InnOvaciOnes de NegOciOs 5(2): 299 -316, 2009

Hacer estimaciones estadísticas
(Making statistical estimations)

Badii, M.H. & K. Cortez
UANL, San Nicolás, N.L., México, mhbadii@yahoo.com.mx

Keywords: Estimations, sampling, statistics

Abstract. The notion of statistical estimation both in terms of point and interval is described. The criteria of a good estimator are noted. The procedures to calculate the intervals for the mean, proportions and the difference among two means as well as the confidence intervals for the probable errors in statistics are provided.

Palabras clave: Estadística, estimación, muestreo

Resumen. En la presente investigación se describen la noción de la estimación estadística, tanto de tipo puntual con de forma de intervalo. Se presentan los criterios que debe reunir un estimador bueno. Se notan con ejemplos, la forma de calcular la estimación del intervalo para la media, la proporción y de la diferencia entre dos medias y los intervalos de confianza para los errores probables.

Introducción

El material sobre teoría de probabilidad constituye la base de la inferencia estadística, rama de la estadística que tiene que ver con el uso de los conceptos de la probabilidad para tratar con la toma de decisiones en condiciones de incertidumbre. La inferencia estadística está basada en la *estimación*, y en la *prueba de hipótesis*. Tanto en la estimación como en la prueba de hipótesis, se hace inferencias acerca de ciertas características de las poblaciones a partir de la información contenida en las muestras (Badii & Castillo, 2007, 2009).

En este trabajo introducimos métodos que nos permiten estimar con precisión razonable la *porción de la población* y la *media de la población*. Calcular la porción exacta o la media exacta sería una meta imposible de obtener, pero, a pesar de ello, seremos capaces de hacer una estimación, de hacer una afirmación con respecto al error que probablemente acompañará a tal estimación, y de poner en marcha algunos controles para evitar lo más que se puede de dicho error. Como tomadores de decisiones, en ocasiones, nos veremos forzados, a confiar en nuestros presentimientos. Sin embargo, en otras situaciones, en las cuales se tenga disponible información y podamos aplicar los conceptos de la estadística, podemos desempeñamos de mejor manera (Badii et al., 2007a, 2007b).

Estimadores

Pueden dividirse los procedimientos de estimación en dos tipos, *estimación puntual* y *estimación por intervalo*. Supongamos que en un ecosistema de pinos se estima la altura media de las plantas mediante un solo número, por ejemplo 8.75 metros, o podríamos afirmar que la altura de los árboles varía en un intervalo de 6.45 a 10.15 metros. El primer tipo se llama estimación puntual, ya que se puede asociar al único número que presenta la estimación, un punto sobre una recta. El segundo tipo se llama estimación por intervalo, porque se tienen dos puntos que definen un intervalo sobre una recta. Consideramos ambos método de estimación.

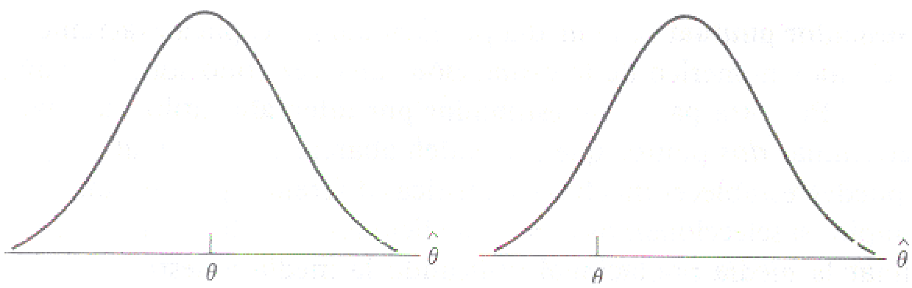
Un procedimiento de estimación puntual utiliza la información de una muestra para llegar a un solo número, o punto, que se estima el parámetro de interés. La estimación real se efectúa mediante un *estimador*. Un *estimador* es una regla que expresa cómo calcular la *estimación*, basándose en la información de la muestra y se enuncia, en general, mediante la media muestral.

Si la media de la distribución de muestreo de un estadístico es igual que la del correspondiente parámetro de la población, el estadístico se llama un *estimador sin sesgo* del parámetro ($\mu_x = \mu$) si no, se llama un *estimador sesgado* ($\mu_x \neq \mu$). Los correspondientes valores de tales estadísticos se llaman *estimaciones sin sesgo* y *sesgadas*, respectivamente.
Por ejemplo, la media muestral:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (1)$$

es un *estimador puntual* de la media poblacional y explica exactamente cómo puede obtenerse el valor numérico de la estimación, una vez conocido los valores muestrales $x_1, x_2, \ldots x_n$. Por otra parte, un estimador por intervalo, utiliza los datos de una muestra para determinar dos puntos que pretenden abarcar el valor real del parámetro estimado.

Si $\hat{\theta}$ es un estimador de un parámetro θ y si la media de la distribución de $\hat{\theta}$ es θ, es decir, si E($\hat{\theta}$) = θ, entonces se dice que $\hat{\theta}$ es no sesgado. Las distribuciones muestrales para un estimador no sesgado y un estimador sesgado, se indican en las Figuras 1 *a* y *b*. Nótese que la distribución muestral para el estimador sesgado está desplazada hacia la derecha de la θ. Este estimador sesgado sobreestimará probablemente $\hat{\theta}$.



a) Estimador no sesgado. b) Estimador sesgado.

Figura 1. Distribución de estimación no sesgado y sesgado.

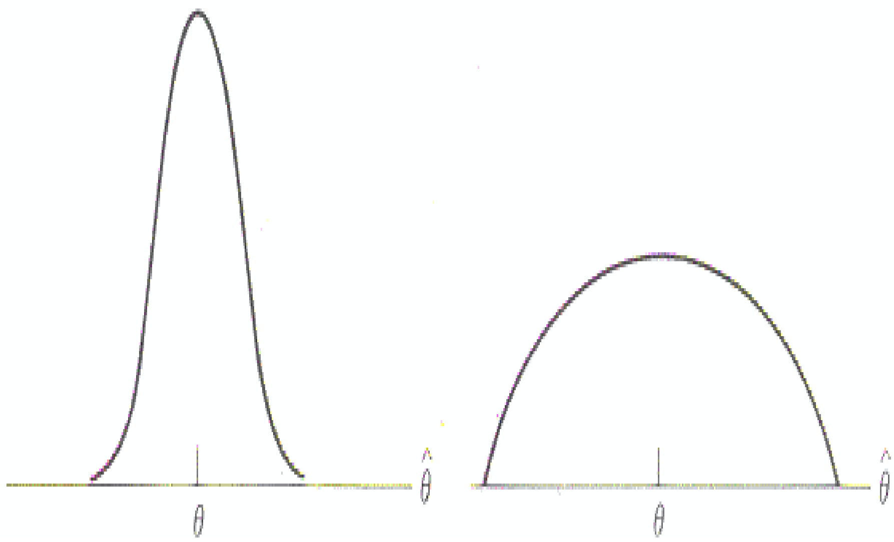
La segunda propiedad deseable de un estimador es que la extensión o alcance (medida por la variancia) de la distribución muestral del estimador, sea lo más pequeña posible. Esto asegura que hay una probabilidad alta de que una estimación individual se encuentra cerca de la θ. Se indican las distribuciones muestrales para dos estimadores, uno con una variancia pequeña y otro con una variancia mayor, en las Figuras 2 *a* y *b*, respectivamente. Si utiliza el término varianza de un estimador, la media de las distribuciones muestrales de variancia se calcula como:

$$\mu_{s^2} = [(n - 1) / n]\, \sigma^2 \qquad (2)$$

Donde, σ^2 es la varianza de la población y *n* es el tamaño de la muestra. La varianza de la muestra es una estimación sesgada de la varianza de la población. Usando la varianza modificada, tenemos:

$$\hat{s}^2 = \frac{n}{n-1} s^2 \qquad (3)$$

Estimaciones Estadísticas



a) Estimador con varianza pequeña. b) Estimador con varianza mayor.

Figura 2. Comparación de la variabilidad de estimadores.

Encontramos $\mu_{s^2} = \sigma^2$, de manera que $\hat{s}^2$ es una estimación sin sesgo de la σ^2. Sin embargo, $\hat{s}$ es una estimación sesgada de la σ.

Si las distribuciones de muestreo de dos estadísticos tienen la misma media (o esperanza), el de menor varianza se llama un *estimador eficiente* de la media, mientras que el otro se llama un *estimador ineficiente.* Los valores correspondientes de los estadísticos se llaman *estimación eficiente* o *estimación ineficiente*, respectivamente. Si consideramos todos los posibles estadísticos cuyas distribuciones de muestreo tienen la misma media, aquel de varianza mínima se llama el *estimador de máxima eficiencia*, o sea, *el mejor estimador*.

La bondad de un estimador por intervalo se analiza de manera muy similar a la de un estimador puntual. Se seleccionan muestras del mismo tamaño, respectivamente, y se determina el intervalo de estimación para cada proceso. Este método generará un gran número de intervalos, en vez de puntos. Una buena estimación por intervalo contendrá, con éxito, el valor real del parámetro para una fracción grande del tiempo. Tal fracción se denomina *coeficiente de confianza* para el estimador; el estimador mismo se llama, a menudo, *intervalo de confianza*.

Criterios de selección del estimador

Algunas estadísticas son mejores estimadores que otras. Afortunadamente, podemos evaluar la calidad de una estadística como estimador mediante el uso de cuatro criterios:

1. **Ausencia de sesgo**. Se refiere al hecho de que una media de muestra es un estimador no sesgado de una media de población porque la media de la distribución de muestreo de las medias de muestras tomadas de la misma población es igual a la media de la población misma. Podemos decir que una estadística es un estimador sin sesgo o imparcial si, en promedio, tiende a tomar valores que están por encima del parámetro de la población que se está estimado con la misma frecuencia y la misma extensión con la que tiende a asumir valores por debajo del parámetro de población que se está estimando.

2. **Eficiencia.** La *eficiencia* se refiere al tamaño del error estándar de la estadística. Si comparamos dos estadísticas de una muestra del mismo tamaño y tratamos de decidir cuál de ellas es un estimador más eficiente, escogeríamos la estadística que tuviera el menor error estándar o la menor desviación estándar de la distribución de muestreo.

3. **Consistencia.** Una estadística es un estimador coherente de un parámetro de la población si al aumentar el tamaño de la muestra, se tiene casi la certeza de que el valor de la estadística se aproxima bastante al valor del parámetro de la población. Si un estimador es coherente, se vuelve más confiable si tenemos tamaños de muestra más grandes.

4. **Suficiencia.** Un estimador es suficiente si utiliza una cantidad de la información contenida en la muestra que ningún otro estimador podría extraer información adicional de la muestra sobre el parámetro de la población que se esta estimando.

Presentamos estos criterios con anticipación para hacerlo consciente del cuidado que los estadísticos deben tener a la hora de seleccionar un estimador. Una estadística de muestra dada no siempre es el mejor estimador de su parámetro de población correspondiente. Considere una población distribuida de manera simétrica, en la que los valores de la mediana y de la media coinciden. En este caso, la media de la muestra sería un estimador sin sesgo de la mediana de la población debido a que asumiría valores que en promedio serían iguales a mediana de la población. También la media de la muestra sería un estimador *consistente* de la mediana de la población puesto que, conforme aumenta el tamaño de la muestra, el valor de la media de la muestra tendrá a acercarse bastante a la mediana de la población. Y la media de la muestra sería un estimador más *eficiente* de la mediana de la población que la mediana de la muestra misma, ya que en muestras grandes, la media de la muestra tiene una desviación estándar menor que la de la mediana de la muestra. Al mismo tiempo, la mediana de la muestra de una población distribuida simétricamente sería un estimador imparcial y consistente de la media de la población, pero *no el más eficiente* estimador porque en muestras grandes su error estándar es mayor que el de la media de la muestra.

Estimación puntual

Como ya se mencionó las medias muestrales, los totales muestrales y las proporciones muestrales tienen distribuciones de muestreo, con propiedades comunes. Las estadísticas mismas son estimadores no sesgadas de sus equivalentes poblacionales, y sus distribuciones de muestreo son aproximadamente normales cuando el tamaño de muestras es grande. Este fenómeno no restringe solamente a las estadísticas discutidas en este trabajo. Muchas otras estadísticas, sobre todo las obtenidas a partir de sondeos de opiniones, tienen distribuciones muestrales que no pueden definir claramente para tamaños de muestra pequeños, pero poseen distribuciones muestrales que tienen forma de montículo, casi aproximadamente normales, cuando el tamaño de muestra es grande. Por tal razón, el procedimiento para evaluar la bondad (es decir, la confiabilidad o exactitud) de cualquier estos estimadores, es lo mismo para cualquier otro estimador. Por ejemplo, supóngase que se desea estimar un parámetro poblacional utilizando un estimador, representado por el símbolo $\hat{\theta}$. Además, supondremos que la distribución muestral de $\hat{\theta}$ es aproximadamente normal, y que $\hat{\theta}$ es un estimador no sesgado de θ, y que se conoce la desviación estándar del estimador, o que se puede aproximarla, y se la presenta por el símbolo σ_{θ}. Si se selecciona una muestra aleatoria de *n* observaciones de la población, y se utiliza $\hat{\theta}$ para calcular una estimación de θ, ¿qué tan exacta será la estimación?

La gráfica de la distribución muestral normal de $\hat{\theta}$, que se tiene en la Figura 3, nos ayudará a contestar esta pregunta. Supóngase que la estimación se encuentra en el punto marcado por una flecha. Tal estimación particular se ubica a la derecha de θ, y por lo tanto, sobrestima θ en una cantidad ($\hat{\theta}$ - θ). El valor absoluto de esta desviación, denotado por $|\hat{\theta} - \theta|$, se llama *error de estimación.*

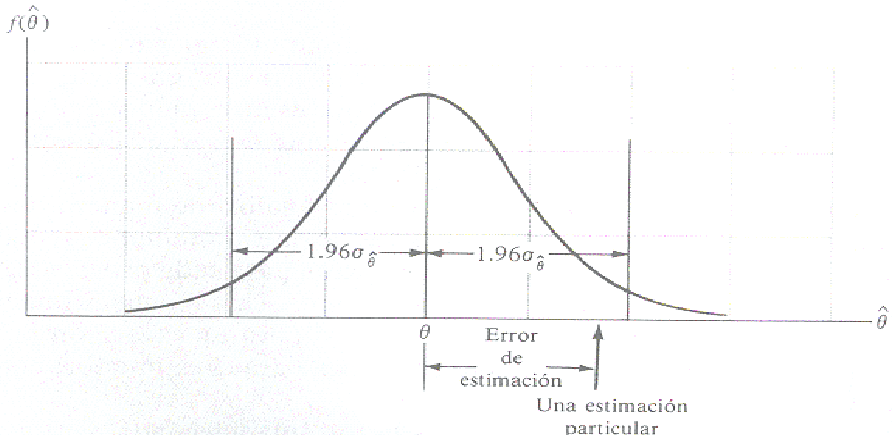


Figura 3. Distribución de muestreo de un estimador "θ".

Ejemplo 1. En una comunidad vegetal, se determinó la tasa de crecimiento anual de una especie de cedro. Se seleccionó una muestra aleatoria de *n* = 50 árboles, y se registró la tasa de crecimiento para cada uno. La media y la desviación estándar de las 50 tasas eran: $\bar{x} =$ 9.1% y s = 0.24. Estime la media de la tasa de crecimiento para la comunidad, y evalúe la exactitud de la estimación.

Solución: Para este ejemplo, el parámetro θ que se desea estimar, es una media poblacional "μ". El estimador puntual $\hat{\theta}$ es la media muestral $\bar{x}$, y la estimación puntual de μ es $\bar{x}$ = 9.1%. Ya que la media muestral satisface el estimador no sesgado de la μ, y su distribución de muestreo es aproximadamente normal, para valores grandes del tamaño de muestra *n*, la cota para el error de estimación es.

$$\text{Cota para el error de estimación} = 1.96\,\sigma_{\bar{x}} = \frac{1.96\sigma}{\sqrt{n}} = \frac{1.96(0.24)}{\sqrt{50}} = 0.07$$

En resumen, la estimación de la media de la tasa de crecimiento en la comunidad, es 9.1%. ¿Qué tan exacta es esta estimación? No lo sabemos realmente. Puede sobrestimar o puede subestimar la media de la tasa de crecimiento. Lo que sí se sabe es que, la probabilidad de que el error sea menor que 0.07%, será aproximadamente 0.95 (Figura 4). Así, la cota para el error de estimación, 0.07%, proporciona una medida de la exactitud para la estimación efectuada en esta comunidad.

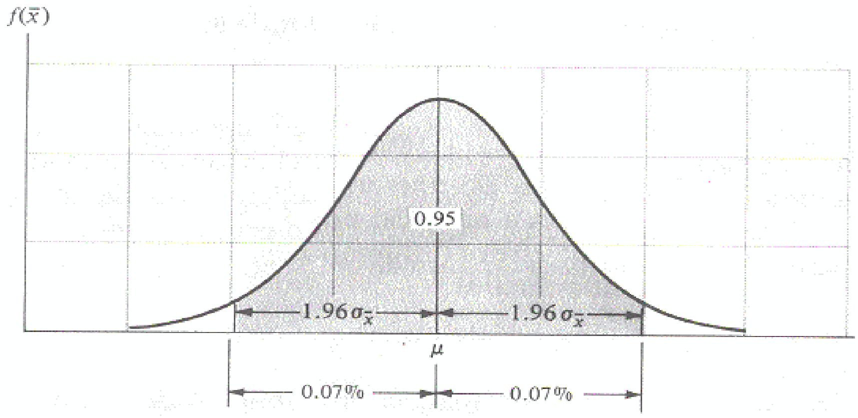


Figura 4. Cota para el error de estimación del Ejemplo 1.

Estimaciones de intervalo

Una estimación de intervalo describe un intervalo de valores dentro del cual es posible que esté un parámetro de población. Consideramos el concepto de intervalo de confianza, cómo se obtiene un tal intervalo para una media poblacional μ a partir de una muestra aleatoria de *n* observaciones.

Se puede determinar un estimador por intervalo, utilizando cualquier estimador puntual que sea no sesgado y que posea una distribución muestral aproximadamente normal. Para ver cómo puede determinarse un intervalo de confianza para "θ", examinemos la distribución muestral de $\hat{\theta}$ en la Figura 5. Supóngase que tuviéramos que sacar una muestra aleatoria de *n* observaciones de la población y utilizar los datos de la muestra para calcular una estimación de "θ". Se indica una estimación puntual particular, indicada por una flecha en la Figura 5, que se cae dentro de $1.96\sigma_{\hat{\theta}}$ de la θ. Puede observarse que el intervalo de ($\hat{\theta} \pm 1.96\sigma_{\hat{\theta}}$), incluye a la θ. La estimación puntual $\hat{\theta}$, caerá dentro de $1.96\sigma_{\hat{\theta}}$ de la θ con una probabilidad igual a 0.95.

Sean θ y σ_θ la media y la desviación estándar (error estándar) de la distribución de muestreo de un estadístico. Entonces, si la distribución de muestreo es aproximadamente normal podemos esperar hallar un estadístico muestral real que esté en los intervalos $\theta_s \pm \sigma_{\theta,}$, $\theta_s \pm 2\sigma_\theta$, o $\theta_s \pm 3\ \sigma_\theta$ alrededor del 68.27%, 95.45% y 99.73% del tiempo, respectivamente. Por esa razón, llamamos a esos porcientos los *intervalos de confianza*. Los números extremos de estos intervalos se llaman entonces los *límites de confianza* (68.27%, 95.45% y 99.73%) o *límites fiduciales*.

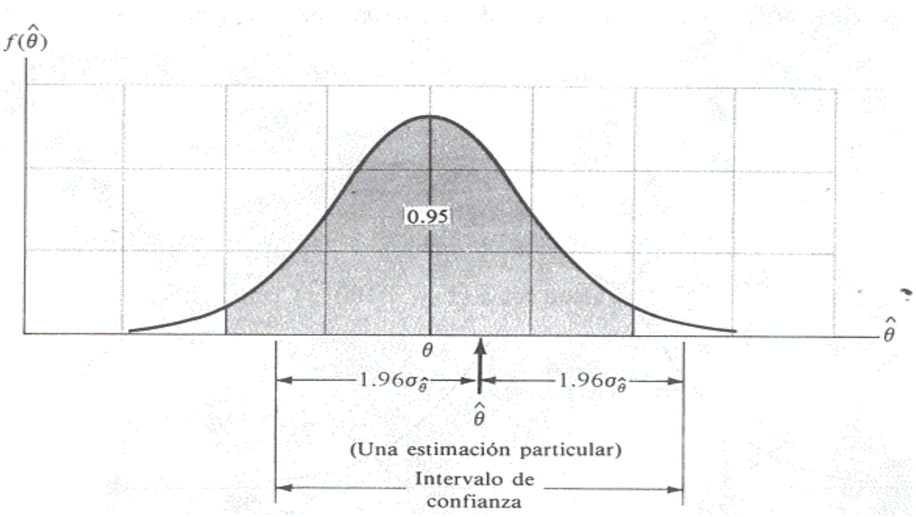


Figura 5. Un intervalo de confianza para la θ.

Análogamente, $\hat{\theta}$ ± 1.96ds y $\hat{\theta}$ ± 2.58ds son los límites de confianza 95% y 99% para la $\hat{\theta}$. El porcentaje de confianza se suele llamar nivel de confianza. Los números (1.96, 2.58, etc.) en los límites de confianza se llaman *coeficientes de confianza* o *valores críticos*, y se denotan por Z_c. De los niveles de confianza podemos deducir los coeficientes de confianza y viceversa (Tabla 1).

Tabla 1. Cálculo de valores de Z_c (coeficiente de confianza) correspondiente a varios niveles de confianza.

Nivel de confianza	99.73	99	98	96	95.45	95	90	80	68.27	50%
Z_c	3.00	2.58	2.33	2.05	2.00	1.96	1.645	1.28	1.00	0.6745

En general, si queremos que el coeficiente de confianza sea igual a (1 - α), se emplea el valor *z* igual a $z_{\alpha/2}$, que limita un área de $\alpha/2$ en el extremo superior de la distribución *z* (Figura 5). Se puede encontrar este valor también en la tabla de la distribución normal en cualquier libro estadístico. Entonces, un intervalo de confianza de (1 - α)100% para θ es

$$\text{I.C.} = \hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$$

Ejemplo 2. Encuentre un intervalo de confianza de 90% para la media de la tasa de crecimiento de árboles de cedro, en una comunidad vegetal.

Solución: Ya hemos notado que el estimador puntual $\bar{x}$ de la media poblacional μ, tiene una distribución de muestreo que satisface las propiedades requeridas. Por lo tanto, un intervalo de confianza de 90% para la media de la tasa de crecimiento μ, es

$$\text{I.C.} = \bar{x} \pm z_{0.05}\sigma_{\bar{x}} = \bar{x} \pm 1.645\frac{\sigma}{\sqrt{n}}$$

Al sustituir $\bar{x}$ = 9.1% y n = 50, y utilizar s = *0.24%* para aproximar la σ, obtendremos:

$$\text{I.C.} = 9.1 \pm (1.645)\frac{0.24}{\sqrt{50}} = 9.1 \pm 0.0558$$

Así se ve que la media de la tasa de crecimiento se encuentra entre 9.0442% y 9.1558%. ¿Podemos asegurar que este intervalo particular contiene a la μ? No, pero si hay expectativas de que así es. Si se utiliza el intervalo de confianza para estimar la μ, la probabilidad de que un intervalo contenga la μ, es 0.90.

A parte de intervalos de confianza bilaterales, es posible determinar también *intervalos de confianza unilaterales* para loa parámetros. Un *intervalo de confianza unilateral inferior* para un parámetro θ estimará que θ es mayor que algún límite inferior de confianza (LIC). Un *intervalo de confianza unilateral superior* estimará que θ es menor que algún limite superior de confianza (LSC). El valor z que hay que utilizar para un intervalo de confianza unilateral de $(1 - \alpha)100\%$, z_{α}, localiza α en un solo extremo de la distribución normal, como se ve en la Figura 6. Se indican límites de confianza unilaterales inferior y superior para θ, en la Tabla 2.

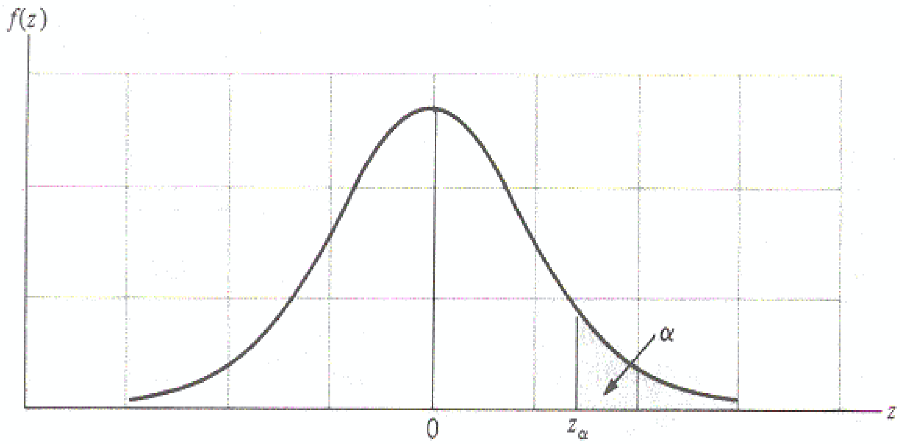


Figura 6. Localización de z_α para un intervalo de confianza unilateral de (1 - α) 100%.

Tabla 2. Límite de confianza unilaterales para la θ.

Coeficiente de confianza	α	z_α	LIC	LSC
0.90	0.10	1.28	$\theta - 1.28\sigma_{\hat{\theta}}$	$\theta + 1.28\sigma_{\hat{\theta}}$
0.95	0.05	1.645	$\theta - 1.645\sigma_{\hat{\theta}}$	$\theta + 1.645\sigma_{\hat{\theta}}$
0.99	0.01	2.33	$\theta - 2.33\sigma_{\hat{\theta}}$	$\theta + 2.33\sigma_{\hat{\theta}}$

Ejemplo 3. El valor promedio de peso y la desviación estándar de camarones cultivados en 40 estanques, eran $\bar{x}$ =10.3 gramos y s = 0.31 gramos. Como solamente interesa el límite superior del peso, hallar un intervalo de confianza unilateral superior de 95% para el peso medio de camarones.

Solución: Puesto que el coeficiente de confianza es 0.95, α = 0.05 y $z_{0.05}$ = 1.645. Por tanto, el intervalo de confianza unilateral de 95%, para μ es:

$$\text{I.C.} = \bar{x} \pm z_{0.05}\sigma_{\bar{x}} = \bar{x} \pm 1.645\frac{\sigma}{\sqrt{n}}$$

Al sustituir $\bar{x}$ = 10.3, n = 40 y s = 0.31, para aproximar la σ, el intervalo de confianza unilateral sera:

$$\text{LSC} = 10.3 + (1.645)\frac{0.31}{\sqrt{40}} = 10.3806$$

La probabilidad de que el intervalo de confianza unilateral contenga a la μ, es 0.95.

Estimar el intervalo de la media

Si el estadístico tiene la media ($\bar{x}$) de la muestra, entonces los límites de confianza para estimar la media μ de la población vienen dados por $\bar{x} \pm 1.96\,\sigma_{\bar{X}}$ (α = 0.05). En términos generales, los límites de confianza para estimar la media de la población vienen dados por $\bar{X} \pm z_c\,\sigma_{\bar{X}}$, donde z_c se puede leer en las tablas correspondientes en los libros estadísticos.

$$\text{I.C.} = \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad (4)$$

Donde:
I.C. = intervalo de confianza.
$z_{\alpha/2}$ = valor z que corresponde al área $\alpha/2$ en el extremo superior de una distribución normal estándar z.
n = tamaño muestral
σ = desviación estándar de la población muestreada

Si el muestreo es de una población finita, el intervalo de confianza se calcula por.

$$\text{I.C.} = \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad (5)$$

En lo que:
N = tamaño de la población conocido.
n = tamaño de la muestra.

Ejemplo 4. Un investigador de un laboratorio de alimentos para cachorros desea estimar la vida media de alimento enlatada en condiciones normales. Se conoce que la desviación estándar de la vida de la población es de seis meses. Suponga que seleccionamos una sola muestra aleatoria de 100 latas con un valor promedio 21 de meses de vida y una desviación muestral de 6 meses. Si el investigador utiliza 10,000 latas al año, ¿cuál es el intervalo de confianza cuando α = 0.05?

Solución: Calcularemos el error estándar de la media haciendo el uso de la ecuación 3.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{100}} = 0.6$$

Con un nivel de 95% de confianza se encuentra la media de la distribución de muestreo

LSC = $\bar{x} + 1.96\sigma_{\bar{x}} = 21 + 1.96(0.6) = 22.18$ meses

y

LIC = $\bar{x} - 1.96\sigma_{\bar{x}} = 21 - 1.96(0.6) = 19.82$ meses

De esta manera podemos informar que la vida media de alimento se encuentra entre 18.82 y 22.18 meses con 95% de confianza.

Ejemplo 5. El departamento de servicio social de una empresa está interesado en estimar el ingreso media anual de 700 familias que viven en una área metropolitana. Se tomó una muestra con las siguientes características.

n = 50 = tamaño de muestra
$\bar{x}$ = $ 11,800 media de la muestra
s = $ 950 desviación estándar de la muestra

En base a estos datos, calcular la desviación estándar estimada de la población.

Solución: La desviación estándar de la población se calcula como:

$$\hat{\sigma} = s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

Ahora podemos estimar el error estándar de la media para un tamaño de población finita como:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} x \sqrt{\frac{N-n}{N-1}}$$

Como estamos calculando el error estándar de la media mediante una estimación de la desviación estándar de la población, volvemos a escribir esta ecuación de modo que quede simbolizada correctamente:

$$\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}_{\bar{x}}}{\sqrt{n}} x \sqrt{\frac{N-n}{N-1}} = \frac{950}{\sqrt{50}} x \sqrt{\frac{700-50}{700-1}} = \$129.57$$

El límite de confianza de 90% se calcula como:

LSC = $\bar{x} + 1.64\hat{\sigma}_{\bar{x}} =$ 11,800 +1.64 ($129.57) = 12,012.50

y

LIC = $\bar{x} - 1.64\hat{\sigma}_{\bar{x}} =$ $11,800 – 1.64 (129.57) = 11,587.50

El informe que podríamos dar al departamento de servicio social sería: con una confianza de 90% estimamos que el ingreso anual promedio de las 700 familias se encuentra entre $11,587.50 y $12,012.50.

Estimar la diferencia entre dos medias

Un problema de igual importancia que la estimación de las medias poblacionales, es la comparación de dos medias. Por ejemplo, quizá se necesite estimar la diferencia entre dos ecosistemas o zonas ecológicas de un país, respecto a la riqueza en biodiversidad.

Para cada uno de estos ejemplos se postulan dos poblaciones, la primera con una media μ_1 y una variancia σ_1^2, y la segunda con una media μ_2 y una variancia σ_2^2. Se toma una muestra aleatoria de n_1 mediciones de la población 1, y n_2 de la población 2, y se supone que las muestras han sido seleccionadas independientemente. Por último, se calcula las estimaciones $\bar{x}_1, s_1^2$, y $\bar{x}_2, s_2^2$ de los parámetros poblacionales, a partir de los datos muestrales.

La diferencia $(\bar{x}_1 - \bar{x}_2)$ entre las medias muestrales es un estimador puntual no sesgado de la diferencia entre las medias poblacionales (μ_1 - μ_2). Como sabemos la distribución muestral del estimador $(\bar{x}_1 - \bar{x}_2)$ será aproximadamente normal para muestras grandes, con una media y una desviación estándar dadas en la siguiente forma:

$$E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2 \qquad (6)$$

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (7)$$

La cota para el error del estimador puntual de (μ_1 - μ_2) = 1.96 $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Se puede obtener un intervalo de confianza bilateral para (μ_1 - μ_2), con "1 - α" coeficiente de confianza, utilizando la siguiente fórmula: *Intervalos de confianza de* (1 - α)100% para (μ_1 - μ_2) en *el caso de muestras grandes*

$$\text{I.C.} = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (8)$$

Se pueden utilizar las varianzas muestrales s_1^2 y s_2^2, cuando se desconocen estos parámetros.

Estimar intervalo de la proporción

Los estadísticos, a menudo, utilizan una muestra para estimar la *porción* de un evento en una población. Las fórmulas para la media y la desviación estándar de la distribución binomial son:

$$\mu = np$$

$$\sigma = \sqrt{npq}$$

En las que:
n = número de ensayos o intentos
p = probabilidad de éxito
q = 1 – *p* = probabilidad de fracaso

Para calcular la media de la distribución de muestreo de la porción de éxito dividimos *np* entre *n* y obtenemos solo el valor *p*. La media, que se encuentra al lado izquierdo de la ecuación se convierte en $\mu_{\bar{x}}$, es decir, en la media de la distribución de muestreo de la proporción de éxitos:

$$\mu_p = p$$

De forma similar podemos modificar la fórmula para la desviación estándar de la distribución binomial, $\sqrt{npq}$, que mide la desviación estándar que existe en el número de éxitos. Para cambiar el número de éxitos a la porción de éxitos, dividimos $\sqrt{npq}$ entre *n* y obtenemos $\sqrt{pq/n}$. En términos estadísticos, la desviación estándar para la porción de éxitos en una muestra se simboliza como:

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}}$$

Ejemplo 6. Considerando que solo 40% de maestros de una universidad prefieren la planificación de sus actividades, se seleccionó una muestra aleatoria de tamaño 75 de población de maestros. Calcular el error estándar estimado de la porción con una precisión de 99%.

Solución: El error estándar estimado de la porción se calcula como:

$$\hat{\sigma}_{\bar{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.4)(0.6)}{75}} = 0.057$$

el límite de confianza se determina como:

LSC = $\bar{p} + 2.58\hat{\sigma}_p$ = 0.4 + 2.58 (0.057) = 0.547 = Límite superior de confianza

y

LIC = $\bar{p} - 2.58\hat{\sigma}_{\bar{x}}$ = (0.4) - 2.58 (0.057) = 0.253 = Límite Inferior de confianza

Por tanto, estimamos, a partir de nuestra muestra de 75 empleados que, con 99% de confianza, creemos que la porción de la población total de maestros que planifican sus propias actividades está entre 0.253 y 0.547.

IC de la diferencia entre dos medias proporcionales

Si el estadístico es la proporción de "éxitos" en una muestra de tamaño "*n*" sacada de una población binomial en la que "*p*" es la proporción de éxitos, entonces los límites de confianza para *p* se dan por $p \pm z_c \, \sigma_p$, donde *p* es la proporción de éxitos en la muestra de tamaño *n.* Si el muestreo es de una población infinita o finita con reposición, los intervalos de confianza son:

$$p \pm z_c \sqrt{\frac{pq}{n}} = p \pm z_c \sqrt{\frac{p(1-p)}{n}} \qquad (9)$$

Si el muestreo es de una población finita de tamaño n_p y sin reposición, los intervalos de confianza son:

$$p \pm z_c \sqrt{\frac{pq}{n}} \sqrt{\frac{n_p - n}{n_p - 1}} \qquad (10)$$

IC para desviaciones estándares y el error probable

Los límites de confianza para la desviación estándar de una población normalmente distribuida (σ), estimados a través de una muestra con desviación estándar igual a la *s*, vienen dados por:

$$s \pm z_c \sigma_s = s \pm z_c \frac{\sigma}{\sqrt{2n}} \qquad (11)$$

Los límites de confianza 50% de los parámetros de población correspondientes a un estadístico *S* vienen dados por $s \pm 0.6745\sigma_s$. La cantidad $0.6745\sigma_s$ se conoce como el *error probable* de la estimación.

Conclusiones

La primera fase de la estadística se trata de coleccionar, ordenar y presentar los datos o hechos numéricos. La segunda parte de la estadística se encarga de analizar, sintetizar (hacer inferencias y realizar interpretación) y finalmente publicar los datos que han sido presentados en forma de grafica y/o de manera tabular. Es precisamente en la sección del análisis estadístico en donde el investigador debe modificar los datos, es decir hacer estimaciones de los datos brutos. Para hacer estimaciones, uno debe estar bien familiarizado con los criterios estadísticos que se debe reunir y considerar en el proceso de la estimación, ya que las estimaciones sesgadas nos conducen a las inferencias y decisiones erróneas. Es precisamente con este punto en la mente que se avoco a realizar la presente investigación.

Referencias

Badii, M.H. & J. Castillo. (eds.). 2007. Técnicas Cuantitativas en la Investigación. UANL, Monterrey.

Badii, M.H. & J. Castillo. 2009. Muestreo Estadístico: Conceptos y Aplicaciones. UANL, Monterrey.

Badii, M.H., J. Castillo, A. Wong & J. Landeros. 2007a. Precisión de los índices estadísticos: técnicas de jacknife & bootstrap. InnOvaciOnes de NegOciOs. 4(1): 63-78.

Badii, M.H., J. Castillo, J. Landeros & K. Cortez. 2007b. Papel de la estadística en la investigación científica. InnOvaciOnes de NegOciOs. 4(1): 107-145.

Badii, M.H., J. Castillos, R. Foroughbakhch & K. Cortez. 2007c. Probability and scientific research. Daena, 2(2): 358-369.

Fuino, Y. 1980. Aproxímate binomial confidence limits. Biométrica 67: 677-681.

Ghosh, B.K. 1979. A comparison of some aproxímate confidence intervals for the bionamial parameter. J. Amer. Statist. Assoc. 74: 894-900.

Pfanzagl, J. 1978. Estimation Confidence Intervals and regions. In kruskal and Tanur 259-267.

Sutton, J. B. 1990. Values of the index of determination at the 5% significance level. Statistician 39: 461-463.