



UNIVERSIDAD TÉCNICA DE MACHALA UNIDAD  
ACADÉMICA DE CIENCIAS EMPRESARIALES  
CARRERA DE ECONOMÍA MENCIÓN EN GESTIÓN EMPRESARIAL

TEMA:  
ANÁLISIS ESTADÍSTICO PARA DETERMINAR LA CONTRIBUCIÓN ANUAL AL  
PRODUCTO INTERNO BRUTO EN LA PARROQUIA EL CAMBIO

TRABAJO PRÁCTICO DEL EXAMEN COMPLEXIVO PREVIO A LA OBTENCIÓN DEL  
TÍTULO DE ECONOMISTA CON MENCIÓN EN GESTIÓN EMPRESARIAL

AUTORA:  
MOLINA CARPIO MAYELI CRISTINA

MACHALA - EL ORO

## CESIÓN DE DERECHOS DE AUTOR

Yo, MOLINA CARPIO MAYELI CRISTINA, con C.I. 0705336485, estudiante de la carrera de ECONOMÍA MENCIÓN EN GESTIÓN EMPRESARIAL de la UNIDAD ACADÉMICA DE CIENCIAS EMPRESARIALES de la UNIVERSIDAD TÉCNICA DE MACHALA, en calidad de Autora del siguiente trabajo de titulación ANÁLISIS ESTADÍSTICO PARA DETERMINAR LA CONTRIBUCIÓN ANUAL AL PRODUCTO INTERNO BRUTO EN LA PARROQUIA EL CAMBIO

- Declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional. En consecuencia, asumo la responsabilidad de la originalidad del mismo y el cuidado al remitirme a las fuentes bibliográficas respectivas para fundamentar el contenido expuesto, asumiendo la responsabilidad frente a cualquier reclamo o demanda por parte de terceros de manera EXCLUSIVA.
  
- Cedo a la UNIVERSIDAD TÉCNICA DE MACHALA de forma NO EXCLUSIVA con referencia a la obra en formato digital los derechos de:
  - a. Incorporar la mencionada obra al repositorio digital institucional para su democratización a nivel mundial, respetando lo establecido por la Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0), la Ley de Propiedad Intelectual del Estado Ecuatoriano y el Reglamento Institucional.
  
  - b. Adecuarla a cualquier formato o tecnología de uso en internet, así como incorporar cualquier sistema de seguridad para documentos electrónicos, correspondiéndome como Autor(a) la responsabilidad de velar por dichas adaptaciones con la finalidad de que no se desnaturalice el contenido o sentido de la misma.

Machala, 18 de noviembre de 2015

---

MOLINA CARPIO MAYELI  
CRISTINA C.I. 0705336485

## 1. Introducción

### 1.1. Contextualización

Para llevar adelante cualquier tipo de investigación, resulta de gran importancia conocer en qué medida los datos de la muestra pueden ser extrapolados a la población a la cual pertenecen, de esta manera mediante la inferencia estadística se pueden llegar a realizar afirmaciones o se tratarán de realizar suposiciones acerca de la población. Una de las estrategias que se llevan a cabo en las investigaciones es el de comparar los resultados obtenidos de dos o más grupos, de esta manera las técnicas paramétricas y no paramétricas que permiten evaluar la diferencia de medias, se encuentran entre las más utilizadas. (Merino Soto & Willson, 2013, pág. 145).

Entre los principales métodos de inferencia estadística se encuentran la estimación de parámetros y las pruebas de hipótesis. Para estimar los parámetros se lo puede realizar de dos formas, de forma puntual o mediante el uso de intervalos de confianza, el punto en común que tiene ambos métodos es el hecho de que tienen como objetivo estimar el valor del parámetro poblacional que se desconoce a partir de los datos de una muestra. En una estimación puntual se proporciona un valor único como estimación del parámetro, mientras que en una estimación por intervalo de confianza se establece un intervalo aleatorio y un nivel de confianza  $1 - \alpha$  que mide la probabilidad de contener al parámetro. (Insunza Caceres & Jiménez Ramírez, 2013, pág. 211).

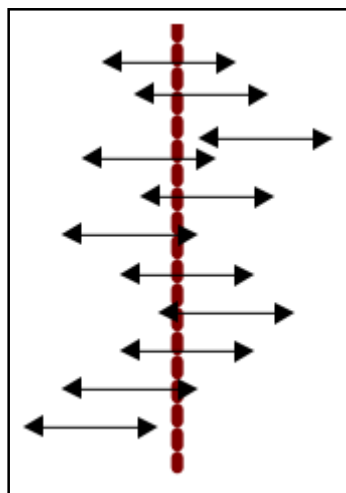
Pero es importante considerar que un estimado puntual, por ser un sólo número, no proporciona por sí mismo información alguna sobre la precisión y confiabilidad de la estimación. El estimado puntual nada dice sobre lo cercano que esta de  $\mu$ . Una alternativa para reportar un solo valor del parámetro que se esté estimando es calcular e informar todo un intervalo de valores factibles, un estimado de intervalo o intervalo de confianza. Un intervalo de confianza se calcula siempre seleccionando primero un nivel de confianza, que es una medida del grado de fiabilidad en el intervalo.

Para el presente caso de análisis, se tratará de establecer la contribución anual al PIB, por parte de los habitantes de la parroquia El Cambio, para lo cual se toma como muestra a 40 familias. Haciendo uso de la inferencia estadística se procederá más adelante a dar respuesta a esta problemática.

El simple cálculo de los intervalos de confianza basado en los estadísticos de la muestra y la comparación de los intervalos de confianza entre muestras es mucho más fácil de interpretar y aplicar que otras pruebas estadísticas. Pero los intervalos de confianza en sí no son un remedio. Sus cálculos, como los de las pruebas paramétricas, incluyen supuestos matemáticos con los que algunos investigadores no cumplen. De todos los acercamientos de la inferencia estadística tradicional o, la estimación de intervalos de confianza sobre la magnitud del efecto según las técnicas de muestreo es la más confiable y menos desacreditada. (Feinsinger, 2012, pág. 3)

Los intervalos de confianza proporcionan una manera de estimar, especialmente con alta probabilidad, un rango de valores sobre los que se encuentran el valor poblacional o parámetro de una variable. Esta probabilidad ha sido fijada por consenso en un 95% en base a supuestos de normalidad, pero rangos entre el 90% y 99% son comúnmente utilizados en la literatura científica. En términos simples, un intervalo de confianza del 95% indica que el valor poblacional se encuentra en un determinado rango de valores con un 95% de certeza. Como regla general, mientras mayor es el tamaño de la muestra, menor es la variabilidad para hacer la estimación del intervalo, lo que lleva a estimadores más precisos. Por el contrario, mientras más certeza se desee respecto a la extrapolación poblacional (por ejemplo, 99%), más amplio será el intervalo. (Madrid Aris & Felipe, 2014, pág. 543).

Un nivel de confianza de 95% implica que 95% de todas las muestras daría lugar a un intervalo que incluye  $\mu$  o cualquier otro parámetro que se esté estimando, y sólo 5% de las muestras producirá un intervalo erróneo. Cuanto mayor sea el nivel de confianza podremos creer que el valor del parámetro que se estima está dentro del intervalo. Una interpretación correcta de la “confianza de 95%” radica en la interpretación frecuente de probabilidad a largo plazo: decir que un evento A tiene una probabilidad de 0.95, es decir que si el experimento donde A está definido se realiza una y otra vez, a largo plazo A ocurrirá 95% de las veces. Para este caso el 95% de los intervalos de confianza calculados contendrán a  $\mu$ .



**Gráfico 1.** Valor verdadero de  $\mu$  (Madrid Aris & Felipe, 2014)

Esta es una construcción repetida de intervalos de confianza de 95% y se puede observar que de los 11 intervalos calculados sólo el tercero y el último no contienen el valor de  $\mu$ . De acuerdo con esta interpretación, el nivel de confianza de 95% no es tanto un enunciado sobre cualquier intervalo en particular, más bien se refiere a lo que sucedería si se tuvieran que construir un gran número de intervalos semejantes.

La diferencia de tamaño entre muestras grandes y pequeñas es importante cuando no se conoce la desviación estándar de la población y es necesario estimarla a partir de la desviación estándar de la muestra. Si el tamaño de la

muestra  $n$  es 30 o menos y se desconoce, debemos utilizar la distribución  $t$ . La distribución  $t$  apropiada tiene  $n - 1$  grados de libertad. Estas reglas también se aplican a la prueba de hipótesis.

Mediante la asignación puntual se asigna el valor del estadístico en la muestra del parámetro poblacional. Sin embargo lo más probable es que exista una cierta discrepancia entre el estadístico y el valor real del parámetro, la estimación por intervalos crea, a partir del valor estadístico, un rango de valores entre los que se espera, con una probabilidad conocida, que se encuentre el valor del parámetro poblacional. La cuestión principal radica en crear el intervalo en el que se debe encontrar el parámetro, para lo cual lo único que se necesita saber es la distribución muestral del estadístico, el mismo que actuará como un estimador. (Rial Boubeta & Varela Mallou, 2008, pág. 347).

Rial, Varela (2008), mencionan entre las características que presenta el intervalo de confianza, las siguientes:

- El intervalo de confianza se forma a partir del valor de la media y de su error típico. Es importante conocer que el error típico de la media es la desviación típica de su distribución muestral.
- Considerando el error típico que sea utilizado para crear el intervalo, este llegará a ser mayor o menor.
- El nivel de confianza, viene a configurarse como la probabilidad de que el intervalo que se ha creado, llegue a incluir dentro de sí, el valor del parámetro poblacional. Si se toma como referencia  $1,96 \sigma_{\bar{x}}$ , el nivel de confianza que le corresponde será  $1 - \alpha = 0,95$ .
- El nivel de riesgo ( $\alpha$ ), se constituye en la probabilidad de que el intervalo no llegue a incluir el valor del parámetro. Al tomar como intervalo  $1,96 \sigma_{\bar{x}}$ , el nivel de riesgo correspondiente es  $\alpha = 0,05$ .
- Al aumentar el número de desviaciones típicas para construir el intervalo, se incrementa el nivel de confianza y se disminuye el riesgo.
- Al aumentar el nivel de confianza, se incrementa la probabilidad de incluir el valor del parámetro en el intervalo.
- De forma habitual se suele tomar como nivel de confianza el 95% o bien el 99%. Es decir se llegan a elaborar intervalos con una probabilidad de 0,95 o bien el 99%. Es decir se elaboran intervalos con una probabilidad de 0,95 o de 0,99 de que incluya el valor del parámetro.

El significado que tiene un intervalo de confianza como procedimiento es relevante, para su comprensión, la adquisición de elementos previos como el de la distribución muestral. Los investigadores cometen muchos errores como

el de pensar que para distintas muestras se obtendrá el mismo intervalo y que el coeficiente de confianza no hace cambiar el intervalo. (Alvarado, Galindo, & Retamal, 2013, pág. 77)

Es importante analizar que un intervalo de confianza explica, con su amplitud, la incertidumbre debida al error muestral. Se señala también que una característica importante del intervalo de confianza es que puede ser utilizado como una prueba de hipótesis e inferir el valor de  $p$  asociado a ésta. (Figuroa, Larios, & Parra, 2011, pág. 129).

De acuerdo al Teorema del Límite Central, la distribución muestral de una estadística (como la media de la muestra) seguirá una distribución normal, siempre y cuando el tamaño de la muestra sea suficientemente grande. Entonces cuando conocemos la desviación estándar de la población podemos calcular un valor o calificación  $z$  y emplear la distribución normal para evaluar probabilidades sobre la media de la muestra. Sin embargo, muchas veces los tamaños de las muestras son muy pequeños, y frecuentemente no conocemos la desviación estándar de la población. Cuando estos problemas ocurren, en estadística se recurre a una distribución conocida como la “ $t$  de student”. (Hernández González & Quesada, 2011, pág. 52)

La  $t$  de Student está relacionada con el estudio de poblaciones muy grandes a partir de una muestra comparativamente muy pequeña. La función surge al querer estimar la media de una determinada variable en cierta población, que se supone normalmente distribuida, pero de la cual se desconoce la varianza, es decir, la tendencia de las muestras a desviarse del valor promedio. (Vecco, Pinedo, & Fernández, 2015, pág. 11)

## 1.2. Indicadores del problema

**Inferencia estadística:** Parte de la estadística que comprende los métodos y procedimientos que por medio de la inducción determina propiedades de una población estadística, a partir de una pequeña parte de la misma.

**Prueba de hipótesis:** Es una aseveración de una población elaborado con el propósito de poner a prueba, para verificar si la afirmación es razonable se usan datos. En el análisis estadístico se hace una aseveración, es decir, se plantea una hipótesis, después se hacen las pruebas para verificar la aseveración o para determinar que no es verdadera. Por tanto, la prueba de hipótesis es un procedimiento basado en la evidencia muestral y la teoría de probabilidad; se emplea para determinar si la hipótesis es una afirmación razonable.

**Hipótesis nula:** Es aquella que representa, lo que tradicionalmente se ha venido aceptando hasta el momento de realizar la prueba de hipótesis, de forma habitual indica que no existe relación entre las variables y que esta puede ser llegada a explicar por el azar. (Madrid Aris & Martínez Lomakin, Estadística para aterrorizados: interpretando intervalos de confianza y valores  $p$ , 2014, pág. 547).

**Hipótesis alternativa:** Es aquella que se acepta si los datos de la muestra, ofrece la suficiente evidencia para tomar la decisión de rechazar la hipótesis nula. (Vara, Díaz, & Batanero, 2011, pág. 76).

**Nivel de significancia:** En estadística, un resultado es estadísticamente significativo cuando no es probable que haya sido debido al azar. Una "diferencia estadísticamente significativa" solamente significa que hay evidencias estadísticas de que hay una diferencia; no significa que la diferencia sea grande, importante, o significativa en el sentido estricto de la palabra. El nivel de significación de un test es un concepto estadístico asociado a la verificación de una hipótesis. En pocas palabras, se define como la probabilidad de tomar la decisión de rechazar la hipótesis nula cuando ésta es verdadera (decisión conocida como error de tipo I, o "falso positivo"). La decisión se toma a menudo utilizando el valor p (o p-valor): si el valor p es inferior al nivel de significación, entonces la hipótesis nula es rechazada. Cuanto menor sea el valor p, más significativo será el resultado. En otros términos, el nivel de significación de un contraste de hipótesis es una probabilidad p tal que la probabilidad de tomar la decisión de rechazar la hipótesis nula - cuando ésta es verdadera - no es mayor que p.

**Intervalos de confianza:** En estadística, se llama a un par o varios pares de números entre los cuales se estima que estará cierto valor desconocido con una determinada probabilidad de acierto. Formalmente, estos números determinan un intervalo, que se calcula a partir de datos de una muestra, y el valor desconocido es un parámetro poblacional. La probabilidad de éxito en la estimación se representa con  $1 - \alpha$  y se denomina nivel de confianza. En estas circunstancias,  $\alpha$  es el llamado error aleatorio o nivel de significación, esto es, una medida de las posibilidades de fallar en la estimación mediante tal intervalo.

El nivel de confianza y la amplitud del intervalo varían conjuntamente, de forma que un intervalo más amplio tendrá más probabilidad de acierto (mayor nivel de confianza), mientras que para un intervalo más pequeño, que ofrece una estimación más precisa, aumenta su probabilidad de error.

Para la construcción de un determinado intervalo de confianza es necesario conocer la distribución teórica que sigue el parámetro a estimar,  $\theta$ . Es habitual que el parámetro presente una distribución normal.

**Valor crítico:** es un punto en la distribución de la prueba que se compara con el estadístico de prueba para determinar si puede rechazarse la hipótesis nula. Si el valor absoluto del estadístico de prueba es mayor que el valor crítico, usted puede declarar significancia estadística y rechazar la hipótesis nula. Los valores críticos están asociados con el nivel de significancia ( $\alpha$ ), así que sus valores se fijan cuando se elige el  $\alpha$  de la prueba.

**Distribución t de student:** Familia de distribuciones de probabilidad que se distinguen por sus grados de libertad individuales; es parecida, en forma, a la distribución normal y se utiliza cuando se desconoce la desviación estándar de la población y el tamaño de la muestra es relativamente pequeño ( $n \leq 30$ ).

**Grados de libertad:** Número de valores de una muestra que podemos especificar libremente, una vez que se sabe sobre dicha muestra.

### **1.3. Objetivo general del reporte**

Conocer el porcentaje de contribución anual al PIB de la población de la parroquia El Cambio, tomando como referencia su media poblacional.

### **1.4. Ventaja competitiva del trabajo**

La estadística es el conjunto de indicadores numéricos que caracterizan diferentes aspectos de la vida social, incluyendo la producción, las relaciones políticas, culturales de la vida cotidiana.....puede definirse como la recopilación, presentación análisis e interpretación de los datos numéricos....la estadística se refiere a las colecciones sistemáticas de datos relativos a un fenómeno". (Mendez, Figueroa, & Lloret, 2006, pág. 131).

Esta ciencia tiene distintas aplicaciones y a través de ella se pueden expresar, mediante indicadores, aspectos de gran utilidad en lo económico, social y natural. La ventaja competitiva se basa en que la economía necesita de la Estadística, con la ayuda de esta se confeccionan los planes de desarrollo de la economía nacional, se supervisa el control de su cumplimiento y se determinan las necesidades de recursos por territorios, así como las reservas con que cuenta la economía a cualquier nivel. Además la estadística constituye un instrumento de suma importancia para que se conozca el comportamiento de la economía a diferentes niveles ya sea en una empresa, municipio, provincia, nación, así como a escala internacional.

El conocimiento de la Estadística en la Economía permite apoyar la toma de decisiones para la aplicación de la política económica que se proponen los países para conducir la sociedad, así como para trazar la estrategia de desarrollo acorde con los programas que se consideran según las condiciones imperantes en cada nación.

Sería un error concebir la Estadística como un simple instrumento para estudiar aspectos parciales. El amplio campo de su aplicación permite incursionar en cada uno de los elementos que componen el complejo sistema socio-económico, así como investigar de una manera integral la relación entre sus principales variables. Es por esto que en el estudio de la economía y la Estadística constituye un elemento de inestimable valor.

Da una caracterización cuantitativa y cualitativa del volumen, composición y dinamismo de las fuerzas productivas y además refleja el comportamiento de las relaciones de producción en cada economía. Estudiar las fuerzas productivas de un país, las condiciones de producción, es el fin de esta ciencia.

Esta ciencia además de reunir los hechos, posibilita analizarlos profundamente y generalizarlos, colocándose en el centro de los fenómenos, convirtiéndose



así en un elemento activo que interviene en la solución de los problemas sociales.

Uno de los aspectos básicos de la Estadística y su relación con la Economía consiste en la elaboración de las metodologías de los sistemas de indicadores globales o macro agregados que deben caracterizar el estudio coherente de la economía nacional. Esto permite conocer como marcha el cumplimiento de los programas económicos a nivel de la economía nacional.

Mediante la estadística se descubren los estímulos del desarrollo de la producción sobre la base del progreso técnico y se estudia como satisfacer cada vez más las necesidades materiales de la sociedad y las posibilidades reales de satisfacción de ellas.

## 2. Desarrollo

**Caso:** En la provincia de El Oro, cantón Machala, parroquia El Cambio, existen aproximadamente 250 familias. Una muestra aleatoria de 40 de estas familias revela que la contribución anual al PIB fue de \$450 y la desviación estándar, de \$75. ¿Cuál será la media poblacional del contexto de las familias, si se sabe que estas pueden ser de \$445 o \$425, teniendo en cuenta un intervalo de confianza del 90%?

Se calcula por medio de la distribución T de student por ser una población finita y no conocer la media poblacional, que es el valor que se quiere calcular, \$450.

En estadística, la probabilidad que asociamos con una estimación de intervalo se conoce como nivel de confianza. Esta probabilidad indica qué tanta confianza tenemos de que la estimación de intervalo incluya al parámetro de población. Una probabilidad más alta implica una mayor confianza. En la estimación, los niveles de confianza que se utilizan con más frecuencia son 90, 95 y 99%, pero existe libertad para aplicar cualquier nivel de confianza.

Este problema requiere el uso de una distribución t, porque así lo pide el ejercicio en sus inicios. Aunque la distribución t está asociada con las estadísticas de muestras pequeñas, es importante recordar que un tamaño de muestra menor que 30 es solo una de las condiciones para usarla. Las otras son que no se conozca la desviación estándar de la población y que la población siga una distribución normal o aproximadamente normal.

La tabla de los valores de la distribución t, difiere en su construcción de la tabla z. La tabla t es más compacta y muestra áreas y valores de t sólo para algunos porcentajes (10, 5, 2 y 1%). Debido a que hay una distribución t diferente para cada número de grados de libertad, una tabla más completa sería bastante grande.

La segunda diferencia de la tabla t es que no se concentra en la probabilidad de que el parámetro de población que se está estimando se encuentre dentro del intervalo de confianza. En lugar de ello, mide la probabilidad de que el

parámetro de población que estamos estimando no esté dentro de nuestro intervalo de confianza (es decir, la probabilidad de que esté fuera).

Si estamos haciendo una estimación a un nivel de confianza del 90%, buscaríamos en la

tabla t en la columna de 0.10 (100% - 90% = 10%). Esta probabilidad de 0.10 del error se representa con el símbolo  $\alpha$ , la letra griega alfa. Encontraríamos los valores t apropiados para intervalos de confianza del 95, 98 y 99% en las columnas con títulos 0.05, 0.02 y 0.01, respectivamente.

La tercera diferencia al utilizar la tabla t es que debemos especificar los grados de libertad que se manejan.

Aplicamos la siguiente formula:

Para la resolución del ejercicio en primer lugar estableceremos los datos del mismo.

**Datos:**

$$\bar{x} = 450$$

$$N = 250$$

$$n = 40$$

$$gl = n - 1$$

$$gl = 40 - 1 = 39$$

Con los datos obtenidos, establecemos el valor de t, que corresponde a  $t = 1.685$ . A continuación se desarrolla el ejercicio con base a la formula inicialmente planteada

$$\bar{x} \pm t \frac{s}{\sqrt{n}} \left( \sqrt{\frac{N-n}{N-1}} \right)$$

$$450 \pm 1.685 \frac{75}{\sqrt{40}} \left( \sqrt{\frac{250-40}{250-1}} \right)$$

$$450 \pm 1.685 \frac{75}{\sqrt{40}} \left( \sqrt{\frac{210}{249}} \right)$$

$$450 \pm 18.35$$

Los puntos extremos del intervalo de confianza son \$431.65 y \$468.35.

Es probable que la media poblacional sea de más de \$431.65 e inferior a \$468.35. En otras palabras, ¿La media de la población puede ser de \$445? Sí, pero no es probable que sea de \$425. ¿Por qué? Porque el valor de \$445 se encuentra dentro del intervalo de confianza y \$425 no pertenece al intervalo de confianza.

### 3. Cierre

La estadística, como parte de las matemáticas empleadas, es la secuencia de razonamientos para estudiar los fenómenos de la naturaleza, y considera que la inferencia estadística es la que permite obtener conclusiones en función de los resultados obtenidos en una muestra en estudio representativa; sin embargo, cabe señalar que hay

dos tipos de inferencias: la que se conoce como prueba de hipótesis y la que concierne a la estimación de intervalos.

Cada prueba de análisis se emplea en función del diseño de investigación, que se haya seleccionado para comprobar si la consecuencia es verificable a partir de la hipótesis general de investigación.

La *t* de Student y la prueba *Z* ponen a prueba la hipótesis de que la diferencia entre dos promedios se debe al azar; la *p* de una correlación indica la probabilidad de que la pendiente de esa recta sea igual a la pendiente de una recta igual a 0, y no indica qué tan importante es esa correlación, mientras que la *p* de la prueba de hipótesis indica cuál es la probabilidad de que la distribución observada se deba al azar.

El intervalo de confianza utilizado en el ejercicio planteado que es del 90% indica la imprecisión de las estimaciones puntuales; esto es si se tomó un número infinito de muestras del mismo tamaño que la que se estudió, responde a la pregunta de: ¿Entre cuál de éstos valores (alto y bajo) se encontraría el 95% de las mediciones de esta diferencia o correlación? Si la magnitud de los valores de los dos extremos es menor a la unidad, se considera un efecto protector; y si la magnitud de estos valores de los dos extremos es mayor a la unidad, el resultado es significativo. Si la magnitud de los valores de uno de los extremos es menor a la unidad y el otro es mayor a la unidad, el resultado es no significativo.

En la resolución del problema se comprobó que el valor obtenido de 425 no se encuentra dentro del intervalo de confianza establecido, planteándose que la media poblacional se encuentre entre \$431.65 e inferior a \$468.35.