PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE MATEMÁTICAS / DEPARTAMENTO DE ESTADÍSTICA

ELM2400 Métodos Estadísticos

Intervalos de Confianza

Profesor: Alexis Peña
Ayudante: Reinaldo González S.

1. Estimación de la Media

Suponga que $X_1, X_2, ..., X_n$ es una muestra aleatoria de una población Normal con media μ y varianza σ^2 y que se quiere obtener un intervalo de confianza del $(1-\alpha)100\,\%$ para μ. El estimador de μ es:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

1.1. σ^2 conocido

Pivote	**Intervalo**
$Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$IC(\mu) = \bar{X} \pm \dfrac{\sigma}{\sqrt{n}} \cdot Z_{1-\alpha/2}$

Nota:
Teorema del Limite Central: Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de una población X con media μ y varianza finita σ^2, entonces:

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow N(0,1) \text{ cuando } n \rightarrow \infty$$

Debido a este Teorema, el intervalo anterior es valido para estimar la media de cualquier distribución especialmente cuando $n \rightarrow \infty$.

1.2. σ^2 desconocido

Se reemplaza σ^2 por su estimador insesgado $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Pivote	Intervalo
$T = \dfrac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$	$IC(\mu) = \bar{X} \pm \dfrac{S}{\sqrt{n}} \cdot t_{n-1,1-\alpha/2}$

Notas:
a) Si $n > 30$ se puede aproximar $t_{n-1,1-\alpha/2}$ por $Z_{1-\alpha/2}$ pues la distribución t-Student converge a la distribución Normal.

b) El intervalo también es útil cuando la población es aproximadamente Normal.

2. Comparación de Dos Medias

Sean $X_1, X_2, ..., X_{n1}$ muestras aleatorias de una población Normal $N(\mu_1, \sigma_1^2)$ e $Y_1, Y_2, ..., Y_{n2}$ muestras aleatorias de una población Normal $N(\mu_2, \sigma_2^2)$. Ambas muestras independientes. Las medias poblacionales μ_1 y μ_2 pueden ser comparadas usando la diferencia $\mu_1 - \mu_2$, cuyo estimador es $\hat{\mu_1} - \hat{\mu_2} = \bar{X} - \bar{Y}$.

2.1. σ_1^2 y σ_2^2 conocidas

Pivote	Intervalo
$Z = \dfrac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$IC(\mu_1 - \mu_2) = (\bar{X} - \bar{Y}) \pm \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}} \cdot Z_{1-\alpha/2}$

2.2. σ_1^2 y σ_2^2 desconocidas pero iguales $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$

Se usa el estimador de σ^2 dado por:

$$\begin{aligned}
\hat{\sigma}^2 &= S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \\
S_1^2 &= \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2 \\
S_2^2 &= \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2
\end{aligned}$$

Pivote	Intervalo
$T = \dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$	$IC(\mu_1-\mu_2) = (\bar{X}-\bar{Y}) \pm S_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}} \cdot t_{n_1+n_2-2,1-\alpha/2}$

2.3. σ_1^2 y σ_2^2 desconocidas y distintas

Pivote	Intervalo
$T = \dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \sim t_v$	$IC(\mu_1-\mu_2) = (\bar{X}-\bar{Y}) \pm \sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}} \cdot t_{v,1-\alpha/2}$

con $v = (S_1^2 n_1 + S_2^2 n_2)^2 \Big/ \dfrac{(S_1^2/n_1)^2}{n_1-1} + \dfrac{(S_2^2/n_2)^2}{n_2-1}$

3. Comparación de Dos Medias para datos Pareados

Sean los pares$(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ donde $X_1, X_2, ..., X_n$ puede ser considerada una muestra aleatoria de una población normal $N(\mu_1, \sigma_1^2)$ e $Y_1, Y_2, ..., Y_n$ una muestra aleatoria de una población normal $N(\mu_2, \sigma_2^2)$. Se supone que X_i e Y_i no son independientes pero si lo son (X_i, Y_i) de (X_j, Y_j), $i \neq j$. Se trata de hacer inferencia respecto a $\mu_1 - \mu_2$.

Pivote	Intervalo
$F = \dfrac{\bar{d}-\mu_D}{S_D/\sqrt{n}} \sim t_{n-1}$	$IC(\mu_1-\mu_2) = \bar{d} \pm \dfrac{S_D}{\sqrt{n}} \cdot t_{n-1,1-\alpha/2}$

$$
\begin{aligned}
\mu_D &= \mu_1 - \mu_2 \\
d_i &= X_i - Y_i \\
\bar{d} &= \frac{1}{n}\sum_{i=1}^{n} d_i = \bar{X} - \bar{Y} \\
S_D^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2
\end{aligned}
$$

Nota:
Si $0 \in IC(\mu_1-\mu_2)$ dados en las secciones 2.1, 2.2, 2.3 y 3, entonces se infiere, con un $(1-\alpha)100\,\%$ de confianza que $\mu_1 = \mu_2$.

4. Estimación de la Varianza

Suponga que $X_1, X_2, ..., X_n$ es una muestra aleatoria de una poblacion normal $N(\mu, \sigma^2)$ y que se quiere obtener un intervalo de confianza del $(1-\alpha)100\,\%$ para σ^2 cuyo estimador es:

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

4.1. μ desconocido

Pivote	Intervalo
$X = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$	$IC(\sigma^2) = \left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}} ; \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right]$

4.2. μ conocido

Pivote	Intervalo
$X = \frac{\sum (X_i - \mu)^2}{\sigma^2} \sim \chi^2_{n}$	$IC(\sigma^2) = \left[\frac{\sum (X_i - \mu)^2}{\chi^2_{n,1-\alpha/2}} ; \frac{\sum (X_i - \mu)^2}{\chi^2_{n,\alpha/2}}\right]$

5. Comparación de Dos Varianzas

Sea $X_1, X_2, ..., X_{n1}$ una muestra aleatoria de una población normal $N(\mu_1, \sigma_1^2)$ e $Y_1, Y_2, ..., Y_{n2}$ otra muestra aleatoria de una población normal $N(\mu_2, \sigma_2^2)$. Ambas muestras independientes con μ_1 y μ_2 desconocidas. Las varianzas poblacionales σ_1^2 y σ_2^2 pueden ser comparadas usando el cuociente cuyo estimador es:

$$\begin{aligned}
\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} &= \frac{S_1^2}{S_2^2} \\
S_1^2 &= \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2 \\
S_2^2 &= \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2
\end{aligned}$$

Pivote	Intervalo
$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$	$IC(\sigma_2^2/\sigma_1^2) = \left[\frac{S_2^2}{S_1^2} \cdot F_{n_1-1,n_2-1,\alpha/2} ; \frac{S_2^2}{S_1^2} \cdot F_{n_1-1,n_2-1,1-\alpha/2}\right]$

Nota:
Si $1 \in IC(\sigma_2^2/\sigma_1^2)$, entonces se infiere, con un $(1-\alpha)100\,\%$ de confianza que las varianzas poblacionales σ_1^2 y σ_2^2 son iguales.

6. Estimación de una Proporción

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población Bernoulli de parámetro p cuyo estimador es:

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad , X_i = 0, 1 \qquad , 0 \leq p \leq 1$$

Usando el Teorema del Limite Central se puede obtener un intervalo aproximado para p si n es grande $(n > 30)$.

$$IC(p) = \hat{p} \pm \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \cdot Z_{1-\alpha/2}$$

7. Comparación de Proporciones

Sean $X_1, X_2, ..., X_{n1}$ e $Y_1, Y_2, ..., Y_{n2}$ muestras aleatorias independientes de poblaciones Bernoulli de parámetros p_1 y p_2 respectivamente. Las proporciones poblacionales pueden ser comparadas a través de un intervalo de confianza aproximado para muestras grandes, dado por:

$$IC(p_1 - p_2) = (\hat{p_1} - \hat{p_2}) \pm \sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}} \cdot Z_{1-\alpha/2}$$

Nota:
Si $0 \in IC(p_1 - p_2)$ entonces se infiere, con un $(1-\alpha)100\,\%$ de confianza, que $p_1 = p_2$.

8. Estimación de Cualquier Parámetro

Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de una población X con función de densidad $f(x, \theta)$ y $\hat{\theta}_{EMV}$ es el estimador máximo verosímil de θ entonces se pueden usar las propiedades de normalidad asintótica de estos estimadores para obtener un intervalo aproximado para θ.

Como $\dfrac{\hat{\theta}_{EMV} - \theta}{1/\sqrt{nI_I(\theta)}}$ tiende a $N(0,1)$ cuando $n \to \infty$. Con $I_I(\theta) = -E\left(\dfrac{\partial^2}{\partial\theta^2}\ln(f(x,\theta))\right)$, entonces el intervalo de confianza asintótico es:

$$IC(\theta) = \hat{\theta}_{EMV} \pm \frac{1}{\sqrt{nI_I(\hat{\theta})}} \cdot Z_{1-\alpha/2}$$