TEMA Nº 4

|  |
| --- |
| ESTIMACIÓN DE INTERVALOS DE CONFIANZA |
| *El fundamento del intervalo de confianza**Intervalo de confianza para media poblacional (Muestras grandes)**Intervalo de confianza para la media en caso de muestras pequeñas - distrib. "t"**Intervalo de confianza para proporción poblacional**Control de ancho de intervalo**Determinación del tamaño apropiado de la muestra* *Propiedades de un buen estimador* |
| Lic. Henry Rukner Reynaga Arce |

Copiright 2000. Estadistica Aplicada a los Negocios y la economía. Allen L Webster

**TEMA 4**

**ESTIMACION DE INTERVALOS DE CONFIANZA**

**4.1 Tipos de estimación, estimación y estimadores**

**Estimación puntual:** Una estimación puntual es un solo número que se utiliza para estimar un parámetro de población desconocido. Ejemplo: al estimar el peso de una persona tan solo digo “apuesto que pesa 75 Kg”**.**

**Estimación de intervalo:** Una estimaciónde intervalo es un intervalo de valores que se utiliza para estimar un parámetro de población. Ejemplo: al estimar el peso de una persona puedo decir “apuesto que su peso está entre los 73 y 78 Kg”

**Estimador:** Cualquier estadístico de muestra que se utilice para estimar un parámetro de población se conoce como estimador, es decir, un estimador es un estadístico de muestra utilizada para estimar un parámetro de la población. La media de la muestra puede ser un estimador de la media de la población, y la porción de la muestra se puede utilizar como estimador de la porción de la población. Un estimador es una regla procedimiento, generalmente expresado en una fórmula que se utiliza para derivar la estimación: ejemplo:

Estimador 

**Estimación:** Cuando observamos un valor especifico de nuestro estimador, nos referimos a ese valor como una estimación. En otras palabras, una estimación es un valor específico observado de una estadística. Hacemos una estimación si tomamos una muestra y calculamos el valor que toma nuestro estimador en esa muestra.

Estimación 

**4.2 Fundamento de un intervalo de confianza**

Un intervalo de confianza tiene un límite inferior de confianza (LIC) y un límite superior (LSC) de confianza. Estos límites se hallan calculando primero la media muestral (), luego se suma cierta cantidad a esta media para obtener el LSC, y la misma cantidad se resta de para obtener el LIC.

La pregunta de rigor es: ¿cómo se puede construir un intervalo y luego argumentar que se puede tener un 95% de confianza en que contiene µ (parámetro), si incluso no se sabe cuál es la media poblacional?. Vale recordar la regla empírica de la distribución normal en una distribución de medias muestrales: “*El 95.5% de caen dentro de dos errores estándar de la media poblacional, por lo tanto la media poblacional esta a máximo dos errores estándar del 95.5% de todas las medias muestrales*”. Por lo tanto al comenzar cualquier media muestral, si se pasa por dos errores estándar por encima de la media y dos errores estándar por debajo de ella, se puede tener un 95.5% de confianza en que el intervalo resultante contenga la media poblacional desconocida.



Todas las muestras consideradas producirán un intervalo que contiene la media poblacional. Entonces, la clave para recordar es esta: como la media poblacional esta a lo mas de dos errores estándar para el 95.5% de todas las medias muestrales, entonces dada una media muestral cualquiera, se puede estar el 95.5% seguros de que el intervalo de dos errores estándar alrededor de dicha media muestral contiene la media poblacional desconocida que vimos en el cuadro anterior.

**Podemos construir nuestro propio intervalo de confianza**

Si cada uno de nosotros desea construir un intervalo a la medida de nuestro objeto de estudio (que sea superior o inferior al 95.5% discutido anteriormente) ¿Cuántos errores estándar debemos mover por encima y por debajo de la media muestra?. Como la tabla Z contiene valores solo para el área que está por encima o por debajo de la media se debe dividir –por ejemplo- el 95% entre 2, produciendo un 0.4750. Luego se halla el valor de Z correspondiente a un área de 0.4750, el cual es Z=1.96. Así para construir un intervalo de confianza del 95%, simplemente se especifica un intervalo de 1.96 errores estándar por encima y por debajo de la media muestral. ***Este valor del 95% es llamado coeficiente de confianza.***

***Coeficiente de confianza:*** es el nivel de confianza que se tiene en el que el intervalo contenga el valor desconocido del parámetro

**4.3 Intervalo de confianza para la media poblacional – Muestras grandes**

El intervalo de confianza se usa con más frecuencia para estimar el valor de la media poblacional el número de circunstancia en las que se puede aplicar para el mundo de los negocios y la economía es casi ilimitado.

Para recordar: El intervalo se forma utilizando la media muestral como una estimación puntual para el cual se adiciona y se resta un cierto valor para obtener los limites superior e inferior del intervalo de confianza respectivamente. Por tanto el intervalo es:

I.C. para estimar µ:

 Notación extendida: 

 Notación reducida: 

Donde:

Z: es el valor z que encontramos en la tabla de distribución normal, para un I.C del 95% es 1.96 para un 90% es 1.645 y para un 99% es 2.575.

= Media muestral.

= Error estándar “mide la tendencia a sufrir error de muestreo en el esfuerzo por estimar ”.

Caso práctico:

Para estimar el gasto promedio de sus clientes, la tienda de batidos toma una muestra de 200 estudiantes de la facultad y encuentra un gasto promedio de Bs.7.5. Se sabe que la desviación estándar de la población es de Bs.1.3. ¿Cuál es el intervalo de confianza del 95% para los gastos promedio de todos los clientes? Interprete sus resultados. Recordemos que 



***Interpretación de un intervalo de confianza***

Primera forma de interpretar: ”Tenemos un 95% de confianza en que la media poblacional real desconocida este entre Bs. 7.32 y 7.68“. Aunque el valor real de la media poblacional sigue siendo desconocido, el vendedor tiene un 95% de confianza en que esté entre estos valores.

Segunda forma de interpretar se puede desarrollar con intervalos de confianza diferentes. Si tomamos otra muestra probablemente producirá un a media muestral diferente debido al error de muestreo. Con una  diferente, en intervalo tendría un límite inferior y superior diferente. Por tanto la segunda interpretación establece que si se construyen todas las combinaciones de n elemento de un una población N para determinar intervalos de confianza, el 95% de los mismos contendría la media poblacional desconocida.

Si una segunda muestra nos da Bs 7.7 de gasto en lugar de Bs. 7.5 estos serían los resultados:



Ahora nuestro vendedor de batidos puede decir que esta 95% seguro de que la media población está comprendida entre Bs. 7.52 y 7.88. Si todos los intervalos posibles se construyeran con base en todas las medias muestrales diferentes, el 95% de ellas contendría la media poblacional desconocida.

Mucha atención: Si trabajamos a un nivel de confianza del 95% significa que el 5% de todos los intervalos estaría errado – no contendría la media poblacional- . Este 5% hallado como (1- coeficiente de confianza), es denominado el valor alfa y representa la población de error. El valor alfa es la probabilidad de que cualquier intervalo dado no contenga la media poblacional.

**Valor alfa :** es la probabilidad de error o la probabilidad de que un intervalo dado no contenga la media poblacional desconocida.

**Intervalo de confianza cuando  es desconocida.**

El caso anterior así como los que hemos trabajado en el tema tres, requiere una suposición poco probable: “la desviación estándar poblacional **** es conocida”. En este caso la desviación estándar de la muestra debe sustituirla.

Intervalo de confianza para estimar µ

Cuando es desconocida.  Donde: 

Caso práctico.

Después de escuchar 60 programas de radio seleccionados aleatoriamente en el lapso de un mes, se reportó un promedio de 42.4 expresiones verbales con signos de violencia. La desviación estándar muestral en este mismo estudio fue de 9.2. ¿Cuál es la estimación al 95% del número promedio de expresiones verbales violentas por programa que los niños escuchan en la radio?



Interpretación: podemos afirmar que tenemos el 95% de certeza que el promedio de expresiones verbales violentas en programas radiales está entre 40 y 44.7.

Nos podemos hacer una nueva pregunta ¿qué pasa si quiero más confianza y elevo el nivel de confianza al 99% con un valor Z de 2.575?. Veamos los resultados.



Los resultados son buenos como malos.

Lo bueno es que nuestro nivel de confiabilidad con 99% nos da mayor certeza de que la media poblacional se encuentre inmersa en nuestro intervalo.

Lo malo es que hemos abandonado gradualmente la **precisión de la estimación** y el intervalo puede hacerse demasiado grande. Suponga que ve un joven bastante alto e infiere que su talla está entre 1 metro y 3 metros, de seguro que la talla de esta persona se encuentra en ese margen sin embargo la apreciación parece burlesca pues no existe una persona que pueda medir 3 metros y por otro lado si midiese 1 metro, para nada seria considerado una persona alta.

**4.4 Intervalo de confianza para la media en el caso de muestras pequeñas – la distribución t**

En los ejemplos estudiados hasta hora el tamaño de muestra siempre fue mayor o igual a 30 observaciones. Sin embargo, no siempre podemos garantizar este tamaño de muestra. Si una compañía de seguros prueba la resistencia de un modelo de automóvil, destruir 30 vehículos o más seria demasiado costoso, por ende la muestra debe ser pequeña.

Cuando tomamos una muestra pequeña, la distribución normal puede no aplicarse. Recordemos que el teorema del límite central asegura normalidad en el proceso solo si la muestra es grande. Cuando se utiliza una muestra pequeña puede ser necesaria una distribución alternativa: ***la distribución t de Student*** (o simplemente distribución t)[[1]](#footnote-1).

Específicamente la distribución t se utiliza cuando se cumplen tres condiciones:

1. La muestra es pequeña.
2. es desconocido
3. La población es normal o casi normal.

Mucha atención: si  es conocida, la distribución Z se usa inclusive si la muestra es pequeña. Si no puede asumirse una población normal se aumenta la muestra hasta llegar a treinta para usar la distribución Z , de no ser posible se deben usar *pruebas no parametricas*.

Al igual que la distribución Z la distribución t tiene una media cero, es simétrica con respecto a la media y oscila entre y . Sin embargo mientras que la varianza de la distribución Z es =1 , la varianza de la distribución t es mayor que 1.Por tanto es mas plan y dispersa que la distribución Z.

Varianza de la distribución t: 

***A tomar muy en cuenta***: la distribución t es una familia de distribuciones cada una con su propia varianza. La varianza depende de los grados de libertad (g.l.), definidos como el número de observaciones que se pueden escoger libremente. Es el número de observaciones menos el número de restricciones impuestas sobre tales observaciones. En donde una restricción es algún valor que tales observaciones deben poseer.

Un ejemplo: se asume que se tienen n=4 observaciones que dan una media de 10. La media de 10 sirve como una restricción y hay n-1=3 grados de libertad. Por tanto, se pueden escoger tres observaciones cualquiera; por ejemplo se pueden escoger 8, 9 y 11 que dan una suma de 28, después de escoger estos tres valores ya no hay libertad para escoger la última observación, por ende el cuarto valor debe ser 12, por al sumar 12+28=40, dividido en tres el número total de observaciones n=4 es resultado será un promedio de 10.

**Grados de libertad**: El número de observaciones menos el número de restricciones impuestas sobre tales observaciones.

**La dispersión y el intervalo en distribuciones t:** Como la varianza en distribuciones t es más grande que la de distribuciones normales también la desviación estándar será mayor, por ende la distribución es mas dispersa . Sin embargo este ancho adicional es necesario debido a que se pierde algo de precisión porque es desconocida y debe estimarse.

**El estadístico t:**



**Error estándar para una distribución t:**

****

**Intervalo de confianza para estimar la media poblacional (muestras pequeñas)**

Notación extendida: ****

Notación reducida: ****

**Ejemplo práctico:**

Papri –k vende vasos de cerveza de 20 onzas cada uno, un grupo de estudiantes de estadística compra un total de 19 vasos, y utilizando su propia tasa de medida estima que el contenido promedio es de 19.6 onzas con una desviación estándar de la muestra de 0.9 onzas (s=0.9). ¿Con un nivel de confianza de 90% los estudiantes piensan que su dinero lo vale? Interprete el intervalo.

****

**4.5 Intervalo de confianza para la proporción poblacional**

 Muchas veces las decisiones dependen de parámetros que son binarios, parámetros con solo dos categorías dentro de las cuales pueden clasificarse las respuestas. En este evento, el parámetro de interés es la proporción poblacional.

En el capítulo 3 vimos que el error estándar de la distribución muestras de las proporciones muestrales era:



Sin embargo esta fórmula contiene P, que es el parámetro que se desea estimar. Por tanto, la proporción muestral p se utiliza como estimador de P.

***Estimación del error estándar de la distribución de las proporciones muestrales.***



El intervalo de confianza para estimar la proporción poblacionales entonces:

Notación extendida: 

Notación reducida: 

Ejemplo práctico:

De los 940 pasajeros que tuvo la Flota Bolpar el fin de semana, 564 registraron solo equipaje de mano. Calcule e interprete el intervalo del 95% para la proporción de todos los pasajeros que llevan solo equipaje de mano.

 p=564/940 = 0.6

******

I.C para estimar la proporción poblacional.



Interpretación: Se confía en que a un nivel de confianza de 95% entre el 56.9% y el 63.1% de los pasajeros de la Flota Bolpar lleva tan solo equipaje de mano en sus viajes.

**4.6 Control de ancho de un intervalo**

Para un investigador es preferible tener un intervalo más estrecho debido a la precisión adicional que proporciona, los métodos principales para lograr un intervalo más preciso son:

1. **Reducir el nivel de confianza**

Como habíamos visto en un caso anterior, al reducir el nivel de confianza por ejemplo de 99% a 95% también el valor z reduce de 2.575 a 1.96 por ende el margen de error disminuirá haciendo que el intervalo sea menor. Lo malo de este método es que al reducir el nivel de confianza también aumentamos la probabilidad de error de 1% (en el caso nc=99%) a 5% (en el caso nc=95%). ¿Existe alguna manera en la que se puede reducir el intervalo sin sufrir una pérdida de confianza?. Si. Incrementando el tamaño muestral.

1. **Incrementar el tamaño muestral**

Como bien sabemos incrementando el tamaño de la muestra podemos reducir el error estándar (). Si aumentamos el tamaño de la muestra lo suficiente, el intervalo de 99% puede presenta un grado de precisión similar al intervalo más estrecho del 95% sin ninguna pérdida de confianza.

Infortunadamente, esta ventaja no se gana sin un precio. El tamaño más grande de la muestra significa más tiempo y más dinero que deben gastarse al recolectar y manejar los datos. De nuevo, debe tomarse una decisión. Se vuelve a la decisión general respecto a que muestra tomar.

**Investigue:**

Propiedades de un buen estimador: Insesgabilidad, eficiencia, consistencia, suficiencia.

1. La distribución t de Student fue desarrollada en 1908 por William S. Gosset (1876 – 1937) quien trapajo como experto cervecero para Guinens Breweries en Dublin, Irlanda. Guiness no permitía que sus empleados publicaran sus investigaciones, de manera que Gosset (a quien le gustaba jugar con números para relajarse) informo por primera vez sobre su distribución t bajo el seudónimo de “Student” para proteger su trabajo. [↑](#footnote-ref-1)