Capítulo 1

Intervalos de confianza

Teorema: Sean $X_1, \ldots, x_n$ variables aleatorias iid con distribución $N(\mu, \sigma^2)$, $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ y $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$. Entonces

1. $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

2. $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim T_{n-1}$, donde T_{n-1} es una t de Student con $n-1$ grados de libertad.

3. $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, donde χ^2_{n-1} es una chi-cuadrado con $n-1$ grados de libertad.

Cuando se obtiene una estimación puntual de un parámetro, es conveniente acompañar dicha estimación por una "medida" de la precisión de la estimación. Un modo de hacerlo es reemplazar la estimación puntual por un intervalo de valores posibles para el parámetro.

Ejemplo Supongamos que tenemos una muestra aleatoria $X_1, X_2, \ldots X_n$ de una distribución $N(\mu, \sigma_0^2)$ con varianza σ_0^2 conocida y media μ desconocida que queremos estimar. Por ser los datos normales, sabemos que

$$\overline{X} \sim N\left(\mu, \frac{\sigma_0^2}{n}\right),$$

entonces

$$\frac{\overline{X}-\mu}{\sigma_0/\sqrt{n}} \sim N(0,1)$$

y, por lo tanto, sabemos que $P\left(-1{,}96 \leq \sqrt{n}\frac{\overline{X}-\mu}{\sigma_0} \leq 1{,}96\right) = 0{,}95$. A partir de esta expresión "despejamos"μ y obtenemos

$$0{,}95 = P\left(-1{,}96 \leq \sqrt{n}\frac{\overline{X}-\mu}{\sigma_0} \leq 1{,}96\right) = P\left(-1{,}96\frac{\sigma_0}{\sqrt{n}} \leq \overline{X}-\mu \leq 1{,}96\frac{\sigma_0}{\sqrt{n}}\right) =$$
$$= P\left(\overline{X}-1{,}96\frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \overline{X}+1{,}96\frac{\sigma_0}{\sqrt{n}}\right).$$

Es decir, que la probabilidad de que el intervalo $\left[\overline{X}-1{,}96\frac{\sigma_0}{\sqrt{n}};\overline{X}+1{,}96\frac{\sigma_0}{\sqrt{n}}\right]$ contenga al verdadero valor del parámetro μ es 0,95. Este intervalo se denomina intervalo de confianza para μ de nivel de confianza 0,95.

Definición: sea $X_1, X_2, \dots X_n$ una muestra aleatoria de de una distribución que depende de un parámetro θ desconocido. Dadas dos funciones de la muestra f y g tales que

$$P(f(X_1, X_2, \dots X_n) \leq \theta \leq g(X_1, X_2, \dots X_n)) = 1-\alpha,$$

el intervalo el intervalo $[f(X_1, X_2, \dots X_n); g(X_1, X_2, \dots X_n)]$ se denomina intervalo de confianza de nivel $1-\alpha$ para el parámetro θ.

Observaciones:

1. No es correcto decir "la probabilidad de que θ pertenezca al intervalo $(a;b)$ es $1-\alpha$" porque θ no es una variable aleatoria. Debemos decir "la probabilidad de que el intervalo $(a;b)$ contenga al parámetro θ es $1-\alpha$"

2. Una vez construído el intervalo a partir de una muestra dada, ya no tiene sentido hablar de probabilidad. En todo caso, tenemos "confianza" de que el intervalo contenga a θ. La confianza está puesta en que el $(1-\alpha)100\,\%$ de las muestras producirán intervalos que contienen a θ.

1.1. Intervalos de confianza para la esperanza de una población normal con varianza conocida

Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria con distribución $N(\mu, \sigma_0^2)$ con σ_0 conocido. Llamamos $z_{\alpha/2}$ al valor que deja un área de $\alpha/2$ en la cola a la derecha de la curva normal estándar (y por lo tanto un área $1-\alpha$ ente $-z_{\alpha/2}$ y $z_{\alpha/2}$). Entonces

$$\left[\overline{X}-z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}};\overline{X}+z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right]$$

es un intervalo de confianza para μ con nivel $1-\alpha$.

Ejemplo 1: Se midió el índice de colesterol (en mg/dl) a 25 hombres adultos. Se obtuvo un promedio muestral de 250 mg/dl. Asumiremos que la distribución del índice de colesterol en los hombres adultos es $N(\mu, \sigma^2)$. Supongamos que se conoce la varianza del índice de colesterol de los hombres adultos siendo su valor 0.01.

1. Hallar un intervalo de confianza de nivel 95 % para la esperanza del índice de colesterol en hombres adultos. Calcular la longitud del intervalo obtenido.

2. ¿A cuántas personas debería realizarse el estudio si se quiere que el error al estimar la esperanza del índice de colesterol en hombres adultos mediante un intervalo de confianza de nivel 99 % sea de longitud menor que 0,005?

1.2. Intervalos de confianza para la esperanza de una población normal con varianza desconocida

Supongamos que tenemos $X_1, X_2, \ldots, X_n$ variables aleatorias con distribución $N(\mu, \sigma^2)$ con σ^2 desconocida. Sabemos que entonces

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

de donde, estandarizando, obtenemos

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Ahora bien, como σ^2 es desconocido $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ no nos sirve para construir un intervalo de confianza para μ así que lo reemplazamos por su estimardor S^2 y vale el siguiente resultado

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} = \sqrt{n}\frac{\overline{X} - \mu}{S} \sim T_{n-1},$$

donde T_{n-1} es la distribución t de Student con $n-1$ grados de libertad.

Si llamamos $t_{n-1;\alpha/2}$ al valor que deja un área de $\alpha/2$ en la cola a la derecha de la función de densidad T_{n-1}, obtenemos:

$$P\left(\overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

y por lo tanto, el intervalo

$$\left[\overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}; \overline{X} + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\right]$$

es un intervalo de confianza para μ con nivel $1 - \alpha$.

Ejemplo 2: Se midieron las tallas (en cm) a los 12 meses de edad de 16 niñas. Se obtuvieron los siguientes valores $\overline{x} = 73,85$ y $s = 2,58$. Se puede suponer que la talla es una variable aleatoria con distribución normal. Construir un intervalo de confianza de nivel 99 % para la talla media de niñas de 12 meses de edad.

1.3. Intervalos de confianza de nivel asintótico para la media

La limitación que tienen los modelos anteriores es que suponen distribución normal. Gracias al Teorema Central del Límite vamos a poder calcular intervalos de confianza para μ aunque no tengamos ninguna información previa sobre la distribución de la variable. Pero estos intervalos sólo van a valer para n "grande". A continuación definimos intervalos de nivel asintótico para un parámetro θ.

Definición: sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de de una distribución que depende de un parámetro θ desconocido. Dadas dos funciones de la muestra f y g tales que

$$\lim_{n\to\infty} P(f(X_1, X_2, \ldots X_n) \leq \theta \leq g(X_1, X_2, \ldots X_n)) = 1 - \alpha,$$

el intervalo el intervalo $[f(X_1, X_2, \ldots X_n); g(X_1, X_2, \ldots X_n)]$ se denomina intervalo de confianza de nivel asintótico $1 - \alpha$ para el parámetro θ.

Consideramos el siguiente modelo: $X_1, X_2, \ldots X_n$ v.a. iid con cualquier distribución. Llamamos $\mu = E(X_i)$ y $\sigma^2 = V(X_i)$. El estimador de μ es $\overline{X}$, por el TCL cumple

$$\overline{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right),$$

estandarizando, obtenemos

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

Para "n grande"vale que $\sigma \approx S$ y por lo tanto

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} = \sqrt{n}\frac{\overline{X} - \mu}{S} \approx N(0,1).$$

Entonces para n grande vale que

$$P\left(\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right) \approx 1 - \alpha$$

y, por lo tanto

$$\left[\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}; \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

es un intervalo de confianza de nivel asintótico $1 - \alpha$ para μ.

Ejemplo 3: Se quere conocer el contenido medio de calorías en las salchicas marca Acme. Con ese fin, se toma una muestra de 50 salchichas y se miden las calorías presentes en cada una de ellas. Se calcula el desvío estandar muestral y el promedio muestral siendo esos números $3,557$ y $83,06$ respectivamente. Hallar un intervalo de confianza de nivel 99 % para el contenido medio de calorías en este tipo de salchichas. Explique de que tipo es el intervalo usado.

1.4. Intervalos de confianza con nivel asintótico para una proporción

Supongamos que tenemos $X_1, X_2, \ldots, X_n$ una muestra aleatoria $X_i \sim Be(p)$. Vimos que los estimadores de la proporción poblacional son

$$\hat{p} = \overline{X}$$

y

$$\widehat{V(X)} = \hat{p}(1-\hat{p}) = \overline{X}.\left(1-\overline{X}\right).$$

Entonces, para n grande $p \approx \overline{X}$. Por otro lado, sabemos que si n es grande por el Teorema Central del Límite resulta que

$$\overline{X} \approx N(p, p(1-p)/n).$$

Luego

$$\frac{\overline{X}-p}{\sqrt{p(1-p)/n}} \approx \frac{\sqrt{n}(\overline{X}-p)}{\sqrt{\overline{X}(1-\overline{X})}} \approx N(0,1).$$

Con lo cual

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}} \leq p \leq \overline{X} + z_{\alpha/2}\frac{\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}}\right) \approx 1-\alpha$$

y, por lo tanto,

$$\left[\overline{X} - z_{\alpha/2}\frac{\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}}; \overline{X} + z_{\alpha/2}\frac{\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}}\right]$$

es un intervalo de confianza para p con nivel de confianza asintótico $1-\alpha$.

Ejemplo 4: Un estudio mostró que de 1010 encuestados elegidos al azar, 750 estaban a favor de la planificación familiar en hospitales públicos. Interesa estimar mediante un intervalo de confianza a la proporción p de individuos a favor de dicha medida.

1. Hallar un intervalo de confianza de nivel $0,95$ para p. ¿Es de nivel exacto o asintótico?
2. Hallar el tamaño de muestra necesario para que la longitud del intervalo al estimar a p mediante un intervalo de confianza de nivel 0,95 sea menor a $0,005$.

1.5. Intervalos de confianza para la varianza de una población normal

Supongamos que tenemos $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población normal y nos interesa calcular un intervalo de confianza para σ^2, la varianza de las mediciones. Vimos que

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1},$$

donde χ^2_{n-1} es una distribución chi-cuadrado con $n-1$ grados de libertad.

Siguiendo la notacion usada con la distribución normal y con la distribución Student, obtenemos que

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

y, por lo tanto,

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

es un intervalo de confianza para σ^2 con nivel $1-\alpha$.

Ejemplo 5: Supongamos que $X_1, X_2, \ldots, X_{31}$ es una muestra aleatoria de una población normal y que $s^2 = 35$. Hallar un intervalo de confianza para σ^2 de nivel 0,95.