Probabilidad y Estadística Año de cursada 1° AÑO

Clase Nº 9: Estimación de parámetro

Contenido: *Parámetros Poblacionales, Estadísticos Muestrales, Media Muestral y Varianza muestral. Estimación puntual y estimación por intervalos. T de Student*

Presentación:

Bienvenidos y bienvenidas a nuestra clase n° 8 de Probabilidad y Estadística. Esta semana, entramos dentro del campo de la inferencia estadística desde la

teoría de muestreo.
En las clases anteriores, desarrollamos varios modelos de distribución de probabilidades que involucra conceptos que han sido desarrollados en el tránsito por esta materia, y nos permitió calcular e interpretar probabilidades de algunos fenómenos en situaciones particulares.

En esta oportunidad, aplicaremos dichos modelos para la caracterización de poblaciones por medio de selección de muestras considerando varios parámetros que la caractericen para extrapolar dichos resultados a la población completa.

Una estimación puntual de un parámetro es un número que puede ser considerado como un valor sensible de la muestra. Se obtiene una estimación puntual seleccionando un estadístico apropiado y calculando su valor con los datos muestrales dados. El estadístico seleccionado se llama estimador puntual de la muestra y se lo suele representar con la letra griega θ.

1. Desarrollo:

Parámetros Poblacionales

Cuando desarrollamos un proceso de investigación el objetivo consiste en obtener resultados a partir de una muestra y luego generalizarlos. Entonces, diremos que la población es conocida si podemos determinar la distribución de probabilidad $f(x)$ de la variable aleatoria asociada X. Por ejemplo, si X es una variable aleatoria cuyos valores son las estaturas de 15000 estudiantes entonces X tiene una distribución de probabilidad$f(x)$.

Si, por ejemplo, X está normalmente distribuida decimos que la población está normalmente distribuida o que tenemos una *población normal.* En forma semejante, si X está binomialmente distribuida, decimos que la población está binomialmente distribuida o *población binomial.*

Existirán ciertos valores que aparecen en $f(x)$, como μ o σ en el caso de la distribución normal o ρ y *n* en el caso de la distribución binomial. Otras cantidades tales como la mediana, momentos, sesgos, entre otros, pueden determinarse en términos de estos. Todas estas cantidades se conocen como ***Parámetros Poblacionales***.

Estadísticos muestrales.

Como hemos mencionado en las primeras clases, la población generalmente resulta inaccesible por varios motivos. Entonces, tomar muestras aleatorias de la población y emplearlas para obtener valores, se torna una herramienta útil que sirven para estimar los parámetros poblacionales.

Retomando el ejemplo antes mencionado:

Ejemplo 1:

Deseamos extraer conclusiones respecto a las estaturas de 15000 estudiantes (población) examinando solo 150 estudiantes (muestra) seleccionados de esta población.

En este caso X es una variable aleatoria cuyos valores son las diferentes estaturas. Para obtener una muestra de tamaño 150 debemos primero escoger un individuo aleatoriamente de la población, para este individuo la variable puede asumir cualquier valor, por ejemplo x_1, de las diferentes estaturas posibles y llamaríamos x_1 el valor de la variable X_1, donde el subíndice 1 se emplea ya que corresponde al primer individuo seleccionado. De la misma manera podemos escoger el segundo individuo para la muestra obteniendo así el valor x_2 y repetir el proceso hasta obtener x_{150} puesto que el tamaño de la muestra es 150.

Cualquier cantidad obtenida de una muestra con el fin de estimar un parámetro poblacional se lo llama *Estadístico muestral*. Estos estadísticos pueden definirse como una función de variables aleatorias $X_1, X_2, \dots, X_n$ donde dicha función $g(X_1, \dots, X_n)$ es otra variable aleatoria.

Generalmente, por cada Parámetro poblacional, hay un estadístico muestral para caracterizarlo. Comúnmente el método para el cálculo de un estadístico muestral es el mismo que el proceso para obtener el parámetro para una población finita, dado que una muestra es un conjunto finito de individuos.

Veamos algunas distribuciones en el muestreo:

- Distribución de la media muestral: Es una distribución de probabilidades de todas la muestras de un determinado tamaño de la población.

  Ejemplo: De una población de 4 elementos, se desea obtener las medias muestrales de cada par de elementos sin repetición.

Elementos	A	B	C	D
X_i	6	6	8	12

Obtenemos entonces todas las posibles muestras

		Media
A	B	6
A	C	7
A	D	9
B	C	7
B	D	9
C	D	10

$\bar{X}$	f_i	p
6	1	0,1667
7	2	0,3333
9	2	0,3333
10	1	0,1667

- Distribución de la proporción muestral:

  Para el cálculo de esta distribución se procede de la misma manera que con la media muestral pero con el estadístico proporción $p = \frac{x}{n}$ donde x es el número de observaciones exitosas o de interés mientras que n es el tamaño de la muestra.

  Ejemplo: Continuando con el ejemplo anterior las observaciones de interés son los elementos A y B por lo que tenemos

		Proporci
A	B	1,0
A	C	0,5
A	D	0,5
B	C	0,5
B	D	0,5
C	D	0

p_i	f_i	p
0	1	0,1667
0,5	4	0,6667
1	1	0,1667

Elementos	A	B	C	D

- Distribución de la varianza muestral
  Se procede de la misma forma que en la media muestral, pero en este caso se calcula la varianza. Al calcular la media de la varianza muestral $E(s^2)$ se obtiene la varianza de la población. Por otro lado la varianza de la varianza muestral $\sigma^2(s^2)$ tiende a cero cuando N se hace muy grande.

Cuando se toma solo un valor como estadístico muestral se denomina **estimación puntual.**

Como corolario de la estimación puntual podemos decir que si bien obtenemos un estadístico muestral que caracteriza a la población, éste no nos dice si es confiable dicho estadístico ya que se basa en un solo punto.

Propiedades de los estadísticos muestrales:

- **Insesgabilidad:** Se considera insesgado cuando la esperanza

  matemática de su distribución en el muestreo coincide con el valor del estadístico.

- **Eficiencia:** Es eficiente cuando su varianza es mínima.
- **Consistencia:** El estadístico muestral se acerca al estadístico poblacional a medida que aumenta el tamaño de la muestra, cuando el tamaño de la muestra es igual al parámetro poblaciónal, el estadístico muestral es insesgado y su varianza es cero.
- **Suficiencia:** El estadístico es suficiente cuando incluye tida la información relevante de la muestra de manera que ningún otro estimador brinde información adicional de la misma.
- **Robustez:** Es robusto si al vulnerar algunos de los supuestos en los que se basa el proceso de estimación, el resultado no cambia de manera significativa y los resultados siguen siendo fiables.

Los estadísticos principales son los siguientes:

- Media muestral:

Denotese $X_1, \dots, X_n$ las variables aleatorias para una muestra de tamaño n , entonces la media muestral se define por:

$$\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$$ donde $\bar{X} = \mu$

- Varianza muestral:

  Si $X_1, X_2, \dots, X_n$ se refiere a las variables aleatorias de una muestra de tamaño n, entonces la varianza muestral se calcula de la siguiente manera:

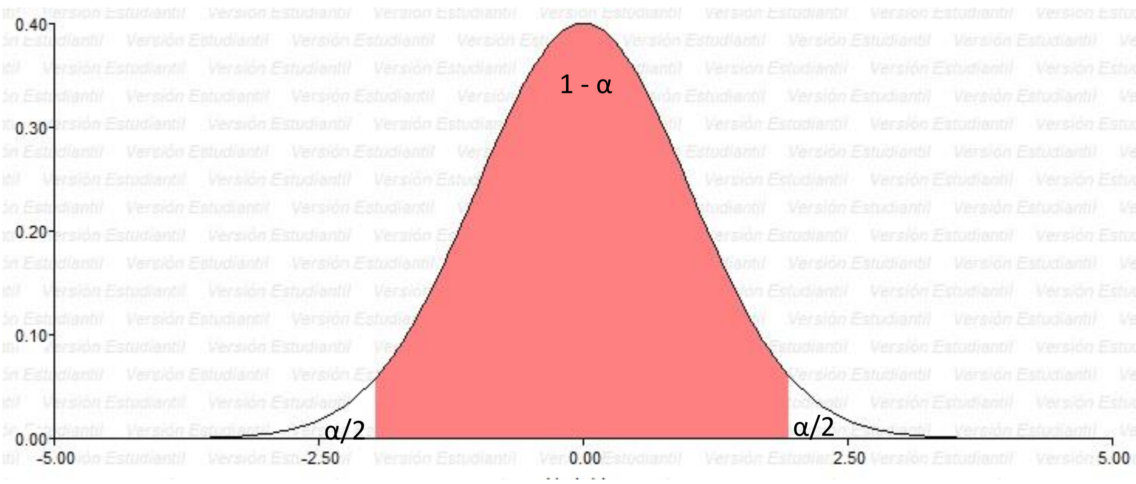
  $$S^2 = \frac{[(X_1-\bar{X})^2+(X_2-\bar{X})+\cdots+(X_n-\bar{X})]}{n}$$ donde $S^2 = \sigma^2$

Estimación por intervalos

Para evitar la consecuencia de la estimación puntual se establece un **intervalo de confianza** que se construyen en torno a la estimación puntual y así aumentar el nivel de confianza de dicho estadístico muestral.

Para armar dicho intervalo se debe determinar la **confianza** que establece la probabilidad del parámetro se encuentre en alguna parte dentro de dicho intervalo. Para calcularla se opera de la siguiente manera:

$confianza = 1 - \alpha$, donde α es un parámetro destinado a realizar cálculos



En la estimación de la media pueden presentarse **tres** situaciones a analizar: *Caso 1: media conocida y varianza conocida*

Para este caso el intervalo se puede calcular de la siguiente forma:

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma^2}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma^2}{\sqrt{n}}$$

Donde Z es un valor que debemos buscar en la tabla Z correspondiente a la distribución normal y $Z_{\alpha/2}\, \sigma^2/\sqrt{n}$ se denomina error de estimación.

Otra forma de calcular Z conociendo $\bar{X}$, μ y σ para luego buscar en la tabla "z" es por medio de la fórmula:

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

Ejemplo:

$\sigma = 2$

En primer lugar establecemos la confianza y α/2 para sumarlos, a dicho resultado lo buscamos en la tabla Z.

n = 49

$1 - 0{,}05 = 0{,}95$ $\frac{0{,}05}{2} = 0{,}025$ $0{,}95 + 0{,}025 = 0{,}975$

$\bar{X} = 10$

Buscamos 0,975 en la tabla Z y obtenemos Z = 1,96

$$10 - 1{,}96\frac{2}{\sqrt{49}} \leq \mu \leq 10 + 1{,}96\frac{2}{\sqrt{49}}$$

$\alpha = 0{,}05$

$$9{,}44 \leq \mu \leq 10{,}56$$

Caso 2: varianza desconocida y muestras mayores o iguales a 30 Para este caso la estimación será:

$$\bar{X} - Z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}$$

Siendo s la desviación estándar muestral.

Ejemplo 2:

$\boldsymbol{s} = \mathbf{3}$

$\boldsymbol{n} = \mathbf{36}$

$\bar{\boldsymbol{X}} = \mathbf{10}$

$\boldsymbol{\alpha} = \mathbf{0,05}$

Como trabajamos con el mismo α el valor Z es el mismo al ejemplo anterior, entonces:

$$10 - 1{,}96\frac{3}{\sqrt{36}} \leq \mu \leq 10 + 1{,}96\frac{3}{\sqrt{36}}$$

$$9{,}02 \leq \mu \leq 10{,}98$$

Caso 3: varianza desconocida y muestras menores a 30

Para este caso no podemos utilizar la distribución de probabilidades normal por lo que debemos utilizar la distribución de t de Student y su fórmula es:

$$\bar{X} - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}$$

Más adelante en esta clase veremos un ejemplo pero primero veamos cómo se define la distribución T de Student

Distribución t de Student:

Es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la 10 muestra es muy pequeña. Es una distribución continua y simétrica que depende del grado de libertad que se asigne a la muestra. Mientras mayor sea la muestra, el valor t tiende a Z de la distribución normal.

Los grados de libertad están dados por el número de valores que pueden ser asignados de forma arbitraria, antes de que el resto de valores tome un valor forzado para obtener un resultado establecido previamente.

Ejemplo:

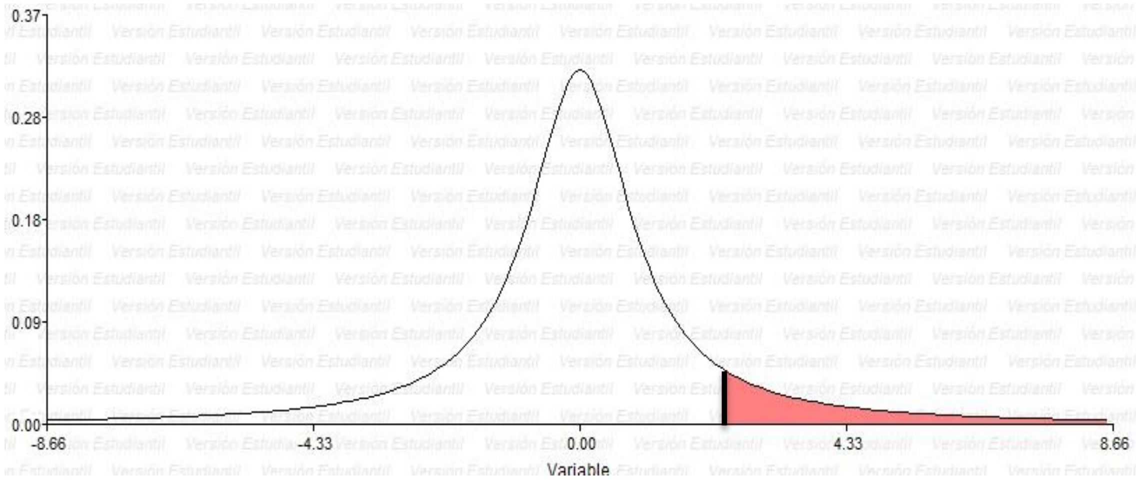
Si tengo una operación como $a_1 + a_2 + a_3 = b$ ¿Cuántos valores puedo dejar libres para llegar al resultado?
Es fácil deducir que solo con un valor libre se puede llegar a la solución ya que sin importar los valores de los otros dos números, siempre que tenga un valor libre podré llegar al resultado deseado.

Por ejemplo $7 + (-8) + a_3 = 13$ siendo $a_3 = 14$

Para este ejemplo tenemos $n - 1$ grados de libertad, es decir, si sumo 3 números tengo libertad en 2 de ellos para asignarle cualquier valor arbitrario ya que el tercero será el que fuerce a llegar al resultado que deseo.

Eventualmente este grado de libertad es el mismo que para la estimación de la media de muestras pequeñas y varianza desconocida.

Para el caso de la distribución t, la tabla **acumula las probabilidades desde** $+\infty$ **hasta t**, es decir, la parte derecha de la gráfica



En la tabla, la primer columna corresponde a los grados de libertad y se simboliza con la letra **v** mientras que α será la parte sombreada de la gráfica.

Ejemplo 3:

Una empresa quiere saber el nivel de ausentismo del personal por semana y para ello selecciona a 10 personas al azar recolectando los siguientes datos:

2, 1, 4, 2, 3, 1, 0, 0, 1, 2

Para un intervalo de confianza del 90% ¿Es posible concluir que el trabajador promedio no tiene ausencias?

Para ello debemos calcular la media muestral y la desviación estándar muestral como estadísticos puntuales

$\bar{X} = 1{,}6 \quad s = 1{,}26$

Ahora se calcula el intervalo de confianza considerando el porcentaje de

confianza

$$1{,}6 - 1{,}833\frac{1{,}26}{\sqrt{10}} \leq \mu \leq 1{,}6 + 1{,}833\frac{1{,}26}{\sqrt{10}}$$

$$0{,}87 \leq \mu \leq 2{,}33$$

Como el límite inferior del intervalo de confianza es mayor que 0 que es el valor que buscamos (sin inasistencias) no es razonable concluir que el trabajador promedio no tiene ausencias.

2. Actividad:

1. La duración de la enfermedad de Alzheimer desde el principio de síntomas hasta el fallecimiento varía de 3 a 20 años; el promedio es 8 años con una desviación estándar de 4 años. El administrador de un gran centro médico al azar selecciona los registros médicos de 30 pacientes de Alzheimer ya fallecidos, de la base de datos del centro médico y anota la duración promedio. Encuentre las probabilidades aproximadas para estos eventos:
   a. La duración promedio es menor a 7 años.
   b. La duración promedio excede de 7 años.
2. Muestras aleatorias de tamaño n se seleccionaron de poblaciones los datos dados aquí. Establezca el intervalo de confianza para cada caso

   a. $n = 36$ $\mu = 10$ $\sigma^2 = 9$ b. $n = 100$ $\mu = 5$ $s = 2$ $\alpha = 0{,}08$ c. $n = 8$ $\mu = 120$ $s = 1$ 98% de confianza
3. Las estaturas de 3000 estudiantes se distribuyen normalmente con una media de 170 cm y desviación típica de 7, 5 cm. Se toman 80 muestras

estudiantes. Establecer el intervalo de confianza que considere el 95% de confianza.

4. **Actividad Integradora.**

Habilitamos en el campus un cuestionario con las respuestas a cada uno de los puntos, para que una vez que resuelvan las actividades, puedan responder el cuestionario. Recuerden que las tablas Z y t estarán como recurso para resolver las actividades.

5. **Cierre:**

Llegamos al final de la clase 8. Abordamos los saberes necesarios para prepararlos para la clase 9 donde vamos a utilizar la estimación puntual y por intervalos en las pruebas de hipótesis.

Les deseamos buenas tareas, cualquier consulta pueden hacerla por los medios habilitados (vía mensajería interna por el campus, o en los foros habilitados para consultas).

Saludos cordiales.

6. **Bibliografía:**

- ✓ Murray, Spiegel, (1977). Probabilidad y Estadística. Naucalpan de Juarez, México.
- ✓ Rusrtom Antonio (2012). ESTADÍSTICA DESCRIPTIVA, PROBABILIDAD E INFERENCIA. Una visión conceptual y aplicada. Santiago de Chile, Chile.
- ✓ Mendenhall W., Beaver R., Beaver B. (2010) Introducción a la probabilidad y estadística 13ª Edición. Santa fe, Mexico