UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

COLEGIO DE CIENCIAS Y HUMANIDADES
PLANTEL VALLEJO
ÁREA DE MATEMÁTICAS
ESTADÍSTICA Y PROBABILIDAD II

UNIDAD III. INFERENCIA ESTADÍSTICA

ESTIMACIÓN PUNTUAL Y POR INTERVALO

En la unidad II se trabajó con muestras tomadas de poblaciones donde sus parámetros se conocían, o bien, se podían tomar todas las muestras posibles de tamaño n.

Lo más común es no conocer la población y el único camino que queda es tomar muestras para poder estudiarla.

En esta unidad se presenta la Estimación o Inferencia Estadística, para mostrar cómo se pueden obtener estimadores de los parámetros desconocidos de una población.

Se muestran requisitos y el procedimiento para calcular los llamados *intervalos de confianza*, para la media y la población, donde esperamos encontrar al verdadero valor poblacional en estudio.

INTRODUCCIÓN

Las razones para efectuar una estimación en una población, en lugar de estudiarla completamente pueden ser, como ya se menciono, que el tamaño de la población sea infinito, que el muestreo sea destructivo, que la población sea finita pero demasiado grande y algunas otras razones como costo y tiempo. Por esto parece ser más práctico tomar muestras. Existen dos maneras básicas de hacer estimaciones: *estimaciones puntuales*; en la cual se estima el parámetro desconocido por un solo valor. *Estimación por intervalos,* donde se estima el parámetro desconocido por medio de un intervalo de números reales, el cual tiene asociado una cierta probabilidad de contenerlo.

Elaborado por: *ELEAZAR GÓMEZ LARA*
Abril 2007

ESTIMACIÓN.

La estadística descriptiva no sólo nos permite obtener un perfil del comportamiento de los datos muestrales; nos permite también obtener estimaciones de parámetros poblacionales, lo que generalmente es lo más importante.

Por una parte, en la población tenemos medidas de tendencia central, de posición y de dispersión que son fijas e invariables.
Estas medidas son llamadas parámetros poblacionales o simplemente parámetros. Por ejemplo, la estatura promedio de la mujer mexicana en la población es constante, así como su desviación estándar, su moda, su mediana etc.

Por otra parte, el cálculo de promedios, medianas, etc. obtenidos en una muestra son estimaciones de esos parámetros. Estas medidas son llamadas parámetros estimados o estimadores.
A diferencia de los parámetros poblacionales, los estimadores muestrales no son únicos, ya que varían al tomar distintas muestras de la misma población.
En su carácter muestral, los estimadores son llamados medidas resumen, estadígrafos o estadísticos.
Los parámetros poblacionales habitualmente se simbolizan con una letra griega y sus estimadores con una letra latina. También es posible estimar distribuciones, conglomerados, etc.

Característica	Parámetro	Estimador
Media o Promedio	μ	$\bar{X}$
Desv. Estándar	σ	s
Varianza	σ^2	s^2
Error Estándar	$\frac{\sigma}{\sqrt{n}}$	$\frac{s}{\sqrt{n}}$
Proporción	π	p
Distribución		Histograma

Las dos aplicaciones principales de la estadística inferencial implican el uso de datos de muestra para:

a) estimar el valor de un parámetro de población,
b) llegar a una conclusión acerca de una población.

En esta unidad presentaremos métodos para estimar valores de los siguientes parámetros de población: medias y proporciones.

Consideremos las temperaturas corporales de un grupo de un grupo de alumnos, con base a estos valores de muestra, queremos estimar la media de todas las temperaturas corporales, podríamos usar una estadística como la mediana de la muestra, la mitad del rango o la moda como un estimado de la media μ de la población, pero la media de la muestra $\overline{x}$ es la que nos da el mejor estimado de la media poblacional. El uso de $\overline{x}$ se basa en un estudio y análisis cuidadosos de la distribución de las diferentes estadísticas que podrían servir como estimadores.

<u>Un estimador</u> es una estadística de muestra que se usa para aproximar un parámetro de población.

<u>Un estimado</u> es un valor o intervalo de valores específico que se usa para aproximar algún parámetro de población.

Dos razones importantes por lo que la media de una muestra es un mejor estimador de una media poblacional son:

a) En muchas poblaciones, la distribución de las medias de muestra $\overline{x}$ tiende ser más uniforme que las distribuciones de otras estadísticas de muestra.
b) Para todas las poblaciones la media de muestra $\overline{x}$ es un estimador no sesgado de la población μ, lo que quiere decir que la distribución de las medias tiende a centrarse alrededor del valor de la media de la población μ.

Por estas y otras razones se utiliza la media como el mejor estimador de la media de la población.

Puesto que la media de muestra $\bar{x}$ es un valor único que corresponde a un punto en la recta real, decimos que es un estimado puntual.

Un estimado puntual es un valor individual (o punto) que se usa para aproximar un parámetro de población.

La media muestral $\bar{x}$ es el mejor estimado puntual de la media de la población μ.

Un intervalo de confianza (o estimación por intervalo) es una gama (o un intervalo) de valores que *probablemente* contiene el valor verdadero del parámetro de población.
A un intervalo de confianza se le asocia un grado de confianza, que es una medida de la certeza que tenemos de que nuestro intervalo contienen el parámetro de población.

El *grado de confianza* es la probabilidad $1-\alpha$ de que el intervalo de confianza contiene el verdadero parámetro de la población (al grado de confianza también se le da el nombre de *nivel de confianza*).

Ahora estamos listos para desarrollar los procedimientos para estimar los parámetros de una distribución muestral.

La estimación comienza con una muestra aleatoria y una distribución de probabilidad asociada. Así los datos son utilizados par estimar e valor de los parámetros de la distribución de probabilidad.

El primer objetivo en estudio estadístico es obtener un sólo número que mejor represente una característica de la población.

Si θ representa el parámetro poblacional que deseamos estimar y $\hat{\theta}$ es la estadística muestral o el estimador puntual de θ. Por ejemplo θ puede ser el peso medio de los paquetes de las cajas de cereal de trigo, μ, o el porcentaje de familias que compran este tipo de cereal, π, así $\overline{X}$, y $\hat{p}$ son ejemplos de estimadores $\hat{\theta}$ de los parámetros poblacionales.

Conociendo la distribución de probabilidad podemos desarrollar ecuaciones matemáticas, llamados estimadores, las cuales pueden ser utilizadas para estimar los parámetros.

Def. Un estimador puntual $\hat{\theta}$, es una función de la muestra de datos que proporciona un estimado de un parámetro desconocido.

La media muestral y la mediana son ejemplos de estimadores de la media poblacional. La media tiene una menor varianza que la mediana para variables normalmente distribuidas. La proporción poblacional es un estimador de la probabilidad de éxitos en la distribución binomial.

Por ejemplo la media muestral $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ es un estimador puntual ($\hat{\theta} = \overline{X}$) de la media poblacional μ ($\theta = \mu$).

Un estimador puntual es un número calculado a partir de una muestra de datos utilizando el estimador puntual. Los estimadores puntuales son variables aleatorias porque son funciones de las observaciones muestrales que a su vez son variables aleatorias.

El valor del estadístico varía de muestra a muestra, y no todos los valores estimados coincidirán con el valor del parámetro a estimar.

Un buen estimador debe ofrecer, al menos en promedio, estimaciones correctas por lo que es deseable que tengan las siguientes propiedades:

Un estimador insesgado, $\hat{\theta}$; es un estimador donde el valor esperado es el valor del parámetro desconocido, matemáticamente

$$E\left(\hat{\theta}\right)=\theta$$

Sesgo.

Se llama sesgo a la diferencia que existe entre un estimador y el parámetro al cual estima. Este sesgo (o error) se presenta cuando hay problemas en la selección de los sujetos que componen la muestra, la calidad de los instrumentos utilizados, la confiabilidad de las respuestas de personas encuestadas, etc. Evidentemente, mientras mayor es el sesgo, peor es la estimación del parámetro de interés. Mientras mayor es la precisión, menor es el sesgo cometido.

Cuando un estimador se "acerca" o "aproxima" cada vez más al parámetro al cual estima, a medida que el tamaño muestral aumenta, se denomina un estimador insesgado.

Finalmente, dado que una medida resumen obtenida en una muestra es al fin y al cabo un sólo valor destinado a estimar un parámetro, y dado además que este estimador no es único, suele llamarse un estimador puntual.

Se utilizará la distribución del estimador para calcular su valor esperado. *Un estimador es insesgado si el valor esperado del estimador es igual al del parámetro.*

Tanto la media como la proporción muestral son estimadores insesgados.

Estimador Eficiente

Una mayor eficiencia indica que el estadístico varía menos de una muestra a otra $\rightarrow$ Mayor precisión en la estimación del parámetro a partir del estadístico estimado en una muestra.

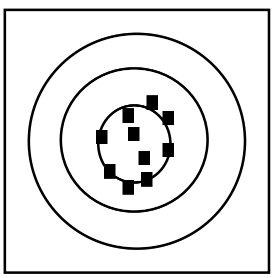
Ej. Comparación de dos estadísticos: Media y mediana. ¿Cuál es más eficiente?

$\sigma^2(\overline{X}) = \frac{\sigma^2}{n}$ y $\sigma^2(\tilde{X}) = 1.57\frac{\sigma^2}{n}$

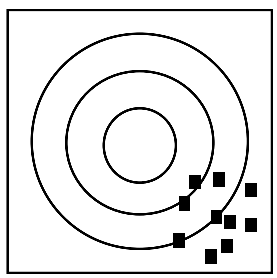
$$\sigma^2(\overline{X}) < \sigma^2(\tilde{X})$$

Ej. Ilustración eficiencia/sesgo

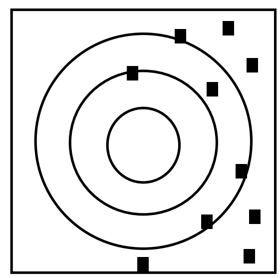
El centro de la diana representa el parámetro a estimar. Hacemos 10 lanzamientos que representan las estimaciones del parámetro en 10 muestras.



Insesgado
Eficiente



Sesgado
Eficiente



Sesgado
Ineficiente

¿Insesgado/ineficiente?

La estimación promedio sería correcta, pero, al tener mucha variabilidad, podría ofrecer estimaciones muy alejadas del valor poblacional

En un conjunto de estimadores insesgados, el de menor varianza, se dice que es el más eficiente. La eficiencia relativa de dos estimadores es definidas como: la razón de las varianzas de los dos estimadores.

Estimador Consistente

Al amentar *n*, aumenta la probabilidad de que el estadístico coincida exactamente con el valor del parámetro en las distintas muestras.

Un estimador consistente es un estimador que se aproxima al verdadero valor del parámetro conforme el tamaño de muestra se incrementa.

Estimador suficiente

Un estimador es eficiente si usa toda la información de la muestra.

La estimación puntual proporciona un "buen número" para representar un sistema o proceso de interés, sin embargo para un mejor entendimiento del proceso que se esta generando en la población también se requiere de una medida de variabilidad.

Por ejemplo el número de autos producidos por día en una fabrica es una medida importante, más sin embargo amplias variaciones hacia arriba o hacia abajo de la media puede resultar en excesivos costos de inventario o a perdidas de ventas.

Es por esto que necesitamos obtener mediciones de tendencia central y de variabilidad, números que contengan ambos, información y error.
La más importante medida de variabilidad son la varianza y la desviación estándar. La desviación estándar de variables aleatorias normales puede utilizarse junto con la media para calcular la probabilidad de que las observaciones estén dentro de intervalos designados.

INTERVALOS DE CONFIANZA PARA LA MEDIA.

Para realizar la estimación por intervalo para la media y la proporción utilizaremos intervalos que combinen mediadas de tendencia central y de variabilidad.

El entendimiento de los intervalos de confianza, puede ser desarrollado considerando los siguientes ejemplos:

Intervalo de confianza.

El intervalo de confianza al $(1-\alpha)$100% es calculado como:

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

El intervalo de confianza es construido de tal forma que $(1-\alpha)$100% de los intervalos incluyen a la media de la población.

$Z_{\alpha/2}$ Es el valor *z* positivo que establece la frontera de un área de $\alpha/2$ en la cola derecha de la distribución normal estándar.

Por ejemplo Deseamos estimar con intervalo de confianza, la tasa de producción por hora para un cierto cereal de trigo, la varianza de la taza de producción por hora es de 100. Una muestra aleatoria de 25 horas de tasas de producción es recolectada, obteniéndose una media muestral de 90 unidades por hora, a partir de esta información deseamos construir un intervalo de 95% de confianza.

La ecuación para este intervalo es $\bar{X} \pm Z_{0.025} \frac{\sigma}{\sqrt{n}}$

Sustituyendo los valores de este ejemplo encontramos que: $90 \pm 1.96 \frac{\sqrt{100}}{\sqrt{25}}$

$$90 \pm 3.92$$

El intervalo es desde 86.08 a 93.92 y podemos interpretarlo como:

Después de muchos intentos, el 95% de los intervalos de confianza construidos utilizando este procedimiento deberán contener la media de la población, tomando muchas muestras de tamaño 25 de esta población.

Cada intervalo es construido a partir una muestra aleatoria centrada en la media muestral $\bar{X}$. La media muestral puede diferir de muestra a muestra, y entonces los intervalos también ser diferentes.

En muchos casos el intervalo incluye la media μ, sin embargo, en el 5% de los caso el intervalo completo esta por arriba o por debajo de la media poblacional μ, recordemos que la media poblacional es siempre la misma, aunque se conozca su valor. Esto es los intervalos pueden variar de uno a otro.

Dependiendo del problema se pueden calcular intervalos de confianza con diferentes probabilidades de error, un intervalo del 99% no contiene a la media de la población en aproximadamente el 1% de los casos pero es más ancho que el intervalo del 95%, en contraste un intervalo al 90%, esquiva a la media de la población en el 10% de los casos siendo más angosto que el intervalo del 95%.

Los intervalos de confianza son útiles porque proporcionan una medida de tendencia central y una medida de la variabilidad.

Mostraremos ahora que el tamaño del intervalo puede ser cambiado al cambiar el riesgo o al cambiar el tamaño de la muestra.

Como los intervalos de confianza son diferentes para cada media muestral $\bar{X}$, ya que los intervalos de confianza están centrados sobre la media de la muestra, y a que la media de la población es siempre la misma, entonces podemos asegurar que el $(1-\alpha)$100% de los intervalos construidos usando una media de muestra $\bar{X}$, contienen a la verdadera media μ, es decir aproximadamente α100% de las muestras pueden no incluir a la media de la población.

Los intervalos de confianza son calculados utilizando la desviación estándar de la población σ y el tamaño *n* de la muestra. La desviación estándar de la media muestral es igual a la desviación estándar de la población dividida por la raíz cuadrada de el tamaño de la muestra tenemos que:

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

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Esta forma enfatiza que el ancho del intervalo de confianza varia inversamente con la raíz cuadrada del tamaño de muestra, por tanto un gran tamaño de muestra proporciona un intervalo de confianza angosto de gran precisión.

Ejemplo 1.- Se han producido varios cientos de marcadores. El gerente está interesado en la longitud de una línea continua que puede ser trazada por cada marcador antes de que se quede sin color, esto define la "vida útil" del marcador. El gerente sabe que la varianza de la producción es de 625 m². 50 marcadores son seleccionados aleatoriamente y probados, obteniéndose una línea continua trazada promedio de 300 m. Construya el intervalo de confianza del 95% para la "vida útil" de este lote de producción.

Solución:

$n = 50$

$\bar{x} = 300$

$\sigma^2 = 625$

$1 - \alpha = 0.95$

$\sigma = 25$

Como el intervalo debe estar centrado en el estimador puntual, debemos de dividir la confianza por dos, es decir $\frac{1-\alpha}{2} = \frac{0.95}{2} = 0.475$, con este valor vamos a la tabla de distribución normal para determinar el valor de $z_{\alpha/2}$, revisando la tabla para encontrar el valor más próximo a 0.475 obtenemos que: $z_{\alpha/2} = 1.96$

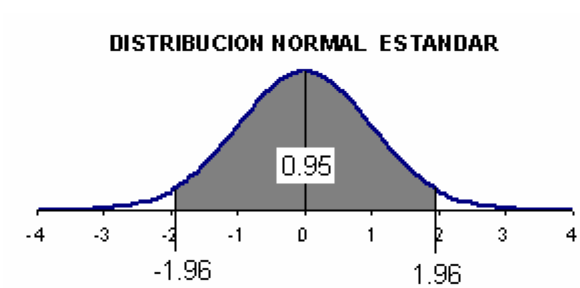
Calculando el error estándar:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{25}{\sqrt{50}} = 3.5355$$

Z	.00	.05	.06	.07	.08	.09
0.0	.00000	.01994	.02[illegible]92	.02790	.03188	.03586
1.7	.45543	.45994	.46080	.46164	.46246	.46327
1.8	.46407	.46784	.46856	.46926	.46995	.47062
1.9	.47128	.47441	.47500	.47558	.47615	.47670
2.0	.47725	.47982	.48030	.48077	.48124	.48169

Sustituyendo en: $\bar{x} - z_{\alpha/2}\sigma_{\bar{x}} \leq \mu \leq \bar{x} + z_{\alpha/2}\sigma_{\bar{x}}$

$$300 - 1.96(3.5355) \leq \mu \leq 300 + 1.96(3.5355)$$



$$300 - 6.9296 \le \mu \le 300 + 6.9695$$

$$293.0705 \le \mu \le 306.9695$$

De acuerdo a la muestra podemos tener la confianza del 95% de que, la “vida útil” media de los plumones se encuentra entre 203.07 m y 306.96 m.

Ejemplo 2.- Estamos interesados en la media de los galones de gasolina consumidos por familias que toman 2 semanas de vacaciones durante el mes de agosto. De una muestra de 180 familias se encontró que: $\overline{x} = 120$ galones y $s = 66$ galones

¿Podemos construir el intervalo de confianza del 90% para el número promedio de galones gastados por todos los viajeros durante estas dos semanas? ¿Por qué si o por qué no? Si es posible hágalo.

Solución:

A pesar de que no conocemos la forma de la distribución de consumo de gasolina y a que no conocemos la varianza de la población, podemos hacer uso de las estadísticas muestrales, ya que la muestra es grande y esto nos permite utilizar el TLC.

$n = 180$

$\overline{x} = 300$

$s = 60$

$1 - \alpha = 0.90$

Dividiendo el nivel de confianza por dos y buscando en la tabla de la distribución normal estándar, obtenemos que para $\frac{1-\alpha}{2} = \frac{0.90}{2} = 0.45$ da un valor de $z_{\alpha/2} = 1.64$

Calculando el error estándar:

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{66}{\sqrt{180}} = 4.9193$$

Z	.00	.03	.04	.05
0.0	.00000	.01197	.01595	.01994
1.4	.41924	.42364	.42507	.42647
1.5	.43319	.43699	.43822	.43943
1.6	.44520	.44845	.44950	.45053
1.7	.45543	.45818	.45907	.45994

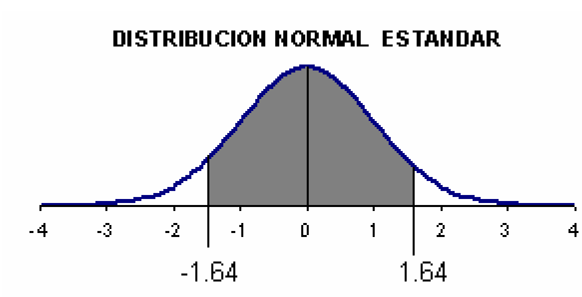
Sustituyendo en:

$$\overline{x} - z_{\alpha/2}\sigma_{\overline{x}} \le \mu \le \overline{x} + z_{\alpha/2}\sigma_{\overline{x}}$$

$$120 - 1.64(4.9193) \le \mu \le 120 + 1.64(4.9193)$$

$$120 - 8.0676 \le \mu \le 120 + 8.0676$$

$$111.9324 \le \mu \le 128.0676$$



De acuerdo a la muestra podemos tener la confianza del 90% de que, el consumo promedio familiar de gasolina en esta dos semanas se encuentra entre 111.93 galones y 128.06 galones.

TAMAÑO DE MUESTRA

Hasta el momento se ha hablado de hacer estimaciones de intervalo para la media, suponiendo que tenemos datos conocidos de una muestra, pero que pasa si no tenemos aún la muestra, ¿Cómo sabemos cuántos elementos debemos seleccionar de la población?. Dado que el margen de error de estimación esta dado por:

$$\varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

podemos obtener una expresión para n despejando de la expresión anterior, por lo que:

$$n = \left(\frac{z_{\alpha/2}\sigma}{\varepsilon}\right)^2$$

El tamaño de muestra debe ser un número entero, pero en ocasiones esta expresión puede producir valores no enteros, para este caso debemos redondear al siguiente entero para asegurar que es del tamaño requerido.

Ejemplo 1.- Una psicóloga ha ideado una nueva prueba de percepción espacial, y quiere estimar el puntaje medio que alcanzan los pilotos del sexo masculino. ¿Cuántas personas deberá probar si quiere que la media de muestra tenga un error de no más de 2 puntos, con una confianza del 95% . De un estudio anterior se sabe que $\sigma = 12.1$.

Para una confianza del 95% sabemos que: $z_{\alpha/2} = 1.96$

Sustituyendo valores:

$\sigma = 12.1$

$\varepsilon = 2$

$1-\alpha = 0.95$

$$n = \left(\frac{1.96(12.1)}{2}\right)^2 = 140.61$$

Dado que el resultado no es entero lo redondeamos al siguiente entero, por que el tamaño adecuado para alcanzar esta confianza y máximo error es de141 o más.

Ejemplo 2.- Se va a realizar una encuesta en un sector del área metropolitana para determinar el ingreso familiar promedio de ese sector. Se desea hacer una estimación de la media con un valor que se encuentre a $150 de la media verdadera con un nivel de confianza del 99%. Se sabe de estudios anteriores que la desviación típica de la población es de $ 3000.

Para una confianza del 9% sabemos que: $z_{\alpha/2} = 2.58$

$\sigma = 3000$ $\quad n = \left(\frac{2.58(3000)}{150}\right)^2 = (51.6)^2$ =2652.56

$\varepsilon = 150$

$1-\alpha = 0.99$ Redondeando al siguiente entero positivo, esto nos indica que el número buscado es 2653 o más elementos de l a población par lograr los objetivos planteados

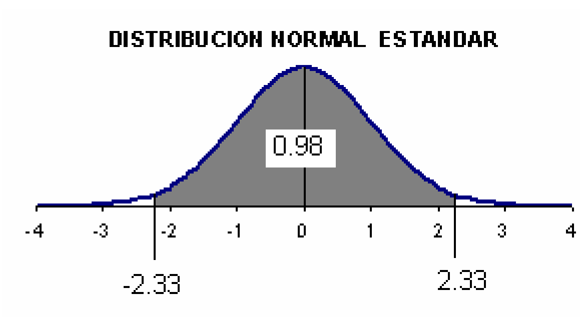
ESTIMACIÓN POR INTERVALO PARA LA PROPORCION.

Al realizar una investigación de mercado para FMC, se averigua que una muestra aleatoria de 1220 hogares contiene 1054 en los que se posee un vehículo. Con base en este resultado, construir un intervalo de confianza del 98% para el porcentaje de todos los hogares en los que se posee un automóvil.

Dado que la población es grande podemos aplicar el Teorema Central del Límite.

Calculamos la proporción muestral de hogares que poseen automóvil;

$$p = \frac{1054}{1220} = 0.8639$$



Calculemos ahora la desviación estándar de la distribución de proporciones como:

$$\sigma_p = \sqrt{\frac{0.8639(1-0.8639)}{1220}} = 0.0098$$

Para una confianza del 98% $z_{0.01} = 2.33$

El intervalo de confianza al 98% es:

$$0.8639 - 2.33(0.0098) \leq \pi \leq 0.8639 + 2.33(0.0098)$$

$$0.8639 - 0.0228 \leq \pi \leq 0.8639 + 0.0228$$

$$0.8411 \leq \pi \leq 0.8867$$

Esperamos que la proporción de hogares que poseen automóvil se encuentre entre 84.11% y 88.67%, con una confianza del 98%.

Una pequeña firma compro un conjunto de 650 partes electrónicas a una gran compañía, y al revisar en una muestra de 80 partes comprobó que 10 de ellas estaban defectuosas. Estime la proporción partes defectuosas para todo el conjunto empleando un intervalo de confianza del a) 90%, b) 95%

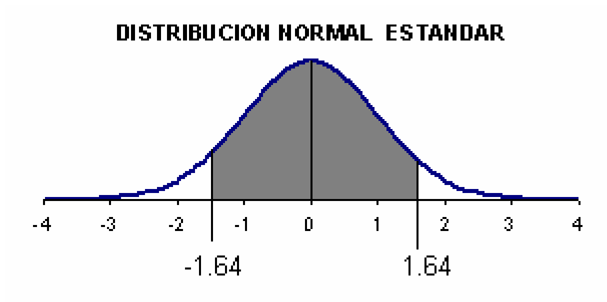
Sol.

Calculemos la proporción muestral de partes defectuosas:

$$p = \frac{10}{80} = 0.125$$

debido la población es finita calculamos la desviación estándar de la distribución de proporciones como:

$$\sigma_p = \sqrt{\frac{0.125(1-0.125)}{80}}\sqrt{\frac{650-80}{650-1}} = 0.0346$$



para una confianza del 90% $z_{0.05} = 1.64$

a) el intervalo de confianza al 90% es:

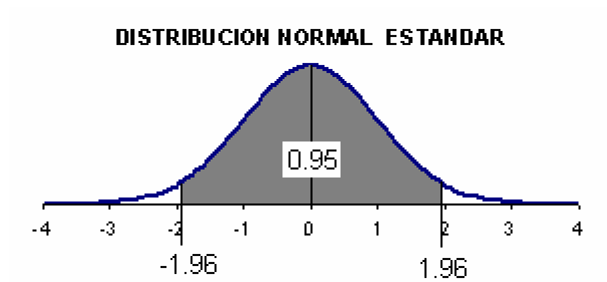
$$0.125 - 1.64(.0346) \le \pi \le 0.125 + 1.64(.0346)$$

$$0.125 - .0567 \le \pi \le 0.125 + .0567$$

$$0.0683 \le \pi \le 0.1817$$

Se tiene una confianza del 90% de que el porcentaje de defectuosos se pueda encontrar dentro del intervalo del 6.83% al 18.17%.

b) para una confianza del 90% $z_{0.25} = 1.96$



el intervalo de confianza al 95% es:

$$0.125 - 1.96(.0346) \le \pi \le 0.125 + 1.96(.0346)$$

$$0.125 - 0.0678 \le \pi \le 0.125 + 0.0678.$$

$$0.0572 \le \pi \le 0.1928$$

Se tiene una confianza del 95% de que el porcentaje de defectuosos se pueda encontrar dentro del intervalo del 5.72% al 19.28%.

TAMAÑO DE MUESTRA PARA ESTIMAR LA PROPORCIÓN

Ahora describiremos el procedimiento para determinar lo grande que debe ser una muestra cuando se desea encontrar el valor aproximado de la proporción de una población. Sabemos que el margen de error de estimación esta dado por:

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

despejando n para obtener el tamaño de muestra:

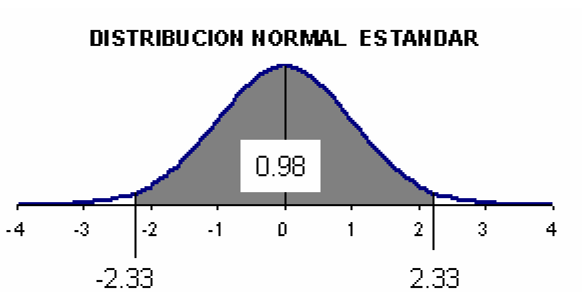
$$n = \frac{z_{\alpha/2}^2\,\hat{p}\hat{q}}{\varepsilon^2}$$

Con esta expresión se debe tener un estimado preeliminar de $\hat{p}$, el cual se puede obtener por medio de un muestreo piloto o con datos históricos o un valor proporcionado por un experto.

En caso de no conocer este valor se debe tomar el valor de 0.5 para $\hat{p}$, que es el valor que nos daría la mejor aproximación al tamaño de muestra buscado, por lo que el tamaño de muestra es:

$$n = \left(\frac{z_{\alpha/2}}{2\varepsilon}\right)^2$$

Ejemplo 1.- Una firma publicitaria afirma que su reciente campaña publicitaria llego al 27% de las familias de cierta zona de la ciudad. La cía. que contrato a esta firma tienen dudas acerca de esta afirmación y desea hacer una encuesta para aclarar la situación. ¿De qué tamaño debe tomar la muestra para tener un 98% de confianza y su estimación no tenga un error mayor a 1.5%?



$\hat{p} = 0.27$
$\in = 625$
$1-\alpha = 0.98$
$z_{.01} = 2.33$

Tenemos que:

$$n = \frac{z_{\alpha/2}\,\hat{p}(1-\hat{p})}{\varepsilon^2}$$

$$n = \frac{(2.33)^2(0.27)(1-0.27)}{(0.015)^2}$$

$$n = 4755.71 \approx 4756$$

El número mínimo de muestra para cumplir con las condiciones establecidas es 4756 o más grande.

PROBLEMAS PROPUESTOS

I.- Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completa la frase:

1) La estimación de tipo puntual es más recomendable que una por intervalo. V F

2) Un intervalo de confianza se construye con el estimador y su desviación típica. V F

3) La formula para obtener un intervalo de confianza es:______________________

4) Un estimador es insesgado cuando su valor esperado coincide con el parámetro poblacional. V F

5) Los errores típicos de estimación se pueden obtener de: ____________________

6) Un estimador es más eficiente que otro cuando su desviación típica es mayor. V F

7) La probabilidad asociada a un intervalo de confianza se relaciona con __________

8) El valor del coeficiente de confianza se calcula con Z y la tabla de distribución normal. V F

9) El intervalo de confianza para estimar la proporción poblaciona l es: _____________

10) A mayor nivel de confianza, mayor ancho de intervalo. V F

11) A mayor tamaño de muestra, mayor ancho de intervalo. V F

12) Un estimador es consistente cuando a medida que aumenta el tamaño de muestra, se aleja del parámetro poblacional. V F

13) La media es un estimador de la desviación estándar de la población. V F

14) El número de elementos de una muestra no depende del nivel de confianza, V F

15) La estimación puntual tiene una alta probabilidad de ser igual al parámetro poblacional, V F

1.- El colesterol de una paciente medido 30 veces dio una media de 256 mg/dl con una desviación estándar de 32 mg/dl. Encontrar el intervalo del 97%.

2.- Mediciones de presión sanguínea de 35 mujeres de edad avanzada tienen una media de 140 mm. de mercurio. Si estos datos se pueden considerar como una muestra tomada al azar de una población normal con $\sigma = 10$ mm. de mercurio, construya un intervalo de confianza del 98% de la media de la población μ.

3.- Durante varios años, se había aplicado una prueba de nivel de matemáticas a todos los alumnos de primer ingreso de cierta universidad. Si 64 estudiantes, seleccionados al azar en este periodo tardaron en promedio 28.5 minutos en resolver la prueba con una varianza de 9.3, construya un intervalo de confianza del 99% del tiempo promedio verdadero que tarda un alumno de primer ingreso en resolver el problema.

4.- La longitud de los cráneos de 10 esqueletos fósiles de una especie de ave extinta tiene una media de 5.68 y una desviación estándar de $s = 0.29$ mm. Suponiendo que estas mediciones están normalmente distribuidas, obtenga el intervalo de confianza del 95% para la longitud media de los cráneos de esta especie de aves.

5.- Un especialista en genética está interesado en la proporción de hombres africanos que presentan un desorden sanguíneo leve. En una muestra aleatoria de 100 de ellos, se encontró que 24 presentaban dicho desorden. Calcule el intervalo de confianza del 99% para la proporción de hombres africanos que tienen este desorden sanguíneo.

6.- Un experto en eficiencia desea determinar el tiempo promedio que tarda el personal de un foso de reparaciones en cambiar un conjunto de 4 neumáticos a un auto de carrera. Determinar el tamaño de muestra requerido para poder afirmar, con el 95% de confianza, que la media de muestra difiere de la media real en cuando mucho dos segundos. Por estudios realizados antes se sabe que la desviación estándar de la población es de 12 segundos.

7.- Un general desea estimar la aptitud física promedio (medida a través de cierta prueba) de miles de soldados que tiene a su cargo, sobre la base de una muestra aleatoria de ellos. El general desea que tal estimación tenga un error de cuando mucho 2 puntos de la prueba, con una confianza mínima del 99%. Si sabe por experiencia que el valor de la desviación estándar es de 15, ¿Cuál es el tamaño mínimo de la muestra de soldados a quienes debe aplicar la prueba?

8.- Una nutrióloga estima, sobre la base de análisis previos, que la desviación estándar del contenido de proteínas por cada lata de atún de cierta marca es de aproximadamente 3.2 gr. ¿Qué tan grande debe ser el tamaño de la muestra de latas que debe analizar para que el error en la estimación del parámetro que desconoce (μ) sea de cuando más 1.5 gr. Con una confianza mínima de a) 95%, b) 99%.

9.- Cierto porcentaje de estudiantes de una universidad considera que hay que cambiar el diseño de las evaluaciones de profesores, debido a que el formato actual las ha convertido en concursos de popularidad y además se presta a venganzas recíprocas.
Suponga que se lleva a cabo una pequeña encuesta piloto en las cafeterías y se observo que el 30% de los encuestados manifestaron estar a favor de una modificación en el diseño de las evaluaciones a docentes. Determine el tamaño de la muestra de estudiantes que se deben encuestar, para tener una confianza del 95% de que el estadístico $\hat{p}$ estime al parámetro P con un margen de error de cuando mucho 3%.

10.- Suponga que en el ejercicio anterior no se realiza ninguna encuesta piloto, y se desea calcular directamente el tamaño de la muestra bajo las mismas condiciones.

11.- Un grupo de cirujanos dentistas, desea averiguar el porcentaje de adolescentes que requieren trabajos de ortodoncia. Determine el tamaño de muestra de adolescentes que deben examinarse, con objeto de que el porcentaje de esa muestra sea representativo del porcentaje verdadero de toda la población adolescente, con un margen de error de ± 4% y un nivel de confianza del 94%.

12.- En un plebiscito realizado entre los habitantes de la ciudad de México, se realiza una encuesta cuyo objetivo es determinar la proporción de habitantes que están a favor de que se construya un segundo piso en una importante avenida de la ciudad. ¿Qué tan grande debe ser la muestra de personas que respondan a esa encuesta si se desea que el máximo error en la estimación sea igual a; a) 3% con 96% de confianza?, b) 0.02 con 95% de confianza?, c) 5% con 90% de confianza?

13.- Un sondeo efectuado con 400 familias de cierta clase social de una ciudad reveló un gasto promedio de 274 pesos en productos de tocador, con una desviación estándar de 45 pesos. Determine los intervalos de confianza del 90%, 95% y 99% para estimar el gasto promedio mensual en productos de tocador,

14.- Si el rendimiento promedio de 80 automóviles de la marca “K”, es de 10.9 km. por litro con una desviación estándar de 4.5 km. Determine los intervalos de estimación del 92%, 96% y 98% para estimar el rendimiento promedio de todos los automóviles de la marca ”K”.

15.- En una muestra aleatoria de 1000 casas en una determinada ciudad, se encuentra que 228 de ellas tienen aire acondicionado. Encuentre los intervalos de confianza del 90%, 96% y 99% para la proporción de hogares de esta ciudad con aire acondicionado.

16.- Se selecciona una muestra aleatoria de 300 ciudadanos y 124 de ellos están de acuerdo con la Reforma Fiscal. Construya los intervalos de confianza del 88%, 95% y 99.5% para estimar la proporción de ciudadanos que están de acuerdo con la Reforma Fiscal.