Primero, lea otras revisiones de la bibliografia. No hay argumento más convincente que el éxito. Pida a un estudiante que ya haya llevado este curso o a su profesor que le muestre una propuesta que haya sido bien recibida. Examine con detenimiento tanto el formato como el contenido de la revisión de la bibliografia. Explore también algunas de las fuentes que mencionamos en este capitulo, sobre todo las que son reseñas de la bibliografia, artículos de revistas científicas y otros trabajos de revisión.

Segundo, cree un tema unificado, o una línea de pensamiento, a seguir durante toda la revisión. No se supone que su revisión de la bibliografia sea una novela, pero casi todas las buenas revisiones se construyen partiendo de un argumento muy general hacia uno más especifico y preparan el escenario para el objetivo de la investigación. Usted debe atraer al lector y crear algún interés respecto a dónde quiere dirigirse con esta investigación porque nadie ha seguido todavía ese camino.

Tercero, utilice un sistema para organizar su material.La mayor parte de las revisiones de la bibliografia se organiza cronológicamente dentro de cada tema. Por ejemplo, si usted va a estudiar las diferencias de género entre los adultos en cuanto a ansiedad y capacidad verbal, organizará todas las referencias por tema (ansiedad y capacidad verbal) y luego, dentro de cada uno de estos temas, iniciará su reviSión con la referencia de fecha más antigua. De este modo, usted avanzará desde los albores hasta lo más reciente y proporcionará cierta perspectiva histórica.

Cuarto, trabaje a partir de un cuadro sinóptico. Si usted es un escritor experimentado y hábil, puede hacer caso omiso de esta sugerencia, pero si apenas está comenzando es recomendable usar esta herramienta que le ayuda a organizar la idea principal de su propuesta antes de iniciar el proceso de redacción propiamente dicho.

Quinto, construya puentes entre las diferentes a reas que revisó. Por ejemplo, si usted va a realizar un estudio intercultural para comparar la forma en que padres provenientes de diferentes culturas disciplinan a sus hijos, es posible que no encuentre mucha bibliografia sobre ese tema especifico, pero ciertamente hay muchos trabajos publicados sobre la educación de los hijos en culturas individuales, y toneladas de referencias sobre la disciplina. Parte del esfuerzo creativo de escribir una propuesta es poder demostrar cómo se juntan los diferentes temas de una manera interesante y potencialmente fructífera.

Sexto, tal vez la practica no hace al maestro, pero tampoco hace daño. Por alguna razon, la mayoria de la gente cree que una persona nace con o sin el talento para escribir. Cualquier escritor exitoso admitirá que para ser un jugador de baloncesto de primera o un consumado violinista es necesario practicar. ¿Es diferente el caso de un escritor? Si usted tiene dudas acerca de esta pregunta, encuentre un escritor profesional y pregúntele cuántas horas al día o la semana practica su oficio. Con mucha frecuencia, resultará que el escritor practica tanto como el deportista o el músico. De hecho, una escritora amiga mía da el siguiente consejo a la gente que desea escribir pero que no tiene una idea precisa del nivel de compromiso que requiere. Ella dice: "Simplemente siéntese frente a su máquina de escribir o procesador de textos y ábrase una vena". Es así de fácil.

Por tanto, la última (pero en realidad la primera) sugerencia es que practique la escritura. A medida que trabaje y averigüe dónde necesita mejorar (obtenga retroalimentación de otros estudiantes y profesores) observará sin duda un cambio positivo.

Todos los que realizan investigación comienzan en algún punto, y casi siempre una revisión bibliográfica pone las ideas y los objetivos en perspectiva. La bibliografia, y todas las herramientas que están disponibles para trabajar con la bibliografía, es su primer y mejor aliado para crear un tratamiento bien investigado y completo de lo que ha ocurrido en el pasado. Una vez identificadas las variables importantes, podrá dirigir su atención a la forma como pueden medir dichas variables, lo cual es el tema del siguiente capitulo.

CAPITULO CUATRO
Muestreo y generalizabilidad

En Neil J. Salkind “Métodos de Investigación”, Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- Por qué un buen muestreo podría ser el paso más importante de un proyecto de investigación
- La importancia de la generalizabilidad
- La diferencia entre una muestra y una población
- El significado de la palabra aleatorio
- Cómo usar una tabla de números aleatorios para seleccionar una muestra
- La diferencia entre muestreo aleatorio simple y muestreo estratificado
- La diferencia entre las estrategias de muestreo probabilisticas y no probabilisticas
- Qué es el error de muestreo y por qué es importante
- Cómo reducir el error de muestreo
- Cómo estimar el tamaño de la muestra

Imagine esto. Se le ha asignado a usted la tarea de medir la actitud general de los estudiantes de educación media hacia el registro irrestricto de sus casilleros en busca de drogas. Usted ya tiene suficiente experiencia en investigación para saber como elaborar algún tipo

de cuestionario asegurándose de que cubra las áreas de contenido importantes y sea fácil de administrar y calificar. Una vez realizado todo el trabajo preliminar, surge la pregunta más importante: ¿a quiénes le pedirá usted que contesten el cuestionario? A todos los estudiantes de todas las escuelas de educación media podrían ser millares? Imposible; seria demasiado costoso. administrar el cuestionario en aquellas escuelas en las que se ha informado de problemas con droga hacerse eso; es muy probable que también haya drogas en las escuelas que no se han identificado como escuelas problema. ¿Qué tal si se pregunta sólo a los estudiantes de ultimo año, ya que se supone que son los que mejor conocen lo que está sucediendo en la comunidad? Tampoco puede hacerse eso porque los estudiantes de los grados inferiores también usan drogas. ¿Qué hacer?

¡Decisiones, decisiones, decisiones! Y decisiones que no se pueden tomar a la ligera éxito del proyecto depende de la forma en que se seleccione a la gente que participara en el estudio, sea que usted vaya a distribuir un cuestionario o administrar una prueba de memoria. En este capítulo examinaremos varias formas de seleccionar a lo proyectos de investigación y la importancia del proceso de selección para estudio. Todo se relaciona con muestras y muestreo.

Poblaciones y muestras

En varios puntos de los primeros cuatro capítulos de este libro usted ha leído acerca de la importancia de inferir los resultados de un experimento de una muestra a una población. Ésta es la base del método inferencial. Si no es posible someter a prueba a todos los miembros de la población, la única opción es seleccionar una **muestra, un subconjunto de la** población. Una buena técnica de muestreo incluye maximizar el grado selecto representa a la población.

Una **población** es un grupo de posibles participantes al cual usted desea generalizar los resultados del estudio. Y el nombre del juego aquí es **generalizabilidad pues** para que un estudio que los resultados de una investigación tengan algún significado más allá de la situación limitada en la que se obtuvieron originalmente es necesario que se pueda generalizar desde una muestra a una población. Si los resultados son generalizables, podrán aplicar se a diferentes poblaciones que tengan las mismas características pero que estén en diferentes. Si los resultados no son generalizables (es decir, si la muestra seleccionada no es una representación exacta de la población), los resultados sólo pueden aplicarse a los integrantes de la misma muestra que participaron en la investigación original y nadie más.

Por ejemplo, supongamos que desea averiguar qué actitudes tienen los estudiantes de educación media hacia el registro de sus casilleros, en las escuelas donde se cuenta con esta ventaja para los alumnos. Se podria administrar el cuestionario a un grupo de estudiantes de último año del área de química. Pero, ¿qué tanto se parecen a la población de estudiantes que asisten a todas las escuelas de educación media del distrito no mucho. O bien, podríamos hacer las preguntas a 10% de las estudiantes de sexo femenino de primer y segundo años de bachillerato de todas las escuelas. Esta selección abarca un grupo mucho más grande que los 30 o 40 estudiantes del grupo del área de química ¿qué tan representativas son ellas? Otra vez, no mucho.

Nuestra tarea, entonces, es idear un plan que asegure que la muestra de estudiantes que seleccionemos sea representativa de *todos* los estudiantes del distrito. Si logramos este objetivo, podremos generalizar los resultados a toda la población con un alto grado de confianza, incluso si sólo usamos un porcentaje pequeño de los miles de estudiantes de bachillerato. En otras palabras, si hacemos correctamente el trabajo (seleccionar una muestra), podremos generalizar los resultados. ¿Cómo sabremos si "hicimos correctamente el trabajo"? Presentaremos algunas pautas en este capitulo, pero una forma de efectuar una especie de autoverificación es hacernos la pregunta: ¿La *muestra que seleccioné de la población tiene al parecer todas las características de la población, en la misma proporción?* ¿Es la muestra, efectivamente, una *minipoblación?*

Para entender el proceso de muestreo, primero necesitamos distinguir entre dos tipos generales de estrategias de muestreo: probabilísticas y no probabilísticas. El **muestreo probabilístico** es un tipo de muestreo en el que se conoce la probabilidad de seleccionar un miembro individual de la población. Si hay 4500 estudiantes en todas las escuelas de bachillerato, y si 1000 de ellos están en el último año, la probabilidad de seleccionar un estudiante de último año como parte de la muestra es de 1000:4500 =0.22.

El muestreo **no** probabilístico es aquel en el que se desconoce la probabilidad de seleccionar cualquier miembro individual de la población. Por ejemplo, si no sabemos cuántos muchachos están inscritos en las escuelas de educación media del distrito, no podremos calcular la probabilidad de seleccionar cualquiera de ellos.

Estrategias de muestreo probabilístico

Las estrategias de muestreo probabilístico son las más utilizadas porque la selección de los participantes está determinada por el azar. Puesto que la decisión de quién entra y quién no entra en la muestra está regida por reglas no sistemáticas y aleatorias, hay una buena posibilidad de que la muestra represente verdaderamente a la población.

Muestreo aleatorio simple

El tipo más común de procedimiento de muestreo probabilístico es el muestreo aleatorio **simple.** Aquí, cada miembro de la población tiene una probabilidad *igual* e *independiente* de ser seleccionado como parte de la muestra. Las palabras clave aquí son *igual* e *independiente. Igual,* porque no existe alguna predisposición a escoger una persona en lugar de otra. *Independiente* porque el hecho de escoger una persona no predispone al investigador en favor o en contra de escoger otra persona dada. Si se muestrea aleatoriamente, las

características de la muestra deberán ser muy parecidas a las de la población.

Por ejemplo, ¿estaríamos realizando un muestreo aleatorio simple si escogiéramos cada quinto nombre del directorio telefónico? No, porque estaríamos violando ambos criterios de igualdad e independencia. En primer lugar, si comenzamos con el quinto nombre de la página 234 del directorio, los nombres 1, 2, 3 y 4 nunca tuvieron una probabilidad igual de ser escogidos. Segundo, si escogemos el número 5 de la lista y luego cada quinto nombre de ahí en adelante, sólo los nombres 10, 15, 20, etc., tendrán la posibilidad de ser escogidos. En otras palabras, el proceso de selección ya no es independiente.

El proceso de muestreo aleatorio simple consta de cuatro pasos:

1. Definir la población de la cual se desea seleccionar una muestra.
2. Listar todos los miembros de la población.
3. Asignar números a cada miembro de la población.
4. Aplicar un criterio para seleccionar la muestra deseada.

Por ejemplo, en la tabla 4.1 se muestra una lista de 50 nombres a los que ya se han asignado números (pasos 1, 2 y 3). No es una población muy grande pero sí excelente para fines ilustrativos. De esta población seleccionaremos una muestra de 10 individuos utilizando una **tabla de números aleatorios.** Veamos cómo funciona esto.

1. Juana	**11.** Susana	**21.** Eduardo.	**31.** Daniela	**41.** Narciso
2. Braulio	12. Nora	22. Jorge	32. Bernardo	42. Penélope
3. Enriqueta	13. Diego	23. Cecilia	33. Diana	43. Inés
4. Laura	14. Juan 5.	24. Gabriela	34. Felipe	44. Débora
5. Miguel	15. Bruno	25. María	35. Federico	45. Carla
6. Sara	16. Lorenzo	26. Celia	36. Manuel	46. Vicente
7. Teresa	17. Roberto	27. Silvestre	37. Donato	47. Gema
8. Joaquina	18. Esteban	28. Felicia	38. Eduardo M.	48. Elena
9. Jaime	19. Samuel	29. Javier	39. Timoteo	49. Alejandro
10. Tomás	20. Mario	30. Enrique	40. Manuel G.	50. Juan D.

Tabla 4.1 He aquí un grupo de *50* nombres que constituye una "población" para nuestros fines. Observe que cada uno está numerado **y** listo para seleccionarse.

Cómo usar una tabla de números aleatorios

Una tabla de números aleatorios es un criterio magnifico, ya que la forma como se generan los números de la tabla carece totalmente de predisposición. Por ejemplo, en la tabla 4.2 hay cantidades casi iguales de dígitos 1, 2, 3, 4, 5, etc. En consecuencia, la probabilidad de seleccionar un número que termine en 1 o en 2 o en 3, etc., es igual. Esto implica que si se relacionan nombres con los números la probabilidad de seleccionar cualquier nombre dado es también igual.

23157	48559	01837	25993
05545	50430	10537	43508
14871	03650	32404	36223
38976	49751	94051	75853
97312	17618	99755	30870
11742	69183	44339	47512
43361	82859	11016	45623
93806	04338	38268	04491
49540	31181	08429	84187
36768	76233	37948	21569

Tabla 4.2 Tabla parcial de números aleatorios. En una tabla de números aleatorios, cabe esperar que habrá cantidades iguales de los dígitos individuales, distribuidos al azar entre todos los números.

Teniendo lo anterior en mente, seleccionemos un grupo de diez nombres utilizando la tabla de números aleatorios de la tabla 4.2. Siga estos pasos.

1. Escoja un punto de partida en algún lugar de la tabla cerrando los ojos y colocando su dedo (o la punta de un lápiz) en cualquier parte de la tabla. Al seleccionar el punto de partida de este modo aseguramos que no se escogerá algún punto de partida (o nombre) especifico.

   Para este ejemplo, el punto de partida fue la primera columna de números, en la última fila (36768); la punta del lápiz quedó en el

cuarto digito, el número 6.

2. El primer número de dos dígitos, entonces es 68 (en negritas en la tabla 4.3). Puesto que la población llega hasta 50, y no hay un nombre en el lugar 68, pasamos por alto este número y consideramos el siguiente número de dos dígitos. Ya que no podemos bajar más en la tabla, pasamos al tope de la siguiente columna y leemos hacia abajo, una vez más seleccionando los primeros dos dígitos. Por comodidad, hemos separado los pares de dígitos en la tabla 4.3.

23157	*48559*	01837	25993
05545	50430	10537	43508
14871	03650	32404	36223
38976	4975 1	94051	75853
97312	1761 8	99755	30870
11742	69183	44339	47512
43361	82859	11016	45623
93806	04338	38268	04491
49540	31 18 1	08429	84187
36768	76233	37948	21569

Tabla 4.3 Punto de partida para la selección de 10 casos utilizando la tabla de números aleatorios. Se puede comenzar en cualquier punto, en tanto ese punto se determine al azar y no se escola intencionalmente.

3. ¡48! ¡Lo logramos! La persona 48 de la lista es Elena, y ella se convierte en la primera de la muestra de 10 miembros.
4. Si seguimos seleccionando números de dos dígitos hasta haber hallado 10 valores entre 01 y 50, habremos seleccionado los nombres de la tabla 4.1 que corresponden a los números que aparecen en negritas en la tabla 4.4. He aquí un desglose de cuáles números funcionaron y cuáles no para los fines de escoger una muestra aleatoria de 10 personas de la población de 50.

- Leyendo hacia abajo en la primera columna de números de dos dígitos, 48, 50, 03, 49 y 17 están bien, porque quedan dentro del intervalo de Ol a 50 (el tamaño de la población) y no se han seleccionado antes,
- 69 y 82 están fuera del intervalo,
- 04 y 31 están bien, y
- 76 está fuera del intervalo

Puesto que no podemos bajar más por la primera columna de números de dos dígitos, hay que subir al siguiente conjunto de números de dos dígitos (en la misma columna de cinco dígitos) en la parte superior de la columna, que comienza con el número 55.

- 55 no está dentro del intervalo,
- 43 está bien,
- 65, 75 y 61 no son aceptables,
- l8 sí,
- 85 no, y (¡por fin!)
- **33 si, y ya tenemos las 10 personas.**

Aqui están:	**Número**	**Nombre**
	48	Elena
	50	JuanD.
	03	Enriqueta
	49	Mejandro
	17	Robefto
	04	Laura
	31	Daniela
	43	Inés
	18	Esteban
	33	Diana

Ahora ya tiene usted una muestra de 10 nombres de una población de 50, seleccionados totalmente al azar. Su muestra se escogió al azar porque la distribución de los números en la tabla parcial de números aleatorios de la tabla 4.2 se generó al azar. ¿Es una simple coincidencia que tres de los primeros cinco números (48, 50, 03, 49, 17) de la tabla parcial de números aleatorios estén agrupados? ¡Absolutamente! Este grupo de cinco es la mejor aproximación y la más representativa de cualquier muestra de cinco de toda la población, dado que cada miembro de la población tiene la misma probabilidad independiente de ser escogido.

23157 48559 01837 25993

05545	50430	10537	43508
14871	03650	32404	36223
38976	4975 1	94051	75853
97312	17618	99755	30870
11742	69183	44339	47512
43361	82859	11016	45623
93806	04338	38268	04491
49540	3118 1	08429	84187
36768	76233	37948	21569

Tabla 4.4 Selección de los números de la tabla de números aleatorios que corresponde a los 10 nombres seleccionados como muestra. Observe que los pares de números individuales están en negritas para poder distinguirlos de los que no se seleccionaron.

El gran supuesto, desde luego, es que los nombres de la población (tabla 4.1) se listaron de manera aleatoria. En otras palabras, los nombres del 01 al 20 no se listaron como los primeros 20 de los 50 porque provienen de un vecindario distinto, tienen mucho dinero, no tienen hermanos, o tienen alguna otra característica que pudiera predisponer de alguna forma la selección.

La regla general es usar un criterio que no tenga relación con lo que se está estudiando. Por ejemplo, si va a realizar un estudio sobre ofrecerse como voluntarios, ¡no va a pedir voluntarios!

Muestreo sistemático

Otro tipo de muestreo aleatorio se denomina muestreo sistemático, en el cual se escoge cada *k-ésimo* nombre de la lista. El término *k-ésimo* representa un número entre O y el tamaño de la muestra que usted quiere seleccionar. Por ejemplo. He aquí cómo se usaría un muestreo sistemático para escoger 10 nombres de la lista de 50 que aparece en la tabla 4.1. Para hacerlo, siga estos pasos.

1. Divida el tamaño de la población entre el tamaño de la muestra deseada. En este caso, 50 dividido entre 10 es 5. Por tanto, se escogerá cada quinto nombre de la lista.
2. Como punto de partida, escoja un nombre de la lista al azar. Haga esto usando el método de "apuntar con los ojos cerrados" o, si la lista está numerada, utilice cualquier digito o par de dígitos del número de serie de un billete que traiga en su cartera. El billete que usamos en este ejemplo tenia como sus dos primeros dígitos 43, y éste será nuestro punto de partida.
3. Una vez determinado el punto de partida, seleccione cada quinto nombre. En este ejemplo, si usamos los nombres de la tabla 4.1 y comenzamos con Inés (#43), la muestra consistirá en Elena (#48), Enriqueta (#3), Joaquina (#8), Diego (#13), Esteban (#18), Cecilia (#23), Felicia (#28), Diana (#33) y Eduardo M. (#38).

El muestreo sistemático es más fácil y menos problemático que el aleatorio, y ésta es una razón por la que muchas veces se prefiere. El muestreo sistemático también es un poco menos deseable. Es evidente que se viola el supuesto de que cada miembro de la población tiene la misma probabilidad de ser seleccionado. Por ejemplo, dado que el punto de partida es Inés (#43), seria imposible seleccionar a Débora (#44).

Muestreo estratificado

Los dos tipos de muestreo aleatorio que acabamos de ver funcionan bien si no nos interesan características específicas de la población (como edad, sexo, grupo étnico, grupo de habilidad, etc.). En otras palabras, si se seleccionara otro conjunto de 10 nombres, supondríamos que, como ambos grupos se escogieron al azar, son efectivamente iguales. Pero, ¿qué sucede si desde un principio los individuos de la población no son "iguales"? En tal caso, querremos asegurarnos de que el perfil de la muestra coincida con el perfil de la población, y esto se hace creando una **muestra estratificada.**

La teoría en que se basa el muestreo (y todo el proceso de inferencia) es a grandes rasgos la siguiente: si usted puede seleccionar una muestra lo más representativa posible de una población, cualesquier observaciones que pueda hacer sobre esa muestra se deberán cumplir también para la población. Hasta aquí vamos bien. Sin embargo, hay ocasiones en que el muestreo aleatorio deja demasiado al azar, sobre todo si no hay garantía de que las distribuciones de los miembros de la población dentro de la muestra sean iguales. En esos casos se utiliza un muestreo estratificado.

Por ejemplo, si la población es 82% católicos, 14% protestantes y 4% judíos, la muestra deberá tener las mismas características *si la religión es una variable importante.* Es crucial entender la última parte del enunciado anterior. Si ésta o aquella característica de la población no tiene relación alguna con lo que se está estudiando, no hay razón para preocuparse por obtener una muestra que siga el mismo patrón de la población y esté estratificada según una de esas variables.

Supongamos que la lista de nombres de la tabla 4.1 representa una población *estratificada* (hombres y mujeres), y que el tema del estudio son las actitudes hacia el aborto. Puesto que las diferencias de sexo o género podrían ser importantes, usted quiere una muestra que refleje las diferencias de género de la población. La lista de 50 nombres consiste en 20 mujeres y 30 hombres, o sea, es 40% mujeres y 60%

hombres. La muestra de 10 deberá reflejar esa distribución y tener 4 mujeres y 6 hombres. He aquí cómo se seleccionaría semejante muestra utilizando muestreo aleatorio estratificado.

1. Los hombres y las mujeres se listan por separado.
2. Cada miembro de cada grupo recibe un número. En este caso, los hombres se numerarian del 01 al 30 y las mujeres del 01 al 20.
3. Usando una tabla de números aleatorios, se seleccionan 4 mujeres al azar de la lista de 20.
4. Usando una tabla de números aleatorios, se seleccionan 6 hombres al azar de la lista de 30.

Aunque podrían presentarse ejemplos sencillos (con un solo estrato o capa) como éste, en muchos casos será necesario estratificar según más de una variable. Por ejemplo, en la figura 4.1 una población de 10000 niños se estratifica según las variables de año (40% de primer año, 40% de tercer año, 20% de quinto año) y lugar de residencia (30% rural, 70% urbano). Se utiliza la misma estrategia: seleccionar 10% (puesto que 1000 es 10% de 10 000) de cada una de las capas estratificadas para producir la muestra de la figura 4.1. Por ejemplo, de los 1200 niños rurales de primer año, se seleccionó aleatoriamente 10%, o 120 niños. De forma similar, se escogieron 140 niños urbanos de quinto año.

	Año			
Residencia	**1**	**3**	**5**	**Total**
Rural	1200	1200	600	3000
	(120)	(120)	(60)	(300)
Urbana	2800	2800	1400	7000
	(280)	(280)	(140)	(700)
Total	4000	4000	2000	10000
	(400)	(400)	(200)	(1000)

Figura 4.1 La selección de una muestra cuando hay más de un estrato consiste en tomar una proporción de cada nivel. Aquí el tamaño de la muestra se indica entre paréntesis debajo del tamaño de cada grupo de la población.

Muestreo por cúmulos

Ultimo tipo de muestreo probabilistico es el **muestreo por cúmulos, en el que se** seleccionan *unidades* de individuos y no los individuos mismos. Por ejemplo, usted podría estar realizando una encuesta de las actitudes de los padres hacia la vacunación. En lugar de asignar aleatoriamente padres individuales a dos grupos (digamos, aquellos a los que se les enviará material informativo y aquellos a los que no se les enviará), usted podría simplemente identificar 30 consultorios pediátricos de la ciudad y luego, utilizando una tabla de números aleatorios, seleccionar 15 para un grupo y designar 15 para el segundo grupo.

El muestreo por cúmulos ahorra mucho tiempo, pero hay que tener la seguridad de que las "unidades" (en este caso los padres que acuden a cada pediatra) sean lo suficientemente homogéneas como para que cualesquier diferencias que haya dentro de la unidad misma no contribuyan a una predisposición. Por ejemplo, si un pediatra se niega a vacunar a los niños antes de cierta edad, eso introduciría una predisposición que usted querría evitar.

Estrategias de muestreo no probabilístico

La segunda categoría principal de estrategias de muestreo, el muestreo no probabilístico; comprende aquellas en las que se desconoce la probabilidad de escoger un solo individuo. En este caso, hay que suponer que los miembros en potencia de la muestra no tienen una probabilidad igual e independiente de ser seleccionados. Examinemos algunos de estos métodos de muestreo.

Muestreo de conveniencia

El muestreo de conveniencia es justo lo que indica su nombre. Un entrenador de fútbol da a cada miembro de su equipo un cuestionario. El público (el equipo) es cautivo, y es una forma muy cómoda de generar una muestra. ¿Fácil? Si. ¿Aleatoria? No. ¿Representativa? Tal vez, pero hasta cierto grado. Quizá usted reconozca este método de muestreo como la razón por la que tantos experimentos en psicología, realizados en Estados Unidos, se basan en resultados obtenidos usando estudiantes universitarios de segundo años; estos estudiantes son un público cautivo y a menudo deben participar para obtener los créditos de un curso.

Muestreo por cuotas

Usted podría encontrarse en una situación en la que necesitara obtener una muestra estratificada según ciertas variables, pero por alguna razón no es posible efectuar un muestreo estratificado. En este caso, el muestreo por cuotas podría ser lo mejor.

En el muestreo **por** cuotas se escogen personas con las características deseadas (digamos, niños rurales de primer año) pero no se selecciona aleatoriamente de la población un subconjunto de todos esos niños, como ocurriría en un muestreo estratificado proporcional. Más bien, el investigador continuaría reclutando niños hasta cumplir con la cuota de 120. El 1760 niño rural de primer año no tiene posibilidad de ser escogido, y ésta es la principal razón por la que esta técnica de muestreo es no probabilística.

He aquí otro ejemplo de sistema por cuotas. Digamos que usted tiene que entrevistar 20 estudiantes universitarios de primer año de ambos sexos. Primero usted podría entrevistar 10 hombres y, sabiendo que la distribución de hombres y mujeres está dividida aproximadamente 50/50, usted entrevistaría las siguientes 10 mujeres que llegaran, y con ello habría acabado. Si bien el muestreo por cuotas es mucho más fácil que el estratificado, también es menos preciso. Imagine cuánto más fácil seria encontrar 10 estudiantes de primer año de sexo masculino que 10 hombres específicos, que es lo que usted tendría que hacer si realizara un muestreo estratificado.

En la tabla 4.5 se listan los diferentes tipos de estrategias probabilísticas y no probabilisticas que hemos visto, y se indica cuándo deben usarse, junto con algunas de sus ventajas y desventajas.

Muestras y error de muestreo

Por más que se esfuerce el investigador, es imposible seleccionar una muestra que represente *perfectamente* a la población. Desde luego, el investigador podría seleccionar toda la población como muestra, pero esto elimina el propósito del muestreo: hacer una inferencia a una población con base en una muestra más pequeña.

Una forma de expresar la falta de congruencia entre la muestra y la población se expresa como el **error de muestreo.** Este error es la diferencia entre las características de la muestra y las características de la población de la cual se seleccionó la muestra. Por ejemplo, la estatura promedio de 10000 alumnos de quinto año es 102 cm. Si usted toma 25 muestras de 100 alumnos de quinto año y calcula la estatura promedio para cada grupo de 100 alumnos, tendrá un promedio para cada grupo, es decir, 25 promedios. Si todos esos promedios son exactamente 102 cm, no habrá error de muestreo. Sin embargo, es casi seguro que no se obtendrá tal resultado. La vida no es tan fácil ni la selección de muestras es así de perfecta. Lo más probable es que usted obtenga valores como 102.5, 102.2, 101.6, 100.1 cm, etc. La cantidad de variabilidad, o la dispersión, de estos valores nos da una idea de la magnitud del error de muestreo. Cuanto mayor es la diversidad de los valores de las muestras, mayor es el error.

Piense un momento en lo que sucedería si toda la población de 10 000 alumnos de quinto año filera la muestra. ¡La estatura promedio seria 102 cm! ¡Perfecto! ¡Cero error! ¿La moraleja? Cuanto mayor es la muestra, menor es el error de muestreo, porque cuanto

Tipo de muestreo	Cuándo debe usarse	Ventajas	Desventajas
Estrategias probabilisticas			
Muestreo aleatorio simple	Cuando los miembros de la población son similares	Asegura un alto grado de representatividad	Tardado y tedioso
Muestreo sistemático	Cuando los miembros de la población son todos similares	Asegura un alto grado de representatividad; no hay que usar una tabla de números aleatorios	Menos verdaderamente aleatorio que el muestreo aleatorio simple
Muestreo aleatorio estatificado	Cuando la población es de naturaleza heteroge nea y contiene varios grupos distintos	Asegura un alto grado de representatividad de todos los estratos de la población	Tardado y tedioso
Muestreo por cúmulos	Cuando la población consiste en unidades más que en individuos	Fácil y cómodo	Posibilidad de que los miembros de las unidades difieran entre sí y reduzcan la efectividad del muestreo
Estrategias de muestreo no probabilisticas			
Muestreo de **conveniencia**	Cuando la muestra es cautivo	Cómodo y económico	La representatividad es dudosa
Muestreo por cuotas	Cuando hay estratos pero no es posible un muestreo estratificado	Asegura cierto grado de representatividad de todos los estratos de la población.	La representatividad es dudosa

Tabla 4.5 Resumen de las diferentes técnicas de muestreo. Al comenzar a considerar la selección de una muestra, hay que tener en cuenta

muchos factores, todos los cuales afectarán el tipo y el tamaño de la muestra que escala.

más grande es una muestra más se aproxima su tamaño al tamaño de la población y es más representativa de esa población.

El proceso exacto para calcular el error de muestreo, que se expresa como un valor numérico, rebasa el alcance de este libro, pero el lector debe tener presente que su propósito al seleccionar una buena muestra es minimizar ese valor. Cuanto más pequeño sea el valor, menor discrepancia habrá entre la muestra y la población.

Pero hay más. Ya sabemos que cuanto más grande es una muestra, más representativa es de una población. Digamos que llega el momento de probar si hay una diferencia entre muestras. Resulta que, cuanto mejor las muestras representan su respectiva población, más exacta es la prueba de las diferencias. En otras palabras, un mejor muestreo da lugar a pruebas más exactas y verdaderas de las diferencias entre poblaciones.

¿Cómo minimizamos el error de muestreo? Sencillo. Utilizamos buenos procedimientos de selección como los que describimos al principio del capitulo y aumentamos el tamaño de la muestra hasta donde es posible y razonable. La siguiente pregunta que usted está a punto de hacer (espero) es, «¿qué tan grande debe ser la muestra?" ¡Qué bueno que preguntó! Pasemos a la última sección del capitulo para entender mejor la respuesta a esa pregunta.

¿Qué tan grande es grande?

Ahora que usted sabe algo acerca de muestreo, ¿qué tantos de esos estudiantes de educación media necesita seleccionar de la población de 4500? Si con 50 basta, ¿no es mejor 500? ¿Y por qué no 1500, si tiene el tiempo y los recursos suficientes para el proyecto?

Usted ya sabe que una muestra demasiado pequeña no es representativa de la población, y que una demasiado grande es una exageración. Muestrear demasiados estudiantes de educación media anularía el propósito del muestreo, pues ya no se estaría aprovechando la utilidad de la inferencia. Hay quienes piensan que cuanto más grande es una muestra, mejor; pero tal estrategia no es lógica desde un punto de vista económico o científico. Una muestra demasiado grande no aumenta la precisión de la prueba de la pregunta en un grado que justifique el costo y el trabajo de obtener una muestra de ese tamaño. Necesitamos un método para calcular el número *real* de estudiantes de educación media que conviene seleccionar para la muestra. Recuerde, cuanto menos representativa de la población sea la muestra, mayor será el error de muestreo presente. Además, cuanto mayor sea el error de muestreo, menos generalizables a la población serán los resultados y menos precisa será la prueba de la hipótesis nula.

Estimación del tamaño de la muestra

Es posible estimar el tamaño de muestra que se necesita para representar de la manera más exacta posible la población, pero esto va a requerir un poco de fe por parte del lector por dos razones. Primera, implica conceptos que están un poco más allá del nivel de este libro. Sin embargo, dado que el lector es inteligente y entusiasta, es probable que entienda los fundamentos de este material. Segundo, y más importante, es preciso estimar ciertos valores, lo cual muchas veces es difícil aun para los expertos.

He aquí los componentes que necesitamos para estimar los tamaños de muestra de dos grupos que se van a comparar entre si.

s =la desviación estándar de la variable dependiente, es decir; qué tanta variabilidad hay en los puntajes para esta variable
t = el valor critico necesario para rechazar la hipótesis nula
D = la magnitud que usted estima tiene la diferencia promedio entre los dos grupos.

Tomemos como ejemplo el caso de un estudio en el que usted está tratando de determinar si un programa de lectura es más eficaz que otro. La variable dependiente es el puntaje de lectura promedio.

La forma como estimamos la desviación estándar de una muestra antes de probar los miembros de la muestra es examinando los puntajes y estadísticas descriptivas de la variable dependiente cuando esta variable se usó en otros estudios. Por ejemplo, digamos que esta prueba de lectura tiene un manual, y que el manual indica que la desviación estándar para alumnos de primero de bachillerato en el pasado ha sido de 15.

A continuación, introducimos el valor critico que refleja la diferencia que usted esperaría si el azar (y no cualquier tratamiento) fuera el único factor que operara. Los valores críticos se pueden encontrar en tablas especializadas. Examinaremos más a fondo este concepto en el capítulo 8.

Por último, necesitamos tener una idea de a cuánto ascenderá la diferencia entre los dos grupos. Si el puntaje que puede obtenerse en la prueba de lectura está entre 20 y 50, ciertamente no esperaremos que la diferencia entre los dos promedios sea de más de 30.

Hay que ser razonables.

Para este ejemplo,

- **la desviación estándar es igual a 15 (supuestamente informada en otro estudio),**
- el valor t es 2 (el valor que obtuvimos de nuestra tabla especial), y
- la diferencia estimada entre los grupos es de 6, que también es una cifra de orden de magnitud.

Dados estos valores, podemos usar la siguiente fórmula para calcular *n,* que es el tamaño de muestra estimado necesario para cada grupo bajo las restricciones que se han establecido:

$$\frac{n = 2s^2 \times t^2}{D_2}$$

donde *n* = tamaño de muestra estimado para cada grupo
s = desviación estándar basada en estudios anteriores
t = valor critico
D = nuestra estimación de la diferencia entre cada grupo

Si sustituimos valores, la fórmula queda así:

$$n = \frac{(2 \times 15^2) \times 2.00^2}{6^2}$$

Esta fórmula revela que, dadas las anteriores restricciones, necesitamos unas 50 personas en cada grupo.
Lo que resulta interesante examinar son los efectos de cambiar ciertos valores (como la desviación estándar y la diferencia estimada) sobre el valor estimado de *n,* el tamaño de muestra que necesitamos. Vea lo que sucede en la tabla 4.6 cuando cambia la diferencia estimada entre los dos grupos.

Efecto **sobre el tamaño de la muestra cuando modificamos la magnitud del error estimado**

s	**t**	**d**	**n**
15	**2**	**6**	**50**
10	2	6	22
5	2	6	6

Efecto sobre el tamaño de la muestra cuando modificamos la magnitud de la diferencia estimado

s	**t**	**d**	**n**
10	**2**	**2**	200
l0	**2**	**4**	*50*
l0	**2**	**6**	22

Tabla 4.6 Podemos ver lo drástico que puede ser el efecto cuando cambiamos lo diferencia estimada entre grupos.

Cuanto mayor es la diferencia estimada (de 2 a 4 a 6), menos grande necesita ser el tamaño de las muestras. ¿Por qué? Porque cuanto mayor es la diferencia entre las muestras, menos representativas tienen que ser éstas de la población para que podamos ver el efecto, si existe. Esto es un poco como decir que es más fácil percibir una diferencia grande entre dos cosas que una diferencia pequeña. En la tabla 4.6 podemos ver lo drástico que puede ser el cambio. Si duplicamos el tamaño estimado de la diferencia (de 2 a 4), reducimos drásticamente el tamaño de muestra estimado necesario, de 200 en cada grupo a 50.

¿Qué es lo más importante en cuanto al tamaño de muestra? Tenga presentes las consideraciones siguientes:

- En general, cuanto mayor es la muestra (dentro de limites razonables), menor es el error de muestreo y mejores son los resultados.
- Si va a utilizar varios subgrupos en su trabajo (como hombres y mujeres, ambos de 10 años de edad, y residentes rurales saludables y no saludables), asegúrese de que su selección inicial de sujetos sea lo bastante grande como para contemplar la división necesaria en grupos de sujetos.
- Si usted va a enviar por correo encuestas y cuestionarios (y ya sabe usted lo que puede sucederles a muchos de ellos), contemple un aumento de 40%-50% en el tamaño de muestra para dar cabida a las pérdidas en el correo y la falta de cooperación de los sujetos.
- Por último, recuerde que lo grande es bueno, pero lo apropiado es mejor. No gaste su dinero arduamente ganado ni su valioso tiempo generando muestras más grandes de lo necesario.

Aunque algunas personas podrían no estar de acuerdo con los temas que usted ha escogido para estudiar, lo que usted escoja es asunto

suyo en tanto presente una justificación razonable de lo que está haciendo. En cambio, su selección de una muestra es cuestión aparte. Existen muchas formas correctas de hacerlo, pero también está la forma incorrecta. Si usted escoge la forma incorrecta (porque es arbitrario y no sigue ningún plan), bien podría sabotear toda su labor de investigación porque sus resultados podrán carecer de generalizabilidad y, por tanto, de utilidad para la comunidad científica.

Ejercicios

1. Suponga que usted es el investigador principal de un estudio longitudinal que está rastreando las preferencias vocacionales desde el bachillerato hasta la edad adulta media. Liste los pasos que seguiria para seleccionar la muestra que usará en el estudio.

2. ¿Por qué es tan útil una tabla de números aleatorios como herramienta para asignar personas a diferentes grupos?

3. ¿Qué diferencia hay entre una estrategia de muestreo probabilístico y uno no probabilistico? Dé un ejemplo de cada una. También, cite las ventajas y desventajas de cada tipo de muestreo.

4. ¿Cuál es la forma más fácil de reducir el error de muestreo? ¿Qué relación hay entre el error de muestreo y la generalizabilidad de los resultados de un estudio? Por último, qué sucede con el error de muestreo a medida que aumenta el tamaño de la muestra? ¿Por qué?

5. Si tiene una población de 10000 niños (50% hombres y 50% mujeres, 70% blancos y 30% de otras razas, y 57% de familias con un solo padre y 43% de familias con ambos padres), ¿qué pasos seguiria para seleccionar una muestra representativa de 150 niños?

6. Utilizando una tabla de números aleatorios, escoja seis nombres de la siguiente lista de 10.

   Miguel
   Susana
   Sara
   Hernán
   Selina
   Enriqueta
   Guillermo
   David
   Sandra
   Eduardo

   ¿Cuántos de los seis esperaría usted que fueran hombres, y cuántos mujeres? ¿Por qué?

7. ¿Qué implicaciones tiene usar una muestra demasiado grande o una demasiado pequeña?

8. A usted le interesa la diferencia entre niños que reciben enseñanza creativa y niños que reciben enseñanza tradicional, en cuanto a la creatividad de los niños. Se utiliza una escala de creatividad que produce un posible puntaje entre 70 y 100. Por investigaciones previas, usted estima que la desviación estándar es 15. Usted tiene acceso a 60 individuos (30 en cada grupo) y estima que la diferencia en la escala de creatividad entre los niños con enseñanza creativa y aquellos con enseñanza tradicional será de 10 puntos. Con las restricciones señaladas, ¿qué tamaño de muestra se necesita para cada grupo?

9. ¿Qué riesgos conlleva aumentar demasiado el tamaño de una muestra?

10. ¿Cuándo debe usarse muestreo por cúmulos y muestreo aleatorio simple?

11. ¿Qué es error de muestreo?

CAPITULO CINCO
Medición del comportamiento

En Neil J. Salkind "Métodos de Investigación", Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- Por qué la medición es una parte importante del proceso de investigación
- Qué incluye el proceso de medición
- Cuáles son los diferentes niveles de medición, y cómo se aplican