CAPÍTULO 2

INFERENCIA ESTADÍSTICA: ESTIMACIÓN

2.1. Introducción.

En muchas investigaciones se está interesado en estudiar una o más poblaciones, las cuales pueden ser caracterizadas por algunos parámetros, es por ello que en múltiples estudios estadísticos se centre la atención sobre dichos parámetros. Por ejemplo, supongamos que se desea conocer el ingreso promedio de los habitantes del Municipio Libertador del Estado Mérida, en dicho caso el parámetro es la media poblacional μ. Obtener el valor del parámetro en general es difícil, porque para ello sería necesario tener toda la información de la población, por ejemplo, el valor de μ puede ser calculado si contamos con el salario de todos los habitantes de la región en estudio, pero evidentemente eso no es posible, bien sea porque no disponemos del tiempo o del dinero necesario para recoger la información.

En tales situaciones se recomienda seleccionar una muestra aleatoria de dicha población y a partir de esos datos calcular el símil de la muestra en la población, conocido como

estadístico, el cual es nos da información sobre el valor del parámetro. En nuestro ejemplo, seleccionamos una muestra aleatoria de trabajadores de la región en estudio, a quienes se les tomaría el sueldo mensual, y a partir de dichos datos se calcula la media muestral $\bar{X}$, el cuál como veremos es el mejor estadístico para estimar la media poblacional μ. Este procedimiento se conoce como **Inferencia Estadística**.

Según Casas(), el objetivo básico de la inferencia estadística es hacer inferencias o sacar conclusiones sobre la población a partir de la información contenida en una muestra aleatoria de la población. Más específicamente, podemos decir que la inferencia estadística consiste en el proceso de selección y utilización de un estadístico muestral, mediante el cual, utilizando la información que nos proporciona una muestra aleatoria, nos permite sacar conclusiones sobre características poblacionales. Es decir, supongase que se tiene una población, la cual se representa por su función de distribución y el parámetro poblacional se denota por θ, que toma valores dentro del espacio paramétrico Θ, el parámetro puede ser cualquiera, por ejemplo, la media μ, la varianza σ^2 , o la proporción poblacional π. Seleccionamos una función de las variables aleatorias muestrales $X_1, X_2, ..., X_n$, que la denotaremos por $\hat{\theta} = g(X_1, X_2, ..., X_n)$ y la utilizaremos para obtener la inferencia sobre el valor del parámetro θ.

Las inferencias sobre el valor de un parámetro poblacional θ se pueden obtener básicamente de dos maneras: a partir de **estimación** o bien a partir de la **prueba de hipótesis**.

- En la **estimación**, basta seleccionar un estadístico muestral cuyo valor se utilizará como estimador del valor del parámetro poblacional.

- En la **prueba de hipótesis**, se hace una hipótesis sobre el valor del parámetro θ y se utiliza la información proporcionada por la muestra para decidir si la hipótesis se

acepta o no.

Ambos métodos de inferencia estadística utilizan las mismas relaciones teóricas entre resultados muestrales y valores poblacionales. Así pues, una muestra es sacada de la población y un estadístico muestral es utilizado para hacer inferencias sobre el parámetro poblacional. En estimación, la información muestral es utilizada para estimar el valor del parámetro θ. En la prueba de hipótesis, primero se formula la hipótesis sobre el valor de θ y la información muestral se utiliza para decidir si la hipótesis formulada debería ser o no rechazada.

En este capítulo nos ocuparemos de la estimación estadística y dejaremos para el capítulo siguiente la prueba de hipótesis.

2.2. Estimación

La estimación estadística se divide en dos grandes grupos: la **estimación puntual** y la **estimación por intervalos**.

- La **estimación puntual** consiste en obtener un único número, calculado a partir de las observaciones muestrales, que es utilizado como estimación del valor del parámetro θ. Se le llama estimación puntual porque a ese número, que se utiliza como estimación del parámetro θ, se le puede asignar un punto sobre la recta real.

- En la **estimación por intervalos** se obtienen dos puntos (un extremo inferior y un extremo superior) que definen un intervalo sobre la recta real, el cual contendrá con cierta seguridad el valor del parámetro θ.

Por ejemplo, si el parámetro poblacional es el salario promedio de los habitantes del Municipio Libertador del Estado Mérida, basándonos en la información proporcionada por

una muestra podríamos obtener una estimación puntual del parámetro μ, que lo denotaremos por $\hat{\mu}$;$\hat{\mu} = 1250$ BsF, sin embargo, el intervalo de estimación para μ sería de la forma $(1200, 1300)$, es decir, de 1200 BsF a 1300 BsF, con un cierto margen de seguridad.

2.2.1. Estimación Puntual

Consideremos una población con función de distribución es $F(x;\theta)$, donde θ es el parámetro poblacional desconocido que toma valores en el espacio paramétrico Θ. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria extraída de dicha población. El **estimador puntual** o simplemente **estimador** del parámetro poblacional θ es una función de las variables aleatorias u observaciones muestrales y se representa por $\hat{\theta} = g(X_1, X_2, ..., X_n)$.

Para una realización particular de la muestra $x_1, x_2, ..., x_n$ se obtiene un valor específico del estimador que recibe el nombre de estimación del parámetro poblacional θ y lo denotaremos por $\hat{\theta} = g(x_1, x_2, ..., x_n)$

Vemos pues que existe diferencia entre estimador y estimación. Utilizaremos el termino estimador cuando nos referimos a la función de las variables aleatorias muestrales $X_1, X_2, ..., X_n$, y los valores que toma la función estimador para las diferentes realizaciones o muestras concretas serán las estimaciones.

El estimador es un estadístico y, por tanto, una variable aleatoria y el valor de esta variable aleatoria para una muestra concreta $x_1, x_2, ..., x_n$ será la estimación puntual. Además como vimos antes, por ser el estimador un estadístico este tiene una distribución de probabilidad que es la distribución muestral del estadístico.

Para clarificar la diferencia entre estimador y estimación consideremos el siguiente ejemplo: supongamos que pretendemos estimar la renta media μ de todas las familias de

una ciudad, para ello parece lógico utilizar como estimador de la media poblacional μ la media muestral $\bar{X}$ siendo necesario seleccionar una muestra aleatoria que supondremos de tamaño n = 80, a partir de la cual obtendríamos la renta media de la muestra, por ejemplo, $\bar{x} = 1500$ BsF. Entonces el estimador de la media poblacional μ será, $\hat{\mu} = \bar{X}$, es decir, el estadístico media muestral $\bar{X}$ y la estimación puntual será $\hat{\mu} = \bar{x} = 1500$ BsF. Observemos que designamos por $\bar{X}$ la variable aleatoria media muestral de las variables aleatorias muestrales $X_1, X_2, , .., X_n$, y por $\bar{x}$ designamos una realización para una muestra específica $x_1, x_2, , .., x_n$, que nos da la correspondiente estimación puntual del parámetro μ, es decir, $\hat{\mu} = \bar{x}$.

Un problema que se consigue un estadístico es que pueden existir varios estimadores para un parámetro, lo que trae como consecuencia que el estadístico tenga que seleccionar entre ellos el mejor. Una manera de hacer esta elección es basandose en las propiedades deseables que un buen estimador debería tener. Veamos a continuación brevemente algunas propiedades que un buen estimados debe poseer.

Propiedades de un Estimador Puntual

1. **Insesgado.** El estadístico $\hat{\theta} = g(X_1, ..., X_n)$ es un estimador insesgado del parámetro θ, si la esperanza matemática del estimador $\hat{\theta}$ es igual al parámetro θ, esto es:

$$E(\hat{\theta}) = \theta \tag{2.2.1}$$

   para todos los valores de θ.

   Es fácil ver que la media muestral $\bar{X}$ es un estimador insesgado de μ, pues $E(\bar{X}) = \mu$. Se deja como ejercicio probar que la varianza muestral dada como $S^{*2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$ no

es insesgados y que la varianza muestral dada como $S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$ si es insesgado.

2. **Eficiente.** En algunas situaciones podemos conseguirnos el caso en que dos estimadores que tenemos a disposición sean insesgados. En ese caso debemos recurrir a otra propiedad que permita diferenciar a dichos estimadores. Una opción seria medir sus eficiencias. Un estimador $\hat{\theta}_1$ es más eficiente que otro estimador $\hat{\theta}_2$ si la varianza del primero es menor que la varianza del segundo. Este criterio parece ser un concepto intuitivamente claro. Evidentemente cuanto más pequeña es la varianza de un estimador, más concentrada está la distribución del estimador alrededor del parámetro que se estima y, por lo tanto, es mejor.

   La mejor ilustración de la eficiencia es los estimadores es la estimación de μ por la media y la mediana muestrales. Si la población está distribuida simétricamente, entonces tanto la media muestral como la mediana muestral son estimadores insesgados de μ. Sin embargo podemos decir que la media muestral es mejor que la media muestral como un estimador de μ, ya que $V(\bar{x}) = \frac{\sigma^2}{n}$ y $V(Med) = 1{,}57076\frac{\sigma^2}{n}$, es decir, la media muestral es más eficiente que la mediana pues $V(\bar{x}) < V(Med)$. Así, concluimos que la media muestral es mejor estimador que la mediana muestral como un estimador de μ.

3. **Consistente.** Hasta ahora hemos considerado propiedades de los estimadores puntuales basados en muestras aleatorias de tamaño n, pero parece lógico esperar que un estimador será tanto mejor cuanto mayor sea el tamaño de la muestra. Así pues cuando el tamaño de la muestra aumenta y por tanto la información que nos proporciona esa muestra es más completa, resulta que la varianza del estimador suele ser menor y la distribución muestral de ese estimador tenderá a encontrarse más concentrada alrededor

del parámetro que pretendemos estimar. Por lo tanto diremos que un estimador insesgado es consistente si su varianza tiende a disminuir a medida que el tamaño de la muestra aumenta. Es decir:

$$V(\hat{\theta}) \to 0 \text{ cuando } n \to \infty \tag{2.2.2}$$

Es fácil ver que $\bar{X}$ es un estimador consistente, pues $V(\bar{X}) = \frac{\sigma^2}{n}$ lo cual tiende a cero cuando n es muy grande.

4. **Suficiente.** Una expresión matemática de esta última propiedad deseable, es bastante complicada. Por fortuna, encontramos que este concepto implica un significado intuitivo preciso. Se dice que un estimador es suficiente si toda la información que contiene la muestra sobre el parámetro esta contenida en el estimador. El significado de la suficiencia reside en el hecho de que si existe un estimador suficiente, es absolutamente innecesario considerar cualquier otro estimador. Puede mencionarse ahora que $\bar{X}, p, S^2, \Delta\bar{X}$ y Δp son estimadores suficientes de los parámetros $\mu, \pi, \sigma^2, \Delta\mu$ y $\Delta\pi$.

Estimadores de Parámetros usados en este curso

En la siguiente tabla se muestran los mejores estimadores de los parámetros más usuales. Dichos estimadores son insesgados, consistentes, eficientes y suficientes. Además se muestra su valor esperado y la varianza.

Parámetro (θ)	**Estimador** $(\hat{\theta})$	$E(\hat{\theta})$	$V(\hat{\theta})$
μ	$\bar{X}$	μ	$\frac{\sigma^2}{n}$
π	p	π	$\frac{\pi(1-\pi)}{n}$
σ^2	S^2	σ^2	-
$\Delta\mu$	$\Delta\bar{X}$	$\Delta\mu$	$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
$\Delta\pi$	Δp	$\Delta\pi$	$\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}$

2.3. Estimación por Intervalo

En la sección anterior, nos hemos ocupado de definir los estimadores puntuales y als propiedades que estos deben poseer. Veíamos que los estimadores eran funciones de las observaciones muestrales, y cuando se calcula el valor del estimador $\hat{\theta}$;para una muestra concreta entonces se tiene la estimación puntual; valor que generalmente difiere del verdadero valor del parámetro θ y, en consecuencia, no nos proporciona suficiente información sobre el parámetro, siendo entonces deseable el acompañar a la estimación del parámetro θ, de alguna medida decl posible error asociado a esta estimación. Es decir, asociado a cada estimación del parámetro daremos un intervalo:

$$[\hat{\theta}_1(X_1, ..., X_n); \hat{\theta}_2(X_1, ..., X_n)]$$

y una medida que nos refleje la confianza que tenemos acerca de que el verdadero valor del parámetro θ se encuentre dentro del intervalo.

Observemos que los extremos del intervalo variarán de manera aleatoria de una muestra a otra, pues dependen de las observaciones de la muestra, luego tanto los extremos del intervalo como la longitud del intervalo serán cantidades aleatorias y, por tanto, no podremos saber con seguridad si el valor del parámetro θ se encuentre dentro del intervalo obtenido cuando

se selecciona una sola muestra. El objetivo que se pretende con los intervalos de confianza es obtener un intervalo de poca amplitud y con una alta probabilidad de que el parámetro θ se encuentra en su interior. Así pues, elegiremos probabilidades cercanas a la unidad, que se representan por $1-\alpha$ y cuyos valores más frecuentes suelen ser 0,90, 0,95 y 0,99.

Luego si deseamos obtener una estimación por intervalo del parámetro poblacional θ desconocido, tendremos que obtener dos estadísticos $\hat{\theta}_1(X_1, ..., X_n)$ y $\hat{\theta}_2(X_1, ..., X_n)$ que nos darán los valores extremos del intervalo, tales que

$$P[\hat{\theta}_1(X_1, ..., X_n) \leq \theta \leq \hat{\theta}_2(X_1, ..., X_n)] = 1 - \alpha \tag{2.3.1}$$

Al valor $1-\alpha$ se le conoce como coeficiente de confianza y al valor $100(1-\alpha)\,\%$ se le llama nivel de confianza.

Observando el intervalo dado en la expresión 2.3.1 se pone de manifiesto:

1. Que se trata de un intervalo aleatorio, pues los extremos dependen de la muestra seleccionada y, por tanto, $\hat{\theta}_1$ y $\hat{\theta}_2$ son variables aleatorias.

2. Que el parámetro θ es desconocido.

3. En consecuencia y antes de seleccionar una muestra no podemos decir que la probabilidad de que el parámetro θ tome algún valor en el intervalo $(\hat{\theta}_1, \hat{\theta}_2)$ es igual a $1-\alpha$, afirmación que no sería correcta después de seleccionar la muestra.

Para una muestra concreta se tendrían unos valores:

$$\hat{\theta}_1(x_,..., x_n) = a \qquad \text{y} \qquad \hat{\theta}_2(x_,..., x_n) = b$$

y no podemos afirmar que

$$P[a \leq \theta \leq b] = 1 - \alpha$$

ya que no tiene sentido alguno, pues a, b y θ son tres valores constantes. Sin embargo, una vez seleccionada la muestra y calculados, los valores de a y b si tiene sentido decir que

- La probabilidad es 1 si $\theta \in [a, b]$
- La probabilidad es 0 si $\theta \notin [a, b]$

Luego, no podemos referirnos a la probabilidad del intervalo numérico sino que nos referiremos al coeficiente de confianza del intervalo, y en consecuencia al nivel de confianza del intervalo, pues la probabilidad ya hemos indicado que, después de extraída la muestra, será 1 o cero.

Para precisar más sobre la interpretación del intervalo de confianza, consideramos un número grande de muestras del mismo tamaño y calculamos los límites inferior y superior para cada muestra, es decir a y b, entonces se obtendrá que aproximadamente en el $100(1-\alpha)\,\%$ de los intervalos resultantes estará en su interior el valor del parámetro θ, y en el $100\alpha\,\%$ restante no estará en su interés el valor del parámetro θ, y en consecuencia al intervalo (a, b) se le llama intervalo de confianza al nivel de confianza del $100(1-\alpha)\,\%$. Es decir, si tomamos 100 muestras aleatorias de tamaño n de la misma población y calculamos los límites de confianza 6 y 8 para cada muestra, entonces esperamos que aproximadamente el 95 % de los intervalos contendrán en su interior el verdadero valor del parámetro p, y el 5 % restante no lo contendrán. Pero como nosotros, en la práctica, sólo tomamos una muestra aleatoria y, por tanto, sólo tendremos un intervalo de confianza, no conocemos si nuestro intervalo es uno del 95 % o uno del 5 %, y por eso hablamos de que tenemos un nivel de confianza del 95 %.

La precisión de la estimación por intervalos vendrá caracterizada por el coeficiente de

confianza $1-\alpha$ y por la amplitud del intervalo. Así pues, para un coeficiente de confianza fijo, cuanto más pequeños sea el intervalo de confianza más precisa será la estimación, o bien para una misma amplitud del intervalo, cuanto mayor sea el coeficiente de confianza mayor será la precisión.

2.3.1. Métodos de construcción de intervalos de confianza

Básicamente existen dos métodos para la obtención de intervalos de confianza de parámetros. El primero, el método pivotal o método del pivote basado en la posibilidad de obtener una función del parámetro desconocido y cuya distribución muestral no dependa del parámetro. El segundo, el método general de Neyman, está basado en la distribución de un estimador puntual del parámetro. En este curso solo construiremos intervalos de confianza con el método de la cantidad pivotal.

Método de la cantidad pivotal

Antes de ver en que consiste el método tenemos que definir cantidad pivotal.

Definición 2.3.1 (Cantidad Pivotal) *Una cantidad pivotal o pivote, es una función de las observaciones muestrales y del parámetro θ, $T(X_1, ..., X_n; \theta)$, cuya distribución muestral no depende del parámetro θ.*

A continuación se presentan algunos ejemplos de cantidad pivotal.

1. $Z = \frac{barX-\mu}{\sigma_{barX}}$ es una cantidad pivotal ya que depende de la muestra a través de $barX$ y del parámetro μ, cuya distribución es la normal estándar, la cual no depende del valor de μ.

2. $W = \frac{(n-1)S^2}{\sigma^2}$ es una cantidad pivotal ya que depende de la muestra a través de S^2 y de σ^2, cuya distribución es la chi-cuadrado, la cual no depende del valor de σ^2.

3. $T = \frac{\Delta\bar{X}-\Delta\mu}{\sigma_{\Delta\bar{X}}}$ es una cantidad pivotal ya que depende de la muestra a través de $\Delta\bar{X}$ y del parámetro $\Delta\mu$, cuya distribución es la t-student, la cual no depende del valor de $\Delta\mu$.

Ahora que sabemos que es una cantidad pivotal, vemos en que consiste el método de la cantidad pivotal.

1. Definir una cantidad pivotal

2. Como la distribución de la cantidad pivotal es conocida, dada un nivel de confianza, se hallan los valores de a y b tales que $P(a \leq T(X_1, ..., X_n; \theta) \leq b) = 1 - \alpha$

3. Como $T(X_1, ..., X_n; \theta)$ es una función del parámetro, se despeja de la desigualdad dicho valor, con lo cuál se obtiene el intervalo de confianza del parámetro deseado.

2.3.2. Intervalos de confianza en poblaciones normales

En esta sección consideramos que la población será normal y obtendremos intervalos de confianza para los parámetros poblaciones en el caso de una muestra y de dos muestras. Aplicaremos el método pivotal, pues en estos casos no existe gran dificultad para obtener una función del parámetro desconocido cuya distribución muestral no dependa del parámetro.

1. **Intervalo de confianza para la media de una población normal**

   Sea $x_1, x_2, ..., x_n$ una muestra aleatoria extraída de una población $N(\mu, \sigma^2)$, con μ desconocido y σ^2 puede ser o no conocida. Estamos interesados en hallar un intervalo

de confianza para μ al nivel de confianza $1-\alpha$. Como σ^2 puede ser o no conocida, veamos cada caso por separado.

a) σ^2 **es conocida.** En principio debemos encontrar un estadístico (cantidad pivotal o pivote) que dependa del parámetro μ y de su estimador y cuya distribución muestral no dependa del parámetro μ. En este caso el estadístico será:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

que se distribuye según una $N(0,1)$.

Ahora, utilizando la tabla de la distribución $N(0,l)$, podemos encontrar dos valores $Z_{\alpha/2}$ y $Z_{1-\alpha/2}$, (la selección de estos dos valores garantiza que la amplitud del intervalo sea mínima) tales que:

$$P(Z_{\alpha/2} \leq Z \leq Z_{1-\alpha/2}) = 1-\alpha \tag{2.3.2}$$

de donde se tiene que

$$P\left(Z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq Z_{1-\alpha/2}\right) = 1-\alpha$$

multiplicando por $\sigma/\sqrt{n}$

$$P\left(Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

restando $\bar{X}$

$$P\left(-\bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Multiplicando por -1

$$P\left(\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{X} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

que es equivalente a

$$P\left(\bar{X} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

como $Z_{\alpha/2} = -Z_{1-\alpha/2}$ se tiene

$$P\left(\bar{X} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Por lo tanto, el intervalo de confianza para la media μ de una población $N(\mu, \sigma^2)$ con σ^2 conocida es:

$$\left[\bar{x} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{x} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \tag{2.3.3}$$

Ejemplo 2.3.1 *De una población $N(\mu, 9)$ se selecciona una muestra aleatoria cuya media es 25. Obtener un intervalo de confianza para la media poblacional μ. Cuando el tamaño de la muestra es $n = 16$ y el nivel de confianza es del 95 %. El intervalo de confianza se obtiene al usar la ecuación 2.3.3, donde $\bar{x} = 25$, $n = 16$*

y $1-\alpha = 0{,}95$*, de este ultimo dato se tiene que:*

$$Z_{1-\alpha/2} = Z_{0{,}975} = 1{,}96$$

Por lo tanto, el intervalo de confianza es

$$\left[25 - 1{,}96\frac{3}{\sqrt{16}}; 25 - 1{,}96\frac{3}{\sqrt{16}}\right]$$

$$[23{,}53; 26{,}47]$$

b) σ^2 **es desconocida.** Cuando la varianza poblaciones es desconocida debemos tomar en cuenta el tamaño de la muestra. Se el tamaño de la muestra es mayor o igual que 30 seguimos usando el intervalo de confianza de la ecuación 2.3.3. Si el tamaño de la muestra es menor que 30, usamos el siguiente estadístico como cantidad pivotal

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

que se distribuye según una t-student con $n-1$ grados de libertad.

Ahora, utilizando la tabla de la distribución t-student, podemos encontrar dos valores $t_{\alpha/2}$ y $t_{1-\alpha/2}$, (la selección de estos dos valores garantiza que la amplitud del intervalo sea mínima) tales que:

$$P(t_{\alpha/2} \leq T \leq t_{1-\alpha/2}) = 1 - \alpha \tag{2.3.4}$$

Procediendo de igual manera al caso anterior se tiene que el intervalo de confianza

con un nivel de confianza $1-\alpha$ para μ con σ^2 desconocido es

$$\left[\bar{x} - t_{1-\alpha/2}\frac{S}{\sqrt{n}}; \bar{x} + t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right] \tag{2.3.5}$$

Ejemplo 2.3.2 *Un fabricante de una determinada marca de vehículos de lujo sabe que el consumo de gasolina de sus vehículos se distribuye normalmente. Se selecciona una muestra aleatoria de 6 carros y se observa el consumo cada 100 km, obteniendo las siguientes observaciones Obtener el intervalo de confianza para el consumo medio de gasolina de todos los vehículos de esa marca, a un nivel de confianza del 90 %.*

Con los datos de la muestra obtenemos la media y la varianza muestral, los cuales son $\bar{x} = 19{,}48$ *y* $S^2 = 1{,}12$*. El intervalo de confianza para la media poblacional cuando* σ^2 *es desconocida tiene la forma dada por la expresión 2.3.5, donde* $\bar{x} = 19{,}48$*,* $S^2 = 1{,}06$*,* $n = 6$ *y* $1-\alpha = 0{,}90$*, de este ultimo dato se tiene que:*

$$T_{1-\alpha/2} = T_{0{,}95} = 2{,}015$$

Por lo tanto, el intervalo de confianza es

$$\left[19{,}48 - 2{,}015\frac{1{,}06}{\sqrt{6}}; 19{,}48 + 2{,}015\frac{1{,}06}{\sqrt{6}}\right]$$

$$[18{,}61; 20{,}35]$$

2. **Intervalo de confianza para la varianza de una población normal**

Cuando se realizan inferencia sobre la varianza de una población normal se debe tomar

en consideración si la media poblacional es o no conocida.

a) μ **es desconocida** Supongamos una población $N(\mu, \sigma^2)$, en donde μ y σ^2 son desconocidos y deseamos obtener un intervalo de confianza para la varianza poblacional σ^2 al nivel de confianza del $100(1-\alpha)\,\%$. Para ello tomamos una muestra aleatoria de tamaño n, (X,, ..., X,) y utilizaremos un estadístico (cantidad pivotal o pivote) que dependa del parámetro σ^2 y de su estimador y cuya distribución muestral no dependa de los parámetros desconocidos. Ese estadístico será:

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

el cual se distribuye según una chi-cuadrado con $n-1$ grados de libertad, χ^2_{n-1}, siendo S^2 la varianza muestral.

Ahora, utilizando la tabla de la distribución chi-cuadrado, podemos encontrar dos valores $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$, (la selección de estos dos valores garantiza que la amplitud del intervalo sea mínima) tales que:

$$P(\chi^2_{n-1,\alpha/2} \leq W \leq \chi^2_{n-1,1-\alpha/2}) = 1-\alpha \tag{2.3.6}$$

de donde se tiene que

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

dividiendo por $(n-1)S^2$

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

Reordenando esta expresión se tiene

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

y el intervalo de confianza para σ^2 al nivel de confianza del $(1-\alpha)\,\%$ sería:

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}; \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right] \tag{2.3.7}$$

b) μ **es conocida** En este caso tal estadístico (cantidad pivotal o pivote) que dependa del parámetro σ^2 y cuya distribución muestral no dependa de σ^2 será:

$$W* = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$$

el cual se distribuye según una chi-cuadrado con n grados de libertad, χ^2_n, pues al ser la media μ conocida no hay que estimarla y el número de grados de libertad es n.

Razonando análogamente al caso anterior, en donde μ era desconocida, llegamos

a obtener el intervalo de confianza:

$$\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n,1-\alpha/2}};\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n,\alpha/2}}\right] \tag{2.3.8}$$

Ejemplo 2.3.3 *El precio de un determinado artículo perecedero en los comercios de alimentación de una ciudad sigue una distribución normal. Se toma una muestra aleatoria de 8 comercios y se observa el precio de ese artículo, obteniendo las siguientes observaciones:*

$$135, 125, 130, 139, 126, 138, 124, 140$$

Obtener al nivel de confianza del 95 %.

a) Un intervalo de confianza para la media poblacional.

b) Un intervalo de confianza para la varianza poblacional.

A partir de las observaciones muestrales obtenemos que $\bar{x} = 131{,}75$ *y* $S^2 = 43{,}07$

a) El intervalo de confianza para la media poblacional cuando σ^2 *es desconocido y* $1-\alpha = 0{,}95$ *viene dado por:*

$$\left[131{,}75 - 2{,}365\frac{6{,}56}{\sqrt{8}}; 131{,}75 + 2{,}365\frac{6{,}56}{\sqrt{8}}\right]$$

$$[126{,}25; 137{,}23]$$

b) El intervalo de confianza para la varianza poblacional cuando μ *es desconocido*

y $1-\alpha = 0{,}95$ *viene dado por:*

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

$$\left[\frac{(8-1)43{,}07}{\chi^2_{7,0,975}}; \frac{(8-1)43{,}07}{\chi^2_{7,0,025}}\right]$$

donde $\chi^2_{7,0,975} = 16{,}015$ *y* $\chi^2_{7,0,025} = 1{,}690$*, por lo tanto el intervalo de confianza es*

$$\left[\frac{(7)43{,}07}{16{,}015}; \frac{(7)43{,}07}{1{,}690}\right]$$

$$[18{,}83; 178{,}39]$$

3. **Intervalo de confianza para la diferencia de medias en poblaciones normales: Muestras independientes**

   Sean $X_{11}, X_{12}, ..., X_{1n_1}$ y $X_{21}, X_{22}, ..., X_{2n_2}$ dos muestra aleatorias independientes extraídas de poblaciones normales, $N(\mu_1, \sigma_1^2)$ y $N(\mu_1, \sigma_1^2)$, respectivamente. Estamos interesados en hallar un intervalo de confianza del $100(1-\alpha)\,\%$ para la diferencia de medias entre las dos poblaciones, $\Delta\mu$. Para hallar dicho intervalo de confianza debemos considerar si las varianzas poblacionales son o no conocidas.

   a) **Varianzas conocidas** En este caso el estadístico (cantidad pivotal o pivote) que depende del parámetro $\Delta\mu$ y de su estimador $\Delta\bar{X}$ y cuya distribución muestral no depende del parámetro es:

$$Z = \frac{\Delta\bar{X} - \Delta\mu}{\sigma_{\Delta\bar{X}}}$$

que se distribuye según una $N(0,1)$, donde $\sigma_{\Delta\bar{X}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Procediendo de manera análoga al caso de una población, se tiene que el intervalo de confianza es

$$\left[\Delta\bar{X} - Z_{1-\alpha/2}\sigma_{\Delta\bar{X}}; \Delta\bar{X} + Z_{1-\alpha/2}\sigma_{\Delta\bar{X}}\right] \tag{2.3.9}$$

b) **Varianzas desconocidas** Cuando las varianzas son desconocidas debemos tomar en cuenta los tamaños de las muestras. Si los tamaños de muestras son mayores que 30, el intervalo de confianza es el de la ecuación 2.3.9. Por el contario si los tamaños de las muestras son menores que 30, debemos estudiar por separado el supuesto de que las varianzas sean iguales o diferentes.

1) **Suponiendo varianzas iguales.** Teniendo en cuenta los resultados obtenidos en el capítulo de distribuciones muestrales, se tiene que una cantidad pivotal es

$$T = \frac{\Delta\bar{X} - \Delta\mu}{S_{\Delta\bar{X}}}$$

que se distribuye según una t-student con v grados de libertad, donde

$$S_{\Delta\bar{X}} = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$v = n_1 + n_2 - 2$$

Por lo tanto, el intervalo de confianza es

$$\left[\Delta\bar{X} - t_{v,1-\alpha/2}S_{\Delta\bar{X}}; \Delta\bar{X} + t_{v,1-\alpha/2}S_{\Delta\bar{X}}\right] \tag{2.3.10}$$

2) **Suponiendo varianzas diferentes.** Si las varianzas se suponen diferentes el estadístico sigue siendo el mismo, pero en este caso

$$S_{\Delta\bar{X}} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$v = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}$$

Por lo tanto, el intervalo de confianza es

$$\left[\Delta\bar{X} - t_{v,1-\alpha/2}S_{\Delta\bar{X}}; \Delta\bar{X} + t_{v,1-\alpha/2}S_{\Delta\bar{X}}\right] \tag{2.3.11}$$

4. **Intervalo de confianza para la diferencia de medias en poblaciones normales: Muestras dependientes**

Ahora tratamos construiremos un intervalo de confianza para la diferencia de dos medias cuando las muestras extraídas de las poblaciones normales no son independientes y las varianzas poblacionales no tienen porqué ser iguales. Es decir, supongamos que obtenemos una muestra aleatoria de n pares de observaciones $(X_1, Y_1)...(X_n, Y_n)$ de poblaciones normales con medias μ_X y μ_Y, en donde $(X_1, ..., X_n)$ indica la muestra de la población con media μ_X, y $(Y_1, ..., Y_n)$ indica la muestra de la población con media μ_Y.

En este caso podemos reducir la información a una sola muestra $(D_1, ..., D_n)$ en donde:

$$D_i = X_i - Y_i \quad , \qquad i = 1, 2, ..., n$$

y por las propiedades de la distribución normal, esta muestra $(D_1, ..., D_n)$ procederá también de una población normal de media:

$$\mu_D = E(D) = E(X - Y) = E(X) - E(Y) = \mu_X - \mu_Y$$

y varianza desconocida σ_D^2.

La varianza poblacional, σ_D^2, se puede estimar por la varianza muestral S_d^2 que sería la varianza de las diferencias que constituyen la muestra:

$$S_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2$$

siendo

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$$

Un estimador puntual de la media poblacional de las diferencias, μ_D, viene dado por $\bar{D}$.

Como la varianza poblacional es desconocida y pretendemos obtener un intervalo de confianza, al nivel de confianza del $100(1-\alpha)\,\%$, para μ_D procederemos de manera análoga al cuando se busco el intervalo de confianza para la media de una población normal cuando σ^2 era desconocida. Así pues, buscaremos un estadístico (cantidad pivotal o pivote) que depende del parámetro μ_D y de su estimador y cuya distribución muestral no depende de los parámetros desconocidos. Ese estadístico es:

$$T = \frac{\bar{D} - \mu_D}{S_{\bar{D}}}$$

que se distribuye según una t-student con $n-1$ grados de libertad, donde $S_{\bar{D}} = \frac{S_d}{\sqrt{n}}$.

Por lo tanto, el intervalo de confianza es

$$\left[\bar{D} - t_{(n-1),1-\alpha/2}\frac{S_d}{\sqrt{n}}; \bar{D} + t_{(n-1),1-\alpha/2}\frac{S_d}{\sqrt{n}}\right] \tag{2.3.12}$$

Ejemplo 2.3.4 *La tabla siguiente muestra el consumo de gasolina por 1.000 km de una muestra aleatoria de 9 carros con dos carburantes X e Y. Si admitimos que los consumos de gasolina se distribuyen normalmente, obtener un intervalo de confianza al nivel de confianza del 99 % para la diferencia de las medias poblacionales.*

Tabla 2.1: Consumo de gasolina por 1000 km, para los modelos X e Y

	Modelo X	Modelo Y	Diferencias d_i	d_i^2
1	132	124	8	64
2	139	141	-2	4
3	126	118	8	64
4	114	116	-2	4
5	122	114	8	64
6	132	132	0	0
7	142	145	-3	9
8	119	123	-4	16
9	126	121	5	25

De la tabla **??** *obtenemos que $\bar{d} = 2$ y $S_d^2 = 26{,}75$. Por lo tanto el intervalo de confianza usando la ecuación 2.3.4 es*

$$\left[2 - t_{8,0,995}\frac{5{,}17}{\sqrt{9}}; 2 + t_{8,0,005}\frac{5{,}17}{\sqrt{9}}\right]$$

como $t_{8,0,995} = 3{,}355$ *se tiene que el intervalo de confianza es*

$$[-3{,}781; 7{,}781]$$

5. **Intervalo de confianza para el cociente de varianzas en poblaciones normales** Sean $X_{11}, X_{12}, ..., X_{1n_1}$ y $X_{21}, X_{22}, ..., X_{2n_2}$ dos muestra aleatorias independientes extraídas de poblaciones normales, $N(\mu_1, \sigma_1^2)$ y $N(\mu_1, \sigma_1^2)$, respectivamente, cuyas varianzas son desconocidas y las medias pueden ser o no conocidas. Estamos interesados en hallar un intervalo de confianza del $100(1-\alpha)\,\%$ para el cociente de as varianzas entre las dos poblaciones, $\frac{\sigma_1^2}{\sigma_2^2}$. Para hallar dicho intervalo de confianza debemos considerar si las medias poblacionales son o no conocidas.

   a) **Medias desconocidas** Teniendo en cuenta la sección del capítulo anterior , en donde estudiamos la distribución del cociente de varianzas cuando las medias poblacionales eran desconocidas, entonces, aquí podemos utilizar como estadístico (cantidad pivotal o pivote) que dependa de los parámetros desconocidos $sigma_1^2$ y σ_2^2 y de sus estimadores y cuya distribución muestral no dependa de los parámetros, el estadístico:

$$F = \frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}/n_1 - 1}{\frac{(n_2-1)S_2^2}{\sigma_2^2}/n_2 - 1} = \frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2}$$

   el cual se distribuye F con $n_1 - 1$ y $n_2 - 1$ grados de libertad, F_{n_1-1,n_2-1},

   Ahora, utilizando la tabla de la distribución F, podemos encontrar dos valores $F_{\alpha/2;n_1-1,n_2-1}$ y $F_{1-\alpha/2;n_1-1,n_2-1}$, (la selección de estos dos valores garantiza que la

amplitud del intervalo sea mínima) tales que:

$$P(F_{\alpha/2;n_1-1,n_2-1} \leq F \leq F_{1-\alpha/2;n_1-1,n_2-1}) = 1-\alpha \tag{2.3.13}$$

de donde se tiene que

$$P\left(F_{\alpha/2;n_1-1,n_2-1} \leq \frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} \leq F_{1-\alpha/2;n_1-1,n_2-1}\right) = 1-\alpha$$

multiplicando por $\frac{S_2^2}{S_1^2}$

$$P\left(\frac{S_2^2}{S_1^2}F_{\alpha/2;n_1-1,n_2-1} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \frac{S_2^2}{S_1^2}F_{1-\alpha/2;n_1-1,n_2-1}\right) = 1-\alpha$$

Invirtiendo cada término y cambiando el orden de la desigualdad de tiene

$$P\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2;n_1-1,n_2-1}} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2;n_1-1,n_2-1}}\right) = 1-\alpha$$

y el intervalo de confianza para $\frac{\sigma_1^2}{\sigma_2^2}$ al nivel de confianza del $(1-\alpha)\,\%$ sería:

$$\left[\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2;n_1-1,n_2-1}}; \frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2;n_1-1,n_2-1}}\right] \tag{2.3.14}$$

b) **Medias conocidas**

En este caso usamos como cantidad pivotal el estadístico

$$F = \frac{\frac{(n_1)S_1^{*2}}{\sigma_1^2}/n_1}{\frac{(n_2)S_2^{*2}}{\sigma_2^2}/n_2} = \frac{S_1^{*2}}{S_2^{*2}}\frac{\sigma_2^2}{\sigma_1^2}$$

el cual se distribuye F con n_1 y n_2 grados de libertad, F_{n_1-1,n_2-1}.

Procediendo de manera análoga al caso anterior obtenemos el siguiente intervalo de confianza:

$$\left[\frac{S_1^{*2}}{S_2^{*2}}\frac{1}{F_{1-\alpha/2;n_1,n_2}};\frac{S_1^{*2}}{S_2^{*2}}\frac{1}{F_{\alpha/2;n_1,n_2}}\right] \tag{2.3.15}$$

donde

$$S_1^{*2}=\frac{1}{n_1}\sum_{i=1}^{n}(x_{1i}-\mu_1)^2 \qquad \text{y} \qquad S_2^{*2}=\frac{1}{n_2}\sum_{i=1}^{n}(x_{2i}-\mu_2)^2$$

Ejemplo 2.3.5 *Supongamos que la distribución de las notas en la asignatura de métodos estadísticos II sigue una distribución normal en los dos grupos existentes. Seleccionada una muestra aleatoria de 21 alumnos del primer grupo y otra de 26 alumnos del segundo grupo, ambas independientes, se obtiene como varianzas 1250 y 900, respectivamente. Obtener un intervalo de confianza para el cociente de las varianzas poblacionales al nivel de confianza del 90 %.*

Como las medias poblacionales son desconocidas utilizaremos la expresión 2.3.14 para hallar el intervalo de confianza. Donde $n_1 = 21$, $n_2 = 26$, $S_1^2 = 1250$ *y* $S_2^2 = 900$. *Usando la tabla F obtenemos que*

$$F_{1-\alpha/2;n_1-1,n_2-1} = F_{0,95;20,25} = \frac{1}{F_{0,05;20,25}} = falta$$

$$F_{\alpha/2;n_1-1,n_2-1} = F_{0,05;20,25} = falta$$

Sustituyendo en la expresión del intervalo se tiene

$$\left[\frac{1250}{900}\frac{1}{F_{1-\alpha/2;n_1-1,n_2-1}};\frac{1250}{900}\frac{1}{F_{\alpha/2;n_1-1,n_2-1}}\right]$$

$$[0{,}69;2{,}89]$$

2.3.3. Intervalos de Confianza para muestras grandes

En la mayoría de las situaciones prácticas la distribución de la población resulta ser desconocida o no es normal, en dicho caso no podríamos utilizar directamente los resultados obtenidos en la sección anterior. Sin embargo, si el tamaño de la muestra es suficientemente grande podemos utilizar el teorema central del limite para poder definir la cantidad pivotal. Consideremos el caso del intervalo de confianza para la media.

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria suficientemente grande procedente de una población con distribución desconocida y varianza σ^2 finita conocida y deseamos obtener un intervalo de confianza al nivel del $100(1-\alpha)\,\%$ para la media, desconocida, μ de la población. Puesto que se cumplen las condiciones del Teorema Central del Límite, podemos decir que el estadístico

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

se distribuye aproximadamente $N(0,1)$. Por lo tanto, dicho estadístico será nuestra cantidad pivotal, con el cual se tiene que

$$P\left(Z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq Z_{1-\alpha/2}\right) \simeq 1 - \alpha$$

y de manera análoga a como procedíamos anteriormente, llegaremos a que el intervalo de confianza al nivel del $100(1-\alpha)\,\%$ será:

$$\left[\bar{x} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{x} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \tag{2.3.16}$$

La diferencia con los intervalos obtenidos anteriormente es que aquellos eran exactos y ahora son aproximados y sólo son válidos para muestras grandes, $n > 30$.

Cuando σ^2 es desconocida se toma como valor aproximado la varianza muestral S^2, y se obtendría como intervalo de confianza:

$$\left[\bar{x} - Z_{1-\alpha/2}\frac{S}{\sqrt{n}}; \bar{x} + Z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right] \tag{2.3.17}$$

Expresiones análogas a las obtenidas anteriormente, se tendrá para el caso de la diferencia de medias poblacionales.

Ejemplo 2.3.6 *De los exámenes realizados a nivel nacional, se extrae una muestra de 75 ejercicios correspondientes a mujeres y otra de 50 ejercicios correspondientes a hombres, siendo la calificación media de la muestra de mujeres 82 puntos con una desviación típica muestra1 de 8, mientras que para los hombres la calificación media fue de 78 con una desviación típica de 6. Obtener el intervalo de confianza al nivel de confianza del 95 % para la diferencia de la puntuación media de las mujeres y la puntuación media de los hombres.*

Como las muestras son suficientemente grandes, pues son mayores que 30 y las poblaciones no son normales podemos obtener un intervalo de confianza aproximado utilizando la expresión 2.3.9 en donde sustituimos las varianzas poblacionales por las varianzas muestrales obteniendo el intervalo:

$$\left[\Delta\bar{X} - Z_{1-\alpha/2}\sigma_{\S\bar{X}}; \Delta\bar{X} + Z_{1-\alpha/2}\S_{\Delta\bar{X}}\right]$$

De donde

$\bar{x_1} = 82,\ S_1 = 8\ y\ n_1 = 75$

$\bar{x_2} = 78,\ S_{=}6\ y\ n_2 = 50$

Por lo tanto,

$$\Delta\bar{x} = \bar{x_1} - \bar{x_2} = 82 - 78 = 4$$

$$S_{\Delta\bar{X}} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{64}{75} + \frac{36}{50}} = 1{,}25$$

Sustituyendo en la expresión del intervalo tenemos:

$$[4 - (1{,}96)(1{,}25); 4 + (1{,}96)(1{,}25)]$$

$$[1{,}55; 6{,}45]$$

2.3.4. Intervalo de Confianza para Proporciones

Intervalo de Confianza para una Proporción

Sea una población binomial $B(1, \pi)$ y una muestra aleatoria de tamaño n de esa población, es decir realizamos n repeticiones del experimento de Bernoulli que estamos considerando, y deseamos obtener un intervalo de confianza al nivel del $100(1 - \alpha)\,\%$ para el parámetro poblacional π. Sólo vamos a considerar el caso en que los tamaños de muestras son grandes.

Como se vio antes el mejor estimador puntual de la proporción poblacional, π, es la proporción muestral, p. Además en el capítulo anterior se demostro que de acuerdo con el Teorema Central del Limite

$$p \to N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

Lo que nos permite decir que el estadístico

$$Z = \frac{p - \pi}{\sqrt{\pi(1-\pi)/n}} \tag{2.3.18}$$

se distribuye aproximadamente $N(0,1)$ cuando n es suficientemente grande.

En consecuencia este estadístico Z lo podemos utilizar como cantidad pivotal o pivote, pues depende del parámetro y de su estimador y su distribución es independiente del parámetro π, pues se trata de una $N(0,1)$. Por tanto, podremos obtener un intervalo de confianza para el parámetro π al nivel del $100(1-\alpha)\,\%$ a partir de la expresión.

$$P\left(Z_{\alpha/2} \leq \frac{p-\pi}{\sqrt{\pi(1-\pi)/n}} \leq Z_{1-\alpha/2}\right) = 1-\alpha$$

Multiplicando cada término de la desigualdad por $\sqrt{\pi(1-\pi)/n}$, restado después p a cada término y multiplicando por - 1, se tiene:

$$P\left(p - Z_{\alpha/2}\sqrt{\pi(1-\pi)/n} \leq \pi \leq p + Z_{\alpha/2}\sqrt{\pi(1-\pi)/n}\right) = 1-\alpha \tag{2.3.19}$$

Pero los límites de la expresión 2.3.19 dependen del parámetro desconocido π. Como n es grande una solución satisfactoria se obtiene sustituyendo π por su estimación p en el límite interior y en el límite superior, resultando:

$$P\left(p - Z_{1-\alpha/2}\sqrt{p(1-p)/n} \leq \pi \leq p + Z_{1-\alpha/2}\sqrt{p(1-p)/n}\right) = 1-\alpha \tag{2.3.20}$$

Luego el intervalo de confianza al nivel de confianza del $100(1-\alpha)\,\%$ para el parámetro π será:

$$\left[p - Z_{1-\alpha/2}\sqrt{p(1-p)/n}; p + Z_{1-\alpha/2}\sqrt{p(1-p)/n}\right)] \tag{2.3.21}$$

Ejemplo 2.3.7 *Se selecciona una muestra aleatoria de 600 familias, a las cuales se les pregunta si poseen o no computador personal en casa, resultando que 240 de esas familias*

contestaron afirmativamente. Obtener un intervalo de confianza al nivel del 95 % para estimar la proporción real de familias que poseen computador personal en casa.

El estimador puntual de π *sabemos que es* $p = \frac{x}{n}$ *y para la muestra concreta de 600 familias la estimación correspondiente será* $p = \frac{240}{600} = 0{,}40$. *Utilizando la Tabla de la distribución normal estándar se tiene que* $Z_{0{,}975} = 1{,}96$. *Sustituyendo en la expresión C2.3.21 tendremos el intervalo de confianza pedido*

$$\left[0{,}40 - 1{,}96\sqrt{0{,}40(1-0{,}40)/600}; 0{,}40 + 1{,}96\sqrt{0{,}40(1-0{,}40)/600}\right]$$

$$[0{,}36; 0{,}44]$$

Intervalo de Confianza para la Diferencia de Proporciones

Ahora estamos interesados en estimar la diferencia entre dos parámetros poblacionales π_1 y π_2, es decir queremos obtener un intervalo de confianza para la diferencia $\Delta\pi = \pi_1 - \pi_2$ de los dos parámetros poblacionales. Para ello se seleccionan dos muestras aleatorias independientes de tamaño n_1 y n_2, de cada una de las dos poblaciones $B(l, \pi_1)$ y $B(1, \pi_2)$, respectivamente. Los estimadores puntuales de los parámetros π_1 y π_2 son p_1 y p_2. Pero a nosotros nos interesa el intervalo de confianza para la diferencia $\Delta\pi = \pi_1 - \pi_2$, para lo cual utilizamos como estimador de esta diferencia, el estadístico $\Delta p = p_1 - p_2$, cuya distribución para muestras grandes (debido al teorea central del limite) es aproximadamente normal, es decir,

$$\Delta p \to N\left(\Delta\pi, \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}\right)$$

Lo que nos permite decir que el estadístico

$$Z = \frac{\Delta p - \Delta\pi}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}} \tag{2.3.22}$$

se distribuye aproximadamente $N(0,1)$ cuando n es suficientemente grande.

Por tanto, también podemos escribir

$$P\left(Z_{\alpha/2} \leq \frac{\Delta p - \Delta\pi}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}} \leq Z_{1-\alpha/2}\right) = 1-\alpha$$

de donde llegaremos a

$$P\left(\Delta p - Z_{1-\alpha/2}\sigma_{\Delta p} \leq \Delta\pi \leq \Delta p + Z_{1-\alpha/2}\sigma_{\Delta p}\right) = 1-\alpha \tag{2.3.23}$$

donde

$$\sigma_{\Delta p} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$$

Pero los límites de la expresión 2.3.23 dependen de los parámetros desconocidos π_1 y π_2. Como n_1 y n_2 son grandes una solución satisfactoria se obtiene sustituyendo cada π por su estimación p en el límite interior y en el límite superior, resultando:

$$P\left(\Delta p - Z_{1-\alpha/2}S_{\Delta p} \leq \Delta\pi \leq \Delta p + Z_{1-\alpha/2}S_{\Delta p}\right) \simeq 1-\alpha$$

donde

$$S_{\Delta p} = \sqrt{\frac{p_1(1-\langle_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Luego el intervalo de confianza al nivel de confianza del $100(1-\alpha)\,\%$ para el parámetro π será:

$$\left[\Delta p - Z_{1-\alpha/2}S_{\Delta p}; \Delta p + Z_{1-\alpha/2}S_{\Delta p}\right)] \tag{2.3.24}$$

Ejemplo 2.3.8 *En una ciudad A se toma una muestra aleatoria de 98 cabezas de familia, de los cuales 48 han sido poseedores de acciones de CANTV. Mientras que en otra ciudad B se selecciona otra muestra aleatoria de tamaño 127 cabezas de familia, de los cuales 21 han sido poseedores de acciones de CANTV. Obtener un intervalo de confianza al nivel del 95 % para la diferencia entre las proporciones de cabezas de familia que han sido poseedores de ese tipo de acciones en ambas ciudades.*

De la información del enunciado se deduce:

$n_1 = 98$, $x_1 = 48$, $p_1 = \frac{48}{98} = 0{,}49$

$n_2 = 127$, $x_2 = 21$, $p_1 = \frac{21}{127} = 0{,}165$

Para el nivel de confianza del 95 %, $\alpha = 0{,}05$, *se tiene* $Z_{0{,}975} = 1{,}96$. *Además*

$$S_{\Delta p} = \sqrt{\frac{0{,}49(1-0{,}49)}{98} + \frac{0{,}165(1-0{,}165)}{127}} = 0{,}118$$

Luego sustituyendo en la expresión 2.3.24 se tiene

$$[0{,}325 - 1{,}96 * 0{,}06; 0{,}325 + 1{,}96 * 0{,}06)]$$

$[0{,}21; 0{,}44)]$

Como el 0 está fuera del rango del intervalo, esto nos indica que es bastante más probable que un cabeza de familia de la ciudad A haya tenido acciones de CANTV que un cabeza de familia de la ciudad B.

2.4. Ejercicios

1. Explique lo que significa margen de error en la estimación puntual.

2. ¿Cuáles son las características del mejor estimador puntual para un parámetro poblacional?.

3. Calcule el margen de error al estimar una media poblacional μ para estos valores.

   a) $n = 30$, $\sigma^2 = 0{,}2$

   b) $n = 30$, $\sigma^2 = 0{,}9$

   c) $n = 30$, $\sigma^2 = 1{,}5$

   ¿Qué efecto tiene una varianza poblacional más grande en el margen de error?.

4. Una muestra aleatoria de 50 observaciones produjo $\bar{x} = 56{,}4$ y $s^2 = 2{,}6$. Dé la mejor estimación para la media poblacional y calcule el margen de error.

5. Estimaciones de la biomasa terrestre, la cantidad total de vegetación que tienen los bosques de la Tierra, son importantes para determinar la cantidad de dióxido de carbono no absorbido que se espera permanezca en la atmosfera de la tierra. Suponga que una

muestra de 75 parcelas de 1 metro cuadrado, elegidas al azr en los bosques de Mérida, produjo una biomasa media de 4.2 kilogramos por metro cuadrado, con una desviación estandar de 1.5 kg/m^2. ¿Cual es el mejor estimador de la biomasa promedio?. Estime la biomasa promedio para los bosques de Merida y el margen de error para su estimación.

6. A la mayoría de los habitantes de un país les encanta participar, o por lo menos ver, un evento deportivo. De una muestra de 1000 personas 780 respondieron que si les gustaba participar o ver un deporte.

   a) Identifique el mejor estimador puntual para la proporcionan de personas que si les gustaba participar o ver un deporte.

   b) Encuentre una estimación puntual para dicha proporción y el margen del error.

   c) La encuesta produce un margen de error de más o menos 3.1 %. ¿Esto concuerda con sus resultados del inciso b? Si no, ¿qué valor de p produce el margen de error dado en la encuesta?.

7. Suponiendo que las poblaciones son normales, encuentre e interprete un intervalo de confianza del 95 % para la media poblacional para estos valores

   a) $n = 36, \bar{x} = 13{,}1, \sigma^2 = 3{,}42$

   b) $n = 64, \bar{x} = 2{,}73, s^2 = 0{,}147$

8. Encuentre e interprete un intervalo de confianza del 90 % para la media poblacional para estos valores

   a) $n = 49, \bar{x} = 11{,}5, s^2 = 1{,}64$

   b) $n = 64, \bar{x} = 15, \sigma^2 = 9$

9. Una muestra aleatoria de $n = 300$ observaciones de una población binomial produjo $x = 263$ éxitos. Encuentre un intervalo de confianza del 90 % para la proporción e interprete el resultado.

10. Una máquina de café llena los vasos con volúmenes distribuidos normalmente con una desviación estándar de 0.11 oz. Cuando se toma una muestra de 23 vasos, se encuentra un volumen promedio de 7.85 oz. Estime el verdadero volumen promedio, de llenado de los vasos con 95 % de confianza.

11. Treinta artículos seleccionados en la producción tienen un costo medio de 180 Bs. Se conoce que la desviación estándar de la población es de 14 Bs. ¿Cuál es el intervalo de confianza al 99 % que considere el verdadero costo medio?.

12. De un lote de 680 máquinas, se estudia una muestra de 72 computadoras de cuarta generación. Se desea conocer cuál puede ser la duración promedio de un componente electrónico en particular, si su vida promedio en la muestra resultó ser de 4300 horas con desviación estándar de 730 horas. Se requiere que la estimación proporcione una confianza del 90 %.

13. Cuando un envasador nuevo se empezó a utilizar en una muestra de 40 envases, se encontró que los frascos de 100 ml eran llenados en promedio con 96 ml con desviación estándar de 8 ml.

    a) Estime entre cuántos mililitros está la verdadera cantidad media envasada con un nivel de confianza del 90 %.

    b) ¿Se podría garantizar que ninguno de los frascos contiene menos de 90 ml.?.

14. El departamento de carnes de una cadena de supermercados empaqueta la carne molida en bandejas de dos tamaños: una esta diseñada para contener más o menos 1 libra de carne, y la otra para casi 3 libras. Una muestra aleatoria de 35 paquetes de las bandejas más pequeñas produjo mediciones de peso con un promedio de 1.01 libras y una desviación estándar de 0.18 libras.

    a) Elabore un intervalo de confianza de 99 % para el peso promedio de los paquetes que vende esta cadena de supermercados en las bandejas de carne pequeñas.

    b) ¿Qué significa la frase çonfianza de 99 %¿.

    c) Suponga que el departamento de control de calidad de esta cadena de supermercados piensa que la cantidad de carne molida en las bandejas pequeñas debe ser en promedio 1 libra. ¿Debe preocupar al departamento de control de calidad el intervalo de confianza del inciso a? Explique.

15. Una muestra aleatoria de 130 temperaturas corporales humanas tuvo una media de 98.25 grados y una desviación estándar de 0.73 grados.

    a) Construya un intervalo de confianza de 99 % para la temperatura corporal promedio de personas sanas.

    b) ¿El intervalo de confianza construido en el inciso a tiene el valor de 98.6 grados, la temperatura promedio usual citada por médicos y otrod? Si no es así, ¿qué conclusiones obtiene?

16. Las especificaciones para una nueva aleación de alta resistencia al calor establecen que la cantidad de cobre en la aleación debe ser menor del 23.2 %. Una muestra de 10 análisis de un lote del producto presenta una media de contenido de cobre de 23 % y una

desviación estándar de 0.24 %. Estime el contenido medio de cobre en este lote, usando un intervalo de confianza del 90 % si se sabe que la cantidad de cobre se distribuye normal.

17. Un muestreo aleatorio de $n = 24$ artículos en un supermercado presenta una diferencia entre el valor real y el valor marcado en éste. La media y la desviación estándar de las diferencias entre el precio real y el precio marcado en los 24 artículos son -37.14 y 6.42 respectivamente. Encuentre un intervalo de confianza para la diferencia media entre el valor real y el marcado por artículo en ese supermercado, suponiendo que dicha diferencia se distribuye normal. Use $1 - \alpha = 0{,}05$

18. La utilidad por cada auto nuevo vendido por vendedor varía de auto a auto y se distribuye normal. La utilidad promedio por venta registrada en la semana pasada fue (en miles de bolívares) $21, 30, 12, 62, 45, 51$. Calcule un intervalo de confianza del 90 % para la utilidad promedio por venta.

19. Un investigador, desea estimar la verdadera proporción de amas de casa que prefieren la marca de detergente Ariel con un nivel de confianza del 95 %. Sabiendo que de una muestra de 150 amas de casa la proporción de amas de casa que les gusta Ariel es 0.47.

20. De entre 2000 piezas se eligen 75 y se encuentra que en 30 hay defectos. Calcule un intervalo de confianza del 90 % para informar a la gerencia.

21. Se tomó una muestra aleatoria de 300 adultos, y 192 de ellos dijeron que siempre votaban en las elecciones presidenciales.

    a) Construya un intervalo de confianza de 95 % para la proporción de venezolanos que afirman votar siempre en las elecciones presidenciales.

b) Una famosa encuestadora afirma que este porcentaje es de 67 %. Con base en el intervalo construido en el inciso a, ¿estaría en desacuerdo con este porcentaje? Explique.

c) ¿Se puede usar la estimación del intervalo del inciso a para estimar la proporción real de venezolanos adultos que votan en la elección presidencial de 2012? ¿Por qué sí o por qué no?.