

**Maestría en Administración y Políticas Públicas**

**Asignatura:**

**Estadística Administrativa**

**Dr. Enrique Antonio Paniagua Molina**

**Actividad 3**

**Edín Pompilio Sánchez López**

**Tapachula, Chiapas. 21 de septiembre de 2015**

**A C T I V I D A D 3**

**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:**

****

**Respuestas:**

**T´=** La casa tiene una chimenea, cuesta más de $ 100 mil pesos y es nueva.

**U´=** Tiene tres o más baños, cuesta más de $ 100 mil pesos y es nueva.

**V´=** Tiene tres o más baños, tiene chimenea y es nueva.

**W´=** Tiene tres o más baños, tiene chimenea y cuesta más de $ 100 mil pesos

**T intersección U=** Vacío.

**T intersección V=** Vacío.

**U´ intersección V=** Cuesta más de $ 100 mil pesos.

**V u W=** Cuesta más de $ 100 mil pesos y es nueva.

**V´u W=** Tiene tres o más baños, tiene chimenea y es nueva.

**T u U=** Tiene tres o más baños y tiene chimenea.

**T u V=** Tiene tres o más baños y cuesta más de $ 100 mil pesos

**V intersección 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**



* **Sub conjunto B**



* **Probabilidad**
  + **Si *x* es la probabilidad que ocurra un número par, \_\_y\_\_ sería la probabilidad que ocurra un número impar. Considerando que y es doblemente probable**
  + **Entonces, encontramos que: \_2y\_+ *x* + \_2y\_ + *x* + \_2y\_+ *x* = 1**
    - **Esto se debe al postulado 2;**
  + **La P(B) sería: x + 2y + x = 2x + 2y; si x= 1/6, y= 1/6; entonces**

**2(1/6) + 2(1/6) = 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**

Para prevalencia de .5:

Z= 1.7506

p= 0.5

q= 1-p = 0.5

i = 0.1

n= (1.7506)2 (0.5) (0.5) = (3.0646) (0.5) (0.5) = **76.61**

(0.1)2 (0.01)

n= **77**

Para prevalencia de .7:

Z= 1.7506

p= 0.7

q= 1-p = 0.3

i = 0.1

n= (1.7506)2 (0.7) (0.3) = (3.0646) (0.7) (0.3) = **64.35**

(0.1)2 (0.01)

n= **64**

**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**

Para prevalencia de .5:

Z= 2.3263 (2.575)

p= 0.5

q= 1-p = 0.5

i = 0.05

N= 350,000

n= (2.3263)2 (350,000) (0.5) (0.5) .

(0.05)2 (350,000-1) + (2.3263)2 (0.5) (0.5)

n= (5.4116) (350,000) (0.5) (0.5) = 473,521.27 = **540.33**

(0.0025) (349,999) + (1.3529) 876.35

n= **540**

Para prevalencia de .7:

Z= 2.3263

p= 0.7

q= 1-p = 0.3

i = 0.05

N= 350,000

n= (2.3263)2 (350,000) (0.7) (0.3) .

(0.05)2 (350,000-1) + (2.3263)2 (0.7) (0.3)

n= (5.4116) (350,000) (0.7) (0.3) = 397,752.60 = **453.98**

(0.0025) (349,999) + (1.1364) 876.13

n= **454**

**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**](Error%20estándar.docx) **de 1.5% con un nivel de confiabilidad del 90%**

n = n´ n´ = s2

1 + (n´/ N) V2

Si:

N = 1,176

s2 = p (1-p) = 0.9 (1-0.9) = 0.09

V2= (se)2 = (0.015)2 = 0.000225

p = 0.9

Entonces:

n´ = 0.09 = 400

0.000225

n = 400 = 400 = 298.48

1 + (400/1176) 1.3401

**n = 298**

**Ejercicio 6.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Estatura (x)** | **f** | **f . x** | **.**  **f . (xi – x)2** |
| **1.52** | **1** | **1.52** | **0.0196** |
| **1.54** | **5** | **7.70** | **0.0720** |
| **1.55** | **4** | **6.20** | **0.0484** |
| **1.58** | **5** | **7.90** | **0.0320** |
| **1.60** | **2** | **3.20** | **0.0072** |
| **1.62** | **4** | **6.48** | **0.0064** |
| **1.64** | **7** | **11.48** | **0.0028** |
| **1.66** | **3** | **4.98** | **0** |
| **1.70** | **5** | **8.50** | **0.0080** |
| **1.71** | **8** | **13.68** | **0.0200** |
| **1.73** | **6** | **10.38** | **0.0294** |
| **1.74** | **5** | **8.70** | **0.0320** |
| **1.77** | **3** | **5.31** | **0.0363** |
| **1.80** | **1** | **1.80** | **0.0196** |
| **1.83** | **1** | **1.83** | **0.0289** |
| **Total** | **60** | **99.66** | **0.3626** |

* **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**

**RESPUESTAS:**

**.**

**Media aritmética= x = Sumatoria (f . x) = 99.66 = 1.66**

**n 60**

**Desviación estándar = Raíz cuadrada de 0.3626 = 0.0777**

**60**

**Varianza = V = el cuadrado de la desviación estándar = (0.0777)2 = 0.0060**

Con base en los resultados obtenidos podemos concluir que la media en la estatura de los 60 trabajadores es de 1.66, correspondiente a 3 trabajadores de los 60; lo que nos indica que 57 trabajadores; es decir, el 94.68%; ostenta diferentes estaturas respecto a la media. Así mismo, la estatura más baja corresponde a un trabajador con 1.52 y la más alta a otro con 1.83; lo que nos indica que la dispersión de la población, con respecto a la media, oscila entre estas dos estaturas. También podemos concluir que 28 trabajadores tienen una estatura por debajo de la media y 29 trabajadores, arriba de la media.

**Control de lectura**

**Capítulo 8.- Selección de la muestra**

¿En una investigación siempre tenemos una muestra?

No siempre, pero en la mayoría de las situaciones sí realizamos el estudio en una 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. Por ejemplo, los estudios motivacionales en empresas suelen abarcar a todos sus empleados para evitar que los excluidos piensen que su opinión no se toma en cuenta. Las muestras se utilizan por economía de tiempo y recursos.

Lo primero: ¿sobre qué o quiénes se recolectarán datos?

Aquí el interés se centra en "qué o quiénes", es decir, en los sujetos, objetos, sucesos o comunidades de estudio (las unidades de análisis), lo cual depende del planteamiento de la investigación. Así, en el caso de que el objetivo sea describir el uso que hacen los niños de la televisión, lo más factible sería interrogar a un grupo de niños. También serviría entrevistar a los padres de los niños. Escoger entre los niños o sus padres, o ambos, dependería no sólo del objetivo de la investigación, sino del diseño de la misma.

¿Cómo se delimita una 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 (Selltiz *et al.,* 1980). 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. Es común que algunos estudios que sólo se basan en muestras de estudiantes universitarios (porque es fácil aplicar en ellos el instrumento de medición, pues están a la mano) hagan generalizaciones temerarias sobre jóvenes que tal vez posean otras características sociales.

Es preferible entonces establecer con claridad las características de la población, con la finalidad de delimitar cuáles serán los parámetros muéstrales. Lo anterior puede ilustrarse con el ejemplo de la investigación sobre el uso de la televisión por los niños. Está claro que en dicha investigación la unidad de análisis son los niños. Pero, ¿de qué población se trata?, ¿de todos los niños del mundo?, ¿de todos los niños de la República Mexicana? Sería muy ambicioso y prácticamente imposible referirnos a poblaciones tan grandes. Así, en nuestro ejemplo, la población se delimitaría con base en lo siguiente:

**Límites de población**

Todos los niños del área metropolitana de la ciudad de México, que cursen 4o, 5° y 6o de primaria en escuelas privadas y públicas del turno matutino.

Al seleccionar la muestra debemos evitar tres errores que pueden presentarse: 1) no elegir a casos que deberían ser parte de la muestra (participantes que deberían estar y no fueron seleccionados), 2) incluir a casos que no deberían estar porque no forman parte de la población y 3) seleccionar casos que son verdaderamente inelegibles (Mertens, 2005).

¿Cómo seleccionar la muestra?

Hasta este momento hemos visto que se debe definir cuál será la unidad de análisis y cuáles son las características de la población. En este inciso hablaremos de la muestra, o mejor dicho de los tipos de muestra, con la finalidad de poder elegir la más conveniente para un estudio. 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.

**Muestra probabilística:** Subgrupo de la población en el que todos los elementos de ésta tienen la misma posibilidad de ser elegidos.

**Muestra no probabilística o dirigida:** Subgrupo de la población en la que la elección de los elementos no depende de la probabilidad sino de las características de la investigación.

¿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 (Kish, 1995).

Para hacer una muestra probabilística es necesario entender los siguientes términos y sus definiciones:

La población, a la que se le suele denominar como ií, es un conjunto de elementos.

La muestra, a la que se le simboliza como *n,* es un subconjunto de la población *N.*

En una población *N* (previamente delimitada por los objetivos de la investigación), nos interesa establecer valores de las características de los elementos de *N.*

Nos interesa conocer valores promedio en la población, lo cual se expresa como:

*Y =* al valor de una variable determinada *(Y)* que nos interesa conocer, digamos un promedio.

También nos interesa conocer:

*V* = la varianza de la población con respecto a determinadas variables (la varianza indica la variabilidad).

Como los valores de la población no se conocen, seleccionamos una muestra *n* además, a través de estimados en la muestra, inferimos valores de la población *[y* será la estimación del valor de *Y,* el cual desconocemos).

En la muestra, *y* es un estimado promedio que podemos determinar. Sabemos que en nuestra estimación habrá una diferencia *(Y* - *y =* ?), es decir, un error, el cual dependerá del número de elementos muestreados. A dicho error se le conoce como error estándar (se).

se = la desviación estándar de la distribución muestral y representa la fluctuación de *y.*

*[se]2* = el error estándar al cuadrado, cuya fórmula nos servirá para calcular la varianza (V) de la población *(N),* así como la varianza de la muestra (n) será la expresión s2.

*s2* = varianza de la muestra, la cual podrá determinarse en términos de probabilidad donde *s2=p{l-p).*

*p =* porcentaje estimado de la muestra, probabilidad de ocurrencia del fenómeno, la cual se estima sobre marcos de muestreo previos o se define, la certeza total siempre es igual a uno, las posibilidades a partir de esto son *"p"* de que sí ocurra y "?" de que no ocurra

*(p + q* = 1). De aquí se deriva 1 *-p.*

Como se habrá podido observar, cuando hablamos de un término de la muestra se simboliza con una letra minúscula *[n, s, se).* Si se trata de un término de la población, se simboliza con una letra mayúscula (JV, *S).*

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. Para lo primero, daremos una fórmula que contiene las expresiones ya descritas. Para lo segundo, requerimos un marco de selección adecuado y un procedimiento que permita la aleatoriedad en la selección.

El tamaño de la muestra

Cuando se hace una muestra probabilística, uno debe preguntarse: dado que una población es de *N,* ¿cuál es el menor número de unidades muéstrales (personas, organizaciones, capítulos de telenovelas, etc.) que necesito para conformar una muestra (n) que me asegure un determinado nivel de error estándar, digamos menor de 0.01?

La respuesta a esta pregunta busca encontrar la probabilidad de ocurrencia de *Y,* así como que mi estimado de *y* se acerque a *Y,* el valor real de la población. Si establecemos el error estándar y lo fijamos en 0.01, sugerimos que esta fluctuación promedio de nuestro estimado *y* con respecto a los valores reales de la población *Y* no sea > 0.01, es decir, que de 100 casos, 99 veces mi predicción sea correcta y que el valor de *y* se sitúe en un intervalo de confianza que comprenda el valor de *Y.*

Resumiendo, para una determinada varianza *[V)* de *Y,* ¿qué tan grande debe ser mi muestra? Ello se determina en dos pasos:

1. *n'* = —- = Tamaño provisional de la muestra1 = varianza de la muestra/varianza de la población

*V2*

2. *n =* 1 + *n'IN*

Pongamos el siguiente caso: en el ejemplo que ya habíamos mencionado en este capítulo, delimitamos una población para un estudio de directores generales, en el cual consideramos a "todos aquellos directores generales de empresas industriales y comerciales que, en 1983, tenían un capital social superior a 30 millones de pesos, con ventas superiores a los 100 millones de pesos y con más de 300 personas empleadas". Con estas características se precisó que la población era de *N* = 1176 directores generales, ya que 1176 empresas reunían las mencionadas características. ¿Cuál es entonces el número de directores generales *(n)* que se debe entrevistar, para tener un error estándar menor de 0.015, y dado que la población total es de 1176?

*N* = tamaño de la población de 1 176 empresas.

*y* = valor promedio de una variable = 1, un director general por empresa.

*se =* error estándar = 0.015, determinado por nosotros

*V2 =* varianza de la población al cuadrado. Su definición *se2:* cuadrado del error estándar

*s2 =* varianza de la muestra expresada como la probabilidad de ocurrencia de *y*

*p* = 0.9

*ri =* tamaño de la muestra sin ajustar

*n =* tamaño de la muestra

Si lo sustituimos, tenemos que:

n = 298 casos

Es decir, para nuestra investigación necesitaremos una muestra de 298 directores generales. Se trata del primer procedimiento para obtener la muestra probabilística: determinar su tamaño con base en estimados de la población. El segundo procedimiento estriba en cómo y de dónde seleccionar a esos 298 sujetos.

Muestra probabilística estratificada

Muestra probabilística estratificada: Subgrupo en el que la población se selecciona una muestra para cada segmento.

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.

El ejemplo anterior de los directores generales de empresa corresponde a una muestra probabilística simple. Determinamos en este caso que el tamaño de la muestra sería de *n* = 298 directivos. Pero supongamos que la situación se complica y que debemos estratificar esta *n* con la finalidad de que los elementos muéstrales o las unidades de análisis posean un determinado atributo. En nuestro ejemplo, este atributo podría ser el giro de la empresa. Es decir, cuando no basta que cada uno de los elementos muéstrales tengan la misma probabilidad de ser escogidos, sino que además es necesario estratificar la muestra en relación con estratos o categorías que se presentan en la población, y que además son relevantes para los objetivos del estudio, se diseña una muestra probabilística estratificada. Lo que aquí se hace es dividir a la población en subpoblaciones o estratos, y se selecciona una muestra para cada estrato.

La estratificación aumenta la precisión de la muestra e implica el uso deliberado de diferentes tamaños de muestra para cada estrato, a fin de lograr reducir la varianza de cada unidad de la media muestral (Kish, 1995). En su libro de muestreo, Kish afirma que, en un número determinado de elementos muéstrales *n* = 2 *nh,* la varianza de la media muestral *y* puede reducirse al mínimo, si el tamaño de la muestra para cada estrato es proporcional a la desviación estándar dentro del estrato.

Esto es,

*Ifh* = - | = *ksh* En donde la muestra *n* será igual a la suma de los elementos muéstrales *nh.* Es decir, el tamaño de *n* y la varianza de *y* pueden minimizarse, si calculamos "submuestras" proporcionales a la desviación estándar de cada estrato. Esto es:

*fH-^-ksn*

En donde *nh* y *Nh* son muestra y población de cada estrato, y *sh* es la desviación estándar de cada elemento en un determinado estrato. Entonces tenemos que:

Siguiendo con nuestro ejemplo de los directores de empresa, la población es de 1176 directores de empresa y el tamaño de muestra es *n =* 298. ¿Qué muestra necesitaremos para cada estrato?

De manera que el total de la subpoblación se multiplicará por esta fracción constante para obtener el tamaño de la muestra para el estrato.

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.** Para dar algunos ejemplos tenemos la tabla 8.3.

En la primera columna se encuentran unidades de análisis que frecuentemente vamos a estudiar. En la segunda, sugerimos posibles racimos donde se encuentran dichos elementos.

**Tabla 8.3 Ejemplo de racimos o *clusters***

**Unidad de análisis Posibles racimos**

Adolescentes Preparatorias

Obreros Industrias

Amas de casa Mercados

Niños Colegios

**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. La unidad muestral (en este tipo de muestra) se refiere al racimo por medio del cual se logra el acceso a la unidad de análisis. El muestreo por racimos supone una selección en dos etapas, ambas con procedimientos probabilísticos. En la primera, se seleccionan los racimos, siguiendo los pasos ya señalados de una muestra probabilística simple o estratificada. En la segunda, y dentro de estos racimos, se selecciona a los sujetos u objetos que van a medirse.

Para ello se hace una selección que asegure que todos los elementos del racimo tienen la misma probabilidad de ser elegidos.

¿Cómo se lleva a cabo el procedimiento de selección de la muestra?

Cuando iniciamos nuestra exposición sobre la muestra probabilística, señalamos que los tipos de muestra dependen de dos cosas: del tamaño de la muestra y del procedimiento de selección. De lo primero hemos hablado con todo detalle, de lo segundo trataremos ahora. Se determina el tamaño de la muestra *n,* pero ¿cómo seleccionar los elementos muéstrales? Se precisa el número de racimos necesario y ¿cómo se seleccionan los sujetos dentro de cada racimo? Hasta el momento sólo hemos dicho que los elementos se eligen de manera aleatoria, pero ¿cómo se hace esto? 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.

Así, en la tabla 8.2, tenemos que, de una población *N* = 53 empresas extractivas y siderúrgicas, se necesita una muestra *n* = 13 de directivos generales de tales empresas. En una lista se numeran cada una de estas empresas. En fichas aparte se sortea cada uno de los 53 números, hasta obtener los 13 necesarios (pueden ser las 13 primeras fichas que se extraigan). Los números obtenidos se verifican con los nombres y las direcciones de nuestra lista, para precisar los que serán participantes del estudio.

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. Los números *random* de la Corporación Rand fueron generados con una especie de ruleta electrónica. Existe una tabla de un millón de dígitos, publicada por esta corporación, cuyas partes se encuentran en los apéndices de muchos libros de estadística.

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. Todo esto lo hemos visto, aunque nos falta exponer sobre algo esencial que precede a la selección de una muestra: el **marco muestra!** Éste constituye un marco de referencia que nos permita identificar físicamente los elementos de la población, la posibilidad de enumerarlos y, por ende, de proceder a la selección de los elementos muéstrales (los casos de la muestra). Normalmente se trata de un listado existente o una lista que es necesario confeccionar *ad hoc,* con los casos de la población.

Los **listados** existentes sobre una población son variados:

Guías telefónicas, listas de miembros de las asociaciones, directorios especializados, listas oficiales de escuelas de la zona, bases de datos de los alumnos de una universidad o de los clientes de una empresa, registros médicos, catastros, nóminas de una organización, etc. En todo caso hay que tener en cuenta lo completo de una **lista,** su exactitud, su veracidad, su calidad y su nivel de cobertura en relación con el problema a investigar y la población que va a medirse, ya que todos estos aspectos influyen en la selección de la muestra.

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. Para establecer correlaciones se considerará como población a todas las personas contratadas durante 10 años. Se relacionan sus datos en la solicitud de empleo con los registros de faltas.

Como no hay una lista elaborada de estos individuos, el investigador decide acudir a los archivos de las solicitudes de empleo. Tales archivos constituyen su marco muestral a partir del cual se obtendrá la muestra. Calcula el tamaño de la población, obtiene el tamaño de la muestra y selecciona sistemáticamente cada elemento *XIK,* cada solicitud que será analizada. Aquí el problema que surge es que en el archivo hay solicitudes de gente que no fue contratada y, por lo tanto, no debe considerarse en el estudio. En este caso, y en otros en los que no todos los elementos del marco de referencia o de una lista aparecen (por ejemplo, nombres en el directorio que no corresponden a una persona física), los especialistas en muestreo (Kish, 1995; Sudman, 1976) no aconsejan el reemplazo con el siguiente elemento, sino simplemente no tomar en cuenta ese elemento, es decir, hacer como si no existiera, y continuar con el intervalo de selección sistemática.

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. A partir de una lista de tiendas de cada cadena competidora, marca sobre un mapa de la ciudad, todas las tiendas de autoservicios, las cuales constituyen una población de racimos, pues en cada tienda seleccionada entrevistará a un número de clientes. El mapa le permite ver la población (tiendas de autoservicio) y su situación geográfica, de manera que elige zonas donde coexistan diferentes tiendas competidoras, para asegurarse de que el consumidor de la zona tenga todas las posibles alternativas. En la actualidad hay mapas de todo tipo: mercadológicos, socioculturales, étnicos, marítimos, entre otros.

Volúmenes

En este caso supongamos que un estudioso del periodismo quiere hacer un análisis de contenido de los editoriales de los tres principales diarios de la ciudad durante los periodos del porfiriato en México, el gobierno sandinista en Nicaragua o el franquismo en España. El investigador va a la Hemeroteca Nacional y encuentra —por ejemplo— que los diarios son encuadernados por trimestre y año, lo cual le proporciona un marco de referencia ideal, a partir de donde seleccionará *n* volúmenes para su análisis. Supongamos, en el caso mexicano, que encuentra que el volumen *X,* que contiene el periódico *El Hijo del Ahuizote* (enero-marzo, 1899), falta en la hemeroteca. ¿Qué hace? Pues redefine la población, manifestando explícitamente que de *N* volúmenes tiene 99% de los elementos y, a partir de este nuevo número de *N,* calcula su muestra *n* y la selecciona.

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.

**¿Cómo y cuáles son las 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 representativos 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. La primera es que, al no ser probabilísticas, no es posible calcular con precisión el error estándar, es decir, no podemos calcular con qué nivel de confianza hacemos una estimación.

La única ventaja de una muestra no probabilística —desde la visión cuantitativa— es su utilidad para determinado diseño de estudio que requiere no tanto una **representatividad** de elementos de una población, sino una cuidadosa y controlada elección de sujetos con ciertas características especificadas previamente en el planteamiento del problema.

Una máxima del muestreo y el alcance del estudio

Ya sea que se trate de un tipo de muestreo u otro, lo importante es elegir a los informantes (o casos) adecuados, de acuerdo con el planteamiento del problema y lograr el acceso a ellos.

Los estudios exploratorios regularmente emplean muestras dirigidas o no probabilísticas, aunque podrían usarse muestras probabilísticas. Las investigaciones experimentales, la mayoría de las veces utilizan muestras dirigidas, porque como se comentó, es difícil manejar grupos grandes (debido a ello se ha insistido que, en los experimentos, la validez externa se consolida mediante la repetición o reproducción del estudio). Los estudios no experimentales descriptivos o correlaciónales-causales deben emplear muestras probabilísticas si quieren que sus resultados sean generalizados a una población.