SESION 15

DISTRIBUCIONES DE MUESTREO

I. CONTENIDOS:
1. Distribución de muestreo.
2. Distribuciones de muestreo de la media
3. Media, mediana y moda, así como su relación con la desviación estándar de las distribuciones de muestreo.
4. Teorema del límite central, métodos de inferencia estadística.

II. OBJETIVOS:
Al término de la Clase, el alumno:
- Comprenderá con exactitud sesgos a la izquierda o derecha, leptocurtosis, platicurtosis y mesocurtosis.

III. PROBLEMATIZACIÓN:
Comenta las preguntas con tu Asesor y selecciona las ideas más significativas.
- ¿De qué tamaño debe ser una muestra representativa de la población de alumnos de tu grupo?
- ¿Cuántas posibles muestras de cinco elementos habrá en una población de mil?

IV. TEXTO INFORMATIVO-FORMATIVO:
1.1. Distribución de muestreo
Las distribuciones de muestreo describen la probabilidad con que ocurren los valores de un estadígrafo que se ha medido en diferentes muestras. Forman la base de la estadística inferencial.
Simbología de estadígrafos y parámetros:

Indicador	Parámetro	Estadígrafo
Tamaño	N	n
Media	μ	$\overline{X}$
Mediana	$\overline{\mu}$	M_e
Moda	$\dot{\mu}$	M_o
Varianza	σ^2	S^2
Desviación estándar	σ	S

Además se tienen, los siguientes símbolos:

$\mu_{\bar{x}}$ = Media de la distribución muestral de medias.

$\sigma^2_{\bar{x}}$ = Varianza de la distribución muestral de medias.

$\sigma_{\bar{x}}$ = Desviación estándar de la distribución muestral de medias.

MUESTREO CON REEMPLAZO

Consiste en extraer elementos de la población, para registrar las categorías de sus variables, y luego reintegrarlos a la población. De este modo, absolutamente todos los elementos de la población tienen la misma oportunidad de participar en la formación de la muestra. Para calcular el número de muestras posibles en el muestreo con reemplazo se eleva el tamaño de la población (N) a la potencia tamaño de la muestra (n)

$$\text{número de muestras diferentes} = \mathrm{N}^{\mathrm{n}}$$

MUESTREO SIN REEMPLAZO

En el muestreo sin remplazo; cada elemento, que se extrae para registrar las categorías de sus variables, no se reintegra a la población. De este modo, en cada nueva elección participan un menor número de elementos de la población. El cálculo del número de muestras posibles en el muestreo sin reemplazo es la combinación de "N" elementos tomados de "n" en "n"

$$\text{número de muestras diferentes} = \frac{N!}{n!(N-n)!}$$

Ejemplo 1 Se tiene una población de 50 objetos y se quiere extraer una muestra de tamaño 3.

a) ¿Cuántas muestras diferentes se pueden extraer con reemplazo?
b) ¿Cuántas muestras diferentes se pueden extraer sin reemplazo?

Para el inciso a)

$$\text{número de muestras diferentes} = \mathrm{N}^{\mathrm{n}} = 50^{3} = 125000$$

Para el inciso b)

$$\text{número de muestras diferentes} = \frac{N!}{n!(N-n)!} = \frac{50!}{3!(50-3)!} = 19600$$

2.1. Distribuciones de muestreo de la media (dmm)

Las distribuciones de muestreo de la media consisten en la descripción de todas medias de las muestras posibles, el número de éstas lo determina si el muestreo se hace con o sin reemplazo.

Ejemplo 2 Los datos siguientes representan el número de hijos profesionistas por familia

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

a) Bajo la condición de muestreo con reemplazo, determinar el total de muestras posibles de tamaño 2 y anotemos las muestras.
b) Bajo la condición de muestreo sin reemplazo, determinar el total de muestras posibles de tamaño 2 y anotemos las muestras
c) Bajo la condición de muestreo con reemplazo, expresar la dmm
d) Bajo la condición de muestreo sin reemplazo, expresar la dmm

Para el inciso a)

Como el muestreo es con reemplazo se usa N = 8 y n = 2 en la fórmula:

$$\text{número de muestras diferentes} = \mathrm{N}^{\mathrm{n}} = 8^{2} = 64$$

Las muestras posibles se pueden expresar formando binas de cada elemento consigo mismo y con todos los demás. En la tabla siguiente se describen las muestras:

(1, 1)	(1, 0)	(1, 2)	(1, 3)	(1, 2)	(1, 4)	(1, 1)	(1, 0)
(0, 1)	(0, 0)	(0, 2)	(0, 3)	(0, 2)	(0, 4)	(0, 1)	(0, 0)
(2, 1)	(2, 0)	(2, 2)	(2, 3)	(2, 2)	(2, 4)	(2, 1)	(2, 0)
(3, 1)	(3, 0)	(3, 2)	(3, 3)	(3, 2)	(3, 4)	(3, 1)	(3, 0)
(2, 1)	(2, 0)	(2, 2)	(2, 3)	(2, 2)	(2, 4)	(2, 1)	(2, 0)
(4, 1)	(4, 0)	(4, 2)	(4, 3)	(4, 2)	(4, 4)	(4, 1)	(4, 0)
(1, 1)	(1, 0)	(1, 2)	(1, 3)	(1, 2)	(1, 4)	(1, 1)	(1, 0)
(0, 1)	(0, 0)	(0, 2)	(0, 3)	(0, 2)	(0, 4)	(0, 1)	(0, 0)

Para el inciso b)
Como el muestreo es sin reemplazo la fórmula es:

$$\text{número de muestras diferentes} = \frac{N!}{n!(N-n)!}$$

Si N = 8 y n = 2

$$\frac{8!}{2!(8-2)!} = \frac{40320}{2(720)} = 28$$

Las diferentes muestras se obtienen combinando cada elemento con los demás que tiene a su derecha. No se debe hacer la combinación con elementos a la izquierda.

(1, 0)	(1, 2)	(1, 3)	(1, 2)	(1, 4)	(1, 1)	(1, 0)
(0, 2)	(0, 3)	(0, 2)	(0, 4)	(0, 1)	(0, 0)	(2, 3)
(2, 2)	(2, 4)	(2, 1)	(2, 0)	(3, 2)	(3, 4)	(3, 1)
(3, 0)	(2, 4)	(2, 1)	(2, 0)	(4, 1)	(4, 0)	(1, 0)

Para el inciso c)
Se suman los elementos de las muestras con reemplazo y el resultado se divide entre dos para obtener las medias de cada muestra.

1	0.5	1.5	2	1.5	2.5	1	0.5
0.5	0	1	1.5	1	2	0.5	0
1.5	1	2	2.5	2	3	1.5	1
2	1.5	2.5	3	2.5	3.5	2	1.5
1.5	1	2	2.5	2	3	1.5	1
2.5	2	3	3.5	3	4	2.5	2
1	0.5	1.5	2	1.5	2.5	1	0.5
0.5	0	1	1.5	1	2	0.5	0

Para el inciso d)
Se suman los elementos de las muestras sin reemplazo y el resultado se divide entre dos para obtener las medias de cada muestra.

0.5	1.5	2	1.5	2.5	1	0.5
1	1.5	1	2	0.5	0	2.5
2	3	1.5	1	2.5	3.5	2
1.5	3	1.5	1	2.5	2	0.5

3.1. Media, mediana y moda, así como su relación con la desviación estándar de las distribuciones de muestreo

En las distribuciones de muestreo de la media es posible calcular la media, la mediana, la moda y la desviación estándar. Si se obtiene la media de la dmm el valor es igual al de la media de la población.

Si el muestreo se hace con reemplazo, o cuando el tamaño de las muestras es menor al 5% del tamaño de la población, la desviación estándar de la dmm es igual a la desviación estándar de la población entre la raíz cuadrada del tamaño de las muestras.

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

Pero si el muestreo se hace sin reemplazo, hay que agregar un factor de corrección que es igual a la raíz cuadrada del cociente de la diferencia del tamaño de la población y el de la muestra entre el tamaño de la población menos uno.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N-1}\right)}$$

La expresión anterior se conoce también como error de muestreo. Es la desviación en la distribución de muestreo de la media. La gráfica de esta distribución se aproxima a la distribución normal conforme aumenta el tamaño de las muestras, y además con ello disminuye el error de muestreo. Esto tiene grandes aplicaciones en los estudios de poblaciones enormes, pues es una medida de la precisión.

Ejemplo 3 Para los datos del ejemplo 2, encuentra la media, la mediana, la moda y la desviación estándar de la dmm, para el muestreo sin reemplazo.
La dmm es:

0.5	1.5	2	1.5	2.5	1	0.5
1	1.5	1	2	0.5	0	2.5
2	3	1.5	1	2.5	3.5	2
1.5	3	1.5	1	2.5	2	0.5

Ordenando los datos: 0, 0.5, 0.5, 0.5, 0.5, 1, 1, 1, 1, 1, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 2, 2, 2, 2, 2, 2.5, 2.5, 2.5, 2.5, 2.5, 3, 3, 3.5
Según los temas vistos en la clase 4
La mediana es 1.5
La moda es 1.5
La media es 1.625
Como se puede comprobar, la mediana y la media de la población coinciden con la media y la mediana de la dmm.
La desviación estándar de la dmm es 0.862150
La desviación estándar de la población es 1.316956
Sustituyendo esta última en la expresión del error de muestreo:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N-1}\right)} = \frac{1.316956}{\sqrt{2}}\sqrt{\left(\frac{8-2}{8-1}\right)} = 0.862150$$

Los últimos dos ejemplos han de ser útiles para comprender mejor el Teorema del límite central.

4.1. Teorema del límite central, métodos de inferencia estadística

El Teorema del límite central establece que si se mantiene un muestreo aleatorio y el tamaño de la muestra es mayor a 30, entonces la media de la distribución de muestreo de la media es igual a la media de la población y la desviación estándar de la distribución muestreo de la media es igual al error de muestreo. Además la distribución de muestreo de la media es aproximadamente normal si el tamaño de la muestra es mayor a 30, por lo que es posible hacer cálculos de probabilidad como se estudió en la clase 13. Este Teorema es posiblemente el concepto más importante de la teoría estadística, su aplicación se extiende al cálculo de los límites de confianza o al tamaño de una muestra, entre otros.

MÉTODOS DE INFERENCIA ESTADÍSTICA

La estimación de parámetros a partir de los estadígrafos del muestreo, se puede dar como una estimación de punto al asociar directamente el estadígrafo con el parámetro correspondiente. Se utiliza en la estimación de punto el error de muestreo para indicar lo precisa que es. Se puede hacer una estimación utilizando la desviación estándar de la muestra en lugar de la desviación estándar de la población, pues a menudo se desconoce.

También existe la estimación de intervalo, en donde se aprovecha la distribución normal que tiene la dmm. Se da un nivel de confianza del 95%, por lo general, de que el valor del parámetro está entre un par de límites que forman el intervalo de confianza.

Una distribución puede tener sesgos, esto ya se explicó en la clase 5, pero la cutosis es el grado de distorsión en el sentido vertical con respecto a la curva normal. Se llama leptocurtosis si se tiene un pico elevado, mesocurtosis si la curva es poco elevada, y platicurtosis si la curva es más bien aplanada. ¡Bien, terminamos!

V. ESTRATEGIAS CENTRADAS EN EL APRENDIZAJE:

A. Realiza las siguientes actividades.

1. Dado un universo de tamaño 20 determina el total de muestras de tamaño 5, 7, 10, 13 con reemplazo.

2. Los datos siguientes representan el ingreso mensual de 6 personas.

5, 500, 3,800 12,000 4,500 3,000

   a) Bajo la condición de muestreo sin reemplazo, determina el total de muestras posibles de tamaño 2 y anótalas.
   b) Anota la distribución de muestreo de la media.
   c) Calcula la media y desviación estándar de la dmm.

3. Estima el error de muestreo de la media y efectúa una estimación de punto apropiada a partir de los datos de las muestras dadas a continuación.

   a) $\bar{x} = 37$ $S = 3.7$
   b) $\bar{x} = 75$ $S = 1.6$
   c) $\bar{x} = 50,000$ $S = 7500$
   d) $\bar{x} = 1$ $S = .05$

Universidad América Latina
Av. Cuauhtémoc 188-E
Fracc. Magallanes
C.P. 39670
Acapulco, Guerrero, México
www.ual.edu.mx

2011

Para cualquier comentario o sugerencia relativa a los **Servicios**, **Personal Docente**, **Administrativo** ó **Guías de Estudio**, favor de comunicarse a los teléfonos:

Dirección General:
01 (33) 47-77-71-00 ext. 1000 con Claudia Ley de 10:00 a 16:00 Hrs.
Coordinación de Asesores:
01 (33) 47-77-71-00 ext. 1013 con el Lic. Miguel Machuca García de 08:00 a 17:00 Hrs.

e-mail: **vicerrectoria@ual.edu.mx**