
TEORÍA Y PRÁCTICA SESIÓN 4: PREPARACIÓN Y CODIFICACIÓN DE DATOS

PEDRO M. VALERO MORA

ESQUEMA

1.	Partes de un cuestionario típico	2
2.	Forma de las preguntas.....	3
3.	¿Cómo deben ser las preguntas cerradas?	5
4.	Codificar las preguntas en el archivo de datos.....	8
5.	Nombres a las variables	9
6.	Codificar valores	12
7.	Valores perdidos.....	16

1. Partes de un cuestionario típico

un poco de organización ayuda

- Número de identificación
- Número de identificación
- Datos a recoger
 - Datos demográficos(Ingresos, Profesión, Nivel de estudios)
 - Información del estudio (Contexto del entrevistado, Conductas, Experiencia o status, Pensamientos y sentimientos)

2. Forma de las preguntas

La manera en la que se pregunta y el proceso de datos

- Dos tipos de preguntas fundamentales
 - Preguntas abiertas
 - Preguntas cerradas
- Preguntas abiertas
 - Ejemplos: 1033 tipo de seguro, 1039 buscador de internet, 1027 bancos/cajas
 - Son más fáciles de elaborar porque no hay que anticipar las posibles contestaciones
 - Son más difíciles de analizar

-
-
- Preguntas cerradas
 - Son más difíciles de elaborar perfectamente porque hay que pensar todas las alternativas
 - De todos modos, son las más utilizadas en la práctica.

3. ¿Cómo deben ser las preguntas cerradas?

Las preguntas bien hechas son fáciles de analizar

- Exhaustividad
 - Ejemplos positivos 106, 105, 102, 803
 - Ejemplos negativos 801-p.ejemplo revistas de viajes-, 803 –aunque quizá no lo pretendan- 809-¿vacaciones culturales?)
 - La exhaustividad es muy difícil de conseguir así que se puede usar la categoría “Otras” pero esto tiene desventajas.
 - Apilamiento: Los sujetos siempre contestan lo mismo o prácticamente lo mismo si no se dan alternativas (por ejemplo, listar sólo dos lugares de compras –pregunta 101- produciría*

eso; una forma de evitar el apilamiento es 816, jugar con el tamaño de las categorías).

–Datos no recogidos acerca de las categorías no incluidas.

- Alternativas excluyentes (ejemplo 1020 no parece considerar que alguien lo tenga ya pero esté pensando en comprar otro)
- Alternativas fáciles de comprender (yo encuentro difícil la pregunta 810)
- Categorías residuales: Hay que usar la categoría otras aunque nos pensemos que no hace falta.
- Anticipar necesidades de análisis (pregunta 102)
- Evitar que la gente no quiera responder (pregunta 102)

ACTIVIDADES

EJERCICIO 3.1 Identificar el mayor número de preguntas en el cuestionario de hábitos de consumo que tengan defectos en alguno de los aspectos considerados.

4. Codificar las preguntas en el archivo de datos

convirtiendo las preguntas a su formato informático

- El objetivo de las siguientes secciones es explicar como llegar a esto a partir del cuestionario.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Col
1	sexo	Numérico	1	0	Sexo del encu	{1, Hombre}...	Ninguno	8
2	raza	Numérico	1	0	Raza del encu	{1, Blanca}...	Ninguno	8
3	región	Numérico	8	2			Ninguno	8
4	feliz	Numérico	1	0	¿Su vida es ex	{0, No procede	0, 8, 9	8
5	vida	Numérico	1	0	¿Su vida es ex	{0, No procede	0, 8, 9	8
6	hermanos	Numérico	1	0	Número de her	{98, No sabe}..	98, 99	8
7	hijos	Numérico	1	0	Número de hijo	{8, Ocho o má	9	8
8	edad	Numérico	2	0	Edad del encu	{98, No sabe}..	0, 98, 99	8
9	educ	Numérico	2	0	Edad de añ	{97, No proced	97, 98, 99	8
10	educpad	Numérico	2	0	Edad de añ	{97, No proced	97, 98, 99	8
11	educmad	Numérico	2	0	Número de añ	{97, No proced	97, 98, 99	8
12	educesp	Numérico	2	0	Número de añ	{97, No proced	97, 98, 99	8

5. Nombres a las variables

¿Cómo se van a llamar las variables?

- Hasta la versión 11 los nombres de variables tenían como máximo 8 caracteres. En la versión 12 se pueden usar nombres más largos.
- No obstante hay que planear los nombres igualmente bien.
- Opciones para los nombres:
 - Usar números a partir del cuestionario (var1, var2, etc.)
 - Método simple de llevar a cabo y el cuestionario explica el archivo de datos.*
 - No obstante, a veces se usan preguntas que no están en el*

cuestionario.

–Las preguntas no dicen nada y los outputs no son fáciles de interpretar.

– Usar abreviaturas

–Simple y la abreviatura ayuda a recordar el sentido de la pregunta.

–No obstante, a menudo se ponen abreviaturas que son muy parecidas y que confunden

– Usar prefijos y sufijos

–El sufijo indica el tema mientras que el prefijo indica la pregunta concreta

–Se puede combinar con números para que tener también la referencia del cuestionario

ACTIVIDADES

EJERCICIO 5.1 Poner nombre a las variables de la primera sección del cuestionario de consumo utilizando las reglas descritas anteriormente. Introducir los nombres en un archivo en blanco en el SPSS.

6. Codificar valores

Poner etiquetas de valores

- Cuando se trabaja con cuestionarios se utiliza casi exclusivamente variables numéricas. Las variables

categorías también se tratan como variables numéricas normalmente

- A cada valor posible de una variable categórica se le asigna un número. Ejemplos:
 - Variables categóricas

<i>Código A</i>	<i>Código B</i>	<i>Código C</i>	<i>Código D</i>	<i>Código E</i>
Ninguna	NIN	N	a	0
Católica	CAT	C	b	1
Protestante	PRO	P	c	2
Budista	BUD	B	d	3
...

– Variables ordinales

<i>Código A</i>	<i>Código B</i>	<i>Código C</i>	<i>Código D</i>	<i>Código E</i>
Muy_frecuentemente	MFR	a	1	5
Frecuentemente	FRE	b	2	4
Ocasionalmente	OCA	c	3	3
Raramente	RAR	d	4	2
Muy_raramente	MRA	e	5	1

- Caso especial: Preguntas de respuesta múltiple.
 - Estas preguntas son especiales porque se puede dar varias respuestas a la misma pregunta.
 - En estas preguntas cada alternativa se convierte en una variable.
 - Hay dos formas de codificarlas:
 - *Método de dicotomías múltiples (cada una de las alternativas es una*

variable con su propio nombre)

–Método de respuestas múltiples (se cuenta el número de respuestas máximo y se crea un número de variables igual a éste. Luego se introducen los códigos para cada respuesta sin ningún orden particular).

ACTIVIDADES

EJERCICIO 6.1 Identificar preguntas de respuesta múltiple en la primera sección del cuestionario de consumo.

EJERCICIO 6.2 Poner nombres de variables y códigos de valores a la primera sección del cuestionario de consumo teniendo en cuenta lo dicho acerca de preguntas de respuesta múltiple

7. Valores perdidos

Hay que pensar en lo que pueda fallar

- Hay dos tipos. Los genéricos o del sistema y los definidos por el usuario.
- Genéricos
 - Se produce cuando se dejan casillas en blanco o se realizan operaciones sin sentido (dividir por cero, etc.)
 - SPSS los simboliza por un punto.
- Definidos por el usuario. Algunos casos puede ser:
 - No responde
 - No sabe

-
-
- Hay un error
 - No es aplicable la pregunta
 - etc.

ACTIVIDADES

EJERCICIO 7.1 Para la primera sección del cuestionario de consumo identificar posibles valores perdidos y códigos correspondientes.