Estimación

El hombre hace estimaciones permanentemente. El gerente de marketing estima las ventas del próximo mes, estima el gerente de producción la cantidad de mano de obra necesaria para un período, el gerente de una entidad bancaria efectúa una estimación sobre el cumplimiento de un cliente para otorgarle de un crédito, estima el ama de casa la cantidad de comida para una cena, etc. En función de esas estimaciones se decide.

En el módulo 1 se define la estadística inferencial con el objetivo final de inferir algo acerca de una población a partir de la información adquirida de una muestra. Y, justamente, la inferencia estadística está basada en la estimación.

En esta lectura, se presenta cómo se puede estimar la media y la proporción de una población, además se estimará el error con que se estima al parámetro, ya que calcular la porción exacta o la media exacta es una meta imposible.

Caso: la jornada laboral de los pilotos de Aerolíneas Argentinas

El promedio de las horas de vuelo de cada piloto de Aerolíneas Argentinas fue de 32,6 horas mensuales, entre enero de 2011 y junio de 2012 (Auditoría General de la Nación, septiembre de 2014). Se supone que esta media se basó en las horas de vuelo de una muestra de 100 pilotos de esa empresa y que la desviación estándar muestral fue 8,5 horas. La población es conocida, 1800 pilotos.

Además, se sabe que los pilotos afiliados al sindicato, dentro de los 100 de la muestra, son 86. Se pide:

1. Realiza una estimación puntual de las horas de vuelo promedio de los pilotos de Aerolíneas Argentinas.
2. Estima la desviación estándar de la población a partir de la desviación estándar de la muestra.
3. Estima el error estándar de la muestra, considerando una población finita.
4. Determina el margen de error con un 95% de confianza.
5. Construye un intervalo de confianza del 95% para la media poblacional de las horas de vuelo de los pilotos.
6. La media en las horas de vuelo de los pilotos de Austral, en ese mismo informe, es de 34,1 horas por mes. Utiliza los resultados del ítem e) para analizar la diferencia entre la cantidad de horas de vuelo de los pilotos en las dos líneas aéreas. Si el costo laboral de Aerolíneas Argentinas es mayor que el de Austral, la información calculada ¿sirve para entender por qué se puede esperar que Aerolíneas Argentinas tenga los costos más elevados?

7

Estima, mediante un intervalo de confianza del 98%, la proporción poblacional de pilotos afiliados al sindicato.

Estimación puntual para la media y para la proporción

Tipos de estimaciones

Ya se estudió que la inferencia estadística utiliza los datos muéstrales para deducir los parámetros poblacionales, es decir, se puede estimar el valor de un parámetro poblacional (de una característica de la población) con los datos de la muestra. A estos datos se los conoce como **estadísticos muéstrales**. De esta manera, se podría estimar la media de la población, su varianza, el desvío estándar y la proporción. Estos son los parámetros más calculados.

Existen dos tipos de estimaciones sobre una población a partir de una muestra:

ESTIMACIÓN PUNTUAL	ESTIMACIÓN POR INTERVALOS

Es cuando al parámetro le asignamos un valor único, un número que se utiliza para estimar un parámetro desconocido de la población.

Si bien la experiencia y los datos históricos ayudan a realizar buenas estimaciones puntuales, hay que considerar, como estimadores puntuales, a los estadísticos que se obtienen de una muestra representativa de la población. Este número, por lo general, se obtiene a través de una fórmula:

- La media de la muestra $\bar{x}$ puede ser un estimador puntual de la media de la población μ.
- La desviación estándar de la muestra s puede ser un estimador puntual de la desviación estándar de la población σ.
- La proporción de la muestra p puede ser un estimador de la proporción de la población p. Al estimador de la proporción de la muestra se lo designa con $\hat{p}$.

ESTIMACIÓN PUNTUAL

ESTIMACIÓN POR INTERVALOS

Se construye un intervalo de estimación para tener información sobre qué tan cerca está la estimación muestral del parámetro poblacional. Este intervalo se crea bajo la premisa de que, con un cierto porcentaje de confianza (que se establece previamente), la media poblacional estará contenida en ese intervalo.

Ejemplos de estimaciones puntuales: el caso de la jornada laboral de los pilotos de Aerolíneas Argentinas, ítems a) y b)

Es de esperarse que las estimaciones puntuales difieran de los correspondientes parámetros poblacionales, justamente, porque son realizadas a partir de una muestra. Estas estimaciones se realizan en la práctica, pero acudiendo a las fórmulas conocidas en la estadística descriptiva.

En el caso presentado ya se dan la media y la desviación estándar muestral. De lo contrario, habría que calcularlas en la muestra y hacer la estimación puntual, a partir de los resultados obtenidos en la muestra.

En el ítem a): se solicita que se realice una estimación puntual de las horas de vuelo promedio de los pilotos de Aerolíneas Argentinas. Las estadísticas dicen que fueron 32,6 horas mensuales en un cierto período de tiempo y que la muestra fue de 100 pilotos. Se puede estimar, a partir de los datos de la muestra, que la media poblacional es de 32,6 horas mensuales. ¿Cómo se obtuvo esta media? Con la fórmula que ya se conoce sobre la definición de media o promedio, aplicada sobre la muestra:

$$\bar{x} = \frac{\sum x_i \cdot f_i}{n} = 32{,}6$$

Por lo tanto, con una estimación puntual, se puede decir que:

$$\mu = 32{,}6$$

Con una estimación puntual a partir de la muestra de 100 pilotos, la media poblacional es de 32,6 horas mensuales.

Del mismo modo se responde a la pregunta b): estimando la desviación estándar de la población a partir de la desviación estándar de la muestra, mediante la fórmula respectiva:

$$s = \sqrt{\frac{(x_i - \bar{x})^2 \cdot f_i}{n-1}} = 8{,}5$$

Por lo tanto, la estimación puntual de la desviación estándar poblacional es:

$$\sigma = 8,5$$

Con una estimación puntual, a partir de la muestra de 100 pilotos, la desviación estándar poblacional es de 8,5 horas mensuales.

Estimación por intervalos de confianza

"Una estimación de intervalo es un rango de valores que se utiliza para estimar un parámetro de la población" (Levin y Rubin, 2004, p. 281).

Una estimación puntual, como se señaló, es insuficiente, porque solo se puede indicar si es correcta o incorrecta. Es más útil, si se acompaña por una estimación del error que podría cometerse en la inferencia. Por eso, es necesario tener en cuenta que, debido a la variabilidad de muestreo, los estadísticos no coincidirán con los parámetros poblacionales.

Por tal motivo se calcula una estimación por intervalo. Este intervalo resulta de sumar y de restar un valor (llamado **margen de error que se designará con E**), al estimador puntual. Por lo tanto, la fórmula general de una estimación por intervalo es:

Estimación puntual = ± E

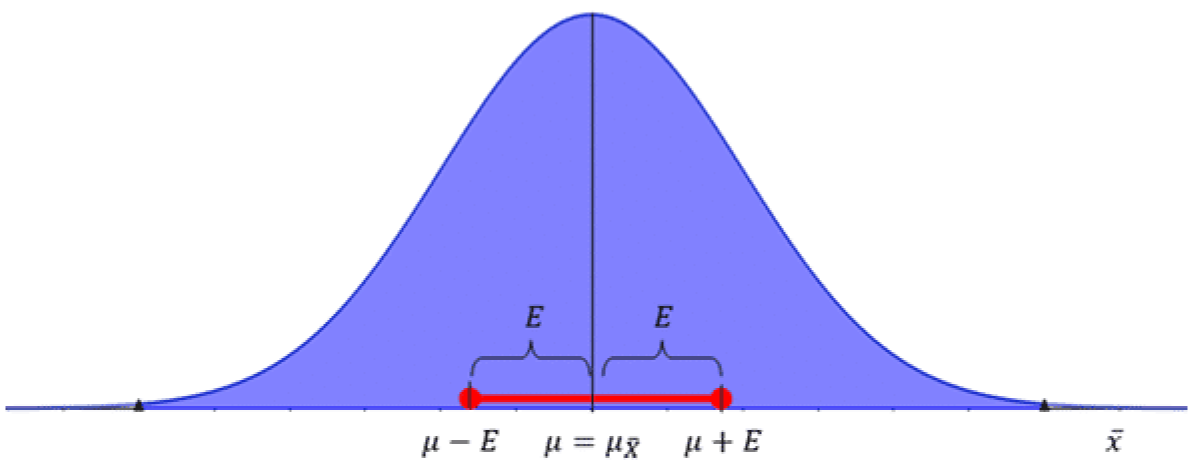
Este margen es el que está dispuesto a aceptar el investigador o la persona que realiza el estudio estadístico. E también se llama **error estimado** o **margen de error**. Por ejemplo, con un E = 0,7, en el caso de las horas de vuelo de los pilotos, se estima que el promedio de horas diario de todos los pilotos está entre 31.9 y 33,3. Además, es muy probable que la media exacta caiga dentro de este intervalo. Pero la medida de esa probabilidad es igualmente estimable, por lo que ese error se puede estimar con más precisión. De esta manera, se tendrá una mejor idea de la confiabilidad de la estimación y se podrán tomar mejores decisiones.

En estadística, a la probabilidad que se asocia con una estimación de intervalo se la conoce como **nivel de confianza**.

Estimación de la media poblacional μ a partir de la media muestral $\bar{x}$ mediante un intervalo de confianza

Error de estimación o margen de error. Confiabilidad

Figura 1: Visualización de un intervalo de confianza para la media en un caso general



Fuente: elaboración propia.

Descripción de la figura. Visualización de un intervalo de confianza para la media en un caso general. Para una muestra cualquiera, su media estará ubicada bajo la curva a una distancia máxima E de la media poblacional. Por lo tanto, si se estima la media poblacional a través de la media de esa muestra, se comete un error que está dado por E.

Si se considera una de las muestras, su media tendrá un valor

$$\bar{x},$$

y se ubicará sobre el eje de las abscisas a una distancia E de la media poblacional . Consecuentemente, si se estima la media poblacional a través de la media de esa muestra, se comete un error que es igual a

$$E = \bar{x} - \mu$$

En realidad, el error es ±, pues está a la derecha y a la izquierda de la media.

Además, como se estudió en la lectura 2, se puede expresar el margen de error E (también llamado error de estimación) en función del error muestral o estándar

$\sigma_{\bar{x}}$, entonces:

$$E = z.\sigma_{\bar{x}} = \bar{x} - \mu$$

Si tenemos en cuenta la fórmula del error estándar de una distribución de medias

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}},$$

la expresión del error de estimación será

$$E = z.\frac{\sigma}{\sqrt{n}}.$$

Como se observa, el error de estimación o margen de error es la función directa de z e inversa de n (tamaño de la muestra), en dónde:

z: definirá el grado de bondad o nivel de confianza,
σ: indica el desvío estándar poblacional y
n: representa el tamaño de la muestra.

El error de estimación es un valor que se puede calcular a priori. Se trata del máximo error permitido entre el valor de la media de una muestra y el valor de la media poblacional.

Si se conoce la desviación estándar de la población, entonces el error muestral queda sujeto a la variación del tamaño de la muestra y del grado de seguridad (confiabilidad) con que se quiera estimar el parámetro poblacional. Los grados de confiablidad más utilizados son de 90%, 95%, 98% y 99%.

Se adopta un valor de confianza del 95% en la estimación de la media poblacional. Esto implica que el valor de z = 1,96 y el área encerrada por la curva normal en el intervalo

$$\mu - 1{,}96 \; y \; \mu + \; 1{,}96$$

es de 0,95. Es decir, el 95% de todas las muestras tienen una media comprendida en ese intervalo.

El nivel de confianza que se establezca a priori define el valor de z. Este dato se extrae de la tabla de distribución normal estándar. Debe ingresarse a dicha referencia por su cuerpo, que son las probabilidades correspondientes a la confianza establecida, y buscar el z correspondiente.

Aclaración importante sobre los valores de z para una confianza determinada

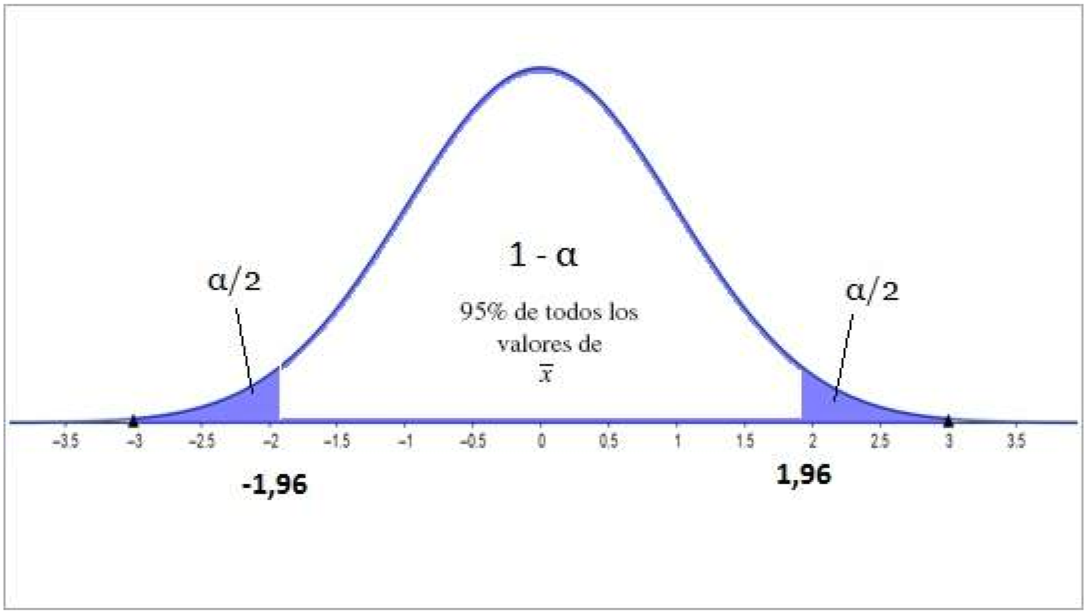
Muchas veces se encontrara, en algunos textos, la expresión $Z_{\alpha/2}$. Es muy utilizada y se refiere a cómo se obtiene que el valor de z correspondiente sea 1,96:

- Primero hay que definir el coeficiente de confianza que es el nivel de confianza expresado como valor decimal. Por ejemplo, 0.95 es el coeficiente de confianza correspondiente al nivel de confianza de 95%.
- Al coeficiente de confianza se lo designa con $(1-\alpha)$, es el área bajo la curva que queda en la cola inferior más el área de la cola superior. Pues al área central la abarca el coeficiente de confianza 0,95. Entonces, $(1-\alpha)=0{,}95$, teniendo en cuenta la simetría de la curva como lo muestra la figura 2.
- En la tabla de la distribución normal estándar, se tendrá que buscar un área con un coeficiente de confianza $(1-\alpha)=0{,}95$ y, por tanto, $\alpha=0{,}05$.
- Pero a este valor de $\alpha=0{,}05$ hay que dividirlo en dos partes iguales para buscar en la tabla el valor de z correspondiente. Por lo que

$$\frac{\alpha}{2} = \frac{0{,}05}{2} = 0{,}025.$$

- Según el tipo de tabla que se utilice, es fácil comprobar que para un área de 0,025 el valor de z para la cola superior es 1,96.

Figura 2: Distribución muestral de la media estandarizada para un intervalo de confianza del 95%



Fuente: elaboración propia.

Descripción de la figura. Distribución muestral de la media estandarizada para un intervalo de confianza del 95%. Se visualiza la ubicación del coeficiente de confianza y el área que ocupa. Además, se muestra el z = 1,96 correspondiente a la confianza del 95%.

Los niveles de confianza más utilizados son los que se muestran en la tabla 1.

Tabla 1: Valores de z para los principales niveles de confianza

Descripción de la figura: se muestran los niveles de confianza más utilizados con los correspondientes valores de z que muestran la tabla de la distribución normal estándar.

Nivel de confianza	Valor de z
90 %	1,645
95 %	1,96
98 %	2,33
99 %	2,575

Fuente: elaboración propia

Aplicación al caso de los pilotos de Aerolíneas Argentinas

Con lo estudiado hasta aquí, ya se pueden responder los apartados c) y d) del caso planteado al comienzo de la lectura, donde se solicita:

> C) Estimar el error estándar de la muestra, considerando una población finita de 1800 pilotos.
>
> Aquí, se aplica a la fórmula del error estándar, el factor de corrección para poblaciones finitas, pues, n > 5 % N

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} . \sqrt{\frac{N-n}{N-1}} = \frac{8,5}{\sqrt{100}} . \sqrt{\frac{1800-100}{1800-1}} = 0,85 \times 0,97 = 0,83$$

d) Determinar el margen de error con un 95 % de confianza.

La fórmula del margen de error es:

$$E = z.\sigma_{\bar{x}} = 1{,}96 \times 0{,}83 = 1{,}63$$

Fórmulas del intervalo de confianza para la media

Con lo estudiado hasta aquí, se pueden generalizar las fórmulas del límite inferior y superior de un intervalo de confianza. Se explica, además, qué sucede cuando la desviación estándar poblacional es conocida o desconocida.

Los casos que no se abordan se estudiarán en otro curso de estadística.

Desviación estándar poblacional conocida

En un caso hipotético, es necesario estimar la media de una población a través de la media de una muestra de tamaño n.

Por lo tanto, para determinar un intervalo de confianza para la estimación de la media poblacional, será necesario:

1. adoptar un nivel de confianza con el cual queda determinado el valor de z,
2. determinar la media de una muestra y utilizarla como estimador puntual de la media poblacional y
3. definir los límites del intervalo sobre la base de la desviación estándar de la población, que es conocida.

$$\text{Límite inferior de confianza: } Lic = \bar{x} - z.\frac{\sigma}{\sqrt{n}}$$

$$\text{Límite superior de confianza: } Lsc = \bar{x} + z.\frac{\sigma}{\sqrt{n}}$$

Nótese que el tamaño de la muestra n es inversamente proporcional al margen de error E. Esto puede manejarlo el investigador: establece el E que está dispuesto a aceptar o establece la muestra mediante una fórmula que se estudiará en la próxima lectura.

> Decir que σ es conocida significa que se cuenta con datos históricos o con otra información que permita obtener una buena estimación de la desviación estándar antes de tomar la muestra que se usará para obtener la estimación de la media poblacional. De manera que, técnicamente esto no significa que σ se conozca con seguridad. Lo que significa es que sólo se obtuvo una buena estimación de la desviación estándar antes de tomar la muestra…
>
> En consecuencia, si un determinado tamaño de muestra da un intervalo demasiado amplio, para que tenga utilidad práctica, se aumenta el tamaño de la muestra. Si n está en el denominador, con un tamaño de muestra mayor se obtendrá un margen de error menor, un intervalo más estrecho y mayor precisión. (Anderson, Sweeney y Williams, 2008, p. 305).

Ejemplo 1

En un restaurante se desea estimar la media del gasto por cliente en una comida. Se tomó una muestra de 50 clientes y se obtuvo una media de $ 345,20. La desviación estándar poblacional se estima en $125,40. Se solicita:

- [ ] Determinar el margen de error para un 98% de confianza.
- [ ] Determinar el intervalo de confianza de 98% para la media poblacional.

Solución

a) $E = z.\sigma_{\bar{x}}$. Según la tabla 1, el valor de *z* para una confianza del 98 % es *z = 2,33.*
Entonces:

$$E = z.\sigma_{\bar{x}} = 2{,}33.\frac{\sigma}{\sqrt{n}} = 2{,}33.\frac{125{,}40}{\sqrt{50}} = 2{,}33 \times 17{,}73 = 41{,}31$$

Es decir que el máximo error permitido para una confianza del 98% es 41,31.

b) $Lic = \bar{x} - z.\frac{\sigma}{\sqrt{n}} = 345{,}20 - 41{,}31 = 303{,}89$

$Lsc = \bar{x} + z.\frac{\sigma}{\sqrt{n}} = 345{,}20 + 41{,}31 = 386{,}51$

Existe un 98% de confianza de que el intervalo [303,89 – 386,51] contenga a la media poblacional.

En este caso no se aplica el factor de corrección para poblaciones finitas, pues la población se considera infinita.

Desviación estándar poblacional desconocida

Si la desviación estándar no se conoce y la muestra es grande, $n \geq 30$, se utiliza la desviación estándar de la muestra como estimador de la desviación estándar poblacional.

Es decir, $\hat{\sigma} = \sigma$.

A $\hat{\sigma}$ se llama estimador de la desviación estándar poblacional.

Si la muestra no es grande, se utiliza otro estimador (que se estudiará en cursos superiores).

Recordar utilizar el factor de corrección para poblaciones finitas, si es necesario.

Ejemplo 2

En una empresa de turismo se realizó un estudio y se encontró que las familias estaban dispuestas a gastar en promedio $ 675185 durante un mes de vacaciones en el transcurso del mes de enero. En el estudio participaron 600 familias y la desviación estándar muestral fue $10500.

1. ¿Con 95% de confianza, cuál es el margen de error?
2. ¿Cuál es el intervalo de confianza de 95% para estimar la media poblacional?

Solución

a) $E = z.\sigma_{\bar{x}}$. Según la tabla 1, el valor de z para una confianza del 95% es $z = 1{,}96$. Además, se utilizará como estimador de la desviación estándar poblacional a $\hat{\sigma} = \sigma = 10500$ y no se aplicará el factor de corrección para poblaciones finitas, pues la población se supone infinita frente a la muestra. Entonces,

$$E = z.\sigma_{\bar{x}} = 1{,}96.\frac{\hat{\sigma}}{\sqrt{n}} = 1{,}96.\frac{10500}{\sqrt{600}} = 1{,}96 \times 428{,}66 = 840{,}17$$

El máximo error permitido para una confianza del 95 % es 840,17.

b) $Lic = \bar{x} - z.\frac{\hat{\sigma}}{\sqrt{n}} = 675185 - 840{,}17 = 674344{,}83$

$Lsc = \bar{x} + z.\frac{\hat{\sigma}}{\sqrt{n}} = 675185 + 840{,}17 = 676025{,}17$

Existe un 95% de confianza de que el intervalo $[674344{,}83 - 676025{,}17]$ contenga a la media poblacional.

Conclusiones importantes

- [ ] Si la población es normal y se utiliza la fórmula

$$\bar{x} \pm z.\frac{\sigma}{\sqrt{n}}$$

el intervalo de confianza que se obtiene es exacto. ¿Qué significa que sea exacto? Que, si esa fórmula se usa repetidas veces para generar intervalos de, por ejemplo, el 95% de confianza, exactamente 95% de esos intervalos contendrán a la media poblacional.

- [ ] Si la población no tiene forma normal, el intervalo de confianza obtenido mediante la fórmula

$$\bar{x} \pm z.\frac{\sigma}{\sqrt{n}},$$

será aproximado. El grado de aproximación dependerá tanto de la distribución de la población como del tamaño de la muestra.

- [ ] Si el tamaño de la muestra es mayor o igual a 30, la fórmula dará un resultado más preciso.
- [ ] En caso de que no se conozca σ y la distribución de la población sea muy sesgada, se recomienda que el tamaño de la muestra sea mayor o igual a 50.
- [ ] En muestras menores, habrá que analizar la simetría de la población para decidir si utilizar la fórmula:

$$\bar{x} \pm z.\frac{\sigma}{\sqrt{n}}.$$

- [ ] Cuando σ es conocida, E es fijo y es igual para todas las muestras del mismo tamaño n.

- [ ] Cuando σ no se conoce, el margen de error varía de una muestra a otra, pues la desviación estándar muestral s varía con la muestra seleccionada.

En otro curso de estadística se estudiará otra distribución para solucionar los casos de muestras menores a 30 y distribuciones no normales con σ desconocida.

Resumen de los casos para la estimación de la media poblacional por intervalos de confianza

El siguiente cuadro funciona como un resumen de los casos planteados para saber las fórmulas que se utilizan cuando la desviación estándar es conocida o desconocida en poblaciones finitas e infinitas.

Figura 3: Fórmulas para intervalos de confianza en la estimación de la media poblacional

	Cuando la población es finita (y $n/N > 0.05$)	Cuando la población es infinita (o $n/N < 0.05$)
Estimación de μ (la media de la población): Cuando σ (la desviación estándar de la población) se conoce	Límite superior: $\bar{x} + z\frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$	$\bar{x} + z\frac{\sigma}{\sqrt{n}}$
	Límite inferior: $\bar{x} - z\frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$	$\bar{x} + z\frac{\sigma}{\sqrt{n}}$
Cuando σ (la desviación estándar de la población) no se conoce $\hat{\sigma} = s$) Cuando n (el tamaño de la muestra) es mayor que 30	Límite superior: $\bar{x} + z\frac{\hat{\sigma}}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$	$\bar{x} + z\frac{\hat{\sigma}}{\sqrt{n}}$
	Límite inferior: $\bar{x} - z\frac{\hat{\sigma}}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$	$\bar{x} - z\frac{\hat{\sigma}}{\sqrt{n}}$

Fuente: elaboración propia, con base en Levin y Rubin, 2004, p. 301.

Descripción de la figura. Fórmulas para intervalos de confianza en la estimación de la media poblacional. Se resumen las fórmulas de los límites para construir un intervalo de confianza para la media poblacional, diferenciando si la desviación estándar poblacional es conocida o no y si la población en finita o no.

Resolución del caso de los pilotos de Aerolíneas Argentinas, ítem e)

Caso A. A. Ítem e) Construye un intervalo de confianza del 95% para la media poblacional de las horas de vuelo de los pilotos.

Se utilizará el margen de error para un 95% de confianza, ya calculado en el ítem d) del caso planteado.

$$E = z.\sigma_{\bar{x}} = 1{,}96 \times 0{,}83 = 1{,}63$$

Por lo tanto, sabiendo que la media muestral de los 100 pilotos de Aerolíneas Argentinas es $\bar{x} = 32{,}6$ horas:

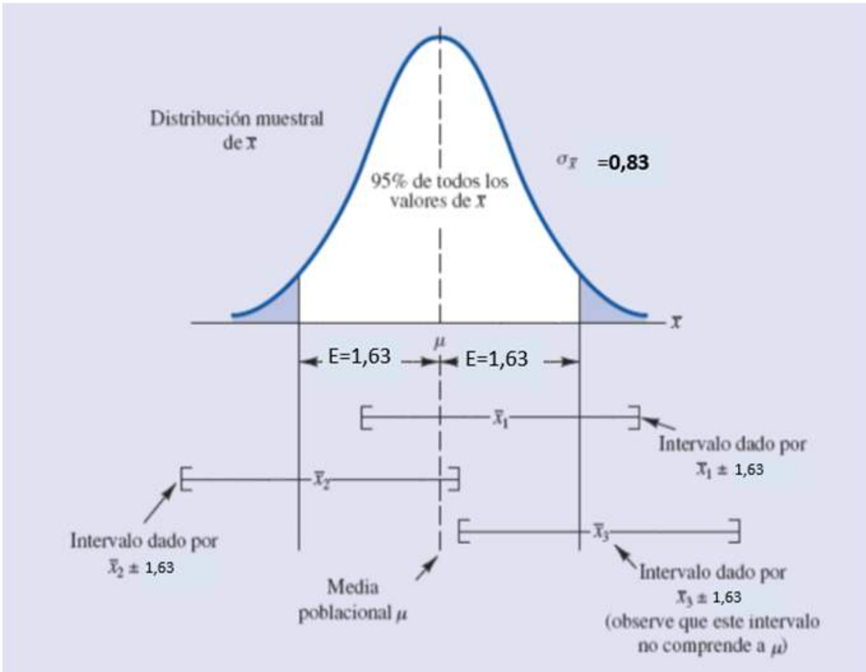
$$Lic = \bar{x} - E = 32{,}6 - 1{,}63 = 30{,}97$$

$$Lsc = \bar{x} + E = 32{,}6 + 1{,}63 = 34{,}23$$

Existe un 95% de confianza de que el intervalo [30,97-34,23] contenga a la media poblacional.

Gráfico del Caso

Figura 4: Interpretación de un intervalo de confianza del 95% para la media poblacional en el caso de los pilotos de Aerolíneas Argentinas



Fuente: elaboración propia, con base en Anderson et al., 2008, p. 303.

Descripción de la figura. Interpretación de un intervalo de confianza del 95% para la media poblacional en el caso de los pilotos de Aerolíneas Argentinas. Se representan, a modo de ejemplo, cómo serían las medias de cada una de las muestras a las que se refiere el teorema del límite central y así se visualiza qué significa el intervalo de confianza, en este caso del 95%, con el margen de error que proporciona el problema.

¿Cómo explicarías a tu jefe que las baterías de un automóvil duran entre 33,56 meses y 35,23 meses, con una confianza del 99%, si has realizado todos los cálculos correctamente?

- [ ] Si se seleccionan muchas muestras aleatorias del mismo tamaño y se calcula un intervalo de confianza para cada una de esas muestras, entonces, en el 99% de los casos la media de la población caerá dentro de dicho intervalo.

- [ ] Si se generan intervalos de confianza del 99%, se puede afirmar que es seguro que el 99% de los intervalos generados contendrán a la media poblacional.

- [ ] Se puede afirmar que el 99% de las baterías seleccionadas en la muestra tendrá una media contenida en el intervalo 33,56 y 35,23 meses.

- [ ] Esta afirmación implica que se tiene una probabilidad de 0,95 de que la vida media de todas las baterías caiga dentro del intervalo establecido para esta muestra.

SUBMIT

La amplitud del intervalo de confianza aumenta a medida que el nivel de confianza aumenta.

- ○ Es verdadero, porque aumenta el error estimado.

- ○ Es falso, porque a mayor confianza menor es el error estimado

SUBMIT

Estimación de la proporción poblacional mediante un intervalo de confianza

Con frecuencia se utiliza una muestra para estimar la proporción de ocurrencias de un evento en una población. Por ejemplo, el gobierno estima, mediante un procedimiento de muestreo, la cantidad de desocupados o la proporción de los votantes del país a un partido determinado.

La proporción se constituye en una población como un nuevo parámetro que se determinará, en la mayoría de los casos, como una inferencia a través del estadístico correspondiente de una muestra.

Para la obtención y análisis del estadístico hay que tener presente que el teorema del límite central también se aplica a las proporciones, como ya se vio. Por otra parte, la distribución de muestras de la proporción tiende a una normal cuando $n \geq 30$.

Acá se debe tener en cuenta que a la media de la distribución muestral de proporciones se la puede designar de distintas formas

$\mu_{\hat{p}}$, $\bar{p}$ o $\hat{p}$, según los distintos autores.

Se recuerda que en la lectura 3, la distribución de las muestras de la proporción tiene:

- una media igual a la proporción poblacional,

$\mu_{\hat{p}} = p$, que también suele escribirse así: $\bar{p} = p$

y el error estándar o la desviación estándar de la distribución muestral de la proporción,

$\sigma_{\bar{p}} = \sqrt{\frac{\bar{p}.\bar{q}}{n}}$ error estándar de la distribución de proporciones muestrales.

Siendo:

$\bar{p}$: la proporción de éxitos de la muestra y

$\bar{q}$: la proporción de fracasos de la muestra, $\bar{q} = 1 - \bar{p}$.

Por otra parte, hay que recordar que la distribución de la proporción surge de una aproximación de la binomial a la normal, por lo que todo lo referente a la estimación por intervalos de la proporción es utilizado cuando el tamaño de la muestra es grande y se cumplen las condiciones: n.p ≥ 5 y n.q ≥ 5.

Error estimado (E) en una estimación por intervalo de confianza de la proporciónHeading

De manera similar a lo analizado para la estimación por intervalos de la media, se inicia obteniendo una estimación por intervalo para la proporción poblacional a partir de la proporción muestral

$$\bar{p}.$$

Siendo

$$\hat{p} \text{ o } \bar{p}$$

el estimador puntual de p, tal como se definió al principio de esta lectura, se dice que la fórmula general de una estimación por intervalo para la proporción será:

$$\hat{p} \pm \ margen\ de\ error\ (E)\ .$$

Se construye la ecuación de E para la distribución muestral de éxitos de la proporción, partiendo de la fórmula de estandarización de p para una distribución normal, pero ahora aplicada a la distribución de las muestras de

$$p:$$

$$E = z.\sigma_{\bar{p}} = \bar{p} - p\ .$$

Siendo

$$p:$$

$$E = z.\sigma_{\bar{p}} = \bar{p} - p\ .$$

la desviación estándar de la distribución muestral de las proporciones o el error estándar,

$$\sigma_{\bar{p}} = \sqrt{\frac{\bar{p}.\bar{q}}{n}}\ .$$

Entonces,

$$E = z.\sqrt{\frac{\bar{p}.\bar{q}}{n}}\ .$$

- El error de estimación es un valor que lo podemos estimar a priori como: el máximo error permitido entre el valor de la proporción de la muestra y el valor de la proporción poblacional.

- Los grados de confiablidad más utilizados son los mismos que los utilizados para estimar la media poblacional y se encuentran en la tabla 1: 90%, 95%, 98% y 99%.

- Si se adoptara un valor de confianza del 95% en la estimación de la proporción poblacional, esto implicaría un valor de z = 1,96 y el área encerrada por la curva normal en el intervalo

$$\bar{p} - 1{,}96\ y\ \bar{p} + 1{,}96 \text{ es de } 0{,}95.$$

- Todas las características sobre la forma de la población y el tamaño de la muestra son aplicables también para la proporción. Por supuesto que el problema de si la desviación estándar es conocida o no, aquí no existe, pues no interviene en la fórmula del intervalo de confianza para la proporción.

- Cuando la relación entre el tamaño de la muestra n y el de la población N es mayor o igual a 0,05, se aplica el factor de corrección para poblaciones finitas. Al igual que para calcular el error estándar y permitido para la media.

Fórmulas del intervalo de confianza para la media

Si comparamos las fórmulas del límite inferior y superior de confianza en el caso con las que se necesitan para estimar la media, solo se deberá cambiar el error estándar poblacional.

Límite inferior de confianza: $Lic = \bar{p} - z.\sqrt{\frac{\bar{p}.\bar{q}}{n}}$.

Límite superior de confianza: $Lsc = \bar{p} + z.\sqrt{\frac{\bar{p}.\bar{q}}{n}}$.

Observa que n es inversamente proporcional al error permitido, igual que en el caso de la estimación por intervalos de la media poblacional. A mayor tamaño de la muestra menor es el error estimado E y el error estándar

$$\sigma_{\bar{p}}$$

Resumen de los casos para la estimación de la proporción poblacional por intervalos de confianza

Se presenta el siguiente cuadro como resumen de los casos planteados para estimar la proporción muestral.

Tabla 2: Fórmulas para intervalos de confianza en la estimación de la proporción poblacional

Descripción de la figura: se resumen las fórmulas de los límites para construir un intervalo de confianza para la proporción poblacional, diferenciando si la población es finita o infinita.

	Cuando la población es finita: n ≥ 0,05 N	Cuando la población es infinita: n < 0,05 N
Estimación de *p* (*proporción de la población*), *cuando* $n \geq 30$ y $\sigma_{\bar{p}} = \sqrt{\frac{\bar{p}.\bar{q}}{n}}$	$Lic: \bar{p} - z.\sqrt{\frac{\bar{p}.\bar{q}}{n}}.\sqrt{\frac{N-n}{N-1}}$ $Lsc: \bar{p} - z.\sqrt{\frac{\bar{p}.\bar{q}}{n}}.\sqrt{\frac{N-n}{N-1}}$	$Lic: \bar{p} - z.\sqrt{\frac{\bar{p}.\bar{q}}{n}}$ $Lsc: \bar{p} - z.\sqrt{\frac{\bar{p}.\bar{q}}{n}}$

Fuente: elaboración propia.

Ejemplo 3

La caja de jubilaciones de una provincia de Argentina ha determinado que de los últimos 100 empleados públicos que deciden iniciar los trámites de jubilación, el 40% lo hace sin recurrir a un gestor.

Se solicita:

1. determinar la proporción poblacional de empleados que encaran personalmente su jubilación y
2. generar un intervalo de confianza que contenga a la proporción poblacional con una bondad del 95%.

Datos: $n = 100$; $\bar{p} = 0,4$; $\bar{q} = 0,6$

a) Estimación puntual: $\bar{p} = p = 0,4$

El 0,4 de los empleados gestionan personalmente su jubilación (40%).

Entonces,

$$Lic: \bar{p} - z.\sqrt{\frac{\bar{p}.\bar{q}}{n}} = 0{,}4 - 1{,}96.\sqrt{\frac{0{,}4 \times 0{,}6}{100}} = 0{,}4 - 1{,}96 \times 0{,}049 = 0{,}4 - 0{,}096$$

$$Lic: 0{,}304$$

$$Lsc: \bar{p} + z.\sqrt{\frac{\bar{p}.\bar{q}}{n}} = 0{,}4 + 0{,}096 = 0{,}496$$

Con un 95% de confianza, el intervalo [0,304; 0,496] contiene a la proporción poblacional.

Resolución del caso de los pilotos de Aerolíneas Argentinas, ítem g)

Se sabe que la proporción de pilotos afiliados al sindicato en la muestra de 100 es de 86 pilotos. Se pide:

G) estimar, mediante un intervalo de confianza del 98%, la proporción poblacional de pilotos afiliados al sindicato.

En este caso, la población se considera finita frente a la muestra, pues esta es mayor al 5% de aquélla.

Datos:

$$\bar{p} = \frac{86}{100} = 0,86$$

$$\bar{q} = 1 - 0,86 = 0,14$$

$n = 100$

$z = 2,33$ para un 98% de confianza

Primero se calcula el margen de error, introduciendo el factor de corrección para poblaciones finitas:

$$E = z.\sqrt{\frac{\bar{p}.\bar{q}}{n}}.\sqrt{\frac{N-n}{N-1}} = 2,33.\sqrt{\frac{0,86 \times 0,14}{100}}.\sqrt{\frac{1800-100}{1800-1}} = 2,33 \times 0,0347 \times 0,9721 = 0,0786$$

$$Lic: \bar{p} - E = 0,86 - 0,0786 = 0,7814$$

$$Lsc: \bar{p} + E = 0,86 + 0,0786 = 0,9386$$

Con un 98 % de confianza, el intervalo [0,7814; 0,9386] contiene a la proporción poblacional.

En la práctica, los tamaños de la muestra de empleados en estimaciones por intervalo de una proporción poblacional suelen ser grandes.

- Es verdadero, porque la estimación puede realizarse siempre que la distribución de muestras de la proporción sea normal.

Es falso, porque la distribución de las muestras de la proporción es binomial.

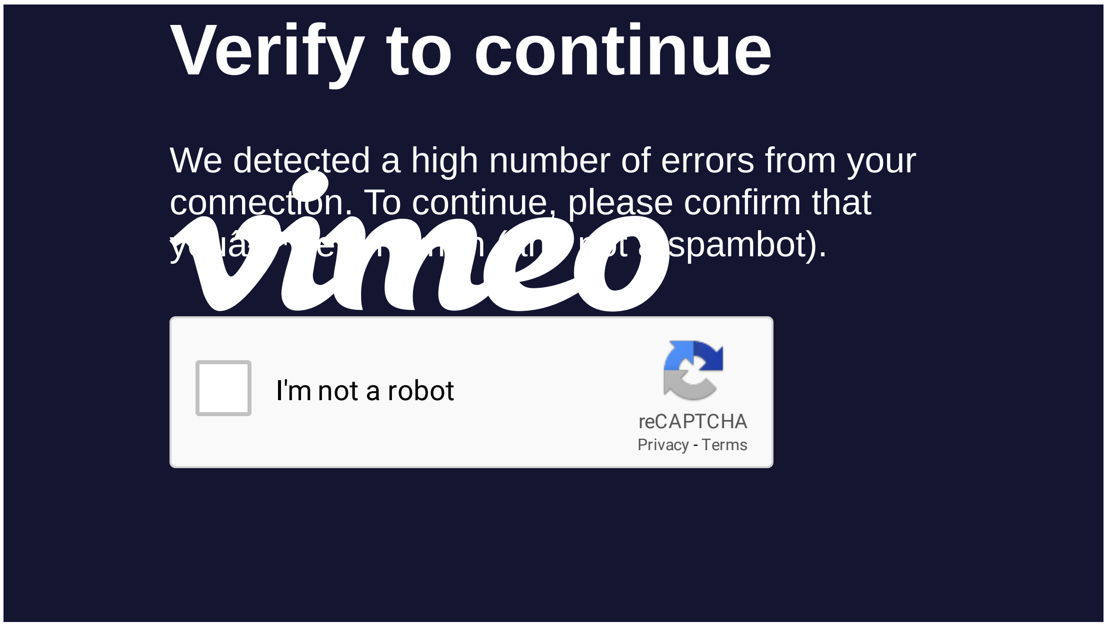
SUBMIT

Video conceptual

Video 1: Teorema central del límite

En este video se enuncia el teorema y sus aplicaciones, se muestra un ejercicio sobre determinación de un intervalo de confianza para la media.

VIMEO



VC - Herramientas Matemáticas III - Estadística I - Módulo 4 (CANVAS)

VER EN VIMEO

Referencias

Levin, R. y Rubin, D. (2004). Estadística para Administración y Economía (7ª edición). México: Pearson.

Anderson, D. R., Sweeney, D. J. y Williams, T. A. (2008). Estimación por intervalo en Anderson, D. R., Sweeney, Dennis J. y Williams, Thomas A. Estadística para Administración y Economía (pp. 300-329). México: Ed. Cengage Learning Editores, S.A.