

Estadística

Estudia las técnicas para recolectar, reunir y ordenar datos y ^{para} inferir resultados a partir de los datos recolectados.

Ejemplo: Supongamos que interesa estudiar la altura de los estudiantes de la facultad. Se toman al azar n estudiantes y a cada uno se le mide la altura. Tenemos n datos tal que asentamos x_1, x_2, \dots, x_n .

Población → conjunto de todos los objetos de interés. Los estudiantes de la facultad

Muestra → una parte de la población. La muestra toma al azar (n se eligen de la población N alumnos al azar como si se eligieran bolillos de una urna con N bolillos). Se hace un muestreo aleatorio simple.

La característica que importa estudiar es la altura de los alumnos. Se define X : "altura de un estudiante tomado al azar".

En estadística la distribución de X es desconocida. Los únicos datos disponibles son las mediciones x_1, x_2, \dots, x_n .

Estadística descriptiva: es la parte de estadística que estudia la muestra, es decir, estudia y ordena la muestra.

Estadística inferencial: estudia como inferir resultados a partir de la muestra sobre la población.

En el ejemplo nos puede interesar $\mu = E(X)$ que es desconocido. Si partimos de x_1, x_2, \dots, x_n se trata de "estimar" un valor de μ . (Se sea de alguna forma obtener algún valor aproximado de μ).

Entonces al tener una muestra, queremos obtener el verdadero valor de μ poniéndole solamente n valores de la población.

Tenemos las mediciones x_1, x_2, \dots, x_n . Podemos asumir que x_i es el resultado de medir la RA X_i : "altura del alumno i " $i=1, 2, \dots, n$

Si las mediciones se hacen bajo las mismas condiciones, podemos asumir que x_1, x_2, \dots, x_n tienen la misma distribución y que es igual a la distribución de la RA X original.

Además supondremos que x_1, x_2, \dots, x_n son independientes.

Entonces se dice que x_1, x_2, \dots, x_n es una muestra aleatoria de la RA X .

Si interesa inferir algún valor para la $\mu = E(X)$ (desconocida), es razonable estimar μ con $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Se busca que la estimación sea este lo más cerca posible al verdadero valor de μ .

Si tomas una muestra de n alumnos, y obtienes otros valores x'_1, x'_2, \dots, x'_n , entonces tendrás otra estimación \bar{x}' para μ .

Podemos considerar que \bar{X} y \tilde{X}' son observaciones diferentes que toma la ra $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, donde ahora X_1, X_2, \dots, X_n son las variables de una muestra aleatoria.

Dicimos que $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ es un estimador puntual de μ y los valores \bar{x} y \tilde{x}' son estimaciones de μ .

En general si tenemos una ra que es una función de la muestra aleatoria, se dice que esa ra es un estadístico.

\bar{X} es un estadístico. Cuando se usa un estadístico para estimar un parámetro desconocido se dice que el estadístico es un estimador puntual.

En el ejemplo $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ es un estimador puntual de μ . Mientras que \bar{x} y \tilde{x}' (valores que toma \bar{X}) son estimaciones puntuales de μ .

Notación: Si queremos a cualquier parámetro desconocido con θ y con $\hat{\theta}$ al estimador puntual de θ . En el ejemplo: $\hat{\mu} = \bar{X}$

Podemos considerar a cualquier estadístico como estimador puntual de un parámetro desconocido.

Entonces necesitamos criterios para considerar que estadístico es "mejor" estimador puntual.

Criterios para evaluar estimadores puntuales:

$$\hat{\mu} = \bar{X} \quad \begin{matrix} \xrightarrow{\bar{x}_1} \\ \xrightarrow{\bar{x}_2} \\ \xrightarrow{\bar{x}_3} \\ \vdots \\ \xrightarrow{\bar{x}_n} \end{matrix}$$

Estimador insesgado: Sea $\hat{\theta}$ un estimador puntual de θ , donde θ es un parámetro desconocido. Si $E(\hat{\theta}) = \theta$ se dice que $\hat{\theta}$ es un estimador insesgado de θ (cuálquier sea el valor de θ) \rightarrow el valor verdadero

Ejemplo: sea X_1, X_2, \dots, X_n muestra aleatoria de una ra X donde $\mu = E(X)$ desconocida. Sea $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Sabemos que $\hat{\mu} = \bar{X}$ es un estimador insesgado de μ , este es: $E(\bar{X}) = \mu$ cuálquier sea el valor de μ .

$$E(\hat{\mu}) = E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu$$

$$\rightarrow \frac{1}{n} \cdot n\mu = \mu = E(\bar{X}) \quad \text{cuálquier sea el valor de } \mu.$$

Sea X_1, X_2, \dots, X_n muestra aleatoria de una ra X con $\mu = E(X)$ y $\sigma^2 = V(X)$ ambos desconocidos. El estadístico que vamos regularmente para estimar σ^2 es:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \hat{\sigma}^2 = S^2 \text{ varianza muestral}$$

$\hat{\sigma}^2 = S^2$ es un estimador insesgado de σ^2 m.s. Sabemos que

$$\left[\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right] \rightarrow E(S^2) = E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right)$$

$$= \frac{1}{n-1} E\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n-1} E\left(\sum_{i=1}^n X_i^2 - n\bar{X}\right) = \frac{1}{n-1} [E(\sum X_i^2) - nE(\bar{X}^2)]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^n E(X_i^2) - nE(\bar{X})^2 \right] \otimes \quad E(\bar{X}^2) = V(\bar{X}) + (E(\bar{X}))^2 \quad V(X) = E(X^2) - (E(X))^2$$

$$= \frac{\sigma^2}{n} + \mu^2 \quad E(X^2) = V(X) + (E(X))^2 = \sigma^2 + \mu^2$$

$$\textcircled{*} \quad = \frac{1}{n-1} \left[\sum (\sigma^2 + \mu^2) - n \left(\frac{\sigma^2}{n} + \mu^2 \right) \right] = \frac{1}{n-1} [n(\sigma^2 + \mu^2) - \sigma^2 - n\mu^2]$$

$$= \frac{1}{n-1} [n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2] = \frac{\sigma^2(n-1)}{n-1} = \sigma^2$$

Observaciones

① si $X_1, X_2, X_3, \dots, X_n$ es una muestra aleatoria de una ra X aleatoria X con $\mu = E(X)$ y $\sigma^2 = V(X)$ desconocida, entonces: $\hat{\mu}^2 = \bar{X}^2$ no es un estimador inexpresivo de μ^2 pues:

$$E(\hat{\mu}^2) = E(\bar{X}^2) = \underbrace{V(\bar{X})}_{\frac{\sigma^2}{n}} + (E(\bar{X}))^2 = \frac{\sigma^2}{n} + \mu^2 \neq \mu^2$$

Ley de un estimador puntual: $b(\hat{\theta}) = E(\hat{\theta}) - \theta$. Entonces en el ej anterior: $b(\hat{\mu}^2) = b(\bar{X}^2) = E(\bar{X}^2) - \mu^2$.

$$= \frac{\sigma^2}{n} + \mu^2 - \mu^2 = \boxed{\frac{\sigma^2}{n}}$$

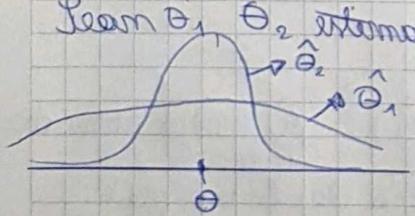
Notar que el \bar{X} tiende a cero cuando el tamaño de la muestra tiende a infinito, entonces \bar{X}^2 es un estimador de μ^2 asintóticamente inexpresivo

General: $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2}$ no es un estimador inexpresivo de σ .

Se verifica si tiene de una $X \sim N(\mu, \sigma)$

→ si $\hat{\theta}$ es un estimador puntual de θ tal que $b(\hat{\theta}) = 0$ entonces $b(\hat{\theta}) = E(\hat{\theta}) - \theta = 0$, o sea $\hat{\theta}$ es un estimador inexpresivo de θ

Sean $\hat{\theta}_1, \hat{\theta}_2$ estimadores inexpresivos de θ . Ora $E(\hat{\theta}_1) = \theta$ y $E(\hat{\theta}_2) = \theta$



Consideremos mejor estimador puntual de θ a aquél que tenga menor varianza. Es decir que si $V(\hat{\theta}_1) > V(\hat{\theta}_2)$, consideramos mejor estimador de θ a $\hat{\theta}_2$.

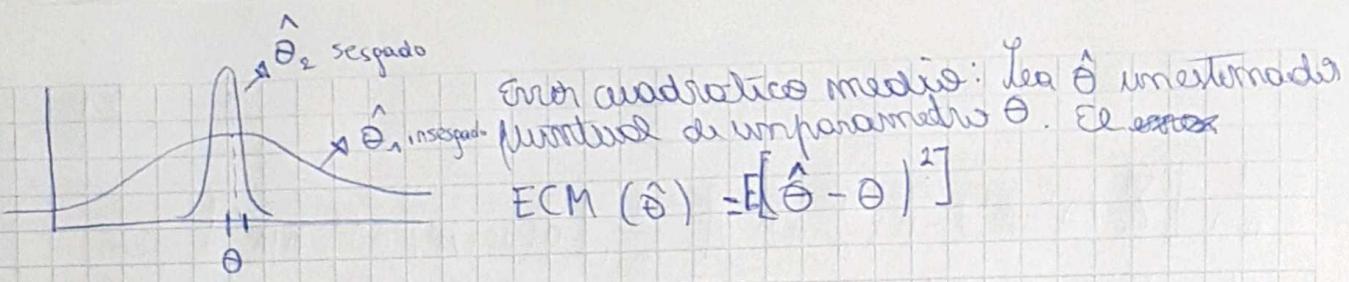
Ej. X_1, X_2, \dots, X_7 muestra aleatoria de ra X con $\mu = E(X)$ y $\sigma^2 = V(X)$
Sean $\hat{\mu}_1 = \bar{X} = \frac{1}{7} \sum_{i=1}^7 x_i$ y $\hat{\mu}_2 = \frac{x_1 + x_7}{2}$

$$E(\hat{\mu}_1) = E(\bar{X}) = \mu \quad E(\hat{\mu}_2) = E\left(\frac{x_1 + x_7}{2}\right) = \frac{1}{2} (E(x_1) + E(x_7)) = \frac{1}{2} 2\mu = \mu$$

Amboos inexpresivos, vean las varianzas.

$$V(\hat{\mu}_1) = V(\bar{X}) = \boxed{\frac{\sigma^2}{7}} \quad V(\hat{\mu}_2) = V\left(\frac{x_1 + x_7}{2}\right) = \frac{1}{4} (V(x_1) + V(x_7))$$

Como $V(\hat{\mu}_1) < V(\hat{\mu}_2)$ y considera que $\hat{\mu}_1$ es mejor estimador puntual que $\hat{\mu}_2$ → $\hat{\mu}_1$ es independiente de $\hat{\mu}_2$.



Observaciones:

① Si $\hat{\theta}$ es un estimador insesgado de θ , entonces $E(\hat{\theta}) = \theta$, o sea $ECM(\hat{\theta}) = E[(\hat{\theta} - E(\hat{\theta}))^2] = V(\hat{\theta})$ y $V(\lambda) = E[(X - E(X))^2] = E(X^2) - (E(X))^2$

② Se puede probar que $ECM(\hat{\theta}) = V(\hat{\theta}) + (b(\hat{\theta}))^2$

En general si $\hat{\theta}_1$ y $\hat{\theta}_2$ son 2 estimadores puntuales de un parámetro θ , se considera mejor estimador de θ aquél que tenga menor ECM .

Ej.: Sea X_1, X_2, \dots, X_n muestra aleatoria de un rango X con $E(X) = M$ y $\sigma^2 = V(X)$.

Consideremos $\hat{\mu}_1 = \frac{1}{7} \sum_{i=1}^7 X_i = \bar{X}$ $\hat{\mu}_3 = \frac{2X_1 - X_2 + X_3}{3}$

$\rightarrow ECM(\hat{\theta}) = V(\hat{\theta}) + (b(\hat{\theta}))^2$

$$E(\hat{\mu}_1) = E(\bar{X}) = M \quad \therefore b(\hat{\mu}_1) = 0$$

$$V(\hat{\mu}_1) = \frac{\sigma^2}{7} \quad \rightarrow ECM(\hat{\mu}_1) = V(\hat{\mu}_1)^2 + 0 = \frac{\sigma^2}{7}$$

$$\begin{aligned} E(\hat{\mu}_3) &= E\left(\frac{2X_1 - X_2 + X_3}{3}\right) = \frac{1}{3}(2E(X_1) - E(X_2) + E(X_3)) \\ &= \frac{1}{3}(2M - M + M) = \frac{2}{3}M \quad \rightarrow b(\hat{\mu}_3) = E(\hat{\mu}_3) - M = \frac{2}{3}M - M = \boxed{-\frac{M}{3}} \end{aligned}$$

$$\begin{aligned} V(\hat{\mu}_3) &= V\left(\frac{2X_1 - X_2 + X_3}{3}\right) = \left(\frac{1}{3}\right)^2 (4V(X_1) + V(X_2) + V(X_3)) = \frac{1}{9}(4\sigma^2 + \sigma^2 + \sigma^2) \\ &= \frac{2}{3}\sigma^2 \end{aligned}$$

$$\rightarrow ECM(\hat{\mu}_3) = \frac{2}{3}\sigma^2 + \left(-\frac{M}{3}\right)^2 = \frac{2}{3}\sigma^2 + \frac{M^2}{9}$$

Notar que $\frac{2}{3}\sigma^2 > \frac{1}{7}\sigma^2 \therefore \frac{2}{3}\sigma^2 + \frac{M^2}{9} > \frac{1}{7}\sigma^2$

$\therefore E(M(\hat{\mu}_3)) > ECM(\hat{\mu}_1)$ ∴ $\hat{\mu}_3$ es mejor estimador de M .

Estimadores consistentes: Sea X_1, X_2, \dots, X_n muestra aleatoria de X .
Sea $\hat{\theta}_n$ un estimador puntual de θ .

Se dice que $\hat{\theta}_n$ es un estimador consistente de θ

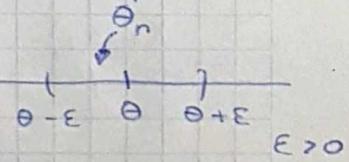
$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| > \varepsilon) \rightarrow 0 \quad \forall \varepsilon > 0$$

$$\text{o } P\left(\lim_{n \rightarrow \infty} |\hat{\theta}_n - \theta| \leq \varepsilon\right) \rightarrow 1 \quad \forall \varepsilon > 0$$

Se puede probar el sig. teorema:

Teorema: Sea $\hat{\theta}_n$ un estimador puntual de θ . Si el sesgo de $b(\hat{\theta}_n) \xrightarrow{n \rightarrow \infty} 0$ ($\theta = E(\hat{\theta}_n) \xrightarrow{n \rightarrow \infty} \theta$) y $V(\hat{\theta}_n) \xrightarrow{n \rightarrow \infty} 0$ entonces $\hat{\theta}_n$ es un estimador consistente de θ

La reciproca no es válida.



Se puede probar que S^2 es un estimador consistente de σ^2
 \bar{X} es un estimador consistente de μ .

Métodos para construir estimadores puntuales

Método de los momentos

Sea X_1, \dots, X_n muestra aleatoria de una r.v. X . Asumimos que se conoce la familia de distribuciones a la que pertenece X .

Sean $\theta_1, \dots, \theta_r$ los parámetros desconocidos de la distribución de X .

Definimos: $E(X^k)$ como el k -ésimo momento poblacional, donde $k=1, 2, 3, \dots, n$. Si $k=1$, entonces tenemos $E(X)$.

$\frac{1}{n} \sum_{i=1}^n X_i^k$ es el k -ésimo momento muestral con $k=1, 2, 3, \dots$

Si $k=1$, tenemos \bar{X}

Se puede probar que: $P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i^k - E(X^k)\right| > \epsilon\right) \rightarrow 0$ $n \rightarrow \infty$

Entonces el método de los momentos, para estimar $\theta_1, \theta_2, \dots, \theta_r$, consiste en plantear:

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n X_i = E(X) \\ \frac{1}{n} \sum_{i=1}^n X_i^2 = E(X^2) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n X_i^r = E(X^r) \end{cases} \quad \begin{array}{l} \text{se resuelve y encuentra estimadores} \\ \text{para } \theta_1, \theta_2, \dots, \theta_r (E(X), E(X^2), \dots, E(X^r)) \\ \text{quedan expresados en función de} \\ \theta_1, \theta_2, \dots, \theta_r \end{array}$$

Ej. Sea X_1, \dots, X_n muestra aleatoria de una r.v. X , donde $X \sim N(\mu, \sigma^2)$. Hallar estimaciones puntuales para μ y σ^2 por el método de los momentos.

$$\theta_1 = \mu, \quad \theta_2 = \sigma^2$$

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_i &= E(X) \quad \text{Sabiendo que } E(X) = \mu \quad \text{y } E(X^2) = V(X) + (E(X))^2 = \sigma^2 + \mu^2 \\ \frac{1}{n} \sum_{i=1}^n X_i^2 &= E(X^2) \quad \begin{cases} \frac{1}{n} \sum_{i=1}^n X_i = \mu & \text{①} \\ \frac{1}{n} \sum_{i=1}^n X_i^2 = \sigma^2 + \mu^2 & \text{②} \end{cases} \quad \begin{array}{l} \text{De ①} \rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \\ \text{De ②} \rightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i - \bar{X}^2 \end{array} \\ & \therefore \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \end{aligned}$$

Observación:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \underbrace{\left(\sum_{i=1}^n X_i^2 - n \bar{X}^2 \right)}_{\sum (X_i - \bar{X})^2} \quad \therefore \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Notar que

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \left(\frac{n-1}{n}\right) S^2 \\ E(\hat{\sigma}^2) &= E\left(\left(\frac{n-1}{n}\right) S^2\right) = \left(\frac{n-1}{n}\right) E(S^2) = \left(\frac{n-1}{n}\right) \sigma^2 \neq \sigma^2 \end{aligned}$$

Introducir el método de los momentos no produce, en general, estimadores insesgados.

Ej) Sea x_1, \dots, x_n muestra aleatoria de una r.v. X , $X \sim p(1)$. Hallar estimador puntual con el método de los momentos.

$$\theta_1 = \lambda$$

$$\frac{1}{n} \sum_{i=1}^n x_i = E(X) \rightarrow \frac{1}{n} \sum_{i=1}^n x_i = \lambda \therefore \boxed{\hat{\lambda} = \bar{x}}$$

Ej): Sea x_1, \dots, x_n una muestra aleatoria de una r.v. X , $X \sim N(0, \sigma^2)$
Hallar estimador de σ^2 por el método de los momentos

$$\frac{1}{n} \sum_{i=1}^n x_i = E(X) \rightarrow \frac{1}{n} \sum_{i=1}^n x_i = 0 \quad X \text{ no tiene variancia } \sigma^2$$

$$\frac{1}{n} \sum_{i=1}^n x_i^2 = E(X^2) = V(X) = \sigma^2 \therefore \boxed{\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2}$$

$$V(X) = E(X^2) - (E(X))^2$$

$$V(X) = E(X^2) - 0^2$$

Ej) Sea x_1, \dots, x_n muestra aleatoria de una r.v. X , donde la densidad de X está dada por:

$$f(x) = \begin{cases} \frac{\theta}{2^\theta} x^{\theta-1} & 0 < x < 2 \\ 0 & \text{cc} \end{cases} \quad \text{Hallar estimador puntual de } \theta \text{ por el método de los momentos.}$$

Densidad: Supongamos que para hallarla planteamos $\int_{-\infty}^{\infty} f(x) dx = 1$

$$1 = \int_0^2 \frac{\theta}{2^\theta} x^{\theta-1} dx = \frac{\theta}{2^\theta} \int_0^2 x^{\theta-1} dx = \frac{\theta}{2^\theta} \left(\frac{x^\theta}{\theta} \right) \Big|_0^2 = \frac{\theta}{2^\theta} \left(\frac{2^\theta}{\theta} - \frac{0^\theta}{\theta} \right) = \frac{\theta}{2^\theta} \cdot \frac{2^\theta}{\theta} = 1$$

$$\text{Planteamos: } \frac{1}{n} \sum_{i=1}^n x_i = E(X) \quad \textcircled{1}$$

$$\text{Calculamos } E(X) = \int_0^2 x \left(\frac{\theta}{2^\theta} x^{\theta-1} \right) dx = \frac{\theta}{2^\theta} \int_0^2 x^\theta dx = \frac{\theta}{2^\theta} \left(\frac{x^{\theta+1}}{\theta+1} \right) \Big|_0^2 \\ = \frac{\theta}{2^\theta} \left(\frac{2^{\theta+1}}{\theta+1} - \frac{0^{\theta+1}}{\theta+1} \right) = \frac{\theta}{2^\theta} \left(\frac{2^{\theta+1}}{\theta+1} - 0 \right) = \frac{2\theta}{\theta+1}$$

$$\textcircled{1} \quad \frac{1}{n} \sum_{i=1}^n x_i = \frac{2\theta}{\theta+1} \rightarrow (\theta+1)\bar{x} = 2\theta \rightarrow \theta\bar{x} + \bar{x} = 2\theta \rightarrow \bar{x} = 2\theta - \theta\bar{x}$$

$$\rightarrow \boxed{\theta = \frac{\bar{x}}{2-\bar{x}}}$$

Método de máxima verosimilitud

Sea x_1, \dots, x_n muestra aleatoria de una r.v. X . Asumimos conocida la familia de distribuciones a la que pertenece X .

Sea θ un parámetro desconocido de la dist. de X .

Supongamos que X es una r.v. discreta.

Supongamos que x_1, x_2, \dots, x_n es una observación de la muestra (x_1, x_2, \dots, x_n) respectivamente).

Planteamos $P(X_1=x_1, X_2=x_2, \dots, X_n=x_n)^{\text{indep}}$

$$= P(X_1=x_1) \cdot P(X_2=x_2) \cdot \dots \cdot P(X_n=x_n) = L(x_1, x_2, \dots, x_n, \theta) \text{ función de verosimilitud}$$

Notar que $P(X=x_i)$ puede escribirse en función de θ . El método de máxima verosimilitud consiste en hallar aquel valor de θ que maximiza la función de verosimilitud $L(x_1, x_2, \dots, x_n, \theta)$ (la cual es una probabilidad).

Ej) Sea X_1, X_2, \dots, X_n muestra de una r.v. $X, X \sim P(\lambda)$. Hallar el EMV de λ .

Planteamos la función de verosimilitud $L(x_1, x_2, \dots, x_n, \lambda) = P(X_1=x_1), P(X_2=x_2), \dots$

$$P(X_n=x_n) = \frac{e^{-\lambda} \lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda} \lambda^{x_2}}{x_2!} \cdots \frac{e^{-\lambda} \lambda^{x_n}}{x_n!}$$

$$= \left(e^{-\lambda} \right)^n \lambda^{\sum_{i=1}^n x_i} = \frac{e^{-\lambda n} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \quad \text{Para facilitar se toma } \ln(L(x_1, x_2, \dots, x_n, \lambda))$$

↓ es estrictamente creciente

$$\rightarrow \ln(L(x_1, x_2, \dots, x_n, \lambda)) = \ln \left(\frac{e^{-\lambda n} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \right) = \ln(e^{-\lambda n}) + \sum_{i=1}^n x_i \ln \lambda - \ln(\prod_{i=1}^n x_i!)$$

$$\rightarrow \ln(e^{-\lambda n}) + \ln(\lambda^{\sum_{i=1}^n x_i}) - \ln(\prod_{i=1}^n x_i!) = (-\lambda n) \cdot 1 + \sum_{i=1}^n x_i \ln \lambda - \ln(\prod_{i=1}^n x_i!)$$

$$\frac{d}{d\lambda} \ln L(x_1, x_2, \dots, x_n, \lambda) \rightarrow -n + \left(\sum_{i=1}^n x_i \right) \frac{1}{\lambda} = 0$$

$$\therefore n = \left(\sum_{i=1}^n x_i \right) \frac{1}{\lambda} \rightarrow \lambda = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \Rightarrow \text{se puede verificar que } L(x_1, x_2, \dots, x_n, \lambda) \text{ tiene en } \lambda = \bar{x} \text{ un punto absoluto}$$

$$\therefore \boxed{\lambda = \frac{1}{n} \sum_{i=1}^n x_i}$$

Ej): Sea X_1, X_2, \dots, X_n muestra aleatoria de una r.v. $X, X \sim B(1, p)$

Supongamos que interesa estimar la proporción de artículos defectuosos fabricados por una máquina en 1 día.

Tomaremos una muestra de artículos de toda la producción. Para cada artículo redefine $x_i = \begin{cases} 1 & \text{si el artículo es defectuoso} \\ 0 & \text{si no} \end{cases}$

$$\rightarrow X_i \sim B(1, p) \quad i = 1, 2, \dots, n$$

$$\text{Le quiere hallar EMV def. } P(X=x_i) = \binom{n}{x_i} p^{x_i} (1-p)^{1-x_i} = p^{x_i} (1-p)^{1-x_i}$$

$$\rightarrow L(x_1, x_2, \dots, x_n, p) = P(X_1=x_1, X_2=x_2, \dots, X_n=x_n)$$

$$= P(X_1=x_1) \cdot P(X_2=x_2) \cdots P(X_n=x_n)$$

$$= p^{x_1} (1-p)^{1-x_1} \cdot p^{x_2} (1-p)^{1-x_2} \cdots \cdot p^{x_n} (1-p)^{1-x_n}$$

$$\rightarrow = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}$$

$$\ln(L(x_1, x_2, \dots, x_n, p)) = \ln(p^{\sum_{i=1}^n x_i}) + \ln((1-p)^{n - \sum_{i=1}^n x_i}) = \sum_{i=1}^n x_i \cdot \ln(p) + (n - \sum_{i=1}^n x_i) \cdot \ln(1-p)$$

$$\frac{d}{dp} \ln(L(x_1, x_2, \dots, x_n, p)) = \left(\sum_{i=1}^n x_i \right) \frac{1}{p} + (n - \sum_{i=1}^n x_i) \frac{1}{1-p} = 0 \quad \text{regla de la cadena}$$

$$\therefore \left(\sum_{i=1}^n x_i \right) \frac{1}{p} = (n - \sum_{i=1}^n x_i) \frac{1}{1-p} \rightarrow \frac{1-p}{p} = \frac{n - \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i} \rightarrow \frac{1}{p} - 1 = \frac{n}{\sum_{i=1}^n x_i} - 1$$

$$\rightarrow \frac{1}{P} = \frac{n}{\sum x_i} \rightarrow P = \frac{\sum_{i=1}^n x_i}{n} = \bar{x} \rightarrow \hat{P} = \frac{1}{n} \sum_{i=1}^n x_i$$

E) b) i) Se seleccionan n=10 chips fabricados por una compañía. Si son defectuosos

$$y \sum_{i=1}^{100} x_i = 5 \therefore \text{la estimación de } P \text{ es } \bar{x} = \frac{1}{100} \sum_{i=1}^{100} x_i = \frac{5}{100} = 0,05$$

(caso continuo)

Si x_1, x_2, \dots, x_n es una muestra aleatoria de una va X continua con densidad $f(x)$ y si x_1, x_2, \dots, x_n es una medición de la muestra aleatoria, entonces se toma:

$L(x_1, x_2, \dots, x_n, \theta) = f(x_1) \cdot f(x_2) \cdots f(x_n)$. Se toma como estimación de θ a aquel valor que maximiza $L(x_1, x_2, \dots, x_n, \theta)$

Ej) Sea x_1, x_2, \dots, x_n muestra aleatoria de una va X , $X \sim \text{Exp}(\lambda)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \rightarrow L(x_1, x_2, \dots, x_n, \lambda) = f(x_1) \cdot f(x_2) \cdots f(x_n)$$

$$= \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \cdots \lambda e^{-\lambda x_n}$$

$$= \lambda^n e^{-\sum_{i=1}^n \lambda x_i} = \lambda^n e^{-\lambda \sum x_i}$$

$$\ln(L(x_1, x_2, \dots, x_n, \lambda)) = \ln(\lambda^n) + \ln(e^{-\lambda \sum x_i}) = n \ln(\lambda) + (-\lambda \sum_{i=1}^n x_i) \quad (n \in \mathbb{N})$$

$$= n \ln(\lambda) - \lambda \sum_{i=1}^n x_i \rightarrow \frac{d}{d\lambda} \ln(L(x_1, x_2, \dots, x_n, \lambda)) = 0$$

$$\rightarrow \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0 \rightarrow \lambda = \frac{n}{\sum_{i=1}^n x_i} \rightarrow \frac{\lambda}{\sum_{i=1}^n x_i} = 1 = \frac{1}{\bar{x}}$$

$$\therefore \hat{\lambda} = \frac{1}{\bar{x}}$$

Se determina la duración en hs de cada una de 10 lámparas donde los siguientes datos:

$$x_1 = 49,5 \quad x_2 = 101,5 \quad x_3 = 100,9,4 \quad x_4 = 173,4 \quad x_5 = 251,6$$

$$x_6 = 315 \quad x_7 = 130,1,4 \quad x_8 = 732,8 \quad x_9 = 660,6 \quad x_{10} = 1456,5$$

Se sospecha que la duración en hs tiene dist. exponencial.

a) Hallar el EMV del.

b) Hallar la estimación de λ para la muestra dada.

X_i : "tiempo de vida en horas de una lámpara i" $X_i \sim \text{Exp}(\lambda)$ (λ desconocida)
se toma una muestra de 10 lámparas

x_i : "duración en horas de la lámpara i" $i = 1, 2, \dots, 10$

x_1, x_2, \dots, x_{10} muestra aleatoria de la va X .

x_1, x_2, \dots, x_{10} es una medición de la muestra aleatoria.

$$a) \hat{\lambda} = \frac{1}{\bar{x}} \quad b) \lambda = \frac{1}{\bar{x}} = \frac{1}{696,52} \text{ es la estimación de } \lambda \text{ para la muestra dada}$$

g) sea X_1, X_2, \dots, X_n muestra aleatoria de una r.a.X $X \sim N(\mu, \sigma^2)$

Hallar EMV para μ y σ^2 $f(x) = \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

$$\begin{aligned} L(x_1, x_2, \dots, x_n, \mu, \sigma^2) &= f(x_1) \cdot f(x_2) \cdots f(x_n) \\ &= \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{1}{2\sigma^2}(x_1-\mu)^2} \cdot \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{1}{2\sigma^2}(x_2-\mu)^2} \cdots \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{1}{2\sigma^2}(x_n-\mu)^2} \\ &= \left(\frac{1}{\sqrt{\sigma^2 2\pi}} \right)^n e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2} = \frac{1}{(\sigma^2 2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2} \\ \ln(L(x_1, x_2, \dots, x_n, \mu, \sigma^2)) &= \ln\left(\frac{1}{(\sigma^2 2\pi)^{n/2}}\right) + \ln\left(e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2}\right) \\ &= -\frac{n}{2} \ln(\sigma^2 2\pi) - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \end{aligned}$$

$$\begin{cases} \frac{d}{d\mu} \ln(L(x_1, x_2, \dots, x_n, \mu, \sigma)) & \textcircled{1} \quad \textcircled{1} - \frac{1}{2\sigma^2} \sum 2(x_i - \mu) (-1) \\ \frac{d}{d\sigma} \ln(L(x_1, x_2, \dots, x_n, \mu, \sigma)) & \textcircled{2} \quad \textcircled{2} \left(\frac{n}{2} \right) \frac{1}{\sigma^2 2\pi} - \frac{1}{2} \sum (x_i - \mu)^2 \left(-\frac{1}{\sigma^4} \right) \end{cases}$$

$$\textcircled{1} \rightarrow \frac{1}{2\sigma^2} \sum 2(x_i - \mu) \cancel{(-1)} \rightarrow \frac{1}{\sigma^2} \sum (x_i - \mu) = 0$$

$$\rightarrow \sum (x_i - \mu) = 0 \rightarrow \sum x_i - \sum \mu = 0 \Rightarrow \sum x_i - n\mu = 0 \rightarrow \hat{\mu} = \frac{\sum x_i}{n} = \bar{x}$$

$$\textcircled{2} \left(-\frac{n}{2} \right) \frac{1}{\sigma^2 2\pi} - \frac{1}{2} \sum (x_i - \mu)^2 \left(-\frac{1}{\sigma^4} \right) \stackrel{0}{\Rightarrow} \frac{n}{\sigma^2} + \frac{1}{\sigma^4} \sum (x_i - \mu)^2 = 0$$

$$\rightarrow \frac{n}{\sigma^2} = \frac{1}{\sigma^4} \sum (x_i - \mu)^2 \rightarrow \frac{\sigma^4}{\sigma^2} = \frac{1}{n} \sum (x_i - \mu)^2 \Rightarrow \sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

$$\rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Importante recordar los resultados,

Propiedades de los EMV \rightarrow son válidas si la muestra es grande
Bajo condiciones muy generales, se puede probar:

① Los EMV son asintóticamente inseguros

② Son consistentes los EMV

③ Para muestras muy grandes, son asintóticamente de menor varianza

④ Propiedad de invariancia: sea $\hat{\theta}$ es el EMV de θ . Sea $g(x)$ una función inyectiva. Entonces el EMV de $g(\theta)$ es $g(\hat{\theta})$

En el ejercicio anterior se encontró el EMV de σ^2 . $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$

Si queremos hallar el EMV de σ , usaremos la propiedad de invariancia.

$$\sigma = \sqrt{\sigma^2} \rightarrow \hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$$

En el ej de la bombilla. b) ii) Se pide el EMY de $(1-p)^6$ $n=100$

$$\sum_{i=1}^{100} x_i = 5 \quad \hat{p} = \bar{x} \quad \text{La estimación para } p \text{ es } \hat{p} = \frac{5}{100} = 0,05$$

$p(p) = (1-p)^6$ (\circ sea x toman 6 chips y ninguno es defectuoso)
el EMY de $(1-p)^6 = g(\hat{p}) = g(\bar{x}) = (1-\bar{x})^6$ y la estimación
para $(1-p)^6$ es $(1-0,05)^6$

En el ej de los lámparas: X : "duración en horas de una lámpara"

$$X_1, X_2, \dots, X_n \text{ m.a. dex. el EMY de } \lambda \rightarrow \hat{\lambda} = \frac{1}{\bar{x}}$$

$$\text{Hallar el EMY de } P(X < 1400) = 1 - e^{-\lambda \cdot 1400}$$

$$\therefore \text{el EMY es } 1 - e^{-\hat{\lambda} \cdot 1400} = 1 - e^{-\hat{\lambda} \cdot 1400} = 1 - e^{-\frac{1}{\bar{x}} \cdot 1400}$$

Intervalos de confianza

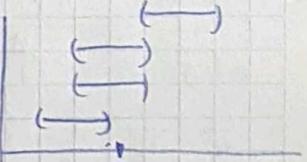
Sea X_1, X_2, \dots, X_n una muestra aleatoria de una r.a.X. Sea θ un parámetro desconocido de la distribución de X . Un intervalo de confianza para θ es un intervalo $(\hat{\theta}_1, \hat{\theta}_2)$ donde $\hat{\theta}_1$ y $\hat{\theta}_2$ son estadísticos. De manera tal que $P(\theta \in (\hat{\theta}_1, \hat{\theta}_2)) = 1 - \alpha$ con $1 - \alpha$ fijo de antemano. $(1 - \alpha) \cdot 100\%$ es el nivel de confianza del intervalo.

Si x_1, x_2, \dots, x_n es una medición de la muestra aleatoria entonces $\hat{\theta}_1$ tomará el valor $\hat{\theta}_1$ y $\hat{\theta}_2$ tomará el valor $\hat{\theta}_2$.

Así para la medición x_1, x_2, \dots, x_n tenemos un intervalo numérico $(\hat{\theta}_1, \hat{\theta}_2)$

Si x'_1, x'_2, \dots, x'_n es otra medición de la muestra aleatoria entonces obtenemos otro intervalo numérico $(\hat{\theta}'_1, \hat{\theta}'_2)$

Fix $\alpha = 0,95$, entonces si tenemos 100 mediciones de la m.a. Nuevas a tener 100 intervalos numéricos de los cuales queremos que 95% contenga a θ



A mayor nivel de confianza más precisión

La precisión de la estimación será mayor con un intervalo de longitud pequeña

→ Intervalos de confianza para la media de una normal con varianza conocida:

Sea x_1, x_2, \dots, x_n m.a. de una r.a.X donde $X \sim N(\mu, \sigma^2)$ con σ^2 conocida. Se quiere un intervalo de confianza para μ en nivel 95%.

Utilizaremos el método del punto

① Se toma un estimador puntual de μ $\hat{\mu} = \bar{x}$

② Se construye a partir de $\hat{\mu} = \bar{x}$ un estadístico

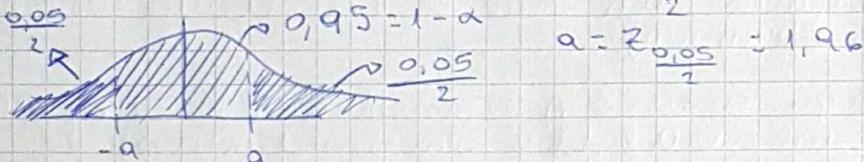
- debe contener a μ
- debe tener distribución conocida

consideramos el estadístico $Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$ notemos que $\bar{X} \sim N(\mu, \sigma^2/n)$ y entonces $Z \sim N(0, 1)$
además Z contiene ap. d^r q se le llama estadístico punto.

Si conocemos la distribución de Z podemos plantear: hallar a tal que $0,95 = P(-\alpha \leq Z \leq \alpha)$. Si $Z \sim N(0, 1)$:

$$P(-\alpha \leq Z \leq \alpha) = \Phi(\alpha) - \Phi(-\alpha) = 2\Phi(\alpha) - 1$$

$$\rightarrow 0,95 = 2\Phi(\alpha) - 1 \rightarrow \Phi(\alpha) = \frac{1 - 0,95}{2} = 0,975 \quad \therefore \boxed{\alpha = 1,96}$$



$$\rightarrow 0,95 = (-1,96 \leq Z \leq 1,96) = \left(-1,96 \leq \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \leq 1,96 \right)$$

$$= P\left(-1,96 \sqrt{\frac{\sigma^2}{n}} \leq \bar{X} - \mu \leq 1,96 \sqrt{\frac{\sigma^2}{n}}\right) = P\left(-\bar{X} - 1,96 \sqrt{\frac{\sigma^2}{n}} \leq -\mu \leq -\bar{X} + 1,96 \sqrt{\frac{\sigma^2}{n}}\right)$$

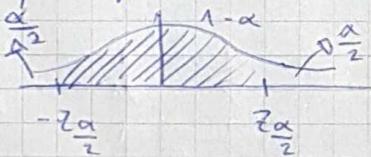
$$= P\left(\bar{X} + 1,96 \sqrt{\frac{\sigma^2}{n}} \geq \mu \geq -1,96 \sqrt{\frac{\sigma^2}{n}} + \bar{X}\right) = P\left(\bar{X} - 1,96 \sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X} + 1,96 \sqrt{\frac{\sigma^2}{n}}\right)$$

$$= P\left(\mu \in \left(\bar{X} - 1,96 \sqrt{\frac{\sigma^2}{n}}, \bar{X} + 1,96 \sqrt{\frac{\sigma^2}{n}}\right)\right) \quad \therefore \left(\bar{X} - \frac{z_{0,05}}{2} \sqrt{\frac{\sigma^2}{n}}, \bar{X} + \frac{z_{0,05}}{2} \sqrt{\frac{\sigma^2}{n}}\right)$$

En general, un intervalo de confianza para μ de nivel $(1-\alpha)100\%$

será:

$$\left(\bar{X} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}, \bar{X} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}\right)$$



$z_{\frac{\alpha}{2}}$ → valor crítico

Se usan tablas de valores críticos → del área a la derecha (el complemento) $P(X \geq x)$

$\frac{\alpha}{2}$

Ejemplo. Se ve que el límite de elasticidad de una barra de acero pertenece al barro de acero esta normalmente distribuida con $\sigma = 100$. Se toman 25 barras de acero, cada una se mide el límite de elasticidad y resulta $\bar{x} = 8439/16$. Cálculo un intervalo de confianza de nivel 90% para la verdadera media de límite de elasticidad.

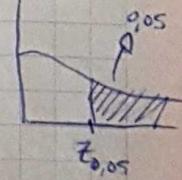
X_i : "límite de elasticidad de una barra de acero" $X_i \sim N(\mu, \sigma^2)$ $\sigma = 100$

X_i : "límite de elasticidad de barra de acero i" $i = 1, \dots, 25$

x_1, x_2, \dots, x_{25} M.A. de la r.a X. $\bar{x} = 8439/16$

$$1 - \alpha = 0,9 \rightarrow \alpha = 0,1 \rightarrow \frac{\alpha}{2} = 0,05 \quad \left(\bar{X} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}, \bar{X} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}\right)$$

$$z_{0,05} = 1,65$$



$$\rightarrow \left(8439 - 1,65 \sqrt{\frac{100^2}{25}}, 8439 + 1,65 \sqrt{\frac{100^2}{25}}\right) = (8406, 8472)$$

$$L = 8472 - 8406 = 66$$

$$\text{Anotamos } 1-\alpha = 0,92 \quad \alpha = 0,08 \quad \frac{\alpha}{2} = 0,04 \quad Z_{0,04} = 1,76$$

$$(8439 - 1,76 \sqrt{\frac{100}{n}}, 8439 + 1,76 \sqrt{\frac{100}{n}}) = (8403,8, 8474,2)$$

$L = 70,4 \rightarrow$ aumentó la longitud

Observación: ① $(\bar{X} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}, \bar{X} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}})$ está centrado en \bar{X}

$$\text{② La longitud } L = \bar{X} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} - (\bar{X} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}) = \boxed{2 z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}}$$

Notar que: si dejamos n fijo y aumenta $1-\alpha$ (disminuye α), entonces $z_{\frac{\alpha}{2}}$ aumenta $\therefore L$ aumenta

Si $1-\alpha$ se deja fijo y aumenta n entonces L disminuye (aumenta la precisión de la estimación)

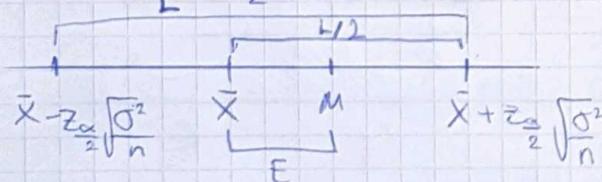
Se puede plantear que tomando n demuestra \forall necesita para que la longitud L sea menor a l . (l se fija de antemano)

$$L = 2 z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \leq l \rightarrow \left(\frac{2 z_{\frac{\alpha}{2}} \sigma}{l} \right)^2 \leq n$$

y: que tomando n demuestra se necesita para que el intervalo de nivel 90% sea menor a la mitad.

$$L = 2 \cdot 1,65 \sqrt{\frac{100}{n}} \leq 33 \rightarrow \left(\frac{2 \cdot 1,65 \cdot 100}{33} \right)^2 \leq n \quad \therefore \boxed{n \geq 100}$$

Si se estima puntualmente μ con \bar{X} , se comete un error E , y nota que ese $E \leq \frac{L}{2}$ con una confianza de $(1-\alpha)100\%$.



Sea X_1, X_2, \dots, X_n m.a de una ra. X , $n \geq 30$. Anotamos $\mu = E(X)$, $\sigma^2 = V(X)$. Se quiere un intervalo de confianza para μ de nivel $(1-\alpha)100\%$.

Se usa el método del punto: se toma un intervalo puntualmente dep. A partir de $\bar{M} = \bar{X}$ se construye un estadístico tal que:

a) debe contener μ

b) debe tener distribución conocida

caso 1: se conoce σ^2
Hollman: Comemos el estadístico $Z = \frac{\bar{X} - M}{\sigma/\sqrt{n}}$ $Z \sim N(0,1)$ por TCL.

Se plantea hallar a tal que

$$1-\alpha = P(-a \leq Z \leq a) \approx \Phi(a) - \Phi(-a) = 2\Phi(a) - 1$$

$\therefore a = z_{\frac{\alpha}{2}}$. Con un planteo análogo al anterior, se llega al intervalo:

$(\bar{X} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}, \bar{X} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}})$. La diferencia es que el nivel de confianza es aproximadamente $(1-\alpha) \cdot 105\%$.

caso 2: no se conoce σ^2 . Se reemplaza σ^2 por $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
 $Z = \frac{\bar{x} - \mu}{\sqrt{\frac{S^2}{n}}}$. Se puede probar que $Z \sim N(0,1)$ por T.C.L. y tenemos

se puede probar que S^2 es estimador consistente de σ^2
con un punto análogo al caso 1, se llega a:

$(\bar{x} - z_{\alpha/2} \frac{\sqrt{S^2}}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sqrt{S^2}}{\sqrt{n}})$ con nivel $(1-\alpha) \cdot 100\%$ aproximadamente

Ej: Se registraron los tiempos utilizados en la compra para 64 clientes seleccionados de azor en un supermercado. La media y la variancia de los 64 tiempos fueron 33 minutos y 256 minutos² respectivamente. Estimar la media real μ del tiempo utilizado en la compra con un intervalo de confianza de nivel 90%.

X: "tiempo de compra en minutos" de un cliente tomado al azar

X_i : "tiempo de compra del cliente i" ($i=1, \dots, 64$) $n=64$

X_1, X_2, \dots, X_{64} m.a. de l.v.a. X

$$\bar{x} = \frac{1}{64} \sum_{i=1}^{64} x_i = 33 \quad S^2 = \frac{1}{64-1} \sum_{i=1}^{64} (x_i - \bar{x})^2 = 256 \quad 1-\alpha = 0,9 \quad \alpha = 0,1 \\ \frac{\alpha}{2} = 0,05$$

$$(\bar{x} - z_{\alpha/2} \frac{\sqrt{S^2}}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sqrt{S^2}}{\sqrt{n}}) \quad z_{\alpha/2} = z_{0,05} = 1,645$$

$$(33 - 1,645 \sqrt{\frac{256}{64}}, 33 + 1,645 \sqrt{\frac{256}{64}}) = (29,71; 36,29) \text{ nivel de confianza } 90\% \text{ aproximadamente}$$

Intervalo de confianza para la media de una muestra con variancia desconocida

Sea X_1, X_2, \dots, X_n muestra aleatoria de una ra X, $X \sim (\mu, \sigma^2)$, σ^2 desconocido.

Se quiere un intervalo de confianza para μ de nivel $(1-\alpha) \cdot 100\%$.

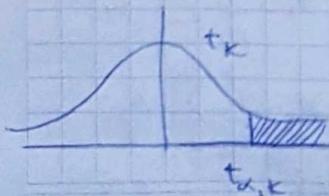
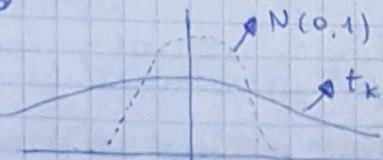
Utilizamos el método de punto. Tomamos el estadístico $T = \frac{\bar{x} - \mu}{\sqrt{\frac{S^2}{n}}}$. Se puede probar que T tiene una distribución llamada "student" con parámetros $n-1$.

Observación: sea T una ra continua. Sea $K \in \mathbb{N}$. T tiene distribución student con K grados de libertad (parámetro K), su densidad es:

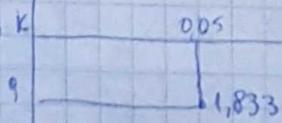
$$f(t) = \frac{1}{\sqrt{\pi K} \Gamma\left(\frac{K}{2}\right)} \cdot \frac{1}{\left[\left(\frac{t^2}{K} + 1\right)^{\frac{K+1}{2}}\right]} \quad -\infty < t < \infty$$

Notación: $T \sim t_K$

$N(0,1)$



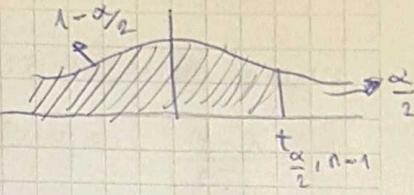
Ejemplo: $t_{0,05}, q=1,833$



T simétrica

Se plantea hallar a tal que $1-\alpha = P(-a \leq T \leq a) = F(a) - F(-a)$

$$\therefore 1-\alpha = 2F(a)-1 \rightarrow F(a) = \frac{1-\alpha}{2} = \frac{1-\frac{\alpha}{2}}{2}$$



Entonces:

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

Despejando μ se llega a: $(\bar{X} - t_{\alpha/2, n-1} \sqrt{\frac{s^2}{n}}, \bar{X} + t_{\alpha/2, n-1} \sqrt{\frac{s^2}{n}})$

Ej. Un fabricante de pelotas desarrolló una nueva fórmula que se puso con 8 granadas. Las velocidades iniciales resultantes en m/s fueron 3005, 2025, 2935, 2965, 2995, 3005, 2937, 2905.

Hallar un intervalo de confianza para la media real de las velocidades iniciales para granadas de este tipo con nivel de confianza 95%. Suponer que las granadas siguen dist. normal

X_i : "vel. inicial de una granada tomada al azar con pelotón nuevo" $X_i \sim N(\mu, \sigma^2)$

X_1, X_2, \dots, X_8 ma. de la var. X . $1-\alpha = 0,95$

$$\bar{X} = 2959 \quad \sigma = 0,05 \quad \alpha = 0,025$$

$$(\bar{X} - t_{\alpha/2, n-1} \sqrt{\frac{s^2}{n}}, \bar{X} + t_{\alpha/2, n-1} \sqrt{\frac{s^2}{n}}) \quad S = 39,1 \quad t_{0,025, 7} = 2,365$$

$$(2959 - 2,365 \frac{39,1}{\sqrt{8}}, 2959 + 2,365 \frac{39,1}{\sqrt{8}}) = (2926,3; 2991,7)$$

Observación: ① ~~Si n es grande~~ En el caso del intervalo que usamos tienen, la longitud es $L = 2t_{\alpha/2, n-1} \sqrt{\frac{s^2}{n}}$ no se pide despejar el n porque no se pide. Aparece en muchos lados.

② En el caso del intervalo \hat{S} (en muestras grandes)

$$(\bar{X} - z_{\alpha/2} \sqrt{\frac{s^2}{n}}, \bar{X} + z_{\alpha/2} \sqrt{\frac{s^2}{n}}) \quad L = 2z_{\alpha/2} \sqrt{\frac{s^2}{n}}$$

En este caso si se pide plantear n tal que $L \leq l$ y se despeja n , porque s^2 es constante (\Rightarrow se toma como un número)

$$\left(\frac{2z_{\alpha/2} s}{l} \right)^2 \leq n$$

Intervalos de confianza para diferencia de medias

Sea $X_{11}, X_{12}, \dots, X_{1n_1}$ ma de la var. X_1 . X_1 y X_2 son indep.
y $X_{21}, X_{22}, \dots, X_{2n_2}$ ma de la var. X_2 .

σ_1^2 y σ_2^2 conocidos. Se quiere un intervalo de confianza para $\mu_1 - \mu_2$ de nivel $(1-\alpha)100\%$.

Por el método del punto: ① $\widehat{\mu_1 - \mu_2} = \bar{X}_1 - \bar{X}_2$

② Considerando $Z = \bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)$

X_1, X_2 indep.

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$NOV(\bar{X}_1 - \bar{X}_2) = V(\bar{X}_1) + V(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\begin{cases} \bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i} \\ \bar{X}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} X_{2i} \end{cases}$$

$$E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2$$

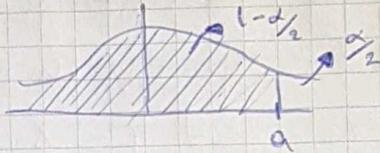
Con $\bar{X}_1 \sim N(\mu_1, \sigma_1^2)$ $\bar{X}_2 \sim N(\mu_2, \sigma_2^2)$ \bar{X}_1 y \bar{X}_2 indep. entonces:

$$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) \rightarrow Z \sim N(0, 1)$$

Le plantea hallar a a tal que $1 - \alpha = P(-a \leq Z \leq a) = \Phi(a) - \Phi(-a) = 2\Phi(a) - 1$

$$\rightarrow \Phi(a) = \frac{1-\alpha}{2} = 1 - \frac{\alpha}{2}$$

$$a = z_{\frac{\alpha}{2}}$$



$$\therefore 1 - \alpha = P\left(-\frac{z_{\alpha/2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq \bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2) \leq \frac{z_{\alpha/2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right)$$

Se despeja $\mu_1 - \mu_2$ y se llega al intervalo:

$$\left(\bar{X}_1 - \bar{X}_2 - \frac{z_{\alpha/2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \bar{X}_1 - \bar{X}_2 + \frac{z_{\alpha/2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right)$$

Ej: X_{1i} : "volumen de llenado en oz. en botella i llenada por máquina 1"
 $i=1, \dots, 12$ $n_1=12$ $\bar{x}_1 = 30,87$ oz.

X_{2i} : "volumen de llenado en oz en botella i llenada por máquina 2"
 $i=1, \dots, 10$ $n_2=10$ $\bar{x}_2 = 30,68$ oz.

$X_{11}, X_{12}, \dots, X_{1n_1}$ m/a de la r.a \bar{X}_1 $\bar{X}_1 \sim N(\mu_1, \sigma_1^2)$

$X_{21}, X_{22}, \dots, X_{2n_2}$ m/a de la r.a \bar{X}_2 $\bar{X}_2 \sim N(\mu_2, \sigma_2^2)$

$$1 - \alpha = 0,9 \quad \alpha = 0,1 \quad \alpha = 0,05$$

$$\frac{z_{\alpha/2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = z_{0,05} = 1,65$$

$$(30,87 - 30,68 - 1,65 \sqrt{\frac{0,10^2}{12} + \frac{0,15^2}{10}}, 30,87 - 30,68 + 1,65 \sqrt{\frac{0,10^2}{12} + \frac{0,15^2}{10}})$$

$$= (0,09837; 0,281620)$$

Observación: 1) Podemos decir que $\mu_1 - \mu_2 > 0$ con una seguridad de 90%
 es decir $\mu_1 > \mu_2$ con una confianza de 90%.

Si el intervalo incluye a cero, no se podría afirmar si $\mu_1 > \mu_2$ o no.

$$\textcircled{1} \quad L = 2 \frac{z_{\alpha/2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Le puede plantear que tamaño $n=n_1=n_2$
 se necesita para que $L \leq l$.

$$\rightarrow 2 \frac{z_{\alpha/2}}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n}}} \leq l \rightarrow \left(\frac{2 z_{\alpha/2}}{l}\right)^2 \left(\frac{\sigma_1^2 + \sigma_2^2}{n}\right) \leq 1$$

Le repite el caso considerando que X_1 y X_2 son normales, y $n_1 \geq 30$ y $n_2 \geq 30$

$$\rightarrow E(X_1) = \mu_1 \quad y \quad E(X_2) = \mu_2 \quad \sigma_1^2 = V(X_1), \quad \sigma_2^2 = V(X_2) \quad \text{desconocido}$$

Le quiere un intervalo de confianza para $\mu_1 - \mu_2$ de nivel $(1-\alpha) 100\%$

$$\textcircled{1} \quad \widehat{\mu_1 - \mu_2} = \bar{X}_1 - \bar{X}_2$$

$$\textcircled{2} \quad Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$s_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_{1i} - \bar{X}_1)^2$$

$$s_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (x_{2i} - \bar{X}_2)^2$$

Se puede probar que $Z \sim N(0,1)$ por TCI y por ser S_1^2 y S_2^2 estimadores consistentes de σ_1^2 y σ_2^2 respectivamente. Se llega al intervalo:

$$\left(\bar{X}_1 - \bar{X}_2 - \frac{z_{\alpha/2}}{2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \bar{X}_1 - \bar{X}_2 + \frac{z_{\alpha/2}}{2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right) \text{ con nivel } (1-\alpha) 100\%.$$

Se repite el caso sabiendo que X_1 y X_2 son normales, pero σ_1 y σ_2 desconocidos.

Se quiere un intervalo de confianza para $\mu_1 - \mu_2$ de nivel $(1-\alpha) 100\%$.

Caso ①: $\sigma_1^2 = \sigma_2^2 = \sigma^2 \rightarrow$ por alguna característica del experimento, etc.

$$1) \widehat{\mu_1 - \mu_2} = \bar{X}_1 - \bar{X}_2 \quad 2) T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \text{ con } n_1+n_2-2 \text{ grados de libertad.}$$

$$\sigma^2 = S_p^2 = \frac{(n_1-1) S_1^2 + (n_2-1) S_2^2}{n_1+n_2-2}$$

$$E(S_p^2) = \sigma^2$$

Se puede probar que $T \sim t_{n_1+n_2-2}$. Se llega al intervalo:

$$\left(\bar{X}_1 - \bar{X}_2 - t_{\frac{\alpha}{2}, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{X}_1 - \bar{X}_2 + t_{\frac{\alpha}{2}, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

Caso ②: $\sigma_1^2 \neq \sigma_2^2$.

$$T^* = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \text{ se puede probar } T^* \approx t_r$$

$$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$V = \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)$$

→ se aproxima al intervalo cercano

$$\frac{\left(\frac{S_1^2}{n_1} \right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2} \right)^2}{n_2-1}$$

$$\left(\bar{X}_1 - \bar{X}_2 - t_{\frac{\alpha}{2}, V} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \bar{X}_1 - \bar{X}_2 + t_{\frac{\alpha}{2}, V} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right)$$

Ejemplo: para comparar 2 tipos de paragolpes, 6 de cada tipo se montaron en un automóvil compuesto y cada auto se estrelló contra un muro de concreto a 5 M/H. Los siguientes son los costos de reparación.

paragolpes 1	407	448	423	465	402	419	$n_1=6$
--------------	-----	-----	-----	-----	-----	-----	---------

paragolpes 2	434	415	412	451	433	429	$n_2=6$
--------------	-----	-----	-----	-----	-----	-----	---------

Hallar intervalos de confianza para diferencia de medios con confianza 99%. Suponiendo que la muestra tienen dist normales y son indep.

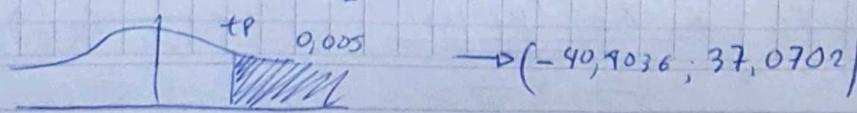
$X_{1,i}$: "costo de rep. de auto i" $i=1, 2, \dots, 6$ $X_i \sim N(\mu_1, \sigma_1^2)$

$X_{2,i}$: "costo de reparación de auto i" $i=1, 2, \dots, 6$ $X_2 \sim N(\mu_2, \sigma_2^2)$

$$1 - \alpha = 99\% \quad \alpha = 0,01 \quad \bar{x}_1 = 427,331 \quad \bar{x}_2 = 429 \quad V = 8,03 \approx 8$$

$$\frac{\alpha}{2} = 0,005$$

$$t_{0,005, V} = 3,355$$



NOTA

$$t_{0,005, 8} = 3,355$$

Ej.: Los siguientes m.a. son predicciones de la capacidad predictiva de calor en millones de celulosas portonelada de experimentos de calor de 2 minas.

Mina 1	8260	8130	8350	8070	8390	$n_1 = 5$	
Mina 2	7950	7890	7800	8190	7920	7890	$n_2 = 6$

Hallar un intervalo de confianza de nivel 99% para $\mu_1 - \mu_2$. Asumir
tamaños iguales y que los datos provienen de poblaciones normales.

X_{i1}: "capacidad productiva de color de la muestra i; de la mina i"
 $i = 1, 2, 3, 4, 5 \quad X_{i1}, X_{i2}, \dots, X_{in_i} \text{ m.s. de la Y} \quad X_i \sim N(\mu_i, \sigma_i^2)$

$X_{21}, X_{22}, \dots, X_{2n_2}$ made la V_2 $X_i \sim N(\mu_1, \sigma_1^2)$ $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$1-\alpha = 99\% \quad \alpha = 0,01 \quad \frac{K}{2} = 0,005 \quad t_{0,005, 5+6-2} = t_{0,005, 9} = 3,25$$

$$\bar{v}_1 = 8230 \quad S_1^2 = 15750 \quad S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = 13066,7$$

$$\bar{x}_1 = 8230 \quad S_1^2 = 15750 \quad S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = 13066,7$$

$$\bar{v}_1 = 8230 \quad S_1^2 = 15750 \quad S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = 13066,7$$

$$\bar{x}_2 = 7940 \quad S_2^2 = 10920 \quad n_1 + n_2 - 2$$

SP = 114,31

⇒ $(65,0410; 514, 958)$ $M_1 > M_2$ con seguridad del 99%.

Il prend de la mimo { tiene manu concedad multitud suya

laminilla 2.

Intervalos de confianza para muestras ~~varias~~ pareadas

Suponemos (x_1, x_2, \dots, x_n) de forma x_1

x_1, x_2, \dots, x_n ma de una var.

$$\mu_1 = E(X_1) \quad \mu_2 = E(X_2) \quad \sigma_1^2 = V(X_1) \quad \sigma_2^2 = V(X_2) \text{ desconvexas}$$

Se quiere un intervalo de confianza para $\mu_1 - \mu_2$ de nivel $(1-\alpha)100\%$.

(X_{ii}, X_{ei}) son 2 mediciones diferentes sobre la misma unidad muestral i .

O sea, ahora X_1 y X_2 son indep. Consideraremos una tercera muestra:

$$D_1 = X_{11} - X_{21}, D_2 = X_{12} - X_{22}, \dots, D_n = X_{1n} - X_{2n}$$

Entonces D_1, D_2, \dots, D_n es una m.a. de la ra $D = X_1 - X_2$
 $\rightarrow E(D) = M_D = E(X_1 - X_2) = E(X_1) - E(X_2) = M_1 - M_2$

$$\sigma_D^2 = \text{Var}(D) = \text{Var}(X_1 - X_2) = \sigma_1^2 + \sigma_2^2 - 2 \text{Cov}(X_1, X_2) \quad \sigma_D^2 \text{ descreve o desvio}$$

Si $D = X_1 - X_2 \sim N(\mu_D, \sigma_D^2)$ entonces tomamos el estadístico p-valor.

$$T = \bar{D} - \mu_D$$

$$\sqrt{\frac{SD}{n}} \sim t_{n-1}$$

$$\text{Se llega al intervalo: } (\bar{D} - t_{\frac{\alpha}{2}, n-1} \sqrt{\frac{s_D^2}{n}}, \bar{D} + t_{\frac{\alpha}{2}, n-1} \sqrt{\frac{s_D^2}{n}})$$

Ej.: Los siguientes son los perdidas semanales promedio de horas hombre debidas a accidentes en 10 plantas industriales. Antes y después de poner en operación cierto programa de seguridad.

	1	2	3	4	5	6	7	8	9	10
Antes	45	73	46	124	33	57	83	34	26	17
Después	36	60	44	119	35	51	77	29	24	11
D	9	13	2	5	-2	6	6	5	2	6

Hallar un intervalo de confianza de 95% para $\mu_1 - \mu_2$ asumiendo que la diferencia de los datos es normal. Se puede deducir que el programa es efectivo?

X_{1i} : "perdida de horas hombre antes del programa en planta i" $i=1, \dots, 10$

X_{2i} : "perdida de horas hombre después del programa en planta i" $i=1, \dots, 10$

$$D = X_1 - X_2 \sim N(\mu_D, \sigma_D^2) \quad \mu_D = \mu_1 - \mu_2 \quad 1-\alpha = 0,95 \quad \alpha = 0,05$$

$$\bar{D} = 5,2 \quad S_D = 4,08 \quad \frac{\alpha}{2} = 0,025$$

$$\left(5,2 - 2,262 \left(\frac{4,08}{\sqrt{10}} \right); 5,2 + 2,262 \left(\frac{4,08}{\sqrt{10}} \right) \right) = (2,2815; 8,1189) \quad t_{0,025, 9} = 2,262$$

$\mu_1 - \mu_2 > 0$ con una confianza de 95%. \therefore es efectivo con una confianza de 95%.

Intervalos de confianza para la variancia de una normal

Sea X_1, X_2, \dots, X_n Ma de la va $X \sim N(\mu, \sigma^2)$ ambos parámetros desconocidos. Se quiere un intervalo de confianza para σ^2 de nivel $(1-\alpha)100\%$.

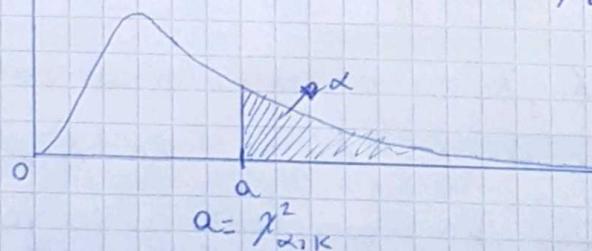
Método del cuadrado:

$$1) \sigma^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$2) X_0 = \frac{(n-1)S^2}{\sigma^2} \quad \text{Se puede probar que } X_0 \text{ tiene distribución chi cuadrado } \chi^2_{n-1} \text{ con parámetros (grados de libertad) } n-1.$$

Observación: Sea X una va continua y $K \in \mathbb{N}$. X tiene distribución χ^2 con K grados de libertad. Su densidad es:

$$f(x) = \frac{1}{2^K \Gamma(\frac{K}{2})} x^{\frac{K}{2}-1} e^{-\frac{x}{2}} \quad x > 0 \quad \text{Notación: } X \sim \chi^2_K$$

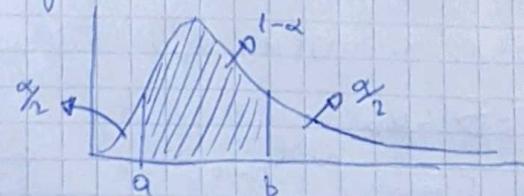


Si K más grande,
se achata la gráfica

Se plantea hallar a y b tal que: $1-\alpha = P(a \leq X_0 \leq b)$. Se puede probar que la mejor elección para a y b es:

$$b = \chi^2_{\frac{\alpha}{2}, n-1} \quad a = \chi^2_{1-\frac{\alpha}{2}, n-1}$$

Se encuentra con esta elección el intervalo con menor longitud



$$\begin{aligned} \therefore 1-\alpha &= P\left(\frac{\chi^2_{1-\frac{\alpha}{2}, n-1}}{(n-1)S^2} \leq \frac{\sigma^2}{(n-1)S^2} \leq \frac{\chi^2_{\frac{\alpha}{2}, n-1}}{(n-1)S^2}\right) = P\left(\frac{\chi^2_{1-\frac{\alpha}{2}, n-1}}{(n-1)S^2} \geq \frac{\sigma^2}{(n-1)S^2} \geq \frac{1}{\chi^2_{\frac{\alpha}{2}, n-1}}\right) \\ &= P\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \geq \sigma^2 \geq \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}}\right) = P\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right) \\ \therefore \left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right) &\rightarrow \text{Notar que: } \left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}}\right) \text{ es} \end{aligned}$$

un intervalo de confianza para σ de nivel $(1-\alpha) 100\%$.
desviación estandar

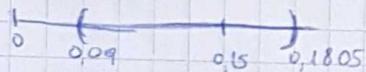
Ej.) X_i : volumen de llenado en onzas de una botella tomada al azar

$X \sim N(\mu, \sigma^2)$ X_i : "volumen de llenado de botella i " ($i=1, 2, \dots, 20$)

X_1, X_2, \dots, X_{20} ma de la ra X $S^2 = 0,0153$ $1-\alpha = 0,95$ $\frac{\alpha}{2} = 0,025$

$$\begin{aligned} \chi^2_{\frac{\alpha}{2}, 19} &= \chi^2_{0,025, 19} = 32,9 \rightarrow \left(\frac{(20-1)0,0153}{32,9}, \frac{(20-1)0,0153}{8,91}\right) \\ \chi^2_{1-\frac{\alpha}{2}, 19} &= \chi^2_{0,975, 19} = 8,91 \quad (0,00884; 0,0326) \end{aligned}$$

El intervalo para σ : $(\sqrt{0,00884}; \sqrt{0,0326}) = (0,09; 0,1805)$



Intervalo de confianza para cociente de varianzas.

Sea X_1, \dots, X_n ma de la X_1

$X_{2,1}, \dots, X_{n,1}$ ma de la X_2

$X_1, X_{2,1}, \dots, X_{n,1}$ indep. y normales con parámetros desconocidos.

Se requiere intervalo de confianza para $\frac{\sigma_1^2}{\sigma_2^2}$

El estadístico pivotal: $F = \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} = \frac{S_2^2}{S_1^2} \left(\frac{\sigma_1^2}{\sigma_2^2}\right)$ → Se pide

señalar que F tiene una distribución llamada Fisher ($\text{o } F$) con parámetros $n_2 - 1$ y $n_1 - 1$ (en ese orden).

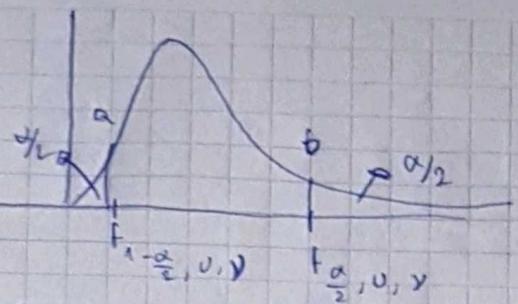
O F tiene dist. Fisher con $n_2 - 1$ grados de libertad en el numerador y $n_1 - 1$ grados de libertad en el denominador.

Observación: Sea X una ra continua. Sean U y V números

X tiene dist. F con U grados de libertad en numerador y V grados de libertad en denominador. X se dice que es:

$$F(x) = \frac{\Gamma(\frac{U+V}{2}) (\frac{U}{V})^{U/2}}{\Gamma(\frac{U}{2}) \Gamma(\frac{V}{2})} \cdot x^{\frac{U}{2}-1} \quad 0 < x < \infty$$

$$\Gamma(\frac{U}{2}) \Gamma(\frac{V}{2}) \left[\frac{U}{V} x + 1 \right]^{\frac{U+V}{2}} \quad \text{Notación } F \sim F_{U,V}$$



$$B = F_{\frac{\alpha}{2}, n_2-1, n_1-1}$$

$$a = F_{1-\frac{\alpha}{2}, n_2-1, n_1-1}$$

$$= \frac{F_{\frac{\alpha}{2}, n_2-1, n_1-1}}{F_{\frac{\alpha}{2}, n_2-1, n_1-1}}$$

Se llega al intervalo: $\left(\frac{s_1^2}{s_2^2} \cdot F_{1-\frac{\alpha}{2}, n_2-1, n_1-1}; \frac{s_1^2}{s_2^2} \cdot F_{\frac{\alpha}{2}, n_2-1, n_1-1} \right)$

$\frac{s_1^2}{s_2^2}$ se encuentra en este intervalo con una confianza $(1-\alpha) 100\%$.

Si λ no está en el intervalo, se puede decir que de los 2 errores grande
intervalos de confianza para una proporción

Supongamos que interesa estimar la proporción P de artículos
defectuosos que produce una máquina en 1 día.

Se toma una muestra de n artículos y se define para cada uno

$$X_i = \begin{cases} 1 & \text{si el art. } i \text{ es defectuoso} \\ 0 & \text{cc} \end{cases} \quad X_i \sim B(\lambda, P)$$

X_1, X_2, \dots, X_n muestreo de X , $X \sim B(n, P) \rightarrow \mu = E(X_i) = P$

Se quiere un intervalo de confianza para P de muestreo $\frac{\sigma^2}{n} = \sqrt{P(1-P)} = \frac{\sqrt{P(1-P)}}{(1-\alpha) 100\%}$
→ muestreo de porcentaje:

$$1) \hat{P} = \bar{X} \quad 2) Z = \frac{\bar{X} - P}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} = \frac{\hat{P} - P}{\sqrt{\hat{P}(1-\hat{P})}}$$

Le puede probar que $Z \sim N(0, 1)$
y por ser \hat{P} un estimador
consistente de P .
Le necesita una muestra grande
∴ se approxima de normal
a la binomial

$$\text{Observación: } Z = \frac{n(\bar{X} - P)}{\sqrt{n\hat{P}(1-\hat{P})}} = \frac{\sum_{i=1}^n X_i - np}{\sqrt{n\hat{P}(1-\hat{P})}}$$

∴ se debe cumplir la regla práctica: $n\hat{P} \geq 10$ y $n(1-\hat{P}) \geq 10$

$$\sum_{i=1}^n X_i \geq 10 \text{ y } n - \sum_{i=1}^n X_i \geq 10$$

Se llega al intervalo: $\left(\hat{P} - \frac{Z_{\alpha/2}}{\sqrt{n}} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}, \hat{P} + \frac{Z_{\alpha/2}}{\sqrt{n}} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \right)$

Observación: ① $L = 2 Z_{\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$. Se puede plantear hallar n
tal que $L \leq l$.

$$\rightarrow 2 Z_{\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq l \rightarrow \left(\frac{2 Z_{\alpha/2}}{l} \right)^2 \hat{P}(1-\hat{P}) \leq n$$

Se necesita una muestra preliminar para calcular \hat{P} . Si no se puede col-
cular \hat{P} entonces plantearnos

$$L = 2 Z_{\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq 2 Z_{\alpha/2} \sqrt{\frac{1}{n}} = \frac{Z_{\alpha/2}}{\sqrt{n}} \leq l \rightarrow \left(\frac{Z_{\alpha/2}}{l} \right)^2 \leq n$$

Ej) En una ma de 500 familias que tienen TV en cierta ciudad se encuentra que 340 están suscriptas a HBO.

a) Encuentre un intervalo de confianza de 95% para la proporción real de familias de dicha ciudad que están suscriptos a HBO.

$X_i \xrightarrow{\text{es}} \text{1 si la familia } i \text{ está suscripta}$

$$\begin{aligned} X_1, \dots, X_{500} &\text{ ma de la var } X \quad X \sim B(1, p) \quad 1 - \alpha = 0,95 \quad \alpha = 0,05 \\ \sum_{i=1}^{500} X_i &= 340 \quad \hat{p} = \frac{340}{500} = 0,68 \quad \rightarrow (0,68 - 1,96 \sqrt{0,68(1-0,68)}, 0,68 + 1,96 \sqrt{0,68(1-0,68)}) \\ n\hat{p} &= 340 \approx 10 \\ n(1-\hat{p}) &= 500 - 340 \approx 10 \end{aligned}$$

$$\rightarrow (0,64; 0,72)$$

b) ¿Qué tam grande se requiere que sea una muestra si queremos tener 95% de confianza de que la estimación de p esté dentro de 0,02

$$\frac{L}{2} \leq 0,02 \rightarrow L \leq 0,02 = 2$$

$$n \geq \left(\frac{2 \cdot 1,96}{0,04}\right)^2 \cdot 0,68(1-0,68) = 2089,8 \quad \therefore n \geq 2090$$

Observación: si n no tiene el valor $\hat{p} = 0,68$ entonces

$$n \geq \left(\frac{2 \cdot 1,96}{0,04}\right)^2 = \left(\frac{1,96}{0,04}\right)^2 = 2401 \quad \therefore n \geq 2401$$

Ley intervalo para diferencias de proporciones

Test de Hipótesis

Una hipótesis estadística es una afirmación sobre un parámetro determinado.

Ej) Se tiene un motor. Interesa la emisión de NO del motor. Se sabe que la emisión media de NO_2 es $\mu = 100 \text{ mg/seg}$

Se diseña un motor nuevo con el objetivo de disminuir μ .

Definimos X : "emisión de NO_2 en motor nuevo"

X_i : "emisión de NO_2 del motor i " ($i=1, 2, \dots, 50$)

X_1, X_2, \dots, X_{50} ma de la var X

$$\bar{x} = 92 \text{ mg/seg} \quad \sigma = \sqrt{V(x)} = 21 \text{ mg/seg}$$

Consideramos 2 hipótesis: $H_0: \mu = 100$, $H_1: \mu < 100$

hipótesis nula

$H_1: \mu < 100$

hipótesis alternativa

Se construye a partir de la ma un estadístico de prueba y se rechaza o no H_0 de acuerdo a la evidencia en contra de H_0 y se toma una decisión: rechazar H_0 (aceptar H_1) o aceptar H_0 (rechazar H_1).

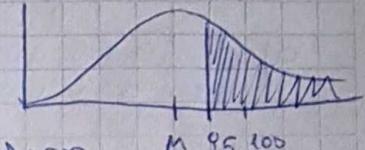
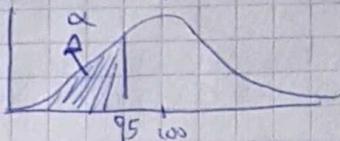
En el ej. el estadístico de prueba es \bar{X} . Si \bar{X} tiene un valor "malo" menor que 100, entonces se tiene evidencia en contra de H_0 y se lo rechaza. Si \bar{X} no es mucho menor que 100 entonces no se tiene evidencia en contra de H_0 y entonces se lo acepta.

Supongamos que tenemos el test

rechazar H_0 si $\bar{X} < 95$	$(-\infty; 95)$ zona de rechazo
aceptar H_0 si $\bar{X} \geq 95$	$(95; +\infty)$ "descartación"

95 es el punto crítico. En el ej. el valor que toma $\bar{x} = 92$ \therefore como el test dañado, se rechaza H_0 .

$$\bar{X} = \frac{1}{50} \sum_{i=1}^{50} X_i \approx N(\mu, \frac{\sigma^2}{50})$$



error de tipo I: rechazar H_0 cuando H_0 es verdadera.

$\alpha = P(\text{error de tipo I})$ es el nivel de significancia del test.

error de tipo II: aceptar H_0 cuando H_0 es falsa.

$\beta = P(\text{error de tipo II})$.

En el ejemplo de los motores: $\alpha = P(\text{rechazar } H_0 / H_0 \text{ es V}) = P(\bar{X} < 95 / H_0 \text{ es V})$

$$= P\left(\frac{\bar{X} - 100}{\sqrt{\frac{21^2}{50}}} < \frac{95 - 100}{\sqrt{\frac{21^2}{50}}}\right) \approx \Phi\left(\frac{95 - 100}{\sqrt{\frac{21^2}{50}}}\right) = 0,046$$

Si el test fuera

rechazar H_0 si $\bar{X} < 93$	$\alpha \approx \Phi\left(\frac{93 - 100}{\sqrt{\frac{21^2}{50}}}\right) = 0,00939$
aceptar H_0 si $\bar{X} \geq 93$	

$$\beta = P(\text{aceptar } H_0 / H_0 \text{ es F}) = P(\bar{X} \geq 95 / H_0 \text{ es F}) = P\left(\frac{\bar{X} - M}{\sqrt{\frac{21^2}{50}}} \geq \frac{95 - M}{\sqrt{\frac{21^2}{50}}}\right) \approx$$

$$\approx 1 - \Phi\left(\frac{95 - M}{\sqrt{\frac{21^2}{50}}}\right) = \beta(\mu). \text{ Si suponemos que } \mu = 94 \text{ seguimos obteniendo}$$

como se comporta el test: $\beta(94) \approx 1 - \Phi\left(\frac{95 - 94}{\sqrt{\frac{21^2}{50}}}\right) = 0,3707$

$$\beta(94,7) \approx 0,46017 \quad \beta(90) \approx 0,04648$$