INSTITUCIÓN EDUCATIVA ABRAHAM REYES

Guía Trabajo

II Periodo Académico

GRADO _7°_ ASIGNATURA: _Estadística__

DOCENTES: _Diana Vileidy García Roldán

__Luz Diana David Segura

Entregar el 3 de junio al correo:

dianagarciar@ieabrahamreyes.edu.co (7-1 y 7-2)

luzdavid@ieabrahamreyes.edu.co (7-3)

TABLA DE DISTRIBUCIÓN DE FRECUENCIA

Una tabla de distribución de frecuencias es un arreglo de filas y columnas en el cual se registra, de manera ordenada y clasificada, la información de una base de datos.

La tabla de distribución de frecuencias está formada por los siguientes elementos:

- **Intervalos de Clase:** Se ubican en la primera columna de la tabla. En cada intervalo se incluye el grupo de datos que está entre el valor que inicia (límite inferior) y el valor en el que termina (límite superior). Tienen la particularidad de ser disjuntos, ordenados e incluir cada dato en uno y un solo intervalo.

- **Frecuencia del Intervalo o absoluta:** Número de individuos o datos que están clasificados en cada intervalo. Se representa con la letra *f*.

- **Frecuencia Relativa:** Relación entre la frecuencia y el total de la muestra o población; se representa con las letras *fr*. Esta relación se expresa mediante el cociente $\frac{f}{n}$ en donde n es el número de datos. Al multiplicar *fr* por 100 se obtiene el porcentaje de los datos que están en dicha clase.

- **Frecuencia Acumulada:** Es la sumatoria del número total de individuos que están en los intervalos anteriores y la frecuencia del intervalo. Se representa por *F*.

- **Frecuencia Relativa Acumulada:** Se representa por *Fr* y se plantea como la relación entre la frecuencia acumulada y el total de datos de la muestra, es decir, $Fr = \frac{F}{n}$.

- **Marca de Clase:** Es el punto medio de cada intervalo y se considera como el dato más representativo de dicho intervalo. Se representa como M_i.

Para elaborar una tabla de distribución de frecuencias se siguen los siguientes pasos:

PRIMERO: Se calcula el número de intervalos, para ello se utiliza la expresión $\#\ Intervalos = \sqrt{n}$, donde n es el total de datos.

SEGUNDO: Se calcula el Rango de la distribución, que es la diferencia entre el dato mayor (D_M) y el dato menor (D_m) de la base de datos:

Rango = $D_M - D_m$

TERCERO: Se calcula el tamaño de cada intervalo:

$$Tamaño = \frac{Rango}{\#\ de\ Intervalos}$$

$$Tamaño = \frac{D_M - D_m}{\sqrt{n}}$$

LUEGO, se construyen los intervalos. Para ello, se toma el dato menor como límite inferior del primer intervalo y a este valor se le suma el tamaño del intervalo para encontrar el límite superior.

Para el segundo intervalo, se toma como límite inferior el límite superior del primer intervalo más uno.

FINALMENTE, se ubican los datos en la tabla comenzando con la frecuencia y, luego, completando las columnas mencionadas al inicio de esta explicación.

EJEMPLO:

Para determinar los factores óptimos de crecimiento de un hongo comestible, un equipo de ingenieros de alimentos realizó un experimento que consistió en cultivar dicho hongo en 50 muestras diferentes y observar, después de 60 días, los cuerpos fructíferos que generó cada cultivo.

Los resultados se presentan a continuación:

123	116	167	198	165	148	169	110	121	100
145	132	145	126	176	189	163	101	120	109
135	127	178	187	180	166	134	129	118	102
167	185	183	177	156	145	167	143	132	121
145	128	119	117	140	121	164	129	132	140

Para elaborar la tabla de frecuencia seguimos los pasos mencionados anteriormente:

PRIMERO: Se calcula el número de intervalos, se debe tener en cuenta el total de datos que para este caso es 50, es decir n=50

$\#\ Intervalos = \sqrt{50} = 7{,}07$, y se aproxima a 7.

SEGUNDO: Se calcula el Rango, para esto se ubica el dato mayor y el dato menor:

$D_M = 198$ y $D_m = 100$

Rango = $D_M - D_m = 198 - 100 = 98$

TERCERO: Se calcula el tamaño de los intervalos:

$$Tamaño = \frac{Rango}{\#\ de\ Intervalos} = \frac{98}{7} = 14$$

CUARTO: Se construyen los 7 intervalos así:

- Primer Intervalo:
  Límite inferior = Dato menor del estudio = 100
  Límite Superior = Límite inferior + tamaño = 100 + 14 = 114
  [100 – 114]

- Segundo Intervalo:
  Límite inferior = Límite superior anterior + 1 = 114 + 1 = 115
  Límite Superior = Límite inferior + tamaño = 115 + 14 = 129
  [115 – 129]

- Tercer Intervalo:
  Límite inferior = Límite superior anterior + 1 = 129 + 1 = 130
  Límite Superior = Límite inferior + tamaño = 130 + 14 = 144
  [130 – 144]

- Cuarto Intervalo:
  Límite inferior = Límite superior anterior + 1 = 144 + 1 = 145
  Límite Superior = Límite inferior + tamaño = 145 + 14 = 159
  [145 – 159]

- Quinto Intervalo:
  Límite inferior = Límite superior anterior + 1 = 159 + 1 = 160
  Límite Superior = Límite inferior + tamaño = 160 + 14 = 174
  [160 – 174]

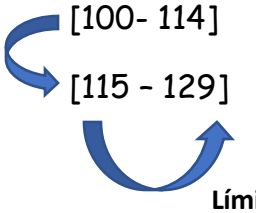
- Sexto Intervalo:
  Límite inferior = Límite superior anterior + 1 = 174 + 1 = 175
  Límite Superior = Límite inferior + tamaño = 175 + 14 = 189
  [175 – 189]

- Séptimo Intervalo:
  Límite inferior = Límite superior anterior + 1 = 189 + 1 = 190
  Límite Superior = Límite inferior + tamaño = 190 + 14 = 204
  [190 – 204]

Se debe tener en cuenta que como los intervalos deben ser disjuntos, el límite superior del primer intervalo debe ser diferente al límite inferior del segundo intervalo (en forme similar con todos), por esto se debe sumar uno para encontrar el límite inferior del intervalo siguiente:

Límite superior anterior + 1

= 114 + 1 = 115



Límite inferior + tamaño

= 115 + 14 = 129

Después, se realiza el conteo para determinar la frecuencia absoluta del intervalo, para esto se ubican la cantidad de datos según lo indique cada intervalo. Por ejemplo, para el primer intervalo que es [100 – 114] hay 5 datos del conjunto que se ubican entre esos valores: 110, 100, 101, 109, 102, por lo tanto, la frecuencia absoluta es 5; y se hace lo mismo con los demás intervalos.

CLASE	***f***
[100 – 114]	5
[115 – 129]	14
[130 – 144]	8
[145 – 159]	6
[160 – 174]	8
[175 – 189]	8
[190 – 204]	1
TOTAL	50

La columna siguiente corresponde a la frecuencia relativa, para hallarla se divide cada una de las frecuencias absolutas entre el total de datos que para este caso es 50.

CLASE	***f***	***fr***
[100 – 114]	5	5/50 = 0.1
[115 – 129]	14	14/50 = 0.28
[130 – 144]	8	8/50 = 0.16
[145 – 159]	6	6/50 = 0.12
[160 – 174]	8	8/50 = 0.16
[175 – 189]	8	8/50 = 0.16
[190 – 204]	1	1/50 = 0.02
TOTAL	50	50/50 = 1

La cuarta columna es para el porcentaje, el cual se determina multiplicando la frecuencia relativa por 100:

CLASE	f	fr	%
[100 – 114]	5	0.1	0.1x100 = 10 %
[115 – 129]	14	0.28	0.28 x 100 = 28 %
[130 – 144]	8	0.16	0.16 x 100 = 16 %
[145 – 159]	6	0.12	0.12 x 100 = 12 %
[160 – 174]	8	0.16	0.16 x 100 = 16 %
[175 – 189]	8	0.16	0.16 x 100 = 16 %
[190 – 204]	1	0.02	0.02 x 100 = 2 %
TOTAL	50	1	100%

La quinta columna corresponde a la frecuencia acumulada, la primera frecuencia acumulada es la primera frecuencia absoluta (5); la segunda es la suma de la primera y la segunda frecuencia (5 + 14 = 19) y así sucesivamente:

CLASE	f	fr	%	F
[100 – 114]	5	0.1	10 %	5
[115 – 129]	14	0.28	28 %	5+14 = 19
[130 – 144]	8	0.16	16 %	19 + 8 = 27
[145 – 159]	6	0.12	12 %	27 + 6 = 33
[160 – 174]	8	0.16	16 %	33 + 8 = 41
[175 – 189]	8	0.16	16 %	41 + 8 = 49
[190 – 204]	1	0.02	2 %	49 + 1 =50
TOTAL	50	1	100%	

La frecuencia relativa acumulada es la división de cada una de las frecuencias acumuladas entre el total de datos:

CLASE	f	fr	%	F	Fr
[100 – 114]	5	0.1	10 %	5	5/50
[115 – 129]	14	0.28	28 %	19	19/50
[130 – 144]	8	0.16	16 %	27	27/50
[145 – 159]	6	0.12	12 %	33	33/50
[160 – 174]	8	0.16	16 %	41	41/50
[175 – 189]	8	0.16	16 %	49	49/50
[190 – 204]	1	0.02	2 %	50	50/50
TOTAL	50	1	100%		

Por último, se calcula la marca de clase. La marca de clase se obtiene promediando los extremos de cada intervalo, es decir, se suman y se dividen entre dos.

Para la primera clase [100 – 114], la marca de clase será:

$$M_i = \frac{100+114}{2} = \frac{214}{2} = 107$$

CLASE	M_i
[100 – 114]	$\frac{100+114}{2} = \frac{214}{2} = 107$
[115 – 129]	$\frac{115+129}{2} = \frac{244}{2} = 122$
[130 – 144]	$\frac{130+144}{2} = \frac{274}{2} = 137$
[145 – 159]	$\frac{145+159}{2} = \frac{304}{2} = 152$
[160 – 174]	$\frac{160+174}{2} = \frac{334}{2} = 167$
[175 – 189]	$\frac{175+189}{2} = \frac{364}{2} = 182$
[190 – 204]	$\frac{190+204}{2} = \frac{394}{2} = 197$
TOTAL	

A continuación, se mostrará la tabla completamente terminada:

CLASE	*f*	*fr*	%	*F*	*Fr*	M_i
[100 – 114]	5	0.1	10 %	5	5/50	107
[115 – 129]	14	0.28	28 %	19	19/50	122
[130 – 144]	8	0.16	16 %	27	27/50	137
[145 – 159]	6	0.12	12 %	33	33/50	152
[160 – 174]	8	0.16	16 %	41	41/50	167
[175 – 189]	8	0.16	16 %	49	49/50	182
[190 – 204]	1	0.02	2 %	50	50/50	197
TOTAL	50	1	100%			

Con base en la tabla anterior se puede concluir, entre otras cosas, que:

- 14 de los hongos dieron entre 115 y 129 cuerpos fructíferos. Este valor fue el más alto registrado en el experimento y equivale a 28%.

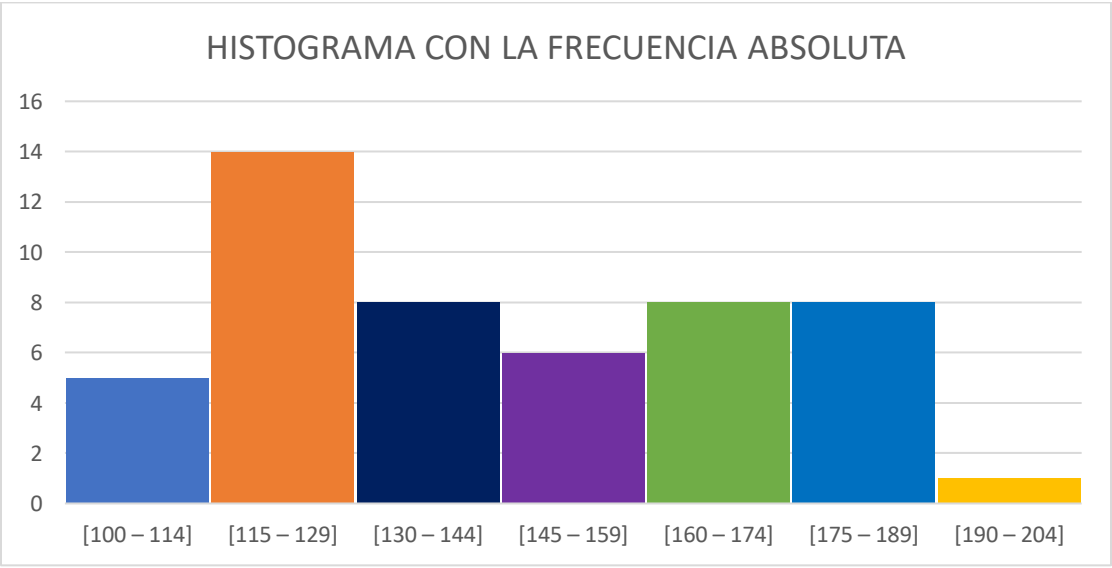
- 8 hongos estuvieron entre 130 y 144 cuerpos fructíferos, igualmente 8 hongos estuvieron entre 160 y 174, lo mismo pasa con los 8 hongos que dieron entre 175 y 189. En cada caso corresponden al 16%.

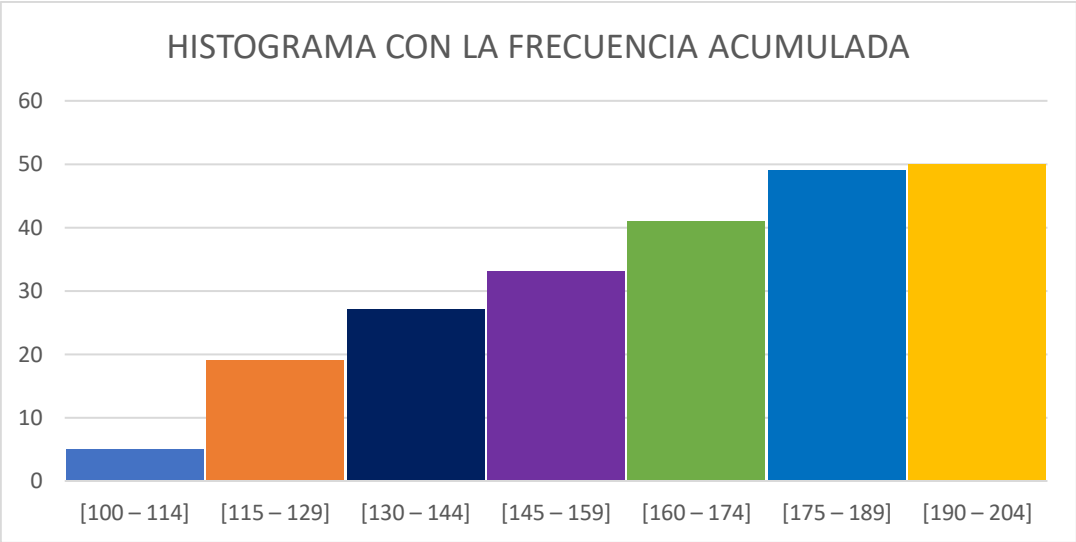
HISTOGRAMA

Un Histograma es una representación gráfica de los datos en una tabla de distribución de frecuencias de una variable cuantitativa. Es un diagrama de barras en el cual cada una de ellas representa la frecuencia.

El histograma se dibuja sobre un plano cartesiano en el cual el eje *X* representa las clases de la variable estudiada y el eje *Y* representa las frecuencias (absoluta o acumulada). La altura de la barra es el valor de la frecuencia.

Por ejemplo, el histograma correspondiente al ejemplo de los cuerpos fructíferos del hongo comestible es el siguiente:





MEDIDAS DE TENDENCIA CENTRAL

Es posible determinar las características de una muestra o población cuando las observaciones allí incluidas se presentan en forma no agrupada. Para ello, se usan, entre otras, las medidas de tendencia central. Las medidas de tendencia central son datos de resumen que indican, de acuerdo con algún criterio, un valor alrededor del cual se distribuyen las observaciones de un estudio estadístico. Las medidas de tendencia central son tres: la media, la mediana y la moda.

LA MEDIA:

La media aritmética o promedio es una de las medidas más utilizadas para la caracterización de una variable. Se ubica en el centro de las observaciones y se representa con $\bar{X}$.

$$\bar{X} = \frac{Suma\ de\ todos\ los\ datos}{Total\ de\ datos}$$

LA MEDIANA:

La mediana es el valor que divide el conjunto de observaciones en dos partes porcentualmente iguales. Se representa con $\tilde{X}$.

Para calcularla, primero se ordena el conjunto de datos de menor a mayor y luego, se ubica entre ellos el dato central.

- Si el número de datos es impar, es exactamente el dato de la mitad
- Si el número de datos es par, le mediana es el promedio de los dos datos centrales.

LA MODA

La moda en un conjunto de datos no agrupados es el dato con mayor frecuencia, es decir, el que más se repite. Se simboliza con $\hat{X}$.

Cuando en un conjunto de datos hay dos datos con una frecuencia alta, se dice que es bimodal.

En el caso de que el conjunto de daros tenga varios datos en los cuales la frecuencia alta se repite, se dice que es polimodal.

Si simplemente no hay ningún dato que se repita, se dice que la moda no existe.

EJEMPLO:

En la sección de urgencias de un hospital se está estudiando la mutación de un nuevo virus respiratorio que se viene presentando en la temporada invernal y afecta mayormente a niños entre 2 y 4 años.

Para ello, se está llevando el registro de la hora de llegada de los niños con los síntomas iniciales, medicarlos y darles manejo ambulatorio. Luego, si el paciente vuelve con los síntomas del virus, se registran las horas transcurridas desde la primera consulta hasta el momento de la nueva atención. A continuación, se muestran los datos de 15 niños en estas condiciones:

8	9	12
9	9	10
12	13	10
11	11	12
10	10	11

Determinar las medidas de tendencia central y escribir una conclusión sobre las mismas.

MEDIA:

$$\bar{X} = \frac{8 + 9 + 12 + 9 + 9 + 10 + 12 + 13 + 10 + 11 + 11 + 12 + 10 + 10 + 11}{15}$$

$$\bar{X} = \frac{157}{15} = 10.47$$

En promedio, el virus muta cada 10.47 horas. Esto significa que es muy probable que si el paciente adquiere el virus y no ha vuelto al servicio de urgencias en 10 horas es porque en dicha persona el virus ha desaparecido.

MEDIANA:

Se organizan los datos de menor a mayor y luego, se ubica el dato de la mitad

~~8~~ ~~9~~ ~~9~~ ~~9~~ ~~10~~ ~~10~~ ~~10~~ 10 ~~11~~ ~~11~~ ~~11~~ ~~12~~ ~~12~~ ~~12~~ ~~13~~

El dato central es 10, lo que indica que el 50% de los pacientes tardan 10 horas o menos para que el virus desaparezca.

MODA:

El dato que más se repite es 10, lo que indica que en la mayoría de los pacientes el virus se tarda 10 horas en desaparecer.

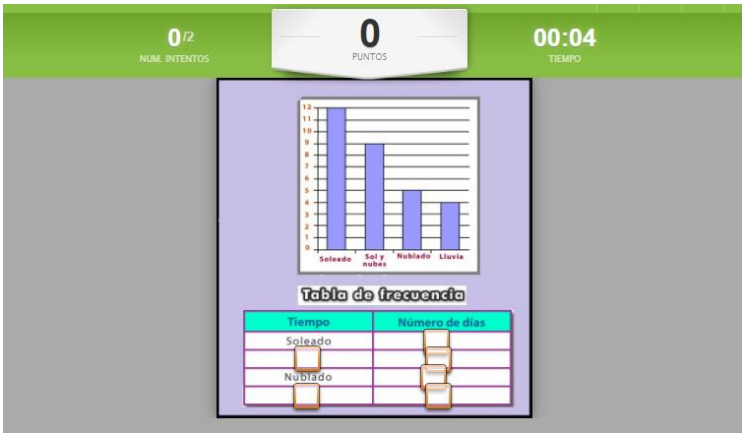
VIDEOS DE APOYO

https://www.youtube.com/watch?v=0DA7Wtz1ddg

https://www.youtube.com/watch?v=eY2xqiT_FF4

PRACTICA EN LOS SIGUIENTES SIMULADORES:

https://es.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/mean-median-basics/e/mean_median_and_mode



https://es.educaplay.com/recursos-educativos/4505999-actividad_3_tablas_de_frecuencia.html

TALLER

NOMBRE Y GRUPO DEL ESTUDIANTE:

_____7°____

NOTA: Cada ejercicio debe tener el proceso como sustentación

Para cada una de las siguientes situaciones realizar:

- Tabla de frecuencia con intervalos
- Histograma
- Calcular las medidas de tendencia central

1. (1.0) Antes de ofrecer la promoción, el dueño decide verificar qué tan conveniente es para su negocio hacer la siguiente oferta:

Para tomar la decisión, mira el siguiente registro del tiempo que se demoró en entregar los últimos domicilios:

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

2. (1.0) El profesor de deportes llevó al salón una báscula para determinar la masa de cada uno de los estudiantes. A continuación, se presentan los resultados en kilogramos:

56	61	53	62	56	46	49	52	61	62
58	59	58	57	51	49	52	53	61	59

3. (1.0) Maribel quiere conocer algunos aspectos de los hábitos de sueño de los estudiantes del grado séptimo. Para ello, les preguntó cuántas horas duermen al día. Los resultados se muestran a continuación:

7 horas	8 horas	10 horas	6 horas	8 horas
8 horas	8 horas	9 horas	8 horas	8 horas
7 horas	9 horas	8 horas	8 horas	8 horas

4. (1.0) Andrés tiene curiosidad de saber cuánto líquido bebe una persona adulta al día. Decidió preguntarles a todos sus profesores cuántos vasos de líquido beben al día. Los resultados fueron:

8	9	7	7	8	10	11	9	10	10	8

5. (1.0) El señor Salamanca tiene un criadero de peces y hoy registró la cantidad de peces adultos que hay en él. Los datos se muestran a continuación:

363	375	431	319	371
353	308	338	318	277

6. **AUTOEVALUACIÓN: _____**

TOMADO DE:

- Caminos del Saber 7°. Editorial Santillana
- Aulas sin Fronteras 7°. MEN