

## **Capítulo: Análisis con SPSS**

Guillermo Alejandro Zaragoza Alvarado<sup>1</sup>, ORCID <https://orcid.org/0009-0006-5466-7486>

<sup>1</sup> Universidad Virtual del Estado de Guanajuato, Posgrados. Guanajuato, México.

### **RESUMEN**

Este capítulo presenta una introducción aplicada al uso de SPSS en el análisis de datos en educación superior y ciencias sociales. En primer lugar, se sitúa la estadística como herramienta clave para la toma de decisiones informadas y se justifica el uso de SPSS como paquete estadístico ampliamente difundido en contextos académicos. A continuación, se describe el entorno de trabajo del programa, destacando la importancia de una correcta definición de variables, la construcción de la base de datos y las operaciones de gestión y transformación de datos (recodificación, creación de índices, tratamiento de valores perdidos y selección de subgrupos). Posteriormente, se abordan las posibilidades de SPSS para la estadística descriptiva (tablas de frecuencias, medidas de tendencia central y dispersión, representaciones gráficas) y para la estadística inferencial (comparación de medias, análisis de asociaciones categóricas, correlación, regresión lineal y regresión logística binaria). Estas técnicas se ilustran mediante un caso aplicado de deserción estudiantil, que muestra cómo utilizar SPSS para identificar factores de riesgo y orientar decisiones institucionales. Finalmente, se proponen acciones metodológicas para integrar SPSS en la enseñanza de la estadística, articulando proyectos, estudios de caso y actividades graduadas que favorezcan el razonamiento estadístico, la interpretación crítica de resultados y la elaboración de informes académicos.

**Palabras clave:** análisis de datos, deserción estudiantil, educación superior, estadística descriptiva, estadística inferencial, SPSS.

### **1. INTRODUCTION**

El presente capítulo tiene como propósito introducir al lector en el uso de SPSS como herramienta para el análisis de datos descriptivo e inferencial en el ámbito de la educación superior y, en general, de las ciencias sociales y de la gestión. Se parte de la idea, ampliamente aceptada en la literatura, de que la estadística es hoy un componente esencial de la formación universitaria y de la toma de decisiones informadas en contextos académicos, económicos y tecnológicos (Amat Abreu, 2017; Levin, 2014; Lind et al., 2016; Mendenhall, 2014). El énfasis no se sitúa únicamente en “hacer cuentas”, sino en desarrollar competencias para interpretar información, evaluar evidencias empíricas y fundamentar decisiones pedagógicas, organizacionales o sociales a partir de datos.

Actualmente, existen distintas investigaciones que han mostrado a muchos estudiantes y futuros profesionales presentando dificultades persistentes en contenidos de estadística, especialmente en lo relativo a la inferencia, la interpretación de hipótesis y el razonamiento con incertidumbre (López-Martín et al., 2019; Rodríguez et al., 2010; Vallecillos, 1999). Estos obstáculos se traducen en una comprensión frágil de conceptos clave y en un uso meramente instrumental de los paquetes estadísticos. Al mismo tiempo, se ha insistido en que una enseñanza de la estadística que conecte con problemas reales y con el uso de software puede favorecer un aprendizaje más significativo y duradero (Sowey, 2020; Shaughnessy, 2006; Verdooy et al., 2015).

En este marco, el objetivo didáctico central del capítulo es doble. Por un lado, familiarizar al lector con el entorno de trabajo de SPSS y con los procedimientos más habituales de análisis de datos que se emplean en la investigación aplicada en educación, administración, economía, informática o ciencias sociales (Barreiro Felpeto et al., 2014; Hernández-Sampieri et al., 2014). Por otro, contribuir a que el lector comprenda qué está haciendo desde el punto de vista estadístico cuando selecciona un procedimiento en SPSS, interpreta una tabla de resultados o redacta conclusiones, evitando el uso acrítico del software como “caja negra”.

De manera transversal, se busca que el uso de SPSS se entienda como parte de un proceso de investigación más amplio: planteamiento de problemas relevantes, diseño de instrumentos, recogida y depuración de datos, análisis, interpretación y toma de decisiones. Este enfoque

es coherente con los trabajos que relacionan el análisis de datos con problemáticas reales de la educación superior, como la deserción estudiantil, la identificación de alumnado en riesgo o la evaluación de estrategias formativas (Korhonen & Rautopuro, 2018; Kori et al., 2018; Lacave et al., 2018; Lázaro, 2020; Meyer & Fang, 2019; Mesa, 2016; Miliszewska et al., 2006; Canedo et al., 2018).

El capítulo se organiza de forma progresiva, de lo más general a lo más específico, con un fuerte énfasis en ejemplos prácticos y en aplicaciones reales. En primer lugar, se presenta brevemente la importancia de la estadística como lenguaje para describir y comprender fenómenos complejos en la sociedad contemporánea y en la educación superior, destacando su papel en la toma de decisiones basada en evidencias (Amat Abreu, 2017; Levin, 2014; Lind et al., 2016; Mendenhall, 2014). Se mostrará cómo, en investigaciones recientes sobre educación, adicción a Internet, procrastinación académica, emprendimiento social o formación continua, el análisis estadístico con apoyo de software es un elemento clave para obtener conclusiones sólidas (Aznar-Díaz et al., 2020; García-González & Ramírez-Montoya, 2020; Gómez et al., 2020; Leyva-Vázquez et al., 2020).

En segundo lugar, se contextualiza el papel de los paquetes estadísticos más habituales (SPSS, R, Statistica, entre otros) y se justifica la elección de SPSS como herramienta central en este capítulo. Se hará énfasis en su amplia difusión en entornos universitarios y profesionales, en la accesibilidad de su interfaz basada en menús y en su idoneidad para cursos en los que el objetivo principal es aplicar correctamente técnicas estadísticas clásicas más que programar (Barreiro Felpeto et al., 2014). Al mismo tiempo, se señalarán de forma crítica sus limitaciones –por ejemplo, la dependencia de licencias comerciales o la menor flexibilidad frente a entornos de programación–, situándolo como una herramienta eficaz pero no exclusiva dentro del ecosistema de recursos para el análisis de datos.

En consecuencia, se describe con claridad qué aprenderá el lector a lo largo del capítulo. Partiendo de un problema de investigación concreto (por ejemplo, el análisis de factores asociados a la deserción estudiantil o al rendimiento académico), el lector será guiado a través de las siguientes etapas:

- Creación e importación de bases de datos en SPSS: definición adecuada de variables, tipos de medida, etiquetas y codificaciones, así como la importación de datos desde hojas de cálculo u otros formatos habituales.
- Gestión y transformación de datos: recodificación de variables, creación de índices y puntuaciones compuestas, selección y filtrado de casos, depuración de errores y manejo de valores perdidos.
- Análisis descriptivo: construcción de tablas de frecuencia, medidas de tendencia central y dispersión, gráficos básicos y resúmenes estadísticos que permitan caracterizar la muestra de estudio de manera rigurosa.
- Análisis inferencial básico: uso de pruebas de contraste de hipótesis para comparar grupos (pruebas t, ANOVA), analizar asociaciones entre variables categóricas (chi-cuadrado) y estudiar relaciones lineales entre variables cuantitativas (correlación y regresión simple), siguiendo los principios expuestos en manuales clásicos de estadística aplicada (Levin, 2014; Lind et al., 2016; Mendenhall, 2014; Verdoy et al., 2015).
- Introducción a análisis más avanzados en SPSS: se ofrecerá una breve puerta de entrada a técnicas como la regresión logística binaria, muy utilizada para identificar factores predictivos de fenómenos como la deserción o el abandono de estudios (Berlanga-Silvente & Vilà-Baños, 2014; Lacave et al., 2018; Lázaro, 2020; Korhonen & Rautopuro, 2018).

De esta manera, el capítulo propondrá orientaciones para presentar e interpretar los resultados de SPSS en informes académicos y de investigación, articulando la dimensión técnica del análisis con su relevancia en la toma de decisiones en contextos educativos y sociales. En esta línea, se invita al lector a entender la estadística no como un conjunto de procedimientos mecánicos, sino como una herramienta al servicio de la comprensión de problemas complejos y del diseño de acciones de mejora, en consonancia con enfoques de

Ciencia, Tecnología y Sociedad que subrayan la responsabilidad social del conocimiento científico (Castro Díaz-Balar, 2004).

## 2. DESARROLLO

### 2.1. ¿Qué es SPSS y para qué sirve?

#### 2.1.1. Definición y breve historia de SPSS

El programa SPSS (acrónimo de Statistical Package for the Social Sciences) es uno de los paquetes estadísticos más difundidos en la investigación aplicada en ciencias sociales, educación, economía, salud y campos afines. Nació originalmente como una herramienta pensada para investigadores en ciencias sociales que necesitaban analizar grandes conjuntos de datos sin recurrir a lenguajes de programación complejos, y con el tiempo ha ido incorporando procedimientos estadísticos cada vez más sofisticados y opciones de gestión de datos avanzadas.

SPSS se comercializa bajo la denominación IBM SPSS Statistics, lo que refleja su integración en una familia más amplia de productos de análisis de datos y business analytics. Sin embargo, su esencia didáctica y aplicada se mantiene: ofrecer un entorno de trabajo relativamente intuitivo que permita al usuario centrarse en el planteamiento de problemas, selección de técnicas estadísticas e interpretación de resultados, más que en los detalles de programación. Este enfoque resulta especialmente adecuado para estudiantes y profesionales de carreras como administración, economía, educación, psicología, trabajo social, ingeniería o ciencias de la salud, en las que la estadística es una herramienta necesaria, pero no el objetivo formativo principal.

La amplia implantación de SPSS en universidades y centros de investigación ha hecho que numerosos manuales de estadística aplicada utilicen capturas de pantalla, ejemplos y ejercicios resueltos con este paquete, lo que facilita articular el aprendizaje de la teoría estadística con la práctica de análisis de datos reales (Verdoy, Beltrán & Peris, 2015; Barreiro Felpeto et al., 2014). En contextos latinoamericanos, su uso también se ha consolidado como apoyo en cursos de estadística en educación superior, particularmente en carreras técnicas y de gestión (Amat Abreu, 2017). En la Figura 1 se observa el archivo de ejemplo *demo.sav* con filas como personas encuestadas y columnas como variables demográficas.

**Figura 1.**

Editor de datos con el archivo *demo.sav*.

	age	marital	address	income	inccat	car
1	55	Marital status	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

#### 2.1.2. Ventajas y limitaciones de SPSS

Desde la perspectiva docente y de investigación aplicada, SPSS presenta una serie de ventajas que explican su popularidad:

a) *Entorno amigable basado en menús y cuadros de diálogo.*

A diferencia de programas que exigen escribir código desde el primer momento, SPSS permite ejecutar la mayoría de los análisis mediante menús desplegables y ventanas en las que el usuario elige variables y opciones con el ratón. Esto reduce la carga cognitiva inicial asociada al aprendizaje de la herramienta y facilita que el foco se sitúe en el sentido estadístico de cada procedimiento.

b) *Amplio repertorio de procedimientos estadísticos.*

El software incluye técnicas descriptivas (tablas, gráficos, medidas resumen) e inferenciales (pruebas t, ANOVA, correlaciones, regresión lineal, regresión logística, análisis no paramétricos, entre otros), que cubren gran parte de las necesidades habituales en trabajos de fin de grado, tesis de posgrado y proyectos de investigación aplicada.

c) *Integración de la gestión de datos con el análisis.*

En un mismo entorno se pueden depurar bases de datos, recodificar variables, crear índices, fusionar archivos y, a continuación, aplicar procedimientos estadísticos, lo que resulta muy conveniente para investigaciones que manejan muestras de tamaño considerable o múltiples fuentes de información.

d) *Presentación de resultados en formato tabular y gráfico.*

SPSS genera tablas y gráficos que pueden exportarse directamente a documentos de texto o de presentación, lo que ahorra tiempo en la elaboración de informes. Este aspecto es especialmente valorado en contextos aplicados donde se requiere comunicar resultados de forma clara a responsables académicos o decisores institucionales.

Sin embargo, es importante subrayar también algunas limitaciones de SPSS, tanto desde el punto de vista técnico como pedagógico:

- Dependencia de licencia comercial. A diferencia de entornos abiertos como R, SPSS requiere licencia y su coste puede limitar su disponibilidad para ciertos grupos de estudiantes o instituciones. Esto plantea, en términos de equidad y acceso, debates relevantes en el marco de la relación entre ciencia, tecnología y sociedad.
- Menor flexibilidad para análisis muy especializados o automatizados. Aunque el programa dispone de múltiples procedimientos, algunos análisis avanzados o altamente personalizados pueden resultar más sencillos en lenguajes de programación estadística. Por ello, en contextos de investigación muy especializados, SPSS suele complementarse con otros recursos.
- Riesgo de uso como “caja negra”. La principal fortaleza didáctica de SPSS –su facilidad de uso por menús– puede volverse una desventaja si se utiliza sin una base conceptual sólida. No es difícil ejecutar una prueba t o una regresión con unos pocos clics, pero interpretar adecuadamente los supuestos, los resultados y las implicaciones de un modelo exige comprender la teoría estadística subyacente.

En la Tabla 1 se muestran, de forma comparativa, algunas de las principales ventajas y limitaciones de SPSS desde la perspectiva de la docencia y la investigación aplicada en educación superior.

**Tabla 1.**

*Ventajas y limitaciones de SPSS en el contexto de la estadística aplicada en educación superior.*

Dimensión	Ventajas de SPSS	Posibles limitaciones	Implicaciones didácticas
Interfaz y modo de uso	Interfaz gráfica intuitiva basada en menús y cuadros de diálogo; permite ejecutar análisis sin escribir código (Barreiro Felpeto et al., 2014).	Puede fomentar un uso mecánico del software (“dar clics”) sin comprender los fundamentos teóricos de los procedimientos.	Facilita el acceso inicial al análisis de datos, pero exige un acompañamiento conceptual para evitar el uso crítico del programa.
Cobertura de procedimientos estadísticos	Incluye un amplio repertorio de técnicas descriptivas e inferenciales (pruebas	Algunos análisis muy especializados o de frontera no están implementados o	Permite cubrir la mayoría de las necesidades de cursos de grado y

	t, ANOVA, correlación, regresión, etc.) (Levin, 2014; Lind et al., 2016; Mendenhall, 2014).	requieren módulos adicionales.	posgrado en estadística aplicada y trabajos de investigación básica.
Gestión y transformación de datos	Integra en un mismo entorno depuración, recodificación, creación de variables e integración de bases de datos (Hernández-Sampieri et al., 2014).	La ventana de datos puede volverse compleja en bases grandes; ciertas tareas de automatización masiva son más cómodas en lenguajes de programación.	Es adecuado para enseñar el ciclo completo: desde la construcción de la base hasta el análisis e interpretación de resultados.
Resultados y comunicación	Genera tablas y gráficos listos para exportar a procesadores de texto y presentaciones; favorece la elaboración ágil de informes (Verdoy et al., 2015).	Las tablas por defecto no siempre siguen formatos de publicación académica; requieren edición para ajustarse a normas específicas.	Favorece que el estudiante se concentre en interpretar resultados y redactar conclusiones, más que en el trabajo mecánico de cálculo.
Accesibilidad y licencias	Muy extendido en universidades y centros de investigación; abundan manuales, tutoriales y ejemplos (Amat Abreu, 2017; Barreiro Felpeto et al., 2014).	Requiere licencia comercial; su coste puede limitar el acceso para ciertos estudiantes o instituciones (Castro Díaz-Balar, 2004).	Plantea reflexiones sobre equidad y acceso al software, y la necesidad de complementar con alternativas libres cuando sea posible.
Flexibilidad y ampliación	Dispone de sintaxis propia para reproducir análisis e incorporar cierto grado de automatización.	Menos flexible que entornos como R para análisis muy personalizados o integración con otros sistemas.	Es útil para introducir la idea de reproducibilidad (guardando sintaxis), y como puente hacia herramientas más programables.
Uso pedagógico y aprendizaje estadístico	Reduce la carga técnica inicial y permite centrarse en la comprensión de conceptos e interpretación de resultados (Amat Abreu, 2017).	Si se usa solo como “recetario” puede reforzar dificultades ya documentadas en inferencia, contrastes de hipótesis e interpretación (López-Martín et al., 2019; Rodríguez et al., 2010; Vallecillos, 1999).	Obliga a diseñar actividades que conecten el uso de SPSS con el razonamiento estadístico y la toma de decisiones (Sowey, 2020; Shaughnessy, 2006).

Esta reflexión crítica sobre el software es coherente con las recomendaciones de la literatura en educación estadística, que subraya la necesidad de formar usuarios capaces de integrar herramientas tecnológicas, comprensión conceptual y razonamiento inferencial de manera equilibrada (Sowey, 2020; Shaughnessy, 2006).

#### 2.1.3. SPSS como herramienta para la toma de decisiones

Más allá de su descripción técnica, es fundamental comprender a SPSS como un instrumento al servicio de la toma de decisiones basada en datos. En investigación educativa, por ejemplo, el uso de paquetes estadísticos ha permitido analizar fenómenos complejos como la

deserción estudiantil, el progreso académico, la equidad de género en áreas STEM o la efectividad de distintas estrategias de enseñanza (Korhonen & Rautopuro, 2018).

En este sentido, SPSS aporta dos contribuciones clave:

- a) Articulación entre estadística descriptiva e inferencia: El programa facilita describir de manera rigurosa una muestra (mediante frecuencias, medias, desviaciones, gráficos, etc.) y, a partir de ahí, aplicar técnicas inferenciales para contrastar hipótesis y estimar parámetros poblacionales con un determinado nivel de confianza (Mendenhall, 2014). La lógica que subyace es la misma que señalan los manuales de metodología de la investigación: se trabaja con datos muestrales, pero el interés último está en tomar decisiones sobre la población de referencia (Hernández-Sampieri et al., 2014).
- b) Modelización de fenómenos complejos: Mediante procedimientos como la regresión lineal o la regresión logística, SPSS permite estudiar la relación entre múltiples variables y construir modelos que ayudan a identificar factores asociados a determinados resultados. En el ámbito educativo, por ejemplo, se han utilizado estos modelos para analizar factores predictivos de la deserción o permanencia en estudios de ingeniería, así como para identificar perfiles de estudiantes en riesgo (Lázaro, 2020).

Ahora bien, es esencial enfatizar que SPSS no “piensa” por el usuario. La calidad de la decisión que se derive de un análisis estadístico depende, en gran medida, de:

- Cómo se ha formulado el problema y las hipótesis.
- La adecuación del diseño de investigación y del proceso de muestreo.
- La calidad de los datos recogidos y de su depuración.
- La elección correcta de las técnicas estadísticas.

La interpretación crítica de los resultados, teniendo en cuenta los supuestos de cada prueba y las limitaciones del estudio.

Como señalan distintos autores en educación estadística, uno de los retos es evitar que el aprendizaje se reduzca a “seguir recetas” en un software, promoviendo en cambio una comprensión profunda del razonamiento estadístico y de su papel en la toma de decisiones responsables en contextos profesionales y sociales (Sowey, 2020; Shaughnessy, 2006; Amat Abreu, 2017).

Para cerrar esta sección, y a modo de puente hacia los apartados posteriores del capítulo, puede plantearse un ejemplo práctico breve:

#### Ejemplo práctico 1. Exploración inicial en SPSS

Supongamos que un grupo de estudiantes de una facultad de educación ha aplicado una breve encuesta a 30 compañeros sobre horas de estudio semanal, uso de Internet con fines académicos, calificación media y percepción de riesgo de abandono de los estudios. Al abrir SPSS, el usuario se encuentra con una ventana similar a la mostrada en la Figura 1.

- a) En la vista de variables, define cada variable (nombre, etiqueta, tipo, nivel de medida).
- b) En la vista de datos, introduce las respuestas de los 30 estudiantes.
- c) Mediante el menú de análisis descriptivo, obtiene la media y desviación estándar de las horas de estudio y de la calificación, así como una tabla de frecuencias de la percepción de riesgo.
- d) Finalmente, reflexiona sobre qué decisiones o preguntas de investigación adicionales podrían plantearse a partir de esta primera exploración (por ejemplo, ¿hay relación entre horas de estudio y calificación?, ¿perciben mayor riesgo de abandono quienes tienen menor rendimiento?).

Este tipo de actividades, desarrolladas con SPSS, permiten que el estudiante experimente el ciclo completo: formulación de preguntas, organización de datos, análisis y reflexión para la toma de decisiones, reforzando así la conexión entre la estadística y problemas reales en educación superior y otras áreas profesionales.

## 2.2. Entorno de trabajo en SPSS

En esta sección se presenta el entorno básico de trabajo de SPSS y se describen sus componentes principales: la vista de datos, la vista de variables, la barra de menús, la ventana de resultados (Output) y el archivo de sintaxis. Conocer bien estos elementos es un

requisito previo para poder aprovechar el potencial del programa en el análisis de datos aplicado.

#### 2.2.1. Vista de datos y vista de variables

Cuando se abre un archivo de datos en SPSS (o se crea uno nuevo), el usuario se encuentra con una estructura matricial donde las filas representan casos (estudiantes, empresas, pacientes, etc.) y las columnas representan variables (edad, sexo, calificación, horas de estudio, etc.). Esta estructura, común a muchos paquetes estadísticos, es especialmente adecuada para organizar información de estudios cuantitativos en educación, negocios o ciencias sociales (Hernández-Sampieri, Fernández-Collado & Baptista-Lucio, 2014; Amat Abreu, 2017).

SPSS ofrece dos formas complementarias de visualizar y editar esa matriz:

##### Vista de datos (Data View)

- Muestra la tabla de datos propiamente dicha: cada fila corresponde a un caso y cada columna a una variable.
- En las celdas se ven los valores numéricos o categóricos tal como han sido introducidos o recodificados (por ejemplo, 1 = hombre, 2 = mujer; 0 = no fuma, 1 = fuma).
- Es el entorno en el que se revisa la consistencia de los registros, se detectan posibles errores de digitación y se realizan correcciones puntuales.

##### Vista de variables (Variable View)

- Presenta información detallada sobre cada variable: nombre, etiqueta descriptiva, tipo de datos, formato, etiquetas de valores, valores perdidos, nivel de medida, entre otros atributos (Barreiro Felpeto et al., 2014; Verdoy et al., 2015).
- Cada fila corresponde a una variable (por ejemplo, sexo, edad, horas\_estudio, nota\_media) y cada columna a una propiedad de esa variable.

En la Figura 2 se propone un esquema de estas dos vistas, destacando el paralelismo entre la fila de la vista de variables y la columna correspondiente en la vista de datos.

**Figura 2.**

*Vista de variables con nombres, etiquetas y tipos*

The screenshot shows the SPSS Data Editor window titled "Untitled - Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. Below the menu is a toolbar with icons for opening files, saving, and other functions. The main area is divided into two tabs: "Data View" (selected) and "Variable View". The "Data View" tab shows a grid of 17 rows and 7 columns. The columns are labeled: Row#, Name, Type, Width, Decimals, Label, and Value. The first three rows contain data: Row 1 has age (Numeric, 8, 2, None); Row 2 has marital (Numeric, 8, 2, None); and Row 3 has income (Numeric, 8, 2, None). Rows 4 through 17 are empty. The "Variable View" tab shows a list of 17 variables, each with its name, type, width, decimals, label, and value. The variables are: 1 age, 2 marital, 3 income, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, and 17. The "Label" and "Value" columns are mostly empty or show "None".

Row#	Name	Type	Width	Decimals	Label	Value
1	age	Numeric	8	2		None
2	marital	Numeric	8	2		None
3	income	Numeric	8	2		None
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						

Entre los campos más relevantes que el usuario debe configurar en la vista de variables se encuentran:

- Name (Nombre): identificador corto de la variable, sin espacios ni caracteres especiales. Se recomienda usar nombres significativos pero concisos, p. ej. horas\_estudio, promedio, riesgo\_desercion.
- Label (Etiqueta): descripción más amplia y legible para informes, p. ej. “Horas de estudio semanal” o “Percepción de riesgo de deserción”.
- Type (Tipo): numérica, cadena (texto) u otras opciones. Incluso las variables categóricas suelen codificarse numéricamente para facilitar el análisis.
- Values (Etiquetas de valores): asignación de significados a los códigos numéricos (1 = Masculino; 2 = Femenino). Esta característica es especialmente útil cuando se prepara el archivo para análisis posteriores y para la presentación de resultados.
- Missing (Valores perdidos): valores que deben tratarse como “dato faltante” (por ejemplo, 99 para “no responde”).
- Measure (Medida): nivel de medición de la variable: nominal, ordinal o escala. Esta clasificación está vinculada a la teoría de la medición y tiene implicaciones directas en la elección del análisis estadístico adecuado (Levin, 2014; Lind et al., 2016; Mendenhall, 2014).

Una adecuada definición de estos elementos reduce errores posteriores y facilita el trabajo tanto descriptivo como inferencial. En términos didácticos, trabajar con la vista de variables permite reforzar conceptos fundamentales como tipo de escala, codificación de categorías y tratamiento de datos perdidos, que suelen ser foco de dificultades para el alumnado (Rodríguez, Montañez & Rojas, 2010; López-Martín, Batanero & Gea, 2019).

**Ejemplo práctico 2. Configuración de variables en un estudio de deserción**  
Supongamos una base de datos con información de 120 estudiantes de una carrera de ingeniería. En la vista de variables se definen, entre otras, las siguientes variables:

- sexo (Type: numérico; Values: 1 = Masculino, 2 = Femenino; Measure: nominal).
  - promedio (Type: numérico; Measure: escala).
  - horas\_estudio (Type: numérico; Measure: escala).
  - deserta (Type: numérico; Values: 0 = Permanece, 1 = Deserta; Measure: nominal).
- Una vez configuradas, la vista de datos mostrará cada fila como un estudiante concreto con sus valores en todas estas variables.

### 2.2.2. Menús principales y tipos de tareas

En la parte superior de la ventana de SPSS se encuentra la barra de menús, que organiza las funciones del programa en categorías lógicas (Archivo, Edición, Datos, Transformar, Analizar, Gráficos, etc.). Esta organización ayuda a estructurar el trabajo del usuario en torno a tres grandes grupos de tareas, que la literatura suele diferenciar:

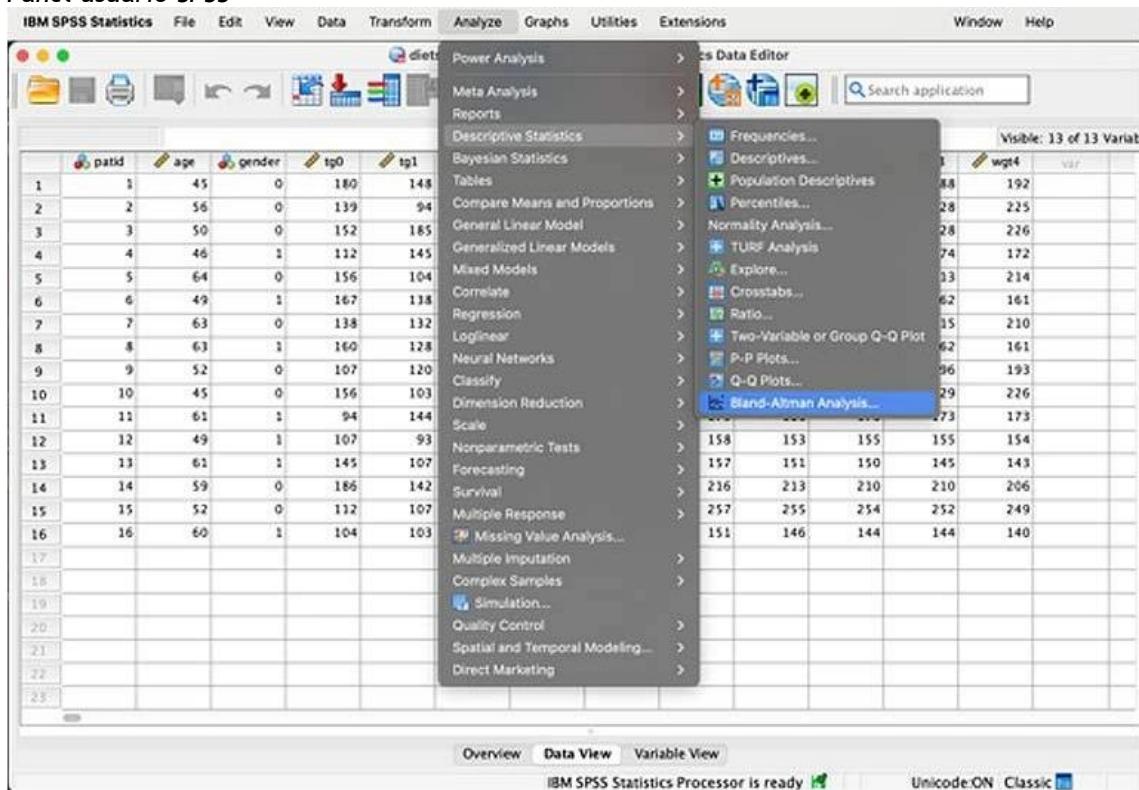
1. gestión y preparación de datos,
2. análisis estadístico propiamente dicho, y
3. presentación de resultados.

A grandes rasgos, los menús más relevantes para el trabajo cotidiano son:

- Archivo (File):
  - Crear nuevos archivos de datos, abrir bases existentes, guardar en formato SPSS (.sav) u otros, importar desde Excel, CSV, etc.
  - También permite guardar y abrir archivos de resultados (.spv) y de sintaxis (.sps).
- Edición (Edit) y Ver (View):
  - Opciones generales de copiar, pegar, buscar, así como personalizar algunos aspectos de la visualización (fuentes, cuadriculas, etc.).
- Datos (Data):
  - Ordenar casos, seleccionar subconjuntos de casos, fusionar archivos, dividir el archivo por grupos, entre otras operaciones sobre la estructura del conjunto de datos.
  - Estas funciones son esenciales cuando se trabajan bases grandes o cuando se requiere analizar subgrupos específicos (por ejemplo, solo estudiantes de primer año, solo mujeres, etc.).
- Transformar (Transform):
  - Incluir comandos como *Recode into Different Variables* o *Compute Variable*.

- Permite recodificar valores, agrupar rangos, crear índices y variables derivadas (por ejemplo, un puntaje total de satisfacción a partir de varios ítems).
- Representa la dimensión de “tratamiento de datos” que los manuales de estadística aplicada consideran clave para la construcción de indicadores e índices (Verdoy et al., 2015; Amat Abreu, 2017).
- Analizar (Analyze):
  - Es el menú central para realizar procedimientos estadísticos:
    - Estadísticos descriptivos (frecuencias, descriptivos, exploración).
    - Comparación de medias (pruebas t, ANOVA).
    - Correlaciones y regresión (lineal, logística, etc.).
    - Pruebas no paramétricas y otros procedimientos más avanzados.
  - La organización de este menú suele seguir, en buena medida, la estructura de los cursos de estadística: primero descriptiva, luego inferencial (Levin, 2014; Lind et al., 2016).
- Gráficos (Graphs) o Gráficos y Tablas (según versión):
  - Generación de gráficos de barras, pastel, histogramas, diagramas de caja, diagramas de dispersión, entre otros.
  - Aunque muchos análisis ya ofrecen gráficos desde el menú Analizar, este menú permite una personalización más fina para la presentación de resultados.
- Ventana (Window) y Ayuda (Help):
  - Gestionan las distintas ventanas abiertas (datos, resultados, sintaxis) y ofrecen acceso a la ayuda contextual y a tutoriales básicos.

**Figura 3**  
**Panel usuario SPSS**



En la Tabla 2 se presenta un resumen de los menús principales y ejemplos de tareas docentes asociadas a cada uno de ellos.

**Tabla 2.**

*Menús principales de SPSS y ejemplos de tareas asociadas en cursos de estadística aplicada.*

Menú de SPSS	Tipo de tarea principal	Ejemplos de uso en docencia / investigación	Comentario didáctico
Archivo (File)	Gestión de archivos de datos	Crear una base nueva, abrir un archivo .sav, importar datos	Permite trabajar con bases de datos reales

	y proyectos	desde Excel/CSV, guardar datos, guardar el <i>Output</i> y la sintaxis.	de investigaciones o encuestas aplicadas en clase.
Edición (Edit)	Edición básica de contenidos y opciones	Copiar/pegar celdas o tablas, buscar y reemplazar valores, deshacer acciones recientes.	Útil para corregir errores de digitación y mostrar al alumnado cómo revisar datos de forma cuidadosa.
Ver (View)	Configuración visual de la interfaz	Mostrar/ocultar cuadriculas, cambiar fuente, resaltar valores seleccionados.	Ayuda a personalizar la interfaz para mejorar la legibilidad durante las prácticas de laboratorio.
Datos (Data)	Organización y selección de casos	Ordenar casos por promedio, seleccionar solo estudiantes de primer año, dividir el archivo por género o cohorte.	Refuerza la idea de población, muestra y submuestras; clave para discutir comparaciones entre grupos.
Transformar (Transform)	Recodificación y creación de nuevas variables	Recodificar “edad” en rangos, crear un índice de satisfacción sumando ítems, generar una variable dicotómica de “aprobado/suspensión”.	Permite trabajar el paso de variables “en bruto” a indicadores más interpretables, conectando teoría de medición y análisis.
Analizar (Analyze)	Aplicación de procedimientos estadísticos	Calcular frecuencias y descriptivos, aplicar pruebas t, ANOVA, chi-cuadrado, correlaciones, regresión lineal o logística.	Es el núcleo del análisis; debe usarse siempre acompañado de discusión sobre supuestos, interpretación y toma de decisiones.
Gráficos (Graphs)	Representación gráfica de resultados	Construir gráficos de barras de rendimiento, histogramas de calificación, diagramas de caja por grupo, diagramas de dispersión horas-nota.	Favorece la interpretación visual de los datos y el desarrollo del razonamiento estadístico a partir de gráficos.
Ventana (Window)	Gestión de ventanas abiertas	Alternar entre archivo de datos, <i>Output</i> y sintaxis; organizar ventanas en cascada.	Ayuda a que el estudiantado entienda la relación entre la base de datos, los análisis realizados y la documentación (sintaxis).
Ayuda (Help)	Acceso a documentación y asistencia contextual	Consultar la ayuda de un procedimiento (por ejemplo, ANOVA o regresión), acceder a ejemplos y descripciones de opciones avanzadas.	Fomenta la autonomía del estudiantado en el uso del software y el hábito de consultar documentación técnica.

Desde una perspectiva didáctica, la barra de menús puede utilizarse como un “mapa” para organizar las prácticas en laboratorio: primero se trabaja con Datos y Transformar para limpiar y preparar la base; luego con Analizar para aplicar técnicas estadísticas; y finalmente con Gráficos y las opciones de exportación para comunicar resultados. Esta secuencia refuerza la idea de que el análisis estadístico no es un acto aislado sino parte de un proceso de investigación más amplio.

### Ejemplo práctico 3. Recorrido básico por los menús en una práctica de aula

En un taller de estadística aplicada, el estudiantado podría seguir estos pasos guiados:

- Usar Archivo para abrir un archivo con datos de estudiantes (variables de rendimiento, dedicación, etc.).
- Revisar la estructura con Datos (ordenar por promedio, seleccionar solo un semestre).
- Emplear Transformar para recodificar la variable “promedio” en tres categorías (bajo, medio, alto).
- Usar Analizar → Descriptive Statistics → Frequencies para describir la distribución del rendimiento.
- Finalmente, con Gráficos, generar un diagrama de barras que muestre el porcentaje de estudiantes en cada categoría de rendimiento.

#### 2.2.3. Ventana de resultados (Output) y archivo de sintaxis

Cada vez que el usuario ejecuta un análisis desde el menú Analizar (o genera ciertos gráficos), SPSS abre –si no estaba ya abierta– la ventana de resultados u Output. Esta ventana cumple una función central: es el lugar donde se visualizan las tablas, gráficos y mensajes que genera el programa como respuesta a las órdenes del usuario (Barreiro Felpeto et al., 2014).

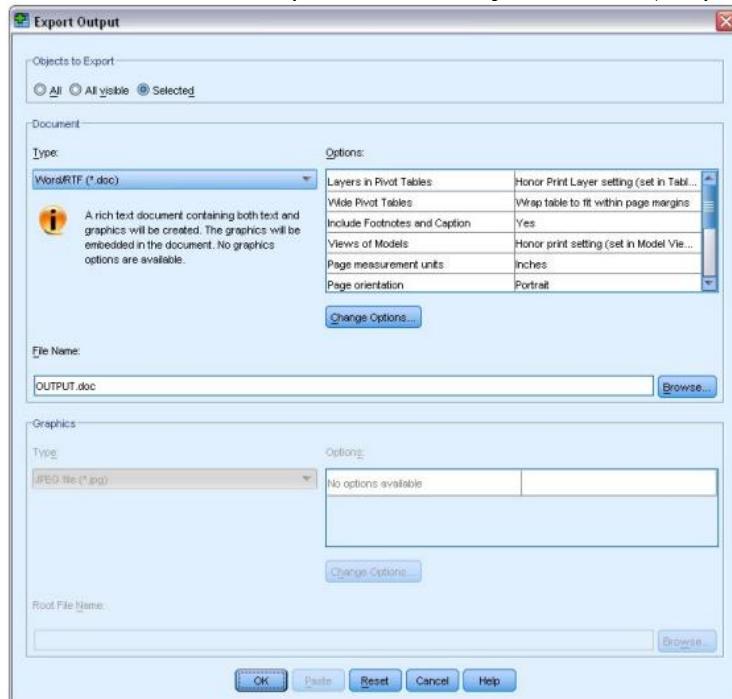
La ventana de Output se organiza en dos paneles:

- Un panel izquierdo en forma de árbol, que actúa como índice de todos los resultados generados en la sesión (análisis descriptivos, pruebas de hipótesis, tablas de regresión, etc.).
- Un panel derecho donde se muestran las tablas, gráficos y notas seleccionados.

Esta estructura facilita la navegación cuando se han realizado múltiples análisis, permitiendo plegar o desplegar secciones, renombrar resultados o eliminar aquellos que no se consideran relevantes para el informe final. La figura 4 reproduce el visor con el árbol de resultados a la izquierda y las tablas/gráficos a la derecha

**Figura 4.**

Visor de resultados con panel de títulos y contenidos (Output) en SPSS.



Desde el Output, el usuario puede:

- Copiar y pegar tablas o gráficos a un procesador de texto (por ejemplo, para un informe de investigación, un trabajo de fin de grado o una presentación).
- Exportar el conjunto de resultados a diferentes formatos (Word, PDF, HTML, etc.).

- Editar algunos aspectos de formato (títulos, alineación, número de decimales), aunque en contextos académicos suele recomendarse adaptar las tablas a las normas de publicación del campo (por ejemplo, normas APA).

Paralelamente, SPSS dispone de un componente menos visible pero muy importante: el archivo de sintaxis o ventana de sintaxis. Cada operación realizada por menús puede representarse también en forma de comandos de texto, que se pueden guardar, editar y ejecutar de nuevo. Esta sintaxis:

- Permite reproducir exactamente los análisis realizados, algo fundamental desde el punto de vista de la transparencia y la calidad de la investigación (Hernández-Sampieri et al., 2014).
- Facilita automatizar tareas repetitivas (por ejemplo, aplicar la misma serie de análisis a distintas submuestras).
- Sirve como puente conceptual hacia otros entornos más programables, donde el análisis se expresa siempre como código.

En estudios que emplean modelos más complejos, como la regresión logística binaria, el uso de sintaxis en SPSS se vuelve especialmente útil para documentar los pasos seguidos y ajustar opciones específicas del procedimiento.

**Figura 5.**

Archivo de datos, sintaxis y visor en SPSS.

Desde un punto de vista pedagógico, introducir al estudiantado en el uso básico de la sintaxis (aunque sea copiando y pegando los comandos generados automáticamente por los menús) contribuye a:

- Reforzar la comprensión de qué exactamente está haciendo el programa cuando ejecuta una prueba o un modelo.
- Promover buenas prácticas de documentación y reproducibilidad, alineadas con las recomendaciones actuales en metodología de la investigación.

#### Ejemplo práctico 4. Guardar y reutilizar sintaxis de un análisis descriptivo

Tras ejecutar un análisis de frecuencias por menús, el estudiante puede ir a la opción “Pegar” en lugar de “Aceptar”. SPSS abrirá una ventana de sintaxis con el comando correspondiente (por ejemplo, FREQUENCIES VARIABLES=...). Esa sintaxis puede guardarse en un archivo .sps y reutilizarse en futuras sesiones, asegurando que se aplican los mismos criterios de análisis a distintos conjuntos de datos.

### 2.3. De la pregunta de investigación a la base de datos en SPSS

El paso de la pregunta de investigación a la base de datos en SPSS constituye un momento clave del trabajo empírico, porque en él se decide qué información será efectivamente recogida, cómo se representará en forma de variables y con qué nivel de precisión se podrá analizar posteriormente. En los manuales de metodología de la investigación se insiste en que toda investigación cuantitativa parte de un problema formulado con claridad, de objetivos bien definidos y de hipótesis o preguntas específicas que orientan la selección de datos

(Hernández-Sampieri, Fernández-Collado y Baptista-Lucio, 2014). Los textos de estadística aplicada subrayan, además, que la calidad del análisis estadístico depende en gran medida de la pertinencia y calidad de las variables registradas (Levin, 2014; Lind et al., 2016; Mendenhall, 2014; Verdoy, Beltrán y Peris, 2015). SPSS no reemplaza esa reflexión conceptual previa: más bien obliga a hacerla explícita, porque cada aspecto del problema de investigación debe traducirse en una columna bien definida de la base de datos.

El proceso suele comenzar con la delimitación de la población de referencia y la determinación de la muestra con la que efectivamente se va a trabajar. Cuando se estudia, por ejemplo, la deserción en ingeniería, el rendimiento académico en estudiantes universitarios o la relación entre uso de Internet y procrastinación académica, la población puede estar constituida por el total de estudiantes matriculados en una titulación, una cohorte de ingreso o un conjunto de asignaturas, mientras que la muestra se concreta en los estudiantes que responden a un cuestionario o cuyos registros académicos están disponibles (Aznar-Díaz et al., 2020; Korhonen y Rautopuro, 2018; Kori, Pedaste y Must, 2018; Lacave, Molina y Cruz-Lemus, 2018; Lázaro, 2020). Esta distinción, que en ocasiones el alumnado percibe como puramente teórica, adquiere un significado muy tangible cuando se abre SPSS: cada fila de la vista de datos corresponde a un individuo de la muestra, pero las inferencias y decisiones que se pretenden tomar se refieren a la población más amplia, lo que pone de relieve la importancia del diseño muestral y de la representatividad.

Una vez establecido el problema y definidos la población y la muestra, es necesario operacionalizar los constructos teóricos en variables observables. Conceptos como “deserción”, “integración académica y social”, “competencia de emprendimiento social” o “uso académico de Internet” deben transformarse en indicadores que puedan medirse mediante ítems de cuestionarios, registros académicos u otros instrumentos (García-González y Ramírez-Montoya, 2020; Canedo, Santos y Leite, 2018; Meyer y Fang, 2019; Gómez et al., 2020). Esta traducción implica decidir qué se medirá, con qué escala de medida y con qué nivel de detalle. SPSS obliga a asignar a cada constructo una o varias variables concretas, lo que invita a preguntarse si la operacionalización es coherente con el marco teórico y si permitirá aplicar las técnicas de análisis previstas. No es lo mismo registrar la “deserción” como una variable dicotómica (permanece/deserta) que como una variable multinivel que distinga entre abandono temporal, cambio de titulación y abandono definitivo, y esa elección condiciona el tipo de análisis (chi-cuadrado, regresión logística, modelos multinomiales) que posteriormente podrán aplicarse (Berlanga-Silvente y Vilà-Baños, 2014; Korhonen y Rautopuro, 2018; Lacave et al., 2018).

El diseño de las variables se refleja directamente en la vista de variables de SPSS. A partir del planteamiento del estudio, el investigador define para cada columna de la base el nombre técnico, la etiqueta descriptiva, el tipo de dato (numérico o cadena), las etiquetas de valores para las variables categóricas, el tratamiento de valores perdidos y el nivel de medida (nominal, ordinal, escala). Esta tarea, que a veces se percibe como “técnica”, tiene una fuerte dimensión conceptual: el nivel de medida asociado a una variable no es una formalidad cualquiera, sino que sintetiza cómo se entiende esa característica y qué operaciones estadísticas son legítimas sobre ella (Levin, 2014; Lind et al., 2016). Por ejemplo, tratar una escala de Likert como ordinal o como aproximadamente de intervalo tiene implicaciones en la elección entre estadísticos no paramétricos y procedimientos paramétricos, y esta discusión puede convertirse en un espacio de reflexión valioso en la enseñanza de la estadística (López-Martín, Batanero y Gea, 2019; Rodríguez, Montañez y Rojas, 2010).

Resulta útil ilustrar este tránsito con un ejemplo concreto. Supóngase que se desea estudiar la relación entre horas de estudio semanal, uso de Internet con fines académicos y calificación media, así como explorar la percepción de riesgo de abandono de los estudios en un grupo de estudiantes universitarios, problemáticas ampliamente abordadas en la literatura reciente (Aznar-Díaz et al., 2020; Korhonen y Rautopuro, 2018; Kori et al., 2018). En la fase conceptual se definen las preguntas de investigación: si existe asociación entre ciertas pautas de estudio y el rendimiento, si quienes perciben mayor riesgo de abandono presentan determinados patrones de uso de la tecnología, o si se observan diferencias entre grupos de género u otras variables sociodemográficas. A partir de estas preguntas se diseñan los instrumentos: un cuestionario que recoge, por ejemplo, sexo, edad, horas de estudio, frecuencia de uso de Internet para tareas académicas, calificación media en la última cohorte y una escala de percepción de riesgo de deserción.

Cuando se traslada esta información a SPSS, cada uno de estos elementos se convierte en una variable específica en la vista de variables. La variable sexo puede definirse como numérica con etiquetas de valores (1 = Masculino, 2 = Femenino) y nivel de medida nominal; edad, horas\_estudio y nota\_media se definen como variables numéricas de medida en escala; la frecuencia de uso de Internet puede registrarse mediante una escala ordinal (por ejemplo, de 1 = nunca a 5 = siempre); y la percepción de riesgo de abandono mediante un ítem o un conjunto de ítems que luego se combinarán en un índice. Paralelamente, en la fase de campo se recoge la información mediante el cuestionario y se procede a la digitación o importación de los datos en SPSS, de modo que cada estudiante ocupa una fila de la vista de datos con sus valores correspondientes en todas las variables. Esta estructura facilita, posteriormente, el cálculo de descriptivos, la comparación entre grupos y la aplicación de modelos de regresión que intenten explicar el rendimiento o el riesgo de abandono a partir de las variables explicativas disponibles (Levin, 2014; Mendenhall, 2014; Verdoy et al., 2015).

En la práctica, este recorrido de la pregunta de investigación a la base de datos en SPSS no es lineal ni exento de dificultades. Es frecuente que durante la exploración preliminar de los datos se detecten problemas de codificación, valores extremos o categorías con muy pocos casos, lo que obliga a replantear algunas decisiones y a recurrir a procedimientos de recodificación o agrupación de categorías mediante el menú Transformar. Asimismo, el análisis descriptivo inicial puede revelar que ciertas variables no aportan información relevante para las hipótesis planteadas, o que sería conveniente construir índices más sintéticos a partir de varios ítems, lo que conecta directamente con las estrategias de tratamiento de datos cuantitativos descritas por Barreiro Felpeto, Guisande González y Vamonde Liste (2014) y con el trabajo de construcción de indicadores en contextos de Ciencia, Tecnología y Sociedad (Castro Díaz-Balar, 2004; Mesa, 2016). Desde una perspectiva didáctica, hacer explícito este ir y venir entre marco teórico, diseño de variables e implementación en SPSS ayuda a que el estudiantado comprenda que la estadística no comienza cuando se abre el menú Analizar, sino mucho antes, en la claridad con que se formula el problema de investigación, se diseña el instrumento y se estructura la base de datos.

#### 2.4. Gestión y transformación de datos en SPSS

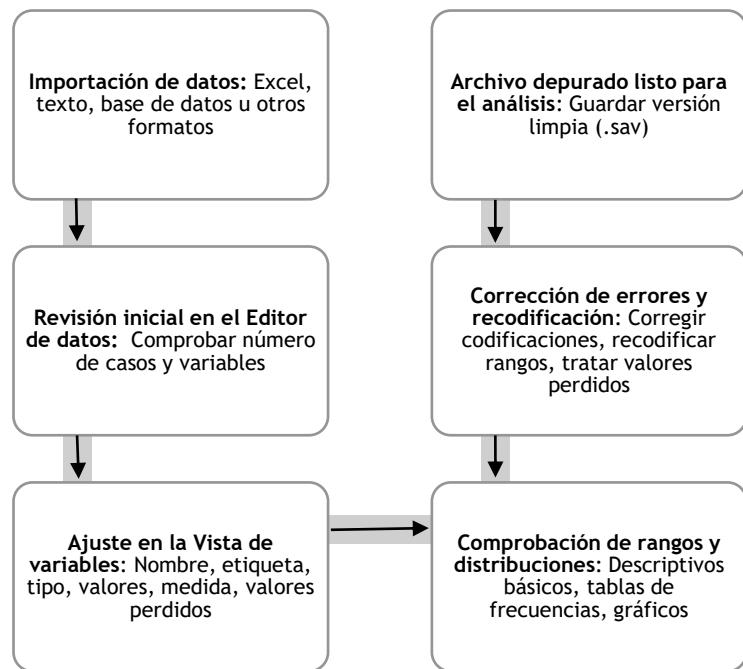
La gestión y transformación de datos en SPSS constituye un puente entre la fase de diseño de la base de datos y el análisis estadístico propiamente dicho. No basta con haber definido correctamente las variables en la vista de variables: es imprescindible depurar la información, recodificar categorías cuando sea necesario, crear nuevas variables derivadas y seleccionar los casos pertinentes para cada análisis. Los manuales de metodología de la investigación y de estadística aplicada coinciden en destacar que una buena parte del tiempo de trabajo en una investigación cuantitativa se dedica precisamente a preparar los datos para que los procedimientos inferenciales puedan aplicarse con sentido. Desde esta perspectiva, SPSS ofrece un conjunto de herramientas integradas que permiten asumir este proceso de preparación de forma sistemática y transparente.

Un primer bloque de tareas se relaciona con la depuración inicial de la base. Una vez introducidos o importados los datos, es habitual realizar un recorrido exploratorio que incluya la comprobación de rangos de valores, la identificación de celdas vacías o codificaciones inconsistentes y la detección de posibles valores atípicos. Este examen puede apoyarse en estadísticos descriptivos y gráficos sencillos, pero también en operaciones de ordenación de casos y filtros temporales. En SPSS, el menú Datos facilita ordenar los registros por una variable determinada –por ejemplo, ordenar estudiantes por promedio o por edad–, lo que permite identificar rápidamente valores imposibles (como edades fuera de rango) o codificaciones erróneas.

En la Figura 6 se propone un esquema de flujo de este proceso de depuración: a partir de la base original, se realiza un recorrido inicial, se corrigen errores manifiestos y se documentan las decisiones adoptadas, antes de pasar a recodificar o crear nuevas variables. Esta forma de trabajo respalda la idea, subrayada por Barreiro Felpeto, et al. (2014), de que el tratamiento de datos cuantitativos no consiste solo en aplicar fórmulas, sino en tomar decisiones fundamentadas sobre la calidad y estructura de la información disponible.

**Figura 6.**

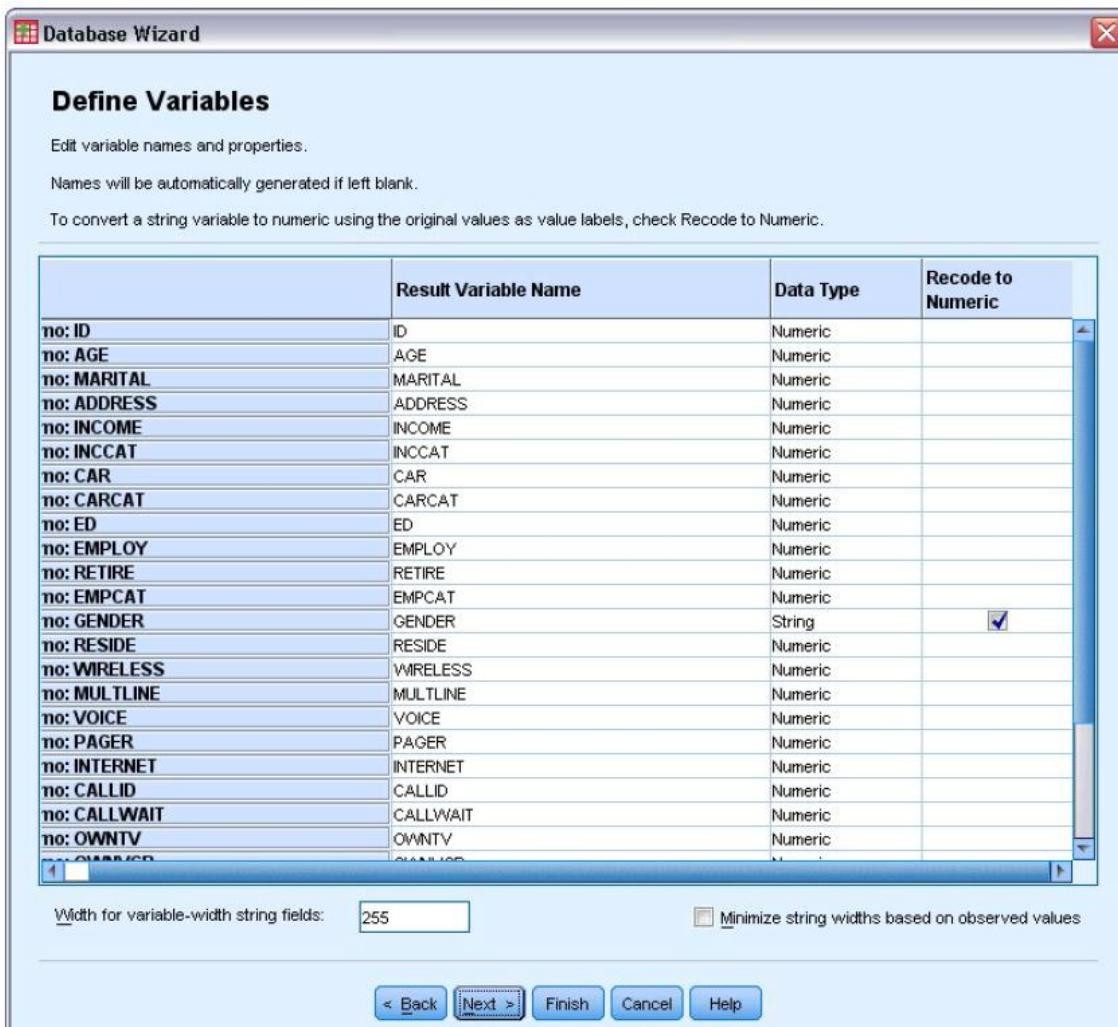
*Flujo de depuración: de la importación al editor de datos*



La recodificación de variables constituye una de las operaciones más frecuentes en SPSS y está estrechamente ligada a las necesidades del análisis. En numerosos estudios en educación y ciencias sociales, los investigadores agrupan categorías, crean rangos o construyen variables dicotómicas a partir de una variable continua o con múltiples categorías originales, por ejemplo al transformar una calificación numérica en niveles de rendimiento (bajo, medio, alto) o al redefinir una variable de deserción en términos de permanece/deserta. El comando de recodificación en SPSS permite especificar cómo se agrupan los valores originales en nuevas categorías y genera una variable distinta, preservando la original para posibles análisis alternativos.

En la Figura 7 se representa el diálogo típico de recodificación en una nueva variable, destacando el espacio donde se introducen los rangos de valores y las etiquetas asignadas a las categorías resultantes. Desde el punto de vista formativo, este procedimiento ofrece una oportunidad para discutir con el estudiantado el sentido de las decisiones de agrupamiento, su impacto sobre la información que se conserva o se pierde y las implicaciones para la interpretación de resultados.

**Figura 7**  
*Cuadro de diálogo para recodificar en una nueva variable*

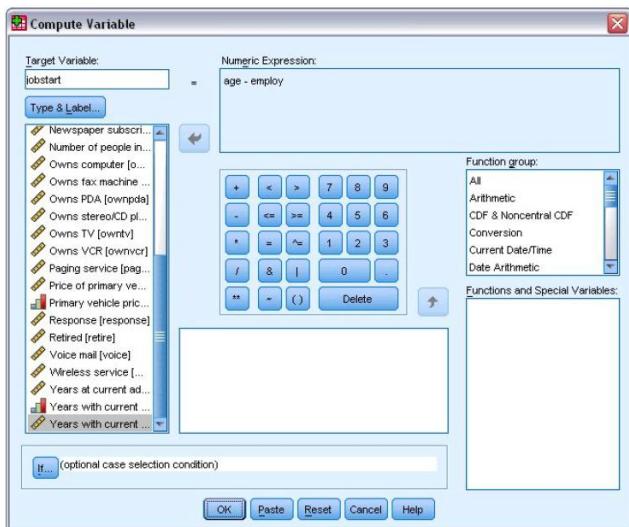


Otra operación central es la creación de variables nuevas mediante combinaciones de variables existentes. En investigaciones educativas y de gestión es frecuente construir índices o puntuaciones totales a partir de ítems de cuestionarios, promediar calificaciones de distintas asignaturas, calcular indicadores como el porcentaje de créditos aprobados o generar variables que sintetizan dimensiones de integración académica y social. SPSS ofrece el comando de cálculo de variables, que permite definir expresiones aritméticas y lógicas para obtener nuevas columnas en la base de datos.

En la Figura 8 se muestra el cuadro de diálogo *Calcular variable* de SPSS, donde se define una nueva variable como combinación de variables existentes, siguiendo la lógica ilustrada en la guía breve de IBM SPSS.

#### Figura 8.

Cuadro de diálogo *Calcular variable*



La gestión de valores perdidos constituye otro aspecto delicado de la transformación de datos. En la práctica, es frecuente que algunos estudiantes no respondan ciertas preguntas o que falten registros en determinadas variables, ya sea por errores de recogida de datos o por omisiones voluntarias. SPSS permite definir códigos específicos para indicar que un valor debe tratarse como perdido —por ejemplo, 99 o 999— y excluir esos registros de los análisis cuando corresponda. Además, en algunos procedimientos se puede elegir entre eliminar casos con datos faltantes de manera lista a lista o caso a caso. La decisión sobre cómo manejar los valores perdidos no es meramente técnica; tiene implicaciones directas en los resultados y en la generalización de las conclusiones, como señalan los manuales de estadística aplicada y metodología.

La figura 9 reproduce el cuadro de diálogo *Valores perdidos* de SPSS, utilizado para definir códigos específicos de datos perdidos en variables numéricas, tal como se ejemplifica en la guía breve.

**Figura 9.**

Cuadro de diálogo de valores perdidos para variables numéricas

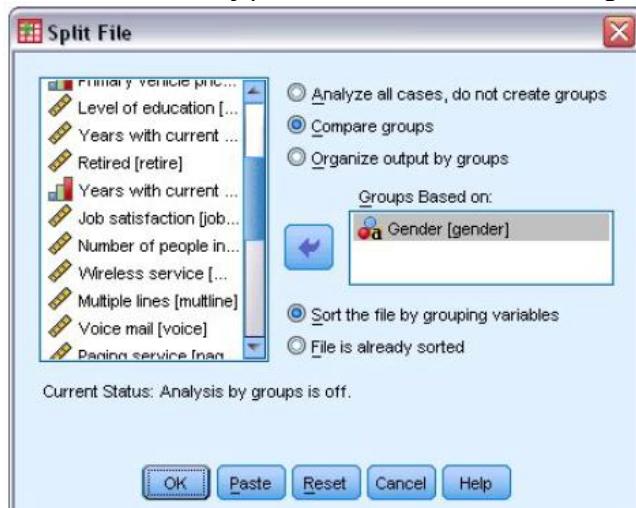


Finalmente, la selección de casos y la segmentación del archivo representan herramientas esenciales para adaptar el análisis a subgrupos específicos de la población estudiada. En investigaciones sobre deserción, por ejemplo, puede ser necesario restringir temporalmente la muestra a una cohorte concreta o analizar por separado el comportamiento de estudiantes de primer año frente a los de cursos superiores. SPSS permite seleccionar casos en función de condiciones lógicas (por ejemplo, analizar solo estudiantes con promedio inferior a cierto umbral, o solo quienes trabajan y estudian a la vez) y también dividir el archivo por grupos, de manera que los análisis descriptivos e inferenciales se ejecuten de forma simultánea para cada categoría de una variable, como género o tipo de institución.

En la Figura 10 se esquematiza el uso conjunto de las opciones de selección de casos y división del archivo para producir resultados desagregados por subgrupos. Este tipo de estrategias facilita el diálogo entre los resultados numéricos y las decisiones de política educativa o de gestión académica, al evidenciar que los fenómenos de rendimiento o deserción pueden comportarse de manera diferente en segmentos específicos de la población estudiada.

**Figura 10.**

*Selección de casos y procesamiento de archivos segmentados*



De manera que, la gestión y transformación de datos en SPSS no debe entenderse como una fase meramente instrumental, sino como un espacio donde se articulan decisiones metodológicas, consideraciones teóricas y criterios de calidad de la información. La posibilidad de recodificar variables, crear indicadores, tratar valores perdidos y seleccionar submuestras, apoyada en un entorno relativamente amigable, hace de SPSS una herramienta adecuada para que el estudiantado experimente con datos reales y tome conciencia de la responsabilidad que implica preparar una base de datos para el análisis estadístico. Desde la perspectiva de la educación estadística, aprovechar este potencial significa diseñar actividades en las que cada transformación se discuta explícitamente, conectando la acción en el software con los principios de medición, inferencia y toma de decisiones que se pretenden desarrollar.

## 2.5. Estadística descriptiva con SPSS

La estadística descriptiva con SPSS constituye el primer acercamiento sistemático a los datos y, en la práctica, suele ocupar un lugar central tanto en la docencia como en la investigación aplicada. Antes de contrastar hipótesis o ajustar modelos inferenciales, es necesario conocer cómo se distribuyen las variables, cuáles son sus tendencias centrales, qué grado de variabilidad presentan y si existen valores extremos o patrones llamativos que requieran una atención especial. Los manuales clásicos de estadística aplicada subrayan que una buena descripción inicial no solo facilita la comprensión del fenómeno estudiado, sino que también previene interpretaciones erróneas en etapas posteriores del análisis. En este sentido, SPSS ofrece un conjunto de procedimientos que permiten calcular y representar de forma relativamente sencilla las principales medidas descriptivas, integrando tablas y gráficos en un entorno coherente con el proceso de investigación (Hernández-Sampieri, Fernández-Collado y Baptista-Lucio, 2014; Barreiro Felpeto, Guisande González y Vamonde Liste, 2014).

En el caso de las variables categóricas, como el género, el tipo de institución, la condición de deserción o permanencia, el hecho de trabajar o no mientras se estudia, o la clasificación del rendimiento en niveles, la herramienta fundamental es la tabla de frecuencias. SPSS permite generar tablas en las que se muestran tanto los recuentos absolutos como los porcentajes asociados a cada categoría, así como porcentajes acumulados cuando tiene sentido ordenarlas. Estas tablas constituyen una base imprescindible para responder preguntas simples pero relevantes, por ejemplo, qué proporción de estudiantes se ubica en cada nivel de rendimiento, cuántos han abandonado los estudios en la cohorte analizada o cómo se

distribuye la muestra según género o turno de estudio, cuestiones habituales en investigaciones sobre deserción y progreso académico.

La Figura 11 muestra un ejemplo de tabla de frecuencias generada por el procedimiento *Frecuencias* de SPSS, similar a las tablas presentadas en la guía breve para variables categóricas. A partir de una tabla como la representada, el lector aprende a redactar interpretaciones sencillas y precisas, del tipo “el 18 % de los estudiantes de la cohorte analizada figura como desertor”, evitando expresiones ambiguas o confusas.

**Figura 11.**

*Tabla de frecuencias generada con el procedimiento Frecuencias*

**Output1 - Viewer**

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

Input

Frequencies

Title Notes Statistics

Frequency Tab

Owns PDA Owns TV

**Frequency Table**

**Owning PDA**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid No	5093	79.6	79.6	79.6
Valid Yes	1307	20.4	20.4	100.0
Total	6400	100.0	100.0	

**Owning TV**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid No	63	1.0	1.0	1.0
Valid Yes	6337	99.0	99.0	100.0
Total	6400	100.0	100.0	

Las variables cuantitativas, por su parte, requieren un conjunto de medidas descriptivas más amplio. En los textos de estadística se insiste en la tríada clásica de tendencia central (media, mediana, moda), dispersión (rango, varianza, desviación estándar) y forma de la distribución (simetría, curtosis), a menudo complementada con medidas de posición como los cuartiles o percentiles. SPSS permite obtener estas medidas mediante procedimientos de descriptivos generales o de exploración, que ofrecen no solo los estadísticos numéricos sino también diagramas de caja e histogramas.

En el contexto de la educación superior, utilizar estas herramientas con variables como la calificación promedio, las horas de estudio semanal, el porcentaje de créditos aprobados o los puntajes en escalas de integración académica y social ayuda a construir una imagen más rica del grupo estudiado. En la Tabla 3 se presenta un modelo de salida de SPSS para descriptivos de calificación, donde se incluyen la media, la desviación estándar, el mínimo, el máximo y el número de casos válidos; este tipo de tabla será retomado más adelante para mostrar cómo se incorporan estos resultados a informes y capítulos de tesis.

**Tabla 3.**

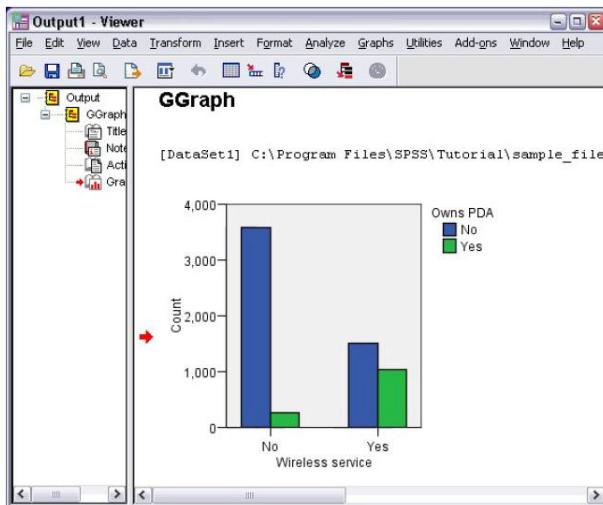
*Estadísticos descriptivos de la calificación promedio*

Variable	N	Mínimo	Máximo	Media	Desv. típ.
Calificación promedio	120	4.0	9.8	7.21	1.12

Los gráficos ocupan un lugar esencial en la estadística descriptiva y, si se emplean adecuadamente, pueden mejorar la comprensión y la retención de los conceptos estadísticos, tal como destacan distintos autores en educación estadística. SPSS dispone de un repertorio variado de representaciones gráficas: diagramas de barras y de pastel para variables categóricas, histogramas y diagramas de caja para variables cuantitativas, así como diagramas de dispersión para explorar relaciones bivariadas. Volviendo al ejemplo de la deserción, un sencillo gráfico de barras que muestre el porcentaje de desertores y de estudiantes que permanecen puede ser más impactante que una tabla, sobre todo al comunicar los resultados a audiencias no especializadas, como autoridades académicas o responsables de programas de apoyo. La Figura 12 reproduce un gráfico de barras obtenido en el Visor de resultados de SPSS, análogo al que se muestra en la guía para representar la distribución de variables categóricas.

**Figura 12.**

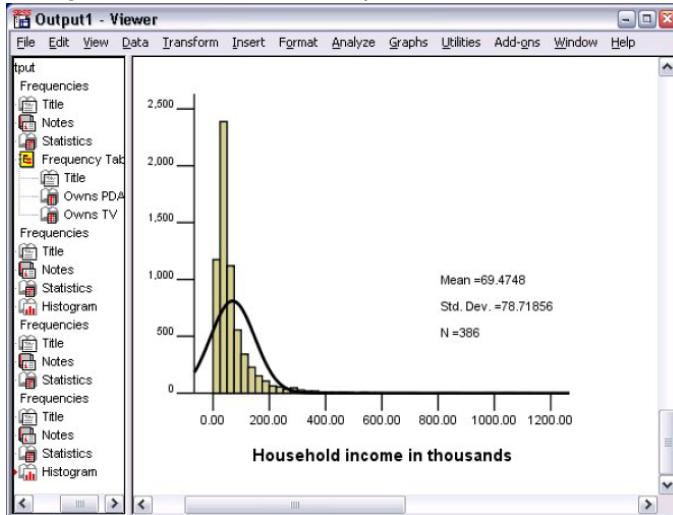
*Gráfico de barras de una variable categórica en el visor*



En el caso de variables continuas, el histograma y el diagrama de caja permiten visualizar la distribución y detectar asimetrías, colas largas o posibles valores extremos. Por ejemplo, un histograma de horas de estudio semanal podría mostrar una concentración fuerte en torno a pocas horas y una cola hacia la derecha con estudiantes que declaran estudiar muchas horas, lo que invitaría a reflexionar sobre la diversidad de perfiles dentro de un mismo grupo. Un diagrama de caja de las calificaciones dividido por género o por tipo de institución puede revelar diferencias en dispersión o mediana que no se aprecian con facilidad en las tablas numéricas. La Figura 13 ilustra un histograma con curva normal para una variable de escala, siguiendo el ejemplo de la guía breve, donde se analiza la distribución de los ingresos del hogar.

**Figura 13.**

*Histograma con curva normal para una variable de escala*



Más allá de generar tablas y gráficos, uno de los aprendizajes clave en esta fase es el de transformar la salida de SPSS en un discurso estadístico coherente. Los resultados descriptivos deben integrarse en el relato de la investigación, no limitarse a una sucesión de valores. En los trabajos de fin de grado, tesis y artículos revisados, es habitual encontrar párrafos donde se presentan las características de la muestra de manera ordenada: primero las variables sociodemográficas básicas, luego indicadores de rendimiento, dedicación o uso de tecnologías, y finalmente las variables específicas del estudio, siempre con una combinación de porcentajes y medidas de tendencia central y dispersión.

SPSS facilita este proceso al permitir copiar tablas y gráficos al procesador de texto, pero la calidad académica del capítulo depende de que la redacción sea clara, precisa y conecte las cifras con las preguntas de investigación. En este sentido, redactar frases como “la media de créditos aprobados es de 34,2 (DE = 9,7), con un mínimo de 12 y un máximo de 60” o “el 27,5 % de la muestra se ubica en el nivel bajo de rendimiento” es parte del aprendizaje del lenguaje estadístico aplicado.

La estadística descriptiva en SPSS también ofrece un espacio privilegiado para abordar dificultades frecuentes del estudiantado. Investigaciones sobre errores en inferencia y comprensión de conceptos estadísticos señalan que muchos alumnos confunden porcentajes y frecuencias, interpretan de forma errónea gráficos de barras y de pastel o tienen dificultades para relacionar medidas de tendencia central con la forma de la distribución (Rodríguez, Montañez y Rojas, 2010; Vallecillos, 1999; López-Martín et al., 2019). Aprovechar los resultados de SPSS para discutir estas dificultades en el aula –por ejemplo, pidiendo que diferentes grupos de estudiantes interpreten la misma tabla o gráfico y contrasten sus conclusiones– contribuye a un aprendizaje más profundo y alineado con los modelos de educación estadística que buscan una retención duradera y un razonamiento crítico (Sowey, 2020; Shaughnessy, 2006).

Conviene subrayar que la estadística descriptiva no es una fase meramente preliminar, sino un componente esencial del análisis. Muchos estudios aplicados en educación y ciencias sociales combinan descripciones detalladas de la muestra con análisis inferenciales, y en ocasiones las decisiones más relevantes se basan precisamente en patrones descriptivos claros, como la concentración de la deserción en determinados semestres, la desigual distribución del rendimiento entre grupos o la elevada variabilidad en determinados indicadores de integración académica. Entender la lógica y el potencial de estas descripciones, y aprender a producirlas e interpretarlas con SPSS, forma parte de la competencia estadística que este capítulo pretende desarrollar.

## 2.6. Estadística inferencial con SPSS

La estadística inferencial con SPSS constituye el núcleo de la toma de decisiones basada en datos, porque permite ir más allá de la simple descripción de una muestra y extraer conclusiones sobre una población más amplia bajo condiciones de incertidumbre. Mientras que la estadística descriptiva resume lo que se observa, la inferencial se ocupa de estimar parámetros y contrastar hipótesis, apoyándose en el cálculo de probabilidades y en el uso de muestras aleatorias. En los manuales de metodología se insiste en que ninguna prueba de hipótesis tiene sentido si no está anclada en un problema de investigación, en objetivos claros y en una estrategia de muestreo razonable. SPSS proporciona un amplio repertorio de procedimientos para contrastar hipótesis y estimar parámetros, pero la responsabilidad de elegirlos adecuadamente e interpretar sus resultados recae siempre en el investigador o en el estudiante que los utiliza.

El punto de partida de la estadística inferencial es el razonamiento en términos de hipótesis nula e hipótesis alternativa, nivel de significación y p-valor. La lógica básica consiste en formular una hipótesis nula que suele representar una situación de “no efecto” o “no diferencia” –por ejemplo, que no hay diferencias entre hombres y mujeres en el promedio de calificaciones, o que el rendimiento medio de estudiantes que asisten a tutorías no difiere del de quienes no asisten–, y una hipótesis alternativa que recoge el efecto o diferencia de interés. SPSS calcula un estadístico de prueba ( $t$ ,  $F$ ,  $\chi^2$ , entre otros) y el p-valor asociado, que expresa la probabilidad de observar, bajo la hipótesis nula, un resultado igual o más extremo que el obtenido con la muestra. Según el criterio fijado (por ejemplo, un nivel de significación de 0,05), se decide si se rechaza o no la hipótesis nula. Sin embargo, estudios sobre dificultades en la comprensión de la inferencia han mostrado que el alumnado suele interpretar erróneamente el p-valor, confundir significación estadística con importancia práctica y tener problemas para conectar las tablas de salida con el contexto del estudio. Por ello, uno de los desafíos al trabajar con SPSS es evitar que el estudiante reduzca la interpretación a “mirar si  $p$  es menor que 0,05”, sin considerar supuestos, tamaño del efecto, intervalos de confianza y sentido sustantivo de los resultados.

En investigaciones educativas y de ciencias sociales, uno de los escenarios más habituales es la comparación de medias entre grupos. SPSS ofrece procedimientos como la prueba  $t$  para muestras independientes o relacionadas y el análisis de varianza de un factor, accesibles desde el menú de comparación de medias. Conceptualmente, la prueba  $t$  para muestras independientes permite contrastar si dos grupos –por ejemplo, estudiantes de primero frente a estudiantes de cursos superiores, o quienes reciben apoyo tutorial frente a quienes no lo reciben– presentan o no diferencias estadísticamente significativas en una variable cuantitativa como el promedio de calificaciones o las horas de estudio (Levin, 2014; Verdoy, Beltrán y Peris, 2015). La prueba  $t$  para muestras relacionadas se utiliza cuando las mediciones provienen del mismo grupo en dos momentos distintos, como antes y después de

una intervención formativa, y permite valorar si la diferencia media entre ambas mediciones difiere de cero de manera significativa. Cuando el número de grupos es mayor de dos, el análisis de varianza de un factor constituye la herramienta principal para comparar medias, por ejemplo, al estudiar diferencias de rendimiento entre tipos de institución (pública, privada, concertada) o entre distintos niveles de dedicación al estudio.

Otro conjunto de situaciones frecuentes en la investigación aplicada se refiere al análisis de la asociación entre variables categóricas. Preguntas como si la deserción está relacionada con el género, el turno de estudio, el tipo de institución o el hecho de trabajar mientras se estudia se formulan naturalmente en términos de tablas de contingencia y pruebas de independencia. SPSS permite construir tablas de doble entrada en las que se muestran frecuencias y porcentajes por filas y columnas, y aplicar la prueba de chi-cuadrado para valorar si la distribución conjunta de las categorías se desvía de lo que cabría esperar bajo independencia (Levin, 2014; Mendenhall, 2014). Esta herramienta ha sido ampliamente utilizada en estudios sobre abandono y progresión académica, así como en análisis de equidad y distribución de oportunidades educativas (Korhonen y Rautopuro, 2018; Lacave, Molina y Cruz-Lemus, 2018; Mesa, 2016).

Las relaciones entre variables cuantitativas se abordan habitualmente mediante correlación y regresión. La correlación de Pearson mide el grado de asociación lineal entre dos variables numéricas, como las horas de estudio y la nota media, el porcentaje de créditos aprobados y la percepción de integración académica, o los puntajes en una escala de competencia de emprendimiento social y el rendimiento (García-González y Ramírez-Montoya, 2020; Kori, Pedaste y Must, 2018; Meyer y Fang, 2019). SPSS genera matrices de correlación en las que se presentan los coeficientes de correlación, los p-valores y el tamaño de la muestra utilizada; la interpretación de estos resultados exige considerar tanto el signo y magnitud del coeficiente como el contexto del estudio y la plausibilidad de relaciones causales. La regresión lineal simple, por su parte, permite modelar una variable dependiente cuantitativa en función de una variable independiente, proporcionando un coeficiente de regresión (pendiente), un intercepto y un coeficiente de determinación ( $R^2$ ) que indica la proporción de variabilidad explicada. En estudios de educación superior se han empleado modelos de este tipo para explorar hasta qué punto variables como el tiempo dedicado al estudio, la asistencia a actividades formativas o determinados indicadores de integración explican diferencias en el rendimiento (Levin, 2014; Verdoy et al., 2015).

Cuando la variable de interés es dicotómica, como permanecer o desertar, aprobar o suspender, empleo o desempleo, la regresión logística binaria se convierte en una herramienta especialmente potente. A diferencia de la regresión lineal, que presupone una variable dependiente continua, la regresión logística modela la probabilidad de que ocurra un determinado evento en función de un conjunto de variables predictoras, como el rendimiento previo, la nota de acceso, la provincia de procedencia o la opción de carrera, entre otras (Berlanga-Silvente y Vilà-Baños, 2014; Korhonen y Rautopuro, 2018; Lacave et al., 2018; Lázaro, 2020). SPSS implementa este procedimiento, permitiendo estimar coeficientes logísticos y odds ratios que se interpretan en términos de incremento o disminución relativa de la probabilidad del evento por unidad de cambio en la variable explicativa.

En todos estos casos, el aporte de SPSS no radica solamente en calcular rápidamente estadísticos y p-valores, sino en proporcionar una estructura de salida que, si se utiliza pedagógicamente, puede ayudar al estudiante a leer y comprender los componentes de un análisis inferencial. No obstante, la literatura en educación estadística recuerda que el uso de software no elimina las dificultades conceptuales de la inferencia. Al contrario, puede enmascararlas si se reduce el trabajo a “seguir pasos” sin conectar los resultados con el contexto, con los supuestos de los modelos y con las decisiones que se pretenden fundamentar (Rodríguez et al., 2010; López-Martín et al., 2019; Sowey, 2020; Shaughnessy, 2006). En este sentido, aprovechar SPSS en la enseñanza de la estadística inferencial implica detenerse a discutir por qué se elige una prueba y no otra, qué significa realmente “rechazar la hipótesis nula”, cómo se interpreta un intervalo de confianza o una odds ratio, y qué importancia tiene el tamaño del efecto más allá de la significación estadística.

Finalmente, la estadística inferencial con SPSS orientada a la toma de decisiones se vuelve especialmente relevante cuando se trabaja con problemáticas complejas como la deserción estudiantil, la progresión académica o la evaluación de programas formativos. Los ejemplos de la literatura muestran que la combinación de comparaciones de medias, análisis de asociaciones categóricas, correlaciones y modelos de regresión proporciona un retrato mucho

más matizado de estos fenómenos que el que podría obtenerse con descripciones aisladas (Korhonen y Rautopuro, 2018; Lacave et al., 2018; Lázaro, 2020; Meyer y Fang, 2019). La Tabla 4, ya mencionada, sintetiza la elección de procedimientos inferenciales en función del tipo de variables, el diseño del estudio y la pregunta de investigación, y puede utilizarse como herramienta de referencia rápida para estudiantes y profesionales. Entender esta lógica de correspondencia entre problema, modelo estadístico y salida de SPSS es un paso fundamental hacia una competencia estadística que integre teoría, práctica de software y responsabilidad en la toma de decisiones.

## 2.7. SPSS en estudios aplicados: caso de la deserción estudiantil

La utilización de SPSS en estudios sobre deserción estudiantil en educación superior permite ilustrar de manera muy clara cómo un paquete estadístico se integra en todo el ciclo de investigación, desde la formulación del problema hasta la toma de decisiones institucionales. La deserción no es solo un indicador académico aislado, sino un fenómeno complejo que afecta a la equidad, la eficiencia del sistema y las trayectorias vitales de los estudiantes, por lo que su análisis exige combinar variables académicas, personales e institucionales en modelos que permitan identificar factores de riesgo y perfiles de alumnado vulnerable. En la literatura reciente se han empleado diseños cuantitativos apoyados en SPSS para estudiar la progresión académica, detectar estudiantes en riesgo y analizar patrones de permanencia y abandono, tanto en contextos presenciales como en titulaciones de informática y otras ingenierías, donde este problema adquiere una relevancia particular (Korhonen y Rautopuro, 2018; Lacave, Molina y Cruz-Lemus, 2018; Lázaro, 2020; Meyer y Fang, 2019; Mesa, 2016).

Un estudio típico de deserción apoyado en SPSS suele comenzar con la definición precisa del evento de interés y de la ventana temporal de observación. Es necesario decidir qué se considerará “desertar”: por ejemplo, no matricularse en ningún curso en un determinado periodo, causar baja definitiva en la titulación o acumular un cierto número de semestres sin actividad académica registrada. Esta decisión se traduce en la construcción de una variable dependiente, generalmente dicotómica, que toma el valor 1 si el estudiante cumple el criterio de deserción y 0 si permanece. En paralelo, se definen las variables explicativas que recogen información previa al ingreso (nota de acceso, tipo de institución de procedencia, opción de carrera), características sociodemográficas (edad, género, situación laboral), y variables de trayectoria (rendimiento en asignaturas clave, número de créditos aprobados por semestre, participación en actividades de apoyo académico). En la Tabla 5 se propone un esquema de estas variables, con sus nombres en SPSS, etiquetas, tipo de medida y codificación, que permite visualizar cómo el marco conceptual del estudio se traduce en columnas concretas dentro de la base de datos.

Una vez construida la base, el primer paso en SPSS consiste en realizar un análisis descriptivo que ofrezca una panorámica de la cohorte estudiada. La estadística descriptiva permite conocer la proporción de estudiantes que han desertado, la distribución por género, la edad media, el rendimiento medio en el primer año, el porcentaje de créditos aprobados y otras características relevantes. Estos resultados suelen presentarse en tablas de frecuencias para variables categóricas y en tablas de descriptivos para variables cuantitativas, acompañadas de gráficos que facilitan la interpretación por parte de decisores institucionales.

El siguiente nivel de análisis consiste en explorar asociaciones bivariadas entre la deserción y posibles factores de riesgo, aprovechando las tablas de contingencia y la prueba de chi-cuadrado. SPSS permite cruzar, por ejemplo, la variable de deserción con el género, el tipo de dedicación (exclusivamente estudio frente a estudio y trabajo), el tipo de institución de procedencia o la pertenencia a determinados grupos de ingreso. El resultado son tablas que muestran frecuencias y porcentajes por filas o columnas, junto con el valor del estadístico  $\chi^2$  y su significación.

Sin embargo, la deserción suele depender de la combinación de múltiples factores, y no es suficiente analizar relaciones dos a dos. En este punto entran en juego los modelos de regresión que SPSS implementa y que permiten estudiar efectos conjuntos, controlando por otras variables. Cuando la variable dependiente es dicotómica (deserta/permanence), la regresión logística binaria se ha consolidado como la herramienta central para modelar la

probabilidad de abandono en función de un conjunto de predictores. A través de este procedimiento es posible estimar, por ejemplo, cómo cambia la probabilidad de deserción al variar la nota media del primer año, la carga de trabajo, la edad de ingreso o el hecho de haber elegido la carrera en primera opción, manteniendo constantes las demás variables incluidas en el modelo (Berlanga-Silvente y Vilà-Baños, 2014; Lacave et al., 2018; Lázaro, 2020). La salida de SPSS para este tipo de análisis se recomienda organizar en varias tablas.

Además de los modelos de regresión logística, en algunos estudios se han utilizado técnicas complementarias, como análisis de supervivencia o modelos de progresión, que permiten considerar el tiempo hasta la deserción y no solo el hecho de que esta ocurra o no. Aunque estos procedimientos pueden requerir módulos específicos o incluso otros paquetes, SPSS ofrece una puerta de entrada a este tipo de análisis al proporcionar funciones de tablas de vida y pruebas de comparación entre curvas de supervivencia.

Un aspecto central del uso de SPSS en estudios de deserción es la traducción de los resultados estadísticos en recomendaciones concretas de política académica y acciones institucionales. El análisis puede revelar, por ejemplo, que la nota de acceso en determinadas materias, la repetición temprana de asignaturas básicas, la incompatibilidad entre trabajo y estudio o la falta de integración académica se asocian de manera significativa con la probabilidad de abandono. La interpretación de los coeficientes de regresión y de las odds ratios permite cuantificar estos riesgos y priorizar intervenciones: reforzar los cursos introductorios, diseñar tutorías específicas para ciertas cohortes, ajustar los criterios de ingreso o crear programas de acompañamiento para estudiantes que trabajan.

Desde la perspectiva de la enseñanza de la estadística, incorporar un caso aplicado de deserción trabajado paso a paso con SPSS tiene un fuerte potencial formativo. No solo permite a los estudiantes ver en acción las herramientas descriptivas, las tablas de contingencia, las pruebas de chi-cuadrado, la regresión logística y las representaciones gráficas, sino que también les ayuda a comprender cómo estos instrumentos se articulan para abordar un problema real de alta relevancia social. Al reconstruir el proceso completo –desde la definición de la variable deserción y la construcción de la base de datos hasta la interpretación de los modelos y la formulación de propuestas de intervención–, el capítulo muestra que la estadística con SPSS no es una colección de técnicas aisladas, sino un recurso para comprender y transformar la realidad educativa.

## **2.8. Acciones metodológicas para enseñar y aprender estadística con SPSS**

Las acciones metodológicas para enseñar y aprender estadística con apoyo de SPSS se sitúan en la intersección entre tres dimensiones: el contenido estadístico, la herramienta tecnológica y los contextos reales de aplicación. La literatura en educación estadística insiste en que la comprensión duradera de los conceptos no se logra solo mediante la exposición magistral, sino a través de tareas en las que el estudiantado analiza datos significativos, explica sus razonamientos, confronta ideas y revisa sus interpretaciones (Shaughnessy, 2006; Sowey, 2020). Si el uso de SPSS se limita a seguir instrucciones paso a paso sin conexión con problemas auténticos, es fácil que se consolide una visión “tecnocéntrica”, donde el objetivo parece ser manejar botones y menús, más que desarrollar pensamiento estadístico. Por el contrario, cuando SPSS se integra en secuencias didácticas que enfatizan la formulación de preguntas, el diseño de estudios, la interpretación crítica de la salida del programa y la comunicación de resultados, se convierte en un catalizador de aprendizajes más profundos, alineados con las demandas de la educación superior y el enfoque de Ciencia, Tecnología y Sociedad (Amat Abreu, 2017; Castro Díaz-Balar, 2004; Mesa, 2016).

Una primera línea de acción consiste en organizar el trabajo de SPSS alrededor de proyectos o estudios de caso que recorran el ciclo completo de investigación cuantitativa. En lugar de proponer ejercicios aislados sobre “cómo hacer una prueba t”, se plantean problemas contextualizados –por ejemplo, analizar factores relacionados con el rendimiento en una asignatura clave, identificar perfiles de estudiantes en riesgo de deserción o evaluar el impacto de una estrategia de formación continua– en los que el estudiantado debe formular preguntas, definir variables, recoger o recibir datos, construir la base en SPSS, realizar

análisis descriptivos e inferenciales y redactar un informe final (Hernández-Sampieri, Fernández-Collado y Baptista-Lucio, 2014; Gómez et al., 2020).

Una segunda línea de trabajo apunta a la explicitación y abordaje de las dificultades frecuentes del alumnado en estadística, especialmente en lo relativo a la inferencia. Diversas investigaciones han documentado errores persistentes en la comprensión de la significación estadística, la interpretación de intervalos de confianza, el razonamiento sobre hipótesis nula y alternativa o la lectura de tablas de salida de software (Rodríguez, Montañez y Rojas, 2010; Vallecillos, 1999; López-Martín, Batanero y Gea, 2019). SPSS puede utilizarse estratégicamente para hacer visibles estas dificultades, por ejemplo proponiendo actividades en las que diferentes grupos analizan la misma salida de una prueba de hipótesis y deben explicar sus conclusiones por escrito, comparar interpretaciones y discutir discrepancias.

Una tercera dimensión metodológica se relaciona con el diseño de actividades graduadas que articulen trabajo guiado, prácticas semiabiertas y proyectos autónomos. En las primeras sesiones, puede ser conveniente utilizar guías detalladas en las que se explica cada paso en SPSS, acompañando la acción en el software con preguntas de reflexión conceptual sobre tipos de variable, niveles de medida, elección de técnicas y lectura de tablas (Barreiro Felpeto, Guisande González y Vamonde Liste, 2014; Verdoy, Beltrán y Peris, 2015). Progresivamente, estas guías pueden volverse menos prescriptivas, dejando mayor espacio a que el estudiantado decida qué procedimientos aplicar y cómo organizar la salida, hasta llegar a proyectos en los que solo se fijan los objetivos generales y algunos criterios de calidad del informe final. En la Tabla 6 se presenta un ejemplo de progresión de tareas con SPSS a lo largo de un curso, donde se vinculan resultados de aprendizaje (describir distribuciones, comparar grupos, modelizar relaciones) con tipos de actividades (ejercicio guiado, estudio de caso, proyecto final) y se indican los menús y procedimientos de SPSS que se espera que el alumnado domine en cada etapa.

**Tabla 6.** *Progresión de tareas con SPSS en un curso de estadística aplicada*

Fase del curso	Resultado de aprendizaje principal	Tipo de actividad	Procedimientos de SPSS implicados	Evidencias / productos esperados
Inicial	Reconocer el entorno de SPSS y diferenciar vista de datos y vista de variables. Identificar tipos de variables y niveles de medida.	Ejercicio guiado en laboratorio, con hoja de ruta detallada.	Apertura/creación de archivo de datos, definición de variables (nombre, etiqueta, tipo, valores, medida). Navegación básica por los menús.	Hoja de SPSS con una base de datos sencilla correctamente estructurada (sin análisis). Breve nota explicativa de las variables.
Inicial	Describir distribuciones de variables categóricas y cuantitativas.	Ejercicios guiados con datos proporcionados.	Descriptive Statistics → Frequencies; Descriptives; Explore. Gráficos básicos (barras, pastel, histogramas, diagramas de caja).	Tablas de frecuencias y descriptivos, gráficos básicos y un párrafo de interpretación para cada variable.
Intermedia	Relacionar el tipo de variable y el diseño con la elección de pruebas de comparación de grupos.	Estudio de caso semiabierto (el docente propone el contexto y las variables).	Compare Means → Independent-Samples T Test, Paired-Samples T Test, One-Way ANOVA.	Informe breve con: descripción de la muestra, elección justificada de la prueba, tabla de salida de SPSS

				seleccionada y conclusiones en lenguaje no técnico.
Intermedia	Analizar asociaciones entre variables categóricas y relaciones lineales entre variables cuantitativas.	Taller en grupos pequeños con discusión plenaria.	Crosstabs (chi-cuadrado), Correlate → Bivariate; gráficos de dispersión.	Tablas de contingencia con porcentajes interpretados, matriz de correlaciones y al menos un gráfico de dispersión comentado.
Intermedia-avanzada	Construir índices o puntuaciones compuestas y gestionar datos (recodificar, seleccionar subgrupos).	Actividades prácticas centradas en una base de datos común para toda la clase.	Transform → Recode into Different Variables; Compute Variable; Data → Select Cases / Split File.	Base de datos enriquecida con nuevas variables derivadas y documento breve que justifique las decisiones de recodificación y selección.
Avanzada	Modelizar fenómenos educativos sencillos mediante regresión (lineal o logística, según el nivel del curso).	Estudio de caso aplicado (p. ej. rendimiento o deserción).	Analyze → Regression → Linear; (opcional) Regression → Binary Logistic.	Informe estructurado que incluya: planteamiento del problema, descripción de variables, tabla(s) de SPSS seleccionadas, interpretación de coeficientes y discusión de implicaciones.
Final	Integrar todo el ciclo de investigación cuantitativa apoyado en SPSS.	Proyecto autónomo (trabajo final de curso) con tema elegido o acotado.	Combinación de todos los procedimientos vistos: gestión de datos, descriptiva, inferencial y presentación de resultados.	Informe final completo (tipo mini-artículo o capítulo de tesis), con anexos de tablas y gráficos generados en SPSS y reflexión sobre limitaciones del estudio.

El trabajo en laboratorio de informática constituye un espacio privilegiado para desplegar estas acciones metodológicas. No se trata únicamente de que el estudiantado “haga prácticas” con SPSS, sino de que en ese entorno puedan experimentar con datos reales, cometer errores, revisar decisiones y observar en tiempo real cómo cambian los resultados cuando se modifican las codificaciones o se seleccionan subgrupos. Los seminarios

investigativos, en los que pequeños grupos trabajan con bases de datos relacionadas con problemas de su entorno –por ejemplo, muestras de estudiantes de su propia institución, datos de evaluaciones internas o encuestas relacionadas con su campo profesional–, refuerzan la percepción de pertinencia del contenido estadístico y permiten integrar discusiones sobre equidad, inclusión y calidad educativa (Amat Abreu, 2017; Mesa, 2016).

Por último, es importante considerar que el uso de SPSS en la formación estadística no solo persigue el dominio de técnicas, sino el desarrollo de competencias transversales como la interpretación de información cuantitativa, la comunicación de resultados y la toma de decisiones fundamentadas. Proponer al estudiantado la elaboración de informes breves en los que deban contextualizar los datos, describir la muestra, presentar resultados descriptivos e inferenciales, valorar la calidad de los datos y reconocer las limitaciones del estudio contribuye a articular la dimensión técnica con la argumentación escrita y el pensamiento crítico (Sowey, 2020; Shaughnessy, 2006). En este sentido, SPSS se convierte en un medio para entrenar la lectura y producción de textos académicos que incorporan tablas y gráficos, algo indispensable en la mayoría de campos profesionales vinculados a la educación, la administración y las ciencias sociales. Integrar estas tareas en un marco de mejora continua –donde los productos se discuten, reciben retroalimentación y se revisan– refuerza la idea de que aprender estadística con SPSS es un proceso gradual, en el que la comprensión conceptual, el manejo de la herramienta y la responsabilidad en el uso de los resultados se construyen de manera conjunta.

## CONCLUSIONS

Este capítulo ha presentado el uso de SPSS como una herramienta para el análisis de datos en el contexto de la educación superior y las ciencias sociales, intentando articular tres dimensiones inseparables: la comprensión conceptual de la estadística, el dominio básico del software y la relación con problemas reales de investigación y toma de decisiones. Partiendo de una introducción donde se resaltó la importancia de la estadística en la formación universitaria y en la sociedad contemporánea, se situó a SPSS dentro del conjunto de paquetes estadísticos disponibles, destacando sus ventajas para la docencia y la investigación aplicada, así como algunas de sus limitaciones y riesgos si se utiliza de manera acrítica.

Se describió el entorno de trabajo del programa, subrayando la relevancia de comprender la estructura de la vista de datos y la vista de variables, la organización de los menús y la función de la ventana de resultados y de la sintaxis. Se insistió en que las decisiones sobre nombres, tipos de variable, etiquetas de valores y niveles de medida no son meramente técnicas, sino que reflejan la manera en que se ha operacionalizado el problema de investigación. El paso de la pregunta de investigación a la base de datos en SPSS se mostró como un tránsito en el que se concretan los constructos teóricos en variables observables, se define la población y la muestra y se anticipan las técnicas de análisis que se usarán más adelante.

La sección dedicada a la gestión y transformación de datos puso de relieve que una parte importante del trabajo estadístico se concentra en la depuración de la información, la recodificación de variables, la construcción de índices y el tratamiento de valores perdidos, así como en la selección y segmentación de casos. Lejos de ser una fase puramente operativa, estas decisiones condicionan la calidad y el alcance de los análisis posteriores. La estadística descriptiva con SPSS se presentó entonces como la primera aproximación sistemática a los datos: tablas de frecuencias, medidas de tendencia central y dispersión, histogramas, diagramas de caja y gráficos de barras constituyen el lenguaje básico con el que se describe la muestra y se identifican patrones, desigualdades y posibles problemas que deben considerarse antes de pasar a la inferencia.

Sobre esa base, la parte dedicada a la estadística inferencial mostró cómo SPSS permite aplicar de forma relativamente accesible procedimientos para comparar medias, analizar asociaciones entre variables categóricas y estudiar relaciones lineales mediante correlación y regresión. Se enfatizó que el núcleo de la inferencia no reside en “obtener p-valores”, sino en formular y contrastar hipótesis bien planteadas, interpretar resultados a la luz de los supuestos de los modelos y valorar tanto la significación estadística como la relevancia práctica y las limitaciones del estudio. La discusión sobre la regresión logística binaria, en particular, señaló su utilidad para modelizar fenómenos dicotómicos como la deserción,

siempre que se integre con un análisis crítico de los factores implicados y sus implicaciones educativas.

El caso aplicado sobre deserción estudiantil sirvió para integrar en un mismo hilo narrativo los distintos procedimientos trabajados: desde la definición de la variable “deserción” y la construcción de la base de datos, pasando por la descripción de la cohorte y las asociaciones bivariadas, hasta la estimación de modelos de regresión orientados a identificar factores de riesgo y a fundamentar decisiones institucionales. Este ejemplo puso de manifiesto que SPSS no es solo un recurso para resolver ejercicios de aula, sino una herramienta que puede contribuir a comprender y afrontar problemas educativos de alto impacto. Finalmente, la sección sobre acciones metodológicas ofreció orientaciones para incorporar SPSS a la enseñanza de la estadística de manera que favorezca el razonamiento estadístico, la interpretación crítica y la elaboración de informes, a través de proyectos, estudios de caso, actividades graduadas y seminarios de investigación.

De esta manera, el capítulo aspira a que el lector no solo aprenda a “manejar SPSS”, sino que desarrolle una competencia estadística aplicada en la que las operaciones realizadas en el software estén siempre conectadas con una comprensión sólida de los conceptos y con la responsabilidad de usar los resultados para tomar decisiones fundamentadas. La invitación final es a seguir profundizando en técnicas más avanzadas de análisis de datos, y, sobre todo, a continuar integrando la estadística y el uso de herramientas como SPSS en una práctica profesional crítica, reflexiva y orientada a la mejora de la educación y de los contextos sociales en los que esta se desarrolla.

## REFERENCES

- Amat Abreu, M. (2017). *La estadística en la educación superior. Aplicaciones en las carreras técnicas de UNIANDES*. Editorial El Siglo.
- Aznar-Díaz, I., Romero-Rodríguez, J. M., García-González, A., & Ramírez-Montoya, M. S. (2020). Mexican and Spanish university students' Internet addiction and academic procrastination: Correlation and potential factors. *PloS one*, 15(5), e0233655. <https://doi.org/10.1371/journal.pone.0233655>
- Barreiro Felpeto, A., Guisande González, C., & Vamonde Liste, A. (2014). *Tratamiento de datos con R, Estadística y SPSS*. Ediciones Díaz de Santos.
- Berlanga-Silvente, V., & Vilà-Baños, R. (2014). Cómo obtener un Modelo de Regresión Logística Binaria con SPSS. *REIRE*, 7(2), 105-118. <https://doi.org/10.1344/reire2014.7.2727>
- Canedo, E. D., Santos, G. A., & Leite, L. L. (2018). An Assessment of the Teaching-Learning Methodologies Used in the Introductory Programming Courses at a Brazilian University. *Informatics in Education*, 17(1), 45-59. <https://doi.org/10.15388/infedu.2018.03>
- Castro Díaz-Balar, F. (2004). *Ciencia, Tecnología y Sociedad. Hacia un desarrollo sostenible en la Era de la Globalización*. Científico-Técnica.
- García-González, A., & Ramírez-Montoya, M. S. (2020). Social entrepreneurship competency in higher education: an analysis using mixed methods. *Journal of Social Entrepreneurship*, 1-19. <https://doi.org/10.1080/19420676.2020.1823872>
- Gómez, G. Á., Moya, J. V., Ricardo, J. E., & Sánchez, C. B. V. (2020). *Evaluating Strategies of Continuing Education for Academics Supported in the Pedagogical Model and Based on Plithogenic Sets* (Vol. 37). Infinite Study. [https://digitalrepository.unm.edu/cgi/viewcontent.cgi?article=1651&context=nss\\_jou\\_rnal](https://digitalrepository.unm.edu/cgi/viewcontent.cgi?article=1651&context=nss_jou_rnal)
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, M. del P. (2014). *Metodología de la Investigación* (6.<sup>a</sup> ed.). McGraw-Hill.
- Korhonen, V., & Rautopuro, J. (2018). Identifying Problematic Study Progression and “At-Risk” Students in Higher Education in Finland. *Scandinavian Journal of Educational Research*, 0(0), 1-14. <https://doi.org/10.1080/00313831.2018.1476407>
- Kori, K., Pedaste, M., & Must, O. (2018). The Academic, Social, and Professional Integration Profiles of Information Technology Students. *ACM Transactions on Computing Education (TOCE)*, 18(4), 20:1-20:19. <https://doi.org/10.1145/3183343>
- Lacave, C., Molina, A. I., & Cruz-Lemus, J. A. (2018). Learning Analytics to identify dropout factors of Computer Science studies through Bayesian networks. *Behaviour & Information Technology*, 1-15. <https://doi.org/10.1080/0144929X.2018.1485053>

- Lázaro, N. (2020). *Acciones tutoriales con TIC atendiendo a factores predictivos de la deserción estudiantil en carreras de Ingeniería Informática* (Tesis doctoral). Universidad de Granada. <https://digibug.ugr.es/handle/10481/64571>
- Levin, R. (2014). *Estadística para administración y economía*. Editorial Pearson.
- Leyva-Vázquez, M., Quiroz-Martínez, M. A., Portilla-Castell, Y., Hechavarria-Hernández, J. R., & González-Caballero, E. (2020). A new model for the selection of information technology project in a neutrosophic environment. *Neutrosophic Sets and Systems*, 32(1), 344-360. [https://digitalrepository.unm.edu/cgi/viewcontent.cgi?article=1494&context=nss\\_journal](https://digitalrepository.unm.edu/cgi/viewcontent.cgi?article=1494&context=nss_journal)
- Lind, D. A., Marchal, W. G., Wathen, S. A., Obón León, M. D. P., & León Cárdenas, J. (2016). *Estadística aplicada a los negocios y la economía*. McGraw-Hill/Interamericana Editores.
- López-Martín, M. D. M., Batanero, C., & Gea, M. M. (2019). ¿Conocen los futuros profesores los errores de sus estudiantes en la inferencia estadística? *Bolema: Boletín de Educação Matemática*, 33(64), 672-693. <https://doi.org/10.1590/1980-4415v33n64a11>
- Mendenhall, R. (2014). *Introducción a la probabilidad y estadística*. Cengage Learning.
- Mesa, L. (2016). El Sistema Educativo Cubano y su carácter inclusivo. *Revista Brasileira de Educação do Campo*, 1(2), 364-380. <https://doi.org/10.20873/uft.2525-4863.2016v1n2p364>
- Meyer, M., & Fang, N. (2019). A qualitative case study of persistence of engineering undergraduates. *International Journal of Engineering Education*, 35(1), 99-108.
- Miliszewska, I., Barker, G., Henderson, F., & Sztendur, E. (2006). The Issue of Gender Equity in Computer Science - What Students Say. *Journal of Information Technology Education*, 5. <https://www.learntechlib.org/d/111535/>
- Rodríguez, N., Montañez, E., & Rojas, I. (2010). Dificultades en contenidos de estadística inferencial en alumnos universitarios. Estudio preliminar. *Revista Electrónica Iberoamericana de Educación en Ciencias y Tecnología*, 2(1), 57-73.
- Shaughnessy, J. M. (2006). Student work and student thinking: An invaluable source for teaching and research. In *3rd International Conference on Teaching Statistics*.
- Sowey, E. R. (2020). Statistics education that inspires long-term retention of learning-a qualitative model. *Statistics Education Research Journal*, 19(1), 15-34. <https://doi.org/10.52041/serj.v19i1.116>
- Universitat de València. (2010). SPSS: Descriptivos. Universitat de València. <https://www.uv.es/innovamide/spss/SPSS/0201a/0201a.wiki>
- Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. *Bulletin of the International Statistical Institute: Proceedings of the Fifty-Second Session of the International Statistical Institute*, 58, 201-204. <https://www.stat.auckland.ac.nz/~iase/publications/5/vall0682.pdf>
- Verdoy, P. J., Beltrán, M. J., & Peris, M. J. (2015). *Problemas resueltos de estadística aplicada a las ciencias sociales*. Publicaciones de la Universitat Jaume I.