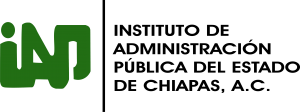
****

**Maestría en Administración y politicas públicas**

**Actividad 3**

* **Control de Lectura capitulo 8**
* **Ejercicios de estadística descriptiva**

**Materia: Estadística Administrativa**

**Dr. Enrique Paniagua Molina**

**Maestrante: L.A. Viridiana Figueroa García**

**21 de septiembre, 2015**

Control de Lectura

Capitulo 8

Selección de la muestra

Sólo cuando queremos realizar un censo debemos incluir en el estudio a todos los sujetos o casos (personas, animales, plantas, objetos) del universo o la población. Las muestras se utilizan por economía de tiempo y recursos.

Para seleccionar una muestra, lo primero que hay que hacer es definir la unidad de 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 y de los alcances del estudio. Estas acciones nos llevarán al siguiente paso, que consiste en delimitar una población.

Para el proceso cuantitativo la muestra es un subgrupo de la población de interés (sobre el cual se recolectarán datos, y que tiene que definirse o delimitarse de antemano con precisión), éste deberá ser representativo de la población. El investigador pretende que los resultados encontrados en la muestra logren generalizarse o extrapolarse a la población.

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. Así, una población es el conjunto de todos los casos que concuerdan con una serie de especificaciones.

La delimitación de las características de la población no sólo depende de los objetivos del estudio, sino de otras razones prácticas. Un estudio no será mejor por tener una población más grande; la calidad de un trabajo investigativo estriba en delimitar claramente la población con base en el planteamiento del problema.

Las poblaciones deben situarse claramente en torno a sus características de contenido, de lugar y en el tiempo.

**¿Cómo seleccionar la muestra?**

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.

**¿Cómo se selecciona una muestra probabilística**?

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. Las muestras probabilísticas tienen muchas ventajas, quizá la principal sea que puede medirse el tamaño del error en nuestras predicciones. Se dice incluso que el principal objetivo en el diseño de una muestra probabilística es reducir al mínimo este error, al que se le llama error estándar.

Las muestras probabilísticas son esenciales en los diseños de investigación transeccionales, tanto descriptivos como correlaciónales-causales (las encuestas de opinión o *surveys,* por ejemplo), donde se pretende hacer estimaciones de variables en la población.

Para una muestra probabilística necesitamos principalmente dos cosas: determinar el tamaño de la muestra *(n)* y seleccionar los elementos muéstrales, de manera que todos tengan la misma posibilidad de ser elegidos.

**Muestra probabilística estratificada**

En ocasiones el interés del investigador es comparar sus resultados entre segmentos, grupos o nichos de la población, porque así lo señala el planteamiento del problema. Por ejemplo, efectuar comparaciones por género (entre hombres y mujeres), si la selección de la muestra es aleatoria, tendremos unidades o elementos de ambos géneros, no hay problema, la muestra reflejará a la población.

Pero a veces, nos interesan grupos que constituyen minorías de la población o universo y entonces si la muestra es aleatoria simple, resultará muy difícil determinar qué elementos o casos de tales grupos serán seleccionados.

**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.*

Muestrear por racimos implica diferenciar entre la unidad de análisis y la unidad muestra! La unidad de análisis indica quiénes van a ser medidos, o sea, los participantes o casos a quienes en última instancia vamos a aplicar el instrumento de medición.

Las unidades de análisis o los elementos muéstrales se eligen siempre aleatoriamente para asegurarnos de que cada elemento tenga la misma probabilidad de ser elegido. Se utilizan tres procedimientos de selección:

**Tómbola.** Muy simple y no muy rápido, consiste en numerar todos los elementos muéstrales del uno al número *n.* Hacer fichas o papeles, uno por cada elemento, revolverlos en una caja, e ir sacando *n* número de fichas, según el tamaño de la muestra. Los números elegidos al azar conformarán la muestra.

**Números *random* o números aleatorios.** El uso de números *random* no significa la selección azarosa o fortuita, sino la utilización de una tabla de números que implica un mecanismo de probabilidad muy bien diseñado.

**Selección sistemática de elementos muéstrales.** Este procedimiento de selección es muy útil e implica elegir dentro de una población *N* un número *n* de elementos a partir de un intervalo *K* Este último *(K)* es un intervalo que se va a determinar por el tamaño de la población y el tamaño de la muestra. De manera que tenemos que *K = N/n,* en donde *K =* un intervalo de selección sistemática, *N* = la población y *n* = la muestra.

**Listados y otros marcos muéstrales.** Las *muestras probabilísticas* requieren la determinación del tamaño de la muestra y de un proceso de selección aleatoria que asegure que todos los elementos de la población tengan la misma probabilidad de ser elegidos.

**Archivos.** Un gerente de reclutamiento y selección de una empresa quiere precisar si algunos datos que se dan en una solicitud de trabajo están correlacionados con el ausentismo del empleado. Es decir, si a partir de datos como edad, género, estado civil, nivel educativo y duración en otro trabajo, es factible predecir la conducta de ausentismo.

**Mapas.** Los mapas son muy útiles como marco de referencia en muestras de racimos. Por ejemplo, un investigador quiere saber qué motiva a los compradores de las tiendas de autoservicio.

**Volúmenes.** El investigador encuentra datos que le proporciona un marco de referencia ideal, a partir de donde seleccionará *n* volúmenes para su análisis.

**Periodos registrados**

En la investigación que se vincula con sesiones terapéuticas, entrevistas con pacientes o reos, etc., se dispone de los listados de archivos que por lo común guardan las instituciones.

El investigador debe buscar posibles listados o bases de datos en cualquier parte para que pueda obtener el tamaño de su población.

**Tamaño óptimo de una muestra**

Las muestras probabilísticas requieren dos procedimientos básicos:

1) la determinación del tamaño de la muestra y

2) la selección aleatoria de los elementos muéstrales.

El primer procedimiento fue descrito en su modalidad más simple en la sección sobre el tamaño de la muestra. Precisar adecuadamente el tamaño de la muestra puede tomarse muy complejo, esto depende del problema de investigación y la población a estudiar.

El tamaño de una muestra depende también del número de subgrupos que nos interesan en una población. Lo óptimo de una muestra depende de cuánto se aproxima su distribución a la distribución de las características de la población. Esta aproximación mejora al incrementarse el tamaño de la muestra. La "normalidad" de la distribución en muestras grandes no obedece a

**Muestras no probabilísticas**

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 re- presentativos de una población determinada. Por ello, para fines deductivos-cuantitativos, donde la generalización o extrapolación de resultados hacia la población es una finalidad en sí misma, las muestras dirigidas en este sentido implican muchas desventajas. Es decir, los datos no pueden generalizarse a ésta, la cual no se consideró en sus parámetros ni en sus elementos para obtener la muestra.

**Muestreo al azar por marcado telefónico** *{Random Digit Dialing)***.** Ésta es una técnica que los investigadores utilizan para seleccionar muestras telefónicas. Involucra identificar áreas geográficas —para ser muestreadas al azar— y sus correspondientes códigos telefónicos e intercambios (los tres dígitos del número telefónico).

**Muestra multietapas o polietápica.** Este concepto significa que para extraer la muestra hemos utilizado diversos procedimientos.

La muestra es un subgrupo de la población puede ser probabilística o no probabilística. Elegir qué tipo de muestra se requiere depende del enfoque y alcances de la investigación, los objetivos del estudio y el diseño.

Ejercicios de Estadística Descriptiva

**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’= Casa con 2 baños

U’= Casa sin chimenea

V’= Casa con valor menor a 100 mil pesos

W´= La casa no es nueva

T ∩ U= Buena casa con tres baños y chimenea óptimas condiciones

T ∩ V= Excelente casa con más de 3 baños precio mayor a 100 mil pesos

U’ ∩ V= La chimenea necesita reparaciones y cuesta más de 100 mil pesos

V U W= Casa nueva con valor mayor a 100 mil pesos

V’ U W= Casa nueva con un valor menor a 100 mil pesos

T U U= Casa con 3 baños y chimenea

T U V= Casa con 3 baños con precio mayor a 100 mil pesos

V ∩ W= Precio mayor a 100 mil pesos, casa nueva

**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

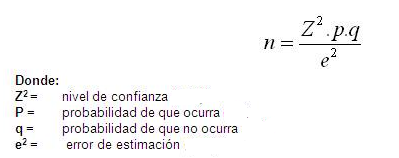
* Si **x** es la probabilidad que ocurra un número par, **2** sería la probabilidad que ocurra un número impar.
* Entonces, encontramos que: 4+ x + 5 + x + 6 + x = 1

Esto se debe al postulado 2

La P(B) sería: 4/9

P(B) = 1/9+2/9+1/9= 4/9

**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*

Donde:

Z2= nivel de confianza

P=probabilidad que ocurra

q=probabilidad que no ocurra

e2= error de estimacion

Z²α= 1.7506

P= .5

.q= 1-p = (1-.5) = .5

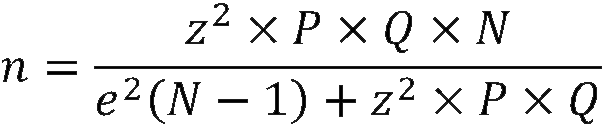
.i= .10

n= (1.7506)²(.5)(.5) = (3.064)(.5)(.5) = (3.064)(.25) =

         (.10)²                   .01                  .01

n= 76.61 = 77

**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



FÓRMULA

Para una prevalencia de 0.5

Z²α= 1.7506

P= .5

.q= 1-p = (1-.5) = .5

.e= .05

n= (2.3263)² (350000)(.5)(.5)             =       (5.4116)(350000)(.25)             =

      (.05)²(350000-1)+ (2.3263)²(.5) (.5)      (0.0025)(349999)+(5.4116)(.25)

n=  473515                 =         473515     =

    (874.9975)+1.3529            876.3504

n= 540.326 = 540

**Para una prevalencia de 0.7**

Z²α= 1.7506

P= .5

.q= 1-p = (1-.7) = .3

.e= .05

n= (2.3263)² (350000)(.7)(.3)             =       (5.4116)(350000)(.21)             =

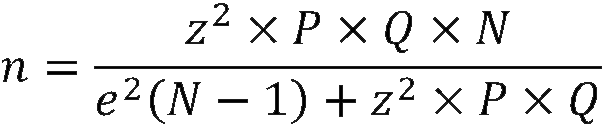
(.05)²(350000-1)+ (2.3263)²(.7) (.3)      (0.0025)(349999)+(5.4116)(.21)

n=  397752.60           =      397752.60     =

    (874.9975)+1.364              876.1339

n= 453.98 = 454

**Ejercicio 5**. De una Población de 1,176 padres de familia de la Cd. de Tuxtla Gutiérrez. Se pretende conocer la aceptación de los programas educativos mediante caricaturas. se pretende obtener una muestra par saber el número de entrevistas y con ello obtener información estadísticamente confiable. Se asume un error estándar de 1.5% con un nivel de confiabilidad del 90%



 Fórmula:

N=1176

Z=1.2815

p=0.5

q= 0.5

i=0.015

n= (1.2815)² (1176)(.5)(.5)             =       (1.6422)(1176)(.25)             =

  (.015)²(1176-1)+ (1.2815)²(.5) (.5)      (0.000225)(1175)+(1.6422)(.25)

n=  482.8068           =      482.8068     =

    (0.2643)+0.2643             0.6746

n=715.48 = 715

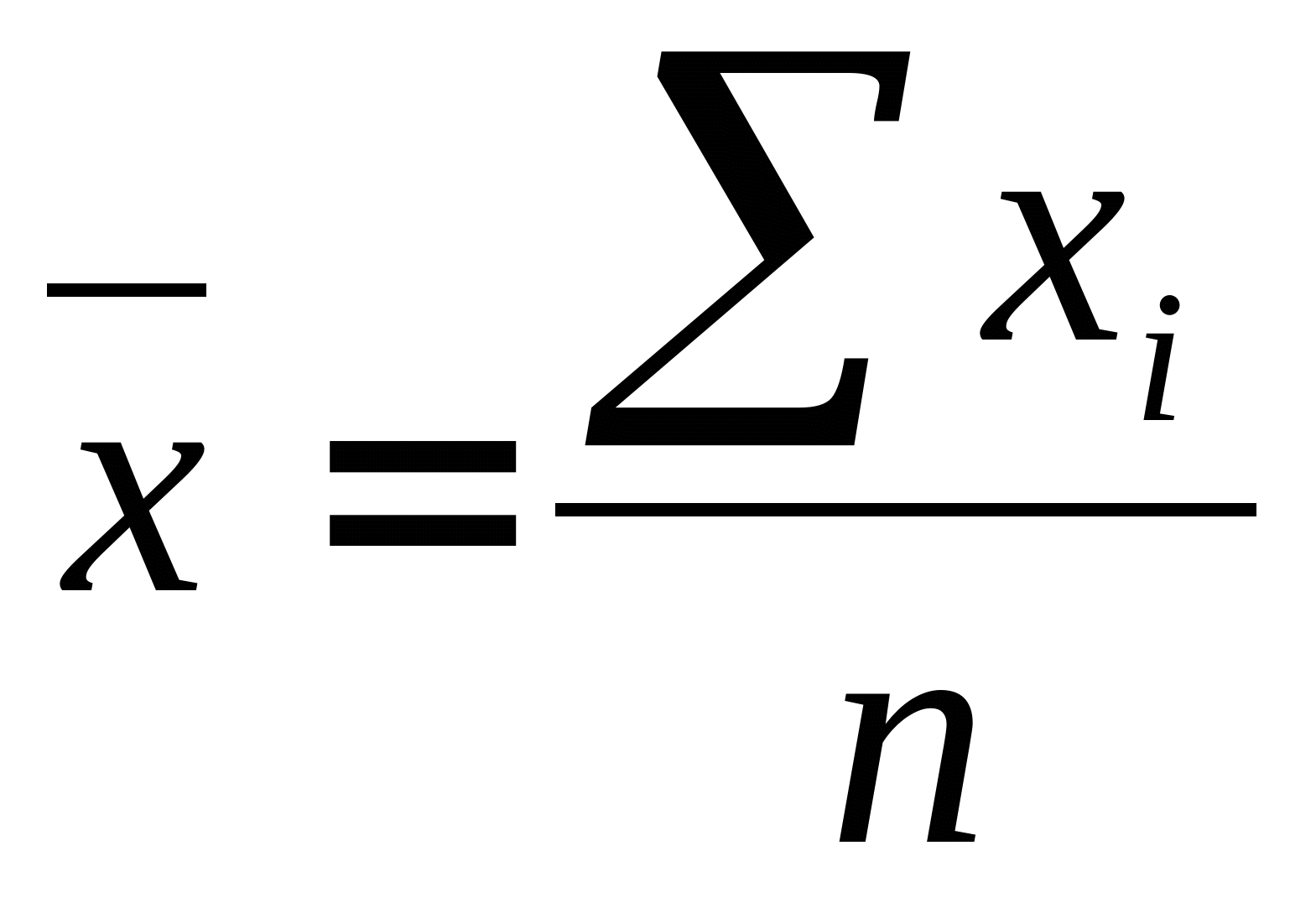
**EJERCICIO 6** Son los resultados de preguntarle la estatura a 60 trabajadores del departamento de limpia municipal de SCLC.

* Obtén la media aritmética (para datos agrupados)
* Obtén la desviación estándar y la varianza (para datos agrupados)
* Interpreta los resultados

|  |  |
| --- | --- |
| **Estatura (x)** | **f** |
| 1.52 | 1 |
| 1.54 | 5 |
| 1.55 | 4 |
| 1.58 | 5 |
| 1.60 | 2 |
| 1.62 | 4 |
| 1.64 | 7 |
| 1.66 | 3 |
| 1.70 | 5 |
| 1.71 | 8 |
| 1.73 | 6 |
| 1.74 | 5 |
| 1.77 | 3 |
| 1.80 | 1 |
| 1.83 | 1 |
| Total | 60 |

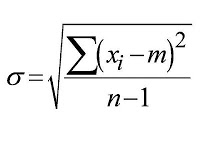
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estaturas (X) | f | Ni | xi \* fi | xi 2 \* fi |
| 1.52 | 1 | 1 | 1.52 | 2.3104 |
| 1.54 | 5 | 6 | 7.7 | 11.858 |
| 1.55 | 4 | 10 | 6.2 | 9.61 |
| 1.58 | 5 | 15 | 7.9 | 12.482 |
| 1.6 | 2 | 17 | 3.2 | 5.12 |
| 1.62 | 4 | 21 | 6.48 | 10.4976 |
| 1.64 | 7 | 28 | 11.48 | 18.8272 |
| 1.66 | 3 | 31 | 4.98 | 8.2668 |
| 1.7 | 5 | 36 | 8.5 | 14.45 |
| 1.71 | 8 | 44 | 13.68 | 23.3928 |
| 1.73 | 6 | 50 | 10.38 | 17.9574 |
| 1.74 | 5 | 55 | 8.7 | 15.138 |
| 1.77 | 3 | 58 | 5.31 | 9.3987 |
| 1.8 | 1 | 59 | 1.8 | 3.24 |
| 1.83 | 1 | 60 | 1.83 | 3.3489 |
| Total | 60 |  | 99.66 | 165.8978 |

**Desviación estándar para datos agrupados:**



**1.- Media Aritmetica**

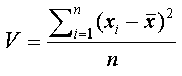
X = 99.66/60= 1.661



**DESVIACION ESTANDAR**

0.077736

**VARIANZA:**



V= 0.006042