ALUMNA:
ELSY MARIA DEARA LOPEZ

DOCENTE:
ABEL ESTRADA DICHI

MATERIA:
BIOESTADISTICA

TRABAJO:
INVESTIGACION

CUATRIMESTRE Y CARRERA:
4* ENFERMERIA

CONCEPTOS BASICOS DE ESTADISTICAS

Moda

La moda es el valor de un conjunto de datos que aparece con mayor frecuencia. Se le obtiene fácilmente a partir de un arreglo ordenado. A diferencia de la media aritmética, la moda no se afecta ante la ocurrencia de valores extremos. Sin embargo, sólo se utiliza la moda para propósitos descriptivos porque es más variable, para distintas muestras, que las demás medidas de tendencia central. Un conjunto de datos puede tener más de una moda o ninguna.

Su símbolo es Mo.

a) Moda para datos no agrupados

Ejemplos

1) datos : $2,\ 4,\ 5,\ 6,\ 7,\ 7,\ 8,\ 7,\ 6 \quad \Rightarrow Mo = 7$

2) datos : $1,\ 1,\ 3,\ 1,\ 1,\ 2,\ 2,\ 4,\ 2,\ 3,\ 2,\ 5,\ 6 \Rightarrow Mo = 1$ y 2

3) datos : $0,\ 0,\ 2,\ 3,\ 4,\ 5 \quad \Rightarrow Mo = 0$

4) datos : $0,\ 1,\ 2,\ 3,\ 4,\ 5 \quad \Rightarrow Mo =$ no existe

b) Moda para datos agrupados

Existe más de una forma de calcular la moda:

Caso a) $$Mo = L_i + \left(\frac{d_1}{d_1 + d_2}\right) \cdot a$$

donde : i es el intervalo de mayor frecuencia absoluta.
L_i es el límite real inferior del intervalo que contiene a la moda.
d_1 es la diferencia entre la frecuencia absoluta del intervalo de la moda y el intervalo anterior : $d_1 = f_i - f_{i-1}$
d_2 es la diferencia entre la frecuencia absoluta del intervalo de la moda y el intervalo posterior : $d_2 = f_i - f_{i+1}$
a es la amplitud del intervalo.

Caso b) $$Mo = L_i + \left(\frac{f_{i+1}}{f_{i-1} + f_{i+1}}\right) \cdot a$$

donde : i es el intervalo de mayor frecuencia absoluta.

Duración	f_i	F_i
$350-399$	4	4
$400-449$	6	10
$450-499$	9	19
$500-549$	**80**	**99**
$550-599$	31	130
$600-649$	20	150
$650-699$	42	192
$700-749$	10	202
$750-799$	8	210
$800-849$	2	212
Total	212	

Caso a): En este caso, el intervalo de mayor frecuencia absoluta es el $4^{to} \Rightarrow i = 4$

$f_4 = 80$

$$Mo = 499,5 + \left(\frac{71}{71+49}\right) \cdot 50$$

$d_1 = 80 - 9 = 71$

$$Mo = 529,08 \text{ horas}$$

$d_2 = 80 - 31 = 49$

$L_4 = 499,5$

$a = 50$

Mediana

La mediana es el valor que divide a la distribución por la mitad. Esto es, la mitad de los caen por debajo de la mediana y la otra mitad se ubica por encima de la mediana. La mediana refleja la posición intermedia de la distribución. Por ejemplo, si los datos obtenidos fueran:

Su símbolo es Me.

24 31 35 35 38 43 45 50 57

La mediana es 38, porque deja cuatro casos por encima (43,45, 50 y 57) y cuatro casos por debajo (35, 35, 31 y 24). Parte a la distribución en dos mitades. En general, para descubrir el caso o puntuación que constituye la mediana de una distribución, simplemente se aplica la fórmula:

$\frac{N+1}{2}$. Si tenemos 9 casos, $\frac{9+1}{2} = 5$,

Entonces buscamos el quinto valor y éste es la mediana. En el ejemplo anterior es 38. Obsérvese que la mediana es el valor observado que se localiza a la mitad de la distribución, no el valor 5. La fórmula no nos proporciona directamente el valor de la mediana, sino el número de caso en donde está la mediana.

La mediana es una medida de tendencia central propia de los niveles de medición ordinal, por intervalos y de razón. No tiene sentido con variables nominales, porque en este nivel no hay jerarquías, no hay noción de encima o debajo. También, la mediana es particularmente útil cuando hay valores extremos en la distribución. No es sensible a éstos. Si tuviéramos los siguientes datos:

24 31 35 35 38 43 45 50 248

La mediana sigue siendo 38.

Media

La media es la medida de tendencia central más utilizada y puede definirse como el promedio aritmético de una distribución. Se simboliza como: X, y es la suma de todos los valores dividida por el número de casos. Es una medida sola mente aplicable a mediciones por intervalos o de razón. Carece de sentido por variables medidas en un nivel nominal u ordinal. Su fórmula es:

$$\overline{X} = \frac{X_1 + X_2 + X_3 + X_k}{N}$$

Por ejemplo, si tuviéramos las siguientes puntuaciones:

8 7 6 4 3 2 6 9 8

La media sería igual a:

$$\overline{X} = \frac{8+7+6+4+3+2+6+9+8}{9} = 5.88$$

La fórmula simplificada de la media es:

$$\overline{X} = \frac{\sum X}{N}$$

El símbolo "$\sum$" indica que debe efectuarse una sumatoria, "X" es el símbolo de una puntuación y "N" es el número total de casos o puntuaciones. En nuestro ejemplo:

$$\overline{X} = \frac{53}{9} \quad =5.88$$

La media sí es sensible a valores extremos. Si tuviéramos las siguientes puntuaciones:

8 7 6 4 3 2 6 9 20

la media sería:

$$\overline{X} = \frac{65}{9} = 7.22$$

Rango

Indica el número de valores que toma la variable. El rango es la diferencia entre el valor máximo y el valor mínimo de un conjunto de datos.

$R = x_{\text{máx}} - x_{\text{mín}}$

Si los datos están agrupados en una tabla de frecuencias, el recorrido es la diferencia entre el límite real superior del último intervalo y el límite real inferior del primer intervalo.

$R = L_{\text{máx}} - L_{\text{mín}}$

Ejemplo:

1) Sea el siguiente conjunto de datos :

12 15 17 23 25 28

$x_{\text{máx}} = 28$ $x_{\text{mín}} = 12$ $R = 28 - 12 = 16$

2) Sea la siguiente tabla:

Peso ($Kg.$)	f_i
$55,0 - 63,0$	5
$63,1 - 71,1$	15
$71,2 - 79,2$	12
$79,3 - 87,3$	5
$87,4 - 95,4$	3
Total	40

$L_{min} = 54,95$ $L_{max} = 95,45$

$R = 95,45 - 54,95$ $R = 40,5\,Kg.$

El rango mide "la dispersión total" del conjunto de datos. Aunque el rango es una medida de dispersión simple y que se calcula con facilidad, su debilidad preponderante es que no toma en consideración la forma en que se distribuyen los datos entre los valores más pequeños y los más grandes.

Varianza

Medida de la variación de una serie de observaciones respecto de la media. Equivale a la dispersión respecto de la media en una serie i- x) 2 /n si corresponde a la población total o sigma (xi- X)2 /(n-1) si corresponde a una muestra de esa población, siendo X la media, n el tamaño de la población o de la muestra y xi cada uno de los valores.

Una medida de dispersión mucho más común, que se calcula al promediar los cuadrados de las desviaciones individuales a partir de la media, es la media de desviaciones cuadráticas o la varianza.

La varianza es una medida de dispersión promedia de un conjunto de datos. Para una población se construye al tomar la diferencia entre cada valor observado y la media poblacional, elevando al cuadrado cada una de estas desviaciones y luego hallando la media aritmética de los valores cuadrados. Para una muestra, una expresión casi análoga se construye con la ayuda de su media.

Para una población

$$\sigma^2 = \Sigma(X - \mu)^2 / N$$

Para una muestra

$$s^2 = \Sigma(X - \bar{X})^2 \; / \; (n - 1)$$

EJEMPLO: Calcule la varianza para una población de N = 5 valores: 2, 2, 4, 7 y 15.

SOLUCION: La tabla que muestra la forma en que la varianza se calcula para datos poblacionales, procedimiento por demás tedioso cuando el número de observaciones es grande. Los programas modernos para computadora efectúan con suma rapidez este tipo de operación.

VALORES OBSERVADOS X	*MEDIA DE LA POBLACION*	*DESVIACIONES*	*CUADRADO DE LAS DESVIACIONES*
2	6	-4	16
2	6	-4	16
4	6	-2	4
7	6	1	1
15	6	9	81
$\Sigma = 30$			$\Sigma = 118$
	$\sigma = 118/5 = 23.6$		

Varianza y Desviación Estándar

Dos medidas de dispersión que se utilizan con frecuencia y que sí toman en consideración la forma en que se distribuyen los valores son la varianza y su raíz cuadrada, la desviación estándar. Estas medidas establecen la forma en que los valores fluctúan con respecto a la media.

Varianza

La varianza se define como el promedio aritmético de las diferencias entre cada uno de los valores del conjunto de datos y la media aritmética del conjunto elevadas al cuadrado. Su símbolo es si estamos trabajando con una muestra y si estamos trabajando con una población.

Su símbolo es S^2 si estamos trabajando con una muestra y σ^2 si estamos trabajando con una población.

a) Varianza para datos no agrupados

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$ donde x_i representa los datos de la muestra.

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N-1}$$ donde x_i representa los datos de la población.

Coeficientes de variación

Un problema que plantean las medidas de dispersión vistas es que vienen expresadas en las unidades en que se ha medido la variable. Es decir, son medidas absolutas y con el único dato de su valor no es posible decir si tenemos una dispersión importante o no. Para solucionar esto, se definen unas medidas de dispersión relativa, independiente de la unidad usada. Estas dispersiones relativas van a permitir además comparar la dispersión entre diferentes muestras (con unidades diferentes). Entre estas medidas hay que destacar el coeficiente de variación de Pearson, definido como el cociente entre la desviación típica y la media aritmética.

$$CV = \frac{s}{|\overline{x}|}. \tag{3.18}$$

Nótese que este coeficiente no se puede calcular cuando $\overline{x} = 0$. Normalmente CV se expresa en porcentaje, multiplicando su valor por 100. Evidentemente, cuanto mayor sea CV, mayor dispersión tendrán los datos.

Ejemplo I–*

(Continuación.)

Calculemos el coeficiente de variación de los ejemplos anteriores.

Ejemplo I–5: $CV = s/|\overline{x}| = 1.16/2.25 = 0.516 \quad 52\,\%$.

Ejemplo I–6: $CV = s/|\overline{x}| = 0.795/8.52 = 0.093 \quad 9\,\%$.

Ejemplo I–8: $CV = s/|\overline{x}| = 0.186/9.873 = 0.019 \quad 2\,\%$.

Asimismo se pueden definir otras medidas de dispersión relativas, como el coeficiente de variación media. Este es similar al coeficiente de variación de Pearson, pero empleando una desviación media en vez de la media aritmética. Se tienen entonces dos coeficientes de variación media dependiendo de que se calcule respecto a la desviación media respecto a la media aritmética o respecto a la mediana.

$$CVM_{\overline{x}} = \frac{D_{\overline{x}}}{|\overline{x}|} \quad ; \quad CVM_{M_e} = \frac{D_{M_e}}{|M_e|}.$$