

Codificación y tabulación de datos e información de mercados

1. Introducción.

Los datos recogidos en la investigación de mercado deben ser elaborados y clasificados con arreglo a ciertos criterios de sistematización, para proceder luego al recuento de estos conforme al sistema más adecuado o factible.

Se utiliza posteriormente el tratamiento estadístico matemático de los datos.

La masa de datos disponibles es de para utilidad si no se procede a la elaboración de estos de acuerdo con los objetivos de la investigación.

2. Objeto de la codificación y tabulación de datos.

Una vez realizado el trabajo de campo y la edición de los datos obtenidos, es necesario efectuar el procesamiento de los datos, es decir, que los mismos se preparan para ser analizados, para ello se apela a dos técnicas de elaboración de los datos: la **codificación** y la **tabulación**.

Codificar datos es asignar números a las modalidades observadas o registradas de las variables que constituyen la base de datos, así como asignar código (valor numérico) a los valores faltantes (aquellos que no han sido registrados u observados).

Ejemplo:

Si la base de datos incluye la variable Sexo, hay que asignar un número a las mujeres y otro a los hombres. Si se trata de variables cuantitativas, hay que definir el número de decimales que van a ser registrados.

Algunas consideraciones que se deben tener en cuenta en la codificación:

- La codificación de preguntas cerradas se suele presentar directamente en el cuestionario.
- En el caso de preguntas abiertas la codificación debe de llevarse a cabo una vez que se ha recolectado la información. En primer lugar, se listan las respuestas de una muestra representativa de los cuestionarios, y posteriormente se agrupan en categorías, a las cuales se les asigna un código. (normalmente alcanzan 2 dígitos).
- A las categorías NS y NC se les suelen asignar códigos separados de los previstos para el sistema de categorías de la variable.
- No es recomendable que existan más de 10 códigos por variable.
- Si una opción de respuesta recoge menos del 5% de las respuestas no es muy útil.
- La opción "Otros" no debe registrar más del 10% de las respuestas.
- Existe un principio que dice que nunca hay que introducir letras, siempre números.

Una variable categórica (por ejemplo, el nombre de la provincia donde ha nacido la persona entrevistada), no se introduce escribiendo el nombre de la categoría, sino un número, es el código numérico que le corresponde.

Así, por ejemplo, **Sevilla** puede tener el código **numérico 8**, mientras que **Cádiz** tiene asignado el **número 3**.

Los números tienen muchas ventajas.

Algunas de ellas:

- Es difícil equivocarse al introducir un número, pero fácil si se teclea un texto. Para el programa de análisis de datos, Cádiz, Cádiz, CADIZ y Cádiz con cuatro valores diferentes.

EDITORIAL TUTOR FORMACIÓN

- Es más rápido escribir “3” que “Cádiz”.
- Algunos programas de ordenador toman decisiones que no deberían tomar. Una frecuente es considerar que, si la variable es literal o de cadena, no se pueden llevar a cabo determinadas acciones. Al introducir la variable como numérica, dejamos que sea el usuario (el ser humano, por tanto) quien tome las decisiones y no una máquina.

Los datos (todos los registros) de todos los encuestados se guardan en archivos informáticos, como se ilustra en la tabla siguiente.

Esta tabla demuestra el caso más común de codificación, donde puede utilizarse más de un registro para cada encuestado.

En esta tabla, las columnas representan los campos; y las filas, los registros.

Ejemplo de un archivo de cómputo: proyecto de la tienda departamental							
	CAMPOS						
	NÚMEROS DE COLUMNA						
REGISTROS	1-3	4	5-6	7-8	26.....	35.....	77
Registro #1	001	1	31	01	6544234553		5
Registro #11	002	1	31	01	5564435433		4
Registro #21	003	1	31	01	4655243324		4
Registro #31	004	1	31	01	5463244645		6
Registro #2701	271	1	31	55	6652354435		5

Estos datos se codificaron de acuerdo con el esquema de codificación especificado en la figura las columnas **1** a **3** representan un solo campo y contienen los números de código de los encuestados 001 a 271.

La columna **4** contiene el número de registro. Esta columna tiene el valor de 1 para todas las filas, porque sólo se presenta el primer registro de los encuestados.

Las columnas **5** y **6** contienen el código del proyecto, que es 31.

Las siguientes dos columnas, **7** y **8**, muestran el código del entrevistador, que varía de 01 a 55 porque se utilizaron 55 entrevistadores.

Las columnas **26** a **35**, que representan cada una un campo, contienen las calificaciones de familiaridad para las 10 tiendas, con valores que van de 1 a 6.

Por último, la columna **77** representa la calificación de la tienda 10 en relación con los precios. Advierta que las columnas **78** a **80** están en blanco.

Hay 10 registros para cada encuestado y 2,710 filas que indican que el archivo contiene los datos de los 271 encuestados.

3. Trabajos previos a la codificación y tabulación de datos.

La codificación es un procedimiento técnico mediante el cual, los datos obtenidos se clasifican en categorías y se traducen en símbolos, ya sean cifras o letras; es decir, se asigna a cada opción de respuestas un número o una letra que permita tabularla rápidamente.

Es importante señalar que la elaboración de un sistema de categorías y la codificación se pueden efectuar en forma simultánea, pero desde un punto de vista lógico, la codificación depende del sistema de categorías o valores que adopte la variable o alternativas que presente la pregunta.



3.1. Edición de datos.

El primer paso es la edición, un análisis e interpretación de datos, un proceso de revisar y ajustar los datos obtenidos de la investigación.

La edición es la revisión de los cuestionarios con el objetivo de incrementar su exactitud y precisión.

Consiste en examinar los cuestionarios para identificar respuestas ilegibles, incompletas, incongruentes o ambiguas.

Las respuestas pueden ser ilegibles si se registraron mal, lo cual es muy frecuente en cuestionarios con un gran número de preguntas no estructuradas.

Los aspectos más importantes de la edición de datos son:

Legibilidad

Las respuestas deben ser claras, de no ser así pueden ser corregidas por la persona que las registró o deducidas a través de otras partes del instrumento.

La primera de todas las variables que debe constar en un archivo de datos es el número de caso o de registro.

En la práctica, lo que hacemos es numerar los cuestionarios y pasar a esa primera variable el número o código numérico asignado a cada uno.

De esta forma, si hay que revisar un cuestionario porque hemos encontrado una ausencia o una incoherencia en el archivo de datos en el registro con el código 71, no hay más que buscar el cuestionario número 71 y resolver el problema.

Algunas variables no tienen por qué encontrarse en el cuestionario de papel, como el código numérico del grupo de trabajo que ha recogido los datos, la zona en donde se aplicó ese cuestionario, el código numérico de la persona que hizo la entrevista...

Es posible que el cuestionario escrito cuente con espacio para indicar esta información, o tal vez no. En cualquier caso, es interesante que sí conste en el archivo de datos.

3.2. Limpieza de datos: verificación de rangos y consistencia.

El proceso de calidad de los datos se hace necesario para mantener los datos, indicadores y reportes, relevantes, precisos y consistentes, dado que en el proceso de recolección y almacenamiento se puede generar datos defectuosos que impacten la legitimidad de los análisis y los resultados.

Con datos limpios se pueden tomar mejores decisiones y asegurar que los resultados obtenidos se asemejen a la realidad del objeto de estudio.

Entre los objetivos del control de la calidad de los datos tenemos:

- Entender y documentar la calidad y confiabilidad de los datos.
- Descubrir en los datos los problemas de calidad que deben ser resueltos.
- Asegurar la estandarización e integración de los datos comunes en las diferentes operaciones estadísticas.
- Especificar las reglas de transformación y validación que deben aplicarse a los datos, para asegurar el nivel de calidad que se requiere en una migración hacia el repositorio de información.

La **verificación de rangos** consiste en comprobar que los códigos de la variable categóricas y numéricas, cualidad, y cantidad, estén entre los admisibles.

Por ejemplo, si el código de posibles respuestas es de 1 a 10, no puede existir un valor menor a 1 y superior a 10 porque estaría fuera del rango.

Podemos aplicar **dos métodos**:

Uno, dejar en blanco la respuesta y dos imputar el dato de dos formas.

- Dando un valor a la pregunta bien de forma manual o automáticamente.

EDITORIAL TUTOR FORMACIÓN

- Rellenar la respuesta con el valor medio de los entrevistados que han respondido correctamente a la pregunta.

Hay que tener en cuenta que muchas personas eligen la opción No sabe/ No contesta porque la pregunta está mal hecha.

Y por supuesto no se debe abusar de implementar la categoría No sabe/No contesta en las preguntas de los cuestionarios.



Puesto que la experiencia demuestra que muchas veces los entrevistados recurren a dicha categoría para ahorrar tiempo y evitar pensar sobre el resto de las categorías.

La **consistencia de los datos** obtenidos hace referencia a la coherencia existente entre un conjunto que datos, como, por ejemplo, detectamos *una persona viuda de diez años*. Indudablemente es una inconsistencia.

En las investigaciones asistidas por ordenador el software se encargará de implementar reglas para evitar inconsistencias.

Con la codificación mediante software se evitarán mayores errores.

- Si el número de inconsistencias es elevado hay que preguntar a los suministradores de información, es decir, a los entrevistadores.
- Siempre hay que crear una categoría de «inválidos».
- Se corrige el dato añadiendo un valor consistente con los demás explicando tal corrección.



Se trata de considerar una forma segura de identificar cada cuestionario para volver a consultarlos y relacionarlos con otros adicionales o posteriores.

Por ejemplo, crear un fichero que resuma las variables geográficas, región, provincia, municipio, y asociar un código que aparece en el cuestionario».

Es necesario para poder hacer una limpieza de los datos, determinar cuáles son cualitativos y cuantitativos en el cuestionario.

- **Análisis de datos cualitativos**

Los datos cualitativos son los que miden una cualidad no definida por número como por ejemplo de dato cualitativo sería el estado civil o la nacionalidad.

EDITORIAL TUTOR FORMACIÓN

- Los **datos cuantitativos** miden una característica mediante un número, como puede ser el peso, altura. A su vez los datos cuantitativos pueden ser discretos y continuos.

En los **datos cuantitativos discretos** los valores son enteros, por ejemplo, los miembros de la unidad familiar y en los **datos cuantitativos continuos** los valores posibles son ilimitados, por ejemplo, el peso y altura.

Los **errores** en la codificación y tabulación de datos se pueden clasificar en erráticos y sistemáticos.

- Los **errores erráticos** son aquellos que se cometen en la manipulación y codificación de la información.

Por ejemplo, al codificar una respuesta se introduce mal el código en el ordenador.

- Los **errores sistemáticos** son los que seguirían produciéndose en caso de que la encuesta se repitiera en las mismas condiciones.

Por ejemplo “las respuestas deliberadamente falsas por parte del entrevistado sobre preguntas acerca de la renta, opiniones políticas, etc.



A sí que hay que evitar en la medida de lo posible la limpieza de los errores porque ocasionarían errores de interpretación, con lo cual los datos finales no serán fiables.

Se entiende por **depuración de datos** de una encuesta al conjunto de técnicas que nos permiten a partir de la información recogida en la encuesta, y, a veces a partir de otra información adicional, corregir una parte de los errores presentes.

Actualmente, la macro depuración de la codificación y tabulación de datos supone una investigación a nivel agregado, a nivel de tablas, estudiando las distribuciones de las variables tratando de identificar áreas de problemas, por ejemplo, cuando bastantes sujetos contestan mal a una pregunta.

La macro depuración, en definitiva, cuida de no desvirtuar la calidad de los datos.

Las características de la macro depuración son:

- Se ignoran los errores cuya magnitud no es significativa o se cancelan en el proceso de agregación.
- No es un conjunto de procedimientos cerrado y una buena planificación es decisiva.
- Tiene en cuenta observaciones o valores sospechosos (outliers), índice de no respuesta, etc.
- El proceso se repite hasta que se eliminan los errores más importantes.

EDITORIAL TUTOR FORMACIÓN

La micro depuración es un enfoque tradicional, en la codificación y tabulación de datos que supone la investigación de los datos a nivel desagregado o microdato para detectar errores y corregirlos mediante imputación manual o automática.

Muchos métodos de micro depuración han sido sustituidos por los de macro depuración cuando el volumen de datos es excesivo.

Un buen diseño de estrategia de depuración debe considerar las siguientes variables:

- El **volumen de la encuesta** o número de cuestionarios. Esto determina los métodos de grabación, tareas de depuración y métodos de imputación.
- **Periodicidad de la encuesta.** La periodicidad permite corregir errores sistemáticos.
- El **grado de calidad.** Cuidar la calidad por los responsables del diseño.
- **Hipótesis** sobre tipos de error. Describir a priori posibles errores, para así detectarlos y depurarlos.
- **Grado de redundancia** de la información. Por ejemplo, pedir edad y fecha de nacimiento.
- El **tipo de datos** del cuestionario. Mayor complejidad de depuración para datos cualitativos (macro depuración ésta más orientada a datos cuantitativos).
- Definir **situaciones imposibles** (un hijo no puede ser mayor que sus padres) e improbables (alguien indica que su sueldo es de 30.000€).



Aplicación práctica

Se pide:

Investigar acerca de TEIDE. (Técnicas de Edición e Imputación de Datos Estadísticos).

Responder a las siguientes cuestiones:

1. ¿Qué es?
2. ¿Para qué se utiliza?

Existen técnicas para tratar valores perdidos: Depuración e Imputación

Técnicas de Depuración

Las estrategias de depuración más usadas son:

- **Listado de valores:** Se trata de buscar en la matriz de datos los valores que están fuera del rango de respuesta. Estos valores se pueden considerar como perdidos, o se puede estimar el valor correcto a partir de otras variables (Imputación).

EDITORIAL TUTOR FORMACIÓN

Ejemplo: En la variable Sexo, cuyos valores son 1 = Hombre y 2 = Mujer, encontramos un 3 en la matriz de datos.

- **Preguntas filtro:** Se trata de comparar el número de respuestas de una categoría filtro y otra categoría filtrada. Si se observa alguna anomalía que no pueda tener solución, se considerará como valor perdido.

Ejemplo: La pregunta filtro A tiene 11 respuestas que conduce a la pregunta filtrada B, mientras que la que conducen a la pregunta filtrada C tiene 9 respuestas. Sin embargo, observamos que en la pregunta B se han dado 14 respuestas (2 más de las que estaban previstas), por lo tanto, no existe coincidencia entre la categoría filtro y la categoría filtrada.

- **Consistencias Lógicas:** Se comprueban las respuestas que puedan ser consideradas contradictorias entre sí.

Ejemplo: Los encuestados que respondieron sobre su estado civil “Soltero” no deberían haber respondido la pregunta “Actividad del / la conyugue”.

- **Nivel de representatividad:** Se realiza un recuento del número de respuestas obtenidas en cada variable. Si el número de preguntas no contestadas es muy elevado, se puede asumir igualdad entre las respuestas y las no respuestas o bien, realizar una imputación de la no respuesta.

Técnicas de Imputación

Esta técnica consiste en sustituir los valores perdidos por valores o respuestas válidas mediante una estimación de éstas.

Existen tres tipos de imputación:

- **Imputación aleatoria:** Este tipo de imputación asume la falta de información a la aleatoriedad de la muestra.

Para realizar la imputación, se analiza la probabilidad de cada valor que aparece en la variable (válidos y perdidos), y se le asignará a cada valor perdido aquellos que tengan una probabilidad igual o menor a esa probabilidad.

Ejemplo: La probabilidad de aparezca el valor A es de 0,012 (1,2%), mientras que la probabilidad del valor B es de 0,357 (35,7%).

Por lo tanto, los valores perdidos que tengan una probabilidad igual o menor a 0,012 se les asignará el valor A, mientras a las que tengan una probabilidad mayor que 0,012 y menor a 0,369 (la suma de probabilidad A: 0,012 y la probabilidad B: 0,357), se les asignará el valor B.

- **Imputación “Hot Deck”:** En este caso, se asume que los valores perdidos no se deben a la aleatoriedad. Para realizar la imputación, se debe buscar las correlaciones entre las variables relacionadas y las variables a imputar, por lo que se usarán valores más cercanos a la variable relacionada.

EDITORIAL TUTOR FORMACIÓN

Ejemplo:

Casos	Variable Valoración del Producto	Variable Sexo	Variable Satisfacción Imputada
1	2	1	2
2	3	2	3
3	9	2	3
4	4	1	4

Como en la variable “Valoración del Producto” hay un valor perdido, se substituye este valor por el valor más cercano, teniendo en cuenta que existe correlación entre las variables “Valoración del producto” y “Sexo”.

- **Imputación de la media de subclases:** Este proceso de imputación guarda cierta similitud con el método “Hot Deck”, pero su uso se centra en variables continuas o de intervalo.

Para poder realizar la imputación, se buscan dos variables que tengan correlación con la variable a imputar, y se calcula la media en cada intervalo de la variable continua, por lo que es la media la que se imputa a los valores perdidos.

Sin embargo, este tipo de imputación tiende a disminuir la Desviación Típica o Estándar, por lo que afectaría a los resultados finales del análisis.

4. Elaboración de un código maestro.

El Codebook o código maestro son listas organizadas de códigos que se deben describen perfectamente cada código, estos pueden ser números o letras.

El libro de códigos contiene las instrucciones para la codificación y la información necesaria sobre las variables en el conjunto de datos.

El libro de códigos orienta el trabajo de los codificadores y ayuda al investigador a identificar y localizar adecuadamente las variables.

Incluso si el cuestionario se codificó de antemano, resulta muy útil preparar un libro de códigos formal.

Por lo general, un libro de códigos contiene la siguiente información:

1. Número de columna
2. Número de registro
3. Número de variable
4. Nombre de la variable
5. Número de la pregunta
6. Instrucciones para la codificación

Para cada encuestado tienen que incluirse los siguientes códigos adicionales: código de proyecto, código del entrevistador, códigos de fecha y hora, y código de validación.

Es muy deseable el uso de códigos de campo fijo, lo cual significa que el número de registros para cada encuestado es el mismo y que para todos los encuestados aparecen los mismos datos en las mismas columnas.

De ser posible, deben usarse códigos estándar para los datos faltantes. Por ejemplo, puede emplearse el código 9 para una variable de una sola columna, 99 para una variable de dos columnas y así sucesivamente.

Los códigos de los valores faltantes deben ser distintos de los códigos asignados a las respuestas legítimas.

La codificación de las preguntas estructuradas es relativamente sencilla porque las opciones de respuestas se determinan con anticipación.

El investigador asigna un código a cada respuesta de cada pregunta, y especifica el registro y la columna adecuados en que deben aparecer los códigos de la respuesta.

EDITORIAL TUTOR FORMACIÓN

Ejemplo:

¿Tiene usted un pasaporte vigente?

1. Sí

2. No (1/54)

Para esta pregunta, la respuesta "Sí" se codifica con el número 1 y la respuesta "No" con el número 2.

Los números entre paréntesis indican que el código asignado aparecerá en el primer registro de este encuestado en la columna 54.

Dado que sólo se permite una respuesta y sólo hay dos respuestas posibles (1 o 2), basta con una columna.

Por lo general, una columna es suficiente para codificar una pregunta estructurada con una sola respuesta, si las respuestas posibles son menos de nueve.

Número de columna	Número de variable	Nombre de la variable	Número de pregunta	Instrucciones de codificación
1-3	1	ID del encuestado		001 a 890 anteponga los ceros que sean necesarios
4	2	Número de registro		1 (igual para todos los encuestados)
5-6	3	Código del proyecto		31 (igual para todos los encuestados)
7-8	4	Código de la entrevista		Como se codificó en el cuestionario
9-14	5	Código de los datos		Como se codificó en el cuestionario
15-20	6	Código de tiempo		Como se codificó en el cuestionario
21-22	7	Código de validación		Como se codificó en el cuestionario
23-24		Espacio en blanco		Deje las columnas en blanco
25	8	Quién compra	I	Jefe de familia = 1 Jefa de familia = 2 Otro = 3 Ingrese el número encerrado en un círculo Valores faltantes = 9
26	9	Familiaridad con tienda 1	IIa	Para la pregunta II partes a a j, Ingrese el número encerrado en un círculo
27	10	Familiaridad con la tienda 2	IIb	No muy familiar = 1 Muy familiar = 6 Valores faltantes = 9
28	11	Familiaridad con la tienda 3	IIc	
35	18	Familiaridad con la tienda 10	IIj	
36	19	Frecuencia: tienda 1	IIIa	Para la pregunta III partes a a j Ingrese el número encerrado en un círculo
37	20	Frecuencia: tienda 2	IIIb	Ninguna = 1 Muy frecuente = 6 Valores faltantes = 9
45	28	Familiaridad con la tienda 10	IIIj	
46-47		Espacio en blanco		Deje estas columnas en blanco en calidad
48	29	Calificación de la tienda 1 en calidad	IVa1	IV a XI para las preguntas
			IVa X	Ingrese el número encerrado en un círculo
57	38	Calificación de la tienda 10 en calidad	IVa10	
58	39	Calificación de la tienda 1 en variedad	IVb1	
67	48	Calificación de la tienda 10 en variedad	IVb10	
68	49	Calificación de la tienda 1 en precios	IVc1	
77	58	Calificación de la tienda 10 en precios	IVc10	
78-80		Espacio en blanco		Deje estas columnas en blanco

EDITORIAL TUTOR FORMACIÓN

En las preguntas que permiten un mayor número de respuestas, debe asignarse una columna separada para cada opción de respuesta posible.

Dichas preguntas incluyen las que se refieren al uso o propiedad de marcas, la lectura de revistas o la preferencia por programas de televisión.

Por ejemplo, suponga que en este ejemplo un encuestado marcó cuenta de ahorros, cuenta de cheques y cuenta de ahorros a plazo fijo.

¿Qué tipo de cuenta tiene <i>actualmente</i> en este banco? (marque con "X" todas las que correspondan).	
Cuenta de ahorros	<input type="checkbox"/> (162)
Cuenta de cheques	<input type="checkbox"/> (163)
Hipoteca	<input type="checkbox"/> (164)
Inversión a la vista	<input type="checkbox"/> (165)
Club de inversión (navidad, etcétera)	<input type="checkbox"/> (166)
Línea de crédito	<input type="checkbox"/> (167)
Cuenta de ahorros a plazo fijo (depósitos a plazos, etcétera)	<input type="checkbox"/> (168)
Seguro de vida con inversión	<input type="checkbox"/> (169)
Crédito para remodelación de casa	<input type="checkbox"/> (170)
Crédito automotriz	<input type="checkbox"/> (171)
Otros servicios	<input type="checkbox"/> (172)

En el registro N° 1, se anotará un número 1 en las columnas número 162, 163, y 168.

En todas las otras columnas (164, 165, 166, 167, 169, 170, 171 y 172) se anotará un 0.

Como sólo hay un registro por encuestado, se omite el registro.

4.1. Codificación de respuestas sobre preguntas cerradas de respuesta única.

En el cuestionario de información, para pasar a la codificación de cuestionarios con preguntas para hacer una encuesta de tipo cerrada pueden ser **dicotómicas**, preguntas cerradas de **respuesta única**, y de **múltiples** opciones de respuesta.

Es decir que las podemos definir como aquellas preguntas que ofrecen una serie de opciones prefijadas en la fase de diseño, pero que sólo se puede elegir una solamente y además las **respuestas son excluyentes**.

1. Definitivas o Dicotómicas, respuestas de Sí o No

Ejemplo:

¿Tiene usted hijos?

1. Si
2. No

Aunque también son posibles otras dos opciones diferente dentro de las preguntas dicotómicas, por ejemplo:

¿Dónde reside?

1. *En España*
2. *Fuera de España*

2. Preguntas de Selección, múltiples, pero deben elegir una.

¿Cuál es su nivel de estudios?

1. *Sin estudios*
2. *Primarios*
3. *Secundarios*
4. *Ciclo Formativo*
5. *Universitarios*
6. *Postgrado*
7. *Doctorado*

La ventaja que tienen las preguntas cerradas de respuesta única es porque **son fáciles de codificar**.

4.2. Codificación de respuestas sobre preguntas cerradas de respuesta múltiple.

Son aquellas preguntas que dan una serie de opciones prefijadas, entre las cuales el entrevistado puede elegir más de una opción.

Las respuestas **No** son excluyentes entre sí.

Por ejemplo:

¿Qué aficiones practica el fin de semana?

- *Lectura,*
- *Ver TV*
- *Ir al cine o teatro*

La Pregunta queda dividida ahora en 6 preguntas convertidas en **dicotómicas**.

¿Lees el fin de semana?

- Sí
- No

¿Ves la televisión el fin de semana?

- Sí
- No

¿Vas al cine el fin de semana?

- Sí
- No

4.3. Codificación de respuestas sobre preguntas abiertas.

Son los más difíciles de tratar.

Las respuestas abiertas deben ser categorizadas antes de ser introducidas en el ordenador.

Para ello, el proceso es:

1. Tomar notas breves de cada respuesta.

Por ejemplo, hemos preguntado por las preferencias de destinos vacacionales.

Alguien respondió *“Me gusta mucho la tranquilidad, el silencio, pero con buenos servicios y comunicaciones”*. *“Me gusta el turismo rural, la montaña, la naturaleza”*.

De aquí se puede anotar, por ejemplo *“tranquilidad / buenos servicios / rural”*.

2. Tras analizar un número suficiente de respuestas (“suficiente” para concluir con cierta seguridad este segundo punto), revisamos nuestras notas y extraemos las dimensiones o variables que se están considerando en las respuestas.

Por ejemplo:

“Existencia de servicios” con los valores (“mencionado” / “no mencionado”), “Entorno” con los valores (“Urbano” / “Rural” / “Mixto” / “Otros”), etc.

En este mismo proceso, se procura no generar variables con muchas categorías, por lo que hay que saber agruparlas convenientemente (por ejemplo, la categoría “rural” puede aglutinar las referencias a “turismo rural”, “vacaciones en el campo”, “senderismo”, “montaña”, etc.).

3. Se pone a prueba el sistema de variables/categorías generado con más respuestas provenientes de otros cuestionarios, realizando las modificaciones que se consideren oportunas.
4. Con el sistema ya robusto, se codifica, es decir se asignan códigos numéricos a las categorías.

Por ejemplo: variable **Servicios**, cuya descripción es “Existencia de servicios en la zona” y los valores son: 1. Mencionado / 0. No mencionado.

5. Desde el primer cuestionario, se codifican las respuestas a esa pregunta.

Si se han considerado tres variables para expresar las respuestas a ese ítem, cada registro o caso debe suministrar tres códigos numéricos.



Aplicación práctica

Se pide:

Elaborar un cuestionario de 12 preguntas sobre un tema de libre elección en el que se recojan preguntas cerradas de respuesta única/múltiple y abiertas.

4.4. Utilización de hojas de cálculo para la creación de tablas de doble entrada para el registro de los datos.

El sistema **SPSS** es un programa estadístico informático muy usado en las empresas de investigación de mercado.

SPSS es uno de los más conocidos paquetes estadísticos, sirve para todo, se puede utilizar en psicología, mercado, en la codificación de datos de redes, etc. también está considerado como el mejor software que existe de programas para tabular y graficar encuestas.

Consiste en un módulo base y módulos anexos que se actualiza constantemente

Podemos encontrar 3 diferentes hojas de cálculo:

- Tabular en Excel, cómo tabular una encuesta en Excel, con Microsoft Office es una aplicación utilizada en tareas financieras y contables, de cómo hacer gráficas de una encuesta.