

2. RECOLECCIÓN Y REVISIÓN DE LOS DATOS

RECOLECCIÓN DE LOS DATOS

Idealmente, los datos que el investigador registra se obtienen directamente de los elementos en estudio mediante la **observación** (observación directa, medición física, medición química) o el **interrogatorio** (entrevistas personales, cuestionarios auto-administrados, diarios). Cuando este es el caso se dice que los datos proceden de **fuerza primaria**. Los datos de **fuerza secundaria** son aquellos que se toman de registros previos, generalmente elaborados con otros propósitos diferentes a los de nuestra investigación. Su utilidad para el análisis dependerá de la calidad de la información y de su accesibilidad administrativa.

Los datos que proceden de una fuerza primaria tienen que ser registrados por el investigador. Ocasionalmente los datos de fuerza secundaria están disponibles en archivos de cómputo y se capturaron de tal manera que están listos para su tabulación, pero lo más frecuente es que el investigador también tenga que buscarlos y registrarlos. En ambos casos, lo más frecuente es que el investigador tenga que registrarlos en alguna forma especialmente diseñada para este objetivo: una **forma de registro de datos**.

FORMA DE REGISTRO DE DATOS

Las formas de registro de datos se tratan con mayor profundidad en los textos dedicados a la medición de la exposición (Armstrong *et al.*, 1992), de la salud (McDowell y Newell, 1996) o sobre cómo elaborar cuestionarios (Fink, 1995). En este capítulo solamente presentamos algunas consideraciones que se tienen que hacer cuando se elaboran.

Antes de empezar a diseñar el instrumento a utilizar es necesario que el investigador enumere todas las variables de interés. En cada una de ellas se deberán definir tres aspectos:

- Su naturaleza, de tal manera que sea específica (que distinga la variable de interés de posibles confusores) y sensitiva (que incluya todas las maneras en que el sujeto se encuentra con la variable de interés).
- La dosis, en sus diferentes expresiones: dosis acumulada, tasa de exposición, exposición promedio, dosis pico.
- El tiempo, o periodo relevante en que se presentó la característica de interés.

Estos tres aspectos de las variables deberán ser interpretados por el investigador y traducidos en preguntas, que en conjunto integrarán la forma de registro de datos. Para algunas variables estas definiciones son muy sencillas y generalmente pueden ser exploradas con una pregunta (sexo, edad en años cumplidos) o se pueden medir directamente en el sujeto en estudio (talla, peso, presión arterial). Otras son engañosamente sencillas, como el alcoholismo, la escolaridad o la ocupación. En algunos casos el proceso puede ser muy complejo y la medición se tiene que realizar de manera indirecta por medio de varias preguntas cuyas respuestas se resumen en un índice. A este grupo pertenecen las mediciones de la calidad de vida, el estrés, la depresión y la inteligencia.

Al pensar en la forma de registro es necesario considerar las siguientes alternativas:

- 1) Utilizar uno ya existente. Esta debería ser nuestra primera opción. Al hacerlo ahorramos tiempo y recursos, además de que la comparación de nuestros resultados con los de otros estudios puede ser más sencilla cuando en ambos trabajos se utilizó el mismo instrumento. Pero es necesario tener cuidado porque un instrumento utilizado previamente no necesariamente es el mejor para una nueva investigación, aunque ya estuviera probado y validado.
- 2) Adaptar uno ya existente. Esta es una opción muy útil cuando ya existe un instrumento previo, pero que tiene alguno de los siguientes inconvenientes:
 - El instrumento es muy extenso para el estudio. Pudiera ser que el instrumento original fue diseñado para una entrevista, pero que se desea utilizar en un cuestionario auto-contestado, o que el tiempo en que se contestará excede el tiempo que se dispone para captar los datos.
 - La población es diferente. Un instrumento desarrollado para un grupo de población urbana puede no ser adecuado para una población rural. Particularmente si los niveles de escolaridad o el uso del lenguaje son diferentes.
 - El instrumento está escrito en otra lengua. En estos casos el instrumento deberá ser traducido.
 - Es necesario extender, reordenar o realizar otros cambios en la manera en que los datos se colectan. Esto es necesario cuando el investigador considera que al instrumento se le deben agregar más preguntas, cuando el orden de las mismas debe ser cambiado, o cuando las preguntas se aplicarán por medio de un equipo de cómputo cuando el instrumento fue diseñado para una entrevista.
- 3) Desarrollar un instrumento propio. Muchos investigadores lo hacen, pero no es un trabajo sencillo que pueda realizarse en poco tiempo. Con frecuencia el desarrollo de un nuevo instrumento toma tiempo y requiere habilidades que no todos los investigadores dominan.

Sin importar que adoptemos uno ya existente, que lo adaptemos o que desarrollemos uno propio, en términos generales el instrumento que utilicemos estará compuesto por preguntas y respuestas. De alguna manera la forma en que queramos registrar la respuesta determinará el tipo de pregunta.

Las preguntas

La selección y el fraseo de las preguntas están determinados por el contexto de la investigación, por lo que es muy importante que tengamos claridad en los diferentes componentes de este contexto antes de decidir qué forma de registro de datos vamos a utilizar. El contexto de la investigación que determina el instrumento incluye los siguientes aspectos:

- Identificar el objetivo general y específicos de la investigación.
- Definir con claridad las variables en estudio.

- Recordar que las preguntas se realizan en un contexto social, cultural y económico.
- Conocer a los encuestados. Es decir, quién responde.
- Identificar a los encuestadores. Es decir, quién pregunta.
- Estandarizar las preguntas. Es decir, cómo se pregunta.
- Estandarizar el formato de las respuestas. Es decir, cómo se responde.
- Conciliar lo que se quiere preguntar con el tiempo que se dispone para preguntar.

Al momento de redactar y seleccionar las preguntas es útil tomar en cuenta las siguientes recomendaciones:

- Realizar preguntas útiles, relacionadas con la investigación que se pretende realizar.
- Hacer preguntas concretas utilizando un lenguaje convencional.
 - Utilizar oraciones completas. En español la sintaxis de la oración tiene la siguiente estructura: primero va el sujeto, luego un verbo conjugado, y termina con un complemento (Cohen, 2010).
 - Utilizar un lenguaje fácil de comprender por los entrevistados, o por quién llena el instrumento. Para lograrlo conviene: no utilizar abreviaturas, no utilizar expresiones coloquiales o en jerga, ni utilizar expresiones técnicas.
- Utilizar períodos de tiempo pertinentes a la importancia de la pregunta.
 - Preguntas y respuestas cortas cuando se quiera ahorrar tiempo, pero sin sacrificar la claridad de las mismas.
 - En caso de temas sensibles al entrevistado se pueden utilizar **preguntas cargadas**. Por ejemplo, si existe interés en preguntar sobre consumo de drogas ilegales podríamos redactar una pregunta en los siguientes términos: “Algunas personas, por experimentar, han probado la marihuana. Ud. ha probado la marihuana en alguna oportunidad de su vida”. Al plantear la pregunta de esta manera facilitamos que los entrevistados la contesten, pero es necesario tener cuidado porque la redacción puede ser tan compleja que llega a dificultar su interpretación por el entrevistado.
- Evitar frases y palabras sesgadas. Estas son las que orientan la respuesta. Por ejemplo, la pregunta “verdad que Ud. quiere a sus hijos” anticipa la respuesta que queremos escuchar.
- Evitar preguntas con dos respuestas.
- Evitar preguntas en negativo.
- Hacer que las preguntas sean revisadas por expertos y por respondientes potenciales.
 - Las preguntas son precisas y sin ambigüedades cuando dos o más personas entrevistadas potenciales están de acuerdo en las palabras usadas para contestarla.
- Recordar que se pueden adoptar o adaptar preguntas que ya han sido utilizadas satisfactoriamente por otros investigadores.

El objetivo al seleccionar, modificar o diseñar una forma de registro de datos será el de obtener una medida que sea el mínimo necesario para cumplir con los objetivos de la investigación, y que el instrumento reúna los tres requisitos siguientes: confiabilidad, validez y objetividad.

- La confiabilidad se refiere a la consistencia en los resultados. Es decir, que cada vez que se mide lo mismo se obtenga el mismo resultado. Por ejemplo, si nos pesamos en una báscula y registramos 70 kg de peso, nos bajamos de la misma y luego nos pesamos nuevamente esperamos que la segunda lectura sea igual (o muy cercana) a 70 kg. Si este es el caso hablamos de una lectura confiable. En caso contrario (que la segunda lectura sea, por ejemplo, de 80 kg) las lecturas no son confiables.
- La validez se refiere a que el instrumento realmente mida la variable que se quiere medir. Por ejemplo, si queremos medir la grasa corporal de una persona tendríamos varias opciones: podríamos pesarlo (sabemos que a mayor grasa corporal mayor peso corporal, pero también la masa muscular incrementa el peso, por lo que el peso total no sería la mejor opción), calcular el índice de masa corporal ($\text{peso}/\text{talla}^2$), medir pliegues cutáneos o medir la impedancia. El último tendría mayor validez que los anteriores.
- La objetividad se refiere al potencial en que los resultados pueden ser influenciados, o sesgados, por los investigadores o por quien registra el dato.

Las respuestas

Las respuestas se clasifican en cerradas o abiertas según sea la forma en que las registramos. Para el caso de las respuestas cerradas el que llena la forma de registro de datos selecciona una o varias opciones entre una lista limitada de respuestas posibles. En las respuestas abiertas estas opciones predefinidas no existen, por lo que quien responde deberá utilizar sus propias palabras.

Respuestas cerradas. Entre las respuestas cerradas podemos encontrar varios modelos: las más frecuentes pueden clasificarse en dicotómicas, múltiples nominales, y múltiples ordinales.

- Las respuestas dicotómicas presuponen sólo dos opciones posibles. Las más frecuentes en estas categorías son aquellas que responden a la pregunta con un SI o un NO. Por ejemplo:

¿Ha tenido usted tres o más evacuaciones líquidas en las últimas 24 horas?

| |
|----|
| No |
| Si |

- Las respuestas múltiples nominales presuponen más de dos opciones posibles, sin que exista entre ellas una relación de orden, por lo que la secuencia en que se anoten generalmente no tiene tanto impacto en la respuesta. Sin embargo no es raro que algunas opciones tengan que situarse después de otras, como es el caso de la opción “otro”. Por ejemplo:

Ayer que regresó del trabajo a su casa, ¿qué medio de transporte utilizó?

| | |
|--|--|
| Automóvil o camioneta | |
| Autobús o minibús | |
| Motocicleta | |
| Otro vehículo de motor | |
| Bicicleta | |
| Otro vehículo | |
| Ninguno de los anteriores, solo caminé | |

- Las respuestas múltiples ordinales presuponen varias opciones posibles que mantienen una relación de orden entre ellas. Por ejemplo:

Qué tan de acuerdo o en desacuerdo está usted con el siguiente enunciado:

Los accidentes son causados por el destino.

| | |
|-----------------------------|--|
| Completamente de acuerdo | |
| Casi de acuerdo | |
| No sabe | |
| Casi en desacuerdo | |
| Completamente en desacuerdo | |

Al considerar el uso de esta opción es necesario tomar en cuenta las siguientes consideraciones:

- a) Utilizara escalas que tengan sentido con los objetivos.
- b) Considerar cinco tipos de respuestas:
 - Aprobación: Definitivamente cierto, cierto, no sé, falso, definitivamente falso.
 - Frecuencia: Siempre, frecuentemente, regularmente, algunas veces, casi nunca, nunca.
 - Intensidad: Nada, muy poco, regular, moderado, severo.
 - Influencia: Gran problema, problema regular, problema pequeño, casi sin problema, sin problema.
 - Comparación: Mucho más que los otros, poco más que los otros, aproximadamente igual que los otros, algo menos que los otros, mucho menos que los otros.
- c) Balancear las respuestas.
- d) Usar escalas de 5 o 7 categorías.
- f) Poner las categorías negativas al principio de la escala.
- g) Mantener las escalas en la misma página y fáciles de completar.
- h) Usar escalas sólo cuando los encuestados pueden ver, o recordar con facilidad, todas las opciones.

Para el caso de cualquiera de los tres modelos de respuestas anteriores (dicotómicas, múltiples nominales, múltiples ordinales) siempre es necesario considerar la opción "otra". Aunque es necesario tener cuidado porque no es raro que la opción "otra" resulte la respuesta más frecuente. Por otra parte, también es necesario tomar en cuenta las respuestas "no sé", "sin datos", "no aplica" o "no quiso contestar" que

llegan a ser importantes durante el análisis aunque no sean categorías de la variable en estudio.

Respuestas abiertas. Con frecuencia se afirma que las respuestas abiertas no son adecuadas para el análisis estadístico, pero esto no es cierto para el caso de las variables cuantitativas que deben registrarse a través de respuestas abiertas, ni para algunas respuestas que cuentan con sistemas de codificación ya establecidos.

- Para el caso de las variables cuantitativas la mejor opción es la respuesta abierta. De esta manera el investigador, o quién responde a la pregunta, podrá registrar el número que corresponda a la pregunta. Por ejemplo:

¿Cuál es su edad en años cumplidos?

Antes de que tuviéramos las facilidades de cómputo que nos brindan los ordenadores, las tabulaciones tenían que hacerse manualmente y se acostumbraba agrupar las posibles respuestas numéricas. Así, para la edad las opciones de respuesta podrían haber sido las siguientes: menores de 1 año, 1 a 4 años, 5 a 9 años, 10 a 14 años. Pero desde que tenemos acceso a computadoras y paquetes estadísticos lo mejor es registrar el dato como tal para después agruparlo durante el análisis, si es que así lo queremos presentar.

- Hay un grupo de variables que ya han sido codificadas con anticipación, para las que incluso existen acuerdos internacionales. Tal es el caso para la codificación de las enfermedades que está normada por la Clasificación Internacional de Enfermedades, actualmente en su décima revisión. En este tipo de respuestas también se incluyen la descripción de los eventos en que los sujetos resultan lesionados por causa externa, las intervenciones quirúrgicas y las ocupaciones laborales. Al registrar estas variables no es raro que sea necesario contestar a dos o más preguntas, como es el caso para la ocupación. Por ejemplo:

En su trabajo usted es (solo seleccione una)

| | |
|------------------------------------|--------------------------|
| Empleado, obrero, jornalero o peón | <input type="checkbox"/> |
| Trabaja por su cuenta | <input type="checkbox"/> |
| Patrón o empresario | <input type="checkbox"/> |

¿A qué se dedica el negocio, empresa o institución donde trabaja?

¿Cuál es el oficio, puesto o cargo que tiene en el trabajo?

¿Cuáles son las funciones o tareas que hace en el trabajo?

Tomando en cuenta las respuestas anteriores, y con el auxilio del catálogo de ocupaciones que esté utilizando se podrán asignar códigos específicos a las diferentes ocupaciones laborales de los sujetos que se estudien.

- Por último, y desde una perspectiva cuantitativa, las respuestas abiertas deben evitarse cuando no existan criterios claros de codificación porque consumen

mucho tiempo durante el llenado del instrumento, su manejo posterior es muy complejo, y generalmente no son adecuadas para el tratamiento estadístico.

REVISIÓN Y CORRECCIÓN DE LOS DATOS RECOGIDOS

Después de que los datos han sido captados en los formatos diseñados para este fin, será necesario que sean revisados con cuidado con la intención de identificar si están completos y han sido llenados de manera correcta. Esta revisión deberá ser realizada, de preferencia, el mismo día en que los datos fueron recogidos. De otra manera podría ser muy difícil, o imposible, hacer las correcciones pertinentes.

Cuando los datos son registrados en un archivo de cómputo, es necesario que la revisión y corrección de datos incluya los guardados en estos medios de almacenamiento. Este tema se trata con mayor amplitud en el Capítulo 4.

REFERENCIAS

Armstrong BK, White E, Saracci R: Principles of Exposure Measurement in Epidemiology. 1ª edición. Oxford University Press: Oxford, 1992

Cohen S: Redacción sin dolor. 5ª edición. Planeta: México, 2010.

Fink A: The survey kit. Tomo 2, How to ask survey questions. 1ª edición. Sage: Thousand Oaks, California, 1995.

McDowell I, y Newell C: Measuring Health: a guide to rating scales and questionnaires. 2ª edición. Oxford University Press: New York, 1996.