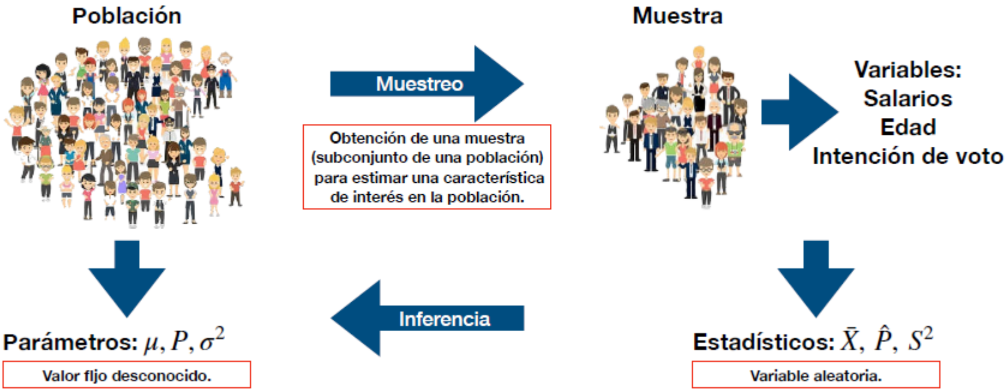
Clase 11: Distribución de la media muestral para poblaciones normales. Teorema del Límite Central (TLC).

Universidad Nacional de Colombia - Sede Medellín

Parámetros y estadísticos



En el proceso de identificar y explicar las características esenciales que permitan describir el comportamiento de un fenómeno, nuestro objetivo es el de establecer de manera aproximada dicho comportamiento usando parte de toda la información acerca del fenómeno.

Algunos ejemplos donde nos interesan parámetros son:

- Estimar la proporción hogares (P) que tienen perros o gatos.
- Estimar el tiempo promedio (μ) que una persona permanece en fila.
- Estimar la variablidad (σ^2) en los diámetros de tuercas de cierto tipo.

Muestra Aleatoria

Cada objeto o individuo seleccionado aporta información acerca de la característica que se quiere medir, la cual varía de individuo a individuo. Así, una muestra no es más que una colección de variables aleatorias.

Definición

Se dice que las variables aleatorias (v.as) $X_1, \ldots, X_n$ forman una muestra aleatoria (ma) de tamaño n, si:

- Las $X_1, \ldots, X_n$ son v.as independientes.
- Todas las $X_1, \ldots, X_n$ tienen la misma distribución de probabilidad.

Es decir, una muestra se puede llamar muestra aleatoria (ma) si son independientes e identicamente distribuidas (iid) de modo que:

- $f(X_1, \cdots, X_n) = \prod_{i=1}^{n} f(x_i)$.

Estadístico muestral

Un estadístico es cualquier cantidad cuyo valor se puede calcular a partir de los datos muestrales, es decir, es una función que depende sólo de la muestra y no depende de parámetros desconocidos. Los estadísticos son también v.as.

- Una buena estimación para μ es $\bar{X} = \frac{1}{n} \sum X_i$.
- Una buena aproximación para σ^2 es $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$.
- Una buena aproximación para P es $\hat{p} = \frac{X}{n}$.

Media y varianza de la media muestral $\bar{X}$

Sea $X_1, \cdots, X_n$ una muestra aleatoria (ma) de una distribución con media μ y varianza σ^2.
La media muestral $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, es un estadístico, por lo tanto tiene media y varianza, así:

$$E\left[\bar{X}\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{n\mu}{n} = \mu\,. \tag{1}$$

$$Var\left[\bar{X}\right] = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var\left[X_i\right] = \frac{\sigma^2}{n}\,. \tag{2}$$

Así, la distribución muestral de $\bar{X}$, tiene media μ y varianza σ^2/n.

Media y varianza del total T

Sea $X_1, \cdots, X_n$ una muestra aleatoria (ma) de una distribución con media μ y varianza σ^2. El estadístico suma Total $T = \sum_{i=1}^{n} X_i$, es un estadístico, por lo tanto tiene media y varianza, así:

$$E[T] = n\mu. \tag{3}$$

$$Var[T] = n\sigma^2. \tag{4}$$

Distribución de $\bar{X}$ y de T

Proposición

Sea $X_1, \cdots, X_n$ una muestra aleatoria de una $\mathbf{N}(\mu, \sigma^2)$, entonces:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{y} \quad T \sim N(n\mu, n\sigma^2).$$

Así,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \quad \text{y} \quad Z = \frac{T - n\mu}{\sqrt{n}\sigma} \sim N(0, 1).$$

Ejemplo de la media muestral

Suponga que el QI de los estudiantes de primer año de Matemáticas es una v.a. normalmente distribuída con media 120 y varianza 100, se seleccionan 25 estudiantes aleatoriamente.

1. ¿Cuál es la probabilidad de que el QI promedio de esta muestra sea superior a 122?
2. ¿Cuál debe ser el tamaño de la muestra que garantice que el 5% de las veces el QI promedio sea superior a 122?

Solución

Sea $X_1, \ldots, X_{25}$ una m.a. que representa los QI de los 25 estudiantes de primer año.
$X_i \sim N(120\,,\,100); \quad i = 1, 2, \ldots, 25$

a) $P(\bar{X} > 122) = P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} > \frac{122-120}{\frac{10}{\sqrt{25}}}\right) = P(Z > 1) = 0.1587$.

b) $P(\bar{X} > 122) = 0.05 \Leftrightarrow P\left(Z > \frac{122-120}{\frac{10}{\sqrt{n}}}\right) = 0.05 \Leftrightarrow P\left(Z > \frac{\sqrt{n}}{5}\right) = 0.05.$

Al buscar en R el cuantil q de manera que a la derecha de él se tenga un área de 0.05, se obtiene que $q = 1.645$. Así que debemos igualar q con $\sqrt{n}/5$, es decir

$$1.645 = \frac{\sqrt{n}}{5}$$
$$n = 67.650625$$

Teorema del límite central (TLC)

Suponga que $X_1, \ldots, X_n$, es una m.a. de una **población cualquiera** con media μ y varianza σ^2. Sea $\bar{X}$ la media muestral (la cual depende de n), entonces cuando $n \to +\infty$, la distribución muestral de $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ es aproximadamente normal estándar.

En otras palabras:

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \overbrace{n \to \infty}^{aprox} N(0, 1).$$

Y para el estadístico Total (T), escribimos:

$$\frac{T - n\mu}{\sqrt{n}\sigma} \overbrace{n \to \infty}^{aprox} N(0, 1).$$

Entre mayor sea n, mejor es la aproximación.

Teorema del límite central (TLC)

Así,

$$P(\bar{X} \leq a) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{a - \mu}{\frac{\sigma}{\sqrt{n}}}\right) \approx P\left(Z \leq \frac{a - \mu}{\frac{\sigma}{\sqrt{n}}}\right).$$

Si σ^2 es desconocida y n grande, se puede reemplazar σ^2 por S^2 y la expresión anterior queda así:

$$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \quad \overset{aprox}{\widetilde{n \to \infty}} \quad N(0\,,\,1)\,.$$

Ejemplo del TLC

Parmenio vende ensaladas de frutas en tres tamaños diferentes: pequeño, mediano y grande, cuyos precios son \$1800, \$2500 y \$4000, respectivamente. De todas las ensaladas vendidas el 30% son pequeñas, 20% medianas y 50% grandes. Durante un día dado se reporta la venta de 225 ensaladas. Calcule la probabilidad de que el ingreso obtenido supere los \$720000.

Solución

Sea X : precio de la ensalada. $\mathcal{A}_X = \{1800, 2500, 4000\}$

x	1800	2500	4000
$p(x)$	0.3	0.2	0.5

Sea $X_1, \ldots, X_{225}$ una m.a. que representa los precios de las 225 ensaladas. Como n es grande y es una m.a, podemos utilizar el T.L.C.

Se pide calcular $P\left(T = \sum_{i=1}^{225} X_i > 720000\right)$.

Calculemos primero $E(X)$ así:

$$E[X_i] = 1800 \times 0.3 + 2500 \times 0.2 + 4000 \times 0.5 = 3040.$$

Ahora calculemos $E(X^2)$ para poder obtener la varianza luego.

$$E\left[X_i^2\right] = 1800^2 \times 0.3 + 2500^2 \times 0.2 + 4000^2 \times 0.5 = 10222000$$

La varianza $Var(X)$ se puede obtener como:

$$Var[X_i] = 10222000 - 3040^2 = 980400 .$$

Ahora si podemos aplicar el TLC para responder la pregunta.

$$\begin{aligned} P(T > 720000) &= P\left(\frac{T - n\mu}{\sqrt{n\sigma^2}} > \frac{720000 - (225)(3040)}{\sqrt{(225)(980400)}}\right) \\ &= P(Z > 2.42) = 0.00776 \end{aligned}$$

Ejemplo del TLC

La resistencia a la compresión del concreto es una v.a con una resistencia media de 2500 psi y una desviación estándar de 50 psi. Si se examinan 36 especimenes de concreto, ¿Cuál es la probabilidad de que la resistencia promedio en esta muestra esté entre 2497 psi y 2505 psi?. Asuma independencia entre las resistencias.

Solución

Sea $X_1, \cdots, X_{36}$ una m.a que representa las resistencias a la compresión de los 36 especímenes. Se sabe que

$$E[X_i] = 2500 \quad \text{y} \quad Var[X_i] = 2500.$$

Como n es grande, es una m.a y no nos dicen nada acerca de la distribución de la población, pero de acuerdo al T.L.C.

$$\begin{aligned}
P(2497 < \bar{X} < 2505) &= P\left(\frac{2497 - 2500}{\frac{50}{\sqrt{36}}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{2505 - 2500}{\frac{50}{\sqrt{36}}}\right) \\
&\approx P(-0.36 < Z < 0.6) \\
&= \Phi(0.6) - \Phi(-0.36) \\
&= 0.366323
\end{aligned}$$

Ejercicio

Se selecciona una m.a. de tamaño 49 de una población con f.m.p. dada por:

x	2	3	7
$p(x)$	$1/3$	$1/3$	$1/3$

¿Cúal es la probabilidad de que $\bar{X}$ esté entre 3.9 y 4.1?