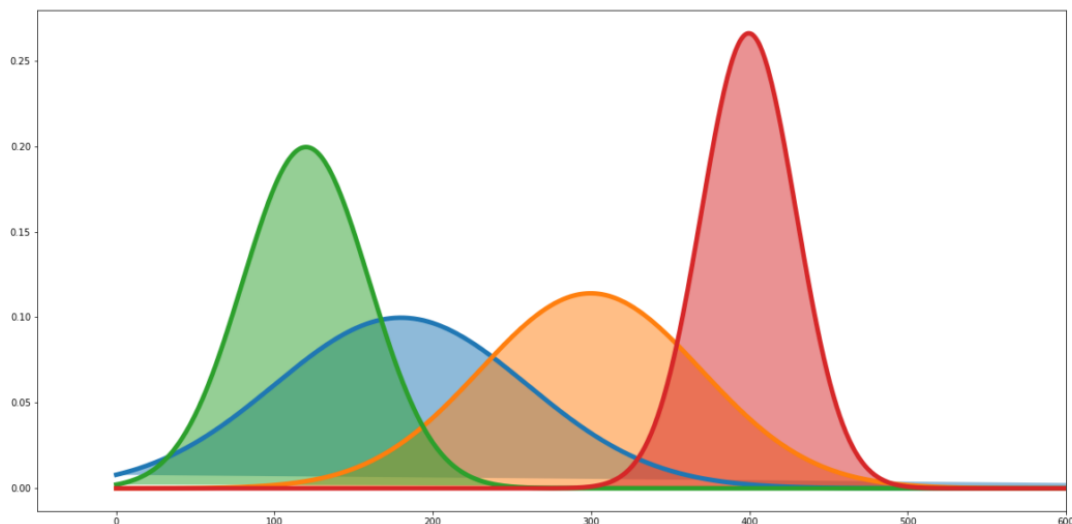


# IDENTIFICACIÓN DE VALORES ATÍPICOS A TRAVÉS DEL USO DE Puntuaciones Z

## DISTRIBUCIÓN NORMAL

La distribución normal es una distribución con forma de campana donde las desviaciones estándar sucesivas con respecto a la media establecen valores de referencia para estimar el porcentaje de observaciones de los datos<sup>1</sup>.

Aunque la forma de la campana puede variar la distribución siempre debe ser simétrica con más casos concentrados en el centro y menos en los extremos<sup>2</sup>.



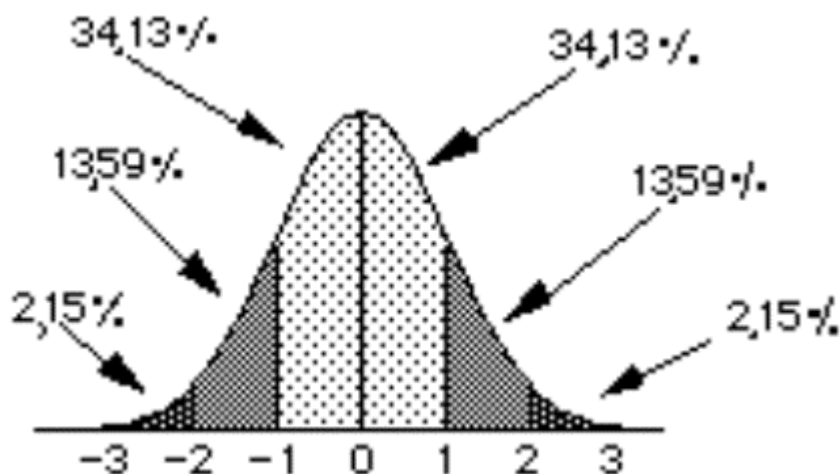
Fuente: <https://machinelearningparatodos.com/como-saber-si-una-variable-sigue-una-distribucion-normal-en-python/>

---

<sup>1</sup> <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/normality/what-is-the-normal-distribution/>

<sup>2</sup> <https://aathosc.tripod.com/PuntajeZ22.htm>

En una distribución perfectamente normal los casos se distribuyen de la siguiente manera<sup>3</sup>:



**El 68,26% del área total de la curva cae entre más menos 1 desviación estándar. Exactamente 95,44% del área cae entre más menos 2 desviaciones estándares y exactamente 99,74% del área cae entre más menos 3 desviaciones estándares.**

La forma de la distribución normal se define por dos parámetros:

- Media. Es el valor que con mayor probabilidad aparecerá en una medida.
- Desviación estándar. Refleja lo abierta o cerrada que es la campana de Gauss correspondiente.

Una distribución muy cerrada corresponde a una serie de medidas muy poco dispersas. Por el contrario, si la distribución es abierta, la desviación estándar es grande.

### ¿Qué es el puntaje Z o puntaje estandarizado?

Los puntajes Z son transformaciones que se pueden hacer a los valores o puntuaciones de una distribución normal, con el propósito de analizar su distancia respecto a la media, expresándolas en unidades de desviación estándar.

Un puntaje Z nos indica la dirección y grado en que un valor individual obtenido se aleja de la media, en una escala de unidades de desviación estándar.

---

<sup>3</sup> <https://aathosc.tripod.com/PuntajeZ22.htm>

La fórmula para transformar un valor de una distribución normal en una unidad de desviación estándar es:

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

**Ejemplo:**

Si la edad promedio del grupo de estudiantes de psicología es de 23 años y la desviación estándar es igual a 4, un estudiante de 27 años se ubicaría 1 desviación estándar ( $Z=1$ ) respecto al promedio.

$$\mu = 23$$

$$\sigma = 4$$

$$x = 27$$

$$Z = \frac{27 - 23}{4} = 1$$

**Puntuaciones Z para identificar valores atípicos<sup>4</sup>**

Los puntajes Z pueden cuantificar lo **inusual** de una observación cuando sus datos siguen la distribución normal. Las puntuaciones Z son el número de desviaciones estándar por encima y por debajo de la media que cae cada valor.

Por ejemplo, una puntuación Z de 2 indica que una observación está dos desviaciones estándar por encima del promedio, mientras que una puntuación Z de -2 significa que está dos desviaciones estándar por debajo de la media. Una puntuación Z de cero representa un valor que es igual a la media.

---

<sup>4</sup> <https://statisticsbyjim.com/basics/outliers/>

Cuanto más lejos de cero está el puntaje Z de una observación, más inusual es. Un valor de corte estándar para encontrar valores atípicos son las puntuaciones Z de  $\pm 3$  o más desde cero. Los puntajes Z más allá de  $\pm 3$  son tan extremos que apenas se puede observar debajo de la curva.

