



# Capítulo 12

## Codificación y tabulación de encuestas

Investigación de mercados  
Prof. Verónica Rosendo Ríos

## CAPÍTULO 12. CODIFICACIÓN Y TABULACIÓN DE ENCUESTAS

---

*“La excelencia es hacer algo común de manera no común”*

BOOKER T. WASHINGTON

# CAPÍTULO 12. CODIFICACIÓN Y TABULACIÓN DE ENCUESTAS

---

## CONTENIDOS

1. Codificación de encuestas
2. Revisión y transcripción de encuestas
3. Depuración de errores
4. Tratamiento de valores perdidos
5. Ajustes estadísticos de los datos



# CAPÍTULO 12. CODIFICACIÓN Y TABULACIÓN DE ENCUESTAS

---

## ★ CODIFICACIÓN:

Código

Campo

Registro

## ★ CODIFICACIÓN:

- ✓ Código de campo fijo.
- ✓ Por ejemplo, el número 9 se puede utilizar como código de valor perdidos.

## ★ CODIFICACIÓN:

- ✓ Codificación de preguntas estructuradas o cerradas
- ✓ Codificación de preguntas no estructuradas o abiertas



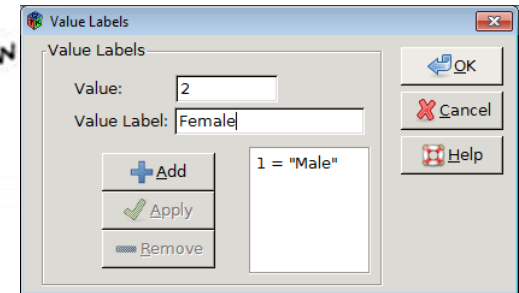
## 1. Codificación de preguntas estructuradas o cerradas

- ✓ Las opciones del encuestado están predeterminadas.

Por ejemplo:

¿Tiene el pasaporte vigente?

1. Sí 2. No



Para esta pregunta, un “sí” se codificaría como 1 y un “no” como 2.

Como solo se permite una respuesta y sólo hay dos respuestas posibles (1 ó 2), una sola columna es suficiente.

## 2. Codificación de preguntas estructuradas o abiertas

- ✓ La codificación de preguntas abiertas o no estructuradas es más compleja. Las respuestas textuales o verbatim de los encuestados se registran en el cuestionario.
- ✓ Posteriormente se generan códigos que se asignan a las respuestas.
- ✓ Una vez que se han desarrollado los códigos, los codificadores deben ser formados de manera que sepan asignar correctamente los códigos al verbatim.



## 2. Codificación de preguntas estructuradas o abiertas

- ✓ Se sugieren las siguientes pautas para codificar preguntas estructuradas y cuestionarios estructurados en general:
- ★ Códigos de categoría: deben ser mutuamente excluyentes y colectivamente exhaustivos.
  - ⊙ Las categorías son mutuamente excluyentes si *cada respuesta se ajusta a uno y solo un código de categoría.*
  - ⊙ No deben *solaparse*.
  - ⊙ Son colectivamente exhaustivas
  - ⊙ La gran mayoría de las respuestas deberían clasificarse en categorías significativas.

## 2. Codificación de preguntas estructuradas o abiertas

- ★ Se sugieren las siguientes **pautas** para codificar preguntas estructuradas y cuestionarios estructurados en general:
  - ✓ También se deberían asignar códigos de categoría para **cuestiones críticas**, incluso si nadie los ha mencionado. Puede ser importante saber que nadie ha mencionado una respuesta en particular.

Por ejemplo, la dirección de una importante empresa de bienes de consumo estaba preocupada por el empaquetado de una nueva marca de jabón. Por tanto, el empaquetado se incluyó como una categoría separada en las respuestas de codificación a la pregunta: "¿Qué es lo que menos te gusta de esta pastilla de jabón?".

- ✓ Los datos deben ser codificados de manera que retengan **el máximo nivel de detalle posible**.

## 2. REVISIÓN Y TRANSCRIPCIÓN DE ENCUESTAS

**Transcripción de datos:** desde los cuestionarios o las hojas de codificación a USB, CD, cintas o directamente en los ordenadores.

- ✓ Si los datos se han recogido mediante **CATI o CAPI**, este paso no es necesario ya que los datos se han introducido directamente en el ordenador a medida que se recogían.
- ✓ Además del *keypunching*, los datos se pueden transferir utilizando *formularios de detección de marcas, escaneo óptico o análisis sensorial computarizado*.
- ✓ El *escaneo óptico*.
- ✓ Ej.: *Código universal de producto (UPC)*.



## 2. REVISIÓN Y TRANSCRIPCIÓN DE ENCUESTAS

### Verificación

- ✓ Si se utiliza la introducción manual de los datos (*keypunching*), *pueden ocurrir errores* y es necesario verificar el conjunto de datos y revisar los cuestionarios.
- ✓ *Una máquina verificadora y un segundo operador.*
- ✓ Cualquier discrepancia entre los dos conjuntos de datos transcritos se investiga para identificar los datos correctos y corregir los errores.
- ✓ Normalmente es suficiente con verificar *solo entre el 25 y el 50%* de los datos *de forma aleatoria.*



## 2. REVISIÓN Y TRANSCRIPCIÓN DE ENCUESTAS

---

### Método de transcripción

La selección del método de transcripción de los datos viene guiada por el tipo de método de encuesta que se haya utilizado y la disponibilidad de equipos adecuados.



### 3. DEPURACIÓN DE ERRORES: LIMPIEZA DE DATOS

---

#### 1. LIMPIEZA DE DATOS

La limpieza de datos incluye:

- ★ Verificaciones de consistencia
- ★ Tratamiento de respuestas perdidas

### 3. DEPURACIÓN DE ERRORES: LIMPIEZA DE DATOS

---

## LIMPIEZA DE DATOS

### 1. Verificaciones de consistencia

- Los datos fuera de rango.
- Respuestas inconsistentes lógicamente.
- Los valores extremos.

## 4. TRATAMIENTO DE VALORES PERDIDOS

---

### 2. TRATAMIENTO DE VALORES PERDIDOS

Los valores perdidos pueden deberse a:

- ✓ Respuestas ambiguas.
- ✓ Falta de respuestas.
- ✓ Respuestas que no han sido registradas correctamente.



El tratamiento de valores perdidos plantea serios problemas, particularmente si la proporción de valores perdidos **supera el 10%**.



## 4. TRATAMIENTO DE VALORES PERDIDOS

### 2. TRATAMIENTO DE VALORES PERDIDOS

Valores perdidos: ¿¿¿Qué hacemos????

1. Sustituir por un valor neutro
2. Sustituir por una respuesta imputada
3. Eliminación de casos (*casewise deletion*)
4. Eliminación por pares (*pairwise deletion*)



[https://cdn.pixabay.com/photo/2018/01/08/08/35/alzheimers-3068938\\_960\\_720.jpg](https://cdn.pixabay.com/photo/2018/01/08/08/35/alzheimers-3068938_960_720.jpg)

## 4. TRATAMIENTO DE VALORES PERDIDOS

---

### 2. TRATAMIENTO DE VALORES PERDIDOS

1. Sustituir por un valor neutro.
2. Sustituir por una respuesta imputada.

Por ejemplo, *el uso de un producto* puede estar relacionado con el *tamaño del hogar* para encuestados que han proporcionado datos a ambas variables. Tomando como referencia el tamaño del hogar, la falta de respuesta al uso del producto puede calcularse en base a la anterior. Sin embargo, este enfoque requiere un esfuerzo considerable y la introducción de sesgos importantes.

## 4. TRATAMIENTO DE VALORES PERDIDOS

---

### 2. TRATAMIENTO DE VALORES PERDIDOS

#### 3. Eliminación de casos (*casewise deletion*)

Los casos o encuestados con valores perdidos son eliminados del análisis.

- ✓ Podría dar como resultado una muestra pequeña.
- ✓ Recopilar datos es costoso y lleva mucho tiempo.
- ✓ La eliminación de casos podría sesgar seriamente los resultados.

## 4. TRATAMIENTO DE VALORES PERDIDOS

---

### 2. TRATAMIENTO DE VALORES PERDIDOS

#### 4. Eliminación por pares (*pairwise deletion*)

### 2. TRATAMIENTO DE VALORES PERDIDOS

- ✓ Los diferentes procedimientos para el tratamiento de valores perdidos pueden dar lugar a **resultados diferentes**, particularmente cuando las respuestas no faltan al azar y las variables están correlacionadas.
  - ✓ Por tanto, debe intentarse que los valores perdidos **sean los menos posibles**.
- El investigador debe considerar cuidadosamente las **implicaciones de los diversos procedimientos** antes de seleccionar un método particular para el tratamiento de los valores perdidos.

## 5. AJUSTES ESTADÍSTICOS DE LOS DATOS

---

### AJUSTES ESTADÍSTICOS DE LOS DATOS

¿¿Qué hacemos???

1. Ponderación.
2. Reespecificación de la variable.
3. Transformación de la escala.
4. Eliminación por pares (*Pairwise deletion*).



## 5. AJUSTES ESTADÍSTICOS DE LOS DATOS

---

### 1. Ponderación

- Por ejemplo, si se lleva a cabo un estudio para determinar *qué modificaciones se deben realizar a un producto específico*, el investigador puede dar mayor importancia a la opinión de *clientes habituales* del producto. Esto puede conseguirse mediante la asignación de ponderaciones (ej.: 3.0 a clientes habituales; 2.0 a clientes medios; 1.0 a clientes esporádicos o no clientes).



### 2. Reespecificación de la variable

- ✓ Transformación de los datos para **crear nuevas variables o para modificar variables ya existentes.**
  - ✓ Por ejemplo, imagine que diez categorías de respuestas de clientes de un producto se transforman en 4 tipos: habitual, medio, esporádico y no-usuario.
- ✓ Dummies.
- ✓ Por ejemplo, en un estudio sobre preferencias del consumidor por comida congelada, los encuestados se clasificaron como habituales, medios, esporádicos y no usuarios y se les asignaron los códigos 4, 3, 2, y 1 respectivamente. Esta codificación no fue significativa para determinados análisis estadísticos. Para realizar estos análisis, el uso del producto fue representado por tres variables ficticias,  $X_1$ ,  $X_2$ ,  $X_3$ , como se muestra a continuación.



## 5. AJUSTES ESTADÍSTICOS DE LOS DATOS

### 2. Reespecificación de la variable

*Ejemplo de uso de la variable dummy*

Categoría de producto	Uso	Código original de la variable	Dummy X <sub>1</sub>	Dummy X <sub>2</sub>	Dummy X <sub>3</sub>
No-usuario		1	1	0	0
Usuario infrecuente		2	0	1	0
Usuario medio		3	0	0	1
Usuario habitual		4	0	0	0

# 5. AJUSTES ESTADÍSTICOS DE LOS DATOS

## 3. Transformación de la escala

- ✓ Manipulación de los valores de la escala para garantizar la comparabilidad con otras escalas o para hacer que los datos sean adecuados para el análisis.
- ✓ Por ejemplo, las variables de imagen se pueden medir en una escala diferencial semántica de siete puntos, las variables de actitud en una escala de calificación continua y las variables de estilo de vida en una escala de Likert de cinco puntos. Por tanto, no es significativo hacer comparaciones entre las escalas de medición entre los encuestados.
- ✓ Un procedimiento más común es la **estandarización**.
- ✓ Para estandarizar una escala  $X_i$ , primero se resta la media,  $X_{\text{bar}}$ , para cada puntuación y luego se divide por la desviación estándar,  $s_x$ . Así, la escala estandarizada tendrá una **media de cero y una desviación típica de 1**.
- La estandarización permite al investigador comparar variables que se han medido utilizando diferentes tipos de escalas.
- Matemáticamente, las puntuaciones estandarizadas,  $z_i$ , pueden obtenerse:

$$Z_i = \frac{(X_i - X_{\text{bar}})}{S_x}$$

- ✓ [https://cdn.pixabay.com/photo/2017/06/26/13/22/cube-2443805\\_960\\_720.png](https://cdn.pixabay.com/photo/2017/06/26/13/22/cube-2443805_960_720.png)





## “Construyendo un Multipetabyte Data System”

• ¿Qué es un petabyte? Para aquellos de ustedes que están familiarizados con el almacenamiento en disco medido en gigabytes (GB), un petabyte es de 1,000,000 de GB. ¿Quién podría necesitar un sistema de datos tan grande? No es de extrañar, el minorista más grande en el mundo - Walmart. Con **más de 800 millones de transacciones vinculadas a más de 30 millones de clientes cada día**, la codificación y el análisis de datos necesarios para dicho sistema están claros. Un diseño de sistema de datos fiable y flexible es crucial para cumplir con los requisitos de datos sustanciales de Walmart. Ya se trate de proveedores que desean ver el movimiento y las ventas de productos en tiempo real, o ejecutivos interesados en inteligencia comercial o planificación de escenarios, el aspecto del diseño de datos del almacén de datos de Walmart es la clave de su éxito. Debido a que prácticamente todas las transacciones se procesan en tiempo real, **la integridad de los datos y la comprobación de errores** son factores de éxito valorados por los ejecutivos de Walmart. Con el tiempo, a medida que aumenta la demanda del análisis de datos más robusto y oportuno, Walmart parece haber realizado las inversiones necesarias para hacer crecer su almacén de datos en un futuro. En el futuro, incluso hay planes para tener data mart-smallers (mercados de datos): sistemas de datos más pequeños y de temas específicos que puedan responder a las necesidades de un área comercial particular.

Fuentes: Zikmund *et al.* (2013). Hayes Weier, Mary, “Hewlett-Packard Data Warehouse Lands in Wal-Mart’s shopping cart”, Intelligent Enterprise (August 4, 2007) [https://cdn.pixabay.com/photo/2018/02/20/16/43/binary-3168094\\_960\\_720.jpg](https://cdn.pixabay.com/photo/2018/02/20/16/43/binary-3168094_960_720.jpg)

## 5. AJUSTES ESTADÍSTICOS DE LOS DATOS

---

### REFERENCIAS BIBLIOGRÁFICAS

- Brown, T. J., y Suter, T. (2012): *MR*. South Western, Cenage Learning. USA.
- Hair, J.; Bush, R., y Ortinau, D. (2006): *Marketing research. Within a changing environment*. Revised International Edition (3<sup>rd</sup> ed.). McGraw-Hill, New York, USA.
- Malhotra, N. K. (1996): *Marketing Research. An Applied Orientation*. 2<sup>nd</sup> ed. Prentice-Hall International. USA.
- Rosendo-Ríos, V., y Pérez del Campo, E. (2013): *Business Research Methods*. ESIC Editorial.
- Rosendo-Ríos, V.; de Esteban, J., y Antonovica, A. (2012): *MR: Development of Theoretical Concepts for Market Research I and II*. South Western, Cenage Learning. USA.
- Zikmund, W. G.; Babin, B. J.; Carr, J. C., y Griffin, M. (2013): *Business Research Methods*. 9<sup>th</sup> Edition. South Western, Cenage Learning. USA.