

UF1781: Tratamiento y análisis de la información de mercados

TEMA 1. Codificación y tabulación de datos e información de mercados

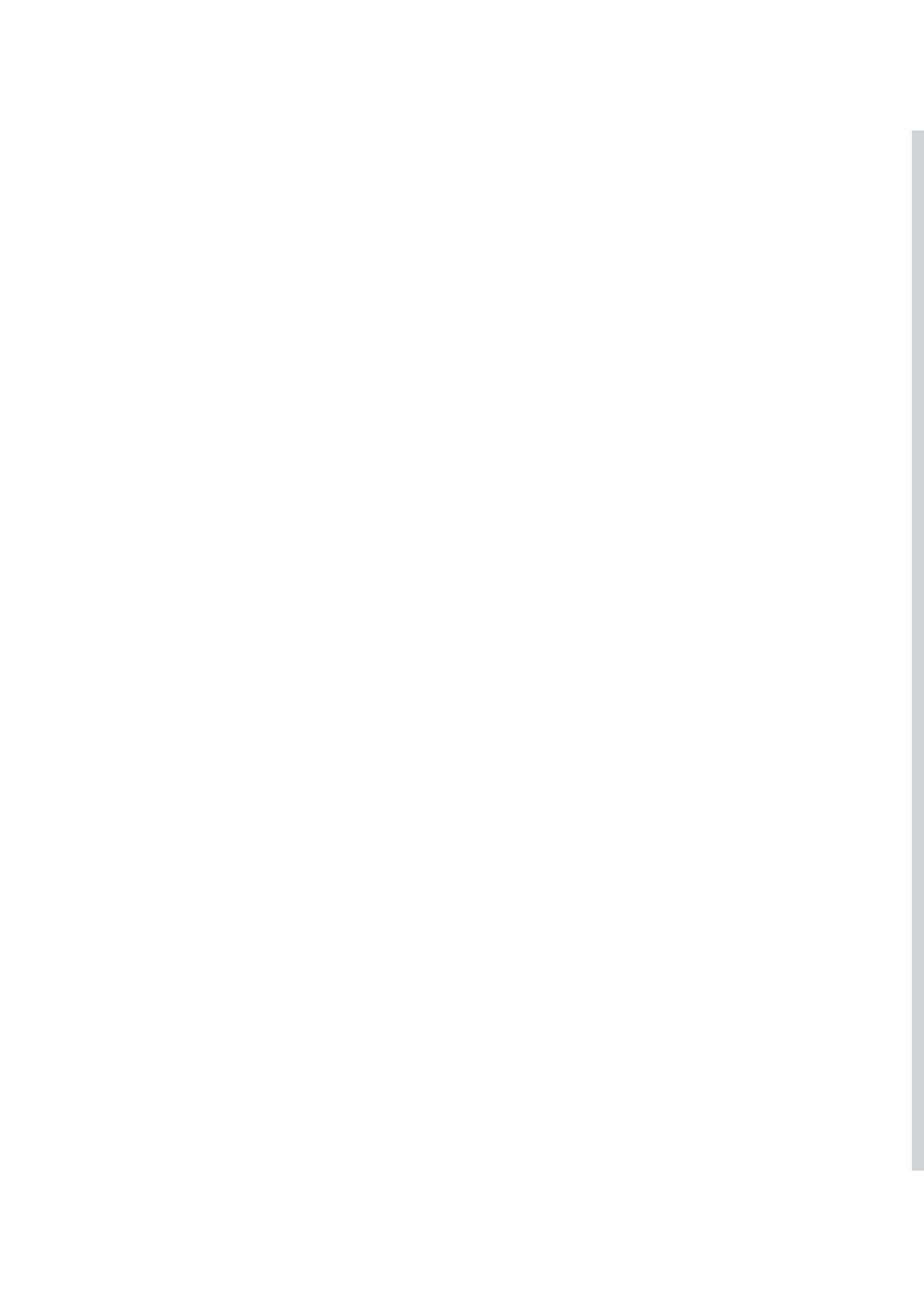
TEMA 2. Análisis estadístico de la información de mercados

TEMA 3. Informes y presentaciones comerciales de la información de mercados

OBJETIVOS

- Aplicar técnicas estadísticas y de tratamiento de datos a la información disponible en el SIM para facilitar su análisis, interpretación y posterior presentación en informes comerciales
- Elaborar informes a partir de la información gestionada por el SIM para su transmisión en los soportes informáticos y/o físicos adecuados y por los canales establecidos





Codificación y tabulación de datos e información de mercados

1. OBJETO DE LA CODIFICACIÓN Y TABULACIÓN DE DATOS

La codificación y tabulación de datos es una de las fases más importantes de la investigación de mercados. En ella se facilita la

- Objeto de la codificación y tabulación de datos
- Trabajos previos a la codificación y tabulación de datos
- Elaboración de un código maestro
- Tabulación de datos

OBJETIVOS:

- Efectuar los trabajos de edición y limpieza de datos
- Describir los métodos de elaboración de códigos maestros y los distintos tipos de tabulación de datos
- Utilizar las hojas de cálculo para realizar tablas de doble entrada

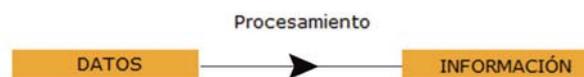
labor de análisis y extracción de conclusiones a partir de los valores obtenidos, mediante la simplificación y ordenación de los mismos.

El **proceso de codificación** supone la agrupación de los datos en categorías que los engloben, expresadas mediante símbolos numéricos. De este modo, se puede procesar y tratar la información de una forma más simple y rápida.

El **proceso de tabulación** consiste en el recuento de las respuestas a los cuestionarios y su posterior disposición en tablas, de modo que se puedan estudiar comportamientos comunes y tendencias de una forma práctica y visual.

Aunque a simple vista pueda parecer una etapa sencilla o de menor importancia, se debe prestar atención a su realización. Los datos recopilados suelen ser cuantiosos y muy diferentes en su adecuación al tema objeto de estudio. Además, durante la recogida de información pueden darse errores, tales como preguntas sin responder, anotaciones inexactas o respuestas diferentes a preguntas similares realizadas a una misma persona (inconsistencia de datos). Estos errores se deberán subsanar durante esta fase.

La base de todos estos errores es la **diferencia existente entre los datos y la información**, aunque a veces se usen como sinónimos. Los *datos* son las respuestas que se pueden obtener acerca del tema de estudio, observadas o recogidas directamente en el entorno de ocurrencia de la materia a investigar. Sin embargo, la *información* hace referencia a las conclusiones extraídas de los datos una vez editados, codificados, tabulados y analizados, cuyo fin último es la toma de decisiones.



Diferencia entre datos e información

Podemos concluir pues, que de la encuesta se recogen datos y que de ellos se obtiene la información útil para la empresa o institución. Por lo tanto, la creación del trabajo de campo y la recogida de datos no tienen significado en la investigación si no se lleva a cabo posteriormente la tarea de codificación, tabulación y análisis de los mismos.

1.1 Breves apuntes sobre la Ley de Protección de Datos

La información que se analiza en una empresa u organización, tanto de carácter primario como secundario, puede contener datos personales o sensibles que deben protegerse contra un mal uso o robo de los mismos. En este sentido, se aprobó la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (LOPD), y se creó la Agencia Española de Protección de Datos (AEPD), que vela por su buen cumplimiento.

En primer lugar, las organizaciones que cuenten con este tipo de ficheros deben cumplir los siguientes requisitos:

- Notificar la existencia de los ficheros al Registro General de Protección de datos, junto con el nombre del responsable del mismo.
- Informar a los titulares de la recogida y posterior almacenamiento de sus datos.
- Asegurarse de la fiabilidad de los datos y su tratamiento ajustado a la finalidad establecida.
- Proteger esos datos frente a robos o el mal uso de los mismos.
- Facilitar a los titulares sus derechos de acceso, rectificación, oposición y cancelación de los datos.

Además, hay que tener en cuenta que la LOPD afecta tanto a archivos de datos físicos como electrónicos, debiéndose seguir unas pautas de actuación diferentes para cada uno de estos ficheros. A continuación se analizarán las características que han de poseer los mismos.

A la hora de crear el fichero, se debe realizar un registro que indique los datos que se van a archivar, de forma que se localicen fácilmente y puedan ser, a su vez, controlados por la AEPD. De esta forma, cada uno de los archivos debe disponer de una ficha que contenga:

- Denominación del fichero.
- Código y localización dentro del archivo (tanto en formato físico como electrónico).
- Origen de los datos.
- Estructura interna del fichero.
- Comunicación de datos prevista (indicar si se efectuarán transferencias de datos, así como sus destinatarios, especialmente si se realizaran a nivel internacional).
- Órganos responsables del fichero.
- Servicios o unidades ante los que ejercer los derechos de acceso, rectificación, cancelación y oposición.
- Nivel de seguridad del fichero (bajo, medio y alto, dependiendo del carácter sensible de los datos).

Existen varios tipos de archivos en función de su operatividad dentro de la empresa. En este sentido, se distingue entre:

- **Archivo administrativo.** Es aquél cuya finalidad es de índole administrativa. Son aquellos ficheros que se constituyen la base de trabajo en la empresa, y que deben consultarse con frecuencia.
- **Archivo histórico.** Está constituido por aquellos ficheros importantes que, habiendo perdido su validez administrativa, se guardan como prueba histórica de forma ordenada y selectiva.

En unidades posteriores se analizarán de forma detallada los métodos de archivos de datos en una empresa u organización.

2. TRABAJOS PREVIOS A LA CODIFICACIÓN Y TABULACIÓN DE DATOS

2.1 Edición de datos

La fase de edición se refiere a la revisión de los datos recogidos en el trabajo de campo, de modo que se puedan detectar respuestas ilegibles o inconsistentes, así como preguntas sin responder. Se pueden distinguir diferentes tipos de edición de datos: de campo e interna.

A. Edición de campo

Se refiere a la revisión que se efectúa **dentro del trabajo de campo**. Tras la finalización de la recogida de datos, los entrevistadores o sus supervisores inspeccionan los documentos de recopilación, a fin de identificar errores por ilegibilidad, multiplicidad u omisión de respuestas.

Este tipo de errores se puede subsanar directamente preguntando al encuestador que haya registrado los datos, o bien recurriendo a la persona de la que se hayan obtenido los mismos, para verificar que se comprende la respuesta, o interrogar con mayor profundidad, si se precisa más información.

La conveniencia de realizar esta fase en el mismo escenario del objeto de estudio deriva de la necesidad de obtener un elevado recuerdo de los datos cedidos por las personas intervinientes en la investigación. El nivel de recuerdo será más elevado cuanto más cercana en el tiempo esté la revisión de los datos, asegurando así una mayor fiabilidad de la edición de campo.

B. Edición interna

Se trata de la revisión que se efectúa fuera del campo de trabajo. Es decir, se realiza una **investigación más profunda de los resultados**, con el fin de descubrir otro tipo de errores no detectables en la edición de campo, tales como las respuestas contradictorias; la falta de orden en las cuestiones, sobre todo en las entrevistas en profundidad; o la existencia de respuestas del tipo “no sabe/no contesta”, que pueden deberse a un verdadero desconocimiento del tema o bien ocultar una información que no se desea compartir con el encuestador.

La subsanación de estos errores es más compleja que en la edición de campo, pudiéndose usar varios procedimientos:

- Insertar en las contestaciones erróneas una **media de las respuestas** obtenidas a esas preguntas en el resto de la muestra objeto de estudio. Por ejemplo, si se está estudiando la edad de los empresarios que exportan sus productos y un individuo de la muestra no ofrece esa información, se puede valorar ese punto con la media de edad del resto de la muestra.
- Darles un **valor diferenciado**, de modo que se distingan los datos erróneos a simple vista, y puedan señalarse en el programa informático de análisis estadístico, de forma que no se tengan en cuenta en los cálculos posteriores. Siguiendo con el ejemplo anterior, a la edad de ese empresario se le puede



Edición interna de los datos de la investigación

dar un valor 999, fácilmente identificable en todos los análisis, tanto manuales como informatizados.

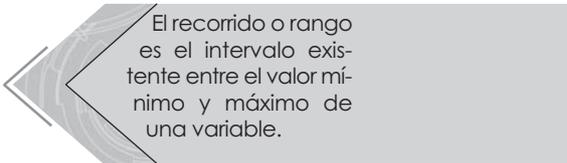
- **Eliminar los valores erróneos**, dejando un espacio en blanco en esa observación concreta. Continuando con el mismo ejemplo, se puede omitir en el cuestionario la pregunta acerca de la edad del empresario que no ofreció ese dato en la investigación.

No obstante, estas prácticas pueden deteriorar o desviar los resultados del análisis, por lo que deben usarse con precaución. Si se descubren varios datos inexactos o erróneos dentro de un mismo cuestionario, se valorará la desestimación completa del mismo.

2.2 Limpieza de datos: verificación de rangos y consistencia

En este punto se estudia la fiabilidad que se le puede dar a la información recogida, teniendo en cuenta que ya se ha realizado la edición de datos. Esta confianza se estudia tanto en función del recorrido o rango en el que puede fluctuar una variable, como en relación a la consistencia de los datos.

En primer lugar, la **verificación de rangos** se refiere a la presencia de observaciones fuera del recorrido de esa variable. Es decir, se



El recorrido o rango es el intervalo existente entre el valor mínimo y máximo de una variable.

RECUERDE

Las observaciones extremas u outliers son las que distan en gran medida del resto de observaciones. Por ejemplo, si se estudia la edad de los individuos que componen un aula de primaria, la edad de los profesores sería una observación extrema, ya que es mucho más alta que la de sus alumnos.

compara cada observación con los valores mínimos y máximos, para asegurarse de que se encuentra dentro de los valores posibles de ocurrencia. Por ejemplo, si se lanza un dado, el resultado no puede ser un siete, pues su rango varía del uno al seis.

Por lo tanto, mediante la verificación de rango se pretenden identificar los valores que no sean válidos, los que sean inusuales (es decir, que aun estando dentro del rango, no son los valores que normalmente se espera que acontezcan), y las observaciones extremas u *outliers*. Además, con esta verificación, puede empezar a distinguirse la distribución o forma que sigue la variable, aspecto muy importante a la hora de analizar los resultados estadísticamente.

En segundo lugar, se realiza la **verificación de consistencia**, relacionando respuestas a preguntas diferentes dentro del mismo estudio. De esta forma, se determinan las combinaciones de valores que son válidas. Por ejemplo, si se realiza una investigación acerca de las empresas andaluzas que exportan, no puede darse el caso de que una empresa indique en una pregunta que posee una larga tradición exportadora hasta la actualidad, y al mismo tiempo responda a otra que el número de países con los que comercia es igual a cero.

En este sentido, la verificación de consistencia se encarga de la detección de combinaciones de datos inválidas o inusuales, además de distinguir la forma de relación entre las variables (lineal, exponencial, etc.), información, de nuevo, muy útil para el análisis estadístico.

3. ELABORACIÓN DE UN CÓDIGO MAESTRO

Una vez los datos se encuentran editados y verificados, estarán listos para su codificación. Esta fase, como vimos anteriormente, conlleva la asignación de un código numérico único a

ACTIVIDAD 1

Busque un ejemplo de error de rango y otro de consistencia de datos.

cada pregunta y respuesta del cuestionario, de modo que se facilite su registro y procesamiento en un programa informático de análisis estadístico.

De esta forma, los datos codificados se encuentran directamente clasificados por respuestas similares, agrupando conjuntos de individuos con las mismas características; aspecto éste de gran utilidad a la hora de extraer conclusiones de la muestra.

Las claves que se emplearán en los cuestionarios variarán según el tipo de encuesta o pregunta de que se trate. Debe recordarse que es interesante crear un **libro de códigos**, en el que se detallan todos los usados en las encuestas analizadas. De este modo, a la hora de presentar los informes oportunos, efectuar revisiones posteriores de las conclusiones extraídas o utilizar esta clave única en investigaciones posteriores, se contará con un escrito en el que se detalla el uso de las mismas.

A continuación, se estudiarán los casos más comunes de cuestionarios y método de codificación que se recomienda usar.

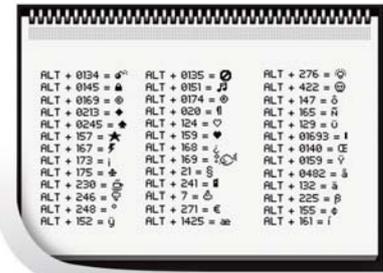
3.1 Codificación de respuestas sobre preguntas cerradas de respuesta única

Las **preguntas cerradas de respuesta única** son aquellas cuestiones que ofrecen varias respuestas, pudiendo el individuo de la muestra elegir solo una de ellas (la que más se ajuste a su situación).

Un **ejemplo** de este tipo de preguntas sería: “¿A qué sector se dedica su empresa?”, siendo las respuestas posibles “Alimentación y bebidas”, “Industrial”, “Calzado y textiles” u “Otros”.

Existen dos tipos de preguntas cerradas: las de **opción múltiple de respuesta**, como la del ejemplo anterior; y las **dicotómicas**, en las que solo hay dos alternativas de contestación, que normalmente suelen ser “Sí” / “No”, o “Verdadero” / “Falso”.

En este tipo de preguntas, los códigos se suelen establecer durante la creación de los enunciados del estudio, de modo que los cuestionarios impresos ya incluyan las respuestas codificadas. Este aspecto es de vital importancia en encuestas asistidas por ordenador, como CAPI o CATI. Sin embargo, mientras se esté codificando, se deben tener en cuenta todas aquellas respuestas



Elaboración de un código maestro



que los individuos de la muestra pueden dar a las preguntas, para que no quede sin especificar en el cuestionario ninguna circunstancia.

Entre sus características principales se puede destacar que:

- Se trata de **códigos excluyentes**, es decir, dos categorías de respuestas no pueden tener la misma clave. Así, se asegura que ninguna pregunta tendrá dos respuestas.
- Son **códigos completos**, cada respuesta tiene una numeración, de modo que no queda ninguna contestación o circunstancia que no esté codificada.

Siguiendo con el ejemplo de este punto, se puede codificar de la siguiente manera:

0100	¿A qué sector se dedica su empresa?	
	0101	Alimentación y bebidas
	0102	Calzados y textiles
	0103	Industrial
0200	¿Se ha planteado la posibilidad de exportar sus productos?	
	0201	Sí
0300	Si se ha planteado la salida al exterior, ¿qué país sería su primera opción?	
	0301	Marruecos
	0302	Francia
	0303	Alemania
	0304	Otros

Como se puede observar, cada pregunta tiene su código, relacionado por las dos primeras cifras a la cuestión a la que responde. Este proceso se lleva a cabo en cada una de las preguntas y respuestas del citado cuestionario.

3.2 Codificación de respuestas sobre preguntas cerradas de respuesta múltiple

Las **preguntas cerradas con respuesta múltiple** son aquellas que ofrecen varias opciones de contestación, pudiendo el encuestado elegir una, varias o todas las respuestas ofrecidas, dependiendo de su contexto personal concreto.

Por **ejemplo**, la pregunta “¿A qué países exporta actualmente su empresa?” tiene como opciones “Marruecos”, “Francia”, “Portugal”, “Reino Unido” y “Otros”, pudiendo el empresario elegir varias alternativas o incluso todas, siempre que sean correctas para su situación.