

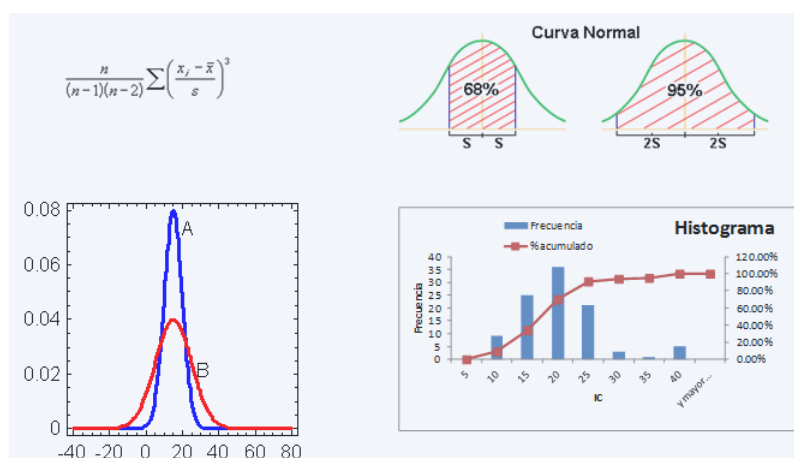
TENDENCIA CENTRAL, VARIABILIDAD y FORMA DE LA DISTRIBUCIÓN

De datos a información

25/05/2012

Conocimiento para todos

JORGE FALLAS



Índice

1. Introducción	1
2. Tendencia central	1
2.1 Moda y mediana	1
2.2 Media aritmética: simple y ponderada	4
2.3 Media geométrica y media armónica	5
2.4 Forma de la distribución y su relación con las medidas de tendencia central y asimetría	7
2.5 Propiedades de la moda, media, y mediana	9
3. Variabilidad	9
3.1 Ámbito y recorrido semiintercuartil	10
3.2 Desviación absoluta de la media	11
3.3 Varianza, desviación estándar y error estándar	12
3.4 Coeficiente de variación	13
3.5 Medidas de variabilidad: comparación y selección	14
3.6 Desviación estándar y la distribución normal	15
3.7 Curtosis	16
4. Estadística descriptiva utilizando XLStatistics	17
5. Resumen estadística descriptiva	18
6. Bibliografía	21
7. Ejercicios	22
Anexo 1: Estadística: software gratuito	26
Anexo 2: Licencia de “Creative Commons”	27
Anexo 2: Abreviaturas y equivalencias	28

El presente documento se distribuye bajo licencia CC BY-NC-SA de “Creative Commons” “reconocimiento-No comercial-Compartir bajo la misma licencia”; la cual permite a otros entremezclar, ajustar y construir con base en su trabajo para fines no comerciales, siempre y cuando se de crédito y licencia de sus nuevas creaciones, en los términos idénticos.

La información independientemente de lo costosa que haya sido crearla, puede ser replicada y compartida a un costo mínimo o nulo. -- Thomas Jefferson

1. Introducción

En el capítulo previo se indicó que una de las tareas de la estadística es ayudarnos a recabar datos fidedignos y confiables que posteriormente debemos ordenar, sistematizar y analizar con el propósito de comprender las relaciones entre los elementos que conforman la realidad económica, ecológica y social y de esta tomar decisiones informadas. En el presente documento trataremos otro aspecto de la estadística descriptiva: las medidas de tendencia central, variabilidad y forma.

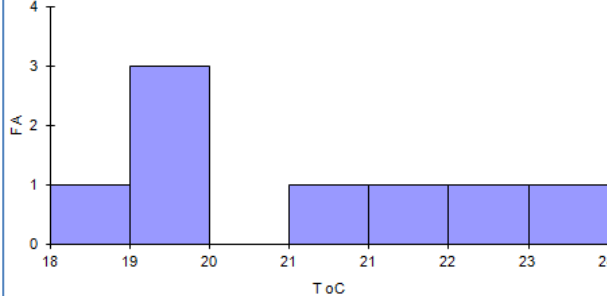
2. Tendencia central

La media, moda y mediana se conocen como promedios o medidas de posición central. Al describir una serie estadística no es recomendable utilizar la palabra "promedio" ya que, para distribuciones no simétricas, cada estimador tiene un valor diferente. Las medidas de tendencia central nos indican, en forma clara y concisa, la ubicación o posición del valor central o el punto de gravedad de la serie estadística. Cada una tiene sus ventajas y limitaciones y al utilizarlas debemos considerar, en primera instancia, la escala de medición de los datos.

2.1 Moda y mediana

La **moda (Mo)** es el **valor o característica más frecuente de la serie estadística**. Esta definición es válida tanto para series cuantitativas como para series cualitativas. A continuación se ilustra el concepto con dos ejemplos:

EJEMPLO A. Datos cuantitativos

¿Cuál es la moda para los siguientes valores de temperatura en °C)? 19,19,22,18,21,19,23,24	 <table border="1"> <caption>Data for Bar Chart</caption> <thead> <tr> <th>T °C</th> <th>FA</th> </tr> </thead> <tbody> <tr><td>18</td><td>1</td></tr> <tr><td>19</td><td>3</td></tr> <tr><td>20</td><td>0</td></tr> <tr><td>21</td><td>1</td></tr> <tr><td>22</td><td>1</td></tr> <tr><td>23</td><td>1</td></tr> <tr><td>24</td><td>1</td></tr> </tbody> </table>	T °C	FA	18	1	19	3	20	0	21	1	22	1	23	1	24	1
T °C	FA																
18	1																
19	3																
20	0																
21	1																
22	1																
23	1																
24	1																
El valor 19 °C se repite tres veces en la serie estadística y por lo tanto es el más frecuente y representa la moda de la serie.	La grafica de frecuencia absoluta muestra claramente que el valor 19 °C es el más frecuente.																

Ejemplo B. Datos cualitativos

¿Cuál es la moda para las siguientes especies?

roble, encino, **roble**, **roble**, jaúl, **roble**, **roble**, jaúl, lloró, mayo, **roble**

La especie **roble** se repite seis veces y por lo tanto es la moda de la serie.

En una curva de distribución de frecuencia para datos agrupados, la moda corresponde al punto más alto de la distribución. Para datos agrupados la moda se calcula utilizando la siguiente fórmula:

$$Mo = Li + (d1/(d1+d2)) \times Ic \quad (1)$$

donde:

Mo = moda para datos agrupados

Li = límite real inferior de clase modal (la clase con mayor frecuencia)

d1 = frecuencia de clase modal menos frecuencia de la clase inmediata anterior

d2 = frecuencia de clase modal menos frecuencia de la clase inmediata superior.

Ic = intervalo de la clase modal

Para una serie estadística agrupada, con los siguientes datos, la moda es:

clase modal = 15,0 - 19,9
 Li = 14,95 cm.
 d1 = 37 - 25 = 12 cm.
 d2 = 37 - 21 = 16 cm.
 Ic = 5 cm.
 $Mo = 14,95 + (12/(12+16)) \times 5$
Mo = 17,1 cm.

La moda tiene escasa o ninguna utilidad desde el punto de vista matemático o de prueba de hipótesis sobre la distribución que caracteriza a los datos. Además, una distribución puede presentar más de un moda (distribución multimodal) o no tener ninguna como se ilustra en la figura 1.

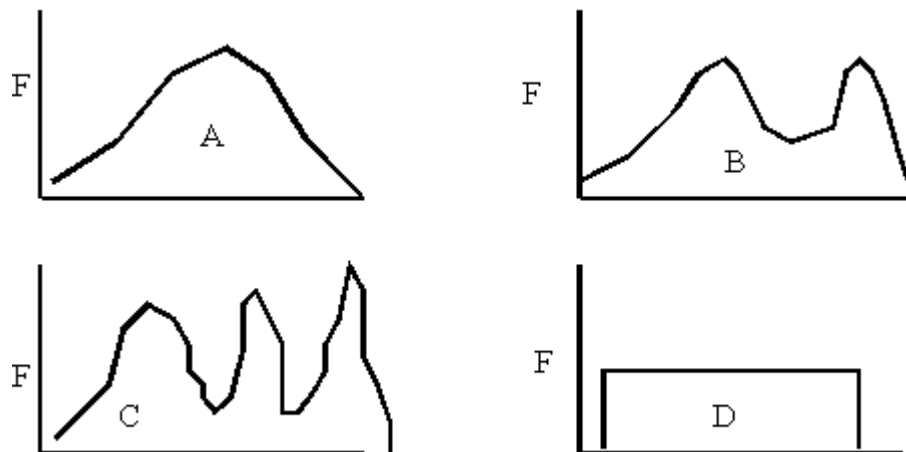


Figura 1: Distribución con una moda (A), bimodal (B), multimodal (C) y sin moda (D).

Entre las ventajas de la moda tenemos que es fácil de calcular y que es la única medida de posición central apropiada para describir atributos (series cualitativas). No es posible determinar ni la mediana ni la media para atributos tales como sexo, especie, tipo de bosque o color de las hojas.

La **mediana (Md)** es el valor que divide la serie estadística ordenada en dos partes iguales, o sea, 50% de las observaciones son mayores y 50% menores que la mediana. La observación mediana se obtiene utilizando la siguiente expresión:

$$Om = (n + 1) / 2 \quad (2)$$

Para los siguientes valores de diámetro (cm), la mediana es la observación número tres: $(5+1)/2=3$, cuyo valor es 8 cm:

5, 6, **8**, 10, 11

Cuando el número de observaciones es par, como por ejemplo 10, 20 o 40, la mediana corresponde al punto medio de los valores medios de la serie por ejemplo: $(10+1)/2 = 5,5$. Para la siguiente serie estadística la mediana es el valor de la observación 5,5 o sea: $(20+21)/2 = 20,5$.

17, 18, 19, 19, 20, 21, 22, 28, 33, 34

Mediana

Para datos agrupados la mediana se calcula utilizando la curva de frecuencia acumulada o la siguiente expresión:

$$Md = Li + \{Ic \times [(0,5 * n) - facum.inf]\} / fmd \quad (3)$$

en donde:

Li = límite real inferior de la clase que contiene a la mediana (50% de las observaciones)

Ic = intervalo de la clase mediana

n = número de observaciones

fmd = frecuencia absoluta de la clase mediana

facum.inf = frecuencia acumulada por debajo de la clase mediana

Una desventaja de la mediana es su insensibilidad a valores extremos. Por ejemplo, las dos series estadísticas que se muestran a continuación tienen la misma mediana, aun cuando los valores que componen cada serie son muy diferentes.

Serie A

Serie B

1,5,7,9,11

1,2,7,10,20

En la figura 2 puede apreciarse que la mediana divide el área bajo la curva de distribución de frecuencia en dos secciones iguales. Las secciones son simétricas (iguales) cuando los datos provienen de una población con una distribución normal.

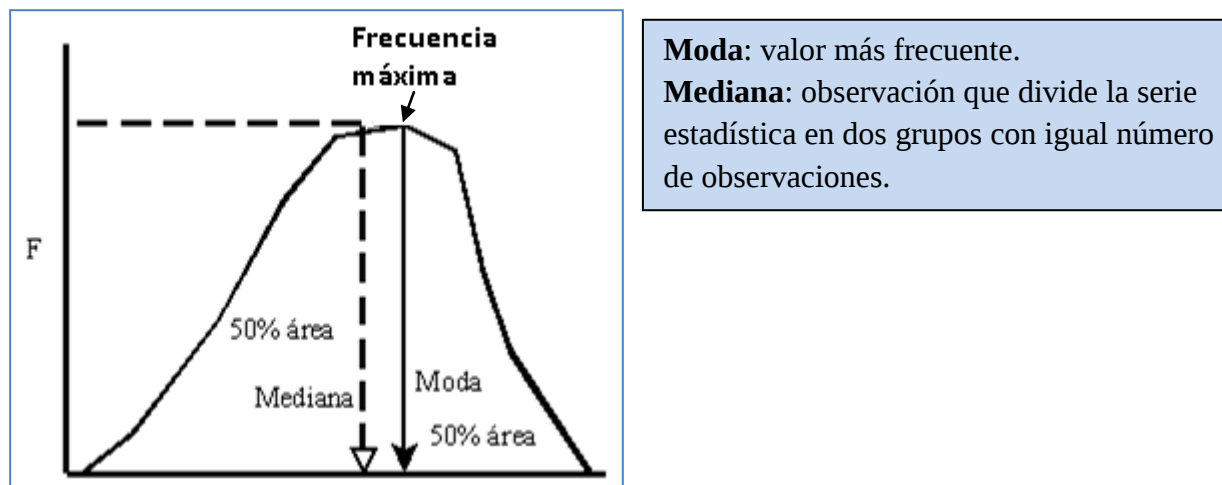


Figura 2: Medidas de tendencia central: moda y mediana.

2.2 Media aritmética: simple y ponderada

La media aritmética simple es posiblemente la medida de tendencia central de uso más frecuente. Dado un grupo de valores $x_1, x_2, x_3, \dots, x_n$; la media es la suma de los valores dividido entre el número de observaciones. Matemáticamente se expresa así:

$$\bar{X} = \sum (x_i) / n \quad (4)$$

en donde:

Σ = símbolo de sumatoria, indica la suma de los valores de x desde i hasta n .

x_i = valor de la observación i

n = número de observaciones

Para los siguientes valores de profundidad de suelo (cm): 5,1; 6,2; 8,4; 10,6 y 12,3; el promedio aritmético es:

$$\bar{X} = (5,1 + 6,2 + 8,4 + 10,6 + 12,3) / 5 = 8,5$$

$$\bar{X} = 8,5 \text{ cm.}$$

Nótese que el promedio se reporta como 8,5 y no como 8, 52; ya que los resultados deben tener la misma precisión que los datos originales.

La media ponderada es una variación de la media aritmética simple y se utiliza cuando las observaciones no tienen el mismo peso o importancia. Su fórmula es:

$$\bar{X} = \sum (W_i * X_i) \quad (5)$$

en donde: W_i es el peso o ponderación dado a la observación X_i .

Por ejemplo, si calculamos el volumen medio de tres parcelas de diferentes dimensiones, el peso o ponderación del volumen por parcela será una función de su área como se muestra a continuación:

Cuadro 1: Calculo de media ponderada.

Parcela	Área m ²	Volumen m ³		Ponderación W_i
		Caso 1	Caso 2	
A	500	10, 000	10, 000	500/ 3500 = 0,14
B	1000	20,000	10, 000	1000/ 3500 = 0,29
C	2000	10,000	20, 000	2000/ 3500 = 0,57
Total	3500	$\bar{X}_w = 12,900$	$\bar{X}_w = 15,700$	Suma $W_i = 1,0$

El efecto de la ponderación en el set de datos es evidente en los resultados obtenidos. La media aritmética simple es igual a 13,330 m³ [(10+20+10)/3] o sea un 3% superior que la media ponderada para el caso 1 A y un 15% menor que la media ponderada del caso 2. El efecto de la ponderación será mayor cuanto mayor sea la diferencia entre el valor de la observación y su respectivo factor de ponderación.

La media aritmética para datos agrupados es un tipo de media ponderada en donde el factor de ponderación es la frecuencia de cada clase. Matemáticamente se expresa así:

$$\bar{X} = \sum (f_i * m_i) / n \quad (6)$$

en donde: m_i representa el punto medio de la clase i y f_i su frecuencia.

La **media aritmética** es el **punto de balance de la distribución**; matemáticamente esto significa que la desviación de valores a su alrededor es siempre cero. Esto es, si calculamos la desviación de cada observación con respecto a la media, la suma de las desviaciones será igual a cero. Por esta razón si cambiamos uno de los valores que componen la serie estadística, la media también cambiará. En otras palabras, la media es sensible a la posición exacta de cada valor en la distribución. Esto es una ventaja con respecto a la moda y la media.

2.3 Media geométrica y media armónica

La media geométrica es igual a la raíz enésima de los productos de la serie estadística y por lo tanto sólo se puede calcular para valores positivos y diferentes de cero. Se utiliza con datos relativos (e.g. porcentajes) y sus valores están relacionados con la distribución log-normal. Su fórmula es:

$$\bar{x} = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 \cdot x_2 \cdots x_n} \quad (7)$$

El logaritmo de la media geométrica es igual a la media aritmética de los logaritmos de los valores de la variable.

$$\log \bar{X}_G = \Sigma (\log x_i) / n \quad (8)$$

El antilogaritmo de la expresión 8 es igual a la media geométrica.

EJEMPLO

Supongamos que la producción de frutos de la parcela A creció un 10% en el año 1, un 30% en el segundo año y en el año 3 decreció en un 20%. ¿Cuál fue el crecimiento medio en la producción de frutos de la parcela durante los tres años? La respuesta correcta no es la media aritmética porque la producción en el primer año debe multiplicarse por 1,10, en el segundo año debe multiplicarse por 1,30 y en el tercer año debe multiplicarse por 0,80 (hubo una reducción de 20% con respecto al valor de referencia). Dado que estamos trabajando con valores relativos, la media geométrica es el estadístico apropiado.

$$\bar{X}_G = (1,3 * 1,2 * 0,8)^{1/3} = 1,045$$

Media aritmética	1,066	crecimiento medio: 6.6%
Media geométrica	1,045	crecimiento medio: 4.5%

Cuando cada valor de la serie estadística no tiene la misma importancia (e.g. representan tamaños de muestra diferentes) se pueden adicionar pesos o ponderaciones (α_i) a cada valor como se muestra a continuación:

$$\bar{x} = \left(\prod_{i=1}^n x_i^{\alpha_i} \right)^{\frac{1}{\sum \alpha_i}} = (x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n})^{\frac{1}{\alpha_1 + \dots + \alpha_n}} \quad (9)$$

La media armónica es igual al número de observaciones dividido entre la sumatoria de los recíprocos de cada observación. Se utiliza para calcular medias de razones o tasas (e.g. velocidad por hora; peso por mes; longitud/día). Su fórmula es:

$$\bar{X}_A = n / \Sigma (1/ x_i) \quad (10)$$

EJEMPLO:

Supongamos que usted midió la distancia recorrida por un animal en 24 horas. Los resultados son los siguientes:

Cuadro 2: Cálculo de media armónica.

Día	Distancia (m)	Velocidad (m/hr)	Sobrestimación
1	4500	187.5	
2	5000	208.3	
3	3000	125.0	
4	2500	104.2	
5	3500	145.8	
Media aritmética		154.2	6.7%
Media armónica		144.5	Valor correcto
Media geométrica		149.3	3.2%

Observe que dada una serie estadística la media aritmética es mayor que la media geométrica y esta a su vez es mayor que la media armónica.

2.4 Forma de la distribución y su relación con las medidas de tendencia central y asimetría

Las características básicas de una distribución de frecuencia son: 1) **tendencia central** y 2) **variabilidad** y 3) **forma**.

Las medidas de tendencia central nos permiten determinar el punto central de la distribución, así como alrededor de cual valor se agrupan las observaciones y la importancia relativa de un valor con respecto a la media. En las secciones previas hemos visto que cada medida de tendencia central tiene un valor específico en una serie estadística. La ubicación o posición de cada estimador es el resultado de la forma de la distribución.

En distribuciones normales (unimodales), la moda, mediana y media tienen el mismo valor y por ende la misma posición en la curva de frecuencia. Esta distribución se denomina simétrica. Para distribuciones con asimetría positiva o hacia la derecha la moda y la mediana se ubican a la izquierda de la media y para distribuciones con asimetría negativa o hacia la izquierda a su derecha. La palabra sesgo se utiliza algunas veces para designar distribuciones asimétricas (Fig. 3).

La posición de la media y la mediana pueden utilizarse para determinar la presencia y dirección de la asimetría. En general, cuanto mayor sea la discrepancia entre ambos estimadores mayor será la intensidad de la asimetría. Por esta razón no siempre la media aritmética es el mejor estadístico para representar la serie estadística. Recordemos que cada estimador brinda diferente información y por lo tanto se debe seleccionar considerando las características de la serie estadística y los objetivos del análisis.

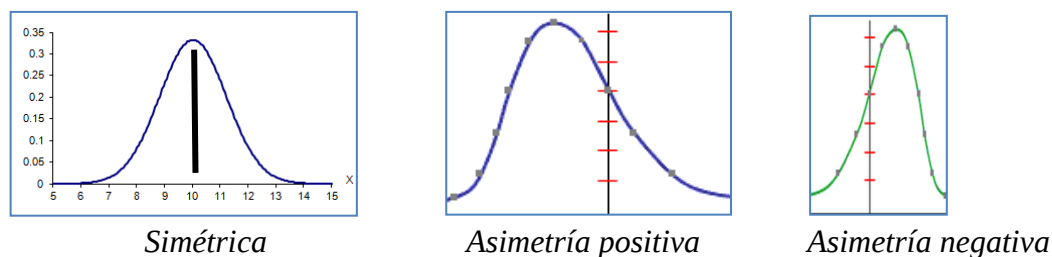


Figura 3: Distribuciones simétrica (normal) y asimétricas (no normal).

Prácticamente todas las distribuciones empíricas son asimétricas. El grado de asimetría es importante en la estadística inferencial ya que las pruebas paramétricas asumen que los datos provienen de una distribución normal (simétrica). Algunas distribuciones con fuerte asimetría son: ingreso familiar; diámetro normal (d) para bosques naturales no intervenidos; eventos hidrometeorológicos extremos (e.g. descarga máxima y mínima, precipitación máxima y mínima); y notas de exámenes muy fáciles o muy difíciles.

La distribución normal se caracteriza por su simetría o balance alrededor de su media. En otras palabras, si dibujamos una línea que conecte la media con el punto más alto de la distribución notaremos que las secciones a la derecha y a la izquierda de esta línea son idénticas. La distribución normal no es la única distribución simétrica (eg. rectangular). El cuadro 1 resume algunas medidas utilizadas para cuantificar la asimetría de una serie estadística.

Cuadro 3: Medidas de asimetría.

Medida de asimetría	Comentario
$\frac{n}{(n-1)(n-2)} \sum \left(\frac{x_j - \bar{x}}{s} \right)^3$	si $m_3 = 0$ distribución simétrica si $m_3 > 0$ asimétrica positiva si $m_3 < 0$ asimétrica negativa Ecuación utilizada por Excel y XLStatistics.
$A_p = \frac{\mu - moda}{\sigma}$	Coeficiente de asimetría de Pearson. Mide la desviación del set de datos con respecto a la simetría expresada como la diferencia entre la media y la moda con respecto a la desviación estándar. Sólo debe utilizarse con distribuciones uniformes, unimodales y moderadamente asimétricas. Para una distribución simétrica la media es igual a la moda y por lo tanto $A_p = 0$. Si $A_p > 0$ la distribución es asimétrica positiva.
$A_B = \frac{Q_{3/4} + Q_{1/4} - 2Me}{Q_{3/4} - Q_{1/4}}$	Coeficiente de asimetría de Bowley. Basado en la posición del primer cuartil (Q25), el segundo cuartil (la mediana, Q50) y el tercer cuartil (Q75). En una distribución simétrica el tercer cuartil estará a la misma distancia de la mediana que el primer cuartil. Por tanto $A_B = 0$. Si la distribución es positiva o a la derecha, $A_B > 0$.
$\gamma_1 = \frac{\mu_3}{\sigma^3}$	Coeficiente de asimetría de Fischer. Donde μ_3 = el tercer momento en torno a la media y σ = la desviación estándar. Si $\gamma_1 > 0$, la distribución es asimétrica positiva. Si $\gamma_1 < 0$, la distribución es asimétrica negativa. Para $\gamma_1 = 0$ la distribución es simétrica.

2.5 Propiedades de la moda, media, y mediana

La moda, mediana y media describen numéricamente la posición del valor central de la distribución. Al seleccionar una de las medidas de posición central se debe considerar el para qué y el porqué de la selección, así como la naturaleza de los datos. En general, la media aritmética simple es el promedio de uso más frecuente; le sigue la mediana la cual es especialmente útil para describir distribuciones muy asimétricas; y por último la moda, la cual es muy poco utilizada con datos cuantitativos pero la única que puede utilizarse con datos cualitativos.

Las principales características de la moda, mediana, y media se resumen en el cuadro 4. La derivación matemática se refiere a que el valor se obtiene de la interacción de todos los valores que componen la serie. La estabilidad muestral indica la estabilidad del estimador ante factores aleatorios (cambio del estimador entre muestras).

Cuadro 4: Características de la moda, mediana y media.

Característica	Medida de posición central		
	Moda	Mediana	Media
Derivación matemática	No	No	Si
Estabilidad muestral	Pobre	Regular	Excelente
Sensibilidad a valores extremos	No	Poco	Muy alta
Aplicación (escala de medición)	Nominal	Intervalo, razón	Razón, Intervalo
Nivel de inferencia	No paramétrico	No paramétrico	Paramétrico

3. Variabilidad

Las medidas de variabilidad expresan cuantitativamente el grado de dispersión o agrupamiento de los datos. Conocer la variabilidad de un set de datos es esencial para describirlo, compararlo, realizar pruebas estadísticas y determinar tamaños de muestra.

Por ejemplo, supongamos que tenemos dos instrumentos para medir alturas, A y B, y deseamos seleccionar el más preciso (el de menor variabilidad). Al realizar una serie de mediciones en un grupo de 25 árboles con una altura media de 15,0 metros, encontramos que la media aritmética para el instrumento A es 14,9 m y para el B 15,1 metros.

Basados en estos datos, podríamos seleccionar cualquiera de los dos instrumentos; ya que sus medias son muy similares y podríamos concluir que un instrumento es tan “bueno” como el otro. Sin embargo, la media no nos brinda información sobre la precisión del instrumento. Si además de la media observamos la distribución de alturas para cada instrumento (Fig. 4), seleccionaríamos el “A”; ya que muestra un menor grado de dispersión y por ende sería el más preciso.

La mayoría de los instrumentos indican su precisión en términos porcentuales (e.g. $\pm 3\%$, $\pm 5\%$), lo cual facilita su evaluación y comparación. El ejemplo anterior ilustra que la media por sí solo no es suficiente para caracterizar una serie estadística y que se requiere de otra medida que nos indique el grado de variabilidad de los datos.

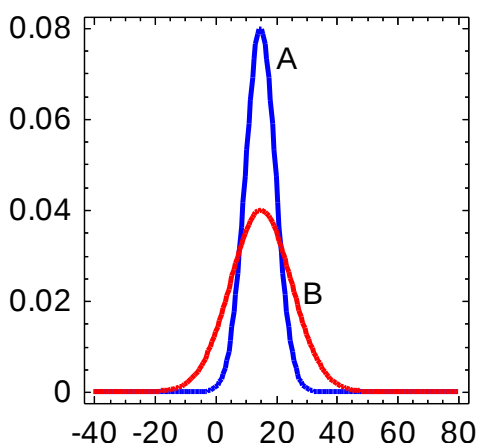


Figura 4: Distribución hipotética de 100 valores de altura medida con dos instrumentos. El instrumento A es preferible sobre el B por que posee menos variabilidad.

El concepto de variabilidad es uno de los aspectos claves sobre los cuales descansa la estadística inferencial; ya que permite estimar el error estadístico asociado a las mediciones. Las medidas de variabilidad expresan el grado de dispersión de los datos con respecto a su centro de gravedad.

3.1 Ámbito y recorrido semiintercuartil

El ámbito es igual a la diferencia entre el valor máximo y mínimo de la serie estadística y al igual que las otras medidas de variabilidad, caracteriza a la distribución en términos de distancia. El ámbito también se conoce como recorrido o rango, nombre que expresa muy claramente la idea de distancia.

El recorrido es muy fácil de calcular e interpretar. Su mayor desventaja es su poca sensibilidad a valores intermedios así como su alta sensibilidad a valores extremos en la serie estadística. Los valores extremos son especialmente importantes porque representan aspectos sobresalientes de los datos o errores de medición. A continuación se ilustra el cálculo del ámbito para tres series estadísticas.

Serie A: 5,6,7,8,9,10,11,12,13,14,15

Recorrido = 15-5=10

Serie B: 20,21,22,23,24,25,26,27,28,29,30

Recorrido = 30-20= 10

Serie C: 150,151,155,154,155,150,157,159,160

Recorrido = 160 – 150=10

El ejemplo anterior ilustra una de las limitaciones del recorrido: no es sensible a los valores intermedios de la serie estadística. El recorrido semiintercuartil se define como aquel valor equidistante entre el primer (Q25) y tercer cuartil (Q75). Matemáticamente se define así:

$$R_s = \frac{C_{75} - C_{25}}{2} \quad (11)$$

No debe confundir R_s , que expresa una distancia, con el valor de los diferentes cuartiles, los cuales expresan una posición a lo largo del eje X (Fig. 5).

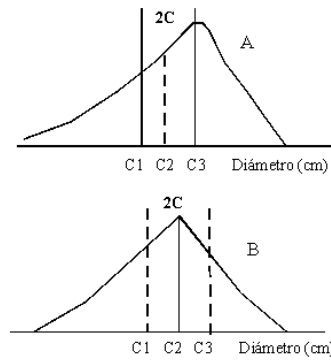


Figura 5: Recorrido intercuartil y semiintercuartil. A. Distribución asimétrica. B. Distribución simétrica.

El recorrido semiintercuartil es menos sensible al efecto de valores extremos, así como al efecto de distribuciones asimétricas; por esta razón es una excelente medida de variabilidad para describir este tipo de distribuciones. Al estar definido por la posición del primer y tercer cuartil, depende de la ubicación de la mediana. Su valor depende del número de observaciones por debajo de C_1 y por encima de C_3 ; sin embargo no nos indica cuán lejos están dichos puntos. Al aumentar el valor máximo de la serie el grado de dispersión de los datos aumenta; sin embargo al no afectar la posición de C_{25} y C_{75} , el valor del recorrido semiintercuartil no cambia.

3.2 Desviación absoluta de la media

La desviación absoluta de la media indica la distancia absoluta de una observación dada con respecto a la media aritmética simple. Simbólicamente se expresa así:

$$D_i = |x - \bar{x}| \quad (12)$$

La suma de las desviaciones absolutas es un índice que permite medir la variabilidad de un set de datos; su fórmula es:

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{N} \quad D_{\bar{x}} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{N} \quad (13)$$

Una característica esencial de toda medida de variabilidad es que su valor sea directamente proporcional al grado de dispersión de los datos, o sea cuanto mayor sea el grado de dispersión mayor debe ser el valor del estadístico y viceversa.

La desviación absoluta para datos agrupados es igual a:

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}|f_1 + |x_2 - \bar{x}|f_2 + \dots + |x_n - \bar{x}|f_n}{N} \quad D_{\bar{x}} = \frac{\sum_{i=1}^n |x_i - \bar{x}|f_i}{N} \quad (14)$$

Donde: f_1, f_2, f_n es igual a la frecuencia de la clase y X_1, X_2, X_n es igual al punto medio de cada clase.

3.3 Varianza, desviación estándar y error estándar

La varianza poblacional es una constante desconocida que se designa con el símbolo sigma cuadrado (σ^2). Matemáticamente, se define como la sumatoria de las desviaciones de la media al cuadrado entre el número de observaciones. Su fórmula es:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{N} \quad \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N} \quad (15)$$

La varianza para datos agrupados es igual a:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \dots + (x_n - \bar{x})^2 f_n}{N} \quad \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{N} \quad (16)$$

Para muestras pequeñas la fórmula 15 tiende a subestimar el valor de la varianza; ya que los valores más alejados de la media y por ende menos frecuentes, son sub representados en la muestra. Sustituyendo n por n-1 la ecuación 15, se transforma en S^2 , el estimador de la varianza poblacional.

$$S^2 = \frac{\sum (x - \bar{X})^2}{n - 1} \quad (17)$$

El efecto de esta corrección es importante cuando el tamaño de la muestra es pequeño. Por ejemplo, cuando $n=30$ la magnitud de la corrección es de aproximadamente un 3% de la varianza. La importancia práctica de la corrección se reduce conforme aumenta el tamaño de la muestra. Desde un punto de vista teórico, el término $n-1$ transforma S^2 en un estimador insesgado; o sea, en promedio es igual al valor poblacional.

La desventaja de la varianza como indicador de variabilidad es que su valor se expresa en unidades cuadradas. Por ejemplo, si la serie estadística “A” tiene una desviación con respecto a la media dos veces mayor que “B”; la varianza de “A” será cuatro veces mayor que la de “B”. Por esta razón la varianza es poco utilizada para describir la variabilidad en series estadísticas. La raíz cuadrada de la varianza se denomina desviación estándar o típica y es la medida de variabilidad de uso más frecuente en trabajos estadísticos. Su fórmula es:

$$S = (S^2)^{0,5} \quad (18)$$

La varianza y la desviación estándar reflejan el grado de agrupamiento o dispersión de los datos alrededor de la media. Cuando los valores tienden a agruparse alrededor de la media, las desviaciones son pequeñas y por ende también lo es la media de las desviaciones cuadráticas. Por otro lado, cuando las observaciones se alejan de la media, la desviación estándar aumenta. En resumen, la desviación estándar es una función de la distancia con que las observaciones se alejan de la media.

La desviación estándar es sensible a valores extremos, aun cuando su grado de sensibilidad es mucho menor comparada con el recorrido. Este aspecto debe tomarse en cuenta cuando se trabaje con distribuciones altamente asimétricas o con valores extremos.

La desviación estándar es también útil para determinar si las diferencias entre dos medias son grandes o pequeñas. Consideremos por ejemplo la diferencia en precipitación de las estaciones “C” y “D”: ¿Son 200 mm una diferencia importante? La respuesta podría ser: sí, no, o no sé y en general parecería bastante subjetiva. Desde un punto de vista estadístico se puede considerar como relevantes aquellas diferencias mayores que $0,5S$ (S = desviación estándar). En este caso la diferencia es aproximadamente $0,9S$ y por ende es importante. Recordemos que para comparar las medias de dos distribuciones, sus desviaciones estándares deben ser muy similares o de lo contrario se debe utilizar la desviación estándar de la media de las varianzas para realizar la comparación. Este concepto se desarrollará en mayor detalle en el capítulo sobre prueba de hipótesis.

El error estándar es directamente proporcional a la desviación estándar e inversamente proporcional a la raíz cuadrada del número de observaciones en la muestra. Su expresión matemática es:

$$E.S. = S / (n)^{0,5} \quad (19)$$

o si lo desea también se puede expresar como un porcentaje con respecto a la media como se muestra a continuación:

$$E.S. (\%) = \frac{E.S.}{\bar{X}} * 100 \quad (20)$$

El error estándar se utiliza con frecuencia como un índice de la precisión con que se estima la media aritmética simple. Además, es útil para determinar el tamaño de muestra requerido para obtener una precisión dada.

3.4 Coeficiente de variación

El coeficiente de variación en porcentaje es igual a la desviación estándar dividida entre la media y multiplicado por cien. Su fórmula es:

$$CV \% = (S/\bar{X}) * 100 \quad (21)$$

El coeficiente de variación se utiliza para comparar la variabilidad de dos o más series estadísticas sin importar los valores de la media o de la desviación estándar. La desviación estándar también se puede utilizar para comparar series estadísticas siempre y cuando las medias de las series sean muy similares. Esto se debe a que la desviación estándar expresa variabilidad con respecto a la media. Los siguientes datos ilustran este punto.

Cuadro 5: Variabilidad expresada como desviación estándar y coeficiente de variación.

Estación	Pt anual (cm)	S_A (cm)	S_B (cm)	CV_A (%)	CV_B (%)
A	230	26	100	11	43
B	540	24	230	4	43
C	500	23	100	5	20
D	520	24	230	5	44

Si comparamos la variabilidad de las estaciones “A” y “B” utilizando la desviación estándar S_A , concluiríamos que la precipitación media anual en la estación A es tan variable como en la estación “B” (la diferencia es de tan solo 2 cm). Por otro lado, si observamos los respectivos coeficientes de variación concluiríamos que “A” es más variable que “B” (11% Vs 4%). ¿Por qué son nuestras conclusiones contradictorias? La respuesta es que ambas estaciones tienen promedios muy diferentes (230 cm Vs. 540 cm).

Ahora bien, si analizamos la desviación estándar S_B de las estaciones “A” y “B” concluiríamos que la estación “B” presenta un mayor grado de variabilidad (203 cm Vs 100 cm). Sin embargo, el coeficiente de variación no apoya nuestra conclusión e indica que ambas estaciones tienen el mismo grado de variabilidad (43%).

Volviendo a nuestro ejemplo, si comparamos la variabilidad de las estaciones “C” y “D” llegaríamos a la misma conclusión indistintamente del estadístico que utilizemos en la comparación. Esto se debe a que las medias de ambas estaciones son muy similares y por lo tanto tenemos un denominador común. La discusión anterior ilustra la importancia de seleccionar el estadístico adecuado para expresar lo que se pretende cuantificar y comparar. Cuando se compara la variación de dos o más series estadísticas se recomienda utilizar el coeficiente de variación y no la desviación estándar.

3.5 Medidas de variabilidad: comparación y selección

De los estadísticos estudiados en las secciones anteriores, la desviación estándar y el coeficiente de variación son las medidas de dispersión de uso más frecuente. Le sigue, por su facilidad de interpretación y uso, el recorrido, y en último lugar tenemos al recorrido semiintercuartil.

La media y la desviación estándar comparten propiedades similares, además la segunda se basa en las desviaciones con respecto a la media. Por lo tanto, ambas se deben utilizar simultáneamente. La mediana y el recorrido semiintercuartil con frecuencia se utilizan conjuntamente; ya que ambas medidas se basan en cuantiles.

Para muchas distribuciones, incluyendo la normal, el recorrido es directamente proporcional al número de observaciones. Esta característica debe considerarse cuando se compare la variabilidad en distribuciones con un número diferente de observaciones. El cuadro 6 resume y compara las características más importantes de las medidas de dispersión estudiadas en el presente capítulo.

Cuadro 6: Comparación de características para medidas de variabilidad

Característica	ESTADÍSTICO		
	Recorrido	Recorrido semiintercuartil	Desviación estándar
Frecuencia de aplicación	alguna	muy poca	muy frecuente
Aplicación en estadística inferencial	muy poca	muy poca	muy frecuente
Estabilidad muestral	pésima	buena	excelente
Uso con pocas observaciones o distribuciones asimétricas	puede llevar a conclusiones erróneas	Se debe interpretar con precaución	adecuado
Asociado a medida de posición central	ninguna	mediana	media aritmética simple
Uso con distribuciones abiertas	no	adecuada	no
Sensible al tamaño de la muestra	sí	no	no

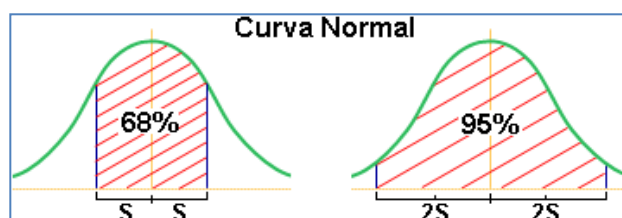
3.6 Desviación estándar y la distribución normal

En las secciones previas se ha definido el concepto de variabilidad y se han presentado varios estadísticos que la cuantifican. Además, hemos dicho que la desviación estándar representa el grado de variabilidad de una serie estadística. A pesar de lo anterior, es también probable que todavía nos preguntemos: ¿Qué representa una desviación estándar? No existe una respuesta simple y concreta a esta pregunta y la única forma de interiorizar el concepto es aplicándolo bajo diferentes contextos y cuantas veces sea posible. Para facilitar su interpretación, como una unidad de medición de distancia, examinaremos su uso en la curva de distribución normal, la distribución de mayor uso en estadística. Conforme nos alejamos de la media hacia la derecha o hacia la izquierda sucede lo siguiente:

$\bar{X} \pm 1S$ contiene aproximadamente 68% de las observaciones

$\bar{X} \pm 2S$ contiene aproximadamente 95% de las observaciones

$\bar{X} \pm 3S$ contiene aproximadamente 99,7% de las observaciones



Por ejemplo, si una muestra de diámetros de un bosque tienen una distribución aproximadamente normal con una media de 66 cm y una desviación estándar de 3,0 cm; entonces se esperaría que el 68% de los árboles se encontraran en el ámbito 63-69 cm.

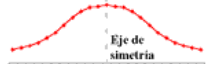
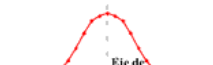
$$\bar{X} + 1S \quad 66 + 3 = 69 \text{ cm}$$

$$\bar{X} - 1S \quad 66 - 3 = 63 \text{ cm}$$

Utilizando el mismo razonamiento podemos decir que aproximadamente 95% de los árboles tienen un diámetro entre 60 y 72 cm y que el 99,7% tendrán un diámetro entre 57 y 73 cm. Las distribuciones para datos experimentales, en la mayoría de los casos, sólo aproximan la distribución normal y por lo tanto las afirmaciones anteriores deben aplicarse a tales distribuciones con precaución.

3.7 Curtosis

La distribución de frecuencia puede caracterizarse en cuanto a su forma, simetría y grado de curtosis. Los primeros dos términos se discutieron en secciones previas, a continuación se trata brevemente el tercero. La palabra curtosis se utiliza para referirnos al grado de achatamiento o agudeza de la distribución de frecuencia; sin embargo expresa realmente la relación entre la altura del centro de la distribución y la longitud y tamaño de sus colas. Según el grado de achatamiento o prominencia de la distribución de frecuencia, la misma puede clasificarse como: (ver figura 6):

 Platicúrtica	Las distribuciones platicúrticas son relativamente planas y de colas cortas. Presentan un reducido grado de concentración de observaciones alrededor de los valores centrales de la serie estadística.
 Mesocúrticas	Las distribuciones mesocúrticas son aquellas que poseen un moderado grado de curtosis; la distribución normal pertenece a esta categoría. Presentan un grado de concentración medio alrededor de los valores centrales de la serie estadística.
 Leptocúrticas	Las distribuciones leptocúrticas son relativamente puntiagudas y de colas largas. Presentan un alto grado de concentración de las observaciones alrededor de los valores centrales de la serie estadística.

El coeficiente de curtosis para una distribución normal es cero; su fórmula es:

$$g_2 = \frac{(1/n) * \sum (x_i - \bar{x})^4 * n_i}{((1/n) * \sum (x_i - \bar{x})^2 * n_i)^2} - 3 \quad (21)$$

Un coeficiente menor que 0 indica que la curva de distribución de frecuencia es plana y con colas cortas (platicúrtica); en tanto que cuando el valor es mayor que 0, indica que la curva es puntiaguda o tienen colas relativamente largas (leptocúrtica).

4. Estadística descriptiva utilizando XLStatistics

1. Active el complemento XLStatistics
2. Abrir el archivo d_h_jaul.xlsx
3. Marcar la columna diámetro

	A
1	diámetro (cm)
2	40
3	10
4	33
5	45
6	33
7	41

4. Hacer un clic sobre SLStatistics y seleccionar 1Num

1Num

5. Hacer un clic sobre *Data & Description*

Information	Data & Description	Tests	Extra Tools
-------------	--------------------	-------	-------------

6. Estadísticos descriptivos para la variable diámetro

Numerical Summaries for diámetro (cm)								
Number	46			Kurtosis	-0.905		Min	10
Mean	35.65217		10	% Tr mean	35.15		Q ₁	22.25
St Dev	15.65065			StdErr Mean	2.307563		Median	33.5
Coeff of Var	0.438982						Q ₃	45
Skew	0.278263						Max	68

Tamaño de muestra: 46

Media aritmética simple: 35,65 cm

Desviación estándar: 15,65 cm

Coefficiente de variación: 0, 43 (43%)

Asimetría: 0,278 (distribución con leve asimetría positiva)

Curtosis: -0,905 (distribución platicúrtica)

Media recortada (se elimina el 10% de los valores extremos). Usted puede cambiar dicho valor (por ejemplo eliminar solo el 5%): 35,15cm

Error estándar de la media: 2,30 cm

Valor mínimo: 10 cm

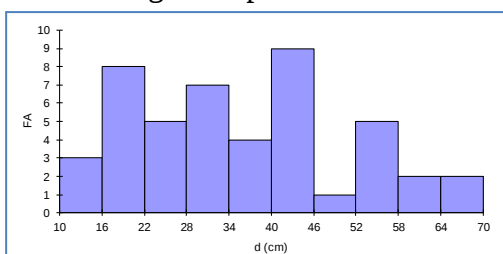
Cuartil 1 (Q25): 22,25 cm

Mediana (Q50): 33,5 cm

Cuartil 3 (Q75): 45 cm

Valor máximo: 68 cm

6. Análisis gráfico para la variable diámetro.



Histograma: # Clases 10, mínimo: 10 cm, máximo: 70 cm.

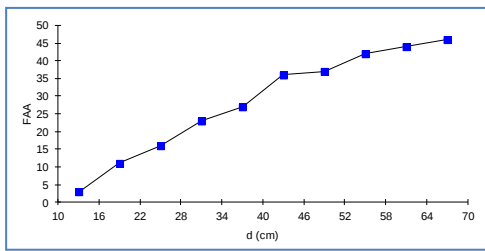


Gráfico de frecuencia absoluta acumulada (FAA)

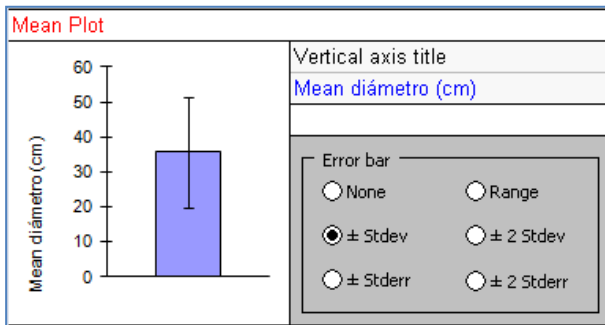


Gráfico de media y desviación estándar.

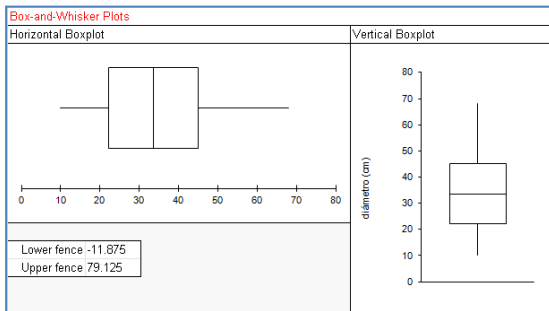


Gráfico de cajas (Box-Whisker).

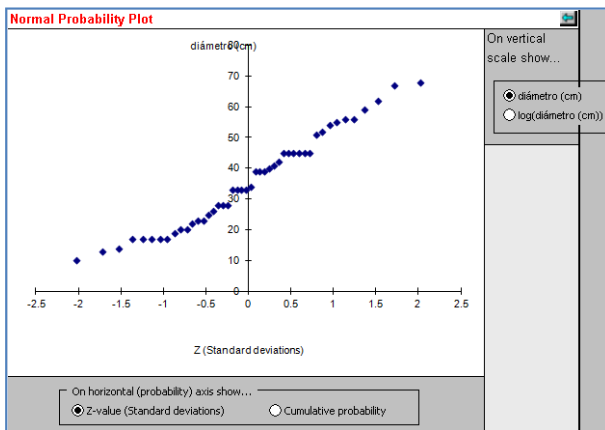


Gráfico de probabilidad normal.

5. Resumen estadística descriptiva

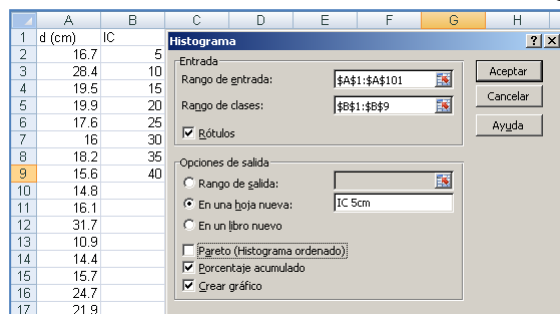
La estadística descriptiva comprende:

1. Crear tablas de frecuencia
2. Elaborar gráficos
3. Calcular estadísticos básicos.

Aun cuando ni Excel ni la versión gratuita “Hoja de cálculo” de “Open Office”¹ son paquetes estadísticos, ofrecen excelentes herramientas para crear gráficos y elaborar tablas de frecuencia. A continuación se muestra cómo crear tablas de frecuencia en Excel y cuáles gráficos poseen dichos programas.

Tablas de frecuencia e histogramas

Seleccione Datos, análisis de datos e Histograma



Seleccione las celdas que contienen los datos que desea analizar (Rango de entrada).

Seleccione las celdas que contienen las clases (Rango de Clases).

Rótulos: Seleccione esta casilla si su rango de entrada y su rango de clases incluye la etiqueta o nombre de la columna (Ej. D (cm) e IC).

Porcentaje acumulado: crea frecuencia acumulada.

Productos

Clases con intervalos de 5cm

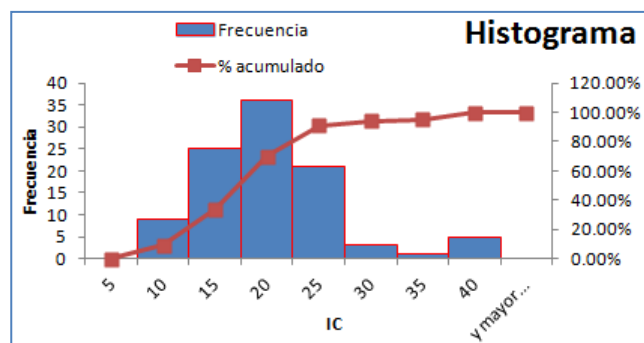
Frecuencia absoluta

Frecuencia acumulada *menor que* (límite superior)

Crear gráfico: El programa crea histograma y línea de frecuencia acumulada

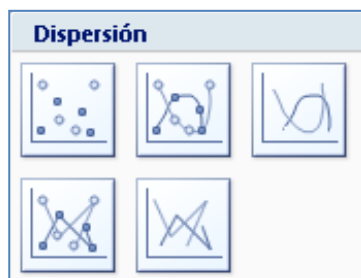
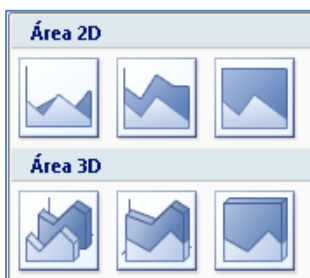
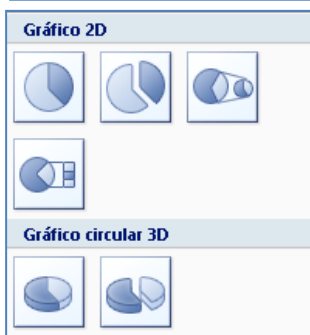
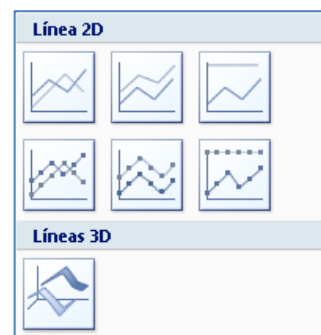
IC	Frecuencia	% acumulado
5	0	0.00%
10	9	9.00%
15	25	34.00%
20	36	70.00%
25	21	91.00%
30	3	94.00%
+35	1	95.00%
40	5	100.00%
y mayor...	0	100.00%

¹ <http://www.openoffice.org/product/>



Histograma de frecuencia absoluta y frecuencia acumulada. **Nota:** Si lo desea puede modificar los colores, fuente y número de decimales.

Menú de gráficos



6. Bibliografía

David H. A. 1998. Early sample measures of variability. Statist. Sci. Volume 13, Number 4: 368-377. Visitado 29 mayo 2012. Disponible en:

http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=euclid.ss/1028905831

Dinov, Ivo D. 2006. Statistics Online Computational Resource. Journal of Statistical Software 16 (1): 1–16. Visitado en 23 mayo 2012. Disponible en <http://www.jstatsoft.org/v16/i11/paper>

McDonald, J.H. 2009. Handbook of Biological Statistics (2nd ed.). Sparky House Publishing, Maryland. Last revised August 18, 2009. Visitado en 23 mayo 2012. Disponible en <http://udel.edu/~mcdonald/statintro.html>

Minium, Eduard W. y Clarke, Robert B. 1998. Elements of statistical reasoning. Wiley; 2 ed. 512 p.

Remington D. Richard y M. Anthony Schork. 2010. Statistics with applications to the biological and health sciences. CreateSpace. 3 ed. 496 p.

Steel, Robert G.D. y J.H. Torrie and Dickey David A. 1996. Principles and Procedures of Statistics: A Biometrical Approach. McGraw-Hill. 672 p.

7. Ejercicios

1 Defina y/o explique los siguientes términos:

moda, mediana, media, distribución simétrica, distribución asimétrica, sesgo positivo, sesgo negativo, Mo, Md, variabilidad, recorrido, recorrido semiintercuartil, recorrido intercuartil, varianza, desviación estándar, estabilidad muestral, coeficiente de variación, error estándar, curtosis.

2. Para los siguientes valores calcule: media aritmética, mediana y moda. ¿Exhiben los resultados algún patrón?

- A) 5,6,1,3,9,8
- B) 3,10,10,9,7,3
- C) 55,75,29,30,40,55,55,50
- D) 20,25,15,10,9,8,13
- E) 4,3,5,4,3,9,12,20,21,22

3. Dado un polígono de frecuencia y una ojiva, ¿cuál es la medida de tendencia central que se puede obtener con mayor facilidad?

4. Para los datos del archivo d_h_jaul.xlsx calcule para el diámetro (cm) y la altura total (m): media aritmética, mediana, moda, recorrido semiintercuartil, desviación estándar, coeficiente de variación.

5. ¿A qué tipo de distribución corresponden los siguientes promedios?

- A) $\bar{X}=52\text{ m}$ Md=58 m Mo=60 m
- B) $\bar{X}=105\text{gr}$ Md=96gr Mo=93gr
- C) $\bar{X}=78\text{ kg}$ Md=78 kg Mo=78 kgr
- D) $\bar{X}=340$ Md=320 Mo=310

6. ¿Es correcta la siguiente afirmación: "Más de la mitad de los costarricenses tienen un salario inferior al ingreso medio"? ¿Por qué?

7. ¿Cuál medida de tendencia central utilizaría usted para describir los siguientes sets de datos:

- A) Subgrupos de suelos
- B) Altura total de bosque tropical
- C) Producción de frutos por árbol
- D) Color de las hojas
- E) Densidad (especies)
- F) Diámetro de bosque natural
- G) Diámetros de plantación forestal

8. ¿Por qué la media es la medida de tendencia central de uso más frecuente?
9. Utilizando los datos del archivo diámetros.xlsx seleccione 5 muestras al azar con reemplazo de tamaño 15(puede utilizar XLStatistics **Sample Selection** **SampSel**). Calcule para cada una la media aritmética, recorrido, recorrido semiintercuartil, desviación estándar, coeficiente de asimetría, coeficiente de variación, mínimo, máximo. Evalúe la estabilidad muestral de cada uno de los estimadores utilizando la escala de Likert.
10. Un ecólogo registró el peso seco de 30 plántulas de cola de pavo (archivo peso_seco.xlsx).
- Elabore una tabla de frecuencia y su respectivo histograma. Elija el intervalo de clase.
 - Calcule la media aritmética simple, mediana, moda, desviación estándar y coeficiente de variación. Evalúe la variabilidad e la serie estadística utilizando la escala de Likert.
 - ¿Qué tipo de distribución muestran los datos? ¿Son los datos normales? Argumente su respuesta.
 - ¿Cuál grafica podría utilizar para evaluar visualmente la normalidad de los datos?
11. ¿Por qué es necesario acompañar a una medida de posición central (promedio) una medida de variabilidad o dispersión? Brinde 3 ejemplos donde pueda apreciarse este aspecto.
12. Utilizando el archivo ppt_mm.xls calcule: moda, mediana, media aritmética, recorrido, desviación semiintercuartil, desviación estándar, coeficiente de variación y error estándar para cada una de las estaciones. ¿Cuál medida de variabilidad utilizaría usted para describir la dispersión de la precipitación en cada una de las de las estaciones? ¿Por qué? ¿Cuál para comparar la variabilidad entre estaciones? Elabore un gráfico que muestre la media y desviación estándar de ambas estaciones. Elabore un gráfico de cajas para ambas estaciones. Evalúe la variabilidad de la serie estadística utilizando la escala de Likert.
14. Dados los siguientes valores:
- A) 10, 10, 10, 10, 10, 10
- B) 6, 6, 8, 10, 10
- C) 6, 8, 10, 12,10
- D) 1004, 1006, 1010, 1012, 1014
- E) 90, 80, 2, 1, 5
- A. ¿Cuál serie estadística muestra menos variabilidad? Ordene los datos según su orden decreciente de variabilidad (inspección visual).
- B. Para cada grupo calcule: promedio, desviación estándar, y coeficiente de variación. Ordene los datos según su orden decreciente de variabilidad.
- C. Compare los resultados obtenidos en A y B.

- D. Sugieren sus resultados que existe alguna relación entre la media y el grado de variabilidad de los datos. Explique.
15. Suponga que se obtienen 2 muestras al azar (1000 m^2) de un bosque húmedo tropical. Para cada muestra se cuantifica diámetro a la altura del pecho (d) y la altura total; luego se calcula el recorrido, el recorrido semiintercuartil, la desviación estándar y el error estándar. ¿Cuáles medidas de variabilidad esperaría usted que fuesen similares? ¿Cuáles muy diferentes? ¿Por qué?
16. Un investigador selecciona y pesa 100 semillas de *Ceiba pentandra*; luego calcula la moda, mediana, media, recorrido, y desviación estándar para sus datos. Posteriormente se da cuenta de que su valor máximo es erróneo y que es algo mayor que el valor utilizado en los cálculos. ¿Cuál es el impacto de este error en cada uno de los estadísticos calculados? Posteriormente, el investigador se da cuenta que uno de los valores por debajo de la media es aún menor. ¿Cuál(es) estadísticos serán afectados por esta corrección?
17. La distribución diamétrica para Ceiba pentandra indica que la clase superior incluye diámetros mayores o iguales que 60 cm. Si se desea calcular un índice de variabilidad, cual recomendaría usted y por qué?
18. ¿Por qué la media y la desviación estándar deban analizarse conjuntamente?
19. Los siguientes estadísticos son reportados para una serie estadística: media= 80, S = 6 y recorrido = 72. ¿Qué opina usted sobre la veracidad de los estadísticos? (¿esperaría usted estos resultados?).
21. El grupo "Estadísticos Asociados" desea conocer la eficiencia de dos métodos para enraizar estacas de pochote. Se muestran datos de altura (cm) al cabo de 10 meses.

	tratamiento A	tratamiento B
Media	100	110
S	20	7
Distribución	aprox. normal	aprox. normal

¿Cuál tratamiento considera usted más efectivo? ¿Es la diferencia en medias suficientemente grande como para ser considerada como importante? ¿Cuál diferencia es aparentemente más importante?

20. Seleccione tres artículos científicos y para cada una de ellos contabilice los estadísticos descriptivos utilizados. Presente sus resultados en forma tabular y gráfica.

21. Analice uno de los periódicos dominicales de mayor circulación. ¿Cuáles son los estadísticos descriptivos más utilizados? En su criterio ¿a qué se debe esto? (Nota: indicar periódico y fecha utilizada).

22. Compare los estadísticos descriptivos que ofrece:

Excel

BioEstat. (<http://www.mamiraua.org.br/downloads/programas.>)

Instat (<http://www.reading.ac.uk/ssc/n/software.htm.>)

Infostat (<http://www.infostat.com.ar/>)

Mstat (<http://www.mcardle.wisc.edu/mstat/#win>)

STATGRAPHICS Centurion (http://www.statgraphics.com/statgraphics_centurion.htm)

SYSTAT (<http://www.systat.com/>)

Anexo 1: Estadística: software gratuito

Si usted desea explorar programas estadísticos gratuitos, le recomiendo visitar los siguientes sitios.

[BioEstat](http://www.mamiraua.org.br/downloads/programas). Análisis estadístico para Windows y Mac. Estadística descriptiva e inferencial paramétrica y no paramétrica, análisis de poder. Interfaz en español. Manual en Portugués. <http://www.mamiraua.org.br/downloads/programas>.

[Diversity](http://www.reading.ac.uk/ssc/n/software/diversity/Diversity.html). Complemento para Excel que permite calcular índices de diversidad. Una buena referencia introductoria al tema es el libro “[Measuring Biological Diversity](#)” de Anne E. Magurran (2003) publicado por Blackwell Science. <http://www.reading.ac.uk/ssc/n/software/diversity/Diversity.html>

[Infostat](http://www.infostat.com.ar/) Análisis estadístico para Windows. Estadística descriptiva e inferencial paramétrica y no paramétrica. La versión en español es muy valorada por los usuarios, especialmente por los estudiantes. Una propiedad casi única de InfoStat es su habilidad de conectarse con R, una plataforma de desarrollo de algoritmos estadísticos de dominio público de gran crecimiento. <http://www.infostat.com.ar/>.

[Instat](http://www.reading.ac.uk/ssc/n/software.htm) Análisis estadístico para Windows. Estadística descriptiva e inferencial paramétrica y no paramétrica. Modulo para aplicaciones climáticas. <http://www.reading.ac.uk/ssc/n/software.htm>.

[LazStats](http://www.statprograms4u.com/LazStats_Features.htm) Análisis estadístico para Windows. Estadística descriptiva e inferencial paramétrica y no paramétrica. http://www.statprograms4u.com/LazStats_Features.htm.

[MacAnova](http://www.stat.umn.edu/macanova/macanova.home.html) Análisis estadístico para Mac, Windows y Linux. Estadística descriptiva e inferencial paramétrica y no paramétrica, análisis de poder. <http://www.stat.umn.edu/macanova/macanova.home.html>.

[Mstat Windows Mac OSX Linux](http://www.mcardle.wisc.edu/mstat/#win) Análisis estadístico para Windows, Mac y Linux. Estadística descriptiva e inferencial paramétrica y no paramétrica. <http://www.mcardle.wisc.edu/mstat/#win>

[OpenEpi](http://www.openepi.com/Menu/OpenEpiMenu.htm) Software especializado en el análisis de datos epidemiológicos. Estadística descriptiva e inferencial paramétrica y no paramétrica <http://www.openepi.com/Menu/OpenEpiMenu.htm>

[OpenSTat](http://www.statprograms4u.com) Análisis estadístico para Windows. Estadística descriptiva e inferencial paramétrica y no paramétrica. <http://www.statprograms4u.com>

[PAST](http://folk.uio.no/ohammer/past/) Análisis estadístico univariado, multivariado, índices de diversidad. Estadística descriptiva e inferencial paramétrica y no paramétrica. WXP, Vista y W7. <http://folk.uio.no/ohammer/past/>

[PSPP](http://www.gnu.org/software/pspp/). Este es un programa para el análisis estadístico, su funcionalidad es similar al programa comercial SPSS <http://www.gnu.org/software/pspp/>

[Remuestreo](http://www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html) Software para análisis estimación y pruebas de hipótesis utilizando remuestreo.
<http://www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html>

[SSC-Stat](http://www.reading.ac.uk/ssc/n/n_sscstat.htm). Complemento para análisis estadístico con Excel.
http://www.reading.ac.uk/ssc/n/n_sscstat.htm

[The R Project for Statistical Computing](http://www.r-project.org/). Gran variedad de análisis, muy poderoso pero requiere de usuarios experimentados. Opera en base a comandos. <http://www.r-project.org/>

[WinIDAMS](http://portal.unesco.org/ci/en/ev.php-url_id=2070&url_do=do_topic&url_section=201.html). Este es paquete de software para la validación, tratamiento y análisis estadístico de datos desarrollado por la Secretaría de la UNESCO en cooperación con expertos de varios países.
http://portal.unesco.org/ci/en/ev.php-url_id=2070&url_do=do_topic&url_section=201.html

Anexo 2: Licencia de “Creative Commons”

Con una licencia de “Creative Commons” usted mantiene sus derechos de autor pero le puede permite a otras personas copiar, modificar y distribuir su obra, siempre y cuando reconozcan la correspondiente autoría y únicamente bajo las condiciones especificadas. Para mayores detalles sobre este tipo de licencia visitar http://creativecommons.org/choose/?lang=es_ES.

Anexo 2: Abreviaturas y equivalencias

Factores de Multiplicación

Factor de multiplicación	Abreviatura	Prefijo	Símbolo
1 000 000 000 000 000 000	10^{18}	exa-	E
1 000 000 000 000 000	10^{15}	peta	P
1 000 000 000 000	10^{12}	tera	T
1 000 000 000	10^9	giga	G
1 000 000	10^6	mega	M
1 000	10^3	kilo	k
100	10^2	hecto	h
10	10^1	deca	da
0,1	10^{-1}	deci	d
0,01	10^{-2}	centí	c
0,001	10^{-3}	ml	m
0,0001	10^{-4}	micro	Φ
0, 000 000 001	10^{-9}	nano	n
0, 000 000 000 001	10^{-12}	pico	p
0, 000 000 000 000 001	10^{-15}	femto	f
0, 000 000 000 000 000 001	10^{-18}	atto	a

Equivalencias

1 tonelada métrica	1 000 Kg
1 kilogramo	1000 gr
1 kilogramo	2,2102 lb
1 hectárea	10 000 m ²
1Km ²	100 ha
1Km ³	1 000 000 m ³
1 m ³	100 lt
1lt	1 000 ml