

Teorema del límite central

2023-11-14

ÍNDICE

La distribución de la media muestral	2
1.1 Distribución de la media muestral para variables normales	2
1.1.1 Caso de desviación típica poblacional conocida	2
1.2 Caso de desviación típica poblacional desconocida. La t de Student	4

1 La distribución de la media muestral

Supongamos que queremos estudiar la media de la altura de unos estudiantes. 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, \dots, x_n de una variable aleatoria normal. La media se define como:

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

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 \bar{X} .

Deberemos distinguir dos casos: cuando la desviación 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 observaciones de la muestra que tenemos debe ser un valor cercano a 172. También parece razonable pensar que observaciones mayores que la media 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 $\bar{x}_1, \dots, \bar{x}_{10}$. 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 \bar{X} y obtenemos que:

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

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 $\frac{11}{\sqrt{10}} = 3.48$ *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:

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

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

- b) ¿Cuál es la probabilidad de que la distancia entre la media muestral (de esta muestra de trescientos estudiantes) y la media poblacional, 172 *cm* sea menor que 1 *cm*?

Por un razonamiento parecido (si la distancia entre dos números a y b ha de ser menor que k , se debe cumplir $|a - b| < k$):

$$P(|\bar{X} - \mu| < 1) = P(-1 < \bar{X} - \mu < 1) = P\left(-\frac{1}{0.635} < \frac{\bar{X} - \mu}{0.635} < \frac{1}{0.635}\right) = P(-1.57 < Z < 1.57)$$

donde Z es una variable aleatoria normal $(0, 1)$. Si buscamos en las tablas de la ley normal $(0, 1)$, vemos que esta probabilidad es igual a 0.8836.

Tenemos así una probabilidad del 0.8836 de obtener un valor para la media muestral que difiera en menos de 1 *cm* del valor real de la media cuando tomamos una muestra de trescientos individuos.

Hay que observar que en ninguna parte hemos utilizado el hecho de que la media fuese exactamente 172 *cm*. Es decir, si sabemos que la variable “altura” sigue una normal con una desviación típica de 11 *cm* y tomamos una muestra de 300 estudiantes, sabemos que la diferencia entre su media y la media poblacional μ (que quizás no conozcamos) será menor de 1 *cm* con una probabilidad del 0.8836.

- c) Consideremos ahora el problema inverso. Supongamos que desconocemos la media μ de la altura de los trescientos estudiantes y queremos estudiar una muestra de manera que la diferencia entre la media de la muestra y la de la población μ sea menor que 1 *cm* con una probabilidad del 0.95. ¿De qué medida tiene que ser nuestra muestra?

Sabemos que la variable estadística tipificada:

$$\frac{\bar{X} - \mu}{\frac{11}{\sqrt{n}}}$$

se distribuye como una normal $(0, 1)$. Por otra parte, si observamos las tablas, nos damos cuenta de que si Z es una normal $(0, 1)$:

$$P(-1.96 < Z < 1.96) = 0.95$$

Por tanto:

$$0.95 = P\left(-1.96 < \frac{\bar{X} - \mu}{\frac{11}{\sqrt{n}}} < 1.96\right) = P\left(-1.96 \frac{11}{\sqrt{n}} < \bar{X} - \mu < 1.96 \frac{11}{\sqrt{n}}\right)$$

Y si imponemos que la diferencia $\bar{X} - \mu$ debe ser menor que 1 *cm*, obtenemos:

$$1.96 \frac{11}{\sqrt{n}} < 1$$

Por tanto, $\sqrt{n} > 11 \cdot 1.96$, y así: $n > (11 \cdot 1.96)^2 = 464.8$. Entonces, si tomamos 465 individuos para llevar a cabo el estudio, sabemos que la diferencia entre la media muestral que obtendremos y la media real será menor de 1 *cm*, con una probabilidad del 0.95. Cuanto mayor sea el tamaño de la muestra, menor será la diferencia entre la media muestral y la poblacional.

Si se multiplican el numerador y el denominador por n , podemos escribir el resultado que hemos visto en este apartado de otra manera.

Si la variable que estudiamos sigue una distribución normal con media μ y desviación típica σ , entonces:

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}$$

sigue una distribución normal estándar.

1.2 Caso de desviación típica poblacional desconocida. La t de Student

En los ejemplos estudiados anteriormente necesitábamos dos cosas:

- Que la variable que se estudiaba fuese normal.
- Que el valor de la desviación típica de la variable fuese conocido.

Estos dos hechos se conocen gracias a estudios previos. A menudo este estudio no se lleva a cabo, pero podemos suponer que la variable es normal. En este caso deberemos hacer una estimación de la desviación típica con la llamada **desviación típica muestral**:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Hay que observar que en el caso de la desviación típica muestral se divide por $n - 1$, no por n .

De manera que en los cálculos del apartado anterior reemplazaremos la σ por la s . Entonces la distribución muestral de la media ya no es una distribución normal, como sucedía cuando en lugar de s conocíamos el auténtico valor σ de la desviación.

Varios estudios realizados por W. S. Gosset, al final del siglo XIX, demostraron que en este caso se obtiene una distribución diferente a la normal, aunque para tamaños lo bastante grandes se parecen bastante. Esta nueva distribución se conoce con el nombre de *t de Student* con $n - 1$ grados de libertad. Esto significa que por cada medida de la muestra, n , en realidad tenemos una distribución diferente.

La **distribución *t de Student* con n grados de libertad**, que denotaremos por t_n es muy parecida a la distribución normal $(0, 1)$: es simétrica alrededor del cero, pero su desviación típica es un poco mayor que la de la normal $(0, 1)$, es decir, los valores que toma esta variable están un poco más dispersos. No obstante, cuanto mayor es el número de grados de libertad, n , más se aproxima la distribución t_n de Student a la distribución normal $(0, 1)$. Consideraremos que podemos aproximar la t_n por una normal estándar para $n > 100$.

Observamos que cuando conocemos el valor auténtico de σ , la variable \bar{X} sigue siempre una distribución normal, pero su varianza depende de n .

El gráfico siguiente representa las funciones de densidad de la *t de Student* para diferentes valores de n y con una línea más gruesa, la densidad de una distribución normal $(0, 1)$.

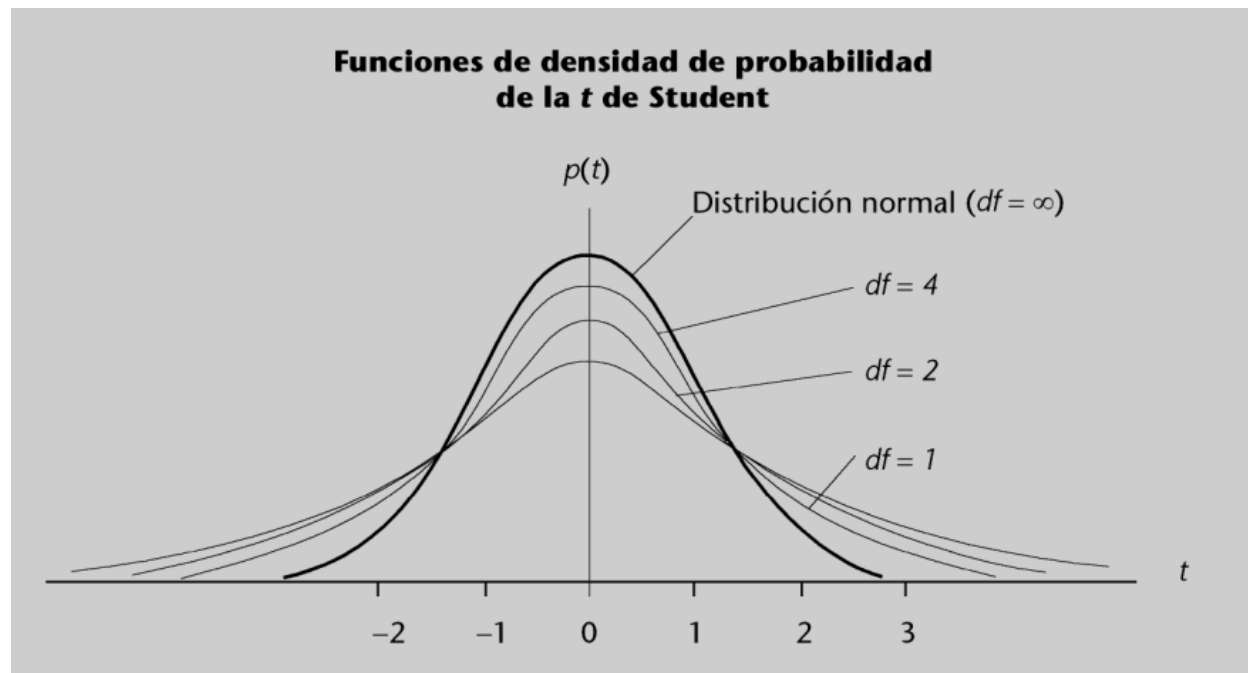


Figure 1: Funciones de densidad de probabilidad de la *t de Student*