

Metodología de la investigación

administración, economía, humanidades y ciencias sociales

César A. Bernal

Tercera
edición

PEARSON



7.9 La población y la muestra objeto de estudio. Determinación

Una vez concebida la idea de investigación, y luego de tener claridad sobre el problema que se va a investigar, plantear los objetivos que se espera lograr, contar con una justificación para desarrollar el estudio, tener un fundamento teórico, plantear la hipótesis o las preguntas de investigación, definir el tipo y el diseño de la investigación, el otro aspecto para tener en cuenta es definir la *población o muestra* con la cual se desarrollará la investigación de interés.

En esta parte de la investigación, el interés consiste en definir quiénes y qué características deberán tener los sujetos (personas, organizaciones o situaciones y factores) objeto de estudio.

En seguida, sólo se plantean algunos conceptos muy relevantes para tener en cuenta en esta etapa del proceso de investigación (la persona interesada en profundizar en cada tema en particular puede consultar expertos y/o material especializado).

Nota: a los aspectos relacionados con la población y muestra, la recopilación de la información, el procesamiento de ésta, así como al análisis y la discusión de resultados, usualmente se les conoce como *estrategias metodológicas* de la investigación.

7.9.1 Población

De acuerdo con Fracica (1988), población es “el conjunto de todos los elementos a los cuales se refiere la investigación. Se puede definir también como el conjunto de todas las unidades de muestreo” (p. 36).

Según Jany (1994), población es “la totalidad de elementos o individuos que tienen ciertas características similares y sobre las cuales se desea hacer inferencia” (p. 48); o bien, unidad de análisis.

Las dos anteriores definiciones son igualmente válidas para el propósito del presente libro.

Por ello, para estos autores una definición adecuada de población debe realizarse a partir de los siguientes términos: *elementos, unidades de muestreo, alcance y tiempo*.

Si desea hacerse un análisis del sector del cuero y el calzado en su país, la población sería:

- Alcance: cinco principales ciudades capitales.
- Tiempo: de 1996 a 2004.
- Elementos: todas las empresas del sector del cuero y el calzado ubicadas en el territorio nacional.
- Unidades de muestreo: todas las empresas del sector del cuero y el calzado en el país.

7.9.2 Marco muestral

Se refiere a la lista, el mapa o la fuente de donde pueden extractarse todas las unidades de muestreo o unidades de análisis en la población, y de donde se tomarán los sujetos objeto de estudio.

7.9.3 Muestra

Es la parte de la población que se selecciona, de la cual realmente se obtiene la información para el desarrollo del estudio y sobre la cual se efectuarán la medición y la observación de las variables objeto de estudio.

Pasos en la selección de una muestra Siguiendo el esquema de Kinnear y Taylor (1993), los siguientes son los pasos para definir una muestra:

1. Definir la población.
2. Identificar el marco muestral.
3. Determinar el tamaño de la muestra.
4. Elegir un procedimiento de muestreo.
5. Seleccionar la muestra.

Variables de la población y su medición Según Fracica (1988), “uno de los aspectos fundamentales para la realización de una investigación es la necesidad de conocer ciertas *características* de la población objeto de estudio”, a las cuales “se les conoce como variables y pueden ser de tipo cuantitativo o cualitativo” (p. 46).

Estas variables se analizan a partir de sus necesidades, ya sea en términos de datos de promedios o totales para las variables cuantitativas, y de proporciones o totales para las variables cualitativas.

EJEMPLO 7.25 Variables cualitativas y cuantitativas de la población

Un estudio busca conocer la opinión de los usuarios respecto a un nuevo producto. La variable opinión es una variable cualitativa y los datos se analizarán en términos de proporciones. Por ejemplo, 20% de los encuestados opinarán favorablemente sobre la calidad del producto.

Un estudio interesado en conocer el volumen de producción del sector automotor y sus variaciones durante los últimos tres años.

La variable volumen de producción es una variable cuantitativa y su análisis se realizará en términos de promedios o de totales; por ejemplo, el volumen de producción del sector automotor de los últimos tres años fue de 30 000 vehículos, promedio anual; y la producción total ascendió a 90 000 vehículos durante los tres años.

7.9.4 Tamaño de la muestra

En la investigación científica, el tamaño de la muestra debe estimarse siguiendo los criterios que ofrece la estadística, y por ello es necesario conocer algunas técnicas o métodos de muestreo.

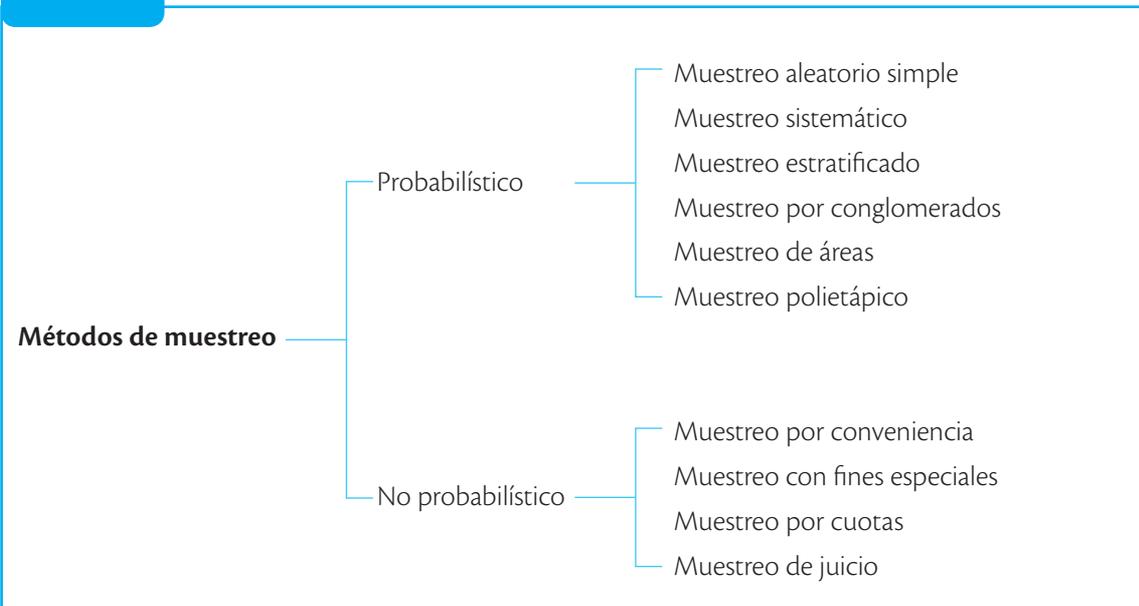
El método de muestreo utilizado para estimar el tamaño de una muestra depende del tipo de investigación que desea realizarse y, por tanto, de las hipótesis y del diseño de investigación que se hayan definido para desarrollar el estudio.

7.9.5 Métodos de muestreo

Existen varias clasificaciones para los métodos de muestreo. Según Weiers (1986), las más usadas son: diseños probabilísticos y no probabilísticos, y diseños por atributos y por variables. El primero de éstos es el más usual.

La Figura 7.13 sintetiza los métodos de muestreo más utilizados en la investigación científica para estimar tamaños de muestra en una población objeto de estudio.

FIGURA 7.13 Métodos de muestreo

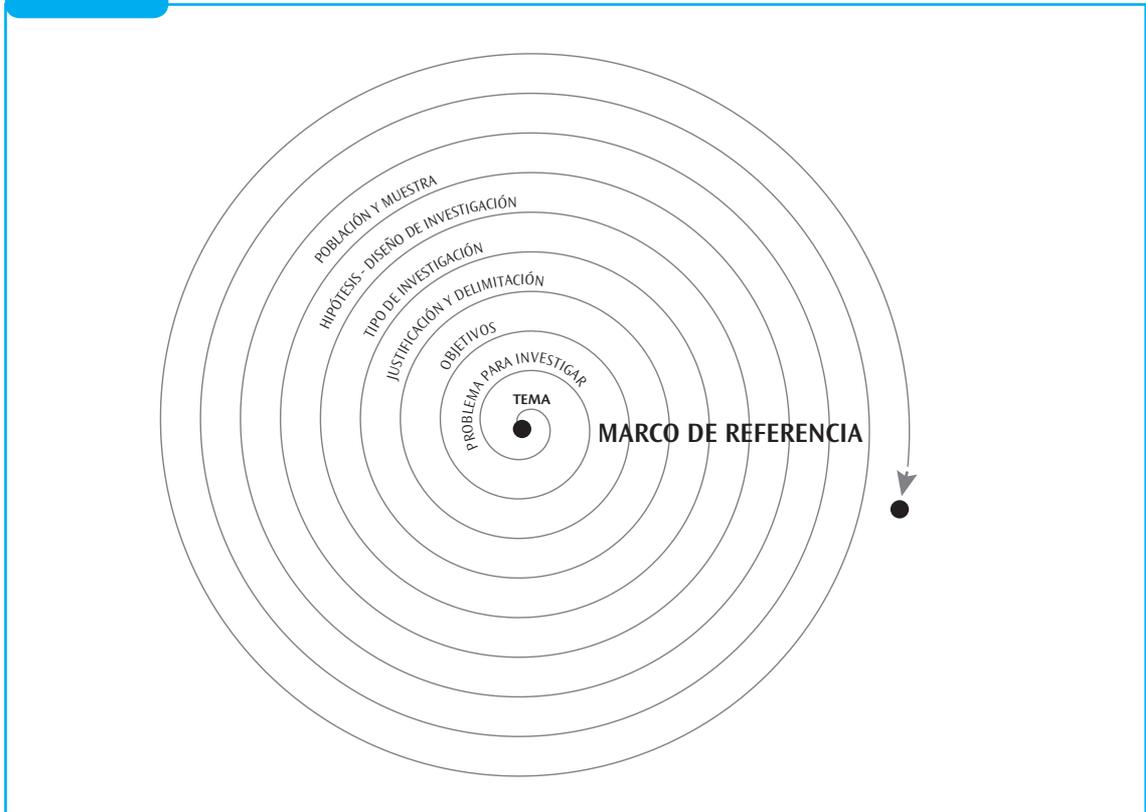


Fuente: Weiers, R. (1986). *Investigación de mercados*. México: Prentice Hall.

Así, de acuerdo con cada método de muestreo, existen criterios diferentes para estimar el tamaño de la muestra.

La Figura 7.14 ilustra la relación entre la población y muestra con los otros componentes de proceso de investigación científica según la espiral del proceso de investigación científica.

FIGURA 7.14 El proceso de investigación. Determinación de la población y la muestra objeto de estudio



Casos y talleres

Cómo estimar el tamaño de una muestra estadística representativa de una población

Muestreo aleatorio simple (MAS)

En investigación, el muestreo aleatorio simple se utiliza cuando en el conjunto de una población, cualquiera de los sujetos tiene la variable o variables objeto de la medición.

Procedimiento para estimar el tamaño de muestra

Para estimar el tamaño de muestra representativo de una población, mediante el muestreo aleatorio simple (MAS), se sigue este procedimiento:

1. Se identifica si se trata de una población infinita (no se conoce el número total de la población N) o finita (se conoce el número total de la población N) y también si la variable objeto del estudio es cuantitativa o cualitativa. Este paso es importante para definir la fórmula por utilizar (la fórmula para estimar la muestra en una población infinita es diferente de la que se utiliza para estimar la muestra en poblaciones finitas) y la forma de estimar el valor del error de estimación que se va a utilizar en la respectiva fórmula (cuando la variable es cualitativa se utiliza un valor en porcentaje no superior a 10%, mientras que en el caso de las variables cuantitativas, el valor del error de estimación también es máximo 10%, pero es un valor absoluto estimado con base en el valor de la media).
2. Se determina la desviación estándar de la población que se puede obtener mediante resultados de estudio previos o mediante la realización de una “muestra piloto”. Cuando se estima la desviación estándar de la población mediante la muestra piloto, se comienza por construir una tabla de frecuencias, resultado de la medición de la variable objeto del estudio (X_i).
3. Se define el nivel de confianza Z que se debe utilizar, el cual se encuentra en la *función de distribución normal tipificada* o *tabla Z*, que hoy día aparece en los libros de estadística y en varios de los libros de metodología de investigación. En este libro se encuentra como Anexo C. Para efectos de investigación, los valores de Z que se van a utilizar corresponden a niveles de confianza iguales o superiores a 90%, que en valores tipificados Z son valores iguales o superiores a 1,96.
4. Se define el error de estimación E (diferencia máxima entre la media muestral y la media poblacional que está dispuesto a aceptar el investigador, dependiendo del nivel de confianza (valores Z) y del valor estimado de la desviación S). En investigación científica, al error de estimación E se le asignan valores menores o iguales a 10% en relación inversa con los valores de Z y S , es decir, cuando se asigna un valor de nivel de confianza pequeño y la desviación estándar es grande, el error de estimación

E^* debe ser pequeño. El valor estimado del coeficiente de S/\bar{X} permite definir cuándo la desviación estándar de una población es grande o pequeña. En los casos en los que el valor de S/\bar{X} son menores o iguales que 30%, se dice que la desviación estándar es pequeña (se entiende que la respectiva población tiende a ser homogénea), mientras que cuando la relación S/\bar{X} es mayor que 30% pero menor que 60%, se dice que la desviación estándar es grande (se considera que la respectiva población tiende a ser heterogénea), pero si la relación S/\bar{X} es superior a 60%, entonces se dice que la población objeto del estudio es heterogénea y en ese caso se recomienda una muestra piloto mayor y asignar nivel de confianza superior a 95% y error de estimación pequeño –menor o igual que 3%.

5. Se calcula el valor del tamaño de muestra representativo según la fórmula correspondiente y se concluye sobre el tamaño mínimo de muestra requerido especificando los respectivos valores de nivel de confianza (Z), desviación estándar, estimado (S) y error de estimación utilizado (E).

Caso 7.1 Introducción de una nueva tarjeta de crédito- Banco JTRQP

Usted(es) ha(n) sido contratado(a)s por una reconocida entidad financiera (Banco JTRQP) con casa matriz en Estados Unidos, que quiere iniciar operaciones de banca en Colombia con motivo de la firma del TLC. El respectivo banco quiere realizar un estudio de mercados que le permita evaluar el gasto promedio mensual (en \$) que actualmente realizan las personas (hombres y mujeres con edades entre 25 y 40 años) por utilización de tarjeta de crédito. Su función es apoyar al gerente de Marketing para estimar el tamaño de muestra representativo con el fin de realizar el respectivo estudio que tendrá lugar en la ciudad de *BGTX*. Por favor, estime(n) el tamaño de la respectiva muestra representativa para el grupo de población objetivo por encuestar.

Procedimiento

1. Se identifica si la población es infinita o finita (para el caso se estimará la muestra en las dos condiciones) y se define la variable objeto de medición (X_i) (en este caso, esa variable se denomina “Gasto promedio mensual por uso de tarjeta de crédito”), la cual es una variable cuantitativa y se define la fórmula del muestreo que se va a utilizar.
2. Se determina el valor de la desviación estándar, para el caso, valor que se obtendrá mediante muestra piloto: se mide la variable objeto del estudio en por lo menos 30 personas con las características de la respectiva población sobre la cual se realizará el estudio. En este caso, para estimar la desviación estándar se ha decidido realizar una muestra piloto a 57 personas con edades entre 25 y 40 años, residentes en la ciudad de *BGTX* y usuarios de tarjeta de crédito. A continuación, se muestra la tabla de frecuencias que se obtuvo de esa muestra piloto.

* El valor del error de estimación E se define de forma diferente para las variables cuantitativas que para las variables cualitativas. En el caso de las variables cuantitativas, el valor de E se obtiene multiplicando el valor del promedio de la respectiva variable por el porcentaje definido para realizar el estudio en función de los valores del nivel de confianza Z y del valor de la desviación estándar estimado S , mientras que para las variables cualitativas, el valor de E se obtiene del valor en porcentaje que se defina, también en función de los valores de Z y S .

TABLA 7.4

Gasto promedio mensual por uso de tarjeta de crédito en personas con edades entre 25 y 40 años en la ciudad de BGTX

No. de encuestas	x_i = gasto promedio/ mes (miles de \$)	No. de encuestas	x_i = gasto promedio/ mes (miles de \$)
1	200	30	1 300
2	450	31	900
3	370	32	1 200
4	1 000	33	350
5	800	34	250
6	400	35	1 100
7	300	36	700
8	250	37	500
9	900	38	700
10	1 100	39	800
11	750	40	500
12	500	41	300
13	300	42	450
14	500	43	800
15	800	44	700
16	750	45	300
17	350	46	250
18	1 200	47	300
19	250	48	750
20	600	49	900
21	750	50	600
22	900	51	1 200
23	450	52	900
24	1 100	53	1 100
25	650	54	1 250
26	350	55	900
27	500	56	700
28	900	57	300
29	450	Total: $n_0 = 57$	$\bar{X} = \sum x_i/n_0 = 663,51$

3. Se estima el valor de la respectiva desviación estándar mediante la siguiente fórmula:

$$S = \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 / n_0 - 1} \implies S = \$312,12$$

Donde:

S : desviación estándar de la muestra piloto de personas con edades entre 25 y 40 años, residentes en la ciudad de *BGTX* y usuarios de tarjeta de crédito.

X_i : gasto promedio de cada una de las personas encuestadas en la muestra piloto.

\bar{X} : gasto promedio del conjunto de las personas encuestadas de la muestra piloto.

n_0 : número de personas entrevistadas en la muestra piloto.

$$\text{Homogeneidad o heterogeneidad de la muestra} = \frac{S}{\bar{X}} \implies \frac{312,12}{663,51} > 30\%$$

En este caso, debido a que el valor de la relación S/\bar{X} es superior al 30%, se considera que la población tiende a ser heterogénea y el valor que se asigne a E (error de estimación) se sugiere que debe ser pequeño, en lo posible menor que 5% (el valor máximo de E es 10%).

4. Estimación del tamaño de la muestra

a. Población infinita

En este caso, se supone que **no** se conoce el número total de personas (N) con edades entre 25 y 40 años, residentes en la ciudad de *BGTX* y usuarios de tarjeta de crédito; entonces, para estimar el tamaño de la muestra se utiliza la fórmula de muestreo aleatorio simple para población infinita, como se muestra a continuación:

$$n = \frac{Z^2 \times S^2}{E^2}$$

Donde:

n : tamaño necesario de la muestra de personas con edades entre 25 y 40 años, residentes en la ciudad de *BGTX* y usuarios de tarjeta de crédito.

Z : nivel de confianza o margen de confiabilidad (para este caso: 96% de confianza, $Z = 2,054$).

S : desviación estándar de la población (estimada mediante una muestra piloto aplicada a 57 personas con edades entre 25 y 40 años, residentes en la ciudad de *BGTX* y usuarios de tarjeta de crédito) $\implies S = 312,12$.

E : error de estimación (en este caso, se tomó un error de estimación pequeño debido a que la población objeto del estudio tiende a ser heterogénea en cuanto a gasto a través del uso de tarjeta de crédito. Entonces, $E = 0,03$ del promedio del gasto en el respectivo uso de tarjeta).

$$n = \frac{Z^2 \times S^2}{E^2}, \implies n = \frac{(2,054)^2 \times (312,12)^2}{(0,03 \times 663,51)^2} \implies n = 1\,038 \text{ personas}$$

5. **En conclusión**, utilizando un nivel de confianza del 96% y un error de estimación del 3% se requiere entrevistar a por lo menos 1 038 personas de entre 25 y 40 años, residentes en la ciudad de *BGTX*, usuarios de tarjeta de crédito, para conocer el gasto promedio mensual en tarjeta de crédito de las personas en el respectivo rango de edad.

b. Tamaño de muestra cuando la población es finita (se conoce N)

En el caso que se conozca el total de la población, es decir, el número total de personas con edades entre 25 y 40 años, residentes en la ciudad de *BGTX* y usuarios de tarjeta de crédito, para estimar el tamaño de la muestra se utiliza la fórmula de muestreo aleatorio simple, como se muestra a continuación:

$$n = \frac{S^2}{\frac{E^2}{Z^2} + \frac{S^2}{N}}$$

Donde:

N : número total de personas entre 25 y 40 años de edad, residentes en la ciudad de *BGTX* y usuarios de tarjeta de crédito. En este caso se supone un número total de 780 000 personas con el respectivo perfil.

n : tamaño necesario de la muestra de personas de entre 25 y 40 años de edad, residentes en la ciudad de *BGTX* y usuarios de tarjeta de crédito.

Z : nivel de confianza o margen de confiabilidad (en este caso: 96% de confianza, $Z = 2,054$).

S : desviación estándar de la población (estimada mediante una muestra piloto aplicada a 57 personas de entre 25 y 40 años de edad, residentes en la ciudad de *BGTX* y usuarios de tarjeta de crédito) $\implies S = 312,12$.

E : error de estimación (en este caso, se tomó un error de estimación pequeño debido a que la población objeto del estudio tiende a ser heterogénea en cuanto a gasto a través del uso de tarjeta de crédito: $E = 0,03^*$ del promedio del gasto en el respectivo uso de tarjeta).

* Recuérdese que para el caso de estimación de tamaños de muestra donde la variable objeto de medición es una variable cuantitativa, el error de estimación se obtiene mediante el porcentaje (máximo 10%) de la media de la muestra piloto de la población.

El procedimiento para la estimación del respectivo tamaño de muestra es el mismo realizado para estimar tamaño de muestra para la población infinita, pero la fórmula es diferente y como sigue:

$$n = \frac{S^2}{\frac{E^2}{Z^2} + \frac{S^2}{N}} \quad n = \frac{(312,12)^2}{\frac{(0,03 \times 663,51)^2}{(2,054)^2} + \frac{(312,12)^2}{780\,000}} \quad n = 1\,036 \text{ personas}$$

En conclusión, utilizando un nivel de confianza del 96% y un error de estimación del 3% para una población de 780 000 personas con edades de entre 25 y 40 años, residentes en la ciudad de *BGTX*, usuarios de tarjeta de crédito, se requiere entrevistar a por lo menos 1 036 personas para conocer el gasto promedio mensual en tarjeta de crédito por parte de las personas en el respectivo rango de edad en la mencionada ciudad.

Caso 7.2 Percepción sobre la gestión del conocimiento

Suponga que usted tiene interés en realizar un estudio cuyo objetivo es conocer la percepción que tienen las personas que laboran en una reconocida compañía del sector farmacéutico en esta ciudad, sobre la forma como en ella se gestiona el conocimiento. Su función es estimar el tamaño de muestra representativa de trabajadores (directivos y no directivos) de la respectiva compañía a quienes para tal propósito se les aplicará una encuesta conformada por 58 enunciados estructurados en escala Likert, donde el encuestado califica cada enunciado según su percepción, en puntuaciones de 1 a 5, siendo:

1 = Pésima; 2 = Mala; 3 = Regular; 4 = Buena; 5 = Excelente.

Procedimiento

1. Se identifica si la población es infinita o finita. Para el caso, se puede saber el número total de personas que trabajan en la compañía en la ciudad, y por ello se dirá que es una población finita. La variable objeto de medición (X_i) –en este caso se denomina “Percepción sobre la forma de gestionar el conocimiento”– es una variable cualitativa* y es un criterio clave para definir el error de estimación**.
2. Se determina el valor de la desviación estándar, para el caso, valor que se obtendrá mediante muestra piloto: se mide la variable objeto del estudio en por lo menos 30 personas con las características de la respectiva población sobre la cual se realizará el estudio. En este caso, para estimar la desviación estándar se aplica la encuesta a una muestra piloto de 40 personas que laboren en la compañía. El siguiente es el respectivo cuadro de frecuencias de la muestra piloto sobre la percepción de las personas encuestadas sobre la forma en que se gestiona el conocimiento en la compañía. En este caso se utilizó la distribución unidimensional, pero el CD que acompaña a este libro ilustra un ejemplo más completo.

* En el CD que acompaña a este libro se ejemplifica la estimación de una muestra representativa.

** Recuérdese que para el caso de estimación de tamaños de muestra donde la variable objeto de medición es una variable cualitativa, el error de estimación se define como un valor porcentaje (máximo 10%) para variables cualitativas estructuradas en escala Likert.

TABLA 7.5 Percepción sobre gestión del conocimiento. Compañía farmacéutica

No. de encuestas	X_i Percepción	No. de encuestas	X_i Percepción
1	3	24	2
2	2	25	1
3	5	26	1
4	3	27	2
5	5	28	1
6	1	29	4
7	2	30	4
8	5	31	4
9	4	32	3
10	3	33	2
11	2	34	2
12	2	35	5
13	2	36	4
14	1	37	4
15	3	38	4
16	2	39	3
17	5	40	3
18	4	41	4
19	4	42	4
20	3	43	1
21	2	44	2
22	2	45	2
23	5	Total $n_0 = 45$	$\bar{X} = \sum x_i / n_0 = 2,92$

La fórmula para la estimación de la desviación estándar es la siguiente:

$$S = \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 / n_0 - 1} \implies S = \$1,29$$

Donde:

S : desviación estándar de la percepción sobre la forma como se gestiona el conocimiento en la compañía por las personas de la muestra piloto.

X_i : percepción de cada una de las personas encuestadas en la muestra piloto sobre la forma como se gestiona el conocimiento en la compañía.

\bar{X} : percepción promedio del conjunto de las personas encuestadas de la muestra piloto sobre la forma como se gestiona el conocimiento en la compañía.

n_0 : número de personas entrevistadas en la muestra piloto.

$$\text{Homogeneidad o heterogeneidad de la muestra} = \frac{S}{\bar{X}} \implies \frac{1,29}{2,92} > 30\%$$

En este caso, debido a que el valor de la relación S/\bar{X} es superior al 30%, se considera que la población tiende a ser heterogénea y el valor que se asigne a E (error de estimación) se sugiere que debe ser pequeño, en lo posible menor que 5% (el valor máximo de E es 10%).

- Se estima el tamaño representativo de la muestra. Se sabe que la compañía tiene una nómina de 7 584 trabajadores, es decir, este es el número total de personas que laboran en la compañía; por tanto, es una población finita y la fórmula de muestreo aleatorio simple, que se debe utilizar para estimar el tamaño de muestra representativa de personas que laboran en la compañía por encuestar para conocer la percepción sobre la forma como se gestiona el conocimiento, se presenta a continuación:

$$n = \frac{S^2}{\frac{E^2}{Z^2} + \frac{S^2}{N}}$$

Donde:

N : número total de personas que laboran en la compañía (7 584).

n : tamaño necesario de la muestra representativa de personas que se debe encuestar para determinar la verdadera percepción que tienen las personas en la compañía sobre la forma como se gestiona el conocimiento en la misma.

Z : nivel de confianza o margen de confiabilidad (para este caso a las directivas de la compañía les interesa ser rigurosas y, por ello, deciden utilizar un nivel de confianza alto: 95% de confianza, para un valor de $Z = 1,96$).

S : desviación estándar de la población (estimada mediante la muestra piloto aplicada a 45 personas $\implies S = 1,29$)

E : error de estimación (en este caso se tomó un error de estimación un valor pequeño, debido a que la población objeto del estudio tiende a ser heterogénea en cuanto a la percepción de la variable objeto de medición; entonces: $E = 0,05^*$).

$$n = \frac{S^2}{\frac{E^2}{Z^2} + \frac{S^2}{N}} \quad n = \frac{(1,29)^2}{\frac{(0,05)^2}{(1,96)^2} + \frac{(1,29)^2}{7\,584}} \quad n = 1\,912 \text{ personas}$$

* Recuérdese que para el caso de estimación de tamaños de muestra donde la variable objeto de medición es una variable cualitativa, el error de estimación se define como un valor porcentaje (máximo 10%).

En conclusión, utilizando un nivel de confianza del 95% y un error de estimación del 5% para el total de trabajadores de la compañía (7 584 personas), se requiere encuestar a por lo menos 1 912 personas para conocer la percepción representativa de los trabajadores sobre la forma como se gestiona el conocimiento en la compañía.

Caso 7.3 Percepción sobre calidad y pertinencia de las lecturas complementarias en un determinado programa académico

Con el propósito de complementar el ejemplo anterior que está desarrollado con una distribución unidimensional (una sola pregunta para todas las personas), éste es más completo y se desarrolla aprovechando el aplicativo Process en el CD que acompaña a este libro y que también ilustra casos tanto del muestreo aleatorio simple, como del muestreo proporcional y del muestreo estratificado.

El director del programa de una determinada facultad desea realizar una investigación cuyo objetivo es evaluar la actitud de los estudiantes de la facultad o programa respecto de la calidad y la pertinencia de las lecturas complementarias. Estas las recomiendan los docentes de las asignaturas o núcleos temáticos correspondientes a los distintos semestres que conforman el programa. El instrumento (cuestionario) que se va a utilizar en la recolección de la información es una encuesta conformada por 50 enunciados en escala Likert, validada tanto por un comité de jueces como por una prueba piloto, con las categorías de respuesta siguientes:

Totalmente de acuerdo	5
Parcialmente de acuerdo	4
Indiferente	3
Parcialmente en desacuerdo	2
Totalmente en desacuerdo	1

TALLER 7.1 Procesamiento de la información apoyado en el programa Process

Para desarrollar estos casos, usted puede apoyarse en la herramienta para el procesamiento de información Process, que se incluye en el CD. A continuación, se describen las instrucciones para definir las encuestas en escala de Likert.

1. Haga clic en Process 3.0, ubicado en la sección Herramientas y ejercicios del CD.
2. En la barra de menú haga clic en **Estadística descriptiva** y luego haga clic en **Procesamiento de encuestas (Escala Likert)**. (En la pantalla 7.1 se muestra el menú de la aplicación).

Pantalla 7.1 Barra de menú Process



3. En la pantalla 7.2 se presentan las condiciones generales para procesar este tipo de encuestas. De acuerdo con cada una de las opciones, ingrese la información que se relaciona a continuación y cuando haya terminado haga clic en el botón **OK**.

Proyecto nuevo

Número de encuestas	35
Número de preguntas por encuesta	50
Número de respuestas para cada pregunta	5

4. Dependiendo de la selección que realice en el cuadro **Número de respuestas para cada pregunta**, el cuadro de diálogo **Rango de respuestas para escalas de Likert** se actualiza, proponiendo una serie de rangos de respuesta y la codificación numérica que utilizará el programa para realizar los cálculos (columna **Códigos**). Si en el cuadro **Número de respuestas para cada pregunta** selecciona las opciones 2, 3 ó 5, el programa ajusta automáticamente el cuadro de diálogo **Rangos de respuesta para escalas de Likert**; en los demás casos digite el texto correspondiente a cada una de las categorías de respuesta. Si las opciones de respuesta que usted requiere en su análisis no concuerdan con las opciones propuestas, usted podrá modificar el texto de las opciones de respuesta propuestas. Para ello, haga clic sobre la categoría que desea modificar y digite el nuevo texto.

Pantalla 7.2 Condiciones iniciales (Hoja de datos)

Si va a ingresar nuevos datos, en el cuadro **Proyecto** seleccione la opción **Nuevo**. Si ha guardado previamente sus datos en un archivo, seleccione la opción **Existente** y ubique el nombre del archivo.

El cuadro **Número de preguntas por encuesta** se relaciona directamente con el cuadro de diálogo **Rangos de respuestas para escalas de Likert**. Si no aparecen opciones de respuesta, dígtelas manualmente.

El programa propone la codificación que se muestra en la columna **Códigos** para tabular las respuestas. Trabaje preferiblemente con éstas para evitar resultados inesperados.

Procedimientos para el cálculo del tamaño de la muestra

1. Se plantea la fórmula estadística apropiada para estimar el tamaño de muestra

$$n = \frac{S^2}{\frac{E^2}{Z^2} + \frac{S^2}{N}} \text{ (Tamaño de población conocido)}$$

$$n = \frac{S^2 \times Z^2}{E^2} \text{ (Cuando no se conoce el total de la población)}$$

Donde:

n : tamaño de la muestra por estimar.

S : desviación estándar de la población.

Z : margen de confianza.

E : error de estimación o diferencia máxima entre la media muestral y la media poblacional que el investigador está dispuesto a aceptar con el nivel de confianza que se defina.

N : número total de la población objeto del estudio.

2. Se estima el valor de S (desviación estándar); el valor de S se puede obtener de dos formas: (a) conocida o estimada a partir de estudios anteriores y (b) estimación mediante una muestra piloto.

En este caso no existen estudios previos, por lo que es necesario realizar una *muestra piloto* para estimar el valor de S ; este valor se estima de dos formas: (a) estimando la desviación estándar para cada una de las preguntas del cuestionario para todos los encuestados y tomar el mayor valor de las diferentes desviaciones por pregunta para estimar el tamaño de la muestra, o (b) estimando la desviación estándar del promedio de total de las encuestas de todos los encuestados, como se ilustra a continuación, mediante la siguiente fórmula:

$$S = \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 / n_0 - 1}$$

Donde:

S : desviación estándar de la población a partir de la muestra piloto.

X_i : puntuación o medición de la variable objeto de estudio (en este ejemplo, medición de la actitud respecto a la lectura complementaria) para cada uno de los estudiantes de la muestra piloto.

\bar{X} : promedio de la población o de la medición de la variable objeto de estudio de cada uno de los estudiantes de la muestra piloto.

n_0 : número de estudiantes de la muestra piloto (los criterios estadísticos indican que la muestra piloto de un estudio la constituyen un número igual o superior a 30 sujetos tomados aleatoriamente de la población y sobre los cuales se recoge la información para estimar el valor de S).

Veamos entonces cómo se estima el valor de S para este caso mediante muestra piloto.

Se aplica la encuesta de las 50 preguntas a un número igual o superior a 30 estudiantes (en este caso se aplicará a 35 estudiantes) seleccionados aleatoriamente y luego se calcula la puntuación promedio de cada una de las encuestas aplicadas a los 35 estudiantes y ese valor es el valor de los X_i . (Para realizar estos cálculos hay que apoyarse en la herramienta de procesamiento de información Process, incluida en el CD).

TALLER 7.2 Procesamiento de encuestas

1. En la **Matriz de respuestas y valores acumulados**, digite los resultados correspondientes a las preguntas de cada una de las encuestas así: la encuesta 1 se diligencia en dirección izquierda-derecha, en el primer renglón, digitando los códigos de las respuestas definidos en la pantalla 7.2. Por ejemplo, si usted seleccionó la opción 2 del cuadro **Número de respuestas para cada pregunta**, las **Categorías- respuesta** son: **Sí, No, Pregunta no contestada**. Por tanto, los códigos que se deben digitar en la matriz son valores entre 0 y 2, pues así se codificó.

Los títulos que aparecen en la parte superior (en el encabezado de cada columna) representan el número de la pregunta; por ejemplo, 1 representa la respuesta codificada a la pregunta R1 (un valor entre 0 y 2); 2 representa el código de respuesta a la pregunta R2, y así sucesivamente, hasta finalizar la totalidad de la encuesta. En la pantalla 7.3 se ilustra la matriz diligenciada.

2. Desplace el cursor de una casilla a otra con las teclas de dirección (flechas derecha-izquierda y arriba-abajo).
3. Para realizar los cálculos, una vez se ingresen las respuestas de todas las encuestas en las casillas de la matriz, haga clic en el botón **Calcular**.

Pantalla 7.3 Matriz de respuesta y valores acumulados

Procesamiento de datos para encuestas tipo Licker:

Matriz de respuestas y valores acumulados:

No	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	Promedi
1	3	2	1	3	2	1	2	3	1	4	2	3	
2	3	1	2	1	4	5	0	3	1	1	2	3	
3	2	1	4	2	3	1	2	3	5	5	4	5	
4	2	4	2	4	3	4	3	2	1	4	2	3	
5	4	5	5	5	4	4	3	3	2	2	1	1	
6	4	1	5	3	5	5	4	3	5	4	3	5	

Tabla de distribución de frecuencias:

Yi	ni	hi	Ni	Hi

Medidas de dispersión:

Calificación máx.

Calificación mín.

Rango

Desv. estándar

Varianza

Tendencia:

Moda

Media

Label1

Análisis de una pregunta:

No. preg.

Digite aquí las respuestas codificadas, de acuerdo con los códigos de la pantalla 7.2.

*Con la opción **Grabar datos como**, usted puede archivar sus datos y recuperarlos posteriormente.*

*Una vez que haya digitado la totalidad de los datos, haga clic en el botón **Calcular**.*

Muy bien, ya se ha procesado parte de la información. La pantalla 7.4 contiene el valor correspondiente a la desviación estándar, la misma que se necesitará posteriormente para calcular el tamaño definitivo de la muestra.

Las operaciones necesarias para determinar el promedio de calificación de cada estudiante ya han sido realizadas por el programa, así como la aplicación de las fórmulas para definir el valor de la media y de la moda.

Desde este punto, usted también puede examinar el comportamiento de una única pregunta. Seleccione la pregunta que desea analizar y haga clic sobre el botón **Analizar**.

Pantalla 7.4 Comportamiento de datos agrupados, desviación estándar y análisis de preguntas

Process 3.0
Estadística descriptiva - Distribuciones - Regresión y correlación - Muestreo - Pruebas de hipótesis - Salir

Hoja de datos(psicologia.txt)

Procesamiento de datos para encuestas tipo Licker:

Matriz de respuestas y valores acumulados:

No	0	R1
29	3.8	
30	4.8	
31	3.1	
32	3	
33	4.5	
34	3.5	
35	4	

Tabla de distribución de frecuencias:

Yi	ni	hi	Ni	Hi
3.00	4	0.11	4	0.11
3.10	1	0.03	5	0.14
3.50	4	0.11	9	0.25
3.80	4	0.11	13	0.36
3.90	2	0.06	15	0.42
4.00	5	0.14	20	0.56
4.10	1	0.03	21	0.59

Medidas de dispersión:

Calificación máx. 4.80
Calificación mín. 3.00
Rango 1.8
Desv. estándar 0.494
Varianza 0.244

Tendencia:
Moda 4.20
Media 3.93
Actitud: muy favorable

Analizar una pregunta:
No. preg. 1
Analizar

Grabar datos Salir Gráfica Calcular

Para analizar el comportamiento de una única pregunta, seleccione el número correspondiente a la pregunta que desea y haga clic sobre el botón **Analizar**.

En la tabla de distribución de frecuencias puede observar el comportamiento de los datos agrupados.

El valor de la desviación estándar servirá posteriormente para calcular el tamaño de la muestra.

Procedimientos para el cálculo del tamaño de la muestra (continuación)

La desviación estándar indica el grado de dispersión de la muestra en relación con la media; es decir, el grado de homogeneidad o heterogeneidad de la muestra. Cuando la S = grande, la población tiende a ser heterogénea y cuando la S = pequeña, la población tiende a ser homogénea.

- Se define el nivel de confianza en valores Z . El nivel de confianza o valor de Z es cualquier valor de la tabla para las puntuaciones Z , que son valores normalizados. Sin embargo, en investigación estadísticamente se suele tomar valores para niveles de confianza igual o superior al 90%. Para este estudio el investigador ha seleccionado un nivel de confianza del 95%, por lo que $Z = 1,96$.
- Se define el valor del error de estimación E . El valor de E se define a partir de la desviación estándar (S) estimada a partir de la muestra piloto, y usualmente se recomienda darle al error de estimación un valor inferior a 10% (cuando la variable objeto de medición es cualitativa sólo se coloca el valor del porcentaje definido, pero cuando la respectiva variable es cuantitativa el valor de E es igual al porcentaje asignado multiplicado por el valor de la media (promedio) de la muestra).

Si la desviación estándar es grande en proporción con la media, el error de estimación (E) debe ser pequeño, tendiendo a cero (0); si el valor de la desviación estándar (S) es pequeño en relación con la media, entonces el valor de error de estimación (E) puede ser cercano pero no superior a 10% del valor de la media o promedio.

En esta investigación, como el valor de la desviación estándar es pequeño e indica que la población de estudiantes tiende a ser homogénea, entonces el valor del error de estimación puede ser $E = 0,06$ ó $E = 6\%$ ó $0,06$.

5. Se estima el tamaño de la muestra.

$$n = \frac{S^2}{\frac{E^2}{Z^2} + \frac{S^2}{N}} \quad n = \frac{(0,49)^2}{\frac{(0,06)^2}{(1,96)^2} + \frac{(0,49)^2}{N}}$$

El valor de N (total de estudiantes) se obtuvo de los archivos que reposan en la Secretaría Académica de la Facultad para el total de alumnos matriculados en el programa objeto de estudio del año en curso. ($N = 1\,200$).

En conclusión, para realizar el estudio de evaluación de la actitud de los estudiantes del programa de psicología, respecto a la calidad y pertinencia de las lecturas complementarias recomendadas por los profesores de las diferentes áreas del conocimiento de la carrera, se requiere una muestra representativa de 214 estudiantes, con un nivel de confianza del 95% y error de estimación del 0,06 ó 6%.

TALLER 7.3 Determinación del tamaño de la muestra (MAS)

Para determinar el tamaño de la muestra con el método de muestreo aleatorio simple, abra la aplicación Process y en la barra de menú haga clic en **Muestreo**, luego haga clic en la opción **Aleatorio simple**.

Pantalla 7.5 Opciones de muestreo



Ahora que ya definió los valores necesarios para realizar estos cálculos, digítelos en cada una de las casillas correspondientes. Una vez complete los datos, haga clic sobre el botón **Calcular**.

Pantalla 7.6 Margen de confiabilidad

El valor correspondiente al margen de confiabilidad se ajusta automáticamente, de acuerdo con el nivel de confianza dado.

Muestreo proporcional

El muestreo proporcional se utiliza cuando la variable objeto de la medición se mide en proporciones o probabilidades de ocurrencia.

Procedimiento para estimar el tamaño de muestra representativo

Para estimar el tamaño de muestra representativa de una población, con muestreo proporcional, igual que en el muestreo aleatorio simple, se procede de la siguiente manera:

1. Se identifica si se trata de una población infinita (no se conoce el número total de la población N) o finita (cuando se conoce el número total de la población N). Recuérdese que este paso es importante para definir la fórmula que se debe utilizar, ya que es distinta la fórmula de tamaño de muestra para poblaciones infinitas que para finitas.
2. Se determina la proporción de la población que tiene la variable objeto de medición, la cual se puede obtener mediante resultados de estudios previos o mediante la realización de una “muestra piloto”. Cuando se estima la proporción de la población mediante la muestra piloto, se comienza por construir una tabla de frecuencias, resultado de la medición de la variable objeto del estudio (P_i).

3. Se define el nivel de confianza Z por utilizar, el cual se encuentra en la *función de distribución normal tipificada* o *Tabla Z*, igual como se hace para el muestreo aleatorio simple. Para efectos de investigación, los valores de Z por utilizar corresponden a niveles de confianza iguales o superiores a 90% que en valores tipificados Z son valores iguales o superiores a 1,64.
4. Se define el valor del error de estimación E (diferencia máxima entre la proporción muestral y la proporción poblacional que el equipo investigador está dispuesto aceptar en función del nivel de confianza definido para el estudio). En muestreo proporcional, el valor de E debe ser menor o igual que 6% ó 0,06, en relación directa con los valores de Z , es decir, cuando se define un valor de nivel de confianza bajo cercano al 90% el error de estimación E debe ser pequeño.
5. Se calcula el valor del tamaño de muestra según la fórmula correspondiente a la población (infinita o finita) y se concluye sobre el tamaño mínimo de muestra representativo especificando los valores asignados al nivel de confianza (Z), la proporción estimada (P) y el error de estimación (E) utilizado.

Para ejemplificar la estimación de un tamaño de muestra representativo con muestreo proporcional, a continuación se presenta el caso 7.4, tanto para población infinita como para población finita y simulando cuando no se conoce el valor de la proporción (P) y se estima por muestra piloto, cuando el valor de P se obtiene de estudios se toma dicho valor y cuando no se conoce P , se puede asignarle a la respectiva proporción el valor de 0,5 ó 50%.

Caso 7.4 Mujeres cabeza de familia cotizantes a salud

Una organización encargada de la protección social en la ciudad está interesada en estimar el tamaño de muestra representativo que le permita conocer la proporción de personas mujeres cabeza de familia de estrato 1 y 2, que son cotizantes para los servicios de salud de los miembros de su familia.

Procedimiento para estimar el tamaño de la muestra representativa para una población infinita

El siguiente es el procedimiento para la estimación de la respectiva muestra cuando la población es infinita y no se tiene referencia de una proporción de mujeres cabeza de familia cotizantes de servicios de salud para los miembros de su familia en su ciudad.

1. Se construye la tabla de distribución de frecuencias para las mujeres cabeza de familia de los estratos 1 y 2, que son cotizantes para los servicios de salud para los miembros de su familia en la ciudad. Para ello, se realizará una muestra piloto de mujeres de la población objeto del estudio para saber la proporción de mujeres cabeza de familia que cotizan servicios de salud. Después se les pregunta a las mujeres seleccionadas si cotizan (1) o no (2) como cabeza de familia.

TABLA 7.6

Proporción de mujeres cabeza de familia en los estratos 1 y 2 que cotizan servicio de salud para los miembros de su familia

No. de encuestas	Cotiza salud	No. de encuestas	Cotiza salud
1	1	32	2
2	1	33	1
3	2	36	2
4	1	35	1
5	2	36	1
6	1	37	1
7	1	38	1
8	1	39	1
9	1	40	1
10	2	41	2
11	2	42	2
12	1	43	2
13	1	44	1
14	1	45	1
15	2	46	1
16	2	47	1
17	1	48	1
18	2	49	1
19	1	50	1
20	1	51	1
21	1	52	2
22	1	53	1
23	1	54	1
24	1	55	1
25	2	56	2
26	2	57	2
27	2	58	2
28	2	59	1
29	2	60	1
30	2	61	1
31	2	62	1

No. de encuestas	Cotiza salud	No. de encuestas	Cotiza salud
63	2	70	1
64	2	71	1
65	1	72	1
66	1	73	2
67	2	74	2
68	2	75	1
69	2		
		Total No. = 75	$P = \frac{\text{Total enc., con 1}}{\text{Total de encuestas}}$ $P = 50 / 75 = 0,666$

2. Se recurre a la fórmula para estimar el tamaño de muestra.

$$n = \frac{Z^2 \times P \times Q}{E^2}$$

Donde:

n : tamaño de muestra por estimar (número de mujeres cabeza de familia de estrato 1 y 2 que es necesario encuestar para conocer la proporción representativa de mujeres del estrato 1 y 2 que cotizan a salud como cabeza de familia).

Z : nivel de confianza o margen de confiabilidad (97%, es decir, $Z = 2,17$).

P : proporción de mujeres del estrato 1 y 2 de la muestra piloto que cotizan a salud como cabeza de familia.

$Q = 1 - P$: proporción de mujeres del estrato 1 y 2 de la muestra piloto que no cotizan a salud como cabeza de familia.

E : error de estimación (diferencia máxima entre la proporción muestral y la proporción poblacional que el equipo investigador está dispuesto a aceptar en función del nivel de confianza definido para el estudio). En este caso, $E = 0,03$ ó 4%. Entonces 3%.

$$n = \frac{Z^2 \times P \times Q}{E^2} \quad n = \frac{2,17^2 \times 0,666 \times 0,334}{0,03^2} \quad n = 1\ 164$$

En conclusión, con un nivel de confianza del 97%, un error de estimación del 3% es necesario encuestar un total de 1 164 mujeres de estrato 1 y 2 cotizantes a salud para conocer la proporción representativa de mujeres de los respectivos estratos, en la ciudad que cotizan a salud como cabeza de familia.

Procedimiento para estimar el tamaño de la muestra representativa para una población finita

El procedimiento para estimar el tamaño de muestra representativo es igual que para las poblaciones infinitas, pero la fórmula es diferente, así:

$$n = \frac{Z^2 \times P \times Q \times N}{E^2(N-1) + Z^2 \times P \times Q}$$

Donde:

n : tamaño de muestra por estimar (número de mujeres cabeza de familia de estrato 1 y 2 que es necesario encuestar para conocer la proporción representativa de mujeres del estrato 1 y 2 que cotizan a salud como cabeza de familia).

Z : nivel de confianza o margen de confiabilidad (97%, es decir, $Z = 2,17$).

P : proporción de mujeres del estrato 1 y 2 de la muestra piloto que cotizan a salud como cabeza de familia.

$Q = 1 - P$: proporción de mujeres del estrato 1 y 2 de la muestra piloto que no cotizan a salud como cabeza de familia.

N : número total de mujeres del estrato 1 y 2 que cotizan a salud en la ciudad.

E : error de estimación (diferencia máxima entre la proporción muestral y la proporción proporcional que el equipo investigador está dispuesto aceptar en función del nivel de confianza definido para el estudio). En este caso $E = 0,03$ ó 3%.

$$n = \frac{Z^2 \times P \times Q \times N}{E^2(N-1) + Z^2 \times P \times Q} \quad n = \frac{2,17^2 \times 0,666 \times 0,334 \times 580\,000}{0,03^2(580\,000 - 1) + 2,17^2 \times 0,666 \times 0,334}$$

$n = 1\,162$ mujeres de los estratos 1 y 2 que cotizan a salud en la respectiva ciudad.

En conclusión, para una población de 580 000 mujeres de los estratos 1 y 2, con un nivel de confianza del 97% y un error de estimación del 3%, es necesario encuestar un total de 1 162 mujeres de estrato 1 y 2 cotizantes a salud para conocer la proporción representativa de mujeres de los respectivos estratos en la ciudad que cotizan a salud como cabeza de familia.

Muestreo estratificado

En investigación científica se utiliza el muestreo estratificado para estimar tamaños de muestra cuando en el estudio la población objeto de la investigación existe alguna variable altamente correlacionada con la variable objeto de medición y esa variable hace que los sujetos sean altamente heterogéneos, por lo que se recomienda agrupar los respectivos sujetos por rangos o estratos. Por ejemplo, alguien quiere conocer la capacidad de compra de un determinado producto y sabe que la variable correlacionada con la capacidad de consumo es el ingreso y éste es distinto para las personas de esa población; entonces, en ese caso, es necesario estratificar los sujetos de la población por rangos de ingreso.

En otro caso, se requiere conocer la percepción sobre la prestación de un servicio de diversión que incluye adultos y niños; entonces, la edad influye en la percepción y por ello se recomienda agrupar la población por rangos de edad.

El procedimiento para estimar el tamaño de muestra en el muestreo estratificado es el siguiente:

1. Se identifica si existe alguna variable correlacionada con la variable objeto de medición que haga que los sujetos de la población sean heterogéneos respecto a esa variable.
2. Se determina si la población es infinita o finita (para el caso se estimará la muestra únicamente para el caso de una población finita) y se define la variable objeto de medición (X_i).
3. Se determina el tamaño total de la muestra (el cual se estima de la misma forma que para el muestreo aleatorio simple) estimando la desviación estándar de la población mediante una muestra piloto; asignando el respectivo nivel de confianza Z que se va a utilizar (niveles de confianza iguales o superiores a 90% que en valores tipificados Z son valores iguales o superiores a 1,64 y el error de estimación E), valores menores o iguales a 10% en relación inversa con los valores de Z y S .
4. Se estratifica la población según la variable correlacionada con la variable objeto de la medición y que hace que los sujetos sean heterogéneos entre sí.
5. Se estima el valor del tamaño de la muestra representativo para cada estrato a partir del tamaño de muestra total de la población según la fórmula correspondiente del muestreo estratificado, especificando los respectivos valores de nivel de confianza (Z), desviación estándar estimada (S) para cada estrato.

Caso 7.5 Inversión en libros de texto por estudiantes universitarios en la ciudad XTG, año 2010

La empresa editorial RGTBF S. A. está interesada en realizar una investigación sobre la adquisición de libros de texto por estudiantes universitarios de los programas de psicología en la ciudad XTG, en el año 2010. Para ello ha decidido estimar un tamaño de muestra de estudiantes vinculados a los programas de psicología (se sabe que los programas son ofrecidos por universidades con valor de matrícula claramente diferenciada), por lo cual considera conveniente estratificar a los estudiantes por estratos de matrículas, así: “universidades tipo A” aquellas con valor de matrícula más alto, “universidades tipo B” las de valor de matrícula medio y “universidades tipo C” las de valor de matrícula más bajo.

Estimación del tamaño de muestra para el total de la población a partir de una muestra piloto

TABLA 7.7 Inversión en libros de texto por parte de estudiantes universitarios en la ciudad XTG, año 2010.

No. de encuestas	Inversión en libro/ semestre (miles de \$)	No. de encuestas	Inversión en libro/ semestre (miles de \$)
1	1 500	33	850
2	200	34	600
3	750	35	600
4	200	36	600
5	600	37	1 100
6	350	38	450
7	120	39	600
8	1 300	40	1 200
9	1 700	41	950
10	250	42	1 300
11	180	43	120
12	450	44	540
13	590	45	610
14	800	46	900
15	900	47	750
16	1 170	48	225
17	200	49	690
18	150	50	350
19	450	51	900
20	600	52	810
21	400	53	450
22	350	54	345
23	270	55	210
24	620	56	190
25	1 600	57	1 100
26	1 330	58	900
27	1 250	59	700
28	245	60	100

No. de encuestas	Inversión en libro/ semestre (miles de \$)	No. de encuestas	Inversión en libro/ semestre (miles de \$)
29	460	61	210
30	700	62	350
31	1500	63	500
32	900	Total No. = 63	$\bar{X} = \sum x_i / No.$ $\bar{X} = 655,3$

El valor de la desviación estándar estimado a partir de una muestra piloto es:

$$S = \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 / n_0 - 1} \implies S = \$ 412,87$$

Para la estimación del total de la muestra piloto de la población se sabe que el total de los estudiantes matriculados en los respectivos programas académicos de las universidades objeto del estudio para el año 2010 es de 9 700 en la mencionada ciudad; entonces, la fórmula para estimar el total de la muestra es:

$$n = \frac{S^2}{\frac{E^2}{Z^2} + \frac{S^2}{N}}$$

Donde:

N : número total de estudiantes matriculados para el año 2010 en los programas de psicología en la ciudad en las universidades objeto del estudio (9 700).

n : tamaño total necesario de la muestra de estudiantes matriculados en los programas de psicología que adquieren libros de texto por período académico para sus estudios.

Z : nivel de confianza o margen de confiabilidad (en este caso: 96% de confianza, $Z = 2,054$).

S : desviación estándar de la población (estimada mediante una muestra piloto, $S = \$412,87$).

E : error de estimación (en este caso, se tomó un error de estimación pequeño, debido a que la población objeto del estudio tiende a ser heterogénea en cuanto a inversión para adquisición de libros; entonces $E = (0,03 \times 655,32)^*$).

$$\text{Homogeneidad o heterogeneidad de la muestra} = \frac{S}{\bar{X}} \implies \frac{412,87}{655,32} > 30\%$$

* Recuérdese que para el caso de estimación de tamaños de muestra donde la variable objeto de medición es una variable cuantitativa, el error de estimación se obtiene mediante el porcentaje (máximo 10%) de la media de la muestra piloto de la población.

El procedimiento para la estimación del respectivo tamaño de muestra es el mismo realizado para estimar el tamaño de muestra para la población infinita, pero la fórmula, diferente, es la siguiente:

$$n = \frac{S^2}{\frac{E^2}{Z^2} + \frac{S^2}{N}} \quad n = \frac{(412,87)^2}{\frac{(0,03 \times 655,32)^2}{(2,054)^2} + \frac{(412,87)^2}{9\,700}} \quad n = 1\,561 \text{ personas}$$

La editorial requiere encuestar a por lo menos 1 561 estudiantes matriculados en los programas de psicología en la ciudad para el período académico de 2010. A partir de este tamaño de muestra, se debe estimar el tamaño de muestra por estrato para cada uno de los tres tipos de universidades, según valor pagado por matrícula.

Para estimar el tamaño de la muestra por tipo de universidad, se utiliza la siguiente fórmula para cada uno de los estratos o rangos en los que se agrupa la población objeto de la investigación (las fórmulas están presentadas de forma general para cualquier número de rangos o estratos)

$$n_A = \frac{(n)(N_A)(S_A)}{(N_A)(S_A) + (N_B)(S_B) + (N_C)(S_C) + (N_D)(S_D) + \dots + (N_X)(S_X)}$$

$$n_B = \frac{(n)(N_B)(S_B)}{(N_A)(S_A) + (N_B)(S_B) + (N_C)(S_C) + (N_D)(S_D) + \dots + (N_X)(S_X)}$$

$$n_C = \frac{(n)(N_C)(S_C)}{(N_A)(S_A) + (N_B)(S_B) + (N_C)(S_C) + (N_D)(S_D) + \dots + (N_X)(S_X)}$$

$$n_D = \frac{(n)(N_D)(S_D)}{(N_A)(S_A) + (N_B)(S_B) + (N_C)(S_C) + (N_D)(S_D) + \dots + (N_X)(S_X)}$$

$$n_X = \frac{(n)(N_X)(S_X)}{(N_A)(S_A) + (N_B)(S_B) + (N_C)(S_C) + (N_D)(S_D) + \dots + (N_X)(S_X)}$$

Donde:

- n : tamaño total de la muestra representativa (para este caso $n = 1\,561$ estudiantes) estimada mediante muestra piloto con nivel de confianza del 97% y un error de estimación del 3%.
- n_A : tamaño de la muestra para el estrato A (tamaño de muestra de estudiantes de psicología de las universidades tipo A).
- n_B : tamaño de la muestra para el estrato B (tamaño de muestra de estudiantes de psicología de las universidades tipo B).
- n_C : tamaño de la muestra para el estrato C (tamaño de muestra de estudiantes de psicología de las universidades tipo C).

- n_C : tamaño de la muestra para el estrato D (para el caso objeto del estudio no hay estrato D).
- n_X : tamaño de la muestra para el estrato X (para el caso objeto del estudio no hay estrato X).
- N_A : total de la población del estrato A (total de estudiantes matriculados en programas de psicología en las universidades tipo A para el año 2010, $N_A = 1\,300$ estudiantes matriculados).
- S_A : desviación estándar de la población del estrato A (la desviación se estima a partir de muestra piloto para la población del respectivo estrato, para este caso la $S_A = \$520$).
- N_B : total de la población del estrato B (total de estudiantes matriculados en programas de psicología en las universidades tipo B para el año 2010, $N_B = 2\,800$ estudiantes matriculados)
- S_B : desviación estándar de la población del estrato B (la desviación se estima a partir de muestra piloto para la población del respectivo estrato, en este caso la $S_B = \$370$)
- N_C : total de la población del estrato C (total de estudiantes matriculados en programas de psicología en las universidades tipo C para el año 2010, $N_C = 5\,600$ estudiantes matriculados).
- S_C : desviación estándar de la población del estrato C (la desviación se estima a partir de muestra piloto para la población del respectivo estrato, en este caso la $S_C = \$150$).
- N_D : total de la población del estrato D (en el caso objeto del estudio no hay estrato D).
- S_D : desviación estándar de la población del estrato D (en el caso objeto del estudio no hay estrato D).
- N_X : total de la población del estrato X (en el caso objeto del estudio no hay estrato X).
- S_X : desviación estándar de la población del estrato X (en el caso objeto del estudio no hay estrato X).

TABLA 7.8

Agrupación de estudiantes de los programas de psicología por tipo de universidad

Tipo de universidad según valor de matrícula	No. de estudiantes matriculados/ tipo de universidad	Desviación estándar por tipo de universidad (miles de \$)
Universidad tipo A	1 300	\$520
Universidad tipo B	2 800	\$370
Universidad tipo C	5 600	\$150

Entonces:

$$n_A = \frac{(n)(N_A)(S_A)}{(N_A)(S_A) + (N_B)(S_B) + (N_C)(S_C)}$$

$$n_A = \frac{(1\,561)(1\,300)(520)}{(1\,300)(520) + (2\,800)(370) + (5\,600)(150)} = 413 \text{ estudiantes}$$

$$n_B = \frac{(n)(N_B)(S_B)}{(N_A)(S_A) + (N_B)(S_B) + (N_C)(S_C)}$$

$$n_B = \frac{(1\ 561)(2\ 800)(370)}{(1\ 300)(520) + (2\ 800)(370) + (5\ 600)(150)} = 634 \text{ estudiantes}$$

$$n_C = \frac{(n)(N_C)(S_C)}{(N_A)(S_A) + (N_B)(S_B) + (N_C)(S_C)}$$

$$n_C = \frac{(1\ 561)(5\ 600)(150)}{(1\ 300)(520) + (2\ 800)(370) + (5\ 600)(150)} = 514 \text{ estudiantes}$$

En conclusión, a fin de conocer la capacidad de compra representativa de los estudiantes de los programas de psicología para adquisición de libros en cada uno de los tipos de universidad según el valor de la matrícula en la ciudad, la editorial con un nivel de confianza del 96% y un error de estimación del 3% debe encuestar por lo menos a 413 estudiantes matriculados en universidades con valor de matrícula considerada alta, 634 estudiantes matriculados en universidades con valor de matrícula considerada media y 514 estudiantes matriculados en universidades con valor de matrícula considerada baja.

RESUMEN

Una vez concebida la idea de investigación, y luego de tener claridad sobre el problema que se va a investigar, hay que plantear los objetivos que se espera lograr, contar con una justificación para desarrollar el estudio, tener un fundamento teórico, plantear la hipótesis o las preguntas de investigación, definir el tipo y el diseño de la investigación, el; otro aspecto para tener en cuenta es definir la *población* o *muestra* con la cual se desarrollará la investigación de interés.

Al respecto, población es la totalidad o el conjunto de todos los sujetos o elementos que tienen ciertas características similares y a los cuales se refiere la investigación. La muestra es la *parte* de la población que se selecciona, de la cual realmente se obtiene la información para el desarrollo del estudio y sobre la cual se efectuarán la medición y la observación de las variables objeto de estudio. Los pasos para definir una muestra son los siguientes:

- Definición de la población.
- Identificación del marco muestral.
- Determinación del tamaño de la muestra.
- Elección de un procedimiento de muestreo.
- Selección de la muestra.

En la población o muestra es necesario conocer ciertas *características*, a las cuales se les denomina como variables y pueden ser de tipo cuantitativo o cualitativo. Estas variables se analizan a partir de sus necesidades, ya sea en términos de datos de promedios o totales para las variables cuantitativas, y de proporciones o totales para las variables cualitativas.

En la investigación científica, el *tamaño de la muestra* debe estimarse siguiendo los criterios que ofrece la estadística, y por ello hay que conocer algunas técnicas o métodos de muestreo. El método de muestreo utilizado para estimar el tamaño de una muestra depende del tipo de investigación que desea realizarse y de las hipótesis y del diseño de investigación que se hayan definido para desarrollar el estudio.

En este libro se ha ilustrado la estimación de tamaño de muestra a partir de tres tipos de muestreo: muestreo aleatorio simple (MAS), muestreo proporcional y muestreo estratificado. Para el muestreo aleatorio simple y para el proporcional se ilustró el cálculo del tamaño de la muestra para el caso de poblaciones finitas e infinitas.

EJERCICIOS DE REPASO Y ANÁLISIS

1. ¿Qué es la población en una investigación? ¿Qué es la muestra en una investigación?
2. ¿Cómo se define el tamaño de la muestra representativa de la población objeto de estudio o investigación que se va a realizar?
3. Defina un tema de investigación, establezca el objetivo general y dé un ejemplo de la estimación o del cálculo del tamaño de muestra para desarrollar la investigación.
4. Suponga que va a hacer una investigación para conocer la opinión de los clientes respecto a la calidad del servicio prestado por el restaurante JPL, al cual asisten diariamente, en promedio, 370 personas. ¿De qué tamaño debe ser la muestra de clientes que se deben entrevistar para que la información obtenida sea representativa?
5. Suponga que va a realizar un diagnóstico sobre la competitividad de las empresas del sector de las artes gráficas. El número de empresas del sector es de 3 900 empresas, discriminadas así: 2 180 son pequeñas empresas, 980 son medianas empresas y 740 grandes empresas. ¿De qué tamaño debe ser la muestra de empresas que se va a diagnosticar por cada tamaño de empresas, si se decide tomar una muestra total de 490 empresas?
6. Revise algún material bibliográfico sobre el tema de muestreo aplicado al campo de su profesión, repase los diferentes tipos de muestreo y haga ejercicios para compartir con lo(a)s compañero(a)s de clases.
7. Plantee situaciones del campo de su disciplina o profesión donde necesite estimar el tamaño de muestra, luego calcule el mismo para las diferentes situaciones planteadas.