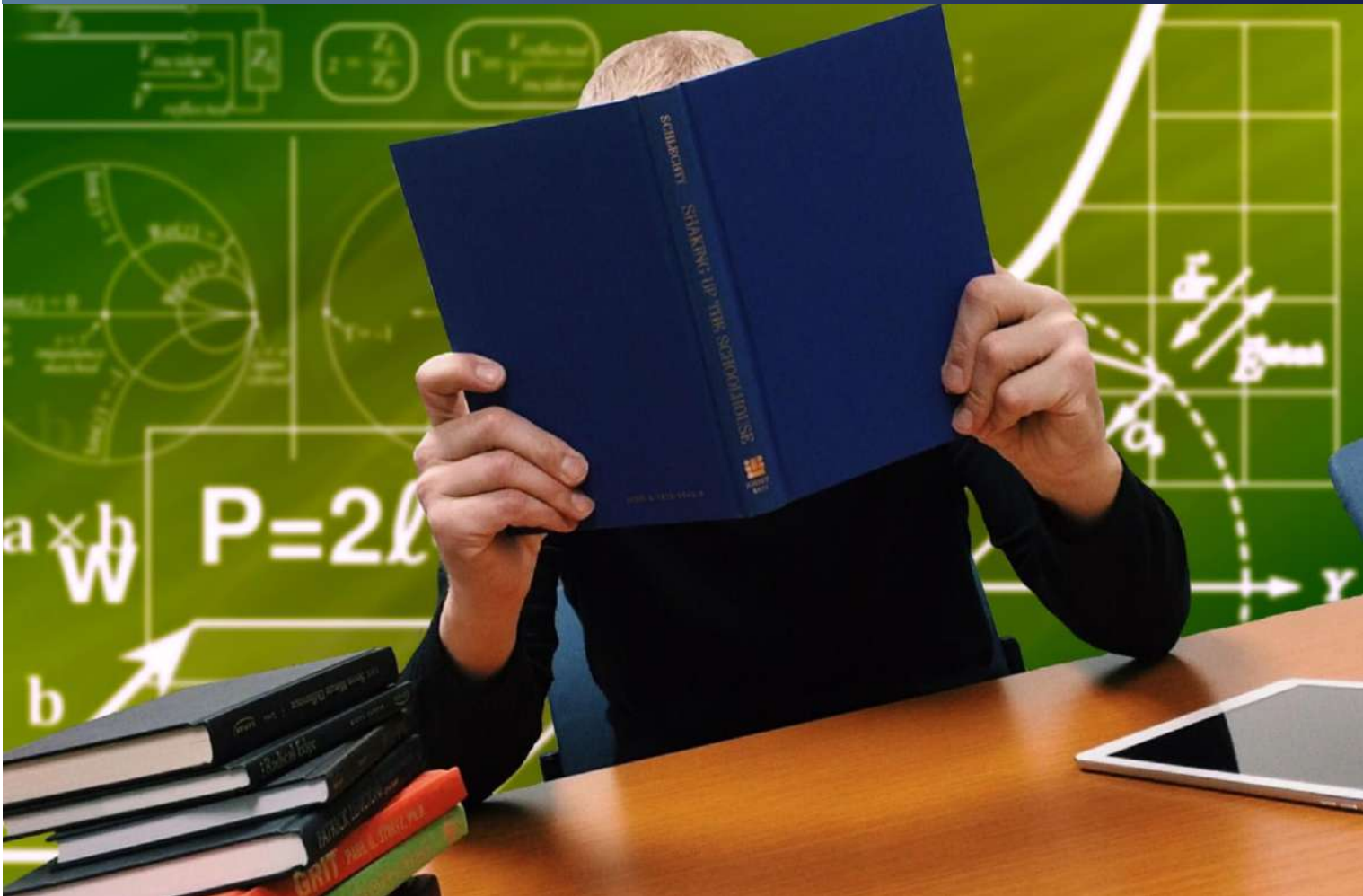


# *Ejercicios y Talleres*



puedes enviarlos a  
[klasesdematematicasymas@gmail.com](mailto:klasesdematematicasymas@gmail.com)

## BIOESTADÍSTICA ESTADÍSTICA I

### ACTIVIDADES:

1. ¿En qué difiere la elaboración de las gráficas de barras, los histogramas y los polígonos de frecuencia?

En el diagrama de barras los rectángulos van separados, en cambio en el histograma los rectángulos se unen en uno de sus extremos. También, el diagrama de barras se usa para representar de forma gráfica datos cuantitativos discretos o datos cualitativos, en cambio el histograma es exclusivo para representar datos cuantitativos continuos.

Los polígonos de frecuencia se diferencian porque se unen los puntos medios de lo que serían los rectángulos del histograma y se unen a través de líneas rectas. Por tanto el polígono de frecuencia muestra en los vértices del polígono la frecuencia absoluta para cada intervalo.

2. Defina cada uno de los términos dados a continuación e ilustre cada uno de ellos por medio de un ejemplo: Media aritmética, mediana, moda, rango, varianza desviación estándar.

**Media aritmética:** Es la suma de todos los valores de la variable dividida entre el número total de elementos.

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_{n-1} + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

**La mediana** o valor mediano será el valor de la variable que separa en dos grupos los valores de las variables, ordenadas de menor a mayor. Por tanto es una cantidad que nos indica orden dentro de la ordenación.

El lugar que ocupa se determina dividiendo el nº de valores entre 2:  $\frac{n}{2}$

Cuando hay un número impar de valores de la variable, la mediana será justo el valor de orden central, aquel cuya frecuencia absoluta acumulada coincida

con  $\frac{n}{2}$ . Es decir:  $N_{i-1} < \frac{n}{2} \leq N_i \Rightarrow Me = x_i$ . Por tanto la mediana coincide con un valor de la variable.

**Moda: Mo.** Será el valor de la variable que más veces se repite, es decir, el valor que tenga mayor frecuencia absoluta. Pueden existir distribuciones con más de una moda: bimodales, trimodales, etc.

**El rango** se suele definir como la diferencia entre los dos valores extremos que toma la variable. Es la medida de dispersión más sencilla y también, por tanto, la que proporciona menos información

**La varianza** se define como la media aritmética de las desviaciones al cuadrado entre cada valor de la variable y la media aritmética. Corresponde

pues, a la media cuadrática de todas las desviaciones de cada valor de la variable con respecto a su media aritmética.

$$S_x^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

**La desviación típica o estándar** se define como la raíz cuadrada de la varianza.

$$S_x = \sqrt{S_x^2}$$

**3.** Un administrador registra el número de entradas de compradores en diez almacenes: 15, 28, 33, 19, 24, 32, 17, 21, 34, 12 calcule: Media, mediana, moda, rango, varianza, desviación estándar.

$$\text{Media: } \bar{X} = \frac{\sum_{i=1}^n x_i}{n} = \frac{15 + 28 + 33 + 19 + 24 + 32 + 17 + 21 + 34 + 12}{10} = \frac{235}{10} = 23.5$$

Para la mediana: 12, 15, 17, 19, 21, 24, 28, 32, 33, 34

$$Me = \frac{21 + 24}{2} = \frac{45}{2} = 22.5$$

Moda: No tiene moda

$$\text{Rango: } D_{\max} - D_{\min} = 34 - 12 = 22$$

Varianza:

$$S_x^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{(12 - 22.5)^2 + (15 - 22.5)^2 + (17 - 22.5)^2 + \dots + (33 - 22.5)^2 + (34 - 22.5)^2}{10}$$

$$S_x^2 = \frac{566.5}{10} = 56.65$$

$$\text{Desviación Estándar: } S_x = \sqrt{S_x^2} = \sqrt{56.65} = 7.52$$

**4.** Una universidad tiene su propio criadero de ratas para experimentos de investigación en el laboratorio. Un muestreo reciente de 50 ratas tomadas del criadero reveló los siguientes pesos para dichas ratas:

320	282	341	324	340	302	336	265	313	317
310	335	318	296	309	308	310	277	288	353
314	298	315	360	275	315	297	330	296	274
250	274	318	287	284	267	292	348	302	297
270	263	269	292	298	343	284	352	345	225

a. Haga una distribución de frecuencias de datos agrupados recuerde que el número de intervalos se puede calcular utilizando la regla de Stungers (número de intervalos =  $1 + 3.3 \log n$ ).

Dato máximo = 360

Dato mínimo = 225

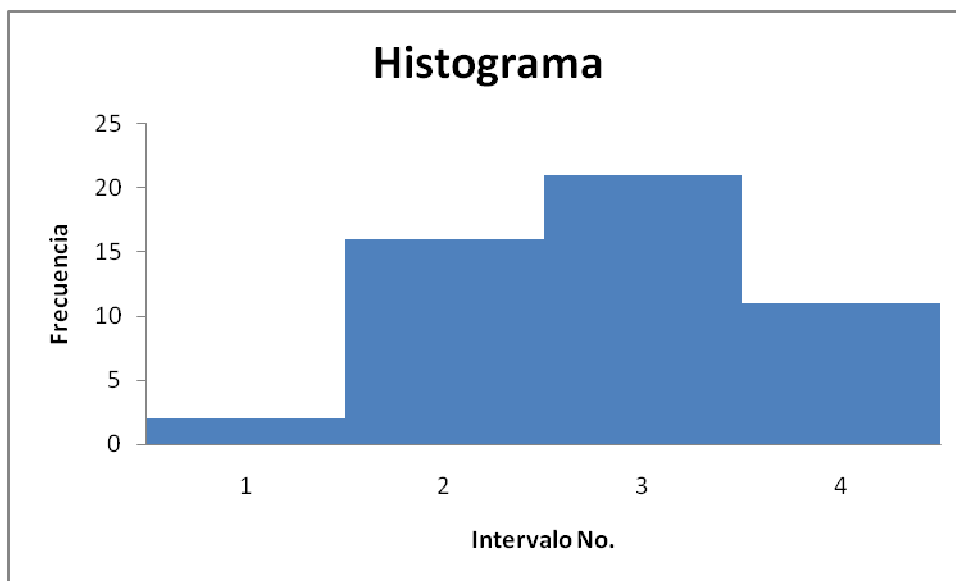
$$\text{Número Intervalos} = m = 1 + 3.3 \cdot \log(50) = 3.25, m = 4$$

$$\text{Rango: } D_{\max} - D_{\min} = 135$$

$$\text{Ancho de clase } c = \frac{\text{Rango}}{m} = \frac{135}{4} = 33.75$$

No	$L_{inf}$	$L_{sup}$	$\hat{y}_i$	$f_i$
1	225	258,75	241,875	2
2	258,75	292,5	275,625	16
3	292,5	326,25	309,375	21
4	326,25	360	343,125	11
<b>Total</b>				<b>50</b>

b. Elabore un histograma de la distribución de frecuencia



c. La distribución es simétrica o asimétrica?

La distribución es asimétrica, se aprecia que la mayoría de los datos están acumulados hacia la derecha de la gráfica.

d. Convierta la distribución de frecuencias en una distribución de frecuencias relativas, de frecuencias acumuladas y de porcentajes acumulados.

No	$L_{inf}$	$L_{sup}$	$\hat{y}_i$	$f_i$	$F_i$	$h_i$	$H_i$
1	225	258,75	241,875	2	2	0,04	0,04
2	258,75	292,5	275,625	16	18	0,32	0,36
3	292,5	326,25	309,375	21	39	0,42	0,78
4	326,25	360	343,125	11	50	0,22	1
<b>Total</b>				<b>50</b>			

$F_i$  = Frecuencia absoluta acumulada

$h_i$  = Frecuencia relativa

$H_i$  = Frecuencia relativa acumulada.

e. A partir de la distribución de frecuencias acumuladas determine  $P_{50}$  y  $P_{75}$

$$P_{50} = L_{\text{inf}} + \left[ \frac{\frac{n}{2} - F_{i-1}}{f_i} \right] \cdot c = 292.5 + \left[ \frac{\frac{50}{2} - 18}{21} \right] \cdot 33.75 = 303.75$$

$$P_{75} = L_{\text{inf}} + \left[ \frac{\frac{3n}{4} - F_{i-1}}{f_i} \right] \cdot c = 292.5 + \left[ \frac{\frac{3 \cdot 50}{4} - 18}{21} \right] \cdot 33.75 = 323.84$$

f. Con la distribución de frecuencias acumuladas, determine el rango percentil de un dato de 275 y el rango percentil de un dato de 318.

$$P_x = L_{\text{inf}} + \left[ \frac{\frac{xn}{100} - F_{i-1}}{f_i} \right] \cdot c = 258.75 + \left[ \frac{\frac{x \cdot 50}{100} - 2}{16} \right] \cdot 33.75 = 275$$

$$\frac{\frac{x \cdot 50}{100} - 2}{16} = \frac{275 - 258.75}{33.75}$$

$$\frac{x \cdot 50}{100} - 2 = \frac{275 - 258.75}{33.75} \cdot 16$$

$$\frac{x \cdot 50}{100} = 7.70 + 2$$

$$x = 15$$

Es el percentil 15.

Para el dato 318

$$P_x = L_{\text{inf}} + \left[ \frac{\frac{xn}{100} - F_{i-1}}{f_i} \right] \cdot c = 292.5 + \left[ \frac{\frac{x \cdot 50}{100} - 18}{21} \right] \cdot 33.75 = 318$$

$$\frac{\frac{x \cdot 50}{100} - 18}{21} = \frac{318 - 292.5}{33.75}$$

$$\frac{x \cdot 50}{100} - 18 = \frac{318 - 292.5}{33.75} \cdot 21$$

$$\frac{x \cdot 50}{100} = 15.86 + 18$$

$$x = 67.7$$

Corresponde al percentil 68

5. Del ejercicio del punto 4 calcule la media, la mediana, la moda, la varianza y la desviación estándar, Que conclusión puede sacar con base en estos datos calculados?

No	$L_{\text{inf}}$	$L_{\text{sup}}$	$\hat{y}_i$	$f_i$	$\hat{y}_i \cdot f_i$	$(\hat{y}_i - \bar{x})^2 \cdot f_i$
1	225	258,75	241,875	2	483,75	7546,06
2	258,75	292,5	275,625	16	4410	12254,49

3	292,5	326,25	309,375	21	6496,875	775,01
4	326,25	360	343,125	11	3774,375	17446,33
<b>Total</b>				<b>50</b>	<b>15165</b>	<b>38021,90</b>

$$\text{Media: } \bar{X} = \frac{\sum_{i=1}^n \hat{y}_i \cdot f_i}{n} = \frac{15165}{50} = 303.3$$

Para la mediana: Me= 303.75. Se calculo anteriormente

Moda: 309.12. Es la marca de clase del intervalo con mayor frecuencia

$$\text{Varianza: } S_x^2 = \frac{\sum (\hat{y}_i - \bar{x})^2 \cdot f_i}{n} = \frac{38021.90}{50} = 760.43$$

$$\text{Desviación Estándar: } S_x = \sqrt{S_x^2} = \sqrt{760.43} = 25.57$$

Los datos tienen una desviación relativamente pequeña respecto a la media. Esto da cuenta de la poca dispersión de los datos, haciendo de la media (303.3) muy confiable. La Media y mediana prácticamente coinciden en sus valores. Hace suponer que la distribución es simétrica.

**6.** Elabore una encuesta sobre un tema de interés relacionado con la carrera que usted está cursando y aplicable en la zona donde se encuentre, al elaborar la encuesta elabore preguntas que utilice características cuantitativas y/o cualitativas donde pueda aplicar los procesos de recopilación, registro, presentación, graficación y entrega de información de datos para una investigación. (trabaje por lo menos una pregunta donde pueda utilizar una tabla de frecuencias con variables continuas y por lo menos aplique 30 encuestas ).

a. Coloque el objetivo de la encuesta.

Determinar el nivel de colesterol de los habitantes de Chía a través de una muestra de 70 personas.

#### Encuesta

1. Sexo: M \_\_\_\_\_ F \_\_\_\_\_
2. Edad: \_\_\_\_\_
3. Bebedor habitual: SI \_\_\_\_\_ NO \_\_\_\_\_
4. Talla \_\_\_\_\_ cm
5. Peso \_\_\_\_\_ Kg
6. Nivel de Colesterol \_\_\_\_\_

Variables Cualitativas: Sexo, Bebedor habitual

Variables Cuantitativas discretas: Edad

Variables cuantitativas continuas: Talla, Peso, Nivel de Colesterol

b. Cuando elabore la encuesta tenga en cuenta las etapas de investigación estadística como son: planeación, recolección y procesamiento. Explíquelas?

En la planeación se determinan los objetivos a seguir, identificando las variables a medir. También se define el número de encuestas y la forma de aplicarlas.

En el proceso de recolección se aplican las encuestas, capacitando a los encuestadores para el desarrollo de la misma

En el procesamiento se realiza el análisis descriptivo de cada una de las variables que se están estudiando.

#### Resultados de la encuesta

TABLA DE LOS DATOS CORRESPONDIENTES DE LA MUESTRA ALEATORIA EN UNA POBLACION MAYOR DE 18 AÑOS						
	EDAD	SEXO	B/H	TALLA	PESO	COLEST
1	78	2	2	142	67	188
2	37	2	2	156	97	212
3	43	2	2	158	60	202
4	46	2	2	158	81	275
5	20	2	2	155	55	358
6	36	2	2	164	67	189
7	42	2	2	159	64	191
8	41	2	2	156	77	223
9	90	2	2	144	54	302
10	39	2	2	159	68	184
11	39	2	2	149	53	183
12	43	2	2	161	70	280
13	35	2	2	173	77	207
14	36	2	2	159	60	219
15	61	2	2	155	80	196
16	75	2	2	150	43	196
17	24	2	2	165	66	182
18	53	2	2	169	66	218
19	37	2	2	160	65	193
20	57	2	2	152	85	260
21	24	2	2	174	64	196
22	22	2	2	159	54	154
23	28	2	2	161	57	239
24	28	2	2	164	75	283
25	21	2	1	164	64	193
26	36	2	1	160	55	217
27	34	2	1	151	57	163
28	35	2	1	158	66	242
29	26	2	1	167	47	179
30	21	2	1	157	57	232
31	71	1	2	154	61	233
32	35	1	2	173	97	211

33	29	1	2	169	81	229
34	50	1	2	167	67	221
35	39	1	2	174	88	214
36	39	1	2	167	82	273
37	47	1	2	164	70	264
38	39	1	2	180	81	238
39	69	1	2	168	71	235
40	41	1	2	166	82	255
41	34	1	1	164	74	198
42	20	1	1	174	51	123
43	63	1	1	173	87	177
44	39	1	1	168	84	262
45	33	1	1	169	59	266
46	21	1	1	175	85	286
47	29	1	1	167	81	195
48	30	1	1	168	59	266
49	33	1	1	165	70	246
50	59	1	1	158	77	247
51	65	1	1	177	63	318
52	47	1	1	164	69	338
53	29	1	1	167	78	371
54	39	1	1	162	81	213
55	75	1	1	156	59	199
56	55	1	1	161	72	239
57	35	1	1	174	72	247
58	46	1	1	173	73	205
59	41	1	1	166	64	361
60	45	1	1	170	83	256

Tabla de distribución de frecuencias para el nivel de colesterol

No	Linf	Lsup	Marca	fi	Fi	hi	Hi
1	123	158,43	140,71	2	2	0,033	0,033
2	158,43	193,86	176,14	11	13	0,183	0,217
3	193,86	229,29	211,57	19	32	0,317	0,533
4	229,29	264,71	247	15	47	0,25	0,783
5	264,71	300,14	282,43	7	54	0,117	0,9
6	300,14	335,57	317,86	2	56	0,033	0,933
7	335,57	371	353,29	4	60	0,067	1
			<b>Total</b>	<b>60</b>			



