REVISTA CIE

Revista de Investigación Educativa | Vol. 2, No. 2
ISSN: 3061-8657 | Julio-Diciembre 2026

ESTADÍSTICA

Distribución Normal

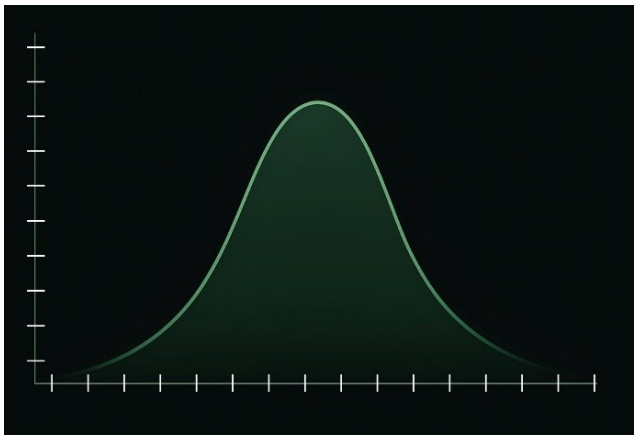
Dr. Gerardo García Maldonado
Profesor de Tiempo Completo e Investigador
SNII Nivel 1
Universidad Autónoma de Tamaulipas
Facultad de Medicina de Tampico "Dr. Alberto Romo Caballero"
gmaldonado@docentes.uat.edu.mx

Dr. Raúl de León Escobedo
Profesor de Tiempo Completo e Investigador
Candidato SNII
Universidad Autónoma de Tamaulipas
Facultad de Medicina de Tampico "Dr. Alberto Romo Caballero"
raul.deleon@docentes.uat.edu.mx

Dr. José Eugenio Guerra Cárdenas
Profesor de Tiempo Completo e Investigador
Universidad Autónoma de Tamaulipas
Facultad de Medicina de Tampico "Dr. Alberto Romo Caballero"
jguerra@docentes.uat.edu.mx

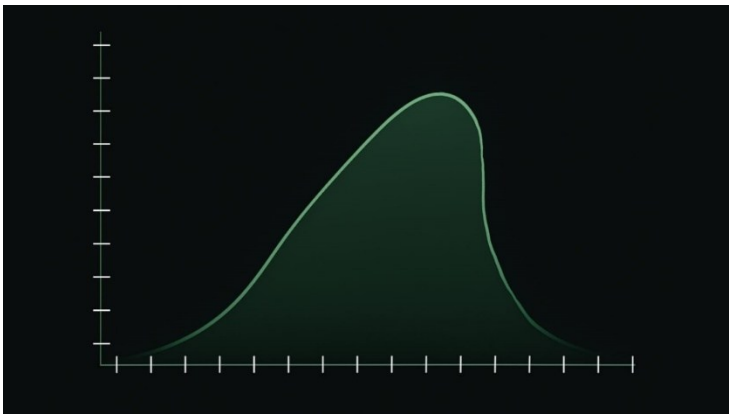
INTRODUCCIÓN

La distribución normal es un modelo teórico ampliamente utilizado que describe cómo se distribuyen los valores numéricos de una variable en un conjunto de datos. Se le conoce comúnmente como curva de campana, debido a que la gráfica de su función de densidad tiene una forma similar a la de una campana.



Teóricamente se asume que los datos de cualquier variable aleatoria, que por cierto siempre son numéricas, ya sea de intervalo o de razón, generalmente se pueden aproximar a una distribución normal. Sin embargo, esto puede en ocasiones no ser así, de ahí que se recomienda realizar procedimientos de ajuste y/o modificar la escala de medición para asegurar la distribución normal en los datos, considerando que, para la utilización de estadística paramétrica, se exige este supuesto.

Por ejemplo, si tenemos a un grupo de personas entre 40/50 años de edad y les practicáramos determinaciones de creatinina sérica, sería probable que la distribución de los datos estuviera sesgada, ya que estos valores aumentan con la edad. Esta realidad daría lugar a una curva de distribución de datos asimétrica con inclinamiento a la derecha.



Para solucionar este contratiempo, podríamos implementar, por ejemplo, un procedimiento estadístico denominado transformación log, lo que permitiría normalizar datos o reducir el sesgo.

La distribución normal es un modelo definido mediante un algoritmo matemático, que se calcula a partir de una fórmula:

$$f(x) = \frac{1}{\sqrt{2}\pi\sigma} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

En donde:

f(x)= Función de una variable x

σ= Desviación típica o estándar poblacional

σ^2= Varianza poblacional

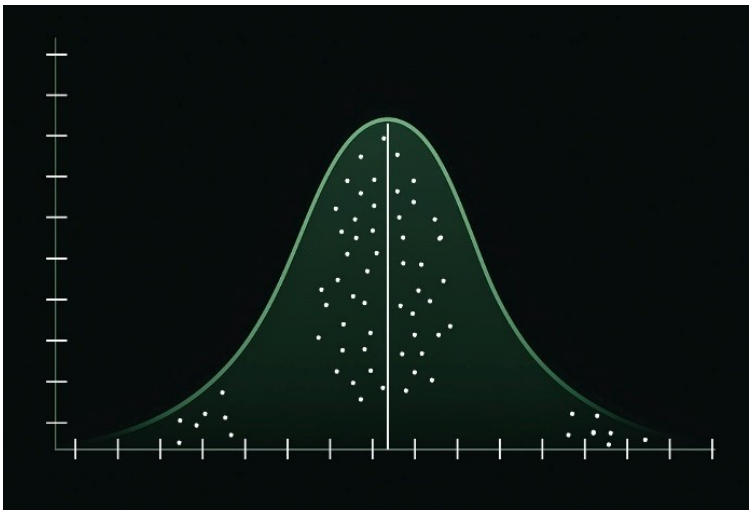
π= 3.1416

e= Base de logaritmo neperiano= 2.7182

μ= Media poblacional de la variable aleatoria X

Este tipo de distribución, también conocida como campana de Gauss, es una de las distribuciones de probabilidad más importantes y comúnmente utilizadas en el campo de las matemáticas, la estadística y diversas disciplinas científicas.

Se caracteriza por su forma simétrica, donde la mayor parte de los valores se concentran alrededor de la media, lo cual refleja también, que la probabilidad de observar valores atípicos o extremos (outliers), disminuya exponencialmente a medida que se alejan de la media.



El alemán Carl Friedrich Gauss (de ahí el nombre de campana de Gauss), introdujo este modelo derivado de sus trabajos y observaciones en campos como la astronomía, física y matemáticas, pero, sobre todo, derivado de sus estudios sobre el denominado método de mínimos cuadrados, la teoría de los números y la distribución de los números primos.

Conviene también destacar que este modelo permite realizar inferencias estadísticas, y lo más importante, la interpretación de los datos dependerá del contexto del problema analizado, ya que describe el comportamiento de una variable aleatoria numérica bajo ciertas condiciones naturales o experimentales.

En el estudio e interpretación de una curva de distribución normal, la media y la varianza son parámetros importantes para determinar la forma y posición de la curva.

- La media define el punto central, alrededor del cual se distribuyen los valores, determinando simetría.
- La varianza influye en la dispersión de los datos, determinando anchura.

La desviación estándar, también juega un papel importante en la estructuración de la forma. Una desviación estándar y una varianza mayor significa que la curva es más ancha y plana, es decir, los valores están más dispersos. Por el contrario, si son menores, implica que los datos están concentrados alrededor de la media, la curva en consecuencia será más picuda por decirlo así. La anchura de la curva se refiere a su apariencia visual, es decir, qué tan amplia o estrecha parece ser.

La variabilidad se refiere a la dispersión de los valores alrededor de la media. Vale la pena señalar que la varianza se utiliza con más frecuencia, ya que sus propiedades permiten realizar:

- Modelado de procesos aleatorios
- Comparaciones de variabilidad.
- Análisis de riesgos.
- Distribuciones multi-variantes.

De hecho, en la fórmula de la distribución normal, la varianza aparece en el denominador del exponente.

Aplicaciones de la distribución normal

- Análisis de datos en casos de distribución simétrica y unimodal.
- Estadística inferencial para pruebas de hipótesis.
- Estructuración de intervalos de confianza.

Limitaciones

- Supuesto de distribución normal de datos, que no siempre es real.
- Sensibilidad a los outliers, lo que afecta precisión.

Otros conceptos importantes son la función de densidad de probabilidad (FDP) y el área bajo la curva, los cuales, si bien están relacionados, no son exactamente lo mismo.

La FDP, describe precisamente la probabilidad de que una variable aleatoria tome un valor específico. Es una función matemática, que asigna una probabilidad a cada valor posible de las variables.

Presenta varias propiedades, pero se destacan las siguientes:

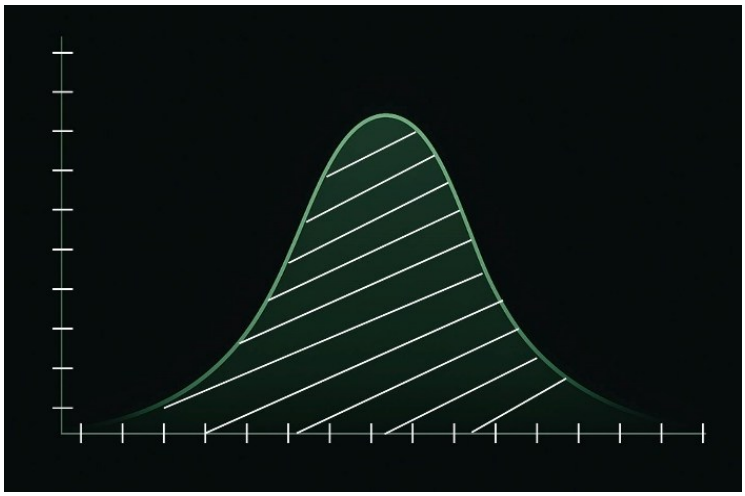
- No negatividad: es decir es no negativa para todos los valores de x.
- Integral igual a 1: la integral sobre todos los valores posible de x es igual a 1.

Ejemplos de interpretación

- Máxima densidad: El punto más alto de la FDP indica el valor más probable de la variable aleatoria.
- Colas de la distribución: Las colas de la distribución indican la probabilidad de valores extremos.
- Simetría: La simetría de la FDP indica si la distribución es simétrica o asimétrica.

Por otro lado, el área bajo la curva de la FDP representa la probabilidad acumulada de que la variable tome un valor dentro de un intervalo específico:

- El área total bajo la curva es igual a 1, lo que representa a la probabilidad total.



En otras palabras, la FDP es la altura de la curva en cada punto, mientras que el área bajo la curva es la integral de la FDP sobre un intervalo específico.

Una variable aleatoria con media (μ) y desviación típica (σ) poblacionales (N), que se distribuye normalmente, se esquematizará de la siguiente manera: N (μ, σ).

Ejemplo

Si la glicemia en ayunas de un conjunto de personas tiene una media aritmética de 110 mg/dl con una desviación típica de 10 mg/dl y se distribuye normalmente; se puede expresar de la siguiente manera: N (110, 10).

La distribución normal es fundamental en muchos campos y se utiliza cuando:

- Los datos siguen una tendencia central, es decir, la mayoría de los valores se agrupan alrededor de una media (el centro).
- Se requieren modelos estadísticos para hacer inferencias sobre poblaciones grandes a partir de muestras.
- Si se cuenta con un gran número de observaciones, tomamos suficientes muestras aleatorias de cualquier distribución con media y varianza finita, la distribución de las medias muestrales tenderá a ser normal (Teorema del Límite Central).

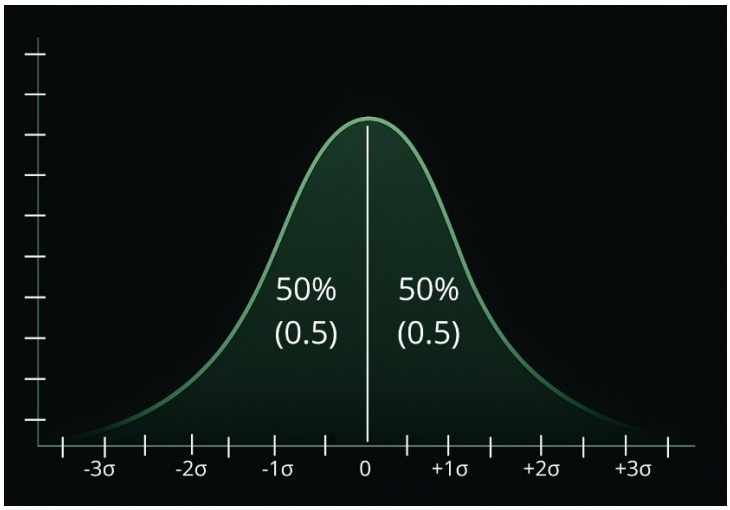
Ejemplos de uso

- Altura humana: En una población dada, la altura de los individuos sigue una distribución normal, donde la mayoría de las personas tienen una altura cercana a una media y solo unas pocas serán extremadamente altas o bajas.
- Puntuaciones de exámenes escolares estandarizados; las puntuaciones de muchos exámenes tienden a distribuirse normalmente. La mayoría de los estudiantes obtienen una puntuación cercana al promedio y los puntajes extremos serán menos frecuentes.

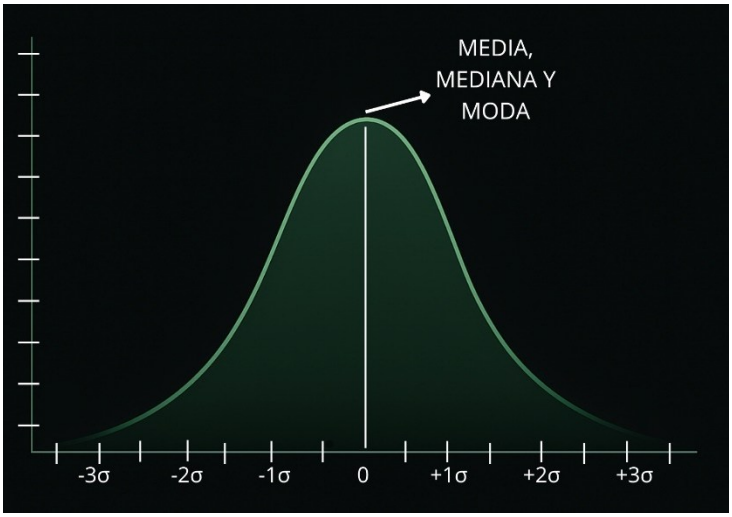
La curva normal, dada su importancia, ha sido estudiada exhaustivamente y algunas de sus propiedades son de gran utilidad para el cálculo de probabilidades.

Para interpretar el área bajo la curva de Gauss, tomemos en cuenta lo siguiente:

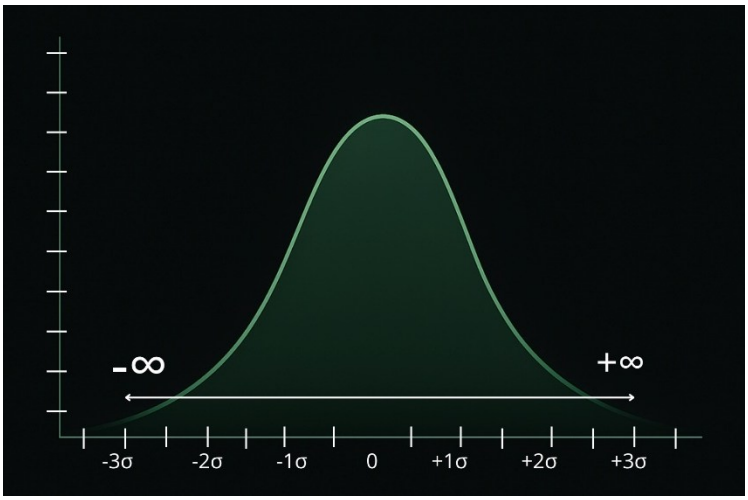
- El área bajo la curva, a ambos lados de la media es el mismo (50% o expresado también en decimales 0.5), de tal manera que el lado derecho es como un reflejo en el lado izquierdo.



- La distribución normal tiene la particularidad que la media, la mediana y la moda son iguales, lo que la convierte en una distribución simétrica, es decir, deben tener el mismo valor, que estarán representados en la punta de la curva.

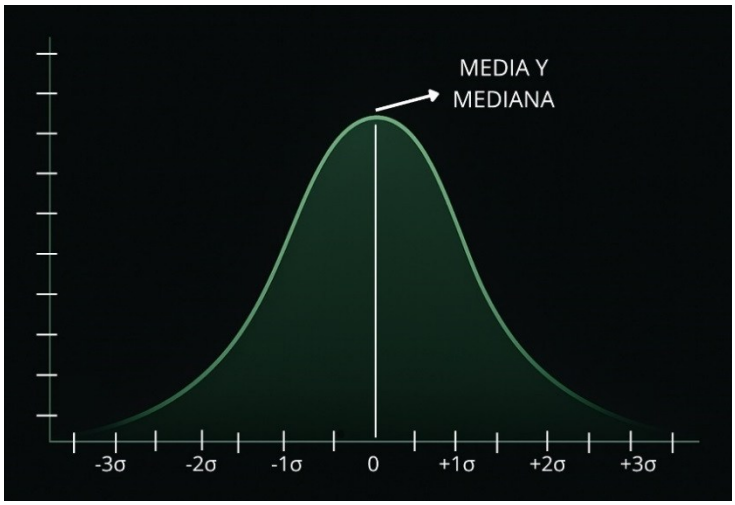


- El área entre la curva y el eje de las abscisas representa la probabilidad de que la variable aleatoria toma un valor cualquiera entre -∞ y +∞, dicha probabilidad debe ser igual a 1.

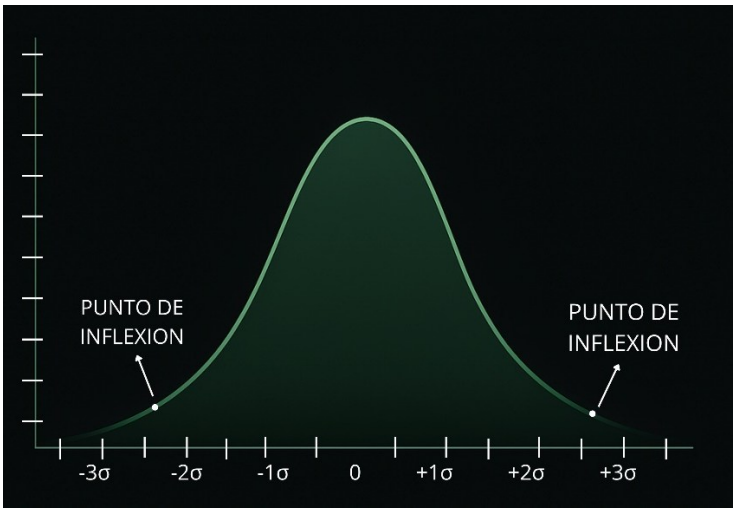


- Los valores de la variable que se alejan de la media aritmética más de 3 desviaciones típicas (σ), tendrán una menor probabilidad de que ocurran.

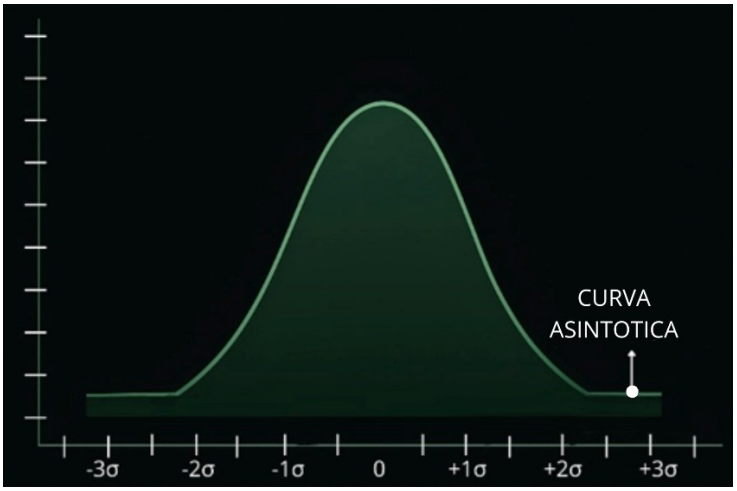
Cualquier valor "muy alejado" de la media es posible, aunque poco probable.



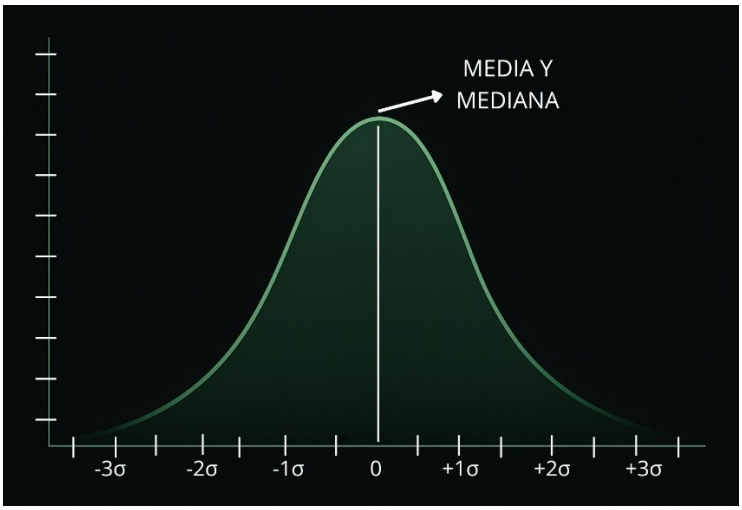
- Los puntos de inflexión (donde pasa de cóncava a convexa) se corresponden con valores de X iguales a (μ - σ) y (μ + σ).



- Las colas son asintóticas, es decir, se acercan, pero nunca llegan a tocar la línea horizontal (también denominada abscisa o eje de la X).



- La mediana divide los datos en dos partes iguales en cuanto a su número, pero si los datos se distribuyen normalmente, si la media y la mediana coinciden, entonces la media aritmética también puede dividir los datos en dos partes iguales en cuanto a su número.



PUNTUACIONES Z

También se les conoce como puntuaciones estándar. Esta medida estandarizada es de utilidad para calcular la probabilidad de que una puntuación ocurra dentro de una distribución normal y es medida en términos de desviaciones estándar. Se considera que también es útil para comparar dos puntuaciones, que pertenezcan a distribuciones normales diferentes.

Para entender un poco más la manera de interpretar la distribución normal, el concepto del área bajo la curva y las puntuaciones z, es necesario, primeramente hacer otras consideraciones respecto a la denominada desviación estándar, la cual es una medida de dispersión para variables numéricas, sobre la que se basan los cálculos de los valores z.

Desviación estándar o típica

- En primer término, habría que señalar que se trata de una medida de variabilidad, cuya función es valorar la dispersión de los datos respecto a su media, lo que permite medir cuanto se desvían los valores individuales de la media de un conjunto de datos.
- En segundo lugar, identifica valores atípicos (outliers).
- Finalmente permite comparar la dispersión de diferentes conjuntos de datos, y calcular intervalos de confianza.

La fórmula es:

$$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$$

Donde:

σ=Desviación estándar o típica

x=Valores individuales

μ=Media poblacional

N= Número de datos

$\sum$ = Sumatoria

Esta medida de variabilidad es interesante, ya que permite hacer comparaciones en medidas diferentes.

Ejemplo: Comparación de edades en dos grupos.

Supongamos que tenemos dos grupos de estudiantes:

Grupo A

- Edades: 18, 19, 20, 21, 22
- Media: 20 años
- Desviación estándar: 1,41 años

Grupo B

- Edades: 18, 22, 25, 28, 30
- Media: 24,6 años
- Desviación estándar: 4,76 años

Análisis

Aunque la media de edad del Grupo B es mayor, lo importante es comparar la desviación estándar. La desviación estándar del Grupo A es baja (1,41 años), lo que indica que las edades están cerca de la media (20 años). Por otro lado, la desviación estándar del Grupo B es alta (4,76 años), lo que indica que las edades están más dispersas y alejadas de la media (24,6 años).

Conclusión

La desviación estándar nos permite comparar la variabilidad de los datos en ambos grupos. En este caso, el Grupo A tiene una mayor homogeneidad en las edades, mientras que el Grupo B tiene una mayor diversidad, de esta manera se proporciona más información, contrario a que si solamente nos enfocamos en una medida de tendencia central como la media. Esto puede ser útil en diversas aplicaciones, como la planificación de programas educativos, programas de salud preventivos, o la identificación de afectaciones en la salud en la población, por solo citar algunos ejemplos.

En síntesis

- Una desviación estándar alta, indica que los datos están muy dispersos y alejados de la media.
- Una desviación estándar baja, indica que los datos están cerca de la media y que hay poca variabilidad.

Ahora bien, es importante señalar que con regularidad las puntuaciones crudas de datos numéricos son transformados a otro tipo de puntuaciones, que, sin lugar a dudas, posibilitan comparaciones e interpretaciones en forma más comprensible.

Una forma de transformar los datos es a través de las denominadas puntuaciones z, también conocidas como valores z o z-scores.

Se considera que esta estrategia es una forma de estandarizar datos dentro de una distribución normal, lo que permitirá indicar a cuántas desviaciones estándar se encuentra un valor determinado, respecto a la media del conjunto de datos.

Interpretación de la puntuación z

Z = 0 el valor está exactamente en la media.

Z > 0 el valor está por encima de la media.

Z < 0 el valor está por debajo de la media.

Fórmula para calcular una puntuación z:

$$z = \frac{x - \mu}{\sigma}$$

Donde:

x = valor individual

μ = media del conjunto de datos

σ = desviación estándar poblacional

Cuanto mayor sea el valor absoluto del z-score, más alejado está el valor de la media.

Ejemplo:

Supongamos que estamos analizando las puntuaciones de un examen.

- La media es μ=70
- La desviación estándar es σ=10

Queremos saber la puntuación z de un estudiante que obtuvo X=85

Aplicamos la fórmula:

$$z = \frac{x - \mu}{\sigma} = \frac{85 - 70}{10} = \frac{15}{10} = 1.5$$

Interpretación del resultado

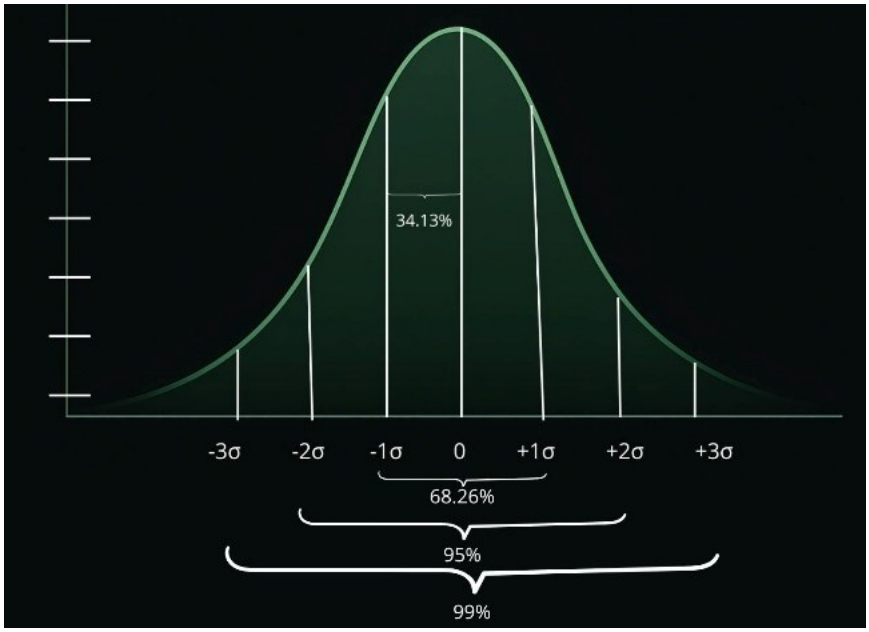
El estudiante que obtuvo 85 puntos, está 1.5 desviaciones estándar por encima de la media. Esto indica un rendimiento notablemente superior al promedio. La puntuación z se obtiene restando el valor de un dato a su media y dividiendo el resultado precisamente entre la desviación estándar.

Este tipo de procedimiento permite detectar valores atípicos e indica, en un valor determinado, a cuántas desviaciones estándar está, ya sea por encima o por debajo de la media aritmética en una distribución. Cuando esta distribución es estrictamente normal, se considera que la puntuación z tiene un valor de 0 y desviación estándar de 1.

Otras consideraciones:

- El área bajo la curva representa el 100% de los datos, la mitad representa el 50% de cada lado.
- En términos de proporción el área bajo la curva es igual a 1.

Cuando una agrupación de valores expresados en términos numéricos, con escala de intervalo presenta una distribución normal, la representación de los mismos como puntuaciones z son de la siguiente manera:



- Una desviación estándar, a la derecha e izquierda de la media contendrá al 68.26% del total de los datos, (cada lado de la curva tendrá en consecuencia el 34.13%).
- Dos desviaciones estándar, a la derecha e izquierda de la media, integrará al 95.44% del total de los datos (cada lado de la curva tendrá el 47.72%). Aunque en la práctica se maneja solo como 95%.
- Tres desviaciones estándar, a la derecha e izquierda de la media mostrará al 99.74% del total de los datos (cada lado de la curva tendrá el 49.87%), es decir, solo 2.15% más que dos desviaciones estándar. Aunque en la práctica se maneja solo como 99%.

Como dato agregado y para ser más exactos, convendrá decir que en la realidad el 95% de los datos está entre 1.96 y -1.96 desviaciones estándar, el 99% entre 2.58 y -2.58 desviaciones estándar y el 99% entre 3.90 y -3.90 desviaciones estándar. Esta aclaratoria es fundamental para procedimientos de estadística inferencial.

En resumen, las puntuaciones z representan una medida estadística que indica a cuántas desviaciones estándar se encuentra un valor, respecto a la media en un conjunto de datos. Esta herramienta es útil para comparar datos de diferentes distribuciones y evaluar su posición relativa en relación con la media.

Ejemplo A:

Supongamos que tenemos un conjunto de calificaciones de un examen aplicado a estudiantes universitarios y se busca calcular la puntuación Z de un estudiante en particular. (Supongamos que no se cuenta en este caso con desviación estándar)

Conjunto de calificaciones de 10 estudiantes: 70, 85, 90, 60, 75, 80, 95, 55, 85, 100

Paso 1: Calcular la media (promedio)

La media (μ) se calcula sumando todos los valores y dividiendo el resultado entre el número total de datos:

$$\mu = \frac{70 + 85 + 90 + 60 + 75 + 80 + 95 + 55 + 85 + 100}{10} = \frac{835}{10} = 83.5$$

Paso 2: Calcular la desviación estándar (σ)

La desviación estándar se calcula utilizando la fórmula:

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{n}}$$

Donde Xi son los valores del conjunto de datos, μ es la media, y n es el número total de datos. Para este conjunto de datos sería lo siguiente:

$$\sigma = \sqrt{\frac{(70 - 83.5)^2 + (85 - 83.5)^2 + (90 - 83.5)^2 + \cdots + (100 - 83.5)^2}{10}} = \frac{\sqrt{4402.5}}{10}$$

$$= \sqrt{440.25} \approx 20.98$$

Paso 3: Calcular la puntuación Z para una calificación específica

La puntuación Z se calcula usando la siguiente fórmula:

$$z = \frac{x - \mu}{\sigma}$$

Donde X es la calificación del estudiante. Si un estudiante obtuvo 95 puntos, entonces:

$$z = \frac{95 - 83.5}{20.98} \approx \frac{11.5}{20.98} \approx 0.55$$

Interpretación

La puntuación Z= 0.55, significa que la calificación cruda de 95 puntos está a 0.55 desviaciones estándar (valor dentro de una desviación estándar) por encima de la media. En otras palabras, el estudiante tiene una calificación ligeramente superior al promedio de la clase (la puntuación está entre la media y una desviación estándar).

APLICACIÓN EN MEDICINA

Ejemplo B:

Crecimiento Infantil

Las puntuaciones Z permiten comparar el peso, la talla y el índice de masa corporal (IMC) de un niño con los valores esperados según su edad y sexo.

Un niño de 5 años pesa 20 kg.

La media del peso para niños de esa edad es 18 kg, con una desviación estándar de 2 kg. (Se cuenta en este caso con dato de desviación estándar)

Se calcula la puntuación Z:

$$z = \frac{x - \mu}{\sigma} = \frac{20 - 18}{2} = \frac{2}{2} = 1.0$$

Interpretación

Un Z = 1.0 significa que el niño en cuanto al peso está a 1 desviación estándar por encima de la media, lo que sugiere que su peso es superior al promedio, pero dentro del rango normal.

Si Z fuera mayor a +2 o menor a -2, indicaría sobrepeso o bajo peso significativo.

Ejemplo C:

Densitometría Ósea

En estudios de densidad mineral ósea (DMO), como la absorciometría de rayos X de energía dual (DXA), la puntuación Z ayuda a comparar la densidad ósea de un paciente con la de individuos de la misma edad y sexo.

Una mujer de 60 años tiene una DMO de 0.85 g/cm².

La media para su grupo de edad es 1.0 g/cm² con una desviación estándar de 0.1 g/cm².

Puntuación Z:

$$z = \frac{0.85 - 1.0}{0.1} = \frac{-0.15}{0.1} = -1.5$$

Interpretación

Un Z = -1.5 sugiere que su densidad ósea es 1.5 desviaciones estándar por debajo de la media, es decir menor al promedio, lo que podría indicar osteopenia incluso valores más bajos sugerirían osteoporosis.

Ejemplo D:

Pruebas de Laboratorio y Valores Clínicos

Las puntuaciones Z también pueden utilizarse para analizar biomarcadores en exámenes de laboratorio, como niveles de glucosa, colesterol o en el análisis de signos vitales como la presión arterial y comparándolos con valores poblacionales de referencia.

Un paciente tiene una presión arterial sistólica de 140 mmHg.

La media poblacional es 120 mmHg con una desviación estándar de 10 mmHg.

Puntuación Z:

$$z = \frac{140 - 120}{10} = \frac{20}{10} = 2.0$$

Interpretación

Un Z = 2.0 indica que la presión arterial del paciente está 2 desviaciones estándar por encima de la media, mayor el promedio, lo que podría sugerir hipertensión arterial sistólica.

REFERENCIAS

Altman DG. Practical Statistics for Medical Research. Boca Raton (FL): CRC Press; 1991.

Bland M. An Introduction to Medical Statistics. 4th ed. Oxford: Oxford University Press; 2015.

Daniel WW, Cross CL. Biostatistics: A Foundation for Analysis in the Health Sciences. 10th ed. Hoboken (NJ): Wiley; 2013.

Devore JL. Probability and Statistics for Engineering and the Sciences. 9th ed. Boston (MA): Cengage Learning; 2015.

Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd ed. Hoboken (NJ): Wiley; 2013.

Kirk RE. Statistics: An Introduction. 5th ed. Belmont (CA): Wadsworth Publishing; 2007.

Martínez González MA, Sánchez-Villegas A, Toledo Atucha E. Bioestadística amigable. 2ª ed. Barcelona: Elsevier; 2020.

Montgomery DC, Runger GC. Applied Statistics and Probability for Engineers. 7th ed. Hoboken (NJ): Wiley; 2018.

Pagano M, Gauvreau K. Principles of Biostatistics. 2nd ed. Belmont (CA): Duxbury Press; 2000.

Sokal RR, Rohlf FJ. Biometry: The Principles and Practice of Statistics in Biological Research. 4th ed. New York: W.H. Freeman and Company; 2012.

Triola MF. Estadística. 13ª ed. Madrid: Pearson Educación; 2019.

Zar JH. Biostatistical Analysis. 5th ed. Upper Saddle River (NJ): Pearson Prentice Hall; 2010.