Distribuciones fundamentales de muestreo

Gustavo Ahumada

Intervalo de confianza para la media poblacional

Ejemplo: Se encuentra que la concentración promedio de zinc que se saca del agua a àrtir de una muestra de mediciones de sinc en 36 sities diferentes es de 2.6 gramos ppor mililitro. Encuentre los intervalos de confianza de 95% y 99% para la concentración de zinc en el río. Soponga que la desvación estándar de la población es de 0.3.

Solución: La estimación puntual de μ es $\bar{x} = 2.6$. El valor z, que deja un area de 0.025 a la derecha y por tanto un área de 0.975 a la izquierda, es $z_{0.025} = 1.96$ De aquí que el intervalo de confianza de 95% es

$$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right),$$

$$2.50 < \mu < 2.7.$$

Para encontrar un intervalo de confianza de 99%, encontramos el valor z que deja un área de 0.005 a la derecha y de 0.995 a la izquierda. Por lo tanto, empleando la table del área bajo la curva normal de nuevo, $z_{0.005} = 2.575$, y el intervalo de confianza de 99% es

$$2.47 < \mu < 2.73.$$

Vemos ahora que se requiere un intervalo más grande para estimar μ con un grado más alto de precisión.

Teorema: Si se emplea $\bar{x}$ como una estimación de μ, podemos tener una confianza de $(1-\alpha)100\%$ de que el error no excederá de $z_{\alpha/2}\sigma/\sqrt{n}$.

Teorema: Si $\bar{x}$ se emplea como estimación de μ, podemos tener $(1-\alpha)100\%$ de confianza de que el error no excederá una cantidad especifica e cuando el tamaño de la muestra es

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$$

Ejemplo: ¿Qué tan grande se requiere una muestra en el ejemplo anterior si se quiere tener 95% de confianza de que nuestra estimación de μ difiera por menos de 0.05?

Solución:

$$n = \left(\frac{(1.96)(0.3)}{0.05}\right)^2 = 138.3.$$

Con frecuencia intentamos estimar la media de una población cuando se desconoce la varianza. **Debe recordar de la unidad 8.8** que si tenemos una muestra aleatoria a partir de una *distribución normal*, entonces la variable aleatoria

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene una distribución t de Student con $n-1$ grados de libertad. Aquí S es la desviación estándar de la muestra. En este caso donde se desconoce σ se puede utilizar T para estimar un intervalo de confianza de μ.

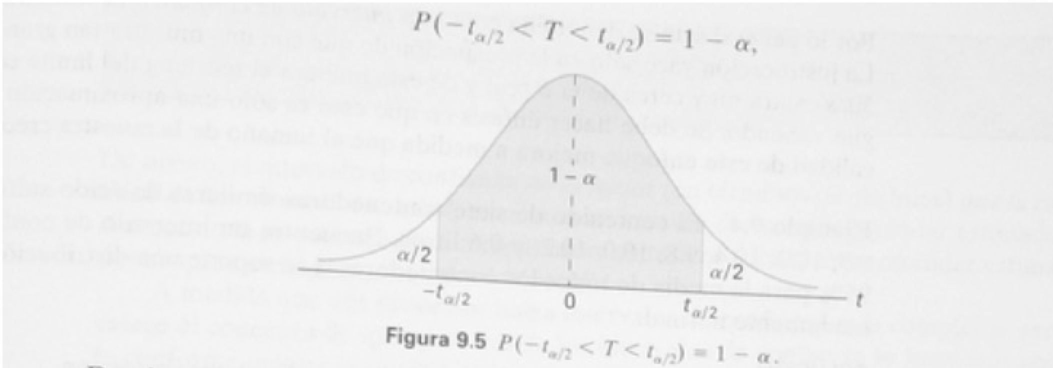

Figure 1: Distribución muestral de diferentes estimadores

El procedimiento es el mismo que cuando se conoce σ excepto que se reemplaza σ por S y la distribución normal estándar por la distribución t.

$$P(\bar{X} - t_{\alpha/2}S/\sqrt{n} < \mu < \bar{X} + t_{\alpha/2}S/\sqrt{n}) = 1 - \alpha.$$

Para nuestra muestra aleatoria de tamaño n, se calcula la media $\bar{x}$ y la desviación estándar s y se obtiene el siguiente intervalo de confianza de $(1-\alpha)100\%$ para μ

$$\bar{x} - t_{\alpha/2}s/\sqrt{n} < \mu < \bar{x} - t_{\alpha/2}s/\sqrt{n},$$

donde $t_{\alpha/2}$ es el valor t con $v = n-1$ grados de libertad, que deja un área a la derecha de $\alpha/2$.

Dos muestras: estimación de la diferencia enrtre medias

Intervalo de confianza para $\mu_1 - \mu_2$; con σ_1 y σ_2 conocidas: Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones con varianzas conocidas σ_1 y σ_2, respectivamente, un intervalo de confinaza de $(1-\alpha)100\%$ para $\mu_1 - \mu_2$ está dado por

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2}{n_2}},$$

donde $z_{\alpha/2}$ es el valor z que deja un área de $\alpha/2$ a la derecha.

Intervalo de confianza para $\mu_1 - \mu_2$; con $\sigma_1 = \sigma_2$ pero desconocidas: Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones con varianzas iguales pero desconocidas, intervalo de confianza de $(1-\alpha)100\%$ para $\mu_1 - \mu_2$ está dado por

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

donde S_p es la estimación de unión de la desviación estándar poblacional y $t_{\alpha/2}$ es el valor t con $v = n_1 + n_1 - 2$ grados de libertad, que deja un área de $\alpha/2$ a la derecha.

$$S_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_1 - 2}$$

Intervalo de confianza para $\mu_1 - \mu_2$; con $\sigma_1 \neq \sigma_2$ pero desconocidas: Si $\bar{x}_1$ y s_1 y $\bar{x}_2$ y s_2 son las medias y las varianzas de muestras aleatorias independientes de tamaño n_1 y n_2 , respectimente,de distribuciones aproximadamente normales con varianzas diferentes y desconocidas, intervalo de confianza de $(1-\alpha)100\%$ para $\mu_1 - \mu_2$ está dado por

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1-1)] + [(s_2^2/n_2)^2/(n_2-1)]},$$

grados de libertad, que deja $\alpha/2$ a la derecha.

Una sola muestra: estimación de una proporción

Intervalo de confianza para p de una muestra grande: Si $\hat{p}$ es la proporción de éxitos en una muestra aleatoria de tamaño n, y $\hat{q} = 1 - \hat{p}$, un intervalo de confianza $(1-\alpha)100\%$ para el parámetro binomial p está dado por

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

donde $z_{\alpha/2}$ es el valor z que deja un área de $\alpha/2$ a la derecha.

Ejemplo: En una muestra aleatoria de $n = 500$ familias que viven en la ciudad de Antofagasta, Chile, se encuentran que $x = 340$ tiene una suscripción a *Netflix*. Encuentre un intervalo de confianza de 95% para la proporción real de familias en esta ciudad con suscripción a *Netflix*

Solución: La estamación puntual de p es $\hat{p} = 340/500 = 0.68$. Con el uso de la tabla del área bajo la curva normal, encontramos que $z_{0.025} = 1.96$. Por lo tanto, el intervalo de confianza de 95% para p es

$$0.68 - 196\sqrt{\frac{(0.68)(0.32)}{500}} < p < 0.68 + 196\sqrt{\frac{(0.68)(0.32)}{500}},$$

que se simplifica a $0.68 < p < 0.72$.

Dos muestras: estimación de la diferencia entre dos proporciones

Intervalo de confianza para $p_1 - p_2$ de una muestra grande: Si $\hat{p_1}$ y $\hat{p_1}$ son las proporciones de éxitos en muestras aleatorias de tamaño n_1 y n_2, y $\hat{q_1} = 1 - \hat{p_1}$ y $\hat{q_2} = 1 - \hat{p_2}$, un intervalo de confianza $(1-\alpha)100\%$ para la diferencia de dos parámetros binomiales $p_1 - p_2$ está dado por

$$(\hat{p_1} - \hat{p_2}) - z_{\alpha/2}\sqrt{\frac{\hat{p_1}\hat{q_1}}{n_1} + \frac{\hat{p_2}\hat{q_2}}{n_2}} < p_1 - p_2 < (\hat{p_1} - \hat{p_2}) + z_{\alpha/2}\sqrt{\frac{\hat{p_1}\hat{q_1}}{n_1} + \frac{\hat{p_2}\hat{q_2}}{n_2}}$$

donde $z_{\alpha/2}$ es el valor z que deja un área de $\alpha/2$ a la derecha.

Ejemplo: Se consideran cierto cambio de proceso de fabricación de partes de las partes de un determinado componente. Se toman muestras del procedimiento existente y del nuevo para determinar si éste tiene como resultado una mejoría. Si se encuentra que 75 de 1500 artículos del procedimiento actual son defectuosos y 80 de 2000 artículos del procedimiento nuevo también lo son, encuentre un intervalo de confianza de 90% para la diferencia real en la fracción de defectuosos entre el proceso actual y el nuevo.

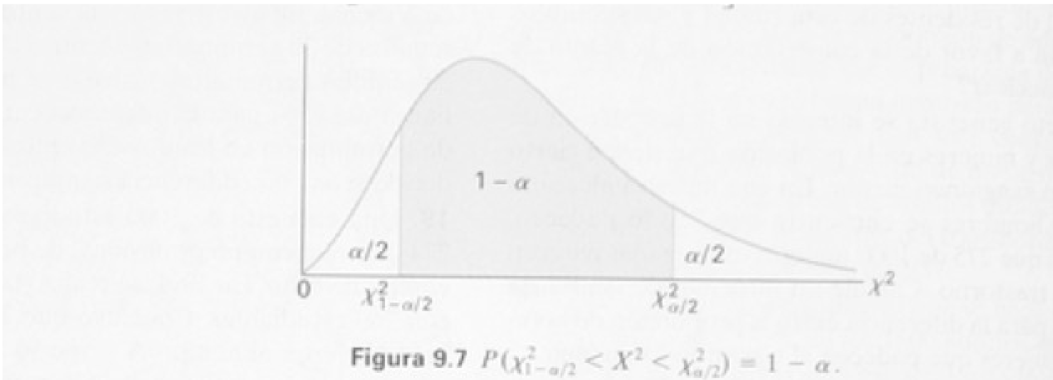

Figure 2: Distribución muestral de diferentes estimadores

Solución: Sean p_1 y p_2 las proporciones reales de defectuosos para los proceso actual y nuevo, respectivamente. De aquí, $p_1 = 75/1500 = 0.05$ y $p_2 = 80/2000 = 0.04$, y la estimación puntual de $p_1 - p_2$ es

$$\hat{p_1} - \hat{p_2} = 0.05 - 0.04 = 0.01$$

Con el uso de la tabla del área bajo la curva normal, encontramos $z_{0.05} = 1.645$. Por lo tanto, al sustituir en esta fórmula obtenemos el intervalo de confianza de 90%

$$0.01 - 1.645\sqrt{\frac{(0.05)(0.95)}{1500} + \frac{(0.04)(0.96)}{2000}} < p_1 - p_2 < 0.01 + 1.645\sqrt{\frac{(0.05)(0.95)}{1500} + \frac{(0.04)(0.96)}{2000}},$$

que se simplifica a $-0.0017 < p_1 - p_2 < 0.0217$. Como el intervalo contiene el valor 0, na hay razón para creer que el nuevo procedimiento producirá una disminución significativa en la proporción de artículos defetusos comparado con el método existente.

Una sola muestra: estimación de la varianza

Si se extrae una muestra de tamaño n de una población normal con varianza σ^2, y se calcula la varianza muestral s^2 obtenemos un valor de la estadística S^2. Esta varianza muestral calculada se usará como estimación puntual de σ^2. Por ello la estadística S^2 se llama estimador de σ^2.

Se puede establecer una estimación por intervalos de σ^2 medianyte el uso de la estadística

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}.$$

La estadística X^2 tiene una distribución ji cuadrado con $n - 1$ grados de libertad cuando las muestran se eligen de una población normal. Podemos escribir

$$P(X^2_{1-\alpha/2} < \chi^2 < \chi^2_{\alpha/2}) = 1 - \alpha,$$

$$P[\chi^2_{1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2}] = 1 - \alpha,$$

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

Intervalo de confianza para σ^2 Si s^2 es la varianza de una muestra aleatoria de tamaño n de una población normal, un intervalo de confianza de $(1-\alpha)100\%$ para σ^2 es

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

donde $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ son los valores χ^2 con $v = n-1$ grados de libertad, que dejan áreas de $\alpha/2$ y $1-\alpha/2$, respectivamente, a la derecha.