



INSTITUTO DE
ADMINISTRACIÓN
PÚBLICA DEL ESTADO
DE CHIAPAS, A.C.

Maestría en
Administración y
Políticas Públicas

ESTADÍSTICA ADMINISTRATIVA

Dr. Enrique Paniagua Molina

Actividad III

- Control de lectura

Capítulos VIII

Metodología de la investigación

- Ejercicios

ALAN FABRIZIO SÁNCHEZ NAVARRO

TAPACHULA, CHIAPAS

21 de Septiembre 2015

Para esta actividad se realizó la lectura del capítulo 8, donde explica el proceso para la selección de la muestra a investigar, teniendo ya definido la población a estudiar, que debido al tiempo y recursos, se seleccionará una muestra para nuestra investigación.

Una vez que se ha definido cuál será la unidad de análisis, se procede a delimitar la población que va a ser estudiada y sobre la cual se pretende generalizar los resultados. Una deficiencia que se presenta en algunos trabajos de investigación es que no describen lo suficiente las características de la población o consideran que la muestra la representa de manera automática. Para evitar errores una adecuada delimitación del universo o población, los criterios que cada investigador cumpla dependen de sus objetivos de estudio, lo importante es establecerlos de manera muy específica, para esto la muestra es, en esencia, un subgrupo de la población. Digamos que es un subconjunto de elementos que pertenecen a ese conjunto definido en sus características al que llamamos población.

Tipos de muestra; Básicamente categorizamos las muestras en dos grandes ramas: las muestras no probabilísticas y las muestras probabilísticas. En estas últimas todos los elementos de la población tienen la misma posibilidad de ser escogidos y se obtienen definiendo las características de la población y el tamaño de la muestra, y por medio de una selección aleatoria o mecánica de las unidades de análisis. En las muestras no probabilísticas, la elección de los elementos no depende de la probabilidad, sino de causas relacionadas con las características de la investigación o de quien hace la muestra. Aquí el procedimiento no es mecánico, ni con base en fórmulas de probabilidad.

Resumiremos diciendo que la elección entre la muestra probabilística y la no probabilística se determina con base en el planteamiento del problema, las hipótesis, el diseño de investigación y el alcance de sus contribuciones.

Muestra probabilística estratificada se utiliza para comparar sus resultados entre segmentos, grupos o nichos de la población, porque así lo señala el planteamiento del problema.

Muestreo probabilístico por racimos En algunos casos, en que el investigador se ve limitado por recursos financieros, por tiempo, por distancias geográficas o por una combinación de éstos y otros obstáculos, se recurre al muestreo por racimos o clusters. En este tipo de muestreo se reducen costos, tiempo y energía, al considerar que muchas veces las unidades de análisis se encuentran encapsuladas o encerradas en determinados lugares físicos o geográficos, a los que se denomina racimos.

Las unidades de análisis o los elementos muestrales se eligen siempre aleatoriamente para asegurarnos de que cada elemento tenga la misma probabilidad de ser elegido. Se utilizan tres procedimientos de selección: Tómbola, Números random o números aleatorios Selección sistemática de elementos muestrales, Archivos, Mapas, Volúmenes, Periodos registrados.

Por otro lado encontramos Las muestras no probabilísticas, también llamadas muestras dirigidas, suponen un procedimiento de selección informal. Se utilizan en muchas investigaciones cuantitativas y cualitativas. La muestra dirigida selecciona sujetos "típicos" con la vaga esperanza de que sean casos representativos de una población determinada.

Muestreo al azar por marcado telefónico, {Random Digit Dialing). Ésta es una técnica que los investigadores utilizan para seleccionar muestras telefónicas. Muestra multi-etapas o poli-etápica Este concepto significa que para extraer la muestra hemos utilizado diversos procedimientos.

En el capítulo se analizaron los conceptos de muestra, población o universo, tamaño de la muestra, representatividad de la muestra y procedimiento de selección. También se presenta una tipología de muestras: probabilísticas y no probabilísticas. Se explica cómo definir a las unidades de análisis, de las cuales se habrán de recolectar los datos. Por lo tanto, para seleccionar una muestra, lo primero que hay que hacer es definir la unidad análisis (personas, organizaciones, periódicos, comunidades, situaciones, eventos, etc.). El sobre qué o quiénes se van a recolectar datos depende del planteamiento del problema a investigar.

Ejercicio 1

Si la señora López compra una de las casas anunciadas para su venta en un diario de TGZ, **T** es el evento de que la casa tiene tres o más baños, **U** es el evento de que tiene una chimenea, **V** es el evento de que cuesta más de \$ 100 mil pesos y **W** es el evento de que es nueva.

- Describa (con palabras) cada uno de los siguientes eventos:
-

$T' =$ La casa tiene una chimenea que cuesta más de \$100 mil pesos y es nueva.

$U' =$ La casa tiene tres o más baños, cuesta más de \$100 mil pesos y es nueva.

$V' =$ La casa tiene tres o más baños, tiene una chimenea y es nueva.

$W' =$ La casa tiene tres o más baños, tiene una chimenea y cuesta más de \$100 mil pesos.

$T \cap U =$ Vacío

$T \cap V =$ Vacío

$U' \cap V =$ La casa cuesta más de \$100 mil pesos.

$V \cup W =$ La casa cuesta más de \$100 mil pesos y es nueva.

$V' \cup W =$ La casa tiene tres o más baños, tiene una chimenea y es nueva.

$T \cup U =$ La casa tiene tres o más baños y una chimenea

$T \cup V =$ La casa tiene tres o más baños y cuesta más de \$100 mil pesos

$V \cap W =$ Vacío

Ejercicio 2

Un dado está arreglado de manera que cada número impar tiene el doble de probabilidad de ocurrir que un número par. Encuentra $P(B)$, donde B es el evento que un número mayor que 3 ocurra en un solo tiro del dado.

Espacio muestral $S = \{1,2,3,4,5,6\}$

Sub conjunto B $B = \{4,5,6\}$

Probabilidad

- Si x es la probabilidad que ocurra un número par, y sería la probabilidad que ocurra un número impar.
- Entonces, encontramos que: $2y + x + 2y + x + 2y + x = 1$
 - Esto se debe al postulado 2
- La $P(B)$ sería: $4/6$

Ejercicio 3

Calcula la muestra para una población desconocida con un 96% de confianza y 10% error. Para una prevalencia de .5 y .7

Fórmula para poblaciones finitas (probabilísticas)
Murray y Larry (2005)

- Tamaño de la muestra para la población infinita o desconocida

$$n = \frac{Z_{\alpha}^2 \cdot p \cdot q}{i^2}$$

$n=?$

i = Error que se prevé cometer si es del= 10%= .10

p = prevalencia esperada del parámetro a evaluar, en caso de desconocerse ($p=0.5$), que hace mayor el tamaño muestral **5 y 7**

$q = 1-p$ (si $p=70$, $q=30$)

z = distribución de gauss para ciertos niveles de confianza

[96%= $z= 2.054$]

$$n = \frac{z_{\alpha}^2 \cdot p \cdot q}{i^2}$$

$$n = \frac{(2.054)^2 (.5)(.5)}{(.10)^2}$$

$$n = \frac{(4.2189) (.25)}{(.01)}$$

$$n = \frac{(1.054725)}{(.01)}$$

$$n = \underline{105.47}$$

$$n = \underline{105}$$

$$n = \frac{(2.054)^2 (.7)(.3)}{(.10)^2}$$

$$n = \frac{(4.2189) (.21)}{(.01)}$$

$$n = \frac{(0.885969)}{(.01)}$$

$$n = \underline{88.59}$$

$$n = \underline{89}$$

Nivel de confianza 90% -> $Z=1.645$

Nivel de confianza 95% -> $Z=1.96$

Nivel de confianza 99% -> $Z=2.575$

Ejercicio 4

Calcula la muestra para una población de 350,000 familias, con un 99% de confianza y 5% error. Para una prevalencia de .5 y .7

- Tamaño de la muestra para la población finita y conocida

$$n = \frac{Z_{\alpha}^2 \cdot N \cdot p \cdot q}{i^2 (N - 1) + Z_{\alpha}^2 \cdot p \cdot q}$$

$n=?$

z = distribución de gauss para ciertos niveles de confianza

[95%= $z=0.05= 1.96$] y [99%= $z=0.01=2.575$]

$N= 350,000$

p = prevalencia esperada del parámetro a evaluar, en caso de desconocerse

($p=0.5$), que hace mayor el tamaño muestral y ($p=0.7$)

$q= 1-p$ (si $p=0.5$; $q=0.5$) (si $p=0.7$, $q=0.3$)

i = Error que se prevé cometer si es del= 5%= **0.05**

$$n = \frac{z_{\alpha}^2 \cdot N \cdot p \cdot q}{i^2 (N-1) + z^2 \cdot p \cdot q}$$

$$n = \frac{(2.575)^2 (350,000)(.5)(.5)}{(0.05)^2 (350,000-1) + (2.58)^2 (.5)(.5)}$$

$$n = \frac{(2.575)^2 (350,000)(.7)(.3)}{(0.05)^2 (350,000-1) + (2.575)^2 (.7)(.3)}$$

$$n = \frac{(6.6306) (350,000)(.25)}{(0.0025) (349,999) + (6.6306)(.25)}$$

$$n = \frac{(6.6306) (350,000)(.21)}{(0.0025) (349,999) + (6.6306)(.21)}$$

$$n = \frac{(6.6306) (87,500)}{(874.9975) + (1.65765)}$$

$$n = \frac{(6.6306) (73,500)}{(874.9975) + (1.392426)}$$

$$n = \frac{580,179.6875}{876.6551}$$

$$n = \frac{487,349.1}{876.3899}$$

$$n = 661.81 \approx \mathbf{662 \text{ Familias}}$$

$$n = 556.08 \approx \mathbf{556 \text{ Familias}}$$

Nivel de confianza 90% -> $Z=1.645$

Nivel de confianza 95% -> $Z=1.96$

Nivel de confianza 99% -> $Z=2.575$

Ejercicio 5

De una población de 1,176 padres de familia de la ciudad de Tuxtla Gutiérrez, se pretende conocer la aceptación de los programas educativos mediante caricaturas. Se pretende obtener una muestra para saber el número de entrevistas y con ello obtener información estadísticamente confiable. Se asume un error standard de 1.5% con un nivel de confiabilidad del 90%

Tamaño de la muestra para población

$$n' = \frac{s^2}{o^2} \qquad n = \frac{n'}{1 + n'/N}$$

Donde:

n : tamaño muestral =?

N : tamaño de la población = **1,176**

se : error estándar = 1.5% = **0.015**

p : % de confiabilidad = 90% = **0.9**

s^2 : varianza muestral = **$p(1-p)$** = 0.9 (1-0.9) = **0.09**

o^2 : varianza poblacional = $(se)^2 = (0.015)^2 =$ **0.000225**

$$n' = \frac{s^2}{o^2} \qquad n' = \frac{0.09}{0.000225} \qquad n' = \underline{400}$$

$$n = \frac{n'}{1 + n'/N} \qquad n = \frac{400}{1 + 400/1176} \qquad n = \frac{400}{1.3401} \qquad n = \underline{298.48} \approx \text{298 entrevistas}$$

Ejercicio 6

Son los resultados de preguntarle la estatura a 60 trabajadores del departamento de limpia municipal de SCLC.

1. Obtén la media aritmética (para datos agrupados) **R:1.66 cms**
2. Obtén la varianza **0.0060419**
3. Y la desviación estándar (para datos agrupados) **0.07773 cms**
4. Interpreta los resultados **Estas medidas establecen la forma en que los valores fluctúan con respecto a la media.**

m	Estatura (xi)	Frecuencia fi	(xi) (fi)	\bar{X}	(xi-x) ²	(xi-x) ² (fi)	
1	1.52	1	1.52	1.66	0.0196	0.0196	
2	1.54	5	7.7	1.66	0.0144	0.072	
3	1.55	4	6.2	1.66	0.0121	0.0484	
4	1.58	5	7.9	1.66	0.0064	0.032	
5	1.6	2	3.2	1.66	0.0036	0.0072	
6	1.62	4	6.48	1.66	0.0016	0.0064	
7	1.64	7	11.48	1.66	0.0004	0.0028	
8	1.66	3	4.98	1.66	0	0	
9	1.7	5	8.5	1.66	0.0016	0.008	
10	1.71	8	13.68	1.66	0.0025	0.02	
11	1.73	6	10.38	1.66	0.0049	0.0294	
12	1.74	5	8.7	1.66	0.0064	0.032	
13	1.77	3	5.31	1.66	0.0121	0.0363	
14	1.8	1	1.8	1.66	0.0196	0.0196	
15	1.83	1	1.83	1.66	0.0289	0.0289	
		60	99.66			0.3626	
			1.661			0.00604333	0.07773888

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_m f_m}{f_1 + f_2 + \dots + f_m} = \sum_{i=1}^m \frac{x_i f_i}{n}$$

Media aritmética= 99,66/ 60= 1,66
La estatura promedio es de 1.66 cms.

• Para el caso de datos agrupados

$$\sigma = \sqrt{\frac{\sum_{i=1}^m f_i \cdot (x_i - \bar{x})^2}{n}}$$

Desviación estándar $0.3626/ 60 = \sqrt{0.006043} = \mathbf{0.07773 \text{ cms}}$

Varianza = $0.7773^2 = \mathbf{0.0060419}$