Estadística y Probabilidades 523250

Intervalos de Confianza

Andrea Fernández & Jean Paul Navarrete

[1] Departamento de Estadística
Universidad de Concepción

Mayo, 2024

Contenidos

Qué hemos visto...

- Hasta ahora solo hemos obtenido estimadores puntuales de los parámetros poblacionales de una distribución de probabilidades de X, $f_X(x;\theta)$ y algunas propiedades importantes.
- Los estimadores son estadísticos (funciones de la muestra aleatoria) y por lo tanto, variables aleatorias. Las distribuciones de los estadísticos las conocemos como distribuciones muestrales.

Estimación intervalar: Objetivos

- Una forma de estimación que involucre tanto al estadístico como su variación en la estimación, es la construcción de un intervalo o rango de valores tal que, bajo ciertas condiciones, contenga al valor real del parámetro θ.
- Necesitamos entonces determinar dos valores numéricos, que constituirán los extremos del intervalo que contendrán el valor del parámetro en cuestión y que, además sea un intervalo de amplitud la menor posible para que sea de utilidad.
- Una estimación por intervalo para un parámetro desconocido proporciona indicaciones tanto del valor numérico del parámetro como también una medida de cuan seguro o que confianza se tiene de ese valor numérico.

Intervalo de confianza: definición

- **Def:**
  Sea X la v.a. de interés en la población, cuya función de probabilidad es $f(x;\theta)$, depende de θ, parámetro desconocido. Dada una muestra aleatoria $\underline{X} = (X_1, \ldots, X_n)$ de$X$, los dos estadísticos $L_1(\underline{X})$ y $L_2(\underline{X})$ forman un intervalo del $(1-\alpha)$ de confianza para θ si

$$P\left(L_1(\underline{X}) \leq \theta \leq L_2(\underline{X})\right) \geq 1-\alpha$$

- No importando cual sea el valor de θ,
  - $1-\alpha$: nivel de confianza.
  - $L_1(\underline{X})$: límite inferior.
  - $L_2(\underline{X})$: límite superior.

Intervalo de confianza: definición

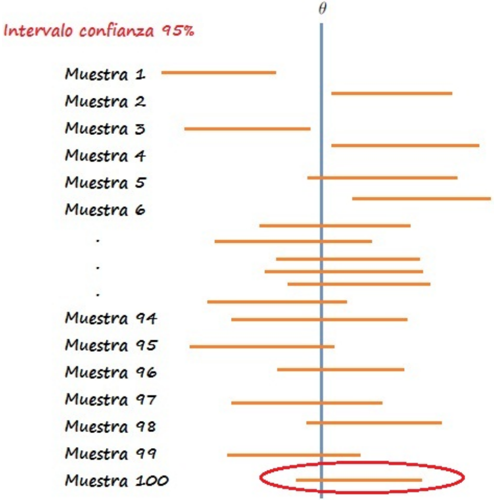
- Por supuesto que para muestras cualquieras, $\underline{x} = (x_1, \ldots, x_n)$, los intervalos obtenidos pueden o no contener al verdadero valor del parámetro. Más aún no sabemos si nuestro intervalo obtenido lo contiene o no.
- Si repetidas muestras del mismo tamaño son tomadas y se calcula

$$(L_1(\underline{x}); L_2(\underline{x}))$$

para cada una, la proporción $1-\alpha$ de los intervalos obtenidos contendrán al verdadero valor de θ.
- Esto nos lleva a decir que con una confianza $(1-\alpha)100\%$ nuestro intervalo contiene a θ.

Intervalo de confianza: definición

- Lo anterior, lo podemos representar gráficamente de la siguiente manera:



Intervalo de confianza: definición

- Para construir los intervalos de confianza, utilizaremos los Pivotes.
- Un pivote es:
  1. Una función de la muestra.
  2. Una función también del parámetro de interés: no tiene más valores desconocidos.
  3. Su distribución de probabilidad es conocida y está completamente definida.

Para la media μ, varianza conocida

- Suponga que $X_1, \ldots, X_n$ corresponde a una m.a de una v.a. $X \sim \mathcal{N}(\mu, \sigma_0^2)$, con σ_0^2 conocida.
- El pivote utilizado para construir el intervalo de confianza para μ está dado por:

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma_0^2}{n}}} \sim \mathcal{N}(0, 1)$$

Para la media μ, varianza conocida

- En base al pivote definido antes, el intervalo de confianza del $(1-\alpha)100\%$ para la media, está dado por:

$$IC(\mu) = \left(\bar{X} - z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}; \bar{X} + z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right)$$

Para la media μ, varianza conocida

- **Obs:**
  Note que en el $IC(\mu)$, éste está centrado en $\bar{X}$ (estimador de la media) y que se desvía en la cantidad $\pm z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}$
- Se conoce como error de estimación a la cantidad

$$e = z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}$$

- Así, se puede deducir que, para un cierto nivel de confianza y nivel de error, el tamaño de la muestra debe ser de

$$n = \left(\frac{z_{1-\alpha/2}\sigma}{e}\right)^2$$

Para la media μ, varianza conocida

- Suponga que $X_1, \ldots, X_n$ corresponde a una m.a de una v.a. X tal que
$$E(X) = \mu \qquad V(X) = \sigma_0^2$$
con σ_0^2 conocida.
- El intervalo de confianza $(1-\alpha)100\%$ para μ está dado por:
$$IC(\mu) = \left(\bar{X} - z_{1-\alpha/2}\frac{S}{\sqrt{n}}; \bar{X} + z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right)$$
- Obs: esta conclusión es valida cuando $n \geq 30$.

Para la media μ, varianza desconocida

- Suponga que $X_1, \ldots, X_n$ corresponde a una m.a de una v.a. $X \sim \mathcal{N}(\mu, \sigma^2)$ ambos parámetros desconocidos.
- Recordemos que, cuando la varianza de la población σ^2 es desconocida,

$$T = \frac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} \sim \mathcal{T}_{(n-1)}$$

lo cual nos sirve de pivote para construir el intervalo de confianza para μ.

Para la media μ, varianza desconocida

- Así, si $X_1, \ldots, X_n$ corresponde a una m.a de una v.a. $X \sim \mathcal{N}(\mu, \sigma^2)$ ambos parámetros desconocidos.
- El intervalo de confianza $(1-\alpha)100\%$ para μ está dado por:

$$IC(\mu) = \left(\bar{X} - t_{1-\alpha/2;(n-1)} \frac{S}{\sqrt{n}}; \bar{X} + t_{1-\alpha/2;(n-1)} \frac{S}{\sqrt{n}} \right)$$

Para la varianza σ^2 de $X \sim \mathcal{N}(\mu, \sigma^2)$

- Sea $X_1, \ldots, X_n$ una m.a. de $X \sim \mathcal{N}(\mu, \sigma^2)$ ambos parámetros desconocidos. Estamos interesados en estimar σ^2
- Recordemos que, cuando calculamos probabilidades sobre la varianza muestral, utilizamos lo siguiente:

$$\frac{(n-1)S_X^2}{\sigma^2} \sim \mathcal{X}^2_{(n-1)}$$

lo cual sirve de pivote para encontrar el intervalo de confianza para σ^2

Para la varianza σ^2 de $X \sim \mathcal{N}(\mu, \sigma^2)$

- Así, si $X_1, \ldots, X_n$ una m.a. de $X \sim \mathcal{N}(\mu, \sigma^2)$ ambos parámetros desconocidos. Estamos interesados en estimar σ^2
- El intervalo de confianza $(1-\alpha)100\%$ para σ^2 está dado por:

$$I.C._{(1-\alpha)100\%}(\sigma^2) = \left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2;(n-1)}} ; \frac{(n-1)S^2}{\chi^2_{\alpha/2;(n-1)}} \right)$$

donde S^2 corresponde a la varianza muestral.

Proporción

- Considere la v.a. $X \sim \mathcal{B}er(P)$ la cual modela los ensayos de éxito-fracaso, es decir

$$X = \begin{cases} 1 & con \quad prob \quad P \\ 0 & con \quad prob \quad 1-P \end{cases}$$

Proporción

- Considere una m.a tamaño n de $X \sim \mathcal{B}er(P)$. Si $n\bar{x} \geq 5$ y $n(1-\bar{x}) \geq 5$, usando el teorema del límite central, sabemos que

$$\bar{X} \sim \mathcal{N}\left(E(X), \frac{V(X)}{n}\right)$$

donde $E(X) = p$ y $V(X) = P(1-p)$

- Luego, el pivote para el intervalo de confianza para P está dado por:

$$Z = \frac{\bar{X} - P}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} \sim \mathcal{N}(0,1)$$

Proporción

- Así, si $X_1, X_2, \ldots, X_n$ es una m.a. de $X \sim \mathcal{B}er(P)$. Si $n\bar{x} \geq 5$ y $n(1-\bar{x}) \geq 5$, entonces
- el intervalo de confianza $(1-\alpha)100\%$ para P está dado por

$$I.C.(P) = \left(\bar{x} \pm z_{1-\alpha/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} \right)$$

Dos poblaciones

- Consideremos las muestras aleatorias:
  i) $X_1, \ldots, X_n$ de $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$
  ii) $Y_1, \ldots, Y_n$ de $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

Dos poblaciones

- Estaremos interesados en estudiar qué ocurre con la diferencia de medas de las v.a. X e Y, así, buscaremos encontrar el intervalos de confianza para $\mu_X - \mu_Y$, lo cual denotaremos como

$$I.C._{(1-\alpha)100\%}(\mu_X - \mu_Y)$$

- Este intervalo de confianza dependerá de la información que se tenga de X e Y respecto a sus varianzas.

Dos poblaciones: varianza conocida

- Suponga que cuenta con una m.a. tamaño n de $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ y con una m.a. tamaño m de $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, donde σ_X^2 y σ_Y^2 son conocidas. Además, las muestras aleatorias son independientes.
- El intervalo de confianza $(1-\alpha)100\%$ para $\mu_X - \mu_Y$ está dado por

$$IC(\mu_X - \mu_Y) = \left((\bar{X} - \bar{Y}) - z_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}; (\bar{X} - \bar{Y}) + z_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \right)$$

Dos poblaciones: varianzas desconocidas e iguales

- Suponga que cuenta con una m.a. tamaño n de $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ y con una m.a. tamaño m de $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, donde σ_X^2 y σ_Y^2 son desconocidas e iguales. Además, las muestras aleatorias son independientes.
- El intervalo de confianza $(1-\alpha)100\%$ para $\mu_X - \mu_Y$ está dado por

$$IC(\mu_X - \mu_Y) = \left(\bar{X} - \bar{Y} \pm t_{1-\alpha/2;(n+m-2)} S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

donde $S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$

Dos poblaciones: varianzas desconocidas y distintas

- Suponga que cuenta con una m.a. tamaño n de $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ y con una m.a. tamaño m de $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, donde σ_X^2 y σ_Y^2 son desconocidas y distintas. Además, las muestras aleatorias son independientes.
- El intervalo de confianza $(1-\alpha)100\%$ para $\mu_X - \mu_Y$ está dado por

$$IC(\mu_X - \mu_Y) = \left(\bar{X} - \bar{Y} \pm t_{1-\alpha/2,(\nu)} \sqrt{\frac{S_X^2}{n} + \frac{S_y^2}{m}}\right)$$

donde $\nu = \frac{(S_X^2/n + S_Y^2/m)^2}{(S_X^2/n)^2/(n-1) + (S_Y^2/m)^2/(m-1)}$

Diferencia de proporciones

- Suponga que cuenta con una m.a. tamaño n de $X \sim \mathcal{B}er(P_X)$ y con una m.a. tamaño m de $Y \sim \mathcal{B}er(P_X)$. Además, las muestras aleatorias son independientes.
- Si $n\bar{x}$ y $n(1-\bar{x}) \geq 5$ y $n\bar{y}$ y $n(1-\bar{y}) \geq 5$, entonces el intervalo de confianza para $P_X - P_Y$ está dado por

$$I.C.(P_X - P_Y) = \left(\bar{x} - \bar{y} \pm z_{1-\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n} + \frac{\bar{y}(1-\bar{y})}{m}}\right)$$

Dos poblaciones: muestras pareadas

- Suponga que cuenta con una m.a. tamaño n de $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ en el tiempo 1 y considere la misma m.a. anterior y ahora mide nuevamente X en un tiempo dos. Interesa saber si ha ocurrido un cambio luego de que X es medida por segunda vez.
- El intervalo de confianza $(1-\alpha)100\%$ para $\mu_D = \mu_X - \mu_Y$ está dado por

$$IC(\mu_D) = \left(\bar{D} \pm t_{1-\alpha/2}(n-1)\sqrt{\frac{S_D^2}{n}} \right)$$

donde $\bar{D} = \frac{\sum D_i}{n}$, $S_D^2 = \frac{\sum (D_i - \bar{D})^2}{n-1}$

Intervalos de confianza para cociente de varianzas

- En los casos anteriores en donde las varianzas poblacionales son desconocidas, se mostraron resultados para los casos en donde las varianzas son distintas o iguales.
- Como las varianzas son desconocidas, se debe de alguna manera verificar si son iguales o distintas. Para ello, se realiza un intervalo de confianza del $(1-\alpha)100\%$ para el cociente de las varianzas. Si este intervalo contiene al valor 1, entonces con una confianza $(1-\alpha)100\%$ entonces se considera que las varianzas son desconocidas, pero iguales.

Cociente de varianzas

- Suponga que cuenta con una m.a. tamaño n de $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ y con una m.a. tamaño m de $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ donde ambas varianzas son desconocidas. Además, las muestras aleatorias son independientes.
- El ntervalo de confianza $(1-\alpha)100\%$ para $\frac{\sigma_Y^2}{\sigma_X^2}$ está dado por

$$I.C.\left(\frac{\sigma_Y^2}{\sigma_X^2}\right) = \left(\frac{S_Y^2}{S_X^2} f_{n-1,m-1;(\alpha/2)}; \frac{S_Y^2}{S_X^2} f_{n-1,m-1;(1-\alpha/2)}\right)$$

Gracias por su atención