Estimación

Dr. Delfino Vargas Chanes

Facultad de Economía
Universidad Nacional Autónoma de México

24 de abril de 2024

Índice

Estimador

Definición 1.1

Un estimador es una regla, usualmente expresada como formula, que nos dice como calcular un valor desconocido de la población, utilizando medidas de la muestra.

- Al estimador se le llama también estadístico.
- La media muestral es un estimador de la media poblacional.
- La estimación puntual, consiste en encontrar un único punto, que sea la mejor aproximación, o mejor *estimador*, de un parámetro poblacional desconocido.
- La estimación puntual es como disparar a una diana. La pistola es el estimador, el estimando es nuestro disparo, y el blanco es el parámetro poblacional.
- ¿Qué es que sea mejor?

Sesgo

Definición 2.1

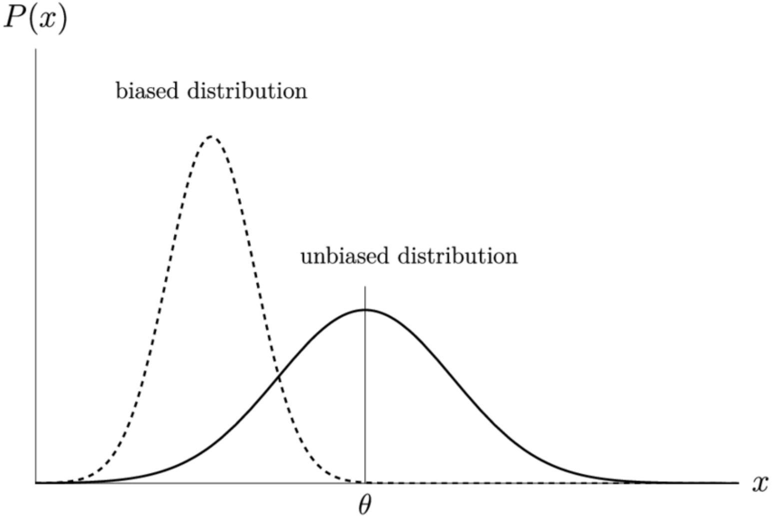
Sea $\hat{\theta}$ el estimador puntual de un parámetro poblacional θ. Decimos que $\hat{\theta}$ es un estimador **insesgado** si $\mathbb{E}[\hat{\theta}] = \theta$. Si ocurre que $\mathbb{E}[\hat{\theta}] \neq \theta$, decimos que $\hat{\theta}$ es un estimador **sesgado**.

Definición 2.2

Denotamos al **sesgo** del estimador puntual $\hat{\theta}$ como $\mathrm{B}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$

- No sabemos si alguien que le da al blanco una vez es un buen tirador.
- Es por ello que hay que ver más de un tiro, hay que construir una distribución de nuestros estimandos y ver que tan cerca está el valor esperado del estimador al verdadero parámetro.

Distribución Sesgada vs Insesgada



Error Cuadrático Medio

Definición 2.3

El error cuadrático medio de un estimador puntual $\hat{\theta}$ se denota como

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$$

Que es una función de su varianza y su esperanza. Lo podemos denotar también como

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + [\mathrm{B}(\hat{\theta})]^2$$

- Se suele buscar estimadores insesgados y con baja varianza.
- Un ejemplo de un estimador insesgado es la media poblacional $\overline{Y}$ de la media poblacional μ, ya que $\mathbb{E}[\overline{Y}] = \mu$

Error de Estimación y Estándar

Definición 2.4

El error de estimación de un parámetro y su estimador es su distancia y se define como

$$\varepsilon = \left|\hat{\theta} - \theta\right|$$

Definición 2.5

Definimos a la desviación estándar, o error estándar, de la distribución muestral del estimador $\hat{\theta}$ se define como

$$\sigma_{\hat{\theta}} = \sqrt{\sigma^2_{\hat{\theta}}}$$

donde $\sigma^2_{\hat{\theta}}$ es la varianza de la distribución muestral de $\hat{\theta}$

Intervalos de Confianza

- El material que vamos a ver es frecuentista
- A diferencia del estimador puntual, un estimador por intervalo busca colocar el parámetro θ dentro de un intervalo.
- Estos se conocen como intervalos de confianza.
- Que tan ancho es un intervalo de confianza es algo aleatorio, y no podemos estar $100\,\%$ seguros que θ caiga dentro de él.
- El intervalo de confianza tiene límite inferior y superior.
- También tiene nivel de significancia, que es la probabilidad que un intervalo de confianza contenga al prámetro θ

Ejemplos

- Suponga usted $x_1, ..., x_n \sim \mathcal{N}(\mu, \sigma^2)$, tal que desconocemos a μ y σ.
- Un estimador es la media muestral $\overline{x}$, o el estimador $\overline{x} - 5$
- Algo que no es un estimador es $\overline{x} - \mu$, ya que no conocemos μ
- Si s^2 es la varianza muestral, $\left[\overline{x} - \frac{2s}{\sqrt{n}},\ \overline{x} + z \cdot \frac{2s}{\sqrt{n}}\right] = \overline{x} \pm \frac{2s}{\sqrt{n}}$ es un estimador por intervalo, ya que s^2 viene de datos.
- Si no conocemos a σ^2, $\left[\overline{x} - \frac{2\sigma}{\sqrt{n}},\ \overline{x} + z \cdot \frac{2s}{\sqrt{n}}\right] = \overline{x} \pm \frac{2\sigma}{\sqrt{n}}$ no es un estimador por intervalo
- Los intervalos son aleatorios
- Explicaremos en el siguiente tema como se relaciona este concepto con la escuela frecientista y la noción de hipótesis.

Intervalo de Confianza

Definición 3.1

Sea θ un parámetro a estimar desconocido. Sean $\hat{\theta}_{\text{inf}}$ y $\hat{\theta}_{\text{sup}}$ los límites inferiores y superiores de un intervalo $[\hat{\theta}_{\text{inf}}, \hat{\theta}_{\text{sup}}]$, llamado intervalo de confianza, con nivel de confianza $(1-\alpha)$ si cumple la propiedad

$$P(\hat{\theta}_{\text{inf}} \leqslant \theta \leqslant \hat{\theta}_{\text{sup}}) = 1-\alpha$$

Datos Normales

- Suponga usted $x_1, \ldots, x_n \sim \mathcal{N}(\mu, \sigma^2)$, y que se conoce σ, pero no μ
- Sabemos que la media estandarizada es una VA con distribución normal estándar

$$z = \frac{x - \mu}{\sigma} = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

- Usando la Definición de intervalo de confianza

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

- Sustituyendo a Z se tiene que

$$P\left(-z_{\alpha/2} < \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha$$

- Donde $z_{\alpha/2}$ son valores criticos, donde $P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$

Datos Normales

- **Tarea Moral:** Vea usted que lo anterior se puede expresar como

$$P\left(\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

- Por lo que el intervalo de confianza con nivel de confianza $(1 - \alpha)$ es

$$\left[\overline{x} - z \cdot \frac{\sigma}{\sqrt{n}},\ \overline{x} + z \cdot \frac{\sigma}{\sqrt{n}}\right]$$

- Por ejemplo, con un nivel de significancia $\alpha = 95\,\% \Rightarrow z = 1.96$

$$\left[\overline{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}},\ \overline{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$$

- Pueden jugar con intervalos de confianza en https://mathlets.org/mathlets/confidence-intervals/

Ejercicio 1

- Se quiere alquilar un apartamento de una habitación sin amueblar en la CDMX, el próximo año, ya que está usted harto de su casere o familia. El alquiler mensual promedio para una muestra aleatoria de 60 apartamentos anunciados en Inmuebles24 es de \$1000. Suponga una desviación estándar poblacional de \$200. Construya un intervalo de confianza del 95 %.
- Recuerde que el intervalo se calcula

$$\left[\overline{x} - z \cdot \frac{\sigma}{\sqrt{n}},\ \overline{x} + z \cdot \frac{\sigma}{\sqrt{n}}\right]$$

- ¿Para que población de departamentos puede usted inferir dada la muestra que se presenta?
- ¿Cuántos apartamentos necesitaría usted muestrar para hacer un intervalo de la media $\pm$ 50\$ con 90 % de confianza?

Solución 1

- El intervalo es $(\$949.39, \$1050.61)$, es decir estamos 95 % seguros que ese intervalo contiene a la media poblacional.

$$\left[1000 - 1.96 \cdot \frac{200}{\sqrt{60}},\ \ 1000 + 1.96 \cdot \frac{200}{\sqrt{60}}\right]$$

- No se puede inferir a todos los departamentos de la CDMX, únicamente aquellos de la muestra presente.
- En ese caso hay que estimar $\pm Z \cdot \frac{\sigma}{\sqrt{n}}$

$$50 = 1.645 \cdot \frac{200}{\sqrt{n}} = \frac{329}{\sqrt{n}}$$

Lo que nos da $n = 43.3$, y hay que redondear a 44

Ejercicio 2

- Matías mantuvo registros cuidadosos de la eficiencia de gasolina de su Chevy. Después de las primeras 100 veces que llenó el tanque, encontró que la media era de 23.4 km por litro de gasolina con una desviación estándar poblacional de 0.9 km por litro de gasolina. Calcula el intervalo de confianza del 95 por ciento para sus km por litro de gasolina.

Solución 2

- El intervalo es $(23.22, 23.56)$, es decir estamos 95 % seguros que ese intervalo contiene a la media poblacional.

$$\left[23.4 - 1.96 \cdot \frac{200}{\sqrt{100}},\;\; 23.4 + 1.96 \cdot \frac{0.9}{\sqrt{100}}\right]$$

Ejercicio 3

- Supongamos que contra un cierto oponente, el número de puntos que anota el equipo de baloncesto de la UNAM sigue una distribución normal con media desconocida θ y varianza 25. Supongamos que durante los últimos 10 juegos entre los dos equipos, la UNAM anotó los siguientes puntos: $59, 62, 59, 74, 70, 61, 62, 66, 62, 75$.
- Compute un intervalo de confianza del 95 % para θ
- Supongamos que tenemos 40 intervalos de confianza con un nivel de confianza del 95 %. ¿Aproximadamente cuántos de ellos esperaríamos que estuvieran equivocados? Es decir, ¿cuántos no contendrían realmente el parámetro que se está estimando?

Solución 3

- El intervalo es $(61.901, 68.099)$, es decir estamos 95 % seguros que ese intervalo contiene a la media poblacional.

$$\left[65 - 1.96 \cdot \frac{5}{\sqrt{10}},\ 65 + 1.96 \cdot \frac{5}{\sqrt{10}}\right]$$

- Un nivel de confianza del 95 % significa que aproximadamente el 5 %, o sea, $\frac{1}{20}$ estará equivocado. Esperaríamos que alrededor de 2 estén equivocados.

Bibliografía

Bernoulli, J. (1713). *Ars coniectandi*. Impensis Thurnisiorum, fratrum.

Bhattacharyya, G. K., y Johnson, R. A. (1977). *Statistical concepts and methods*. Wiley.

Bickel, P. J., y Doksum, K. A. (2015). *Mathematical statistics: basic ideas and selected topics, volumes i-ii package*. CRC Press.

Fair, R. C. (1978). A theory of extramarital affairs. *Journal of political economy*, *86*(1), 45–61.

Groeneveld, R. A. (1998). A class of quantile measures for kurtosis. *The American Statistician*, *52*(4), 325–329.

Wackerly, D., Mendenhall, W., y Scheaffer, R. L. (2014). *Mathematical statistics with applications*. Cengage Learning.

¡Gracias por su atención!