

Protegiendonos de los peligros del promedio

🕒 Created	@Jul 11, 2020 10:44 AM
🏷️ Tags	

Podremos encontrar casos donde dos grupos de datos distintos tengan el mismo promedio, pero sus datos son muy diferentes uno del otro. No es lo mismo un grupo de datos donde su desviación es menor a 1, que aquel donde sus datos tienen una desviación de 4 o 6 puntos.

La formula del coeficiente de variación nos es útil al momento de evaluar estos casos:

$(\text{desviación estándar})/(\text{promedio}) * 100 = \text{coeficiente}$

Si el coeficiente es **mayor al 25% entonces los datos no son homogéneos**, varían mucho.

Dentro de R podemos sacar la desviación estándar con la función `sd` y el promedio con `mean`.

¿La Desviacion estandar es lo mismo que el coeficiente de variacion?

No. **El coeficiente**, el cual se expresa en % (por eso se multiplica por 100) nos dice finalmente que tan desviados están los datos frente al promedio. Por eso la fórmula es: Desviación (SD) sobre promedio (mean) * 100 . En ciencias exactas cómo movimientos de partículas, si se pasa de 10% los datos están desviados. En lo no exacto, si se pasa de 25% están desviados. No por tratarse de números como precios de un producto hablaríamos de algo exacto porque los precios suben y bajan debido al comportamiento humano el cual mueve oferta y demanda y por eso el tope es el 25% para poder decir que los datos no están desviados del promedio. Y cuando los datos están desviados es mejor describir

una situación con una mediana (la mitad de ... se encuentra por encima de). El coeficiente nos salva de los peligros del promedio.

La **desviación es la raíz cuadrada de la varianza**. La varianza calcula las distancias de cada dato frente al promedio. Pero cuando tenemos por ejemplo las desviaciones de dos conjuntos de datos (dos datasets) y una desviación (sd) es 2,3 y la otra es 4,8, pues simplemente podríamos decir que el dataset con 4,8 de desviación tiene los datos más desviados frente al promedio que el otro dataset, pero es el coeficiente de variación el que nos dice que tan desviados están los datos y en este ejemplo, los dos conjuntos de datos podrían tener un coeficiente sobre 25% y llevarnos a preferir describir los datos con una mediana o un cuartil y no con el promedio. Por ejemplo, con las notas de los estudiantes, yo como profesora prefiero un promedio más bajo, cómo 4,0 (sobre 5,0) con un coeficiente que no se pase de 25%, lo que significa que todos los alumnos tuvieron notas cercanas a ese promedio por encima o por debajo (grupo homogéneo) frente a otro grupo con un promedio de 4,5 pero con un coeficiente que supera el 25%, lo cual significa que tuve alumnos destacados que lograron notas finales de 5,0 y otros alumnos que perdieron la materia con 2,9 o menos (grupo heterogéneo).