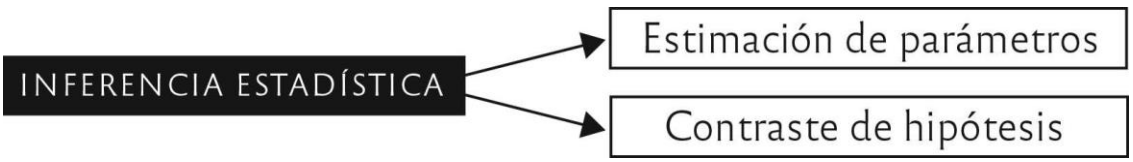
Capítulo 2

INFERENCIA ESTADÍSTICA. TEORÍA DE LA ESTIMACIÓN

1. Introducción

Hasta ahora hemos desarrollado metodologías que permiten efectuar una adecuada exploración de los datos, lo cual constituye una etapa fundamental en el comienzo de cualquier análisis estadístico. También conceptos importantes respecto al cálculo de probabilidades y diversos modelos teóricos de probabilidad para variables aleatorias discretas y continuas nos permitieron estudiar las distribuciones de resultados obtenidos a partir de un conjunto de datos.

Estos conocimientos son fundamentales para el correcto entendimiento y aplicación de los métodos de inferencia estadística que tienen por objetivo expandir los resultados provenientes de una muestra aleatoria a la población de la cual fue extraída. La inferencia estadística se aplica básicamente con dos objetivos alternativos:



a) La **estimación de parámetros**: cuando desconocemos los parámetros de una población y, con el fin de aproximarnos al conocimiento de esos valores, tomamos una muestra a partir de la cual estimamos los parámetros.

b) El **contraste de hipótesis**: también la realizamos a partir de una muestra tomada de la población de interés; pero, en este caso, buscamos poner a prueba (o contrastar) alguna hipótesis referida ya sea al valor de un parámetro o a las características de la población (si tiene cierta forma su distribución, si hay o no independencia entre dos variables, entre otros).

En este capítulo desarrollaremos uno de los temas básicos de la inferencia estadística: la estimación de parámetros. Los objetivos son:

- Comprender el concepto de estimador y de estimación puntual.

- Conocer las propiedades de los buenos estimadores y los métodos para obtenerlos.
- Comprender el concepto de estimación por intervalos y sus diferencias con la estimación puntual.
- Determinar el tamaño de muestra en distintas situaciones.
- Calcular e interpretar intervalos de confianza para algunos parámetros.

Recordemos que toda medida resumen que se calcula para describir características poblacionales se llama **parámetro**. Esta es una cantidad fija que generalmente no se conoce y debe de ser estimada.

En efecto, para aproximarnos al conocimiento de los parámetros desconocidos tomamos una muestra aleatoria de la población y, a partir de ella, estimamos el valor de los parámetros de interés. Los **estimadores** son aquellos que, justamente, estiman parámetros poblacionales; es decir que, aunque no coincidan exactamente con el parámetro, si la muestra fue correctamente seleccionada, deberían asumir valores bastante próximos a los mismos.

La Tabla 2.1 resume los principales parámetros y estimadores que pueden calcularse, según el tipo de variable analizada.

Tipo de variable	Medidas	Parámetros	Estimadores
		Población de tamaño *N*	**Muestra de tamaño *n***
Variable cuantitativa	Media	μ	$\overline{X}$
	Varianza	σ^2	S^2
Variable cualitativa	Proporción	P	$\hat{p}$

Tabla 2.1. Parámetros y estimadores

Ejemplo

Si se calcula la edad promedio del total de los/as clientes/as de un banco que utilizan servicios en línea, la media resultante es un parámetro poblacional. Si se toma una muestra de 100 clientes/as del banco, se obtiene la información de su edad y, en base a los datos recogidos, se calcula la edad media, este promedio es un estimador muestral. Ahora bien, ¿para qué nos sirve calcular la edad promedio de los/as 100 clientes/as que fueron seleccionados en la muestra? Es la única información con la que contaremos para decir algo acerca de toda la clientela de la entidad bancaria. Debido a esto, la utilizaremos para estimar al parámetro poblacional (edad promedio de todos los clientes del banco), que prácticamente se desconoce.

2. Estimación puntual

Como mencionamos anteriormente, el problema de la estimación de parámetros se debe a que en muchos casos la población que desea estudiarse es muy grande, es infinita, o resulta difícil acceder a su conocimiento total por diversos motivos (alta dispersión, por costo, entre otros). Cuando se desea conocer el valor de algún parámetro y resulta inviable el estudio de toda la población, surge la necesidad de estimar el valor de ese o esos parámetros y, para ello, se recurre al muestreo. En efecto, se toma una muestra aleatoria de la población y se calculan estimadores a partir de dicha muestra, que servirán para proporcionar una idea de los valores posibles de los parámetros poblacionales desconocidos.

Es natural pensar que si queremos conocer algo sobre la media poblacional μ recurramos a estimadores como la media muestral o la mediana de la muestra. Sin embargo, se podría definir otra combinación de las observaciones muestrales como estimador del parámetro y entonces podríamos preguntarnos, ¿cuál es el mejor estimador de μ?

Existen algunos criterios que determinan ciertas **propiedades** deseables para poder elegir un estimador. Así, utilizaremos el estimador de un parámetro que cumpla con todas o la mayoría de esas propiedades. Existen también ciertos métodos de estimación que proporcionan los mejores estimadores cuando se conocen algunas características de la población.

Si simbolizamos con θ a un parámetro cualquiera de la población, con $\hat{\theta}$ al estimador de ese parámetro y, además, tomamos una muestra de *n* observaciones de una población determinada, podemos definir formalmente al estimador $\hat{\theta}$ del parámetro θ como cualquier función de las observaciones muestrales. Por ejemplo, la media muestral $(\overline{X})$ es un estimador de la media poblacional μ.

Recordemos que cada observación muestral es una variable aleatoria, por lo tanto, un estimador, por ser función de variables aleatorias, es también una variable aleatoria.

A cada valor particular de un estimador, es decir, al valor que asume para una muestra

determinada, se lo llama estimación.

Ejemplo

Cuando se toma una muestra de 100 clientes del banco y se calcula su edad media, se obtiene una estimación particular de μ (edad promedio de todos los clientes del banco).

Dado que un estimador puede asumir cualquier valor dentro del rango posible determinado por su distribución de probabilidad, una estimación puntual no permite realizar afirmaciones ciertas ni probabilísticas acerca del verdadero valor del parámetro, o de la confianza que puede depositarse en el estimador. Esto último se logra mediante la **estimación por intervalos**, que toma como punto de partida estimadores puntuales que cumplan con propiedades deseables; las cuales desarrollaremos más adelante.

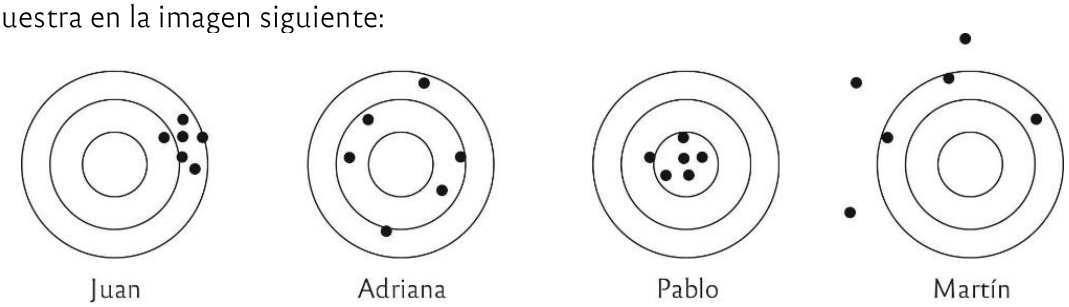
Si consideramos que definimos al estimador como cualquier función de las observaciones muestrales, pareciera obvio que pueden resultar tantos estimadores como funciones uno pueda imaginar. En realidad, no ocurre así, porque antes de elegir un estimador se presta atención a sus propiedades. Hay estimadores cuyas propiedades otorgan la confianza en que cada estimación particular que se realice con ellos proporcionará un resultado no muy alejado del parámetro.

A continuación, desarrollaremos las propiedades deseables de los estimadores o **propiedades de los buenos estimadores**.

3. Propiedades de los buenos estimadores

Para introducir el tema, realizaremos una analogía con un juego de tiro al blanco. Nuestra intención es destacar el razonamiento empleado para determinar la bondad de un estimador. Supongamos que se realiza un concurso de tiro al blanco y se quiere seleccionar un/a representante para participar en él. En la prueba final, cada tirador/a deberá realizar un solo tiro; quien pegue más tiros cerca del blanco será el/la ganador/a.

Se presentan cuatro postulantes y cada uno/a tira seis veces al blanco, tal como se muestra en la imagen siguiente:



De acuerdo con los resultados, ¿a quién elegiríamos como representante en el campeonato? Muy probablemente a Pablo, porque, si bien no siempre acierta, en promedio sus tiros dan en el blanco y, además, su dispersión no es muy grande. Adriana presenta un promedio cercano al blanco, pero tienen mucha dispersión. Es razonable decir que es más probable que Pablo acierte en el blanco en lugar de Adriana. Juan, por su parte, presenta poca dispersión, pero en promedio sus tiros pegan lejos del blanco. Martín no se acerca en promedio al blanco y, además, su dispersión es muy grande.

Si pensamos cada tirador/a como un estimador, cada tiro como una estimación y el blanco como el parámetro que se desea alcanzar, inmediatamente podemos definir las propiedades que nos permitirán elegir el/la mejor representante para el concurso. En las secciones siguientes definimos las propiedades deseables de los estimadores.

3.1. Insesgabilidad

Como explicamos, los estimadores son variables aleatorias. En consecuencia, para cada una de las muestras posibles pueden asumir valores diferentes, y cada valor o intervalo de valores posibles tiene asociada una probabilidad. Un estimador tiene, entonces, una distribución de probabilidad y, por lo tanto, puede calcularse su esperanza. El hecho de que la esperanza del estimador sea igual al parámetro a estimar no asegura que de una estimación surgirá un resultado igual al parámetro. Solo se trata de una propiedad teórica que afirma que, si se toman todas las muestras posibles, el promedio de los valores del estimador (su valor esperado) será igual al parámetro. En cada estimación, el valor del estimador puede resultar más o menos alejado de ese valor esperado. ¿De qué depende que no resulte muy alejado? De la variabilidad, es decir, de la desviación estándar (o el error estándar) del estimador. En efecto, siempre que el valor esperado no esté muy lejos del parámetro, será preferible entre dos estimadores posibles aquél que tenga menor variabilidad; puesto que, en ese caso, las probabilidades de que en cada muestra el estimador esté cercano al parámetro serán mayores.

Formalmente, decimos que un estimador es **insesgado** cuando su valor esperado es igual al parámetro que se desea estimar, es decir:

Si existiera una diferencia entre la esperanza del estimador y el parámetro a estimar, esa diferencia se denomina sesgo (k).

$$k = E(\hat{\theta}) - \theta$$

Por lo tanto, si $E(\hat{\theta}) = \theta + k$, entonces $\hat{\theta}$ es un estimador sesgado de θ, siendo k la magnitud del sesgo.

Si consideramos dos estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ del mismo parámetro θ y graficamos sus distribuciones de probabilidad, podemos obtener una situación como la que se muestra en el Gráfico 2.1.

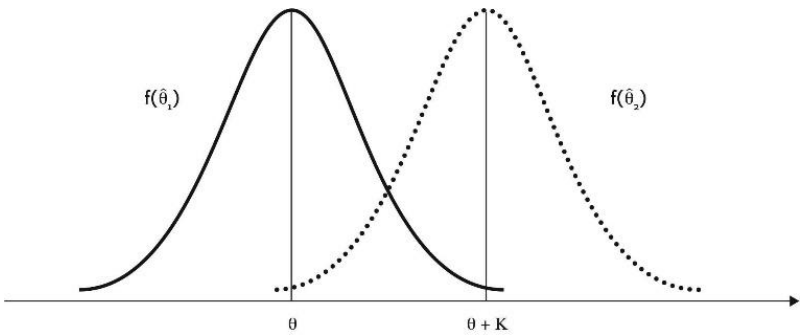
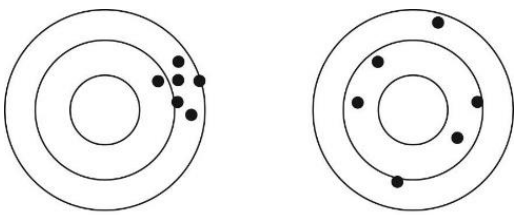


Gráfico 2.1. Distribuciones de probabilidad de los estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$

Podemos observar que $\hat{\theta}_1$ es un estimador insesgado porque su esperanza coincide con el valor del parámetro, en tanto que $\hat{\theta}_2$ es un estimador sesgado de θ.

Volviendo al ejemplo del tiro al blanco, podemos decir que Adriana es mejor que Juan, ya que su promedio está cerca del blanco (es insesgado). Juan presenta poca variabilidad, pero es sesgado (es muy probable que pegue cerca pero no del blanco).



En relación a los estimadores, la media muestral $\overline{X}$ es un estimador insesgado de la media poblacional μ ya que $E(\overline{X}) = \mu$. Para la proporción poblacional (P), la proporción muestral $\hat{p}$ es un estimador que cumple con la propiedad de insesgabilidad.

Es sesgado, en cambio, el estimador de la varianza poblacional calculado mediante la varianza muestral con la fórmula:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n}$$

Ya que $E(S^2) = \sigma^2.\left(\frac{n-1}{n}\right)$, lo que significa que posee sesgo.

Para corregir el sesgo se sugiere utilizar como estimador de la varianza poblacional, la varianza muestral corregida:

$$S_c^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1}$$

La varianza muestral no corregida S^2, sin embargo, es un estimador asintóticamente insesgado; si bien presenta un sesgo, el sesgo tiende a desaparecer al aumentar el tamaño de la muestra.

3.2. Consistencia

Intuitivamente, un buen estimador sería aquel que se acerca al valor del parámetro cuando crece el tamaño de la muestra, es decir, el error estándar tiende a cero al incrementarse el tamaño muestral. La consistencia es una propiedad deseable y muy importante de los estimadores cuando se usan muestras grandes. Si estas son pequeñas, no tiene relevancia el hecho de que un estimador sea o no consistente.

Se dice que un estimador $\hat{\theta}$ es consistente si:

$$\lim_{n\to\infty} P\left(\left|\hat{\theta} - \theta\right| < e\right) = 1$$

Es decir que, al tomar muestras grandes tenemos la seguridad de que el estimador se aproximará al parámetro. En otras palabras, un estimador es consistente si existe una probabilidad igual a 1 de que la diferencia (error) entre el estimador y el parámetro sea inferior a un número arbitrario e.

La varianza muestral es un estimador consistente de la varianza poblacional y la media muestral lo es de la media poblacional. Por otro lado, la mediana muestral no es un estimador consistente de la media poblacional, aunque sí de la mediana poblacional. No obstante, como estimador de la media poblacional, es insesgado si la distribución poblacional es simétrica. Esto significa que, al tomar muestras grandes, la media de la muestra se aproximará a la media poblacional y la mediana de la muestra se aproximará a la mediana poblacional.

Existe otra condición de consistencia conocida como consistencia en error cuadrático que implica que tanto el sesgo como la varianza del estimador tienden a cero a medida que aumenta el tamaño de la muestra.

3.3. Eficiencia

Esta propiedad indica que un buen estimador es aquel de mínima varianza con respecto a otros estimadores del mismo parámetro. La varianza de un estimador es muy importante ya que proporciona una idea del grado de confianza que se puede tener en el mismo.

Ante dos estimadores insesgados (o por lo menos consistentes) $\hat{\theta}_1$ y $\hat{\theta}_2$ del mismo parámetro θ, se dice que $\hat{\theta}_1$ es más **eficiente** si $V\left(\hat{\theta}_1\right) < V\left(\hat{\theta}_2\right)$. Es claro que, cuanto menor sea la varianza de un estimador es más probable que asuma valores cercanos al parámetro (siempre que sea insesgado) y, por lo tanto, se trate de un mejor estimador.

El concepto de eficiencia al que nos referimos aquí es el de **eficiencia relativa**, porque estamos contrastando la eficiencia de dos estimadores. Un ejemplo interesante surge de comparar la media y la mediana muestrales como estimadores de la media poblacional. Sabemos que la varianza de la media muestral es $V = \frac{\sigma^2}{n}$. En cambio, la

varianza de la mediana muestral es $V(me) = 1{,}25332\,\frac{\sigma^2}{n}$. Esto significa que hay mayor probabilidad de que el valor de la media muestral se encuentre cerca del de la media poblacional ya que presenta menor error estándar. Por lo tanto, aunque ambos son estimadores insesgados de la media poblacional, la media muestral es más eficiente.

Si consideramos las distribuciones de probabilidad de dos estimadores insesgados $\hat{\theta}_1$ y $\hat{\theta}_2$ del mismo parámetro θ (tal como se muestra a continuación), es evidente que los valores de $\hat{\theta}_1$ están mucho más concentrados alrededor del parámetro que los valores de $\hat{\theta}_2$.

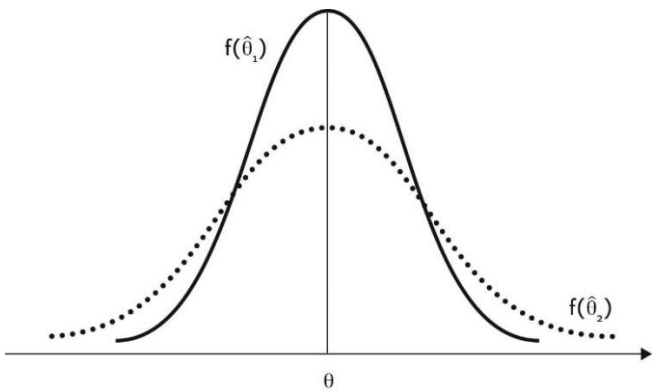
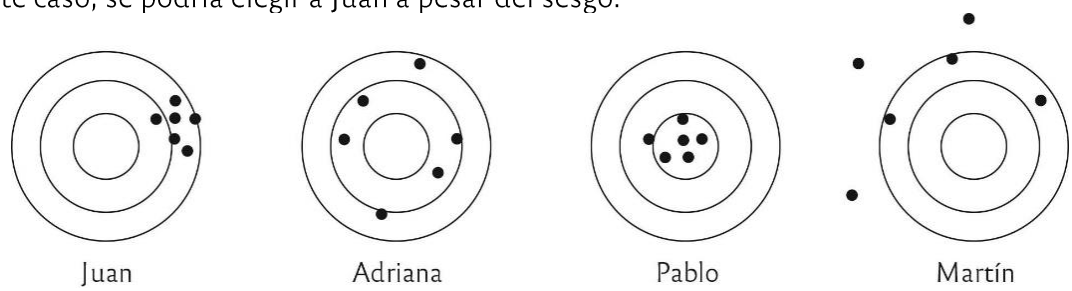


Gráfico 2.2. Distribuciones de probabilidad de los estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$

Un investigador confiará más en un estimador muestral que tenga menor posibilidad de alejarse del verdadero valor del parámetro poblacional.

En el ejemplo del tiro al blanco, Martín es el que presenta mayor variabilidad, por lo que sería el menos eficiente. Si comparamos a Juan y Adriana, observamos que el primero es sesgado y la segunda no. No obstante, Juan tiene menor variabilidad que Adriana. En este caso, se podría elegir a Juan a pesar del sesgo.



Existe también el concepto de **eficiencia absoluta** (conocido como la acotación de Rao-Cramer). Este permite establecer, dada una distribución poblacional, cuál es la varianza mínima de un estimador del parámetro, lo que no desarrollaremos en este manual.

3.4. Suficiencia

Un estimador es **suficiente** cuando utiliza toda la información de la muestra con

respecto al parámetro a estimar. Si en una población se desea estimar la media poblacional, se toma una muestra de tamaño *n* y se define como estimador de la media poblacional la siguiente función de las observaciones muestrales:

$$\hat{\mu} = \frac{x_1 + x_2}{2}$$

Es decir, el promedio de las dos primeras observaciones. Podemos comprobar que, aunque este estimador es insesgado, no utiliza la información proporcionada por las *n*-2 observaciones restantes. Por ello, decimos que no es un estimador suficiente.

Podemos encontrarnos con otras propiedades (que no estudiaremos en este manual), como la **distribución asintóticamente Normal**, que refiere a aquellos estimadores cuya distribución es Normal al aumentar el tamaño muestral (como ocurre con todos los casos donde se aplica el teorema central del límite), o como los **estimadores robustos**, que no se ven muy afectados ante pequeños desvíos respecto del cumplimiento de ciertos supuestos.

Actividades de aprendizaje

Le proponemos realizar las siguientes actividades para profundizar los temas estudiados.

Actividad 1

Sean x_1, x_2 y x_3 valores que representan el tiempo de fabricación de una pieza, pertenecientes a una muestra de tamaño 3, tomada del total de piezas fabricadas en una empresa, con media μ y varianza σ^2. Considerando los siguientes estimadores del promedio del tiempo de fabricación, ¿cuál de ellos elegiría y por qué?

$\hat{\mu}_1 = 1/3\, x_1 + 1/3\, x_2 + 1/3\, x_3$
$\hat{\mu}_2 = 1/9\, x_1 + 3/9\, x_2 + 5/9\, x_3$
$\hat{\mu}_3 = 6/10\, x_1 + 4/10\, x_3$

Actividad 2

A continuación, se presentan los diámetros de una muestra de 5 artículos fabricados en una empresa: 6,33 cm; 6,37 cm; 6,36 cm; 6,32 cm; 6,37 cm.
Suponiendo que los diámetros se distribuyen en forma normal:

a) Calcular una estimación con un estimador insesgado y eficiente:
   1) de la media poblacional
   2) de la varianza poblacional

1. Determinar un estimador insesgado pero ineficiente del diámetro promedio poblacional.

2. Determinar un estimador insuficiente de la media poblacional.

Actividad 3

Sea *X* el número de éxitos en *n* repeticiones de un experimento binomial con probabilidad de éxito igual a *p*. El estimador de *P* es $\hat{p} = \frac{x}{n}$, o sea la frecuencia relativa de éxitos. Verificar que este estimador es insesgado.

4. Método de estimación puntual por máxima verosimilitud

Uno de los métodos más utilizados para obtener estimadores puntuales es el **método de máxima verosimilitud**, que fue introducido por Fisher en la década de 1920. Este método consiste en seleccionar, de entre todos los valores posibles del parámetro, el valor estimado del parámetro que maximiza la probabilidad de una muestra (a posteriori de haberla extraído).

Se trata de un método muy utilizado para obtener estimadores porque comprueba que los estimadores máximos verosímiles gozan de la mayoría de las propiedades enunciadas en el punto anterior (en algunos casos son sesgados, pero el resto de las propiedades siempre están presentes). Esto significa que, haber encontrado un estimador máximo verosímil, es haber encontrado un buen estimador del parámetro desconocido.

Ejemplo

Para entender mejor el concepto supongamos que se realiza una encuesta de opinión a una muestra aleatoria de 10 personas, seleccionadas con reposición. Se les formula una única pregunta que será respondida por sí o por no.

Sean $x_1, x_2, \dots, x_{10}$ las variables aleatorias correspondientes a la respuesta tales que:

$$x_i = \begin{cases} 1 \;\; si\ la\ respuesta\ es\ Sí \\ 0\ si\ la\ respuesta\ es\ No \end{cases}$$

Para $i = 1, 2, \dots, 10$ y $p = P(x_i = 1)$.

Observemos que las variables aleatorias x_i son independientes y cada una de ellas tiene distribución Binomial o distribución Bipuntual $(1,p)$. Entonces, la función de probabilidad conjunta de las observaciones será:

$$f(x_1, x_2, \dots, x_{10}) = p^{x_1}.(1-p)^{1-x_1}.p^{x_2}.(1-p)^{1-x_2}\dots.p^{x_{10}}.(1-p)^{1-x_{10}}$$

Si en la muestra obtenida se observan 3 respuestas por sí ($x_i = 1$), tendríamos:

$$f(x_1, x_2, \dots, x_{10}) = p^3.(1-p)^7$$

La pregunta es: ¿qué valor de p hace que los valores muestrales obtenidos sean los más probables? Por medio de la distribución Binomial podemos calcular cuál es la probabilidad de obtener tres respuestas favorables en una muestra de tamaño 10, para algunos de todos los posibles valores de *p*, tal como se muestra en la siguiente tabla:

p	**P(*x*=3,10,*p*)**
0,0	0,0000
0,1	0,0574
0,2	0,2013
0,3	0,2668
0,4	0,2150
0,5	0,1172
0,6	0,0425
0,7	0,0090
0,8	0,0008
0,9	0,0000
1,0	0,0000

Evidentemente, no elegiríamos $p = 0$ o $p = 1$ como estimador de *p* en la población. Por un lado, en la muestra hubo 3 que contestaron afirmativamente, por lo que p no puede ser igual a cero. Por otro lado, hay 7 que contestaron negativamente, luego *p* no puede ser igual a 1.

Del resto de valores posibles de *p*, advertimos que 0,3 es el que proporciona la mayor probabilidad de que en una muestra de 10 ocurran 3 opiniones favorables. Así, 0,3 es el valor del estimador máximo verosímil o de máxima probabilidad de *p*.

Visto de otra forma, buscamos el valor de *p* que hace máxima a la función $f(x_1, x_2, \dots, x_{10})$ o, equivalentemente, al $lnf(x_1, x_2, \dots, x_{10})$, ya que el logaritmo es una función monótona creciente.

Vamos a maximizar la siguiente función de p aplicando algunos pasos matemáticos:

$$lnf(x_1, x_2, \dots, x_{10}) = 3.lnp + 7.ln(1-p)$$

Derivamos esta función respecto de *p* e igualamos a cero la derivada primera:

$$\frac{\partial lnf(x_1, x_2, \dots, x_{10})}{\partial p} = \frac{3}{p} - \frac{7}{1-p}$$

$$\frac{\partial lnf(x_1, x_2, \dots, x_{10})}{\partial p} = \frac{3(1-p) - 7p}{p.(1-p)}$$

$$\frac{\partial lnf(x_1, x_2, \dots, x_{10})}{\partial p} = \frac{3 - 10p}{p.(1-p)}$$

$$\frac{3 - 10\hat{p}}{\hat{p}.(1-\hat{p})} = 0$$

$$\hat{p} = \frac{3}{10}$$

Podemos verificar que la derivada segunda es menor a cero, con lo que comprobamos que $0{,}30$ es el valor de *p* que maximiza la función.

En general, es posible obtener estimadores máximos verosímiles en forma analítica a partir de la función de probabilidad conjunta de una muestra, que depende de la forma de la distribución poblacional, de las observaciones muestrales y de los parámetros poblacionales.

Desarrollaremos ahora el caso de la distribución Normal, con media μ y varianza σ^2. La distribución conjunta de las observaciones muestrales (llamada **función de verosimilitud**) que depende de estas observaciones y de los parámetros desconocidos es:

$$L(\mu, \sigma^2) = f(x_1, x_2, \dots, x_n; \mu, \sigma^2) = \left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]^n e^{\frac{-1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}}$$

Esta función se plantea una vez que se obtuvo la muestra y, por lo tanto, las x_i son conocidas. Las incógnitas son μ y σ^2. Si se busca el máximo de $L(\mu, \sigma^2)$ con respecto a ambos parámetros, se obtienen los valores de μ y σ^2 que otorgan máxima probabilidad a la muestra. Esos valores son los estimadores máximos verosímiles.

Para encontrar este máximo, se toma el logaritmo natural de $L(\mu, \sigma^2)$y luego se deriva con respecto a μ y σ^2**.** Se igualan a cero esas derivadas, se encuentran los valores que satisfacen cada ecuación y luego se verifican las condiciones de segundo orden. Tomando logaritmo natural en la expresión anterior y aplicando propiedades de los logaritmos obtenemos:

$$ln(L) = \frac{-n}{2} ln(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}.ln\ e$$

En esta función, derivamos con respecto a μ y σ^2:

$$\frac{\partial ln(L)}{\partial \mu} = 0 - \frac{1}{2\sigma^2}.2\sum_{i=1}^{n}(x_i-\mu)(-1)$$

$$\frac{\partial ln(L)}{\partial \sigma^2} = \frac{-n}{2} \cdot \frac{2\pi}{2\pi\sigma^2} - \frac{1}{2}\left(\frac{-1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

Simplificando:

$$\frac{\partial ln(L)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$$

$$\frac{\partial ln(L)}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2 = \frac{-n\sigma^2 + \sum_{i=1}^{n}(x_i - \mu)^2}{2(\sigma^2)^2}$$

Al imponer la condición de anular estas derivadas, obtenemos dos ecuaciones; siendo los valores de μ y σ^2 que las verifican los estimadores que buscamos:

$$\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(x_i - \hat{\mu}) = 0$$

$$\frac{-n\hat{\sigma}^2 + \sum_{i=1}^{n}(x_i - \hat{\mu})^2}{2(\hat{\sigma}^2)^2} = 0$$

Despejando obtenemos:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \hat{\mu})^2}{n}$$

De esta forma, hemos comprobado que los estimadores máximos verosímiles de la media y varianza en una población normal son respectivamente la media muestral y la varianza muestral sin corregir.

Los estimadores máximos verosímiles tienen, en general, las propiedades deseables planteadas más arriba. Es por ello que suelen ser los mejores estimadores de los diferentes parámetros en casi todos los casos. No obstante, el estimador máximo verosímil de la varianza es un estimador sesgado. Por eso, en lugar de usar directamente el estimador máximo verosímil, en general se corrige para que desaparezca el sesgo; esto es, dividiendo por *n*-1 en lugar de *n*.

Siguiendo estos pasos es posible obtener el estimador máximo verosímil de los parámetros de cualquier distribución partiendo de su función de probabilidad o densidad. A continuación, encontrará ejercicios para aplicar lo desarrollado hasta el momento.

Actividades de aprendizaje

Le proponemos realizar las siguientes actividades para profundizar los temas estudiados.

Actividad 4

Una compañía aseguradora ha comprobado que el número de siniestros de determinado tipo ocurridos semanalmente se ajusta a un modelo Poisson. Se desea obtener la estimación máximo verosímil del promedio de siniestros semanales (λ) del modelo, sabiendo que $f(x_i, \lambda) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$

Actividad 5

Sabiendo que la función de densidad en el caso de la distribución exponencial es:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \; para \; x \geq 0 \\ 0 \;\; para \; otro \; valor \end{cases}$$

Encuentre el estimador máximo verosímil de λ.

Actividad 6

Para la siguiente función de verosimilitud o probabilidad conjunta L(*P*), determine el estimador máximo verosímil de *P* (proporción poblacional de éxitos).

$$L(p) = C_n^X p^X (1-p)^{n-X} \qquad \text{donde } X = \sum x_i$$

5. Estimación por intervalos

En el apartado anterior, desarrollamos los conceptos vinculados con la definición de estimadores, sus propiedades deseables y los métodos para obtener buenos estimadores. Además, planteamos algunos estimadores que poseen dichas propiedades y que son utilizados en los problemas de aplicación más frecuentes.

Ahora bien, las estimaciones realizadas en forma puntual no llevan asociada idea alguna acerca del grado de aproximación que puede existir entre el valor del estimador y el del parámetro que se está estimando. Dicho de otra forma, no es posible cuantificar el error que puede cometerse al afirmar que el parámetro desconocido se estima igual a esa función de las observaciones muestrales definidas por cada estimador, tampoco saber sobre la confianza que puede depositarse en la sospecha de que el valor estimado se encuentra relativamente cerca del parámetro desconocido.

Precisamente, la diferencia existente entre el valor del estimador en una muestra particular (estimación) y el verdadero valor del parámetro desconocido se llama **error de muestreo**. En efecto, si θ es el parámetro que se desea estimar y $\hat{\theta}$ es su estimador puntual, la expresión:

$$\left|\hat{\theta} - \theta\right| \leq e$$

indica que el valor absoluto del error cometido al realizar la estimación no excederá a *e*. El error es aleatorio, porque depende de $\hat{\theta}$ que es una variable aleatoria. Esto significa que no es posible saber con exactitud en cuánto nos equivocaremos al estimar un parámetro a partir de una muestra, por ello la expresión plantea que el error no supere a una magnitud *e*.

Teniendo en cuenta el conocimiento de las distribuciones de probabilidad de algunos estadísticos (funciones de los estimadores y los parámetros) podemos establecer un intervalo aleatorio $(\hat{\theta} - e; \hat{\theta} + e)$ cuya amplitud es igual al doble del error máximo de estimación y que posea una elevada probabilidad de que el parámetro θ esté contenido en él.

Con una estimación por intervalos se pierde precisión en la estimación al referirse la misma a un rango de valores y no a un valor puntual, pero se gana en el conocimiento del error que puede cometerse al realizarla y del grado de confianza de que la afirmación sea verdadera.

5.1. Definición de intervalo de confianza a través de un ejemplo

Supongamos que necesitamos estimar la media poblacional teniendo en cuenta lo que sabemos de la distribución de la media muestral. Debemos, entonces, estimar la media de una variable en cierta población de la cual se conoce que tiene distribución Normal, con varianza poblacional σ^2. Realizaremos la estimación a partir de una muestra aleatoria de tamaño *n*.

Ya sabemos que el estimador puntual de μ es $\overline{X}$. Por tratarse de una población normal (no importa cuál sea *n*), se conoce que la distribución de $\overline{X}$ es también normal, con esperanza μ y varianza σ^2/n. El estadístico es:

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

Es función del estimador, del parámetro a estimar y de σ, pero, como suponemos conocida esta última, μ es la única incógnita de la expresión. Además, *n* es el tamaño de la muestra y obtendremos el valor particular del estimador $\overline{X}$ a partir de la muestra. El conocimiento de la distribución de probabilidad del estadístico nos permite afirmar que:

$$P\left(z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right) = 1 - \alpha$$

Y se lee: existe una probabilidad igual a $1 - \alpha$ de que el estadístico esté comprendido entre $z_{\alpha/2}$ y $z_{1-\alpha/2}$ (valores de la variable normal estandarizada). Así, por ejemplo, si $1 - \alpha = 0{,}95$; entonces $z_{\alpha/2}$ = -1,96 y $z_{1-\alpha/2}$ = 1,96 dado que *P*(-1,96 < z < 1,96) = 0,95. Como puede observarse $z_{\alpha/2} = -z_{1-\alpha/2}$.

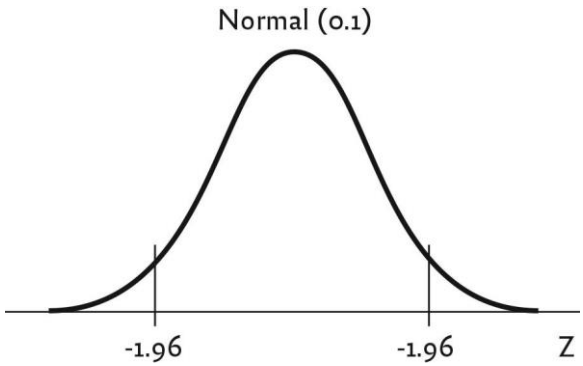


Gráfico 2.3. Distribución de probabilidad de la variable Z $(1-\alpha=0{,}95)$

Despejando el error aleatorio en la expresión anterior queda:

$$P\left(-z_{1-\alpha/2}.\sigma/\sqrt{n} < \overline{X}-\mu < z_{1-\alpha/2}.\sigma/\sqrt{n}\right) = 1-\alpha$$

Ahora despejamos el valor de μ:

$$P\left(\overline{X}-z_{1-\alpha/2}.\sigma/\sqrt{n} < \mu < \overline{X}+z_{1-\alpha/2}.\sigma/\sqrt{n}\right) = 1-\alpha$$

Y debe leerse como: existe una probabilidad igual a $1-\alpha$ de que el intervalo aleatorio $\left(\overline{X}-z_{1-\alpha/2}.\sigma/\sqrt{n}\,;\overline{X}+z_{1-\alpha/2}.\sigma/\sqrt{n}\right)$ contenga a μ.

Una vez tomada la muestra y realizada la estimación de μ (habiendo obteniendo una media muestral particular), el intervalo deja de ser aleatorio y toma dos valores particulares llamados límite inferior (LIC) y superior (LSC) de confianza. Entonces, el nivel $1-\alpha$ ya no expresa una probabilidad (porque no hay ninguna variable aleatoria) sino un **nivel de confianza** del $(1-\alpha).100\ \%$, de que el intervalo obtenido contiene al parámetro desconocido.

$$LIC = \overline{X}-z_{1-\alpha/2}.\sigma/\sqrt{n}$$

$$LSC = \overline{X}+z_{1-\alpha/2}.\sigma/\sqrt{n}$$

El gráfico que sigue puede ayudar a comprender el significado de los límites de confianza en una estimación. De acuerdo al ejemplo planteado, se trata de una distribución Normal con media μ y desviación estándar σ. El gráfico exhibe la distribución de la media muestral para muestras de tamaño *n* que, por lo tanto, también es Normal, con media μ, y desviación estándar $\sigma/\sqrt{n}$.

El intervalo señalado en el gráfico encierra el $0{,}95$ de probabilidad debajo de la curva normal. Esto es, el $95\ \%$ de las posibles muestras aleatorias de tamaño *n* tienen una media dentro de ese intervalo.

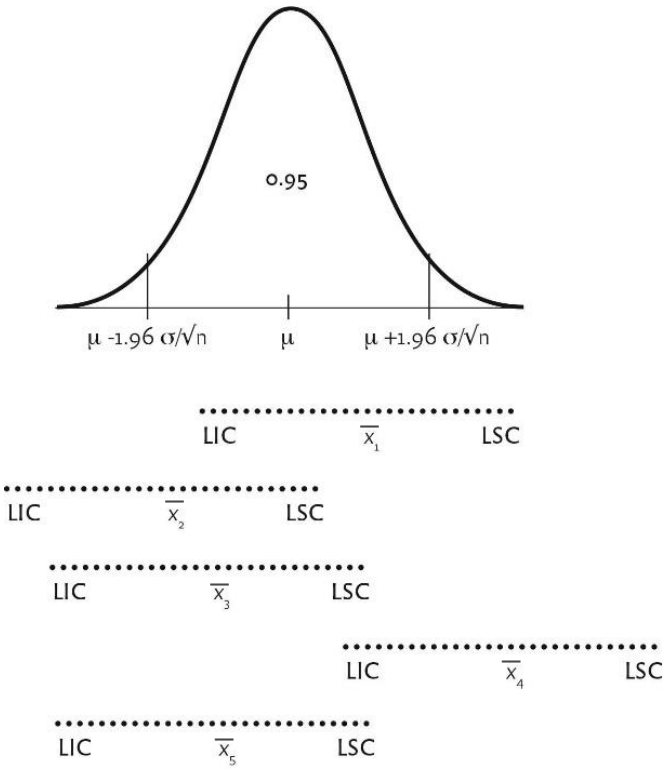


Gráfico 2.4. Intervalos de confianza de cinco muestras

Como podemos observar en el Gráfico 2.4, en las líneas trazadas debajo de la curva normal (correspondientes a los posibles resultados de cinco muestras), siempre que la media muestral caiga dentro del intervalo del 0,95 de probabilidad, los límites de confianza contendrán en su interior la media poblacional μ. Es decir, el intervalo construido encerrará la verdadera media (aunque esta sigue siendo desconocida). Pero cuando la media muestral cae fuera del intervalo del 95 %, los límites de confianza obtenidos no contendrán la verdadera μ.

Luego, cuando tomemos una muestra aleatoria, y a partir de la media de esa muestra construyamos un intervalo de amplitud igual al doble de $1{,}96.\sigma/\sqrt{n}$, obtendremos una confianza del 95 % en que el valor del parámetro desconocido μ está dentro de ese intervalo.

En este ejemplo gráfico solo en la cuarta muestra resulta un intervalo que no contiene a μ, en las restantes, es verdadera la afirmación de que μ se encuentra entre el LIC y el LSC.

Para precisar el caso planteado más arriba acerca del tiempo de demora para entregar pedidos, supongamos que la varianza poblacional asume un valor de 4. Tomamos una muestra de 49 pedidos, y encontramos una media muestral de 8 días. Luego, para un 95 % de confianza:

$$LIC = 8 - 1{,}96.\frac{2}{\sqrt{49}} = 7{,}44$$

$$LSC = 8 + 1{,}96.\frac{2}{\sqrt{49}} = 8{,}56$$

Esto se interpreta de la siguiente manera: existe un 95 % de confianza en que la demora promedio para entregar un pedido se encuentra entre 7,44 y 8,56 días. La diferencia entre el LSC y el LIC determina la amplitud del intervalo, la cual está inversamente relacionada con la precisión (mayor amplitud implica menor precisión y viceversa).

5.2. Planteo general de la estimación por intervalos

Teniendo en cuenta el ejemplo anterior, podemos describir el procedimiento general para la construcción de intervalos de confianza:

a) Establecer cuál es el parámetro θ desconocido, y qué se conoce de la población.

b) Buscar un estimador puntual, $\hat{\theta} = g(x_1, x_2, \dots, x_n)$, como función de las observaciones muestrales en una muestra de tamaño *n*.

c) Plantear un estadístico como función del estimador y del parámetro $h(\hat{\theta}, \theta)$. Dicho estadístico debe permitir despejar algebraicamente el parámetro (única incógnita de la expresión) y tener una distribución de probabilidad conocida.

d) En estas condiciones, y fijando un nivel de confianza $1 - \alpha$, determinar un intervalo para el estadístico:

$$P(k_1 < h(\hat{\theta}, \theta) < k_2) = 1 - \alpha$$

donde k_1 y k_2 se obtienen a partir de la distribución de probabilidad del estadístico y el nivel de confianza.

e) Despejar el parámetro. Se obtiene entonces un intervalo aleatorio que tiene una probabilidad igual a $1 - \alpha$ de contenerlo.

f) Por último, ya con los valores particulares de una muestra, realizar la estimación del parámetro y obtener los límites inferior y superior de confianza entre los cuales se estima, con una confianza igual a $(1 - \alpha)\ 100\ \%$, que está contenido el parámetro θ

6. Intervalos de confianza para la media, la proporción y la varianza poblacional

Estudiaremos, a continuación, la estimación por intervalos de los principales parámetros poblaciones de interés en nuestra investigación. Comenzaremos con la estimación de la media poblacional, para luego continuar con la estimación de la

proporción poblacional y de la varianza poblacional.

6.1. Intervalo de confianza para la media poblacional

En lo que sigue explicaremos cómo obtener intervalos de confianza para la media de una población aplicando el método general descripto anteriormente. Se plantean distintos casos, según el conocimiento que se tenga de algunos parámetros poblacionales, el tamaño de la muestra, y los supuestos que pueden realizarse acerca de la distribución poblacional.

En los casos siguientes suponemos un muestreo aleatorio simple (con reemplazo). Si se trata de poblaciones finitas y el muestreo es sin reemplazo, debe corregirse el error estándar de la media muestral multiplicando $\sigma/\sqrt{n}$ por el factor de corrección de la varianza $\sqrt{\frac{N-n}{N-1}}$.

Consideraremos los siguientes casos:

- Intervalos de confianza para la media poblacional con varianza poblacional conocida
- Intervalos de confianza para la media poblacional con varianza poblacional desconocida

6.1.1. Varianza poblacional conocida

Este primer caso coincide con el desarrollado en el ejemplo. Siendo μ el parámetro a estimar y conociendo la varianza poblacional, advertimos de inmediato que el estimador puntual será la media muestral y el estadístico a utilizar será:

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

siempre que se trate de una muestra extraída de una población normal (por aplicación del teorema de muestreo en poblaciones normales) o se trate de muestras grandes ($n > 30$) extraídas de cualquier población (por aplicación del teorema central del límite).

Obtenemos los límites para un nivel de confianza $1-\alpha$:

$$LIC = \overline{X} - z_{1-\alpha/2}.\sigma/\sqrt{n}$$

$$LSC = \overline{X} + z_{1-\alpha/2}.\sigma/\sqrt{n}$$

Si la población no es normal y la muestra es menor que 30, no puede usarse este estadístico. Algunas alternativas son utilizar la desigualdad de Tchebycheff para obtener el intervalo de confianza, o bien realizar alguna transformación a la variable para que la distribución poblacional de la variable transformada se aproxime a la normal.

Si tomamos la muestra sin reemplazo, el intervalo queda:

$$LIC = \overline{X} - z_{1-\alpha/2}.\sigma/\sqrt{n}.\sqrt{\frac{N-n}{N-1}}$$

$$LSC = \overline{X} + z_{1-\alpha/2}.\sigma/\sqrt{n}.\sqrt{\frac{N-n}{N-1}}$$

Actividades de aprendizaje

Resuelva la siguiente actividad para aplicar los temas desarrollados.

Actividad 7

El dueño de una estación de servicio desea saber la cantidad de nafta que vende diariamente, en promedio, por cliente/a. Toma una muestra al azar de 36 clientes/as y encuentra que, en promedio, vendió 15 litros de nafta. Sabe, por estudios anteriores, que la población se distribuye aproximadamente normal con una desviación estándar de 2 litros. Encontrar:

a) La estimación puntual de la media poblacional.

b) Un intervalo de confianza del 95 % para la media de la venta diaria de combustible en dicha estación de servicio.

c) Un intervalo de confianza del 99 % (explique la diferencia con el intervalo obtenido en el inciso anterior).

6.1.2. Varianza poblacional desconocida

Si se desconoce la varianza poblacional, no podemos aplicar el estadístico planteado en el punto anterior; ya que existirían dos parámetros desconocidos: μ y σ^2. En este caso se aplica el estadístico:

$$\frac{\overline{X}-\mu}{S/\sqrt{n}}\ t_{n-1},$$

donde *S* es la desviación estándar corregida, estimada a partir de la muestra (recuerde que esta expresión fue deducida en el capítulo 1).

Es importante recordar que, para aplicar la distribución *t*, necesariamente se parte de una distribución poblacional normal. Es decir, bajo los supuestos de que desconocemos la varianza poblacional y la población es normal, podemos construir intervalos de confianza utilizando el estadístico *t*, cualquiera sea el tamaño muestral.

Si la población no es normal, podemos **i**ntentar alguna transformación de la variable original para aproximarla a una normal. Si las muestras son lo suficientemente grandes ($n > 100$), y teniendo en cuenta que *S* es un estimador consistente de σ^2, podemos

usar el estadístico con distribución Normal; reemplazando la varianza poblacional por la muestral. Esto solo puede hacerse con muestras lo suficientemente grandes porque, de lo contrario, con poblaciones no normales y varianza poblacional desconocida, los intervalos resultantes no poseen la confianza esperada. Si la muestra no es mayor que 100 y no puede hacerse una transformación adecuada de la variable, deberá recurrirse a la desigualdad de Tchebycheff. Aunque este procedimiento no es aconsejable, porque los intervalos resultantes son poco precisos.

A continuación, en la Tabla 2.2 presentamos un cuadro resumen de los estadísticos a aplicar para construir intervalos de confianza para μ.

	σ^2 conocida	σ^2 desconocida
Población normal	Para cualquier n $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\ N(0,1)$	$\frac{\overline{X}-\mu}{S/\sqrt{n}}\ t_{n-1}$
Población no normal	$n \geq 30$ por teorema central del límite $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\ N(0,1)$	$n \geq 100$ $\frac{\overline{X}-\mu}{S/\sqrt{n}}\ N(0,1)$
	$n < 30$ – Desigualdad de Tchebycheff o métodos no paramétricos.	$n < 100$ – Desigualdad de Tchebycheff o métodos no paramétricos.

Tabla 2.2. Estadísticos para intervalos de confianza de la media poblacional

Actividades de aprendizaje

Le proponemos realizar las siguientes actividades para profundizar los temas estudiados.

Actividad 8

Un artículo de investigación analiza el ratio contable entre el activo corriente y el pasivo corriente como medida de la solvencia para empresas que distribuyeron dividendos el año pasado y para empresas que no lo hicieron. Presentamos a continuación la salida del programa InfoStat correspondiente al ratio contable para una muestra de empresas que no distribuyeron dividendos. Suponga que el ratio de solvencia tiene distribución Normal.

Medidas resumen

Resumen	Solvencia
n	69,00
Media	1,028
D.E.	0,163
Mín	1,010
Máx	1,072
Mediana	1,050

a) Construir e interpretar el intervalo del 90 % de confianza para el ratio promedio (μ) de la población de empresas que no distribuyen dividendos.

b) Analizar cómo influiría en la precisión del intervalo un aumento del tamaño de muestra, pasando de 69 a 150 casos.

Actividad 9

La página Consumer Review publica en Internet estudios y evaluaciones de diversos productos. La evaluación de una muestra aleatoria de sistemas de sonido, en una escala de evaluación de 1 a 5 (siendo 5 la mejor puntuación), arrojó los siguientes valores:

Medidas resumen

Variable	n	Media	D.E.	Mín	Máx	Mediana	Q1	Q3
Evaluación	20	3,99	0,81	2,14	4,88	4,19	4,00	4,50

a) Estimar la media poblacional, con una confianza de 99 %, asumiendo que la variable evaluación tiene distribución Normal. Interpretar.

b) Comparar los resultados obtenidos en el punto a con los suministrados por InfoStat identificando cada uno de los elementos que aparecen en la salida y el cómo se calculan.

Intervalos de confianza

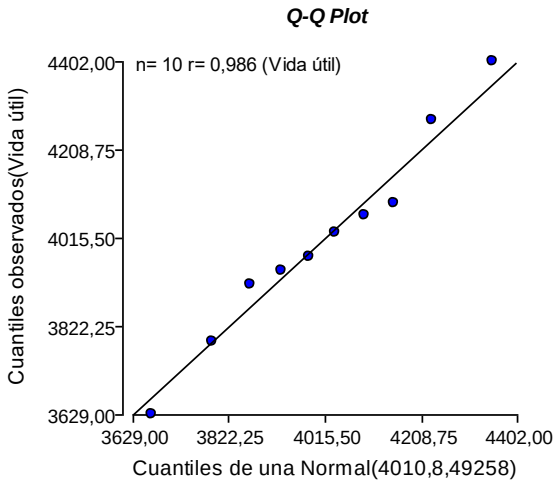
Bilateral
Estimación paramétrica

Variable	Parámetro	Estimación	E.E.	n	LI(99%)	LS(99%)
Evaluación	Media	3,99	0,18	20	3,47	4,51

Actividad 10

El estudio de la duración de focos de luz implica la destrucción de los mismos. Es por ello que se ha analizado una muestra aleatoria de $n = 10$ focos de una nueva marca lanzada al mercado. En base a la información disponible:

Vida Útil (en h)
4402
4066
3788
4028
3973
3629
4275
3944
4090
3913



a) Estimar la vida útil media en la población.

b) Indicar qué estadístico correspondería utilizar para estimar por intervalos la duración poblacional media de esta nueva marca. Verificar que se cumplen los supuestos necesarios.

c) ¿Qué puede decir de la duración promedio de esta nueva marca, a un nivel de confianza del 95 %?

Actividad 11

Los/as agentes de ventas de una empresa presentan semanalmente un informe que enumera los/as clientes/as contactados/as durante la semana. En una muestra de 65 informes semanales la media muestral es 19,5 clientes/as por semana y la desviación estándar muestral es 5,2. Suponga que el número de clientes/as contactados/as semanalmente se distribuye en forma normal.
Estimar con una confianza del 90 % la media poblacional del número de clientes/as contactados/as semanalmente por el personal de ventas.

Actividad 12

El número medio de horas de vuelo de los/as pilotos de la aerolínea *A* es 49 horas por mes. Suponga 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 es 8,5 horas. Considere un nivel de confianza del 95 %.

a) ¿Qué valores asume el error aleatorio?

b) Estimar la media poblacional de las horas de vuelo de los/as pilotos (nivel de confianza: 95 %).

Actividad 13

Para determinar el rendimiento anual de ciertas acciones, un grupo de inversores/as tomó una muestra de $n = 50$. La media y desviación estándar resultaron $8{,}71\ \%$ y $2{,}1\ \%$ respectivamente. Suponiendo que el rendimiento de esta clase de acciones se distribuye en forma aproximadamente normal, estimar su verdadero rendimiento anual promedio usando un intervalo de confianza del $90\ \%$.

Actividad 14

Un estudio presentó datos estadísticos sobre la cantidad de minutos de programa en media hora de transmisión de un canal de cable. Los datos (en minutos) son los siguientes:

21,06	23,82	20,02	23,36	21,23	21,91	22,19	20,62	21,52	21,20
21,66	21,52	22,37	22,24	20,30	22,20	23,44	23,86	23,14	22,34

Suponga que la población es aproximadamente normal. Estimar puntualmente y por intervalos (con una confianza de $99\ \%$) la cantidad media de minutos de programa en media hora de transmisión.

Actividad 15

La película *Harry Potter y la piedra filosofal* fue un éxito desde su estreno. En una muestra de 100 cines, se encontró que la media de la recaudación bruta en el primer fin de semana fue de USD 25.467 por cine, con una desviación estándar de USD 4.980. Estimar la recaudación media poblacional por cine con un nivel de confianza del $95\ \%$.

6.2. Intervalos de confianza para la proporción poblacional

Sea una población dicotómica con N elementos, de los cuales X tienen una determinada propiedad. Luego, $P = X/N$ es el parámetro (desconocido) que pretendemos estimar a partir de una muestra de tamaño n. El estimador de p es $\hat{p} = \frac{x}{n}$, donde $\mathbf{x}$ es el número de elementos con la característica de interés (éxitos) en la muestra. Recuerde que $\mathbf{x}$ tiene distribución binomial, con esperanza $\boldsymbol{n.p}$ y varianza $\boldsymbol{n.p.}(1\text{-}p)$.

Además, si $n.p \geq 5$ y $n.(1\text{-}p) \geq 5$, por el teorema central del límite, es posible utilizar el estadístico:

$$\frac{\hat{p} - P}{\sqrt{P\,(1-P)/n}}\ N(0,1)$$

Deben cumplirse ambas desigualdades. Si alguna de ellas no se cumple, la aproximación normal no es adecuada.

Al despejar *p*, obtenemos un intervalo con probabilidad $1-\alpha$ de contener al parámetro.

$$P\left(\hat{p}-z_{1-\alpha/2}\sqrt{P\,(1-P)/n}\right)<P\left(\hat{p}+z_{1-\alpha/2}\sqrt{P\,(1-P)/n}\right)=1\text{-}\alpha$$

Un problema es que los límites de confianza contienen el parámetro desconocido. No obstante, podemos lograr una buena aproximación de manera más sencilla. Esto es, como *n* es grande, los términos en los que *n* o *n* al cuadrado están en el denominador son próximos a cero. Al despreciar esos términos, resultan los límites:

$$LIC=\hat{p}-z_{1-\alpha/2}\cdot\sqrt{\hat{p}\,(1-\hat{p})/n}$$

$$LSC=\hat{p}+z_{1-\alpha/2}\cdot\sqrt{\hat{p}\,(1-\hat{p})/n}$$

Actividades de aprendizaje

Le proponemos realizar las siguientes actividades para profundizar los temas desarrollados.

Actividad 16

El jefe de producción de una fábrica de radiadores decide estimar el verdadero porcentaje de productos defectuosos que sale de la planta, ante los reiterados rechazos de estos productos que recibe por parte de sus clientes. En caso de que dicho porcentaje supere el 15 % está dispuesto a realizar los ajustes que sean necesarios en el proceso de producción. Con este fin, toma una muestra aleatoria de 120 radiadores y encuentra que 11 de ellos tienen algún defecto. ¿Debe el jefe de producción realizar algún ajuste? Use $1-\alpha=0{,}90$.

Actividad 17

La auditora de una gran empresa encuentra que, de 1.500 cuentas por cobrar controladas, 450 tienen su saldo vencido.

a) Obtener una estimación puntual de *p* (proporción verdadera de todas las cuentas por cobrar que están vencidas).

b) Construir un intervalo de confianza del 95 % para *P*.

Actividad 18

En una muestra de usuarios de Internet (jóvenes entre 15 y 17 años), el 9% votó por Google como el sitio de Internet más popular. Suponga que en este estudio participaron 1.400 jóvenes. ¿Cuál es el error máximo de estimación y la estimación por

intervalo de la proporción poblacional de quienes consideran que este sitio es el más popular? Trabajar con un nivel de confianza del 99,73 %.

Actividad 19

Un grupo de docentes de la Facultad de Ciencias Económicas desea conocer el porcentaje de estudiantes que se dedicarían a la docencia luego de egresar. Una muestra aleatoria de 100 estudiantes arrojó que 39 elegirían la docencia.

a) Obtener alguna conclusión a partir de los datos siguientes:

Intervalos de confianza
Estimación paramétrica

Variable	Parámetro	estimación	E.E.	n	LI(99%)	LS(99%)
docencia	proporción(>0)	0,39	0,05	100	0,26	0,52

b) Un menor nivel de confianza, ¿mejoraría la precisión de la estimación? Justificar la respuesta.

6.3. Intervalos de confianza para la varianza en poblaciones normales

Recordemos que, al estimar la varianza poblacional, el estimador insesgado es la varianza muestral corregida y que el estadístico es:

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

Este estadístico requiere el cumplimiento del supuesto de distribución Normal de los datos. Luego, a partir de la distribución χ^2_{n-1}, construimos el intervalo de confianza:

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

Los valores $\chi^2_{\frac{\alpha}{2}}$ y $\chi^2_{1-\frac{\alpha}{2}}$ corresponden a una distribución χ^2 con n-1 grados de libertad, lo cual deja a la izquierda de $\chi^2_{\frac{\alpha}{2}}$ una probabilidad igual a $\alpha/2$ y a la derecha de $\chi^2_{1-\frac{\alpha}{2}}$ una probabilidad similar.

Al despejar σ^2, obtenemos los límites de confianza:

$$LIC = \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}$$

$$LSC = \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}}$$

Actividades de aprendizaje

Le proponemos realizar las siguientes actividades para profundizar los temas estudiados.

Actividad 20

El Departamento de Control de Calidad de una envasadora de lubricantes desea conocer la varianza al llenar las latas de 4 l. Se sabe que la variable contenido tiene distribución Normal. Interprete la salida que se muestra a continuación:

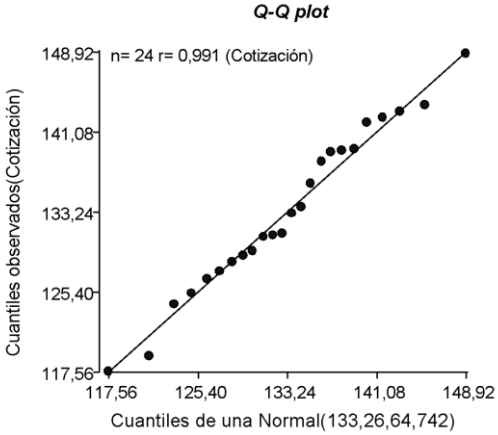
Intervalos de confianza
Estimación paramétrica

Variable	Parámetro	Estimación	E.E.	n	LI(90%)	LS(90%)
contenido	Varianza	0,02	0,01	27	0,01	0,03

Actividad 21

Un agente de bolsa debe asesorar a un nuevo inversor acerca del precio de las acciones de un banco, no solo en su promedio de cotización sino también en su variabilidad. Para ello, computó los valores diarios de cotización durante los primeros 24 días del mes anterior, obteniendo:

Día	Cotización	Día	Cotización	Día	Cotización
1	142,07	9	128,90	17	117,56
2	142,42	10	133,14	18	125,27
3	128,32	11	139,40	19	127,35
4	129,36	12	133,71	20	130,98
5	139,23	13	148,73	21	131,09
6	130,76	14	138,19	22	143,09
7	135,95	15	126,68	23	124,11
8	119,17	16	139,02	24	143,64



Elaborar un pequeño informe para el inversor en base a la información suministrada. Trabajar con un nivel de confianza del 95 %.

Actividad 22

Se efectuaron las siguientes observaciones sobre la resistencia a la fractura de placas base con 18 % de acero:

69,5	71,9	72,6	73,1	73,3	73,5	75,5	75,7	75,8	76,1	76,2
76,2	77,0	77,9	78,1	79,6	79,7	79,9	80,1	82,2	83,7	93,7

Estimar la desviación estándar poblacional de la distribución de resistencia a la fractura (nivel de confianza: 98 %). ¿Bajo qué condiciones es válido el cálculo realizado?

7. Determinación del tamaño de muestra para la media y para la proporción poblacional

Una de las primeras preguntas que surge cuando se desea realizar una estimación de parámetros a partir de una muestra es, naturalmente, ¿de qué tamaño debe ser la muestra?

La pregunta acerca del tamaño necesario de muestra debe considerar: ¿Cuál es el máximo error que se está dispuesto a tolerar? ¿Cuál es el nivel de confianza deseado para las estimaciones? ¿Cuál es la varianza de la población bajo estudio? ¿Se trata de muestreo con o sin reemplazo?

En lo que sigue responderemos a esta cuestión de cómo calcular el tamaño de muestra para estimar la media de una población. Comenzamos por el caso más general.

7.1. Muestreo con reemplazo o en poblaciones infinitas. Estimación de la media y de la proporción poblacional

Supongamos que queremos **estimar la media de una población (μ)** y que se conoce su varianza. Si el nivel de confianza deseado es de $1-\alpha$, entonces:

$$P\left(z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right) = 1-\alpha$$

Aquí, $\left|\overline{X}-\mu\right|$ es el error de estimación que hemos llamado *e*. Al despejar, obtenemos:

$$P\left(|e| < z.\sigma/\sqrt{n}\right) = 1-\alpha$$

Donde $z = z_{1-\alpha/2}$. Podemos leer esta expresión como: existe una alta probabilidad $(1-\alpha)$ de que el error de estimación sea, como máximo, igual a $z.\frac{\sigma}{\sqrt{n}}$. Es decir, e sería el error máximo de estimación.

$$|e| = z.\frac{\sigma}{\sqrt{n}}$$

$$n = \frac{z^2\sigma^2}{e^2}$$

Este es el tamaño de muestra necesario para que *e* sea el máximo error de estimación. Un *n* mayor reducirá el error máximo de estimación.

Esta fórmula permite conocer de qué tamaño debe ser la muestra para que el error no supere a *e*, eligiendo *z* de acuerdo con el nivel de confianza deseado. El tamaño muestral depende, entonces, del nivel de error aceptable (*e*), del nivel de confianza deseado (reflejado en *z*) y de la dispersión de la variable en la población (σ^2). La aplicación de la fórmula que analizamos presenta algunos inconvenientes:

- La necesidad de conocer la varianza poblacional. Esta situación es superada en la práctica utilizando una varianza conocida por experiencias anteriores vinculadas con la misma variable o con alguna variable *proxy* (otra variable con comportamiento similar a la que se tiene en estudio). También es posible estimar la varianza a partir de una muestra piloto, o bien utilizando el conocimiento de un experto sobre la forma de la distribución poblacional y de los valores mínimo y máximo que pueden encontrarse en la muestra.
- Por otra parte, el error en la fórmula está expresado en forma absoluta, en las unidades de la variable en cuestión. Para evaluar cuándo un cierto nivel de error se considera elevado o aceptable, es necesario relacionarlo con los valores posibles de la variable y de su media. Este aspecto puede resolverse realizando los cálculos con una medida de error relativo, tal como analizaremos más adelante.
- En tercer lugar, cuando realizamos una investigación, generalmente procuramos estimar parámetros de varias variables: ¿a cuál de ellas hay que referirse para calcular el tamaño de la muestra? Esta cuestión generalmente se resuelve seleccionando la variable más relevante, o la que se supone de mayor variabilidad, con el fin de no subestimar el tamaño de la muestra.

Si queremos **estimar una proporción**, el mismo razonamiento previo conduce a la fórmula:

$$n = \frac{z^2 p(1-p)}{e^2}$$

puesto que se trata de una población dicotómica en la que se desea estimar la proporción de éxitos mediante una muestra.

Nuevamente nos surge el problema de tener que colocar en la fórmula un valor de p que previamente se desconoce, dado que es el parámetro que se desea estimar. En este caso, observamos que si $p = 1\text{-}p = 0{,}5$, el producto $p.(1\text{-}p)$ alcanza su valor máximo $(0{,}25)$. Este es el que debemos colocar en la fórmula si no hay ningún conocimiento acerca del posible valor de *p*.

Tengamos en cuenta que, dada la naturaleza dicotómica de la variable, cuando debamos precisar el error máximo (e) también lo haremos como una proporción.

Ejemplo

Supongamos que se quiere conocer qué tamaño debe tener una muestra si en una población muy grande se desea estimar la proporción de personas adultas que no han completado la educación media. Además:

- No se tiene ninguna información acerca de esa posible proporción. Por lo tanto, se usará una $p = 0{,}50$.
- Se desea un nivel de confianza en la estimación del $95\ \%$, por lo tanto, $z = 1{,}96$.
- Se pretende que el error no supere a $0{,}05$ (es decir, que la estimación resulte en un valor de p estimado $\pm\, 0{,}05$).

Aplicamos la fórmula: $n = \frac{(1{,}96)^2 0{,}50.0{,}50}{0{,}05^2} \qquad n \geq 385$

Luego, tenemos que será necesario tomar una muestra no menor a 385 casos para satisfacer las exigencias planteadas. Notemos que se coloca $\geq$ y no $=$, porque consideramos un tamaño mínimo de muestra.

Este tamaño de muestra nos puede parecer excesivo, pero de este orden son las muestras necesarias cuando se quieren estimar proporciones. No ocurre lo mismo cuando se desean estimar medias poblacionales para variables cuantitativas, ya que pueden tener varianzas pequeñas (como suele ocurrir en biología o ciencias naturales). En esos casos, sobre todo si se trata de poblaciones normales, a veces es posible trabajar con muestras pequeñas. Si existieran elementos de juicios previos para saber algo acerca de p, colocaríamos el valor en la fórmula y así la muestra resultaría menor que si utilizamos el valor $0{,}50$.

En las fórmulas planteadas para determinar el tamaño muestral podemos observar que el tamaño de la muestra depende, como hemos visto, de z, de la varianza y del error máximo aceptable. Es conveniente reflexionar acerca de estos elementos:

- La variable z es el valor extraído de la tabla de la distribución Normal. Determina la confianza de las estimaciones que se realizarán $(1 - \alpha)$ y, por lo tanto, el riesgo α de equivocarse al afirmar que cierto intervalo contiene al valor del parámetro.
- La varianza, σ^2 o $p.(1-p)$ indica una característica de la población con respecto a la variable de interés. La mayor o menor dispersión en la población incide en el tamaño muestral.
- El valor e es el error máximo aceptable medido en términos de la variable en

cuestión (si es cuantitativa), o en términos de proporción (si la variable es dicotómica).

Si en las fórmulas planteadas despejamos *e*, observamos que el error máximo depende del riesgo aceptado, de la varianza y del tamaño de la muestra.

Actividades de aprendizaje

Le proponemos realizar las siguientes actividades para profundizar los temas estudiados.

Actividad 23

El ratio precio/ganancia (P/G) en las acciones de la Bolsa de Nueva York tenía una desviación estándar de 10 en el primer trimestre del año pasado. Si se desea estimar la media poblacional del ratio P/G en todas las acciones de la Bolsa de Nueva York:

a) ¿Cuántas acciones habrá que tomar en la muestra si se está dispuesto a correr un riesgo del 1 % de cometer un error de estimación de 2 o menos?

b) ¿Cuál será el tamaño de la muestra si se aumenta el error de estimación a 5?

c) ¿Cuál sería el tamaño de *n* (con el error del punto a) si se aceptara un riesgo del 10 %?

Analice los valores obtenidos en cada caso y explique las diferencias.

Actividad 24

Un informe presenta los tiempos de traslado al trabajo que son requeridos en las cinco ciudades más grandes de Argentina. Suponga que se utiliza una muestra aleatoria piloto de los habitantes de Buenos Aires y se conoce la desviación estándar poblacional: 6,25 minutos. Si desea estimar la media poblacional del tiempo que necesitan en Buenos Aires para trasladarse al trabajo, con un error de estimación de 2 minutos, ¿cuál debe ser el tamaño de la muestra? Suponga que el nivel de confianza es del 95 %.

Actividad 25

Se quiere estimar la incidencia de la hipertensión arterial en el embarazo. En los siguientes casos, ¿cuántas mujeres embarazadas habrá que observar para estimar (con una confianza del 95 %) dicha incidencia con un error del 2 %?:

a) Sabiendo que en un sondeo previo se ha observado un 9 % de mujeres hipertensas.
b) Sin ninguna información previa.

Explique la diferencia entre ambos resultados.

Actividad 26

La Asamblea de Pequeños y Medianos Empresarios (APYME) necesita realizar un estudio sobre la proporción de pymes exportadoras que han sido beneficiadas con la devaluación del peso argentino. El Ministerio de la Producción estima que el 65 % de ellas han aumentado su nivel de ingresos por esta razón. ¿Cuántas empresas deberá consultar esta organización si desea realizar una estimación precisa de dicho porcentaje?, definiendo un error del 3 % y tomando dos niveles de confianza:

a) 95,45 %
b) 99,73 %

Explique el efecto que tuvo sobre el tamaño de la muestra el cambio en el nivel de confianza.

Actividad 27

Se tiene previsto realizar una consulta popular con el fin de conocer la opinión de las personas acerca de la anticipación de las elecciones. ¿A cuántos/as votantes se deberá entrevistar en un sondeo previo si se desea cometer un error no superior al 10 % y depositar una confianza del 95 % en las conclusiones?

Actividad 28

Suponga que una investigadora desea realizar un estudio sobre el gasto de las familias en la ciudad de Córdoba. Como no cuenta con información referida a esa variable utilizará como desviación estándar muestral la referida al ingreso familiar, obtenida en la encuesta permanente de hogares, que es $S = \$2.800$.

a) Encontrar el tamaño de muestra adecuado teniendo en cuenta que la investigadora indica que el máximo error muestral no debe ser mayor que $500 por arriba o debajo de la verdadera media del gasto, y que se debe tomar un nivel de confianza del 95,45 %.

b) Indicar cuál es el valor del riesgo si se toma una muestra de tamaño 500 y el error máximo es el del inciso a. Analizar.

c) ¿Cuál sería el error si el nivel de confianza es de 95,45 % y el tamaño de la muestra es de 500 personas? Analizar.

7.1.1. Determinación del tamaño de la muestra teniendo en cuenta el error relativo

Determinar el máximo error para una estimación en **forma absoluta** puede constituir un serio inconveniente, especialmente cuando desconocemos la magnitud aproximada de los parámetros a estimar, sean estos medias o proporciones. Si, en cambio, expresamos el error en **forma relativa**, como por ejemplo un 10 % de la media o un 5 %

de la verdadera proporción, superaríamos dicho inconveniente y mantendríamos un error máximo acotado según las necesidades. No obstante, aparecen otros inconvenientes a la hora de aplicar estas fórmulas. Simbolizando con ε la proporción del parámetro deseada como error máximo, tenemos que:

$$e = \varepsilon . \mu$$

$$e = \varepsilon . P$$

Por lo tanto, para estimar la media:

$$n = \frac{z^2 \sigma^2}{\varepsilon^2 . \mu^2}$$

Para estimar la proporción:

$$n = \frac{z^2 P(1-P)}{\varepsilon^2 . P^2}$$

Como el **coeficiente de variación** (CV) de una variable es σ/μ, resulta que:

$$n = \frac{z^2 (CV)^2}{\varepsilon^2}$$

Estas fórmulas son adecuadas para calcular el tamaño de la muestra cuando conocemos el coeficiente de variación si se trata de una variable cuantitativa, o tenemos alguna idea del valor de *P* si la variable es dicotómica. Así, por ejemplo, si $\varepsilon = 0{,}10$, significa que el máximo error aceptable es un $10\ \%$ del valor del parámetro, sea este una media o una proporción.

En el caso de la media, suele resultar una ventaja el uso del coeficiente de variación en lugar de la varianza, ya que puede ser más estable que la varianza. Es posible utilizar los valores obtenidos con anterioridad para ese coeficiente, o de otras variables con un comportamiento similar.

7.2. Determinación del tamaño de muestra para poblaciones finitas

Cuando trabajamos con una población finita y extraemos la muestra sin reemplazo, es necesario aplicar el factor de corrección para las varianzas. Para el caso de la media, tenemos que:

$$n = \frac{z^2 \sigma^2}{e^2} . \frac{N-n}{N-1}$$

Como *n* aparece en ambos miembros es necesario despejar. Llamando n_∞ al resultado de calcular el tamaño de muestra con reemplazo, nos queda:

$$n = n_\infty . \frac{N-n}{N-1}$$

Y, al despejar, obtenemos:

$$n = \frac{n_\infty . N}{(N-1) + n_\infty}$$

Podemos comprobar que se obtiene exactamente la misma fórmula para el caso de la proporción.

El tamaño de muestra así calculado es menor al del muestreo con remplazo. Sin embargo, para que esa reducción tenga un efecto interesante, *n* debe ser grande comparado con *N* (tamaño de la población). La evidencia empírica indica que debe aplicarse la fórmula corregida solo cuando el tamaño de la muestra calculado resulta superior al 5 % del tamaño de la población. De lo contrario, no es necesaria la corrección, ya que la reducción en el costo (por menor tamaño muestral) será imperceptible.

Actividades de aprendizaje

Le proponemos realizar las siguientes actividades para profundizar los temas estudiados.

Actividad 29

Una compañía de transporte local de pasajeros necesita establecer una línea desde un determinado barrio hasta el centro de la ciudad. Dicha empresa quiere estimar la proporción de usuarios/as que utilizarían esta nueva ruta con una confianza del 95 % y un error máximo de ± 0,02. ¿Cuántas personas, sobre una población de 8.000 potenciales usuarios/as, debería entrevistar la compañía a fin de tomar la decisión de implementar el nuevo servicio? Analizar las diferencias entre *n* y n_∞.

Actividad 30

El mantenimiento de cuentas de crédito puede resultar demasiado costoso si el promedio de compra por cuenta es menor a un cierto nivel. El gerente de una empresa quiere conocer la cantidad promedio mensual de compras realizadas por clientes/as que usan cuentas de crédito, con un error de no más de $500 y una confianza aproximada de 95 %.
Si el gerente sabe, además, que la desviación estándar de las cuentas de crédito es de $1.500, ¿cuántas unidades debe seleccionar de su archivo de 4.000 cuentas? ¿Por qué los valores de *n* y n_∞ son prácticamente iguales?

Actividad 31

Una compañía de televisión por cable, que cuenta con 5.000 suscriptores/as en una determinada ciudad, quiere estimar la proporción de los que estarían dispuestos/as a

comprar un pack especial de programación. Dicha empresa quiere tener un 95 % de confianza de que su estimación es correcta, con un error aproximado de ± 0,05. La experiencia previa en otras ciudades indica que cerca del 30 % de personas suscriptas adquieren el pack. ¿Qué tamaño de muestra es el adecuado a estos requerimientos si el muestreo se realiza sin reposición?

Actividad 32

Se desea estimar la resistencia media a la tracción de una remesa de 1.000 alambres de acero. Se conoce de estudios anteriores que la desviación estándar de la resistencia a la tracción es de aproximadamente 9,07 kg. ¿Qué cantidad de alambres de acero se deberán muestrear, sin reposición, de la remesa si se desea trabajar con una confianza del 95 % de cometer un error de muestreo no mayor a 3 kg?

Actividad 33

Una supervisora del proceso de empacado de café en sobres tomó una muestra aleatoria de tamaño 12 en la planta empacadora. El peso neto de dichos sobres de café fue el siguiente:

Gramos por sobre	Cantidad de sobres
15,7	1
15,8	2
15,9	2
16	3
16,1	3
16,2	1

Suponiendo que el peso del café empacado tiene distribución normal, estimar el peso promedio por sobre utilizando un nivel de confianza del 98%.

Actividad 34

En una muestra aleatoria de 15 alumnos/as del 5º año de un colegio secundario se encontró que 5 tenían decidida la carrera universitaria a seguir. Estimar la proporción de estudiantes secundarios que tienen decidido qué carrera seguir, a un nivel del 99 %.

Actividad 35

El dueño de una radio quiere conocer la proporción de gente a la que le gustan los programas deportivos. A cuántas personas deberá encuestar si:

a) - El error muestral no debe ser mayor a un 2 % de la proporción.

- El nivel de confianza es del 95 %.
- La proporción de gente que gusta de estos programas es aproximadamente 0,60.

b) No se conoce nada acerca de la proporción de gente que gusta de este tipo de programas.

Actividad 36

Indicar si las siguientes afirmaciones son verdaderas o falsas, justificando en todos los casos su respuesta:

a) Un intervalo de confianza del 99 % para la media siempre contiene el valor desconocido de la media poblacional.

b) Si el tamaño de la muestra y la varianza poblacional permanecen constantes, y se disminuye el nivel de confianza del intervalo, entonces su amplitud aumenta.

c) Si el tamaño de la muestra y el nivel de confianza son fijos, y la variabilidad poblacional es mayor a la supuesta originalmente, entonces el nuevo intervalo de confianza que se obtiene será menos preciso.

d) Mientras más observaciones se toman, menor amplitud tendrán los intervalos que se construyan.

e) Si se tiene una muestra grande, se puede usar la distribución Normal para plantear intervalos de confianza.

f) Si disminuye el nivel de confianza, disminuye la amplitud del intervalo y disminuye también el error estándar del estimador.

Actividad 37

Una pequeña industria dedicada a la fabricación de pilas recibe reclamos de sus clientes/as porque las mismas duran menos de 1.000 h. El fabricante pretende mostrar una estimación de la verdadera duración promedio de sus pilas. Se conoce que la desviación típica poblacional es de 120 h y que la variable posee distribución Normal.
La información disponible es:

Estadística Descriptiva

Resumen	Pilas
n	81
Media	1203,38
Var(n-1)	12780,14
E.E.	12,56
Mín	905
Máx	1523
Mediana	1201
Q1	1121
Q3	1267

a) Indicar cuál sería la estimación al nivel del 90 %.

b) Si el error máximo tolerado fuera de 47,04 h, ¿qué cantidad de pilas debería haber seleccionado? Se conoce que la producción diaria de la empresa es de $N = 345$.

Actividad 38

Se tomó una muestra de 61 empresas exportadoras de la industria alimenticia que arrojó los siguientes volúmenes de exportación anual:

Exportaciones (miles de $)	Cantidad de empresas
10-50	9
50-90	19
90-130	21
130-170	7
170-250	5

Con un 90 % de confianza, estimar:

a) La exportación media anual de las empresas de esta rama industrial.

b) El porcentaje de empresas que no han alcanzado más de $90.000 de exportación anual.

8. Respuesta a las actividades

Actividad 1

El mejor estimador cumplirá las propiedades de los buenos estimadores: insesgadez, suficiencia, consistencia y eficiencia. Como la muestra es pequeña, ($n = 3$) no analizaremos la consistencia.

	Insesgadez $E(\hat{\theta}) = \theta$	**Suficiencia** (toda la inform. muestral)	**Eficiencia** (estimador con la menor varianza en relación a otro)
$\hat{\mu}_1 = \frac{1}{3}x_1 + \frac{1}{3}x_2 + \frac{1}{3}x_3$	si (1)	si	es más eficiente que $\hat{\mu}_2$ y $\hat{\mu}_3$ $V(\hat{\mu}_1) < V(\hat{\mu}_2) < V(\hat{\mu}_3)$
$\hat{\mu}_2 = \frac{1}{9}x_1 + \frac{3}{9}x_2 + \frac{5}{9}x_3$	si (2)	si	es más eficiente que $\hat{\mu}_3$ $V(\hat{\mu}_2) < V(\hat{\mu}_3)$
$\hat{\mu}_3 = \frac{6}{10}x_1 + \frac{4}{10}x_3$	si (3)	no	Es el que posee menos eficiencia

(1)	$V(\hat{\mu}_1)=V(⅓\ x_1+⅓\ x_2+⅓\ x_3)$
$E(\hat{\mu}_1)=E(⅓\ x_1+⅓\ x_2+⅓\ x_3)$	$V(\hat{\mu}_1)=V(⅓\ x_1)+V(⅓\ x_2)+V(⅓\ x_3)$
$E(\hat{\mu}_1)=E(⅓\ x_1)+E(⅓\ x_2)+E(⅓\ x_3)$	$V(\hat{\mu}_1)=1/9.V(x_1)+1/9.V(x_2)+1/9.V(x_3)$
$E(\hat{\mu}_1)=⅓.E(x_1)+⅓.E(x_2)+⅓.E(x_3)$	$V(\hat{\mu}_1)=1/9\sigma^2+1/9.\sigma^2+1/9.\sigma^2$
$E(\hat{\mu}_1)=⅓\mu+⅓.\mu+⅓.\mu$	$V(\hat{\mu}_1)=3/9\sigma^2$
$E(\hat{\mu}_1)=3/3\mu$	$V(\hat{\mu}_1)=0{,}3333\sigma^2$
(2)	$V(\hat{\mu}_2)=V(1/9\ x_1+3/9\ x_2+5/9\ x_3)$
$E(\hat{\mu}_2)=E(1/9\ x_1+3/9\ x_2+5/9\ x_3)$	$V(\hat{\mu}_2)=V(1/9\ x_1)+V(3/9\ x_2)+V(5/9x_3)$
$E(\hat{\mu}_2)=E(1/9\ x_1)+E(3/9x_2)+E(5/9x_3)$	$V(\hat{\mu}_2)=1/81V(x_1)9/81.V(x_2)+25/81.V(x_3)$
$E(\hat{\mu}_2)=1/9.E(x_1)+3/9(x_2)+5/9.E(x_3)$	$V(\hat{\mu}_2)=1/81\sigma^2+9/81.\sigma^2+25/81.\sigma^2$
$E(\hat{\mu}_2)=1/9\ \mu+3/9\ .\ \mu+5/9\ .\ \mu$	$V(\hat{\mu}_2)=35/81\ \sigma^2$
$E(\hat{\mu}_2)=9/9\ \mu$	$V(\hat{\mu}_2)=0{,}432\ \sigma^2$
(3)	$V(\hat{\mu}_3)=V(6/10x_1+4/10x_3)$
$E(\hat{\mu}_3)=E(6/10x_1+4/10x_3)$	$V(\hat{\mu}_3)=V(6/10x_1)+V(4/10\ x_2)$
$E(\hat{\mu}_3)=E(6/10x_1)+E(4/10\ x_2)$	$V(\hat{\mu}_3)=36/100.V(x_1)+16/100.V(x_2)$
$E(\hat{\mu}_3)=6/10.E(x_1)+4/10.E(x_2)$	$V(\hat{\mu}_3)=36/100\sigma^2+16/100.\sigma^2$
$E(\hat{\mu}_3)=6/10\mu+4/10.\mu$	$V(\hat{\mu}_3)=51/100\sigma^2$
$E(\hat{\mu}_3)=10/10\mu$	$V(\hat{\mu}_3)=0{,}51\sigma^2$

Actividad 2

a) 1. El estimador insesgado y eficiente de la media poblacional es: $\overline{X}=6{,}35$

$$\hat{\mu}_1 = \overline{X}$$

$$E\ (\hat{\mu}_1) = E\ (\overline{X}) = \mu \text{ (se verifica que es insesgado)}$$

$$Var\ (\hat{\mu}_1) = Var\ (\overline{X}) = \frac{\sigma^2}{n} = \frac{\sigma^2}{5}$$

a) 2. El estimador insesgado y eficiente de la varianza poblacional es:
$S^2 = 0{,}00055$

b) El estimador insesgado pero ineficiente del diámetro promedio poblacional es:

$$\hat{\mu}_2 = 3/7\ X_1 + 1/7\ X_2 + 1/7\ X_3 + 1/7\ X_4 + 1/7\ X_5$$

$E\ (\hat{\mu}_2) = E\ (3/7\ X_1 + 1/7\ X_2 + 1/7\ X_3 + 1/7\ X_4 + 1/7\ X_5)$

$E\ (\hat{\mu}_2) = E\ (3/7\ x_1) + E\ (1/7\ x_2) + E\ (1/7\ x_3) + E\ (1/7\ X_4) + E\ (1/7\ X_5)$

$E\ (\hat{\mu}_2) = 3/7\ E(x_1) + 1/7\ E\ (x_2) + 1/7\ E\ (x_3) + 1/7\ E\ (x_4) + 1/7\ E\ (x_5)$

$E\ (\hat{\mu}_2) = 3/7\ \mu + 1/7\ \mu + 1/7\ \mu + 1/7\ \mu + 1/7\ \mu$

$E\ (\hat{\mu}_2) = 7/7\ \mu$ (se verifica que es insesgado)

$V(\hat{\mu}_2) = V(3/7\ X_1 + 1/7\ X_2 + 1/7\ X_3 + 1/7\ X_4 + 1/7\ X_5)$

$V(\hat{\mu}_2) = 9/49.V(X_1) + 1/49.V(X_2) + 1/49.V(X_3) + 1/49.V(X_4) + 1/49\ V(X_5)$

$V(\hat{\mu}_2) = 9/49\ \sigma^2 + 1/49\ \sigma^2 + 1/49\ \sigma^2 + 1/49\ \sigma^2 + 1/49\ \sigma^2$

$V(\hat{\mu}_2) = 13/49\ \sigma^2$

$V(\hat{\mu}_2) = 0{,}265\ \sigma^2$

Observamos que $V(\hat{\mu}_1) < V(\hat{\mu}_2)$, por lo que $\hat{\mu}_1$ es más eficiente.
El estimador propuesto es ineficiente

c) Determinar un estimador insuficiente de la media poblacional.

$Mo = 6{,}37$

ya que no utiliza toda la información disponible en la muestra.

Actividad 3

Sea X: número de éxitos en n repeticiones de un experimento binomial con probabilidad de éxito igual a *p*. El estimador de *P* es $\hat{p} = \frac{x}{n}$. Verificar que este estimador es insesgado.

$E(\hat{p}) = E\left(\frac{x}{n}\right)$ siendo x: cantidad de éxitos

$E(\hat{p}) = 1/n\ (np)$

$E(\hat{p}) = P$

Actividad 4

X = número de siniestros de un determinado tipo una compañía de seguros por semana.

$$f(x_i, \lambda) = \frac{e^{-\lambda}.\lambda^x}{x_i!}$$

$$L(\lambda) = f(x_1, x_2, x_3, \ldots x_n, \lambda)$$

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}.\lambda^x}{x_i!}$$

Al aplicar propiedades de las potencias de igual base:

$$L(\lambda) = \frac{e^{-\lambda n}.\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

Tomando logaritmo natural en ambos miembros:

$$lnL(\lambda) = \ln\frac{e^{-\lambda n}.\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

Aplicamos las propiedades del logaritmo[1]:

[1] log (a · b) = log (a) + log (b)
log (a / b) = log (a) - log (b)
log (ac) = c log (a)

$$lnL(\lambda) = \ln e^{-\lambda n} + \ln \lambda^{\sum_{i=1}^{n} x_i} - \ln \prod_{i=1}^{n} x_i!$$

$$lnL(\lambda) = -\lambda n \ln e + \sum_{i=1}^{n} x_i \ln \lambda - \ln \prod_{i=1}^{n} x_i!$$

$$lnL(\lambda) = -\lambda n + \sum_{i=1}^{n} x_i \ln \lambda - \ln \prod_{i=1}^{n} x_i!$$

$$\frac{\partial lnL(\lambda)}{\partial \lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i - 0$$

Para verificar si existe un máximo o mínimo relativo hay que igualar la derivada primera a cero:

$$\frac{\partial lnL(\lambda)}{\partial \lambda} = 0$$

$$-n + \frac{1}{\hat{\lambda}} \sum_{i=1}^{n} x_i - 0 = 0$$

$$\frac{1}{\hat{\lambda}} \sum_{i=1}^{n} x_i = n$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$$

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

Luego, se verifica que la derivada segunda sea menor a cero para comprobar el cumplimiento de las condiciones de segundo orden.

Actividad 5

$$f(x_1, x_2, x_3, \ldots x_n, \lambda) = L(\lambda)$$

$$f(x_1, x_2, x_3, \ldots x_n, \lambda) = \lambda^n e^{-\lambda \sum x} \quad \text{para } x \geq 0$$

$$\ln L(\lambda) = \ln (\lambda^n e^{-\lambda \sum x})$$

$$\ln L(\lambda) = [\ln \lambda^n + (-\lambda \sum x_i) \ln e]$$

$$\ln L(\lambda) = [n \ln \lambda + -\lambda \sum x_i]$$

$$\frac{\partial lnL}{\partial \lambda} = n \frac{1}{\lambda} - \sum x_i \rightarrow \frac{\partial L}{\partial \lambda} = 0$$

$$n \frac{1}{\hat{\lambda}} - \sum x_i = 0$$

$$\frac{1}{\hat{\lambda}} = \frac{\sum x_i}{n}$$

$$\frac{1}{\hat{\lambda}} = \overline{X} \rightarrow \hat{\lambda} = 1/\overline{X}$$

Luego, se verifica que la derivada segunda sea menor a cero para comprobar el cumplimiento de las condiciones de segundo orden.

Actividad 6

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Actividad 7

a) La estimación puntual de la media poblacional es: $\bar{X} = 15$

b) Como *X* se distribuye aproximadamente normal y se conoce σ, el estadístico a utilizar es *z*:

$\text{LIC} = \bar{X}$ - $z\,\sigma/\sqrt{n}$

$\text{LIC} = 15 - 1{,}96 \text{ x } 2/\sqrt{36} = 14{,}34$

$\text{LSC} = \bar{X} + z\,\sigma/\sqrt{n}$

$\text{LSC} = 15 + 1{,}96 \text{ x } 2/\sqrt{36} = 15{,}65$

c) $\text{LIC} = 15 - 2{,}576 \text{ x } \frac{2}{\sqrt{36}} = 14{,}14$

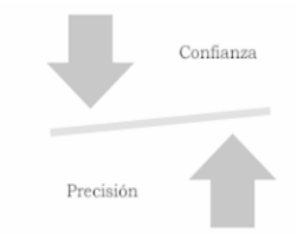
$\text{LSC} = 15 + 2{,}576 \text{ x } \frac{2}{\sqrt{36}} = 15{,}85$

La diferencia con respecto al intervalo calculado en el inciso anterior está en la amplitud del intervalo. En la medida que aumentamos el nivel de confianza, la longitud o amplitud del intervalo se hace más grande y se reduce la precisión en la estimación por intervalo.

Longitud = LSC - LIC

Longitud = $\bar{X} + z\,\sigma/\sqrt{n}$ - $(\bar{X}$ - $z\,\sigma/\sqrt{n})$

Longitud = $2\,z\,\sigma/\sqrt{n}$



Si quiero que la longitud sea pequeña y que tenga mucha confianza puedo aumentar *n*.
Si $1\text{-}\alpha$ aumenta → *z* aumenta → longitud aumenta

Actividad 8

Datos:
X_i = ratio contable de solvencia de la empresa *i* que no distribuyó dividendos.
$n = 69$
$1\text{-}\alpha = 0{,}90$

Resumen	Solvencia
n	69,00
Media	1,028
D.E.	0,163
Min.	1,010
Max.	1,072
Mediana	1,050

a) $\text{LIC} = \bar{X}$ - $t_{\alpha/2}\,S/\sqrt{n}$

$\text{LSC} = \bar{X} + t_{1-\alpha/2}\,S/\sqrt{n}$

$\text{LIC} = 1{,}028 - 1{,}668 \text{ x } 0{,}163/\sqrt{69} = 0{,}9953$
$\text{LSC} = 1{,}028 + 1{,}668 \text{ x } 0{,}163/\sqrt{69} = 1{,}0607$

Existe un 90 % de confianza en que el ratio contable promedio para empresas que no distribuyeron dividendos se encuentre entre 0,9953 y 1,0607.

b) Si se aumentan los casos de 69 a 150 casos:
$\text{LIC} = 1{,}028 - 1{,}655 \text{ x } 0{,}163/\sqrt{150} = 1{,}006$
$\text{LSC} = 1{,}028 + 1{,}655 \text{ x } 0{,}163/\sqrt{150} = 1{,}050$

Al aumentar el tamaño de la muestra observamos que la amplitud del intervalo se reduce, con lo cual la precisión de la estimación aumenta.

Actividad 9

Medidas resumen

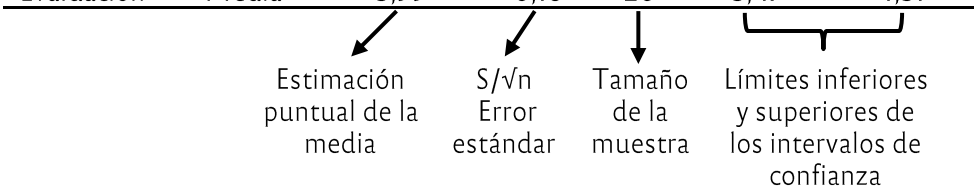
Variable	n	Media	D.E.	Mín	Máx	Mediana	Q1	Q3
Evaluación	20	3,99	0,81	2,14	4,88	4,19	4,00	4,50

a) Como X se distribuye aproximadamente normal, la muestra es pequeña ($n < 30$) y no se conoce σ, el estadístico a utilizar es t
$\text{LIC} = \bar{X} - t\, S/\sqrt{n}$
$\text{LIC} = 3{,}99 - 2{,}861 \text{ x } 0{,}81/\sqrt{20} = 3{,}47$
$\text{LSC} = \bar{X} + t\, S/\sqrt{n}$
$\text{LSC} = 3{,}99 + 2{,}861 \text{ x } 0{,}81/\sqrt{20} = 4{,}51$

Existe un 99 % de confianza en que la evaluación promedio incluya el verdadero valor del parámetro μ.

b) Intervalos de confianza
Bilateral
Estimación paramétrica

Variable	Parámetro	Estimación	E.E.	n	LI (99%)	LS (99%)
Evaluación	Media	3,99	0,18	20	3,47	4,51



Actividad 10

a) $\overline{X} = 4010{,}80$ h

b) En este caso, como tenemos una población que se distribuye normal (Q-Q Plot) con varianza poblacional desconocida y la muestra es pequeña ($n = 10$), el estadístico a utilizar es:

$$t = \frac{\underline{x} - \mu}{S/\sqrt{n}}\ t_{n-1}$$

c) $n = 10$
$S = 221{,}94$
$\bar{X} = 4010{,}8$
$LIC = \bar{X}\text{-}\ t_{\alpha/2}\ s/\sqrt{n}$
$LIC = 4.010{,}8 - 2{,}262 \text{ x } \frac{221{,}94}{\sqrt{10}} = 3852{,}04$
$LSC = \bar{X} + t_{1-\alpha/2}\ s/\sqrt{n}$
$LIC = 4.010{,}8 + 2{,}262 \text{ x } \frac{221{,}94}{\sqrt{10}} = 4169{,}56$
Con un 95 % de confianza la duración promedio de los focos de la nueva marca se encuentra entre las 3.852,04 y 4.169,56 horas de vida útil.

Actividad 11

X = Número de informes semanales de contacto de clientes/as
Datos: $\bar{X} = 19{,}5$, $n = 65$ y $S = 5{,}2$
$LIC = 19{,}5 - 1{,}669 - 5{,}2/\sqrt{65} = 18{,}42$
$LSC = 19{,}5 + 1{,}669\ 5{,}2/\sqrt{65} = 20{,}58$
Existe un 90 % de confianza en que el intervalo asociado a los informes semanales construido incluya el verdadero valor del parámetro **μ.**

Actividad 12

a) *X*: número de horas de vuelo de los/as pilotos de la aerolínea *A* por mes.
$\bar{X} = 49,\ S = 8{,}5$ (asimilable a σ), $n = 100$ y $1 - \alpha = 0{,}95$
El error aleatorio o error de estimación es:
$|e| = z.S/\sqrt{n}$
$|e| = 1{,}96.8{,}5\sqrt{100}$
$|e| = +/- 1{,}666$
b) $LIC = \bar{X}$ - z $\sigma/\sqrt{n}$
$LIC = 49 - 1{,}96 \text{ x } 8{,}5/\sqrt{100} = 47{,}33$
$LSC = \bar{X}+$ z $\sigma/\sqrt{n}$
$LSC = 49 + 1{,}96 \text{ x } 8{,}5/\sqrt{100} = 50{,}67$
Con un 95 % de confianza las horas de vuelo promedio de los/as pilotos se encuentra entre 47,33 y 50,67 horas.

Actividad 13

X = Rendimiento anual de ciertas acciones (expresado en %)

$n = 50$, $\bar{X} = 8{,}71$ y $S = 2{,}1$
$\text{LIC} = 8{,}71 - 1{,}677 \text{ x } 2{,}1/\sqrt{50} = 8{,}21$
$\text{LSC} = 8{,}71 + 1{,}677 \text{ x } 2{,}1/\sqrt{50} = 9{,}21$
Existe un $90\ \%$ de confianza en que el intervalo asociado a los rendimientos anuales construido incluya al verdadero valor del parámetro μ.

Actividad 14

X = minutos de programa en media hora de transmisión
$n = 20$
$1 - \alpha = 0{,}99$
$S = 1{,}12$
$\bar{X} = 22$
$\text{LIC} = 22 - 2{,}861 \text{ x } 1{,}12/\sqrt{20} = 21{,}28$
$\text{LSC} = 22 + 2{,}861 \text{ x } 1{,}12/\sqrt{20} = 22{,}72$
La cantidad media de minutos de programa en media hora de transmisión se encuentra entre $[21, 28; 22, 72]$ con un $99\ \%$ de confianza.

Actividad 15

X = USD de recaudación bruta por cine en el primer fin de semana de proyección de la película *Harry Potter y la piedra filosofal*.
Aunque se desconoce si la población es normal y se desconoce la varianza, el tamaño de muestra con el que se trabaja es lo suficientemente grande. Por lo tanto, procede el TCL como el cumplimiento de la propiedad de consistencia de la varianza (es decir la varianza muestral tiende a ser muy parecida a la varianza poblacional). Esos dos supuestos nos dan la posibilidad de trabajar con el estadístico z.

Tenemos así $\bar{x}$ que es 25.467 y S de 4.980 que aproximamos a σ.
$\text{LIC} = 25467 - 1{,}96 \,.\, 4980/\sqrt{100} = 24490{,}92$
$\text{LSC} = 25467 + 1{,}96 \,.\, 4980/\sqrt{100} = 26443{,}08$

Actividad 16

$\hat{p} = \frac{11}{120} = 0{,}092$, $1 - \alpha = 0{,}90$, $n = 120$
p = proporción de rechazos en relación a la producción total de una fábrica de radiadores
$\text{LIC} = \hat{p} - z\sqrt{\hat{p}.(1-\hat{p})/n} = 0{,}092 - 1{,}645\sqrt{0{,}092.(1-0{,}0912)/120} = 0{,}0486$
$\text{LSC} = \hat{p} + z\sqrt{\hat{p}.(1-\hat{p})/n} = 0{,}092 + 1{,}645\sqrt{0{,}092.(1-0{,}092)/120} = 0{,}1354$
Con una confianza del $90\ \%$, el intervalo construido contiene valores por debajo del valor establecido. Por ello, no existe evidencia para afirmar que se esté incumpliendo la

meta y por ende no se deben realizar ajustes.

Actividad 17

a) $\hat{p} = \frac{450}{1500} = 0{,}3$

b) $\text{LIC} = \hat{p}\text{- z}\sqrt{\hat{p}.(1-\hat{p})/n} = 0{,}3 - 1{,}96\,.\sqrt{0{,}3.(1-0{,}3)/1500} = 0{,}2768$
$\text{LSC} = \hat{p}\text{+ z}\sqrt{\hat{p}.(1-\hat{p})/n} = 0{,}3 + 1{,}96\,.\sqrt{0{,}3.(1-0{,}3)/1500} = 0{,}3232$

Actividad 18

p = proporción de usuarios/as de Internet de jóvenes entre 12 y 17 años que votan por Google como el sitio de Internet más popular
$n = 1.400;\ 1\text{-}\alpha = 0{,}9973;\ \hat{p} = 0{,}09$
$|e| = z\sqrt{\hat{p}.(1-\hat{p})/n} = 3\sqrt{0{,}09.(1-0{,}09)/1400}$
$e = +/- 0{,}0229$
$\text{LIC} = \hat{p}\text{- } z\sqrt{\hat{p}.(1-\hat{p})/n} = 0{,}09 - 3\sqrt{0{,}09} = 0{,}06705$
$\text{LSC} = \hat{p} + z\sqrt{\hat{p}.(1-\hat{p})/n} = 0{,}09 + 3\sqrt{0{,}09\,.(1-0{,}09)/1400} = 0{,}11295$

Actividad 19

a) Interpretación de los datos y conclusiones
Intervalos de confianza
Estimación paramétrica

Variable	Parámetro	Estimación	E.E.	n	LI (99%)	LS (99%)
Docencia	Proporción (>0)	0,39	0,05	100	0,26	0,52

Estimación puntual de la proporción

$\sqrt{\hat{p}.\frac{(1-\hat{p})}{n}}$ Error estándar

Tamaño de la muestra

Límites inferiores y superiores de los intervalos de confianza

La estimación puntual de la proporción arroja un valor de 0,39. Al trabajar con intervalos de confianza con un nivel de confianza del 99 % encontramos que el LIC es 0,26 y el LSC es de 0,52. Esto implica que 99 veces de cada 100 encontraremos que el intervalo calculado contendrá el verdadero valor del parámetro poblacional.

b) Con un menor nivel de confianza, sí mejoraría la precisión de la estimación.
Recordemos que la precisión aumenta en la medida en que disminuye la longitud.
Longitud = LSC - LIC
Longitud = $\hat{p} + z\sqrt{\hat{p}.(1-\hat{p})/n}$ - $(\hat{p}$ - $z\sqrt{\hat{p}.(1-\hat{p})/n})$

Longitud = $2\ z\sqrt{\hat{p}.(1-\hat{p})/n}$
Si quiero que la longitud sea pequeña:
Al disminuir $1\text{-}\alpha \rightarrow z$ disminuye → Longitud disminuye

Actividad 20

Intervalos de confianza
Estimación paramétrica

Variable	Parámetro	Estimación	E.E.	n	LI (90%)	LS (90%)
Contenido	Varianza	0,02	0,01	27	0,01	0,03

Estimación puntual de la varianza | Error de la varianza | Tamaño de la muestra | Límites inferiores y superiores de los intervalos de confianza

Actividad 21

Datos:
$s^2 = 64{,}7425$; $\bar{X} = 133{,}25$ y $n = 24$
Para la varianza:

$$\text{LIC} = \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}} = \frac{23*64{,}7425}{38{,}0756} = 39{,}11$$

$$\text{LSC} = \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} = \frac{23*64{,}7425}{18{,}5856} = 80{,}11$$

El estimador puntual elegido para representar la variabilidad en los precios de las cotizaciones es s^2, que en este caso es igual a $64{,}72$ pesos al cuadrado. Dado que el estimador puntual puede no asumir el verdadero valor del parámetro, se suele calcular una estimación por intervalos de confianza al sumar y restar al estimador puntual un margen de error. Como se conoce que la distribución es aproximadamente normal (a través del análisis del Q - Q plot donde los puntos están cercanos a la línea de $45°$ y con un $r = 0{,}991$) y si se toma un nivel de confianza del $95\ \%$ podemos trabajar con la estimación por intervalos. El intervalo de confianza para la varianza es $[39{,}11;\ 127{,}39]$.
Para la media:
Como el estimador puntual puede no asumir el verdadero valor del parámetro, calculamos la estimación por intervalos de confianza con un nivel de confianza del $95\ \%$. Como sabemos que la población subyacente es aproximadamente normal y estamos en presencia de una muestra pequeña ($n < 30$), trabajaremos con el estadístico *t*.
$\text{LIC} = \bar{X}$ - $t\ S/\sqrt{n} = 133{,}25$ - 2,06866 x $8{,}05/\sqrt{24} = 129{,}85$
$\text{LSC} = \bar{X} + t\ S/\sqrt{n} = 133{,}25 + 2{,}06866$ x $8{,}05/\sqrt{24} = 134{,}65$

Informe para el inversor:
A partir del comportamiento de las acciones se puede concluir lo siguiente:
La rentabilidad media de la acción oscila entre $\$129{,}85$ y $\$134{,}65$ (el $95\ \%$ de los intervalos contiene a la rentabilidad promedio). Sin embargo, la rentabilidad que brinda el título valor tiene aparejada una variabilidad o riesgo. En este caso, la variabilidad para el período analizado oscila entre $\$39{,}11$ y $\$80{,}11$ (con un nivel de confianza $95\ \%$).
Este activo financiero, con el nivel de confianza establecido, generaría rentabilidad; aunque es de destacar que para tomar la decisión de invertir se debería considerar el costo de oportunidad como el nivel de inflación y otras inversiones financieras (como por ejemplo la tasa de rentabilidad de plazo fijo) para decidir su conveniencia.

Actividad 22

$n = 22;\ \bar{X} = 77{,}33,\ 1\text{-}\alpha = 0{,}98,\ s^2 = 25{,}37$

$$\text{LSC} = \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} = \frac{(22-1)25{,}37}{8{,}897} = 59{,}88$$

$$\text{LIC} = \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}} = \frac{(22-1)25{,}37}{38{,}932} = 13{,}68$$

El intervalo $[13{,}68;\ 59{,}88]$ contendrá la varianza poblacional con una confianza del $98\ \%$.
Para que el cálculo sea válido se requiere que la población subyacente sea normal.

Actividad 23

a) El número de acciones que habrá que tomar en la muestra si está dispuesto a correr un riesgo del $1\ \%$ de cometer un error de estimación de 2 o menos es:
X: razón precio/ganancia (P/G) en las acciones de la Bolsa de Nueva York
Y sabemos que $\sigma = 10$
Error de estimación (e): $|e| \leq 2$

$$n = \frac{z^2.\sigma^2}{e^2} = \frac{2{,}576^2.10^2}{2^2} = 165{,}89$$

$n \geq 166$, la muestra debe ser de al menos 166 acciones

b) Si aumenta el error de estimación a 5 y se mantiene el riesgo:
Error de estimación ≤ 5

$$n = \frac{z^2.\sigma^2}{e^2} = \frac{2{,}576^2.10^2}{5^2} = 26{,}54$$

$n \geq 27$, la muestra debe ser de al menos de 27 acciones

c) Si se aceptara un riesgo del $10\ \%$ y se tolera un error de estimación de 2 o menos.
Error de estimación ≤ 2

$$n = \frac{z^2.\sigma^2}{e^2} = \frac{1{,}645^2.10^2}{2^2} = 67{,}65$$

$n \geq 68$, la muestra debe ser de al menos de 68 acciones.

Se puede observar que, si se está dispuesto a tolerar un error de estimación más alto, el tamaño de muestra se puede reducir. Lo mismo sucede al disminuir el nivel de confianza de la estimación. Es decir, para niveles de confianza más altos y error de estimación más bajo se requiere un mayor tamaño de muestra.

Actividad 24

X = tiempo de transporte hacia el trabajo en las ciudades más grandes de Argentina

$$n = \frac{z^2.\sigma^2}{e^2} = \frac{1{,}96^2.6{,}25^2}{2^2} = 37{,}51$$

$n \geq 38$, el tamaño de la muestra debe ser al menos de 38.

Actividad 25

a) Sabiendo que en un sondeo previo se ha observado un 9 % de hipertensas

Error de estimación es relativo en relación a la estimación de p y es del 2 % (con una confianza del 95 %).

$$n = \frac{z^2 p(1-p)}{e^2} = \frac{1{,}96^2.0{,}09(0{,}91)}{0{,}02^2} = 786{,}57$$

$n \geq 787$

b) Sin ninguna información previa.

$$n = \frac{z^2 p(1-p)}{e^2} = \frac{1{,}96^2.0{,}5(0{,}5)}{0{,}02^2} = 2401$$

$n \geq 2401$

Actividad 26

P = proporción de empresas pymes exportadoras que han incrementado sus ingresos producto de la devaluación del peso argentino

$\hat{p} = 0{,}65$

$|e| = 0{,}03$

a) $1\text{-}\alpha = 0{,}9545$

$Z_{0.97725} = 2$

$$n = \frac{Z^2\ \hat{p}\ (1-\hat{p})}{e^2} = \frac{(2)^2 0{,}65.0{,}35}{(0.03)^2} = 1011{,}11$$

b) $1\text{-}\alpha = 0{,}9973$.

$Z_{0.99865} = 3$

$$n = \frac{Z^2 \hat{p}(1-\hat{p})}{e^2} = \frac{(3)^2 0{,}65.0{,}35}{(0.03)^2} = 2275$$

El tamaño de la muestra aumenta al aumentar el nivel de confianza.

Actividad 27

X_i: número de opiniones a favor del candidato *i* con anterioridad a las elecciones
$\hat{p}$ = proporción de opiniones a favor del candidato *i* con anterioridad a las elecciones

$$n = \frac{Z^2 \hat{p}(1-\hat{p})}{e^2} = \frac{(1{,}96)^2 0{,}50.0{,}5}{(0.10)^2} = 96{,}04$$

$n \geq 97$

Actividad 28

a) El tamaño de muestra adecuado si la investigadora indica que el máximo error muestral no debe ser mayor que $\$500$ por arriba o debajo de la verdadera media del gasto y toma un nivel de confianza de $95{,}45\ \%$ es de 141.
Asimilamos el dato de la encuesta permanente de hogares (S= $\$2.800$) a σ
$Z_{0{,}97725}\ \sigma/\sqrt{n} \leq |e|$

$$n = \frac{z^2.\sigma^2}{e^2} = \frac{2^2.2800^2}{500^2} = 125{,}44$$

$n \geq 126$, el tamaño de la muestra debe ser al menos de 126

b) El valor del riesgo si se toma una muestra de tamaño 500 y el error máximo es el del inciso a, se relaciona con la probabilidad de riesgo que es α
$Z.2800/\sqrt{500} = 500 \rightarrow z = 3.99 \rightarrow 1\text{-}\alpha = 0{,}9999 \rightarrow \alpha = 0{,}00002$

c) El error, si el nivel de confianza es de $0{,}9545$ y el tamaño de la muestra es de 500 personas, es:
$e = 2.2800/\sqrt{500}$
$e \leq 250{,}44$

Actividad 29

$\hat{p}$= proporción de usuarios de la nueva ruta
$1\text{-}\alpha = 0{,}95$
$|e| = 0{,}02$
n_∞= con reposición
n = sin reposición

$$n_\infty = \frac{Z^2 \hat{p}(1-\hat{p})}{e^2} = \frac{1{,}96^2 0{,}5(1-0{,}5)}{0{,}02^2} = 2401$$

$$n = \frac{n_\infty N}{(N-1)+n_\infty} = \frac{2401.8000}{(8000-1)+2401} = 1846{,}92$$

$n \geq 1.847$, el tamaño de la muestra debe ser al menos de 1.847 usuarios/as potenciales
Al trabajar con una población finita el tamaño de muestra puede reducirse

considerando el tamaño de la población en su cálculo. Esto sucede siempre y cuando la muestra sea relativamente grande en comparación con el tamaño poblacional.

Actividad 30

X_i = $ gastados en las cuentas de crédito $\mathbf{i}$
$\sigma_x = \$1500$
$e = +/- 500$
$1\text{-}\alpha = 0{,}95$
$N = 4000$
$n_\infty =$ con reposición
$n =$ sin reposición

$$n_\infty = \frac{Z^2\sigma^2}{e^2} = \frac{1{,}96^2 1500^2}{500^2} = 34{,}57$$

$$n = \frac{n_\infty N}{(N-1)+n_\infty} = \frac{34{,}57.4000}{(4000-1)+34{,}57} = 34{,}28$$

$n \geq 35$
Los valores de n y n_∞ son prácticamente iguales porque el tamaño de muestra es relativamente pequeño en relación al tamaño de la población. Como si se tratara de una población infinita ya que las probabilidades de selección no se ven prácticamente afectadas.

Actividad 31

$\hat{p}$ = proporción de usuarios que estarían dispuesto a comprar un pack especial de programación.
$\hat{p} = 0{,}30$
$1\text{-}\alpha = 0{,}95$
$|e| = 0{,}05$
n_∞ = con reposición
n = sin reposición

$$n_\infty = \frac{Z^2\hat{p}(1-\hat{p})}{e^2} = \frac{1{,}96^2 0{,}3(1-0{,}3)}{0{,}05^2} = 322{,}69 \approx 323$$

$$n = \frac{n_\infty N}{(N-1)+n_\infty} = \frac{323\ .5000}{(5000-1)+323} = 303{,}45$$

$n \geq 304$, el tamaño de muestra es de 304, como mínimo, si el muestreo se realiza sin reposición.

Actividad 32

$\sigma^2 = 9{,}07$ y $e = +/- 3$

$N = 1000$
$1\text{-}\alpha = 0{,}95 \rightarrow z = 1{,}96$

$$n_\infty = \frac{Z^2\sigma^2}{e^2} = \frac{1{,}96^2 9{,}07}{3^2} = 35{,}11$$

$$n = \frac{n_\infty N}{(N-1)+n_\infty} = \frac{35{,}11.1000}{(1000-1)+35{,}11} = 33{,}95$$

$n \geq 34$
La diferencia entre el muestreo C/R y el S/R no es significativo ya que $n < 5\%\ N$ (la muestra es una fracción pequeña de la población).

Actividad 33

X = peso neto en gramos de un sobre de café
$n = 12$, $1\text{-}\alpha = 0{,}98$
$\overline{X} = 15{,}97$, $S = 0{,}1497$
$LIC = 15{,}97 - 2{,}718 \,.\, 0{,}1497/\sqrt{12} = 15{,}85$
$LSC = 15{,}97 + 2{,}718 \,.\, 0{,}1497/\sqrt{12} = 16{,}09$

Actividad 34

$\hat{p}$ = proporción de alumnos/as del 5º año del colegio secundario que saben qué carrera seguir
$\hat{p} = 5/15 = 0{,}33$
$n = 15$
$1\text{-}\alpha = 0{,}99 \rightarrow z = +/- 2{,}576$
$LIC = \hat{p} - z\sqrt{\hat{p}.(1-\hat{p})/n} = 0{,}33 - 2{,}576.\sqrt{0{,}33.(1-0{,}33)/15} = 0{,}017$
$LSC = \hat{p} + z\sqrt{\hat{p}.(1-\hat{p})/n} = 0{,}33 + 2{,}576.\sqrt{0{,}33.(1-0{,}33)/15} = 0{,}643$

Actividad 35

a) $e = 0{,}02\ (0{,}60) = 0{,}012 \quad p = 0{,}60$

$$n = \frac{z^2.p(1-p)}{e^2} = \frac{1{,}96^2.0{,}60.(1-0{,}60)}{0{,}012^2} = 6402{,}67$$

$n \geq 6403$

b) Si no se conoce nada acerca de la proporción de gente que gusta de este tipo de programas, se considera $p = 0{,}50$

$$n = \frac{z^2.p(1-p)}{e^2} = \frac{1{,}96^2.0{,}50.(1-0{,}50)}{0{,}012^2} = 6669{,}44$$

$n \geq 6670$

Actividad 36

a) Falso, un nivel de confianza del 99 % implica que, de 100 intervalos, 99 contienen el verdadero valor del parámetro poblacional

b) Falso. Si el tamaño de la muestra y la varianza poblacional permanecen constantes y se disminuye el nivel de confianza del intervalo, entonces su error disminuye. Por lo tanto, disminuye la amplitud del intervalo.

c) Verdadero. Aumenta el error de estimación, por lo tanto, la amplitud del intervalo es mayor, lo que equivale a una menor precisión.

d) Verdadero. Mientras más observaciones se toman, menor error de estimación, por lo tanto, menor amplitud tendrá el intervalo que se construya.

e) Falso. Se puede usar solo para construir intervalos de confianza para la media y la proporción cuando se desconoce la distribución de la variable en la población y se aplica el Teorema Central del Límite.

f) Falso. Si disminuye el nivel de confianza, disminuye el error de estimación.

Actividad 37

a) X = horas de duración de las pilas
$n = 81$
$1 - \alpha = 0{,}90$
$\sigma = 120$
$\overline{X} = 1203{,}8$
Como X se distribuye aproximadamente normal y se conoce σ, el estadístico a utilizar es z
$\text{LIC} = \overline{X} - z\,\sigma/\sqrt{n}$
$\text{LIC} = 1203{,}38 - 1{,}645 \text{ x } 120/\sqrt{81} = 1181{,}47$
$\text{LSC} = \overline{X} + z\,\sigma/\sqrt{n}$
$\text{LSC} = 1203{,}38 + 1{,}645 \text{ x } 120/\sqrt{81} = 1225{,}31$
El intervalo comprendido entre $1181{,}47$ y $1225{,}31$ contiene la duración promedio de las pilas, con un 90 % de confianza.

b) Si el error máximo tolerado fuera de $47{,}04$ horas y $N = 345$
$\sigma_x = 120$
$e = 47{,}04$
$1\text{-}\alpha = 0{,}90$
$N = 345$
n_∞ = con reposición
n = sin reposición

$$n_\infty = \frac{Z^2\sigma^2}{e^2} = \frac{1{,}65^2 120^2}{47{,}04^2} = 17{,}5$$

$$n = \frac{n_\infty N}{(N-1)+n_\infty} = \frac{17{,}5.345}{(345-1)+17{,}5} = 16{,}7$$

$n \geq 17$

Actividad 38

a) La población se distribuye aproximadamente normal, con una muestra de tamaño 61 y σ desconocida. Se trabajará con el estadístico t para construir el intervalo con un nivel de confianza de 0,90.

X = exportaciones en miles de $

$n = 61$

$\overline{X} = 98{,}52$

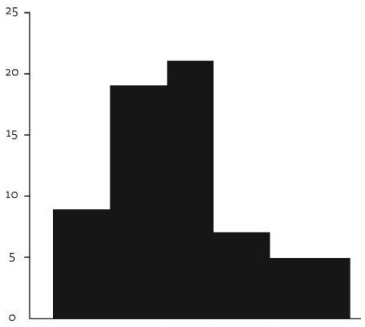
$S = 48{,}50$

$1\text{-}\alpha = 0{,}90$

$\text{LIC} = 98{,}52 - 1{,}671.\ 48{,}5/\sqrt{61} = 88{,}14$

$\text{LSC} = 98{,}52 + 1{,}671.\ 48{,}5/\sqrt{61} = 108{,}89$

El intervalo comprendido entre $88.140; $108.890 contiene el verdadero valor del parámetro con un 90 % de confianza.



Exportaciones en miles de pesos

b) Para analizar el porcentaje de empresas que no han alcanzado más de 90.000 pesos de exportación anual.

Intervalo de confianza para la proporción verificamos que $n.\hat{p}>5$ y $n.\hat{q}>5$

$1\text{-}\alpha = 0{,}90 \rightarrow z = +/- 1{,}645$

$\hat{p} = 28/61 = 0{,}33$ empresas que han alcanzado hasta 90.000 pesos

$\text{LIC} = 0{,}459 - 1{,}645.\ \sqrt{0{,}459.\ (1-0{,}459)/61} = 0{,}354$

$\text{LSC} = 0{,}459 + 1{,}645.\ \sqrt{0{,}459.\ (1-0{,}459)/61} = 0{,}564$

Entre el 35,4% y el 56,4% de las empresas no han alcanzado más de $90.000 de exportaciones, con una confianza del 90%.