Estadística

Estudia las técnicas para recolectar, resumir y ordenar datos y 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 anotamos $x_1, x_2, \dots, x_n$.

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

Muestra → una parte de la población. La muestra se toma al azar. (se eligen de la población n alumnos al azar como si se eligieran bolillas de una urna con N bolillas). 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, esto es, estudia y ordena la muestra.

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

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

Entonces al tener una muestra, no obtendremos el verdadero valor de μ pues tenemos 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 va 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 va 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 va 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 de μ esté lo más cerca posible al verdadero valor de μ.

Si tomara otra muestra de n alumnos, y obtenemos otros valores $x_1', x_2', \dots, x_n'$ y entonces tendremos otra estimación $\bar{x}'$ para μ

Podemos considerar que $\bar{x}$ y $\bar{x}'$ son dos valores diferentes que tomó la va $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ donde ahora $X_1, X_2, \ldots, X_n$ son las variables de una muestra aleatoria.

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

En general si tenemos una va que es una función de la muestra aleatoria, se dice que esa va 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 $\bar{x}'$ (valores que tomó $\bar{X}$) son estimaciones puntuales de μ.

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

Podemos considerar 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:

$$\hat{\mu} = \bar{X} \to \bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_n$$

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. (Cualquiera sea el valor de θ) → el valor verdadero

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

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

$$\to \frac{1}{n} \cdot n\mu = \boxed{\mu = E(\bar{X})}$$ cualquiera sea el valor de μ.

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

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

$\hat{\sigma}^2 = S^2$ es un estimador insesgado de σ^2 pues. Notar que

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

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

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

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

$$V(X) = E(X^2) - (E(X))^2$$
$$E(X^2) = V(X) + (E(X))^2 = \sigma^2 + \mu^2$$

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

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

Observaciones

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

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

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

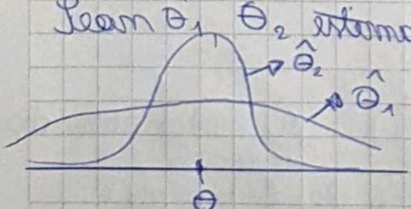
$$= \frac{\sigma^2}{n} + \mu^2 - \mu^2 = \boxed{\frac{\sigma^2}{n}}$$

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

En general: $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ no es un estimador insesgado de σ.
Lo sería si viene 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 insesgado de θ

Sean $\hat{\theta}_1$, $\hat{\theta}_2$ estimadores insesgados de θ. O sea $E(\hat{\theta}_1) = \theta$ y $E(\hat{\theta}_2) = \theta$



Consideramos mejor estimador puntual de θ a aquel 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 v.a. 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 $\hat{\mu}_2 = \frac{X_1 + X_7}{2}$

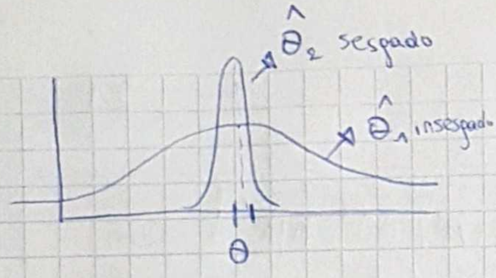
$E(\hat{\mu}_1) = E(\bar{X}) = \mu$ $\quad E(\hat{\mu}_2) = E\left(\frac{X_1+X_7}{2}\right) = \frac{1}{2}\left(E(X_1) + E(X_7)\right) = \frac{1}{2} 2\mu = \mu$

Ambos insesgados, se ven las varianzas.

$V(\hat{\mu}_1) = V(\bar{X}) = \boxed{\frac{\sigma^2}{7}}$ $\quad V(\hat{\mu}_2) = V\left(\frac{X_1+X_7}{2}\right) \underset{\text{indep.}}{=} \frac{1}{4}\left(V(X_1) + V(X_7)\right)$

$= \frac{1}{4}(\sigma^2 + \sigma^2) = \boxed{\frac{\sigma^2}{2}}$

Como $V(\hat{\mu}_1) < V(\hat{\mu}_2)$, se considera que $\hat{\mu}_1$ es mejor estimador puntual que $\hat{\mu}_2$



Error cuadrático medio: Sea $\hat{\theta}$ un estimador puntual de un parámetro θ. El ~~error~~

$$ECM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

Observaciones:

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

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 aquel que tenga menor ECM

Ej: Sea $X_1, X_2, \ldots, X_7$ muestra aleatoria de una va X con $E(X) = \mu$ y $\sigma^2 = V(X)$.
Consideramos $\hat{\mu}_1 = \frac{1}{7}\sum_{i=1}^{7} X_i = \bar{X}$ $\qquad \hat{\mu}_3 = \frac{2X_1 - X_7 + X_3}{3}$

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

$E(\hat{\mu}_1) = E(\bar{X}) = \mu$ $\qquad \therefore b(\hat{\mu}_1) = 0$
$V(\hat{\mu}_1) = \frac{\sigma^2}{7}$ $\rightarrow ECM(\hat{\mu}_1) = V(\hat{\mu}_1)^2 + 0 = \frac{\sigma^2}{7}$

$$E(\hat{\mu}_3) = E\left(\frac{2X_1 - X_7 + X_3}{3}\right) = \frac{1}{3}\left(2E(X_1) - E(X_7) + E(X_3)\right)$$
$$= \frac{1}{3}(2\mu - \mu + \mu) = \frac{2}{3}\mu \rightarrow b(\hat{\mu}_3) = E(\hat{\mu}_3) - \mu = \frac{2}{3}\mu - \mu = \boxed{-\frac{\mu}{3}}$$

$$V(\hat{\mu}_3) = V\left(\frac{2X_1 - X_7 + X_3}{3}\right) = \left(\frac{1}{3}\right)^2\left(4V(X_1) + V(X_7) + V(X_3)\right) = \frac{1}{9}(4\sigma^2 + \sigma^2 + \sigma^2)$$
$$= \frac{2}{3}\sigma^2$$

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

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

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

Estimadores consistentes: Sea $X_1, X_2, \ldots X_n$ 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
si $P(|\hat{\theta}_n - \theta| > \varepsilon) \xrightarrow[n\to\infty]{} 0 \quad \forall \varepsilon > 0$

ó $P(|\hat{\theta}_n - \theta| \leq \varepsilon) \xrightarrow[n\to\infty]{} 1 \quad \forall \varepsilon > 0$

$\hat{\theta}_n$ · $\theta - \varepsilon$ · θ · $\theta + \varepsilon$ · $\varepsilon > 0$

Se puede probar el sig. teorema:

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

La recíproca no es válida

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

Metodos para construir estimadores puntuales

Metodo de los momentos

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

Sean $\theta_1 \dots \theta_r$ r parametros desconocidos de la distribución de X.

Definimos: $E(X^k)$ como el k-ésimo momento poblacional, donde $k=1,2,3,\dots$
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: $P\left(\left|\frac{1}{n}\sum_{i=1}^{k} X_i^k - E(X^k)\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0$

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

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} x_i = E(x) \\ \frac{1}{n}\sum_{i=1}^{n} x_i^2 = E(x^2) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_i^r = E(x^r) \end{cases}$$

queda un sist de r ecuaciones con r incógnitas

→ Se resuelve y encuentra estimadores para $\theta_1, \theta_2, \dots, \theta_r$ ($E(x), E(x^2), \dots, E(x^r)$ quedan expresados en función de $\theta_1, \theta_2, \dots, \theta_r$)

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

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

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} X_i = E(X) \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = E(X^2) \end{cases}$$

Sabemos que $E(X) = \mu$ y $E(X^2) = V(X) + (E(X))^2 = \sigma^2 + \mu^2$

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} x_i = \mu & (1) \\ \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \sigma^2 + \mu^2 & (2) \end{cases}$$

De (1) → $\boxed{\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}}$

De (2) → $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu^2 = \frac{1}{n}\sum_{i=1}^{n} x_i - \bar{X}^2$

$$\boxed{\therefore \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{X}^2}$$

Observación:

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

Notar que

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

$$E(\hat{\sigma}^2) = E\left(\left(\frac{n-1}{n}\right)S^2\right) = \left(\frac{n-1}{n}\right)E(S^2) = \left(\frac{n-1}{n}\right)\sigma^2 \neq \sigma^2$$

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

Ej) Sea $X_1, \dots X_n$ muestra aleatoria de una v.a X, $X \sim P(\lambda)$. Hallar estimador puntual con el método de los momentos.

$\theta_1 = \lambda$

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

Ej): Sea $X_1, \dots X_n$ una muestra aleatoria de una v.a X, $X \sim N(0, \sigma^2)$. Hallar un estimador de σ^2 por el método de los momentos.

$$\frac{1}{n}\sum_{i=1}^{n} x_i = E(X) \rightarrow \frac{1}{n}\sum_{i=1}^{n} x_i = 0$$ X no sirve porque no aparece σ^2

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

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

Ej) Sea $X_1, \dots, X_n$ muestra aleatoria de una v.a X, donde la densidad de X está dada por:

$$f(x) = \begin{cases} \frac{\theta}{2^\theta} x^{\theta-1} & 0 < x < 2 \\ 0 & cc \end{cases}$$

Hallar estimador puntual de θ por el método de los momentos.

Observación: Supongamos que para hallar θ planteamos $\int_{-\infty}^{\infty} f(x)\,dx = 1$

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

Planteamos: $\frac{1}{n}\sum_{i=1}^{n} X_i = E(X)$ ①

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

$$= \frac{\theta}{2^\theta}\left(\frac{x^{2+1}}{2+1} - \frac{x^{\theta+1}}{\theta+1}\right) = \frac{\theta}{2^\theta}\left(\frac{2^{\theta+1}}{\theta+1} - 0\right) = \frac{2\theta}{\theta+1}$$

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

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

Método de máxima verosimilitud

Sea $X_1 \dots X_n$ muestra aleatoria de una v.a 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 v.a 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 $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \overset{\text{indep}}{=}$

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

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

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

Planteamos la función de verosimilitud $L(x_1, x_2, \ldots, x_n, \lambda) = P(X_1=x_1) \cdot P(X_2=x_2) \ldots$
$P(X_n = x_n) = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} \cdot \ldots \cdot \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$

$= \frac{(e^{-\lambda})^n \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} = \frac{e^{-\lambda n}\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$ Para facilitar se toma $\ln(L(x_1, x_2, \ldots, x_n, \lambda))$
es estrictamente creciente

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

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

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

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

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

Ej): Sea $X_1, X_2, \ldots, X_n$ muestra aleatoria de una va X, $X \sim B(1, P)$

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

Tomamos una muestra de n artículos de toda la producción. Para cada artículo se define
$X_i = \begin{cases} 1 & \text{si el artículo es defectuoso} \\ 0 & cc \end{cases}$ $i = 1, 2, \ldots, n$

$\rightarrow X_i \sim B(1, P)$

Se quiere hallar EMV de P. $P(X=x_i) = \binom{1}{x_i} P^{x_i}(1-P)^{1-x_i} = P^{x_i}(1-P)^{1-x_i}$

$\rightarrow L(x_1, x_2, \ldots, x_n, P) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$
$= P(X_1 = x_1) \cdot P(X_2 = x_2) \ldots P(X_n = x_n)$
$= P^{x_1}(1-P)^{1-x_1} \cdot P^{x_2}(1-P)^{1-x_2} \cdot \ldots \cdot P^{x_n}(1-P)^{1-x_n}$

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

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

regla de la cadena

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

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

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

Ej) b) i) Se seleccionan n=100 chips fabricados por una compañía. 5 son defectuosos y $\sum_{i=1}^{100} X_i = 5$ ∴ la estimación de P es $\bar{x} = \frac{1}{100}\sum_{i=1}^{100} x_i = \frac{5}{100} = 0{,}05$

Caso continuo

Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una va X continua con densidad $f(x)$ y si $x_1, x_2, \ldots, x_n$ es una medición de la muestra aleatoria, entonces se toma:

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

Ej) Sea $X_1, X_2, \ldots, X_n$ muestra aleatoria de una va X, $X \sim Exp(\lambda)$

Hallar el EMV de λ

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

$\rightarrow L(x_1, x_2, \ldots, x_n, \lambda) = f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_n)$

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

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

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

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

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

∴ $\hat{\lambda} = \frac{1}{\bar{X}}$

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

$x_1 = 49{,}5$ $x_2 = 1015$ $x_3 = 1009{,}4$ $x_4 = 173{,}4$ $x_5 = 251{,}6$

$x_6 = 315$ $x_7 = 1301{,}4$ $x_8 = 732{,}8$ $x_9 = 660{,}6$ $x_{10} = 1456{,}5$

Se sospecha que la dur en hs tiene dist exponencial.

a) Hallar el EMV de λ.

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

X: "tiempo de vida en horas de una lamparita" $X \sim Exp(\lambda)$ λ desconocido

Se toma una muestra de 10 lámparas

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

$X_1, X_2, \ldots, X_{10}$ muestra aleatoria de la va X.

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

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

g) Sea $X_1, X_2, \ldots, X_n$ muestra aleatoria de una v.a X. $X \sim N(\mu, \sigma^2)$

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

$$L(x_1, x_2, \ldots, x_n, \mu, \sigma^2) = f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_n)$$

$$= \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{1}{2\sigma^2}(x_1-\mu)^2} \cdot \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{1}{2\sigma^2}(x_2-\mu)^2} \cdot \ldots \cdot \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{1}{2\sigma^2}(x_n-\mu)^2}$$

$$= \left(\frac{1}{\sqrt{\sigma^2 2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum (x_i-\mu)^2} = \frac{1}{(\sigma^2 2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1} (x_i-\mu)^2}$$

$$\ln(L(x_1, x_2, \ldots, x_n, \mu, \sigma^2)) = \ln\left(\frac{1}{(\sigma^2 2\pi)^{n/2}}\right) + \ln\left(e^{-\frac{1}{2\sigma^2}\sum (x_i-\mu)^2}\right)$$

$$= 0 - \frac{n}{2}\ln(\sigma^2 2\pi) - \frac{1}{2\sigma^2}\sum (x_i-\mu)^2$$

$$\begin{cases} \frac{d}{d\mu} \ln L(x_1, x_2, \ldots, x_n, \mu, \sigma) \quad ① \\ \frac{d}{d\sigma} \ln L(x_1, x_2, \ldots, x_n, \mu, \sigma) \quad ② \end{cases}$$

① $-\frac{1}{2\sigma^2}\sum 2(x_i-\mu)(-1)$

② $\left(-\frac{n}{2}\right)\frac{1}{\sigma^2 2\pi} 2\pi - \frac{1}{2}\sum (x_i-\mu)^2 \left(-\frac{1}{\sigma^4}\right)$

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

$\to \sum (x_i-\mu) = 0 \to \sum x_i - \sum \mu = 0 = \sum x_i - n\mu = 0 \to \boxed{\hat{\mu} = \frac{\sum x_i}{n} = \bar{X}}$

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

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

$\to \boxed{\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} (X_i-\bar{X})^2}$ Importante acordarse los resultados

Propiedades de los EMV → son válidos si la muestra es grande

Bajo condiciones muy generales, se puede probar:

① Los EMV son asintóticamente insesgados

② Son consistentes los EMV

③ Para muestras muy grandes, son asintóticamente los de menor varianza

④ Propiedad de invarianza: 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 ej 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, usamos la propiedad de invarianza.

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

En el ej de la binomial: b) ii) Se pide el EMV de $(1-p)^6$ $n=100$

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

$g(p) = (1-p)^6$ (o sea se toman 6 chips y ninguno es defectuoso)

el EMV 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 las lamparas: X: duración en horas de una lampara

$X_1, X_2, \ldots, X_n$ m.a. de X. el EMV de $\lambda \rightarrow \hat{\lambda} = \frac{1}{\bar{X}}$

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

$\therefore$ el EMV es $\widehat{1 - e^{-\lambda \cdot 1400}} = 1 - e^{-\hat{\lambda} \cdot 1400} = 1 - e^{-\frac{1}{\bar{X}} 1400}$

Intervalos de confianza

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una va X. Sea θ un parametro 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 $P(\theta \in (\hat{\Theta}_1, \hat{\Theta}_2)) = 1 - \alpha$

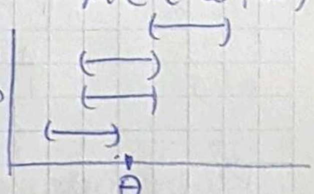
con $1-\alpha$ fijado de antemano. $(1-\alpha) \cdot 100\%$ es el nivel de confianza del intervalo.

Si $x_1, x_2, \ldots, x_n$ es una medición de la muestra aleatoria entonces $\hat{\Theta}_1$ tomará el valor θ_1 y $\hat{\Theta}_2$ tomará el valor θ_2

O sea para la medición $x_1, x_2, \ldots, x_n$ tenemos un intervalo numerico (θ_1, θ_2)

Si $x_1', x_2', \ldots, x_n'$ es otra medición de la muestra aleatoria entonces obtendremos otro intervalo numerico (θ_1', θ_2') $\mu \in (-\infty, \infty)$

Si $1-\alpha = 0{,}95$, entonces si tenemos 100 mediciones de la m.a. vamos a tener 100 intervalos numericos de los cuales aprox. 95 van a contener a θ



~~A mayor nivel de confianza, más precisión~~

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

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

Sea $X_1, X_2, \ldots X_n$ m.a. de una v.a X donde $X \sim N(\mu, \sigma^2)$ con σ^2 conocido. Se quiere un intervalo de confianza para μ con nivel 95%

Utilizaremos el método del pivote

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

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

- debe contener a μ
- debe tener distribución conocida

consideramos el estadístico $Z = \dfrac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$ notar que $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ y entonces $Z \sim N(0,1)$

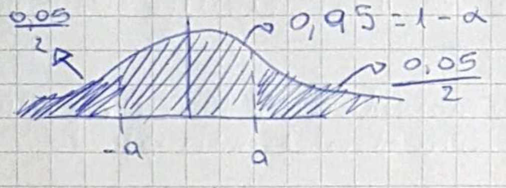
además Z contiene a μ. A Z se lo llama estadístico pivote.

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

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

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

$a = z_{\frac{0{,}05}{2}} = 1{,}96$



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

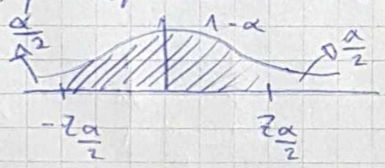
$= P\left(-1{,}96\sqrt{\frac{\sigma^2}{n}} \leq \bar{X} - \mu \leq 1{,}96\sqrt{\frac{\sigma^2}{n}}\right) = P\left(-\bar{X} - 1{,}96\sqrt{\frac{\sigma^2}{n}} \leq -\mu \leq -\bar{X} + 1{,}96\sqrt{\frac{\sigma^2}{n}}\right)$

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

$= P\left(\mu \in \left(\underbrace{\bar{X} - 1{,}96\sqrt{\frac{\sigma^2}{n}}}_{\hat{\theta}_1}, \underbrace{\bar{X} + 1{,}96\sqrt{\frac{\sigma^2}{n}}}_{\hat{\theta}_2}\right)\right) \therefore \left(\bar{X} - z_{\frac{0{,}05}{2}}\sqrt{\frac{\sigma^2}{n}}, \bar{X} + z_{\frac{0{,}05}{2}}\sqrt{\frac{\sigma^2}{n}}\right)$

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

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



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

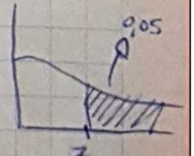
Se usan tablas de valores críticos $\rightarrow$ del área a la derecha (el complemento de $\frac{0{,}05}{2}$) $P(X \geq x)$

Ejemplo. Se sabe que el límite de elasticidad de un tipo particular de barra de acero está normalmente distribuido con $\sigma = 100$. Se toman 25 barras al azar, a cada una se le mide el límite de elasticidad y resulta un $\bar{x} = 8439\,lb$. Calcule un intervalo de confianza de nivel 90% para la verdadera media de límite de elasticidad.

X: "límite de elasticidad de una barra de acero" $X \sim N(\mu, \sigma^2)$ $\sigma = 100$
X_i: "límite de elasticidad de barra de acero i" $i = 1, \ldots, 25$
$X_1, X_2, \ldots, X_{25}$ m.a. de la v.a. X. $\bar{x} = 8439\,lb$

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

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



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

$L = 8472 - 8406 = 66$

Si tomamos $1-\alpha=0,92$ $\quad \alpha=0,08 \quad \frac{\alpha}{2}=0,04 \quad z_{0,04}=1,76$

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

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

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

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

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

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

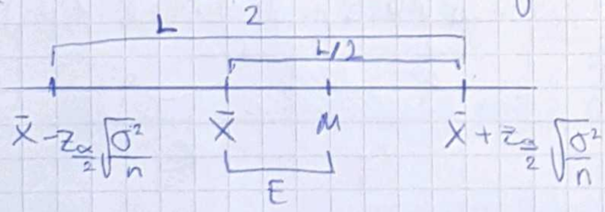
Se puede plantear que tamaño n de muestra se necesita para que la longitud L sea menor a "l". ("l" se fija de antemano)

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

ej: que tamaño de muestra se necesita para que el intervalo de nivel 90% se reduzca a la mitad.

$$L=2\cdot 1,65\sqrt{\frac{100^2}{n}}\le 33 \longrightarrow \underbrace{\left(\frac{2\cdot 1,65\cdot 100}{33}\right)^2}_{100}\le n \quad \therefore \boxed{n\ge 100}$$

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



Sea $X_1, X_2, \dots X_n$ m.a de una v.a. X, $n\ge 30$. 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 pivote: se toma un estimador puntual de μ.
A partir de $\hat\mu=\bar X$ se construye un estadístico tal que:
a) debe contener a μ

b) debe tener distribución conocida

Caso 1: se conoce σ^2

~~Hallamos a~~ Tomamos el estadístico $Z=\frac{\bar X-\mu}{\sqrt{\frac{\sigma^2}{n}}}$ $\quad Z\approx N(0,1)$ por TCL.

Se plantea hallar a tal que

$$1-\alpha=P(-a\le Z\le a)\underset{TCL}{\approx}\Phi(a)-\Phi(-a)=2\Phi(a)-1$$

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

$\left(\bar X - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\ ,\ \bar X + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right)$. La diferencia es que el nivel de confianza es aproximadamente $(1-\alpha)\cdot 100\%$.

NOTA

caso 2: no se conoce σ^2. Se reemplaza σ^2 por $S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$

$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}}$. Se puede probar que $Z \approx N(0,1)$ por TCL. y también se puede probar que S^2 es ~~aproximadamente~~ estimador consistente de σ^2

Con un planteo análogo al caso 1, se llega a:

$\left(\bar{X} - z_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}}, \bar{X} + z_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}}\right)$ con un nivel $(1-\alpha).100\%$ aproximadamente

Ej: Se registraron los tiempos utilizados en la compra para 64 clientes seleccionados al azar en un supermercado. La media y la varianza de los 64 tiempos fueron 33 minutos y 256 minutos^2 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 la v.a. X

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

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

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

Intervalo de confianza para la media de una normal con varianza desconocida

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

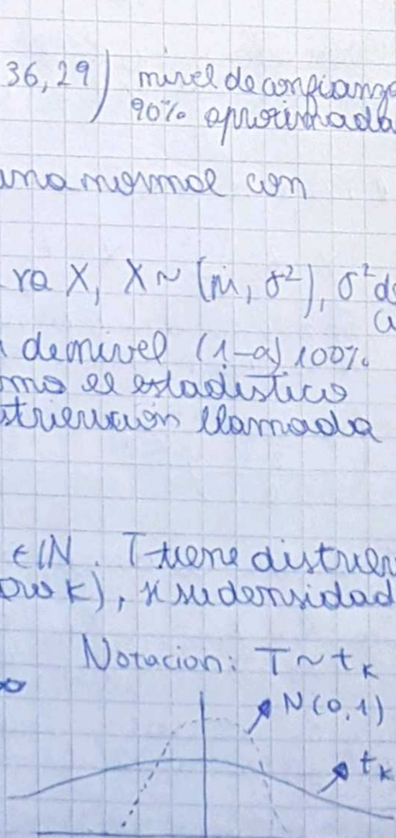
Se quiere un intervalo de confianza para μ de nivel $(1-\alpha)100\%$. Utilizamos el método del pivote. Tomamos el estadístico

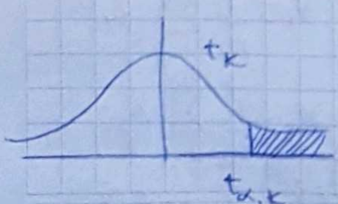
$T = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}}$. Se puede probar que T tiene una distribución llamada "student" con parámetro $\underline{n-1}$.

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

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

Notación: $T \sim t_k$





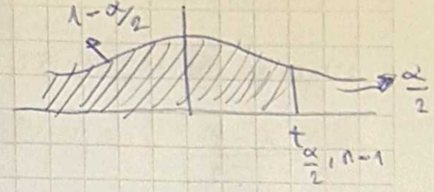
Ejemplo: $t_{0,05,9} = 1,833$

k	0,05
9	1,833

T simétrica

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

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



Entonces:

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

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

Ej. Un fabricante de pólvora desarrolló una nueva fórmula que se probó con 8 granadas. Las velocidades iniciales resultantes en 10 por segundo 3005; 2925, 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: "velocidad inicial de una granada tomada al azar con pólvora nueva" $X \sim N(\mu, \sigma^2)$

X_i: "vel. inicial de la granada i" $i = 1, 2, \dots, 8$

$X_1, X_2, \dots, X_8$ m.a. de la va X. $1-\alpha = 0{,}95$ $\bar x = 2959$ $\alpha = 0{,}05$ $\alpha = 0{,}025$

$\left(\bar X - t_{\frac{\alpha}{2},n-1}\sqrt{\frac{S^2}{n}} \; ; \; \bar X + t_{\frac{\alpha}{2},n-1}\sqrt{\frac{S^2}{n}}\right)$ $S = 39{,}1$ $t_{0{,}025,\underset{n-1}{7}} = 2{,}365$

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

Observaciones: ① ~~[illegible]~~ En el caso del intervalo que usamos recién, la longitud es $L = 2t_{\frac{\alpha}{2},n-1}\sqrt{\frac{S^2}{n}}$ no se puede despejar el n porque no se puede. Aparece en muchos lados.

② En el caso del intervalo: (en muestras grandes)

$\left(\bar X - z_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}} \; ; \; X + z_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}}\right)$ $L = 2 z_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}}$

En este caso sí se puede plantear hallar n tal que $L \le l$ y se despeja n, porque S^2 es consistente (se lo toma como un número)

$$\left(\frac{2 z_{\frac{\alpha}{2}} S}{l}\right)^2 \le n$$

Intervalos de confianza para diferencia de medias

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

$X_1 \sim N(\mu_1, \sigma_1^2)$ $X_2 \sim N(\mu_2, \sigma_2^2)$

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 pivote: ① $\widehat{\mu_1 - \mu_2} = \bar X_1 - \bar X_2$

$$\begin{cases} \bar X_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i} \\ \bar X_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i} \end{cases}$$

② Consideramos $Z = \frac{\bar X_1 - \bar X_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

X_1 y X_2 indep.

$V(\bar X_1 - \bar X_2) = V(\bar X_1) + V(\bar X_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

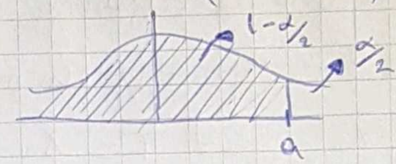
$E(\bar X_1 - \bar X_2) = E(\bar X_1) - E(\bar X_2) = \mu_1 - \mu_2$

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:

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

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

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



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

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

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

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

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

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

$X_{11}, X_{12}, \ldots, X_{1n_1}$ ma de la va X_1 $X_1 \sim N(\mu_1, 0{,}1^2)$

$X_{21}, X_{22}, \ldots, X_{2n_2}$ ma de la va X_2 $X_2 \sim N(\mu_2, 0{,}15^2)$

$1-\alpha = 0{,}9$ $\alpha = 0{,}1$ $\alpha = 0{,}05$ $z_{\frac{\alpha}{2}} = z_{0{,}05} = 1{,}65$

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

$= (0{,}09837 \ ; \ 0{,}281620)$

Observacion: 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.

② $L = 2 z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ Se puede plantear que tamaño $n = n_1 = n_2$ se necesita para que $L \le l$.

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

Se repite el caso sin decir que X_1 y X_2 son normales, y $n_1 \ge 30$ y $n_2 \ge 30$

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

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

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

② $Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ $S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2$

$S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2$

Se puede probar que $Z \approx N(0,1)$ por TCL 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:

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

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

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

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

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: $\sigma_1^2 \neq \sigma_2^2$.

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

$$\nu = \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}}$$ $\rightarrow$ se aproxima al entero más cercano

Se llega al intervalo:

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

Ejemplo: para comparar 2 tipos de paragolpes, 6 de cada tipo se montaron en un automóvil compacto. 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 intervalo de confianza para la diferencia de medias con confianza 99%. Asumir que la muestra tienen dist normales y son indep.

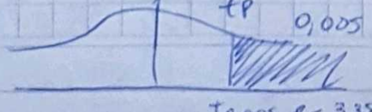
X_{1i}: "costo de rep. de auto i" $i = 1, 2, \ldots, 6$ $\quad X_1 \sim N(\mu_1, \sigma_1^2)$

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

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

$\frac{\alpha}{2} = 0{,}005$ $\quad S_1 = 24{,}4513$ $\quad S_2 = 14{,}2126$

$t_{0{,}005,\nu} = 3{,}355$



$\rightarrow (-40{,}9036 \; ; \; 37{,}0702)$

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

NOTA

Ej: Los siguientes m.a: son mediciones de la capacidad productiva de calor en millones de calorías por tonelada de especímenes de carbón de 2 minas.

mina 1	8260	8130	8350	8070	8340		$n_1=5$
mina 2	7950	7890	7900	8140	7920	7840	$n_2=6$

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

X_{1i}: "capacidad productiva de calor de la muestra i de la mina 1"
$i=1,2,3,4,5$ $X_{11}, X_{12}, \dots, X_{1n_1}$ m.a. de la v.a $X_1 \sim N(\mu_1, \sigma_1^2)$

X_{2i}: " " " " " mina 2" $i=1,2,\dots,6$

$X_{21}, X_{22}, \dots, X_{2n_2}$ m.a de la v.a $X_2 \sim N(\mu_1, \sigma_2^2)$ $\sigma_1^2=\sigma_2^2=\sigma^2$

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

$\bar{x}_1=8230$ $s_1^2=15750$ $S_p^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}=13066{,}7$

$\bar{x}_2=7940$ $s_2^2=10920$

$S_P=114{,}31$

$\rightarrow (65{,}0410 ; 514{,}958)$ $\mu_1>\mu_2$ con seguridad del 99%.

El carbón de la mina 1 tiene mayor capacidad productiva que de la mina 2.

Intervalo de confianza para muestras ~~[illegible]~~ pareadas

Supongamos $X_{11}, X_{12}, \dots, X_{1n}$ m.a de una v.a X_1
$X_{21}, X_{22}, \dots, X_{2n}$ m.a de una v.a X_2.

$\mu_1=E(X_1)$ $\mu_2=E(X_2)$ $\sigma_1^2=V(X_1)$ $\sigma_2^2=V(X_2)$ desconocidas

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

(X_{1i}, X_{2i}) son 2 mediciones diferentes sobre la misma unidad muestral i.

O sea, ahora X_1 y X_2 no son indep. Consideramos una tercer muestra:

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

Entonces $D_1, D_2, \dots, D_n$ es una m.a. de la v.a $D=X_1-X_2$

$\rightarrow E(D)=\mu_D=E(X_1-X_2)=E(X_1)-E(X_2)=\mu_1-\mu_2$

$\sigma_D^2=V(D)=V(X_1-X_2)=\sigma_1^2+\sigma_2^2-2\,Cov(X_1,X_2)$ σ_D^2 desconocida

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

$$T=\frac{\bar{D}-\mu_D}{\sqrt{\frac{S_D}{n}}} \sim t_{n-1}$$

Se llega al intervalo: $\left(\bar{D} - t_{\frac{\alpha}{2}, n-1}\sqrt{\frac{S_D}{n}} ; \bar{D} + t_{\frac{\alpha}{2}, n-1}\sqrt{\frac{S_D}{n}}\right)$

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

~~Antes~~	1	2	3	4	5	6	7	8	9	10
Antes	45	73	46	124	33	57	83	34	26	17
Despues	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 nivel 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" $i = 1, \dots, 10$
$X_{11}, \dots, X_{1\,10}$ ma de X_1
X_{2i}: "perdida de horas hombre despues del programa en planta i" $i = 1, \dots, 10$
$X_{21}, \dots, X_{2\,10}$ ma de X_2

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

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

$t_{0{,}025, 9} = 2{,}262$

$\left(5{,}2 - 2{,}262\left(\frac{4{,}08}{\sqrt{10}}\right) ; 5{,}2 + 2{,}262\left(\frac{4{,}08}{\sqrt{10}}\right)\right) = (2{,}2815 ; 8{,}1184)$

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

Intervalo de confianza para la varianza de una normal

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

Metodo del pivote:

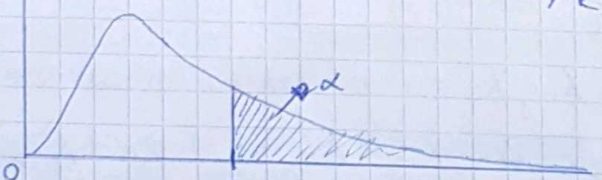
1) $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

2) $X_0 = \frac{(n-1)S^2}{\sigma^2}$ Se puede probar que X_0 tiene distribución ji cuadrado (χ^2) con parametro (o 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 si su densidad es:

$$f(x) = \begin{cases} \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} & x > 0 \\ 0 & cc \end{cases}$$

Notación: $X \sim \chi^2_k$



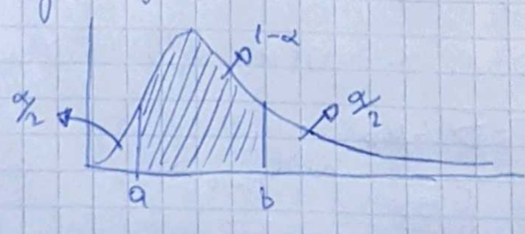
si k más grande, se achata la grafica

$a = \chi^2_{\alpha, k}$

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

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

Se encuentra con esta elección el intervalo con menor longitud



NOTA

$\therefore 1-\alpha = P\left(\chi^2_{1-\frac{\alpha}{2},n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2},n-1}\right) = P\left(\frac{1}{\chi^2_{1-\frac{\alpha}{2},n-1}} \ge \frac{\sigma^2}{(n-1)S^2} \ge \frac{1}{\chi^2_{\frac{\alpha}{2},n-1}}\right)$

$= P\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} \ge \sigma^2 \ge \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}\right) = P\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right)$

$\therefore \left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} ; \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right) \rightarrow$ Notar que: $\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}} ; \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}}\right)$ es un intervalo de confianza para σ de nivel $(1-\alpha)100\%$.
desviación estandar

Ej.) X: "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" $i = 1, 2, \dots, 20$

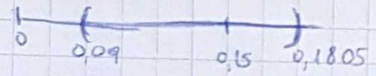
$X_1, X_2, \dots, X_{20}$ m.a de la v.a X $S^2 = 0{,}0153$ $1-\alpha = 0{,}95$ $\frac{\alpha}{2} = 0{,}025$

$\chi^2_{\frac{\alpha}{2},19} = \chi^2_{0{,}025,19} = 32{,}9$

$\chi^2_{1-\frac{\alpha}{2},19} = \chi^2_{0{,}975,19} = 8{,}91$ $\rightarrow \left(\frac{(20-1)\,0{,}0153}{32{,}9} ; \frac{(20-1)\,0{,}0153}{8{,}91}\right)$

$(0{,}00884 ; 0{,}0326)$

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



Intervalo de confianza para cociente de varianzas.

Sea $X_{11}, \dots, X_{1n}$ m.a. de v.a X_1
$X_{21}, \dots, X_{2n}$ m.a de v.a X_2
X_1 y X_2 son indep. y normales con parametros desconocido.

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

El estadístico pivotes: $F = \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} = \frac{S_2^5}{S_1^2}\left(\frac{\sigma_1^2}{\sigma_2^2}\right)$. Se puede probar que F tiene una distribución llamada Fisher (o F) con parametros n_2-1 y n_1-1 (en ese orden)

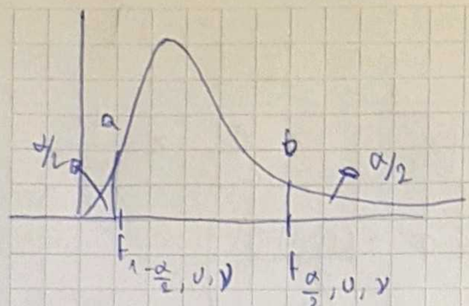
O F tiene dist Fisher con n_2-1 grados de libertad en el numerador y n_1-1 grados de libertad en el denominador.

Observación: Sea X una v.a continua. Sean U y V numeros naturales

X tiene dist F con U grados de libertad en numerador y V grados de libertad en denominador si su densidad es:

$$f(x) = \frac{\Gamma\left(\frac{U+V}{2}\right)\left(\frac{U}{V}\right)^{U/2} \cdot x^{\frac{U}{2}-1}}{\Gamma\left(\frac{U}{2}\right)\Gamma\left(\frac{V}{2}\right)\left[\left(\frac{U}{V}\right)x+1\right]^{\frac{U+V}{2}}} \quad 0 < x < \infty$$

Notación $F \sim F_{U,V}$



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

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

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

Se llega al intervalo: $\left(\dfrac{S_1^2}{S_2^2} \cdot F_{1-\frac{\alpha}{2}, n_2-1, n_1-1} \; ; \; \dfrac{S_1^2}{S_2^2} \cdot F_{\frac{\alpha}{2}, n_2-1, n_1-1} \right)$

$\dfrac{\sigma_1^2}{\sigma_2^2}$ se encuentra en este intervalo con una confianza $(1-\alpha)\,100\%$

Si 1 no está en el intervalo, se puede decir cual de los 2 es mas grande