

SUBSECRETARÍA DE POSGRADO · FHyCS · UNJu · JUNIO 2026

UNIDAD 4 · MATERIAL COMPLEMENTARIO DE LECTURA*Análisis Exploratorio, Visualización y uso de IA
para el trabajo con Bases de Datos*

Dr. Javier Altamirano · Mg. Ana María Chalabe

Facultad de Humanidades y Ciencias Sociales – UNJu

Introducción

Este material de lectura acompaña el cuarto encuentro del curso y tiene como propósito desarrollar en profundidad los conceptos trabajados durante la clase sincónica. Su lectura previa o posterior a la sesión en vivo le permitirá afianzar los fundamentos del análisis exploratorio de datos, la visualización estadística y el uso de inteligencia artificial como herramienta metodológica.

Mientras las Unidades anteriores se concentraron en el diseño, construcción y depuración de bases de datos, esta unidad da el paso siguiente: una vez que la base está lista, ¿cómo se analiza? ¿cómo se presentan sus resultados de manera clara y rigurosa? ¿qué rol puede jugar la IA en ese proceso?

El material está organizado en cuatro secciones: (1) análisis exploratorio y tipos de variable; (2) tablas de frecuencia y visualización gráfica; (3) introducción a los prompts y herramientas de IA; y (4) análisis de variables cualitativas con asistencia de IA.

UNIDAD 04 — Análisis Exploratorio, Visualización y uso de IA para el trabajo con Bases de Datos

1. Análisis exploratorio y tipos de variable

El análisis exploratorio de datos es la primera etapa de cualquier proceso de investigación cuantitativa. Su objetivo es conocer la estructura y distribución de los datos antes de aplicar técnicas estadísticas más complejas. No se trata de probar hipótesis: se trata de comprender qué tiene la base, cómo están distribuidos los valores de cada variable, y cuáles son los patrones, tendencias o anomalías presentes.

El punto de partida del análisis exploratorio es siempre identificar correctamente el tipo de variable con el que se trabaja, porque el tipo de variable determina qué indicadores son adecuados, qué tabla de frecuencia corresponde y qué gráfico es apropiado.

1.1 Clasificación de variables

Las variables se clasifican en dos grandes grupos, cada uno con sus subtipos:

**Variable
cualitativa
nominal**

Categorías sin orden inherente. El análisis se basa en frecuencias y porcentajes. Ejemplos: sexo (macho/hembra), especie, tipo de suelo, tratamiento aplicado.

Variable cualitativa ordinal	Categorías con un orden lógico definido que debe respetarse en el análisis. Ejemplos: estadio de desarrollo (huevo < larva < pupa < adulto), nivel de severidad (leve/moderado/grave), escala de Likert.
Variable cuantitativa discreta	Valores numéricos enteros, generalmente conteos. Ejemplos: número de huevos, cantidad de individuos capturados, número de especies por parcela.
Variable cuantitativa continua	Valores numéricos que pueden tomar cualquier valor dentro de un rango, incluyendo decimales. Ejemplos: peso, talla, temperatura, porcentaje de cobertura.

Regla práctica: si podés ordenar las categorías con sentido (leve es menor que moderado), la variable es ordinal. Si el orden es arbitrario (rojo no es "mayor" que azul), es nominal. Esta distinción cambia la tabla y el gráfico que corresponde usar.

1.2 Por qué importa identificar el tipo de variable

Usar el tipo incorrecto de tabla o gráfico produce resultados estadísticamente incorrectos o visualmente engañosos. Un histograma aplicado a una variable nominal no tiene sentido: los intervalos de clase solo aplican a variables continuas. Un gráfico de barras que no respeta el orden natural de las categorías de una variable ordinal oculta información relevante.

La identificación correcta del tipo de variable es, en sí misma, una decisión metodológica que debe quedar documentada en el informe de investigación.

2. Tablas de frecuencia y visualización gráfica

Una vez identificado el tipo de variable, el análisis exploratorio sigue un orden lógico: primero se calcula la tabla de frecuencia correspondiente, y luego se selecciona el gráfico más adecuado para representarla visualmente.

2.1 Tablas de frecuencia según tipo de variable

Cualitativa nominal y ordinal	Frecuencia absoluta (n): cantidad de casos en cada categoría. Frecuencia relativa (f): proporción sobre el total (n/N). Frecuencia porcentual (%): $\text{relativa} \times 100$. Para ordinales: se agrega frecuencia acumulada (suma progresiva).
Cuantitativa discreta	Mismos indicadores que para nominales/ordinales. Se suman medidas de tendencia central (media, mediana, moda), dispersión (rango, desvío estándar) y posición (cuartiles, percentiles).
Cuantitativa continua	Los datos se agrupan en intervalos de clase de igual longitud antes de calcular frecuencias. Incluye todos los estadísticos de la discreta, más el histograma y el box-plot como representaciones gráficas específicas.

2.2 Elección del gráfico correcto

La elección del gráfico no es estética: es metodológica. Cada tipo de variable tiene gráficos adecuados e inadecuados.

Nominal — una variable	Diagrama circular (torta) o barras. El circular es adecuado cuando hay pocas categorías (idealmente 2–5) y se quiere mostrar proporciones del total.
Nominal — dos variables cruzadas	Barras 100% apiladas. Permite comparar la distribución de una variable dentro de cada categoría de la otra.
Ordinal	Barras con las categorías ordenadas según su jerarquía natural, nunca en orden alfabético.
Discreta	Barras o columnas. Para comparación por grupos, columnas apiladas (abundancias por árbol y estación, por ejemplo).
Continua — distribución	Histograma (datos agrupados en intervalos) y polígono de frecuencia (línea que une los puntos medios de cada intervalo).
Continua — comparación entre grupos	Diagrama de caja y bigotes (box-plot). Muestra mediana, cuartiles y valores atípicos simultáneamente.
Dos variables continuas	Diagrama de dispersión. Permite visualizar la relación entre ambas variables y calcular el coeficiente de correlación.

Pregunta clave antes de elegir el gráfico: ¿qué quiero mostrar? ¿la distribución de una variable? ¿la relación entre dos? ¿la comparación entre grupos? La respuesta a esa pregunta, combinada con el tipo de variable, determina el gráfico correcto. Un gráfico de torta con 10 categorías es ilegible; un histograma con variables nominales es incorrecto.

2.3 El objetivo como guía del análisis

Un principio fundamental del análisis exploratorio es que toda tabla y todo gráfico deben responder a un objetivo específico. El objetivo orienta qué variable analizar, desde qué perspectiva y con qué comparación.

Por ejemplo, el objetivo "determinar el porcentaje de machos eclosionados de la especie X" orienta a calcular la frecuencia porcentual de la variable Sexo y representarla con un diagrama circular. El objetivo "evaluar la influencia de la temperatura en el porcentaje de machos eclosionados" orienta a cruzar Sexo y Temperatura en una tabla de contingencia y representarla con barras 100% apiladas.

3. Introducción a los prompts y herramientas de IA

La inteligencia artificial generativa —como Claude, ChatGPT o Copilot— puede asistir en el análisis exploratorio de datos: leer una base de datos, calcular frecuencias, sugerir gráficos, interpretar resultados y generar tablas en formato descargable. Sin embargo, su uso efectivo requiere saber cómo comunicarle lo que se necesita. Esa habilidad se llama prompting.

3.1 ¿Qué es un prompt?

Un prompt es la instrucción que se le escribe a una herramienta de IA para que realice una tarea. Su calidad determina directamente la calidad del resultado: una instrucción vaga produce respuestas genéricas e imprecisas; una instrucción clara, con contexto y formato esperado bien definidos, produce resultados precisos y reutilizables.

La IA no adivina intenciones: responde exactamente a lo que se le pide. Aprender a construir buenos prompts es una habilidad metodológica, no solo técnica: es parte del diseño de la estrategia de análisis.

3.2 Los cuatro componentes de un prompt efectivo

Contexto	Informar a la IA quién es el usuario y en qué marco trabaja. Ejemplo: "Soy investigadora y trabajo con datos de pacientes pediátricos con asma..."
Tarea	Indicar con verbos concretos qué debe hacer. Evitar verbos vagos como "analizá" sin especificar qué. Ejemplos: calculá, generá, clasificá, resumí, graficá.
Formato esperado	Especificar cómo debe presentarse el resultado: tabla, archivo Excel, texto, lista numerada, gráfico. Sin esta indicación, la IA elige el formato que le parece mejor.
Restricciones	Indicar qué debe excluir, qué variables no son relevantes, si debe ignorar valores vacíos, o si se prefiere que solo describa sin interpretar.

3.3 Prompt débil vs. prompt bien armado

El mismo objetivo —calcular la frecuencia de una variable— produce resultados muy distintos según cómo se formula el prompt. Comparar ambas versiones permite internalizar la diferencia:

Prompt débil	"Haceme un análisis de la base de datos." → Sin variable específica, sin formato, sin contexto. La IA inventa qué hacer y el resultado es genérico.
Prompt bien armado	"Soy investigadora. Tengo datos de 25 pacientes pediátricos (Asma.xlsx). La variable Sexo es nominal. Generame: 1) tabla de frecuencia absoluta, relativa y porcentual; 2) el gráfico más adecuado con su justificación. Si hay valores vacíos, indicámelos antes de continuar." → Contexto + tarea numerada + formato + restricción.

Regla práctica: si el prompt tiene menos de dos líneas, probablemente le esté faltando algo. El tiempo invertido en construir un buen prompt se recupera en la precisión y usabilidad del resultado. Tip adicional: pedirle a la IA que justifique la elección del gráfico (no solo que lo genere) convierte el uso de la IA en una experiencia de aprendizaje metodológico, no solo de producción de resultados.

3.4 Documentar el uso de IA como parte de la metodología

En investigación, todo proceso metodológico debe quedar registrado. El uso de IA no es la excepción. El informe final de la evaluación del curso incluye una sección específica de "Informe de uso de IA", donde se deben documentar:

- Los prompts utilizados en cada etapa del análisis.
- La herramienta de IA empleada (Claude, ChatGPT, Copilot, etc.).
- Cómo se verificaron y ajustaron los resultados obtenidos.
- Las limitaciones o errores detectados en las respuestas de la IA.

Esta documentación convierte el uso de IA en un acto transparente y reproducible, coherente con los estándares de integridad científica.

4. Análisis de variables cualitativas con IA

Hasta aquí el material se enfocó en variables cuantitativas y categóricas con datos estructurados. Pero muchas investigaciones en ciencias sociales y humanas trabajan con datos cualitativos: respuestas abiertas a preguntas de encuestas, fragmentos de entrevistas, observaciones de campo, documentos. La IA es especialmente útil en este tipo de análisis, donde el volumen de datos hace inviable la lectura y codificación manual.

4.1 El desafío de las respuestas abiertas

Una base de datos con 200 respuestas a una pregunta abierta (¿qué cambiarías en las materias de investigación?) no puede analizarse simplemente contando palabras. Requiere identificar patrones, agrupar respuestas por categorías temáticas y construir una tipología que dé cuenta de la diversidad de perspectivas presentes.

Antes de pedirle a la IA que analice los datos, es necesario definir: ¿cuál es el objetivo del análisis? ¿qué tipo de categorías se buscan? ¿desde qué marco teórico o metodológico se interpretan las respuestas?

4.2 El pipeline metodológico

El proceso completo de análisis cualitativo asistido por IA sigue una secuencia de cinco pasos:

Paso 1 Objetivo	Definir con precisión qué se quiere saber de los datos. El objetivo guía todo el análisis y debe formularse antes de abrir la base.
Paso 2 Prompt de metodología	Pedirle a la IA que proponga una metodología de análisis adecuada al objetivo, con búsqueda de antecedentes en la web. Este prompt genera la tipología de perfiles o categorías temáticas que se usará.
Paso 3 Revisión metodológica	Evaluar críticamente la metodología propuesta por la IA. ¿Es coherente con el objetivo? ¿Las categorías tienen sentido para el contexto de la investigación? Ajustar si es necesario.
Paso 4 Prompt de análisis	Aplicar la metodología revisada a los datos reales, pidiéndole a la IA que clasifique cada respuesta según los perfiles o categorías definidas, y que genere las tablas de resultados.

Paso 5 Tablas y gráficos

Solicitar la producción de tablas de frecuencia, distribución de perfiles, propuestas más frecuentes, nube de palabras y gráficos correspondientes. Verificar que los resultados sean coherentes con los datos originales.

La IA no reemplaza el juicio metodológico del investigador: lo asiste. La decisión de qué categorías usar, cómo interpretarlas y qué significa cada patrón encontrado sigue siendo responsabilidad del investigador.

Si los resultados de la IA parecen incorrectos o imprecisos, no hay que descartar el prompt: hay que corregirlo y volver a ejecutarlo. El refinamiento iterativo del prompt es parte normal del proceso.

4.3 Verificación y triangulación

Los resultados que genera la IA deben siempre verificarse. En análisis cualitativo, esto implica releer una muestra de respuestas y confirmar que la clasificación propuesta es coherente con el contenido real de los datos. Si hay discrepancias, deben documentarse y ajustarse.

Una buena práctica es triangular los resultados de la IA con al menos dos lecturas independientes de una selección de respuestas. Esta verificación mejora la confiabilidad del análisis y da solidez metodológica al informe.

5. Síntesis y conexión con el proyecto propio

Los contenidos de esta unidad forman parte directa de las competencias necesarias para completar el proyecto final del curso. La evaluación final requerirá presentar un informe que incluya el análisis exploratorio de la base de datos propia, con tablas de frecuencia, visualizaciones adecuadas al tipo de variable, e informe de uso de IA documentando todo el proceso.

Tema de la unidad	Aplicación en el proyecto propio
Tipo de variable	Identificar correctamente cada variable del propio conjunto de datos.
Tabla de frecuencia	Construir la tabla adecuada para al menos tres variables de distintos tipos.
Elección del gráfico	Justificar la elección de cada gráfico en función del tipo de variable y el objetivo.
Prompting	Usar prompts estructurados para replicar el análisis con IA y documentarlos.
Análisis cualitativo	Si la base incluye preguntas abiertas, aplicar el pipeline metodológico completo.

El Trabajo Práctico 4 (TP4) brindará la primera oportunidad de aplicar estos contenidos sobre la base de datos Asma.xlsx: análisis tradicional en Excel primero, luego replicación con IA. Ese ejercicio preparatorio antecede directamente la entrega del proyecto final, donde el mismo proceso se aplica sobre los datos propios de cada investigación.