

BIOESTADÍSTICA

Alfredo de Jesús Celis de la Rosa Vanessa Labrada Martagón

3ª edición

$$\operatorname{tg}(\alpha - \beta) = \frac{\operatorname{tg}\alpha - \operatorname{tg}\beta}{1 + \operatorname{tg}\alpha \operatorname{tg}\beta}$$



Manual Moderno

1 Características de los datos

$$\sum_{i=1}^N x_i$$

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

$$\frac{\sum_{i=1}^N x_i}{N}$$

CARACTERÍSTICAS DE LOS DATOS

Los datos con los cuales se realizan los trabajos estadísticos varían entre individuos y se obtienen a partir de elementos que en conjunto conforman un universo.

Con frecuencia **dato** e **información** se utilizan como sinónimos. Sin embargo, por información entendemos los datos procesados de manera significativa para el receptor, con valor real y perceptible para tomar decisiones presentes y futuras, los cuales se nos presentan en forma de indicadores. La información así planteada se obtiene como resultado o producto del proceso que se muestra en la **figura 1-1**.

Los datos no son útiles o significativos como tales, sino hasta que son procesados y convertidos en información. De alguna manera, la información es el conocimiento derivado del análisis de los datos. Ésta es la diferencia básica entre datos e información. Hay que hacer notar que la información obtenida en un proceso puede servir como dato para otro proceso.

Universo

En estadística **universo** o **población** se definen como el conjunto de valores por los cuales existe algún interés. El total del universo o población se representa con la letra mayúscula *N*.

Las poblaciones pueden definirse especificando una regla (o reglas). Éstas pueden ser: características de individuos, límites geográficos, grupos ya existentes, límites de tiempo, etc. Por ejemplo: residentes de Guadalajara, asistentes a un paseo escolar, derechohabientes del IMSS, enfermos de cólera.

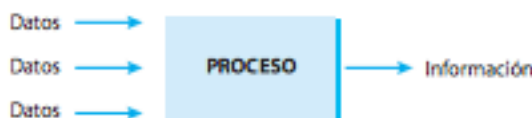


Figura 1-1. Generación de información a partir de datos.

Elementos del universo

Los **elementos del universo** pueden ser personas, lugares o cosas, ya sean individuos únicos o agrupados. Por ejemplo: los pacientes encamados son elementos que conforman parte del universo definido como hospital, pero también el personal, el mobiliario y los diversos servicios que en él se prestan pueden ser elementos del mismo conjunto. Para individualizarlo, cada elemento de la población se identifica con un número progresivo, que inicia en 1 y termina en *N*. Esta identificación de cada sujeto se representa como un subíndice y se le conoce como valor “*i*” o valor *i*-ésimo.

VARIABLES

Normalmente, el interés del investigador se dirige a las características de los elementos que conforman el universo. A dichas características se les designa como **variables**. Ejemplos de variables pueden ser: talla, peso, sexo, temperatura corporal, condición social y escolaridad. Si el valor de la variable no puede predecirse con anticipación, se le denomina **variable aleatoria** y para representarla se utilizan letras mayúsculas (*X*, *Y*, *Z*). Así pues, la variable aleatoria “edad” se puede representar con la letra *X* y las variables aleatorias “sexo” y “escolaridad”, con *Y* y *Z*. Los valores individuales de una variable aleatoria se representan con letras minúsculas (*x*, *y*, *z*) y un subíndice *i*-ésimo que identifica el elemento del conjunto que posee la característica.

Ejemplo explicativo 1-1

Un investigador está interesado en los valores de hemoglobina en sangre de los trabajadores de un taller dedicado al mantenimiento de automóviles. Para ello, decide estudiar a todos los trabajadores que laboran en el taller. Después de recabar los datos de interés, decide organizarlos de la manera que se presenta en el **cuadro 1-1**.

De acuerdo con los datos reportados, el universo se compone de cinco trabajadores; por tanto, *N* = 5. En el **cuadro 1-1**, el investigador representa la variable “sexo” con la letra *X*, la variable “edad en años cumplidos” con la letra *Y*, y la variable “hemoglobina en sangre” con la letra

Cuadro 1-1. Algunas características de trabajadores que laboran en un taller de automóviles

Nombre del trabajador	Número progresivo que identifica al trabajador	Sexo	Edad en años cumplidos	Hemoglobina en sangre (mg/dL)
	i	X	Y	Z
Miguel Domínguez	1	Masculino	45	12
Domitila Hernández	2	Femenino	23	13
Manuel Benítez	3	Masculino	32	11
Jesús Ortiz	4	Masculino	18	15
Sergio Martínez	5	Masculino	21	14

Z. Identifica a Miguel Domínguez con el valor *i*-ésimo 1, "sexo" x_1 = masculino, "edad en años cumplidos" y_1 = 45 y "hemoglobina en sangre (mg/dl)" z_1 = 12. Para el caso de Domitila Hernández, los valores x_2 , y_2 , z_2 son femenino, 23 y 13, respectivamente. Lo mismo se hace con el resto de los trabajadores hasta completar todos los elementos que conforman el universo definido por el investigador.

sificar los elementos en estudio. Cuando se trata de variables cuantitativas la decisión suele ser sencilla, pero no cuando la variable es cualitativa o cuando una variable cuantitativa se presenta agrupada. En cualquier caso, se espera que la escala de la variable permita clasificar a todos los elementos, sin excepción, y que cada elemento sea contado sólo una vez en relación con esa variable: en otras palabras, la escala ha de ser exhaustiva y excluyente.

Definición operativa

Todas las variables que se utilicen en cualquier trabajo de estadística han de ser definidas con claridad, de tal manera que se eviten confusiones, se facilite la búsqueda y análisis de los datos, y se garantice la comparabilidad de los resultados con los obtenidos en otros estudios previos o posteriores. Esto es particularmente importante cuando las variables pueden definirse de maneras diferentes.

Ejemplo explicativo 1-2

El sarampión es una enfermedad viral caracterizada por síntomas prodrómicos (fiebre, conjuntivitis, coriza, tos y manchas de Koplik en la mucosa bucal) que del tercer al séptimo días presenta en cara erupción exantemática que se generaliza al resto del cuerpo y desaparece de 4 a 7 días después. Durante el periodo de incubación, enfermedad y convalecencia, se presentan modificaciones inmunes características del proceso morboso. El término "enfermo de sarampión" puede definirse operativamente de diversas maneras, entre las que se pueden encontrar las siguientes:

- Pacientes con manchas de Koplik en mucosa bucal.
 - Pacientes con exantema maculopapular de tres o más días de duración, fiebre y cualquiera de las tres siguientes: tos, coriza o conjuntivitis.
 - Pacientes con IgM específica para virus del sarampión.
- La selección de una de éstas como definición operativa de sarampión tiene implicaciones importantes en el desarrollo de la investigación. El equipo de trabajo tendrá que utilizar una de ellas, u otra que se adapte a sus necesidades, y precisarla de tal manera que quienes conozcan su trabajo sepan a qué se refiere con el término "enfermo de sarampión".

Una vez que la variable ha sido definida operativamente, se debe especificar la escala de valores que se utilizará para cla-

Típos de variables

Las variables se pueden clasificar en cualitativas y cuantitativas. Cuando sus características se expresan como categorías se dice que se trata de variables cualitativas, mientras que cuando se expresan como valores se les identifica como variables cuantitativas.

1. Las **variables cualitativas** proporcionan **datos nominales** (en los que se tiene, o no se tiene, la característica de interés) y **datos ordinales** (en los que la característica es graduable).
 - a) Una **variable nominal** es aquella cuya característica se define por un nombre y no implica ser más o menos que la característica definida por un nombre diferente. Por ejemplo, "sexo" es una variable nominal, ya que ser "masculino" no significa ser más o menos que "femenino"; la ocupación también es una variable nominal, ya que ser "ingeniero" o "abogado" no significa ser más o menos que "médico". Aquellas variables nominales que se conforman de dos categorías (nacional, extranjero; con diarrea, sin diarrea, etc.) se designan como dicotómicas.
 - b) Las **variables ordinales** son aquellas cuyas características pueden recibir algún orden subjetivo. Su característica principal es que, al ser clasificadas de alguna manera, se puede asumir que se es más o menos que las otras, aunque se desconozca qué tanto más o qué tanto menos. En relación con el dolor, por ejemplo, el paciente puede decir que le duele "poco" o "mucho" y quien lo interroga puede asumir con seguridad que "mucho" significa más dolor que "poco", aunque no se podría saber qué tanto es "mucho" ni qué "distancia" existe entre "poco" y "mucho". Aun se podría asignar-

le una graduación subjetiva más detallada y no por eso dejaría de ser ordinal. Tal es el caso cuando se le pide al paciente que ubique, entre el "0" y el "100", qué tanto dolor siente. En este caso, al igual que al clasificarlo como "mucho" o "poco", se le asigna un orden en el cual "25" significa más que "15" y menos que "35", pero se desconoce qué tan grande es la distancia entre "15" y "25" y no se puede asumir que sea la misma que existe entre "25" y "35", ni que el "15" en una persona corresponda al "15" en otra. Las variables ordinales, al igual que las nominales, también pueden dicotomizarse sin que cambie su escala de medición. Por ejemplo, al definir el comportamiento como "bueno" o "malo" se divide la variable en dos categorías, en las cuales "bueno" representa algo más deseable (o indeseable) que "malo".

2. Las **variables cuantitativas** también permiten diferenciar entre los individuos, pero además señalan cuán grandes son las diferencias observadas. Las observaciones cuantitativas brindan **datos discretos** (en los que sólo se admiten valores individuales en números enteros) y **datos continuos** (en los cuales es posible un número infinito de fracciones entre dos puntos de la escala).

a) Se definen como **variables discretas** aquellas cuyas valores en la escala están separados entre sí por una cantidad determinada, por ejemplo, el número de consultas otorgadas por médico en un día o el conteo de linfocitos en sangre. A diferencia de las variables ordinales, la "distancia" absoluta entre 5 y 7 consultas es la misma que entre 105 y 107 consultas, y también se puede saber qué tantas más son 100 consultas en relación con 10 consultas. Un rasgo distintivo de estas variables es que la unidad no puede fraccionarse, porque pierde su naturaleza. Así, si se parte por la mitad un paciente, éste deja de serlo para convertirse en dos mitades de cadáver.

b) Las **variables continuas** son aquellas en las cuales la escala de medición se puede dividir en una cantidad infinita de valores entre dos puntos cualquiera. Entre éstas se encuentran las medidas de longitud, peso, tiempo y volumen. Por ejemplo: entre 0 y 100 metros existe un número infinito de valores que pueden caracterizar al elemento en estudio, pero también es infinito el número de valores que se encuentran entre 0 y 10 metros o entre 0 y 10 centímetros. Al igual que con las variables discretas, la "distancia" absoluta entre dos puntos se mantiene a lo largo de la escala. Algunas variables parecen no respetar la última característica; tal es el caso de las titulaciones que se reportan como 1:1, 1:2, 1:4, 1:8, etc. Sin embargo, mediante una transformación matemática (logarítmica para el caso) se puede observar que sí se apegan a las características señaladas para las variables continuas.

Las variables cuantitativas también pueden clasificarse según tengan o no en su escala un valor de cero absoluto en variables de **intervalo** (no tienen cero absoluto) y variables de **razón** (sí lo tienen). Por ejemplo, la temperatura medida en grados centígrados tiene un valor de 0° C, pero éste es arbitrario y no es un valor absoluto porque existen otros valores por debajo

de esa temperatura. La edad en años cumplidos, por otra parte, sí tiene un valor de 0 absoluto, porque nadie llega a tener menos de cero años de edad. La temperatura en grados Kelvin también tiene una escala de razón, porque no existe una temperatura por debajo de 0° Kelvin. La distinción de estas dos escalas resulta importante para la interpretación de una razón. Por ejemplo: supongamos un niño que pesaba 50 kg y ahora tiene 60 kg. Podemos decir correctamente que el aumento del peso fue de 20%. Por otra parte, si un líquido que tenía una temperatura de 50° C llega a los 60° C no podremos afirmar lo mismo, porque la escala de temperatura en grados Celsius no tiene un cero absoluto.

Ejemplo explicativo 1-3

Una investigadora estaba interesada en identificar algunas condiciones de la madre que pudieran relacionarse con el bajo peso de los niños al nacer. Para ello, decidió estudiar las variables que se muestran en el **cuadro 1-2**, en el cual las variables se clasifican según su escala de medición.

Regularmente, la clasificación de las variables es una tarea sencilla, pero en ocasiones genera algunas dificultades menores. Por ejemplo, el tiempo es una variable continua porque entre un instante y otro cualquiera existe una cantidad infinita de divisiones, pero la edad (al menos como frecuentemente se registra) es una variable discreta: un niño que cumple 10 años sigue reportando la misma edad durante todo el año hasta su siguiente cumpleaños, fecha a partir de la cual empezará a decir que tiene 11 años. Por otra parte, el número de moléculas de glucosa (o de cualquier otra sustancia) es una variable discreta, porque al dividirse la molécula deja de ser glucosa, pero la concentración de glucosa en sangre es una variable continua, porque la dilución (mg/dl) está medida en una escala que tiene un número infinito de divisiones.

Cuadro 1-2. Algunas características de mujeres durante el embarazo

Variable	Escala de medición
· Edad de la madre (en años cumplidos)	Discreta, de razón
· Talla de la madre (en centímetros)	Continua, de razón
· Estado civil (soltera, casada, unión libre, otro)	Nominal
· Escolaridad (menos de primaria, primaria, secundaria, preparatoria o más)	Ordinal
· Temperatura (en grados Celsius)	Continua, de intervalo
· Ocupación durante el embarazo (según la Clasificación Mexicana de Ocupaciones)	Nominal
· Exposición al humo de tabaco durante el embarazo (no, fumadora pasiva, fumadora activa)	Ordinal
· Hemoglobina en sangre (mg/dl) en la primera consulta prenatal	Continua, de razón
· Número de consultas prenatales antes del tercer trimestre del embarazo	Discreta, de razón

Ejercicios

1. En las variables que se presentan a continuación, indique de qué tipo de variable se trata mediante las siguientes abreviaturas: N para una variable cualitativa nominal, O para una variable cualitativa ordinal, D para una variable cuantitativa discreta y C para una variable cuantitativa continua.

Variable	
1. Concentración de triglicéridos en sangre (mg/dl)	()
2. Diabetes (sí, no)	()
3. Colesterol total (mg/dl)	()
4. Creatinina ($\mu\text{mol/L}$)	()
5. Número de cigarrillos fumados el día anterior	()
6. Índice de masa corporal (peso/talla ²)	()
7. Número de consultorios en la clínica	()
8. Presión arterial diastólica (mm Hg)	()
9. Intensidad del dolor (escala de 0 a 10)	()
10. Opinión del servicio (bueno, regular, malo)	()
11. Edad en años cumplidos	()
12. Uso de drogas antihipertensivas (sí, no)	()
13. Presencia de cefalea (sí, no)	()
14. Sexo (masculino, femenino)	()
15. Porcentaje de grasa corporal	()
16. Depósito de agua (aljibe, balde, alberca, tina, otro)	()

2. Observe el cuadro que se le presenta a continuación. Los datos que contiene se refieren a las características de seis pacientes.

Identificación (número progresivo)	Índice de masa corporal (IMC)	Colesterol HDL	Glucosa en sangre	Fumador
1	24	52	110	sí
2	27	50	95	no
3	25	57	103	no
4	27	48	115	sí
5	26	49	100	no
6	29	56	120	no

Primero identifique la variable IMC con la letra W, la variable Colesterol-HDL con la letra X, la variable glucosa en sangre con la letra Y, y la variable Fumador con la letra Z. Luego escriba cuáles son los valores individuales que representan las siguientes referencias:

Referencia	Valor	Referencia	Valor
W ₁	()	Y ₃	()
W ₂	()	Y ₄	()
W ₃	()	Y ₅	()
X ₂	()	Z ₄	()
X ₃	()	Z ₅	()
X ₄	()	Z ₆	()

REFERENCIAS

Camel, F. (1991). *Estadística Médica y Planificación de la Salud* (Vol. 1, 1a. ed.). Mérida, Venezuela: Universidad de los Andes.

Nunnally, J. C., & Bernstein, I. (1997). *Psychometric Theory* (3a. ed.). EUA: McGraw-Hill.

- Es necesario extender, reordenar o realizar otros cambios en la manera en que los datos se colectan. Esto es necesario cuando el investigador considera que al instrumento se le deben agregar más preguntas, cuando el orden de éstas debe ser cambiado o cuando el instrumento original fue diseñado para una entrevista y las preguntas se aplicarán por medio de un equipo de cómputo.
- 3) Desarrollar un instrumento propio. Muchos investigadores lo hacen, pero no es un trabajo sencillo que pueda realizarse en poco tiempo. Con frecuencia el desarrollo de un nuevo instrumento toma tiempo y requiere habilidades que no todos los investigadores dominan.

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

Las preguntas

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

- Identificar el objetivo general y los específicos de la investigación.
- Definir con claridad las variables en estudio.
- Recordar que las preguntas se realizan en un contexto social, cultural y económico.
- Conocer a los encuestados, es decir, quién responde.
- Identificar a los encuestadores, es decir, quién pregunta.
- Estandarizar las preguntas, es decir, cómo se pregunta.
- Estandarizar el formato de las respuestas, es decir, cómo se responde.
- Conciliar lo que se quiere preguntar con el tiempo que se dispone para hacerlo.

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

- Realizar preguntas útiles, relacionadas con la investigación que se pretende realizar.
- Hacer preguntas concretas utilizando un lenguaje convencional.
 - Utilizar oraciones completas. En español, la sintaxis de la oración tiene la siguiente estructura: primero va el sujeto, luego un verbo conjugado y termina con un complemento (Cohen, 2010).
 - Utilizar un lenguaje fácil de comprender para los entrevistados o para quien llene el instrumento. Para ello, conviene no utilizar abreviaturas, expresiones coloquiales o en jerga o expresiones técnicas.

- Utilizar periodos de tiempo pertinentes a la importancia de la pregunta.
 - Preguntas y respuestas cortas cuando se quiera ahorrar tiempo, pero sin sacrificar su claridad.
 - En caso de temas sensibles al entrevistado, se pueden utilizar **preguntas cargadas**. Por ejemplo, si existe interés en preguntar sobre consumo de drogas ilegales, podríamos redactar una pregunta en los siguientes términos: "Algunas personas, por experimentar, han probado la mariguana. ¿Usted ha probado la mariguana en alguna oportunidad de su vida?". Al plantear la pregunta de esta manera, facilitamos que los entrevistados la contesten, pero es necesario tener cuidado, porque la redacción puede ser tan compleja que llega a dificultar su interpretación por el entrevistado.
- Evitar frases y palabras sesgadas. Éstas son las que orientan la respuesta. Por ejemplo, la pregunta "¿Verdad que usted quiere a sus hijos?" anticipa la respuesta que queremos escuchar.
- Evitar preguntas con dos respuestas.
- Evitar preguntas en negativo.
- Hacer que las preguntas sean revisadas por expertos y por respondientes potenciales.
 - Las preguntas son precisas y sin ambigüedades cuando dos o más personas entrevistadas potenciales están de acuerdo con las palabras usadas para contestarla.
- Recordar que se pueden adoptar o adaptar preguntas que ya han sido utilizadas satisfactoriamente por otros investigadores.

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

- La confiabilidad se refiere a la consistencia en los resultados. Es decir, que cada vez que se mida lo mismo se obtenga el mismo resultado. Por ejemplo, si nos pesamos en una báscula y registramos 70 kg, nos bajamos de la misma y luego nos pesamos nuevamente, esperamos que la segunda lectura sea igual (o muy cercana) a 70 kg. Si éste es el caso, hablamos de una lectura confiable. En caso contrario (que la segunda lectura sea, por ejemplo, de 80 kg), las lecturas no son confiables.
- La validez se refiere a que el instrumento realmente mida la variable que se quiere medir. Por ejemplo, si queremos medir la grasa corporal de una persona, tendríamos varias opciones: podríamos pesarla (sabemos que a mayor grasa corporal mayor peso corporal, pero también la masa muscular aumenta el peso, por lo que el peso total no sería la mejor opción), calcular el índice de masa corporal (peso/talla²), medir pliegues cutáneos o medir la impedancia. El último tendría mayor validez que los anteriores.
- La objetividad se refiere al grado potencial en que los resultados pueden ser influidos o sesgados por el investigador o quien registra el dato.

Las respuestas

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

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

- Las respuestas dicotómicas presuponen sólo dos opciones posibles. Las más frecuentes en esta categoría son SÍ o NO. Por ejemplo:
¿Ha tenido usted tres o más evacuaciones líquidas en las últimas 24 horas?

No
Sí

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

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

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

- Las respuestas múltiples ordinales presuponen varias opciones posibles que mantienen una relación de orden entre ellas. Por ejemplo:
¿Qué tan de acuerdo o en desacuerdo está usted con el siguiente enunciado?: Los accidentes son causados por el destino.

Completamente de acuerdo	
Casí de acuerdo	
No sabe	
Casí en desacuerdo	
Completamente en desacuerdo	

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

- Utilizar escalas que tengan sentido con los objetivos.
- Considerar cinco tipos de respuestas:
 - Aprobación: Definitivamente cierto, cierto, no sé, falso, definitivamente falso.

- Frecuencia: Siempre, frecuentemente, regularmente, algunas veces, casi nunca, nunca.
 - Intensidad: Nada, muy poco, regular, moderado, severo.
 - Influencia: Gran problema, problema regular, problema pequeño, casi sin problema, sin problema.
 - Comparación: Mucho más que los otros, poco más que los otros, aproximadamente igual que los otros, algo menos que los otros, mucho menos que los otros.
- Balancear las respuestas.
 - Usar escalas de 5 o 7 categorías.
 - Poner las categorías negativas al principio de la escala.
 - Mantener las escalas en la misma página y fáciles de completar.
 - Usar escalas sólo cuando los encuestados pueden ver o recordar con facilidad todas las opciones.

Para el caso de cualquiera de los tres modelos de respuestas anteriores (dicotómicas, múltiples nominales, múltiples ordinales), siempre es necesario considerar la opción "otra". Es preciso tener cuidado, porque no es raro que la opción "otra" resulte la respuesta más frecuente. Por otra parte, también es necesario tomar en cuenta las respuestas "no sé", "sin datos", "no aplica" o "no quiso contestar" que llegan a ser importantes durante el análisis aunque no sean categorías de la variable en estudio.

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

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

¿Cuál es su edad en años cumplidos?

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

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

o más preguntas, como en el caso de la ocupación. Por ejemplo:

En su trabajo usted es (sólo seleccione una)

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

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

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

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

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

- Por último, y desde una perspectiva cuantitativa, las respuestas abiertas deben evitarse cuando no existan criterios claros de codificación, porque consumen mucho tiempo durante el llenado del instrumento, su manejo posterior es muy complejo y generalmente no son adecuadas para el tratamiento estadístico.

Revisión y corrección de datos recogidos

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

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

REFERENCIAS

Armstrong, B. K., White, E., & Saracci, R. (1992). *Principles of Exposure Measurement in Epidemiology*. Oxford: Oxford University Press.

Cohen, S. (2010). *Redacción sin dolor* (5a. ed.). México: Planeta.

Fink, A. (1995). *The survey kit* (Vol. 2, How to ask survey questions). Thousand Oaks, California: Sage.

McDowell, I., & Newell, C. (1996). *Measuring Health: a guide to rating scales and questionnaires* (2a. ed.). New York: Oxford University Press.

es fácil olvidar si el elemento que se está señalando ha sido o no contado, por lo que con frecuencia se cometen errores que son difíciles de corregir; c) el número de clasificaciones simultáneas que se puede realizar es limitado (difícilmente más de 3 o 4); y d) es imposible realizar sumatorias de valores no agrupados.

Tarjetas simples

Consisten en un rectángulo de papel grueso (p. ej., bristol o cartulina) del cual se ha recortado una esquina. Las características de los individuos registrados se anotan en pequeños espacios indicados en una de sus caras, de preferencia con letras grandes y siempre en la misma localización de la tarjeta. La esquina recortada facilita el acomodo previo al conteo (véase figura 3-2).

Para la tabulación procedemos así:

1. Acomodamos todas las tarjetas en un solo fajo, teniendo cuidado de colocar la esquina recortada en la misma posición para todas las tarjetas. Cuando hacemos esto, todas las tarjetas nos mostrarán el mismo frente de la hoja.
2. Identificamos la localización de la variable a tabular (presencia de vómito, estado civil, etc.).
3. Separamos las tarjetas en dos o más montones según las categorías de la variable de interés (diarrea sí, diarrea no; masculino, femenino, etc.).
4. En caso necesario, cada montón de tarjetas puede separarse nuevamente en atención a otras variables.
5. Una vez completadas las separaciones necesarias, se procede a contar el número de tarjetas de cada montón.

Este procedimiento presenta varias ventajas: a) puede ser utilizado en la mayoría de las investigaciones que se realizan en el campo de la salud, b) facilita la verificación de conteos y disminuye el potencial de error, c) el número de clasificaciones simultáneas no está limitado a nuestra memoria, ni a la complejidad de un cuadro tabular, d) las tarjetas son relativamente económicas y e) el mismo investigador puede elaborarlas. No obstante, presenta algunas desventajas, entre las que encontramos: a) dificultad para obtener sumatorias de variables cuantitativas, b) facilidad para perder una o varias tarjetas y c) dificultad para manejar grandes cantidades de datos.

Masculino	Diarrea: Sí
Vómito: No	Sopa: Sí
Ensalada: Sí	

Figura 3-2. Modelo de tarjeta simple.

Tarjetas con perforaciones marginales

Son tarjetas de tamaño variable, cuya característica más importante son sus perforaciones marginales (figura 3-3). Al igual que las tarjetas simples, una de sus esquinas debe recortarse con la finalidad de facilitar su acomodo. En una cara de la tarjeta se anota, anexo a una perforación marginal, la categoría de interés. Si el elemento al cual corresponde la información contenida en la tarjeta posee la característica señalada, se realiza un recorte entre la perforación marginal y el borde de la tarjeta. El recorte debe ser lo suficientemente grande para evitar que exista algún "gancho" que impida el manejo de las tarjetas.

Para contabilizar los datos se procede de la siguiente manera:

1. Las tarjetas se acomodan atendiendo al recorte de la esquina.
2. Se pasa un punzón o aguja de tejer (según el número de tarjetas) a través de las perforaciones que corresponden a la variable y categoría de interés.
3. Tomando de cada extremo el instrumento punzante utilizado, se procede a levantar y agitar suavemente las tarjetas. Aquellas tarjetas que tengan la perforación recortada caerán y corresponderán a los individuos que presentan la característica de interés.
4. El procedimiento se puede repetir a partir del paso 1 el número de veces necesarias para completar las clasificaciones simultáneas de interés.
5. Al final, sólo resta contar el número de tarjetas separadas en la última clasificación.

Este tipo de tarjetas permite manejar con facilidad grandes volúmenes de información. Al igual que las tarjetas simples, facilitan la verificación de conteos y la realización de un gran número de clasificaciones simultáneas. Sin embargo, ya que han de elaborarse en imprenta, son caras y, al igual que con las tarjetas simples, existe la dificultad para obtener sumatorias de variables cuantitativas y la posibilidad de perder una o varias tarjetas.

Sexo masculino	
Presentó diarrea	<input type="checkbox"/>
Presentó vómito	<input type="checkbox"/>
Hipertermia > 37° C	<input type="checkbox"/>
Presentó prurito	<input type="checkbox"/>
Comió sopa	
Comió ensalada	
Comió guisado	<input type="checkbox"/>
Comió postre	<input type="checkbox"/>

Figura 3-3. Modelo de tarjeta con perforación marginal.

existentes, más bien se escribe sobre algunos que los autores han tenido oportunidad de utilizar.

PROCESADORES DE TEXTO

Estos programas son los más genéricos de todos los que se utilizan. Básicamente sirven para escribir las propuestas iniciales, reportes finales del trabajo y formas para captar datos (cuestionarios, cédulas de captura). Algunos incluyen utilerías capaces de elaborar cuadros y gráficos. Otros incluso permiten escribir fórmulas como las que se presentan en este libro. Entre estos programas destaca Word para Windows.

ADMINISTRADORES DE BASES DE DATOS

Cumplen una función muy importante durante el trabajo estadístico: ayudan a capturar datos en los archivos de cómputo en los que se almacenan. Además de generar la estructura de la base y de permitir capturar los datos, estos programas también permiten editar y transformar datos, así como generar nuevas variables y asignarles valores a partir de los existentes. Pueden ayudar a generar pantallas de captura que faciliten el trabajo de introducir datos, limitando así, el número de errores que se pueden cometer. La verificación de la captura se puede realizar comparando los datos en pantalla con la forma en la que se registraron. Si existe algún error, el programa permite corregirlo. Si alguien cuenta con un poco de experiencia en programación, también puede utilizar estos programas para tabular datos y realizar operaciones estadísticas, como el cálculo de proporciones, promedios y desviaciones estándar, o pruebas de chi-cuadrada y t de Student, por ejemplo. El potencial de estos programas en el análisis estadístico es enorme, pero la limitación más importante reside en que es necesario conocer los programas con mucha profundidad y saber programar. Entre los programas de esta categoría están dBase, Fox y Access.

HOJAS DE CÁLCULO

Su estructura de celdas, construidas a partir de columnas y renglones, además de las funciones que incluyen, las hace muy útiles para el trabajo estadístico. Entre estos programas, los más conocidos son Excel y Lotus. Su primera aplicación suele ser la captura de datos; es mucho más sencilla que la que se puede hacer en los administradores de bases de datos, pero

menos versátil. Para capturar datos en una hoja de cálculo, se procede a identificar las columnas con las variables y los renglones con los registros. En las celdas del primer renglón se anotan los nombres de las variables y, a partir del segundo renglón, se capturan los datos que corresponden a cada elemento del grupo. Por brevedad, en la captura suelen utilizarse códigos y no etiquetas. Por ejemplo, en vez de escribir "masculino" en la columna de sexo se puede anotar "1" y en lugar de "femenino", "2". De esta manera se ahorra mucho tiempo y se reduce el número de errores.

Ejemplo explicativo 4-1

Recuérdese el cuadro 1-1 en el que se presentaron los datos de un grupo de trabajadores de un taller de automóviles. En una hoja de cálculo, los mismos datos tendrían la presentación del cuadro 4-1, en el cual los renglones (identificados con un número en la primera columna) y las columnas (identificadas con letras mayúsculas en el primer renglón) definen las referencias de cada celda. En el renglón 1 se identifican las variables: "registro" en la columna "A", "sexo" en la columna "B", "edad" en la columna "C" y hemoglobina con la etiqueta "Hb" en la columna "D". Bajo la etiqueta "registro" se anota el número *i*-ésimo que corresponde a la captura, generalmente un número progresivo, en sustitución del nombre del sujeto o elemento del conjunto que se estudia. Bajo la etiqueta "sexo" el código "1" significa "varón", mientras que el "2" se refiere a "mujer". En las celdas debajo de "edad" y "Hb", se anotan los valores que corresponden a cada lectura.

La verificación de datos puede realizarse de varias maneras. En primer lugar, los datos en la pantalla se pueden comparar con la forma en la cual se registraron los datos. Otro procedimiento más ventajoso por su facilidad es la utilización de filtros, a partir de los cuales se pueden ordenar los datos según diferentes criterios. Por ejemplo, si los datos del cuadro 4-1 fueran muy extensos, entonces se podría pedir al programa que los ordenara según los valores anotados en la columna del sexo. Si por algún motivo alguien hubiera anotado un tres en esa columna, se podría identificar al principio o al final de ella después de ordenar los registros según un criterio ascendente o descendente, respectivamente. Los filtros reconocen todos los criterios o secuencias de datos capturados en cada

Cuadro 4-1. Representación de la captura de datos del cuadro 1-1 en una hoja de cálculo

	A	B	C	D	E
1	registro	sexo	edad	Hb	
2	1	1	45	12	
3	2	2	23	13	
4	3	1	32	11	
5	4	1	18	15	
6	5	1	21	14	