datos

3.1 Determinar el Rango :

$$R= V.Max - V.Min$$

El rango es el Valor máximo menos el mínimo

El V.Max = 132 (de la tabla de datos)

El V.Min = 63 (de la tabla de datos)

Entonces :

$$R = 132 - 63$$

$$R = 69$$

3.2 Determinar el N° de Intervalos de Clase

$$K = 1 + 3.32 \text{ Log } n$$

(*) Redondeando el entero superior e inferior según convenga

Donde

$n = N^\circ$ de observaciones

$$n = 84$$

Entonces :

$$K = 1 + 3.32 \text{ Log } n$$

$$K = 1 + 3.32 \text{ Log } (84)$$

$$K = 1 + 3.32 * 1.9243$$

$$K = 1 + 6.38$$

$$K = 7.38$$

(*) Redondeando el entero superior e inferior según convenga.

$$K = 7$$

3.3 Determinar el Ancho de Clase

R Rango

$$W = \text{-----} = \text{-----}$$

K N° de Intervalos de clase

Reemplazando el valor del rango (obtenido en 3.1) y el N° de Intervalos (obtenido en 3.2) en la formula 3.3 del ancho de clase

Entonces :

$$69$$

$$W = \text{-----} = 9.86 = 10$$

$$7$$

RANGO NUEVO R1

Si :

R

W = -----

K

Entonces despejando

$R1 = W * K$

Reemplazando :

$R1 = 10 * 7$

$R1 = 70$

Comparando Rango original (R) con Rango Nuevo (R1)

$EXCESO = R1 - R = 70 - 69 = 1$ El exceso se agrega al limite superior

$R = 69$

63-----69-----132

$R1 = 70$

+1

63-----70-----133

IMPORTANTE

- Cuando hay un exceso de 1 le agregamos al limite superior
- Cuando el exceso es 2 repartimos en partes proporcionales en ambos limites
- Cuando el exceso es 3 se reparte 1 a la izquierda (se le resta al limite inferior) y 2 a la derecha (se le agrega al limite superior)
- El exceso no puede ser mayor a 5

3.4 Calculando el Limite Inferior y Superior de los Intervalos

1er Intervalo

Limite Inferior = Valor mínimo (del conjunto de datos)

Limite Inferior = 63

Limite Superior = Limite Inferior + Ancho de Clase

Limite Superior = 63 + 10

Limite Superior =73

Entonces :

1er Intervalo : [63– 73 >

2do Intervalo

Limite Inferior = Limite superior anterior (1er intervalo)

Limite Inferior = 73

Limite Superior =Limite Inferior +Ancho de Clase

Limite Superior =73 + 10

Limite Superior =83

Entonces :

2do Intervalo : [73 – 83 >

3er Intervalo

Limite Inferior = Limite superior anterior (2do intervalo)

Limite Inferior = 83

Limite Superior =Limite Inferior +Ancho de Clase

Limite Superior =83 + 10

Limite Superior =93

3er Intervalo : [83 – 93 > y así sucesivamente hasta cubrir todos los datos.

Finalmente :

• Tabla de Distribución de Frecuencias

1er intervalo

Se contabiliza las observaciones de robos entre 63 y 72.999 ...nuevos soles y se coloca en fi (frecuencia absoluta) del primer intervalo [63 – 73 > en este caso se contabilizo 4 robos entre 63 y 72.999 ...nuevos soles

2do intervalo

Se contabiliza las observaciones de robos entre 73 y 82.999 ...nuevos soles y se coloca en fi (frecuencia absoluta) del segundo intervalo [73 – 83 > en este caso se contabilizo 6 robos entre 73 y 82.999 ...nuevos soles

Y así sucesivamente hasta el 7mo intervalo.

Intervalos	fi
[63– 73 >	4
[73– 83 >	6
[83– 93 >	8
[93– 103 >	40
[103– 113 >	14
[113– 123 >	6
[123– 133 >	6

INTERPRETACION

Existe 4 delincuentes que robaron entre 63 y 72 nuevos soles

Existe 6 delincuentes que robaron entre 73 y 82 nuevos soles

Existe 8 delincuentes que robaron entre 83 y 92 nuevos soles

Existe 40 delincuentes que robaron entre 93 y 102 nuevos soles

Existe 14 delincuentes que robaron entre 103 y 112 nuevos soles

Existe 6 delincuentes que robaron entre 113 y 122 nuevos soles

Existe 6 delincuentes que robaron entre 123 y 132 nuevos soles

*** Se puede concluir que el mayor numero de robos a transeúntes y comerciantes en el Centro de Lima se focaliza en aquellas personas cuyo monto de efectivo fluctúa entre 93 y 102 nuevos soles.**

ELABORACION DE LA TABLA DE

DISTRIBUCION DE FRECUENCIAS

CASO 3

VARIABLE CUANTITATIVA DISCRETA (Entera)

Paso 1: Observación

En el Agustino en la zona de mayor pobreza extrema se realizo una encuesta a los hogares afectadas por violencia familiar para averiguar el numero de personas que vive en cada vivienda obteniendo los siguientes resultados.

Paso 2 : Recopilación

Al momento de visitar las viviendas en la zona de mayor pobreza extrema se eligió el asentamiento humano " 10 de Octubre "

5	6	4	7	5	4	7	4	5	7

Paso 3 : Organización de los Datos

En este caso no construye intervalos de clases; tenemos que observar el dato menor (4) y el dato mayor (9) para comenzar a ordenarlos

Tabla de Distribución de Frecuencias

N° de personas	fi (Viviendas)
4	8
5	10
6	8
7	13
8	5
9	6

INTERPRETACION

Hay 8 viviendas que están integradas por 4 personas

Hay 10 Viviendas que están integradas por 5 personas

Hay 8 viviendas que están integradas por 6 personas

Hay 13 viviendas que están integradas por 7 personas

Hay 5 viviendas que están integradas por 8 personas

Hay 6 viviendas que están integradas por 9 personas

*** Se puede concluir que en el asentamiento humano 10 de octubre las viviendas se encuentran ocupadas en su mayoría por un excesivo numero de personas por ejemplo existen 13 viviendas donde habitan 7 personas y en otras 9 personas. Siendo el hacinamiento en zonas de pobreza extrema una de las principales causas de violencia familiar.**

ELABORACION DE LA TABLA DE

DISTRIBUCION DE FRECUENCIAS

CASO 4

VARIABLE CUALITATIVA O CATEGORICA

Ejemplo 1

Paso 1: Observación

En la Comisaria del distrito de San Roque a 20 jóvenes que fueron detenidos por el consumo de drogas se le consulto las razones por las cuales se iniciaron en el consumo de PBC e indicaron lo siguiente :

PF = PROBLEMAS FAMILIARES CU = CURIOSIDAD

ES = ENTORNO SOCIAL OT = OTROS

Paso 2 : Recopilación

PF	ES	ES	CU	PF
CU	PF	ES	PF	ES
ES	PF	CU	PF	OT
PF	OT	ES	OT	PF

Paso 3 : Organización de los Datos

En este caso no construye intervalos de clases; tenemos que ordenar los datos de acuerdo a sus características o cualidades.

Tabla de Distribución de Frecuencias

Consumo de Drogas (Causa)	N° Jóvenes
Problemas familiares	8
Entorno social	6
Curiosidad	3
Otros	3

INTERPRETACION

En el distrito de San Roque los jóvenes se inician en el consumo de PBC por los siguientes factores:

8 Jóvenes por Problemas familiares

6 Jóvenes por el Entorno Social

3 Jóvenes por Curiosidad

3 Jóvenes por otros factores

*** Se puede concluir que en el distrito de San Roque la iniciación en el consumo de Pasta Básica de Cocaína por parte de la población juvenil se explica principalmente por los problemas familiares que son determinativos y explicativos del consumo de PBC. Otro factor explicativo importante es el entorno social (grupos de desarrollo social). El consumo de droga genera criminalidad potencial de tal manera que para evitar el desarrollo delincriminal el núcleo familiar debe ofrecer un clima de estabilidad.**

Ejemplo 2

Paso 1: Observación

En el Perú de acuerdo a datos del Compendio de Estadísticas Policiales 2004 publicado en la Web de la Policía Nacional del Perú : www.pnp.gob.pe el numero de delitos se encuentran clasificado (en categorías). Presentando características específicas los delitos se distribuyen de acuerdo al siguiente cuadro. Toda aquella variable que se presenta con categorías clasificaciones y propiedades se le denomina en la Estadística Variable Cualitativa , de cualidad o Variable Categórica.

Paso 2 : Recopilación

La recopilación se obtuvo de fuentes públicas

PERU : PRINCIPALES DELITOS REGISTRADOS EN EL 2004

TIPO DE DELITO (CATEGORIA)	N° DE DELITOS
Contra la Vida Cuerpo y Salud	19,260
Contra la Familia	5,433
Contra la Libertad	10,160
Contra el Patrimonio	114,574
Contra la Fe Pública	2,114
Contra la Seguridad Pública	5,719

FUENTE : Compendio de Estadísticas Policiales

ELABORACION : ESTADISTICA APLICADA /JMR

INTERPRETACION

En el Perú en el año 2004 19,260 delitos fueron Contra la Vida Cuerpo y Salud

En el Perú en el año 2004 5,433 delitos fueron Contra la Familia

En el Perú en el año 2004 10,160 delitos fueron Contra la Libertad

En el Perú en el año 2004 114,574 delitos fueron Contra el Patrimonio

En el Perú en el año 2004 2,114 delitos fueron Contra la Fe Pública

En el Perú en el año 2004 5,719 delitos fueron Contra la Seguridad Pública

*** Se puede concluir que en el Perú los delitos se concentran en delitos contra el patrimonio y contra la vida señalando que la política de seguridad pública debe orientarse a focalizar e implementar estrategias para reducir el número de estos delitos.**

FRECUENCIA ABSOLUTA, FRECUENCIA ABSOLUTA ACUMULADA, FRECUENCIA RELATIVA Y FRECUENCIA RELATIVA ACUMULADA

FRECUENCIA ABSOLUTA

Es el número de datos que caen dentro de cada intervalo de clase. Se denota por f_i y debe cumplir con la siguiente propiedad :

$$f_i \geq 0$$

$$f_1 + f_2 + f_3 + \dots + f_n = n$$

$$f_i \leq n$$

La suma de las frecuencias absolutas es equivalente al número de observaciones del conjunto de datos.

FRECUENCIA ABSOLUTA ACUMULADA

Aquel valor que se obtiene acumulando las frecuencias absolutas. // Aquella que resulta de sumar sucesivamente las frecuencias absolutas.

$$F1 = f1$$

$$F2 = f1 + f2$$

$$F3 = f1 + f2 + f3$$

$$Fn = f1 + f2 + f3 + \dots + fn = n$$

FRECUENCIA RELATIVA

Cociente entre la frecuencia absoluta y el numero total de datos . // Cociente entre la frecuencia absoluta y el numero de observaciones .debe cumplir con las siguientes propiedades:

f i Frecuencia Absoluta f i

$$hi = \frac{fi}{n} = \frac{fi}{\text{N}^\circ \text{ de Observaciones}}$$

n N° de Observaciones f i

PROPIEDADES

$$hi \geq 0$$

$$\sum hi = 1$$

FRECUENCIA RELATIVA ACUMULADA

Aquel valor que se obtiene acumulando las frecuencias relativas. // Aquella que resulta de sumar sucesivamente las frecuencias relativas.

$$H1 = h1$$

$$H2 = h1 + h2$$

$$H3 = h1 + h2 + h3$$

$$Hn = h1 + h2 + h3 + \dots + hn = 1$$

EJEMPLO DE APLICACIÓN

En un operativo en el distrito de Villa María del Triunfo se efectuó la detención de 50 consumidores de drogas observándose que sus edades fluctúan entre 12 años y 46 años los datos ya tabulados aparecen en la siguiente distribución de frecuencias se pide calcular :

- La frecuencia absoluta acumulada
- La Frecuencia Relativa y Frecuencia Relativa Acumulada
- Calcular la marca de clase

TABLA DE DISTRIBUCION DE FRECUENCIAS

CONSUMO DE DROGAS (EDADES ENTRE 12 Y 46 AÑOS)

DISTRITO VILLA MARAI DEL TRIUNFO

N= 50 Observaciones

Solución :

- La frecuencia absoluta acumulada

$$F1 = 12$$

$$F2 = 12 + 6 = 18$$

$$F3 = 12 + 6 + 5 = 23$$

$$F4 = 12 + 6 + 5 + 9 = 32$$

$$F5 = 12 + 6 + 5 + 9 + 8 = 40$$

$$F6 = 12 + 6 + 5 + 9 + 8 + 6 = 46$$

$$F7 = 12 + 6 + 5 + 9 + 8 + 6 + 4 = 50$$

- La frecuencia relativa

$$h1 = f1/ n = 12/50 = 0.24$$

$$h2 = f2/ n = 6/50 = 0.12$$

$$h3 = f3/ n = 5/50 = 0.10$$

$$h4 = f4/ n = 9/50 = 0.18$$

$$h5 = f5/ n = 8/50 = 0.16$$

$$h6 = f6/ n = 6/50 = 0.12$$

$$h7 = f7/ n = 4/50 = 0.08$$

Intervalo abierto

Intervalo cerrado

Frecuencia Absoluta

Limite Superior (LS)

Limite Inferior (LI)

Valor máximo = 46

Valor mínimo = 12

Valor mínimo = 63

Valor máximo = 132

f i = n

hi = 1

Frecuencia relativa acumulada

H1 = 0.24

H2 = 0.24 + 0.12 = 0.36

H3 = 0.24 + 0.12 + 0.10 = 0.46

H4 = 0.24 + 0.12 + 0.10 + 0.18 = 0.64

H5 = 0.24 + 0.12 + 0.10 + 0.18 + 0.16 = 0.80

H6 = 0.24 + 0.12 + 0.10 + 0.18 + 0.16 + 0.12 = 0.92

H7 = 0.24 + 0.12 + 0.10 + 0.18 + 0.16 + 0.12 + 0.08 = 1.00

c) Calcular la marca de clase

X 1 = (Li 1 + Ls 1) / 2 = 12+17 / 2 = 14.5

X 2 = (Li 2 + Ls 2) / 2 = 17+ 22 /2 = 19.5

X 3 = (Li 3 + Ls 3) / 2 = 22 + 27 /2 = 24.5

X 4 = (Li 4 + Ls 4) / 2 = 27+ 32 /2 = 29.5

X 5 = (Li 5 + Ls 5) / 2 = 32+ 37/ 2 = 34.5

X 6 = (Li 6 + Ls 6) / 2 = 37 + 42 /2 = 39.5

X 7 = (Li 7 + Ls 7) / 2 = 42 + 47 /2 = 44.5