ESTADÍSTICA: MEDIDAS DESCRIPTIVAS I

AUTOR: PABLO VARGAS PEREIRA

FEBRERO: 2021

San Marcos

Contenido

Introducción

Las medidas de tendencia central son el tema, en estadística, más conocido. Con múltiples aplicaciones a lo largo de la historia, estas medidas corresponden a un análisis descriptivo de los datos, con el cuidado en no caer en interpretaciones erróneas por tratar de inferir algo que no corresponde a los datos.
Además, en esta lectura se encuentran las medidas de variabilidad, que corresponden a un complemento descrptivo de los datos que analizan de forma más profunda y completa.

Medidas descriptivas

Hasta este punto, se han abordado todos los los métodos de tabulación con las frecuencias absolutas, relativas, acumuladas, con variables discretas y continuas. Además, se abordaron los métodos de graficación con lo que se representan de una forma ilustrativa los datos que se habían representado en las tablas. Sin embargo, todos estos, presentan algunas limitaciones, ya que no pueden describir con una mayor profundidad el conjunto de datos.

Las medidas de tendencia central, de acuerdo con Anderson, Sweeney, Williams (2016) "son medidas descriptivas que señalan hacia donde tienden a concentrarse los valores contenidos en un conjunto de datos. Su resultado debe ser un valor típico o representativo de la muestra o población, el cual es utilizado para describir o analizar un fenómeno" (p. 3).

Estas medidas son abstractas, sin embargo, son de gran utilidad en la vida cotidiana, por lo que es responsabilidad del aplicador, utilizarla de forma correcta, para que los lectores las puedan comprender sin llegar a vacíos en la interpretación.

Por ejemplo, en Costa Rica, de acuerdo con la periodista Pilar Cisneros en un reportaje en Octubre 2020 realiza una comparación entre los salarios del sector público, con respecto al del sector privado. La periodista menciona que el sector público gana, en promedio, tres veces más que en el sector público. Esto no es una mentira, pues comparando directamente los resultados de los promedios se puede ver esta relación, sin embargo, es necesario que se homogenicen los datos, ya que en el sector privado también se encuentra el sector informal, cuyos ingresos no son contabilizados por hacienda, inflando la cantidad de personas pero disminuyendo el dato correspondiente al salario.

Por razones como la expuesta en el ejemplo anterior se debe tener un cuidado especial y sobre todo "en el análisis de muchos fenómenos también se necesita conocer la manera en que los valores de una serie se dispersan entre sí. Para ello se acude a otro tipo de medidas descriptivas, las medidas de dispersión o de variabilidas" (Anderson, Sweeney, Williams, 2016, p. 3).

Con estas medidas se evidencia que tanto varían los datos entre si y con respecto a las medidas de tendencia central y su interpretación dependería de que tan grande o pequeña sea esta variación.

I. Medidas de tendencia central

Media Aritmética

De acuerdo con la Universidad Nacional Autonoma de México [UNAM], en el material para el curso de Estadística de la facultad de Estudios Superiores, la media aritmética es la medida de tendencia central más utilizada y es igual a lo que conocemos como promedio.

Se define matemáticamente, con el símbolo $\bar{x}$ y se calcula mediante la fórmula:

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

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

En prosa, la fórmula indica: *la suma de todos los datos, dividido entre la cantidad de datos*.

Ejemplo 1

Considere los datos correspondientes a la cantidad de estudiantes por sección en un colegio con 7 secciones de sétimo.

32, 25, 29, 30, 35, 33, 30

El promedio de la cantidad de estudiantes en el nivel de sétimo corresponde a:

$$\bar{x} = \frac{32 + 25 + 29 + 30 + 35 + 33 + 30}{7}$$

$$\bar{x} = 30{,}57$$

Nótese que el promedio no indica que en todas las secciones del nivel de sétimo se cuenta con 30 estudiantes, este es el cuidado que se debe tener con la interpretación de este.

Otro significado que se le puede plantear para la media aritmética sería el punto de equilibrio de los datos, ya que este se posiciona de manera que, si se colocara en una balanza, esta se podría mantener en enquilibrio hipotético con los datos que se encuentran a sendos lados de esta.

Ejemplo 2

En una empresa muy pequeña, de venta de servicios, se tienen los siguientes salarios por mes:

200 000, 400 000, 150 000, 180 000, 1 000 000, 900 000

¿Cuál es el promedio de los salarios en la empresa?

$$\bar{x} = \frac{200\,000 + 400\,000 + 150\,000 + 180\,000 + 1\,000\,000 + 900\,000}{6}$$

$$\bar{x} = 471\,666{,}6$$

Se puede notar que 4 de los 6 salarios corresponden a una cantidad menor o igual a 400 000 colones. Por lo que deducir que el promedio del salario de todos los trabajadores es 471 000 resulta desproporcionado con respecto a la realidad contextual de los datos. Por lo que no se debe fiar unicamente de esta medida de tendencia central.

Promedio ponderado para datos discretos

Este se utiliza para datos que se encuentran agrupados. El significado sigue sin sufrir una variante en importancia, pero se debe tomar en consideración ese agrupamiento. Generalmente este se representa por medio de la frecuencia absoluta.

Matemáticamente, si se tiene un total de n datos, pero cada uno de estos se encuentra agrupado por frecuencias absolutas f_i entonces el promedio ponderado se calcula de la siguiente manera:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i \cdot x_i}{\sum_{i=1}^{n} f_i}$$

$$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \cdots + f_n \cdot x_n}{f_1 + f_2 + \cdots + f_n}$$

Para ejemplificar la formula se presenta la siguiente situación.

Ejemplo 3

Un docente, después de calificar sus examenes, realiza una tabla de frecuencia que resuma todos los datos obtenidos por sus estudiantes, quedando de la siguiente manera.

Notas	Frecuencia Absoluta
60	8
70	2
80	5
90	4
100	1
Total	**20**

El docente quiere determinar el promedio de las notas de los estudiantes.

$$\bar{x} = \frac{8 \cdot 60 + 2 \cdot 70 + 5 \cdot 80 + 4 \cdot 90 + 1 \cdot 100}{8 + 2 + 5 + 4 + 1}$$

$$\bar{x} = 74$$

De esta forma el docente puede determinar que el promedio de las notas sería 74.

Ejemplo 4

Las universidades tienen un sistema de créditos por materia, de manera que se identifique cuanto tiempo se le debe dedicar a esta para asegurar el éxito. Un estudiante matriculó 4 materias: Investigación de 6 créditos. Estadística de 4 créditos. Historia de 3 créditos y Sociología de 3 créditos. Al finalizar el semestre obtuvo las notas que se presentan en la tabla.

Materia	**Nota**
Investigación	85
Estadística	80
Historia	95
Sociología	90

El estudiante quiere calcular el promedio de sus notas, por lo que debe contemplar el valor de los créditos.

$$\bar{x} = \frac{6 \cdot 85 + 4 \cdot 80 + 3 \cdot 95 + 3 \cdot 90}{6 + 6 + 3 + 3}$$

$$\bar{x} = 86{,}56$$

Se realizará el cálculo del promedio simple, para contrastar la diferencia:

$$\bar{x} = \frac{85 + 80 + 95 + 90}{4}$$

$$\bar{x} = 87{,}5$$

Se puede observar una diferencia entre el promedio ponderado y el promedio simple, puesto que en la primera se toma en consideración el número de créditos, que en este caso se toma en consideración como si el valor de la nota se repitiera 6, 4 o 3 veces alterando el promedio con respecto al simple que solamente toma en cuenta la nota una sola vez.

Promedio ponderado para datos contínuos

En este apartado se toma en consideración el tipo de dato continuo, ya que este en la tabla de frecuencias tiene un punto relevante por trabajar.

Los datos están comprendidos en intervalos, no están de forma discreta, por lo que se requiere utilizar el concepto de marca de clase M_c.

$$M_c = \frac{límite\ inferior + límite\ superior}{2}$$

Este será el dato que se considere para el cálculo del promedio, tal y como lo muestra la fórmula

$$\bar{x} = \frac{\sum_{i=1}^{n} M_{c_i} \cdot f_i}{\sum_{i=1}^{n} f_i}$$

$$\bar{x} = \frac{M_{c_1} \cdot f_1 + M_{c_2} \cdot f_2 + \cdots + M_{c_n} \cdot f_n}{f_1 + f_2 + \cdots + f_n}$$

Ejemplo 5

Se realizaron mediciones de altura en un grupo de 20 personas de una empresa como parte del control en salud ocupacional. Los datos se resumieron en la tabla.

Variable Altura	Frecuencia Absolulta
$[1.40\ ,1.50[$	2
$[1.50\ ,1.60[$	5
$[1.60\ ,1.70[$	6
$[1.70\ ,1.80[$	7
Total	**20**

Se requiere realizar el cálculo del promedio de las alturas.

Para ello se requiere calcular las marcas de clase.

$$M_{c_1} = \frac{1.50 + 1.60}{2} = 1.55$$

$$M_{c_3} = \frac{1.70 + 1.80}{2} = 1.75$$

$$M_{c_2} = \frac{1{,}60 + 1{,}70}{2} = 1.65$$

$$M_{c_4} = \frac{1.80 + 1.90}{2} = 1.85$$

Ahora, con las marcas de clase se procede a calcular el promedio.

Se puede notar que no se conoce, con base en la información de la tabla, cuál es el dato exacto en cada uno de los casos, por eso se toma el punto medio, sin embargo, se evidencia que se tiene ya un error en el cálculo, que puede ser considerable, con respecto a la precisión y la exactitud del proceso realizado.

$$\bar{x} = \frac{2 \cdot 1.55 + 5 \cdot 1.65 + 6 \cdot 1.75 + 7 \cdot 1.85}{2 + 5 + 6 + 7}$$

$$\bar{x} = 1{,}74$$

Este corresponde al promedio de las alturas en la empresa.

Mediana para datos discretos

Este concepto corresponde al valor central que se localiza en una serie ordenada de datos. (UNAM, sf) El procedimiento difiere de la cantidad de datos, por lo que se abordará de forma separada.

<u>Cantidad de datos impar</u>

Si la cantidad de datos n es impar la mediana M_e corresponde al dato que se encuentra justamente en el centro y se obtiene buscando la posición central.

$$M_e = \frac{n}{2}$$

Al ser impar, el resultado de la división siempre tendrá un $,5$ por lo que se redondea hacia arriba y esa sería la posición en la que se encuentra la mediana o el dato central.

Ejemplo 6

Se tienen los siguientes datos de un test spsicométrico a los que se les desea calcular la mediana. Los datos son:

350, 430, 210, 500, 220, 660, 710

Como primer punto se deben ordenar los datos de forma ascendente.

210, 220, 350, 430, 500, 660, 710

Luego se calcula la $M_e = \frac{7}{2} = 3{,}5$ esto indica que el punto medio se encuentra en la posición 4. Por lo tanto, $M_e = 430$

Cantidad de datos par

Si la cantidad de datos n es par la mediana M_e corresponde al dato que se encuentra en el promedio de los dos datos centrales y se obtiene buscando la posición central.

$$M_e = \frac{n}{2}$$

Al ser par, el resultado de la división siempre será exacto, por lo que se promedio la posición dada por la fórmula y la inmediata superior.

Ejemplo 7

El mismo test se aplicó, pero esta vez 8 personas: 110, 350, 430, 210, 500, 220, 660, 710.

De igual forma se ordenan los datos de forma ascendiente:

110, 210, 220, 350, 430, 500, 660, 710

Ahora se realiza el cálculo de la mediana.

$$M_e = \frac{8}{2} = 4$$

Por lo que se deben elegir los puntajes que se encuentran en la posición 4 y 5, que corresponde a 350 y 430. Con ellos se calcula un promedio simple.

$$M_e = \frac{350 + 430}{2} = 555$$

Por lo tanto la mediana del problema corresponde a 555.

Mediana para datos contínuos

Cuando se tienen datos agrupados en clases, normalmente continuos, no se evidencia claramente, desde la tabla de frecuencias, cuáles son los datos. Por lo que el procedimiento depende mucho de la interpretación que se brinde de la tabla.

Clase de la mediana

La clase mediana es la primera de las clases que presenta una frecuencia porcentual acumulada igual o superior al 50%.

Una vez definida la clase mediana, se utiliza el límite inferior $(lím\ inf)$, el límite superior $(lím\ \sup)$, la frecuencia relativa (f_r) y la frecuencia relativa acumulada (F_{R_A}) de la clase anterior; con esto se aplica la siguiente fórmula para determinar el valor de la mediana.

$$M_e = lím\ inf + \frac{(lím\ sup - lím\ inf) \cdot (0{,}5 - F_{R_A})}{f_r}$$

Ejemplo 8[1]

Considere la siguiente tabla que muestra el tiempo, en minutos, que se invierte para atender a un paciente en una cita de psicología.

Variable Tiempo	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Acumulada %
141 - 157	2	0,04	2	4%
157 - 173	13	0,26	15	**30%**
173 - 189	18	**0,34**	32	64%
189 - 205	14	0,28	46	92%
205 - 221	3	0,06	49	98%
221 - 237	1	0,02	50	100%
Total	**50**	**1**		

Claramente se puede observar que la primera clase con una frecuencia porcentual acumulada igual o superior al 50% es la tercera (173 - 189) con un 64%, por lo que a esta clase se le aplica la fórmula brindada.

$$M_e = 173 + \frac{(189 - 173) \cdot (0{,}5 - 0{,}3)}{0{,}34}$$

$$M_e = 182{,}4$$

Nota: con colores se muestra de donde provienen los datos que se utilizaron para la fórmula.

[1] Tomado de UNAM (sf). Medidas de Tendencia Central y de dispersión. **http://www.cuautitlan.unam.mx**

Moda para datos contínuos

La moda (M_o) es el valor con mayor frecuencia absoluta (f_a) en una serie de datos. Dada una tabla de frecuencias se tiene que

$$M_o = máx\{f_{a_1}, f_{a_2}, f_{a_3}, \dots, f_{a_n}\}$$

Nótese que con esta definición puede existir más de una moda en un conjunto de datos, estas pueden denominarse bimodal (2 modas), trimodal (tres modas), o multimodal (más de tres modas). También puede darse el caso de que todas las frecuencias absolutas sean iguales, por lo que en ese caso se considera que no presenta una moda definida.

Ejemplo 9

Considere los siguientes datos: 3, 4, 5, 5, 5, 4, 3, 2, 1, 2, 5. Con base en estos se puede determinar que $M_o = 5$ puesto que es el dato que tiene una mayor frecuencia absoluta, con respecto a los demás datos.

Para el caso de los datos con caracter continuo, se define la clase modal como aquella que presenta mayor frecuencia absoluta, sin embargo si se desea tener un dato, se puede tomar la marca de clase como la moda.

Ejemplo 10

Tomando por referencia la tabla de frecuencias del ejemplo 8, se puede determinar que la clase modal es 173 - 189, puesto que tiene una frecuencia absoluta de 18, mas, si se desea saber un valor moda, este corresponde a la marca de clase.

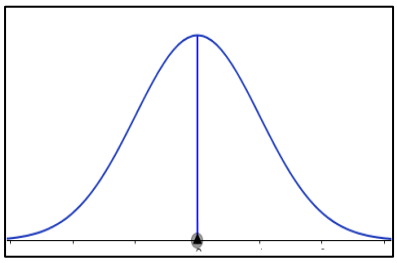
$$M_o = M_c = \frac{lím\ sup + lím\ inf}{2} = \frac{173 + 189}{2} = 181$$

Simetría

Distribución normal

Se le considera una distribución en forma de campana, por la forma en la representa gráficamente. Esta occurre cuando la media aritmética, la mediana y la moda son el mismo valor, o están muy cercanos entre si.

$$\bar{x} = M_e = M_o$$

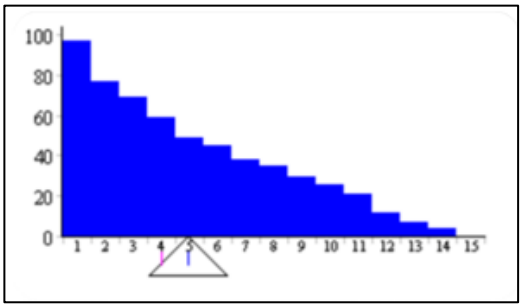


También se puede observar en los gráficos de barras, polígono de frecuencias o histogramas.

<u>Sesgo a la derecha</u>

Se considera que el conjunto de datos tiene un sesgo hacia la derecha si se evidencia el cumplimiento de la desigualdad:

$$\bar{x} > M_e > M_o$$

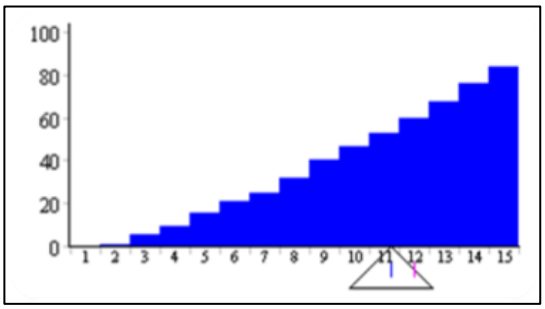


Fuente: UNAM (sf) Medidas de tendencia central y medidas de dispersión

<u>Sesgo a la izquierda</u>

Se considera que el conjunto de datos tiene un sesgo hacia la derecha si se evidencia el cumplimiento de la desigualdad:

$$\bar{x} < M_e < M_o$$



Fuente: UNAM (sf) Medidas de tendencia central y medidas de dispersión

Conocer el sesgo de los datos es de suma importancia en la interpretación de los datos, puesto que, de no hacerlo, se pueden cometer errores.

Ejemplo 11

Considere la siguiente tabla correspondiente a salarios en una empresa de servicios.

Salarios por mes	Frecuencia Absoluta
400 000	15
800 000	12
3 800 000	2
6 900 000	3
Total	**32**

Se calculan las medidas de tendencia central

$$\bar{x} = \frac{15 \cdot 400000 + 12 \cdot 800000 + 2 \cdot 3800000 + 3 \cdot 6900000}{32} = 1\,371\,875$$

$$M_e = 800\,000$$

$$M_o = 400\,000$$

Durante una auditoría, se requiere hacer un recorte en los salarios para dismuir los gastos, por lo que se toma en consideración que el promedio de los trabajadores es de 1 371 875 colones por mes.

Se decide hacer un rebajo a los trabajadores del 15% del salario para todos por igual. Los trabajadores molestos buscan ayuda y se dan cuenta de que los datos presentan un sesgo a la derecha, ya que se cumple que:

$$\bar{x} > M_e > M_o$$

$$1\,371\,875 > 800\,000 > 400\,000$$

Esto les indica que, aunque el promedio de los salarios es más alto, la moda, es decir el salario que más se repite en la empresa es de 400 000 colones, por lo que hacer un rebajo generalizado para todos no es conveniente, ya que la mayor cantidad de trabajadores no tienen el nivel de ingresos que dictamina el promedio.

II. Medidas de dispersión

Considere las siguientes conjuntos de datos

Grupo 1	Grupo 2
20, 22, 24	15, 16, 23

Si se calcula, en ambos casos el promedio, se puede notar que en ambos casos es igual.

Grupo 1	Grupo 2
20, 22, 24	10, 26, 30
$\boldsymbol{n = 3;\ \bar{x} = 22}$	$\boldsymbol{n = 3;\ \bar{x} = 22}$

De esta forma resulta imposible determinar las diferencias entre ambos grupos de datos, por lo que resulta necesario utilizar otro tipo de medida necesaria para diferenciarlas.

Rango de datos

Se define como rango de los datos a la diferencia entre el valor máximo y el valor mínimo.

$$R = máx - mín$$

Si se aplica esta nueva definición para los datos del ejemplo desarrollado en este partado se tiene lo siguiente:

Grupo 1	Grupo 2
20, 22, 24	10, 26, 30
$n = 3;\ \bar{x} = 22$	$n = 3;\ \bar{x} = 22$
$R = 24 - 20 = 4$	$R = 30 - 10 = 20$

Ahora si se pueden iniciar notar las diferencias, sin embargo, como lo menciona UNAM (sf) sería muy burda, pues toma en consideración los valores más extremos y no da una visualización completa de lo que pasa con los demás datos. Ya que como se puede observar, en el grupo 1 los datos están muy agrupados al rederor de la media aritmética y en el grupo dos están más dispersos con respecto a la media.

Por tanto, es necesaria "una medida que valore la variabilidad o distancia promedio de los datos con respecto a su media" (UNAM, sf. p. 20).

Varianza

Esta medida de variabilidad por sí misma no tiene significado, ya que lo que ejecuta es el cuadrado de la suma de cada diferencia entre el promedio y los datos. Se representa con la letra s y se calcula mediante la fórmula:

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

Para el ejemplo que se desarrolla se tiene el siguiente procedimiento.

$\boldsymbol{x_i}$	$\boldsymbol{x_i - \bar{x}}$	$\boldsymbol{(x_i - \bar{x})^2}$	$\boldsymbol{x_i}$	$\boldsymbol{x_i - \bar{x}}$	$\boldsymbol{(x_i - \bar{x})^2}$
20	$20 - 22 = -2$	4	10	$10 - 22 = -12$	144
22	$22 - 22 = 0$	0	26	$26 - 22 = 4$	16
24	$24 - 22 = 2$	4	30	$30 - 22 = 8$	64
Sumas	**0**	**8**	**Sumas**	**0**	**224**

Se evidencia la importancia de que las diferencias estén elevadas al cuadrado, ya que si no lo estuvieran, por la definición de que el promedio es un punto de equilibrio.

Por otra parte, una vez caculada la suma de todas las diferencias al cuadrado, se realiza la división con respecto a $n - 1$ datos.

Grupo 1	**Grupo 2**
20, 22, 24	10, 26, 30
$n = 3;\ \bar{x} = 22$	$n = 3;\ \bar{x} = 22$
$R = 24 - 20 = 4$	$R = 30 - 10 = 20$
$s^2 = \frac{8}{2} = 4$	$s^2 = \frac{20}{2} = 10$

De esta forma se denota una gran diferencia, mas, si se le agrega el siguiente concepto se termina de interpretar la diferencia entre ambos grupos.

Desviación estándar

La UNAM (sf) la define como: "proporciona la variabilidad promedio de los datos con respecto a su media" (p. 23). Resulta de resolver raíz cuadrada a la varianza. Se representa con el símbolo s y se calcula:

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

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

No existe una tipificación universal sobre qué tan grande o pequeña debe ser una desviación estándar, esto depende del contexto del problema.

Para el ejemplo desarrollado se tiene que

Grupo 1	**Grupo 2**
20, 22, 24	10, 26, 30
$n = 3;\ \bar{x} = 22$	$n = 3;\ \bar{x} = 22$
$R = 24 - 20 = 4$	$R = 30 - 10 = 20$
$s^2 = \frac{8}{2} = 4$	$s^2 = \frac{20}{2} = 10$
$s = \sqrt{4} = 2$	$s = \sqrt{10} = 3{,}16$

De esta forma se puede determinar que el grupo 2 varía más que el grupo 1, con respecto a la media aritmética.

Coeficiente de variación

El coeficiente de variación es una medida de variabilidad relativa de una serie de datos y se obtiene dividiendo la desviación estándar de los datos entre su media. Suele calcularse en forma porcentual y con esto se colabora a la comparación de grupos.

$$c.v. = \frac{s}{\bar{x}} \cdot 100\%$$

Concluyendo en el ejemplo se realiza el cálculo del coeficiente de variación.

Grupo 1	**Grupo 2**
$20, 22, 24$	$10, 26, 30$
$n = 3;\ \bar{x} = 22$	$n = 3;\ \bar{x} = 22$
$R = 24 - 20 = 4$	$R = 30 - 10 = 20$
$s^2 = \frac{8}{2} = 4$	$s^2 = \frac{20}{2} = 10$
$s = \sqrt{4} = 2$	$s = \sqrt{10} = 3{,}16$
$c.v = \frac{2}{22} \cdot 100 = 9\%$	$c.v = \frac{3{,}16}{22} \cdot 100 = 14{,}3\%$

Estadarización

De acuerdo con el Ministerio de Educación Pública de Costa Rica [MEP], (2018):

> La estandarización ($V.E.$) consiste en una estrategia que se utiliza para realizar comparaciones con datos que pertenecen a diferentes contextos o diferentes magnitudes. Para que los datos sean comparables se les debe estandarizar: para cada uno se calcula su diferencia respecto al promedio y el resultado se divide por la desviación estándar. (p. 2)

$$V.E = \frac{x_i - \bar{x}}{s}$$

Un dato estandarizado puede ser positivo, negativo o incluso nulo.

Ejercicio 1[2]

Considere la siguiente información:

A dos secciones de octavo año se les aplicó la misma prueba en iguales condiciones. La nota (de 1 a 100) más alta obtenida por un estudiante de cada sección, así como la media aritmética y la desviación estándar de las notas obtenidas por los estudiantes

[2] Tomado de MEP. (2018). Coeficiente de variación y estandarización. https://www.reformamatematica.net/wp-content/uploads/2018/10/20181015-Solucionario-EyP-Coeficiente-de-variaci%C3%B3n-y-estandarizaci%C3%B3n.pdf

de cada sección, se muestran en la siguiente tabla:

Sección	Media Aritmética	Desviación estándar	Nota más alta
8 -1	50	10	80
8 -2	52	8	78

De acuerdo con la información anterior, considere las siguientes proposiciones:

I. La nota más alta obtenida por un estudiante da la sección 8 - 2, tiene una mejor posición relativa que la nota más alta obtenida por un estudiante de la sección 8-1, con respecto a las notas de su correspondiente sección.

II. Las notas de los estudiantes de la sección 8-1 presentan mayor variabilidad relativa que las notas de los estudiantes de la sección 8- 2

¿Cuál o cuáles de ellas son verdaderas?

Ejercicio 2[3]

Considere la siguiente Información:

Dos trabajadores que desempeñan el mismo puesto, uno en la empresa N y el otro en la empresa T, tienen un salario de ¢744 000 y ¢820 000 respectivamente. En la empresa N, el salario promedio es ¢688 000, con una desviación estándar de ¢46 500, mientras en la empresa T, el salario promedio es ¢818 000, con una desviación estándar de ¢58 900.

De acuerdo con la Información anterior, considere las siguientes proposiciones:

I. Los salarios en la empresa T poseen una menor variabilidad relativa que los salarios en la empresa N.

II. El salario del trabajador de la empresa N ocupa una mejor posición relativa, con respecto a los salarios en su empresa, que el salario del trabajador de la empresa T, con respecto a los salarios en su empresa.

De ellas, ¿cuál o cuáles son verdaderas?

[3] Tomado de: MEP. (2018). Coeficiente de variación y estandarización. https://www.reformamatematica.net/wp-content/uploads/2018/10/20181015-Solucionario-EyP-Coeficiente-de-v8ariaci%C3%B3n-y-estandarizaci%C3%B3n.pdf

Ejercicio 3[4]

Un pediatra obtuvo la siguiente tabla sobre los meses de edad de 50 niños de su consulta en el momento de caminar por primera vez.

Meses	Niños
9	1
10	4
11	9
12	16
13	11
14	8
15	1

Calcular las medidas de tendencia central para los datos anteriores.

Ejercicio 4

Los salarios anuales de 4 individuos son 15,000, 16,000, 16,500, y 40,000

A) Calcular las medidas de tendencia central para los datos.

B) ¿Qué tipo de simetría presentan los datos?

C) Calcular la desviación estándar.

[4] Tomado de: https://sites.google.com/site/medidasddetendenciacentral/6-tarea

Ejercicio 5

Considere la siguiente situación:

La atleta nacional Andrea Vargas compitió en el mundial de atletismo en el 2018 y en el 2019, obteniendo los siguientes resultados.

Año 2018	Año 2019
Posición general: 5	Posición general: 6
Tiempo en 100 metos: 10,023	Tiempo en 100 metros: 9.998
Promedio general: 10, 112	Promedio general: 9,9998
Desviación estándar: 1,34	Desviación estándar: 2, 0001

De acuerdo con los datos, en qué año obtuvo el mejor rendimiento deportivo.

Solución 1[5]

A partir de la información de la tabla se evaluarán las dos proposiciones planteadas:

Proposición I: “La nota más alta obtenida por un estudiante da la sección 8 - 2, tiene una mejor posición relativa que la nota más alta obtenida por un estudiante de la sección 8-1, con respecto a las notas de su correspondiente sección”

Lo primero que hay tener claro, es que las notas de la sección 8-1 no se pueden comparar en términos absolutos con las de la sección 8-2, debido a que son dos distribuciones con diferente comportamiento. Es por esto que se deben estandarizar los datos (notas) para que puedan ser comparables.

Nota más alta de la (Sección 8-1):

$$\frac{80 - 50}{10} = 3{,}0$$

Nota más alta (Sección 8-2):

$$\frac{78 - 52}{8} = 3{,}25$$

Como el valor estandarizado de la sección 8-2 (3,25) es mayor que el del sección 8-1 (3,0), entonces se concluye que la proposición I es verdadera.

Proposición II: “Las notas de los estudiantes de la sección 8-1 presentan mayor variabilidad relativa que las notas de los estudiantes de la sección 8- 2”

La variabilidad relativa corresponde al coeficiente de variación por lo que se calculará el coeficiente en ambos casos.

Sección 8 - 1

$$c.v = \frac{10}{50} \cdot 100 = 20\%$$

Sección 8 – 2

$$c.v = \frac{8}{52} \cdot 100 = 15{,}38\%$$

Como el coeficiente de variación de la sección 8-1 es mayor que el de la sección 8-2

[5] Tomado de: MEP. (2018). Coeficiente de variación y estandarización. https://www.reformamatematica.net/wp-content/uploads/2018/10/20181015-Solucionario-EyP-Coeficiente-de-v8ariaci%C3%B3n-y-estandarizaci%C3%B3n.pdf

entonces se concluye que la proposición II es verdadera.

Solución 2[6]

Lo primero que hay tener claro, es que los salarios de la empresa N y la empresa T no son comparables en términos absolutos debido a que las distribuciones de datos presentan un comportamiento diferente.

De acuerdo a la información presentada en el ítem se evaluarán las dos proposiciones planteadas.

Proposición I: “Los salarios en la empresa T poseen una menor variabilidad relativa que los salarios en la empresa N.”

Al ser varibiliad relativa se calculará el coeficiente de variación de la empresa T

$$c.v = \frac{58900}{818000} \cdot 100 = 7{,}2\%$$

Para la empresa N

$$c.v = \frac{46500}{688000} \cdot 100 = 7{,}76\%$$

Como el coeficiente de variación de la empresa T es mayor al de la empresa N, se concluye que la proposición I es falsa.

Proposición II: “El salario del trabajador de la empresa N ocupa una mejor posición relativa, con respecto a los salarios en su empresa, que el salario del trabajador de la empresa T, con respecto a los salarios en su empresa”

Para que los salarios sean comparables se les debe estandarizar, con esta medida se elimina el efecto de la unidad de medida de los datos debido a que se modifica su posición, llevándolos a un estándar comparativo. Un dato estandarizado puede ser positivo, negativo o incluso nulo.

Empresa T

$$\frac{820000 - 818000}{58900} = 0{,}034$$

[6] Tomado de: MEP. (2018). Coeficiente de variación y estandarización. https://www.reformamatematica.net/wp-content/uploads/2018/10/20181015-Solucionario-EyP-Coeficiente-de-v8ariaci%C3%B3n-y-estandarizaci%C3%B3n.pdf

Empresa N

$$\frac{744000 - 688000}{46500} = 1{,}204$$

Como el salario estandarizado de la empresa N es un valor mayor al de la empresa T, se concluye que la proposición II es verdadera.

Solución 3

Meses	Niños
9	1
10	4
11	9
12	16
13	11
14	8
15	1

Se realizará el cálculo del promedio

$$\bar{x} = \frac{1 \cdot 9 + 4 \cdot 10 + 9 \cdot 11 + 16 \cdot 12 + 11 \cdot 13 + 8 \cdot 14 + 1 \cdot 15}{50}$$

$$\bar{x} = 12{,}2$$

Cálculo de la mediana

Al ser 50 datos, se debe tomar en consideración los datos en la posiciones 25 y 26, para promediarlos. Sin embargo, ambas posiciones tienen al número 12, por lo que la mediana es 12.

Cálculo de la moda

El dato con mayor frecuencia absoluta es 12.

Solución 4

Los salarios anuales de 4 individuos son 15 000, 16 000, 16 500, y 40 000

Promedio

$$\bar{x} = \frac{15\,000 + 16\,000 + 16\,500 + 40\,000}{4} = 21\,875$$

Mediana

Al ser 4 elementos, se deben promediar los datos dos y tres. Por lo que

$$M_e = \frac{16000 + 16500}{2} = 16250$$

Moda

Estos elementos no presentan una moda.

Simetría

Debido a que se cumple que

$$\bar{x} > M_e$$

Entonces se tiene una simetría hacia la derecha.

Desviación estándar.

Se calculan las restas entre cada dato y el promedio

$$\begin{aligned} 15\,000 - 21\,875 &= -6875 \\ 16\,000 - 21\,875 &= -5875 \\ 16\,500 - 21\,875 &= -5375 \\ 40\,000 - 21\,875 &= 18\,125 \end{aligned}$$

Luego se elevan al cuadrado, los resultados de la resta.

$$\begin{aligned} (-6875)^2 &= 47\,265\,625 \\ (-5875)^2 &= 34\,515\,625 \\ (-5375)^2 &= 28\,890\,625 \\ (18\,125)^2 &= 328\,515\,625 \end{aligned}$$

Seguidamente se suman los resultados anteriores

$$439\,187\,500$$

Luego se divide entre $n - 1$ datos, y como se tienen 4 datos se debe dividir entre 3.

$$146\,395\,833,33$$

Por último, se calcula la raíz cuadrada.

$$s = 12\ 099{,}41$$

<u>Solución 6</u>

Año 2018	**Año 2019**
Posición general: 5	Posición general: 6
Tiempo en 100 metos: 10,023	Tiempo en 100 metros: 9.998
Promedio general: 10, 112	Promedio general: 9,9998
Desviación estándar: 1,34	Desviación estándar: 2, 0001

En primera instancia por la posición entre las corredoras parece que el año que le fue mejor fue el 2018, sin embargo, se debe realizar el procedimiento de estandarización para determinar si le fue mejor.

Para el 2018

$$V.E = \frac{10{,}112 - 10{,}023}{1{,}34}$$
$$V.E = 0{,}06641$$

Para el 2019

$$V.E = \frac{9{,}9998 - 9{,}998}{2{,}0001}$$
$$V.E = 0{,}00089$$

Por lo tanto el año que le fue de una mejor manera, pese a la posición obtenida, fue el 2019.

Referencias bibliográficas

Anderson, D. Sweeney, D. Williams, T. (2016). Estadística para negocios y economía. https://gc.scalahed.com/recursos/files/r157r/w13118w/Estad%20para%20Neg_1aEd_03.pdf

MEP. (2018). Coeficiente de variación y estandarización. https://www.reformamatematica.net/wp-content/uploads/2018/10/20181015-Solucionario-EyP-Coeficiente-de-variaci%C3%B3n-y-estandarizaci%C3%B3n.pdf

Universidad Nacional Autónoma de México. (sf). Medidas de tendencia central y de dispersión. http://www.cuautitlan.unam.mx

San Marcos

www.usanmarcos.ac.cr

San José, Costa Rica