

PROCESAMIENTO DE INFORMACIÓN

El procesamiento de información se compone de las actividades que se realizan desde que se termina de aplicar las encuestas (o cualquier otro instrumento) hasta que se tiene una base de datos organizada para hacer análisis estadístico y sacar conclusiones.

Estas actividades consisten en revisar cuestionarios, codificar respuestas, capturar datos, detectar errores y corregirlos y, por último, organizar la base de datos.

Si estas actividades no se llevan a cabo de forma correcta, la investigación de mercados puede dar resultados falsos o confusos, aunque el cuestionario esté bien diseñado.

En investigación de mercados, normalmente se siguen estas etapas:

1. **Revisión y depuración de cuestionarios.**
2. **Codificación de las respuestas.**
3. **Captura de datos en una base.**
4. **Limpieza y depuración de la base de datos.**
5. **Organización final para el análisis.**

Analicemos estas etapas.

1. Revisión y depuración de cuestionarios (u otros instrumentos)

En este paso debemos revisar que los cuestionarios:

- Estén completos (no falten demasiadas respuestas).
- Sean legibles (marcas claras, números entendibles).
- Tengan coherencia básica (por ejemplo, alguien que pone 10 años y dice que trabaja tiempo completo... no es una respuesta real).

Debemos separar los cuestionarios llenos en su mayoría de los cuestionarios con errores graves (que quizá no se usen). Este paso es importante porque evita que introduzcamos al sistema datos mal hechos que pueden afectar los resultados (promedios raros, porcentajes equivocados, etcétera).

2. Codificación de las respuestas

Aquí convertimos las respuestas en números o códigos que el programa pueda usar. En investigación de mercados, casi siempre se hace esto con las preguntas de opción múltiple y escalas.

Ejemplo de codificación:

Pregunta: *Sexo*

- 1 = Hombre
- 2 = Mujer

Pregunta: Bebida que consume al estudiar

- 1 = Café
- 2 = Agua
- 3 = Refresco
- 4 = Té

Pregunta: Nivel de satisfacción con el servicio

- 1 = Muy insatisfecho
- 2 = Insatisfecho
- 3 = Neutral
- 4 = Satisfecho
- 5 = Muy satisfecho

Diccionario de códigos

Es recomendable hacer un pequeño documento donde se anote:

- Nombre de la variable (SEXO, BEBIDA, SATISFACCION).
- Qué significa.
- Qué código corresponde a cada respuesta.

Esto ayuda a que cualquiera que use la base la entienda.

3. Captura de datos en una base

Una vez codificadas las respuestas se capturan en un archivo que puede ser en:

- Excel, Google Sheets o similar.
- Programas estadísticos: SPSS, R, Stata, Jamovi, etcétera.

Vaciar la información de forma correcta es esencial para un correcto procesamiento, para esto es importante saber que:

- Cada fila = una persona / un cuestionario.
- Cada columna = una pregunta o variable.

Por ejemplo:

| ID | Sexo | Edad | Bebida | HorasEstudio | Satisfacción |
|----|------|------|--------|--------------|--------------|
| 1 | 1 | 18 | 1 | 3 | 4 |
| 2 | 2 | 19 | 2 | 2 | 3 |
| 3 | 1 | 20 | 1 | 4 | 5 |

Debemos capturar con mucho cuidado y tener estas precauciones:

- No “recorrer” columnas (un error típico es poner una respuesta en la columna equivocada).
- Revisar una parte de los datos (por ejemplo, 10 - 20%) comparando con los cuestionarios físicos para asegurarse de que se capturó bien.

4. Limpieza y depuración de la base de datos

Ya que está todo capturado, toca limpiar:

a) **Datos faltantes.** Las celdas pueden estar vacías porque:

- El encuestado no contestó o se saltó la pregunta.
- Hubo error de captura.

Luego tendremos que tomar una decisión respecto a estos faltantes:

- Dejar el dato como faltante (muchos programas aceptan “NA” o similar).
- En algunos análisis, excluir esos casos para esa variable.

b) **Valores extraños o imposibles.** Elementos como:

- Edad = 150.
- Horas de estudio = 40 horas diarias.
- En escala de 1 a 5 aparece un 9.

Ante estos datos se puede tomar las siguientes decisiones:

- Si fue error de captura → se corrige.
 - Si fue error del encuestado → se puede marcar como dato inválido o faltante.
- c) **Coherencia entre variables:** Por ejemplo, si alguien responde “no ha usado nunca el servicio” y luego califica la calidad del servicio, hay contradicción. Puede que haya que excluir ese cuestionario o esas respuestas.
- d) **Registros duplicados:** Por ejemplo, encontrar dos filas exactamente iguales que parecen ser la misma persona. Esto se corrige o se marca para que no distorsione los resultados.

5. Organización final para el análisis.

Cuando la base ya está limpia, se da un último paso: prepararla para el análisis estadístico. Se procede a:

- **Recodificar variables**

Agrupar categorías si hay muy pocas observaciones. Por ejemplo: si solo 3 personas dijeron “refresco”, se puede agrupar en “otras bebidas”.

- **Crear nuevas variables, si se necesitan**

Por ejemplo, con la respuesta de “ingreso mensual” crear rangos:

Bajo = 1 / Medio = 2 / Alto = 3

Se pueden sumar varios ítems de satisfacción para crear un índice total.

- **Poner etiquetas claras**

Asignar en el programa nombre y descripción de cada variable para saber qué es.

- **Guardar versiones de la base**

Hacer un archivo con la base original (solo capturada), generar otro archivo con la base limpia, crear otra base final para análisis.

Referencia:

Mita Arancibia, Erick Gregorio. (s.f.) Investigación de Mercados. Recuperado de:
https://usfx.bo/Documentos/RepositorioLibros/Investigacion_de_mercados.pdf