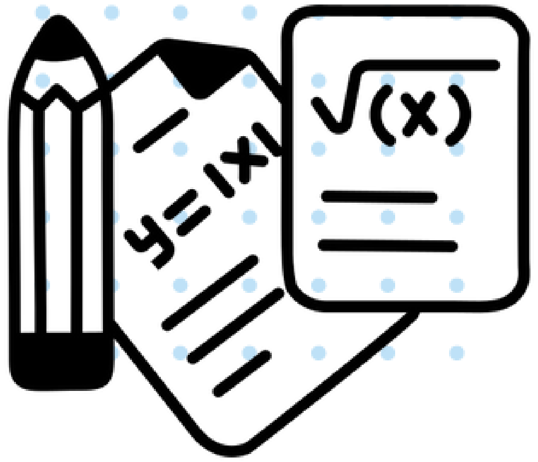
Lección 2: Momentos estadísticos



Tiempo de ejecución: 2 horas

PLANTEAMIENTO DE LA SESIÓN	MATERIALES
El análisis a través de las técnicas estadísticas permite modelas y entender el comportamiento de una variable aleatoria. Usualmente en un conjunto de datos, cada una de las características o columnas se pueden entender empleando el concepto de variable aleatoria para determinar qué comportamiento está modelado por los datos y saber las mejores técnicas matemáticas y de ciencia de datos que se pueden aplicar tanto para el tratamiento de los datos como para el análisis.	

Los momentos estadísticos de una variable aleatoria ayudan a determinar su comportamiento, en especial cuando los datos tienen una distribución de probabilidad poco conocida o desconocida. Cada uno de los momentos muestra diversas características de la variable aleatoria y ayuda a describirla.

Momentos respecto a la media

Dada una variable aleatoria X con función de probabilidad o densidad f(x) podemos definir una función de X que sea igual a la diferencia entre la variable y su media aritmética, elevada a un exponente entero no negativo.

$$z(x) = (x - \mu)^k$$

El valor esperado de z(x) es el k-ésimo momento de la variable X respecto a la media y se le llama: μ_k

$$\mu_k = \{\sum_x (x - \mu)^k f(x) \qquad si\ x\ es\ discreta\ \int_{-\infty}^{\infty} (x - \mu)^k f(x)dx \qquad si\ x\ es\ continua$$

Reemplazando se tiene que, para k=0

$$\mu_0 = E(x - \mu)^0 = 1$$

Es decir, si se calcula el momento para cuando k=0 al elevar la esperanza de la variable a cero el resultado es uno.

Para k=1:

$$\mu_1 = E(x - \mu)^1 = E(x) - \mu = 0$$

En el caso de k=1 se dice que es el primer momento estadístico de una variable. Como la esperanza de una variable es igual a la media, al restar tenemos cero. Quiere decir que el primer momento estadístico de una variable es la media (promedio). Como se vio anteriormente, ayuda a describir hacia donde tienden los datos.

Segundo momento estadístico (k=2):

$$\mu_2 = E(x - \mu)^2 = \sigma^2$$

La definición del segundo momento estadístico respecto a la media es llamado la varianza. Esta mide la dispersión de sus valores respecto al valor central . Para calcular la varianza se puede emplear la linealidad de la esperanza que permite extraer el valor central:

$$\sigma^2 = E\left|(x - \mu)^2\right| = E\left|x^2\right| - \mu^2 = E\left|x^2\right| - E|x|^2$$

Es decir, se puede calcular la varianza como la media de los cuadrados menos el cuadrado de la media. El problema de la varianza es que está expresada en unidades de la media al cuadrado, por lo que su interpretación no es clara para el contexto físico de la variable. Para solucionar el problema se suele emplear la raíz cuadrada de la varianza que se le conoce como desviación estándar.
Esta mide la dispersión en las mismas unidades que las de la media.

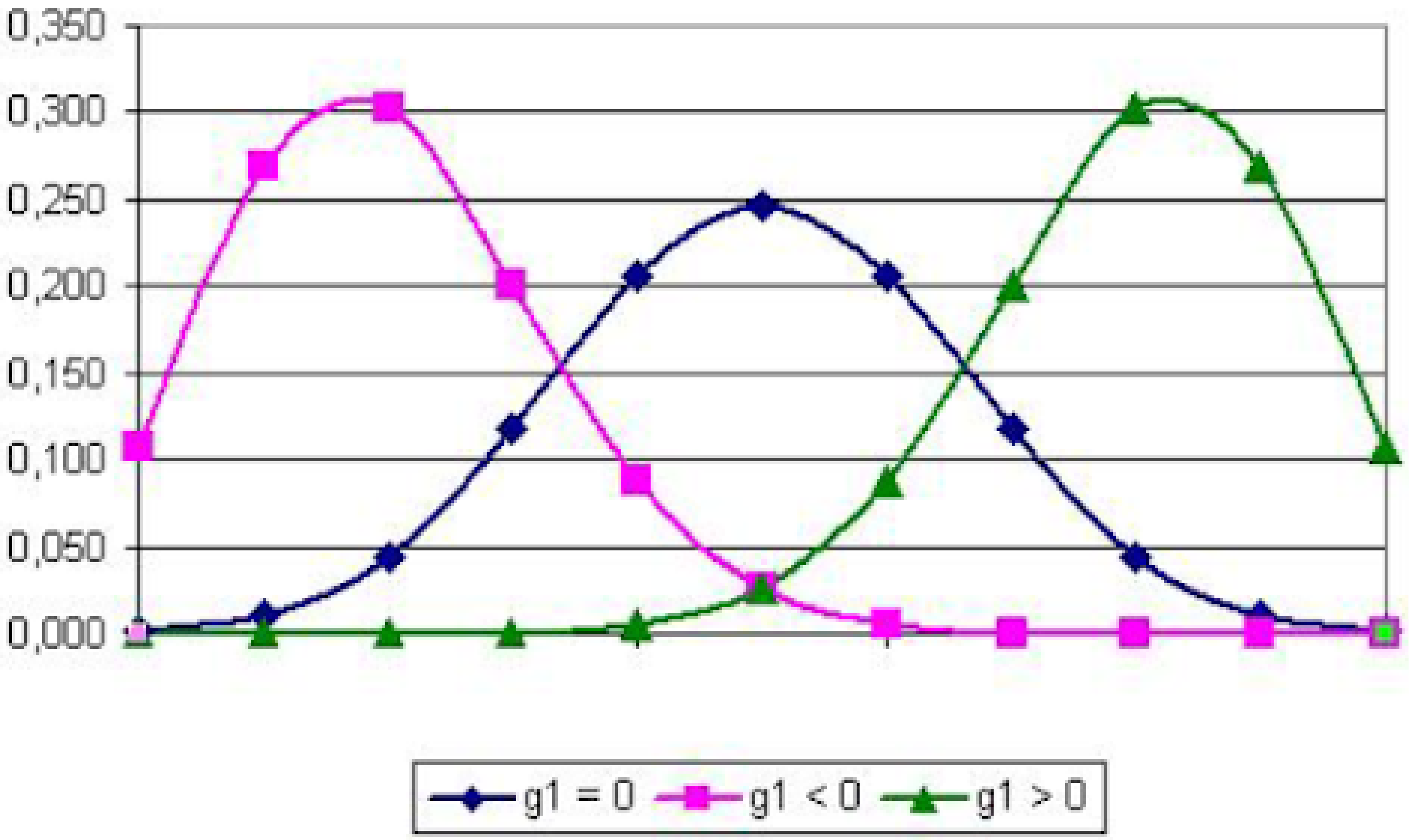
Tercer momento estadístico (k=3)

$$\mu_3 = E[(x - \mu)^3]$$

El tercer momento respecto de la media mide la asimetría de la distribución, es decir, si existen o no observaciones muy extremas en algún sentido con frecuencias razonablemente altas. Si la asimetría es negativa, la variable toma valores muy bajos con mayor frecuencia que valores muy altos y se dice que tiene una cola izquierda pesada o que es asimétrica hacia la izquierda. Si la asimetría es positiva, la variable toma valores muy altos con mayor frecuencia que valores muy bajos y se dice que tiene una cola derecha pesada o que es asimétrica hacia la derecha. Si la asimetría es cero, los valores bajos y altos de la variable tienen probabilidades iguales (el ejemplo más típico de variable simétrica es la variable normal).

La asimetría tiene el mismo problema que la varianza y la covarianza en cuanto a sus unidades de medida y, por ello, normalmente se utiliza una medida adimensional de la asimetría que es el coeficiente de asimetría, g1 , que se calcula como el cociente entre el tercer momento y el cubo de la desviación típica.

$$g_1 = \frac{\mu_3}{\sigma^3}$$


0,350
0,300
0,250
0,200
0,150
0,100
0,050
0,000
g1 = 0
g1 < 0
g1 > 0

Cuarto momento estadístico (k=4)

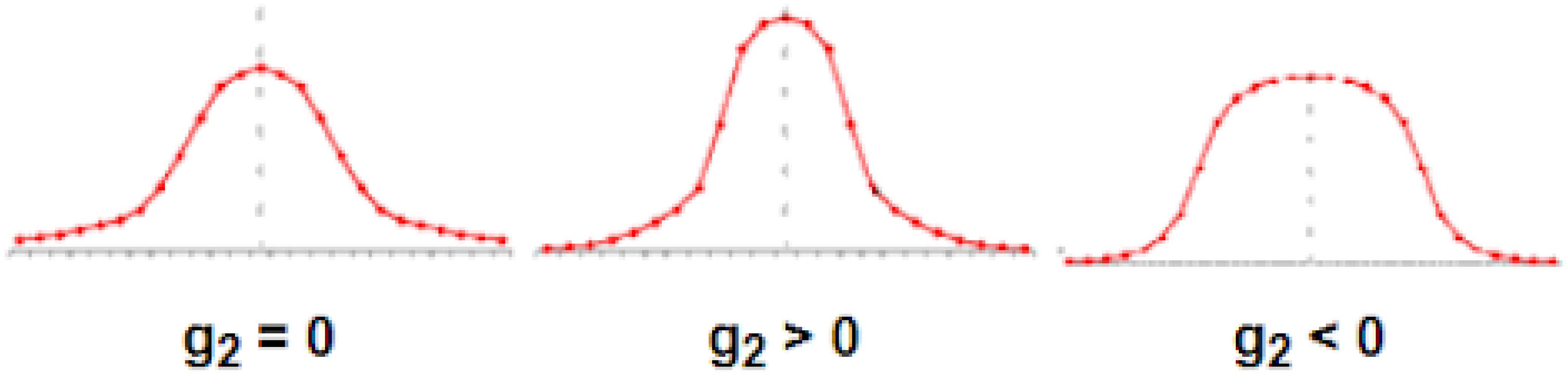
$$\mu_4 = E[(x - \mu)^4]$$

El cuarto momento respecto de la media mide la curtosis de la distribución, es decir, la forma de la distribución de probabilidad. Al representar gráficamente variables con curtosis pequeña, platicúrticas, se observan curvas o histogramas con colas cortas y aspecto aplanado o en meseta; si la variable tiene curtosis grande, es decir, si es leptocúrtica, su gráfica ser alta y estilizada, con colas largas y pesadas.

La curtosis de una variable siempre es positiva y se mide en la unidades de la variable elevadas a potencia 4. Por tanto, nuevamente se nos plantean los problemas relacionados con las unidades de medida y las escalas y necesitamos una medida adimensional de la curtosis.

Esta medida adimensional de la curtosis es el coeficiente de curtosis, g2 , que se calcula como el cociente entre el cuarto momento y el cuadrado de la varianza, al que se le resta 3 unidades. Esta corrección se debe a que, sin ella, las variables normales tendrían coeficiente de curtosis igual a 3; al restar 3 conseguimos que el coeficiente de curtosis de la variable normal sea 0 y que las variables platicúrticas tengan coeficiente de curtosis negativo y la leptocúrticas positivo, lo cual es más mnemotécnico que la distinción entre curtosis pequeña y grande.

$$g_2 = \frac{\mu_4}{\sigma^4} - 3$$



$g_2 = 0$
$g_2 > 0$
$g_2 < 0$