

## **SESIÓN 7. ANÁLISIS DE DATOS**

### **7.1. FLUJO DE PROCESAMIENTO DE DATOS**

#### **7.1.2. CAPTURA DE DATOS**

### **7.2. IDENTIFICACIÓN DE LOS INSTRUMENTOS ACEPTABLES**

#### **7.3. EDICIÓN**

#### **7.4. CODIFICACIÓN**

#### **7.5. PERFORACIÓN Y VERIFICACIÓN**

#### **7.6. DEPURACIÓN DEL CONJUNTO DE DATOS**

#### **7.7. GENERACIÓN DE NUEVAS VARIABLES**

#### **7.8. FLUJOS ALTERNATIVOS DE PROCESAMIENTO**

### **7.1. FLUJO DE PROCESAMIENTO DE DATOS**

#### **7.1.2. CAPTURA DE DATOS**

La captura de datos primarios en el campo; por entrevistas, observación, o ambos, es la parte más débil en el proceso de investigación. El personal aceptablemente bien calificado trabaja con cuidado en las etapas previas para garantizar la precisión de los resultados. Sin

embargo, los frutos de su trabajo pueden perderse si el personal de capo (tomadores de datos) es

capacitado y supervisado en forma inadecuada. La administración del personal durante las labores

constituye una tarea difícil, porque son normalmente trabajadores de tiempo parcial con poca motivación en su trabajo. Este lo realizan en lugares donde no pueden ser observados en muchas

ubicaciones separadas.

Los pasos finales de un proyecto de investigación de mercado son: analizar los datos, interpretar los

resultados y entregar un informe por escrito. El moderno equipo electrónico de procesamiento de

datos permite al investigador analizar y tabular gran cantidad de datos de manera rápida y barata. El

producto final de investigación son las conclusiones y las recomendaciones del investigador entregadas por escrito.

La tarea fundamental en el procesamiento de datos es convertir los datos en bruto del instrumento

de recolección de datos en una forma legible por el computador. Luego podemos hacer uso de los

procedimientos computarizados de análisis de datos para extraer la información de los datos.

Antes

de ver cómo se lleva a cabo este proceso, existen conceptos y términos básicos que se necesita

entender:

a. **Caso:** Un caso se refiere a una unidad específica de análisis para el estudio. Con mucha frecuencia, la unidad de análisis es quien responde al cuestionario, por tanto cada encuestado se consideraría un caso y el número total de casos sería equivalente al tamaño de la muestra.

- b. *Representación computarizada de los datos*: Los datos que suministra un instrumento de investigación deben convertirse a una forma legible por el computador. Históricamente, la tarjeta de computador es el instrumento básico utilizado. Actualmente casi siempre los datos se ingresan directamente al computador en el momento de la recolección de los mismos.
- c. *Paquete de datos*: está compuesto por todas las tarjetas de computador necesarias para representar los datos que aparecen en todos los cuestionarios.
- d. *Almacenamiento de datos*: una vez que se han ingresado todos los datos al computador, el investigador puede utilizar la capacidad de éste para almacenar los datos del paquete de datos en un archivo de disco. De esta manera, se evitan los problemas que representarían si se eliminase o perdiese el paquete de datos.
- e. *Matriz de datos*: El almacenamiento en un archivo puede considerarse como una manera de formar una matriz de datos. Cada fila de la matriz representa un caso y cada columna representa una variable. El número total de las filas es igual al número de casos, y el número total de columnas es igual al número de variables.

#### **FLUJO DEL PROCESAMIENTO DE DATOS**

La secuencia tradicional de las funciones que deben realizarse en el procesamiento de datos incluyen:

4

*Decisión sobre si debe utilizarse o no el instrumento de recolección de datos para el análisis*

*Edición*

*Codificación*

*Perforación del ingreso de datos al computador y verificación*

*Conversión del paquete de datos a una forma legible para el computador*

*Depuración del conjunto de datos;*

*Generación de nuevas variables según sea necesario*

*Ponderación de los datos de acuerdo con el plan de muestreo y resultados*

*Almacenamiento del conjunto de datos.*

Con los sistemas directos de entrada de datos algunos de estos pasos son innecesarios o se realizan

simultáneamente con el ingreso de los mismos. Sin embargo, en los sistemas basados en el computador es necesaria la función realizada en cada uno de estos pasos y su comprensión.

## **7.2. IDENTIFICACIÓN DE LOS INSTRUMENTOS ACEPTABLES**

Al recibir un instrumento de recolección de datos de campo, el investigador debe examinarlos para

determinar si son aceptables para su uso en el estudio. Los criterios exactos para determinar un

instrumento como no aceptable varía de un estudio a otro, pero los que se presentan a continuación

son característicos:

*Una parte significativa del instrumento se deja sin contestar o sin respuesta a los elementos claves;*

*El claro que a partir de las respuestas dadas el encuestado no entendió el objetivo que se buscaba para completar el instrumento;*

*Las respuestas muestran muy poca varianza; por ejemplo, las respuestas a una serie de preguntas de actitud son todas e, en una escala de 7 puntos. Esto demuestra que el encuestado*

*no está realizando su tarea con seriedad.*

*El instrumento se ha llenado con un elemento equivocado de la muestra. Por ejemplo, el estudio*

requiere que los encuestados sean mujeres que trabajan, y un hombre completa el instrumento.

El instrumento está físicamente incompleto. Por ejemplo, es posible que no se haya incluido una página.

El instrumento se recibe después de una fecha límite para su entrega. Si se esperara hasta que

todos los instrumentos se devolvieran, posiblemente el estudio no se completaría a tiempo.

Si el investigador considera que el defecto particular en cuestión puede corregirse dentro de unas

limitaciones razonables de tiempo y costos, el instrumento puede devolverse al campo. De no ser

así, se elimina.

Todos los instrumentos de recolección de datos deben estar sujetos a este tipo de examen preliminar

antes de enviarse a través del resto del flujo de procesamiento de datos. Ante todo, deben establecerse los criterios para aceptar o rechazar un instrumento antes de recibirlo.

5

### 7.3. EDICIÓN

La edición significa una revisión de los instrumentos de recolección de datos para asegurar la máxima exactitud y la mínima ambigüedad. Es importante que la edición se haga consecuentemente.

En un estudio pequeño, una persona puede realizar la función de edición y es probable que la consecuencia sea alta. En un estudio grande que requiere muchos editores, se necesita un supervisor de edición para asegurar que se mantenga la consecuencia entre editores. Esta persona

tendría que verificar las muestras de diferentes editores en diversas secciones del instrumento.

Alternativamente, cada editor podría hacerse responsable de una sección diferente del instrumento y

editar todos los instrumentos para esta sección.

Al realizar la función de edición es necesario preocuparse de:

a. *Legibilidad*: los datos deben ser legibles para que posteriormente puedan codificarse en forma

adecuada. El editor elimina la ambigüedad de los datos registrados con el propósito de que el codificador sepa exactamente lo que tiene que hacer;

b. *Totalidad*: Las preguntas no respondidas pueden tratarse de tres maneras. En primer lugar, el

editor puede ponerse en contacto con el entrevistador para tratar de determinar si el encuestado no respondió la pregunta o si el entrevistador olvidó registrar la respuesta. Al hacer esto, se corre el riesgo de que el entrevistador no recuerde esta entrevista específica correctamente. Por otra parte, puede volverse a contactar al encuestado para que responda una pregunta específica. El segundo enfoque es clasificar este dato en particular como dato faltante. Finalmente, si el editor considera que faltan demasiados elementos de los datos, puede volverse a enviar todo el instrumento al campo o eliminarlo del estudio.

c. *Consecuencia*: Debe hacerse una verificación preliminar sobre la consecuencia (o consistencia) de los datos. Por ejemplo, el editor puede verificar si los encuestados que aseguran que compran gasolina con tarjeta de crédito, verdaderamente tengan tarjetas de crédito. El editor puede solicitar al entrevistador que resuelva alguna inconsecuencia, clasifique las respuestas a estas preguntas como datos faltantes o elimine el instrumento del estudio.

d. *Exactitud*: El editor necesita estar atento a cualquier evidencia de inexactitud en los datos. En este caso, el área más importante se relaciona con el posible sesgo o trampa del entrevistador. Tales actividades pueden detectarse buscando un patrón común de respuestas en los instrumentos de un entrevistador o de quien registra las respuestas en particular.

e. *Clarificación de respuesta*: Algunas veces las repuestas a preguntas abiertas son difíciles de interpretar con claridad. Las palabras de quien registra las respuestas pueden abreviar demasiado la respuesta o algunas palabras pueden ser ambiguas. El editor puede asignar un significado a la respuesta o solicitar al entrevistador qué quiso decir. El riesgo de error es alto en ambos casos. Obviamente, un buen trabajo de campo inicial puede evitar que surjan muchos problemas. También se presentan dificultades cuando no se contestan las preguntas en la forma en la cual lo requieren las instrucciones del instrumento. Este problema se presenta especialmente en las encuestas por correo.

## 7.4. CODIFICACIÓN

6  
Comprende la asignación de un símbolo numérico a una determinada columna de la tarjeta de computador, para representar una respuesta específica en un instrumento de recolección de datos.

### *Preguntas cerradas y preguntas abiertas*

Para preguntas estructuradas o cerradas, el esquema de codificación generalmente se determina antes de emprender el trabajo de campo. Se pueden imprimir hasta los códigos reales de los instrumentos de recolección de datos. Por ejemplo, una clasificación del sexo puede aparecer en el instrumento de la siguiente manera:  
Las preguntas abiertas representan un problema más complejo para su codificación. En este caso, el entrevistador registra las respuestas del encuestado literalmente. Hay dos métodos generales para este problema. El primero es la preparación de un esquema de codificación relativamente bien elaborado antes de terminar el trabajo de campo. Para poder lograrlo, el investigador debe guiarse por los resultados de estudios anteriores o por consideraciones teóricas dominantes. La principal tarea del investigador es capacitar a los codificadores, de manera que conviertan las respuestas literales en la categoría de codificación correcta.  
El segundo enfoque se limita a esperar a que los instrumentos regresen del campo antes de desarrollar un esquema de codificación. En este caso, el investigador hace una lista de 50 a 100 de las respuestas a una pregunta específica. Posteriormente examina esta lista y decide cuáles son las categorías apropiadas para resumir los datos. Entonces, el investigador capacita a los codificadores con base en este esquema y los alerta para que estén al tanto de otras respuestas que se presentan con alguna frecuencia. Si esto sucede, es posible que sea necesario volver atrás y revisar el

esquema de codificación y, por tanto, tener que volver a codificar todos los instrumentos en esta pregunta.

*Reglas y convenciones para la construcción del código:*

1. *Establecer categorías de códigos mutuamente excluyentes y colectivamente exhaustivas:*

Es

fácil lograr que las categorías sean colectivamente exhaustivas agregando, por ejemplo, la categoría de código "otro", "sin información" o "ninguno" a las principales categorías para la variable en cuestión. Por "mutuamente excluyentes" se entiende que cada respuesta debe corresponder a una y sólo una categoría de código. Las categorías no deben superponerse; esta es la regla fundamental en toda codificación. Los otros elementos de esta sección se describen mejor como convenciones útiles, no como reglas.

2. Si existe incertidumbre acerca de los posibles usos para una variable particular bajo análisis,

deben codificarse los datos de tal manera que retengan muchos detalles. Es posible combinar las categorías de códigos en el momento del análisis si no se requiere tal detalle. Sin embargo es imposible ampliar los códigos para el análisis, si éstos se han registrado con muy poco detalle. Por consiguiente, podría desearse codificar con más detalle del que se pretende usar en el análisis. Un ejemplo de esto sería la codificación de la edad exacta del encuestado, y luego la combinación de las edades en categorías al momento del análisis.

3. Al escoger los puntos de quiebre para variables continuas, considere las siguientes preguntas:

¿En cuántas categorías debe dividirse la variable dado el plan de análisis desarrollado?

7

¿Las categorías deben tener intervalos iguales, o deben construirse de tal forma que cada categoría tenga aproximadamente el mismo número de casos?

¿Las categorías extremas deben estar abiertas con el propósito de lograr una amplia gama de puntajes extremos o deben ser fijos los intervalos?

Al tomar estas decisiones, el investigador debe reconocer que el uso de intervalos iguales permite

que se realice posteriormente un análisis estadístico más fácil y que el número de categorías seleccionadas afecte el detalle de la información retenida de los instrumentos. Cuantas más categorías se tengan, mayor será el detalle. Al final, el número y definición de las categorías seleccionadas debe satisfacer al usuario de la investigación.

*Respuestas múltiples*

El problema de las respuestas múltiples surge en dos contextos. El primero se presenta cuando el

investigador espera recibir una sola respuesta. En este caso, la decisión podría ser seleccionar una

de las respuestas con base en una prioridad establecida, por ejemplo, la primera respuesta que se

escribió. Alternativamente, pueden desarrollarse categorías de códigos con el propósito de representar las combinaciones de las respuestas.

En situaciones donde el investigador espera más de una respuesta, existen dos procedimientos

posibles. El primero es tratar cada una de las respuestas posibles como una variable por separado,

en una columna distinta. Por ejemplo, una pregunta puede solicitar al encuestado que enumere los

deportes que practica. Las respuestas para cada deporte serían del tipo "sí" o "no" y se asignarían en

una columna única. La segunda opción se presenta cuando el investigador tiene un número

específico de respuestas esperadas. El investigador puede asignar una columna separada a las primeras respuestas, otra columna a las segundas respuestas, etc. Al hacer el análisis, el investigador podría examinar cada columna de respuesta por separado o combinarlas para lograr respuestas totales. Debe tenerse precaución al combinar datos de este tipo; es posible que un atributo colocado en segundo o tercer lugar por un gran número de encuestas resulte con más menciones totales que otro atributo que tuvo más menciones en la primera respuesta. Una suma simple de las menciones a través de las columnas podría confundir a los gerentes. Debe consultárseles acerca de la importancia que le atribuyen a los segundos, terceros o demás niveles de mención antes de desarrollar un esquema de combinación. Es necesario reconocer que la codificación es una tarea potencialmente tediosa. Después de trabajar ardua y creativamente para desarrollar el esquema de codificación, el proceso real de codificación puede parecer bastante aburrido. Además, con frecuencia esta tarea no es bien remunerada, lo cual genera problemas con la calidad del personal. Los errores de codificación ocurren con frecuencia a menos que el investigador supervise muy de cerca la codificación.

#### *Libro de códigos*

Un libro de códigos es donde se documenta toda la información necesaria acerca de las variables

en un conjunto de datos. Este libro tiene tres funciones:

- Sirve como guía para los codificadores
- Ayuda a que los investigadores localicen las variables que desean utilizar en un determinado análisis de datos

8

- Permite que se realice una identificación adecuada de las categorías de variables a medida que se interpreta el resultado del computador.

El contenido del libro de códigos varía. Para un estudio muy simple sólo se puede escribir el número

de columnas apropiada y perforarse el número en el instrumento. Un cuestionario completamente

precodificado puede servir como libro de códigos.

En estudios más complejos, es útil tener un libro de códigos que contenga más información.

Por

tanto, el libro debe contener:

- número de la pregunta
- número de la variable
- columnas apropiadas de la tarjeta
- formato (posición decimal implícita)
- nombre de la variable
- definiciones de la categoría

#### *Codificación real*

Una vez que se han establecido las categorías de código, se realiza la codificación real. En este

caso, los codificadores escriben los códigos apropiados en el sigio designado del papel.

## 7.5. PERFORACIÓN Y VERIFICACIÓN

Los perforadores reciben las hojas de codificación terminadas y perforan el número exacto que aparece en las hojas en las tarjetas de computador. Así se convierten los datos de instrumento de recolección de datos a huecos en tarjetas de computador. El resultado de este proceso debe ser un paquete de datos exactos.

### ***Convertir un paquete de datos en un formato legible por el computador***

El paquete de datos debe introducirse en el computador de tal manera que lo puedan utilizar los programas de análisis de datos del computador. Un elemento que debe incluirse como una variable en el archivo de datos es el número de identificación del caso. Esto ayudará a facilitar el paso siguiente, que consiste en la depuración de los datos.