

# **Informe Encuesta Metodología y Técnicas Cuantitativas**

Metodología Cuantitativa

15/2/23

# **Tabla de contenidos**

# Presentación y objetivos del informe

El siguiente informe tiene una finalidad principalmente pedagógica. Específicamente está pensado como un mecanismo para ayudar a alcanzar los siguientes objetivos que forman parte del [programa](#) de la materia “Metodología y Técnicas Cuantitativas” de la UNAJ:

- Lograr un conocimiento mínimo de la existencia y pertinencia de las técnicas de análisis de datos *básicas* y una *habilidad* mínima en la ejecución de las mismas.
- *Lograr un* conocimiento mínimo de la existencia y pertinencia de (otras) técnicas de análisis de datos más específicas y (usualmente) más *complejas*.

Aparte de la distinción entre técnicas básicas y complejas la diferencia en cuanto a los objetivos pedagógicos es que entre ambos objetivos es que las últimas sólo se aspira a conocerlas mientras que en las primeras se espera, también, que se adquiera cierta habilidad en su ejecución. Expresado de otro modo, de las técnicas consideradas complejas se espera que se sepa de su existencia y qué tipos de problemas ayuda a solucionar. En cambio, en las técnicas básicas se espera lo anterior pero que además se adquiera la habilidad de poder ejecutar las mismas.

Para lograr lo anterior durante la cursada de la materia se realiza una encuesta mediante un [formulario de google](#) que contestan los mismos estudiantes. Luego el contenido de sus respuestas se analiza con el [lenguaje R](#) y con el mismo se escribe y publica este informe.

En cuanto al estilo del texto será deliberadamente informal aunque con 3 excepciones con fines principalmente pedagógicos:

- Se incluirán piezas de código
- Se hará uso de citas y referencias (estilo APA)
- Los conceptos importantes aparecerán traducidos al inglés entre paréntesis.

Por otro lado, el texto estará acompañado de los siguientes elementos visuales:

# 1 Diseño de la encuesta

En el diseño de la encuesta que contestan los estudiantes se pueden distinguir 2 grandes procesos. Uno referido al diseño del cuestionario y otro referido al diseño de la selección de los casos. El primero se preocupa por **qué se pregunta** y el segundo por **a quienes**. En esta sección la mayoría de las veces sólo se explicitará pero pocas veces se justificará las decisiones tomadas en ambos procesos. En otras palabras, la palabra “diseño” le queda grande a ambos procesos para esta encuesta. Lamentablemente lo mismo puede afirmarse de varias encuestas académicas o profesionales. La justificación particular para la falta de atención metodológica a estos puntos es que en este caso se trata de una actividad con fines principalmente pedagógicos más que académicos o profesionales.

## 1.1 Diseño del cuestionario

El **cuestionario**, como muchos cuestionarios que se usan en la práctica profesional, está diseñado con la lógica de módulos. Como veremos más adelante en esta instancia cada módulo tiene una justificable lógica interna pero lo que no tiene este cuestionario es una coherencia interna que explique la relación entre los diferentes módulos.

El diseño de un cuestionario mediante la estrategia de módulos suele ser útil tanto para el encuestado como para el investigador. Al encuestado lo ayuda a ordenarse visualmente especialmente si se trata de un cuestionario largo. Al investigador lo ayuda tanto para el etiquetamiento (o nombramiento) como para el orden de las variables en la base de datos.

El cuestionario por ahora cuenta con los siguientes módulos:

1. Identificación
2. Demográfico
3. Composición del hogar
4. Cuidados
5. Vivienda
6. Uso del tiempo
7. Inseguridad alimentaria

8. Preferencias sociales
9. Origen social
10. Trabajo actual
11. Ingresos del hogar
12. Académico UNAJ
13. Expectativas materia cuantitativa
14. Redes sociales entre estudiantes de la materia

Claramente no parece haber mucha relación entre los diferentes módulos. A diferencia de algunas encuestas ómnibus en donde este problema se presenta porque cada investigador que forma parte de la investigación agrega su propio “paquete de preguntas” para su particular investigación en este caso la justificación es que cada módulo tiene una función pedagógica particular.<sup>1</sup>

## 1.2 Diseño de la selección de casos

En cuanto al criterio utilizado para la selección de los casos se puede afirmar que se trata de una muestra de los estudiantes de la materia “Metodología y Técnicas Cuantitativas” de la carrera de Trabajo Social de la Universidad Nacional Arturo Jauretche. Esto ya afirma algo pero se puede especificar aún más.

Antes que nada, si bien es una encuesta que potencialmente les llega a todos los estudiantes que cursan la materia, la misma efectivamente sólo llega a una muestra de los mismos. Esto se produce principalmente por la no-respuesta de algunos estudiantes que comienzan la cursada pero no responden la encuesta. El factor “deserción” no parece afectar tanto a la selección de casos dado que la encuesta se realiza al principio de la cursada. Esto hace que la encuesta, si es representativa de algo, lo sea de la población intermedia que queda conformada entre los inscriptos a la materia y los que finalmente la regularizan. Obviamente esto no alcanza para decir mucho acerca de la representatividad de la encuesta sobre poblaciones mayores compuestas por:

- Los estudiantes de la carrera de Trabajo Social de la UNAJ.
- Los estudiantes de la UNAJ
- Los estudiantes del sistema universitario nacional

---

<sup>1</sup>En una investigación ómnibus varios investigadores (o clientes) comparten el mismo diseño de selección de casos y de esa manera, juntos, pueden llegar a más casos ya que entre todos amortizan estos costos que suelen ser grandes en encuestas grandes presenciales. A cambio, más allá de un algunos módulos comunes (p.e. el demográfico), cada investigador o cliente agrega su módulo de particular interés.

## 2 Limpieza y consistencia de los datos

El producto de los procesos de producción y registro de los datos suele cristalizarse en una (o varias) bases de datos. Estas pueden (y suelen) contener diferentes tipos de errores por lo que se considera una buena práctica realizar un proceso de limpieza y preparación para recién después comenzar el proceso estricto del análisis de los mismas.

En esta sección veremos algunos ejemplos tanto de limpieza, consistencia y construcción de nuevas variables. Aquí veremos ejemplos de los casos mas sencillos. Procesos como el pegado (*joint*) de variables, necesario cuando los datos se encuentran en diferentes archivos o bases de datos, no se verán aquí.

### 2.1 Limpieza

La idea de limpieza (*cleaning*) viene de usar la metáfora de dato sucio (*dirty*). Un dato sucio no necesariamente es un dato incorrecto aunque sí se trata de un tipo de dato más incómodo de trabajar ya que dificulta el posterior análisis.

La tarea de la limpieza de una base de datos suele implicar un tiempo, especialmente en bases de datos con muchas variables (muchas columnas). Si bien es una práctica recomendada siempre hay que tener en cuenta la “escala” del trabajo a realizar porque que en algunas situaciones puede que sea más simple corregir algunas cuestiones de manera artesanal o a mano en el proceso mismo de la publicación de los análisis. Esto suele ser particularmente cierto en algunos de los siguientes escenarios y sus posibles combinaciones:

- Encuestas con muchas preguntas en donde se sabe que se van a analizar una sola vez o que se analiza sólo una pequeña parte de la información disponible.
- Base de datos y procesos de análisis que luego no se van a compartir (los resultados se comparten pero estos no son replicables por terceros)

De manera complementaria el proceso de limpieza se amortiza considerablemente cuando los datos se van a analizar más de una vez y cuando se requiere un grado de transparencia que exige que la investigación sea enteramente replicable.

### 2.1.1 Renombre de las variables

La tarea básica de limpieza (*cleaning*) que aquí se hará será el **renombrar de todas las variables**. La razón de esta operación es que, al menos si se trabaja con google forms (aunque algo similar suele suceder con otros sistemas de formularios online), los nombres de las variables son el texto de la propia pregunta del formulario. Esto incomoda un poco el análisis de los datos por la gran extensión de algunas preguntas y esa incomodidad se traduce en una mayor propensión al error. En este sentido para trabajar con los nombres de las variables suele ser recomendable:

- Eliminar los espacios entre las palabras agregando algún símbolo que las pegue o una (“\_”, “-” o cualquier otro),
- Pasar todas las letras a minúscula (o mayúscula)
- Renombrar el nombre original con un nombre más corto y recordable. Una opción recomendable es que cuando en la encuesta haya variables que corresponde a un mismo módulo (p.e. módulo vivienda) se inicien por un mismo prefijo (viv\_habitaciones, viv\_inodoro, etc.).

### 2.1.2 Orden de las variables

Otra tarea diferente pero relacionada con la anterior tiene que ver con el **orden de las variables** en la base de datos. En algunos sistemas el orden de las variables tiene que ver con el orden cronológico en que se fueron construyendo las preguntas (p.e. en una encuesta) o con el orden en que luego se fueron construyendo variables complejas o recategorizaciones. En cualquier caso, lo que se debe tratar de lograr es que las variables que tienen relación temática entre sí, no sólo tengan un prefijo que las una sino que se encuentren visualmente cercanas en la base de datos. Esta recomendación es más importante cuanto más variables tenga la base de datos.

### 2.1.3 Etiquetado de las variables

Antes se habló de renombrar y ordenar las variables. Sin embargo, en las ciencias sociales y en especial en aquellas disciplinas donde se encuentren difundidos programas estadísticos como el SPSS, SAS y Stata se usa la distinción entre **nombre de la variable y etiqueta de la misma**. El primero es como el nombre real de la variable y así lo entiende el mismo programa. La segunda es como un alias o un metadato que permite una interpretación más humana del significado de la variable. Muchas veces el contenido de la etiqueta se acerca a la pregunta original. Ahora bien, las variables, especialmente las que se suelen denominar discretas o categóricas, aparte de tener un nombre pueden contener una serie finita de categorías.

Cuando se trata de variables del tipo “Indique su cantidad de hijos” si la respuesta de la base de datos es un “2” se entiende que la persona ha respondido que tiene 2 hijos. Lo mismo si aparece un “1” o un “3”. Se dice que estos son los **valores** de las variables. Pero el problema comienza, siguiendo una vieja tradición del análisis de datos (ver “De donde vienen las etiquetas”) cuando como respuesta a variables del tipo “Indique su género” nos encontramos con símbolos (más precisamente numerales) como “1”, “2”, etc. en la base de datos. Estos símbolos se suelen llamar **códigos** y, sin información externa, no hay manera de saber que significan. Para eso vuelven en nuestra ayuda las etiquetas

Así como hay etiquetas para los nombres de las variables también lo hay para sus etiquetas.

La etiqueta es especialmente útil en los contextos de presentación de los análisis sea tanto en formato de tablas y gráficos. La razón es que el lector de los informes puede no estar al tanto ni del cuestionario original ni de la propia base de datos. Simplemente es un lector de un informe que desea, razonablemente, que en vez de algo como “n\_hijos” aparezca algo como “Cantidad de hijos” o que en vez de “3” en los análisis de la variable “Género” aparezca algo como “género no binario”.

#### **2.1.4 Tipo de dato de la variable**

Otro punto a destacar es lo que a veces se suele denominar “nivel de medición” de la variable. Este término se suele utilizar más en programas o lenguajes específicos para análisis de datos pero tienen una similitud con el proceso usual en una planilla de cálculo (excel, google sheet, etc) de aplicarle “formato” a la celda para indicarle si se trata de texto o un número y si se trata de número de que tipo (porcentaje, fecha, decimal, etc.).

Este tipo de metainformación suele ser útil para que el programa detecte que tipo de gráfico es apropiado o que tipo de cálculo realizar. De todos modos los programas vienen cada vez con mejores heurísticas para adivinar, sin previa indicación de que tipo de dato se trata.