

Programa de Bioestadística

Unidad I: Introducción a la estadística. Definición de estadística.

Análisis estadístico. Datos Variables. Organización de datos cuantitativos Tablas de intervalos. Gráficos.

Unidad II: Distribución de frecuencia. Organización de datos cualitativos. Escala de medición. Tabla de distribución de frecuencia. Serie simple, serie de intervalos. Gráficos.

Unidad III: Indicadores. Cifras absolutas. Cifras relativas. Elementos y clasificación. Tasas, tipos de tasas. Propiedades. Razones.

Unidad IV: Medidas de posición. Introducción. Media aritmética. Propiedades. Elección de una medida de posición adecuada. Modo. Mediana. Cuartiles. Deciles. Percentiles.

Unidad V: Medidas de posición. Introducción. Rango. Desviación media. Varianza. Propiedades. Desviación estándar. Coeficiente de variación. Utilidad de las medidas de dispersión. Control de calidad. Aplicaciones prácticas.

Unidad VI: Análisis de Regresión y Correlación. Introducción. Residuales. Coeficientes de determinación. Análisis de correlación lineal. Interpretación de la correlación y regresión.

Unidad VII: Concepto de Probabilidad. Definición. Ejemplos. Probabilidad Condicional, Sucesos Independientes y Dependientes. Sucesos mutuamente excluyentes.

Unidad I

Introducción a la Bioestadística

“La Estadística aplicada a la Biología es la Bioestadística”

El propósito de un estudio estadístico es extraer conclusiones de la naturaleza de una población. Al ser la población grande y no poder ser estudiada en su integridad en la mayoría de los casos las conclusiones obtenidas deben basarse en el examen de solamente una parte de estas, lo que nos lleva a la justificación, necesidad y definición de las diferentes técnicas de muestreo.

La primera técnica de la estadística consistió esencialmente en la organización de datos en su presentación gráfica en el cálculo de cantidad representativa del conjunto, con el objeto de que los aspectos sobresalientes del mismo fueron rápidos y fácilmente aprehensibles. Esta parte de la metería es la que en terminología moderna se conoce como estadística descriptiva. El papel de la estadística es cuantificarla incertidumbre que es indispensable para las conclusiones finales.

Entonces organizar y procesar datos es lo que se llama estadística descriptiva y comprende la toma de decisiones a partir de un gran volumen de datos encaminando una parte de ellos se llama estadística inferencial o inferencia estadística.

Variable: una característica se clasifica como variables si, tal y como se observa se encuentran que toma diferentes valores en diferentes personas, lugares o cosas.

Variable cualitativa: son aquellas que no pueden ser medidas por un número sino por unas características.

Variable cuantitativa: son aquellas que pueden medirse en forma usual. Las mediciones realizadas en variables cuantitativas llevan información con respecto a la cantidad. Estos pueden ser:

Variables cuantitativas discretas: son los que provienen de conteo (números enteros).

Variables cuantitativas continuas: son las que provienen de mediciones (números reales, admite decimales).

Variable independiente: la variable independiente de un experimento es aquella que es controlada en forma sistemática por el operador o investigador.

En la mayor parte de los experimentos, el científico está interesado en determinar el efecto que tiene una variable sobre alguna o más variables.

Variable dependiente: la variable dependiente en un experimento es la medida dada por un investigador para determinar el efecto de la variable independiente.

Ejemplo: se entrega una cantidad cierta de alcohol (variable independiente).

Efectos que produce (Variable dependiente).

Escala de medición: el tipo de escala de medición utilizado para reunir los datos ayuda a determinar el tipo de prueba de inferencia estadística a emplearse.

Escalas nominales: representa el nivel mínimo de medición y se utiliza con frecuencia para variables de naturaleza cualitativa y no cuantitativa. Tiene la propiedad de equivalencia. No tiene los atributos matemáticos de magnitud, intervalos iguales o cero absolutos.

Ejemplo: sexo, tipo de religión, medicina que le fue suministrada, etc.

Escalas ordinales: representa el siguiente nivel de medida, el cual es relativamente bajo de acuerdo con las propiedades de magnitudes. Con esta escala ordenamos los objetos medidos según posea más, menos o la misma cantidad de la variable medida.

Ejemplo: nivel de introducción de una persona (1°, 2°, 3°, etc.)

Escalas de intercales: representa un nivel superior de medición con respecto a la escala ordinal. Por las propiedades de magnitud e igualdad de intervalos entre las unidades adyacentes, pero no tiene el cero absoluto.

Escala de Razón o Proporción: tiene todas las propiedades de una escala de intervalos y, además, posee un cero absoluto.

Ejemplo: escala de temperatura Kelvin.

Actividad

Redondee cada uno de los siguientes números la aproximación indicada.

- a) 3256 aproximadamente a las centenas
- b) 5,781 aproximadamente a las décimas
- c) 0,0045 aproximadamente a las milésimas
- d) 46,7385 aproximadamente a las centésimas
- e) 125,9995 aproximadamente a 2 cifras decimales
- f) 3502378 aproximadamente a millones
- g) 148,45 aproximadamente a unidades
- h) 0,000098501 aproximadamente a millonésimas
- i) 2184,73 aproximadamente a decenas
- j) 43,87500 aproximadamente a las centésimas

Recordatorio:

3,75 impar, pasa a 3,80

3,65 par, queda como está

Altura (pulgadas)	N° de estadísticas de estudiantes
60 - 62	5
63 - 65	18
66 - 68	42
69 - 71	27
72 - 74	8
	Total 100

$$x_i = \frac{L_i + L_s}{2}$$

L_i = Límite inferior x_i = Marca de clase

L_s = Límite superior $f(x_i)$
= Frecuencia absoluta

$F(x_i)$ = Frecuencia absoluta acumulada $h(x_i)$
= Frecuencia relativa

$H(x_i)$ = Frecuencia relativa acumulada

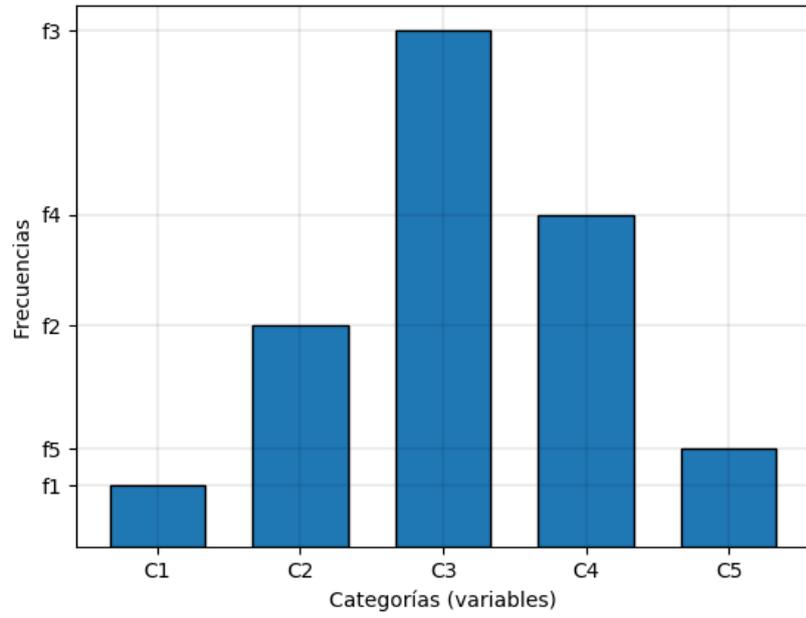
Ejemplo:

Altura (pulgadas)	x_i	$f(x_i)$	$F(x_i)$	$h(x_i)$	$H(x_i)$	%	%Ac
60 - 62	61	5	5	0.05	0.05	5	5
63 - 65	64	18	23	0.18	0.23	18	23
66 - 68	67	42	65	0.42	0.65	42	65
69 - 71	70	27	92	0.27	0.92	27	92
72 - 74	73	8	100	0.08	1.00	8	100
Total		100					

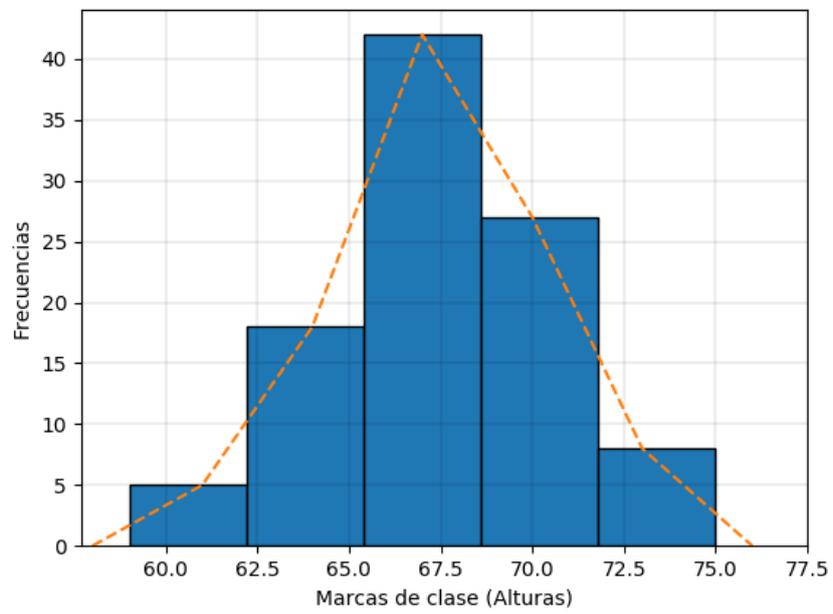
Gráficas

Diagrama de Barras (Cualitativas)

Frecuencias



Histograma



Preguntas:

¿Cuál es el límite de la quinta clase?

¿Cuál es el límite

de la octava clase?

¿Cuál es la marca de clase de la séptima clase?

¿Cuáles son los límites reales de la última clase?

¿Cuál es el tamaño de intervalo de clase?

¿Cuál es la frecuencia de la cuarta clase?

¿Cuál es el porcentaje de tubos que no sobrepasa las 600 horas?

¿Cuál es el porcentaje de tubos cuya duración es mayor o igual a 900 horas?

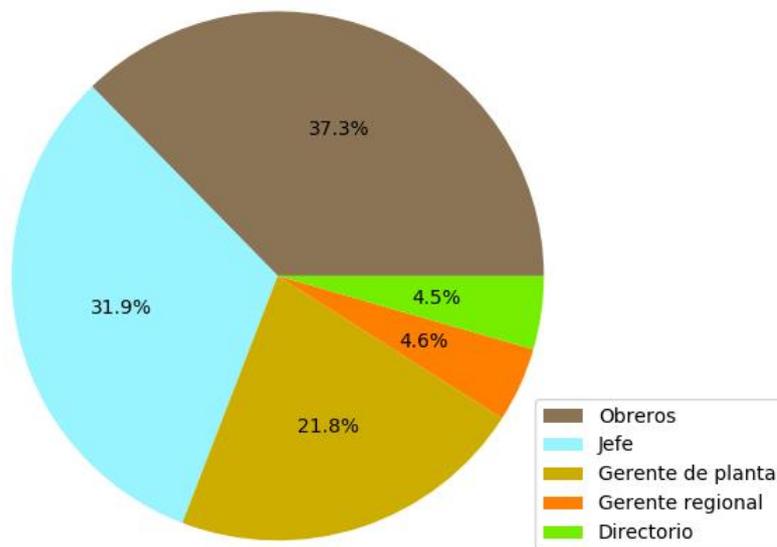
¿Cuál es el porcentaje de tubos que dura al menos 500 horas, pero menos de 1000 horas?

¿Cuál es la frecuencia relativa de la sexta clase?

Ejemplo:

Cargo	Cantidad	$F(x_i)$	$h(x_i)$	$H(x_i)$	%	%Ac
Obrero	474	474	0,37	0,37	37,29	37,29
Jefe	405	879	0,32	0,69	31,86	69,16
Gerente de planta	277	1156	0,22	0,91	21,79	90,95
Gerente Regional	58	1214	0,05	0,96	4,56	95,52
Directorio	57	1271	0,04	1,00	4,48	100,00
Total	1271		1,00		100,00	

Gráfico de Cargos



Unidad II

Distribución de frecuencia

Reglas Empíricas para la construcción de intervalos

1. Las clases han de ser excluyentes.
2. Obtener el rango o amplitud de los datos. Es la diferencia entre el valor máximo y el valor mínimo.

$$Ra = X_{max} - X_{min}$$

3. Determinar el número de intervalo (m) y la amplitud de los intervalos de clase (c).

$$m \cong 1 + 3,3 \cdot \log N \qquad c > \frac{Ra}{m}$$

$$m \in \mathbb{N}$$

4. Aunque no tiene que ser necesariamente así, es conveniente que la amplitud de los intervalos sea constante.
5. Determine el límite superior del último intervalo y el límite inferior del primero.
6. Determine la marca de clase de cada intervalo, que no es otra cosa que el punto medio del intervalo.
7. Todos los datos de una clase quedan representados por la marca de clase. De esta manera todos los cálculos se realizan como si en lugar de tener n valores distintos de una clase, tuviéramos n veces la marca de clase.
8. Promediando el límite superior de una clase con el inferior de la siguiente se obtiene la frontera de clase o verdadero límite de la clase.

La siguiente tabla representa el pH de una sustancia en diferentes concentraciones de una disolución.

4,47	4,47	3,48	5,00	3,42	3,78
3,10	3,57	4,20	4,50	3,60	3,75
4,50	2,85	3,70	4,20	3,20	4,05
4,90	5,10	5,30	4,16	4,56	3,84
3,50	5,20	4,71	3,70	4,78	4,14
4,14	4,80	4,10	3,83	3,60	2,98
4,32	5,10	4,30	3,90	3,96	3,54
4,80	4,30	3,39	4,47	3,19	
3,10	4,70	3,69	3,30	2,85	
4,68	4,04	4,44	5,43	3,04	

Diagrama de tallos y de hojas

2	85 – 85 – 98
3	4 – 10 – 41 – 49 – 20 – 30 – 30 – 39 – 42 – 48 – 50 – 54 – 54 – 57 – 60 – 60 – 69 – 70 – 70 – 756 – 78 – 83 – 90 – 96
4	4 – 5 – 10 – 14 – 14 – 16 – 20 – 20 – 30 – 30 – 32 – 44 – 47 – 47 – 47 –50 – 50 – 56 – 68 – 70 – 71 – 78 – 80 – 80 – 90
5	0 – 10 – 10 – 20 – 30 – 43

$$Ra = X_{max} - X_{min} \qquad m \cong 1 + 3,3 \cdot \log N \qquad C > \frac{Ra}{m}$$

$$Ra = 5,13 - 2,85 \qquad m \cong 1 + 3,3 \cdot \log 57 \qquad C > \frac{2,58}{7}$$

$$Ra = 2,58 \quad m \cong 7 \quad C \cong 0,37$$

pH de una sustancia	x_i	$f(x_i)$	$F(x_i)$	$h(x_i)$	$H(x_i)$	%	%Ac
2,85 – 3,22	3,04	8	8	0,14	0,14	14,04	14,04
3,22 – 3,54	3,40	8	16	0,14	0,28	14,04	28,07
3,59 – 3,96	3,78	9	25	0,16	0,44	15,79	43,86
3,96 – 4,33	4,14	12	37	0,21	0,65	21,05	64,91
4,33 – 4,7	4,52	8	45	0,14	0,79	14,04	78,95
4,7 – 5,07	4,88	7	52	0,12	0,91	12,28	91,23
5,07 – 5,44	5,26	5	57	0,09	1,00	8,77	100,00
				1,00		100	

Unidad III

Cifras Absolutas

Las frecuencias absolutas referente a los hechos que interesan en el área de la salud podrán ser: nacimiento, enfermedades, etc. Sus distribuciones según ciertas características tienen numerosas aplicaciones en la planificación, organización y ejecución de servicios de salud de la comunidad, en ciertas cosas constituye el resultado más adecuado y útil para tales fines.

El uso de las frecuencias absolutas tiene sus limitaciones ya que no siempre facilitan la mejor apreciación de un fenómeno, sobre todo cuando sea resumir y comparar datos especialmente si las cifras absolutas a comparar son de gran magnitud es más apropiado el uso de cifras relativas.

Cifras relativas

En forma general puede decirse que un número relativo es una cantidad que está referida a otra que se usa como base de comparación.

Clasificación:

Tasas: por razones de comodidad es preferible ampliar el valor de la frecuencia para convertirlas en números enteros, y en vez de calcularla para un instante hacerlo para un periodo determinado.

Matemáticamente expresan la relación 1 a 1 entre dividendos y divisor, necesariamente deben mantener estrecha vinculación causal entre sí, por la naturaleza de sus contenidos.

En bioestadística es el cociente que resulta al relacionar las frecuencias absolutas de los atributos (numerador) ocurridos en un determinado lugar y tiempo. Con el total o parte de la población donde sucedieron los hechos (denominador) en el mismo lugar y tiempo.

Factor de ampliación: este siempre hace referencia al denominador, el valor real de las tasas siempre es muy inferior a la unidad por lo que se prefiere convencionalmente referirla a una base de 1000, 10000, 100000 habitantes.

Tasas Brutas o Crudas: también se las sabe llamar tasa general.

Tasas Crudas

$$\frac{\text{Número de hechos Ocurridos en la población}}{\text{Población Total}} \times \text{Factor de Ampliación}$$

Tasa Bruta de Natalidad

$$\frac{\text{Número de nacidos vivos}}{\text{Población estimada a mitad de año}} \times 1000$$

Tasa Bruta de Mortalidad

$$\frac{\text{Número de defunciones}}{\text{Población estimada a mitad de año}} \times 1000$$

Tasa de Morbilidad

$$\frac{\text{Número de enfermos de una afección}}{\text{Población estimada a mitad de año}} \times 100000$$

Tasas Específicas: son tasas específicas las de la mortalidad producida por una determinada enfermedad o bien la tasa de mortalidad de un determinado sexo.

$$\frac{\text{Número de hechos ocurridos en un Grupo específico de la población}}{\text{Población de ese grupo específico}} \times \text{Factor de Ampliación}$$

Tasa de Letalidad:

$$\frac{\text{Número de defunciones por una enfermedad}}{\text{Número de casos de esa enfermedad}} \times 100$$

Tasa de Prevalencia:

$$\frac{\text{Número de enfermos de una afección}}{\text{Estimación de la población}} \times 100; 1000; 10000; 100000$$

Ejemplo: En el pueblo de San Cristóbal de 13748 habitantes, hay una epidemia al parecer desencadenada por la abrupta proliferación de ratas, que enfermaron a 137 lactantes de 0 a 2 años, 2390 niños de 5 años, 1988 niños de 3 a 4 años. ¿Cuál es la probabilidad de enfermar de los menores de 5 años?

$$\frac{2390 + 137}{137 + 2390 + 1988} = \frac{2527}{4515} = 0,56 \times 100 = 56$$

Proporciones

Frecuencias relativas

Las frecuencias relativas que establecen una relación entre 2 cifras se llaman proporciones y se obtienen dividiendo cada partida por el total y por lo general se lo multiplica por 100.

Enfermedades	Números
Cáncer	783 – 783/3792 = 0,20x100 = 20
Cardiovascular	625 – 625/3792 = 0,16x100 = 16
Lesiones basculares	331 – 331/3792 = 0,08x100 = 8
Otras enfermedades	2053 – 2053/3792 = 0,54x100 = 54
TOTAL	3792 100

Razones: matemáticamente es la relación establecida entre 2 guarismos (cifras) donde una actúa como numerador, y otro como denominador. Relaciona características distintas de un mismo o dos fenómenos diferentes.

No necesitan, ni suelen llevar factor de ampliación ya que el guarismo de menor valor puede colocárselo en el denominador sin que ello afecte la interpretación.

Ejemplo: Si tenemos 450000 habitantes y entre ellos hay 5000 enfermeros.

$$\frac{450000}{5000} = 90$$

Tenemos 90 habitantes por cada enfermo.

Durante la semana de febrero, en un pueblo de 1500 habitantes, 190 personas asistieron a una fiesta de cumpleaños donde se les sirvió una comida preparada por varias personas de lugar. Unos días después 59 de los invitados se enfermaron por un cuadro clínico en que predominaba la diarrea.

De los 50 hombres presentes, 37 se enfermaron, 77 de los invitados tenían entre 20 y 40 años de edad y 53 se enfermaron. Calcule:

- a) La tasa de ataque entre los invitados a la fiesta.
- b) La tasa de ataque entre los hombres.
- c) La tasa de ataque entre las mujeres.
- d) La tasa de ataque para el grupo de entre 20 a 40 años.

Unidad IV

Medidas de Posición o de Centralización

Sumatorio:

$$\sum_{i=1}^N x_i = x_1 + x_2 + \cdots + x_N$$

Productoria:

$$\prod_{i=1}^N x_i = x_1 \cdot x_2 \cdots x_N$$

$$\sum_{i=1}^N x_i \cdot y_i = x_1 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_{N-1} \cdot y_{N-1}$$

$$\sum_{i=1}^N x^i = x^0 + x^1 + \cdots + x^N + x^{N+1}$$

$$\begin{aligned} \sum_{i=1}^K \sum_{j=1}^N x_i \cdot y_j &= x_1 \cdot y_1 + x_2 \cdot y_1 + \cdots + x_K \cdot y_1 + x_1 \cdot y_2 + x_2 \cdot y_2 + \cdots + x_K \cdot y_2 + x_1 \cdot y_N \\ &\quad + x_2 \cdot y_N + \cdots + x_K \cdot y_N \end{aligned}$$

Ejercicio:

1-

a) $\sum_{j=1}^4 (x_j - 3)^2$

b) $\sum_{j=1}^N a$

c) $\sum_{k=1}^5 f_k x_k$

d) $\sum_{j=1}^5 (x_j - a)$

Desarrollo

$$a) \sum_{j=1}^4 (x_j - 3)^2 = (x_1 - 3)^2 + (x_2 - 3)^2 + (x_3 - 3)^2 + (x_4 - 3)^2$$

$$b) \sum_{j=1}^N a = a_1 + a_2 + a_3 + \dots + a_N = Na$$

$$c) \sum_{k=1}^5 f_k x_k = f_1 x_1 + f_2 x_2 + f_3 x_3 + f_4 x_4 + f_5 x_5 =$$

$$d) \sum_{j=1}^3 (x_j - a) = (x_1 - a) + (x_2 - a) + (x_3 - a) = x_1 - a + x_2 - a + x_3 - a \\ = x_1 + x_2 + x_3 - 3a$$

2-

$$a) x_1^2 + x_2^2 + x_3^2 + \dots + x_{10}^2 = \sum_{i=1}^{10} x_i^2$$

$$b) (x_1 + y_1) + (x_2 + y_2) + \dots + (x_8 + y_8) = \sum_{i=1}^8 (x_i + y_i)$$

$$c) f_1 x_1^3 + f_2 x_2^3 + f_3 x_3^3 + \dots + f_{20} x_{20}^3 = \sum_{i=1}^{20} f_i x_i^3$$

$$d) a_1 b_1 + a_2 b_2 + a_3 b_3 + \dots + a_N b_N = \sum_{i=1}^N (a_i b_i)$$

$$e) f_1 x_1 y_1 + f_2 x_2 y_2 + f_3 x_3 y_3 + f_4 x_4 y_4 = \sum_{i=1}^4 (f_i x_i y_i)$$

Demostración:

$$\sum_{j=1}^N (ax_j + by_j - cz_j) = a \sum_{j=1}^N (x_j) + b \sum_{j=1}^N (y_j) - c \sum_{j=1}^N (z_j)$$

$$\begin{aligned}
& \sum_{j=1}^N (ax_j + by_j - cz_j) = \\
&= (ax_1 + by_1 - cz_1) + (ax_2 + by_2 - cz_2) + \dots + (ax_N + by_N - cz_N) \\
&== ax_1 + by_1 - cz_1 + ax_2 + by_2 - cz_2 + \dots + ax_N + by_N - cz_N = \\
&= ax_1 + by_1 - cz_1 + ax_2 + by_2 - cz_2 + \dots + ax_N + by_N - cz_N = \\
&= ax_1 + by_1 - cz_1 + ax_2 + by_2 - cz_2 + \dots + ax_N + by_N - cz_N = \\
&= (ax_1 + ax_2 + \dots + ax_N) + (by_1 + by_2 + \dots + by_N) = \\
&= -(cz_1 + cz_2 + \dots + cz_M) = \\
&= a(x_1 + x_2 + \dots + x_N) + b(y_1 + y_2 + \dots + y_N) \\
&- c(z_1 + z_2 + \dots + z_M) == a \sum_{j=1}^N (x_j) + b \sum_{j=1}^N (y_j) - c \sum_{j=1}^N (z_j)
\end{aligned}$$

Parámetros de posición:

Promedios y medidas de Centralización: un promedio es un valor, que es típico o representativo de un conjunto de datos. Como tales valores tienden a situarse en el centro del conjunto de datos ordenados según su magnitud, los promedios se conocen también como medida de centralización.

Se pueden definir varios tipos de medida de centralización, las más comunes son media aritmética, mediana y modo o moda, media geométrica y media armónica. Cada una de ellas tiene su ventaja e inconvenientes dependiendo del caso.

Media:

Para datos simples \rightarrow Frecuencia = 1

x	f
x_1	1
x_2	1
\vdots	\vdots
x_N	1

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$+ f_N = N$$

x	f
x_1	f_1
x_2	f_2
\vdots	\vdots
x_N	f_N

$$\sum_{i=1}^N f_i = N$$

Formula de la Media Aritmética

$$\bar{x} = \frac{\sum_{i=1}^N x_i f_i}{\sum_{i=1}^N f_i} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_N f_N}{f_1 + f_2 + \dots + f_N}$$

Mediana: la mediana de una colección de datos ordenados en orden de magnitud el valor medio o la media aritmética de los dos valores medios.

Ejemplo 1: 3,4, 4, 5, 6, 8, 8, 8, 10

La mediana es: $\tilde{x} = 6$

Ejemplo 2: 5, 5, 7, 9, 11, 12, 15, 18

$$\tilde{x} = \frac{9 + 11}{2} = 10$$

Para datos agrupados

	x_i	f_i	F_i
$[x_1 - x_2]$	$(x_1+x_2)/2$		
$[x_2 - x_3]$	$(x_2+x_3)/2$		
\vdots	\vdots	\vdots	\vdots
$[x_{k-1} - x_k]$	$(x_{k-1}+x_k)/2$	f_k	F_k

$x_i = \text{marca de clase}$

$f_i = \text{frecuencia absoluta}$

$F_i = \text{frecuencia absoluta acumulada}$

Mediana

$$\tilde{x} = L_i + \frac{\frac{N}{2} - (\sum f_{i-1})}{f_i} C$$

$L_i =$ límite real inferior de la clase mediana (es decir la clase que contiene a la mediana).

$N =$ número total de datos.

$\sum f_{i-1} =$ suma de todas las frecuencias de todas las clases por debajo de la clase mediana.

$f_i =$ frecuencia de la clase mediana.

$C =$ tamaño del intervalo de la clase mediana.

Ejemplo:

Peso (libras)	x_i	f_i	F_i
118 – 126	122	3	3
127 – 135	131	5	8
136 – 144	140	9	17
145 – 153	149	12	29
154 – 162	158	5	34
163 – 171	167	4	38
172 – 180	176	2	40
TOTAL		40	

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^N x_i f_i}{\sum_{i=1}^N f_i} = \frac{122.3 + 131.5 + 140.9 + 149.12 + 158.5 + 167.4 + 176.2}{40} \\ &= \\ &= \frac{366 + 655 + 1260 + 1788 + 790 + 668 + 352}{40} = 146,975 \cong 147 \text{ libras} \end{aligned}$$

$$\frac{N}{2} = \frac{40}{2} = 20$$

$$\tilde{x} = 144,5 + \frac{20 - 17}{12} 9 = 144,5 + \frac{3}{12} 9 = 144,5 + 2,25 = 146,75$$

$$\cong 146.8 \text{ libras}$$

Modo o Moda

La moda de una serie de números es aquel valor que presenta la mayor frecuencia, es decir, es el valor más común (repetitivo). La moda puede existir o no, incluso si existe puede no ser única.

Ejemplo 1: 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18.

La moda es: $\hat{x} = 9$

Ejemplo 2: 3, 5, 8, 10, 12, 15, 16 → No tiene moda → Frecuencia = 1

Ejemplo 3: 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, 9 ⇒ $\hat{x} = 4,7$ ⇒ *Bimodal*

En el caso de los datos agrupados tenemos lo siguiente:

$$\hat{x} = L_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) C \qquad \hat{x} = 144,5 + \left(\frac{3}{3 + 7} \right) 9 \Rightarrow \hat{x} = 147,2 \text{ libras}$$

L_i = límite inferior real de la clase modal.

Δ_1 : Exceso de la frecuencia modal sobre la frecuencia de la clase contigua inferior

Ejemplo: $(12 - 9) = 3$ en el caso anterior.

Δ_2 : Exceso de la frecuencia modal sobre la frecuencia de la clase contigua superior.

Ejemplo: $(12 - 5) = 7$ en el caso anterior.

C = ancho o tamaño del intervalo.

Relación empírica entre Media, Mediana y Moda

Media – Moda = 3 veces la diferencia de media – mediana

$$\bar{x} - \hat{x} = 3(\bar{x} - \tilde{x})$$

$$\bar{x} - \hat{x} = 3(\bar{x} - \tilde{x}) = 3\bar{x} - 3\tilde{x} \Rightarrow \hat{x} = \bar{x} - 3\bar{x} + 3\tilde{x}$$

$$\hat{x} = -2\bar{x} + 3\tilde{x}$$

$$\hat{x} = (3.145,8) - (2.147,0)$$

$$\hat{x} = 146,4$$

Altura (pulgadas)	x_i	f_i	F_i	$x_i f_i$
60 – 62	61	5	5	305
63 – 65	64	18	23	1152
66 – 68	65	42	65	2814
69 – 71	70	27	92	1890
72 – 74	73	8	100	584
TOTAL		100		

$$\bar{x} = \frac{6745}{100} = 67,45 \cong 67,5$$

$$\tilde{x} = L_i + \left(\frac{\frac{N}{2} - \sum f_{i-1}}{f_i} \right) C \quad \rightarrow \quad L_i = \frac{66 + 65}{2} = 65,5$$

$$\frac{N}{2} = \frac{100}{2} = 50 \quad y \quad C = 3$$

$$\tilde{x} = 65,5 + \left(\frac{50 - 23}{42} \right) 3 = 67,42$$

$$\hat{x} = L_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) C = 65,5 + \left(\frac{24}{24 + 15} \right) 3 = 67,34$$

Quartiles, Deciles y Percentiles

Si una serie de datos se coloca en orden de magnitud, el valor medio (o la media aritmética de los dos valores) que divide el conjunto de datos en dos partes iguales es la mediana. Por extensión, de esa idea, se puede pensar en aquellos valores que divide a los datos en 4 partes iguales. Estos valores, representados por Q_1 , Q_2 , Q_3 se llaman primer, segundo y tercer cuartil, respectivamente el valor de Q_2 es igual al de la mediana.

Análogamente los valores que dividen los datos en 10 partes iguales se llama deciles y se representan con D_1 , D_2 , D_3 , ..., D_9 , mientras que los valores que dividen los datos en 100 partes iguales se llaman percentiles y se representan con P_1 , P_2 , ..., P_{99} . El quinto decil y el percentil 50 corresponde a la mediana.

Los percentiles P_{25} y P_{75} corresponden al primer y tercer cuartil. En conjunto cuartiles, deciles y percentiles y otros valores obtenidos por subdivisiones análogas de los datos se llaman cuantiles.

Quartil

$$Q_1 = L_i + \left(\frac{\frac{N}{4} - \sum f_{i-1}}{f_i} \right) C \rightarrow \frac{N}{4} \text{ es la clase en la que se encuentra el } Q_1$$

$$Q_2 = L_i + \left(\frac{\frac{2.N}{4} - \sum f_{i-1}}{f_i} \right) C$$

$\rightarrow \frac{2.N}{4}$ es la clase en la que se encuentra el Q_2

$$Q_3 = L_i + \left(\frac{\frac{3.N}{4} - \sum f_{i-1}}{f_i} \right) C \rightarrow \frac{3.N}{4} \text{ es la clase en la que se encuentra el } Q_3$$

Decil

$$D_n = L_i + \left(\frac{\frac{N.n}{10} - \sum f_{i-1}}{f_i} \right) C$$

$\rightarrow \frac{N.n}{10}$ es la clase en la que se encuentra el D_n

Percentil

$$P_n = L_i + \left(\frac{\frac{N.n}{100} - \sum f_{i-1}}{f_i} \right) C$$

$\rightarrow \frac{N.n}{100}$ es la clase en la que se encuentra el P_n

Ejemplo:

Calificación	Número de estudiantes
90 – 100	9
80 – 89	32
70 – 79	43
60 – 69	21
50 – 59	11
40 – 49	3
30 – 39	1
Total	120

Calificación	X_i	(f_i)	$F(x_i)$	$h(x_i)$	$H(x_i)$	%	$\%_{Ac}$	$x_i f_i$
30 – 39	34,5	1	1	0,01	0,01	0,83	0,83	34,5
40 – 49	44,5	3	4	0,03	0,04	2,50	3,33	133,5
50 – 59	54,5	11	15	0,09	0,13	9,17	12,50	599,5
60 – 69	64,5	21	36	0,18	0,30	17,50	30,00	1354,5
70 – 79	74,5	43	79	0,36	0,66	35,83	65,83	3203,5
80 – 89	84,5	32	111	0,27	0,93	26,67	92,50	2704
90 – 99	94,5	9	120	0,08	1,00	7,50	100,00	850,5
Total		120		1,00		100,00		8880

$$\bar{x} = \frac{8880}{120} = 74 \quad \frac{N}{2} = \frac{120}{2} = 60$$

$$\tilde{x} = 69,5 + \left(\frac{60 - 36}{43} \right) 9$$

$$\tilde{x} = 74,52 \quad L_i = \frac{69 + 70}{2} = 69,5$$

$$\hat{x} = 69,5 + \left(\frac{(43 - 21)}{(43 - 21) + (43 - 32)} \right) 9$$

$$\hat{x} = 69,5 + \left(\frac{22}{22 + 11} \right) 9$$

$$\hat{x} = 75,5$$

$$Q_1 = L_i + \left(\frac{\frac{N}{4} - \sum f_{i-1}}{f_i} \right) C$$

$$\frac{N}{4} = \frac{120}{4} = 30$$

$$Q_1 = 59,5 + \left(\frac{30 - 15}{21} \right) 9$$

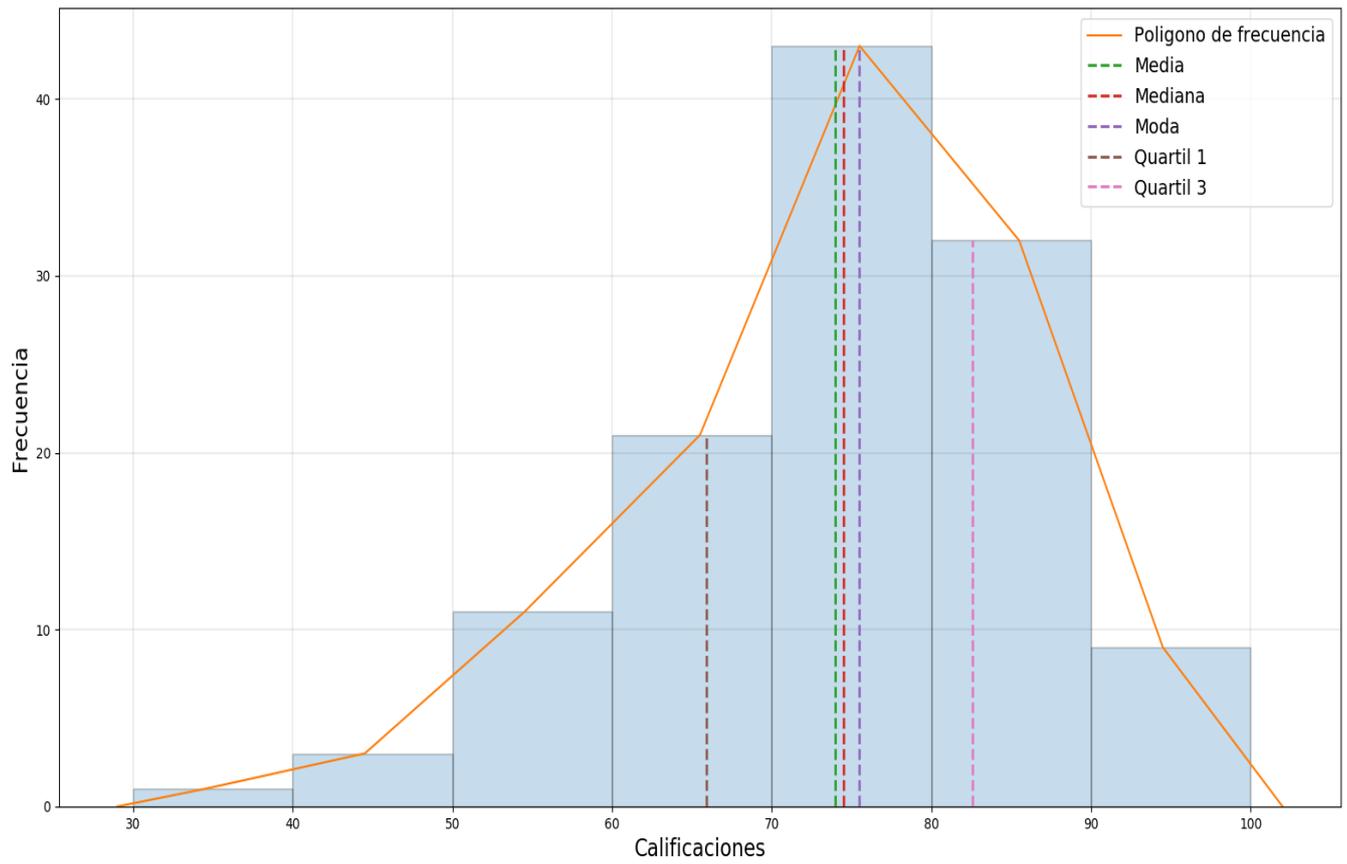
$$Q_1 = 65,93$$

$$Q_3 = L_i + \left(\frac{\frac{3N}{4} - \sum f_{i-1}}{f_i} \right) C$$

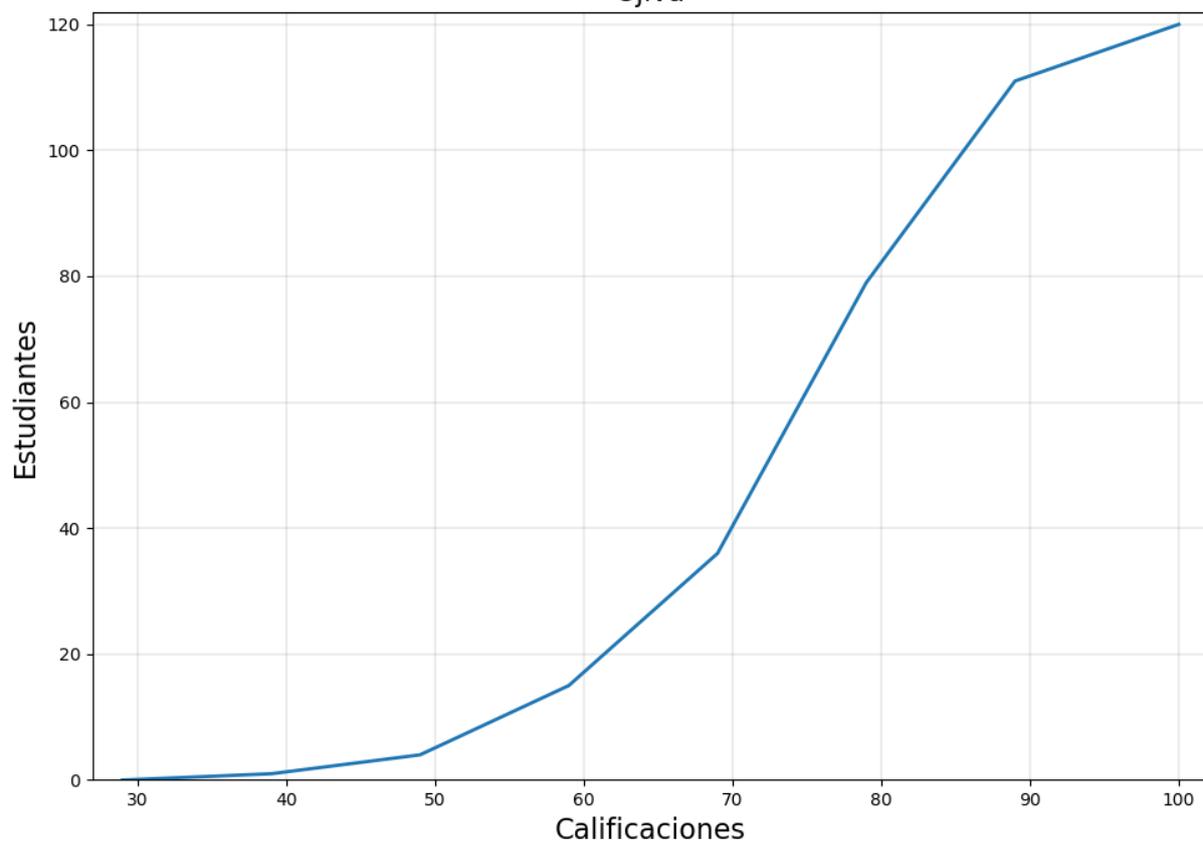
$$\frac{3N}{4} = \frac{3 \cdot 120}{4} = 90$$

$$Q_3 = 79,5 + \left(\frac{90 - 79}{32} \right) 9$$

$$Q_3 = 82,6$$



Ojiva



Unidad V

Parámetros de dispersión

Dispersión o variación: al grado en que los datos numéricos tienden a extenderse alrededor de un valor medio se llama variación o dispersión de los datos; se utilizan distintas medidas de dispersión o variación, las más empleadas son rango, la desviación media y la desviación típica.

Rango: el rango de un conjunto de números es la diferencia entre el mayor y el menor de todos ellos. Ejemplo:

$$2, 3, 3, 5, 5, 5, 8, 10, 12 = 10$$

Desviación de cada lectura:

$x_1, x_2, x_3, x_4, \dots, x_n$ lectura de cada medición

$$\bar{x} = \frac{x_1, x_2, x_3, x_4, \dots, x_n}{N}$$

Desviación = lectura – media aritmética

$$d_i = x_i - \bar{x} = \begin{cases} > 0 \rightarrow x_i \text{ derecha de } \bar{x} \\ < 0 \rightarrow x_i \text{ izquierda de } \bar{x} \\ = 0 \rightarrow \text{Coinciden } x_i = \bar{x} \end{cases}$$

Teorema

$$\begin{aligned} \sum_{i=1}^N d_i &= 0 & \sum_{i=1}^N d_i &= d_1 + d_2 + d_3 + d_4 + \dots + d_N = \\ & & &= (x_1 - \bar{x}) + (x_2 - \bar{x}) \dots = x_1 - \bar{x} + x_2 - \bar{x} + \dots = \\ & & &= (x_1 + x_2 + \dots) - (N\bar{x}) = \sum x_i - N\bar{x} = \sum x_i - N \frac{\sum x_i}{N} = \\ & & &= \sum x_i - \sum x_i = 0 \end{aligned}$$

Ejemplo:

$$x_1 = 3 \qquad d_1 = 3 - 5 = -2$$

$$x_2 = 4 \qquad d_2 = 4 - 5 = -1$$

$$x_3 = 6 \qquad d_3 = 6 - 5 = 1$$

$$x_4 = 7 \qquad d_4 = 7 - 5 = 2$$

$$x_5 = 5 \qquad d_5 = 5 - 5 = 0$$

$$\bar{x} = 5$$

$$\sum_{i=1}^5 d_i = -2 - 1 + 1 + 2 + 0 = 0$$

Desviación media $\bar{d} = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{N} = \frac{1}{N} \sum_{i=1}^N |x_i - \bar{x}|$

$$\bar{d} = \frac{|-2| + |-1| + |1| + |2| + |0|}{5} = 1,2 \quad \text{desviación media}$$

Para datos agrupados:

x_i	f_i
x_1	f_1
x_2	f_2
x_3	f_3
\vdots	\vdots
x_k	f_k

$$\sum_{i=1}^k (x_i - \bar{x}) f_i = (x_1 - \bar{x}) f_1 + (x_2 - \bar{x}) f_2 + (x_3 - \bar{x}) f_3 + \dots + (x_k - \bar{x}) f_k = 0$$

$$\bar{d} = \frac{\sum_{i=1}^N |x_i - \bar{x}| f_i}{\sum_{i=1}^k f_i}$$

Rango semiintercuartil:

El rango semiintercuartil, o desviación cuartil, de un conjunto de datos se denota Q y está definido por

$$Q = \frac{Q_3 - Q_1}{2}$$

Donde Q_1 y Q_3 son el primero y tercer cuartiles en los datos. Algunas veces se usa el rango intercuartil $Q_3 - Q_1$; sin embargo, el rango semiintercuartil es más usado como medida de dispersión.

Rango percentil:

El rango percentil 10-90 de un conjunto de datos está definido por

$$\text{Rango percentil } 10 - 90 = P_{90} - P_{10}$$

donde P_{10} y P_{90} son los percentiles 10 y 90 en los datos. El rango semipercentil 10-90, $\frac{1}{2}(P_{90} - P_{10})$, también puede usarse, pero no es muy común.

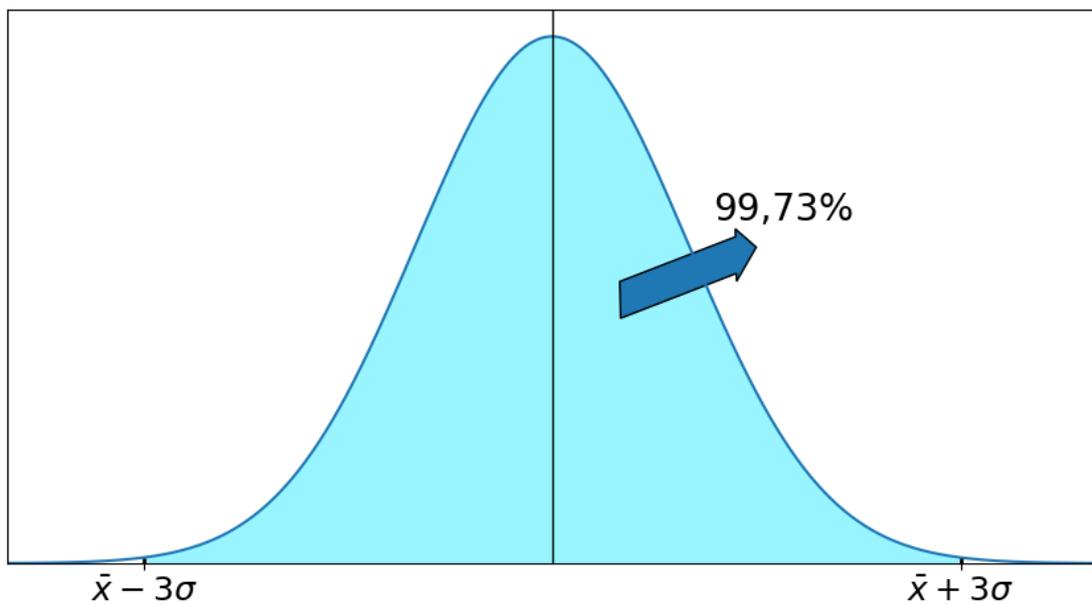
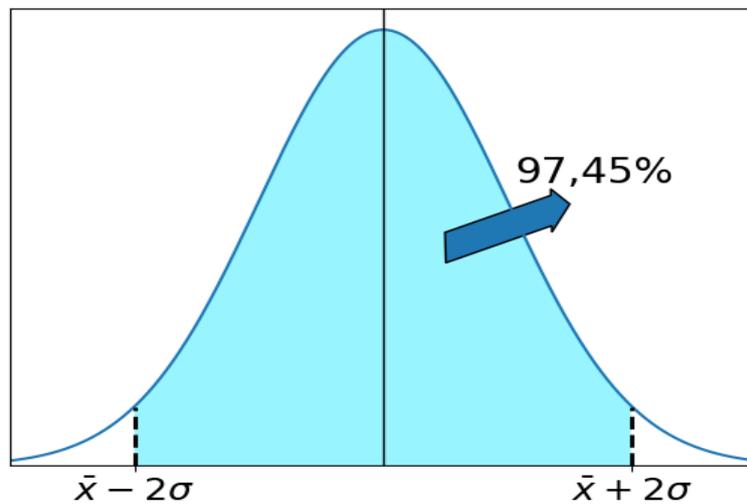
Varianza

La varianza de un conjunto de datos se define como

$$V_{(x)} = \frac{\sum_{i=1}^N f_i (X_i - \bar{X})^2}{N} \quad \text{para datos simples}$$

Donde $N = \sum_{i=1}^N f_i$. Esta fórmula es útil para datos agrupados.

Punto de inflexión: punto en que la curva cambia de concavidad.



Informe $\bar{x} \pm \sigma$

$(\bar{x} - \sigma, \bar{x} + \sigma)$

Coefficiente de Variación: el coeficiente de variación es útil para comparar distribuciones donde las unidades pueden ser diferentes. El coeficiente de variación es independiente de las unidades utilizadas. Un inconveniente del coeficiente de variación es que deja de ser útil cuando la media aritmética está próxima a cero.

Fórmula:

$$C_v = \frac{\sigma}{\bar{x}} \quad \%C_v = \frac{\sigma}{\bar{x}} * 100$$

Ejercicios:

x_i	f_i	$x_i f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})f_i$	$ x_i - \bar{x} f_i$	$(x_i - \bar{x})^2 f_i$
99,5	5	497,5	-129,584	-647,92	647,92	83960,06528
139,5	48	6696	-89,584	-4300,032	4300,032	385214,0667
179,5	265	47567,5	-49,584	-13139,76	13139,76	651521,8598
219,5	458	100531	-9,584	-4389,472	4389,472	42068,69965
259,5	281	72919,5	30,416	8546,896	8546,896	259962,3887
299,5	128	38336	70,416	9013,248	9013,248	634676,8712
339,5	35	11882,5	110,416	3864,56	3864,56	426709,257
379,5	7	2656,5	150,416	1052,912	1052,912	158374,8114
Total	1227	281086,5		0,432	44954,8	2642488,02

$$\bar{X} = \frac{281086,5}{1227} = 229,084 \text{ mg} \qquad \bar{d} = \frac{44954,8}{1227} = 36,638 \text{ mg}$$

$$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2 f_i}{\Sigma f_i} = \frac{2642488,02}{1227} = 2153,617 \text{ mg}^2$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{2153,617 \text{ mg}^2} = 46,407 \text{ mg}$$

$$Cv = \pm \frac{46,407 \text{ mg}}{229,084 \text{ mg}} = 0,20 \times 100 \% = 20 \%$$

- 1) Hallar la desviación media, varianza, desviación estándar y coeficiente de variación de las alturas de 100 estudiantes de una universidad. La altura está en pulgadas.

Alturas (pulgadas)	$f_i(x_i)$	x_i	$x_i f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})f_i$	$ x_i - \bar{x} f_i$	$(x_i - \bar{x})^2 f_i$
60 – 62	5	61	305	-6,45	-32,25	32,25	208,0125
63 – 65	18	64	1152	-3,45	-62,1	62,1	214,245
66 – 68	42	67	2814	-0,45	-18,9	18,9	8,505
69 – 71	27	70	1890	2,55	68,85	68,85	175,5675
72 – 74	8	73	584	5,55	44,4	44,4	246,42
Total	100		6745		0	226,5	852,75

$$\bar{x} = \frac{6745 \text{ pulgadas}}{100} = 67,45 \text{ pulgadas}$$

$$\bar{d} = \frac{226,50 \text{ pulgadas}}{100} = 2,265 \text{ pulgadas}$$

$$\sigma^2 = \frac{852,75}{100} = 8,52 \text{ pulgadas}^2$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{8,52 \text{ pulgadas}^2} = \pm 2,92 \text{ pulgadas}$$

$$Cv = \frac{2,92 \text{ pulgadas}}{67,45 \text{ pulgadas}} \cdot 100 = 0,043 \cdot 100 = 43\%$$

2)

Máximo de carga (Toneladas cortas)	N° de cables	x_i	$x_i f_i$	$x_i - \bar{x}$	$(x_i - \bar{x}) f_i$	$ x_i - \bar{x} f_i$	$(x_i - \bar{x})^2 f_i$
9,3 – 9,7	2	9,5	19	-1,59	-3,18	3,18	5,07
9,8 – 10,2	5	10	50	-1,09	-5,46	5,46	5,96
10,3 – 10,7	12	10,5	126	-0,59	-7,10	7,10	4,20
10,8 – 11,2	17	11	187	-0,09	-1,56	1,56	0,14
11,3 – 11,7	14	11,5	161	0,41	5,72	5,72	2,33
11,8 – 12,2	6	12	72	0,91	5,45	5,45	4,95
12,3 – 12,7	3	12,5	37,5	1,41	4,23	4,23	5,95
12,8 – 13,2	1	13	13	1,91	1,91	1,91	3,64
Total	60		665,5		0,00	34,60	32,25

$$\bar{x} = \frac{665,5 \text{ toneladas}}{60} = 11,09 \text{ toneladas}$$

$$\bar{d} = \frac{34,6 \text{ toneladas}}{60} = 0,58 \text{ toneladas}$$

$$\sigma^2 = \frac{32,23}{60} = 0,54 \text{ toneladas}^2$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{0,54 \text{ toneladas}^2} = \pm 0,73 \text{ toneladas}$$

$$Cv = \frac{0,73 \text{ toneladas}}{11,09 \text{ toneladas}} \cdot 100 = 0,07 \cdot 100 = 7\%$$

Unidad VI

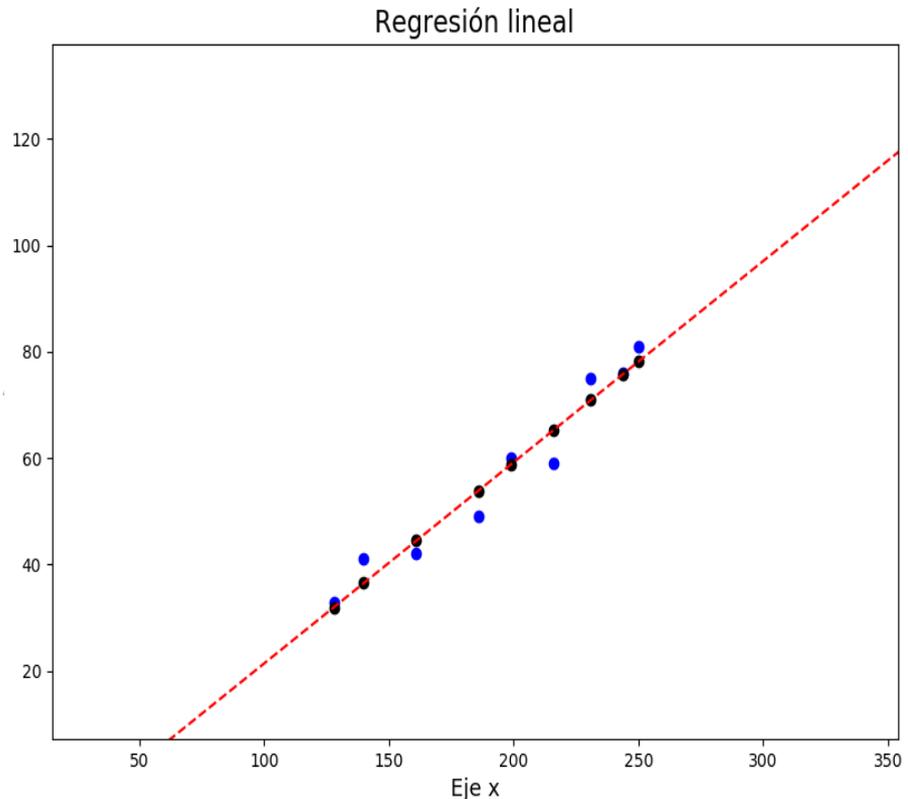
Análisis de Regresión y Correlación

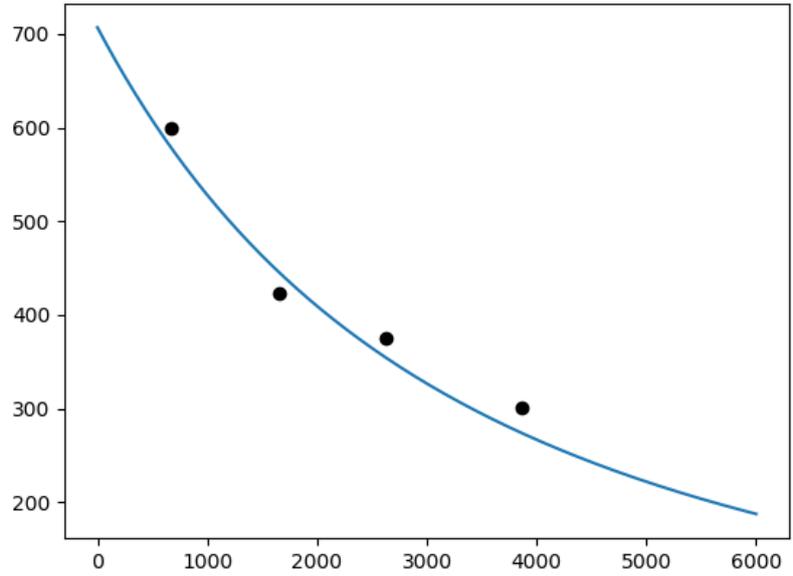
Relación entre dos variables: muy a menudo se encuentra en la práctica que existe una relación entre dos o más variables. Por ejemplo: los pesos de los hombres adultos dependen en cierto modo de su estatura, las circunferencias de los círculos dependen de sus radios y la presión de una masa dada de gas depende de su temperatura y volumen. Se desea frecuentemente expresar esta relación mediante una ecuación matemática que ligue las variantes.

Curva de ajuste: para llegar a determinar una ecuación que relacione las variables, un primer paso que nos sirve de ayuda es la colección de datos que muestre los correspondientes valores de las variables consideradas.

$$y = a_0 + a_1x$$

Altura	Pesos
x	y
x_1	y_1
x_2	y_2
x_3	y_3
\vdots	\vdots
x_n	y_n





N° individuos

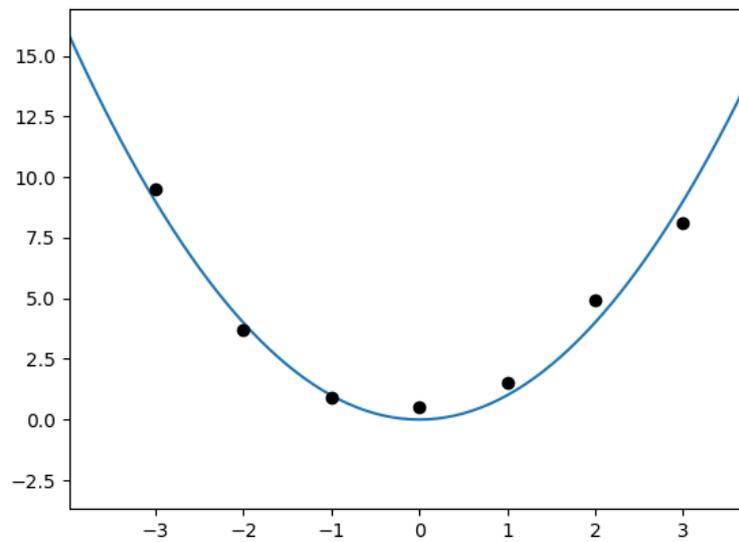
En un Gas

Presión	Volumen
x	y
x_1	y_1
x_2	y_2
x_3	y_3
\vdots	\vdots
x_n	y_n

$$y = \frac{1}{a_x}$$

N° de mediciones

$$y = a_0 + a_1x + a_2x^3$$

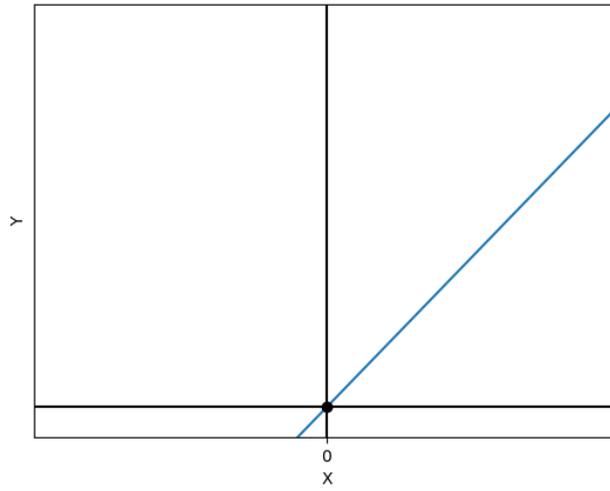


$a_0 =$ ordenada al origen

$\operatorname{tga} = a_1$ pendiente

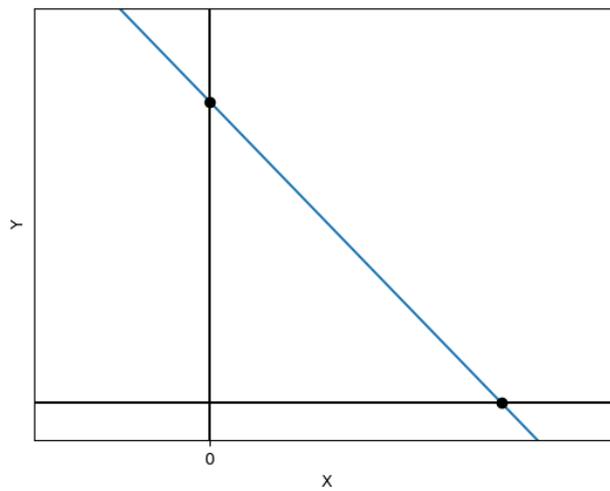
1. $\hat{\alpha} < 90^\circ$ $a_1 > 0$

$a > 0$



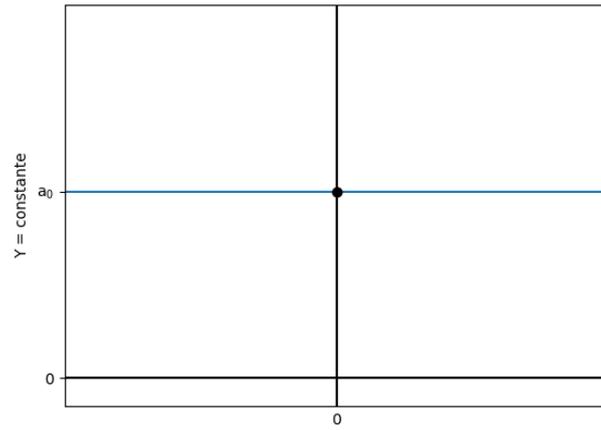
2. $\hat{\alpha} > 90^\circ$ $a_1 < 0$

$a < 0$



3. $\hat{\alpha} = 0^\circ$ $a_1 = 0$

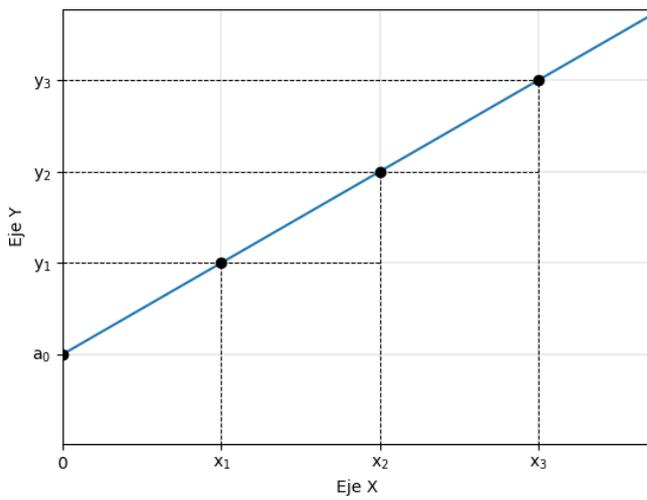
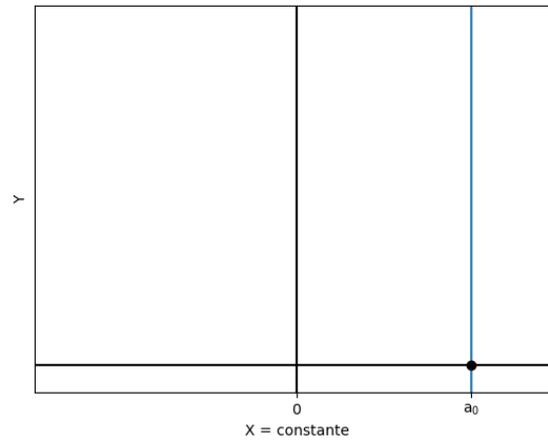
$$y = a_0$$



$$4. \hat{\alpha} = 90^\circ$$

$$a_1 \rightarrow \infty$$

$$x = a_0$$



$$\tan \hat{\alpha} = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

$$\tan \hat{\alpha}' = \frac{\Delta y}{\Delta x} = \frac{y - y_2}{x - x_2}$$

$$\hat{\alpha} = \hat{\alpha}' \rightarrow \tan \hat{\alpha} = \tan \hat{\alpha}'$$

$$\frac{y - y_2}{x - x_2} = \frac{y_2 - y_1}{x_2 - x_1} = a_1 \frac{y - y_2}{x - x_2} = a_1$$

$$y - y_2 = a_1(x - x_2)$$

$$y - y_2 = a_1x - a_1x_2$$

$$y = a_1x + (-a_1x_2 + y_2) \rightarrow -a_1x_2 + y_2 = a_0$$

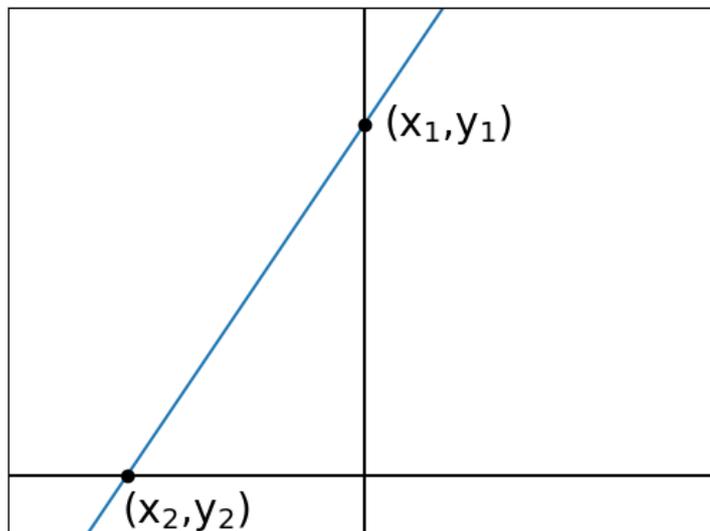
$$y = a_1x + a_0 \quad \text{Función lineal en Forma Explícita}$$

Ecuación general

$$Ax + By + C = 0$$

$$By = -Ax - C$$

$$y = -\frac{Ax}{B} - \frac{C}{B} \rightarrow a_1 = -\frac{A}{B}; \quad a_0 = -\frac{C}{B}$$



$$\frac{y - y_2}{x - x_2} = \frac{y_2 - y_1}{x_2 - x_1}$$

$$\frac{y - 0}{x - a} = \frac{0 - b}{a - 0}$$

$$\frac{y}{x - a} = -\frac{b}{a}$$

$$y = -\frac{b}{a}(x - a)$$

$$y = -\frac{b}{a}x + b$$

$$\frac{b}{a}x + y = b \quad \text{Dividiendo miembro a miembro por } b$$

$$\boxed{\frac{x}{a} + \frac{y}{b} = 1} \quad \text{Ecuación en forma Segmentaria}$$

Ejemplo:

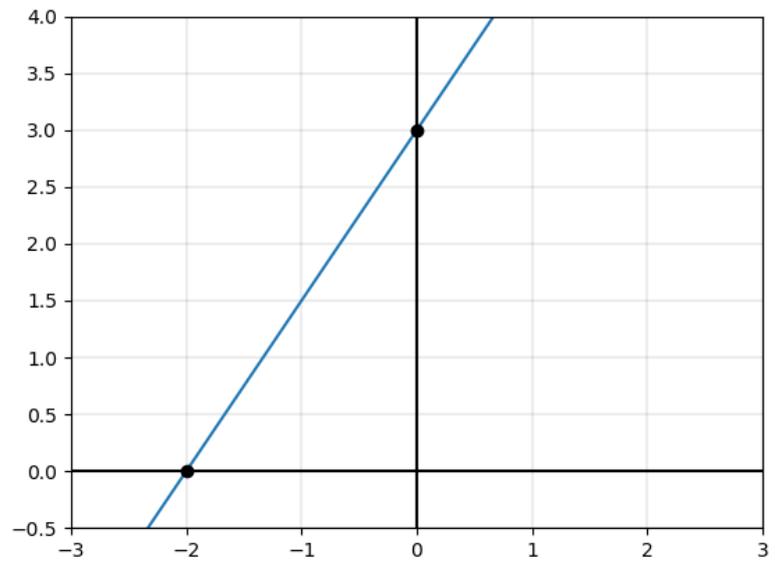
$$\frac{x}{-2} + \frac{y}{3} = 1 \text{ Ecuación segmentaria}$$

$$-\frac{y}{3} = \frac{x}{2} + 1$$

$$y = \frac{3x}{2} + 3 \text{ Forma explícita}$$

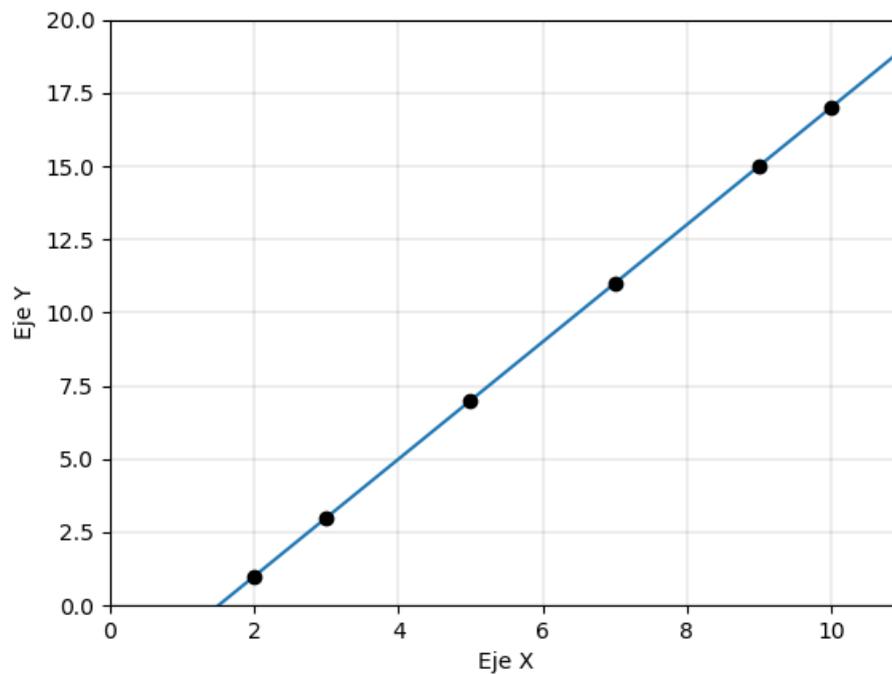
$$\frac{3}{2}x - y + 3 = 0$$

$$3x - 2y + 6 = 0 \text{ Ecuación general}$$



Ejercicios: Construir una línea recta que aproxime los datos de la tabla y hallar la ecuación de esta recta.

x	2	3	5	7	9	10
y	1	3	7	11	15	17



$$\frac{y - y_2}{x - x_2} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{3 - 1}{3 - 2} = 2$$

$$y - 3 = 2(x - 3)$$

$$y - 3 = 2x - 6$$

$$y = 2x - 6 + 3$$

$$\boxed{y = 2x - 3}$$

$$y - 2x = -3$$

$$\boxed{-\frac{1}{3}y + \frac{2}{3}x = 1}$$

$$-\frac{1}{3}y = 1 - \frac{2}{3}x$$

$$\boxed{y = -3 + 2x} \quad \boxed{-3 - y + 2x = 0}$$

Regresión:

A veces, basándose en los datos muestrales, se desea estimar el valor de una variable “y” correspondiente a un valor dado de una variable “x”.

Esto puede conseguirse estimando el valor “y” de la curva de mínimos cuadrados que ajusta los datos muestrales. La curva resultante se llama **curva de regresión** de y sobre x, puesto que Y se estima a partir de x.

Si se desea estimar el valor de x a partir de uno dado de y, se utiliza la curva de regresión de x sobre y, que proviene de intercambiar las variables en el diagrama de dispersión de modo que x sea la variable dependiente e y la independiente.

Recta de Regresión de mínimos cuadrados

$$y = a_0 + a_1x ?$$

$$y_1 = a_0 + a_1x_1$$

$$y_2 = a_0 + a_1x_2 \quad \Sigma y_i = Na_0 + a_1 \Sigma x_i$$

⋮

$$y_N = a_0 + a_1x_N$$

$$x_1y_1 = a_0x_1 + a_1x_1^2$$

$$x_2y_2 = a_0x_2 + a_1x_2^2 \quad \Sigma x_iy_i = a_0 \Sigma x_i + a_1 \Sigma x_i^2$$

⋮

$$x_Ny_N = a_0x_N + a_1x_N^2$$

$$\frac{\sum y_i - a_1 \sum x_i}{N} = a_0$$

$$\frac{\sum x_i y_i - a_1 \sum x_i^2}{\sum x_i} = a_0$$

$$\frac{\sum y_i - a_1 \sum x_i}{N} = \frac{\sum x_i y_i - a_1 \sum x_i^2}{\sum x_i}$$

$$\sum x_i (\sum y_i - a_1 \sum x_i) = N (\sum x_i y_i - a_1 \sum x_i^2)$$

$$\sum x_i \sum y_i - a_1 \sum x_i^2 = N \sum x_i y_i - N a_1 \sum x_i^2$$

$$N a_1 \sum x_i^2 - a_1 \sum x_i^2 = N \sum x_i y_i - \sum x_i \sum y_i$$

$$a_1 (N \sum x_i^2 - \sum x_i^2) = N \sum x_i y_i - \sum x_i \sum y_i$$

$$a_1 = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{N \sum x_i^2 - (\sum x_i)^2}$$

$$\frac{\sum y_i - N a_0}{\sum x_i} = a_1$$

$$\frac{\sum x_i y_i - a_0 \sum x_i}{\sum x_i^2} = a_1$$

$$\frac{\sum y_i - N a_0}{\sum x_i} = \frac{\sum x_i y_i - a_0 \sum x_i}{\sum x_i^2}$$

$$\sum x_i^2 (\sum y_i - N a_0) = \sum x_i (\sum x_i y_i - a_0 \sum x_i)$$

$$\sum x_i^2 \sum y_i - N a_0 \sum x_i^2 = \sum x_i \sum x_i y_i - a_0 \sum x_i^2$$

$$\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i = N a_0 \sum x_i^2 - a_0 \sum x_i^2$$

$$\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i = a_0 (N \sum x_i^2 - \sum x_i^2)$$

$$a_0 = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{N \sum x_i^2 - (\sum x_i)^2}$$

Ejemplo:

Datos	x	y	x^2	xy	y^2
1	65	68	4225	4420	4624
2	63	66	3969	4158	4356
3	67	68	4489	4556	4624
4	64	65	4096	4160	4225
5	68	69	4624	4692	4761
6	62	66	3844	4092	4356
7	70	68	4900	4760	4624
8	66	65	4356	4290	4225
9	68	71	4624	4828	5041
10	67	67	4489	4489	4489
11	69	68	4761	4692	4624
15	71	70	5041	4970	4900
Total	800	811	53418	54107	54849

$$y = a_0 + a_1x$$

$$a_1 = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{N\sum x_i^2 - (\sum x_i)^2} = \frac{12.54107 - 800.811}{12.53418 - 800^2} = 0,476$$

$$a_0 = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{N\sum x_i^2 - (\sum x_i)^2} = \frac{53418.811 - 800.54107}{12.53418 - 800^2} = 35,82$$

$$\boxed{y = 35,82 + 0,476x}$$

$$x = b_0 + b_1y$$

$$b_1 = \frac{N\sum y_i x_i - \sum y_i \sum x_i}{N\sum y_i^2 - (\sum y_i)^2} = \frac{12.54107 - 800.811}{12.54849 - 811^2} = 1,04$$

$$b_0 = \frac{\sum y_i^2 \sum x_i - \sum y_i \sum x_i y_i}{N\sum y_i^2 - (\sum y_i)^2} = \frac{54849.800 - 811.54107}{12.54849 - 811^2} = -3,38$$

$$\boxed{x = -3,38 + 1,04y}$$

Formula-Producto-Momento para el coeficiente de correlación lineal

$$r = \frac{\Sigma XY}{\sqrt{(\Sigma X^2) \cdot (\Sigma Y^2)}}$$

$$\bar{x} = \frac{800}{12} = 66,67$$

$$X = x - \bar{x}$$

$$\bar{y} = \frac{811}{12} = 67,58$$

$$Y = y - \bar{y}$$

$x - \bar{x}$	$y - \bar{y}$	$X \cdot Y$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
-1,67	0,42	-0,7014	2,7889	0,1764
-3,67	-1,58	5,7986	13,4689	2,4964
0,33	0,42	0,1386	0,1089	0,1764
-2,67	-2,58	6,8886	7,1289	6,6564
1,33	1,42	1,8886	1,7689	2,0164
-4,67	-1,58	7,3786	21,8089	2,4964
3,33	0,42	1,3986	11,0889	0,1764
-0,67	-2,58	1,7286	0,4489	6,6564
1,33	3,42	4,5486	1,7689	11,6964
0,33	-0,58	-0,1914	0,1089	0,3364
2,33	0,42	0,9786	5,4289	0,1764
4,33	2,42	10,4786	18,7489	5,8564
		40,3332	84,6668	38,9168

$$r = \frac{\Sigma XY}{\sqrt{(\Sigma X^2) \cdot (\Sigma Y^2)}} = \frac{40,33}{\sqrt{84,67 \cdot 38,92}} = 0,70$$

Hallar las rectas de regresión lineal “y” sobre “x” y “x” sobre “y” y el coeficiente de correlación lineal, de la siguiente tabla:

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

Datos	x	y	x^2	xy	y^2
1	1	1	1	1	1
2	3	2	9	6	4
3	4	4	16	16	16
4	6	4	36	24	16
5	8	5	64	40	25
6	9	7	81	63	49
7	11	8	121	88	64
8	14	9	196	126	81
Total	56	40	524	364	256

$$y = a_0 + a_1x$$

$$a_1 = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{N\sum x_i^2 - (\sum x_i)^2} = \frac{8.364 - 56.40}{8.524 - 56^2} = 0,64$$

$$a_0 = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{N\sum x_i^2 - (\sum x_i)^2} = \frac{524.40 - 56.364}{8.524 - 56^2} = 0,54$$

$$\boxed{y = 0,54 + 0,64x}$$

$$x = b_0 + b_1y$$

$$b_1 = \frac{N\sum y_i x_i - \sum y_i \sum x_i}{N\sum y_i^2 - (\sum y_i)^2} = \frac{8.364 - 40.56}{8.256 - 40^2} = 1,5$$

$$b_0 = \frac{\sum y_i^2 \sum x_i - \sum y_i \sum x_i y_i}{N\sum y_i^2 - (\sum y_i)^2} = \frac{256.56 - 40.364}{8.256 - 40^2} = -0,5$$

$$\boxed{x = -0,5 + 1,5y}$$

$$r = \frac{\sum XY}{\sqrt{(\sum X^2) \cdot (\sum Y^2)}}$$

$$\bar{x} = \frac{56}{8} = 7$$

$$X = x - \bar{x}$$

$$\bar{y} = \frac{40}{8} = 5$$

$$Y = y - \bar{y}$$

$x - \bar{x}$	$y - \bar{y}$	$X.Y$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
-6	-4	24	36	16
-4	-3	12	16	9
-3	-1	3	9	1
-1	-1	1	1	1
1	0	0	1	0
2	2	4	4	4
4	3	12	16	9
7	4	28	49	16
		84	132	56

$$r = \frac{\Sigma XY}{\sqrt{(\Sigma X^2) \cdot (\Sigma Y^2)}} = \frac{84}{\sqrt{132 \cdot 56}} = 0,98$$

Problema:

La siguiente tabla muestra la edad “x” y la presión sanguínea “y” de 12 mujeres.

- Hallar el coeficiente de correlación entre x e y.
- Determinar la ecuación de regresión de mínimos cuadrados de “y” sobre “x”.
- Estimar la presión sanguínea de una mujer de 45 años.

Edad (x)	56	42	72	36	63	47	55	49	38	42	68	60
Presión sanguínea (y)	147	125	160	118	149	128	150	145	115	140	152	155

Unidad VII

Probabilidad

Supóngase un suceso E, evento de un total de N casos posibles todos igualmente factibles que pueden presentarse en H de los casos posibles. Entonces la probabilidad de la aparición del suceso llamada ocurrencia viene dada por:

$$p = P(E) = \frac{h}{n} \rightarrow \frac{\text{casos favorables}}{\text{casos posibles}}$$

$$q = P(\text{no } E) = 1 - \frac{h}{n} = 1 - p$$

$$p + q = 1$$

$$q = P(\tilde{E}) = P(\text{No } E)$$

Ejemplo: Sea E el suceso que aparezcan los números de 3 o 4 en una sola tirada de un dado.

$$P\{E\} = \frac{2}{6} = \frac{1}{3} \qquad q = 1 - p$$

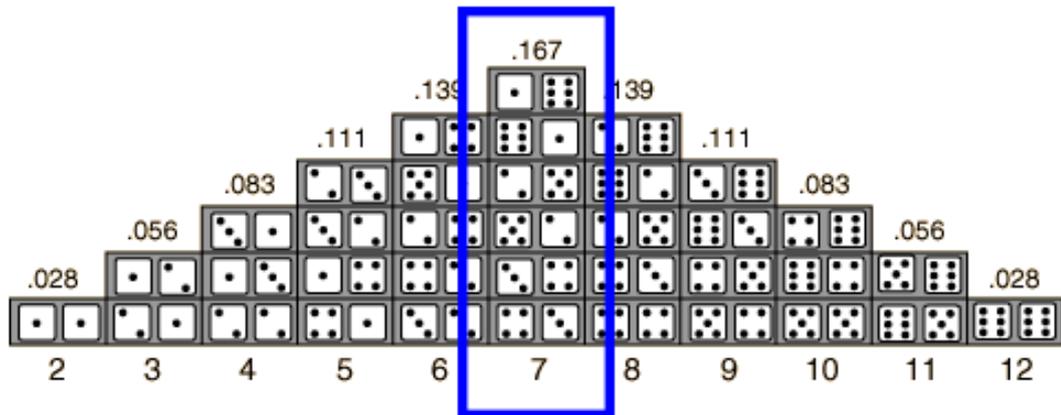
$$q = 1 - \frac{1}{3} = \frac{2}{3}$$

$$P\{\tilde{E}\} = \frac{4}{6} = \frac{2}{3}$$

La aparición de un número impar en una tirada de un dado equilibrado.

$$P\{E\} = \frac{3}{6} = \frac{1}{2} \cdot 100 = 50\%$$

La obtención de 7 puntos en una sola tirada de un par de dados.



$$P\{E\} = \frac{6}{36} = \frac{1}{6}$$

Probabilidad Condicional. Sucesos independientes y dependientes

Sea E_1 y E_2 dos sucesos, la probabilidad de que ocurra E_2 dado que ha ocurrido E_1 .

$$P\{E_1|E_2\}$$

Se llama probabilidad condicional de E_2 dado que E_1 sea presentado.

Esto se llama probabilidad condicional.

Si la ocurrencia o no ocurrencia de E_1 no afecta la probabilidad de la ocurrencia de E_2 denotado:

$$P\{E_1|E_2\} = P\{E_2\}$$

Se dice que E_1 y E_2 son sucesos independientes. Si no ocurre esto los sucesos se dicen dependientes.

$$E_1, E_2$$

$$P\{E_1 E_2\} = P\{E_1\} \cdot P\{E_1|E_2\} \text{ Dependiente}$$

$$P\{E_1 E_2\} = P\{E_1\} \cdot P\{E_2\} \text{ Independiente}$$

Ejemplos:

- 1- Sea E_1 y E_2 , respectivamente, los sucesos “cara en el quinto lanzamiento” y “cara en el sexto lanzamiento” de una moneda.

$$P\{E_1 E_2\} = \frac{1}{2} \frac{1}{2} = \frac{1}{4}$$

- 2- Supóngase una caja que contenga 3 bolas blancas y 2 bolas negras. Sea E_1 el suceso de que “la primera bola extraída sea negra” y E_2 el suceso de que “la segunda bola extraída sea negra” en extracciones sin reemplazamiento.

$$P\{E_1 E_2\} = \frac{2}{5} \frac{1}{4} = \frac{1}{10}$$

Sucesos mutuamente excluyentes

Se dice que dos o más sucesos se dicen mutuamente excluyentes si la ocurrencia de uno cualquiera de ellos imposibilita la ocurrencia de los otros.

$$P\{E_1 E_2\} = 0$$

Si $E_1 + E_2$ denota el suceso de que ocurra E_1 o E_2 o ambos

$$P\{E_1 + E_2\} = P\{E_1\} + P\{E_2\} - P\{E_1 E_2\}$$

En particular, si son excluyentes

$$P\{E_1 + E_2\} = P\{E_1\} + P\{E_2\}$$

Ejemplo:

- 3- Si E_1 es el suceso “extracción de un As de una baraja de cartas” y E_2 es el suceso “extracción de un Rey”. Cuál es la probabilidad de extracción de un As o un Rey en una sola extracción.

$$P\{E_1 E_2\} = \frac{4}{52} + \frac{4}{52} = \frac{2}{13}$$

Problemas:

1. La aparición de al menos de una cara en 2 lanzamiento de una moneda.
2. La aparición de cruz del siguiente lanzamiento de una moneda, si en 100 lanzamientos precisos aparecieron 56 caras.
3. De una caja que contiene 6 bolas rojas, 4 blancas y 5 azules se extraen una al azar, determine la probabilidad de que sea:
 - a. Roja

- b. Blanca
 - c. Azul
 - d. No roja
 - e. Roja o blanca
4. Un dado se lanza 2 veces. Hallar la probabilidad de obtener 4, 5 o 6 en el primer lanzamiento y 1, 2, 3 o 4 en el segundo lanzamiento.

BIBLIOGRAFIA

- Estadística Elemental, Jhon Freund – Gary a. Simon, Editorial: Prentice Hall, 1994
- Estadística, Murria R. Spiegel, Editorial: Serie Schaum- Mc Graw- Hill, 1995
- Probabilidad, Lipschutz, Editorial: Serie Schsum- Mc Graw – Hill, 2003
- Probabilidad y Estadística, Spiegel, Editorial: Serie Schaum- Mc Graw- Hill, 2014
- Estadística Aplicada a la Administración y a la Economía. Kazmier. Editorial: Serie Schaum- Mc Graw – Hill, 2006
- Fundamentos de Bioestadística, Marcello Pagano-Kimberlee Gauvreau Editorial: Thomson – Learning, 2003
- Estadística aplicada a Ciencias de la Salud – Liliana G. Bordón. Editorial PAPPEL, 2007

Guía de Problemas:

TRABAJO PRÁCTICO NRO. 1

- a) Redondear cada uno de los siguientes números a la aproximación indicada.
- a) 3256 aproximando centenas
 - b) 5,781 aproximando décimas
 - c) 0,0045 aproximando milésimas
 - d) 46,7385 aproximando centésimas
 - e) 125,9995 aproximando dos cifras decimales
- b) ¿Cuántas cifras significativas hay en cada una de las siguientes cantidades, suponiendo que los números se han registrado exactamente?
- a) 2,54 cm
 - b) 0,004500 yardas
 - c) 3.510.000 fanegas
 - d) 378 individuos
- c) Escribir cada uno de los siguientes números utilizando la notación científica. Suponer todas las cifras significativas a menos que se indique lo contrario.
- a) 0,000317
 - b) 428.000.000 (4 cifras significativas)
 - c) 21.600,00
 - d) 0,000009810
 - e) 732 millares
 - f) 1,80 decenas de millar

TRABAJO PRÁCTICO N° 2

CLASIFICACIÓN DE VARIABLES

Ejercicio 1.1. Clasificar las siguientes variables:

1. Preferencias políticas (izquierda, derecha o centro).
2. Marcas de cerveza.
3. Velocidad en Km / h.
4. El peso en Kg.
5. Nivel educativo (primario, secundario, superior).
6. Años de estudios completados.
7. Tipo de enseñanza (privada o pública).
8. Número de empleados de una empresa.
9. La temperatura de un enfermo en grados Celsius.
10. La clase social (baja, media o alta).
11. La presión de un neumático en N / cm².

Ejercicio 1.2. Colocar verdadero o falso, en caso de éste último justificar su respuesta.

- a). - La estadística es una ciencia para el tratamiento de los datos obtenidos.
(V - F)
- b). - El objeto de la Estadística es tomar decisiones con base científica
(V-F)
- c). - La población de la familia rosáceos es la clase angiospermas.
(V - F)
- d). - Todo estudio biométrico debe tener un marco de referencia
(V - F)
- e). - Las mujeres con síndrome de Turner, no son un parámetro en la población femenina.
(V - F)

- *Estadígrafos
La población.
- * Es toda magnitud cuyo valor es constante en toda
- * Estadística descriptiva
- * Es el resultado de hacer una medición.
- * Variable
- * Está relacionada con el error sistemático de las Mediciones.
- *Parámetro
- * Es lo que se estudia en este primer tema.
- * Dato
- * Es toda magnitud que permite diferenciar entre sí a Los componentes de una misma población.
Matemáticamente, se expresa como una función

TRABAJO PRÁCTICO N° 3

CUADROS Y GRÁFICOS ESTADÍSTICOS

Ejercicio 1. Sobre base de datos de la SAGPyA – Dirección de Forestación, en el marco “Captura de Carbono y Desarrollo Forestal Sustentable en la Patagonia Argentina”. CENIT CEPAL – UDESA. Informó que la superficie forestada en el marco de los regímenes del Período 1995 – 1999 fueron los siguientes:

- Neuquén: 890,08; 2.295,75; 2.704,10; 2.031,14; 3.135,56 Hectáreas
- Río Negro: 98,77; 395,60; 506,89; 751,90; 611,68 Hectáreas
- Chubut: 0.00; 154,50; 778,00; 1.397,49; 1.148,55 Hectáreas
- Santa Cruz: 8,00; 4,99; 0,00; 3,00; 0,00; Hectáreas

1. a). -Construir un cuadro estadístico que refleje en valores absolutos el SUB-TOTAL de la Región Patagónica por año.
2. b). - Porcentaje % por año de forestación y el total en los 5 años.

Ejercicio 2. En el año 1998 se publicó el Primer Inventario de Bosques Nativos, por la Secretaría de Ambiente y Desarrollo Sustentable, de Regiones Forestales en Hectáreas.

- En la Selva Misionera (Tierras forestales 914,823; Bosques Rurales 538,558)
- En la Selva Tucumano Boliviana (Tierras Forestales 3.697,483; Bosques Rurales 29,352)
- Bosque Andino Patagónico (Tierras Forestales 1.985,495; Bosques Rurales 0,00)
- Parque Chaqueño (Tierras Forestales 22.040,637; Bosques Rurales 1.327,347)

2. a).-Mediante un cuadro estadístico expresar estos datos colocando las cifras en totales de Hectáreas por zonas, y total del país. El porcentaje sobre el total del área forestal.

Ejercicio 3.- Las siguientes son las ganancias semanales en dólares de treinta vendedores: 425, 440, 610, 518, 324, 482, 624, 390, 468, 457, 509, 561, 482, 480,520, 654,768, 777, 675,

UNIVERSIDAD NACIONAL DE FORMOSA

FACULTAD DE CIENCIAS DE LA SALUD

Carrera: Técnico en Laboratorio– Lic. en Bromatología Cátedra: Bioestadística

453, 440, 450, 654, 655, 800, 876, 880,883, 778,789. Elabore una gráfica de tronco y hoja con las clasificaciones de tronco 3, 4, 5, 6, 7, 8; las decenas se deben usar como hojas.

Ejercicio 4.- Según datos del informe final de “Captura de Carbono y Desarrollo Forestal Sustentable en la Patagonia Argentina, arroja la Superficie de las Plantaciones Forestales durante el año 1998, en porcentaje %.

Provincia o región: Misiones 32,58 %; Corrientes 27,89; Entre Ríos 14,97 %; Buenos Aires 11,77 %; Noroeste (Jujuy, Salta y Tucumán): 2,57 %;

Centro (Córdoba, La Pampa y Santa Fe); 6,04 %; Sur (Chubut, Mendoza, Neuquén y Río Negro); 4,18. Totalizando un total de 780,396 hectáreas de superficie de plantaciones.

4. a). -Confeccione un cuadro estadístico con los datos proporcionados, es decir, teniendo en cuenta la Provincia o Región, el % porcentaje plantado y las Hectáreas de plantaciones forestales.

Ejercicio 5.- La siguiente es la distribución de la facturación mensual de 200 cuentas de un Laboratorio Municipal:

Importe en Dólares	Frecuencia
0.00 - 19.99	22
20.00 - 39.99	47
40.00 - 59.99	66
60.00 - 79.99	35
80.00 - 99.99	21
100.00 - 119.99	9
Total	

5. (a) Trace un histograma de esta distribución.

5. (b) Trace una gráfica de barras de esta distribución.

Ejercicio 6.- Convierta la distribución del ejercicio anterior en una distribución acumulativa de “menos de” y trace una ojiva.

Ejercicio 7.- Las siguientes son las alturas en centímetros de dieciséis estudiantes de Licenciatura en Bromatología: 172, 182, 177, 174, 166, 158, 170, 178, 163, 161, 171, 201, 166, 172, 191, 167. Elabore una gráfica de tronco y hoja con las clasificaciones de tronco 15, 16, 17, 18, 19 y 20.

Ejercicio 8.- Los siguientes son los pesos en libras de veinte solicitantes de empleo en un departamento de guardabosques de una región de E.E. U.U.:

225, 182, 194, 210, 205, 172, 181, 198, 164, 176, 180, 193, 178, 193, 208, 186, 183, 170, 186, 188. Elabore una gráfica de tronco y hoja con las clasificaciones de tronco 16, 17, 18, 19, 20, 21 y 22.

Ejercicio 9.- La siguiente es la distribución de los pesos de 150 mujeres que ingresaron a una universidad:

Peso en libras	Frecuencia
90 - 99	6
100 - 109	25
110 - 119	46
120 - 129	37
130 - 139	22
140 - 149	7
150 - 159	3
160 - 169	3
170 - 179	0
180 - 189	1
total	

- (a). Trace un histograma de esta distribución.
 (b). Trace un polígono de frecuencias de esta distribución.

Ejercicio 10.- La siguiente tabla muestra el número de trabajadores agrícolas y no agrícolas en Argentina en los años (1860 – 1970).

Año	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970
Trabajadores agrícolas (millones)	3,7	4,9	6,2	6,9	8,6	9,9	10,9	11,6	11,4	10,5	8,8	6,8
Trabajadores no agrícolas (millones)	1,7	2,8	4,3	6,1	8,8	13,4	18,2	25,8	31,0	38,4	42,9	52,2

Representar los datos utilizando (a) gráfico de línea; (b) diagrama de barras, (c) diagramas de barras compuestas.

Ejercicio 11.- Diseñar un pictograma apropiado para mostrar la variación en número de (a) trabajadores agrícolas, (b) trabajadores no agrícolas de los datos del problema anterior. ¿Puede diseñar un pictograma que muestre la variación de (a) y (b)?

Ejercicio 12.- Colocar verdadero o falso, en caso de éste último justificar su respuesta.

- El mejor método para influenciar al lector es el método tabular. V F
- En el método gráfico se pierde precisión en aras de la sencillez. V F
- Se debe siempre indicar la fuente de donde se tomaron los datos. V F
- Las líneas con las cuales se trazan los gráficos deben tener todo el mismo grosor. V F
- Cuando se usa escala en un gráfico, siempre debe colocarse el cero de la escala F

TRABAJO PRÁCTICO N° 4

TABLAS DE DISTRIBUCIÓN DE FRECUENCIAS

Ejercicio 1.- La puntuación final en matemáticas de 80 estudiantes de StateUniversity se registran en la tabla.

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

Con relación a esta tabla, encontrar:

- (a). -la puntuación más alta, (b). -la puntuación más baja, (c). -el rango.
- (d). - Las puntuaciones de los cinco estudiantes de mayor puntuación.
- (e). -Las puntuaciones de los cinco estudiantes de menor puntuación.
- (f). -La puntuación del décimo estudiante de mayor puntuación.
- (g). -¿Cuántos estudiantes obtuvieron puntuación de 75 o mayor?
- (h). -¿Cuántos estudiantes obtuvieron puntuación menor de 85?
- (i). - ¿Qué porcentaje de estudiantes obtuvo una puntuación mayor que 65 pero no mayor que 85?
- (j). - ¿Qué puntuación no tiene ningún estudiante?

UNIVERSIDAD NACIONAL DE FORMOSA

FACULTAD DE CIENCIAS DE LA SALUD

Carrera: Técnico en Laboratorio– Lic. en Bromatología Cátedra: Bioestadística

Ejercicio 2.- Elabore una distribución de las siguientes cantidades de óxidos de azufre (en toneladas emitidas por una planta industrial en 80 días). Para hacerlo tome 6 clases.

15.8	26.4	17.3	11.2	23.9	24.8	18.7	13.9	9.0	13.2
22.7	9.8	6.2	14.7	17.5	26.1	12.8	28.6	17.6	23.7
26.8	22.7	18.0	20.5	11.0	20.9	15.5	19.4	16.7	10.7
19.1	15.2	22.9	26.6	20.4	21.4	19.2	21.6	16.9	19.0
18.5	23.0	24.6	20.1	16.2	18.0	7.7	13.5	23.5	14.5
14.4	29.6	19.4	17.0	20.8	24.3	22.5	24.6	18.4	18.1
8.3	21.9	12.3	22.3	13.3	11.8	19.3	20.0	25.7	31.8
25.9	10.5	15.9	27.5	18.1	17.9	9.4	24.1	20.1	28.5

2. a). -Identificar la variable en estudio.
2. b). -Calcular las frecuencias absolutas acumuladas, las frecuencias relativas y las frecuencias relativas acumuladas.
2. c). -Grafique las frecuencias relativas acumuladas (ojiva)

Ejercicio 3.-

Para agrupar datos sobre el número de días lluviosos registrados por una estación climatológica para el mes de agosto durante los últimos seis días, un meteorólogo usa las clases:

- 0 - 5
- 6 - 11
- 12 - 16
- 18 - 24
- 24 - 30

Explique en dónde pueden surgir las dificultades.

Ejercicio 4

Las siguientes son las calificaciones que obtuvieron 40 estudiantes en un examen de Biología, en el cursillo de ingreso a la carrera Ingeniería Forestal.

75 89 66 52 90 68 83 94 77 60
 38 47 87 65 97 49 65 72 73 81
 63 77 31 88 74 37 85 76 74 63
 69 72 91 87 76 58 63 70 72 65

- a). -Confeccione una tabla de distribución de frecuencias en 8 clases.
- b). -Incluir frecuencias absolutas, frecuencias absolutas acumuladas, frecuencias relativas y frecuencias relativas acumuladas.
- c). - Representar gráficamente las frecuencias absolutas.

Ejercicio 5

La tabla muestra una distribución de frecuencias de la duración de 400 tubos de radio comprobados en la L&M TUBE COMPANY. Con referencia a esta tabla determinar:

Duración (horas)	Números de tubos
300 - 399	14
400 - 499	46
500 - 599	58
600 - 699	76
700 - 799	68
800 - 899	62
900 - 999	48
1000 - 1099	22
1100 - 1199	6
Total	400

- (a) Límite superior de la quinta clase
- (b) Límite inferior de la octava clase
- (c) Marca de clase de la séptima clase
- (d) Límites reales de la última clase
- (e) Tamaño del intervalo de clase
- (f) Frecuencia de la cuarta clase
- (g) Frecuencia relativa de la sexta clase
- (h) Porcentaje de tubos cuya duración no sobrepasa las 600 horas
- (i) Porcentaje de tubos cuya duración es mayor o igual a 900 horas
- (j) Porcentaje de tubos cuya duración es al menos de 500 horas, pero menor de 1000 horas

Ejercicio 6

Para los datos del problema anterior construir:

1. Un histograma
2. Un polígono de frecuencias correspondientes a la distribución
3. Una distribución acumulada relativa o porcentual

Ejercicio 7

Una encuesta efectuada en un hotel de Nueva York indicó que 40 huéspedes llegaron en los medios de transporte siguientes: automóvil, automóvil, autobús, avión, tren, autobús, autobús, avión, automóvil, avión, avión, autobús, avión, automóvil, automóvil, tren, tren, automóvil, automóvil, automóvil, automóvil, avión, avión, automóvil, autobús, automóvil, autobús, automóvil, avión, automóvil, avión, , avión, automóvil, automóvil, autobús, autobús, , tren, automóvil, autobús y automóvil.

Estructure una distribución categórica que demuestre las frecuencias correspondientes a los diferentes medios de transporte.

TRABAJO PRÁCTICO N° 5

MEDIDAS DE POSICIÓN Y ORDEN

Ejercicio 1.-

En diecinueve páginas de un informe, un mecanógrafo cometió: 0, 0, 1, 2, 0, 3, 1, 0, 0, 0, 0, 1, 0, 0, 4, 1, 0, 0 y 2 errores. Encuentre:

- a). - La media,
- b). - La mediana

Ejercicio 2.-

Los siguientes valores son las retenciones porcentuales de azufre de 42 áreas de vertientes del noroeste de Estado Unidos, registrados en 1989.

18.3	41.9	32.4	33.2	25.9
60.6	19.0	54.7	48.8	43.0
57.4	44.2	56.2	59.6	43.5
17.8	37.6	66.2	49.7	39.0
53.7	39.8	21.8	43.2	24.3
42.0	19.3	42.1	53.9	28.2
47.8	34.3	33.0	33.2	42.2
31.7	38.0	45.3	41.7	45.7
37.0	51.0			

- a). - Elabore una gráfica de tronco y hoja y úsela para encontrar la Media aritmética, Mediana y Modo.
- b). - Grafica un Histograma y un polígono de frecuencias acumuladas.

Ejercicio 3.-

Los siguientes valores son los diámetros de una parcela experimental de Pinustaeda, plantados en la provincia de Misiones. Los datos están expresados en centímetros.

138	142	113	126	135
142	159	157	140	157
121	128	142	164	155
139	143	158	140	118
142	146	123	130	137

2.a). -Obtenga la mediana directamente ordenando los datos de acuerdo a su tamaño.

2.b). - Obtenga la mediana elaborando primero un gráfico de tronco y hoja.

2.c). -Calcule la Media aritmética.

2.d). -Interprete cada una de las medidas calculadas.

Ejercicio 4.-

Encuentre la moda (si existe) de cada uno de los siguientes conjuntos de lecturas de alturas de los árboles dados en cm.

1. 144; 145; 146; 146; 148; 146; 146; 145; 147; 145; 144

2. 146; 149; 146; 141; 146; 149; 147; 147; 149; 149; 145

3. 167; 151; 175; 144; 152; 148; 156; 169; 143; 177; 161

¿Es adecuada la medida peticionada? ¿Cuál propondría Ud.?

Ejercicio 5.-

A continuación, se les presentan los pesos en Kg. De un grupo de trabajadores de una empresa Forestal.

52	54	62	66	78	56	78	50	54	50	60	68	66	66	65
72	64	63	62	68	61	78	68	78	66	76	64	72	74	52
70	62	60	80	78	88	78	74	86	84	90	82	76	74	57
84	90	80	94	90	86	84	80	81	82	98	96	92	97	72

Elabore:

1. Una distribución de frecuencias de clase completa, para determinar el número de clases, utilice su criterio.
2. Realice un Histograma y un polígono de frecuencias
3. Calcule la Media, la Mediana, la Moda, Q1, Q2, Q3, D5, P25, P50, P75.
4. Extraiga conclusiones de cada medida calculada.

Ejercicio 6.-

Los siguientes son los números de pasajeros de 46 autobuses con vista panorámica en Boston, Massachusetts:

Números de personas	Frecuencia
15 o menos	5
16 - 20	8
21 - 25	11
26 - 30	13
31 - 35	6
36 - 40	3
Total	

- a). - Encuentre si es posible, la media y la mediana.
- b). - El cuartil primero (Q_1),
- c). - El cuartil setenta y cinco (Q_3)
- d). - El percentil, 40
- e). - El percentil, 80
- f). - Interprete la significancia de cada uno.
- g). - Realice un gráfico e incluya las medidas de tendencia Central.
- h). - Interprete cada uno.

Ejercicio 7.

Representar gráficamente la moda e interpretar su significado, del ejercicio 5 y 6.

TRABAJO PRÁCTICO N°6

MEDIDAS DE DISPERSIÓN

- 1) Frecuencia absoluta de niveles de colesterol en la sangre de 1227 varones estadounidenses con edades entre 55 y 64 años, periodo de 1976 – 1980.

Niveles de colesterol (mg/100 ml)	N° de Varones n_i
80 – 119	5
120 – 159	48
160 – 199	265
200 – 239	458
240 – 279	281
280 – 319	128
320 – 359	35
360 – 399	7
TOTAL	1227

- Completar la tabla de distribución de frecuencias, construir el Histograma y el polígono de frecuencia.
- Calcular; Media aritmética, Mediana, Modo. Interpretar los resultados obtenidos.
- Calcular el Q_1 y Q_3 . Interpretar los resultados.
- Mostrar la gráfica de frecuencias acumuladas u ojiva de datos.
- Desviación de cada lectura.
- Desviación media.
- Varianza y desviación Típica.
- Coefficiente de variación.

- 2) Abajo aparece una distribución de frecuencia con un resumen de las presiones arteriales sistólicas en reposo de una muestra de 35 pacientes con isquemia cardiaca, o interrupción del flujo sanguíneo al corazón.

Presión arterial (mm Hg)	Número de pacientes
115 – 124	4
125 – 134	5
135 – 144	5
145 – 154	7
155 – 164	5
165 – 174	4
175 – 184	5
TOTAL	35

- Completar la tabla de distribución de frecuencias, construir el Histograma y el polígono de frecuencia.
- Calcular; Media aritmética, Mediana, Modo. Interpretar los resultados obtenidos.
- Calcular el Q_1 y Q_3 . Interpretar los resultados.
- Mostrar la gráfica de frecuencias acumuladas u ojiva de datos.
- Desviación de cada lectura.
- Desviación media.
- Varianza y desviación Típica.
- Coeficiente de variación.

TRABAJO PRÁCTICO N° 7

CORRELACIÓN LINEAL

Ejercicio 1.-

- (a). Construir una línea recta que aproxime los datos de la tabla.
- (b). Hallar la ecuación de esta recta

X	2	3	5	7	9	10
Y	1	3	7	11	15	17

Ejercicio 2.-

En el problema 1, hallar:

- (a). Y cuando $X = 4$;
- (b). Y cuando $X = 15$;
- (c). Y cuando $X = 0$;
- (d). Y cuando $X = 7,5$;
- (e). Y cuando $X = 0$;
- (f). el incremento de Y correspondiente al incremento unitario de X

Ejercicio 3.-

Hallar: a) la pendiente,

- b) la ecuación,
- c) la intersección de Y y;
- d) la intersección de X de la recta que pasa por los puntos (1, 5) y (4, -1)

Ejercicio 4.-

Hallar la ecuación de la recta que pasa por el punto (4, 2) y es la paralela a la recta $2X + 3Y = 6$

Ejercicio 5.-

- (a). Construir una línea recta que aproxime los datos de la Tabla
- (b). Hallar la ecuación de esta recta

X	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9

Ejercicio 6.-

Ajustar una recta de mínimos cuadrados a los datos del Problema 8 utilizando (a) X como variable independiente,

- (b) X como variable dependiente.

Ejercicio 7.-

La tabla muestra las respectivas alturas X y Y de una muestra de 12 padres y sus hijos primogénitos.

1. Construir un diagrama de dispersión.
2. Hallar la recta de regresión de mínimos cuadrados de Y sobre X
3. Hallar la recta de regresión de mínimos cuadrados de X sobre Y

Altura X del padre (pulg)	65	63	67	64	68	62	70	66	68	67	69	71
Altura Y del hijo (pulg.)	68	66	68	65	69	66	68	65	71	67	68	70

TRABAJO PRÁCTICO N° 8

PROBABILIDAD

- Determinar o estimar la probabilidad p de cada uno de los eventos siguientes:
 - Al lanzar una vez un dado obtener un número non (Número que no es divisible entre 2).
 - Al lanzar dos veces una moneda obtener por lo menos una cara.
 - Al sacar una carta de una baraja, bien barajada, con 52 cartas obtener un as, un 10 de diamantes o un 2 de espadas.
 - Al lanzar una vez un par de dados su suma sea siete.
 - Si en 100 lanzamientos de una moneda se obtuvieron 56 caras, en el siguiente lanzamiento obtener una cruz.
- Un experimento consiste en lanzar una moneda y un dado. Si E_1 es el evento en que se obtenga “cara” al lanzar la moneda y E_2 es el evento en que se obtenga “3 o 6” al lanzar el dado, expresar en palabras cada uno de los eventos siguientes:
 - \tilde{E}_1
 - \tilde{E}_2
 - E_1E_2
 - $\Pr\{E_1\tilde{E}_2\}$
 - $\Pr\{E_1|E_2\}$
 - $\Pr\{\tilde{E}_1|\tilde{E}_2\}$
- De una caja que contiene 6 pelotas rojas, 4 pelotas blancas y 5 pelotas azules se extrae, de manera aleatoria una pelota. Determinar la probabilidad de que la pelota extraída sea:
 - roja
 - blanca
 - azul
 - no sea roja
 - sea roja o blanco
- Un dado se lanza dos veces. Encontrar la probabilidad de obtener un 4, un 5 o un 6 en el primer lanzamiento y un 1,2,3 ó 4 en el segundo lanzamiento.
- De una baraja, bien barajada, con 52 cartas se extraen dos cartas. Encuentre la probabilidad de que las dos sean ases si la primera carta:
 - se devuelve a la baraja
 - no se devuelve a la baraja

UNIVERSIDAD NACIONAL DE FORMOSA

FACULTAD DE CIENCIAS DE LA SALUD

Carrera: Técnico en Laboratorio– Lic. en Bromatología Cátedra: Bioestadística

6. Encuentre la probabilidad de que en dos lanzamientos de un dado se obtengan por lo menos un 4.