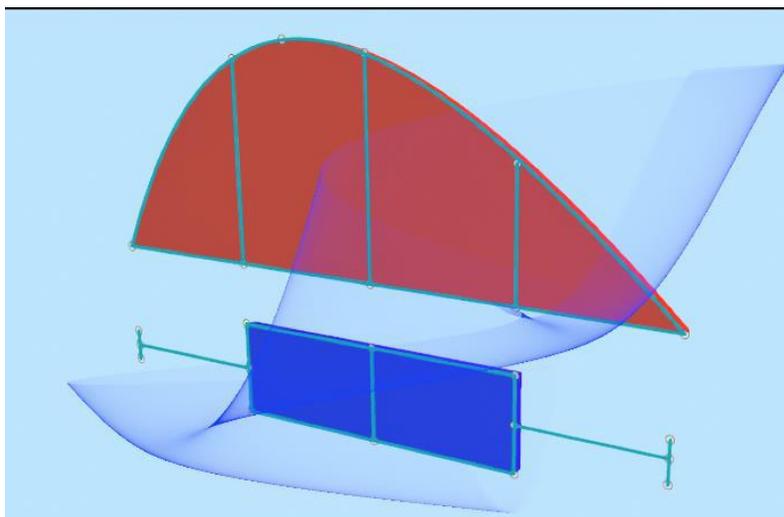


REFORMA DE LA EDUCACIÓN MATEMÁTICA EN COSTA RICA



Colección *Preparación Matemáticas Bachillerato*



ESTADÍSTICA Y PROBABILIDAD: **Medidas estadísticas**

Material complementario

Costa Rica

2017

Contenido

Índice de conceptos	3
I. Introducción	4
1. Uso de medidas estadísticas de posición: medidas de tendencia central	4
La media aritmética o promedio	4
La mediana	4
La moda	5
Problema 1. Enfermos por picadura de mosquito	5
Problema 2. Consumo promedio anual de electricidad por abonado residencial	7
Problema 3. Rendimiento en combustible de los vehículos	8
2. Relación gráfica entre las medidas de tendencia central	11
Problema 4. Comparación del rendimiento entre grupos	12
Problema 5. Años de experiencia en labores docentes	14
3. Otras medidas estadísticas de posición: medidas de orden	15
Los cuartiles	15
El máximo y el mínimo:	17
Problema 6. Entrenamiento para los 100 metros planos	17
Problema 7. Rendimiento en combustible de los vehículos	19
4. Representación de la distribución de datos mediante un diagrama de cajas	21
Diagrama de cajas	21
Problema 8. Vida útil de los bombillos	22
Problema 9. Defunciones por accidentes de tránsito según provincia	24
5. Media aritmética ponderada o promedio ponderado	25
Promedio ponderado	25
Problema 9. Ganancia por venta de zapatos	25
Problema 10. Nota promedio de un curso	26
Problema 11. Insecticida contra cucarachas	27
6. Uso de medidas estadísticas de variabilidad	29
Problema 12. Enfermos por picadura de mosquito	29
Problema 13. Contrato de trabajo con empresa turística	29
El recorrido	31
Problema 14. Constancia en el lanzamiento del martillo	32
El recorrido intercuartílico	34
Problema 15. Rendimiento en combustible de los vehículos	34
Variancia poblacional:	36

Variación muestral:	36
Desviación estándar:	36
Problema 16. Enfermos por picadura de mosquito	37
Problema 17. Comparación de notas en Español	40
Problema 18. Contrato de trabajo con empresa turística	44
II. Bibliografía	47
III. Créditos	47

Índice de conceptos

[Cuartiles](#) (p. 15)

[Desviación estándar](#) (p. 36)

[Diagrama de cajas](#) (p. 21)

[Máximo](#) (p. 17)

[Media aritmética o promedio](#) (p. 4)

[Mediana](#) (p. 4)

[Mínimo](#) (p. 17)

[Moda](#) (p. 5)

[Promedio ponderado](#) (p. 25)

[Recorrido](#) (p. 31)

[Recorrido intercuartílico](#) (p. 34)

[Relación gráfica entre las medidas de tendencia central](#)s (p. 11)

[Variación muestral](#) (p. 36)

[Variación poblacional](#) (p. 36)

I. Introducción

El presente documento ha sido elaborado por el Proyecto Reforma de la Educación Matemática en Costa Rica. Su propósito básico consiste en apoyar a estudiantes para la preparación de las Pruebas Nacionales de Matemáticas en el análisis de las medidas estadísticas.

Este documento complementa el Mini MOOC denominado *Medidas estadísticas* de la colección *Preparación Matemáticas Bachillerato* (<http://minimoocs.reformamatematica.net>) diseñado por este mismo proyecto.

Se analizan diferentes conocimientos vinculados con el uso de las medidas estadísticas de posición y variabilidad para la resolución de problemas de acuerdo con las temáticas incluídas en los Programas de Estudios de Matemáticas para la Educación Diversificada. Se incluyen además

Esperamos que el material sea de provecho para todos aquellos lectores que lo utilicen.

1. Uso de medidas estadísticas de posición: medidas de tendencia central

La media aritmética o promedio

Corresponde al valor numérico que se obtiene de sumar todos los datos y dividirlos por el total de datos. La fórmula es:

$$\text{media aritmética: } \frac{\text{Suma de datos}}{\text{Total de datos}}$$

[Volver al índice de conceptos](#)

La mediana

Divide al grupo de datos en subconjuntos y se ubica en la mitad, cumple la siguiente propiedad: el 50% de los datos toma un valor numérico menor o igual que la mediana y otro 50% tiene un valor numérico mayor o igual que la mediana.

Por ejemplo, si se tienen n datos ordenados de menor a mayor $X_1 \leq X_2 \leq \dots \leq X_n$, (el subíndice representa la posición de cada dato, de menor a mayor, X_1 es el menor dato, le sigue X_2 , así sucesivamente) entonces la mediana se determina por:

$$\text{Mediana: } \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{si el } n \text{ es impar} \\ \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} & \text{si el } n \text{ es par} \end{cases}$$

Nota: Si n es impar, entonces $n+1$ es un número par y entonces $\frac{n+1}{2}$ es un número entero, por esta razón el dato $X_{(\frac{n+1}{2})}$ corresponde al dato que se encuentra en la posición $\frac{n+1}{2}$.

Por otro lado, si n es par, entonces $\frac{n}{2}$ es un número entero y también su $\frac{n}{2} + 1$ que sería el número entero consecutivo, entonces $X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}$ representa la suma de los datos que están en estas posiciones

[Volver al índice de conceptos](#)

La moda

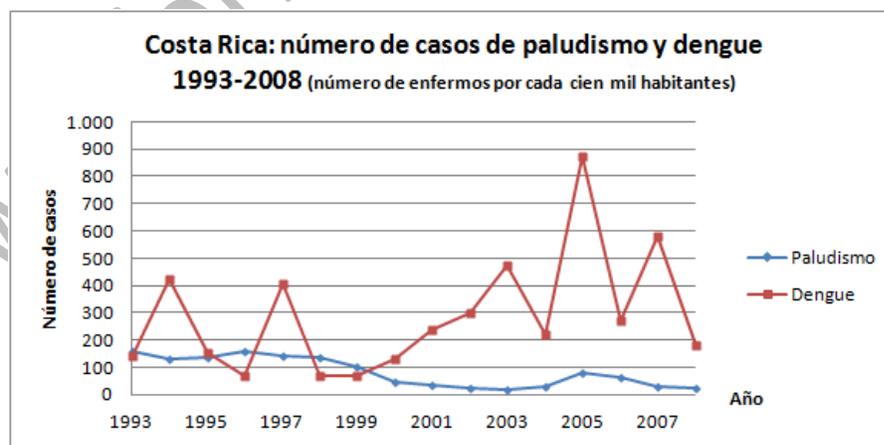
Corresponde al valor numérico que más se repite dentro de un grupo de datos. En un grupo de datos puede presentarse una única moda, más de una moda, o podría no existir este valor.

Nota: Las medidas estadísticas: media aritmética, mediana y moda se les llama medidas de tendencia central, porque se ubican hacia el centro de la distribución de los datos.

[Volver al índice de conceptos](#)

Problema 1. Enfermos por picadura de mosquito

Existen diversas enfermedades que se adquieren por la picadura de mosquitos, dos de las más conocidas son el paludismo y el dengue. La siguiente gráfica muestra el número de casos de estas enfermedades que se presentó en el país entre 1993 y 2008.



Fuente: <http://www.estadonacion.or.cr/index.php/estadisticas/costa-rica/compendio-estadistico/estadisticas-sociales>

Considere las siguientes proposiciones que se relacionan con la lectura realizada al gráfico:

- I. Para el período 1993-2008, el número promedio anual de enfermos por dengue es superior al número promedio anual de enfermos por paludismo.

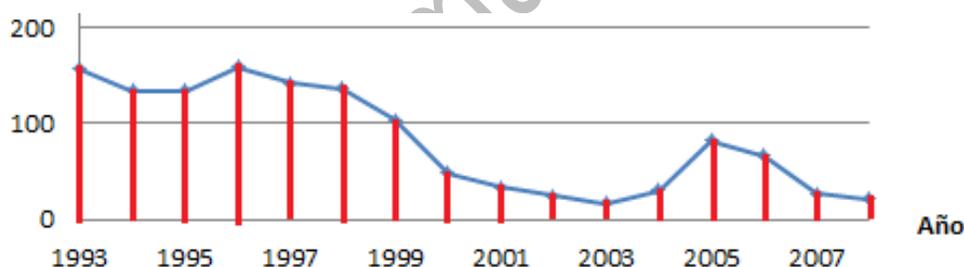
- II. Para el período 1993-2008, el número promedio anual de enfermos por paludismo fue menor de 100 casos por cada cien mil habitantes.

Determine si las proposiciones anteriores son verdaderas o falsas.

Solución:

Hay que destacar que la representación gráfica anterior muestra un índice de salud que corresponde al número de enfermos por cada cien mil habitantes para un período de 18 años (de 1993 hasta el 2008). Según esta representación en la mayoría de los años analizados el número de enfermos por dengue superó al de enfermos por paludismo (esto último se puede observar ya que hay una mayor cantidad de años en el período en donde la gráfica del dengue está por encima que la gráfica del paludismo), estas diferencias fueron más grandes después de 1999. Entonces para este período del estudio, al sumar el número de enfermos por dengue se obtiene un valor mayor que la suma del número de enfermos por paludismo. Por ello es claro que el número promedio anual de enfermos por dengue superó al número promedio anual de enfermos por paludismo.

Por otro lado, si observamos solamente el comportamiento de la curva correspondiente a los casos de paludismo, vemos que entre 1993 y 1999 se presentaron más de cien casos por cada cien mil habitantes, pero después de ese año el número de casos fue menor a cien como se muestra en la siguiente imagen.

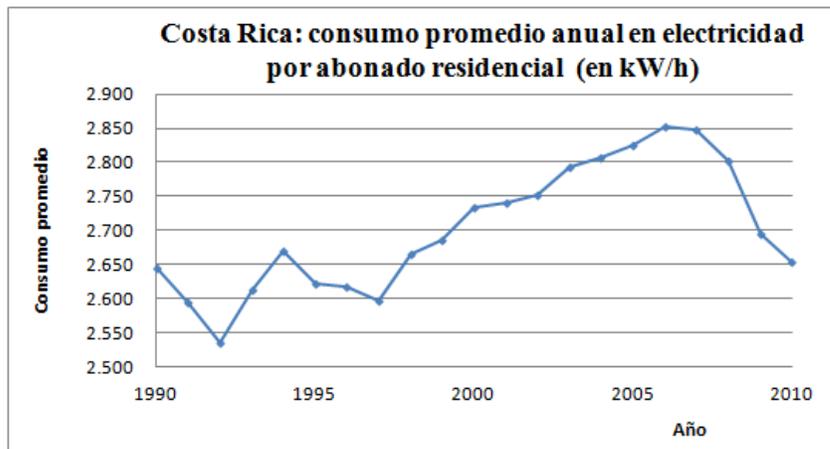


Bajo el supuesto que el gráfico está construido con una escala adecuada, al extraer el promedio de las magnitudes de los segmentos que se representan con rojo que corresponden al número anual de casos de paludismo por cada cien mil habitantes, se puede observar visualmente que este promedio debería ser menor de 100.

Por lo anterior se puede decir que las dos proposiciones son verdaderas.

Problema 2. Consumo promedio anual de electricidad por abonado residencial

Observe la siguiente representación gráfica.



Fuente: <http://www.estadonacion.or.cr/index.php/estadisticas/costa-rica/compendio-estadistico/estad-ambientales>

Analice las siguientes afirmaciones vinculadas con la información que presenta el gráfico anterior:

- I. El consumo total de electricidad entre los abonados residenciales fue aproximadamente igual en 1990 y en el 2010.
- II. El máximo consumo de electricidad presentado en una vivienda fue de 2850 kW/h.

Determine cuál o cuáles de estas afirmaciones se puede asegurar con certeza que sean verdaderas:

Solución:

Es evidente que en 1990 había mucho menos abonados residenciales que en el 2010. Por esta razón, no es suficiente saber que el consumo promedio anual de electricidad (en kW/h) fue el mismo para 1990 y 2010 (como lo presenta la gráfica) para concluir que el consumo total también haya sido aproximadamente igual en esos años, puesto que ello depende del número total de abonados residenciales en cada uno de dichos años. Hay que recordar que el promedio de cada año se determina por la fórmula:

$$\text{Promedio} = \frac{\text{Consumo total en electricidad de los abonados residenciales}}{\text{Total abonados residenciales}}$$

Entonces si los consumos promedio por abonado residencial de 1990 y 2010 son iguales y el número total de abonados residenciales son diferentes, también el consumo total de electricidad de los abonados residenciales en ambos años son diferentes.

Por otro lado, la gráfica ilustra el consumo promedio anual de electricidad y ello no permite asegurar el valor del máximo consumo de electricidad en una vivienda particular, basta con analizar la fórmula de cálculo del promedio, para identificar que con la información del gráfico

no es posible determinar el consumo total de electricidad por vivienda ni tampoco el acumulado para todas las viviendas.

Por lo anterior, ambas proposiciones son falsas.

Problema 3. Rendimiento en combustible de los vehículos

Pilar y Beatriz discuten acerca del rendimiento de sus automóviles con respecto al gasto de combustible. Cada una de ellas indica que su vehículo es más económico.



<http://www.diariomotor.com/>

Deciden medir el rendimiento de sus vehículos en los días laborales (5 días por semana). Al iniciar cada día llenan completamente el tanque de combustible y miden el kilometraje recorrido. Al día siguiente vuelven a realizar el proceso para determinar los litros de combustible consumidos. Después de dos semanas, tienen información correspondiente a 10 días. Al dividir los kilómetros recorridos por los litros consumidos obtuvieron el rendimiento en kilómetros por litro de combustible.

Rendimiento en kilómetros por litro de los vehículos de Pilar y Beatriz en 10 días laborales.

Vehículo de Pilar			Vehículo de Beatriz		
Km recorridos	Litros consumidos	Rendimiento (Km por litro)	Km recorridos	Litros consumidos	Rendimiento (km por litro)
28,62	2,37	12,1	43,15	3,96	10,9
28,68	2,17	13,2	34,84	2,54	13,7
30,05	2,40	12,5	35,99	3,76	9,6
30,50	2,62	11,6	37,77	3,12	12,1
30,73	2,05	15,0	38,85	2,96	13,1
28,42	2,14	13,3	40,92	2,88	14,2
27,85	2,13	13,1	38,77	4,07	9,5
27,26	2,33	11,7	38,01	2,71	14,0
32,96	2,44	13,5	39,36	3,04	12,9
29,10	2,02	14,4	40,84	3,36	12,2

Información ficticia utilizada con fines didácticos

Sin embargo, no saben cómo analizar esta información para realizar una comparación que permita identificar cuál de los dos autos tuvo un mejor rendimiento en las dos semanas observadas. Utilice las medidas de tendencia central para ayudar a Pilar y Beatriz a resolver el problema que enfrentan.

Solución:

En este problema vemos el efecto de la variabilidad en las situaciones de la cotidianidad, pues aunque tenemos los mismos autos y se someten a recorridos similares, el rendimiento varía de un día a otro, lo cual puede deberse a múltiples causas además del funcionamiento del vehículo: embotellamientos vehiculares, estilos de conducción, estado de las carreteras, entre otras. El reto para resolver el problema consiste en resumir la información por medio de alguna estrategia que permita comparar el rendimiento entre los autos.

En estos casos las medidas de posición constituyen una buena alternativa. Regularmente, las personas recurren al promedio o media aritmética de datos, la cual se define de la siguiente manera:

En el problema que estamos analizando, para el auto de Pilar la suma de los rendimientos es 130,4 km/L. Entonces la media aritmética del rendimiento sería: $\frac{130,4}{10}$ km/L = 13,04 km/L.

Para el auto de Beatriz la suma de los rendimientos es 122,2 km/L. Entonces la media aritmética del rendimiento sería: $\frac{122,2}{10}$ km/L = 12,22 km/L.

De acuerdo con lo anterior, para los diez días observados, el rendimiento promedio del auto de Beatriz superó el rendimiento promedio del auto de Pilar.

Sin embargo, desde el punto de vista estadístico no es adecuado recurrir a una sola medida de posición, sino que conviene analizar otras medidas que nos ayuden a complementar el mensaje. En este sentido mediana permite complementar el mensaje que comunica la media aritmética. Seguidamente se presenta la definición de la mediana:

Para nuestro problema, si ordenamos los rendimientos de menor a mayor en cada caso se tendría lo siguiente:

Rendimiento en kilómetros por litro			
Auto de Pilar		Auto de Beatriz	
1	11,6	1	9,5
2	11,7	2	9,6
3	12,1	3	10,9
4	12,5	4	12,1
5	13,1	5	12,2
6	13,2	6	12,9
7	13,3	7	13,1
8	13,5	8	13,7
9	14,4	9	14,0
10	15,0	10	14,2

Como puede notarse en cada caso $n = 10$ datos, que corresponde a un valor par, entonces la mediana se calcula mediante la fórmula:

$$\frac{X_{(\frac{10}{2})} + X_{(\frac{10}{2}+1)}}{2} = \frac{X_5 + X_6}{2}$$

Nota: Como n es 10, entonces la mediana se encuentra entre los datos que están en la posición $\frac{10}{2} = 5$ y en la posición $\frac{10}{2} + 1 = 5 + 1 = 6$. En el caso del auto de Pilar, en la posición cinco está el número 13,1 y en la posición seis está el número 13,2.

La mediana en el rendimiento del auto de Pilar es $\frac{13,1+13,2}{2}$ km/L = 13,15 km/L. Esto quiere decir que en la mitad de los días observados el rendimiento en el auto de Pilar fue de 13,15 km/L o menos, y en la otra mitad de días el rendimiento fue mayor o igual que ese valor.

La mediana en el rendimiento del auto de Beatriz es $\frac{12,2+12,9}{2}$ km/L = 12,55 km/L. Entonces en la mitad de los días observados el rendimiento en el auto de Beatriz fue de 12,55 km/L o menos, y en la otra mitad de días el rendimiento fue mayor o igual que ese valor.

También la mediana en el rendimiento del auto de Pilar toma un valor más alto que la correspondiente al auto de Beatriz, para los diez días observados.

Finalmente se tiene que para los datos del problema no hay modas.

En resumen, se tiene lo siguiente:

Rendimiento en kilómetros por litro de los autos de Pilar y Beatriz, información recabada por día, en 10 días consecutivos

Medida estadística de posición	Rendimiento de auto de Pilar Km/L	Rendimiento de auto de Beatriz Km/L
Media aritmética	13,04	12,22
Mediana	13,15	12,55
Moda	No tiene	No tiene

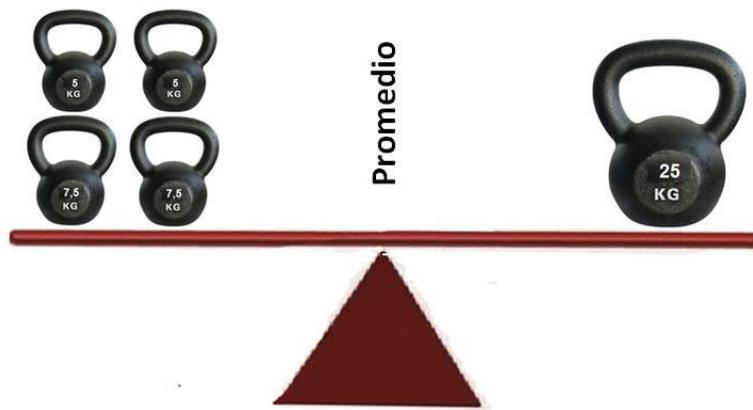
Tanto la media aritmética como la mediana de los rendimientos fueron superiores para el auto de Pilar, lo que da la impresión que este auto tiene un mayor rendimiento, no puede asegurarse que sea así en definitiva, porque se basa en resultados de unos pocos días, podría ser que si se observan más días la situación cambie.

[Volver al índice de conceptos](#)

2. Relación gráfica entre las medidas de tendencia central

La media aritmética, la mediana y la moda tienen una posición estratégica en el eje x cuando se visualizan gráficamente. Cuando se tiene un número grande de datos y se analiza su patrón de variabilidad entre las medidas de tendencia central (media aritmética, mediana o moda), existe una relación que se vincula directamente con la asimetría de la distribución de los datos.

Si la distribución de los datos es asimétrica (a la derecha o a la izquierda), los valores extremos provocan que la media aritmética se aparte un poco del lugar donde está la mayor concentración de datos



Observe que el peso promedio de estas cinco pesas es $\frac{5,0+5,0+7,5+7,5+25,0}{5}$ kg = 10 kg. Sin embargo, este valor de 10 kg no es representativo de la tendencia central en los pesos de las cinco pesas. En este caso el peso promedio se ve influenciado por la pesa de 25 kg, la cual provoca que el promedio sea más grande que el común de las pesas. Por esta razón, se dice que el promedio o media aritmética no es una medida adecuada para representar la tendencia central de los datos cuando existe fuerte asimetría positiva (a la derecha como en el caso de las pesas) o negativa (a la izquierda).

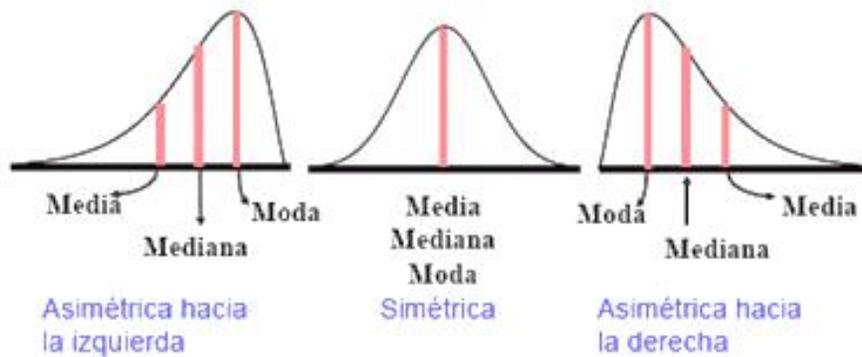
En cambio, independientemente del tipo de asimetría que presente la distribución de los datos, la mediana, siempre se ubica en el centro, de modo que no más del 50% de los datos son menores y no más del 50% son mayores.

Observe que para el caso de las pesas, al existir cinco datos, la mediana en peso estaría dada por el valor: $Mediana = X_{\left(\frac{5+1}{2}\right)} = X_3 = 7,5$ kg.



Por su parte la moda, se ubica donde hay más concentración de datos; sin embargo, es común que la moda tienda a variar mucho entre muestras diferentes. En nuestro ejemplo hay dos modas que son 5 kg y 7,5 kg.

En términos generales, para muestras suficientemente grandes en donde se presenta una única moda, la relación entre estas tres medidas de tendencia central se visualiza en el siguiente esquema:



Nota: La relación anterior es útil cuando los datos han sido resumidos previamente en una distribución de frecuencias y se tiene un patrón general de la forma de la distribución, pero cuando se tienen pocos datos la moda podría tener un comportamiento diferente del que se muestra en el esquema, por esta razón es conveniente concentrarse solamente en la comparación entre la mediana con la media aritmética.

Analice el siguiente problema que se vincula con los conceptos anteriores:

Problema 4. Comparación del rendimiento entre grupos

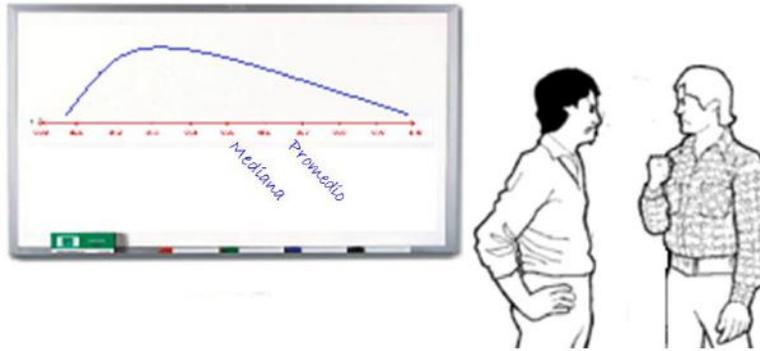
Dos profesores analizan los resultados de un examen en un curso universitario, para los cuales se presentan a continuación dos medidas de tendencia central.

Calificaciones (en escala de 0 a 100) de un examen parcial de los grupos A y B

Medida estadística	Grupo A	Grupo B
Media aritmética	85	77
Mediana	76	81

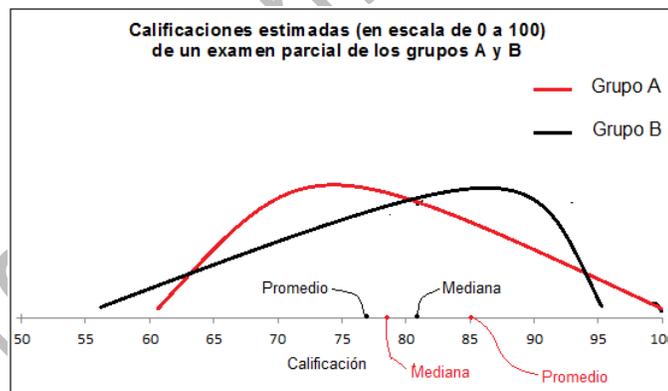
El profesor del grupo A se muestra muy orgulloso porque la calificación promedio de su grupo fue de 85, mientras que en el grupo B que corresponde al profesor fue apenas de 77.

Sin embargo, el profesor del grupo B le indica que, en términos generales, su grupo tuvo un mejor rendimiento debido a que la mayoría de sus estudiantes tuvieron una calificación por encima de 80, lo cual no ocurrió en el grupo A. El docente del grupo A le responde que eso no puede ser posible y que está manipulando la información. Usted debe decidir si el profesor del grupo B tiene o no tiene la razón.



Solución:

Con tan poca información resulta muy difícil determinar cuál de los grupos tuvo un mejor rendimiento. Además, el concepto de mejor rendimiento podría ser ambiguo, debido a que depende de qué quiere decir *un mejor rendimiento*. Sin embargo, las medidas estadísticas son claras en cuanto a su interpretación. Por ejemplo, en términos relativos, la media aritmética establece que en el grupo A la suma de las notas es mayor que en el grupo B. Por otro lado, la mediana señala que el 50% de los estudiantes del grupo B lograron calificaciones mayores o iguales a 81, lo que es un buen indicador para señalar que la mayoría de los estudiantes (más de la mitad) tuvo notas mayores o iguales a 81. Mientras que la mediana del grupo A señala que la mitad de los estudiantes tuvo calificaciones mayores o iguales a 76. Con esta información el docente del grupo B tiene razón en su afirmación. Un esquema muy básico que podría aproximar la distribución de calificaciones sería similar al siguiente:



En este esquema se observa que en el grupo B hay más estudiantes que tuvieron notas mayores que 80. Pero se debe tener claro que esto es simplemente una aproximación de la distribución de las calificaciones. Las distribuciones reales pueden ser diferentes, pero la afirmación del profesor B es correcta.

En este problema, las diferencias en el promedio pueden obedecer a la presencia de valores extremos hacia uno de los lados. Posiblemente, unas pocas calificaciones muy altas provocan que la calificación promedio del grupo A se incremente, mientras que unas pocas calificaciones muy bajas en el grupo B podrían haber provocado que la calificación promedio del grupo B disminuyera.

Problema 5. Años de experiencia en labores docentes

Considere la información del cuadro y las proposiciones dadas.

En el cuadro siguiente se muestran los años de experiencia que tienen 196 docentes de educación primaria que fueron encuestados en el año 2009.

Distribución de los docentes encuestados según número de años de experiencia en labores de docencia. Año 2009

Años de experiencia	Cantidad de docentes	Porcentaje de docentes
De 1 a menos de 5	28	14,3
De 5 a menos de 10	31	15,8
De 10 a menos de 15	37	18,9
De 15 a menos de 20	40	20,4
De 20 a menos de 25	48	24,5
De 25 a menos de 30	12	6,1
Total	196	100

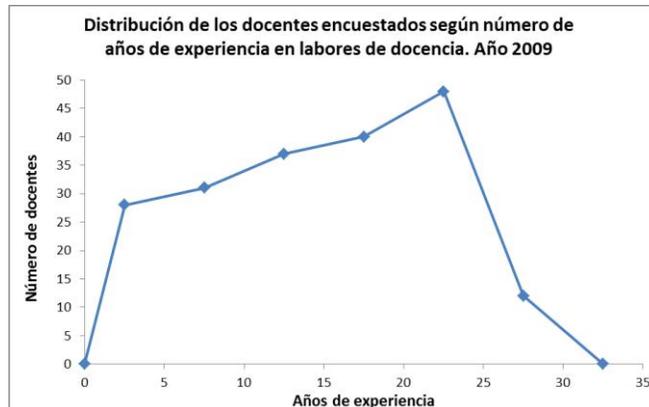
Con base en los datos del cuadro anterior analice las siguientes proposiciones:

- I. El número promedio de años de experiencia de los docentes encuestados es un valor entre 20 y 25 años.
- II. El valor de la mediana de la experiencia de los docentes encuestados se encuentra entre 15 y 20 años.

Determine cuál o cuáles de las proposiciones anteriores son verdaderas.

Solución:

Los datos sobre los años de experiencia de los docentes fueron resumidos en el cuadro que se presenta, por lo que se sabe cuáles fueron los datos individuales. Sin embargo, al observar el ordenamiento de los datos y las frecuencias correspondiente, el cuadro permite identificar una asimetría negativa.



En este sentido, de acuerdo con la relación que debe existir entre el promedio y la mediana, se supone que el número promedio de años de experiencia de los profesores debe ser menor que la mediana de los años de experiencia. Por otro lado, de acuerdo con la definición de la mediana, esta medida acumula el 50% de los datos, si analizamos los porcentajes acumulados, puede notarse que hasta la clase de 15 a 20 años de experiencia, se ha acumulado este 50%, lo cual significa que la mediana en los años de experiencia está en esta clase.

Años de experiencia	Cantidad de docentes	Porcentaje de docentes	Porcentaje acumulado
De 1 a menos de 5	28	14,3	14,3
De 5 a menos de 10	31	15,8	30,1
De 10 a menos de 15	37	18,9	49,0
De 15 a menos de 20	40	20,4	69,4
De 20 a menos de 25	48	24,5	93,9
De 25 a menos de 30	12	6,1	100,0
Total	196	100	

Por lo tanto la primera proposición sería falsa y la segunda verdadera.

Nota: Los análisis anteriores evidencian que en distribuciones asimétricas, positivas o negativas el promedio o media aritmética se ve afectado por la presencia de los valores extremos, entonces no es un buen indicador del comportamiento central de los datos. La mediana puede ser un mejor indicador de esta tendencia central. Pero si se tienen las tres medidas, lo ideal es realizar un análisis integral.

A pesar de lo anterior, desde un punto de vista estadístico la media aritmética sigue teniendo una gran importancia en los análisis más elaborados.

[Volver al índice de conceptos](#)

3. Otras medidas estadísticas de posición: medidas de orden

En algunos análisis estadísticos se requiere destacar una medida o valor hasta el cual se acumula un porcentaje o un cierto porcentaje de datos. En este sentido existen diversas medidas estadísticas destinadas a determinar estos valores. Unos de los más conocidos son los cuartiles:

Los cuartiles

Son tres valores que dividen al grupo de datos en cuatro partes. Tienen una función similar a la de la mediana. El primer cuartil o C_1 cumple que el 25% de los datos son menores o iguales a C_1 y el 75% son mayores o iguales a C_1 . El segundo cuartil o C_2 corresponde a la mediana. El tercer cuartil o C_3 cumple que el 75% de los datos son menores o iguales a C_3 y el 25% son mayores o iguales a C_3 .

Existen diferentes fórmulas de cálculo para los cuartiles, normalmente se presentan pequeñas diferencias en el resultado de aplicar estas fórmulas, pero esas diferencias no son importantes para los problemas que se aplican en secundaria, siempre que se utilice una misma fórmula al hacer comparaciones.

Al igual que con la mediana, lo primero que se debe considerar para calcular un cuartil es que los datos estén ordenados de menor a mayor. Entonces si se tienen n datos y los representamos con $X_1, X_2, X_3, \dots, X_n$ se cumple que:

$$X_1 \leq X_2 \leq X_3 \leq \dots \leq X_n$$

Una de las fórmulas más utilizadas para determinar los cuartiles es la siguiente:

Si k toma los valores 1, 2 o 3, entonces se calcula el valor $\frac{k \cdot (n+1)}{4}$:

- a) Si $\frac{k \cdot (n+1)}{4}$ es un entero, el cuartil número k sería el dato ordenado que se encuentre en esa posición, o sea:

$$\text{Cuartil } k: C_k = X_{\frac{k \cdot (n+1)}{4}}$$

- b) Si $\frac{k \cdot (n+1)}{4}$ no es entero, se determina la parte entera de este número, digamos M (corresponde al mayor número entero inferior a $\frac{k \cdot (n+1)}{4}$). Se dice que el cuartil k es un promedio entre los datos ordenados que están en la posición M y $M + 1$, o sea:

$$\text{Cuartil } k: C_k = \frac{X_M + X_{M+1}}{2}$$

Nota: De acuerdo con lo anterior el primer cuartil acumula el 25% de datos menores, el segundo cuartil acumula el 50% de los datos menores y el tercer cuartil acumula el 75% de los datos menores.

Nota: Observe que el segundo cuartil coincide con la definición de la mediana.

Nota: además de los cuartiles existen otras medidas similares tales como: los quintiles los cuales dividen el conjunto de datos en cinco grupos, los deciles que dividen el conjunto de datos en diez grupos, los percentiles que dividen el conjunto de datos en cien partes. Todas estas medidas cuartiles, quintiles, deciles y percentiles se les llama cuantiles. Para efectos de la educación secundaria solamente se analizan los cuartiles.

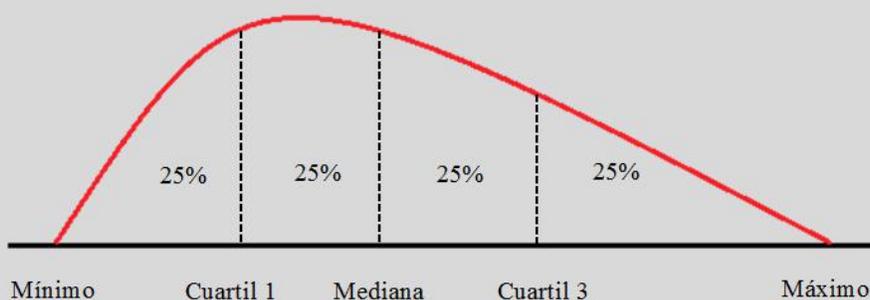
En muchos casos la identificación del valor más bajo o del más alto es de interés práctico, este hecho provoca que estos valores sean identificados también como medidas estadísticas.

El máximo y el mínimo:

Corresponde al mayor valor y al menor valor de grupo de datos respectivamente.

Nota: A las medidas estadísticas: mediana, cuartiles (en general a todos los cuantiles) y valores máximo y mínimo se les llama medidas estadísticas de orden, porque ordenan el conjunto de datos.

Cuanto estos valores se analizan en una representación gráfica de una variable continua, se puede interpretar el porcentaje de datos encerrados entre las cuantiles como el área bajo la curva:



Problema 6. Entrenamiento para los 100 metros planos

Un atleta de alto rendimiento se entrena para participar internacionalmente en los 100 metros planos (competencia atlética en la cual los corredores deben correr 100 metros en una pista en el menor tiempo posible). En los últimos días realizó 50 competencias de entrenamiento que fueron cronometradas. Debido a la variabilidad de los resultados, se observan tiempos de recorrido muy buenos, pero otros no tanto. Junto con su entrenador coinciden en que posiblemente un 25% de los tiempos no responden a su verdadero potencial debido a circunstancias técnicas especiales: una mala salida (salió tarde), una caída o resbalón durante la competencia, un roce con un compañero, mal estado de la pista, entre otras. Por esta razón, deciden eliminar el 25% de los tiempos más altos alcanzados en estas competencias, y analizar el resto de los resultados.



<http://caliescribe.com/deporte/mundial-atletismo-menores-2015>

Los tiempos obtenidos por este atleta en segundos para las 50 competencias fueron los siguientes:

12,38	10,29	10,79	10,76	9,95
10,41	11,53	9,97	10,30	9,98
12,10	10,63	9,91	10,51	9,94
10,24	10,10	12,50	10,95	12,46
10,35	10,57	10,83	10,63	10,48
10,43	10,18	10,78	10,20	10,00
10,70	10,24	9,95	10,09	12,18
9,86	11,94	10,85	10,84	10,25
10,03	10,30	10,43	10,29	10,13
9,92	13,54	12,51	10,30	10,31

Ayude al atleta y a su entrenador a determinar el valor hasta el cual se acumula el 75% de los mejores tiempos logrados por este (el 25% de los tiempos son de mayor magnitud) y al mismo tiempo determine el mejor tiempo de competencia logrado.

Solución:

Como puede notarse las fórmulas de los cuartiles pueden ayudarnos a resolver este problema. Específicamente, se requiere determinar el tercer cuartil relacionado con los datos de estos tiempos. Pero hay que tomar en cuenta que para determinar cualquiera de estos cuartiles, se requiere ordenar los datos de menor a mayor.

Posición	Tiempo								
1	9,86	11	10,09	21	10,30	31	10,57	41	10,95
2	9,91	12	10,10	22	10,30	32	10,63	42	11,53
3	9,92	13	10,13	23	10,30	33	10,63	43	11,94
4	9,94	14	10,18	24	10,31	34	10,70	44	12,10
5	9,95	15	10,20	25	10,35	35	10,76	45	12,18
6	9,95	16	10,24	26	10,41	36	10,78	46	12,38
7	9,97	17	10,24	27	10,43	37	10,79	47	12,46
8	9,98	18	10,25	28	10,43	38	10,83	48	12,50
9	10,00	19	10,29	29	10,48	39	10,84	49	12,51
10	10,03	20	10,29	30	10,51	40	10,85	50	13,54

Se tiene $n = 50$ datos. Este número es par, además, como buscamos calcular el tercer cuartil, se tiene que $k = 3$,

Al aplicar la fórmula $\frac{k \cdot (n+1)}{4}$ se tiene que $\frac{3 \cdot 51}{4} = 38,25$, entonces el mayor número entero inferior a 38,25 es 38, esto quiere decir que $M = 38$, con lo cual el tercer cuartil está entre el dato que está en la posición 38 ($X_{38} = 10,83$) y el dato de la posición 39 ($X_{39} = 10,84$):

$$\text{Cuartil 3: } C_3 = \frac{X_{38} + X_{38+1}}{2} = \frac{10,83 + 10,84}{2} = 10,835$$

Esto quiere decir que en el 75% de las competencias, los tiempos fueron menores o iguales que 10,835 segundos; mientras que en el otro 25% de las competencias, los tiempos fueron mayores o iguales a 10,835 segundos.

Para responder el problema les podemos decir al atleta y a su entrenador que valoren las competencias en las cuales el tiempo obtenido fue menor o igual a 10,835 segundos.

Por otro lado, para determinar el mejor tiempo de competencia basta con identificar el valor más bajo o tiempo mínimo obtenido en las 50 competencias, este valor fue de 9,86 segundos.

Nota: Anteriormente se han analizado diferentes medidas estadísticas de posición (algunas de tendencia central y otras de orden), cada una de ellas tiene una interpretación diferente, la comparación entre las diferentes medidas permite generar un análisis más integral del mensaje que comunican los datos para la resolución de un problema.

Problema 7. Rendimiento en combustible de los vehículos

Para ejemplificar las medidas de orden, volvamos a analizar el problema 9, en donde Pilar y Beatriz discutían sobre el rendimiento de combustible de sus vehículos (en kilómetros recorridos por litro). Para los 10 días observados, los datos de los rendimientos de estos vehículos ordenados de menor a mayor son:

Rendimiento en kilómetros por litro			
Auto de Pilar		Auto de Beatriz	
1	11,6	1	9,5
2	11,7	2	9,6
3	12,1	3	10,9
4	12,5	4	12,1
5	13,1	5	12,2
6	13,2	6	12,9
7	13,3	7	13,1
8	13,5	8	13,7
9	14,4	9	14,0
10	15,0	10	14,2

Analice el problema de acuerdo con la información que proporcionan las medidas de orden para indicar cuál de los vehículos tiene un mejor rendimiento.

Solución:

Para responder el problema se debe determinar las siguientes medidas de orden: mínimo, primer cuartil, mediana, tercer cuartil y máximo (hay que recordar que el segundo cuartil es la mediana)

Según la fórmula de los cuartiles, el primer cuartil: $k = 1$, como $n = 10$ entonces debemos calcular la expresión $\frac{k \cdot (n+1)}{4} = \frac{1 \cdot 11}{4} = 2,75$, debido a que no es entero entonces la parte entera 2,75 es $M = 2$, quiere decir que el primer cuartil es el promedio de los datos dos y tres:

Para el auto de Pilar: Cuartil 1: $C_1 = \frac{X_2 + X_3}{2} = \frac{11,7 + 12,1}{2}$ km/L = 11,9 km/L, lo que significa que en el 25% de los días el rendimiento fue menor o igual a 11,9 km/L, y en el 75% de los días el rendimiento del vehículo fue mayor o igual a este valor.

Para el rendimiento del auto de Beatriz: Cuartil 1: $C_1 = \frac{X_2 + X_3}{2} = \frac{9,6 + 10,9}{2}$ km/L = 10,25 km/L, la interpretación es similar.

Tercer cuartil: $k = 3$, como $n = 10$ entonces debemos calcular la expresión $\frac{k \cdot (n+1)}{4} = \frac{3 \cdot 11}{4} = 8,25$, debido a que no es entero entonces la parte entera de 8,25 es $M = 8$. Quiere decir que el tercer cuartil es el promedio de los datos siete y ocho:

Para el rendimiento del auto de Pilar:

Cuartil 3: $C_3 = \frac{X_8 + X_9}{2} = \frac{13,5 + 14,4}{2}$ km/L = 13,95 km/L, lo que significa que en el 75% de los días el rendimiento fue menor o igual a 13,95 km/L y en el 25% de los días el rendimiento del vehículo fue mayor o igual a este valor.

Para el rendimiento del auto de Beatriz:

Cuartil 3: $C_3 = \frac{X_8 + X_9}{2} = \frac{13,7 + 14,0}{2}$ km/L = 13,85 km/L, la interpretación es similar.

Para determinar la mediana, como $n = 10$ y representa un número par, entonces la mediana se encuentra los datos $X_{\frac{10}{2}}$ y $X_{\frac{10}{2}+1}$ (revisar fórmula de la mediana) entonces:

$$\text{Mediana} = \frac{X_5 + X_6}{2}$$

La mediana en el rendimiento del auto de Pilar es $\frac{13,1 + 13,2}{2}$ km/L = 13,15 km/L. Esto quiere decir que en la mitad de los días observados el rendimiento del auto de Pilar fue de 13,15 km/L o menos y en la otra mitad de días el rendimiento fue mayor o igual que ese valor.

La mediana en el rendimiento del auto de Beatriz es $\frac{12,2 + 12,9}{2}$ km/L = 12,55 km/L. Entonces en la mitad de los días observados el rendimiento del auto de Beatriz fue de 12,55 km/L o menos y en la otra mitad de días el rendimiento fue mayor o igual que ese valor.

En cuanto a los valor máximo y valor mínimo se tiene que para el auto de Pilar el mayor rendimiento fue de 15,0 km/L y el menor 11,6 km/L. Para el vehículo de Beatriz el mayor rendimiento fue de 14,2 km/L y el menor 9,5 km/L.

Para una mejor comparación, las medidas estadísticas anteriores pueden ser resumidas en un cuadro (se incluyó también la mediana que fue calculada en el problema 9):

Rendimiento en kilómetros por litro de los autos de Pilar y Beatriz, información recabada por día, en 10 días consecutivos

Medida estadística de orden	Rendimiento de auto de Pilar Km/L	Rendimiento de auto de Beatriz Km/L
Primer cuartil	11,9	10,25
Mediana	13,15	12,55
Tercer cuartil	13,95	13,85
Máximo	15,0	14,2
Mínimo	11,6	9,5

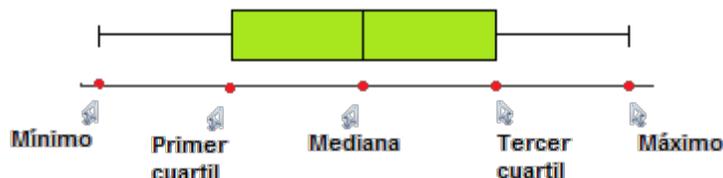
En conclusión, en las medidas estadísticas de orden analizadas, el auto de Pilar mostró un mayor rendimiento, por ello puede concluirse que el vehículo de Pilar tiene un mejor rendimiento en el consumo de combustible por kilómetro recorrido en los diez días en que se recolectaron los datos, según la información de los 10 días muestreados.

[Volver al índice de conceptos](#)

4. Representación de la distribución de datos mediante un diagrama de cajas

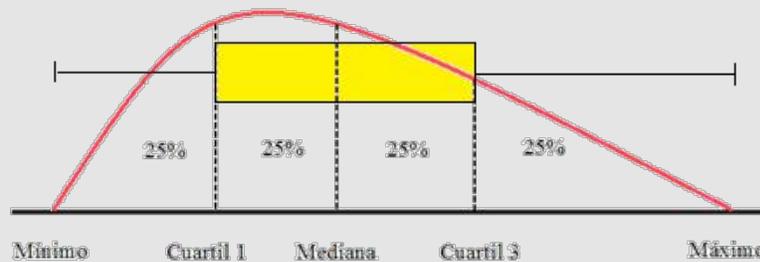
Diagrama de cajas

Como se mencionó anteriormente, las medidas estadísticas de orden tienen una interpretación gráfica que favorece una visualización práctica sobre el patrón de la distribución de los datos. En este sentido los diagramas de cajas se convierten en una importante herramienta estadística para llevar a cabo esta visualización. Un diagrama de cajas, también llamado gráfico de bigotes, consiste en una representación estadística en la cual se incluyen las principales medidas de orden que hemos analizado previamente. Cada caja debe estar en escala según el valor de la medida estadística en el eje x. Entre el primer cuartil y el tercer cuartil se dibuja una caja tal como se muestra. Esta caja se une con el mínimo y el máximo por medio de segmentos. Finalmente la mediana se representa con una línea que divide la caja.



De acuerdo con lo anterior, un diagrama de cajas proporciona información de estas cinco medidas de orden, pero además, son un indicador de la forma aproximada que tiene la distribución de datos. Sin embargo, no se puede perder de vista que estos diagramas son simplemente un resumen de los datos, por medio de cinco medidas. Es importante mencionar que hay que ser muy cuidadosos al interpretar la información que comunican.

Nota: Observe que por medio de diagrama de cajas se puede visualizar en forma aproximada la distribución general de los datos, tal como se hace con un polígono de frecuencias.



Observe que la distribución de los datos representada por la curva muestra una ligera asimetría positiva, esto se evidencia también en diagrama de cajas, debido a que la caja está sesgada hacia la izquierda.

Problema 8. Vida útil de los bombillos

Suponga que le envían a comprar unos bombillos a una ferretería, la persona que le atiende indica que de acuerdo con las características del bombillo que usted está solicitando tiene dos marcas con el mismo precio: A y B. Le pregunta que cuál de ellas desea. Usted recuerda que en un artículo de una revista que leyó recientemente salió publicado la siguiente representación gráfica sobre la vida útil de estos bombillos:

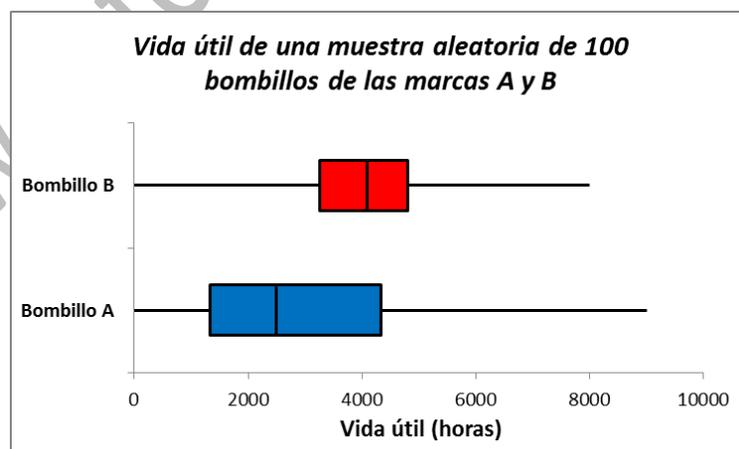
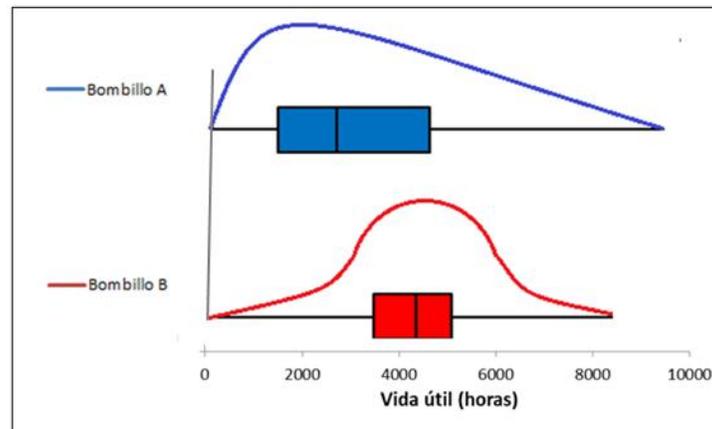


Imagen propia elaborada con fines didácticos

Utilice esta información para decidir qué marca de bombillos compraría y argumente esta decisión.

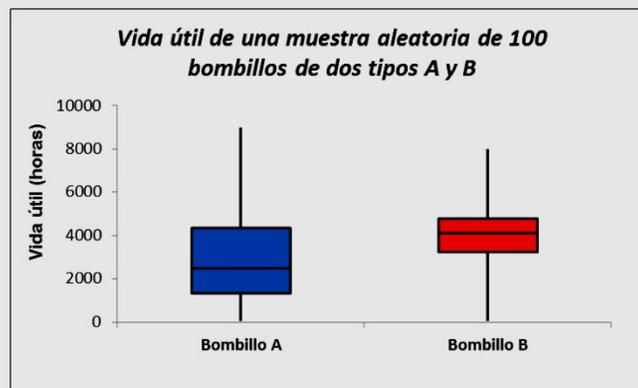
Solución

Note que en la caja correspondiente, a cada tipo de bombillo se puede visualizar el tipo de asimetría que presenta la distribución de los datos de acuerdo con la representación gráfica que apareció en el problema 3.



La información que proporciona el diagrama de cajas es evidencia suficiente para decidir adquirir los bombillos B debido a que existe mayor posibilidad de que tenga una más duración.

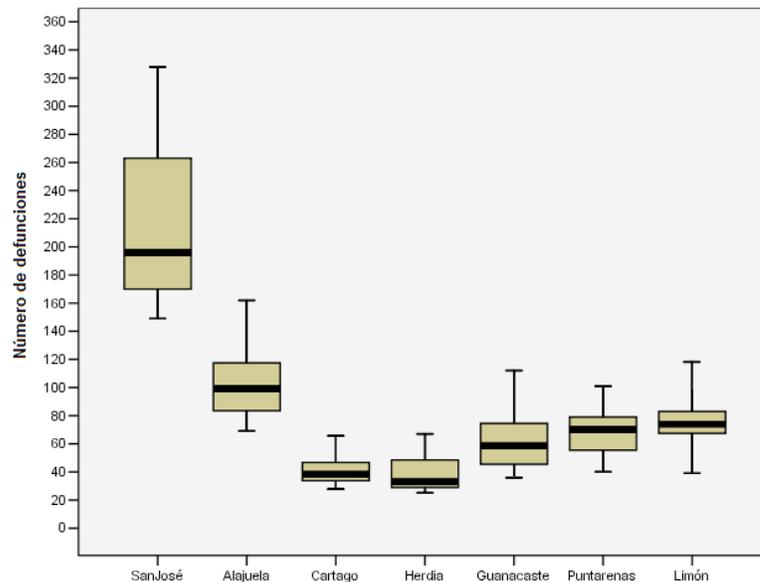
Nota: Los diagramas de cajas pueden expresarse en forma vertical tal como se muestra a continuación:



Problema 9. Defunciones por accidentes de tránsito según provincia

Considere el siguiente diagrama de cajas.

Costa Rica: Número anual de defunciones por accidentes de tránsito
Período 1991-2010



Fuente: <http://www.estadonacion.or.cr/index.php/estadisticas/costa-rica/compendio-estadistico/estadisticas-sociales>

De la información proporcionada por el diagrama, ¿cuál de las siguientes proposiciones son verdaderas:

- I. Limón y Cartago son las provincias en las que se presentó la menor cantidad de muertes por años debidas a accidentes de tránsito en este período.
- II. Para San José, en más de la mitad de los años del período 1991-2010 se presentó más de 220 defunciones anuales por accidentes de tránsito.
- III. Heredia fue la provincia en que se presentó la menor cantidad de defunciones por accidentes de tránsito por año en este período.

Solución

Al observar las cajas que representan al número anual de defunciones por accidentes de tránsito en nuestro país en el período 1991-2010, se identifica que es en Heredia donde se presentó la menor cantidad de defunciones por accidentes de tránsito en dicho período: tiene menor mediana que todas las demás provincias y el tercer cuartil y valor máximo son similares a Cartago que sería la provincia que le sigue en este análisis de valores absolutos.

En este sentido se tiene que la proposición I sería falsa, debido a que las provincias con menor cantidad de defunciones son Heredia y Cartago. En San José la mediana es menor de 200 defunciones anuales, por ello no es cierto que “*en más de la mitad de los años del período 1991-2010, presentó más de 200 defunciones anuales por accidentes de tránsito*”, debido a

que el porcentaje de años en los que hubo más de 220 defunciones es mucho menor que la mitad.

Por lo tanto la proposición III es la única correcta.

[Volver al índice de conceptos](#)

5. Media aritmética ponderada o promedio ponderado

Hasta el momento los datos con que hemos trabajado no presentan diferencias en cuanto al valor o peso relativo de cada uno. En estos casos, se dice que los datos son simples porque todos tienen el mismo valor o peso dentro del análisis. No obstante, esto no siempre es así: podemos enfrentar situaciones donde los datos con que trabajamos tienen diferente valor o peso. Observe los siguientes problemas:

Promedio ponderado

Cuando se requiere determinar el promedio de datos que tienen pesos diferentes, es decir no todos ellos tienen el mismo valor para efectos del problema, entonces es necesario ponderar según el peso relativo (ponderación) de cada dato. En estos casos se aplica la fórmula:

$$\text{Promedio ponderado: } \frac{\text{Suma de (datos} \cdot \text{ponderación)}}{\text{Suma de (ponderaciones)}}$$

Problema 9. Ganancia por venta de zapatos

En una zapatería se promedian ventas diarias de 8 pares de zapatos de hombre y 12 de mujer. La ganancia media en los zapatos de hombre es de ₡2000, mientras que en los zapatos de mujer la ganancia media es de ₡2500. Con esta información ayude al gerente de la zapatería a determinar la ganancia promedio por par de zapatos.

Solución

De acuerdo con la fórmula del promedio ponderado, la ganancia promedio por par de zapatos debe calcularse de la siguiente manera

$$\frac{8 \cdot \text{₡}2000 + 12 \cdot \text{₡}2500}{20} = \text{₡}2300.$$

Entonces este sería el monto solicitado.

Problema 10. Nota promedio de un curso

Suponga que usted y una amiga matriculan un curso que tiene el siguiente sistema de evaluación: tres exámenes cortos (o quices), dos exámenes parciales y un examen final. Los exámenes cortos tienen un valor de 5% cada uno, los exámenes parciales valen 20% cada uno y el examen final tiene un valor de 45%. Las calificaciones obtenidas por usted y su amiga son:

Evaluación	Calificación personal	Calificación de amiga
Primer examen corto	75	89
Segundo examen corto	84	93
Tercer examen corto	88	98
Primer examen parcial	71	87
Segundo examen parcial	73	87
Examen final	91	74
Suma total	482	528

Considerando estas calificaciones ¿cuál fue la nota promedio de cada uno?

Solución:

Si se consideran las calificaciones en forma simple, su amiga obtuvo notas más altas con excepción del último examen. Si calculamos el promedio simple de las calificaciones se obtiene: $\frac{482}{6} = 80,3$ para usted y la correspondiente a su amiga $\frac{528}{6} = 88,0$.

Sin embargo, determinar un promedio simple constituye un error para efectos de cálculo, porque el peso o valor relativo es diferente para los distintos tipos de evaluación. En estos casos cada dato debe ponderarse de acuerdo con el peso relativo (ponderación) de cada evaluación, de acuerdo con la fórmula:

$$\text{Promedio ponderado: } \frac{\text{Suma de (datos} \cdot \text{ponderación)}}{\text{Suma de (ponderaciones)}}$$

Debido a que las ponderaciones corresponden al valor porcentual de cada evaluación, entonces la calificación promedio puede determinarse tal como se muestra en el siguiente cuadro:

Evaluación	Calificación personal	Calificación de amiga	Ponderación	Nota · ponderación (personal)	Nota · ponderación (amiga)
Primer examen corto	75	89	5	$75 \cdot 5 = 375$	$89 \cdot 5 = 445$
Segundo examen corto	84	93	5	$84 \cdot 5 = 420$	$93 \cdot 5 = 465$
Tercer examen corto	88	98	5	$88 \cdot 5 = 440$	$98 \cdot 5 = 490$
Primer examen parcial	71	87	20	$71 \cdot 20 = 1420$	$87 \cdot 20 = 1740$
Segundo examen parcial	73	87	20	$73 \cdot 20 = 1460$	$87 \cdot 20 = 1740$
Examen final	91	74	45	$91 \cdot 45 = 4095$	$74 \cdot 45 = 3330$
Suma total			100	8210	8210

En las últimas dos columnas se incluyó el producto de cada calificación por la ponderación y al final de la columna aparece la suma, entonces su calificación promedio es $\frac{8210}{100} = 82,1$, que es igual a la de su amiga. Entonces se concluye que la nota final del curso es 82,1 para ambos.

Este promedio ponderado se utiliza también cuando los datos han sido agrupados en una distribución de frecuencias. Observe el siguiente problema:

Problema 11. Insecticida contra cucarachas

En un laboratorio químico se está realizando un experimento con un nuevo insecticida contra cucarachas.



www.freepik.es

Desean determinar el tiempo promedio de vida después de que el insecto ha estado en contacto con dicho insecticida. El experimento consiste en fumigar un sector del piso y exponer varias cucarachas que cruzan por el sector que fue rociado. Luego se mide el tiempo hasta que la cucaracha muere. El asistente de laboratorio anota el número de cucarachas que han muerto cada cinco minutos y forma la siguiente distribución:

Tiempo de sobrevivencia de una muestra de 110 cucarachas que fueron expuestas a un nuevo insecticida

Tiempo de sobrevivencia	Número de cucarachas
De 0 a menos de 5 minutos	7
De 5 a menos de 10 minutos	32
De 10 a menos de 15 minutos	46
De 15 a menos de 20 minutos	18
De 20 a menos de 25 minutos	5
De 25 a menos de 30 minutos	2
Total	110

Información ficticia utilizada con fines didácticos

El asistente de laboratorio debe entregar un informe en donde además del cuadro anterior debe incluir el tiempo promedio de sobrevivencia; sin embargo, no sabe cómo determinar este valor. Ayude al asistente de laboratorio a determinar este promedio de modo que pueda concluir el informe correctamente.

Solución:

El asistente del laboratorio enfrenta un problema típico de los análisis estadísticos. Aunque se expusieron 110 cucarachas al insecticida, no se tienen los 110 datos correspondientes al tiempo de sobrevivencia de cada una de ellas, entonces no es posible determinar el tiempo de sobrevivencia promedio utilizando la fórmula de cálculo simple. En estos casos únicamente se sabe que siete cucarachas murieron en menos de cinco minutos, pero no se sabe cuánto tiempo duró cada una, igualmente ocurre con las otras clases.

Para resolver el problema de la ausencia de los datos simples, se busca un valor que represente a cada una de las clases o agrupamientos, el punto medio de la clase (corresponde al promedio entre el límite inferior y el límite superior de esta) constituye el mejor representante. Esto quiere decir que el valor 2,5 segundos sería el representante de la primera clase, 7,5 segundos el representante de la segunda clase y así sucesivamente. Con estos valores entonces se puede determinar un promedio ponderado de la siguiente forma:

Tiempo de sobrevivencia	Número de Cucarachas (f)	Punto medio (x) (representante de clase)	$x \cdot f$
De 0 a menos de 5 minutos	7	2,5	17,5
De 5 a menos de 10 minutos	32	7,5	240
De 10 a menos de 15 minutos	46	12,5	575
De 15 a menos de 20 minutos	18	17,5	315
De 20 a menos de 25 minutos	5	22,5	112,5
De 25 a menos de 30 minutos	2	27,5	55
Total	110		1315

Si se aplica la fórmula del promedio ponderado se tiene que el tiempo promedio de sobrevivencia de las cucarachas una vez que fueron expuestas al insecticida fue de $\frac{1315 \text{ segundos}}{110} = 11,95$ segundos aproximadamente.

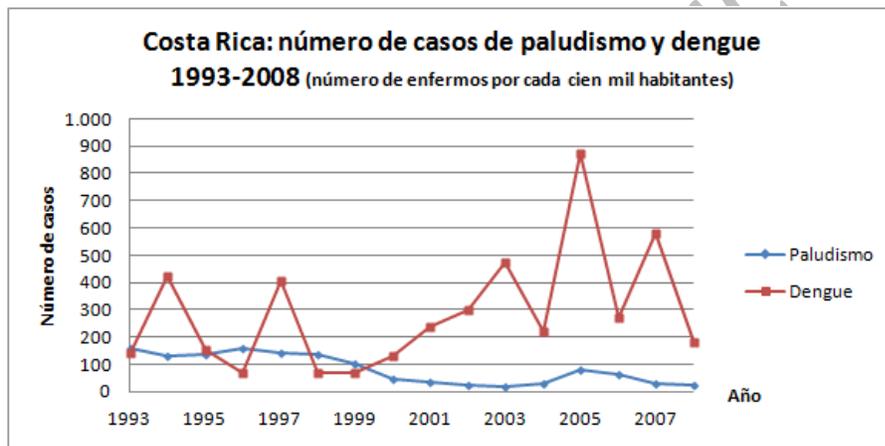
[Volver al índice de conceptos](#)

6. Uso de medidas estadísticas de variabilidad

Los análisis anteriores se han concentrado en la posición de los datos, pero debido a que el propósito básico de la Estadística consiste en analizar su patrón de variabilidad, es común encontrar situaciones para las cuales se hace necesario medir esta variabilidad. Al respecto observe el siguiente problema.

Problema 12. Enfermos por picadura de mosquito

Suponga que se analizan dos enfermedades que se vinculan con la picadura de mosquitos: el paludismo y el dengue. La siguiente gráfica muestra el número de casos de estas enfermedades en el país entre 1993 y 2008.



Fuente: <http://www.estadonacion.or.cr/index.php/estadisticas/costa-rica/compendio-estadistico/estadisticas-sociales>

De acuerdo con esta información, ¿en cuál de las variables se presenta mayor variabilidad: en el número anual de enfermos por paludismo, o en el número anual de enfermos por dengue?

Solución:

A simple vista puede observarse que el número de enfermos por paludismo fue menos variable en este período e incluso presentó una importante disminución. Por su parte, el número de enfermos por dengue fue muy inestable con grandes variaciones entre años y con una tendencia al aumento. En general se puede concluir que el número anual de enfermos por dengue fue mucho más variable durante el periodo 1993-2008.

Problema 13. Contrato de trabajo con empresa turística

Don Juan está actualmente sin empleo, pero acaba de recibir dos ofertas que le parecen atractivas y que generan un salario competitivo dentro de su especialidad laboral que es el turismo. Sin embargo, el salario varía en los meses dependiendo de que la temporada turística sea alta, media o baja. Por esta razón, debido a que don Juan tiene gastos mensuales fijos no le conviene tener muchas diferencias en su ingreso entre un mes y otro.



El siguiente cuadro resume los salarios mensuales (en colones) que le ofrecen las empresas:

Salario mensual en colones ofrecido a don Juan por las empresas A y B, según el mes del año

Mes	Temporada	Empresa A	Empresa B
Enero	Alta	1 300 000	1 050 000
Febrero	Alta	1 200 000	1 050 000
Marzo	Media	1 000 000	850 000
Abril	Media	800 000	850 000
Mayo	Baja	600 000	650 000
Junio	Baja	600 000	650 000
Julio	Alta	1 100 000	1 050 000
Agosto	Media	700 000	850 000
Setiembre	Baja	500 000	650 000
Octubre	Baja	500 000	650 000
Noviembre	Media	800 000	850 000
Diciembre	Alta	1 100 000	1 050 000

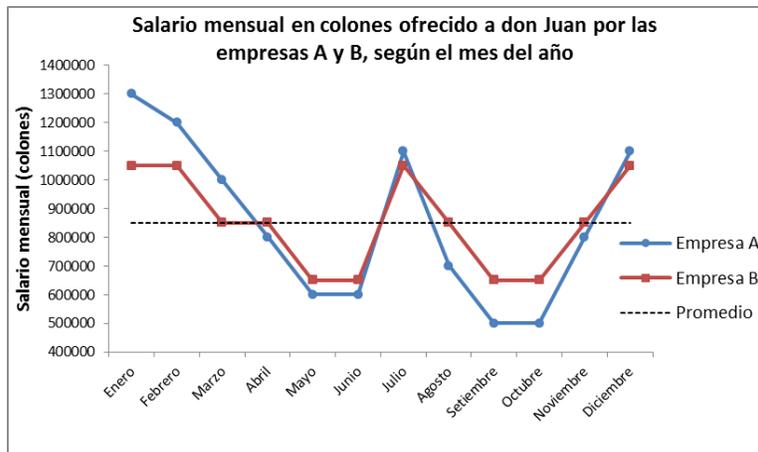
Información ficticia utilizada con fines didácticos

Ayude a don Juan a tomar una decisión sobre la empresa que debería seleccionar, según sus necesidades e intereses.

Solución:

Lo primero que debería realizar don Juan consiste en determinar el ingreso promedio mensual que se le está ofreciendo en cada empresa. Para ello debe sumar los salarios de los 12 meses y dividirlo por 12. En ambos casos el total es de 10 200 000 colones, y entonces su salario promedio mensual será de $\frac{10\,200\,000}{12}$ colones = 850 000 colones.

Una vez determinado el salario promedio, se debe analizar la variabilidad, ya que don Juan requiere que no haya muchas variaciones en su salario mensual a lo largo del año, debido a los gastos fijos que debe cubrir mensualmente. En este sentido, basta con observar los datos para darse cuenta que en la empresa B los salarios son más estables a lo largo del año, presentan menos diferencias entre las temporadas y entre los meses, por ello la recomendación sería que acepte el empleo de la empresa B. Para facilitar el análisis la información del cuadro puede ser representada mediante un gráfico de línea como el siguiente:



Puede notarse que la línea correspondiente a los salarios ofrecidos por la empresa A presenta mayor variabilidad en relación con el salario promedio, debido a que en términos generales se aleja más del promedio que la línea correspondiente a los salarios ofrecidos por la empresa B. Por esta razón es que le conviene seleccionar la empresa B.

En estos dos problemas resultó relativamente fácil determinar cuál de los grupos de datos tenía menos variabilidad; sin embargo, esto no siempre ocurre debido a que muchas veces las diferencias de variabilidad no se logran visualizar a partir de los datos y se debe recurrir a análisis más elaborados para resolver estos problemas. Para simplificar este trabajo se han diseñado indicadores para medir la variabilidad de los datos. Las medidas estadísticas de variabilidad más utilizadas son:

- Recorrido
- Recorrido intercuartílico
- Variancia y desviación estándar

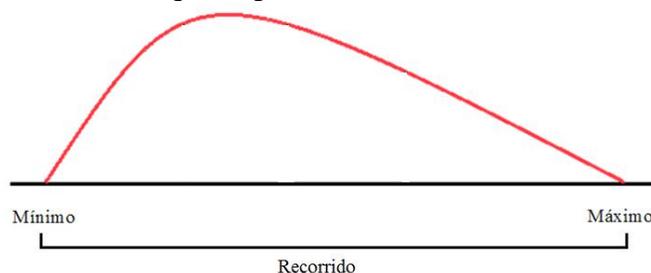
Al igual que con las medidas de posición, las medidas de variabilidad tienen una connotación muy particular y conceptualmente miden diferentes relaciones entre los datos.

[Volver al índice de conceptos](#)

El recorrido

Se define por: $Recorrido = \text{máximo} - \text{mínimo}$

Mide la mayor diferencia numérica que se presenta entre los datos.



Problema 14. Constancia en el lanzamiento del martillo

María y Raquel son dos lanzadoras de martillo que se preparan para participar en diferentes eventos internacionales. En las competencias sobre lanzamiento de martillo, el atleta se ubica en un círculo de 2,15 m de diámetro, el cual está instalado dentro de una jaula de seguridad, protegida con redes y tiene solamente espacio sobre el cual debe salir el martillo. En el proceso de clasificación se realizan solamente tres lanzamientos y se toma como válido el mayor. La técnica debe practicarse mucho porque un alto porcentaje de los lanzamientos son nulos debido a que el atleta se sale del círculo al hacer el lanzamiento o no logra enviar el martillo a la zona de competencia.



<https://commons.wikimedia.org/wiki/File:EVD-martillo-011.jpg>

Producto de lo anterior, además de que es muy importante que los atletas lancen a gran distancia el martillo, también deben ser muy constantes (haya poca variabilidad en los lanzamientos) debido a que, muchas veces solamente logran obtener un lanzamiento válido (de los tres posibles).

De 10 lanzamientos de entrenamiento se generaron los siguientes resultados (en metros):

María	69,38	66,72	70,46	65,52	66,18	69,39	71,75	76,04	70,12	52,32
Raquel	76,21	70,76	64,61	60,88	74,63	72,45	65,09	58,24	56,34	72,62

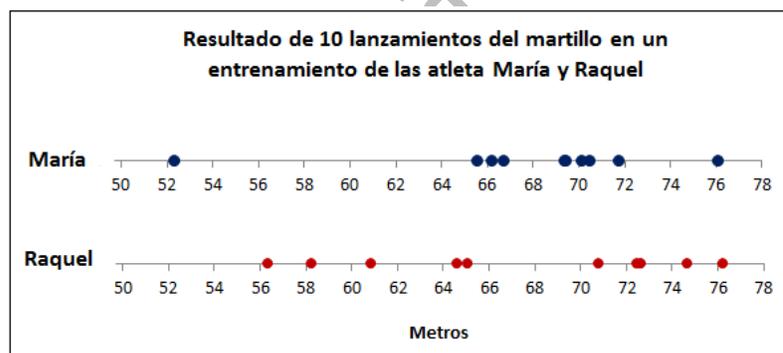
De acuerdo con esta información, analice si el recorrido será una buena medida para determinar cuál de las atletas tuvo lanzamientos menos variables.

Solución:

Para determinar cuál de las atletas es más consistente en los lanzamientos de entrenamiento (son menos variables), al utilizar el recorrido se puede caer en un grave error. Observe que para este caso se tienen los siguientes resultados:

Atleta	Recorrido en los lanzamientos
María	$76,04 - 52,32 = 23,72$
Raquel	$76,21 - 56,34 = 19,87$

El valor del recorrido pareciera indicar que los lanzamientos de María son más variables, pero observando los datos se puede notar que, con excepción del lanzamiento de 52,32 metros, el resto de lanzamientos de María son menos variables entre sí que los lanzamientos de Raquel. Esto también se puede observar mediante una representación gráfica aproximada de los datos.



Hay que recordar que el recorrido si interpreta como la mayor diferencia entre lanzamientos: para María fue de 23,72 metros y para Raquel fue de 19,87 metros.

Nota: El ejemplo anterior muestra que se debe tener cuidado al comparar la variabilidad entre grupos de datos por medio del recorrido, debido a que no necesariamente cuando un grupo de datos tiene un mayor recorrido significa que sea el más variable. El recorrido se ve muy afectada por la presencia de uno o más valores que sea extremadamente grande o extremadamente pequeño (o ambos).

El error que se podría cometer consiste en creer que el grupo de datos con mayor recorrido es el más variable, esto no siempre ocurre, lo que el recorrido nos indica es la mayor diferencia entre dos datos del conjunto.

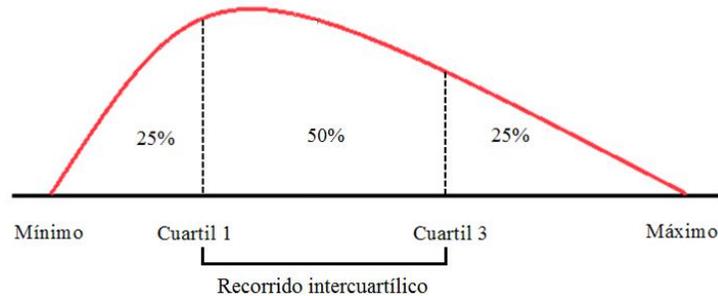
[Volver al índice de conceptos](#)

El recorrido intercuartílico

Permite corregir el problema de los valores extremos que presenta el cálculo del recorrido, porque para su cálculo se considera únicamente el 50% de los datos centrales o intermedios. Por esta razón, su fórmula de cálculo es:

$$\text{Recorrido intercuartílico} = \text{Tercer cuartil} - \text{Primer cuartil}$$

Observe que con esta medida se indica la longitud del ámbito o rango en el cual varía el 50% de los valores centrales de la distribución de datos.



Una forma visual muy práctica para analizar la variabilidad por medio de la articulación del recorrido y el recorrido intercuartílico son los diagramas de caja. Aunque un diagrama de caja incluye medidas estadísticas de orden (mínimo, cuartiles y máximo), permite valorar visualmente la variabilidad de los datos por medio de estas figuras.

Problema 15. Rendimiento en combustible de los vehículos

Para ejemplificar la interpretación del recorrido intercuartílico, volvamos a analizar el problema 3, en donde Pilar y Beatriz discutían sobre el rendimiento de combustible de sus vehículos (en kilómetros recorridos por litro). Para los 10 días observados, los datos de los rendimientos de estos vehículos ordenados de menor a mayor son:

Rendimiento en kilómetros por litro

Auto de Pilar		Auto de Beatriz	
1	11,6	1	9,5
2	11,7	2	9,6
3	12,1	3	10,9
4	12,5	4	12,1
5	13,1	5	12,2
6	13,2	6	12,9
7	13,3	7	13,1
8	13,5	8	13,7
9	14,4	9	14,0
10	15,0	10	14,2

Analice la variabilidad en el rendimiento de estos vehículos desde según la información que proporciona el recorrido intercuartílico.

Solución:

Para determinar el recorrido intercuartílico de ambos rendimientos, se puede recurrir a las medidas de orden que fueron calculadas en el problema 12, los valores se resumen en el siguiente cuadro:

Rendimiento en kilómetros por litro de los autos de Pilar y Beatriz, información recabada por día, en 10 días consecutivos

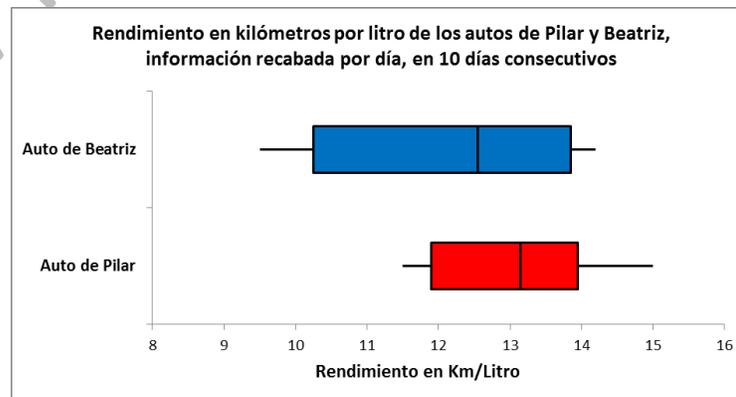
Medida estadística de orden	Rendimiento de auto de Pilar Km/L	Rendimiento de auto de Beatriz Km/L
Primer cuartil	11,9	10,25
Mediana	13,15	12,55
Tercer cuartil	13,95	13,85
Máximo	15,0	14,2
Mínimo	11,6	9,5

De acuerdo con esta información el recorrido y el recorrido intercuartílico vienen dados por:

Medida	Rendimiento de auto de Pilar Km/L	Rendimiento de auto de Beatriz Km/L
Recorrido	3,4	4,7
Recorrido intercuartílico	2,05	3,6

Para interpretar estas medidas podemos decir que la mayor diferencia en el rendimiento del auto en dos días muestreados fue de 3,4 Km/L para el auto de Pilar y de 4,7 Km/L para el auto de Beatriz. Por otro lado, el 50% de los valores intermedios del rendimiento variaron en un rango de 2,05 Km/L para el auto de Pilar y de 3,6 para el auto de Beatriz. Ambos valores son mayores en el auto de Beatriz.

Estos valores pueden ser analizados en forma visual por medio en un diagrama de cajas:



En general puede notarse que en el período observado el auto de Beatriz refleja una mayor variabilidad en el rendimiento, solamente en el último 25% de los datos el rango fue mayor en las mediciones del auto de Pilar; sin embargo, hay que tener claro que este análisis de variabilidad es parcial porque se basa solamente en las medidas de orden.

Aunque el recorrido y el recorrido intercuartílico se complementan para ilustrar el patrón de variabilidad de un conjunto de datos, sus fórmulas de cálculo no incluyen todos los datos, por lo que podrían no ser suficientemente ilustrativas para todos los problemas. Por esta razón se han definido medidas de variabilidad más completas.

[Volver al índice de conceptos](#)

Variación poblacional:

En términos sencillos, cuando se trabaja con todos los datos de una población (n datos), para el cálculo de la variancia se determina el cuadrado de la diferencia de cada dato con respecto al promedio $(dato - promedio)^2$ y se aplica la siguiente fórmula:

$$Variación = \frac{\text{Suma } (dato - promedio)^2}{n}$$

Variación muestral:

Cuando se trabaja con los elementos de una muestra (n datos) en vez de toda la población, la fórmula de la variancia viene dada por:

$$Variación = \frac{\text{Suma } (dato - promedio)^2}{n - 1}$$

[Volver al índice de conceptos](#)

Desviación estándar:

La variancia se calcula con la intención de generar un indicador general de la variabilidad de todos los datos (se elimina el efecto de valores negativos); sin embargo, las unidades de medida que corresponden al contexto del problema quedan al cuadrado. Por esta razón si se extrae la raíz cuadrada de la variancia se genera otra medida de variabilidad que se llama desviación estándar, la cual tiene las mismas unidades de medida que los datos originales.

$$\text{Desviación estándar poblacional} = \sqrt{\frac{\text{Suma } (dato - promedio)^2}{n}}$$

$$\text{Desviación estándar muestral} = \sqrt{\frac{\text{Suma } (dato - promedio)^2}{n - 1}}$$

Nota: El cálculo de la variancia tiene una particularidad que no la tiene ninguna de las otras medidas utilizadas previamente, la cual consiste en que su fórmula de cálculo varía si los datos corresponden a toda la población o a una muestra. La diferencia entre la variancia de una muestra o de una población consiste en que para la población la suma de las desviaciones al cuadrado se divide entre el número de datos y en una muestra se divide entre el número de datos menos uno. Las razones por las que se hace esta distinción obedecen a principios de teoría estadística que no conviene discutir en el presente documento (dado que está vinculado con propiedades de los estimadores muestrales para estimar los valores poblacionales desconocidos).

Para hacer patente esta diferencia los libros de texto utilizan diferentes símbolos que incluye también la media aritmética o promedio (aunque se utiliza la misma fórmula). Si se tiene un grupo de datos X_1, X_2, \dots, X_n que corresponden a variable cuantitativa, entonces se utiliza los símbolos:

- μ representa la media aritmética de una población
- \bar{X} representa la media aritmética de una muestra
- σ^2 representa la variancia de una población
- S^2 representa la variancia de una muestra

Para la variancia se cumple que:

$$\sigma^2 = \frac{\text{Suma (dato - promedio)}^2}{n} = \frac{\text{Suma } (X_j - \mu)^2}{n}$$

$$S^2 = \frac{\text{Suma (dato - promedio)}^2}{n - 1} = \frac{\text{Suma } (X_j - \bar{X})^2}{n - 1}$$

Incluso se tienden a utilizar otras formas de simbolizar lo anterior; pero no es necesario que el estudiante profundice en estos temas, lo que tiene que tener presente corresponde a diferenciar si los datos pertenecen a toda la población o una muestra, lo cual debe extraerse del contexto donde se plantea el problema, es decir, la redacción del problema planteado debe dar las pautas para que el estudiante pueda establecer si los datos pertenecen a una muestra o a una población.

[Volver al índice de conceptos](#)

Problema 16. Enfermos por picadura de mosquito

Volvamos al problema 12, en donde se analizó el número de enfermos por paludismo y dengue que se presentaron entre 1993 y 2008.

Año	1993	1994	1995	1996	1997	1998	1999	2000
Paludismo	157	134	134	159	141	137	104	49

Dengue	143	422	153	67	407	70	68	129
---------------	-----	-----	-----	----	-----	----	----	-----

Año	2001	2002	2003	2004	2005	2006	2007	2008
Paludismo	34	25	17	30	82	66	27	22
Dengue	235	299	473	221	874	274	583	184

Fuente: <http://www.estadonacion.or.cr/index.php/estadisticas/costa-rica/compendio-estadistico/estadisticas-sociales>

Por medio de la desviación estándar en el número de defunciones, determine ¿en cuál de las variables se presenta mayor variabilidad: en el número anual de enfermos por paludismo o en el número anual de enfermos por dengue?

Solución:

Para determinar la desviación estándar de los datos en cada caso, primeramente hay determinar el número promedio de enfermos por año. Según el cuadro que se muestra abajo, en este período de 16 años se presentaron 1318 enfermos por paludismo y 4602 enfermos por dengue, entonces el promedio de enfermos para el período es:

Promedio anual de enfermos por paludismo

$$\frac{1318}{16} = 82,375$$

Promedio anual de enfermos por dengue

$$\frac{4602}{16} = 287,625$$

De acuerdo con la fórmula de la variancia, para cada año debe calcularse la diferencia del número de defunciones con respecto al promedio y luego elevar al cuadrado. Por ejemplo:

Para paludismo:

Para 1993 se tiene $157 - 82,375 = 74,625$, entonces $(74,625)^2 = 5568,89063$

Para 1994 se tiene $134 - 82,375 = 51,625$, entonces $(51,625)^2 = 2665,14063$

Para 1995 se tiene $134 - 82,375 = 51,625$, entonces $(51,625)^2 = 2665,14063$

Del mismo modo, se calculan los valores de los otros años:

Para dengue:

Para 1993 se tiene $143 - 287,625 = -144,625$, entonces $(-144,625)^2 = 20916,3906$

Para 1994 se tiene $422 - 287,625 = 134,375$, entonces $(134,375)^2 = 18056,6406$

Para 1995 se tiene $153 - 287,625 = -134,625$, entonces $(-134,625)^2 = 18123,8906$

En forma idéntica se calculan los otros valores. En el cuadro de abajo aparecen todos los cálculos. Para determinar la variancia la suma de todos los cuadrados debe dividirse por el número de años menos uno, bajo el supuesto de que la información corresponde a un período de 16 años, la cual representa una muestra de la situación de estas enfermedades.

Para el caso del paludismo, como la suma de las diferencias al cuadrado es de 44401,75, entonces la variancia viene dada por:

$$\text{Variancia para paludismo} = \frac{44401,75}{15} = 2960,11667$$

Para el caso del dengue, la suma de las diferencias al cuadrado es de 724487,75, entonces la variancia viene dada por:

$$\text{Variancia para dengue} = \frac{724487,75}{15} = 48299,1833$$

Hay que recordar que la desviación estándar es la raíz cuadrada de la variancia, entonces se tendría que la desviación estándar en el número anual de enfermos por paludismo para el período 1993-2008 viene dada por $\sqrt{2960,11667}$ que equivale aproximadamente a 54,4 enfermos. Mientras que para este mismo período la desviación estándar en el número anual de enfermos por dengue fue de $\sqrt{48299,1833}$ equivale aproximadamente a 219,8 enfermos. Es claro entonces que para el período 1993-2008, el número anual de enfermos por paludismo fue menos variable que el número anual de enfermos por dengue. El siguiente cuadro muestra todos los cálculos:

Año	Número de enfermos		Diferencia respecto al promedio		Diferencias al cuadrado	
	Paludismo	Dengue	Paludismo	Dengue	Paludismo	Dengue
1993	157	143	74,625	-144,625	5568,89063	20916,3906
1994	134	422	51,625	134,375	2665,14063	18056,6406
1995	134	153	51,625	-134,625	2665,14063	18123,8906
1996	159	67	76,625	-220,625	5871,39063	48675,3906
1997	141	407	58,625	119,375	3436,89063	14250,3906
1998	137	70	54,625	-217,625	2983,89063	47360,6406
1999	104	68	21,625	-219,625	467,640625	48235,1406
2000	49	129	-33,375	-158,625	1113,89063	25161,8906
2001	34	235	-48,375	-52,625	2340,14063	2769,39063
2002	25	299	-57,375	11,375	3291,89063	129,390625
2003	17	473	-65,375	185,375	4273,89063	34363,8906
2004	30	221	-52,375	-66,625	2743,14063	4438,89063
2005	82	874	-0,375	586,375	0,140625	343835,641
2006	66	274	-16,375	-13,625	268,140625	185,640625
2007	27	583	-55,375	295,375	3066,39063	87246,3906
2008	22	184	-60,375	-103,625	3645,14063	10738,1406
Total	1318	4602			44401,75	724487,75
Promedio	82,375	287,625				
Variancia					2960,11667	48299,1833
Desviación estándar					54,4069542	219,770752

Nota: Como puede notarse, el cálculo de la variancia es tedioso, sobre todo en aquellos casos en que hay muchos datos. Realizar todos estos cálculos es de poco interés práctico desde un punto

de vista estadístico. Lo adecuado es utilizar una calculadora científica con funciones específicas que generan el valor de la desviación estándar, ya sea correspondiente a una muestra o a una población. En este sentido, el estudiante debe limitarse a incluir los datos a la calculadora y solicitar la medida requerida.



Se debe revisar el manual de usuario de la calculadora para determinar la forma en que se incluyen los datos y en que la máquina devuelve los resultados. En el caso de la desviación estándar, las calculadoras emplean diferentes símbolos para el caso de poblaciones o muestras, por ello debe identificar cada símbolo y su significado. Sobre este tema se profundiza en uno de los videos del curso.

Problema 17. Comparación de notas en Español

Las notas en Español durante el primer trimestre de los dos grupos de décimo año en cierto colegio, vienen dadas por:

Grupo 1	40	67	68	69	70	70	70	72	74	74	75	75	75	76	81	81	82	84	87	98
Grupo 2	48	51	53	68	70	70	71	72	75	75	75	78	79	79	82	89	91	92	93	

El profesor responsable de atender estos grupos desea conocer en cuál de ellos se generó el mejor rendimiento y en cuál las notas fueron más variables. Colabore con el profesor de Español para resolver el problema. Para ello analice y compare los valores del recorrido, recorrido intercuartílico y la desviación estándar.

Solución:

Debemos observar que las notas ya fueron ordenadas de menor a mayor, por lo que es posible determinar las medidas estadísticas de orden con más facilidad.

Posición	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
Grupo 1	40	67	68	69	70	70	70	72	74	74	75	75	75	76	81	81	82	84	87	98	1488
Grupo 2	48	51	53	68	70	70	71	72	75	75	75	78	79	79	82	89	91	92	93		1411

La media aritmética de las calificaciones o calificación promedio es

Grupo 1	$\frac{1488}{20} = 74,40$
Grupo 2	$\frac{1411}{19} = 74,26$

Como en el grupo 1 hay 20 estudiantes (número par) y en el grupo 2 hay 19 estudiantes (número impar), las fórmulas de cálculo de la mediana son diferentes:

- Para el grupo 1 la mediana se calcula por: $\frac{X_{\frac{20}{2}} + X_{\frac{20}{2} + 1}}{2} = \frac{X_{10} + X_{11}}{2}$ (el promedio de los datos 10 y 11).

- Para el grupo 2 sería el dato $X_{\frac{19+1}{2}} = X_{10}$ (el dato 10).

Grupo 1	$\frac{74 + 75}{2} = 74,5$
Grupo 2	75

Esto quiere decir que también la mediana de las calificaciones es similar en ambos grupos. Para el grupo 1 la mediana significa que el 50% de los estudiantes tuvo una nota de 74,5 o menos y el otro 50% tuvo una nota de 74,5 o más. Del mismo modo se interpreta la mediana del grupo 2. Además, el hecho de que la mediana y el promedio tengan pocas diferencias, es un indicador de que la distribución de las notas es bastante simétrica en ambos casos.

En cuanto a la moda, el grupo 1 tiene dos modas 70 y 75 (con frecuencia de tres estudiantes cada una), mientras que en el grupo 2 es 75 (con frecuencia de tres estudiantes).

Nota: en los casos en que hay pocos datos, como ocurre en este problema, la moda no es una buena medida de tendencia central, porque tiene un comportamiento irregular tal como se indicó antes.

De acuerdo con la fórmula de cálculo de los cuartiles, para determinar el valor del primer cuartil, con $n = 20$, su posición es $\frac{k \cdot (n+1)}{4} = \frac{1 \cdot 21}{4} = 5,25$, esto quiere decir que en el grupo 1, primer cuartil se encuentra entre los datos cinco y seis y se cumple que:

$$C_1 = \frac{X_5 + X_6}{2} = \frac{70 + 70}{2} = 70.$$

Esto significa que el 25% de los estudiantes del grupo tuvo una nota de 70 o menos, y el 75% tuvo una nota de 70 o más.

Para el grupo 2, $n = 19$, su posición es $\frac{k \cdot (n+1)}{4} = \frac{1 \cdot 20}{4} = 5$, entonces el primer cuartil corresponde al dato cinco:

$$C_1 = X_5 = 70.$$

Se interpreta de la misma forma que en el grupo 1.

En cuanto al tercer cuartil, para el grupo 1 la posición es $\frac{k \cdot (n+1)}{4} = \frac{3 \cdot 21}{4} = 15,75$, entonces el tercer cuartil se encuentra entre los datos 15 y 16. Entonces:

$$C_3 = \frac{X_{15} + X_{16}}{2} = \frac{81 + 81}{2} = 81.$$

Entonces, significa que el 75% de los estudiantes tuvo una nota de 81 o menos y el 25% tuvo una nota de 81 o más.

Para el grupo 2, la posición es $\frac{k \cdot (n+1)}{4} = \frac{3 \cdot 20}{4} = 15$, entonces el tercer cuartil corresponde al dato 15, o sea:

$$C_3 = X_{15} = 82.$$

Se interpreta en forma similar al correspondiente del grupo 1.

Por último la nota más baja del grupo 1 fue de un 40 y en el grupo 2 fue un 48. Por su parte, la nota más alta del grupo 1 fue un 98 y del grupo 2 fue un 93.

El siguiente cuadro presenta un resumen de las principales medidas de posición:

Calificaciones del primer trimestre en la materia de Español de los grupos 1 y 2

Medida	Grupo 1	Grupo 2
Promedio	74,40	74,26
Mediana	74,50	75,00
Moda	70,00 y 75,00	75,00
Primer cuartil	70,00	70,00
Tercer cuartil	81,00	82,00
Mínimo	40,00	48,00
Máximo	98,00	93,00

En resumen, las medidas estadísticas de posición muestran que el rendimiento es similar en ambos grupos en cuanto a las calificaciones del primer trimestre.

Para analizar la variabilidad se tiene que el recorrido en cada grupo fue:

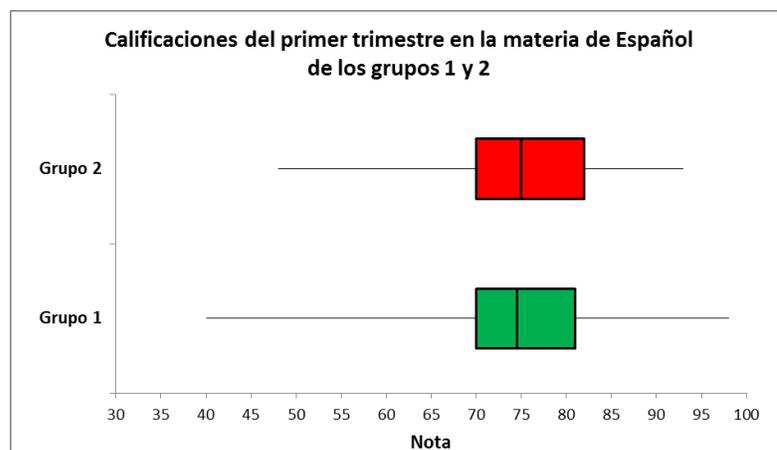
Grupo 1	$98 - 40 = 58$
Grupo 2	$93 - 48 = 45$

Quiere decir que en el grupo 1 la mayor diferencia en las calificaciones fue de 58 puntos, mientras que en el grupo 2 fue de 45 puntos. Por su parte, el recorrido intercuartílico viene dado por:

Grupo 1	$81 - 70 = 11$
Grupo 2	$82 - 70 = 12$

El 50% de las calificaciones intermedias varían en un rango entre 11 puntos para el grupo 1 y 12 puntos para el grupo 2.

Con estos valores y el de la mediana se puede construir un diagrama de cajas.



Nota: De acuerdo con lo anterior, se podría creer que la variabilidad en las calificaciones es similar en ambos grupos. Sin embargo, hay que tener presente que el recorrido intercuartílico es un indicador de variabilidad que nos muestra el rango de variación del 50% de los valores centrales. Entonces este es el mensaje que debe prevalecer al interpretar el recorrido intercuartílico.

Seguidamente se presenta el cálculo de la suma de desviaciones al cuadrado para el grupo 1

<i>Dato</i>	<i>Dato - promedio</i>	<i>(Dato - promedio)²</i>
40	40 - 74,4 = -34,4	(-34,4) ² = 1183,36
67	67 - 74,4 = -7,4	(-7,4) ² = 54,76
68	-6,4	40,96
69	-5,4	29,16
70	-4,4	19,36
70	-4,4	19,36
70	-4,4	19,36
72	-2,4	5,76
74	-0,4	0,16
74	-0,4	0,16
75	0,6	0,36
75	0,6	0,36
75	0,6	0,36
76	1,6	2,56
81	6,6	43,56
81	6,6	43,56
82	7,6	57,76
84	9,6	92,16
87	12,6	158,76
98	23,6	556,96
Total	0	2328,8

Entonces la variancia del grupo 1 sería: $\text{Variancia} = \frac{2328,8}{20} = 116,44$

Para el grupo 1 la desviación estándar de las calificaciones viene dada por $\sqrt{116,44} \approx 10,79$. Entonces la desviación estándar de las calificaciones del grupo 1 es 10,79 puntos. Se puede repetir el procedimiento para el grupo 2:

<i>Dato</i>	<i>Dato – promedio</i>	<i>(Dato – promedio)²</i>
48	48 – 74,26 = -26,26	(-26,26) ² = 689,59
51	51 – 74,26 = -23,26	(-23,26) ² = 541,03
53	-21,26	451,99
68	-6,26	39,19
70	-4,26	18,15
70	-4,26	18,15
71	-3,26	10,63
72	-2,26	5,11
75	0,74	0,55
75	0,74	0,55
75	0,74	0,55
78	3,74	13,99
79	4,74	22,47
79	4,74	22,47
82	7,74	59,91
89	14,74	217,27
91	16,74	280,23
92	17,74	314,71
93	18,74	351,19
Total	0	3057,73

Entonces la variancia del grupo 2 sería:

$$\text{Variancia} = \frac{3057,73}{19} = 160,93$$

Y la desviación estándar es $\sqrt{160,93} \approx 12,69$ puntos.

Puede notarse que al utilizar una medida de variabilidad mucho más precisa, las calificaciones del grupo 2 se muestran más variables que las del grupo 1, la desviación estándar es prácticamente dos unidades mayor.

Nota: si utilizamos la calculadora, no sería necesario realizar los cálculos repetitivos y tediosos al determinar la variancia.

Problema 18. Contrato de trabajo con empresa turística

Retomemos nuevamente el problema 13 donde don Juan debe escoger entre dos empresas de turismo que le ofrecen salario que varía de un mes a otro. Recordemos que, en igualdad de condiciones, debe escoger la empresa que le ofrezca los salarios menos variables.

Mes	Temporada	Empresa A	Empresa B
Enero	Alta	1 300 000	1 050 000
Febrero	Alta	1 200 000	1 050 000
Marzo	Media	1 000 000	850 000
Abril	Media	800 000	850 000
Mayo	Baja	600 000	650 000
Junio	Baja	600 000	650 000
Julio	Alta	1 100 000	1 050 000
Agosto	Media	700 000	850 000
Setiembre	Baja	500 000	650 000
Octubre	Baja	500 000	650 000
Noviembre	Media	800 000	850 000
Diciembre	Alta	1 100 000	1 050 000

Dado que el salario promedio que ofrecen ambas empresas es de 850 000 colones, entonces determine la empresa que le ofrece los salarios más variables con base el cálculo de la desviación estándar y ayude a don Juan a tomar la decisión.

Solución:

Si no se utiliza la calculadora para determinar directamente la desviación estándar, entonces se deben realizar los siguientes cálculos:

Mes	Empresa A	Empresa B	Diferencias A	Diferencias B	Diferencias al cuadrado A	Diferencias al cuadrado B
Enero	1300000	1050000	450000	200000	202500000000	40000000000
Febrero	1200000	1050000	350000	200000	122500000000	40000000000
Marzo	1000000	850000	150000	0	22500000000	0
Abril	800000	850000	-50000	0	2500000000	0
Mayo	600000	650000	-250000	-200000	62500000000	40000000000
Junio	600000	650000	-250000	-200000	62500000000	40000000000
Julio	1100000	1050000	250000	200000	62500000000	40000000000
Agosto	700000	850000	-150000	0	22500000000	0
Setiembre	500000	650000	-350000	-200000	122500000000	40000000000
Octubre	500000	650000	-350000	-200000	122500000000	40000000000
Noviembre	800000	850000	-50000	0	2500000000	0
Diciembre	1100000	1050000	250000	200000	62500000000	40000000000
Total	10200000	10200000	0	0	870000000000	320000000000

Debido a que los datos corresponden a todos los salarios del año, entonces la fórmula de cálculo empleada corresponde a la variancia de la población. Entonces la variancia y la desviación estándar para los salarios ofrecidos por cada empresa serían:

Empresa	Variancia	Desviación estándar
A	$\frac{870\,000\,000\,000}{12} = 72\,500\,000\,000$	269 258,24
B	$\frac{320\,000\,000\,000}{12} = 26\,666\,666\,666,67$	163 299,32

El valor de la desviación estándar deja en evidencia lo que se había observado visualmente en el problema 20: los salarios de la empresa B son menos variables y debería seleccionar dicha empresa.

[Volver al índice de conceptos](#)

II. Bibliografía

Ministerio de Educación Pública (2015). Curso bimodal para el II Ciclo: Estadística mucho más que procedimientos y técnicas. Unidad didáctica Estadística. San José, Costa Rica: autor.

Ministerio de Educación Pública (2012). Programas de Matemáticas para la Educación General Básica y el Ciclo Diversificado. San José, Costa Rica: autor.

III. Créditos

Material complementario Estadística y Probabilidad: Medidas estadísticas, es parte del MiniMOOC Medidas estadísticas, una actividad del Proyecto Reforma de la Educación Matemática en Costa Rica.

Este proyecto del Ministerio de Educación Pública es apoyado por Asociación Empresarial para el Desarrollo y por la Fundación Costa Rica - Estados Unidos de América para la Cooperación.

Autor del presente documento

Edwin Chaves Esquivel

Revisores de este documento

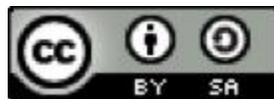
Ángel Ruiz, Edison De Faria, Johanna Mena, Keibel Ramírez, Luis Hernández, Xinia Zúñiga.

Director general del proyecto Reforma de la Educación Matemática en Costa Rica

Ángel Ruiz

Para referenciar este documento

Ministerio de Educación Pública, Proyecto Reforma de la Educación Matemática en Costa Rica (2017). *Material complementario Estadística y Probabilidad: Medidas estadísticas*, San José, Costa Rica: autor.



Material complementario Estadística y Probabilidad: Medidas Estadísticas por Ministerio de Educación Pública de Costa Rica, Proyecto Reforma de la Educación Matemática en Costa Rica, se encuentra bajo una Licencia Creative Commons Atribución-CompartirIgual 3.0 Unported.