

# Una revisión de distintas formulaciones de la técnica de muestreo

FRANCISCO SÁNCHEZ ESPINOZA<sup>1</sup>

## Resumen

En esta oportunidad se presentan los resultados de una recopilación y revisión de distintas formas de obtención de una muestra a partir de una población determinada. Como se sabe, si fuera necesario obtener información directamente de las entidades de estudio muchas ocasiones de investigación no serían abordables, lo cual sucede mayormente cuando la población o universo de estudio es grande. Para hacer posible este tipo de investigaciones se utiliza el muestreo, método con el cual es posible acortar el tiempo y los recursos financieros que la investigación requiere. Se exponen seis textos, en cada uno se contiene una formulación distinta, en los cuales se encuentra esta necesidad de tomar muestras para aplicar cuestionarios y obtener información como fuente primaria de investigación.

*Palabras clave:* población, muestra, información.

## Introducción

Se inicia exponiendo el formato de muestreo utilizado por Francisco Javier Barranco Sáiz en su libro *Técnicas de marketing político*, el cual dedica el capítulo 2, denominado “Investigación del mercado político”, a la obtención de información por fuentes primarias. Presenta cuatro ejemplos, uno para cada una de las fórmulas que expone.

El segundo texto es el de Roberto Hernández Sampieri, su título es *Metodología de la investigación*, en el capítulo 8, “¿Cómo seleccionar una muestra?”, expone un par de ejemplos.

---

1. Doctor en Sociología, profesor e investigador de la Facultad de Derecho y Ciencias Sociales de la Universidad Autónoma de Puebla. Candidato al sni. Profesor en la licenciatura y maestría en Ciencias Políticas. Correo electrónico: frasaes\_7@hotmail.com.

Otro formato de muestreo lo expone María Isabel Pérez Enríquez en su libro *El impacto de las migraciones y expulsiones indígenas de Chiapas*, en el anexo aborda el muestreo, el cual es aplicado en el estudio como forma de obtener información.

Felipe Pardinás en su libro *Metodología y técnicas de investigación en ciencias sociales*, expone en el capítulo 7 otro formato de muestreo, ofreciendo dos ejemplos de aplicación.

En un quinto texto revisado, David Aaker y Georges Day en su libro *Investigación de mercados* dedican el capítulo 11, “Tamaño de la muestra y teoría estadística”, al muestreo, exponiendo el tipo probabilístico y dando una fórmula para obtener el tamaño de la muestra, además dan un ejemplo que va variando para exponer distintos niveles de error, confianza, tamaño poblacional y desviación estándar.

Por último, Iván Ximilitl Islas en su tesis de grado “Organización campesina y comercialización de café: estudio en dos municipios de la sierra norte de Puebla”, el autor expone una fórmula distinta para tomar el tamaño de la muestra que utiliza en su investigación.

En la séptima parte, después de exponer las formulaciones se les enfrenta con fines de evaluación, para lo cual se destacan sus características y se establece la preferencia de algunos formatos sobre otros, mostrándose la primacía de uno de ellos.

Finalmente se muestran en una bibliografía los textos que en esta ocasión se han revisado.

## 1. Barranco Sáiz

Francisco Barranco<sup>2</sup> expone una forma de muestreo basada en una serie de elementos y en la aplicación de cuatro fórmulas.

Los elementos necesarios para el diseño de una muestra son los siguientes:

1. Tipos de población:
  - a) Infinita, es mayor a 100,000 entidades.
  - b) Finita, es menor o igual a 100,000 entidades.
2. Coeficiente de fiabilidad:
  - 1) 95.5 %
  - 2) 99.7 %
3. Error de muestreo (E):

Puede escogerse dentro de un margen de 0 a 100%, por supuesto, se prefieren errores pequeños. El 0 % de error implicaría encuestar al 100 % de la población.
4. Porcentajes:

Los porcentajes o proporcionalidades son dos valores, p y q que sumados dan como resultado 100%. Se pueden obtener a partir de una encuesta piloto que puede ser tomada en 100 a 200 entidades. Cuando no es posible, sobre todo por

---

2. Barranco, Francisco J. (1994), *Técnicas de marketing político*, México, Red Editorial Iberoamericana.

cuestiones financieras, hacer esa encuesta piloto, entonces se asigna un valor de 50 a cada una de las proporciones.<sup>3</sup>

Con los dos primeros elementos, tipos de población y coeficiente de fiabilidad, se escoge una de cuatro fórmulas; los otros dos elementos se encuentran en la constitución de la fórmula. Entonces, las posibles fórmulas son las siguientes:

- 1) Población infinita con coeficiente de fiabilidad de 99.7%

$$n = \frac{9pq}{E^2}$$

- 2) Población infinita con coeficiente de fiabilidad de 95.5%

$$n = \frac{4pq}{E^2}$$

- 3) Población finita con coeficiente de fiabilidad de 99.7%

$$n = \frac{9pqN}{E^2(N-1) + 9pq}$$

N es el tamaño de la población o universo.

- 4) Población finita con coeficiente de fiabilidad de 95.5%

$$n = \frac{4pqN}{E^2(N-1) + 4pq}$$

A continuación se presenta un ejemplo de aplicación de cada una de las fórmulas.<sup>4</sup>

Para el caso de una población infinita con un coeficiente de fiabilidad de 99.7%, considérese que un partido político desea hacer un sondeo a nivel nacional,<sup>5</sup> con el objeto de conocer la imagen que tiene entre los futuros electores. Para ello encuesta, inicialmente, a 150 votantes potenciales, obteniendo que 60 tienen intención de votar-le. ¿A cuántos tendrá que encuestar, si fija un coeficiente de fiabilidad de 99.7% y un error de muestreo de  $\pm 1.5\%$ ?

Los valores de p y q se obtendrán del pretest realizado:

$$p = \frac{60}{150} * 100 = 40\%$$

$$q = 100 - 40 = 60\%$$

3. Cuando ya se ha encuestado con anterioridad a la misma población, entonces es posible tomar los valores de los porcentajes de esa encuesta.

4. Estos cuatro ejemplos se han adecuado para su exposición a partir de Barranco, *Técnicas*, pp. 58-64.

5. Se supone que al tratarse de un sondeo nacional, la población a encuestar excede a las 100,000 entidades de estudio.

Entonces los elementos son:

Población infinita

|                            |      |   |
|----------------------------|------|---|
| Coefficiente de fiabilidad | 99.7 | % |
| Error                      | 1.5  | % |
| p                          | 40   | % |
| q                          | 60   | % |

Sustituyendo en la expresión se tiene lo siguiente:

$$n = \frac{9pq}{E^2}$$

$$n = \frac{9 * 40 * 60}{1.5^2} = \frac{21600}{2.25} = 9600$$

Se deberá entrevistar a 9,600 individuos de la población en estudio.

Para ejemplificar el caso de una población infinita con coeficiente de fiabilidad de 95.5% se considera que se desea realizar una encuesta en una región,<sup>6</sup> con objeto de analizar las aspiraciones del electorado, de cara a una futura confrontación electoral. ¿A cuántos posibles votantes se deberá entrevistar, si el coeficiente de fiabilidad a utilizar es de 95.5%, el error de muestreo del +- 4 %, y se considera que las condiciones de muestreo son desfavorables?

Entonces los elementos son:

Población infinita

|                            |      |   |
|----------------------------|------|---|
| Coefficiente de fiabilidad | 95.5 | % |
| Error                      | 4    | % |
| p                          | 50   | % |
| q                          | 50   | % |

Sustituyendo en la expresión se tiene lo siguiente:

$$n = \frac{4pq}{E^2}$$

$$n = \frac{4 * 50 * 50}{4^2} = \frac{10000}{16} = 625$$

Ahora se ejemplifica el caso de una población finita con un coeficiente de fiabilidad de 99.7%, se trata del caso en que un partido político desea hacer una encuesta de opinión entre los integrantes de un colegio profesional, formado por 87,520 colegiados, ¿a cuántos deberá entrevistar si desea trabajar con un coeficiente de fiabilidad de 99.7%, un error de muestreo del +- 4.5 % y condiciones desfavorables?

6. Como en el caso anterior, se supone que las entidades en estudio son más de 100,000.

Entonces los elementos son:

|                            |         |   |
|----------------------------|---------|---|
| Población finita           |         |   |
| Coefficiente de fiabilidad | 99.7    | % |
| Error                      | 4.5     | % |
| p                          | 50      | % |
| q                          | 50      | % |
| N                          | 87, 520 |   |

Sustituyendo en la fórmula:

$$n = \frac{9pqN}{E^2(N-1) + 9pq}$$

$$n = \frac{9 * 50 * 50 * 87520}{4.5^2 * (87520 - 1) + 9 * 50 * 50} = \frac{1969200000}{1794759.75} = 1097.19$$

Se entrevistará a 1,907 integrantes del colegio.<sup>7</sup>

Para el caso de la población finita con un coeficiente de fiabilidad de 95.5%, considérense los datos del ejemplo anterior, pero sabiendo que se ha llevado a cabo un pretest a 180 colegiados, de los que 105 denotan afinidad ideológica hacia nuestro partido. El coeficiente de fiabilidad a aplicar será de 95.5 % y el error de muestreo del +- 4 %.

Primero se calculan las proporciones:

$$p = \frac{105}{180} * 100 = 58.33 \%$$

$$q = 100 - 58.33 = 41.67 \%$$

Por lo tanto los elementos son los siguientes:

|                            |         |   |
|----------------------------|---------|---|
| Población finita           |         |   |
| Coefficiente de fiabilidad | 95.5    | % |
| Error                      | 4       | % |
| p                          | 58.33   | % |
| q                          | 41.67   | % |
| N                          | 87, 520 |   |

Aplicando la fórmula se tiene lo siguiente:

$$n = \frac{4pqN}{E^2(N-1) + 4pq}$$

$$n = \frac{4 * 58.33 * 41.67 * 87520}{4^2 * (87520 - 1) + 4 * 58.33 * 41.67} = \frac{850908333.9}{1410026.444} = 603.47$$

---

7. El resultado correcto es 1,097, en el citado texto (Barranco, *Técnicas*), se da como resultado 1,799.

### Tablas de Harvard

Este método, constituido por cuatro fórmulas predeterminadas para obtener el tamaño de  $n$ , tiene gran sencillez. Solamente debe tenerse cuidado en el caso de que el valor escogido para el coeficiente de error sea de 0 y la población sea infinita, ya que en ese caso, cualquiera que sea el índice de confianza, el resultado de la fórmula se indefiniría; no sucede lo anterior cuando la población es finita, ya que en ese caso  $n$  es igual a  $N$ , resultado que es correcto, ya que si se desea un error de 0 se tendría que encuestar a toda la población. El caso de indefinición en la fórmula no se puede resolver; sin embargo, se sabe que un error de 0 requiere que se iguale  $n$  a  $N$ , es decir, el tamaño de la muestra es el total de la población, por lo que quedaría salvado el problema, no estadísticamente pero sí en términos prácticos.

La sencillez de las fórmulas permite la elaboración de las llamadas tablas de Harvard. Se trata de cuatro tablas, una para cada caso de fórmula: población infinita con coeficiente de confianza de 95.5% y otra para 99.7%; una más para población finita con coeficiente de confianza de 95.5 % y otra para 99.7 %.<sup>8</sup>

Tabla 1  
Población infinita con coeficiente de confianza de 95.5%

| Límites de error en porcentajes | Valores posibles de $p$ y $q$ ( $p + q = 100$ ) |        |         |         |         |         |         |         |         |         |         |         |         |           |
|---------------------------------|---|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-----------|
|                                 | 1/99  | 2/98   | 3/97    | 4/96    | 5/95    | 10/90   | 15/85   | 20/80   | 25/75   | 30/70   | 35/65   | 40/60   | 45/55   | 50/50     |
| 0,1                             | 39.600  | 78.400 | 116.400 | 153.600 | 190.000 | 360.000 | 510.000 | 640.000 | 750.000 | 840.000 | 910.000 | 960.000 | 990.000 | 1.000.000 |
| 0,2                             | 9.900   | 19.600 | 29.100  | 38.400  | 47.500  | 90.000  | 127.500 | 160.000 | 187.500 | 210.000 | 227.500 | 240.000 | 247.500 | 250.000   |
| 0,3                             | 4.400   | 8.711  | 12.933  | 17.067  | 21.111  | 40.000  | 56.667  | 71.111  | 83.333  | 93.333  | 101.111 | 106.667 | 110.000 | 111.111   |
| 0,4                             | 2.475   | 4.900  | 7.275   | 9.600   | 11.875  | 22.500  | 31.875  | 40.000  | 46.875  | 52.500  | 56.875  | 60.000  | 61.875  | 62.500    |
| 0,5                             | 1.584   | 3.136  | 4.656   | 6.144   | 6.600   | 13.400  | 20.400  | 25.600  | 30.000  | 33.600  | 36.400  | 38.400  | 39.600  | 40.000    |
| 0,6                             | 1.100   | 2.178  | 3.233   | 4.267   | 5.278   | 10.000  | 14.167  | 17.778  | 20.833  | 23.333  | 25.278  | 26.667  | 27.500  | 27.778    |
| 0,7                             | 808   | 1.600  | 2.376   | 3.135   | 3.878   | 7.347   | 10.408  | 13.061  | 15.306  | 17.143  | 18.577  | 19.592  | 20.204  | 20.408    |
| 0,8                             | 619   | 1.225  | 1.819   | 2.400   | 2.969   | 5.625   | 7.969   | 10.000  | 11.719  | 13.125  | 14.219  | 15.000  | 15.469  | 15.625    |
| 0,9                             | 489   | 968    | 1.437   | 1.896   | 2.346   | 4.444   | 6.296   | 7.901   | 9.259   | 10.370  | 11.235  | 11.852  | 12.222  | 12.346    |
| 1,0                             | 396   | 784    | 1.164   | 1.536   | 1.900   | 3.600   | 5.100   | 6.400   | 7.500   | 8.400   | 9.100   | 8.600   | 9.900   | 10.000    |
| 1,5                             | 176   | 348    | 517     | 683     | 844     | 1.600   | 2.267   | 2.844   | 3.333   | 3.733   | 4.044   | 4.267   | 4.400   | 4.444     |
| 2,0                             | 99  | 196    | 291     | 384     | 475     | 900     | 1.275   | 1.600   | 1.875   | 2.100   | 2.275   | 2.400   | 2.475   | 2.500     |
| 2,5                             | 63  | 125    | 186     | 246     | 304     | 576     | 816     | 1.024   | 1.200   | 1.344   | 1.456   | 1.536   | 1.584   | 1.600     |
| 3,0                             | 44  | 87     | 129     | 171     | 211     | 400     | 517     | 711     | 833     | 933     | 1.011   | 1.067   | 1.100   | 1.111     |
| 3,5                             | 32  | 64     | 95      | 125     | 155     | 294     | 416     | 522     | 612     | 686     | 743     | 784     | 808     | 816       |
| 4,0                             | 25  | 49     | 73      | 96      | 119     | 225     | 310     | 400     | 469     | 525     | 569     | 600     | 619     | 625       |
| 4,5                             | 20  | 39     | 57      | 76      | 94      | 178     | 252     | 316     | 370     | 415     | 449     | 474     | 489     | 494       |
| 5,0                             | 16  | 31     | 47      | 61      | 76      | 144     | 204     | 256     | 300     | 336     | 364     | 384     | 396     | 400       |
| 6,0                             | 11  | 22     | 32      | 43      | 53      | 100     | 142     | 178     | 208     | 233     | 253     | 267     | 275     | 278       |
| 7,0                             | 8   | 16     | 24      | 31      | 39      | 73      | 104     | 131     | 153     | 171     | 186     | 196     | 202     | 204       |
| 8,0                             | 6   | 12     | 18      | 24      | 30      | 56      | 80      | 100     | 117     | 131     | 142     | 150     | 155     | 156       |
| 9,0                             | 5   | 10     | 14      | 19      | 23      | 44      | 63      | 79      | 93      | 104     | 112     | 119     | 122     | 123       |
| 10,0                            | 4   | 8      | 12      | 15      | 19      | 36      | 51      | 64      | 75      | 83      | 91      | 96      | 99      | 100       |
| 15,0                            | 2   | 3      | 5       | 7       | 8       | 16      | 23      | 28      | 33      | 37      | 40      | 43      | 44      | 45        |
| 20,0                            | 1   | 2      | 3       | 4       | 5       | 9       | 13      | 16      | 19      | 21      | 23      | 24      | 25      | 25        |
| 25,0                            | 0,6   | 1      | 2       | 2       | 3       | 6       | 8       | 12      | 12      | 13      | 15      | 15      | 16      | 16        |

8. Las tablas pueden ser consultadas en Barranco, *Técnicas*, pp. 62-65.

El manejo de la tabla incluye en las columnas distintos valores posibles para los porcentajes, p y q. Por otro lado, incluye en las filas distintos valores para el límite de error. El tamaño de la muestra está dado en el cruce de la fila con la columna.

*Tabla 2*  
Población infinita con coeficiente de confianza de 99.7%

| Límites de error en porcentaje | Valores posibles de p y q (p + q = 100) |         |         |         |         |         |           |           |           |           |           |           |           |           |
|--------------------------------|---|---------|---------|---------|---------|---------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
|                                | 1/99                                    | 2/98    | 3/97    | 4/96    | 5/95    | 10/90   | 15/85     | 20/80     | 25/75     | 30/70     | 35/65     | 40/60     | 45/55     | 50/50     |
| 0,1                            | 89.100                                  | 176.400 | 261.900 | 345.600 | 427.850 | 810.000 | 1.147.500 | 1.440.000 | 1.687.500 | 1.890.000 | 2.047.500 | 2.160.000 | 2.227.500 | 2.250.000 |
| 0,2                            | 22.275                                  | 44.100  | 65.475  | 86.400  | 106.875 | 202.500 | 286.875   | 360.000   | 421.875   | 472.500   | 511.875   | 540.000   | 556.875   | 562.500   |
| 0,3                            | 9.900                                   | 19.600  | 29.100  | 38.400  | 47.500  | 90.000  | 127.400   | 160.000   | 187.500   | 210.000   | 227.500   | 240.500   | 247.500   | 250.000   |
| 0,4                            | 5.569                                   | 11.025  | 16.369  | 21.600  | 26.719  | 50.625  | 71.719    | 90.000    | 105.469   | 118.125   | 127.969   | 135.000   | 139.219   | 140.625   |
| 0,5                            | 3.564                                   | 7.056   | 10.476  | 13.824  | 17.100  | 32.400  | 45.000    | 57.600    | 67.500    | 75.600    | 81.900    | 86.400    | 89.100    | 90.000    |
| 0,6                            | 2.475                                   | 4.900   | 7.275   | 9.600   | 11.875  | 22.500  | 31.875    | 40.000    | 46.875    | 52.500    | 56.875    | 60.000    | 61.875    | 62.500    |
| 0,7                            | 1.818                                   | 3.600   | 5.345   | 7.053   | 8.724   | 16.531  | 23.418    | 29.388    | 34.439    | 38.571    | 41.786    | 44.082    | 45.459    | 45.918    |
| 0,8                            | 1.392                                   | 2.756   | 4.092   | 5.400   | 6.680   | 12.656  | 17.930    | 22.500    | 27.367    | 29.531    | 31.992    | 33.750    | 34.805    | 35.156    |
| 0,9                            | 1.100                                   | 2.178   | 3.233   | 4.267   | 5.278   | 10.000  | 14.167    | 17.778    | 20.833    | 25.278    | 26.667    | 26.667    | 27.500    | 27.778    |
| 1,0                            | 891                                     | 1.764   | 2.619   | 3.456   | 4.275   | 8.100   | 11.475    | 14.400    | 16.875    | 18.900    | 20.475    | 21.600    | 22.275    | 22.500    |
| 1,5                            | 396                                     | 784     | 1.164   | 1.356   | 1.900   | 3.600   | 5.100     | 6.400     | 7.500     | 8.400     | 9.100     | 9.600     | 9.900     | 10.000    |
| 2,0                            | 223                                     | 441     | 655     | 864     | 1.069   | 2.025   | 2.869     | 3.600     | 4.219     | 4.725     | 5.119     | 5.400     | 5.569     | 5.627     |
| 2,5                            | 143                                     | 282     | 419     | 553     | 684     | 1.296   | 1.836     | 2.304     | 2.700     | 3.024     | 3.276     | 3.456     | 3.564     | 3.600     |
| 3,0                            | 99                                      | 196     | 291     | 384     | 475     | 900     | 1.275     | 1.600     | 1.875     | 2.100     | 2.275     | 2.400     | 2.475     | 2.500     |
| 3,5                            | 73                                      | 144     | 241     | 282     | 349     | 661     | 937       | 1.176     | 1.378     | 1.543     | 1.671     | 1.763     | 1.818     | 1.837     |
| 4,0                            | 56                                      | 110     | 164     | 216     | 267     | 506     | 717       | 900       | 1.055     | 1.181     | 1.280     | 1.350     | 1.392     | 1.406     |
| 4,5                            | 44                                      | 87      | 129     | 171     | 211     | 400     | 567       | 711       | 833       | 933       | 1.011     | 1.067     | 1.100     | 1.111     |
| 5,0                            | 36                                      | 71      | 105     | 138     | 171     | 324     | 459       | 576       | 675       | 756       | 819       | 864       | 891       | 900       |
| 6,0                            | 25                                      | 49      | 73      | 96      | 119     | 225     | 319       | 400       | 469       | 525       | 569       | 600       | 619       | 625       |
| 7,0                            | 18                                      | 36      | 53      | 71      | 87      | 165     | 234       | 294       | 344       | 386       | 418       | 441       | 455       | 459       |
| 8,0                            | 14                                      | 28      | 41      | 54      | 67      | 127     | 179       | 225       | 264       | 295       | 320       | 338       | 348       | 352       |
| 9,0                            | 11                                      | 22      | 32      | 43      | 53      | 100     | 142       | 178       | 208       | 233       | 263       | 267       | 275       | 278       |
| 10,0                           | 9                                       | 18      | 26      | 35      | 43      | 81      | 115       | 144       | 169       | 189       | 205       | 216       | 223       | 225       |
| 15,0                           | 4                                       | 8       | 12      | 15      | 19      | 36      | 51        | 64        | 75        | 84        | 91        | 96        | 99        | 100       |
| 20,0                           | 2                                       | 4       | 7       | 9       | 11      | 20      | 29        | 36        | 42        | 47        | 51        | 54        | 56        | 56        |
| 25,0                           | 1                                       | 3       | 4       | 6       | 7       | 3       | 18        | 23        | 27        | 30        | 33        | 35        | 36        | 36        |
| 30,0                           | 1                                       | 2       | 3       | 4       | 5       | 9       | 13        | 16        | 19        | 21        | 23        | 24        | 25        | 25        |
| 35,0                           | 0,7                                     | 1       | 2       | 3       | 3       | 7       | 9         | 12        | 14        | 15        | 17        | 18        | 18        | 18        |
| 40,0                           | 0,6                                     | 1       | 2       | 2       | 3       | 5       | 7         | 9         | 11        | 12        | 13        | 14        | 14        | 14        |

El manejo de la tabla incluye en las columnas distintos valores posibles para los porcentajes, p y q. Por otro lado, incluye en las filas distintos valores para el límite de error.

*Tabla 3*  
Población finita con coeficiente de confianza de 95.5%

| Amplitud del universo | $p = q = 50$ |           |           |           |           |            |
|-----------------------|--------------|-----------|-----------|-----------|-----------|------------|
|                       | $\pm 1\%$    | $\pm 2\%$ | $\pm 3\%$ | $\pm 4\%$ | $\pm 5\%$ | $\pm 10\%$ |
| —                     | —            | —         | —         | —         | 222       | 83         |
| 1.000                 | —            | —         | —         | 385       | 286       | 91         |
| 1.500                 | —            | —         | 638       | 441       | 316       | 94         |
| 2.000                 | —            | —         | 714       | 476       | 333       | 95         |
| 2.500                 | —            | 1.250     | 769       | 500       | 345       | 96         |
| 3.000                 | —            | 1.364     | 811       | 517       | 353       | 97         |
| 3.500                 | —            | 1.458     | 843       | 530       | 359       | 97         |
| 4.000                 | —            | 1.538     | 870       | 541       | 364       | 98         |
| 4.500                 | —            | 1.607     | 891       | 549       | 367       | 98         |
| 5.000                 | —            | 1.667     | 909       | 556       | 370       | 98         |
| 6.000                 | —            | 1.765     | 938       | 566       | 375       | 98         |
| 7.000                 | —            | 1.842     | 949       | 574       | 378       | 99         |
| 8.000                 | —            | 1.905     | 976       | 580       | 381       | 99         |
| 9.000                 | —            | 1.957     | 989       | 584       | 383       | 99         |
| 10.000                | 5.000        | 2.000     | 1.000     | 588       | 383       | 99         |
| 15.000                | 6.000        | 2.143     | 1.034     | 600       | 390       | 99         |
| 20.000                | 6.667        | 2.222     | 1.053     | 606       | 392       | 100        |
| 25.000                | 7.143        | 2.273     | 1.064     | 610       | 394       | 100        |
| 50.000                | 8.333        | 2.381     | 1.087     | 617       | 397       | 100        |
| 100.000               | 9.091        | 2.439     | 1.099     | 621       | 398       | 100        |
| $\infty$              | 10.000       | 2.500     | 1.111     | 625       | 400       | 100        |

El manejo de la tabla incluye en las columnas distintos valores posibles para el coeficiente de error. Por otro lado, incluye en las filas distintos valores para la amplitud del universo. Los valores de  $p$  y  $q$  en todos los casos son de 50. En esta tabla y en la siguiente cuando no aparece dato en el cruce entre fila y columna es porque la amplitud de la muestra es superior a la mitad de la población y no parece viable el muestreo.



*Tabla 4*  
Población finita con coeficiente de confianza de 99.7%

| Amplitud del universo | $p = q = 50$ |           |           |           |           |
|-----------------------|--------------|-----------|-----------|-----------|-----------|
|                       | $\pm 1\%$    | $\pm 2\%$ | $\pm 3\%$ | $\pm 4\%$ | $\pm 5\%$ |
| 500                   | —            | —         | —         | —         | —         |
| 1.000                 | —            | —         | —         | —         | 474       |
| 1.500                 | —            | —         | —         | 726       | 563       |
| 2.000                 | —            | —         | —         | 826       | 621       |
| 2.500                 | —            | —         | —         | 900       | 662       |
| 3.000                 | —            | —         | 1.364     | 958       | 692       |
| 3.500                 | —            | —         | 1.458     | 1.003     | 716       |
| 4.000                 | —            | —         | 1.539     | 1.041     | 735       |
| 4.500                 | —            | —         | 1.607     | 1.071     | 750       |
| 5.000                 | —            | —         | 1.667     | 1.098     | 763       |
| 6.000                 | —            | 2.903     | 1.765     | 1.139     | 783       |
| 7.000                 | —            | 3.119     | 1.842     | 1.171     | 798       |
| 8.000                 | —            | 3.303     | 1.905     | 1.196     | 809       |
| 9.000                 | —            | 3.462     | 1.957     | 1.216     | 818       |
| 10.000                | —            | 3.600     | 2.000     | 1.233     | 826       |
| 15.000                | —            | 4.091     | 2.143     | 1.286     | 849       |
| 20.000                | —            | 4.390     | 2.222     | 1.314     | 861       |
| 25.000                | 11.842       | 4.592     | 2.273     | 1.331     | 869       |
| 50.000                | 15.517       | 5.056     | 2.381     | 1.368     | 884       |
| 100.000               | 18.367       | 5.325     | 2.439     | 1.387     | 892       |
| $\infty$              | 22.500       | 5.625     | 2.500     | 1.406     | 900       |

El manejo de la tabla incluye en las columnas distintos valores posibles para el coeficiente de error. El valor considerado para p y q es de 50. Por otro lado, incluye en las filas distintos valores para la amplitud del universo.

## 2. Hernández Sampieri

Otra forma de tomar el tamaño de una muestra la expone Roberto Hernández.<sup>9</sup> Dada una población de tamaño N, debe preguntarse por el mínimo número de unidades muestrales que deberán ser tomadas en cuenta y que serán n, para que se tenga un error estándar menor a .01.

Al tomar el tamaño de una muestra se desea que el valor de n en una variable, representado por  $\bar{y}$ , se acerque al valor de la misma variable en N, representado por  $\bar{Y}$ .

El tamaño de la muestra se obtiene en dos pasos. El primero es aplicar la siguiente fórmula:

$$n = \frac{S^2}{V^2}$$

9. Hernández Sampieri, Roberto. "Metodología de la investigación." Mc Graw-Hill, México, 1988.

O lo que es lo mismo:

$$\text{Tamaño provisional de } n = \frac{\text{Varianza de la muestra}}{\text{Varianza de la población}}$$

Dado que se trata de una medida provisional, debe ajustarse, si se conoce el tamaño de la población, lo cual se hace con la fórmula:

$$n = \frac{n'}{1 + \frac{n'}{N}}$$

A continuación se presentan ejemplos para ilustrar la aplicación de las fórmulas.

Se parte de considerar a todos los directores generales de empresas industriales y comerciales que en 1983 tenían un capital social superior a 30 millones de pesos, además de que sus ventas hayan sido superiores a 100 millones de pesos y/o con más de 300 personas empleadas. Eso dio como resultado a 1,176 directores generales, es decir, en el mismo número de empresas se cumplían las características enunciadas. La pregunta es a cuántos directores se deberá entrevistar para tener un error estándar menor a 1.5%.<sup>10</sup>

Es decir:

N Población 1176

Se Error estándar .015

V<sup>2</sup> Varianza de la población. Su definición (Se) cuadrado del error estándar.

S<sup>2</sup> Varianza de la muestra. Expresada como probabilidad de ocurrencia, es decir, S<sup>2</sup> = p (1-p)

Sustituyendo en la fórmula:

$$n' = \frac{S^2}{V^2}$$

Y como,

$$S^2 = p (1 - p) = .9 (1 - .9) = .09$$

$$V^2 = (.015)^2 = .000225$$

10. El ejemplo se ha adecuado a partir de Hernández, *Metodología*, p. 211. Adopto señalar el error estándar como 1.5 % y no, como se hace en el texto citado, 0.015 %. También debe considerarse que en el texto se señala "V" cuando se refiere al error estándar elevado al cuadrado, por lo que debiera señalarse como "V<sup>2</sup>".

Entonces:

$$n = \frac{S^2}{V^2} = \frac{.09}{.000225} = 400$$

Si ahora se sustituye en la fórmula de ajuste:

$$n = \frac{n}{1 + \frac{n}{N}}$$

$$n = \frac{400}{1 + \frac{400}{1176}} = 298$$

Entonces se deberá entrevistar a 298 directores generales.

Para un ejemplo adicional<sup>11</sup> se plantea que una radiodifusora local desea hacer un estudio para saber cómo usan la radio los adultos de la ciudad: a qué hora la escuchan, qué prefieren, qué opinan del noticiario, etc. La ciudad tiene 2'500,000 adultos, la población a considerar son todos los sujetos, hombres o mujeres, que tengan más de 21 años, y que vivan en una casa o departamento propio o rentado de la ciudad. Se desconoce el número de sujetos con dichas características, por lo que se toma en cuenta un mapa de la ciudad, denotando que tiene 5,000 cuadras. La pregunta es en cuántas cuadras deberá levantarse la encuesta si se desea un error estándar no mayor a 1.5 % y con una probabilidad de ocurrencia de 50%.

Entonces se toma en cuenta la fórmula:

$$n = \frac{S^2}{V^2} = \frac{p(1-p)}{.015} = \frac{.5(1-.5)}{.000225} = \frac{.25}{.000225} = 1111.11$$

Ajustando la muestra:

$$n = \frac{n}{1 + \frac{n}{N}}$$

$$n = \frac{1111.11}{1 + \frac{1111.11}{5000}} = 909.0902$$

---

11. El ejemplo se ha adecuado a partir de Hernández, *Metodología*, pp. 214-215. Debe entenderse en el texto referenciado que se trabajará con un error estándar no mayor a 0.015 y no, como dice, a 0.15, ya que ésa es la notación que utiliza en el otro ejemplo que presenta. Asimismo, al efectuar operaciones con la primera fórmula se obtiene un error estándar de 0.000225 y no, como dice el texto, 0.00025.

### 3. Pérez Enríquez

María Isabel Pérez Enríquez en su libro *El impacto de las migraciones y expulsiones indígenas de Chiapas*,<sup>12</sup> elabora una encuesta que abarca, en la aplicación de cuestionarios, los municipios de san Pedro Chenalhó y san Andrés Larráinzar.

Para el “cálculo del muestreo” se sigue la siguiente fórmula:

$$n/cn = \frac{N t_2 s_2}{N d_2 + t_2 s_2}$$

Para un ejemplo, considérense los siguientes datos, con el fin de obtener el número de cuestionarios que deben aplicarse.

| POBLACIÓN DEL MUNICIPIO |        |
|-------------------------|--------|
| San Andrés              | 11,050 |

Fuente:Elaboración del autor a partir de Pérez Enríquez, María Isabel. "El impacto de las migraciones y expulsiones indígenas de Chiapas." UNACH, Tuxtla Gutiérrez, 1998. P. 13.

Además, debe tenerse en cuenta que la autora considera una probabilidad de 95% del nivel de confianza en relación a la “t” de Student, considerando a  $t = 1.96$  y el nivel de precisión de un 20% del promedio premuestral para la población, donde  $S_2$  es la varianza de la variable. Con la fórmula y sustituyendo valores en ella:

$$n/cn = \frac{N t_2 s_2}{N d_2 + t_2 s_2}$$

$$n/cn = \frac{11050 * 1.96 * (50 * 50)}{11050 * 20 + 1.96 * (50 * 50)}$$

$$n/cn = \frac{54145000}{225900} = 239.68 \approx 240$$

Es decir, deben aplicarse 240 cuestionarios en san Andrés.

El texto indica que las variables a considerar para el muestreo son demográficas, económicas, de salud, políticas e ideológicas (educación familiar, escolaridad, religión). Asimismo, considera homogéneas a las comunidades indígenas al interior de los parajes, y dice que se aplicó el muestreo simple-aleatorio cualitativo y cuantitativo, de acuerdo a la asesoría proporcionada por el maestro en Estadística, Guillermo Pe-láez Gramajo.<sup>13</sup>

12. Pérez Enríquez, María Isabel (1998), *El impacto de las migraciones y expulsiones indígenas de Chiapas*, Tuxtla Gutiérrez, UNACH.

13. Pérez, *Impacto*, p. 183.

En cuanto a la variable cuantitativa considerada apunta: “Fue necesario ver una variable cuantitativa, que de acuerdo con la relevancia del análisis, estuvo constituida por la tenencia de la tierra, y en particular el número de hectáreas por familia, en contraste con variables cualitativas de orden político, organizativo y religioso”.<sup>14</sup>

Por lo tanto, la fórmula es:

$$n/cn = \frac{N t_2 s_2}{N d_2 + t_2 s_2}$$

En donde:

|                |                                       |
|----------------|---------------------------------------|
| n/cn           | Muestra cuantitativa                  |
| N              | Tamaño de la población                |
| S <sub>2</sub> | Considerando la varianza máxima (p=q) |
| t <sub>2</sub> | Distribución “t” de Student (1.96)    |
| d <sub>2</sub> | Nivel de precisión o error aceptado.  |

Por último, debe considerarse que la autora señala que bajo la forma de muestreo que presenta se utiliza la varianza máxima (p = q) y el nivel de confianza de 95%.

#### 4. Felipe Pardinás

Para estimar el tamaño de la muestra Pardinás<sup>15</sup> utiliza la siguiente fórmula:

$$n = \frac{N z_1^2 - \frac{\alpha}{2} \sigma^2}{z_1 - \frac{\alpha}{2} \sigma^2 + (N - 1) e^2}$$

N = tamaño de la población.

$z_1 - \frac{\alpha}{2}$  = nivel de confianza de la prueba que diseñamos.

$\sigma^2$  = variancia de la población, conocida o estimada por otros estadísticos.

e = imprecisión que podemos tolerar.

14. Ibídem, p. 183.

15. Pardinás, Felipe (1985), *Metodología y técnicas de investigación en ciencias sociales*, México, Siglo XXI editores, p. 178.

En la fórmula Pardinás incluye el término  $z_1^2 - \frac{\alpha}{2}$ , en tanto que después habla del nivel de confianza como  $z_1 - \frac{\alpha}{2}$ , en realidad se trata del mismo elemento. Asimismo, remite al lector a la tabla 8, pero en el libro no hay tablas.

Para ejemplificar con la fórmula tómanse los datos siguientes:

|            |        |
|------------|--------|
| N          | = 125  |
| z de .95   | = 1.96 |
| e          | = 1.7  |
| $\sigma^2$ | = 4    |

$$n = \frac{N z_1^2 - \frac{\alpha}{2} \sigma^2}{z_1^2 - \frac{\alpha}{2} \sigma^2 + (N-1) e^2} = \frac{125 * 1.96^2 * 4^2}{1.96^2 * 4^2 + 124 * 1.7^2} = \frac{7683.2}{419.83} = 18.3 \approx 18$$

La fracción  $\frac{n}{N}$ , aquí  $\frac{18}{125} (100) = 14.4\%$  es el porcentaje de la población requerido por la muestra.

Supóngase que se eleva N a 1250 (125\*(10))

$$n = \frac{N z_1^2 - \frac{\alpha}{2} \sigma^2}{z_1^2 - \frac{\alpha}{2} \sigma^2 + (N-1) e^2} = \frac{1250 * 1.96^2 * 4^2}{1.96^2 * 4^2 + 1249 * 1.7^2} = \frac{76832}{3671.07} = 21$$

$$\frac{n}{N} (100) = \frac{21}{1250} (100) = 1.68$$

Con esto, Pardinás advierte atinadamente que no debe optarse por establecer un porcentaje de población como tamaño de muestra.

## 5. David Aaker

Para tomar el tamaño de la muestra es necesario conocer las características que tiene una población: la media, la varianza y la desviación estándar.<sup>16</sup>

La población de votantes, de ciudadanos, de habitantes, etc., tiene una media en sus preferencias electorales, sus actitudes políticas y sus valoraciones en general

16. Véase Aaker, David y George S. Day (1989, 2ª ed.), *Investigación de mercados*, México, McGraw-Hill. El texto dedica el capítulo 11 especialmente al muestreo.

acerca de lo político. Se trata del promedio simple, entendido como la suma de las observaciones dividida entre el número de entidades tomadas en cuenta.

El mismo tipo de población de entidades tiene diferencias en sus opiniones y preferencias, esas diferencias pueden cuantificarse con la varianza, la cual es la sumatoria de las diferencias de cada observación respecto a la media o promedio elevadas al cuadrado, y dividida, la sumatoria, entre el número de entidades menos la unidad. La desviación estándar es la raíz cuadrada de la varianza. Las unidades en las que está medida la varianza están elevadas al cuadrado, dicha elevación responde a la necesidad de evitar que la sumatoria de las diferencias de las observaciones respecto a la media sea cero, por lo que son unidades de medición que no se pueden usar en el análisis, las unidades de desviación estándar evitan ese problema retornando las unidades de expresión a las unidades de medición originales. Si, por ejemplo, la unidad de medida original de las observaciones fueran kilogramos, entonces las unidades de varianza estarán en kilogramos al cuadrado, que es una medida abstracta, para retornar las medidas a su forma original, en kilogramos, es necesario obtener las unidades de desviación estándar.

El problema es cuando la media de la población no es conocida y debe ser estimada a partir de una muestra, por lo que la media, la varianza y la desviación estándar muestrales tienen alguna distancia respecto de esas mismas medidas en la población, y ese es un problema que no puede evitarse cuando por razones de tiempo y recursos económicos debe sacrificarse tener las medidas poblacionales sustituyéndolas por las muestrales.

Desde luego, no todas las muestras posibles de una población tienen las mismas magnitudes en su media, varianza y desviación estándar. Intuitivamente, es razonable pensar que la variación en la media será más grande a medida que la varianza de la población también lo sea. En extremo, si no hay variación en la población, tampoco habrá variación en la media. También es razonable pensar que, a medida que aumenta el tamaño de la muestra, la variación en la media disminuirá. Cuando la muestra es pequeña, se necesitan sólo uno o dos valores extremos para afectar sustancialmente a la media muestral, generando de este modo una media relativamente grande o pequeña. A medida que aumenta el tamaño de la muestra, estos valores extremos tendrán un menor impacto cuando aparezcan, porque serán promediados con más valores. La variación en la media es medida por su error estándar, el cual es:

$$ES_x = \frac{\overline{ES} * X}{\sqrt{n}}$$

ES, es el error estándar.

Entonces, el error estándar de la media muestral depende de n, el tamaño de la muestra, por lo que si n se altera el error estándar cambiará de acuerdo con ello, aun cuando el error estándar de x permanezca inalterado.

La variable X tiene una distribución de probabilidad, la media de la muestra también la tiene. Se acostumbra suponer que la variación de la media muestral de muestra

a muestra seguirá la distribución normal, también llamada curva de campana o campana de Gauss. Tal supuesto no es tan extremo como puede parecerlo. La distribución de las medias muestrales será normal si la distribución de la población es normal, o si el tamaño de la muestra aumenta. La última condición constituye el teorema del límite central.

La media de la muestra es usada para estimar la media de la población desconocida. Debido a que la media de la muestra varía de muestra a muestra, no es igual a la media de la población. Existe un error de la muestra. Es útil proporcionar una estimación de intervalo en torno de la media de la muestra, el cual refleje el juicio acerca del alcance de este error muestral:

Media muestral  $\pm$  error de la muestra = estimación de intervalo de la media poblacional

Así:

$$\bar{X} \pm \text{Error muestral} , \quad \text{ó} \quad \bar{X} = \pm \frac{ZDSX}{\sqrt{n}}$$

Donde:

Z = 2 para un nivel de confianza de 95%

Z = 5/3 para un nivel de confianza de 90%

DSX = Desviación estándar de la población (será usada la muestral si la poblacional es desconocida)

n = Es el tamaño de la muestra

De este modo, el tamaño de la estimación del intervalo dependerá de tres factores. El primero es el nivel de confianza. Si estamos dispuestos a tener menos confianza de que la estimación del intervalo incluya a la media de la población verdadera y desconocida, entonces el intervalo será más pequeño. El segundo factor es la desviación estándar de la población. Si hay poca variación de la población, entonces la estimación del intervalo de la media poblacional será más pequeña. El tercero es el tamaño de la muestra. Conforme el tamaño de la muestra aumenta, el error de la muestra se ve reducido y el intervalo se volverá más pequeño.

Mediante esos conceptos es que se puede finalmente especificar cómo se determina el tamaño de una muestra. Para proceder al análisis deben especificarse dos cosas: el tamaño del error de la muestra que se desea y el nivel de confianza.

Esta especificación dependerá de la intercompensación entre el valor de la información más exacta y el costo por un incremento en el tamaño de la muestra. Para un nivel de confianza dado, un error muestral más pequeño tendrá un costo, en términos de un tamaño de la muestra más grande. Similarmente, para un error de la muestra dado, un error muestral más pequeño tendrá un costo, en términos de un tamaño de la muestra más grande.

Usando la fórmula general para la estimación del intervalo:



$$\bar{X} \pm \text{Error muestral}, \text{ ó } \bar{X} \pm \frac{ZDS}{\sqrt{n}}$$

Además, se sabe que:

$$\text{Error de la muestra} = \frac{ZDS}{\sqrt{n}}$$

Dividiendo toda la expresión entre el error de la muestra y multiplicando por la raíz cuadrada de n:

$$\sqrt{n} = \frac{ZDS}{\text{Error Muestral}}$$

Y elevando al cuadrado ambos lados de la ecuación, obtenemos una expresión para el tamaño de la muestra:

$$n = \frac{Z^2 DS^2}{(\text{Error de la Muestra})^2}$$

Por lo tanto, si se ha decidido con qué nivel de confianza trabajar, y por consiguiente Z, y si también se ha decidido el error de la muestra que se permitirá, entonces el tamaño necesario de la muestra queda especificado por la fórmula anterior.

El procedimiento que se acaba de mostrar supone que la desviación estándar de la población o de la muestra es conocida. En la mayor parte de las situaciones prácticas, no es conocida y debe ser estimada usando uno de varios enfoques disponibles. Un método es usar una desviación estándar de la muestra obtenida de una encuesta anterior comparable o de una encuesta piloto. Otro enfoque es tomar una situación del “peor de los casos”. Por ejemplo, considerar que la mitad de la población tiene la opinión exactamente opuesta a la que tendría la otra mitad de la población; obviamente, la muestra se torna más grande.

A manera de ejemplo supóngase que se quiere estar 95% seguros de que el error muestral al estimar la media de la población no exceda de 0.3, además se conoce que la desviación estándar de la población es de 1.49, entonces:

$$n = \frac{Z^2 DS^2}{(\text{Error de la Muestra})^2}$$

$$n = \frac{Z^2 DS^2}{(\text{Error de la Muestra})^2} = \frac{2^2 * 1.49^2}{0.3^2} = \frac{8.8804}{0.09} = 98.67 \approx 99$$

Si se varía el nivel de confianza a 90 % entonces bajará el tamaño de n:<sup>17</sup>

$$n = \frac{Z^2 DS^2}{(\text{Error de la Muestra})^2}$$

$$n = \frac{Z^2 DS^2}{(\text{Error de la Muestra})^2} = \frac{5/3^2 * 1.49^2}{0.3^2} \frac{6.16694}{0.09} = 68.52 \approx 69$$

Si ahora se varía el error permitido a .5 entonces n bajará, aunque se sostenga el nivel de confianza de 95%:

$$n = \frac{Z^2 DS^2}{(\text{Error de la Muestra})^2}$$

$$n = \frac{Z^2 DS^2}{(\text{Error de la Muestra})^2} = \frac{2^2 * 1.49^2}{0.5^2} \frac{8.8804}{0.25} = 35.52 \approx 36$$

En cuanto al tamaño de la población Aaker apunta que el cálculo del tamaño de la muestra es determinado de la misma manera, con la misma fórmula, independientemente de si el tamaño de la población es de 1, 000 o de 100, 000.

## 6. Gómez Aguilar y Ximitl Islas

La fórmula que presenta Iván Ximitl<sup>18</sup> en su trabajo, la cual es tomada de Roberto Gómez,<sup>19</sup> es la siguiente:

$$n = \frac{N \sum_{i=1}^k N_i S_i^2}{N^2 V + \sum_{i=1}^k N_i S_i^2}$$

Donde:

N Tamaño de la población

$S_i^2$  Varianza del distrito i

Para otras literales debe considerarse la fórmula:

$$V = \frac{d^2}{Z^2}$$

17. El resultado que el texto incorrectamente señala es 65.5.

18. Ximitl Islas, Iván (2006), "Organización campesina y comercialización de café: estudio en dos municipios de la sierra norte de Puebla". Tesis de maestría en Estrategias para el Desarrollo Agrícola Regional, Colegio de Postgraduados, Puebla.

19. Gómez Aguilar, Roberto (1979), "Introducción al muestreo". Tesis de maestría en Ciencias en Estadística, Colegio de Postgraduados, Chapingo.

La asignación de tamaño de muestra a cada uno de los estratos se hace con la siguiente fórmula:

$$n_i = \frac{N_i}{N} n$$

Como ejemplo podemos considerar que en la referenciada tesis de maestría en Estrategias para el Desarrollo Regional Agrícola del Colegio de Postgraduados, cuyo título es “Organización campesina y comercialización de café: estudio en dos municipios de la sierra norte de Puebla”, Iván Ximitl toma en cuenta, entre sus fuentes primarias, la aplicación de cuestionarios. En Puebla localiza 54,725 predios dedicados al cultivo de café. Hay cinco Distritos de Desarrollo Rural: Huauchinango, Libres, Tehuacan, Teziutlán y Zacatlán. De ese territorio se consideran los municipios en que predominan grupos étnicos: nahuas en Cuetzalan y totonacos en Huehuetla.

La variable que se considera para el muestreo es la superficie por predio en términos de la varianza, tal como aparece en el cuadro siguiente:<sup>20</sup>

| Municipio              | Promedio | N    | Desviación Estándar | Suma   | Mínimo | Máximo | Varianza |
|------------------------|----------|------|---------------------|--------|--------|--------|----------|
| Cuetzalan del Progreso | 0.898    | 5490 | 1.232               | 4931.3 | 0.2    | 4.0    | 1.518    |
| Huehuetla              | 0.742    | 1754 | 0.584               | 1302.2 | 0.3    | 6.0    | 0.341    |

Se considera una precisión del 15 % y una confiabilidad del 90 %, los datos completos se presentan a continuación:

|                     |           |                                |   |       |
|---------------------|-----------|--------------------------------|---|-------|
| Población           | Región    | N                              | = | 493   |
| Población estrato 1 | Cuetzalan | N <sub>1</sub>                 | = | 305   |
| Población estrato 2 | Huehuetla | N <sub>2</sub>                 | = | 188   |
| Varianza estrato 1  | Cuetzalan | (S <sup>2</sup> ) <sub>1</sub> | = | 1.518 |
| Varianza estrato 2  | Huehuetla | (S <sup>2</sup> ) <sub>2</sub> | = | 0.341 |
| Precisión           |           | d                              | = | 0.15  |
| Confiabilidad       |           | Z                              | = | 1.64  |

Sustituyendo los valores en la fórmula:

$$n = \frac{N \sum_{i=1}^k N_i S_i^2}{N^2 V + \sum_{i=1}^k N_i S_i^2} = \frac{493 * [(305 * 1.518) + (188 * 0.341)]}{493^2 * \frac{0.15^2}{1.64^2} + [(305 * 1.518) + (188 * 0.341)]} = 101$$

20. El cuadro en la tesis da la siguiente referencia: Ramírez Valverde, Benito (2003), “Proyecto: Estudio sobre estrategias para el desarrollo sustentable de la sierra Nor-oriente de Puebla con la participación de productores, organizaciones e instituciones como respuesta a las condiciones de pobreza y marginación”. Primer informe semestral. Fondo mixto de investigación científica y tecnológica Conacyt-Gobierno del estado de Puebla, Puebla, México.

La asignación de tamaño de muestra a cada uno de los estratos:

$$n_1 = \frac{N_1}{N} n = \frac{305}{493} * 101 = 62$$

$$n_2 = \frac{N_2}{N} n = \frac{188}{493} * 101 = 39$$

Entonces:

$$62 + 39 = 101$$

## 7. A manera de conclusiones

La técnica de Hernández Sampieri contiene dos fórmulas, primero debe obtenerse el tamaño de muestra y luego la muestra ajustada por el tamaño de la población.

En cuanto a los conceptos, se utilizan la “varianza de la muestra” y “varianza de la población”. Como se ha dicho,  $S^2$  o sea la “varianza de la muestra”, es tomada en términos de probabilidad de ocurrencia, es decir, la probabilidad de que un evento ocurra es  $p$ , la probabilidad de que no ocurra es  $1 - p$ . Esta cuestión es equivalente a lo que Barranco maneja como los “porcentajes”  $p$  y  $q$ .

Asimismo,  $V^2$  es la “varianza de la población” pero es tomada, para todo efecto práctico, como el cuadrado del error estándar, el cual define el investigador. Se trata del máximo error que se está dispuesto a aceptar. Barranco lo maneja como “error de muestreo”, el cual define como el error que se comete por el hecho de extraer un grupo pequeño de individuos de un grupo mayor.

Pérez Enríquez indica que considera solamente la “varianza” máxima,  $S_2$ , es decir, cuando la probabilidad de ocurrencia de un fenómeno es de 50% y la probabilidad de que no ocurra es de 50%. Asimismo, utiliza sólo una confianza de 95%. Los términos correspondientes en Barranco son los “porcentajes”, para la “varianza” de Pérez; y el “coeficiente de fiabilidad” para la “confianza”.

Por otro lado, Pérez habla del “nivel de precisión”,  $d$ , que es el “error” aceptado en Barranco.

En ambos autores  $N$ , simboliza el tamaño de la población, si bien en Pérez se utiliza en su única fórmula, en tanto que en Barranco se utiliza en dos de sus cuatro fórmulas.

Una diferencia fundamental en ambos autores es que Pérez utiliza la distribución  $t$  para el nivel de confianza, en tanto que Barranco sólo tiene dos opciones para el nivel de confianza, 99.5 y 99.7%.

Para Felipe Pardinás “ $e$ ” es la “imprecisión” que se puede tolerar, en tanto que en Barranco  $E$  es el “error” de muestreo. Asimismo,  $N$  es la “población” en ambos autores.

También Pardinás utiliza el nivel de confianza  $z_1^2 - \frac{\alpha}{2}$  y la varianza  $\sigma^2$ .

Un asunto notorio en Pardinas es que cuando presenta el ejemplo que se ha incluido, en el que después establece variaciones en  $N$  y se vuelve a calcular el tamaño de  $n$ , en el primer momento  $n$  es 14% de  $N$  y, en el segundo momento, es 1.68%, él dice que: “Advertimos que dar a  $n$  prejuiciadamente un porcentaje fijo de la población no es un buen criterio para calcular  $n$ .”<sup>21</sup>

En tanto que Pérez dice:

Aún cuando se muestra una ligera diferencia en cuanto a la población total en ambos municipios, tratamos de encuestar al 10% del total de las familias, en las comunidades seleccionadas para el estudio. Realizamos 721 encuestas para tratar de precisar las condiciones socioeconómicas de los municipios de origen en donde se dieron las expulsiones [...]”<sup>22</sup>

Es decir, no tiene sentido tomar el tamaño de muestra si se ha establecido prejuiciadamente una cifra.

Ahora bien, Pardinas hace una reflexión muy atinada en torno a la existencia de varias fórmulas de muestreo:

Existen varias fórmulas para calcular el tamaño de una muestra porque aunque todas ellas demandan como factores un grado de precisión elegido por el investigador —diferencia entre parámetro y estadístico encontrado—, simbolizado por  $e$ ; el número de veces en que no deseamos equivocarnos dada la distribución de la estimada que buscamos —nivel de confianza—; el error estándar de esa estimada y la prueba que deseamos de uno o dos extremos; las fórmulas del tamaño de la muestra varían, porque varían los parámetros que buscamos.<sup>23</sup>

En Aaker se utiliza el “nivel de confianza”, señala dos niveles de desviación estándar, 95 y 90% con  $Z$  igual a 2 y 5/3 respectivamente, en Barranco también se usa el concepto, pero las fórmulas son para 99.7 y 95.5%. Usa el “error de la muestra” y en Barranco haya su correspondiente en el mismo sentido. Asimismo, usa la “desviación estándar de la población”, pero aclara que si no se conoce la poblacional deberá usarse el de la muestra.

El cuestionamiento de la desviación estándar o de la varianza en cuanto a qué referente debe elaborarse la medición, es abordado de la siguiente manera por Aaker:

Generalmente, un instrumento de encuesta o un experimento no se basará sólo en una pregunta. Algunas veces, cientos de ellas pueden estar involucradas. Normalmente no valdrá la pena pasar por todo este proceso en todas las preguntas. Un enfoque razonable sería tomar unas cuantas preguntas representativas y determinar el tamaño de la muestra a partir de ellas. Deberían incluirse las más cruciales con la varianza esperada más alta.<sup>24</sup>

---

21. Pardinas, *Metodología*, p. 180.

22. Pérez, *Impacto*, p. 13.

23. Pardinas, *Metodología*, p. 178.

24. Aaker, *Investigación*, p. 334.

Es decir, se hace respecto de alguna variable, la cual escoge el investigador.

Por otro lado, Aaker corrobora lo dicho por otros autores cuando critican que no debe manejarse un porcentaje de la población como muestra:

Debería notarse que el cálculo del tamaño de la muestra es completamente independiente del tamaño de la población. Una mala concepción común es que una “buena” muestra debería tener un porcentaje relativamente alto con respecto a la estructura muestral incluida. En realidad, el tamaño de la muestra será determinado de la misma manera, independientemente de si el tamaño de la población es de 1,000 o de 100,000.<sup>25</sup>

Entonces, la lógica de cálculo es la misma cuando se trata de una población pequeña o de una grande.

Ximitl utiliza la “población” en el mismo sentido de Barranco. Además utiliza la “varianza” y el tamaño de cada estrato. Asimismo, usa la “precisión” y “confiabilidad”; la primera refiere al nivel de “error” que se permite y, la segunda al “nivel de confianza” en Barranco.

Esta fórmula considera la existencia de estratos, es decir, la población está dividida. Por ello escribe Ximitl: “Al analizar la información se determinó que el esquema de muestreo adecuado a estos datos era un muestreo estratificado aleatorio, y debido a las diferencias en los tamaños de la población de cada estrato se decidió por la asignación proporcional a cada uno de los municipios”.<sup>26</sup>

Entonces, desde el tamaño general de la muestra se toma en cuenta el tamaño de cada uno de los estratos.

Este autor muestra en su trabajo que es posible obtener la varianza a partir de cifras oficiales.

Por su parte Barranco propone una serie de cuatro fórmulas, son sencillas en su manejo y distinguen entre tipos de población, sin tener que ajustar el resultado. Ellas no exigen el conocimiento del tamaño de los estratos, ni siquiera estratificar a la población. Además se han creado tablas que facilitan aún más el manejo de la herramienta estadística de muestreo, sin siquiera usar una fórmula. En los casos de población infinita no se requiere conocer el tamaño de N.

Si bien el formato sólo permite el manejo de dos niveles de confianza, el error aceptado puede variarse en la fórmula.

---

25. *Ibidem*, p. 332.

26. Ximitl, *Organización*, p. 64.

## Bibliografía

- Aaker, D. y G. Day (1989, 2ª ed.), *Investigación de mercados*, México, McGraw-Hill.
- Barranco, F. (1944), *Técnicas de marketing político*, México, Red Editorial Iberoamericana.
- Gómez, R. (1979), "Introducción al muestreo". Tesis de maestría en Ciencias en Estadística, Colegio de Postgraduados, Chapingo.
- Hernández, R. (1988), *Metodología de la investigación*, México, McGraw-Hill.
- Pardinas, F. (1985), *Metodología y técnicas de investigación en ciencias sociales*, México, Siglo XXI Editores.
- Pérez, M. (1998), *El impacto de las migraciones y expulsiones indígenas de Chiapas*, Tuxtla Gutiérrez, UNACH.
- Ramírez, B. (2003), "Proyecto: Estudio sobre estrategias para el desarrollo sustentable de la sierra Nor-oriental de Puebla con la participación de productores, organizaciones e instituciones como respuesta a las condiciones de pobreza y marginación". Primer informe semestral. Fondo mixto de investigación científica y tecnológica Conacyt-Gobierno del estado de Puebla, Puebla.
- Ximitl, I. (2006), "Organización campesina y comercialización de café: estudio en dos municipios de la sierra norte de Puebla". Tesis de maestría en Estrategias para el Desarrollo Agrícola Regional, Colegio de Postgraduados, Puebla.