Teorema del límite central

2023-11-13

ÍNDICE

1 La distribución de la media muestral

Supongamos que queremos estudiar la media de la altura de los estudiantes de una universidad. De entre ellos hemos seleccionado una muestra al azar, los hemos medido y hemos calculado la media de las alturas de los estudiantes de la muestra. Ahora queremos ver cómo se comporta esta media muestral.

Veremos que si sabemos que la variable que se estudia es normal, entonces la media muestral también es normal, pero con desviación típica menor. También veremos que si la variable no es normal, pero la muestra es lo bastante grande, la media también será aproximadamente normal.

1.1 Distribución de la media muestral para variables normales

Supongamos que tenemos una muestra $x_1, ..., x_n$ de una variable aleatoria normal. La media se define como:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Esta media depende de la muestra. Normalmente tendremos solo una muestra, pero podríamos tomar muchas diferentes, de manera que a cada una le correspondería una media diferente. Esto nos da pie a hablar de la distribución muestral de la media. Para indicar que se trata de una variable aleatoria, la denotaremos por $overline X$.

Deberemos distinguir dos casos: cuando la deviación típica de la variable que medimos es conocida y cuando es desconocida.

1.1.1 Caso de desviación típica poblacional conocida

La desviación poblacional es la desviación real de la variable, que en este caso suponemos conocida. Cuando calculamos la desviación a partir de muestras, hablamos de *desviación muestral*.

Supongamos que en un estudio anterior se había demostrado que las alturas de los estudiantes seguían una distribución normal de media 172 cm y desviación típica de 11 cm.

Intuitivamente vemos que la media de las observacioens de la muestra que tenemos debe ser un valor cercano a 172. También parece razonable pensar que observaciones mayores que la meida poblacional, 172, se compensarán con valores menores, y que cuanto mayor sea la muestra, más cercano será el valor de la media muestral a 172.

Pensemos ahora que tenemos una muestra de cien estudiantes. Hacemos diez grupos de diez estudiantes y hacemos la media aritmética para cada grupo. Obtenemos diez valores, correspondientes a las diez medias $\overline{x}_1, ..., \overline{x}_{10}$. Parece razonable pensar que la media de estos nuevos datos sería también 172. Por otra parte, también parece razonable pensar que estos nuevos valores sean más cercanos a 172 que los datos originales, ya que en cada una de las medias se nos habrán compensado valores grandes con valores pequeños.

Si la variable que estudiamos sigue una distribución normal con media μ y desviación típica σ conocidas, entonces la media muestral es también normal con la misma media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$, donde n es el tamaño de la muestra. Por tanto, tipificamos la variable $\overline{X}$ y obtenemos que:

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

sigue una distribución normal estándar.

La demostración de este resultado es consecuencia de una importante propiedad de las variables aleatorias normales. La propiedad es la siguiente: si X e Y son variables aleatorias independientes con leyes

$$N(\mu_1, \sigma_1^2) \text{ y } N(\mu_2, \sigma_2^2)$$

respectivamente, entonces $X + Y$ tiene una ley:

$$N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

En el ejemplo, la variable que recoge todas las posibles medias de cada grupo de diez estudiantes sigue una distribución normal de media 172 cm y desviación típica $\dfrac{11}{\sqrt{10}} = 3.48\ cm$. Observamos que cuanto mayor es la muestra, menor resulta la desviación típica y, por tanto, hay menor dispersión.

Este cociente que nos da la desviación típica de la media muestral se conoce como *error estándar*.

Si σ es la desviación típica de la población y n el tamaño de la muestra, se define el **error estándar de la media muestral** como:

$$\frac{\sigma}{\sqrt{n}}$$

Ejemplo de error estándar de una media muestral

Consideremos las alturas de los estudiantes. Supongamos que sabemos que se trata de una variable aleatoria normal de media 172 cm y desviación típica 11 cm y que hemos tomado una muestra de trescientos estudiantes al azar. Entonces podemos contestar preguntas del tipo siguiente:

a) ¿Cuál es la probabilidad de que la media sea menor que 170 cm?

   La distribución de la media muestral es normal de media 172 cm y desviación típica:

$$\frac{11}{\sqrt{300}} = 0.635$$

   Tipificamos la variable para obtener una normal $(0,1). Debemos calcular:

$$P(\overline{X} < 170) = P\left(\frac{\overline{X} - 172}{0.635} < \frac{-2}{0.635}\right) = P(Z < -3.149) = 0.0008$$

   ya que Z es una variable aleatoria normal $(0,1)$.