PROBABILIDAD y ESTADÍSTICA II

UNIDAD Nº3

Licenciatura en Enseñanza de la Matemática
Año 2011
Mg. Lucía C. Sacco

UNIDAD N°3 Estimaciones de parámetros.

Procedimientos de la Estadística inferencial: la estimación. Estimación de parámetros. Propiedades deseables de un estimador: insesgado, eficiente o con varianza mínima, coherencia y suficiencia. Interpretación gráfica de propiedades. Valor estimado de una parámetro.

Estimación puntual. Métodos de estimación puntual.

Estimación por intervalos. Intervalos de confianza. Intervalos de confianza para las medias con varianza conocida.

Propósitos:

Brindar oportunidades para la construcción de herramientas que permitan:

- Comprender los fundamentos teóricos y la lógica subyacente de la inferencia estadística en una de sus grandes ramas: la estimación de parámetros.
- Diferenciar las formas de estimación de parámetros poblacionales teniendo en cuenta las condiciones de los buenos estimadores y reconociendo las particularidades de cálculo en distintos casos -media, varianza y proporción poblacional-.
- Aplicar conceptos y procedimientos centrales de la inferencia estadística en la resolución de casos y problemas.

INFERENCIA ESTADÍSTICA

Los dos principales procedimientos de la estadística inferencial son la estimación (puntual o por intervalos) y las pruebas de hipótesis. Comúnmente, los parámetros de una población son desconocidos, siendo necesario estimar el valor de éstos o, si no, efectuar indagaciones (pruebas de hipótesis) para comprobar si los valores a ellos atribuidos pueden ser considerados como verdaderos.

En esta Unidad veremos, entre otras cosas, cómo es posible obtener, para cada parámetro de interés, el **mejor estimador** de ese parámetro.

Estimación de parámetros — **INFERENCIA ESTADÍSTICA** — *Prueba de hipótesis estadísticas*

UNIDAD Nº3 **UNIDAD Nº4**

INFERENCIA ESTADÍSTICA

Estimar o predecir el valor del parámetro

Tomar la decisión acerca del valor de un parámetro, con base en alguna idea preconcebida acerca de cuál podría ser su valor

Los procedimientos estadísticos son importantes porque dan dos tipos de información:

- **métodos para hacer la inferencia**
- **una medida numérica de la bondad o confiabilidad de la inferencia.**

Población y muestra

El procedimiento que generalmente se sigue en cualquier investigación consiste en obtener resultados a partir de una muestra y luego generalizarlos a la población objetivo.

Una población cualquiera queda perfectamente especificada por ciertas medidas denominadas **parámetros poblacionales**.

Un parámetro poblacional es una medida que se calcula teniendo en cuenta todos los elementos que componen una cierta población.

Por ejemplo, si el ingreso promedio de las familias de la ciudad de San Nicolás se calcula considerando el ingreso de todas las familias que habitan en la ciudad, este ingreso promedio es un parámetro poblacional.

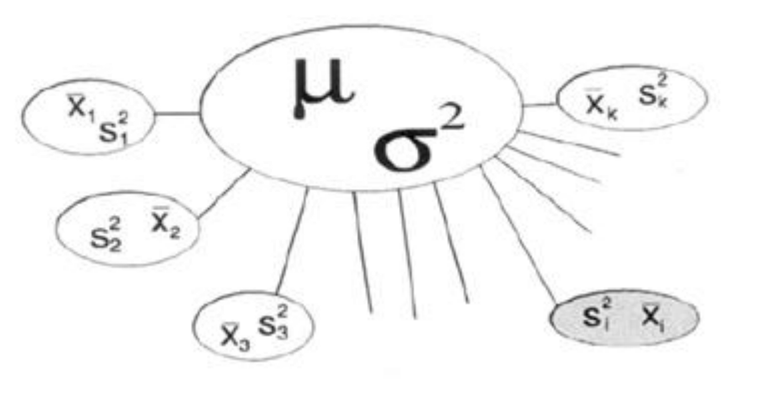
Es evidente que los parámetros poblacionales son generalmente imposibles de calcular. En la práctica, casi siempre se trabaja con muestras.

Las medidas calculadas a partir de las observaciones muestrales, se conocen con el nombre de **estadísticos muestrales**.

Un estadístico muestral es una medida que se calcula teniendo en cuenta solamente los elementos que integran una muestra determinada.

Estimación de parámetros

El objetivo de la estadística consiste en hacer inferencias acerca de los parámetros de una población teniendo en cuenta la información contenida en la muestra.



Ahora bien, como en general los parámetros poblacionales son desconocidos, existe una amplia gama de técnicas estadísticas que tienen como objetivo la estimación de estos parámetros a través de estadísticos muestrales adecuados a cada caso en particular.

Estimación puntual

Estimación por intervalo

Estimadores

Un estimador es una función que permite calcular valores aproximados al del parámetro, además, se lo considera una variable aleatoria ya que puede tomar valores que pertenecen a un intervalo de números reales

Es posible definir muchos estadísticos para estimar un parámetro desconocido. Por ejemplo, puede elegirse la **media muestral** para estimar el valor de la **media poblacional**, o también la **mediana muestral**.

Por ejemplo, la media muestral se obtiene sumando todas las observaciones de la muestra y dividiendo esta suma por el tamaño de la muestra.

Cualquier persona podría definir otra combinación de las observaciones muestrales como estimador del parámetro μ y entonces cabría la pregunta:

¿Cuál es el "mejor" estimador de μ?

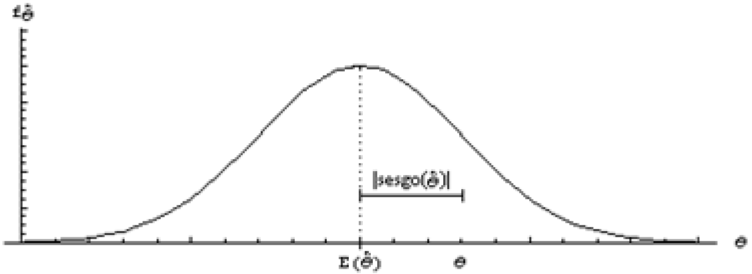
Un problema importante que debió resolver la teoría estadística, fue el de determinar el mejor estimador de cada parámetro en particular.

Propiedades de los estimadores

1. La distribución muestral de $\hat{\theta}$ debe tener una media igual al parámetro estimado θ, en este caso se dice que el **estimador es insesgado**.

   Si se usa la media muestral $\overline{X}$ para estimar la media poblacional μ, se sabe que la $\mu_{\overline{X}} = \mu$, por lo tanto la media es un estimador insesgado.

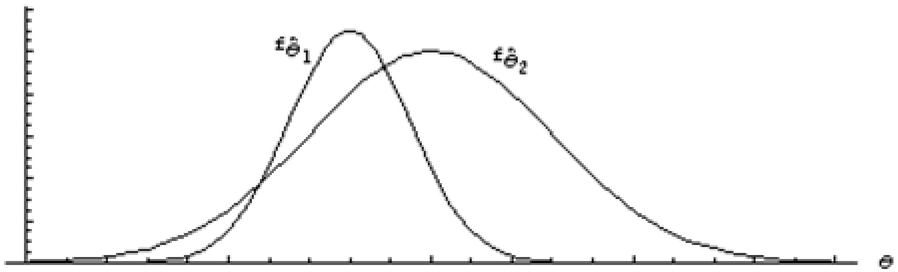
En este gráfico podemos apreciar que es deseable que el valor esperado del estimador coincida con el parámetro estimado. Denominamos sesgo a la diferencia $E(\hat{\theta}) - \theta$ Por eso cuando el sesgo de un estimador es cero, se lo denomina "insesgado".



Propiedades de los estimadores

2. La varianza del estimador debe ser la menor posible, en este caso se dice que el **estimador es eficiente o con varianza mínima**.

Otra característica importante que analizamos fue la varianza. Es deseable que la varianza de un estimador sea pequeña, para que la variabilidad respecto de su valor esperado sea pequeña.



En el ejemplo graficado, la varianza de $\hat{\theta}_1$ es más pequeña que la de $\hat{\theta}_2$ Vemos que su variabilidad respecto de su valor esperado es menor.

Propiedades de los estimadores

- **Coherencia:** Una estadística es un estimador coherente de un parámetro de población, si al aumentar el tamaño de la muestra se tiene casi la certeza de que el valor de la estadística se aproxima bastante al valor del parámetro de la población. Si un estimador es coherente se vuelve más confiable si tenemos tamaños de muestras más grandes.

- **Suficiencia:** Un estimador es suficiente si utiliza una cantidad de la información contenida de la muestra que ningún otro estimador podría extraer información adicional de la muestra sobre el parámetro de la población que se está estimando. Es decir se pretende que al extraer la muestra el estadístico calculado contenga toda la información de esa muestra. Por ejemplo, cuando se calcula la media de la muestra, se necesitan todos los datos. Cuando se calcula la mediana de una muestra sólo se utiliza a un dato o a dos. Esto es solo el dato o los datos del centro son los que van a representar la muestra. Con esto se deduce que si utilizamos a todos los datos de la muestra como es en el caso de la media, la varianza, desviación estándar, etc.; se tendrá un estimador suficiente.

Propiedades de los estimadores

Ejemplo 1

Consideremos una población compuesta por 5 escuelas rurales en las que se ha registrado el número de maestros obteniendo: 2, 3, 6, 8, 11.

Analizar cómo el estimador media muestral cumple con todas las propiedades de un buen estimador.

- **Insesgabilidad**
- **Insesgabilidad de mínima varianza (eficiencia)**
- **Consistencia**
- **Distribución asintóticamente normal**

UNIDAD Nº3

Estimación puntual

La estimación puntual es un proceso mediante el cual se estima el parámetro en un punto, dando un valor específico como estimación.

Los dos métodos tradicionales de estimación puntual de parámetros son el de **mínimos cuadrados** y el de **máxima verosimilitud**.

El método de mínimos cuadrados fue trabajado en la Unidad Nº2 cuando hablamos del análisis de regresión.

El objetivo de la estimación puntual es seleccionar sólo un número, basados en datos de la muestra, que represente el valor más razonable de θ

Una **estimación puntual** de un parámetro θ es un sólo número que se puede considerar como el valor más razonable de θ. La estimación puntual se obtiene al seleccionar un estadístico apropiado y calcular su valor a partir de datos de la muestra dada. El estadístico seleccionado se llama **estimador puntual** de θ.

Estimación puntual

Ejemplo 2

La longitud de los tornillos que produce una determinada máquina es una variable normal, pero no sabemos cuánto vale el parámetro (media poblacional) μ de esa distribución normal.

Podemos hacer el experimento de tomar 10 tornillos, calcular el promedio de sus longitudes, y usar ese promedio como estimación de μ .

Ejemplo 3

En el futuro habrá cada vez más interés en desarrollar aleaciones de Mg de bajo costo, para varios procesos de fundición. En consecuencia, es importante contar con métodos prácticos para determinar varias propiedades mecánicas de esas aleaciones. Examine la siguiente muestra de mediciones del módulo de elasticidad obtenidas de un proceso de fundición a presión 44.2 43.9 44.7 44.2 44.0 43.8 44.6 43.1

Suponga que esas observaciones son el resultado de una muestra aleatoria. Se desea estimar la varianza poblacional σ^2

Estimación por intervalos

Una estimación por intervalos de un parámetro desconocido θ es un intervalo de la forma $I \leq \theta \leq u$, donde los puntos extremo l y u dependen del valor numérico de la estadística $\hat{\theta}$ para una muestra en particular y de la distribución de muestreo de $\hat{\theta}$.

De la distribución de muestreo de $\hat{\theta}$ es posible determinar los valores de l y u tales que la siguiente proposición sea verdadera:

$$P(I \leq \theta \leq u) = 1 - \alpha \qquad ; \qquad 0 < \alpha < 1$$

Por tanto se tiene una probabilidad de $1-\alpha$ de seleccionar una muestra que produzca un intervalo que contiene el valor verdadero de θ El intervalo resultante $I \leq \theta \leq u$ se conoce como **Intervalo de Confianza del** $100(1-\alpha)$ **por ciento.**

Las cantidades l y u se denominan límites de confianza inferior y superior y $1-\alpha$ es el **coeficiente de confianza** (o ***nivel de confianza***, que es una medida del grado de fiabilidad en el intervalo).

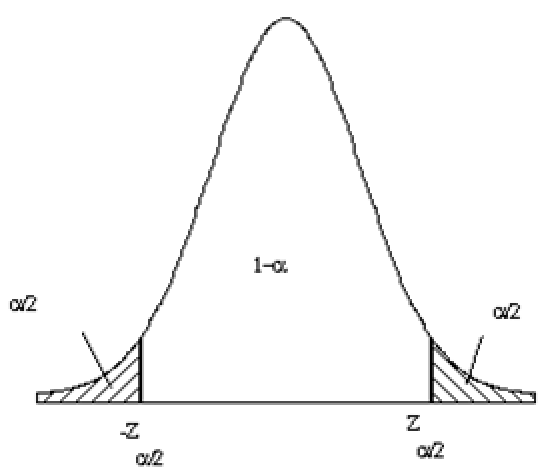
Un intervalo del tipo $I \leq \theta \leq u$ recibe el nombre más apropiado de **Intervalo de Confianza Bilateral**. También existen intervalos de confianza Unilaterales $\theta \geq I$ y $\theta \leq u$ donde los límites de confianza se eligen de modo que $P(\theta \geq I) = 1-\alpha$ y $P(\theta \leq u) = 1-\alpha$

Intervalos de confianza para las medias

Para estimar la media μ de una característica de la población, es necesario, primero saber si la varianza de la población es conocida o no lo es.

Para estimar la media μ de una característica de la población, cuando se considera conocida la varianza de esa población, se toma una muestra de tamaño "n" y se le calcula la media muestral $\overline{X}$

Por el Teorema del Límite Central, se sabe que de la distribución de la media muestral se obtiene que $Z = \dfrac{\overline{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ tenga una distribución como una normal estándar, con media $E(\overline{X}) = \mu$ y varianza $\dfrac{\sigma^2}{n}$.



Teniendo en cuenta que la distribución de la media muestral sigue o tiende a una Distribución Normal, y considerando la varianza como conocida, el intervalo de confianza debe abarcar el área de $(1-\alpha)\%$ entre sus límites superior e inferior en dicha distribución. Cada límite es expresado en unidades de desviación típica y esas unidades esta expresadas por $Z_{\alpha/2}$. En consecuencia, el intervalo de confianza bilateral del 100 $(1-\alpha)\%$ para μ dado esta dado por $I = \left[\overline{X} - Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} ; \overline{X} - Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right]$

Estimación por intervalos

Ejemplo 4

Supongamos que el director de investigaciones de mercado de una fábrica automotriz necesita hacer una estimación de la vida promedio de las baterías que su compañía produce. Selecciona aleatoriamente 200 usuarios y resulta tener una vida promedio de sus baterías de 36 años.

Encontrar el intervalo dentro del cual es probable que esté la media de población desconocida.

Ejemplo 5

Un vendedor de partes del automotor, mayorista, necesita una estimación de la vida media que puede esperar de los limpiadores de parabrisas en condiciones normales de manejo. La administradora ya ha determinado que la desviación estándar de la vida útil de la población es de 6 meses. Se selecciona una sola muestra de 100 limpiadores y se obtienen que $\bar{x} = 21 meses$, encontrar el intervalo de confianza de un 95%

Ejemplo 6

Se quiere estimar el ingreso medio anual de 700 familias que vive en una sección de 4 manzanas. Si se toma una muestra de 50 familias y se hallan los siguientes resultados $\bar{x} = \$11800$ y $s = \$950$ Encontrar un intervalo con un nivel de confianza del 90% en el que pueda encontrarse la media poblacional.

Conclusiones

Como síntesis, podemos decir que para reducir la amplitud de un intervalo de confianza y en consecuencia aumentar su precisión, debemos reducir el error estándar de la media muestral $\bar{x}$ que es $\sigma/_n$. *Esto puede lograrse solamente disminuyendo la variabilidad de los datos ya sea homogeneizando el material experimental o, si esto no puede llevarse a cabo, aumentando el tamaño de la muestra.*

Es costumbre utilizar coeficientes de confianza del 90%, 95% y 99% Por este motivo es posible considerar la siguiente tabla que resume los valores de probabilidad de la distribución normal estandarizada para estos niveles de confianza.

Coeficiente de confianza	z
0,90	1,645
0,95	1,960
0,99	2,576