Estadígrafos de Dispersión

Rango, Desviación, Varianza, Coeficiente de Variación

Lectura e interpretación de datos

Los estadígrafos de posición no son suficientes para observar las características de un grupo de datos, agrupados o no agrupados. Por ejemplo, podemos tener el caso de dos grupos de alumnos de III semestre de una misma escuela que tienen como grupo una misma nota o calificación. Sin embargo, al analizar los datos por grupo podemos observar que, en primer grupo, los valores estuvieron en posiciones diferentes en cada uno de los parciales con respecto a las calificaciones del otro grupo (tal vez las mejores calificadas concentradas más en los últimos parciales). Es decir, a pesar que los dos grupos tengan al final del semestre un mismo promedio de calificación, el comportamiento o variación de las calificaciones durante el semestre para cada uno de ellos ha sido completamente diferente.

Concluyendo, las medidas de posición no nos dan mucha información del comportamiento de las calificaciones de los grupos durante el semestre. Entre las medidas de dispersión empleadas con mayor frecuencia por los investigadores están la varianza, la desviación estándar o típica, y el coeficiente de variación.

Estas medidas permiten evaluar el grado de homogeneidad, dispersión, diseminación o variabilidad de un conjunto de datos. Estas medidas son: la amplitud o rango, la varianza, la desviación estándar y el coeficiente de variabilidad.

- **RANGO o AMPLITUD**

Iniciemos recordando qué es el rango y cómo se calcula.

Existen dos tipos de rango:

El rango es la diferencia entre el puntaje más alto y el menor de los datos. Llamado también Rango excluyente. Por ejemplo, cuando nos dicen que la temperatura más alta durante el presente año en nuestra ciudad capital de Saltillo, durante este verano ha sido de 40 grados centígrados y la más fría de 23 grados centígrados, podemos calcular el rango de temperatura mediante la siguiente fórmula:

$$R = V_M - V_m$$

De donde:

R = Rango
V_M = Dato mayor
V_m = Dato menor

Para este ejemplo tendríamos, $R = 40 - 23 = 17$; traduciendo, la temperatura pudiese oscilar hasta 17 grados centígrados durante el verano.

Jack Levin menciona que "una de las principales ventajas del rango es su fácil y rápido cálculo, que se convierte en una desventaja, ya que como el rango depende solamente de dos valores de los puntajes, nos da solo un índice no procesado de la dispersión de la distribución".

EJEMPLO 1:

Calculemos el rango excluyente de las siguientes puntuaciones obtenidas en una muestra:

a) $99, 38, 44, 67, 89, 76, 50, 32, 89, 92, 23, 76, 89, 43, 12$

$$R = V_M - V_m = 99 - 12 = 87$$

a) 0, 43, 56, 98, 76, 54, 34, 89

$$R = V_M - V_m = 98 - 0 = 98$$

b) $17,\ 42,\ 40,\ -10,\ -5$

$$R = V_M - V_m = 40 - (-10) = 40 + 10 = 50$$

El rango incluyente es la diferencia entre el límite superior real del intervalo que abarca la mayor puntuación, y el límite inferior real del intervalo que abarca la puntuación menor (GENE.V.GLASS, 1996).

EJEMPLO 2:

Si tenemos los siguientes valores como resultado de los pesos (expresados en Kg) obtenidos entre 10 estudiantes de determinado grupo:

$$45, 67, 49, 58, 86, 68, 95, 48, 73, 64$$

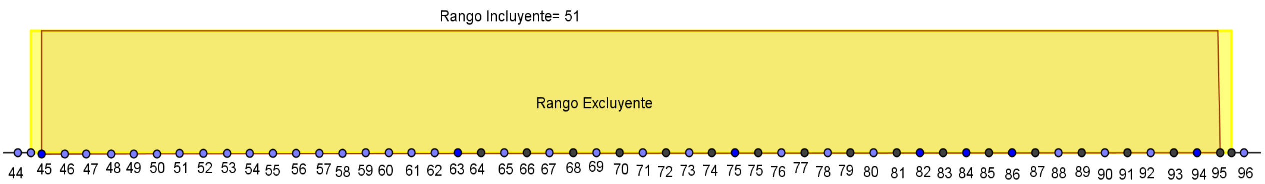
El peso más bajo es $45\ kg$ y el más alto $95\ kg$; 45 se encuentra en el intervalo $(44.5, 45.5)$, (L_i, L_s) y 95 en el intervalo $(94.5, 95.5)$.

El peso real de los estudiantes: $44.5\ kg$ y $95.5\ kg$.

Por lo que el rango incluyente será: $R = L_sRV_M - L_iRV_m = 95.5 - 44.5 = 51\ kg$.

Comparado con el resultado del <u>Rango Excluyente</u> que será: $R = V_M - V_m = 95 - 45 = 50$

Existe una diferencia entre ambos de una unidad. Veámoslo gráficamente:



En el futuro, si no se especifica si el rango es incluyente o excluyente, es porque nos referimos a cualquiera de los dos, ya que el rango es la medida más imperfecta de variabilidad.

- **LA VARIANZA**

La varianza y la desviación estándar son medidas de dispersión útiles y muy utilizadas, ya que nos proporcionan la forma en que los valores fluctúan alrededor del promedio.

La varianza es el promedio de las desviaciones respecto del promedio elevados al cuadrado.

La varianza se puede representar como σ^2, se lee sigma cuadrado. Y se calcula utilizando:

Si $x_1,\ x_2,\ x_3, \ldots, x_n$ son n datos y ▯ su promedio, la varianza es:

$$\sigma^2 = \frac{(x_1 - \bar{\boldsymbol{x}})^2 + (x_2 - \bar{\boldsymbol{x}})^2 + (x_3 - \bar{\boldsymbol{x}})^2 + \cdots + (x_n - \bar{\boldsymbol{x}})^2}{n}$$

El resultado de la varianza nos dará datos grandes respecto a los datos originales y se expresa en términos de los datos originales, elevados al cuadrado (QUINTANILLA, OCHOA & VARGAS, 2008).

La varianza:

- ✓ Se expresa en unidades cuadráticas respecto a la variable de procedencia, esto permite darle una interpretación más real.
- ✓ Es un número no negativo.
- ✓ De un grupo constante de datos es igual a cero; es decir, todas las observaciones o datos recabados son igual a cero.
- ✓ Indica el grado de dispersión de los datos.
- ✓ Mientras mayor es la varianza en un conjunto de observaciones o datos, mayor será la dispersión de los datos.

❖ DESVIACIÓN ESTÁNDAR O TÍPICA

La desviación típica es una medida de desviación íntimamente relacionada con la varianza y se denota por "s" o "σ"; se define como la raíz cuadrada positiva de esta última.

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

La desviación típica es una magnitud no negativa.

EJEMPLO 3:

Calcular la varianza y la desviación estándar de los siguientes valores obtenidos en una muestra:

$11, 12, 13, 9, 10$; así que $x_1 = 11, x_2 = 12, x_3 = 13, x_4 = 9$ y $x_5 = 10$

Calculando el promedio de dichos datos obtendremos: $\bar{x} = \frac{11+12+13+9+10}{5} = 11$

Al sustituir estos valores en la fórmula de la varianza, se obtiene:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2}{n}$$

$$\sigma^2 = \frac{(11-11)^2 + (12-11)^2 + (13-11)^2 + (9-11)^2 + (10-11)^2}{5} =$$

$$\sigma^2 = \frac{(0)^2 + (1)^2 + (2)^2 + (-2)^2 + (-1)^2}{5} =$$

$$\sigma^2 = \frac{0+1+4+4+1}{5} = \frac{10}{5} = 2$$

$$\sigma^2 = 2$$

$$\sigma = \sqrt{2} = 1.4142$$

De donde concluimos que la varianza $= 2$ y la desviación estándar es 1.4142

❖ **VARIANZA Y DESVIACIÓN ESTÁNDAR PARA DATOS AGRUPADOS**

Sean $y_1, y_2, y_3, \ldots, y_n$ datos agrupados en k intervalos, y sean $x_1, x_2, x_3, \ldots, x_n$ las marcas de clase o puntos medios de los k intervalos; entonces la varianza de los s n datos está dada por la siguiente fórmula:

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i}$$

Y la desviación estándar es:

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

Estadígrafos de Dispersión

La desviación estándar es de suma utilidad en determinadas distribuciones, donde se conocen los porcentajes aproximados de puntuaciones situadas a una, dos, tres o más desviaciones, con respecto a la media (Levin, 2006).

Te recomiendo el tutorial:

http://youtu.be/t3Wnx4UO0pQ

Referencia:
Este video fue tomado de **ngeniat** (2011) **Cuartile datos agrupados**
Recuperado: 17/06/15 A partir de: http://youtu.be/t3Wnx4UO0pQ

EJEMPLO 4:

En una prueba de aptitud matemática aplicada a un grupo de 50 estudiantes en una escuela preparatoria, se obtuvieron las siguientes calificaciones: $n = 50$. Calcular la varianza y la desviación estándar.

88	74	77	69	79
33	86	78	66	69
38	65	65	49	75
44	39	63	78	70
77	79	84	75	97
90	64	89	82	71
99	68	74	73	54
56	62	78	91	63
78	85	81	81	82
72	86	66	90	76

Estadígrafos de Dispersión

Es el ejemplo que hemos venido trabajando desde la lección 1. Aquí retomaremos el cuadro construido hasta la lección 3.

i	$L_s - L_i$	f_i	x_i	F_i	h_i	H_i	$x_i f_i$
1	98 – 89	6	93.5	6	0.12	0.12	561
2	88 – 79	11	83.5	17	0.22	0.34	918.5
3	78 – 69	16	73.5	33	0.32	0.66	1176
4	68 – 59	10	63.5	43	0.2	0.86	635
5	58 – 49	3	53.5	46	0.06	0.92	160.5
6	48 – 39	2	43.5	48	0.04	0.96	87
7	38 – 29	2	33.5	50	0.04	1	67
		$\sum = 50$		$\sum = 50$	$\sum = 1$		3605

De donde hemos obtenido la $media = 72.1$ y la $moda = 73.54$.

Para calcular la varianza, utilizaremos la siguiente fórmula.

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \overline{\boldsymbol{x}})^{\mathbf{2}}}{\sum_{i=1}^{n} f_i}$$

$$\sigma^2 = \frac{f_1(x_1 - \overline{\boldsymbol{x}})^2 + f_2(x_2 - \overline{\boldsymbol{x}})^2 + f_3(x_3 - \overline{\boldsymbol{x}})^2 + f_4(x_4 - \overline{\boldsymbol{x}})^2 + f_5(x_5 - \overline{\boldsymbol{x}})^2 + f_6(x_6 - \overline{\boldsymbol{x}})^2 + f_7(x_7 - \overline{\boldsymbol{x}})^2}{f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7}$$

Las marcas de clase son:

$x_1 = 93.5, x_2 = 83.5, x_3 = 73.5, x_4 = 63.5, x_5 = 53.5, x_6 = 43.5, x_7 = 33.5$, sustituyendo:

$$\sigma^2 = \frac{6(93.5 - 72.1)^2 + 11(83.5 - 72.1)^2 + 16(73.5 - 72.1)^2 + 10(63.5 - 72.1)^2 + 3(53.5 - 72.1)^2 + 2(43.5 - 72.1)^2 + 2(33.5 - 72.1)^2}{6 + 11 + 16 + 10 + 3 + 2 + 2}$$

$$\sigma^2 = \frac{3(457.96) + 11(129.96) + 16(1.96) + 10(73.96) + 3(345.96) + 2(817.96) + 2(1489.96)}{50}$$

$$\sigma^2 = \frac{1373.88 + 1429.56 + 31.36 + 739.6 + 1037.88 + 1635.92 + 2979.92}{50} = 184.5624$$

Por lo que la desviación estándar será:

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

$$\sigma = \sqrt{184.5624}$$

$$\sigma = 13.58$$

- **¿QUÉ REPRESENTA LA DESVIACIÓN ESTÁNDAR?**

Recordemos, la desviación estándar representa la variabilidad promedio de una distribución, ya que mide el promedio de desviaciones de la media.

Mientras mayor sea la dispersión de los datos alrededor de la media en una distribución, mayor será el valor de la desviación estándar.

Por ejemplo, si en el procesamiento de datos obtenidos por las calificaciones del primer parcial de II "A" y II "B", se tiene que la desviación estándar del grupo A es de 1.45 y la desviación estándar del grupo B es de 3.50, quiere decir que existe mayor variabilidad en las calificaciones del grupo "B".

- **COEFICIENTE DE VARIACIÓN:**

Es un estadígrafo que refleja la dispersión sin depender de la magnitud de las observaciones; es decir, siendo un valor relativo.

Se define como el cociente de la desviación estándar entre su media.

$$C.V = \frac{\sigma}{\bar{x}}$$

De donde:
$C.V$ = Coeficiente de variación
σ = Desviación estándar
$\bar{x}$ = Media

Como resultaba imposible realizar una comparación directa de dos o más medidas de dispersión, Karl Pearson contribuyó de manera importante a la estadística y desarrolló esta medida conocida como coeficiente de variación.

Ahora viene una parte delicada de la estadística, la interpretación de los datos.

Estadígrafos de Dispersión

Para todos los temas estudiados en esta unidad se recomiendan los siguientes tutoriales:

Si tienes dudas puedes revisar los TUTORIALES RECOMENDADOS:

#	DATOS NO AGRUPADOS	
1	Moda	http://youtu.be/h2tdhAgLLAw
2	Mediana	http://youtu.be/6wagmDkaaXk
3	Quartil	http://youtu.be/p_xWntf4ydY
4	Deciles	http://youtu.be/qkFeTckQqLM
	DATOS AGRUPADOS	
5	Moda	http://youtu.be/ISbnLcFFrNY
6	Mediana	http://youtu.be/CSPp6yG82kk
7	Media	http://youtu.be/-ZnUSLlUj9A
8	Varianza	http://youtu.be/VkiWEtnkV38
9	Desviación Típica	http://youtu.be/t3Wnx4UO0pQ
10	Quartil	http://youtu.be/oVf_tNan_Xs
11	Percentil	http://youtu.be/p_xWntf4ydY

Estadígrafos de Dispersión

Referencias:

1. Este video fue tomado de **unicoos** (2011) Estadistica 01 SECUNDARIA (4ºESO) matematicas Recuperado: 17/06/15 A partir de: http://youtu.be/h2tdhAgLLAw
2. Este video fue tomado de ingeniat (2011) SEP RM2S R1.9.2 Mediana Recuperado: 17/06/15 A partir de: https://www.youtube.com/watch?v=6wagmDkaaXk&feature=youtu.be
3. Este video fue tomado de Sergio Arias (2011) Medidas de posicion ¼ Recuperado: 17/06/15 A partir de: https://www.youtube.com/watch?v=p_xWntf4ydY&feature=youtu.be
4. Este video fue tomado de **Sergio Arias** (2011) Medidas de posicion 2/4 Recuperado: 17/06/15 A partir de: https://www.youtube.com/watch?v=qkFeTckQqLM&feature=youtu.be
5. Este video fue tomado de **unicoos** (2011) Estadistica 02 SECUNDARIA (4ºESO) matematicas intervalos Recuperado: 17/06/15 A partir de:https://www.youtube.com/watch?v=ISbnLcFFrNY&feature=youtu.be
6. Este video fue tomado de **ingeniat** (2011) UDEM Estadística para negocios Moda Recuperado: 17/06/15 A partir de: https://www.youtube.com/watch?v=CSPp6yG82kk&feature=youtu.be
7. Este video fue tomado de **Abel Esteban Ortega Luna** (2012) Tabla de distribución de frecuencias con intervalos Recuperado: 17/06/15 A partir de: https://www.youtube.com/watch?v=-ZnUSLlUj9A&feature=youtu.be
8. Este video fue tomado de **ingeniat** (2011) SEP MATE 2S BI A1.9 R1.9.1 ¿La media o la mediana? Recuperado: 17/06/15 A partir de: https://www.youtube.com/watch?v=VkiWEtnkV38&feature=youtu.be
9. Este video fue tomado de **ingeniat** (2011) UDEM Estadística para negocios Medidas de desviación Recuperado: 17/06/15 A partir de: https://www.youtube.com/watch?v=t3Wnx4UO0pQ&feature=youtu.be
10. Este video fue tomado de **Luis Lázaro** (2010) Cuartile datos agrupadosRecuperado: 17/06/15 A partir de: https://www.youtube.com/watch?v=oVf_tNan_Xs&feature=youtu.be