



Estimación por intervalo.

Esquema

- [1. El intervalo de confianza para \$\mu\$, cuando se muestrea una distribución normal con varianza conocida](#)
 - [2. Intervalo de confianza para \$\mu\$ cuando se muestrea una distribución normal de varianza desconocida.](#)
 - [3. Intervalos de confianza para la diferencia de medias cuando se muestrean dos distribuciones normales independientes.](#)
 - [4. Intervalo de confianza para \$\sigma^2\$ cuando se muestrea una distribución normal con media desconocida.](#)
 - [5. Intervalos de confianza para el cociente de dos varianzas cuando se muestrean dos distribuciones normales independientes.](#)
 - [6. Intervalos de confianza para el parámetro de proporción \$p\$ cuando se muestrea una distribución binomial.](#)
-

Pongamos un ejemplo para introducir el concepto de estimación por intervalo.

Una compañía decide que los últimos 36 meses han sido típicos en relación a la demanda de un producto, con base en los datos muestrales el valor calculado para la media muestral es de:

$$\bar{x} = 200$$

En otras palabras 200 es un valor estimado puntual de un parámetro desconocido, el cual representa la demanda promedio de este producto por la compañía. Este estimador, ¿implica que la demanda promedio desconocida no sea mayor de 250 ni menor de 150?. Esto no es posible saberlo, ya que no se tiene ninguna indicación del posible error en el estimado puntual.

Supóngase que la desviación estándar de la media muestral es 60. De acuerdo con el teorema central del límite

$$\bar{X} \rightarrow N(\mu, 60)$$

De esta forma la probabilidad de que

$$P\left(|\bar{X} - \mu| < 120\right) = 0,95, \quad (2\sigma = 120)$$

lo que equivale a que

$$P\left(\bar{X} - 120 < \mu < \bar{X} + 120\right) = 0,95$$

Dado que \bar{X} es una variable aleatoria, el intervalo

$$\left(\bar{X} - 120, \bar{X} + 120\right)$$

es un intervalo aleatorio y la probabilidad de que dicho intervalo contenga al verdadero valor de μ , es de 0,95. En otras palabras si se obtienen muestras repetidas de la misma población y del mismo tamaño y se calculan los valores específicos para dicho intervalo; entonces debe esperarse que en un 95% de los casos estos intervalos contengan al valor de la media desconocida μ .

Por otro lado, para los datos obtenidos de una sola muestra

$$\bar{x} = 200$$

sería incorrecto decir que la probabilidad de que μ se encuentre en el intervalo

$$\left(\bar{x} - 120, \bar{x} + 120\right) = (80, 320)$$

es de 0,95. Sin embargo, la probabilidad de 0,95 para el intervalo aleatorio sugiere que la confianza en que el intervalo (80, 320) contenga el valor de la media desconocida μ es alta. Sólo en este sentido se permite asignar un grado de confianza a la proposición μ se encuentra en el intervalo (80, 320) igual a la probabilidad del intervalo aleatorio

$$\left(\bar{X} - 120, \bar{X} + 120\right)$$

Así cuando se escribe

$$P(80 < \mu < 320) = 0,95$$

no se está formulando ninguna proposición probabilística en sentido clásico, sino más bien un grado de confianza.

En general, la construcción de un intervalo de confianza para el parámetro desconocido θ consiste en encontrar un estadístico suficiente T y relacionarlo con otra variable aleatoria $X=f(T;\theta)$, en donde X involucra a θ pero la distribución de X no contiene a θ , así como tampoco a ningún otro parámetro desconocido. Entonces se seleccionan dos valores x_1 y x_2 tales que:

$$P(x_1 < X < x_2) = 1 - \alpha$$

en donde $1-\alpha$ recibe el nombre de coeficiente de confianza. Mediante cálculo algebraico se puede modificar el contenido entre paréntesis y expresarlo:

$$P[h_1(T) < \theta < h_2(T)] = 1 - \alpha$$

El intervalo de confianza de θ se obtiene sustituyendo en $h_1(T)$ y $h_2(T)$ los estimadores calculados a partir de la muestra, dando lugar a lo que se llama intervalos de **confianza bilaterales**. Al seguirse el mismo procedimiento pueden calcularse los intervalos de **confianza unilaterales**.

$$P[g_1(T) < \theta] = 1 - \alpha \quad \text{o bien}$$

$$P[\theta < g_2(T)] = 1 - \alpha$$

Intervalos de confianza para μ cuando se muestrea una distribución normal con varianza conocida.

El intervalo de confianza se calcula en base al mejor estimador de μ , la media muestral \bar{X} .

Por tanto,

$$P[g_1(\mu) < \bar{X} < g_2(\mu)] = 1 - \alpha$$

de manera que

$$\int_{-\infty}^{g_1(\mu)} f(\bar{x}, \mu) d\bar{x} = \alpha / 2$$

y

$$\int_{g_2(\mu)}^{\infty} f(\bar{x}, \mu) d\bar{x} = \alpha / 2$$

en donde f es la función de densidad de la distribución de muestreo de \bar{X} .

Dado que

$$\bar{X} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

y por tanto

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

sigue una normal $N(0; 1)$ se sigue:

$$P\left[g_1(\mu) < \bar{X} < g_2(\mu)\right] = P\left[\frac{g_1(\mu) - \mu}{\sigma/\sqrt{n}} < Z < \frac{g_2(\mu) - \mu}{\sigma/\sqrt{n}}\right] = 1 - \alpha$$

Pero

$$P\left[Z_{\alpha/2} < Z < Z_{1-\alpha/2}\right] = 1 - \alpha$$

se sigue que

$$\frac{g_1(\mu) - \mu}{\sigma/\sqrt{n}} = Z_{\alpha/2} \quad y \quad \frac{g_2(\mu) - \mu}{\sigma/\sqrt{n}} = Z_{1-\alpha/2}$$

Despejando $g_1(\mu)$ y $g_2(\mu)$ se obtiene:

$$g_1(\mu) = \mu + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad y \quad g_2(\mu) = \mu + Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Teniendo en cuenta que $Z_{\alpha/2}$ es igual a $-Z_{1-\alpha/2}$ el intervalo queda de la forma:

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

Al manipular las desigualdades de dentro del paréntesis, se tiene:

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

Si se reemplaza la variable aleatoria por los datos calculados a partir de una muestra de tamaño n , un intervalo de confianza del $100(1-\alpha)\%$ para μ , es:

$$\bar{x} \pm Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Tamaño de la muestra. Supóngase, en las mismas condiciones de este ejercicio, que deseamos estimar el tamaño de la muestra de manera que, con una probabilidad de $1-\alpha$, la media muestral \bar{X} se encuentre a una distancia inferior a ε unidades de la media poblacional μ . De la expresión:

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

se obtiene

$$\varepsilon = Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

al resolver para n se obtiene:

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

Una objeción muy severa a esta formula es que se requiere el conocimiento de la varianza poblacional σ^2 . Si este no se conoce, una estimación muy burda para la desviación típica es igual a la sexta parte del recorrido de las observaciones, ya que para distribuciones unimodales la gran mayoría de las observaciones se encuentran a una distancia de 3 desviaciones estándar de la media.

[\[Volver al principio\]](#)

Intervalo de confianza para μ cuando se muestrea una distribución normal de varianza desconocida.

Cuando se muestrea una población normal $N(\mu, \sigma)$, el estadístico

$$T = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

sigue una t-Student con n-1 grados de libertad, donde

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)}$$

Por lo tanto es posible determinar el cuantil $t_{1-\alpha/2, n-1}$ de T, para el cual

$$P(-t_{1-\alpha/2, n-1} < T < t_{1-\alpha/2, n-1}) = 1 - \alpha$$

Al sustituir T por su valor se obtiene

$$P\left(-t_{1-\alpha/2, n-1} < \frac{\bar{X} - \mu}{S / \sqrt{n}} < t_{1-\alpha/2, n-1}\right) = 1 - \alpha$$

de donde

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

La probabilidad de que este intervalo aleatorio contenga a la media poblacional es de $1-\alpha$.

Para una muestra particular de tamaño n, a partir de los valores estimados de \bar{x} y de s^2 , un intervalo de confianza del $100(1-\alpha)\%$ para μ es:

$$\bar{x} \pm t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}}$$

[\[Volver al principio\]](#)

Intervalos de confianza para la diferencia de medias cuando se muestrean dos distribuciones normales independientes.

Sean X_1, X_2, \dots, X_{n_x} e Y_1, Y_2, \dots, Y_{n_y} dos muestras aleatorias de dos distribuciones normales independientes.

Si suponemos que las **varianzas poblacionales son distintas pero conocidas**, entonces se tiene:

$$\bar{X} \rightarrow N\left(\mu_x, \frac{\sigma_x^2}{n_x}\right)$$

$$\bar{Y} \rightarrow N\left(\mu_y, \frac{\sigma_y^2}{n_y}\right)$$

De donde

$$\bar{X} - \bar{Y} \rightarrow N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right)$$

y por tanto

$$Z = \frac{\bar{X} - \bar{Y} - \mu_x - \mu_y}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \rightarrow N(0,1)$$

De esta forma es posible encontrar el cuantil $Z_{1-\alpha/2}$, tal que:

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

Sustituyendo Z por su valor y operando

$$P\left(\bar{X} - \bar{Y} - Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} < \mu_x - \mu_y < \bar{X} - \bar{Y} + Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right) = 1 - \alpha$$

Un intervalo de confianza para una muestra particular de tamaño n, viene dado por:

$$\bar{x} - \bar{y} \pm Z_{1-\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

Para el caso en el que el muestreo se lleve a cabo sobre dos poblaciones normales independientes con **varianzas iguales pero desconocidas**

$$T = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2} \left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{S_p \sqrt{\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}}$$

donde

$$S_p = \sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}}$$

sigue una t-Student con $n_x + n_y - 2$ grados de libertad.

Un intervalo de confianza, para una muestra particular al $100(1-a)\%$ viene dado por:

$$\bar{x} - \bar{y} \pm t_{1-\alpha/2, f} S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

Donde s_p es el valor obtenido para la muestra particular de S_p .

En el **supuesto de que σ_x^2 y σ_y^2 sean desconocidas y distintas** y haya que estimarlas a partir de S_x^2 y S_y^2 el problema se complica, en tal caso se tiene:

$$T = \frac{(\bar{X} - \bar{Y} - (\mu_x - \mu_y))}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}}$$

sigue una distribución t-Student con f grados de libertad, en donde f es la aproximación de Welch:

$$f = \frac{\left(\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}\right)^2}{\frac{\left(\frac{S_x^2}{n_x}\right)^2}{n_x + 1} + \frac{\left(\frac{S_y^2}{n_y}\right)^2}{n_y + 1}}$$

expresado en número entero.

[\[Volver al principio\]](#)

Intervalo de confianza para σ^2 cuando se muestrea una distribución normal con media desconocida

La inferencias con respecto a la varianza σ^2 cuando se muestrea una población normal se formula con base a

$$\frac{(n-1)S^2}{\sigma^2} \rightarrow \chi_{n-1}^2$$

Entonces es posible determinar los valores cuantiles

$$\chi_{\alpha/2, n-1}^2 \quad \text{y} \quad \chi_{1-\alpha/2, n-1}^2$$

tales que:

$$P\left(\chi_{\alpha/2, n-1}^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{1-\alpha/2, n-1}^2\right) = 1 - \alpha$$

y de aquí

$$P\left[\frac{1}{\chi_{\alpha/2, n-1}^2} > \frac{\sigma^2}{(n-1)S^2} > \frac{1}{\chi_{1-\alpha/2, n-1}^2}\right] = 1 - \alpha$$

Con base en los datos de una muestra se puede calcular un intervalo de confianza al $100(1-\alpha)\%$

$$\left[\frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2}, \frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2} \right]$$

[\[Volver al principio\]](#)

Intervalos de confianza para el cociente de dos varianzas cuando se muestrean dos distribuciones normales independientes.

Sean X_1, X_2, \dots, X_{n_x} variables aleatorias independientes $N(\mu_x, \sigma_x)$

Sean Y_1, Y_2, \dots, Y_{n_y} variables aleatorias independientes $N(\mu_y, \sigma_y)$

Si X e Y son independientes

$$\frac{\frac{S_x^2}{\sigma_x^2}}{\frac{S_y^2}{\sigma_y^2}} \rightarrow F_{n_x-1, n_y-1}$$

Es decir, el cociente sigue una F con n_x-1, n_y-1 grados de libertad.

Entonces puede escribirse:

$$P\left(a < \frac{S_x^2 / \sigma_x^2}{S_y^2 / \sigma_y^2} < b\right) = 1 - \alpha$$

donde a y b son los valores cuantiles

$$\alpha = \frac{1}{f_{1-\alpha/2, n_y-1, n_x-1}} \quad \text{y} \quad b = f_{1-\alpha/2, n_x-1, n_y-1}$$

La probabilidad se puede expresar así:

$$P\left[\frac{aS_Y^2}{S_X^2} < \frac{\sigma_Y^2}{\sigma_X^2} < \frac{bS_Y^2}{S_X^2}\right] = 1 - \alpha$$

De esta manera un intervalo de confianza del $100(1-\alpha)\%$ para σ_Y^2/σ_X^2 está dado por

$$(as_Y^2 / s_X^2, bs_Y^2 / s_X^2)$$

[\[Volver al principio\]](#)

Intervalos de confianza para el parámetro de proporción p cuando se muestrea una distribución binomial.

Puede demostrarse que el estimador de máxima verosimilitud del parámetro p , denotado por

$$\hat{P} = X/n$$

donde X es binomial con parámetro n y p . Nótese que \hat{P} es un estimador insesgado de p , ya que

$$E(\hat{P}) = \frac{1}{n} E(X) = \frac{1}{n} np = p$$

La varianza viene dada por:

$$Var(\hat{P}) = Var(X/n) = \frac{1}{n^2} [np(1-p)] = p(1-p)/n$$

Recuérdese que para n grande, la variable aleatoria

$$\frac{X - np}{\sqrt{np(1-p)}} \rightarrow N(0,1)$$

entonces puede demostrarse que la distribución de

$$\frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} \rightarrow N(0,1)$$

De esta forma, la probabilidad

$$P \left[\hat{p} - Z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + Z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] = 1 - \alpha$$

Para una muestra particular de tamaño n, un intervalo de confianza al 100(1- α)% viene dado por

$$\left[\hat{p} - Z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

donde el estimador de \hat{p} se obtiene de la muestra:

$$\hat{p} = \frac{x}{n}$$

Tamaño de muestra: Con respecto al muestreo de una distribución binomial, un problema que surge de forma frecuente, es el estimar el tamaño de la muestra necesario para que con una confiabilidad del 100(1- α)% el estimado del parámetro de proporción se encuentre a no más de ϵ unidades de p. Es decir

$$P \left(\left| \frac{X}{n} - p \right| < \epsilon \right) = 1 - \alpha$$

de donde se sigue que

$$\epsilon = Z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

Despejando

$$n = \frac{Z_{1-\alpha/2}^2 p(1-p)}{\epsilon^2}$$

Se ha obtenido n en función de p, el parámetro que se desea estimar. Dado que este no se conoce de antemano, lo que de manera generalizada se hace es escoger el valor más conservador de n. Esto ocurre cuando la cantidad p(1-p) es máxima, es decir cuando p=1/2.

[\[Volver al principio\]](#)