Probabilidad y estadística

MAT 041, Primer semestre

Francisco Cuevas Pacheco

5 de diciembre de 2022

Recuerdo estimación puntual

Recordemos el problema de inferencia:

Observación

Sea $X_1, \ldots, X_N$ una muestra aleatoria proveniente de una distribución $f(x; \theta)$. Deseamos encontrar $\hat{\theta}$, que muestre alguna característica de los datos.

Recuerdo estimación puntual

Recordemos el problema de inferencia:

Observación

Sea $X_1, \ldots, X_N$ una muestra aleatoria proveniente de una distribución $f(x;\theta)$. Deseamos encontrar $\hat{\theta}$, que muestre alguna característica de los datos.

Problema

El problema con la estimación puntual es que, en general, $P(\hat{\theta} = \theta) = 0$.

Es por esto que dada una muestra aleatoria $\underline{X} = (X_1, X_2, \ldots, X_n)$ y una probabilidad dada $(1 - \alpha)$, resulta más adecuado encontrar estadísticas muestrales $L(\underline{X})$ y $U(\underline{X})$ tales que uno pueda establecer que

$$P(L(\underline{X}) \leq \theta \leq U(\underline{X})) = 1 - \alpha$$

Estimación Intervalar

Es por esto que dada una muestra aleatoria $\underline{X} = (X_1, X_2, \ldots, X_n)$ y una probabilidad dada $(1-\alpha)$, resulta más adecuado encontrar estadísticas muestrales $L(\underline{X})$ y $U(\underline{X})$ tales que uno pueda establecer que

$$P(L(\underline{X}) \leq \theta \leq U(\underline{X})) = 1 - \alpha$$

Esto quiere decir que $IC(\theta) = [L(\underline{X}) \ ; \ U(\underline{X})]$ **captura** al verdadero parámetro con probabilidad $1 - \alpha$.

Lectura equivocada

El verdadero parámetro se encuentra en $L(\underline{X}) \leq \theta \leq U(\underline{X})$ con un α porciento de confianza.

Construción del intervalo

Antertiormente hemos visto que un estadístico $T(\underline{X})$ nos permitía tener una estimación del parámetro buscado, además este estadístico al ser ua función de $\underline{X}$ corresponde a ser una variable aleatoria y por lo tanto tiene una función de densidad asociada. Usaremos entonces la dfistribución de $T(\underline{X})$ para construir nuestro intervalo.

El punto de partida para crear intervalos de confianza es mediante una **Cantidad Pivotal**

Definición

*Una **cantidad pivotal** es una estadística o expresión aleatoria $Q(\underline{X};\theta)$ que cumple tres requisitos,*

1. *Q depende de la muestra aleatoria $\underline{X} = (X_1, X_2, \ldots, X_n)$.*
2. *Q depende del parámetro θ.*
3. *La distribución de Q no depende de θ.*

- La idea de los requisitos 1 y 2 es relacionar el parametro θ que se quiere estimar con la información acerca de ese parámetro contenida en la muestra.
- El requisito 3 permite *pivotear*, esto es despejar θ en el cálculo de probabilidades.

Ejemplo

Considere la muestra aleatoria $X_1, X_2, \ldots, X_n$ de una población con distribución normal de media μ y varianza σ^2. Entonces, se sabe que $\bar{X}$ es un buen estimador puntual de μ pero no es una cantidad pivotal para estimar μ mediante un intervalo de confianza porque no se cumplen los requisitos 2. y 3. De hecho, su distribución depende de μ ya que $\bar{X} \sim N(\mu; \sigma^2/n)$. Sin embargo, estandarizando se obtiene una cantidad pivotal para μ dada por,

$$Q(\underline{X}; \mu) = Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0; 1)$$

Otro ejemplo de cantidad pivotal usada para estimar con intervalo de confianza la varianza σ^2 de una población normal es

$$Q(\underline{X}; \sigma^2) = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1},$$

donde $S^2 = \sum_{i=1}^{n}(X_i - \hat{X})^2/(n-1)$ es el estimador insesgado de σ^2.

El método usado para obtener el intervalo de estimación $IC(\theta)$ con confianza $(1-\alpha)100\,\%$ se llama **método de la cantidad pivotal** y está basado en el hecho que si $Q(\underline{X};\theta)$ es una cantidad pivotal para θ con distribución muestral conocida y si a, b son números reales tales que

$$P(a \leq Q(\underline{X};\theta) \leq b) = 1-\alpha$$

Entonces, *pivoteando* o despejando θ de ambas desigualdades se obtiene,

$$P(L(\underline{X}) \leq \theta \leq U(\underline{X})) = 1-\alpha$$

Ilustraremos este método en las siguientes secciones.

Ejemplo: Construcción de un intervalo de confianza

Intervalo de confianza para la media poblacional en el caso normal:

Sea $X_1, \ldots, X_n$ una m.a. desde una distribución $N(\mu, \sigma^2)$. Como $\overline{X}$ es el mejor estimador de μ, entonces si se conoce σ^2, se tiene que

$$Z = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \sim N(0, 1)$$

Por lo tanto Z es una cantidad pivotal.

Intervalo de confianza para la media poblacional en el caso normal:

Sea $X_1, \ldots, X_n$ una m.a. desde una distribución $N(\mu, \sigma^2)$. Como $\overline{X}$ es el mejor estimador de μ, entonces si se conoce σ^2, se tiene que

$$Z = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \sim N(0,1)$$

Por lo tanto Z es una cantidad pivotal.

Luego se deben buscar los valores apropiados de q_1 y q_2 tales que

$$P\left[q_1 \leq \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \leq q_2\right] = 1 - \alpha = \gamma$$

Despejando

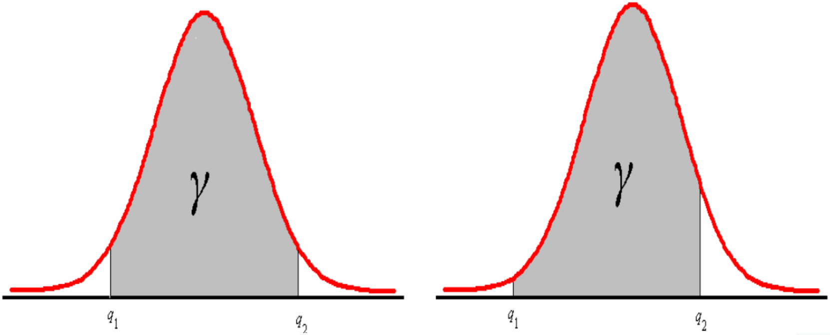
$$\mathbb{P}\left[q_1 \leq \frac{(\overline{X}-\mu)\sqrt{n}}{\sigma} \leq q_2\right] = \gamma$$

$$\mathbb{P}\left[q_1\sigma \leq (\overline{X}-\mu)\sqrt{n} \leq q_2\sigma\right] = \gamma$$

$$\mathbb{P}\left[\frac{q_1\sigma}{\sqrt{n}} \leq (\overline{X}-\mu) \leq \frac{q_2\sigma}{\sqrt{n}}\right] = \gamma$$

$$\mathbb{P}\left[\frac{q_1\sigma}{\sqrt{n}} - \overline{X} \leq -\mu \leq \frac{q_2\sigma}{\sqrt{n}} - \overline{X}\right] = \gamma$$

$$\mathbb{P}\left[\overline{X} + \frac{q_2\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + \frac{q_1\sigma}{\sqrt{n}}\right] = \gamma$$



Como se observa en la figura la elección q_1 y q_2 no es única, pero tiene sentido de que busquemos el intervalo que minimice la distancia entre q_1 y q_2, esto se logra cuando se produce igualdad de probabilidades en las colas.

Logitud minimal

Deseamos un intervalo de confianza que sea de longitud mínima. Para esto, planteamos el siguiente problema

$$\begin{gathered} \min q_2 - q_1 \\ s.a \\ F_Z(q_2) - F_Z(q_1) = \gamma \end{gathered}$$

Calculamos el Lagrangiano

$$\mathcal{L}(q_1, q_2, \lambda) = q_2 - q_1 - \lambda(F_Z(q_2) - F_Z(q_1) - \gamma)$$

Y calculamos su gradiente

$$\begin{cases} \frac{\partial}{\partial q_1}\mathcal{L}(q_1, q_2, \lambda) = -1 + \lambda f(q_1) \\ \frac{\partial}{\partial q_2}\mathcal{L}(q_1, q_2, \lambda) = 1 - \lambda f(q_2) \\ \frac{\partial}{\partial \lambda}\mathcal{L}(q_1, q_2, \lambda) = F_Z(q_2) - F_Z(q_1) - \gamma \end{cases}$$

Logitud minimal

Calculamos el Lagrangiano

$$\mathcal{L}(q_1, q_2, \lambda) = q_2 - q_1 - \lambda(F_Z(q_2) - F_Z(q_1) - \gamma)$$

Resolviendo $\nabla\mathcal{L}(q_1, q_2, \lambda) = 0$ se obtiene que

$$\lambda = \tfrac{1}{f(q_2)}$$
$$f(q_1) = f(q_2)$$
$$F_Z(q_2) - F_Z(q_1) = \gamma$$

Logitud minimal

Calculamos el Lagrangiano

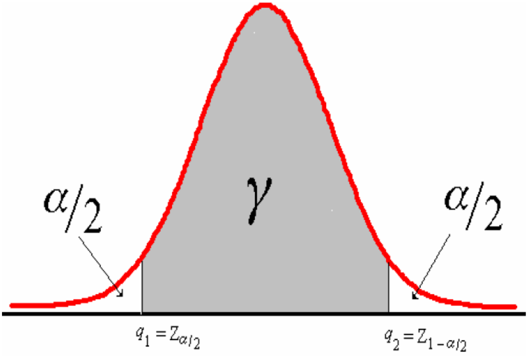
$$\mathcal{L}(q_1, q_2, \lambda) = q_2 - q_1 - \lambda(F_Z(q_2) - F_Z(q_1) - \gamma)$$

Resolviendo $\nabla\mathcal{L}(q_1, q_2, \lambda) = 0$ se obtiene que

$$\lambda = \tfrac{1}{f(q_2)}$$
$$f(q_1) = f(q_2)$$
$$F_Z(q_2) - F_Z(q_1) = \gamma$$

De donde se concluye que $q_1 = -q_2$

Para el caso normal se tiene que



esto es

$$q_1 = \phi^{-1}(\alpha/2) = Z_{\alpha/2}, \qquad q_2 = \phi^{-1}(1-\alpha/2) = Z_{1-\alpha/2}$$

Además, notamos que $-Z_{\alpha/2} = Z_{1-\alpha/2}$.

De la última idea podemos extraer que:

$$P\left[Z_{\alpha/2} \leq \frac{(\overline{X}-\mu)\sqrt{n}}{\sigma} \leq Z_{1-\alpha/2}\right] = 1-\alpha$$

$$P\left[Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \overline{X} \leq -\mu \leq Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} - \overline{X}\right] = 1-\alpha$$

$$P\left[\overline{X} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

Como $Z_{\alpha/2} = -Z_{1-\alpha/2}$ entonces:

$$P\left[\overline{X} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

De la última idea podemos extraer que:

$$P\left[Z_{\alpha/2} \leq \frac{(\overline{X}-\mu)\sqrt{n}}{\sigma} \leq Z_{1-\alpha/2}\right] = 1-\alpha$$

$$P\left[Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \overline{X} \leq -\mu \leq Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} - \overline{X}\right] = 1-\alpha$$

$$P\left[\overline{X} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

Como $Z_{\alpha/2} = -Z_{1-\alpha/2}$ entonces:

$$P\left[\overline{X} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

Con lo anterior se concluye que el intervalo de confianza del $(1-\alpha)\,\%$ para la media poblacional es:

$$IC_{1-\alpha}(\mu) := \left[\overline{X} \pm Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

1. Límites Aleatorios en función de $\hat{\theta}$ y α : Los límites aleatorios $L(\underline{X})$ y $U(\underline{X})$ dependen de la muestra a través del estimador puntual $\hat{\theta}$ y dependen del nivel de confianza $(1-\alpha)$.

Comentarios respecto a los Intervalos de Confianza para una media μ

1. Límites Aleatorios en función de $\hat{\theta}$ y α : Los límites aleatorios $L(\underline{X})$ y $U(\underline{X})$ dependen de la muestra a través del estimador puntual $\hat{\theta}$ y dependen del nivel de confianza $(1-\alpha)$.

2. Calidad del Intervalo de Confianza:. Una manera de evaluar la calidad de un intervalo de confianza o un criterio para seleccionar un intervalo de estimación es que su *longitud* sea mínima para un tamaño de muestra n y para una confianza $(1-\alpha)100\,\%$ dados.

3. Error Estándar y Margen de Error. En el caso de la media μ de una población normal se sabe que $\hat{\mu} = \bar{X}$ es un estimador insesgado porque $E(\bar{X}) = \mu$ y es consistente porque el error cuadrático medio $ECM(\hat{\mu}) = Var(\bar{X}) = \sigma^2/n \to 0$ cuando $n \to \infty$. La raíz cuadrada de este error es $\sigma/\sqrt{n}$ y se llama *error estándar* de estimación de la media.
   El *margen de error* o simplemente *error* de estimación de μ corresponde al producto entre el error estándar y el cuantil de la normal y está dado por,
   $$\varepsilon = \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}$$
   Así, el intervalo de confianza $IC(\mu) = \bar{X} \pm \varepsilon$ da cuenta de la incertidumbre que se tiene con el estimador puntual de μ.

4. Relación entre el tamaño de muestra, la precisión y la confianza. El margen de error cometido al estimar la media μ de una población normal dado por $\varepsilon = (\sigma/\sqrt{n})Z_{1-\alpha/2}$ es una expresión que muestra la relación (compromiso) que existe entre la precisión de la estimación ε, el tamaño de la muestra n y la confianza $(1-\alpha)100\,\%$.
   Fijando dos de esas características se determina la tercera. Por ejemplo, el tamaño de muestra necesario para estimar μ con σ conocida, con un margen de error ε y con confianza $(1-\alpha)100\,\%$ es
   $$n = \left(\frac{\sigma}{\varepsilon} Z_{1-\alpha/2}\right)^2$$

Caso σ^2 desconocido

En el ejemplo anterior hemos considerado el caso en que la varianza σ^2 es conocida para poder construir un intervalo de confianza para la media μ, pero cuando la varianza es desconocida no podemos utilizar el intervalo anterior.

Caso σ^2 desconocido

En el ejemplo anterior hemos considerado el caso en que la varianza σ^2 es conocida para poder construir un intervalo de confianza para la media μ, pero cuando la varianza es desconocida no podemos utilizar el intervalo anterior.

Si σ^2 no es conocido se debe estimar.

Caso σ^2 desconocido

En el ejemplo anterior hemos considerado el caso en que la varianza σ^2 es conocida para poder construir un intervalo de confianza para la media μ, pero cuando la varianza es desconocida no podemos utilizar el intervalo anterior.

Si σ^2 no es conocido se debe estimar.

Luego, nos interesa encontrar una cantidad pivotal para $\bar{X}$ bajo este contexto.

Caso σ^2 desconocido

Para el caso σ^2 conocido se tiene que:

$$\sqrt{n}\frac{(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$$

Caso σ^2 desconocido

Para el caso σ^2 conocido se tiene que:

$$\sqrt{n}\frac{(\bar{X}-\mu)}{\sigma} \sim N(0,1)$$

Estimamos σ^2 mediante S^2, luego:

Caso σ^2 desconocido

Para el caso σ^2 conocido se tiene que:

$$\sqrt{n}\frac{(\bar{X}-\mu)}{\sigma} \sim N(0,1)$$

Estimamos σ^2 mediante S^2, luego:

$$\sqrt{n}\frac{(\bar{X}-\mu)}{S} \sim t(n-1)$$

Caso σ^2 desconocido

Para el caso σ^2 conocido se tiene que:

$$\sqrt{n}\frac{(\bar{X}-\mu)}{\sigma} \sim N(0,1)$$

Estimamos σ^2 mediante S^2, luego:

$$\sqrt{n}\frac{(\bar{X}-\mu)}{S} \sim t(n-1)$$

...y repetimos lo anterior!

Resumiendo

Para proveer una estimación intervalar para la media μ de una muestra proveniente de una distribución normal existen dos escenarios posibles:

Resumiendo

Para proveer una estimación intervalar para la media μ de una muestra proveniente de una distribución normal existen dos escenarios posibles:

- Varianza σ^2 conocida:
$$IC_{1-\alpha}(\mu) = \left[\overline{X} \pm \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}\right]$$
- Varianza σ^2 desconocida:
$$IC_{1-\alpha}(\mu) = \left[\overline{X} \pm \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1)\right]$$

Ejemplo práctico

TC Auditores contrató un psicólogo laboral para medir el grado de satisfacción en el trabajo de sus empleados. Dieciocho de estos fueron seleccionados y sometidos a un test que entregó un nivel de satisfacción promedio de 78.2 puntos en la escala $0 - 100$. La empresa cree que el grado de satisfacción de sus empleados sigue una distribución normal con varianza 144.

Encuentre un intervalo de confianza del 95 % para estimar la satisfacción media de todos los empleados.

Solución:

X : Nivel de satisfacción en el trabajo de los empleados de la empresa TC Auditores medido en la escala 0-100.

Los datos del enunciado son que $X \sim N(\mu; 144)$, $\bar{X} = 78,2$, $\sigma = 12$ y $\alpha = 0,05$, entonces

Solución:

X : Nivel de satisfacción en el trabajo de los empleados de la empresa TC Auditores medido en la escala 0-100.

Los datos del enunciado son que $X \sim N(\mu; 144)$, $\bar{X} = 78,2$, $\sigma = 12$ y $\alpha = 0,05$, entonces

$$\begin{aligned} IC(\mu) &= \overline{X} \pm \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} \\ IC(\mu) &= 78,2 \pm \frac{12}{\sqrt{18}} \cdot 1,96 \\ &\approx 78,2 \pm 5,5 \\ &\approx \left[72,7 \ ; \ 83,7\right] \end{aligned}$$

Por tanto, hay un 95 % de confianza de que el nivel medio de satisfacción laboral de los empleados de TC Auditores esté entre 72,7 y 83,7 puntos.

Solución:

X : Nivel de satisfacción en el trabajo de los empleados de la empresa TC Auditores medido en la escala 0-100.

Los datos del enunciado son que $X \sim N(\mu; 144)$, $\bar{X} = 78,2$, $\sigma = 12$ y $\alpha = 0,05$, entonces

$$\begin{aligned} IC(\mu) &= \overline{X} \pm \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} \\ IC(\mu) &= 78,2 \pm \frac{12}{\sqrt{18}} \cdot 1,96 \\ &\approx 78,2 \pm 5,5 \\ &\approx \left[72,7 \ ; \ 83,7\right] \end{aligned}$$

Por tanto, hay un 95 % de confianza de que el nivel medio de satisfacción laboral de los empleados de TC Auditores esté entre 72,7 y 83,7 puntos.

Intervalo para la varianza

Considere que $X_1, \ldots, X_N$ es una muestra aleatoria proveniente de una distribución Normal de media μ y varianza σ^2. Asumiendo que σ^2 es desconocido, deseamos encontrar un intervalo de confianza para σ^2

Necesitamos una cantidad pivotal

$$\frac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi^2(n-1)$$

Intervalo para la varianza

Considere que $X_1, \ldots, X_N$ es una muestra aleatoria proveniente de una distribución Normal de media μ y varianza σ^2. Asumiendo que σ^2 es desconocido, deseamos encontrar un intervalo de confianza para σ^2

Necesitamos una cantidad pivotal

$$\frac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi^2(n-1)$$

Ejercicio para el lector

Intervalo para el ratio de varianzas

Sea $X_1, \ldots, X_n$ una muestra aleatoria de tamaño n e $Y_1, \ldots, Y_m$ una muestra aleatoria de tamaño m. Asumiendo una distribución normal, con medias desconocidas y varianzas σ_X^2 y σ_Y^2, encuentre un intervalo de confianza para el ratio σ_X^2/σ_Y^2

Intervalo para el ratio de varianzas

Sea $X_1, \ldots, X_n$ una muestra aleatoria de tamaño n e $Y_1, \ldots, Y_m$ una muestra aleatoria de tamaño m. Asumiendo una distribución normal, con medias desconocidas y varianzas σ_X^2 y σ_Y^2, encuentre un intervalo de confianza para el ratio σ_X^2/σ_Y^2

Ejercicio para el lector