Estadística
Inferencia Estadística

Laureate International Universities

Problemas en Inferencia Estadística

POBLACIÓN → $X \approx F(\theta)$ ← desconocido

A partir de una M.A.S. $X_1, X_2, \ldots, X_n$ queremos estimar el valor de θ

Estimar : Asignar un valor a algo que desconocemos

¿Cómo?

A partir del modelo F y de la muestra

- Estimación puntual: Un solo valor
- Estimación por intervalos: Un rango de valores
- Contraste de hipótesis: Rechazo o aceptación de ciertas hipótesis sobre el parámetro

Intervalo de confianza

Sea $X_1,\ldots, X_n$ una m.a.s de una población X con distribución F (θ), y T_1 , T_2 dos estadísticos tales que:

$$P(T_1(X_1,\ldots,X_n) \leq \theta \leq T_1(X_1,\ldots,X_n)) = 1-\alpha$$

Entonces,

$$[T_1(X_1,\ldots,X_n), T_2(X_1,\ldots,X_n)]$$

Es un intervalo para estimar θ con un nivel de confianza de $1-\alpha$

Queremos que esta probabilidad esté cercana a 1, es decir que **α** sea pequeño

Universidad Europea de Madrid

Nivel de Confianza

El concepto de confianza puede interpretarse de la siguiente manera:

Si se repitiera el experimento un número suficiente de veces, en $1-\alpha$ de cada 100 casos se confiaría que el parámetro θ pertenece al intervalo.

Por ejemplo, si elegimos $1-\alpha = 0.95$, de cada 100 estimaciones, el 95% serían buenas y el 5% malas.

Universidad Europea de Madrid

Métodos de construcción de intervalos: Método de la cantidad pivotal

El método de la cantidad pivotal consiste en encontrar un estadístico (pivote) que dependa del parámetro y de la muestra, $C_{\theta}(X_1,\ldots,X_n)$ pero que su distribución NO dependa de θ.

Para construir el intervalo, buscamos dos constantes K_1, K_2 tales que:

$$P(K_1 \le C_{\theta}(X_1,\ldots,X_n) \le K_2) = 1-\alpha$$

De entre todos los posibles buscamos lo de menor longitud.

Esto no siempre es fácil, por lo que se construyen intervalos simétricos, que muchas veces coinciden con los de longitud mínima.

Universidad Europea de Madrid

Intervalos para un población Normal

Consideremos un población con distribución NORMAL

1. Intervalos para la media μ

1.1 Varianza conocida

Consideremos la siguiente cantidad pivotal para estimar la media

$$\overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \approx N(0,1)$$

$$P(K_1 \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le K_2) = 1-\alpha \Rightarrow P\left(K_1\frac{\sigma}{\sqrt{n}} \le \overline{X}-\mu \le K_2\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Universidad Europea de Madrid

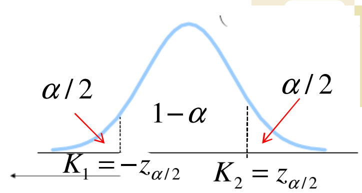
Intervalo para la media en una población Normal

$$P\left(K_1 \frac{\sigma}{\sqrt{n}} - \bar{X} \le -\mu \le K_2 \frac{\sigma}{\sqrt{n}} - \bar{X} \right) = 1 - \alpha$$

$$P\left(\bar{X} - K_2 \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} - K_1 \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

El intervalo queda: $I = \left[\bar{X} - K_2 \frac{\sigma}{\sqrt{n}}, \bar{X} - K_1 \frac{\sigma}{\sqrt{n}} \right]$

Sólo queda calcular las constantes:



Intervalo para la media en una población Normal

1.1 Varianza conocida $I = \left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$

1.2 Varianza desconocida

Si σ es deconocido, no podemos utilizar la cantidad pivotal anterior, puesto que depende del valor de la varianza.

Pero conocemos, por el teorema de la t-student (ver transparencias modelos continuos 2) que

$$\frac{\bar{X} - \mu}{\hat{S} \ / \sqrt{n}} \approx t_{n-1}$$

Intervalo para la media en una población Normal

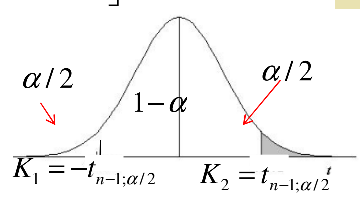
$$P\left(K_1 \le \frac{\overline{X}-\mu}{\hat{S}\ /\sqrt{n}} \le K_2\right) = 1-\alpha \Rightarrow P\left(K_1 \frac{\hat{S}}{\sqrt{n}} \le \overline{X}-\mu \le K_2 \frac{\hat{S}}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(K_1 \frac{\hat{S}}{\sqrt{n}} - \overline{X} \le -\mu \le K_2 \frac{\hat{S}}{\sqrt{n}} - \overline{X}\right) = 1-\alpha$$

$$P\left(\overline{X} - K_2 \frac{\hat{S}}{\sqrt{n}} \le \mu \le \overline{X} - K_1 \frac{\hat{S}}{\sqrt{n}}\right) = 1-\alpha$$

El intervalo queda: $I = \left[\overline{X} - K_2 \frac{\hat{S}}{\sqrt{n}}, \overline{X} - K_1 \frac{\hat{S}}{\sqrt{n}}\right]$

Sólo queda calcular las constantes:



Fórmulas de los intervalos de confianza

Una población Normal

1. Intervalos para la media μ

1.1 Varianza conocida $I = \left[\overline{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$

1.2 Varianza desconocida $I = \left[\overline{X} - t_{n-1;\alpha/2} \frac{\hat{S}}{\sqrt{n}}, \overline{X} + t_{n-1;\alpha/2} \frac{\hat{S}}{\sqrt{n}}\right]$

2. Intervalo para la varianza σ² $I = \left[\frac{(n-1)\hat{S}^2}{\chi^2_{n-1;\alpha/2}}, \frac{(n-1)\hat{S}^2}{\chi^2_{n-1;1-\alpha/2}}\right]$

Ejemplos

1. Se ha llevado a cabo un estudio sobre el tiempo medio de funcionamiento de un componente electrónico y se ha encontrado en una muestra de 100 componentes una vida media de 26 meses de funcionamiento. Si se sabe que la desviación típica de la vida de los componentes es de 8 meses. Calcular un intervalo de confianza al 99% para la vida media de los componentes.

Tenemos que calcular un intervalo para la media con varianza conocida (σ=8), sabiendo que $n = 100, \bar{x} = 26, \alpha = 0.01$

Utilizamos la fórmula

$$I = \left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] = \left[26 - z_{0.01/2} \frac{8}{\sqrt{100}}, 26 + z_{0.01/2} \frac{8}{\sqrt{100}} \right]$$

Ejemplos

Operamos y buscamos el valor en las tablas, obteniendo:

$$I = [26 - z_{0.005} 0.8, 26 + z_{0.005} 0.8] = [23.94, 28.06]$$

Es decir, los componentes electrónicos tendrán, con una confianza del 99%, una vida media de entre 23.94 y 28.06 meses.

Ejemplos

2. Se desea estudiar la característica X que sigue una distribución Normal. Para estimar su varianza, se tomó una muestra de 20 elementos obteniéndose una cuasivarianza de 3.4. Obtener un intervalo de confianza para la varianza al 95%.

Tenemos que calcular un intervalo para la varianza, sabiendo

$$n = 20, \hat{s}^2 = 3.4, \alpha = 0.05$$

Utilizamos la fórmula

$$I = \left[\frac{(n-1)\hat{S}^2}{\chi^2_{n-1;\alpha/2}}, \frac{(n-1)\hat{S}^2}{\chi^2_{n-1;1-\alpha/2}} \right] = \left[\frac{(20-1)3.4}{\chi^2_{19;0.05/2}}, \frac{(20-1)3.4}{\chi^2_{19;1-0.05/2}} \right] =$$

$$\left[\frac{64.6}{32.9}, \frac{64.6}{8.91} \right] = [1.96, 7.25]$$

Intervalos de confianza para proporciones

Intervalo para la proporción p

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

Ejemplo:

En una muestra de 500 familias se observa que 345 de ellas tienen ADSL en casa. Calcular un intervalo para la proporción de familias con internet al 95%

Comparación de poblaciones

Dos poblaciones Normales $X \approx N(\mu_1, \sigma_1), Y \approx N(\mu_2, \sigma_2)$

Se toman dos m.a.s. de tamaños n y m respectivamente.

3. Intervalos para la diferencia de medias μ_1-μ_2

3.1 Varianzas conocidas

$$I = \left[\bar{x} - \bar{y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, \bar{x} - \bar{y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right]$$

3.2 Varianzas desconocidas pero iguales $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$I = \left[\bar{x} - \bar{y} - t_{n+m-2;\alpha/2}\hat{S}_p\sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{x} - \bar{y} + t_{n+m-2;\alpha/2}\hat{S}_p\sqrt{\frac{1}{n} + \frac{1}{m}} \right]$$

donde $$\hat{S}_p^2 = \frac{(n-1)\hat{S}_1^2 + (m-1)\hat{S}_2^2}{n+m-2}$$

Universidad Europea de Madrid

Fórmulas de los intervalos de confianza

Dos poblaciones Normales $X \approx N(\mu_1, \sigma_1), Y \approx N(\mu_2, \sigma_2)$

Se toman dos m.a.s. de tamaños n y m respectivamente.

3. Intervalo para el cociente de varianzas

$$I = \left[\frac{\hat{S}_1^2 / \hat{S}_2^2}{F_{n-1,m-1;\alpha/2}}, \frac{\hat{S}_1^2}{\hat{S}_2^2} F_{m-1,n-1;\alpha/2} \right]$$

Universidad Europea de Madrid

Observaciones

Podemos comparar las medias de dos poblaciones mediante el **intervalo de confianza de la diferencia de medias**.

Si el intervalo contiene al 0, entonces aceptamos como válido ese valor, por lo que ambas medias podrían considerarse iguales, y decimos que: **no existen diferencias significativas entre las medias**

$$\mu_1 - \mu_2 = 0 \Leftrightarrow \mu_1 = \mu_2$$

Universidad Europea de Madrid

Observaciones

Podemos comparar las varianzas de dos poblaciones mediante el **intervalo de confianza del cociente de varianzas**

Si el intervalo contiene al 1, entonces aceptamos como válido ese valor, por lo que ambas varianzas podrían considerarse iguales, y decimos que: **no existen diferencias significativas entre las varianzas**

$$\frac{\sigma_1^2}{\sigma_2^2} = 1 \Leftrightarrow \sigma_1^2 = \sigma_2^2$$

Universidad Europea de Madrid