IEDT PABLO NERUDA BARRANQUILLA – 2020 GUÍA DE TRABAJO DE ESTADISTICA N° 2

ESTUDIANTE: _____ CURSO: _____ FECHA: _____

**OBJETIVO:** Desarrollar tablas de distribución de frecuencia para orientar la realización de un análisis estadístico de la información.
**DBA:** Interpreta, valora y analiza críticamente los resultados y las inferencias presentadas en estudios estadísticos.

**INDICADOR(ES):**
- Elaborar tablas de distribución de frecuencia a partir de información dada.
- Elaborar gráficos estadísticos.
- Extraer información estadística a partir del análisis de tablas y gráficas.
- Hace inferencias sobre los parámetros basadas en los estadígrafos calculados.
- Hace análisis críticos de las conclusiones de los estudios presentados en medios de comunicación o en artículos científicos.

1. SITUACIÓN PROBLÉMICA

Observa y analiza las tablas siguientes.

| Intervalos reales | | Marcas de clase | Frecuencia absoluta | Frecuencia acumulada | Frecuencia relativa | Frecuencia rel. acumulada |
|---|---|---|---|---|---|---|
| Límite inferior | Límite superior | $x_i$ | $f_i$ | $fa_i$ | $fr_i$ | $fra_i$ |
| 40.5 | 46.5 | 43.5 | 3 | 3 | 0.010000 | 0.010000 |
| 46.5 | 52.5 | 49.5 | 9 | 12 | 0.030000 | 0.040000 |
| 52.5 | 58.5 | 55.5 | 23 | 35 | 0.076667 | 0.116667 |
| 58.5 | 64.5 | 61.5 | 43 | 78 | 0.143333 | 0.260000 |
| 64.5 | 70.5 | 67.5 | 62 | 140 | 0.206667 | 0.466667 |
| 70.5 | 76.5 | 73.5 | 63 | 203 | 0.210000 | 0.676667 |
| 76.5 | 82.5 | 79.5 | 53 | 256 | 0.176667 | 0.853333 |
| 82.5 | 88.5 | 85.5 | 26 | 282 | 0.086667 | 0.940000 |
| 88.5 | 94.5 | 91.5 | 16 | 298 | 0.053333 | 0.993333 |
| 94.5 | 100.5 | 97.5 | 2 | 300 | 0.006667 | 1.000000 |

| | Frecuencia absoluta | Frecuencia absoluta acumulada | Frecuencia relativa | Frecuencia relativa acumulada |
|---|---|---|---|---|
| Pesos (Kg) | $f_i$ | $F_i$ | $h_i$ | $H_i$ |
| 53 | 1 | 1 | 2,00% | 2,00% |
| 55 | 1 | 2 | 2,00% | 4,00% |
| 56 | 1 | 3 | 2,00% | 6,00% |
| 57 | 2 | 5 | 4,00% | 10,00% |
| 58 | 2 | 7 | 4,00% | 14,00% |
| 59 | 2 | 9 | 4,00% | 18,00% |

Con base en la observación y el análisis responde:

a. ¿Cuál es la variable en cada tabla? ¿Cuál es la población?
b. ¿Consideras que la segunda tabla está completa? Explica.
c. ¿Qué diferencia observas entre ambas tablas?
d. ¿Cuál es la temática que se aborda en ambas tablas? ¿Cómo las titularías?
e. ¿Qué elementos estadísticos reconoces en las tablas?

2. LECTURA: Lee, analiza, resume y consigna en tu cuaderno.

TABLAS DE DISTRIBUCIÓN DE FRECUENCIAS

Las tablas de distribución de frecuencias se utilizan cuando se recolectan datos, con ellas se pueden representar los datos de manera que es más fácil analizarlos.
Se pueden elaborar tablas de distribución de frecuencias para datos no agrupados y para datos agrupados. Estas últimas se utiliza cuando se tienen muchos datos.

Para elaborar tablas de distribuciones de frecuencia se debe tener en cuenta lo siguiente:
Cuando hay muchos datos se agrupan en clases.

Clase es cada uno de los grupos en que se dividen los datos. Para determinar cuántas clases crear, se puede utilizar la siguiente fórmula (fórmula de Sturges)

Número de clases = 1 + 3,322 log n donde n es el número total de datos.

También se puede aplicar la formula siguiente: Número de clases = raíz cuadrada de n (número total de datos).
Si al aplicar la fórmula se obtiene un número decimal, se aproxima al siguiente entero.

El intervalo de clase o el ancho de la clase (tamaño de la clase o rango) es el espacio que hay entre el límite superior y el límite inferior de la clase, los cuales corresponden a los valores extremos de la clase. Para obtener el ancho de clase se utiliza la siguiente fórmula:

Ancho de clase = (dato superior – dato inferior)/ número de clases

La frecuencia absoluta es el número de veces que se repite cada dato. Cuando se agrupan los datos, es el número de datos que tiene cada clase. Se simboliza con $f_j$.

La marca de clase es el punto medio de la clase. Se obtiene dividiendo entre dos la suma de los valores extremos de cada clase.

El rango es la diferencia entre el valor mayor y el valor menor en estudio de una distribución de datos.

La frecuencia absoluta acumulada es la frecuencia total hasta el límite superior de cada clase. Se simboliza con $F_j$.

La frecuencia relativa de un dato da información sobre qué parte de la población o de la muestra en estudio corresponde a la característica analizada. Se obtiene dividiendo la frecuencia absoluta entre el número total de datos y se puede expresar como una fracción, como un decimal o como un porcentaje. Se simboliza con $f_j$ / n donde n es el número de datos.

La frecuencia acumulada relativa es la frecuencia relativa total hasta el límite superior de cada clase. Se simboliza con $F_j$ / n donde n es el número total de datos.

Ejemplo:

Suponga que un investigador desea determinar cómo varía el peso de un grupo de estudiantes de primer semestre de una universidad. Selecciona una muestra de 50 estudiantes y registra sus pesos en kilogramos. Los datos obtenidos fueron los siguientes:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 65 | 63 | 65 | 63 | 69 | 67 | 53 | 58 | 60 | 61 |
| 64 | 65 | 64 | 72 | 68 | 66 | 55 | 57 | 60 | 62 |
| 64 | 65 | 64 | 71 | 68 | 66 | 56 | 59 | 61 | 62 |
| 63 | 65 | 63 | 70 | 67 | 66 | 57 | 59 | 61 | 62 |
| 64 | 64 | 63 | 69 | 67 | 66 | 58 | 60 | 61 | 62 |

Para determinar el número de veces que aparece cada dato (frecuencia absoluta), se utiliza el diagrama de tallo y hojas. Se traza una línea y a la izquierda se escriben las cifras anteriores a las unidades que tengan los datos, a la derecha de la línea se escriben la cifra de las unidades para cada uno de los datos. Este diagrama facilita determinar la cantidad de veces que se repite un dato y los valores de los datos con el fin de escribirlos de manera ordenada en la tabla.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ~~64~~ | ~~65~~ | ~~64~~ | ~~71~~ | ~~68~~ | ~~66~~ | ~~56~~ | ~~59~~ | ~~61~~ | ~~62~~ |
| ~~63~~ | ~~69~~ | ~~67~~ | ~~53~~ | ~~58~~ | ~~65~~ | ~~63~~ | ~~65~~ | ~~60~~ | ~~61~~ |
| ~~57~~ | ~~59~~ | ~~61~~ | ~~62~~ | ~~63~~ | ~~65~~ | ~~63~~ | ~~70~~ | ~~67~~ | ~~66~~ |
| ~~64~~ | ~~65~~ | ~~64~~ | ~~72~~ | ~~68~~ | ~~66~~ | ~~55~~ | ~~57~~ | ~~60~~ | ~~62~~ |
| ~~67~~ | ~~66~~ | ~~58~~ | ~~60~~ | ~~61~~ | ~~62~~ | ~~64~~ | ~~64~~ | ~~63~~ | ~~69~~ |

| | |
|---|---|
| 5 | 6 9 3 8 7 9 5 7 8 |
| 6 | 4 5 4 8 6 1 2 3 9 7 5 3 5 0 1<br>1 2 3 5 3 7 6 4 5 4 8 6 0 2 7<br>6 0 1 2 4 4 3 9 |
| 7 | 1 0 2 |

Luego, se organiza la información en la tabla, de la siguiente manera:

## Tabla de datos no agrupados

| | Frecuencia absoluta | Frecuencia absoluta acumulada | Frecuencia relativa | Frecuencia relativa acumulada |
|---|---|---|---|---|
| Pesos (Kg) | $f_i$ | $F_i$ | $h_i$ | $H_i$ |
| 53 | 1 | 1 | 2,00% | 2,00% |
| 55 | 1 | 2 | 2,00% | 4,00% |
| 56 | 1 | 3 | 2,00% | 6,00% |
| 57 | 2 | 5 | 4,00% | 10,00% |
| 58 | 2 | 7 | 4,00% | 14,00% |
| 59 | 2 | 9 | 4,00% | 18,00% |

Así sucesivamente…

Para construir la tabla de datos agrupados se debe calcular primero lo siguiente:

**Número de clases o intervalos:**

$$k = 1 + 3{,}322 \log_{10} n$$

$$k = 1 + 3{,}322 \log_{10} 50 = 6{,}64 \approx 7$$

**Rango:**

Rango = (dato máximo – dato mínimo)

Rango = 72 – 53 = 19

**Ancho de clase (tamaño del intervalo)**

$$l = \text{Rango} / k$$

$$l = 19 / 7 = 2{,}71 \approx 3$$

Al construir la tabla de datos agrupados con la información del ejemplo, se tiene:

## Tabla de datos agrupados

| | Frecuencia absoluta | Frecuencia absoluta acumulada | Frecuencia relativa | Frecuencia relativa acumulada |
|---|---|---|---|---|
| Pesos (Kg) | $f_i$ | $F_i$ | $h_i$ | $H_i$ |
| 53 - 55 | 2 | 2 | 4,00% | 4,00% |
| 56 - 58 | 5 | 7 | 10,00% | 14,00% |
| 59 - 61 | 9 | 16 | 18,00% | 32,00% |
| 62 - 64 | 15 | 31 | 30,00% | 62,00% |
| 65 - 67 | 12 | 43 | 24,00% | 86,00% |
| 68 - 70 | 5 | 48 | 10,00% | 96,00% |
| 71 - 73 | 2 | 50 | 4,00% | 100,00% |
| | 50 | | 100,00% | |

3. ACTIVIDAD N° 1 (Modelación):
A partir de la información anterior, construye tabla de distribución de frecuencia (datos agrupados y no agrupados) con base en el conjunto de datos siguiente.

| | | | | | | |
|---|---|---|---|---|---|---|
| 20 | 30 | 18 | 13 | 14 | 26 | 27 |
| 15 | 10 | 21 | 12 | 13 | 23 | 23 |
| 17 | 28 | 26 | 21 | 19 | 11 | 23 |
| 16 | 27 | 19 | 22 | 16 | 13 | 24 |
| 19 | 29 | 15 | 23 | 16 | 27 | 29 |
| 18 | 21 | 14 | 29 | 24 | 21 | 28 |

4. ACTIVIDAD N° 2 (Aplicación)
4.1. Al finalizar el primer periodo del año lectivo 2020, se realizó un control para reconocer el número de áreas perdidas por un curso de 10º del colegio Pablo Neruda. A continuación se muestra la información.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 7 | 6 | 5 | 0 | 0 | 0 | 4 | 2 | 2 | 2 |
| 0 | 2 | 2 | 8 | 6 | 5 | 3 | 3 | 4 | 1 |
| 1 | 0 | 0 | 0 | 2 | 5 | 9 | 2 | 1 | 1 |
| 2 | 8 | 7 | 4 | 5 | 6 | 1 | 0 | 8 | 7 |

a) Determina: población, variable, tipo de variable.
b) Construye la tabla de distribución de frecuencia (datos no agrupados).
c) Si el cuadro correspondiera al informe final de 2020, conociendo las directrices del colegio, cuántos estudiantes perderían el año sin recuperación, cuántos podrían acceder a recuperación y cuántos ganarían el año sin necesidad de recuperación.
d) Cuánto es el porcentaje de estudiantes que no pierden áreas y de cuánto el porcentaje de estudiantes que pierden más de cuatro áreas.

4.2. Queremos hacer un estudio estadístico del número de Técnicos Superiores en Electricidad (TSE) que existen en las empresas eléctricas de una determinada ciudad. Para ello se ha encuestado a 50 empresas y se han obtenido los siguientes datos:

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 4 | 2 | 3 | 1 | 2 | 4 | 2 | 3 | 0 | 2 | 2 | 2 | 3 | 2 | 6 | 2 | 3 | 2 | 2 | 3 | 2 | 3 | 3 | 4 |
| 3 | 3 | 4 | 5 | 2 | 0 | 3 | 2 | 1 | 2 | 3 | 2 | 2 | 3 | 1 | 4 | 2 | 3 | 2 | 4 | 3 | 3 | 2 | 2 | 1 |

Se pide:
a) ¿Cuál es la población objeto de estudio?
b) ¿Qué variable estamos estudiando?
c) ¿Qué tipo de variable es?
d) Construir la tabla de frecuencias (datos no agrupados)
e) ¿Cuál es el número de empresas que tiene como máximo 2 TSE?
f) ¿Cuántas empresas tienen más de 1 TSE, pero como máximo 3?
g) ¿Qué porcentaje de empresas tiene más de 3 TSE?

5. CONCEPTUALIZACIÓN (complemento). Lee, analiza, resume y consigna en tu cuaderno.

Para elaborar una tabla de distribución de frecuencia para datos agrupados, es relevante lo siguiente:
Rango = R
Número de intervalos = m
Amplitud = A
Las expresiones para calcular lo anterior se presentan a continuación:
R= Xmáx – Xmín, donde Xmáx y Xmín corresponden respectivamente al dato más grande y más pequeño del conjunto de datos.
m = raíz cuadrada de N ó m = 1 + 3,3 log N, donde N corresponde al total de datos.
A = R / m
Es conveniente tener en cuenta lo siguiente para las aproximaciones (en el caso de obtener valores decimales para m y/o A).
m Se puede aproximar a la parte entera o al entero siguiente.
A Solamente se puede aproximar al entero siguiente.
<u>NUEVO RANGO</u>
En algunos casos, luego de hallado los elementos necesarios para construir la tabla de distribución de frecuencia, se puede presentar el caso que los valores para m y A (luego de aproximaciones) alteren el rango. Para este caso es necesario encontrar un nuevo rango modificando los límites iniciales (Xmáx y Xmín). Para verificar si es necesario hacer esto, realizo:

R’= m.A
Si R’ = R entonces es posible iniciar la tabla, pues el rango no se ha alterado.
Si R’ ≠ R entonces debo:
R’ – R = D
D = a + b
X’máx = Xmáx + (a ó b)
X’mín = Xmín – (a ó b)
Luego mis límites para la elaboración de la tabla serán X’máx y X’mín.
Ejemplo:
Se tiene la siguiente muestra

| | | | | | |
|---|---|---|---|---|---|
| 30 | 22 | 26 | 25 | 32 | 24 |
| 22 | 25 | 19 | 30 | 19 | 24 |
| 24 | 30 | 23 | 21 | 23 | 22 |
| 23 | 25 | 19 | 20 | 26 | |
| 23 | 20 | 20 | 29 | 23 | |

R = Xmax - Xmin = 32 - 19 = 13
k = 1+ 3,3 log (28) = 5,77... se redondea a 5.
c = 13/5 = 2,6... se redondea c'=3
R' = 5 * 3 = 15
(R'-R) = 15 - 13 = 2 D = 1 + 1
X'min = Xmin - 1 = 19 - 1 = 18
X'max = Xmax + 1 = 32 + 1 = 33

Luego mis límites para la elaboración de la tabla serán: X'min = 18 X'max = 33

Obsérvese un ejemplo con datos con cifras decimales.
La tienda CABRERA’S Y ASOCIADOS estaba interesada en efectuar un análisis de sus cuentas por comprar. Uno de los factores que más interesaba a la administración de la tienda era el de los saldos de las cuentas de crédito. Se escogió al azar una muestra aleatoria de 30 cuentas y se anotó el saldo de cada cuenta (en unidades monetarias) como sigue:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 77.97 | 13.02 | 17.97 | 89.19 | 12.18 | 8.15 | 34.40 | 43.13 | 79.61 | 90.99 |
| 43.66 | 29.75 | 7.42 | 93.91 | 20.64 | 21.10 | 17.64 | 81.59 | 60.94 | 43.97 |
| 32.67 | 43.66 | 51.69 | 53.40 | 68.13 | 11.10 | 12.98 | 38.74 | 70.15 | 25.68 |

N=30
X1 = valor mínimo = 7.42
Xn= valor máximo = 93.91
R = valor mayor – valor menor = Xn – X1 = 93.91 – 7.42 = 86.49
K=1+3.322(log N)
K = 1 + 3.322 (log 30)
= 1 + 3.322 (1.477)
= 1+ 4.9069
= 5.9069 ~6 aproximado al siguiente entero

$$C = \frac{R}{K} = \frac{86.49}{6} = 14.415$$

| ***Clases*** | ***P.M. Xi*** | ***fi*** | ***fr*** | ***fa↓*** | ***fa↑*** | ***fra↓*** | ***fra↑*** |
|---|---|---|---|---|---|---|---|
| 7.420 – 21.835 | 14.628 | 10 | 0.33 | 10 | 30 | 0.33 | 1.00 |
| 21.835 – 36.250 | 29.043 | 4 | 0.13 | 14 | 20 | 0.46 | 0.67 |
| 36.250 – 50.665 | 43.458 | 5 | 0.17 | 19 | 16 | 0.63 | 0.54 |
| 50.665 – 65.080 | 57.873 | 3 | 0.10 | 22 | 11 | 0.73 | 0.37 |
| 65.080 – 79.495 | 72.288 | 3 | 0.10 | 25 | 8 | 0.83 | 0.27 |
| 79.495 – 93.910 | 86.703 | 5 | 0.17 | 30 | 5 | 1.00 | 0.17 |
| Total | XXX | 30 | 1.00 | XXX | XXX | XXX | XXX |

6. MODELACIÓN.

A partir de los datos siguientes, construyo tabla de distribución de frecuencia (datos agrupados)

40 55 19 51 62 15 20 44 60 60 45 15 21 31 13
44 41 43 51 35 50 33 25 16 61 14 14 59 59 59
20 23 25 29 29 59 58 54 50 49 39 27 37 23 24
58 27 28 57 32 32 34 57 56 35 35 54 36 43 46
52 50 49 42 43 46 40 39 31 48

7. GLOSARIO

Población: es el conjunto de elementos, individuos o entes sujetos a estudio y de los cuales queremos obtener un resultado.
Variable: es la característica que estamos midiendo.
Variable cualitativa: Es aquella que expresa un atributo o característica, ejemplo: Rubio, moreno, etc.
Variable cuantitativa: Es aquella que podemos expresar numéricamente: edad, peso, etc. Esta a su vez la podemos subdividir en: Variable discreta, aquella que entre dos valores próximos puede tomar a lo sumo un número finito de valores. Ejemplos: el número de TSE de una familia, el de obreros de una fábrica, el de alumnos de la universidad, etc. Variable continua: la que puede tomar los infinitos valores de un intervalo. En muchas ocasiones la diferencia es más teórica que práctica, ya que los aparatos de medida dificultan que puedan existir todos los valores del intervalo. Ejemplos, peso, estatura, distancias, etc.
Muestra: Conjunto de elementos que forman parte de población. La muestra representa a esta población.
Tamaño muestral: Es el número de elementos u observaciones que tomamos. Se denota por n ó N.
Dato: Cada uno de los individuos, cosas, entes abstractos que integran una población o universo determinado. Dicho de otra forma, cada valor observado de la variable.
Frecuencia absoluta: Llamaremos así al número de repeticiones que presenta una observación. Se representa por fi.
Frecuencia relativa: Es la frecuencia absoluta dividida por el número total de datos, se suele expresar en tanto por uno. La suma de todas las frecuencias relativas, siempre debe ser igual a la unidad.
Frecuencia absoluta acumulada: es la suma de los distintos valores de la frecuencia absoluta tomando como referencia un individuo dado. La última frecuencia absoluta acumulada es igual al nº de datos.
Frecuencia relativa acumulada: es el resultado de dividir cada frecuencia absoluta acumulada por el número total de datos, se la suele representar con la notación: Hi

8. COMPROMISO

| | |
|---|---|
| a. | Consulto el enlace siguiente: https://app.emaze.com/@AWCZTRIO#1<br>- Resumo a través de un mapa conceptual la información de la presentación. |
| b. | Consulta y resume en tu cuaderno:<br>- Gráficas para datos agrupados y no agrupados.<br>- Parámetros estadísticos. |

BIBLIOGRAFÍA

JOHNSON Robert y KUBY Patricia. Estadística elemental, lo esencial. 2° edición. International Thomson Editores S.A.
CHAO Lincoln L. Estadística para las ciencias administrativas. Tercera edición. Editorial Mc Graw Hill
PORTUS Lincoyan. Curso práctico de estadística. Editorial Mc Graw Hill.