

capítulo 7

Recolección de los datos cuantitativos. Segunda parte

INTRODUCCIÓN

En este capítulo se tratan algunos temas que por cuestiones de espacio no fueron incluidos en el capítulo 9 (Recolección de los datos cuantitativos) del texto impreso de *Metodología de la investigación*, 6ª edición, o bien, no se profundizó en ellos. Fundamentalmente, esos temas son los siguientes: procedimiento “paso a paso” para construir un instrumento de medición, el escalograma de Guttman (escala de actitudes), pruebas e inventarios, contextos de administración de cuestionarios, análisis de contenido, observación cuantitativa, y obtención y análisis de datos secundarios (comprendido el metaanálisis) y un comentario sobre las fuentes de error en las mediciones.

PROCEDIMIENTO PARA CONSTRUIR UN INSTRUMENTO DE MEDICIÓN “PASO A PASO”

En el capítulo 9, en la sección “¿Qué procedimiento se sigue para construir un instrumento de medición?”, se expuso en términos generales el proceso para desarrollar un instrumento o sistema de recolección de datos. Ahora revisaremos con mayor profundidad cada una de las fases.

Fase 1: Redefiniciones fundamentales

Al comenzar a desarrollar el o los instrumentos de medición, vale la pena una última reflexión para cerciorarnos de cuáles son nuestras variables y algunas otras precisiones.

Para tal efecto, sugerimos contestar las preguntas que se muestran en la tabla de la página 2.

Una vez que se responden estas preguntas, se cuenta con las redefiniciones más importantes para proseguir.

Fase 2: Revisión de la literatura enfocada en instrumentos de medición

Si al elaborar el marco teórico efectuamos una revisión adecuada de la literatura, esta fase suele ser innecesaria. Pero puede ocurrirle a quien se inicia en la investigación, que al consultar las fuentes primarias no se haya enfocado en los instrumentos de medición utilizados. En esta situación, resulta indispensable evaluar nuevamente la literatura, centrándose en tales instrumentos, lo cual resulta sencillo, porque —supuestamente— tenemos las referencias a la mano.

Cuando se valoran los diferentes instrumentos utilizados en estudios previos, conviene identificar —en cada caso— el propósito del instrumento tal como fue concebido por el investigador o investigadores, y analizar si es similar al nuestro.

¿Qué va a ser medido? (identificación y lista de variables).

Variable o concepto 1: _____.

Variable o concepto 2: _____.

Variable o concepto 3: _____.

Variable o concepto k: _____.

¿Qué o quiénes van a ser medidos? _____.

¿Cuándo (fechas precisas)? _____.

¿Dónde (lugar específico)? _____.

¿Cuál es nuestro propósito al recolectar los datos? (debe ser coherente con el planteamiento)

_____.

Nuestras definiciones operacionales son:

Variable o concepto 1: _____.

Variable o concepto 2: _____.

Variable o concepto 3: _____.

Variable o concepto k: _____.

¿Qué tipo de datos queremos obtener? (Por ejemplo: respuestas verbales, conductas observadas, contestaciones escritas, estadísticas epidemiológicas, proyecciones internas, marcas de sistema electrónico, mediciones del desempeño de una maquinaria, indicadores económicos, etcétera).

Este es el primer punto que tomamos en cuenta para ver qué instrumentos pueden ser útiles. Asimismo, se recomienda considerar los siguientes elementos:

- Las variables que mide el instrumento o herramienta: ¿cuántas y cuáles?
- Los dominios de contenido de las variables (dimensiones, componentes o indicadores): ¿cuáles y qué tan extensos son? (Sharma y Petosa, 2014, Mertens, 2010 y Punch, 2009).
- La muestra: ¿a qué o a quiénes fue administrado? ¿Qué perfil o características poseen las unidades medidas?
- Las condiciones de aplicación: ¿en qué lugar y contexto fue utilizado? ¿Cómo y cuándo fue administrado? ¿Por quién o quiénes se aplicó? ¿Eran individuos con capacidades y entrenamiento para implementarlo, calificarlo e interpretarlo? ¿Qué tiempo requiere para ser aplicado, codificado y calificado?
- La información sobre la confiabilidad: ¿de qué tipo es? ¿Cómo fue establecida? ¿Qué procedimientos se utilizaron para calcularla? ¿Cómo se reportó? ¿Los resultados son aceptables y significativos?
- La información sobre la validez: ¿qué evidencia de contenido, criterio y constructo se aportó? ¿Se estableció suficiente evidencia sobre la validez total? (de Vet, Terwee, Mokkink y Knol, 2011).
- La información sobre la objetividad: ¿se aplicó de manera estandarizada?
- Los recursos necesarios para administrarlo y su costo: ¿qué recursos se requieren para aplicarlo? (por ejemplo: informáticos).

Sobre la base de tales elementos, se seleccionan uno o varios instrumentos que pueden ser apropiados para nuestra investigación.

Fase 3: Identificación del dominio de las variables que se medirán y sus indicadores

Esta fase, prácticamente simultánea a la anterior, se refiere a la precisión del dominio de las variables que mediremos, fundamentándonos en las definiciones conceptuales de estas y la revisión de instrumentos

previos de recolección de datos. Es decir, debemos señalar puntualmente los componentes, dimensiones o factores que teóricamente integran cada variable. Incluso, se sugiere hacerlo gráficamente como en la figura 7.1., con la “satisfacción laboral”).

● **Figura 7.1** Ejemplo de establecimiento de factores con la variable satisfacción laboral.



Una vez ubicados tales factores es necesario establecer los indicadores de cada uno de ellos. Por ejemplo, en el caso de la “satisfacción en torno a las condiciones de trabajo”, estos podrían ser: satisfacción sobre la jornada laboral, horarios de entrada, salida y alimentación, lugar, transporte, iluminación, servicio médico y vacaciones. Cuando se trata de la “satisfacción sobre el desempeño personal en el trabajo”: satisfacción respecto a la productividad, eficiencia, eficacia, calidad del servicio, desperdicio y calidad factual.

Para cada indicador se generan uno o más ítems tal como se comentó en el capítulo 9 del texto impreso, lo que, recordemos, se denomina *operacionalización* de las variables (la secuencia: variable → factores, dimensiones o componentes → indicadores → ítems o valores). La operacionalización es la base de las definiciones operacionales de las variables.

Una vez que se ha operacionalizado la variable (se cuenta con ítems o valores concretos), se especifica qué instrumento puede medirla y se construye su definición operacional, la cual implica especificar y describir el procedimiento para medir la variable a través del instrumento.

Para reforzar lo anterior paso por paso, Solís (2013), Hernández-Sampieri *et al.* (2013) y Sjoberg y Nett (1980) consideran que las siguientes etapas conforman la medición de una variable:

1. Definición nominal de la variable: identificarla y nombrarla.
2. Definición conceptual de la variable.
3. Dimensionalización de la variable: determinar sus factores, dimensiones o componentes.
4. Definición de indicadores empíricos e índices de cada factor o dimensión de la variable (que es lo que en este apartado denominamos identificación del dominio de la variable que se debe medir y sus indicadores).
5. Identificación, desarrollo o generación de ítems o valores de cada indicador o índice.¹

¹ Recordemos que a veces se sigue la secuencia completa: variable → dimensiones → indicadores → ítems o valores, mientras que en otras ocasiones se transita de la variable directamente al <<<<ítem, escala o valor. Por ejemplo, temperatura → lectura de un valor en una escala de un aparato (termómetro); o bien, presión arterial (aplicación de un procedimiento y lecturas de escalas en un esfigmomanómetro: mmHg).

Operacionalización Paso de una variable teórica o abstracta a indicadores e ítems o valores empíricos verificables.



6. Definición instrumental de la variable: especificar los instrumentos que se utilizarán para recopilar los datos pertinentes de los indicadores empíricos y de los ítems o valores requeridos para cada indicador.
7. Definición operacional de la variable: determinar el procedimiento detallado que se seguirá para aplicar el instrumento de recolección de datos y medir la variable.

Para ilustrar la operacionalización, tomemos el ejemplo de Lee y Guerin (2009), el cual se trata en la obra impresa. Recordemos que esta investigación tuvo como objetivo “identificar si la satisfacción respecto de la calidad del diseño ambiental del interior de áreas de trabajo u oficinas afectaba significativamente la satisfacción general del espacio de trabajo por parte de sus ocupantes y su desempeño laboral”.

La satisfacción respecto de la calidad del diseño ambiental del interior del área de trabajo fue medida aplicando a los participantes la encuesta IEQ (*Indoor environmental quality*) desarrollada por el Centro para el Medio Ambiente de la Universidad de California en Berkeley.

Se establecieron las dimensiones de tal satisfacción y los indicadores de ellas, que fueron operacionalizados utilizando ítems escalados bajo el sistema de Likert con siete puntos que van de “muy satisfecho” (+3) a “muy insatisfecho” (−3); lo mismo se aplicó a la satisfacción general sobre el espacio de trabajo. Las dimensiones consideradas y sus indicadores se muestran en la tabla 7.1.

● **Tabla 7.1** Dimensiones e indicadores de la variable "satisfacción de la calidad del diseño ambiental del interior del área de trabajo".

Dimensión o criterio	Indicadores medidos (satisfacción sobre...)
Diseño de la oficina	<ul style="list-style-type: none"> • Espacio aprovechable para el trabajo individual, guardar/almacenar • Nivel de privacidad visual • Facilidad para interactuar con compañeros de trabajo
Mobiliario	<ul style="list-style-type: none"> • Confort • Capacidad/adaptabilidad • Colores y texturas del piso, muebles y acabados
Temperatura	<ul style="list-style-type: none"> • Temperatura ambiental
Ventilación	<ul style="list-style-type: none"> • Calidad del aire en el espacio personal (respirable, ambiente ventilado, rancio/no rancio, olores agradables...)
Iluminación	<ul style="list-style-type: none"> • Cantidad de luz • Confort visual (deslumbrante/oscuras, reflejos, contraste, etcétera)
Acústica	<ul style="list-style-type: none"> • Ruido. • Privacidad de sonidos (aislamiento)
Limpieza y mantenimiento	<ul style="list-style-type: none"> • Limpieza del espacio • Mantenimiento del edificio • Servicio de limpieza y mantenimiento del espacio de trabajo

El desempeño, como ya se mencionó en el texto impreso, se midió perceptualmente preguntándoles a los participantes el grado en que cada criterio facilita/interfiere con la capacidad de terminar el trabajo. Desde luego, hubiera sido ideal agregar datos directos del desempeño que son estándares de toda clase de ocupaciones y organizaciones (porque las empresas hacen distintas y muy variadas valoraciones del desempeño laboral y también dependen del puesto) tales como asistencia-ausentismo (número de días laborados), productividad e índice de asuntos resueltos en el último mes (número de asuntos encargados/número de asuntos resueltos), etcétera.

En ocasiones, las dimensiones o factores son pruebas (caso frecuente en ciencias físicas o de la salud). Por ejemplo, la medición fisiológica del sueño según lo determinado por la polisomnografía, (del griego *poly*, “muchos”; *somno*, “sueño”; y *graphy*, “escritura”) que no es otra cosa que una prueba de múltiples parámetros usada en el estudio del sueño.

El estudio polisomnográfico incluye tres estudios básicos: electroencefalografía (EEG), electrooculografía (EOG) y electromiografía (EMG). Los otros parámetros analizados son: electrocardiografía, oximetría de pulso, esfuerzo respiratorio, CO₂ trascutáneo, registro de sonidos para evaluar ronquidos, EMG de extremidades y monitorización continua por video (Hall *et al.*, 2008).

Para reforzar este punto se recomienda que en el curso de la investigación se discutan en clase ejemplos de operacionalización de variables de la ciencia o disciplina en la cual se enmarca la asignatura (por ejemplo, en ingenierías, ver cómo se operacionalizan variables como presión, aceleración, volumen, masa, etc.; en medicina, presión arterial, crecimiento de un tumor cancerígeno, obesidad, etc.; en administración, productividad de los trabajadores —dependiendo del tipo de organización—, rentabilidad de la empresa, etc.; en psicología, autoestima, depresión, personalidad, etcétera).

Fase 4: Decisiones claves

Una vez evaluados varios instrumentos (al menos un par en el caso de estudiantes de licenciatura o pregrado) que se utilizaron previamente, es necesario tomar una serie de decisiones fundamentales, que agrupamos en tres rubros:

1. Utilizar un instrumento de medición ya elaborado, adaptarlo o desarrollar uno nuevo

En el primer caso, debemos elegir un instrumento que obviamente se encuentre disponible (“esté a la mano” y podamos utilizarlo sin violar ningún derecho de autor) y sea posible su aplicación (al respecto, debemos cuestionarnos si en realidad podemos administrarlo). Asimismo, debemos seleccionar que: *a*) se haya generado lo más moderno posible; *b*) se cite con mayor frecuencia y amplitud en la literatura; *c*) sea bien evaluado por expertos de asociaciones académicas y profesionales; *d*) demuestre mayor confiabilidad, validez y objetividad; *e*) se pueda aplicar más a nuestro planteamiento y propósito de recolección de datos; *f*) se adecue mejor a nuestra muestra y contexto, y *g*) se adapte en mayor medida a las aptitudes que tenemos (que estemos más calificados para utilizarlo). Solo deben seleccionarse instrumentos que reporten fiabilidad, validez y objetividad ya que no es posible confiar en una forma de medir que carezca de este tipo de evidencia clara y precisa. Cabe señalar que algunos instrumentos se han validado para contextos iberoamericanos. Por ejemplo, la Escala Clima-UNI con 96 ítems para medir el clima organizacional, puede ser aplicada tal cual al medio laboral mexicano (Hernández-Sampieri, Méndez y Contreras, 2013 y Hernández-Sampieri, 2005). Su confiabilidad promedio es de 0.95 y posee evidencia de validez de todo tipo. Desde luego, si se quisiera administrar en otra nación latinoamericana, debe volver a validarse y, en el caso de España, tendría que adecuarse. Si se trata de una herramienta escrita (como un cuestionario), es ineludible revisar la redacción.

En la segunda situación, el instrumento elegido se adapta al ambiente de nuestro estudio; es decir, se realizan algunos ajustes a su presentación, instrucciones, ítems (preguntas, categorías de observación, etc.). También se revisan cuidadosamente el lenguaje y la redacción. Este es el caso de instrumentos que se traducen y validan. Por ejemplo, el instrumento Children’s Knowledge of Abuse Questionnaire-Revised (CKAQ-R), versión en español,² fue adaptado de su homólogo, el cual fue generado en Estados Unidos y se tradujo como Cuestionario del Conocimiento del Abuso Infantil-Revisado.

Si ninguna de las dos opciones anteriores resulta satisfactoria, tenemos que desarrollar uno propio. Este puede elaborarse a partir de varios instrumentos previos, de nuevas concepciones e ítems, o bien, una mezcla de estos. Para tal efecto, idealmente debe generarse un conjunto de ítems y/o categorías para los indicadores de cada dimensión, con el apoyo de un grupo de expertos (en una o varias sesiones del tipo denominado “tormenta de ideas”). Posteriormente, los ítems se deben someter a evaluación de otros expertos y redactar la introducción e instrucciones. Por ejemplo, Hernández-Sampieri (2005), construyó su instrumento mediante la selección de ítems de varios cuestionarios y escalas desarrolladas en Estados Unidos, y otras más que habían sido aplicadas en México, además de incluir nuevos reactivos.

² Este instrumento es parte del ejemplo “El abuso sexual infantil” tratado en la sexta edición impresa de Metodología de la investigación (capítulos cuantitativos), cuyo reporte del estudio original se presenta en forma de artículo académico en los ejemplos de este centro de recursos en línea o página web.

Una vez más, recordemos que el tránsito de la variable a los ítems es progresivo, como se muestra en el siguiente ejemplo:

Ejemplo

Variable	Dimensión	Indicadores	Ítems
Intención del voto para las próximas elecciones para alcalde	Intención de voto por partido	Intención de voto favorable	En las próximas elecciones para alcalde, ¿por qué partido piensa votar? (con sus opciones de respuesta)
		Intención de voto desfavorable	En las próximas elecciones para alcalde, ¿por qué partido nunca votaría? (con sus opciones de respuesta)
	Intención de voto por candidato	Intención de voto favorable	En las próximas elecciones para alcalde, ¿por qué candidato piensa votar? (con sus opciones de respuesta)
		Intención de voto desfavorable	En las próximas elecciones para alcalde, ¿por qué candidato nunca votaría? (con sus opciones de respuesta)

2. Resolver el tipo de instrumento, formato y/o procedimiento

La segunda decisión es simultánea a la primera cuando se utiliza un instrumento ya elaborado o adaptado, pues este ya posee sus características y tiene un formato y procedimiento propios. Sin embargo, cuando se desarrolla uno nuevo, el investigador debe decidir de qué tipo será (cuestionario, aparato, procedimiento, hoja de observación, escala de actitudes, prueba estandarizada, etc.) y cuál será su formato y procedimiento de aplicación, lo cual depende de la clase y características físicas (tamaño, colores, tipo de fuente, especificaciones de papel y aspectos similares). Por ejemplo, no es lo mismo un cuestionario escrito en papel que uno incluido en una página web, e incluso que un cuestionario en tableta. O bien, para medir la masa corporal hay distintos escáneres.

3. Determinar el contexto de administración

Este paso implica decidir en qué espacio físico se implementará (si en un aula de capacitación, en un salón de clases, en la calle, en los hogares, de manera telefónica, en una cámara Gesell, en un laboratorio, en el consultorio, en una sala ambulatoria, en un acelerador de partículas, etc.). Desde luego, la naturaleza de los datos que se buscan y el tipo de instrumento elegido nos ayudan en esta tarea. Por ejemplo, ya sabemos que si deseamos datos escritos y vamos a utilizar un cuestionario, este puede administrarse en distintos contextos, mientras que si nuestro interés es observar, lo podemos hacer en un lugar público, en un salón o en una cámara Gesell. Cuando se comentó sobre las formas de aplicar cuestionarios se profundizó en el tema, pero en el caso de pruebas estandarizadas estas suelen administrarse en lugares que permitan a los participantes concentrarse, como aulas universitarias, salones de hoteles y despachos privados. En cambio, la observación depende de los hechos o conductas que se pretende registrar.

Fase 5: Construcción del instrumento

En esta etapa se generan todos los ítems o reactivos (o escalas de valores) y categorías del instrumento, de acuerdo con la naturaleza de este y su formato. Recordemos que un ítem es, por ejemplo, una pregunta

de un cuestionario, una categoría de un sistema de observación, un valor en la escala de un dispositivo o una estadística en un sistema de registro. Al final de esta etapa el producto es una primera versión del instrumento. Una parte muy importante de ella es cómo vamos a registrar los datos y el valor que les otorgaremos (codificación).

Fase 6: Prueba piloto

Esta prueba consiste en administrar el instrumento a personas, objetos o fenómenos con características semejantes a las de la muestra objetivo de la investigación (unidades). Se somete a prueba no solo el instrumento de medición, sino también las condiciones y contexto de aplicación y los procedimientos involucrados. Se debe verificar que las instrucciones se comprenden y que los ítems (lecturas o equivalentes) funcionen de manera adecuada. En el caso de cuestionarios, se debe evaluar el lenguaje y la redacción. Los resultados se utilizan para calcular la confiabilidad inicial y, de ser posible, la validez tentativa del instrumento de medición. La prueba piloto se realiza con una pequeña muestra (inferior a la muestra definitiva).

Recordemos que el tamaño de la muestra para la prueba o estudio piloto se puede calcular utilizando *STATS**, donde el tamaño de la población o universo es el tamaño de la muestra calculado y el nivel de confianza puede ser de 95 o 99%. Es decir, y valga la redundancia, se extrae una muestra de la propia muestra de la investigación.

Por ejemplo, Hernández-Sampieri (2005) aplicó el cuestionario del clima organizacional a una prueba piloto en dos muestras de 30 y 32 participantes, respectivamente. La confiabilidad inicial de estas pruebas fue de 0.91 y 0.92.

En ocasiones, el instrumento se somete a varias pruebas piloto y se depura paulatinamente (en el caso de dispositivos, se calibran). Núñez (2001) generó una primera versión de su escala del sentido de vida (conformada por 135 reactivos), la aplicó a un grupo y realizó ajustes sobre los resultados; posteriormente, una segunda versión (180 ítems) fue administrada a otro grupo piloto similar al primero; se volvieron a efectuar cambios en el instrumento, hasta que llegó a una versión final, la cual contenía 99 reactivos.

Cabe señalar que cuando se desarrolla un nuevo instrumento, es conveniente incluir un elevado número de ítems, para elegir a los que más contribuyan a su confiabilidad, validez y objetividad. Desde luego, sin excedernos ni caer en redundancias excesivas, es decir que el número o cantidad de reactivos debe ser manejable. Además, la inclusión de cada ítem debe estar justificada (Neuman, 2009). Por ello, se recomienda que los reactivos se generen mediante una o varias sesiones con expertos (DeVellis, 2012 y Gall, Gall y Borg, 2003), los cuales pueden ser profesores universitarios, consultores de empresas y especialistas en el tema de estudio. Cuando no encontramos expertos, nuestra revisión de la literatura tiene que ser exhaustiva y profunda.

Como ya se comentó, la prueba piloto evalúa todo el procedimiento de medición. Así, por ejemplo, en el experimento de Naves y Poplawsky (1984), se evaluó la conducta de aceptación-avoidancia de los participantes hacia el supuesto individuo con capacidad mental distinta (que recordemos era un actor) mediante observación. Cada interacción se filmó y se analizó. Durante la prueba piloto se descubrió que la iluminación era insuficiente y no se podían registrar ciertas conductas no verbales; ante ello, la prueba se perfeccionó.

En el caso de instrumentos para aplicar mediciones en personas, parte fundamental de la prueba piloto consiste en conversar con los participantes para recoger sus opiniones con respecto al instrumento y al contexto de aplicación (por ejemplo, si lo consideran largo o corto, comprensible, no intrusivo, etc.). Una vez, al comenzar una prueba piloto de un instrumento para evaluar al superior inmediato, uno de los participantes inquirió: “En el cuestionario se pregunta sobre nuestro jefe, pero, ¿a cuál de ellos se refieren? Porque en estos momentos en el área de producción tenemos dos superiores, el de siempre, nuestro supervisor regular; pero como estamos fabricando un nuevo modelo de automóvil, también tenemos un superior de este proyecto”. Sin la prueba piloto, al levantar los datos no se hubiera sabido a qué superior se referían, y la confusión hubiera sido inevitable.



A continuación incluimos algunas recomendaciones para tal prueba en el caso de instrumentos escritos (cuestionario, examen o prueba estandarizada).³

- Es aconsejable que se otorguen espacios para comentarios de los participantes.
- Solicitar a los participantes que señalen ambigüedades, opciones o categorías no incluidas, términos complejos y redacción confusa.
- Tomar en cuenta y evaluar lo que escribieron los participantes durante todo el proceso.
- Estar atento a que no se presenten “patrones tendenciosos” por efecto del instrumento en: a) el orden de las preguntas, b) el orden de las opciones de respuesta, c) tendencias en las respuestas (por ejemplo: que siempre tiendan a estar de acuerdo o en contra de todo, sin que realmente sea su opinión; o que respondan “no sé” a una buena parte de las preguntas; que dejen de contestar a varios ítems, etcétera).
- Simular las condiciones reales de administración de la mejor manera posible (por ejemplo, si el cuestionario va a enviarse por correo, la prueba piloto debe hacerse por este medio: se les envía y, posteriormente, pero de manera inmediata, se les reúne en grupo para comentar o, al menos, se simulan las condiciones en el grupo: a los participantes se les entrega el cuestionario en un sobre postal, el cual abren y leen, etcétera).

Fase 7: Versión final

Así, con los resultados de la prueba piloto, el instrumento de medición preliminar se modifica, ajusta y mejora, dependiendo de su tipo: se calibra el dispositivo, se reconsideran escalas o categorías, se eliminan o agregan ítems, se cambian palabras, se otorga más tiempo para responder, etc. Finalmente, se tiene la versión definitiva administrar (si se trata de cuestionarios, se incluye un diseño gráfico atractivo para los participantes y de acuerdo con el medio: impreso en papel, página web, tableta o PC, etcétera).

Fase 8: Entrenamiento del personal

Las personas del equipo de investigación que van a administrar y/o calificar el instrumento, ya sea que se trate de operadores, entrevistadores, observadores, encuestadores o individuos que dirijan las sesiones para aplicarlo, deben recibir un entrenamiento en varias cuestiones:

- a) La importancia de su participación y su papel en el estudio.
- b) El propósito del estudio.
- c) Las personas e instituciones que patrocinan y/o realizan la investigación.
- d) La calendarización y tiempos del estudio (fechas claves de entrega de información).
- e) La confidencialidad que deben guardar respecto de la investigación y sobre todo de los participantes. Esta no puede violarse bajo ninguna circunstancia.
- f) La manera en que deben vestirse si van a entrevistar o tener contacto con los participantes.
- g) El respeto y la sensibilidad que deben tener hacia los participantes (por ejemplo, no discriminar ni utilizar lenguaje que pudiera resultar ofensivo).
- h) No discutir ni molestarse con los participantes (ya que nos están haciendo un favor).
- i) Cómo establecer un ambiente de confianza.
- j) Usar gafetes y cartas de identificación.
- k) El instrumento de medición (estructura, orden, ítems, instrucciones para el personal y para los participantes) y sus condiciones de administración.

Durante el entrenamiento, el personal que aplicará y/o calificará las respuestas o mediciones debe realizar varios ensayos a la vista de todos (de entrevista, observación, aplicación de prueba, implementación de un dispositivo o procedimiento, etc.), porque ello la permitirá corregir aquellas situaciones que puedan afectar la objetividad y el correcto desempeño en la recolección de los datos.

El entrenamiento puede llevarse a cabo mediante una o más sesiones, según sea el caso. Se recomienda —al menos— una reunión previa al inicio de la recolección de datos y una después del levantamiento

³ Algunas de estas sugerencias fueron adaptadas de Mertens (2010) y Koretz (2009).

de estos. Durante la sesión o sesiones se motiva al personal, se escuchan sus sugerencias y se responden inquietudes.

Fase 9: Autorizaciones

Antes de recolectar los datos, es necesario gestionar varias autorizaciones, las cuales dependen del contexto de la investigación. Por ejemplo, en estudios con materiales delicados (sustancias radioactivas o peligrosas), de parte de organismos públicos que supervisan su manejo y utilización. Si se trata de estudios en empresas, se necesita el permiso por parte de los representantes de las organizaciones a las cuales pertenecen los participantes (directivos y líderes sindicales, directores y maestros en las escuelas —si son menores de edad, de sus tutores,—, funcionarios gubernamentales, etcétera).

Para este fin conviene producir una breve presentación del estudio en un programa para tal efecto (Power Point, Flash, Prezi, etc.) e idealmente en algún formato multimedia o de video (que pueda reproducirse utilizando un programa para PC, DVD u otros dispositivos) y presentar el protocolo o proyecto escrito que de cualquier manera debemos elaborar.

En segundo término, es ineludible obtener el consentimiento de los propios participantes, en particular por escrito. Desde luego, en encuestas en la vía pública o en los hogares, a veces esto resulta complicado y la autorización suele ser verbal o implícita. El asunto es más delicado cuando se trata de menores de edad y personas con capacidades diferentes. En este caso, el consentimiento tienen que otorgarlo los tutores y las personas que se integrarán a la muestra. No obtenerlo o lograrlo mediante engaños es una clara *violación a la ética de la investigación*. No puede filmarse a alguien sin su autorización, salvo conductas públicas manifiestas (consultar capítulo 2 de este centro de recursos en línea).

Fase 10: Administración del instrumento

Este es el momento de la verdad, en el cual todo nuestro trabajo conceptual y de planeación se confronta con los hechos. La administración depende del tipo de instrumento que hayamos seleccionado y la experticia de todo el equipo que participa. Debemos aclarar que una cuestión es el tipo de instrumento en sí y otra el contexto de administración. En la literatura sobre recolección de datos cuantitativos todavía hay cierta confusión al respecto. Por ejemplo, ciertos autores desarrollan el tema de los “cuestionarios” en las secciones de *surveys* o encuestas, otros en el apartado sobre entrevistas, algunos más los consideran en sí mismos como un tema o los incluyen en medición. Debido a estas diferencias, hemos decidido ver en el capítulo 9 del libro y en este capítulo del centro de recursos en línea cada instrumento y los contextos más comunes en que se usan.

CONTEXTOS DE ADMINISTRACIÓN DE LOS CUESTIONARIOS

En el capítulo 9 del libro (dentro del apartado “¿En qué contextos puede administrarse o aplicarse un cuestionario?”), se comentaron los principales contextos de administración de los cuestionarios. Ahora, en esta sección reforzaremos algunos aspectos sobre el tema.

Autoadministración grupal

Cuando el cuestionario se aplica grupalmente, es importante verificar algunas cuestiones, que se incluyen en la tabla 7.2.

Autoadministración por envío

En el caso de autoadministración por envío mediante el correo o mensajería, el cuestionario se remite junto con una carta explicativa firmada por el investigador o investigadores, la cual cumple las funciones de la introducción (con los elementos comentados previamente en el capítulo 9 del libro: propósito del estudio, motivadores, agradecimiento, tiempo de respuesta, etc., excepto las instrucciones que suelen incluirse en el instrumento). Si la carta está membretada con el logotipo del instrumento, mejor aún.

● **Tabla 7.2** Listado de puntos a verificar al administrar cuestionarios en grupo.⁴

1. ¿Tenemos suficientes cuestionarios?	Sí _____	No _____
2. ¿Hemos diseñado alguna medida para que quienes no puedan asistir a la sesión respondan al cuestionario?	Sí _____ ¿Cuál? _____	No _____
3. ¿Se notificó formalmente a los participantes potenciales de la fecha, hora y lugar en que se aplicaría el cuestionario?	Sí _____ ¿Cómo? (carta, correo electrónico, memo, mensaje) _____	No _____
4. ¿Se verificó que el lugar donde se aplicará el cuestionario presenta las condiciones adecuadas de espacio e iluminación?	Sí _____ ¿Quién verificó? _____	No _____
5. ¿Se aisló el lugar de fuentes potenciales de ruido u otras distracciones?	Sí _____	No _____
6. ¿Quiénes van a leer las instrucciones y asistir a los participantes a lo largo de la sesión?	Persona(s): _____	No hay designados _____
7. ¿Las instrucciones incluyen cómo responder al cuestionario?	Sí _____	No _____
8. ¿Se contempló un tiempo razonable para responder dudas e inquietudes de los participantes antes de que comiencen a contestar el cuestionario?	Sí _____	No _____
9. Quiénes van a leer en voz alta las instrucciones, ¿tienen una voz nítida y suficientemente fuerte para que todos le escuchen y su lectura será pausada?	Sí _____	No _____
10. ¿Se verificará que todos hayan respondido al cuestionario?	Sí _____	No _____
11. ¿Quién dará las gracias a los participantes por su cooperación?	Persona(s): _____	No hay designados _____
12. ¿Quién entregará o enviará las cartas de agradecimiento o equivalentes a los participantes y a quienes facilitaron la sesión? ¿A través de qué medios?	Persona(s): _____ Medios: _____	No hay designados _____

⁴ Adaptado de McMurtry (2005).

Se recomienda que los cuestionarios sean breves y concisos de acuerdo al planteamiento y los objetivos de medición. Si al hablar de otros instrumentos autoadministrados se comentó que las instrucciones deben ser precisas y claras, esta condición resulta particularmente importante en estos casos, ya que las posibilidades de realimentación y resolución de dudas se reducen al mínimo. La carátula, además de lo que se señaló previamente, debe contener la fecha exacta de envío. En las instrucciones es necesario agregar la fecha en que se requiere sea devuelto y la forma de regresar el cuestionario contestado, paso a paso. De ser posible, resulta aconsejable designar a una persona para que atienda dudas y comentarios del instrumento y el estudio, por medio telefónico, correo electrónico y/o una red social (whatsApp, twitter o facebook, por ejemplo). Obviamente, esta persona tiene que proporcionar sus datos completos. Ofrecer a los participantes un resumen de los resultados, una vez que concluya la investigación, es una práctica recomendable (este informe se puede enviar por correo electrónico, mostrar en una página web o equivalente).

Asimismo, el paquete para cada individuo potencial incluye dos sobres *personalizados*: uno que contiene el cuestionario y la carta, y el otro para que devuelva el cuestionario cumplimentado. Desde luego, este último con los datos completos del remitente (destinatario final) y con el porte de regreso o la guía de paquetería prepagada (es necesario cubrir todos los gastos generados en este proceso). Un diseño original de los sobres puede ser de gran ayuda, al menos para que sean abiertos.

Es fundamental contactar vía telefónica y/o correo electrónico al futuro encuestado, para motivarlo a que conteste el cuestionario. Una vez que se reciba su respuesta, es preciso agradecerle su cooperación. Algunas personas se niegan a participar en investigaciones, porque fueron tratadas con descortesía una vez que se obtuvo de ellas lo que se deseaba.

Los cuestionarios autoadministrados pueden ser procesados de forma casi inmediata si se usa codificación por lectura óptica u otros medios electrónicos. Es decir, si el papel del cuestionario cubre ciertos requisitos y es respondido con un lápiz o bolígrafo especial o en computadora, celular o tableta, se evita la codificación, puesto que el sistema lee las respuestas y automáticamente las envía a la base de datos correspondiente.

Vinuesa (2005) señala que la encuesta por correo postal permite una selección muestral de los participantes de acuerdo con su perfil sociodemográfico, de compra, estilo de vida, etc., y de individuos concretos (profesionales, miembros de alguna asociación, etcétera).

Algunas desventajas residen en que nunca podremos estar seguros de quién respondió el cuestionario y la ausencia de un encuestador impide asegurar la franqueza de las respuestas.

Por correo electrónico. Se trata de un procedimiento similar, pues lo único que cambia es el medio de contacto y el cuestionario puede ser contestado en el dispositivo de quien responde (PC, laptop, tableta, celular, etc.). La carta, carátula, instrucciones y el cuestionario son enviados a través de un correo electrónico (*e-mail*).

Por medio de una página de internet o servidor de la web. Esta vía es similar a las dos anteriores en cuanto a la mecánica. Sin embargo, en este caso se le pide al participante (por contacto telefónico, correo electrónico, mensaje en celular u otro medio; o bien, mediante una red social en internet) que acceda a un sitio web o hipervínculo (locación URL única),⁵ donde se localiza el cuestionario, el cual se puede contestar en el momento o por etapas; otra modalidad puede ser que se “descargue” o “baje” el cuestionario para guardarlo como archivo en la computadora y posteriormente, una vez contestado, se envía por correo u otro mecanismo electrónico.⁶

Los cuestionarios utilizados en medios electrónicos regularmente se elaboran en un programa de texto e imagen, o se escanean (si están impresos con anterioridad) y “se anexan” al correo electrónico (como un “archivo adjunto”), o se pueden colocar o “subir” al sitio web, aunque para este segundo caso lo más común es que se elaboren especialmente para tal medio. En ambas situaciones, las posibilidades de diseño del instrumento son amplísimas.

⁵ El acceso puede ser público o limitado mediante contraseñas (ID) o números PIN.

⁶ Entre los principales programas para el uso de encuestas electrónicas podemos mencionar a: WebCT, GoogleDocs y Blackboard.

Las limitaciones de los estudios que utilizan el correo electrónico y la web residen en que no todas las personas poseen computadora e internet (sobre todo en América Latina) y algunos individuos (por ejemplo, los mayores de 60 años) se resisten a utilizar estos recursos, porque es una tecnología relativamente reciente y desconocida para gran parte de ellos. Varios estudios han encontrado que la mayoría de los usuarios de internet son personas educadas, jóvenes y de estratos medios o altos (y en Estados Unidos,

● **Tabla 7.3** Ventajas y desventajas de administrar cuestionarios mediante internet.⁷

Ventajas	Desventajas
Costo relativamente bajo.	Dificultad de acceso a directorios.
Codificación automática.	No se sabe quién contesta.
Normalmente se aplican rápidamente (pero no es una regla).	Generalmente la muestra es no probabilística. No hay una elevada representatividad o poder de generalización.
Instrumentos con diseños sumamente atractivos e interactivos (por ejemplo, posibilidad de incorporar gráficas multimedia, sonido y videos).	Riesgo de que quien responde se concentre más en la forma que en el fondo.
Elevado control del proceso de envío y recepción de cuestionarios.	

predominantemente blancos) (Nathan, 2008).

Algunas ventajas y desventajas de aplicar cuestionarios vía internet se mencionan en la tabla 7. 3.

La administración de cuestionarios por internet es útil en estudios exploratorios y descriptivos (Nathan, 2008).

En cuanto a las tasas de respuesta por internet y correo, la situación es variable y las investigaciones a este respecto arrojan distintos resultados. Fricker y Schonlau (2002), en una primera revisión de estudios que recolectaron datos a través de la web (muestras probabilísticas o censos), reportan índices de entre 38 y 44%, que son las más altas las que se conocen entre empleados, y un estudio entre estudiantes con una tasa de 8%. En otra revisión de la literatura, consideraron investigaciones en las cuales preguntaban a los encuestados si preferían contestar vía internet o mediante el correo, oportunidad en la que descubrieron preferencias mucho mayores por el segundo medio (seis investigaciones con porcentajes que oscilaron entre 73 y 96%, mientras que por internet el porcentaje varió entre 4 y 27%). Solo se conoce un estudio donde la elección por internet fue de 80% y por correo tradicional 20%, y una entre dentistas con tasas de 84% y 16%, respectivamente, lo que podría indicar que personas con alto nivel educativo y tecnológico prefieren la web para responder a cuestionarios. En una tercera revisión de trabajos donde se compararon el correo electrónico y el correo tradicional, en ocho estudios se obtuvo una tasa de respuesta mayor o mucho mayor vía este último (las muestras incluyeron profesores y empleados federales). Solo en el caso de colaboradores de la empresa telefónica AT&T la tasa fue de 38% por correo y 68% por e-mail). En otros tres estudios, en dos se prefirió el primer modo. En definitiva, todavía no hay claridad sobre cuál tipo de contexto de administración es mejor en términos de calidad de la información obtenida y las tasas de respuesta, lo que significa que se requiere mayor investigación al respecto (Fowler, 2014 y Carley-Baxter, Peytchev y Black, 2010).

Desde luego, como menciona Mertens (2010), una tasa de devolución de cuestionarios cumplimentados por correo o de manera electrónica por encima de 50% es muy favorable.

Asimismo, el tópico o tema está relacionado con la tasa de respuesta y puede interactuar con el medio elegido (Carley-Baxter, Peytchev y Black, 2010). En ello también influye el contexto histórico, social y económico del momento en el que se lleva a cabo el levantamiento de datos. Por ejemplo, si el entorno está caracterizado por violencia y delincuencia, preguntar sobre temas de seguridad pública y criminalidad puede ser delicado debido a que las personas pueden atemorizarse al responder o pensar que hay “trasfondo” en la recolección de respuestas. Preguntar sobre asuntos de salud en el momento de una epidemia,

⁷ Tourangeau, Conrad y Couper (2013); Sue y Ritter (2012); Nathan (2008) y Couper (2003).

también resulta inconveniente. Otro caso de estos se presenta cuando se inquiera sobre niveles socioeconómicos o temas relacionados cuando se ha concretado una reforma fiscal o acciones de fiscalización.

Las encuestas de casa en casa son problemáticas si hay robos en la ciudad o comunidad, pues las personas no permiten el acceso a sus domicilios y ni siquiera desean contestar en las puertas de sus hogares. La desconfianza es mayúscula.

Las tasas de respuesta por cualquier medio se incrementan cuando: 1) quienes responden al cuestionario reciben beneficios económicos, 2) los instrumentos no son extensos, 3) se produce un seguimiento sistemático de las respuestas y persistente a casos de no respuesta, 4) se establece una vinculación personal con los participantes y 5) hay contacto antes del envío (Häder, 2012; McCarty, House, Harman y Richards, 2006 y Cook, Heath y Thompson, 2001 y 2000).

De acuerdo con varios experimentos y otros tipos de estudios, en encuestas donde el cuestionario es contestado en dispositivos pequeños se producen efectos negativos en las respuestas (Peytchev y Hill, 2010). Si algunas opciones no aparecen en la pantalla, algunos participantes señalan no haber visto todas las categorías o ignorar aquellas que necesitan de recorrer la pantalla. Lo mismo en las preguntas, varios sujetos contestan con base en la parte de la interrogación que se percibe primero. En ocasiones, si los teclados son pequeños, los encuestados no proporcionan toda la información requerida, incluso evitan contestar ciertas preguntas. En ocasiones, con pantallas y teclados miniatura surgen sesgos al contestar (por ejemplo, marcar regularmente la primera opción). Este mediocre panorama puede mejorar en el futuro cuando las personas de todas las edades se familiaricen con estas tecnologías. Es necesario efectuar más investigación sistemática en relación con el tema.

Otra ventaja de la autoadministración por envío (correo tradicional, correo electrónico y página de internet) es que cuando se hacen preguntas personales o de mayor carga emotiva, el sujeto puede contestar de manera más relajada y sincera, puesto que no está frente a otra persona.

Entrevistas interactivas y telefónicas

Una posibilidad novedosa son las *entrevistas interactivas* (algunas representan modalidades telefónicas, otras se trata de los denominados “medios inteligentes” de correo electrónico o de sitios web), en donde un sistema se contacta vía telefónica o por correo electrónico con los participantes potenciales y administra el cuestionario o lo envía. Son mecanismos con reconocimiento de voz, lectura óptica y dictado digital. El problema es —hasta el momento— que la mayoría de las personas se percata de que no es otro ser humano con quien entra en contacto y suele negarse a responder. Además, la saturación de correos electrónicos, llamadas telefónicas y sitios web hacen difícil captar la atención de los participantes potenciales. Si se utilizan estos sistemas, se aconseja que los cuestionarios sean muy breves, no más de diez preguntas. Desde luego esta situación irá modificándose y cada vez serán más los estudios que utilicen tales tecnologías.

Para administrar cuestionarios vía *entrevista telefónica* se dispone de varias tecnologías, además de las ya comentadas de reconocimiento de voz y dictado digital, como el CATI (*Computer-Assisted Telephone Interviewing*), la cual permite que el entrevistador se siente frente a su computadora u ordenador, cuyo sistema selecciona números telefónicos generados al azar y los marca automáticamente (Outwater, 2008 y Boslaugh, 2007). Una vez que contesta la persona indicada, él comienza a leer las preguntas en el monitor y anota las respuestas (desde luego, mediante el teclado o el mouse), las cuales son capturadas y codificadas de manera automática. El sistema gestiona el desarrollo de la entrevista, ya que remite a las opciones adecuadas (en caso de preguntas condicionadas, como por ejemplo: “¿Tiene usted una cuenta en este banco, sí o no?” Si la respuesta es “sí”, el sistema continúa con la siguiente pregunta enlazada: “¿Qué servicios usa...?” Pero si la respuesta fue un “no”, puede concluir con un muchas gracias..., o pasar a otras preguntas). El entrevistador puede utilizar diademas con audífonos y micrófono. O bien, el sistema tiene la facilidad de reconocimiento de voz y de capturar directamente la respuesta. Es una interfase con el *Random Digit Dialing* (Sue y Ritter, 2012 y Outwater, 2008).

Algunas recomendaciones para las entrevistas telefónicas son las siguientes:⁸

⁸ Mertens (2010); Kuusela, Callegaro y Vehovar (2009) y The Psychology Research Handbook (2006). Esta última fuente es una excelente introducción para el alumno que comienza a adentrarse en las encuestas telefónicas.

● **Tabla 6.3** Ejemplos adicionales de encuesta en áreas de salud (continuación).

1. Reclutar entrevistadores con habilidades de comunicación verbal (voz clara, capacidad de realimentar, manejo de conflictos verbales, agilidad mental para responder, etcétera).
2. Asignar a una persona que monitoree permanentemente a los entrevistadores (supervisor).
3. Enviar una carta, telefonema o correo electrónico previo, en que se indique el objetivo de la entrevista, la persona o institución que realiza la encuesta y el día y hora en que se efectuará la comunicación telefónica.
4. Llevar a cabo la comunicación en el día y hora acordados.
5. El entrevistador debe identificarse y recordarle al entrevistado el propósito del estudio; asimismo, debe asegurarse de que es un buen momento para la comunicación.
6. Entre la comunicación previa y la entrevista telefónica no debe pasar más de una semana (programar llamadas adecuadamente).
7. El entrevistador debe asegurarse de que está hablando con la persona correcta o que posee el perfil adecuado según la definición de la muestra.
8. Indicar el tiempo que tomará la entrevista.
9. Utilizar un cuestionario breve con preguntas preferentemente estructuradas (cerradas) y sencillas. Más de 15 o 17 preguntas suelen complicar la situación.
10. El entrevistador debe vocalizar correctamente y a la misma velocidad de su interlocutor.
11. Anotar casos de rechazos y las razones.
12. En el entrenamiento, simular las condiciones de aplicación (igual en la prueba piloto).
13. Establecer metas de comunicaciones telefónicas por hora.
14. Si se pretende grabar la entrevista, debe interrogarse al participante si está de acuerdo con ello.

Desde luego, estas recomendaciones se aplican a una encuesta telefónica cuyos tiempos de ejecución son extensos. Sin embargo, en ocasiones tenemos que realizar sondeos inmediatos para obtener tendencias en la opinión pública y algunas de estas recomendaciones no son pertinentes. Por ejemplo, después de una catástrofe (como un magnicidio, acto terrorista o desastre natural), una noticia mundial (la elección de un nuevo pontífice, una invasión de un país o un acuerdo de paz) o local (un triunfo electoral, un nuevo impuesto o el rechazo a un mandatario). Tales fueron los casos de las encuestas que se realizaron en los días posteriores al asesinato de J. F. Kennedy (Sheatsley y Feldman, 1964), las que se efectuaron después de los ataques terroristas del 11 de septiembre de 2001 en Nueva York (University of Southern California y Bendixen & Associates, 2002), las que se realizaron con motivo de los atentados ocurridos en Madrid el 11 de marzo de 2004 (Michavila, 2005), las ulteriores a las explosiones del 7 de julio de 2005 en Londres (COMPAS, 2005; The Harris Poll, 2005 y British Broadcasting Corporation, 2005), la encuesta telefónica nacional que se implementó en México luego del brote del virus de la influenza humana (Consulta Mitofsky, 2009) y la que se aplicó también en México después de la elección del papa Francisco (BGC-Excelsior, 2013).

Finalmente, es necesario destacar la importancia de no realizar investigaciones que requieran enviar el cuestionario en épocas complejas del año (vacaciones de verano o invierno: en Navidad la saturación es impresionante) o que lo sean para la población en estudio (por ejemplo, a fiscalistas y contadores de empresas durante momentos de cierres contables y pago de impuestos; a las personas de edad avanzada en épocas de frío extremo, etcétera).

Un apunte: comparación entre contextos de entrevistas personales

Jaffe, Pasternak y Grifel (1983) hicieron un estudio para comparar, entre otros aspectos, las respuestas obtenidas en dos puntos: en el hogar y en los lugares de venta. El estudio se interesaba en la conducta del comprador y los resultados concluyeron que se pueden obtener datos exactos en ambos puntos, aunque la entrevista en los lugares de compraventa es más barata.

A la misma conclusión llegaron Hernández-Sampieri, Cuevas y Méndez (2009), quienes entre 2007 y 2009 hicieron ocho encuestas para conocer la intención del voto y las tendencias electorales en varios municipios de México, y encontraron resultados similares a las que arrojaron las entrevistas en el hogar y en sitios públicos (parques, mercados, centros comerciales, etcétera).

Aplicaciones multimodales de cuestionarios

El ideal de la aplicación de cuestionarios es la versión multimodal o su implementación utilizando varios contextos de administración (por ejemplo, en encuestas), la cual se ha popularizado en los últimos años y combina diferentes maneras de recolectar la información en un mismo proyecto de investigación que incluye datos cuantitativos y cualitativos. Por ejemplo, utilizar la encuesta tradicional por entrevista personal para obtener datos cuantitativos básicos y respuestas a unas cuantas preguntas abiertas, junto con otras formas complementarias como la encuesta telefónica (tanto a teléfonos fijos como móviles o celulares) para realizar preguntas muy cortas y cerradas, la encuesta por internet para penetrar ciertos aspectos mediante preguntas abiertas y otras más complejas, y la administración de cuestionarios grupales con la finalidad de lograr profundidad en determinadas respuestas. Aunque, desde luego, las preguntas centrales son las mismas.

Se recomienda este enfoque cuando se dispone de un presupuesto suficiente y se pretende elucidar problemas asociados con la potencial cobertura deficiente que implica utilizar un solo método o incrementar las tasas de respuesta (Link, 2008). También, cuando se intenta acceder a diversas poblaciones con distintas características, como disponibilidad de teléfono e internet, dispersión geográfica y nivel socioeconómico y educativo. Por ejemplo, una encuesta que sea administrada a diversos tipos de personas del sector salud: funcionarios o ejecutivos (que como tienen poco tiempo para responder el cuestionario puede ser autoaplicado), enfermeras y empleados de apoyo (administración grupal), médicos y pacientes actuales (mediante entrevistas personales), pacientes dados de alta (telefónica e internet), etcétera.

Dillman, Smyth y Christian (2009) sugieren que cuando se implemente una encuesta multimodal, se limite a ciertas preguntas y que las opciones de respuesta sean las mismas entre los varios formatos (y se reduzcan categorías o se disponga de las alternativas básicas del método más sencillo, por ejemplo, el cuestionario para la versión telefónica), se eviten cambios en la estructura esencial de las preguntas para no modificar el estímulo, las instrucciones sean equivalentes en los diferentes modos y se efectúen pruebas piloto con todos los instrumentos, para asegurar que siempre se obtenga la misma información de manera válida y confiable, lo cual depende del planteamiento del problema y lo que pretenda el investigador.

ESCALOGRAMA DE GUTTMAN

En el libro *Metodología de la investigación*, 6ª edición, concretamente en las escalas de actitudes, como técnicas para recolectar datos, se comentaron el escalamiento tipo Likert y el diferencial semántico, por lo que quedó pendiente el escalograma de Guttman. Ahora lo consideraremos.⁹

Este método para medir actitudes fue desarrollado por Louis Guttman (1916-1987).¹⁰ Se basa en el principio de que algunos ítems indican en mayor medida la fuerza o intensidad de la actitud. *La escala está constituida por afirmaciones*, las cuales poseen las mismas características que el escalamiento tipo Likert. Sin embargo, *el escalograma garantiza que la escala mida una dimensión única*. Es decir, cada afirmación mide la misma dimensión de la misma variable. A esta propiedad se le conoce como *unidimensionalidad* (Reckase, 2009).

Algunos autores consideran que el escalograma, más que un método de medición de actitudes, es una técnica para determinar si un conjunto de afirmaciones reúne los requisitos de un tipo particular de escala (Engelhard, 2008; Shye, 2008 y Edwards, 1983).

Para construir el escalograma *es necesario desarrollar un conjunto de afirmaciones pertinentes al objeto de actitud, las cuales deben variar en intensidad* (Trochim, 2008). Por ejemplo, si pretendiéramos medir la actitud hacia la calidad del trabajo en el nivel gerencial, la afirmación “la calidad tiene que vivirse en todas las actividades del trabajo y en el hogar”, es más intensa que “la calidad debe vivirse sólo en las actividades más importantes del trabajo”. Dichas afirmaciones *se aplican a una muestra a manera de prueba piloto*. Una

⁹ Para profundizar en esta escala, se sugiere consultar a DeVellis (2012); Abdi (2010); Roberts y Jowell (2008); Fabrigar y MacGregor (2006); Dunn-Rankin et al. (2004); Cliff, Kyats y Keats (2002); Weller y Romney (1990); Black y Champion (1976); Guttman (1976); Lingo (1976); Dotson y Sumers (1976) y Nie et al. (1975).

¹⁰ Guttman hizo importantes aportaciones a la medición, además del escalograma, entre otras cuestiones, fue quien conceptualizó en 1944 los niveles de medición nominal, ordinal, de intervalos y de razón (Katz, 1988 y Stevens, 1951).

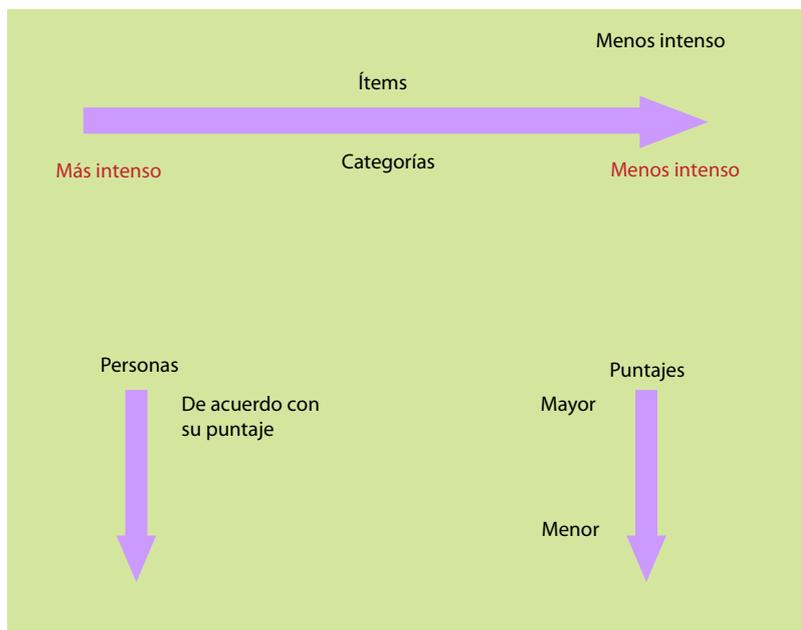
vez administradas, se procede a su análisis. Cabe mencionar que *las categorías de respuesta* para las afirmaciones *pueden variar entre dos* (“de acuerdo-en desacuerdo”, “sí-no”, etc.) *o más categorías* (por ejemplo, las mismas categorías que en el caso de Likert).

Técnica de Cornell

La manera más conocida de analizar los ítems o las afirmaciones y desarrollar el escalograma es la *técnica de Cornell* (Guttman, 1976), en la cual se procede a:

1. *Obtener el puntaje total* de cada participante o sujeto en la escala.
2. *Ordenar a los sujetos o encuestados* de acuerdo con su puntaje total (del puntaje mayor al menor, de manera vertical descendente).
3. *Ordenar las afirmaciones* de acuerdo con su intensidad (de mayor a menor y de izquierda a derecha).
4. *Construir un cuadro donde se crucen los puntajes de los participantes ordenados con los ítems y sus categorías, de manera jerarquizada*. Así, tenemos un cuadro donde los sujetos constituyen los renglones y las categorías de los ítems forman las columnas, lo cual se representa en la figura 7.2.
5. *Analizar el número de errores o rupturas en el patrón ideal de intensidad* de la escala.

● **Figura 7.2** Manera de disponer los resultados para el análisis de ítems mediante la técnica de Cornell.



Ejemplo

Suponga que aplicamos una escala con cuatro ítems o afirmaciones a 14 participantes, cuyo objeto de actitud es “la calidad dentro de la organización” (la variable es la conciencia sobre la calidad por parte de los empleados).

La calidad de la organización, definida como “el grado en que los empleados de una empresa se orientan a cumplir con las especificaciones y requerimientos de los clientes en cuanto a sus productos y servicios” (en precio, tiempo de entrega, cumplimiento de estándares del mercado, etcétera).

Cada afirmación tiene dos categorías de respuesta (“de acuerdo” y “en desacuerdo”, codificadas como uno y cero, respectivamente).

Afirmación A: “La calidad tiene que vivirse siempre en todas las actividades del trabajo y en el hogar”.

Afirmación B: “La calidad tiene que vivirse siempre en todas las actividades del trabajo”.

Afirmación C: “La calidad tiene que vivirse solamente en las actividades más importantes del trabajo”.

Afirmación D: “La calidad tiene que vivirse solamente en la actividad más importante del trabajo”.

Los resultados se muestran en la tabla 7.4.

● **Tabla 7.4** Ejemplo de la técnica de Cornell para el análisis de ítems o afirmaciones.

Sujeto	AFIRMACIONES								PUNTUACIONES TOTALES
	A		B		C		D		
	DA (1)	ED (0)	DA (1)	ED (0)	DA (1)	ED (0)	DA (1)	ED (0)	
1	X		X		X		X		4
2	X		X		X		X		4
3	X		X		X		X		4
4		X	X		X		X		3
5		X	X		X		X		3
6		X	X		X		X		3
7		X		X	X		X		2
8		X		X	X		X		2
9		X		X	X		X		2
10		X		X	X		X		2
11		X		X		X	X		1
12		X		X		X	X		1
13		X		X		X		X	0
14		X		X		X		X	0

DA = de acuerdo o "1" ED = en desacuerdo o "0"

Ahora bien, como se observa en la tabla 7.4, los participantes están ordenados por su puntuación en la escala total. Asimismo, las frases deben ordenarse por su intensidad (en el ejemplo, *A* tiene mayor intensidad que *B*, *B* mayor que *C*, y a su vez, *C* mayor que *D*), y sus categorías se encuentran jerarquizadas de acuerdo con su valor, de izquierda a derecha. Hay que recordar que si la afirmación es negativa, la codificación se invierte (si a la frase *la calidad es poco importante para el desarrollo de una empresa* recibe la respuesta “de acuerdo” se codificará con cero, y si es “en desacuerdo” con uno). En el ejemplo de la tabla 7.4 encontramos cuatro afirmaciones positivas.

Los sujetos que estén “de acuerdo” con la afirmación *A*, que es la más intensa, muy probablemente también lo estén con las afirmaciones *B*, *C* y *D*, ya que su intensidad es menor. Los individuos que respondan “de acuerdo” a la afirmación *B* tenderán a estar “de acuerdo” con *C* y *D* (afirmaciones menos intensas), pero no necesariamente con *A*. Quienes estén “de acuerdo” con *C* lo más probable es que se encuentren “de acuerdo” con *D*, pero no necesariamente con *A* y *B*.

Debe observarse que el individuo número 1 estuvo “de acuerdo” con las cuatro afirmaciones. Los casos 2 y 3 respondieron de igual forma. Las puntuaciones de estos equivalen a cuatro (1 + 1 + 1 + 1). Los participantes 4, 5 y 6 obtuvieron una puntuación de tres, pues estuvieron “de acuerdo” con tres afirmaciones y así sucesivamente (los últimos dos sujetos estuvieron “en desacuerdo” con todas las afirmaciones).

En forma ideal, las personas que obtienen una puntuación total de cuatro en esta escala particular, es porque respondieron “de acuerdo” a las cuatro afirmaciones. Los individuos que alcanzan una puntuación total de tres respondieron que están “en desacuerdo” con la primera afirmación, pero “de acuerdo” con las demás afirmaciones. Quienes reciben una puntuación de dos manifiestan estar “en desacuerdo” con los dos primeros ítems, pero “de acuerdo” con los dos últimos. Del mismo modo, los participantes con puntuación de uno respondieron “en desacuerdo” a las tres primeras afirmaciones y “de acuerdo” a la última. Por último, aquellos que estuvieron “en desacuerdo” con las cuatro afirmaciones, recibieron una puntuación total de cero.

Los sujetos se escalan de manera perfecta, sin que nadie rompa el patrón de intensidad de las afirmaciones: si están “de acuerdo” con la afirmación más intensa, también lo están con las menos intensas. En otro ejemplo, si estoy de acuerdo con la afirmación “*podría casarme con una persona de nivel económico diferente al mío*”, seguramente estaré de acuerdo con la afirmación “*podría viajar en un*

automóvil con una persona de nivel económico diferente al mío” (casarse es mucho más intenso que viajar).

Cuando los individuos se escalan de manera perfecta respecto de las afirmaciones, se puede decir que los ítems en realidad varían gradualmente en intensidad. Es la prueba empírica de que están escalados por su intensidad. Se le denomina *reproductividad* al grado en que un conjunto de afirmaciones o ítems escalan perfectamente según su intensidad. Esto significa que el patrón de respuesta de una persona en relación con todos los ítems puede reproducirse con exactitud si tan solo se conoce su puntuación total en la escala completa (Roberts y Jowell, 2008 y Black y Champion, 1976).

La reproductividad ideal se logra cuando nadie rompe el patrón de intensidad de la escala. Sin embargo, en la realidad sólo algunas cuantas escalas del tipo de Guttman reúnen la reproductividad ideal; la mayoría contienen inconsistencias o rupturas del patrón de intensidad. El grado en que se alcanza el patrón perfecto de intensidad de la escala o *reproductividad* se determina mediante el análisis del número de personas o casos que rompen dicho patrón (Abdi, 2010), que es el quinto paso para construir el escalograma de Guttman.

Análisis del número de errores o rupturas del patrón ideal de intensidad de la escala

Un error es una inconsistencia en las respuestas de una persona a una escala; es un rompimiento con el patrón ideal de intensidad de esta. La tabla 7.5 muestra tres ejemplos de error encerrados en círculos que, como puede verse, son inconsistencias del patrón ideal. El segundo participante respondió “de acuerdo” a los ítems más intensos o fuertes y “en desacuerdo” al ítem menos intenso. El cuarto sujeto manifestó estar “de acuerdo” con las afirmaciones B y C (supuestamente más intensas), pero “en desacuerdo” con D (supuestamente menos intensa). El quinto estuvo “en desacuerdo” con los ítems menos intensos, pero “de acuerdo” con el más intenso. Son inconsistencias o errores. Si un escalograma presenta diversos errores significa que los ítems en realidad no tienen distintos niveles de intensidad (o que los participantes no entienden o son deshonestos).

● **Tabla 7.5** Ejemplos de errores o inconsistencias en un escalograma de Guttman.

Sujeto	AFIRMACIONES								PUNTUACIONES TOTALES
	A		B		C		D		
	DA	ED	DA	ED	DA	ED	DA	ED	
1	X		X		X		X		4
2	X		X		X			X	3
3		X	X		X		X		3
4		X	X		X			X	2
5	X			X		X		X	1
6				X		X		X	0

DA = de acuerdo o “1” ED = en desacuerdo o “0”

Los errores se detectan mediante el análisis de las respuestas que rompen el patrón; para ello, se establecen los *puntos de corte* en el cuadro donde se cruzan las afirmaciones y sus categorías con las puntuaciones totales. En el ejemplo de la tabla 7.4, los “puntos de corte” serían los que se muestran en la tabla 7.6.

● **Tabla 7.6** Ejemplo de establecimiento de los puntos de corte en la técnica de Cornell.

Sujeto	AFIRMACIONES								PUNTUACIONES TOTALES
	A		B		C		D		
	DA (1)	ED (0)	DA (1)	ED (0)	DA (1)	ED (0)	DA (1)	ED (0)	
1	X		X		X		X		4
2	X		X		X		X		4
3	X		X		X		X		4
4		X	X		X		X		3
5		X	X		X		X		3
6		X	X		X		X		3
7	X			X	X		X		2
8	X			X	X		X		2
9	X			X	X		X		2
10	X			X	X		X		2
11	X			X		X	X		1
12	X			X		X	X		1
13	X			X		X		X	0
14	X			X		X		X	0

DA = de acuerdo o "1" ED = en desacuerdo o "0"

No se aprecia *ninguna* inconsistencia. En cambio, la tabla 7.7 muestra *cuatro* inconsistencias o errores; *las respuestas están desubicadas respecto de los puntos de corte*, es decir, rompen el patrón de intensidad (los errores están encerrados en un círculo.)

● **Tabla 7.7** Ejemplo de errores respecto a los puntos de corte.

Sujeto	AFIRMACIONES								PUNTUACIONES TOTALES
	A		B		C		D		
	DA (1)	ED (0)	DA (1)	ED (0)	DA (1)	ED (0)	DA (1)	ED (0)	
1	X		X		X		X		4
2	X		X		X		X		4
3	X		X		X		X		4
4		X	X		X		X		3
5		X	X		X		X		3
6		X	X		X		X		3
7	(X)			X	X		X		3
8	X			X	X		X		2
9	X			X	X		X		2
10	(X)			X	X		(X)		2
11	X			X		X	X		1
12	(X)			X		X		X	1
13	X			X		X		X	0
14	X			X		X		X	0

----- = Puntos de corte (líneas punteadas)

Como ya se mencionó, cuando el número de errores es excesivo la escala no presenta *reproductividad* y no puede aceptarse. La reproductividad se determina mediante un *coeficiente*, cuya fórmula es:

$$\text{Coeficiente de reproductibilidad} = 1 - \left(\frac{\text{Número de errores}}{(\text{número de ítems})(\text{número de sujetos})} \right)$$

En este caso, el número total de respuestas es igual al número de ítems o afirmaciones multiplicado por el número de sujetos. Por lo tanto, la fórmula directa es:

$$\text{Coeficiente de reproductibilidad} = 1 - \left(\frac{\text{Número de errores}}{(\text{número de ítems})(\text{número de sujetos})} \right)$$

En el ejemplo de la tabla 7.7, el coeficiente de reproductividad es:

$$Cr = 1 - \frac{4}{(14)(4)}$$

$$Cr = 1 - 0.071$$

$$Cr = 0.929 = 0.93$$

El coeficiente de reproductividad oscila entre 0 y 1; cuando equivale a 0.90 o más nos indica que el número de errores es tolerable, además de que la escala es unidimensional y se acepta. Cuando es menor que 0.90 no se acepta la escala. Originalmente, Guttman recomendó administrar un máximo de 10 a 12 ítems o afirmaciones a un mínimo de 100 personas (Black y Champion, 1976).

Una vez determinado el número de errores aceptable mediante el *coeficiente de reproductividad*, se procede a aplicar la escala definitiva (si dicho coeficiente fue de 0.90 o más, esto es, si el error permitido no fue mayor a 10%) o a realizar ajustes en la escala (reconstruir ítems, eliminar los que generen errores, etc.). Estos cinco pasos son una especie de prueba piloto para demostrar que la escala es unidimensional y funciona.

Codificación de respuestas

Cuando se aplica la versión definitiva de la escala, los resultados *se codifican dicotómicamente o usando los mismos valores* que en la escala de Likert, lo que depende del número de categorías de respuesta que se incluyan (Abdi, 2010; Fabrigar y MacGregor, 2006 y Carmines y Woods, 2003). Al igual que la escala Likert y el diferencial semántico, todos los *ítems* deben tener *el mismo número de categorías* de respuesta, requisito de *todas* las escalas de actitud. Asimismo, se considera una respuesta inválida si se marcan dos o más opciones ante una misma afirmación. El escalograma de Guttman *es una escala estrictamente ordinal*, pero que se suele usar como si fuera de *intervalo*. Puede *aplicarse* mediante *entrevista* (con uso de tarjetas que contengan las opciones o categorías de respuesta) o *autoadministrarse*.

PRUEBAS E INVENTARIOS

En las ciencias y disciplinas del siglo XXI, gran cantidad de estudios utilizan para sus mediciones pruebas e inventarios de muy diversos tipos. Estos instrumentos se aplican en seres humanos.

Clasificación de las pruebas e inventarios

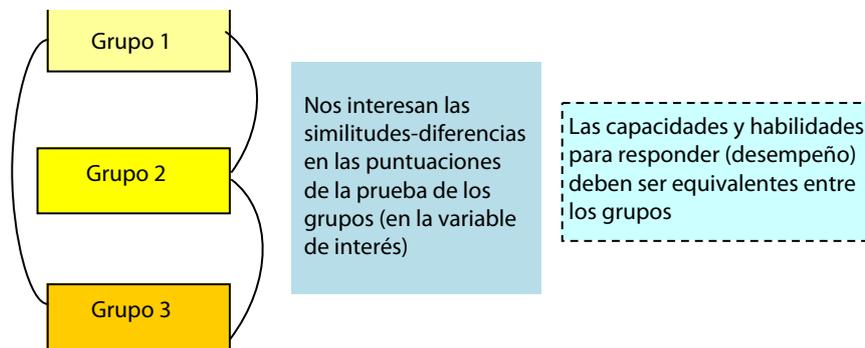
Fundamentalmente, las pruebas e inventarios pueden clasificarse en dos tipos: *estandarizados* y *no estandarizados*.

Los primeros se caracterizan esencialmente porque las instrucciones para administrarse y calificarse son uniformes, y cuentan con ciclos específicos de desarrollo y reconfirmación. Los segundos, usualmente se generan mediante un proceso menos riguroso y su aplicación es limitada (como una prueba de un maestro para evaluar el aprendizaje de sus alumnos en una asignatura). Desde luego, los *no* estandarizados pueden irse robusteciendo (validándose en diferentes contextos) hasta convertirse en pruebas estandarizadas.

En teoría, un inventario estandarizado puede aplicarse a cualquier persona del tipo para el cual fue elaborado (Mertens, 2010), y el tiempo que se brinda para contestarlo en ocasiones es flexible y en otras no. Cabe destacar que solo podemos comparar los resultados entre grupos a los que se les administró una prueba estandarizada (por ejemplo, en experimentos), cuando no existen diferencias entre los grupos que puedan generar sesgos en las respuestas (sería incorrecto comparar las respuestas a un test de razonamiento matemático entre niños de diez años y niñas de seis). Por supuesto que, cuando contrastamos grupos

(hombres frente a mujeres, entre distintos niveles socioeconómicos, jerarquías en la empresa, ocupaciones, etc.) intentamos detectar si hay o no diferencias, pero solo nos interesan las desigualdades entre las variables medidas o los resultados de la prueba. En cambio, no nos deben interesar las discrepancias entre grupos en las capacidades para el desempeño de esta, porque supuestamente las capacidades son las mismas (salvo que el objetivo del estudio sea considerar disparidades en la ejecución de un inventario). El esquema que se pretende lograr cuando se comparan grupos podría esquematizarse como en la figura 7.3.

● **Figura 7.3** Interés al comparar grupos con la misma medición.



Por ejemplo, si queremos comparar a varios pilotos y saber cuál es el mejor, los autos deberán ser los mismos, la pista igual, el combustible, equipo humano, etc. Si probamos autos y pilotos, la pista será la misma, las condiciones ambientales también, etc. Imaginemos que queremos comparar la inteligencia de cuatro grupos, uno de españoles, otro de peruanos, un tercero de italianos y un cuarto de germanos, y la prueba está en alemán. En este caso, los resultados reflejarán la inteligencia o el conocimiento del idioma? Más bien lo segundo.

¿Qué pruebas estandarizadas se encuentran disponibles?

En la actualidad existe una amplia diversidad de pruebas e inventarios desarrollados por diversos investigadores para medir un sinnúmero de variables. Estas pruebas, que tienen su propio procedimiento de aplicación, codificación e interpretación, se encuentran disponibles en diversas fuentes secundarias y terciarias, así como en centros de investigación y difusión del conocimiento. Hay *pruebas* para medir habilidades y aptitudes (como habilidad verbal, razonamiento, memoria, inteligencia, percepción, habilidad numérica), personalidad, intereses personales, valores, desempeño escolar, motivación, aprendizaje, clima laboral en una organización, inteligencia emocional, etc. También se puede disponer de *pruebas clínicas* para detectar conductas anormales, pruebas para seleccionar personal, pruebas para conocer las percepciones u opiniones de las personas sobre diversos tópicos, pruebas para medir la autoestima y, en fin, otras muchas variables internas y del comportamiento.

El reto que implica el empleo de estas pruebas es que la mayoría se han desarrollado en contextos muy diferentes al iberoamericano, por lo que en ocasiones su empleo resulta inadecuado, inválido y poco confiable. Cuando se utilice una prueba estandarizada como instrumento de medición, es conveniente que *se seleccione una desarrollada o adaptada por algún investigador para el mismo ámbito de nuestro estudio, que sea válida y confiable* (debemos tener información sobre este aspecto).

En caso de queelijamos una prueba diseñada en otro contexto, será necesario adaptarla y efectuar pruebas piloto para calcular su validez y confiabilidad, así como ajustarla a las condiciones de nuestra investigación. Recordemos que cada vez que se administra un instrumento resulta indispensable demostrar que es válido y confiable para el contorno en el cual se va a aplicar.

Como ya comentamos, Nuñez (2001) desarrolló una prueba estandarizada para medir el sentido de vida (bajo la concepción de Victor Frankl), una versión con 59 ítems, mediante el escalamiento tipo Likert, con el reporte de una confiabilidad alfa-Cronbach de 0.91, y una solución de ocho factores,

ubicados en cuatro cuadrantes de dos ejes: *homo faber* (éxito-fracaso) y *homo patiens* (plenitud-desesperación). Revisó, antes de crear su prueba, 15 instrumentos como el PIL, Logo Test y Song Test, Mile Test, Life Purpose Questionnaire Test, etc. Comenzó con un instrumento de 180 reactivos que depuró hasta llegar a sus 59 finales. Este es un excelente ejemplo de una adaptación al medio latinoamericano de una prueba estandarizada sobre el sentido de vida. Los ítems definitivos se incluyen al terminar esta sección del capítulo.

En la actualidad encontramos gran variedad de estas pruebas en formato digital (internet, CD, DVD y USB) que hacen mucho más fácil y divertido contestarlas; un ejemplo es la prueba *Emotional IQ Test* (prueba de inteligencia emocional) de Meyer, Salovey y Caruso (1999). En la web hay diversos sitios donde se localizan estas pruebas: en su buscador teclee entre comillas la variable a medir y la palabra “prueba o *test*”, y encontrará un sinnúmero de estos instrumentos para recolectar datos, pero adáptelos/contextualícelos de acuerdo con lo señalado en el capítulo 9 del libro impreso. También busque en los sitios recomendados en el capítulo 3 de este centro de recursos en línea.

Un tipo de pruebas estandarizadas bastante difundido lo constituyen las *pruebas proyectivas*, las cuales presentan estímulos a los participantes para que respondan a ellos; después se pueden analizar e interpretar las respuestas tanto cuantitativa como cualitativamente. Estas pruebas evalúan proyecciones de las personas, como, por ejemplo, la personalidad y sus trastornos (Rodríguez, 2007).

Dos pruebas proyectivas muy conocidas son el *test de Rorschach* (que presenta manchas de tinta en tarjetas o láminas blancas numeradas a los sujetos y estos relatan sus asociaciones e interpretaciones en relación con ellas) (Choca, 2012; Weiner, 2009 y Karson, 2006) y el *Test de Apercepción Temática* o TAT (que, con un esquema similar al de Rorschach, muestra a los participantes cuadros que evocan pasajes o historias, y ellos deben elaborar una interpretación, que es analizada por expertos).¹¹ Por ejemplo, Pirkko (2008) examinó mediante esta prueba historias de 40 niños para evaluar sus fantasías y construcción de la realidad, mientras que Blankenship *et al.* (2006) lo utilizaron para medir la necesidad de logro. Por su parte, Gau y James (2013) implementaron el TAT para confirmar los valores presentes en los espectadores de deportes (diversión, socialización, estatus, moral, ritual, epistémico, espiritualidad y estético, entre otros).

Dichas pruebas se estructuran sobre un procedimiento estandarizado, aunque también se consideran instrumentos cualitativos para recolectar datos, dada la profundidad de información que recogen de las personas en sus propias palabras. Cabe señalar que, regularmente, su aplicación es estandarizada, no así su interpretación. Por su parte, el análisis es más bien de corte cualitativo.

Muchas pruebas estandarizadas (como las proyectivas) requieren de un entrenamiento considerable y un conocimiento profundo de las variables por parte del investigador que las administra e interpreta. No pueden aplicarse con superficialidad e indiscriminadamente. La manera de emplear, codificar, calificar y explicar los resultados de las pruebas estandarizadas es tan variada como los tipos existentes.

Algunos ejemplos de las pruebas, se incluyen en la tabla 7.8.

Hay miles de pruebas e inventarios en las fuentes secundarias y bases de datos que se mencionan en los apéndices 1 y 2 de esta página web o centro de recursos en línea, así como páginas en internet como:

- Tests and Measures in the Social Sciences: Tests available in compilation volumes (pruebas de diversas clases): <http://libraries.uta.edu/helen/test&meas/testmainframe.htm>
- PsycInfo: <http://www.apa.org/psycinfo/>
- TestLink (ETS Test Collection Advanced Search o Tests on Microfiche) y The Mental Measurements Yearbook (también en versión manual de base datos). Se ingresa por medio de las páginas de las universidades (debe estar suscrita su institución).
- ERIC (mencionada ampliamente): <http://www.eric.ed.gov/>
- Psych Tests, BBC Surveys and Psychology Tests o Queendom.com.
- Amazon.

¹¹ Para profundizar en la técnica y ver ejemplos de su aplicación se recomienda a: Teglasi (2010); Husain (2009); Sheldon et al. (2007); Clemence (2006); Langan-Fox y Grant (2006); Blankenship et al. (2006) y Weertman et al. (2006).

● **Tabla 7.8** Ejemplo de pruebas estandarizadas.

Variable general	Variable específica	Instrumento/referencias o ejemplos de aplicación
Inteligencia	Coficiente de inteligencia o IQ de adultos	Escala de inteligencia de Wechsler (WAIS) (Grégoire, 2013; Lichtenberger y Kaufman, 2009; Hartman, 2009 y Kaufman, 2009).
	Coficiente de inteligencia o IQ de niños	Escala de inteligencia para niños de Wechsler (WISC-IV) (Grégoire, 2013).
	Inteligencia sensorio-motriz de bebés (6 meses a 2 años)	Escala argentina de Inteligencia Sensorio-motriz (Caballero y Contini, 2008 y Oiberman, Mansilla y Orellana, 2001).
Memoria	Memoria (funcionamiento cognitivo)	Escala de la memoria Wechsler (WMS-IV) (Chu, Lai, Xu y Zhou, 2012).
Personalidad	Personalidad en prácticamente todas sus dimensiones	Inventario Multifacético de la Personalidad (MMPI), buscar las actualizaciones más recientes (versión adultos y versión adolescentes) (Tarescavage et al., 2013; Greene, 2009; Toomey, Kucharski y Duncan, 2009; Elliott, 2008 y Dao, Prevatt y Horne, 2008).
Aptitudes mentales	Nivel general de aptitudes mentales de niños	Escala Stanford-Binet para niños (Oerbeck, Johansen, Lundahl y Kristensen, 2012; Wallace, 2009 y Newton et al., 2008). En el caso de niños autistas se sugiere a Coolican, Bryson y Zwaigenbaum (2008).
	Aptitud cognitiva (nivel mental) para adolescentes y adultos	Escala Binet-Simon o Standfor-Binet (Pierson, Kilmer, Rothlisberg y McIntosh, 2012; Wallace, 2009; Terman, 2008 y Roid, 2003).
Machismo	Machismo general	Escala de machismo (Pardo, Weisfeld, Hill y Slatcher, 2013).
Actitud hacia el colegio	Actitud de adaptación al colegio en estudiantes de educación media y superior	Inventario de adaptación al colegio WPS de Brown y Leigh (2009).
Creencias familiares	Creencias familiares de los padres	Inventario de las creencias familiares de los padres (Fischer y Corcoran, 2007).
	Creencias familiares de los adolescentes	Inventario de las creencias familiares de los adolescentes (Fischer y Corcoran, 2007).
Dolor	Dolor físico	Cuestionarios del dolor McGill [MPQ] (McClellan et al., 2012; Ljunggren, Strand y Johnsen, 2007 y Hyunjeong et al., 2008).
		Escala visual análoga (VAS) y Escala de calificación verbal (VRS) (Aicher, Peil, Peil y Diene, 2011).
Percepción en el matrimonio	Satisfacción de la relación marital	Escala de felicidad marital [MHS] (Fischer y Corcoran, 2007). Escala-cuestionario del amor en el matrimonio y la relación (Pardo, Weisfeld, Hill y Slatcher, 2013).
	Actitud hacia el matrimonio	Escala de actitud e integración hacia el matrimonio (Riggio y Weiser, 2008).
Amor y sexo	Percepciones sobre el amor y el sexo	Escala del amor y el sexo (Hendrick y Hendrick, 2002).

● **Tabla 7.8** Ejemplo de pruebas estandarizadas (continuación).

Variable general	Variable específica	Instrumento/referencias o ejemplos de aplicación
Percepción de la vida	Satisfacción con la vida	<p>Escala de satisfacción con la vida (SWLS) (Wiest, Schüz y Wurm, 2012; Lopez y Snyder, 2003; Vassar, 2008 y Gouveia et al., 2009).</p> <p>Batería de escalas sobre satisfacción con la vida de pacientes depresivos y con adicciones: 1) Escala multidimensional sobre la satisfacción con la vida (breve), 2) Escala de construcción positiva de la vida/sentirse contento/bienestar y 3) Inventario de la depresión de Beck (Büssing, Heusser y Götz, 2011).</p>
Conciencia ecológica	Conciencia ecológica	Escala de nueva conciencia ecológica de Ellis y Thompson (Robinson, Shaver y Wrightsman, 1999).
Calidad de la práctica de la enfermera	Intervenciones para asistir al paciente en estado de estrés	Inventario de intervenciones para asistir al paciente en situaciones estresantes de Edelman y Mandle (Stanhope y Knollmueller, 2001).
Dependencia	Dependencia interpersonal	Inventario de dependencia interpersonal de Hirschfield (IDI) (Shahar, 2008 y Tzeng, 1993), así como también la subescala del Inventario del estilo personal de Robins (PSI) (Shahar, 2008).
Lealtad del cliente	Lealtad hacia la marca	<p>Escala de Wunderman de experiencia con la marca (Morel, 2003).</p> <p>Encuesta de conciencia y lealtad a la marca (Panchal, Khan y Ram, 2012).</p>
Jubilación	Actitud hacia la jubilación	Escalas para medir las actitudes hacia la jubilación usando diferencial semántico y escalamiento de Thurstone (Rodríguez, 2006).
Inteligencia emocional	Inteligencia emocional y personalidad	<p>Veinte cuestionarios de personalidad (rasgos) (Carter, 2009), así como pruebas de inteligencia emocional (EI) (Stough, Saklofske y Parker, 2009). Prueba audiovisual de la inteligencia emocional [para PC o dispositivo electrónico] (Zysberg y Tell, 2013).</p> <p>En España se adaptó el Meta Mood Scale-24 para medir la inteligencia emocional (Pérez y Castejón, 2007).</p>
Trastorno alimentario	Trastorno o desorden alimentario	Inventario del desorden alimenticio II (Zysberg y Tell, 2013).
Mascotas	Actitud hacia las mascotas	Escala de actitud hacia las mascotas (Morovati et al., 2008).
Comportamiento de compra	Comportamiento de compra compulsiva	Compulsive Buying Scale (Escala de compra compulsiva) (Koran et al., 2007).
Religiosidad	Creencias, asistencia a la iglesia y participación o membresía en organizaciones religiosas	Diversas escalas y cuestionarios, considerando un metaanálisis con 63 estudios (Zuckerman, Silberman y Hal, 2013).

Además, se recomiendan: Urbina (2014), Yan, von Davier y Lewis (2014) (pruebas estandarizadas en dispositivos electrónicos), Cohen, Swerdlik y Sturman (2012), Mertens (2010), Kubiszyn y Borich (2009), Barrett (2009), Anastasi y Urbina (2009), Groth-Marnat (2009), Abell, Springer y Kamata (2009), Thorndike y Thorndike-Christ (2009), Kaplan y Saccuzzo (2008) y Mertens y McLaughlen (2004) (particularmente pruebas para personas con capacidades diferentes), Kamphaus y Reynolds (2003), Thorndike y Hagen (1989), Cronbach (1984), Norton (1980), Woelfel y Danes (1980) y, desde luego, las obras clásicas de Torgerson (1985) y Nunnally (1970). Se trata de autores especializados sobre medición y manejo de pruebas estandarizadas.

Además, existen métodos propios de las diferentes ciencias sociales, como el análisis de redes, para evaluar la forma en que se manifiesta la comunicación en un sistema social (quién se comunica con quién, quiénes distorsionan la información, cómo fluye la comunicación, quiénes son los líderes comunicativos, etc.). Para este análisis se sugiere a Rogers y Kincaid (1981), Wellman y Berkowitz (1997), Wellman (1999), Knoke y Yang (2007) y Hassanien, Abraham y Snasel (2009). Un ejemplo de esta clase de investigación es la de Crossley (2008), quien analizó la red social de los “actores claves” involucrados en el “círculo central” del movimiento punk en el Reino Unido.

Asimismo, los sistemas de medición fisiológica [por ejemplo, Jagroop, Persaud y Mikhailidis (2011), generaron un método más rápido para medir el colesterol en las plaquetas humanas], escalas multidimensionales que miden a los sujetos en varias dimensiones [como los sistemas Galileo y ALSCAL de Woelfel y Fink (1980), así como Borg y Groenen (2005), para medir la distancia psicológica entre conceptos en individuos (como por ejemplo: “patria”, “madre”, “presidente”, “nación” y cualquier otro), en las que se consideran dimensiones cognitivas y emocionales]; y en fin, otros métodos que escapan del alcance introductorio de este libro.

En el caso de ingenierías y ciencias físicas vale la pena señalar que cada año se busquen actualizaciones de libros clásicos como: *Instrumentation and measurement in electrical engineering*, de Roman Malaric (Brown Walker Press), *Measurement and instrumentation principles*, de Alan S. Morris (Butterworth-Heinemann), *Measurement and instrumentation: Theory and application*, de Alan S. Morris y Reza Langari (Butterworth-Heinemann), *Theory and design for mechanical measurements*, de Richard S. Figliola y Donald E. Beasley (John Wiley and Sons) e *Instrumentation for process measurement and control*, de Norman A. Anderson (CRC Press), solo por mencionar algunos; además, es útil recurrir a revistas científicas. Normalmente los profesores que imparten asignaturas específicas en su institución conocen varios de estas fuentes.

Ejemplo de ítems de prueba estandarizada (Sentido de vida Celaya)¹²

Los ítems o afirmaciones son calificados a través de una escala de Likert que va desde “totalmente de acuerdo” hasta “totalmente en desacuerdo”.

1. En mi familia nos escuchamos, nos demostramos amor y ocasionalmente analizamos los desacuerdos.
2. Mi religión ha fomentado el crecimiento de mi capacidad de amar y desarrollarme.
3. Dedico parte de mi vida a realizar actividades desinteresadamente.
4. Estoy convencido de que estoy bien; me acepto y me amo.
5. En la búsqueda de la felicidad intento lograr satisfactores materiales.
6. Honestamente tengo conciencia de que amo las cosas y utilizo a las personas.
7. La afectividad es muy importante en mi vida.
8. Mi alto nivel de energía lo empleo para triunfar.
9. Mi religión me ha proporcionado razones para vivir con rectitud, para estar bien con Dios.
10. Acepto que me traten indignamente, con tal de tener seguridad económica.
11. Promuevo que las personas que dependen de mí tengan aprendizajes significativos.
12. Me gusta todo lo que es fácil y que no requiere pensar mucho.
13. Asisto a congresos, conferencias, tomo cursos y diplomados.
14. Me gusta comunicarme conmigo y sentir que mi cabeza sabe lo que tiene mi corazón.
15. Le agrado a la mayoría de las personas.



16. Si me dieran la oportunidad, podría hacer algo que realmente ayudara a otros.
17. Repentinamente mi estado de ánimo cae en la desesperanza.
18. Valoro todo lo que favorece mis posibilidades de estar bien.
19. Puedo vislumbrar la posibilidad de cambiar y de crecer en mi persona.
20. Me gusta comprometerme en obras de tipo social.
21. Contagio mi optimismo.
22. Mis problemas personales influyen en mi trabajo.
23. En mi proyecto de vida establezco metas relacionadas con un aprendizaje importante.
24. En mi familia muchas veces hemos rezado juntos.
25. Siento que mi cuerpo pierde fortaleza interna.
26. He buscado activa y libremente a Dios.
27. A mis familiares les disgusta comprometerse en obras de tipo social.
28. En mi familia mis padres acostumbraban devaluarse uno a otro.
29. Me siento insatisfecho con mi trabajo.
30. Mis sentimientos me impiden ver con claridad la realidad.
31. Tengo la sensación de que “algo” le falta a mi vida.
32. A la persona que amo la acepto incondicionalmente.
33. Mi meta al educar no sólo es informar, sino afinar la conciencia.
34. En estos momentos me doy cuenta de que no puedo ocuparme de nada porque no tengo voluntad de hacerlo.
35. Promuevo una educación que tenga una relación más estrecha con la vida.
36. Invierto parte de mis ingresos en adquirir revistas científicas, buenos libros, videos, documentales, enciclopedias.
37. Con frecuencia siento que algo malo me va a ocurrir.
38. Mis opiniones políticas no están bien definidas.
39. Renuncio a mi libertad para obtener satisfactores que me convengan.
40. Soy muy sensible al sufrimiento de los demás.
41. Trato de estar con personas divertidas para que me ayuden a salir de mi aburrimiento.
42. En mi familia se carece de sentimientos de lealtad y de pertenencia.
43. Me siento lleno de vida y vitalidad.
44. Disfruto la lluvia y no me importa mojarme.
45. Me parece que sufro más que los demás cuando rompo una relación.
46. Prefiero mantenerme a cierta distancia de la gente y quedarme en mi soledad.
47. Me considero una persona optimista.
48. La vida es como una “pizza” llena de ingredientes, apetecible y deliciosa.
49. Me siento capaz para trabajar eficientemente.
50. Soy instrumento de las circunstancias y el destino.
51. No pasa un día en que no dedique unos minutos para contemplar la belleza de la naturaleza.
52. En mi trabajo confío en mis dotes y en mi capacidad de adaptarme para hacer frente a los desafíos que se presentan.
53. Admiro a personas como Martin Luther King, Nelson Mandela y la madre Teresa de Calcuta.
54. Diariamente trato en mi trabajo de dar lo mejor de mí mismo.
55. Tengo necesidad de protagonismo, de ser notado.
56. Me gusta echarme boca arriba en el césped y sacar figuras de las nubes.
57. Hay en mi interior una gran riqueza y sensibilidad.
58. Puedo permitirme vivir sentimientos delicados o tiernos.
59. Me siento capaz de hacer bien mi trabajo.

ANÁLISIS DE CONTENIDO

¿Qué es y para qué sirve el análisis cuantitativo de contenido?

Este método tiene mayores implicaciones que simplemente recoger contenidos, pues es, al mismo tiempo, un cierto tipo de diseño de investigación y/o técnica de análisis, pero lo incluimos en este capítulo porque innegablemente su proceso comienza con la recopilación de contenidos de comunicaciones en sus múltiples formas: verbales, no verbales, escritas y audiovisuales.

De acuerdo con la definición clásica de Berelson (1971), el **análisis de contenido** es una técnica para estudiar y analizar la comunicación de una manera objetiva, sistemática y cuantitativa. Stalnaker (2014) y Krippendorff (2013) extienden la definición del análisis de contenido a un método de investigación para hacer inferencias válidas y confiables de datos sobre contenidos con respecto a su contexto. Krippendorff y Bock (2008), así como Weber (1990) señalan que tales inferencias implican a la fuente de los mensajes, a los mensajes en sí mismos y a sus receptores o destinatarios.

Esta herramienta es una técnica muy útil para analizar los procesos de comunicación y sus características en muy diversos contextos (Neuendorf, 2001). El análisis de contenido puede aplicarse virtualmente a cualquier forma de comunicación [páginas web (como las de anuncios publicitarios o personales) y a toda clase de blogs, mensajes en redes sociales de internet (whatsApp, twitter, facebook, etc.), programas televisivos o radiofónicos, artículos en prensa o periódicos electrónicos, libros, poemas, conversaciones, pinturas, discursos, cartas, melodías, leyes y reglamentos, etc.]. Por ejemplo, para analizar la personalidad de alguien es conveniente evaluar sus escritos; conocer las actitudes de un grupo de personas mediante la valoración de sus expresiones verbales (desde pandillas hasta movimientos de protesta y pacientes que sufren una determinada enfermedad); indagar sobre las preocupaciones de un pintor o un músico mediante el estudio de su material; compenetrarse con los valores de una cultura o averiguar las intenciones de un publicista o propagandista (como las campañas nazis de Hitler y Goebbels durante la Segunda Guerra Mundial en todos los medios de comunicación colectiva de la época).

Análisis de contenido Técnica para estudiar la comunicación de una manera objetiva, sistemática y que cuantifica los contenidos en categorías.



Usos del análisis de contenido

Riffe, Lacy y Fico (2014); Weber (1990) y Berelson (1971) señalan varios usos del análisis de contenido, entre los que se destacan:

1. *Describir tendencias y características* del contenido de la comunicación.
2. *Develar diferencias* en el contenido de la comunicación (entre personas, grupos, instituciones, países, medios informativos, etcétera).
3. *Comparar mensajes, niveles y medios* de comunicación.
4. *Intentar predecir* los posibles efectos del contenido de la comunicación en la audiencia destinataria.
5. *Auditar el contenido* de la comunicación y compararlo con estándares u objetivos.
6. *Construir y aplicar estándares de comunicación* (políticas, normas, etcétera).
7. *Exponer técnicas publicitarias y de propaganda*.
8. *Medir la claridad* de los mensajes.
9. *Descubrir estilos* de comunicación.
10. *Identificar intenciones, apelaciones y características de comunicadores*.
11. *Descifrar mensajes ocultos* y otras aplicaciones a la inteligencia militar y a la seguridad política.
12. *Revelar "centros" de interés y atención* de una persona, un grupo, un medio informativo o una comunidad.
13. *Determinar el estado psicológico* de personas o grupos.
14. *Obtener indicios del desarrollo verbal* (por ejemplo, en la escuela, como resultado de la enseñanza, el aprendizaje de conceptos).
15. *Anticipar respuestas* a comunicaciones.



16. *Reflejar actitudes, valores y creencias de personas, grupos o comunidades.*
17. Cerrar preguntas abiertas.

Asimismo, puede resultar útil para examinar redes sociales (quiénes se comunican con quiénes —perfiles—, para qué —propósitos—, con qué frecuencia, etc.). Por ejemplo, podría estudiarse una red de jóvenes inscritos en facebook, twitter, whatsApp, Hi5 o mySpace mediante el análisis de sus páginas (mensajes escritos y visuales).

¿Cómo se realiza el análisis de contenido?

El análisis de contenido se efectúa por medio de la codificación, es decir, el proceso en virtud del cual las características relevantes del contenido de un mensaje se transforman en unidades que permitan su clasificación o categorización precisa. Lo importante del mensaje se convierte en algo susceptible de describir y analizar. Para codificar es necesario definir el *universo*, las *unidades de análisis* y las *categorías de análisis* (Krippendorff y Bock, 2008).

Universo

El universo podría ser la obra completa de Paulo Coelho o J. K. Rowling; las emisiones de un noticiario o informativo en televisión durante un mes; los editoriales publicados en un día por cuatro diarios (por ejemplo, en España, El Mundo, El País, ABC y La Vanguardia); todos los capítulos de tres telenovelas; las canciones completas de un grupo musical o solista (U2, Joan Manuel Serrat, Shakira, Coldplay, La Oreja de Van Gogh, Motel, etc.); los escritos de un conjunto de estudiantes durante un ciclo escolar; los discursos pronunciados por varios contendientes políticos en el mes previo a la elección; los mensajes de un círculo de amigos intercambiados en una red social en internet durante una semana vacacional; los escritos de un grupo de pacientes en psicoterapia en un periodo específico; las conversaciones grabadas de 15 parejas que participan en un experimento sobre conflictos maritales; los escritos de una secta terrorista, etc. El universo, como en cualquier investigación cuantitativa, debe delimitarse con precisión. Algunos autores, como Sánchez Aranda (2005), se refieren al universo como el contexto de la unidad, el cuerpo más largo del contenido que puede examinarse al caracterizar una o más unidades de registro.

Unidades de análisis

Las unidades de análisis o registro constituyen segmentos del contenido de los mensajes que son caracterizados para ubicarlos dentro de las categorías. Se predeterminan de acuerdo al planteamiento del problema, a diferencia de las unidades en el análisis cualitativo, las cuales surgen de manera emergente. Sánchez-Aranda (2005, p. 221), las define como “El cuerpo de contenido más pequeño en que se cuenta la aparición de una referencia, ya sean palabras o afirmaciones que nos interesa localizar”. Riffe, Lacy y Fico (2014) y Berelson (1971) mencionan cinco unidades importantes de análisis:

1. *La palabra.* Es la unidad de análisis más simple, aunque, como señalan Franzosi (2009) y Kerlinger y Lee (2002), puede haber unidades más pequeñas, como letras, fonemas o símbolos. Así, se puede medir cuántas veces aparece una palabra en un mensaje (por ejemplo, número de veces que en los informativos o noticiarios televisivos de fin de año se menciona al presidente de la República o el número de alusiones de un grupo psiquiátrico al término “depresión”).
2. *El tema.* Se define a menudo como un enunciado respecto de algo. Los temas suelen ser más o menos generales.¹² Kerlinger y Lee (2002) utilizan un excelente ejemplo para ello: “Las cartas de adolescentes o estudiantes de colegios superiores pueden ser estudiadas en sus expresiones de autorreferencia”. Este sería un tema extenso, pero se definirían los temas como cualesquiera de las oraciones que usen “yo”, “mi” y otros términos que indiquen referencia al “yo” del escritor. Así, se analizaría qué tanta autorreferencia contienen las cartas.

¹² El significado del término “temas” en el análisis cuantitativo del contenido se refiere a tópicos y es distinto al significado de “temas emergentes” en el análisis cualitativo.

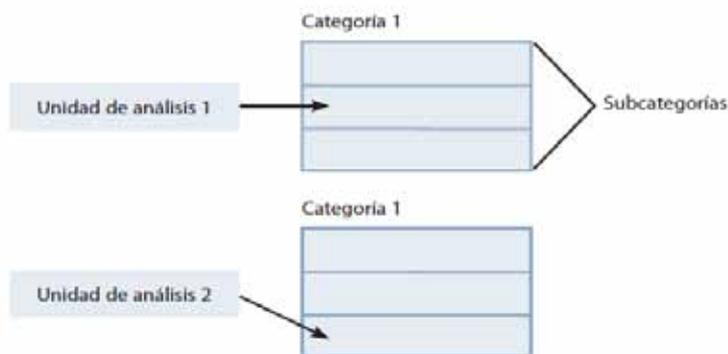
Si los temas son complejos, el análisis del contenido es más difícil, en especial si se complica al incluirse más de una oración simple.

3. *El ítem.* Es una unidad de registro muy utilizada y puede definirse como la unidad total empleada por los productores del material simbólico (Stan, 2009 y Weber, 1990). Ejemplos de ítems son un libro, un editorial, una emisión de un programa de radio o televisión, un discurso, una ley, una fotografía, una carta amorosa, una conversación telefónica, un mensaje en una red social de internet, una canción o la respuesta a una pregunta abierta. En este caso, lo que se analiza es el material simbólico total.
4. *El personaje.* Se trata de un individuo. Por ejemplo, un paciente, un héroe o heroína de una serie televisiva, un líder histórico, un profesor, un avatar, un joven que publica una página en facebook, etc. Aquí lo que se analiza es el personaje.
5. *Medidas de espacio-tiempo.* Son unidades físicas, como el centímetro-columna (por ejemplo, en la prensa), la línea (en escrito), el minuto (en una conversación o en radio), el mensaje (enviado por celulares o móviles), el periodo de 20 minutos (en una interacción), el cuadro (en televisión), cada vez que se haga una pausa (en un discurso, una entrevista o una interacción entre individuos), una página (red social en la web), etc. Krippendorff y Bock (2008) y Sánchez-Aranda (2005) lo denominan unidad de enumeración. En este caso, la unidad física (por ejemplo, cada minuto) es ubicado en categorías (si tenemos 10 minutos de grabación, se colocan diez unidades en las categorías).

Algunos contenidos típicos de herramientas cualitativas de recolección de datos (como las grabaciones o transcripciones de grupos de enfoque) pueden evaluarse mediante análisis cuantitativo del contenido.

Las unidades se insertan, colocan o caracterizan en categorías y/o subcategorías, lo cual podría represen-

● **Figura 7.4** Agrupación de unidades de análisis en categorías.



tarse tal como se muestra en la figura 7. 4.

En cuanto a qué unidad debe seleccionarse, ello depende de los objetivos y las preguntas de investigación. Sin embargo, Krippendorff (2013) y Berelson (1971) sugieren lo siguiente:

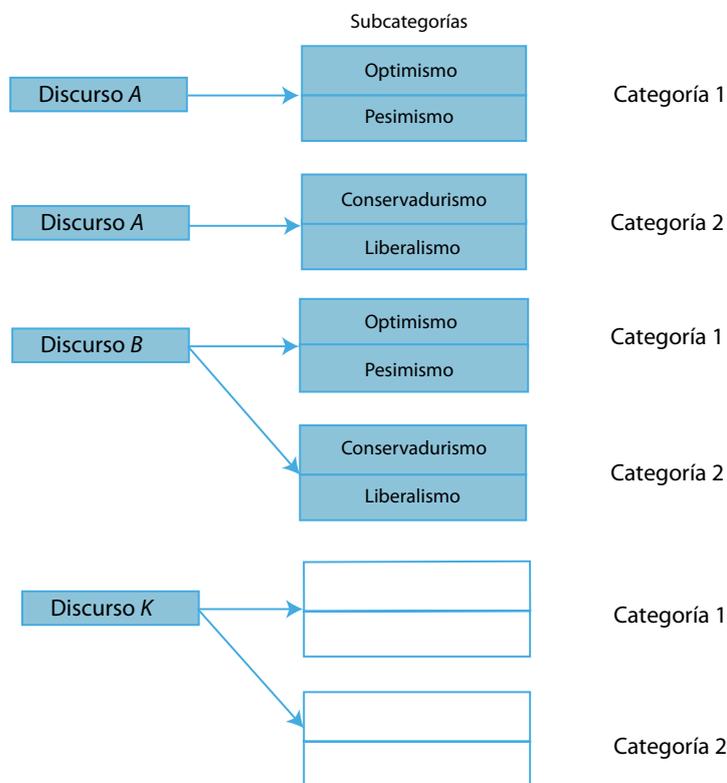
- a) En un solo estudio es posible utilizar más de una unidad de análisis.
- b) Los cálculos de palabras y las unidades amplias, como el ítem y las medidas de espacio-tiempo, son más adecuados en los análisis que dan mayor importancia a los asuntos definidos.
- c) Las *unidades* amplias y las más definidas son válidas para aceptar o rechazar una categoría.
- d) Por lo general, las *unidades* amplias requieren de menos tiempo para su codificación que las unidades pequeñas que se refieren a las mismas categorías y materias.
- e) Debido a que los temas o las oraciones agregan otra dimensión al asunto, la mayoría de las veces son más difíciles de analizar que las palabras y las *unidades* amplias.
- f) El tema es adecuado para el análisis de significados y las relaciones entre estos.

Categorías

Las **categorías** son los niveles donde se caracterizan las unidades de análisis. Como menciona Holsti (1969), son las “casillas o cajones” en las cuales se clasifican las unidades de análisis. También se pre-

determinan de acuerdo con el planteamiento del problema. Prasad (2008), Sánchez Aranda (2005) y Weber (1990) las definen como características o aspectos que presenta la comunicación con la que trabajamos (en cuanto a referencias). Por ejemplo, un discurso podría clasificarse como optimista o pesimista, como liberal o conservador. Un personaje de una caricatura se catalogaría como bueno, neutral o malo. En ambos casos, la unidad de análisis se categoriza. Veamos esquemáticamente la categorización en la figura 7. 5.

● **Figura 7.5** Ejemplos de unidades de análisis y su clasificación en categorías.



Es decir, cada unidad de análisis se incluye o encasilla en uno o más sistemas de categorías. Por ejemplo, en un estudio citado por Krippendorff (1980) se analizaron 2 430 actos de violencia televisada. En cada acto, el personaje principal (unidad de análisis) se categorizó como:

- Bueno, neutral o malo (sistema 1).
- Asociado con hacer cumplir la ley, no relacionado con el cumplimiento de la ley o como un delincuente o criminal (sistema 2).

En este caso, la unidad de análisis o registro es el comportamiento del personaje durante el acto televisivo. Esto podría representarse como se muestra en la figura 7. 6.

● **Figura 7.6** Ejemplo con dos categorías y seis subcategorías.



Como ya se comentó, la selección de categorías depende del planteamiento del problema.

Tipos de categorías

Krippendorff (2013) señala cinco tipos básicos de categorías:

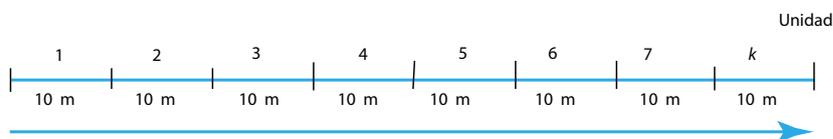
1. *De asunto o tópico.* Se refieren a cuál es el asunto, la materia o la temática tratada en el contenido (¿de qué trata el mensaje o la comunicación?).

Ejemplo

Analizar las problemáticas que más preocupan a los ciudadanos detectadas en los mensajes enviados por correo electrónico al alcalde o presidente municipal durante el último mes.

Categoría:	Inseguridad
Subcategorías:	Asaltos/robos al transitar o caminar en las calles o lugares públicos abiertos, como parques Asaltos/robos al transitar en automóviles Asaltos/robos al viajar en transporte público Asaltos/robos a los hogares Asaltos/robos en lugares públicos cerrados, como restaurantes, cines, etcétera Secuestros Violaciones Etcétera
Categoría:	Deficiencias en los servicios públicos que presta el municipio
Subcategorías:	El servicio de recolección de la basura no ocurre todos los días El servicio de recolección de la basura no se realiza más o menos a la misma hora todos los días, es decir, es irregular Suspensión por más de un día del servicio de suministro del agua Cortes frecuentes en el suministro de agua Etcétera

Imaginemos que pretendemos investigar el grado en que una determinada telenovela presenta consumo de alcohol entre sus personajes y las situaciones y contextos en los cuales ocurre, así como su finalidad (reducir estrés, inducir al romance, socializar, etc.). Cada periodo de 10 minutos de la telenovela se va a considerar la unidad de análisis, o podría ser todo el capítulo.



Categoría:	Relación del personaje con el consumo de alcohol
Subcategorías:	El personaje consume bebida alcohólica El personaje no consume bebida alcohólica
Categoría:	Finalidad de consumir alcohol
Subcategorías:	Reducir estrés o relajarse Socializar con otros personajes Transformar su carácter en uno más asertivo, violento o agresivo Transformar su carácter en uno más bondadoso o simpático Inducir al romance a otro personaje Etcétera

2. *De dirección.* Estas categorías se refieren a cómo es tratado el asunto (¿positiva o negativamente?, ¿favorable o desfavorablemente?, ¿nacionalista o no nacionalista?, etcétera).

Ejemplo

Comparar la manera en que dos noticieros televisivos comentan sobre el presidente de Estados Unidos.

Categoría:	Tono en el tratamiento respecto del presidente de Estados Unidos
Subcategorías:	Favorable Neutral Desfavorable

3. *De valores.* Este tipo de categorías indican qué valores, intereses, metas, deseos o creencias se revelan.

Ejemplo

Al estudiar la compatibilidad ideológica de matrimonios, se podría analizar la ideología de cada cónyuge, pidiéndoles un escrito sobre temas que puedan reflejar valores (sexo, actitud hacia la pareja, significado del matrimonio).

Categoría:	Ideología del esposo
Subcategorías:	Muy tradicional Más bien tradicional Neutral Más bien liberal Muy liberal

4. *De receptores.* Estas categorías se relacionan con el destinatario de la comunicación (¿a quién van dirigidos los mensajes?)

Ejemplo

Analizar a quiénes encausa más sus críticas el rector o director general de la universidad en los discursos del último año dirigidos a toda la comunidad estudiantil.

Categoría:	Receptores a quienes se dirigen las críticas del rector o director
Subcategorías:	Profesores de tiempo completo Profesores que no están de tiempo completo (de asignaturas, por honorarios o tiempo parcial) Alumnos en general Alumnos de últimos semestres Alumnos de semestres intermedios Alumnos de primeros semestres Directivos Empleados administrativos Padres de familia A toda la comunidad

5. *Físicas.* Son categorías para ubicar la posición y duración o extensión de una unidad de análisis. De posición son, por ejemplo, la sección, página y cuadrante (en prensa escrita o electrónica), y el horario (en televisión y radio); de duración, los minutos (en una interacción, un comercial televisivo, un programa de radio, un discurso), los centímetros/columna (en prensa), los cuadros o escenas en una película, etc. No se deben confundir las unidades de espacio-tiempo con las categorías físicas. Las primeras son unidades de análisis y las segundas constituyen categorías. Veámoslo en la figura 7.7 con un ejemplo simple.

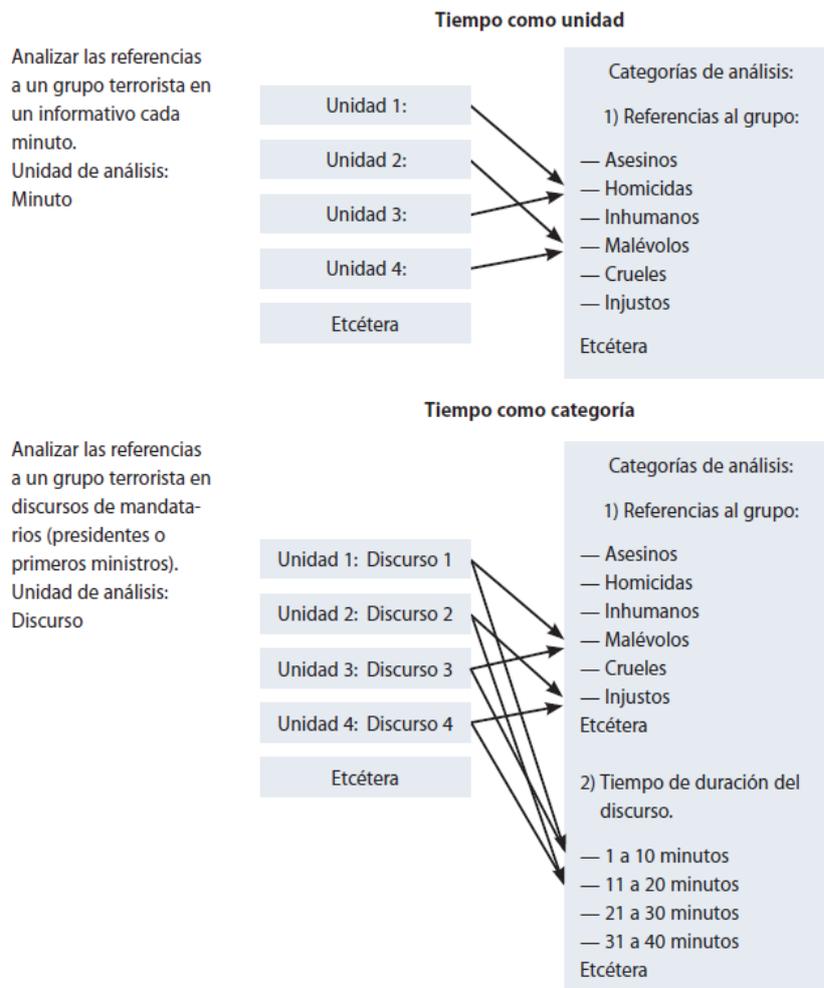
Un sexto tipo de categorías es la “intensidad”, que indica más allá de la dirección, grados de *fuerza* de la categoría. Por ejemplo, diversos gabinetes, despachos o departamentos de comunicación social de organismos públicos (como sería el caso de ministerios o secretarías), examinan diariamente las críticas que hacen los editorialistas de periódicos al titular de la dependencia (ministro o secretario), y determinan si la crítica es “muy severa”, “severa” y “no severa”.

Requisitos de las categorías

Cuando se lleva a cabo un análisis de contenido se suelen determinar varias categorías, que deben cumplir los siguientes *requisitos*:

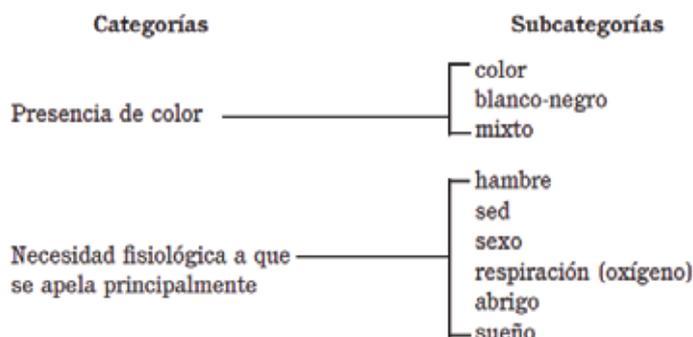
1. Las categorías y subcategorías deben ser *exhaustivas*, es decir, abarcar todas las posibles subcategorías

● **Figura 7.7** Ejemplo del tiempo como unidad de análisis y como categoría.



de lo que se va a codificar. Por ejemplo, la categoría “ideología del esposo” no podría prescindir de la subcategoría “neutral”.

- De preferencia, las subcategorías deben ser *mutuamente excluyentes*, de tal manera que una unidad de análisis clasifique en una y sólo una de las subcategorías de cada categoría. Por ejemplo, un personaje de una serie televisiva no puede ser “bueno” y “malo” a la vez. Si ejerciera ambos roles, habría que generar la categoría “en momentos malo y en otros bueno”.
- Obviamente, una unidad de registro puede encajar en una subcategoría de una categoría y en otras subcategorías de otras *categorías*. Por ejemplo, al analizar comerciales televisivos podríamos tener, entre otras categorías, las siguientes:



Una unidad de análisis (un comercial) puede ubicarse en una subcategoría de “presencia de color” y en otra subcategoría de la categoría “necesidad fisiológica a que se apela” (por ejemplo: color y sed), pero no puede situarse en dos subcategorías de la misma categoría, como “hambre” y “sed” (porque la categoría es la necesidad a que se apela *principalmente*), a menos que generáramos la subcategoría “hambre y sed”.

En ciertos casos especiales, al analista le llega a interesar una categoría donde las subcategorías no sean mutuamente excluyentes. Tal sería el caso de que la categoría fuera “apelaciones del comercial” (estatus socioeconómico, salud, juventud, belleza, amor romántico, etcétera).

Un comercial podría utilizar más de una apelación (salud y belleza, por ejemplo).

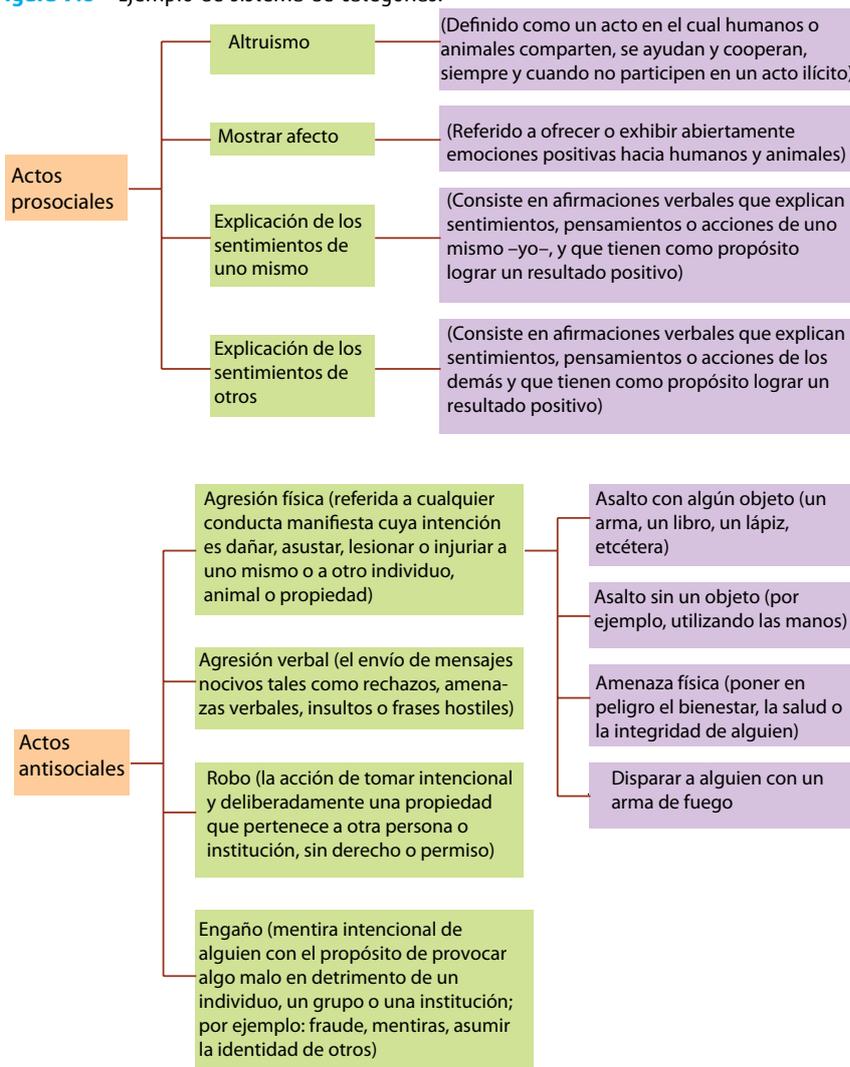
- Las categorías y subcategorías *deben derivarse del marco teórico y de una profunda evaluación de la situación.*

Ejemplo

Análisis de contenido de series televisivas

Para ejemplificar el análisis de contenido y específicamente la generación *a priori* de categorías, se acudirá a un estudio de Greenberg *et al.* (1980). La investigación consistió en un análisis de contenido de las series televisadas por las tres grandes cadenas de Estados Unidos: ABC, CBS y NBC. Se evaluaron diversos programas durante tres periodos de 1975 a 1978, para determinar el grado en que la televisión estadounidense mostraba actos prosociales y antisociales como medida de la violencia televisada. Las categorías y subcategorías que se utilizaron son las que se muestran en la figura 7. 8.

● **Figura 7.8** Ejemplo de sistema de categorías.



El estudio consideró como unidad de análisis la conducta (cada vez que se presentaba un comportamiento de los personajes era codificado).

Cuando se crean las categorías, estas deben definirse con precisión y es necesario explicitar qué van a comprender en cada caso y qué se excluirá.

El análisis de contenido consiste en asignar cada unidad a una o más categorías. En realidad, el producto de la codificación son frecuencias de categorías. Mediante el programa de análisis estadístico, se cuenta cuántas veces se repite cada categoría o subcategoría (cuántas unidades de análisis se clasificaron en cada una de las categorías). Por ejemplo, Greenberg et al. (1980, p. 113) encontraron los resultados que se muestran en la tabla 7.9.

● **Tabla 7.9** Ejemplo de los resultados de la codificación del estudio de acuerdo con la estudio de Greenberg et al. (1980).

Incidencia de actos antisociales en los tres periodos						
	Año 1*		Año 2*		Año 3*	
	f	%	f	%	f	%
I. Agresión física						
A. Asalto con un objeto	466	(15.7)	248	(10.8)	370	(13.6)
B. Asalto sin un objeto	111	(3.7)	159	(6.9)	177	(6.5)
C. Amenaza física	180	(6.1)	233	(10.1)	135	(5.0)
D. Disparar	106	(3.6)	75	(3.2)	74	(2.7)
E. Otras	128	(4.3)	171	(7.4)	130	(4.8)
II. Agresividad verbal	1 629	(55.0)	1 099	(47.6)	1 464	(54.0)
III. Robo	61	(2.1)	72	(3.1)	44	(1.6)
IV. Engaño	283	(9.5)	251	(10.9)	319	(11.8)
Total	2 964	(100.0)	2 308	(100.0)	2 713	(100.0)
Horas analizadas	(68.5)		(58)		(63)	

*Año 1 incluyó de octubre de 1975 hasta que se grabó un episodio de cada una de las series existentes (1976), año 2 igual, pero en 1976-1977, y año 3 igual, pero de 1977-1978.
f = número de casos o unidades.
% = porcentajes.

A continuación mostramos un ejemplo de análisis de contenido, a manera de ejercicio, que se hace cotidianamente en la mayoría de los países para clasificar las películas y regular el ingreso (por el factor *edad*) a las salas cinematográficas.

Ejercicio

Imagine el lector que le asignan la tarea de clasificar películas u obras cinematográficas para evaluar si son adecuadas para el público de todas las edades, si sólo son apropiadas para adolescentes y adultos, o únicamente son aptas para adultos (mayores de 18 años). Por supuesto, la unidad de análisis es el filme o cinta, pero: ¿qué categorías podrían utilizarse para clasificar las unidades?

Algunas de las categorías para clasificar las películas, cintas o filmes podrían ser:

- Contenido de sexo.
- Contenido de violencia.
- Contenido de lenguaje ofensivo, soez o procaz (insultos, groserías, majaderías).
- Contenido sobre adicciones y drogas ilegales.
- Contenido de horror (no terror sino, por ejemplo, escenas de muertos —un holocausto que muestre cadáveres— o la filmación de la muerte de una persona).

Cabrían al menos tres preguntas sobre estas categorías: ¿son demasiado generales?, ¿qué significa cada una de ellas?, ¿son suficientes o podrían establecerse otros criterios? Las respuestas podrían ir en varios sentidos. La primera es indudable: “sí”, tales categorías son demasiado generales. Por ello, su significado es poco preciso. Y alguien podría agregar otra categoría: contenido de discriminación, y aun otras categorías.



La categoría “contenido de sexo” podría englobar las siguientes subcategorías primarias: a) desnudos, b) presentación de actos sexuales, c) degradación de valores mediante el sexo, y otros conceptos. En la subcategoría desnudos, podrían agruparse otras subcategorías (secundarias), por ejemplo: 1) eróticos ligados al acto sexual (uno o varios actores y actrices desnudos, comprometidos en acciones sexuales), 2) eróticos no ligados al acto sexual (uno o varios actores y actrices simplemente desnudos), 3) no eróticos ni ligados a una acción sexual (por ejemplo, una escena donde aparezca alguna escultura desnuda, como el David, de Miguel Ángel, o un bebé desnudo). La presentación de actos sexuales podría abarcar las siguientes subcategorías secundarias: 1) sexo sugerido (escenas donde los actores insinúan que habrá un acto sexual, pero no se ve ninguna acción; por ejemplo, entran a una recámara, sonríen, se miran entre sí, cierran la puerta...), 2) sexo implícito (escenas de conducta sexual manifiesta, en la que se observa una acción pero esta no es clara ni precisa; por ejemplo, una pareja que está en una cama, entre las sábanas y supuestamente tiene una relación sexual) y 3) sexo explícito (conducta sexual manifiesta y escenificada, en la cual los actores se encuentran desnudos y se observa con claridad y precisión la acción sexual). La degradación de valores podría clasificarse en (subcategorías secundarias): 1) el sexo se presenta como una forma de entrega y amor de la pareja para el desarrollo armónico de la relación y el crecimiento humano de sus integrantes, 2) el sexo se presenta como una pulsión instintiva, 3) el sexo se presenta como una forma de degradación humana o humillación.

La violencia podría ser de distintos tipos (subcategorías primarias): a) física, b) verbal o c) psicológica. En el caso de la violencia física, esta podría llevarse a cabo con armas o sin ellas (subcategorías secundarias). La violencia física podría provocar ciertos efectos: 1) ningún efecto, 2) heridas reversibles que no requieren hospitalización, 3) heridas reversibles que requieren hospitalización, 4) heridas irreversibles que requieren hospitalización y 5) muerte (subcategorías secundarias). La violencia también podría asociarse al consumo de drogas o no y al contenido sexual o no (subcategorías secundarias).

El contenido sobre adicciones y/o drogas podría abarcar las subcategorías primarias: a) producción, b) distribución y c) consumo. Esta última podría subdividirse en subcategorías secundarias: 1) consumo de alcohol, 2) tabaco, 3) estupefacientes de daño significativo pero menor (por ejemplo, marihuana) y 4) estupefacientes de daño severo a todo el organismo (por ejemplo, cocaína, heroína, entre otras). Asimismo, otras subcategorías secundarias del consumo podrían estar relacionadas con sus consecuencias: 1) daños reversibles a los sistemas del organismo, 2) daños irreversibles y 3) fallecimiento. Y así, las demás categorías.

De esta forma, cada película se clasificaría en las categorías y subcategorías. Podríamos señalar que la presencia de ciertas subcategorías secundarias otorga un estatus u otro a la cinta en cuestión, por ejemplo:

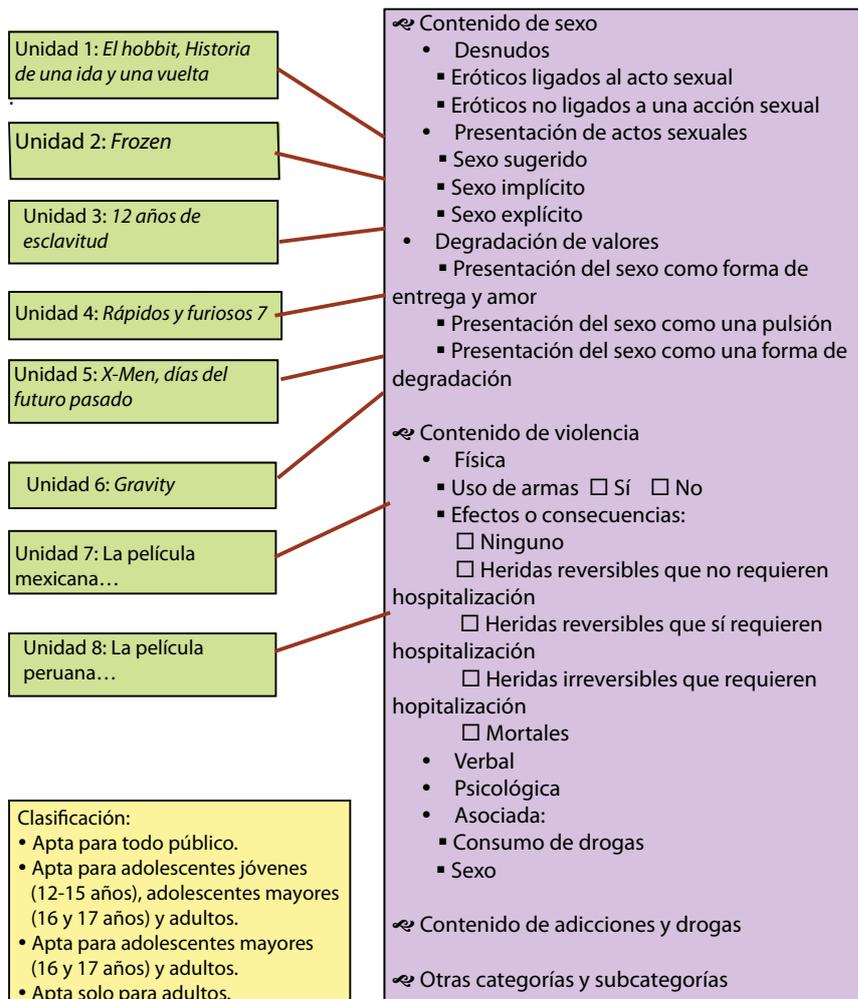
Películas aptas para todo público

- Sólo se permiten desnudos no eróticos; si alguna escena presenta desnudos eróticos (ya sea que estén o no ligados al acto sexual), el filme se clasificará, según sea el caso, como “apto para adolescentes jóvenes (12-15 años)”, o “apto para adolescentes mayores (16 y 17 años) y adultos”.
- Únicamente se permite sexo sugerido; si alguna escena presenta sexo implícito o sexo explícito, el filme será clasificado según sea el caso.
- Imposible vincular el sexo con valores, mucho menos presentarlo como una pulsión instintiva y menos aún como una forma de degradación humana o humillación (subcategorías secundarias).
- La violencia únicamente puede presentarse en un contexto fantástico caricaturizado.
- No se permite ningún contenido sobre adicciones y/o drogas.

Así, con cada grupo de edad: apta para adolescentes jóvenes (12-15 años); apta para adolescentes mayores (16 y 17 años) y adultos; y apta solamente para adultos o mayores de edad.

Este proceso de categorización podría representarse gráficamente de la forma como se muestra en la figura 7.9, de la página siguiente

● **Figura 7.9** Ejemplo de un proceso de categorización de películas.



Para complementar el ejercicio le sugerimos que, junto con otros compañeros, lo termine y genere un sistema para clasificar películas (simplemente mostramos algunas categorías por las cuales puede comenzar; desde luego, podría no estar de acuerdo con ellas o con las subcategorías mencionadas, o con los criterios para otorgar un estatus u otro a la cinta). Una vez que desarrolle el sistema, vaya al cine, analice algunos filmes de estreno y otórgueles su propia clasificación de acuerdo con sus criterios.

¿Cuáles son los pasos para llevar a cabo el análisis de contenido?

Ya hemos mencionado tres:

1. Definir con precisión el universo y extraer una muestra representativa.
2. Establecer y definir las unidades de análisis.
3. Establecer y definir las categorías y subcategorías que representan las variables de la investigación.

Los demás pasos son:

4. *Seleccionar los codificadores.* Los codificadores son las personas que deben asignar las unidades de análisis a las categorías. Deben tener un nivel educativo profesional (estudiantes de licenciatura, como mínimo).
5. *Elaborar las hojas de codificación.* Estas hojas contienen las categorías y subcategorías, y los codificadores anotan en ellas cada vez que una unidad entra en una subcategoría. En la figura 7. 10 se mues-

● **Figura 7.10** Elementos de una hoja o registro de codificación.

CODIFICADOR: 1		MATERIAL A ANALIZAR:	CAPÍTULO DE LA SERIE "La teoría del Big Bang"
FECHA: 29-11-2014		DURACIÓN: 60 MINUTOS	
CATEGORÍAS		FRECUENCIAS	TOTALES
ACTOS PROSOCIALES (1)	ALTRUISMO (1.1)		2
	MOSTRAR AFECTO (1.2)		5
	EXPLICACIÓN DE LOS PROPIOS SENTIMIENTOS (1.3)		3
	EXPLICACIÓN DE LOS SENTIMIENTOS DE OTROS (1.4)		0
ACTOS ANTISOCIALES (2)	AGRESIÓN FÍSICA (2.1)	Asalto con objeto (2.1.1)	0
		Asalto sin objeto (2.1.2)	0
		Amenaza física (2.1.3)	0
		Disparar (2.1.4)	0
		Otros (2.1.5)	0
	AGRESIÓN VERBAL (2.2)		3
	ROBO (2.3)		0
	ENGAÑO (2.4)		1
		TOTAL	14

Pueden agregarse páginas para comentarios del codificador

tran los elementos de una hoja o registro de codificación, que puede estar impresa en papel o elaborada en un programa para cualquier dispositivo electrónico (PC o tableta, por ejemplo). Incluso puede ser una hoja en Excel.

Las *hojas o registros de codificación* pueden incluir elementos más precisos para describir el material. Por ejemplo, para analizar artículos de periódicos: nombre del periódico, fecha de publicación, género periodístico (reportaje, entrevista, editorial, artículo de fondo, carta, inserción pagada, artículo general), signatario del texto, sección donde se publicó, página, volumen de publicación (año, número o equivalente), ubicación en los cuadrantes de páginas,

X	

tamaño (en centímetros/columna), nombre del codificador, día de codificación, etc. Una hoja puede servir para una o varias unidades, según nos interese o no tener datos específicos de cada unidad (normalmente se prefiere tener la información por unidad).

Asimismo, las *categorías y subcategorías* deben *codificarse con* sus respectivos valores numéricos. En la hoja de codificación de la figura 7.10, 1 (uno) significa acto prosocial y 2 (dos) acto antisocial (así, 1.1 es “altruismo”, 1.2 “mostrar afecto”, etc.). Recordemos que puede haber categorías secundarias y terciarias o sistemas aún más desglosados.

6. *Entrenar a los codificadores.* Esta capacitación incluye que los codificadores se familiaricen y competan con las variables, comprendan las categorías y subcategorías, así como las definiciones de ambas. Además, debe entrenárseles en la manera de codificar, y discutir ampliamente con ellos las diferentes condiciones en que puede manifestarse o estar presente cada categoría y subcategoría. Asimismo, tienen que entender en forma cabal en qué consiste la unidad de análisis.

Parte del entrenamiento implica realizar una prueba piloto con el sistema de categorías y subcategorías, es decir, realizar un análisis de contenido en una muestra representativa del universo del material de estudio (por ejemplo, unos programas de las series televisivas que se analizarán, segmentos de los libros de un autor —capítulos de cada obra—, partes de los discursos de oradores —páginas de cada discurso—. Como se puede apreciar, en este caso, en lugar de muestrear con individuos, se realiza con unidades de significado). Así, podemos descubrir categorías y/o subcategorías no contempladas previamente. La prueba piloto también nos sirve para evaluar el nivel de confiabilidad de los codificadores, lo que nos conduce a la siguiente etapa.

7. *Determinar la confiabilidad de los codificadores.* Todos los codificadores realizan análisis de contenido provisional de una parte representativa del material (este debe ser el mismo para todos los codificadores), con la finalidad de determinar el grado de acuerdo o consenso entre ellos. Si no hay consenso, no puede efectuarse una evaluación confiable de la muestra o universo de contenido.

Para llevar a cabo la evaluación se calcula la confiabilidad de cada codificador (individual) y la confiabilidad entre codificadores. El cálculo de la *confiabilidad* individual de los codificadores depende de que sean uno o varios.

- a) *Si se dispone de un solo codificador* (porque el material es reducido) se observan las diferencias de la codificación del mismo mensaje hecha por el codificador en dos tiempos diferentes. Si las diferencias son muy pequeñas, el codificador es individualmente confiable. Este tipo de confiabilidad se llama confiabilidad intracodificador, la cual mide la estabilidad de la prueba y reprueba de un codificador a través del tiempo.

Otro método consistiría en que el codificador trabaje una parte representativa del material y después, un experto externo y el investigador, revisen su trabajo y apliquen a su codificación (resultados) la siguiente fórmula:

$$\text{Confiabilidad individual} = \frac{\text{Número de unidades de análisis catalogadas correctamente por el codificador}}{\text{Número total de unidades de análisis}}$$

Por ejemplo, un mensaje que conste de 20 unidades y se logren correctamente las 20, la confiabilidad será de $1 = \left(\frac{20}{20}\right)$, que es el máximo de confiabilidad. Si el codificador sólo pudo codificar en forma correcta 15 de las 20, la confiabilidad sería de $0.75 = \left(\frac{15}{20}\right)$.

- b) *Si se dispone de varios codificadores*, la confiabilidad individual puede determinarse así: se solicita a todos los codificadores que codifiquen el mismo material, se toman los resultados de todos

menos los de uno y se compara la codificación de este con la del resto. Así se procede con cada codificador.

También puede aplicarse a todos los codificadores la fórmula para calcular la confiabilidad individual: quien se distancie del resto se considera un caso poco confiable. Si varios muestran valores bajos, se practica con otro material hasta alcanzar índices mínimos de .90 (para la individual).

Ejemplo

Codificador A	Codificador B	Codificador C	Codificador D
0.89	0.93	0.92	0.67

El codificador D tiene baja confiabilidad.

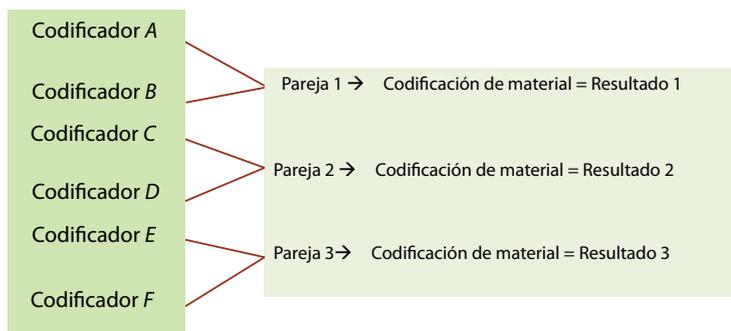
El *cálculo de la confiabilidad intercodificadores* se realiza por pares de codificadores (parejas). Se pide a cada pareja que codifique el material, se comparan los resultados obtenidos por las parejas, se cuenta el número de acuerdos entre ellas, se determina el número de unidades de análisis y se aplica la siguiente fórmula:

$$\text{Confiabilidad entre parejas} = \frac{\text{Número total de acuerdos entre dos parejas}}{\text{Número total de unidades de análisis codificadas}}$$

Después se suman los resultados de esta fórmula y se divide entre el número de comparaciones, que depende del número de parejas. Veámoslo con un ejemplo:

Ejemplo

Las parejas codifican el material:



Se determina el número de acuerdos entre las parejas (un acuerdo consiste en que dos parejas codifican en la misma categoría/subcategoría a una misma unidad de análisis).

Pareja	Número de unidades de análisis codificadas	Número de acuerdos entre parejas
1	18	Entre parejas 1 y 2 = 17
2	18	Entre parejas 1 y 3 = 16
3	17	Entre parejas 2 y 3 = 16

Se aplica la fórmula de confiabilidad entre parejas:

$$C_{1y2} = \frac{17}{18} = 0.94$$

$$C_{1y3} = \frac{16}{18} = 0.89$$

$$C_{2y3} = \frac{16}{18} = 0.89$$

Debe observarse que no hubo consenso total entre cuántas unidades de análisis podían distinguirse en el material (la pareja 3 distinguió 17 unidades, y las parejas 1 y 2 distinguieron 18). En este caso, para aplicar la fórmula de confiabilidad entre parejas, se toma en cuenta el máximo de unidades de análisis codificadas por alguna de ellas.

Pareja	Número de unidades de análisis codificadas por alguna de las parejas. Si fuera:	Número de acuerdos entre parejas
A	25	21
B	22	

Se toma el máximo

En este caso, la fórmula de confiabilidad sería:

$$C_{AB} = \frac{21}{25} = 0.84$$

Se obtiene la *confiabilidad total* (que es la suma de las confiabilidades entre parejas sobre el número de comparaciones). En nuestro ejemplo sería:

$$C_T = \frac{C_{1y2} + C_{1y3} + C_{2y3}}{3}$$

$$C_T = \frac{0.94 + 0.89 + 0.89}{3} = 0.906$$

No es conveniente tolerar una *confiabilidad* menor de 0.85 (ni total, ni entre dos parejas) y de ser posible debe superar el 0.89. Al igual que con otros instrumentos de medición, la confiabilidad oscila entre cero (0 = nula confiabilidad) y uno (1 = confiabilidad total).

En el análisis de contenido, una confiabilidad baja puede deberse a que las categorías o unidades de análisis no se definieron con claridad y precisión, a un deficiente entrenamiento o a la inhabilidad de los codificadores. Cuando se obtiene una baja confiabilidad tiene que detectarse y corregirse el problema.

Asimismo, es conveniente calcular la confiabilidad a la mitad de la codificación (con 50% del material codificado) y al completar esta.

8. Llevar a acabo la codificación. Esta fase implica asignar las unidades al sistema de categorías y contar las frecuencias de repetición de las unidades en las categorías (número de unidades que corresponden a cada categoría).
9. Vaciar los datos de las hojas o registros de codificación y obtener totales de cada categoría. Este vaciado puede ser de las hojas o registros a la matriz de SPSS, Minitab o programa equivalente, o bien, mediante el empleo de dispositivos ópticos.
10. Realizar los análisis estadísticos apropiados. Se calculan estadísticas dependiendo del nivel de medición de las categorías (frecuentemente no paramétricas).

El *análisis de contenido* transforma sistemáticamente material verbal, textual o visual en datos cuantitativos, mediante un conjunto de códigos y reglas previamente establecidos (Stark y Roberts, 2005). Una de sus aplicaciones más socorridas es la evaluación de los mensajes transmitidos a través de los medios de comunicación colectiva. Por ejemplo, Taylor (2009) lo implementó para determinar cómo son retratadas las víctimas de feminicidio y sus perpetradores en los periódicos de Florida (durante un periodo de cinco años). Huhn, Oettler y Peetz (2009) exploraron los tópicos que son tratados por la prensa de Costa Rica, El Salvador y Nicaragua al referirse a la violencia. Asimismo, Stein (2009) investigó la forma en que las organizaciones sociales de Estados Unidos utilizan internet (mediante una encuesta entre usuarios y análisis de contenido de páginas web de 86 organizaciones durante un periodo de cuatro meses en 2006). Auster y Michaud (2013) analizaron el contenido de la mercadotecnia de los parques temáticos de Disney en internet.

El análisis de contenido también puede aplicarse creativamente a una enorme variedad de investigaciones. Por ejemplo, DeCicco (2007) estudió el contenido de los sueños de un grupo de mujeres y demostró su relación significativa con el descubrimiento de sí mismas. Tian (2009) inspeccionó las apelaciones publicitarias utilizadas y los valores culturales reflejados en los anuncios en los periódicos y la televisión de China. Smith (2008) desarrolló un libro sobre múltiples aplicaciones del análisis de contenido para efectuar inferencias sobre la personalidad y la motivación de las personas, así como respecto a características y experiencias de grupos. Por su parte, Harpster, Adams y Jarvis (2009), del FBI, y un departamento de policía, examinaron los indicadores verbales de denuncias hechas sobre homicidios a través del teléfono de urgencias 911, para determinar la inocencia o culpabilidad de la persona que llama. Entre sus resultados descubrieron seis dimensiones lingüísticas que pueden ser útiles para analizar las llamadas telefónicas sobre homicidios y apoyar las investigaciones.

Algunos ejemplos adicionales se muestran en la tabla 7. 10.

Finalmente, hemos de señalar que con el desarrollo de las nuevas tecnologías de comunicación, el análisis de contenido se mantiene vigente y en expansión.¹³

OBSERVACIÓN

¿Qué es y para qué sirve la observación cuantitativa?

Esta técnica de recolección de datos consiste en registrar de manera sistemática, válida y confiable los comportamientos o conductas que se manifiestan, de acuerdo a un conjunto de reglas y procedimientos predeterminados derivados del planteamiento del problema de investigación (Yoder, 2013 y McKechnie, 2008). Puede utilizarse como instrumento de medición en muy diversas circunstancias. Polster y Collins (2007) y Haynes (1978) mencionan que es el método que más utilizan quienes se orientan conductualmente. Por ejemplo, es útil para determinar la eficacia de diferentes estilos de liderazgo de los profesores en la enseñanza, así como para analizar conflictos familiares, eventos masivos (como la violencia en los estadios de fútbol o el rechazo a algún gobierno en manifestaciones públicas), la aceptación de un producto en un supermercado, el comportamiento de personas con capacidades mentales distintas, el efecto de ciertas conductas de atención y cuidado con los pacientes por parte de las enfermeras sobre el apego al tratamiento médico prescrito, los riesgos en el trabajo, la adaptación de un equipo humano al uso de cierta maquinaria, etcétera.

Desde luego, la observación no es adecuada como instrumento de recolección en todos los estudios. Tal sería el caso de los procesos cognitivos o cuestiones internas de los individuos (McKechnie, 2008).

La **observación cuantitativa**, como método para recolectar datos, es similar al análisis de contenido. Por ello, en este apartado algunos conceptos sólo se mencionarán, pero no se profundizará en ellos, pues se trataron en el apartado sobre análisis de contenido.

Un aspecto importante que se debe destacar es que en la observación el instrumento de recolección de datos es una o varias perso-



Observación cuantitativa Registro sistemático, válido y confiable de comportamiento manifiesto. Se recolecta información sobre la conducta más que de percepciones.

¹³ Para revisar técnicas específicas de análisis de contenido aplicadas a los medios de comunicación masiva actuales, recomendamos a Divakaran (2009), y respecto de la investigación turística a Stepchenkova, Kirilenko y Morrison (2009).

● **Tabla 7.10** Ejemplos adicionales de análisis de contenido.

Título de la investigación	Referencia	Breve síntesis
Análisis de contenido de la cobertura en las noticias de las atletas asiáticas	Yu (2009)	El propósito del estudio fue entender cómo los diarios impresos y electrónicos cubrían en las noticias a las deportistas asiáticas que habían ganado medallas en las olimpiadas de Sydney, Atenas y Beijing y si enfatizaban sus características femeninas y qué expectativas y estereotipos señalaban sobre ellas. Se analizaron 266 artículos noticiosos entre el 1 de enero de 2000 y el 31 de agosto de 2008. Uno de los descubrimientos es que la mayoría se refería a buceo y maratón, a pesar de que las atletas son también competitivas en otros deportes. Los artículos provenían de periódicos de China, Canadá, Estados Unidos, Inglaterra, Japón, Francia, India y Malasia, principalmente.
Los prejuicios asociados con la edad y que generan discriminación en los procesos de reclutamiento en las empresas	Álvarez y Hernández-Sampieri (2014)	Como parte de un proyecto de investigación mixta más amplio, en una primera etapa los investigadores analizaron el contenido de ofertas de empleo publicadas durante 2013 en Celaya, Guanajuato, México, en dos muestras probabilísticas (una extraída de una bolsa de trabajo en línea y la otra seleccionada de los dos principales periódicos locales (n1=368 y n2=337), con el fin de detectar si había o no discriminación por edad y género en el caso de diferentes empleos que no tuvieran limitaciones físicas. Descubrieron que en tales anuncios se presentaba discriminación en ambas variables demográficas. La segunda fase, cualitativa y en proceso, implica entrevistar en profundidad a una muestra de ejecutivos y empleados para descubrir las razones de tal discriminación.
Un análisis de contenido de las evaluaciones-intervenciones de la violencia entre parejas íntimas	Hays y Emelianchik (2009)	Aproximadamente 30% de los individuos de diferentes culturas son víctimas de violencia por parte de su pareja, a lo largo de sus vidas. Las investigadoras efectuaron un análisis de contenido de 38 evaluaciones e intervenciones de tal violencia, descubriendo de manera inductiva siete temas (y sus respectivas limitaciones): contexto de la relación, formas de abuso, indicadores inmediatos de riesgo, dinámica familiar, grado de apoyo, estructura de evaluación e información psicométrica. Posteriormente, los temas fueron considerados como categorías y desplegados en subcategorías para analizar su frecuencia y evaluar el contenido de todas las valoraciones-asesorías. Por ejemplo, la categoría "forma de abuso" se dividió en las siguientes subcategorías (categorías secundarias): emocional, físico, sexual y conflicto en la relación. A su vez, cada una, se desplegó en categorías más específicas o terciarias (abuso emocional: verbal, maltrato, falta de respeto, control, miedo, celos, acoso y acecho). Los resultados sirvieron para analizar la naturaleza de la evaluación-intervención de la violencia en la pareja y sus limitaciones, así como recomendar mejores prácticas cuando se trabaja con esta.
¿La voz de quienes carecen de ella? : Un análisis de contenido cuantitativo del canal satelital insignia de noticias en inglés Al-Jazeera	Figenschou (2010)	Al-Jazeera es el primer canal satelital que transmite noticias en inglés desde Medio Oriente con el propósito de equilibrar el flujo de noticias occidentales y proporcionar otra visión sobre el acontecer global (la perspectiva de los países árabes). La investigación consideró 60 noticieros y definió a la unidad de análisis como la noticia (n=1 324). Entre otras categoría se incluyeron: la región (global o internacional, Asia, África, Norteamérica, Sudamérica, Europa, Oceanía, Medio Oriente), el formato (noticia breve, análisis, entrevista fuera del estudio, entrevista en el estudio, otras), región general (sur o países no desarrollados y norte o países desarrollados), fuente de la noticia (establecida, élites independientes, personas ordinarias o generales, empleados y miembros del staff, otros), género de la fuente y temáticas. Se calcularon coeficientes de confiabilidad intercodificadores por categoría. Los resultados dan un panorama claro de cómo el canal maneja la información noticiosa.


Tabla 7.10 Ejemplos adicionales de análisis de contenido (continuación).

Título de la investigación	Referencia	Breve síntesis
Misoginia en la música rapera: un análisis de contenido sobre la prevalencia y significados	Weitzer y Kubrin (2009)	<p>Esta investigación se llevó a cabo para evaluar la manera en que se representa a las mujeres en una muestra de 403 canciones de rap y si se manifiesta misoginia en ellas. Una categoría fue la forma como se refieren al género femenino en las melodías; en ella, las subcategorías consideradas fueron: nombrar y avergonzar a la mujer (49% de frecuencia), tratar a la mujer como "objeto sexual" (67%), desconfiar de la mujer (47%), legitimar la violencia contra ella (18%) y prostitución y proxenetismo (20%).</p> <p>Un requisito para que las canciones se incluyeran en la muestra fue que formaran parte de un álbum con certificación de platino (al menos un millón de copias vendidas), debido a que así se aseguraba que hubieran sido escuchadas por un segmento sustancial de la población (se consideraron 130 álbumes producidos entre 1992 y 2000).</p> <p>La confiabilidad entre codificadores fue de 73.4% para la categoría misoginia.</p>
Imágenes de guerra: Un análisis de contenido de la cobertura fotográfica en la guerra en Kosovo	Nikolaev (2009)	<p>De acuerdo con documentos oficiales estadounidenses e internacionales, el conflicto de Kosovo fue una guerra civil entre la población serbia respaldada por el gobierno de Yugoslavia y la Organización Terrorista Albanesa (KLA). De esta versión partió el trabajo con una pregunta de investigación sencilla pero muy significativa: ¿los medios impresos estadounidenses contaron esta historia correctamente? La hipótesis general afirmaba que la cobertura fotográfica de la guerra había sido desequilibrada con un sesgo evidente contra los serbios.</p> <p>Se consideraron todas las 365 fotografías (unidades de análisis) dedicadas a la guerra de Kosovo por las revistas Newsweek, Time y U. S. News & World Report en un periodo de cinco meses de 1999. La confiabilidad entre codificadores fue sumamente significativa (0.97).</p> <p>Al final, los resultados estadísticos apoyaron la hipótesis.</p>
El contenido de las definiciones de los niños: la distinción oral-escrita	Marinellie (2009)	<p>El estudio se propuso investigar el contenido de las definiciones de palabras proporcionadas por niños, tanto en el modo oral como escrito. Treinta infantes (cuyo promedio de edad era de nueve años y dos meses) definieron sustantivos comunes en ambos modos. Las definiciones fueron codificadas por el uso de términos categóricos (por ejemplo, manzana: fruta; autobús: vehículo) y atributos (por ejemplo, manzana: roja). Los resultados indicaron que los niños muestran diferentes tipos de conocimiento de las palabras en los dos modos, pero que la forma escrita es más útil para producir términos categóricos (para encajar los sustantivos en categorías). Las definiciones orales tienden a ser más extensas.</p> <p>Los resultados sirven para planear mejor la enseñanza de conceptos en niños de educación básica.</p>

nas, porque los equipos (de filmación, para captar movimientos, etc.) solo registran lo que los sentidos humanos van a percibir (Bentzen, 2008 y Anastas, 2005).

En la observación cuantitativa, a diferencia de la cualitativa, las variables que se observan son especificadas y definidas antes de comenzar la recolección de los datos. Se enfoca en información que pueda ser evaluada por medio de los sentidos (datos visuales, auditivos, producto del tacto y el olfato, etc.). Asimismo, los observadores registran lo que perciben de acuerdo con un sistema y reglas que se aplican invariablemente, las cuales les ayudan a minimizar su efecto sobre los registros e interacciones con los participantes observados. Esta circunstancia es particularmente crítica porque los seres humanos somos selectivos en nuestras observaciones e inducimos cierto grado de variación (Ratcliff y Lassiter, 2007).

Pasos para construir un sistema de observación

Los pasos para construir un sistema de observación son:

1. *Definir con precisión el universo de aspectos, eventos, conductas o variables que se observarán.* ¿Qué y

quién o quiénes se van a observar y por cuánto tiempo? Por ejemplo, si nuestro interés es conocer los recursos con que cuentan las escuelas de un distrito escolar, debemos precisar lo que consideramos como “recurso escolar”. Un universo de aspectos a observar podría ser el comportamiento verbal y no verbal de un grupo de alumnos durante un semestre en una asignatura. Otro universo lo podrían constituir las conductas de un grupo de trabajadores durante sus sesiones en círculos de calidad o equipos para la calidad, en el periodo de un año. O bien, las conductas agresivas de un grupo de esquizofrénicos en sesiones terapéuticas en cierto periodo. El universo obviamente depende de las variables que conformen el planteamiento del problema.

2. *Especificar las circunstancias de la observación.* ¿Dónde y cuándo se conducirán las observaciones? (lugar y ubicación temporal). Por ejemplo, el aula en la clase de economía durante este semestre desde el inicio y hasta que concluya el ciclo, antes del periodo de exámenes; todas las sesiones de los grupos para la calidad del departamento de producción, en el lugar donde se reúnen normalmente; las próximas 10 sesiones terapéuticas de ciertos pacientes en una cámara Gesell.¹⁴ Tales circunstancias dependen del planteamiento del problema y la situación. Los lugares pueden ser muy variados: una sala de emergencias, un consultorio médico, una tienda, un aula escolar, un sitio de una comunidad, etc. Resulta obvio que vamos a observar en lugares adecuados y evitar los que son inapropiados (por ejemplo, intentar registrar conductas criminales en un desfile familiar no sería lo común, lo lógico). A veces se arregla o adapta un ambiente, de tal modo que sea lo suficientemente estable y se pueda aprovechar para realizar comparaciones entre conductas por medio de observaciones repetidas (tanto en el tiempo como entre individuos). Además, es indispensable definir si la observación se realizará “en vivo” (mientras ocurren las conductas, aspectos o eventos) o posteriormente (se filmarán los comportamientos y hechos que después se analizarán). Asimismo, se debe especificar el tipo de equipo necesario para efectuar las tareas de observación (si observamos células cancerosas, requeriremos de un potente microscopio; si se trata de planetas, un telescopio de muy largo alcance; el policía que observa a un grupo de pandilleros requerirá de binoculares y un micrófono especial que pueda registrar las conversaciones; con tiburones podría necesitarse un equipo que nos logre aislar de los escualos, etcétera).

Un caso maravilloso de observación por equipo altamente especializado lo ofrece el proyecto “Impacto Profundo” de la Administración Nacional Aeronáutica y Espacial (NASA, por sus siglas en inglés). La misión envió una nave espacial para colisionar al cometa *Tempel 1* (lo cual implicó una investigación sumamente avanzada, el diseño de tecnología nunca antes vista, complejos cálculos matemáticos y proyecciones). Entre algunos de los objetivos del proyecto, podemos citar los siguientes: analizar la composición de los cometas y meteoros similares e indagar más sobre los orígenes del universo. La colisión calculada se produjo el 4 de julio de 2005 a 23 000 mph (NASA, 2005). La observación se realizó mediante cámaras fotográficas y de video que se ubicaron en la propia nave *Deep Impact* (nave de colisión) y en la nave *Flyby*, así como mediante algunos de los telescopios más desarrollados del mundo. Un proyecto ideal para la observación del espacio. Además, ahora los seres humanos sabemos que somos capaces de impactar cualquier meteorito o cometa que amenace nuestro planeta. Conocer la composición de estos objetos permitirá diseñar el proyectil adecuado.

Otro caso es la observación permanente de la superficie de Marte mediante la unidad robot rodante *Curiosity* con el fin de estudiar al planeta y sus cambios ambientales, que nos ayudará, entre otras cuestiones, a evaluar la posibilidad de habitarlo y generar teoría sobre nuestro propio cambio climático global (NASA, 2014).

3. *Incluir todo el universo o extraer una muestra representativa de aspectos, eventos o conductas a observar.* Un repertorio que sea suficiente para observar las conductas, eventos o cuestiones sobre las variables de interés.
4. *Establecer y definir las unidades de observación.* Las unidades deben ser captadas por los sentidos. Por ejemplo, observar cada vez que se presenta una conducta agresiva de un determinado paciente; ana-

¹⁴ Este recinto, donde los participantes son colocados para observación, cuenta con dos secciones separadas por un cristal de una sola vista o unidireccional (que permite ver de una sección a la otra, pero no de esta hacia la primera). Una sección es regularmente una sala o mesa de trabajo (desde donde el cristal parece ser un espejo o un vidrio opaco) y la otra, la zona de observación, integrada por butacas, sillas o un aula escolar, en la cual se ve y se escucha todo lo que ocurre en la sala o mesa de trabajo (además, se filma la acción).

lizar cada minuto si el alumno está o no atento a la clase de economía (la materia se imparte dos veces a la semana, dos horas cada clase, de 7:00 a 9:00 horas), etc. El concepto de unidad de análisis es el mismo que en la recopilación y el análisis de contenido, sólo que en la observación cuantitativa se

● **Tabla 7.11** Algunos ejemplos de unidades de análisis o registro.

Estudio	Unidades de observación/medición
Evaluar si un programa de televisión con contenido violento genera conductas agresivas en los niños.	Agresiones físicas de los niños (cada vez que se manifiesta una agresión, como golpear a otro niño, empujarlo intencionalmente, etcétera).
Determinar la conducta de compra en una tienda de ropa por parte de cierto tipo de clientes.	Episodio de compra: en el caso de cada cliente, desde que ingresa a la tienda y hasta que sale de ella: ¿qué prendas o accesorios compró?, ¿de qué colores, texturas?, ¿cuál es su perfil?, etcétera.
Analizar la atención que se presta al tablero de avisos de la compañía.	episodio de interacción: cada vez que un empleado pasa junto al tablero de avisos, se detiene o no a verlo, cuánto tiempo lo observa y qué información capta más su atención.
Evaluar el efecto del consumo de alcohol en el comportamiento agresivo de los jóvenes.	stancia en un bar o cervecería: número de copas o tragos que consumen y número de comportamientos agresivos y sus características (registro progresivo de todo el tiempo de permanencia en el lugar).
Inferir actitudes de fervor religioso en la iglesia.	Cada periodo de tres minutos: registrar conductas que reflejen el fervor religioso (persignarse, arrodillarse, etcétera).
Deducir conductas de aprobación por parte de los pacientes sobre el trato que reciben de las enfermeras.	Manifestaciones de aprobación (cada vez que el enfermo sonríe a la enfermera, la mira afectuosamente, asiente favorablemente, etcétera).
Determinar el apego de los pacientes a ciertas prescripciones de los tratamientos (medicación y ejercicio), mediante filmación.	Conductas de apego (cada vez que el paciente consume el medicamento a la hora prescrita y el número de minutos que se ejercita de acuerdo a lo que el médico le solicitó).
Analizar explosiones solares: magnitud, duración y naturaleza.	La observación mediante diferentes equipos resulta permanente y la unidad es cada explosión (todo su ciclo).

trata de conductas, eventos o aspectos manifiestos. Algunos ejemplos se presentan en la tabla 7.11.

5. *Establecer y definir las categorías y subcategorías de observación.* Estas categorías son similares a las puntualizadas para el análisis de contenido. La observación también consiste en asignar unidades a categorías y subcategorías de observación. En esta, como en las demás técnicas de medición, se asume que las variables (representadas por las categorías y subcategorías) han sido definidas conceptual y operacionalmente de forma adecuada, para que el observador realice correctamente las tareas de discriminación, que es su función central. En la tabla 7.12 se muestran ejemplos.

● **Tabla 7.12** Ejemplos de definiciones de categorías.

Variable	Categorías	Definición
Conducta agresiva hacia los demás	0 a k conductas agresivas	Comportamientos que se manifiesten con la intención aparente de injuriar o dañar a otra persona, ya sea mediante insultos verbales o signos no verbales y acciones físicas.
Conducta de compra (ropa)	0 a k artículos adquiridos	Compra de artículos a crédito o al contado.

● **Tabla 7.12** Ejemplos de definiciones de categorías (continuación).

Variable	Categorías	Definición
Amabilidad en el servicio cuando el cliente tiene alguna necesidad	<ul style="list-style-type: none"> • Amable • Neutral • Grosero 	<p>Amable: atender al cliente sonriéndole y haciendo contacto visual con él. Preguntarle por sus necesidades, resolver sus inquietudes, no interrumpirlo cuando esté hablando, dar solución a las necesidades pertinentes, mantenerse con él hasta que la solución solicitada se implemente, preguntar si requiere algo más e inquirirle si está satisfecho.</p> <p>Neutral: no sonreír al cliente, hacer contacto visual solo una vez o no mantener tal contacto. No demostrarle que se le está escuchando, no permanecer con él hasta que la solución se haya implementado y no verificar su satisfacción con esa solución.</p> <p>Grosero: no sonreír al cliente, mostrarse molesto por sus peticiones, no establecer contacto visual, no preguntarle por sus necesidades, no resolver sus inquietudes, interrumpirlo cuando habla, no solucionar las necesidades pertinentes, no mantenerse con él hasta que la solución solicitada se haya implementado, no preguntar si requiere algo más ni inquirirle si está satisfecho.</p>
Oportunidad en la atención al cliente (en una sucursal bancaria, en una taquilla en el cine, puesto de venta de boletos en un estadio de fútbol)	0 a k minutos y segundos	Tiempo que el cliente espera en fila (línea o cola).

Ejemplo



De unidades y categorías

En el caso del experimento citado en el capítulo 7 (al hablar de la manipulación de variables independientes en experimentos) para probar la hipótesis “A mayor grado de información sobre la deficiencia mental que el sujeto normal maneje, mostrará menor evitación en la interacción con el deficiente mental” (Naves y Poplawsky, 1984). Las unidades de análisis eran lapsos de 10 segundos. La interacción entre la persona normal y el actor que hacía el papel de “deficiente mental” (individuo con capacidad mental distinta) duraba tres minutos. La variable dependiente fue “evitación de la interacción” y las categorías fueron cuatro (Naves y Poplawsky, 1984, pp. 107-109):

1. Distancia física: se refiere a si el participante experimental aumenta o disminuye su distancia hacia el interlocutor a partir de la distancia que inicialmente ocuparía; esta distancia inicial estuvo delimitada por los asientos que el actor y el sujeto debían ocupar y, según la teoría, es la distancia en la que dos extraños en una situación de comunicación pueden interactuar de manera cómoda. Las dimensiones o subcategorías que este indicador adquiere son: acercarse (afiliación) con valor de “2”, mantenerse estático, con valor de “1”; y alejarse (evitación) del actor (deficiente mental), con valor de “0”; mediante inclinaciones corporales, o bien, al modificar por completo su distancia.
2. Movimientos corporales que denotan tensión: esta categoría intenta captar los movimientos que el sujeto realiza: si efectuaba movimientos que indicaran tensión (evitación) el valor era de “0”; ningún movimiento con valor de “1”; y movimientos que señalan que se experimenta relajación (afiliación), con valor de “2”. En este indicador específicamente se analizan si los movimientos de pies y piernas están a un ritmo acelerado, si hay ademanes con brazos y manos (como rascarse, sentir picor, etc.) y la postura en general del participante.
3. Conducta visual del sujeto: según se estipula en esa investigación, este factor adquiere dos dimensiones o subcategorías:
 - a) Dirigida hacia el interlocutor (afiliación) con valor de “1”.
 - b) Dirigida hacia cualquier otra parte (evitación) con valor de “0”.
4. Conducta verbal: este indicador está compuesto por el contenido verbal del discurso del sujeto hacia el deficiente y se orienta primordialmente a la forma del discurso; incluye dos modalidades:

- a) Frases u oraciones completas (afiliación), con valor de “1”.
- b) Frases dicotómicas y silencios (evitación), con valor de “0”.

La modalidad de frases dicotómicas incluye respuestas monosilábicas (sí, no), murmullos, sonidos guturales; los silencios que se interpretan como respuestas dicótomas (de evitación) son aquellos no naturales en el discurso, cuando expresamente el individuo se queda en silencio. La conducta verbal se mide a través del diálogo que sostenga la persona con el deficiente mental; es decir, en respuesta al guión que el actor interpreta ante cada sujeto (que es idéntico para todos) y en las intervenciones que el mismo individuo realice.

En este caso, además de medir la conducta verbal en cuanto a su forma, convendría medirla también en cuanto a su contenido; es decir, determinar si las frases emitidas por el participante en respuesta a lo expresado por el deficiente revelan un contenido positivo o negativo.

Los periodos de observación, las unidades y categorías varían de acuerdo con el planteamiento del problema e hipótesis.

Las subcategorías pueden ser escalas de actitudes

Las subcategorías pueden ser tratadas como escalas del tipo Likert, Guttman o diferencial semántico.

Ejemplo

Categoría		Subcategoría												
Atención del alumno		Elevada (3) Meida (2) Baja (1) Nula (0)												
Disciplina del alumno		Completa (3) Aceptable (2) Media (1) Indisciplina (0)												
0 bien, agresividad del niño	Muy alta (5) Alta (4) Regular (3) Baja (2) Inexistente (1)													
0, al hablar del desempeño laboral observado en coordinadores de escuelas o asociaciones:														
Productividad	<table border="1" style="text-align: center;"> <tr> <td>:-</td><td>:-</td><td>:-</td><td>:-</td><td>:-</td><td>:-</td> </tr> <tr> <td>(5)</td><td>(4)</td><td>(3)</td><td>(2)</td><td>(1)</td><td>(0)</td> </tr> </table>	:-	:-	:-	:-	:-	:-	(5)	(4)	(3)	(2)	(1)	(0)	Improductividad
:-	:-	:-	:-	:-	:-									
(5)	(4)	(3)	(2)	(1)	(0)									
Actitud de servicio	<table border="1" style="text-align: center;"> <tr> <td>:-</td><td>:-</td><td>:-</td><td>:-</td><td>:-</td><td>:-</td> </tr> <tr> <td>(5)</td><td>(4)</td><td>(3)</td><td>(2)</td><td>(1)</td><td>(0)</td> </tr> </table>	:-	:-	:-	:-	:-	:-	(5)	(4)	(3)	(2)	(1)	(0)	Actitud de no servicio
:-	:-	:-	:-	:-	:-									
(5)	(4)	(3)	(2)	(1)	(0)									
Cumplimiento-incumplimiento														
Etcétera														

6. *Diseñar la manera de registrar las observaciones* (mediante un formulario, hoja de codificación o guía de observación).¹⁵ Este instrumento, en el cual se ingresan o anotan los datos, está prediseñado sobre la base de la revisión de la literatura o construido especialmente para el estudio. Su formato es similar al que se presentó en el apartado sobre análisis de contenido. Anastas (2005) indica que en ocasiones se utilizan sistemas de observación probados debido a: a) los costos (desarrollar uno requiere tiempo, varias pruebas y recursos), b) existen diversos instrumentos válidos y confiables para un amplio reper-

¹⁵ Recordemos: en papel o como documento de PC, puede ser una hoja de Excel.

 **Figura 7.12** Formato de registro de conductas inapropiadas.

CONDUCTAS INAPROPIADAS EN EL SALÓN DE CLASE O AULA															
FECHA:								HORA:							
CLASE:								PROFESOR:							
NÚMERO DE ESTUDIANTES:								GRADO:							
Conductas inapropiadas (CI):								Número de estudiantes involucrados (N):							
Agresión física								1-Individuo							
Ofensa verbal _____								2 - Diada _____							
Insulto no verbal (señas) _____								3 - Grupo pequeño (3 a 5) _____							
Hablar en clase _____								4 - Grupo grande (6 a 15) _____							
Ruido provocado								5 - Toda la clase							
Desobediencia															
No enfocado en la clase (distráido)															
1 - Otra															
Estudiantes															
Núm.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CI															
N															
Núm.	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
CI															
N															
Núm.	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
CI															
N															
Núm.	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
CI															
N															

analizar cuáles son las que se presentan con mayor frecuencia y el número de estudiantes involucrados (Creswell, 2005).

Los números (columnas) representan episodios de conducta inapropiada en la clase (cada vez que ocurre). En la fila “CI” (conducta inapropiada) se coloca la letra inicial de la conducta (ejemplo: A, agresión física, O, ofensa verbal; H, hablar en clase...), y “1” si son otras. Asimismo, en la fila “N” se indica el número de estudiantes involucrados en el episodio.

La hoja de codificación, formato o guía debe ser acompañada de las instrucciones para su llenado y los registros, así como la mención de las unidades de observación y el tiempo que durará esta. Debe evitarse un instrumento que sea tan complejo que sea difícil de codificar y utilizarse en la realidad.

7. Determinar el papel de los observadores. ¿Qué tan cercanos deben estar el observador y los observados?

A este respecto mucho se ha escrito y los papeles varían desde una distancia prudentemente lejana hasta una cercanía completa (McCurdy y Uldam, 2014).

Williams (2008), Polster y Collins (2007) y Anastas (2005) plantean dos roles fundamentales de la relación observador-observados: *a*) participante (donde hay relación entre el observador y las personas observadas) y *b*) no participante (la relación es mínima —simplemente por el hecho de que los individuos saben que están siendo observados— o puede ser nula, si los individuos no lo saben).

La ética de la investigación marca que siempre que va a observarse a un ser humano, se debe obtener su consentimiento; asimismo, no se deben invadir sus esferas privadas (baños, recámaras, despachos, etc.). Anastas (2005) está de acuerdo con este principio, aunque marca una excepción: cuando se observan conductas inocuas que pudieran ocurrir de manera natural en situaciones públicas y cuando las personas observadas no son identificadas o afectadas por los observadores. En última instancia, si el investigador trata de evitar la “reactividad” en el proceso (que la presencia de los observadores provoque reacciones poco o nada naturales en los participantes), una vez que se efectúan las observaciones debe solicitársele a cada sujeto su autorización, mantener su nombre en el anonimato, nunca mostrar las filmaciones o registros y respetar su negativa, si es el caso.

Jorgensen (1989) sugiere tres actuaciones: *a*) *participante activo*: el observador interactúa plenamente con los observados, tiene un papel de miembro, es uno más de ellos (por ejemplo, si se trata de observar el compromiso de trabajadores en los grupos de calidad, el observador es un miembro del grupo con funciones específicas), *b*) *participante privilegiado*: no es un miembro del grupo, pero tiene acceso a información amplia y privada; los observados lo conocen y le tienen confianza; y *c*) *observador limitado*; no tiene más que el papel de observador externo. Rogers y Bouey (2005) los clasifican en el siguiente continuo:

Participante completo: total interacción; su rol primario es participar y generar la interacción	Participante observador: su rol primario es participar, pero casi a la par observa	Observador participante: su rol primario es observar, aunque tiene ciertas participaciones	Observador completo: nula participación; se limita a observar
--	--	--	---

McKechnie (2008) y Patton (2001) sugieren que el observador debe ubicarse en varios continuos para establecer su papel (que es una ubicación aproximada y no exacta):

Papel como evaluador-observador		
Observador participante (participa activamente en tareas o situaciones)	Observador que participa sólo parcialmente	Observador no participante (externo)
Papel ante los demás (observados, miembros del ambiente, comunidad o contexto)		
Abierto (los demás conocen su papel y saben que están siendo observados)	Intermedio (algunos conocen su papel, y saben que están siendo observados, y otros no)	Encubierto (nadie conoce su papel ni que está siendo observado)
Duración de la observación (estancia del observador)		
Breve (una hora, un día)		Larga (meses, años)
Enfoque de la observación		
Enfocado (un elemento, un significado, una interacción en especial)		General (holístico: todos los elementos o las unidades)

Sin que de ninguna manera sea una norma, cuando la observación cuantitativa trabaja con personas, regularmente es no participante, encubierta o intermedia, de corta estancia y enfocada, mientras que la observación cualitativa es participante, abierta, prolongada y general.

8. *Seleccionar a los observadores-codificadores.* Puesto que son las personas que codificarán la conducta, deben conocer las variables, categorías y subcategorías. Las habilidades más importantes de un buen observador son:
 - Selectividad para registrar lo que observa, de acuerdo con reglas.
 - Uso de todos los sentidos, particularmente ante lo que no se esperaba encontrar.
 - Capacidad para aplicar el sistema de registro que involucra el procesamiento de información.
9. *Entrenar a observadores-codificadores* (en las variables, categorías, subcategorías, unidades de análisis y el procedimiento para codificar, así como sobre las diferentes maneras en que puede manifestarse una categoría o subcategoría de conducta). Debido a que, en algún grado, la observación involucra hacer juicios, el entrenamiento debe apoyar la estandarización de procedimientos y, por ende, la objetividad.
10. *Efectuar la prueba piloto y calcular la confiabilidad de los observadores* (intraobservador e interobservadores). En la prueba piloto se evalúa todo lo relativo al proceso de observación. Desde la iluminación ambiental (ejemplo, cámara Gesell), el sonido y los equipos de filmación, hasta el lugar donde se coloquen los observadores, las instrucciones y la guía de observación o registro.

Gracias a la prueba piloto podemos detectar problemas potenciales y corregirlos (por ejemplo, si vamos a observar conductas agresivas de niños y nada más hemos trabajado con adultos, nos percatamos de que la agresión no se manifiesta igual en ambos casos. De este modo, durante la prueba piloto podemos registrar comportamientos agresivos no contemplados, que se agregarán como subcategorías en el instrumento de observación definitivo). Esto se hace con una muestra del repertorio de conductas o actos que se considerarán. Con respecto a la confiabilidad, recordemos que debe demostrarse y no simplemente asumirse, y se determina para evaluar si los observadores catalogan o clasifican coherentemente (confiabilidad individual) y si registran entre sí de manera similar los mismos eventos o conductas (confiabilidad interobservadores). Los procedimientos y las fórmulas pueden ser las mismas que las vistas en el apartado sobre el análisis de contenido; lo único que cambia es la palabra “codificadores”, “codificación”, “codificadas”, por “observadores”, “observación”, “observadas”. Por ejemplo:

$$\text{Confiabilidad individual} = \frac{\text{Número de unidades de análisis catalogadas correctamente por el observador}}{\text{Número total de unidades de análisis}}$$

$$\text{Confiabilidad entre parejas} = \frac{\text{Número total de acuerdos entre dos parejas}}{\text{Número total de unidades de análisis codificadas}}$$

Haynes (1978, p. 160) proporciona otra fórmula para calcular la confiabilidad entre observadores o el grado de acuerdo interobservadores (A_o).

$$A_o = \frac{Ia}{Ia + Id}$$

donde Ia es el número total de acuerdos e Id es el número total de desacuerdos entre observadores, respectivamente. Un “acuerdo” se define como la codificación de una unidad de análisis en una misma categoría/subcategoría por distintos observadores. Se interpreta como cualquier *coeficiente de confiabilidad* (0 a 1).

Cuando se trata de establecer la confiabilidad es muy importante evitar que los observadores “copien” o vean el trabajo de sus compañeros, ya que esto puede ocurrir.

Bakeman y Quera (2011); Grinnell, Williams y Unrau (2009), así como Anastas (2005) consideran como una confiabilidad interobservadores mínima aceptable un puntaje de 0.85 u 85%. Si se elige

un instrumento ya desarrollado, al igual que en otros métodos de medición, debe demostrarse que en los estudios donde se aplicó resultó válido y confiable, y adaptarse a las condiciones de nuestra investigación.

Para establecer la validez de contenido se analiza el dominio de nuestras categorías y subcategorías en contraste con el universo de posibles conductas que pueden manifestarse. Muchas veces, la validez de criterio se establece usando otra técnica (por ejemplo, la “hostilidad”, que se infiere de observar actos agresivos, puede validarse mediante una escala que mida tal variable).

1. Llevar a cabo la codificación por observación.
2. Vaciar los datos de las hojas o formatos de codificación y obtener los totales de cada categoría.
3. Realizar los análisis apropiados.

La observación es una excelente herramienta de recolección de los datos en ciertas poblaciones, como los niños y las personas con capacidades distintas. Por ejemplo, Rapport *et al.* (2009) utilizaron la observación para analizar los procesos y patrones de atención en el aula por parte de niños con trastorno por déficit de atención con hiperactividad (TDAH). Sutherland *et al.* (2013) desarrollaron un sistema de observación directa de la interacción del profesor con sus alumnos infantes para evaluar la eficacia del primero y estudiar los patrones de vinculación, así como analizar conductas problemáticas. Asimismo, la observación se utiliza como un método para recolectar datos sobre el comportamiento de los pacientes (desde su conducta social hasta las reacciones ante los tratamientos y sus expresiones). Por ejemplo, Mendez *et al.* (2013) generaron e implementaron un inventario para observar las conductas de enfermos de Alzheimer y demencia frontotemporal.

También puede ser una herramienta en estudios longitudinales. Mejía *et al.* (2009) observaron durante un periodo de poco más de dos años para comprender qué factores inciden en la conducta grupal de un equipo de trabajo comunitario dedicado a la prevención del VIH. Grabaron en video 18 sesiones del grupo.

CODIFICACIÓN EN EL ANÁLISIS DE CONTENIDO Y LA OBSERVACIÓN

Una vez recolectados los datos del análisis de contenido o la observación, todos deben codificarse, es decir, asignar códigos (generalmente numéricos) a las categorías y/o subcategorías. Por ejemplo, en el caso del estudio de Naves y Poplawsky (1984), la codificación es la que se muestra en la tabla 7.13 (que corresponde al primer ejemplo de un formato de registro o codificación, figura 7. 11).

Además, la matriz de SPSS o equivalente sería la de la tabla 7.14 (desde luego, solo se incluyen 10 casos a manera de ejemplo).

RECOPIACIÓN Y ANÁLISIS DE DATOS SECUNDARIOS

La recopilación y análisis de datos secundarios se fundamenta en información disponible recolectada por otros analistas distintos del investigador. Esta técnica para recabar los datos puede subdividirse en: a) datos estadísticos provenientes de fuentes oficiales y b) datos para metaanálisis. En ambos casos se acude a bases de información que pueden encontrarse en una o varias fuentes y que son útiles para los propósitos del estudio (Schmidt y Hunter, 2014 y Vartanian, 2010). La diferencia entre estos dos tipos estriba en que los primeros recogen estadísticas como datos para analizar (por ejemplo, tasas de defunción o mortalidad y nacimiento, número de hectáreas de riego, cifras de egreso en los distintos niveles del sistema escolar, número de camas de los hospitales y pacientes atendidos, número de monumentos históricos por tipo, producto interno bruto, número de instalaciones deportivas y tamaño en metros cuadrados, extensión de la red de alcantarillado, índice de inflación, monto de la deuda externa pública y privada, unidades producidas por una o varias fábricas, digamos de autos, etc.); en cambio, el metaanálisis recopila datos de estudios efectuados por otros sobre el planteamiento del problema, los evalúa y puede volver a analizarlos e integrarlos en una base mayor de información; o bien, sumarlos y compararlos (Ringquist, 2013).

Krysiak (2005) considera tres tipos de datos secundarios: datos obtenidos del nivel “micro”, datos

● **Tabla 7.13** Ejemplo de codificación en el caso del estudio de Naves y Poplawsky (continuación).

Variable	Categorías	Subcategorías	Códigos	Columnas
- Tratamiento experimental	- Grupo cultural		1	1
	- Grupo sociopsicológico		2	
- Conducta	- Distancia física	- Alejamiento	0	2
		- Acercamiento	2	
		- Estático	1	
	- Movimientos corporales	- Tensión	0	3
	- Relajación	2		
	- Ninguno	1		
	- Conducta visual	- Al sujeto	1	4
		- A otra parte	0	
	- Conducta verbal	- Frases completas	1	5
		- Frases dicótomas o silencios	0	
- Codificador	- LRE		1	6
	- MRM		2	
	- APY		3	

Desde luego, Naves y Poplawsky (1984) para las categorías de “conducta de evitación” obtenían esta codificación cada 10 segundos (cada unidad de análisis), y sumaban el número de 2 (dos) y 1 (uno) y lo transformaban en porcentaje. Aquí suponemos que toda la interacción con el supuesto deficiente mental puede categorizarse y subcategorizarse. Recordemos que eran 30 participantes.

● **Tabla 7.14** Matriz o vista de los datos para la codificación de la tabla 7.13 (Naves y Poplawsky).

	Columna 1 Tratamiento experimental	Columna 2 Distancia física	Columna 3 Movimientos corporales	Columna 4 Conducta visual	Columna 5 Conducta verbal	Columna 6 Codificador
S ₁	1	0	0	0	0	1
S ₂	1	0	0	0	0	1
S ₃	2	2	0	1	1	2
S ₄	1	1	0	0	0	2
S ₅	2	2	2	1	1	2
S ₆	1	0	1	1	0	3
S ₇	2	2	2	1	1	3
S ₈	2	2	2	1	1	3
S ₉	2	1	2	0	1	1
S ₁₀	1	2	1	0	0	2

agregados y datos cualitativos micro.

- a) *Datos obtenidos del nivel “micro”*. Son derivados de unidades de análisis individuales como personas, organizaciones u hogares (número de habitantes por vivienda, número de trabajadores de la empresa, preferencias de votantes por partidos políticos y candidatos en un distrito electoral, satisfacción laboral de los trabajadores, etc.). La mayoría de las veces los datos se recolectan por medio de encuestas. Por ejemplo, los censos económicos (vistos en el capítulo de encuestas o *surveys* de este centro de recursos en línea) generan esta clase de datos, al igual que los de población.

La encuesta nacional de adicciones de estudiantes (implementada anualmente en varios países latinoamericanos) es otro caso y recolecta —entre otros datos— los siguientes:¹⁶

1. Personas económicamente activas en el sistema familiar.
2. Distribución del tiempo libre del estudiante.
3. Convivencia del alumno en el hogar (padres y otros parientes).
4. Estructura del sistema matrimonial de los tutores del estudiante (padres casados, divorciados, separados, etcétera).
5. Ocupación de las personas económicamente activas del hogar.
6. Personas con quienes los estudiantes se comunican más dentro del hogar.
7. Actividades deportivas del estudiante (tipos y periodicidad de cada actividad deportiva, tiempos de dedicación, entre otros).
8. Conducta de fumar (inicio, frecuencia de la conducta de fumar, consumo diario de cigarrillos).
9. Ingesta de alcohol (inicio, periodicidad, consumo semanal).
10. Percepción proyectiva del consumo de estupefacientes o drogas ilegales (se responde a la pregunta: *¿A cuántos de tus compañeros de escuela has visto consumir droga?, ¿a cuántos de tus amigos, fuera de la escuela?*).
11. Conocimiento del tipo de droga que se consume (cuestionamiento: *Señala la droga que sabes consumir en este plantel... Señala la droga que sabes se consume en tu vecindario...*).
12. Preferencia de consumo en el entorno inmediato por tipo de estupefaciente (fuera de la escuela).
13. Percepción del encuestado respecto de si es un problema general el consumo de estupefacientes.
14. Percepción respecto de si se pueden conseguir drogas o estupefacientes dentro de la escuela.
15. Percepción respecto de si se consume droga o sustancias ilegales dentro de la escuela.
16. Percepción respecto de si el consumo de estupefacientes es un problema en su entorno.
17. Preferencia de consumo en la escuela por tipo de estupefaciente.
18. Actitud ante el consumo de estupefacientes por una amistad.
19. Manifestación de haber consumido algún estupefaciente.
20. Edad de inicio de consumo.
21. Figura o tipo de persona que invitó a experimentar el consumo (amigo, tutor, etcétera).
22. Tipo de estupefaciente o droga consumida por primera vez.
23. Consumo de diferentes clases de estupefacientes.
24. Preferencia individual de consumo por tipo de estupefaciente.
25. Periodicidad del consumo.
26. Motivos del consumo.
27. Acciones recomendadas por los encuestados para reducir el consumo, evitarlo o prevenirlo.
28. Conciencia del consumo de estupefacientes.
29. Fuente más impactante de conocimiento del problema social que implica el consumo de estupefacientes.
30. Comentarios abiertos (cualitativo).

El investigador, aunque no generó estos datos, tiene acceso directo a ellos (a la información individual original; a los cuestionarios, registros, matriz de datos, etc.) y, por lo tanto, los puede volver a analizar (incluso a veces puede identificar a los encuestados o sus perfiles). Por ética, regularmente sólo debe proporcionarse la matriz de datos y respetar el principio de confidencialidad.

- b) *Datos agregados*. Consisten en datos individuales que ya se han procesado y reunido con otros en cifras estadísticas, tablas o cuadros. En estos, las características de las unidades de análisis no pueden ser identificadas (por ejemplo: el investigador recibe el dato sobre el promedio de satisfacción laboral en una escala o la tasa de desempleo y obviamente no puede reanalizar los datos, sólo utilizarlos para los fines que convengan).

¹⁶ Por ejemplo, en México la publica periódicamente el Consejo Nacional Contra las Adicciones (CONADIC).



Estos datos se pueden obtener de:

1. Oficinas y agencias gubernamentales y oficiales (Ministerio de Educación de Argentina; Ministerio de Educación, Cultura y Deporte de España; Dirección de Impuestos y Aduanas Nacionales de Colombia; Instituto Nacional de Estadística de Bolivia; Ministerio de Desarrollo e Inclusión Social del Perú; Banco Central de Venezuela; Consejo Nacional de Ciencia y Tecnología de México, etcétera).
2. Organizaciones privadas (Cámara de Comercio de Santiago de Chile, Consejo de Fundaciones Privadas de Guatemala, Fundación Cultural Española para el Fomento de la Artesanía, Centro de Estudios Económicos del Sector Privado en México, Fundación Ford, entre otras).
Algunas bases de datos privadas contienen archivos periodísticos, como por ejemplo: U.S. News Archives on the Web (periódicos de Estados Unidos), Reuters y Europrensa (Universidad Complutense de Madrid), International News Archives on the Web (varios países, incluidos diarios latinoamericanos), etc. Los periódicos más importantes del mundo poseen su propia base de datos.
3. Organizaciones no gubernamentales (Agencia de Cooperación en Ingeniería Ambiental o ACIA, Pro Amazonia “Selva Para la Vida”, Alianza Cívica en México, Asociación Bienestar y Desarrollo en España, Asilo de Ancianos de Celaya y otras).
4. Organizaciones internacionales (Comisión Económica para América Latina y el Caribe; Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, UNESCO; Organización Mundial de la Salud, OMS; Fondo de las Naciones Unidas para la Infancia, UNICEF; Fondo Monetario Internacional, etcétera).

Las primeras tres fuentes genéricas varían en cada país, por lo que resultaría poco práctico intentar mencionarlas. Por ello, cada estudiante o profesor debe tener en mente cuáles son y qué tipo de información pueden proporcionar.¹⁷

De los datos resulta fundamental conocer:¹⁸

- El nombre y la filiación institucional de quien los recolectó.
- El periodo de recopilación.
- La muestra (método de selección, tamaño, características...).
- El método de recolección, incluyendo el instrumento, procedimiento, confiabilidad y validez, etcétera.
- Localización (disponibilidad).

Y de ser posible:

- Procedimientos de codificación.
 - Análisis practicados.
 - Sugerencias para el análisis.
- c) *Datos cualitativos micro*. Son producidos por una persona y recabados por otra. Al investigador le llega el dato (cartas recopiladas por otros, transcripciones de entrevistas hechas por psicólogos, grabaciones de las llamadas de los pasajeros que iban en los aviones secuestrados por terroristas en los atentados a las torres gemelas del World Trade Center en Nueva York en 2001, etc.). Tales datos pueden ser reanalizados por el investigador.

Las ventajas de los datos secundarios residen en que su costo es generalmente el más barato de todos los métodos de recolección y a veces son la única opción para el investigador (Smith, 2008; Sieppert, McMurtry y McClelland, 2005 y Encyclopedia of Evaluation, 2004); además, pueden replicarse estudios con base en ellos. Gran cantidad de datos secundarios se localiza mediante internet. Sus desventajas potenciales consisten en que a veces no son accesibles, no están estandarizados,

¹⁷ Para el caso de España recomendamos a Berganza y García (2005) y a Corbetta (2003) y desde luego una búsqueda en los disparadores genéricos de internet. Para las organizaciones no gubernamentales latinoamericanas, por ejemplo, en el caso del medio ambiente, la siguiente dirección en internet incluye una gran cantidad de ellas: <http://www.ecoportail.net/content/view/full/158>.

¹⁸ Se aplica también a los datos “micro”.

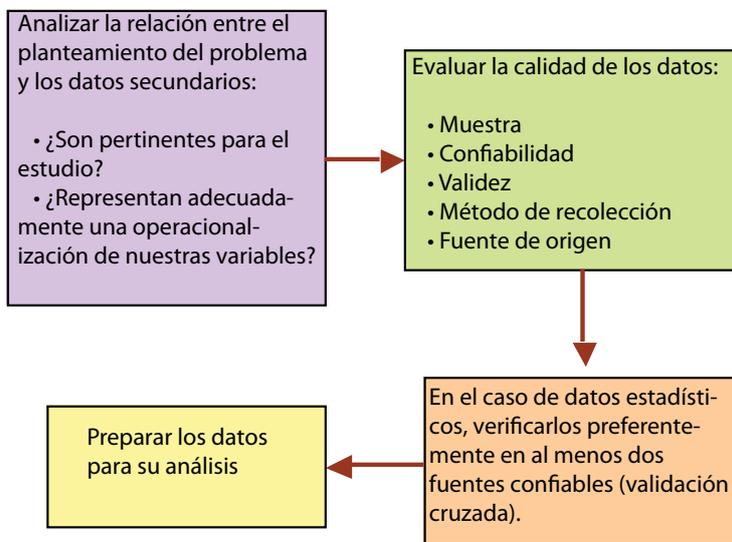
pueden haberse omitido variables importantes para nuestro estudio y es factible que den una falsa impresión de ciertos subgrupos (los promedios de ingreso nacional per cápita, por ejemplo, suelen resultar engañosos).

Corti (2008) presenta las siguientes dificultades y retos cuando se deben reutilizar y compartir datos secundarios:

1. Consideraciones y limitaciones que impone la necesaria ética al reportar datos.
2. Cobertura y representación de los datos y su contexto.
3. Desconocimiento de los métodos usados para recolectarlos.
4. Falta de tiempo para familiarizarse con los datos generados.
5. Falta de infraestructura para acceder a los datos de interés.
6. Interpretación incorrecta del sentido de los datos.
7. Violación de los derechos de autor y propiedad intelectual.

El proceso para utilizar datos secundarios en una investigación se resume en la figura 7.13.¹⁹

● **Figura 7.13** Proceso esencial para utilizar datos secundarios.



Para el manejo de archivos se recomienda consultar a Webb, Campbell y Schwartz (2000).

El metaanálisis

Esta clase de procedimiento de recolección y análisis de datos implica sintetizar resultados de investigación por medio de varios estudios que consideran el mismo planteamiento y variables (Schmidt y Hunter, 2014 y Wiersma y Jurs, 2008). El término metaanálisis fue acuñado en 1976 por el estadístico norteamericano Gene V. Glass, mientras era profesor de la Universidad de Colorado en Boulder (Encyclopedia of Health Services Research, 2009).

Los pasos de un metaanálisis son muy similares a los de otros tipos de indagación, con la salvedad de que la unidad de análisis ya *no* es el participante, evento o caso, sino el estudio realizado por otros investigadores. El reto es resumir los datos a partir de los diferentes estudios de manera muy bien estructurada, que evite errores y facilite su ulterior análisis.

Cabe señalar que el metaanálisis será más preciso entre más homogéneos sean los estudios que se consideran, en cuanto a la preguntas de investigación, variables, método, instrumentos de medición y análisis estadístico. Por lo tanto, el primer paso consiste en ponderar si es posible combinarlos (nos referimos solo a estudios cuantitativos). Tal consideración nos puede llevar a un sinfín de opciones, desde

¹⁹ Adaptado de Corti (2008), Sieppert et al. (2005) y Krysik (2005).



descubrir que hay varias investigaciones que son totalmente compatibles y pueden mezclarse, hasta encontrar que algunas solo tienen un ítem en sus cuestionarios o un indicador cuyos resultados nos sean de utilidad.

Existen diferentes propuestas para homogeneizar estadísticamente los resultados de distintos estudios, entre las que se destacan tres, que se conocen con el nombre de modelo de efectos fijos, modelo de efectos aleatorios y, por último, el modelo bayesiano (Molinero, 2003).

En el *modelo de efectos fijos*, los estudios se combinan y se asume que *no* existe heterogeneidad entre ellos; por lo tanto, todos constituyen estimaciones de un efecto real cuya magnitud se desea conocer. Así pues, las inferencias realizadas se condicionan a las investigaciones incluidas.

En el *modelo de efectos aleatorios* las inferencias se fundamentan en suponer que los estudios seleccionados para el análisis constituyen una muestra aleatoria del universo de estudios posibles, y sus resultados son más conservadores pues tienen en cuenta una fuente extra de variación, ya que se incluyen dos fuentes potenciales: la existente *dentro* de los estudios y la variación *entre* ellos (Molinero, 2003, p. 3).

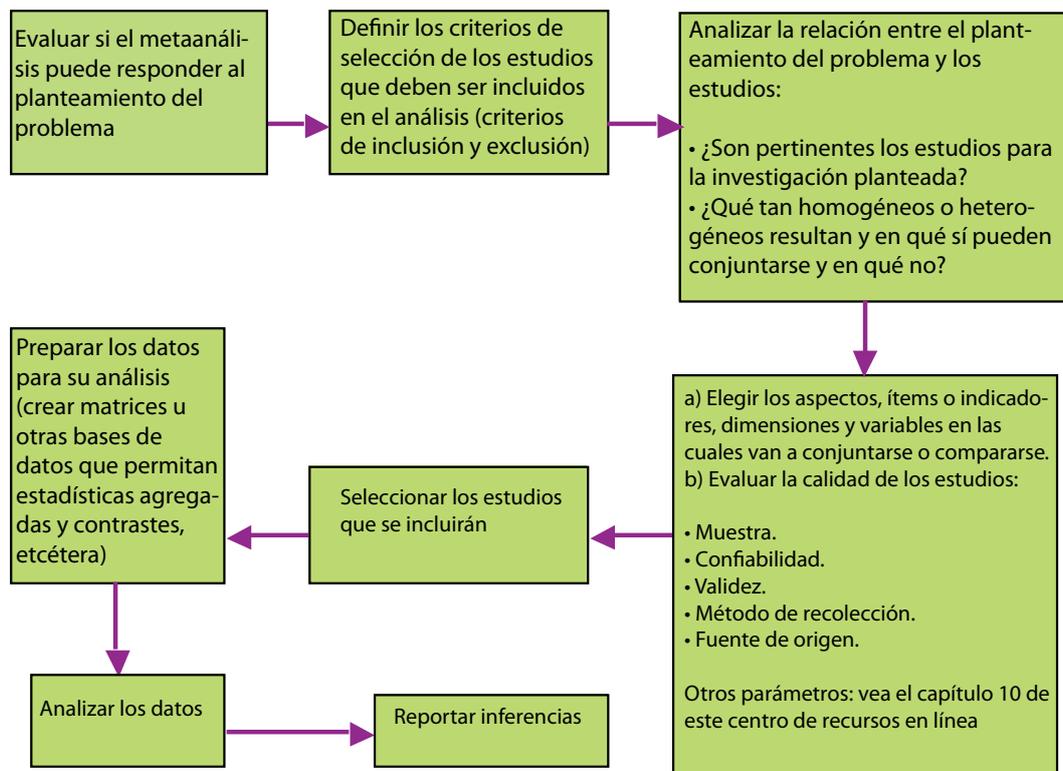
La utilización de modelos bayesianos requiere de modelos estadísticos complejos de varianza.

El metaanálisis constituye un intento de resolver contradicciones entre los resultados de diferentes investigaciones y evaluar patrones comunes en distintos estudios. Es un método sumamente importante para generar teoría.

Como todo método inferencial, tiene sus dificultades y limitaciones. Uno de los principales retos es que al ser un enfoque que generalmente se ejecuta de forma retrospectiva, tiene una posibilidad mayor de introducir sesgos (error sistemático). Otros desafíos son los que se derivan del procedimiento de combinar datos de múltiples estudios, entre los cuales puede haber diferencias en cuanto a método y muestreo (Card, 2011 y Smith, 2008).

El procedimiento general de un metaanálisis es similar al de los datos secundarios, pero se agregan algunos elementos como el análisis mismo (figura 7.14).

● **Figura 7.14** Proceso esencial del metaanálisis.



Algunas recomendaciones para el metaanálisis las proporcionan Schmidt y Hunter (2014), Harpe (2009) y Céspedes (1995):

1. Realizar una cuidadosa revisión de la literatura de los estudios que son adecuados potencialmente para integrarse como unidades de análisis.
2. Generar parámetros precisos de inclusión y exclusión de estudios, con índices de calidad para que los estudios elegidos sean los apropiados.
3. Seleccionar investigadores independientes que revisen cuidadosamente, clasifiquen, codifiquen, cuantifiquen y finalmente evalúen y elijan el grupo de estudios que será incluido en el metaanálisis (MA). Primero, cada investigador valora de manera autónoma, mientras que la selección final se hace en equipo.
4. Evaluar minuciosamente la homogeneidad-heterogeneidad de las unidades o investigaciones a considerar.
5. Elaborar una guía de los datos de los estudios escogidos que deben ser considerados para ser procesados mediante el metaanálisis.
6. Combinar los resultados obtenidos y asegurar la calidad de los datos y su procesamiento estadístico adecuado (puede ser en matrices, procedimientos estadísticos, bases de datos comparativas, etcétera).
7. Preferentemente utilizar datos provenientes de muestras aleatorias.

El análisis estadístico general del modelo de efectos fijos (*no* existe heterogeneidad, hay homogeneidad o las investigaciones que se estiman muestran un efecto uniforme, debido al uso de un mismo tratamiento), utiliza el método de Mantel-Haentzel-Peto.

Si los estudios son heterogéneos al mostrar el efecto ante un tratamiento (modelo de efectos aleatorios), se utiliza el método de Cochran-Dersimonian-Laird (Céspedes, 1995).

En general, se obtiene para las estadísticas calculadas, el tamaño del efecto (Nugent, 2009; Wiersma y Jurs, 2008 y Glass, 1976), el cual se revisó en el capítulo 10 del libro impreso. Recordamos que es una medida de la “fuerza” de la diferencia de las medias u otros valores considerados (Creswell, 2013a). Resulta ser una medida en unidades de desviación estándar. Su cálculo fue ejemplificado con la prueba *t*.

Existen diversos programas computacionales para el metaanálisis, entre los cuales se encuentran los siguientes:

- RevMan®: The Cochrane Collaboration’s Review Manager. The Cochrane Collaboration (tiene versión en español).
- BUGS y WinBUGS.
- SAS® (programa ya comentado, en su página existe un subprograma para metaanálisis).
- STATA®.
- SPSS®, entre otras aplicaciones del programa SPSS MACRO.
- MetaWin.

Y otros muchos que puede buscar vía disparador (o motor de búsqueda) en internet.

EJEMPLOS DE ESTUDIOS BASADOS EN DATOS SECUNDARIOS

Shu-li *et al.* (2008) efectuaron una investigación para desarrollar un modelo que predijera la permanencia de personas ancianas en su hogar (en lugar de acudir a asilos, por ejemplo), utilizando datos secundarios de 9 879 individuos de la base de información pública de Estados Unidos: “The National Home and Hospice Care Survey 2000 (NHHCS 2000)”. Descubrieron que el principal predictor de la permanencia es que la persona viva con familiares. Asimismo, Smith (2009) se fundamentó en análisis de datos secundarios para evaluar las inequidades en la educación británica entre escuelas.

Hernández-Sampieri *et al.* (2008) realizaron un estudio de mercado utilizando datos secundarios con la finalidad de indagar sobre la conveniencia de abrir clínicas de cirugía estética y plástica en México. Mediante búsquedas en diversas fuentes de internet, medios impresos (revistas y diarios de circulación nacional) y bases de datos, lograron ubicar a las principales clínicas que operaban a nivel nacional y en las regiones y ciudades más pobladas del país (Ciudad de México, Monterrey, Guadalajara, Tijuana y Puebla),



establecer los precios de sus servicios, determinar el mercado potencial y sus características (por ejemplo, número de cirugías que se practicaron durante el último año, interés por practicarse una cirugía plástica, preocupación por la estética personal...), evaluar las apelaciones de su publicidad y otras cuestiones. Una encuesta hubiera arrojado resultados más precisos, pero también hubiese sido más cara, de tal modo que se hizo una investigación de acuerdo con el presupuesto del inversionista español. Incluso, los autores le proporcionaron un análisis del marco jurídico mexicano para abrir y operar tales clínicas.

En derecho, el análisis de datos secundarios es muy importante, porque se consideran resoluciones judiciales o juicios previos y la indagación va mucho más allá de una simple revisión de antecedentes.

El metaanálisis es muy socorrido en todos los campos. Tal es el caso de las ciencias de la salud. En epidemiología, por ejemplo, la proliferación de estudios con resultados contradictorios puede afrontarse con metaanálisis, como ocurre con los informes conflictivos sobre el efecto de la dieta en el riesgo del cáncer de mama (Thacker y Stroup, 2007). Asimismo, para determinar la eficacia de un procedimiento quirúrgico o contrastar resultados discordantes de experimentos clínicos (Davis, Li y Leucht, 2009).

Lo y Lo (2013) implementaron un metaanálisis que incluía a 24 estudios para determinar la eficacia de la enseñanza a través del idioma inglés como método de instrucción. Por su parte, Lin, Niddam y Hsu (2014), a partir de considerar que la resonancia magnética funcional ha sido ampliamente usada para investigar la representación del cerebro asociada con el dolor dental evocado por la estimulación pulpal eléctrica y que los estudios experimentales han sido heterogéneos y con muestras pequeñas, además de que la representación común del cerebro sobre el dolor dental sigue siendo esquiva, efectuaron un metaanálisis sobre seis investigaciones relacionadas con el dolor dental ($n = 87$) y sometieron a prueba tres hipótesis: 1) el dolor dental está asociado a la red del “núcleo” relacionado con el dolor; 2) la activación cerebral vinculada con el dolor es organizada somatotípicamente en la corteza somatosensorial; y 3) el dolor dental se encuentra asociado con la red cognitiva-afectiva relacionada con el dolor.

Sofi *et al.* (2014) llevaron a cabo un metaanálisis sobre la correlación entre el insomnio y los padecimientos cardiovasculares, incluyendo en su muestra a 13 estudios ($n = 122\ 501$ individuos).

Dos de los colaboradores de esta edición, Roberto Hernández-Sampieri y Sergio Méndez Valencia se encuentran en proceso de realización de un metaanálisis sobre estudios que vinculan empíricamente el clima laboral y la cultura organizacional, a partir de investigaciones usadas como ejemplos en la quinta y sexta ediciones de *Metodología de la investigación*.

Para profundizar en el metaanálisis en la sociología se recomienda a Roelfs *et al.* (2013) y un ejemplo en turismo puede leerse en Castro-Nuño, Molina-Toucedo y Pablo-Romero (2013). Asimismo, para detallar los aspectos de muestreo en el metaanálisis se sugiere a Valentine, Pigott y Rothstein (2010), y para errores de muestreo a Marín-Martínez y Sánchez-Meca (2010).

Estudio para el diagnóstico municipal

En las anteriores ediciones de esta obra se integraba un ejemplo para diagnosticar y caracterizar a un municipio con el fin de elaborar su plan de desarrollo. Regularmente, estos diagnósticos son estudios mixtos, pero una parte muy importante de sus insumos lo constituyen los datos secundarios, que se generan con base en indicadores, los cuales han sido desarrollados por diferentes organizaciones dedicadas a la investigación municipal (por ejemplo: el Instituto Vasco de Estadística; el Centro de Investigación y Desarrollo Económico de México (CIDE); el Instituto de Estadística de la Comunidad de Madrid; el Sistema Nacional de Indicadores Municipales (SINIM) de la Unidad de Información Municipal del Gobierno de Chile; el Instituto Nacional de Estadística, Geografía e Informática en México (INEGI); el Instituto Brasileiro de Geografía y Estadística; el Sistema Nacional de Información Municipal, CEDEMUN, de la Secretaría de Gobernación de México; la Federación Nacional de Departamentos de Colombia; el Departamento de Estadística del Ayuntamiento de Barcelona; etcétera).

Algunos indicadores se aplican a cualquier municipio y son relevantes para todas las entidades poblacionales (número de habitantes, ingreso per cápita, ingresos totales del municipio, gasto total del municipio, inversión total, etc.), mientras que otros indicadores resultan importantes sólo para ciertos municipios (por ejemplo, los indicadores relativos a la producción agrícola son irrelevantes para un municipio completamente urbano). Por ello, en cada diagnóstico municipal deben elegirse los indicadores (datos secun-

darios) significativos para la entidad en cuestión.

También, cada provincia, departamento, región o estado genera indicadores a su nivel, lo mismo que las naciones.

En este centro de recursos en línea: Material complementario ® Documentos ® Documento 3: “Indicadores”, se incluyen ejemplos de indicadores secundarios municipales que pudieran recolectarse para realizar un diagnóstico de un municipio, con el fin de que los estudiantes que apenas comienzan en estos menesteres de la metodología de la investigación los puedan comprender sin mayores complicaciones (se cubren unas cuantas áreas, no su totalidad ni mucho menos). Tampoco son los únicos, ya que hay miles de ellos. Están agrupados por áreas generales (las cuales también pueden fluctuar según el autor u otros criterios). Determinados indicadores pueden formar parte de dos o más áreas. Sin polemizar (porque cuando se sopesan pueden generar discusiones severas), se presentan en el documento mencionado y fueron en particular desarrollados para un municipio del Estado de México. Sería interesante como ejercicio que en cada institución se propusieran los que serían los más adecuados para su municipio.

Errores en las mediciones

Esta parte se escribe porque es una pregunta que se nos ha hecho constantemente sobre los tipos de errores en la medición.

Cualquier instrumento de medición tiene un grado de error (recordemos lo señalado en el capítulo 9 del libro y al inicio de este capítulo).

Imaginemos que realizamos la evaluación de la autoestima en un grupo de estudiantes. Si la mido un día y la vuelvo a medir al día siguiente, y posteriormente a la semana, podemos esperar que las puntuaciones sean las mismas en las tres mediciones. La autoestima como otros atributos *no* varían a través del tiempo, a menos que “algo” los haga variar (un estímulo, la maduración de la persona, un suceso de vida, entre otros). Es decir, si los estudiantes obtuvieron puntuaciones estadísticamente diferentes en las pruebas, su desempeño y resultados se debieron a la influencia de otro motivo distinto a su propia autoestima. Estas otras razones provocan error. El grado en que un instrumento se encuentra libre de error indica su confiabilidad. Entre más fiable resulta un instrumento, el investigador puede acercarse a una verdadera estimación del atributo considerado. Tal “exactitud” se obtiene al minimizar las fuentes de error en la medida de lo posible y con la estimación del grado de error que permanece. Pues bien, hay dos tipos de errores que pueden ocurrir: sistemáticos y no sistemáticos. Los primeros maximizan o minimizan el desempeño de un instrumento y afectan la confiabilidad y la validez (Mertens, 2010). En el ejemplo de la autoestima, un método para desarrollarla o fortalecerla (por ejemplo, motivacional o terapéutico) puede ser visto como una influencia sistemática en la variable o atributo. El efecto del error sistemático en las respuestas es constante y puede predecirse. Pero los errores *no* sistemáticos (que conciernen a los investigadores) varían de situación en situación (aplicaciones distintas) y en consecuencia *no pueden predecirse* (o es sumamente complejo), pero *sí prevenirse*.

Estos errores pueden caer en tres categorías: *a*) personales (dependen de cada individuo, como el estrés, el cansancio o la motivación), *b*) condiciones de administración del instrumento (instrucciones distintas a quienes responden, temperatura ambiental, lenguaje, etc.) y *c*) modificaciones en los instrumentos o tareas (por ejemplo, cambios en ítems o las conductas a observar). Normalmente, la confiabilidad es calculada mediante una estadística que compara el desempeño de los mismos individuos en tiempos diferentes o partes distintas del instrumento y que oscila entre cero y uno, su fin es precisamente ayudar al investigador a identificar el grado de error que se presenta en la medición, en su consistencia interna; mediante la fiabilidad y un análisis del investigador se deben hacer transparentes las fuentes de error.