

TEMA 1

Conceptos básicos y organización de datos

- 1.1. INTRODUCCIÓN
- 1.2. LA INVESTIGACIÓN EN PSICOLOGÍA
- 1.3. CONCEPTO Y FUNCIONES DE LA ESTADÍSTICA: DESCRIPCIÓN E INFERENCIA
- 1.4. VARIABLES: MEDICIÓN Y CLASIFICACIÓN
- 1.5. DESCRIPCIÓN DE VARIABLES: DISTRIBUCIÓN DE FRECUENCIAS Y REPRESENTACIÓN GRÁFICA
 - 1.5.1. Descripción de variables cualitativas
 - 1.5.2. Descripción de variables ordinales o cuasicuantitativas
 - 1.5.3. Descripción de variables cuantitativas
- 1.6. TENDENCIA CENTRAL, VARIABILIDAD Y FORMA DE UNA VARIABLE: APROXIMACIÓN GRÁFICA
- 1.7. RESUMEN
- 1.8. EJERCICIOS
- 1.9. SOLUCIONES A LOS EJERCICIOS

1.1. INTRODUCCIÓN

En la actualidad, la Estadística se aplica en casi todas las disciplinas. En las ciencias sociales y de la salud su uso ha aumentado exponencialmente en los últimos 30 años. Así, por ejemplo, se utiliza en estudios epidemiológicos (Medicina), en estudios toxicológicos relacionados con la eficacia de los medicamentos (Farmacia), en estudios genéticos y de impacto ambiental (Biología), en muestreos en las prospecciones petrolíferas o hidráulicas (Geología), en los censos de población e información demográfica (Sociología), y en estudios sobre la optimización del coste-beneficio (Economía). En Psicología se utiliza para cualquier cuestión relacionada con la medición de variables psicológicas y con la evaluación, ya sea diagnóstica, de tratamientos, de programas educativos, sociales, etc...

Se puede hacer una distinción entre Estadística **teórica** y **aplicada**; la primera se ocupa de los aspectos matemáticos formales y normativos, y la segunda constituye la aplicación a un campo concreto, como los ejemplos vistos. La estadística aplicada ha recibido distintas denominaciones según su campo de aplicación, tales como bioestadística, psicoestadística o socioestadística. Algunos autores han propuesto para la estadística aplicada la denominación de análisis de datos (Botella, Suero y Ximénez, 2012; Garriga et al., 2009; Merino et al., 2007), término cuyo uso se está extendiendo y que da nombre a este libro.

A pesar de su diversidad de aplicaciones, esta disciplina no es popular entre los estudiantes de ciencias sociales y de la salud, debido posiblemente a la imagen de la Estadística como una rama de las matemáticas de difícil comprensión y ajena a nuestro día a día. Sin embargo, diariamente estamos sometidos a un bombardeo de datos estadísticos. El no ser capaz de distinguir una interpretación rigurosa de unos datos de una defectuosa, hace que se sea vulnerable a la manipulación. En ocasiones, las estadísticas presentadas en distintos medios (de comunicación, políticos, publicidad, entorno laboral...) son incorrectas o engañosas, ya sea por falta de preparación o por voluntad de «maquillar» los resultados. De ahí, la frase atribuida a Benjamín Disraeli (primer ministro del Reino Unido) «hay tres tipos de mentiras: las mentiras, las grandes mentiras y las estadísticas». La Estadística nos proporciona las herramientas necesarias para valorar de manera crítica la información que recibimos.

En este primer tema introductorio se aborda el papel que juega la Estadística en el análisis de los datos en Psicología, se define el concepto de

variable, así como su clasificación y notación simbólica y se analizan los distintos niveles de medida de una variable (nominal, ordinal, de intervalo y de razón). Asimismo, se explica cómo organizar la información en una matriz de datos, y cómo a partir de ella se construye una distribución de frecuencias y se realizan representaciones gráficas para considerar, de un solo vistazo, las características del fenómeno estudiado.

Objetivos del tema:

- Ubicar la materia Análisis de Datos en el plan de estudios del grado en Psicología.
- Establecer el papel del análisis de datos en Psicología.
- Diferenciar y manejar los conceptos básicos, la nomenclatura y las definiciones centrales de la estadística, a fin de poder aplicarlos en el estudio formal de la materia.
- Manejar con soltura las distintas denominaciones y clasificaciones de las variables.
- Entender la importancia de la medición en el ámbito psicológico, distinguiendo entre las distintas escalas o niveles de medida (nominal, ordinal, de intervalo y de razón), y conociendo las relaciones que pueden establecerse en cada una de ellas.
- Saber elaborar, a partir de un conjunto de datos, una distribución de frecuencias, adquiriendo y desarrollando la capacidad para recopilar, organizar, presentar, e interpretar datos numéricos.
- Aplicar las técnicas de representación gráfica adecuadas en función de los datos disponibles (diagrama de barras, diagrama de sectores, histograma y diagrama de líneas).
- Entender, desde una perspectiva gráfica, la tendencia central, la variabilidad y la forma de una variable.

1.2. LA INVESTIGACIÓN EN PSICOLOGÍA

A lo largo de la historia, el hombre se ha servido de diversas formas de conocimiento, tales como el mito, el sentido común o el folclore popular. Con la aparición de la ciencia moderna en el siglo xvii, el método científico

pasó a ser la fuente de conocimiento más utilizada (Navas, 2001). Las ciencias se distinguen entre sí por su objeto de estudio, pero tienen en común el método científico. Al igual que la Biología tiene como objeto de estudio la vida, y la Química las propiedades de las sustancias, la Psicología se sirve del método científico para acercarse a su objeto de estudio: la conducta.

El **método científico** es un procedimiento estructurado que utiliza la ciencia para la ampliación de sus conocimientos. El método científico se caracteriza por ser sistemático y replicable. **Sistemático** porque es un proceso que tiene unas etapas definidas y **replicable** porque los datos obtenidos mediante su uso tienen que poder ser replicados o refutados (en las mismas circunstancias) por cualquier investigador interesado. El método científico, por tanto, proporciona una manera de actuar para afrontar una investigación, a través de las siguientes fases interdependientes:

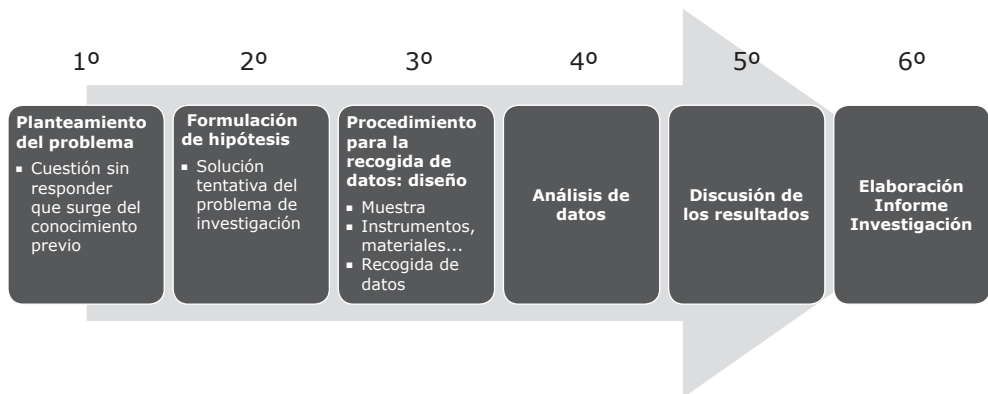


Figura 1.1. Fases de una investigación con el método científico

En primer lugar se define un problema, que puede surgir de teorías ya establecidas, de la lectura de la bibliografía o de la experiencia directa con los hechos. En la mayoría de los casos surgen de lagunas o contradicciones en investigaciones anteriores. A partir de ese problema se plantea una hipótesis, que no es más que una solución tentativa al problema planteado. Las siguientes tres fases tratan de contrastar si la hipótesis planteada es compatible con los hechos. Para ello, es necesario establecer un procedimiento adecuado de recogida de información, analizar los datos obtenidos y discutir los resultados en busca de conclusiones. Por último, hay que elaborar un informe de la investigación que se ha realizado para dar a conocer los resultados obtenidos (Fontes et al., 2010).

Esta asignatura se ocupa de las fases cuarta y quinta de una investigación: el análisis de los datos y la interpretación de los resultados obtenidos. Por tanto, aquí se trata de aprender a procesar los datos recogidos en una investigación con el fin de obtener la información que se precisa para contrastar la hipótesis formulada, y poder dar respuesta al problema planteado. En las asignaturas **Fundamentos de Investigación y Diseños de Investigación y Análisis de Datos** se tratarán de manera detallada el resto de las fases de una investigación científica, así como los posibles diseños a utilizar y el análisis correspondiente a cada uno de ellos.

Ejemplo 1.1. Diversos estudios ponen de manifiesto el efecto que la *ansiedad ante los exámenes* puede tener en la calificación obtenida en la Prueba de Acceso a la Universidad (PAU). Un equipo investigador ha diseñado un programa de tratamiento para paliar este efecto, que combina técnicas de estudio con técnicas de relajación. Para comprobar la eficacia del tratamiento en el examen de Lengua de la PAU se ha seleccionado a una muestra de 40 estudiantes con este problema de ansiedad, que participaron voluntariamente en el estudio. De ellos, la mitad se ha asignado aleatoriamente al grupo 1 (sin tratamiento) y la otra mitad al grupo 2 (que pasará el tratamiento). Al finalizar el curso académico, se recogieron datos sobre las variables relevantes de la investigación, además de algunas variables sociodemográficas, como *sexo*, *nivel de estudios de la madre*, opción de *bachillerato* elegido y *horas de estudio semanales*.

¿Cómo relacionaría los datos de este ejemplo con las fases de una investigación?

Solución:

Las dos primeras fases de la investigación son la definición del problema y la deducción de hipótesis contrastables. En el ejemplo, el problema objeto de estudio es valorar la eficacia del programa de tratamiento que se ha aplicado; para ello, se comprobará si el tratamiento influye en la calificación obtenida en la prueba de Lengua de la PAU. Una hipótesis es una predicción del resultado de la investigación, por lo que, en este caso, como hipótesis, el grupo investigador espera que su tratamiento sea eficaz, lo que operativamente significa que el grupo 2, (que ha pasado el tratamiento que combina técnicas de estudio con técnicas de relajación), tenga un mayor rendimiento en el examen que el grupo 1, que no ha pasado dicho tratamiento.

En la siguiente fase se encontraría la determinación de un plan de trabajo o procedimiento para la recogida de datos, es decir la elección de un diseño de investigación. Aquí, el investigador decide escoger como muestra a 40 estudiantes con problemas de ansiedad ante los exámenes asignándolos de manera aleatoria a los grupos 1 y 2 (para que reciban o no el tratamiento) comparando después sus resultados. Comparar los resultados conlleva el análisis de los datos obtenidos y la discusión de dichos resultados. En esta investigación en concreto se analizarían las calificaciones obtenidas en el examen de Lengua de la PAU por ambos grupos para comprobar si realmente el grupo 2 que ha recibido el tratamiento obtiene mejores puntuaciones que el grupo 1. Para ello se calcularía y se compararía la media de ambos grupos (su cálculo se verá en el Tema 2). Además, en otras investigaciones con otros objetivos, podría ser interesante plantear otro tipo de análisis, como cuantificar la relación entre el nº de horas estudiadas y la calificación en el examen de Lengua, o realizar pronósticos en el rendimiento en el examen en función de la ansiedad ante los exámenes y el nº de horas estudiadas (ambos procedimientos se estudiarán en el Tema 5 del programa). Por último, para difundir los resultados de la investigación se elabora un informe.

En este texto se explicarán de manera detallada los análisis de datos básicos que pueden ser necesarios realizar, tanto en la investigación psicológica como en el ejercicio profesional. El análisis de datos constituye una parte integral no solo de la actividad investigadora, sino también en la práctica profesional. En este sentido, resulta crucial tener unos conocimientos básicos de Estadística para evaluar los resultados de una investigación, y en general para leer de forma crítica las publicaciones de carácter psicológico (ya sean artículos científicos, libros, informes de investigación o notas de prensa).

1.3. CONCEPTO Y FUNCIONES DE LA ESTADÍSTICA: DESCRIPCIÓN E INFERENCIA

La **Estadística** es la rama de las matemáticas que se encarga del estudio de determinadas características en una población, recogiendo los

datos, agrupándolos, organizándolos en tablas, representándolos gráficamente y analizándolos para sacar conclusiones de dicha población.

Teniendo en cuenta las funciones de la Estadística, podemos considerar dos grandes áreas: la Estadística Descriptiva y la Estadística Inferencial.

Mediante la **Estadística Descriptiva** se organizan y resumen conjuntos de observaciones cuantificadas procedentes de una muestra o de la población total. Este resumen puede hacerse mediante tablas, gráficos o valores numéricos. Así, se dispone de distintos procedimientos que nos permiten estudiar las características de una o más variables:

- En el caso de una variable, podemos recurrir a estadísticos que nos indicarán cuáles son los valores más habituales de esa variable (índices de **tendencia central**), hasta qué punto esos valores son similares o diferentes entre sí (estadísticos de **variabilidad**), en qué grado las observaciones se reparten equilibradamente por encima y por debajo de la tendencia central (estadísticos de **asimetría**) y cómo de apuntada es la distribución de las puntuaciones de la variable (estadísticos de **curtosis**). Estos conceptos se abordarán de manera intuitiva al final de este tema, y de manera formal en los Temas 2 y 3.
- En el caso de dos variables podemos utilizar índices que nos indiquen hasta qué punto están ambas variables relacionadas entre sí (**índices de asociación**), así como procedimientos que nos permitirán predecir el valor de una variable en función de otra (**ecuaciones de regresión**). Los Temas 4 y 5 abordarán de manera detallada ambos procedimientos.

Mediante la **Estadística Inferencial** se realizan inferencias acerca de una población basándose en los datos obtenidos a partir de una muestra. Estas generalizaciones de la muestra a la población se basan en el cálculo de probabilidades. Los últimos temas de este texto tratarán sobre **probabilidad e inferencia estadística**.

En una investigación cualquiera, lo habitual es que se desee conocer un **parámetro** o característica de los elementos de una **población**; sin embargo, la población suele ser demasiado extensa para estudiarla al completo (conllevaría un coste inabordable). Por este motivo, se realiza un muestreo con el que se obtiene un conjunto de elementos que representan a la población y se estudia la característica deseada en la **muestra** mediante **estadísticos** que se utilizarán para estimar los parámetros de la población.

En este sentido, en el Ejemplo 1.1 es de esperar que el investigador esté interesado en estudiar si el tratamiento es útil para tratar a los estudiantes con ansiedad ante los exámenes en general. Por tanto, su población objetivo serían los estudiantes que padecen ansiedad ante los exámenes. Dado que no es posible acceder a todos los estudiantes con este problema, escoge una muestra de 40 que son los que realmente participan en la investigación.

Es importante distinguir entre población y muestra: una población es el conjunto de todos los elementos que cumplen una determinada característica objeto de estudio y una muestra es un subconjunto cualquiera de una población. Estos elementos pueden ser personas, animales o cosas que cumplan una definición compartida por la población. Por ejemplo, una población podrían ser los niños con Trastorno por Déficit de Atención e Hiperactividad (TDAH) de la Comunidad de Madrid. En este caso, los elementos de esta población son personas y las características que tienen en común son presentar un diagnóstico de TDAH, ser niños y residir en la Comunidad de Madrid. Una muestra es un subconjunto de una población, pero, ¿por qué elegir un subconjunto y no trabajar con la población completa? Pues en la gran mayoría de casos, es una cuestión de viabilidad, ya que habitualmente no es posible trabajar con la población completa. En el ejemplo que nos ocupa, el número de niños con TDAH puede ser demasiado grande como para trabajar con todos ellos, además de que puede haber familias (elementos de la población) que no deseen participar en el estudio. Por tanto, lo habitual es trabajar con muestras. Para asegurar la representatividad de la muestra se han establecido algunas técnicas de muestreo, como se verá en el Tema 9 de este manual.

En una investigación siempre se trata de caracterizar a la población; en el ejemplo planteado nos puede interesar conocer la inteligencia de los niños con TDAH. En este caso, utilizaríamos un test de inteligencia, por ejemplo el WISC-IV; al no tener acceso a la población completa de niños con TDAH se extrae una muestra de dicha población para obtener el nivel de inteligencia de cada niño de la muestra.

Es preciso distinguir entre parámetro y estadístico. Un parámetro es un índice medido en una población que la describe de alguna manera, mientras que un estadístico es un índice medido en una muestra. Utilizando la estadística inferencial se pronostica el valor de los parámetros poblacionales a partir de los estadísticos muestrales. Así, en el ejemplo de los niños

con TDAH se calcula la media en inteligencia de los niños de la muestra, que es el estadístico \bar{X} , para pronosticar el valor medio en inteligencia de la población, que es el parámetro μ (el valor que realmente nos interesa). Habitualmente los parámetros se representan por letras griegas (μ para la media, σ_x^2 para la varianza y n para la proporción) y los estadísticos por letras latinas (\bar{X} para la media, S_x^2 para la varianza y P para la proporción).

Población: es el conjunto de todos los elementos que cumplen una determinada característica objeto de estudio.

Muestra: es un subconjunto cualquiera de una población.

Parámetro: es una propiedad descriptiva (una medida) de una población. Se denota con letras griegas.

Estadístico: es una propiedad descriptiva (una medida) de una muestra. Se denota con letras latinas.

1.4. VARIABLES: MEDICIÓN Y CLASIFICACIÓN

En el Ejemplo 1.1, para llevar a cabo su estudio, los investigadores pueden registrar los datos de los 40 estudiantes en las variables *sexo*, *nivel educativo de la madre*, *bachillerato elegido*, *nivel de ansiedad ante los exámenes*, *horas de estudio semanales*, *grupo asignado* y *calificación en el examen de Lengua de la PAU*.

Una variable es el conjunto de valores resultantes de medir una característica de interés sobre cada elemento individual de una población o muestra.

Para representar a las variables se utilizan letras latinas mayúsculas. Para referirnos a un valor cualquiera de la variable X se utiliza el subíndice i (X_i), siendo n el número de elementos que componen la muestra, por lo que, de manera genérica, se designa la variable como:

$$X_i \text{ siendo } i = 1, 2, 3, \dots, n$$

Cuando se trata de objetos físicos, el proceso de medición es directo y generalmente sencillo porque es cuestión de seguir unas reglas prescritas expresadas mediante determinadas escalas. Así por ejemplo, es fácil medir la estatura de una persona asignando el número correspondiente de la cinta métrica a la distancia que hay desde sus pies hasta su cabeza. Cuando se trata de medir la timidez de un estudiante en una situación de interacción social, medir ya no es tan sencillo. El reto al que se enfrenta la Psicología es su necesidad de medir en muchas ocasiones variables que no son directamente observables.

Medición es el proceso por el cual se asignan números a objetos o sucesos según determinadas reglas.

El proceso de medición es previo al análisis de datos y especifica el procedimiento de asignación de números a los valores de la variable. Por ejemplo, a los dos valores de la variable *sexo* (hombre y mujer) se les puede asignar los números 1 y 2, y al *peso de una rata* se le puede asignar el número en gramos que da la balanza. Para medir variables psicológicas en muchas ocasiones se utilizan test psicológicos diseñados para ese fin. Su aplicación proporciona una puntuación para cada persona en esa variable.

Otro ejemplo podría ser la valoración de la calidad de vida de un paciente, medida a través de una pregunta que forma parte de un test amplio y que se incluye en bastantes investigaciones sobre salud:

¿Cómo calificarías tu calidad de vida?

- A) Muy mala.
- B) Regular.
- C) Normal.
- D) Bastante buena.
- E) Muy buena.

La regla consiste en asignar un número a cada una de las opciones de respuesta. Así se podría asignar un 1 a escoger la opción «muy mala», un 2 a «regular», un 3 a «normal», un 4 a «bastante buena» y un 5 a «muy buena».

En Psicología se utilizan diferentes escalas de medida en función de la variable a medir, entendiendo como escala de medida el conjunto de reglas o modelos desarrollados para la asignación de números a las variables. Un ejemplo de escala de medida es la escala centígrada de temperatura, que se basa en asignar 0° a la temperatura de congelación del agua y 100° a la de ebullición.

En función de las relaciones matemáticas que puedan verificarse empíricamente entre los distintos valores de una variable y, siguiendo la clasificación de Stevens (1946), pueden distinguirse cuatro tipos de niveles o escalas de medida: nominal, ordinal, de intervalo y de razón.

En la escala nominal solo distinguiremos la igualdad o desigualdad entre dos valores, la escala ordinal añade la posibilidad de establecer un orden, en la escala de intervalo se usa una unidad y tienen sentido las diferencias y, por último, en la escala de razón se pueden comparar dos medidas mediante un cociente.

Para cada tipo de variable existen unos procedimientos estadísticos apropiados para hacer el mejor uso de la información que contienen los valores de las variables.

A) ESCALA NOMINAL

La escala de medida nominal consiste en la asignación, puramente arbitraria de números o símbolos a cada uno de los valores de la variable. Por tanto, la única relación que se tiene en cuenta es la de **igualdad** (y la **desigualdad**), que implica la pertenencia o no a una categoría determinada. En la escala nominal los valores de la variable se denominan categorías.

Usando una escala nominal podemos decidir si un sujeto es igual o diferente a otro, pero no podemos establecer relaciones de orden respecto a esa variable, ni de cantidad. Por ejemplo, si utilizamos la variable enfermedad, distinguiendo entre: (1) «sanos» y (2) «enfermos», carece de sentido establecer relaciones entre estos dos números del tipo $1 + 1 = 2$, ya que sería considerar algo así como que dos personas «sanas» es igual a una persona «enferma».

En las variables nominales se puede asignar a cada valor de la variable cualquier tipo de símbolo. En el ejemplo anterior, en lugar de números podríamos haber utilizado (S) para designar a los «sanos» y (E) a los «enfermos».

A las variables que presentan un nivel de medida nominal se les denomina **variables cualitativas o categóricas**. Las variables cualitativas se clasifican además, en función del número de categorías que presentan. Si una variable presenta solo dos categorías se dice que es una **variable dicotómica** (por ejemplo, el *sexo*); si presenta más de dos categorías se dice que es una **variable politómica** (por ejemplo, el *estado civil*).

Ejemplo 1.2. ¿Qué variables de las que aparecen en el Ejemplo 1.1 pueden considerarse nominales? De ellas, ¿hay alguna dicotómica?

En el Ejemplo 1.1. son variables nominales el *grupo*, el *sexo* y el *bachillerato elegido*.

- La variable *grupo* se utiliza para distinguir a los estudiantes que han recibido el programa de entrenamiento (que combina técnicas de estudio con técnicas de relajación) de los estudiantes que no han recibido dicho tratamiento.
 - Es una variable cualitativa porque pertenecer a un grupo u otro no indica que se posea en mayor o menor grado la característica medida (*grupo*) simplemente que son grupos distintos.
 - Es una variable dicotómica porque únicamente puede adoptar dos valores distintos: grupo 1 y grupo 2.
- La variable *sexo* se utiliza para distinguir a hombres y mujeres.
 - Es una variable cualitativa porque pertenecer a un grupo u otro no indica que se posea la característica en mayor o menor grado, únicamente se distingue entre los distintos valores de la variable.
 - Es una variable dicotómica porque únicamente puede adoptar dos valores distintos: hombre y mujer.
- La variable *bachillerato* se utiliza para distinguir entre los estudiantes que han elegido las distintas opciones posibles de bachillerato.
 - Es una variable cualitativa porque elegir una opción determinada de bachillerato no significa tener un valor mayor o menor de la variable.
 - Es una variable politómica porque puede adoptar más de dos valores distintos.

En ocasiones se categorizan variables que podrían medirse a un nivel superior; en este caso, decimos que una variable se ha dicotomizado si se han establecido dos categorías, y politomizado si se han establecido más de dos categorías. Un ejemplo sería la variable *peso* del roedor de un experimento: aunque podríamos medir exactamente su peso en gramos, puede resultar útil en una investigación dicotomizar la variable *peso* clasificando a las ratas en peso alto y bajo, o politomizarla, estableciendo tres o más niveles de peso.

B) ESCALA ORDINAL

En la escala ordinal se asignan números a objetos para indicar la extensión relativa en que se posee una característica. Los datos pueden utilizarse para jerarquizar u ordenar las observaciones, pero sin indicar la distancia que hay entre las posiciones. Cuando se asignan números es sólo para indicar el **orden** de las posiciones de lo que se está clasificando.

Esta escala no solo permite la identificación y diferenciación de los sujetos sino que además permite establecer relaciones del tipo «mayor que» o «menor que», aunque no se plantea una distancia entre unas medidas y otras. En este caso, la asignación de números a las distintas categorías no puede ser completamente arbitraria, debe hacerse atendiendo al orden existente entre éstas.

Un ejemplo sería la variable *severidad de la enfermedad*, que podría adoptar tres valores: 1 leve, 2 moderado y 3 grave. Podemos decir que no es lo mismo padecer una enfermedad con una intensidad leve o grave, y que la intensidad de la enfermedad en el caso de grave es mayor.

Las variables ordinales también reciben el nombre de **cuasicuantitativas**.

Ejemplo 1.3. ¿Qué variables de las que aparecen en el Ejemplo 1.1 pueden considerarse ordinales?

La única variable ordinal de las que aparecen en el Ejemplo 1.1. es la variable *nivel de estudios de la madre*. Se puede considerar que esta variable adopta cinco valores: Primarios, Enseñanza Secundaria Obligatoria (ESO), Bachillerato, Grado universitario y Posgrado

universitario, que podríamos codificar con los números 1, 2, 3, 4 y 5 respectivamente (por supuesto, serían posibles otras codificaciones alternativas). En este caso, los números no solo indican que son diferentes, sino también un mayor o menor nivel de estudios. Sin embargo, las distancias entre los distintos niveles de estudio no tienen por qué ser iguales. Por tanto, únicamente se verifican las relaciones de igualdad-desigualdad y orden.

C) ESCALA DE INTERVALO

Las escalas de intervalos son aquellas que ordenan los objetos según la magnitud del atributo que representan y proveen intervalos iguales entre las unidades de medida. Con la escala de intervalo, los números asignados a los objetos, no solo permiten decidir si un objeto es igual o diferente a otro o si posee en mayor o menor grado la característica de interés; además, la distancia entre los distintos valores consecutivos de la variable es la misma.

La *inteligencia* medida con un test es un ejemplo de escala de intervalo. Si cuatro personas (A, B, C y D) han obtenido 80, 90, 150 y 160 puntos en un test de inteligencia, podemos decir que la diferencia en inteligencia entre A y B es la misma que entre C y D ($90-80 = 160-150$), ya que el test proporciona una unidad de medida estable. Sin embargo, no se puede afirmar que D sea el doble de inteligente que A aunque tenga el doble de puntuación en el test, ya que para realizar una afirmación de ese tipo sería necesario que el cero de la escala fuera absoluto. En este caso es arbitrario porque obtener un cero en un test de inteligencia no refleja ausencia de la característica medida, no significa que no se posea ni un ápice de inteligencia. Por convención, las puntuaciones obtenidas de test psicológicos se consideran que están medidas en una escala de intervalo.

Como se ha visto en el ejemplo, lo que caracteriza a una escala de intervalo es la existencia de una **unidad de medición** común y constante. En la escala de intervalo el origen es **arbitrario**, y no refleja en ningún momento ausencia de la magnitud que estamos midiendo.

Ejemplo 1.4. ¿Qué variables de las que aparecen en el Ejemplo 1.1 pueden considerarse de intervalo?

En el Ejemplo 1.1. la variable *ansiedad ante los exámenes* es una variable de intervalo porque se trata de una variable que se ha medido con un test psicológico. Así, se puede afirmar que hay igualdad o desigualdad de ansiedad en las distintas puntuaciones del test, que las puntuaciones más altas indican mayor ansiedad que las puntuaciones más bajas y que la distancia en ansiedad entre, por ejemplo, las puntuaciones 14 y 16 es la misma que entre las puntuaciones 18 y 20. Lo único que no se puede admitir en esta variable es que un estudiante que haya obtenido un 0 en el test de ansiedad ante los exámenes no posea en absoluto esta característica ya que el 0 en esta escala es un valor arbitrario, que no refleja ausencia de la variable medida. Lo mismo puede argumentarse de la variable calificación en Lengua

D) ESCALA DE RAZÓN

En la escala de razón los números asignados a los objetos admiten como válidas las relaciones de igualdad-desigualdad, orden, suma, resta, multiplicación y división.

Se caracteriza porque tiene todas las características de una medida de intervalo y, además, se le puede asignar un punto de origen verdadero de valor cero, es decir, el valor cero de esta escala significa ausencia de la magnitud que estamos midiendo. Dado que el cero ya no es arbitrario, sino un valor **absoluto**, se puede afirmar que A tiene dos, tres o cuatro veces la magnitud de la propiedad presente en B. La *altura* y el *peso* son dos ejemplos típicos de escala de razón. Por ejemplo, si una rata de laboratorio pesa 350 gramos y otra 175, podemos afirmar que la 1ª rata pesa el doble que la segunda.

Ejemplo 1.5. ¿Qué variables de las que aparecen en el Ejemplo 1.1 pueden considerarse de razón?

La variable *número de horas de estudio semanales* es una variable de razón. Sus puntuaciones admiten como válidas todas las relaciones: las puntuaciones pueden ser iguales o diferentes, las puntuaciones mayores indican mayor nº de horas estudiadas, la distancia entre

un alumno que ha estudiado 2 horas y otro que ha estudiado 4 es la misma que entre otros dos alumnos que hayan estudiado 6 y 8 horas respectivamente, y si un alumno ha estudiado 6 horas, podemos afirmar que ha estudiado el doble de horas que otro alumno que ha estudiado 3 horas. Esta última relación se puede verificar porque en esta escala el valor 0 es absoluto: si no se estudia ninguna hora se trata de una ausencia completa de la característica medida.

Hay que tener en cuenta que en muchas ocasiones el nivel de medida de una variable va a depender de cómo se haya definido. Por ejemplo, la variable *calificación obtenida en el examen de Lengua de la PAU* puede suscitar dudas razonables sobre su nivel de medida. Si la variable se define como el nivel de conocimientos de Lengua necesarios para ingresar en la universidad se trataría de una variable de intervalo porque con esta interpretación el cero sería arbitrario (obtener un cero en el examen no significa ausencia total de los conocimientos necesarios sino que se ha obtenido un rendimiento nulo en las preguntas en concreto con las que se ha construido el examen). Sin embargo, si en lugar del nivel de conocimientos, nos interesa simplemente contar el número de aciertos (definiendo la variable como el número de aciertos obtenidos en el examen de Lengua de la PAU) se trataría de una variable de razón, ya que aquí el cero sí es absoluto e indicaría ausencia absoluta de preguntas acertadas.

Es muy importante, por tanto, la **definición operativa** de una variable (cómo se define y se registra) porque puede determinar su nivel de medida. La mayoría de las variables psicológicas se considera que están medidas en una escala de intervalo. Así, si la variable *perseverancia*, que es un rasgo de personalidad, se ha medido mediante una prueba psicológica o test, su nivel de medida es de intervalo. Sin embargo, si se define *perseverancia* como el número de intentos o ensayos que realiza una persona para conseguir un objetivo se trata de una escala de razón. Si la variable *discriminación visual* sólo puede tomar dos valores (discrimina/no discrimina) estamos en una escala nominal. Si definimos *discriminación visual* como, por ejemplo, número de veces que una persona discrimina en 20 ensayos, se trataría de una escala de razón.

Las variables medidas en escala de intervalo y de razón son **variables cuantitativas**. Las variables cuantitativas se clasifican, además, en función de los valores numéricos que pueden asignarse en continuas y discretas.

Una **variable continua** es aquella para la que, dados dos valores, siempre se puede encontrar un tercer valor que esté incluido entre los dos primeros. Un ejemplo de variable continua es el *peso*, ya que entre los valores 79 y 80 kg. se pueden considerar uno, dos, tres o todos los decimales que se quiera. Una **variable discreta** es aquella que adopta valores aislados. Por tanto, fijados dos valores consecutivos, no se puede tomar ninguno intermedio. Un ejemplo de variable discreta es el *número de hijos* (huelga decir que se pueden tener dos hijos o tres, pero nunca un valor intermedio entre ambos).

En la Tabla 1.1. se resumen los tipos de variables, las escalas de medida, las características básicas de cada una de ellas, las relaciones válidas que admiten, y algunos ejemplos.

Tabla 1.1. Resumen de las escalas de medida.

Tipo de variable	Escala de Medida	Características básicas	Relaciones válidas	Ejemplos
Cualitativa <ul style="list-style-type: none"> ▪ Dicotómica ▪ Politómica 	→ Nominal	Los números identifican y clasifican objetos	Relaciones del tipo «igual que» o «distinto que»	Sexo, estado civil, raza, diagnóstico clínico.
Cuasicuantitativa	→ Ordinal	Además, los números indican las posiciones relativas de los objetos	Además, relaciones del tipo «mayor que» o «menor que»	Dureza, posición en el ranking de la ATP, grado de satisfacción.
Cuantitativa <ul style="list-style-type: none"> ▪ Discreta ▪ Continua 	Intervalo	Además, hay una unidad de medición común	Además, igualdad o desigualdad de diferencias	Temperatura en grados centígrados, inteligencia.
	Razón	Además, el punto cero es absoluto	Además, igualdad o desigualdad de razones	Longitud, peso, altura, tiempo de reacción.

1.5. DESCRIPCIÓN DE VARIABLES: DISTRIBUCIÓN DE FRECUENCIAS Y REPRESENTACIÓN GRÁFICA

En el apartado anterior ha quedado de manifiesto que en Psicología se trabaja con valores de variables que pueden ser nominales, ordinales, de

intervalo o de razón, con las características propias de cada escala. En cualquier caso, una vez que el investigador ha recabado la información a través del proceso de medida y recogido los datos correspondientes, dispone de un listado o base, comúnmente llamado matriz de datos. La generación de una base de datos supone la codificación previa de las observaciones, la introducción de los datos en algún programa informático, la depuración de los datos ya grabados (detección y tratamiento de los errores de grabación y valores perdidos), y eventualmente la realización de transformaciones de variables que faciliten su posterior tratamiento estadístico. Hay muchos programas estadísticos que se pueden utilizar para organizar y analizar los datos. En concreto, en el curso virtual de la asignatura hay disponibles tutoriales sobre el uso de Excel para hacer distribuciones de frecuencia, gráficos y diversos análisis.

Codificar datos es asignar números a las variables cualitativas y cuasicuantitativas, y registrar los valores de las variables cuantitativas que constituyen la base de datos, así como asignar un código (que puede ser un espacio en blanco o un valor numérico) a los valores perdidos (aquellos que no han sido registrados u observados). En la matriz de datos, los casos se sitúan en las filas y las variables en las columnas.

En la Tabla 1.2 se muestran los datos de los 40 estudiantes en las variables *sexo*, *nivel de estudios de la madre*, *bachillerato elegido*, *puntuación en un test de ansiedad ante los exámenes*, *calificación obtenida en el examen de Lengua de la PAU* y *horas de estudio semanales*, del Ejemplo 1.1. La codificación de las variables se hace en función de sus características:

- La variable *ID* es una variable de identificación que asigna un número a cada estudiante. Se considera una variable nominal, ya que estos números únicamente sirven para identificar a cada estudiante.
- La variable *grupo* (cualitativa y dicotómica) se ha codificado asignando el valor 1 a los estudiantes sin tratamiento y el 2 a los estudiantes con tratamiento.
- La variable *sexo* es una variable cualitativa y dicotómica. Dado que es nominal, para codificarla es posible asignar cualquier número a estos dos valores siempre y cuando se asigne un número diferente a hombres y mujeres. En la Tabla 1.1 a los hombres se les asigna el valor 1 y a las mujeres el valor 2.

- La variable *nivel de estudios de la madre* es una variable ordinal que puede adoptar 5 valores distintos: Primarios, ESO, Bachillerato, Grado universitario y Posgrado universitario. Para codificarla, además de asignar un número diferente a cada valor, hay que tener en cuenta que los números deben cumplir la condición de orden (no se puede asignar al nivel de estudios Primarios un número mayor que el asignado a Bachillerato, por ejemplo). Los números asignados a los distintos valores son: 1 Primarios, 2 ESO, 3 Bachillerato, 4 Grado universitario y 5 Posgrado universitario.
- La variable *Bachillerato elegido* es una variable cualitativa y politómica, que puede adoptar los valores Ciencias, Humanidades y Ciencias Sociales y Arte. Al ser nominal el único requisito para codificarla es asignar un número diferente a cada una de las modalidades. Así, se ha asignado el valor 1 a los estudiantes que han elegido el Bachillerato de Ciencias, el 2 a Humanidades y Ciencias Sociales y el 3 al Bachillerato de Artes.
- En las tres últimas columnas de la tabla se sitúan las tres variables cuantitativas de la investigación. La variable *ansiedad ante los exámenes* recoge las puntuaciones obtenidas en un test diseñado para tal efecto. De manera similar se codifican la *calificación en el examen de Lengua de la PAU* y el *número de horas de estudio semanales*, recogiendo los valores correspondientes a estas variables.

Una vez que los datos están codificados es preciso realizar una depuración de la base de datos, que conlleva el procesamiento de los datos perdidos y de los valores atípicos. Los **datos perdidos** son valores que no han sido registrados, habitualmente porque el participante no ha conseguido ese dato. Existen procedimientos de imputación de datos, basados en los valores válidos de otros casos que se utilizan en ocasiones en variables cuantitativas. Un **dato atípico** es un valor muy diferente al resto de valores de la misma variable. Suelen ser ocasionados por errores al introducir los datos o por valores extremos. Los datos atípicos distorsionan los resultados de los análisis, y por esta razón hay que identificarlos y tratarlos de manera adecuada, generalmente excluyéndolos del análisis. La Tabla 1.2 se basa en un ejemplo ficticio con fines didácticos, en el que no se han introducido datos perdidos.

Si los datos han sido registrados manualmente en un software es recomendable hacer un control de calidad de la grabación de los mismos,

Tabla 1.2. Datos recogidos en la investigación del ejemplo 1.1.

ID	Grupo	Sexo	Nivel de estudios de la madre	Bachillerato	Ansiedad ante los exámenes	Calificación PAU Lengua	Horas de estudio semanales
1	1	1	3	2	5	6	7
2	1	1	2	1	13	4	11
3	1	2	3	2	4	9	16
4	1	1	2	2	15	4	5
5	1	2	2	2	3	8	14
6	1	1	3	1	10	7	10
7	1	1	4	1	7	7	12
8	1	2	1	2	25	1	10
9	1	1	2	3	15	4	2
10	1	2	4	2	5	8	15
11	1	2	3	2	12	5	10
12	1	1	3	1	17	4	10
13	1	1	2	2	30	3	15
14	1	1	1	3	9	5	9
15	1	2	2	2	12	5	9
16	1	1	4	2	4	7	8
17	1	1	1	2	8	6	14
18	1	1	3	1	19	4	8
19	1	1	3	2	15	6	18
20	1	1	2	2	17	4	8
21	2	1	3	1	4	6	4
22	2	2	1	2	14	4	8
23	2	2	4	3	9	7	10
24	2	1	5	1	8	8	12
25	2	1	5	2	5	10	16
26	2	2	2	1	3	8	15
27	2	1	5	1	10	7	13
28	2	2	3	2	7	7	10
29	2	2	4	1	5	7	12
30	2	2	3	3	5	8	18
31	2	2	1	1	5	8	14
32	2	2	3	2	12	6	9
33	2	2	2	3	17	6	11
34	2	1	2	2	3	4	3
35	2	1	1	3	10	6	10
36	2	1	2	1	10	5	9
37	2	1	4	2	6	7	14
38	2	2	1	2	2	8	10
39	2	1	5	1	9	5	9
40	2	2	4	2	22	2	4

revisando la codificación de un porcentaje de los casos, habitualmente un 5% - 10% del total.

Una vez depurada, la base de datos se utiliza para extraer la información relevante. Si tenemos muy pocos datos es posible que la simple inspección visual de los mismos sea suficiente para describir el fenómeno estudiado. Pero esto no es nada frecuente. Habitualmente el número de datos es elevado, por lo que se hace necesario organizar la información mediante una **distribución de frecuencias**.

Una distribución de frecuencias es una tabla en la que se resume la información disponible de una variable. Se sitúan los valores de la variable por filas y en las columnas se dispone el número de ocurrencias por cada valor, porcentajes, etc. La finalidad de las agrupaciones en frecuencias es facilitar la lectura de la información que contienen los datos. Además de la organización de los datos, la distribución de frecuencias cumple dos funciones fundamentales: ofrecer la información necesaria para realizar representaciones gráficas y facilitar los cálculos para obtener los estadísticos que serán objeto de estudio en los próximos temas.

1.5.1. Descripción de variables cualitativas

La descripción de una variable cualitativa consiste básicamente en una distribución de frecuencias y en su representación gráfica mediante un diagrama de barras o de sectores.

En la quinta columna de la Tabla 1.2 aparece el *Bachillerato elegido* por los participantes. Sin embargo, la simple inspección visual de estos datos no es suficiente para que el investigador se haga una idea precisa de cuántos estudiantes han elegido cada una de las modalidades de Bachillerato existentes, por lo que es necesario construir una distribución de frecuencias.

En la distribución de frecuencias de variables cualitativas habitualmente se muestran las frecuencias absolutas, las frecuencias relativas y los porcentajes.

Para construir la tabla de distribución de frecuencias se inspeccionan en primer lugar los valores que toma la variable. En este caso se trata de una variable de carácter cualitativo (nominal) que puede adoptar tres valores distintos. En la primera columna se especifican los valores que adopta la

variable X o el número asignado a ese valor (en la Tabla 1.3 se muestran ambos). En la segunda columna aparece la **frecuencia absoluta** (n_i) que es el número de observaciones en cada categoría. En la siguiente columna aparece la **frecuencia relativa o proporción** de cada categoría (p_i), que se obtiene dividiendo la frecuencia absoluta, n_i , entre el número total de observaciones, que se representa por n . La frecuencia relativa también se expresa en términos de **porcentaje** (P_i) para lo cual hay que multiplicar cada una de las proporciones por cien (cuarta columna).

Tabla 1.3. Distribución de frecuencias de la variable *Bachillerato elegido*.

X	n_i	p_i	P_i
1. Ciencias y Tecnología	13	0,325	32,5
2. Humanidades y CC Sociales	21	0,525	52,5
3. Artes	6	0,15	15
Σ	40	1	100

Pues bien, ahora sí podemos hacernos una idea de la distribución de los estudiantes según el Bachillerato que han elegido; sabemos que el más demandado es el de Humanidades y Ciencias Sociales (un 52,5% de los estudiantes lo eligen) y que el menos demandado es el de Artes (elegido por un 15% del total de estudiantes).

Los dos gráficos más habituales en la descripción de variables cualitativas son los gráficos de barras y los gráficos de sectores. En los gráficos de barra los distintos valores de la variable se sitúan en el eje horizontal y las frecuencias o los porcentajes en el eje de ordenadas. Cada barra representa una categoría de la variable a representar, siendo su altura igual a su frecuencia (o porcentaje). En los gráficos de sectores cada sector representa una categoría de la variable y su ángulo central debe ser proporcional a su frecuencia (o porcentaje). En la Figura 1.2 se muestra el diagrama de barras y el diagrama de sectores de la variable *Bachillerato elegido*. El diagrama de barras se ha construido sobre las frecuencias absolutas de la variable y el diagrama de sectores sobre los porcentajes.

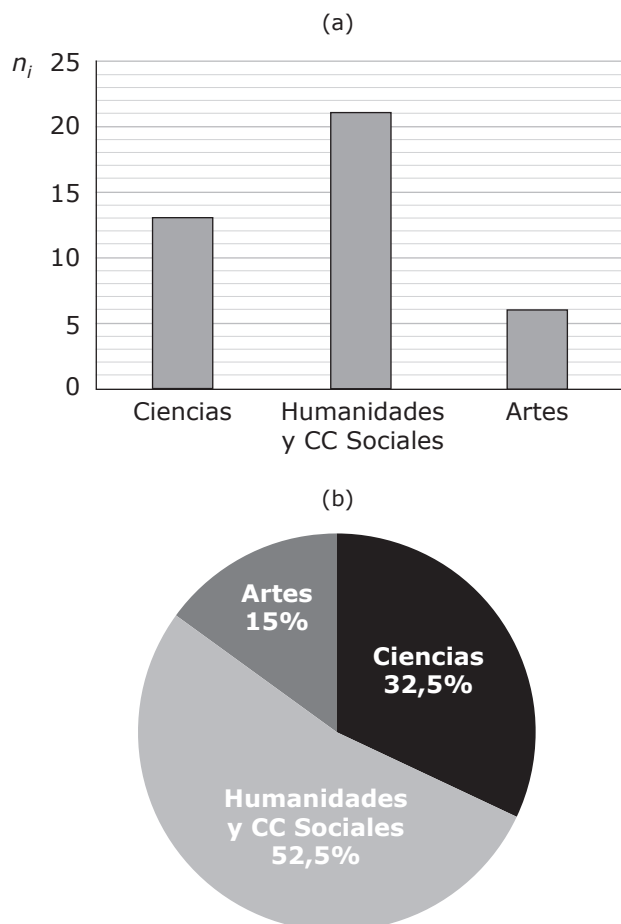


Figura 1.2. Diagrama de barras (a) y diagrama de sectores (b) de la variable *Bachillerato elegido*.

Como se verá en el Tema 2, el único índice apropiado para variables cualitativas es la moda.

Ejemplo 1.6. Se muestra a continuación la distribución de frecuencias de la variable *estado civil* de una determinada muestra. ¿Cuál es la proporción de personas casadas?

X	n_i
Soltero	6
Casado	24
Divorciado	6
Viudo	4
	40

Hay 24 personas casadas (frecuencia absoluta). La proporción o frecuencia relativa de las personas casadas será:

$$p_i = \frac{n_i}{n} = \frac{24}{40} = 0,6$$

1.5.2. Descripción de variables ordinales o cuasicuantitativas

En el caso de variables ordinales se procede de la misma manera, aunque con los valores situados en la tabla de acuerdo a un determinado orden. Por ejemplo, la variable *nivel de estudios de la madre* presenta los valores: Primarios, ESO, Bachillerato, Grado universitario y Posgrado universitario. En la distribución de frecuencias hay que preservar este orden, ya sea empezando por el valor más bajo o más alto de la variable:

Tabla 1.4. Distribución de frecuencias de la variable *nivel de estudios de la madre*.

X	n_i	p_i	P_i	n_a	p_a	P_a
1. Primarios	7	0,175	17,5	7	0,175	17,5
2. ESO	11	0,275	27,5	18	0,450	45
3. Bachillerato	11	0,275	27,5	29	0,725	72,5
4. Grado universitario	7	0,175	17,5	36	0,900	90
5. Posgrado universitario	4	0,1	10	40	1	100
Σ	40	1	100			

En esta tabla se han añadido tres columnas más: la **frecuencia absoluta acumulada** (n_a), la **frecuencia relativa acumulada o proporción acumulada** (p_a) y el **porcentaje acumulado** (P_a), para cada una

de las categorías de respuesta. Para obtener estos valores, simplemente hay que ir acumulando (sumando), desde la categoría de menor valor de la variable a la de mayor valor, las frecuencias absolutas, proporciones o porcentajes, de cada categoría de respuesta. Por ejemplo, la frecuencia absoluta acumulada en el caso de *Bachillerato* es 29, resultado de sumar las frecuencias de los valores anteriores ($7 + 11 = 18$) y la suya propia ($18 + 11 = 29$), indicando que 29 personas presentan un nivel de estudios de Bachillerato o inferior. En las variables nominales carece de sentido el cálculo de las frecuencias acumuladas, ya que sus valores no establecen un orden determinado.

Los conceptos explicados hasta el momento son:

Frecuencia absoluta (n_i): número de veces que se repite cada uno de los valores de una variable. La suma de todas las frecuencias absolutas representa el total de la muestra (n).

Proporción o frecuencia relativa (p_i): cociente entre la frecuencia absoluta de cada valor de la variable (n_i) y el número total de observaciones (n). Formalmente $p_i = n_i/n$.

Porcentaje (P_i): valor de la frecuencia relativa (p_i) multiplicado por cien. Formalmente $P_i = p_i \times 100$

Frecuencia absoluta acumulada (n_a): número de veces que se repite cada valor o cualquiera de los valores inferiores.

Proporción acumulada o frecuencia relativa acumulada (p_a): cociente entre la frecuencia absoluta acumulada y el total de observaciones. Formalmente $p_a = n_a/n$.

Porcentaje acumulado (P_a): valor de la frecuencia relativa acumulada multiplicado por cien. Formalmente: $P_a = p_a \times 100$.

Al igual que las variables cualitativas, las variables ordinales generalmente se representan con un diagrama de barras o un diagrama de sectores. El diagrama de barras también se puede realizar sobre las frecuencias, proporciones o porcentajes acumulados, siempre teniendo en cuenta que es necesario respetar el orden de los valores de la variable representada. En este caso, se ha elegido un diagrama de barras que contiene en el eje horizontal la tabla con los datos que representa el gráfico.

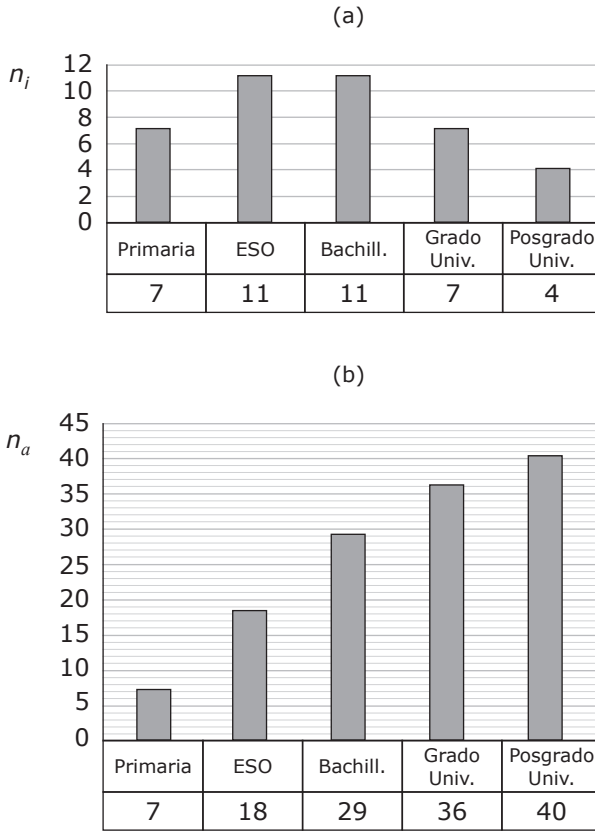


Figura 1.3. Diagrama de barras (a) y diagrama de barras acumulado (b) de la variable *nivel de estudios de la madre*.

Algunos índices apropiados para este tipo de variables son la mediana y la moda (explicados en el Tema 2) y la amplitud intercuartil (explicada en el Tema 3).

1.5.3. Descripción de variables cuantitativas

Al trabajar con variables cuantitativas puede suceder que el número de valores que tome la variable sea reducido (como la variable *nº de hijos*, que habitualmente no adopta valores mayores de 4) o sea muy amplio (como las variables *ansiedad ante los exámenes* y *horas de estudio semanales* de la Tabla 1.2). En el primer caso, para elaborar la distribución de

frecuencias se procede de la forma indicada para variables ordinales y en el segundo será necesario agrupar la variable en intervalos.

La variable *ansiedad ante los exámenes* de la Tabla 1.2 forma parte de este segundo caso. El estudiante con menor puntuación en el test de ansiedad ante los exámenes tiene una puntuación igual a 2 y el que tiene una puntuación mayor ha obtenido una puntuación de 30. Si se actúa de la misma manera que en los ejemplos anteriores, para hacer la distribución de frecuencias (utilizando una fila para cada valor) tendríamos una tabla con una gran cantidad de filas, la mayoría de ellas con una frecuencia absoluta de 0 o de 1, por lo que esta distribución, así presentada, no resultaría útil. En estos casos se recurre a la agrupación en **intervalos**, que consiste en formar grupos de valores consecutivos de la variable. Para ello, se sitúa cada uno de estos grupos en una fila, y se calculan las frecuencias de cada grupo o intervalo de valores, y no de cada valor de la variable.

En primer lugar, hay que decidir qué número de intervalos tendrá la distribución de frecuencias. Siempre habrá varias posibilidades pudiendo optar desde establecer un número muy pequeño de intervalos muy amplios hasta muchos intervalos de muy pequeña **amplitud**. A la hora de tomar esta decisión hay que tener presente que al establecer intervalos siempre se pierde información, ya que ahora la frecuencia no estará referida a un solo valor de la variable, sino a todos los contenidos en el intervalo. Por tanto, esta decisión dependerá del tratamiento que el investigador quiera dar a la variable en su estudio, tratando de encontrar el equilibrio entre la precisión que necesite y la manejabilidad de los datos.

En el Ejemplo 1.1 unos intervalos de amplitud 5 pueden ser apropiados para la variable *ansiedad ante los exámenes* (ver Tabla 1.5.). Así, el primer intervalo contendrá las puntuaciones comprendidas entre 1 y 5, el segundo las puntuaciones comprendidas entre 6 y 10, y así sucesivamente hasta llegar al último intervalo que contiene las puntuaciones comprendidas entre 26 y 30. Estos valores constituyen los **límites aparentes** del intervalo. Para cada intervalo existe un límite inferior y un límite superior.

Tabla 1.5. Distribución de frecuencias con los datos agrupados en intervalos de la variable *ansiedad ante los exámenes* del Ejemplo 1.1.

X	n_i	p_i	n_a	p_a
1 – 5	13	0,325	13	0,325
6 – 10	12	0,3	25	0,625
11 – 15	8	0,2	33	0,825
16 – 20	4	0,1	37	0,925
21 – 25	2	0,05	39	0,975
26 – 30	1	0,025	40	1
Σ	40	1		

La variable *ansiedad ante los exámenes* adopta 29 valores distintos (del 2 al 30). Dado que 29 no es múltiplo de 5, o el intervalo inferior empieza en un valor que no es un valor observado de la variable, o el intervalo superior termina en un valor que no es uno de los valores de la variable. En este caso se ha empezado el primer intervalo (1–5) con el valor 1, que no es un valor que aparezca en la Tabla 1.2 de datos, pero esta distribución también podría empezar en el intervalo 2–6 y terminar en el intervalo 27–31.

Estos límites aparentes tienen la misma unidad de medida que los valores de la variable. Esto es, si los datos son enteros, entonces los límites aparentes son enteros. Si los datos contienen decimales, los límites aparentes tendrán el mismo número de decimales que los datos recogidos. En nuestro ejemplo, los datos son números enteros, por lo que los límites aparentes no contienen decimales.

Con los límites aparentes en la distribución existe discontinuidad entre un intervalo y el siguiente, ya que el límite superior de un intervalo no coincide con el límite inferior del siguiente intervalo. Con los **límites exactos** de una distribución no existe discontinuidad entre un intervalo y el siguiente, ya que el límite superior exacto de un intervalo coincide con el límite inferior exacto del intervalo siguiente. El **Límite Inferior Exacto (LIE)** se calcula restando al valor del límite inferior aparente media unidad de medida y el **Límite Superior Exacto (LSE)** se calcula sumando al valor del límite superior aparente media unidad de medida.

Por tanto, los límites exactos del intervalo 1-5 son 0,5-5,5, los del intervalo 6-10 son 5,5-10,5 y así sucesivamente, de forma que el límite

superior exacto de un intervalo coincide con el límite inferior exacto del siguiente (ver Tabla 1.6).

A partir de los límites aparentes o de los límites exactos se calcula el punto medio del intervalo, que es la semisuma del límite superior e inferior del intervalo. Como se verá en los Temas 2 y 3, el punto medio del intervalo es el valor que se utilizará para el cálculo de algunos índices estadísticos con distribuciones agrupadas en intervalos. Con estos datos, completamos la distribución de frecuencias de la variable *ansiedad ante los exámenes* del Ejemplo 1.1.

Tabla 1.6. Distribución de frecuencias con los datos agrupados en intervalos de la variable *ansiedad ante los exámenes* del Ejemplo 1.1

X Límites aparentes	X Límites exactos	X Punto medio	n_i	p_i	n_a	p_a
1 – 5	0,5 – 5,5	3	13	0,325	13	0,325
6 – 10	5,5 – 10,5	8	12	0,3	25	0,625
11 – 15	10,5 – 15,5	13	8	0,2	33	0,825
16 – 20	15,5 – 20,5	18	4	0,1	37	0,925
21 – 25	20,5 – 25,5	23	2	0,05	39	0,975
26 – 30	25,5 – 30,5	28	1	0,025	40	1
Σ			40	1		

En este ejemplo, el cálculo de los límites exactos de los intervalos es muy sencillo porque la unidad de medida de la variable *ansiedad ante los exámenes* es 1. Esto es así porque sus valores son números enteros (sin decimales). Por tanto, sumar y restar media unidad de medida al límite superior e inferior, respectivamente, supone sumar y restar 0,5 (que es la mitad de 1). Sin embargo, cuando los límites aparentes contienen decimales, la unidad de medida de la variable ya no será 1, y el número que habrá que sumar y restar para calcular los límites exactos dependerá del número de decimales que contienen los valores de la variable. Así:

- Si los límites aparentes son enteros, la unidad de medida de la variable es 1, y su mitad es 0,5, que es la cantidad que habrá que restar al límite inferior y sumar al límite superior para calcular los límites exactos.
- Si los límites aparentes son números con un decimal, la unidad de medida de la variable es 0,1, por lo que la cantidad a sumar y restar para calcular los límites exactos será 0,05.

- Si los límites aparentes son números con dos decimales, la unidad de medida de la variable es 0,01, por lo que la cantidad a sumar y restar para calcular los límites exactos será 0,005.
- Y así sucesivamente...

Por ejemplo, si se mide el tiempo que se emplea en ejecutar una determinada tarea, y los valores resultantes oscilan ente 3,01 segundos y 3,30 segundos, se podría establecer una distribución de frecuencias con 6 intervalos, como se muestra en la Tabla 1.7:

Tabla 1.7. Límites aparentes de la variable *tiempo empleado en ejecutar una determinada tarea.*

X
3,01 – 3,05
3,06 – 3,10
3,11 – 3,15
3,16 – 3,20
3,21 – 3,25
3,26 – 3,30

En ese caso nuestra unidad de medida es 0,01, ya que los valores de la variable contienen dos decimales. Por eso, para calcular los límites exactos hay que sumar y restar la mitad de esta unidad de medida que es 0,005. Así, los límites exactos serían:

Tabla 1.8. Límites de la variable *tiempo empleado en ejecutar una determinada tarea.*

X Límites aparentes	X Límites exactos
3,01 – 3,05	3,005 – 3,055
3,06 – 3,10	3,055 – 3,105
3,11 – 3,15	3,105 – 3,155
3,16 – 3,20	3,155 – 3,205
3,21 – 3,25	3,205 – 3,255
3,26 – 3,30	3,255 – 3,305

Los nuevos conceptos que han aparecido son:

Intervalo: cada uno de los grupos de valores que ocupan una fila en una distribución de frecuencias.

Límites aparentes: son los valores que delimitan el grupo de valores que constituyen un intervalo. Para cada intervalo existe un Límite Inferior Aparente (LIA), que es el valor menor del intervalo y un Límite Superior Aparente (LSA), que es el valor mayor incluido en el intervalo.

Límites exactos o reales: son aquellos que no presentan discontinuidad entre un intervalo y el siguiente. Para cada intervalo existe un Límite Inferior Exacto (LIE) y un Límite Superior Exacto (LSE). El límite inferior exacto es el valor que resulta de restar al límite inferior aparente media unidad de medida. El límite superior exacto es el valor que resulta de sumar al límite superior aparente media unidad de medida.

Punto medio del intervalo (PM): es la suma de los límites exactos o de los límites aparentes de un intervalo dividido entre dos. Formalmente:

$$PM = \frac{LIE + LSE}{2} \quad \text{ó} \quad PM = \frac{LIA + LSA}{2}$$

Amplitud del intervalo: es la diferencia entre el límite superior exacto y el límite inferior exacto.

A un intervalo que no tiene límite inferior o límite superior se le denomina **intervalo abierto**. Por ejemplo, si en la variable *ansiedad ante los exámenes* del Ejemplo 1.1. hubiera dos sujetos con una puntuación de 41 y 43 respectivamente, se puede optar por establecer el intervalo abierto «más de 30», en lugar de añadir los tres intervalos correspondientes 31-35, 36-40 y 41-45, dos de ellos con frecuencia nula.

Los gráficos más habituales para representar a una variable cuantitativa discreta son el diagrama de barras y el diagrama de líneas. En el caso de variables cuantitativas continuas agrupadas en intervalos en lugar del diagrama de barras se utiliza el **histograma**.

El histograma es una extensión del diagrama de barras que dibuja los rectángulos unidos entre sí, indicando de este modo que existe continuidad en los valores de las variables. Un histograma, es por tanto, un gráfico de variable continua dividida en intervalos en los que se eleva un rectángulo con área proporcional a su frecuencia. El histograma puede construirse sobre frecuencias absolutas, frecuencias relativas o porcentajes, ya sean o no acumulados. En la Figura 1.4 se muestra un histograma (a) y un histograma acumulado (b) de la variable *ansiedad ante los exámenes*.

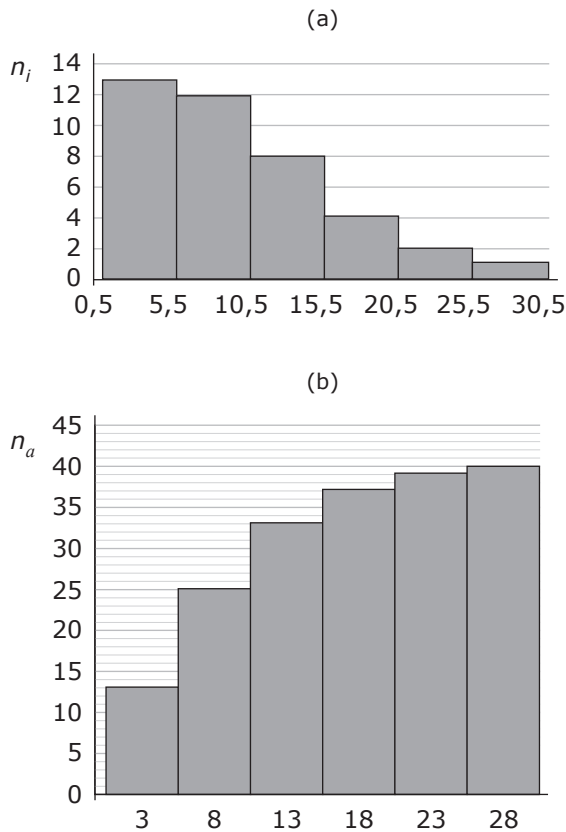


Figura 1.4. Histograma (a) e histograma acumulado (b) de la variable *ansiedad ante los exámenes*.

En el eje horizontal de un histograma se sitúan los límites exactos de los intervalos o su punto medio. El histograma (a) se ha realizado sobre

los límites exactos de los intervalos y el histograma acumulativo (b) se ha realizado sobre los puntos medios de los intervalos.

El **diagrama de líneas** se construye situando un punto a una altura proporcional a la frecuencia en cada valor o en el punto medio de cada intervalo (si la variable está agrupada en intervalos). Finalmente se unen los puntos para formar una línea. A este gráfico también se le denomina polígono de frecuencias. En la Figura 1.5 se muestra el diagrama de líneas de la variable ansiedad ante los exámenes en proporciones (a) y en proporciones acumuladas (b).

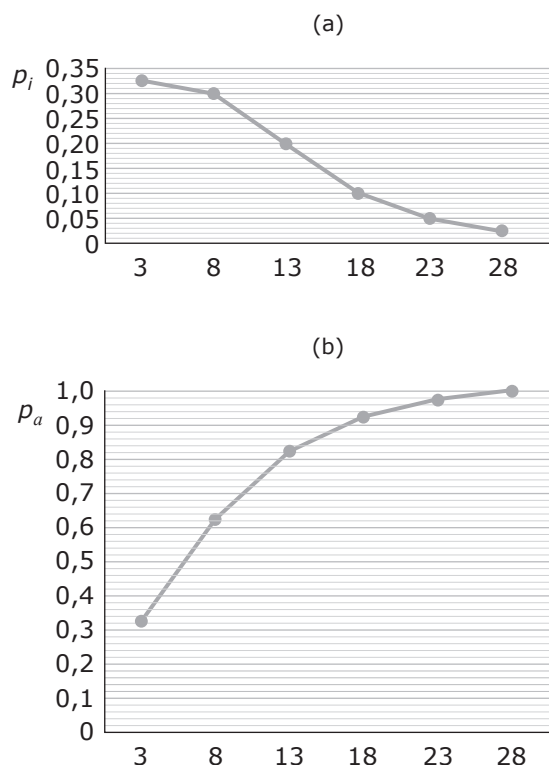


Figura 1.5. Diagrama de líneas (a) y diagrama de líneas acumulativo (b) de la *variable ansiedad ante los exámenes*.

Ejemplo 1.7. Construye una distribución de frecuencias con los datos de la variable *calificación obtenida en el examen de Lengua de la PAU* de la Tabla 1.2.

Hay variables con un número de valores determinado, que hace posible utilizar una distribución de frecuencias con o sin intervalos. Así sucede con la variable *calificación obtenida en el examen de Lengua de la PAU* del Ejemplo 1.1. Se puede trabajar con la distribución de frecuencias sin agrupar, tal y como aparece en la Tabla de la izquierda o con la distribución de frecuencias agrupada de la Tabla de la derecha, en la que se ha elegido agrupar los datos en 5 intervalos de amplitud 2.

(a)				(b)			
X	n_i	p_i	P_i	X	n_i	p_i	P_i
1	1	0,025	2,5	1 - 2	2	0,05	5
2	1	0,025	2,5	3 - 4	9	0,225	22,5
3	1	0,025	2,5	5 - 6	12	0,3	30
4	8	0,200	20	7 - 8	15	0,375	37,5
5	5	0,125	12,5	9 - 10	2	0,05	5
6	7	0,175	17,5	Σ	40	1	100
7	8	0,200	20				
8	7	0,175	17,5				
9	1	0,025	2,5				
10	1	0,025	2,5				
Σ	40	1	100				

Finalmente, para describir una variable cuantitativa se utilizan algunos índices estadísticos que se verán en los próximos temas, los más frecuentes son la media (Tema 2) y la desviación típica (Tema 3).

1.6. TENDENCIA CENTRAL, VARIABILIDAD Y FORMA DE UNA VARIABLE: APROXIMACIÓN GRÁFICA

En el apartado anterior se ha explicado cómo describir cualquier tipo de variable mediante una tabla de datos (su distribución de frecuencias) y la representación gráfica más adecuada. En los dos próximos temas se explicará cómo describir las variables mediante los índices estadísticos adecuados. Estos índices se utilizan para medir la **tendencia central**, **variabilidad** y **forma** de la distribución de una variable. Pero, antes de

calcular estos índices se tratará de ver gráficamente qué característica de la variable pretenden evaluar. Para hacerlo, se utilizarán curvas suavizadas, que son histogramas basados en un gran número de observaciones, cuyos ángulos se han suavizado. Así, si disponemos de los datos de una muestra en una variable X (Figura 1.6. A) y hacemos esos intervalos más pequeños (Figura 1.6. B), y más pequeños aún (Figura 1.6. C), al trazar un diagrama de líneas sobre los puntos medios de esos intervalos, la línea resultante será una curva.

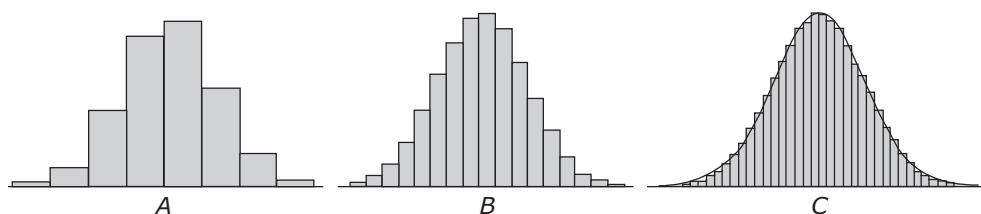


Figura 1.6. Histograma de una variable (A) disminuyendo la amplitud de los intervalos (B), y disminuyendo aún más su amplitud (C).

A) TENDENCIA CENTRAL

La tendencia central de una distribución se refiere al lugar donde se centra una distribución particular en la escala de valores.

La Figura 1.7 podría representar, por ejemplo, la estatura medida en un grupo de 1000 hombres nacidos en 1950 (A) y en otro grupo de hombres nacidos en 1990 (B). Se puede apreciar, atendiendo al eje horizontal en el que aparece la estatura en centímetros que, en líneas generales, los hombres nacidos en 1990 son más altos que los nacidos en 1950. Eso no significa que todos los nacidos en 1990 sean más altos (se puede observar que hay solapamiento entre las curvas).

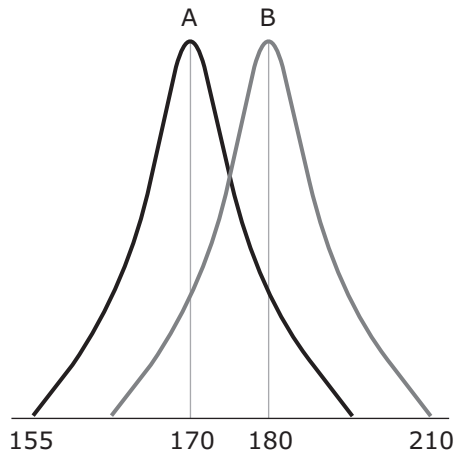


Figura 1.7. Ejemplo de dos distribuciones con tendencias centrales distintas.

A la vista de los gráficos, se puede afirmar que la tendencia central de los dos grupos es distinta (las curvas no se solapan completamente) y que el grupo B (el nacido en 1990) tiene una estatura promedio mayor que el grupo A (porque la curva del grupo B está situada a la derecha, en puntuaciones más altas de estatura).

Esta centralidad o tendencia central puede cuantificarse mediante unos índices conocidos como estadísticos de tendencia central, que se explicarán en el próximo tema.

B) VARIABILIDAD

Esta propiedad se refiere al grado de concentración de los valores entre sí o con respecto a un valor central de la distribución. Una distribución de frecuencias es **homogénea** (tiene poca variabilidad) si los valores de la distribución están cercanos al promedio y es **heterogénea** (tiene mucha variabilidad) si los valores se dispersan mucho con respecto al promedio.

En la Figura 1.8 el grupo A representa, por ejemplo, las puntuaciones en inteligencia medidas en un grupo de niños de distintos colegios de la geografía española mientras que el grupo B representa las puntuaciones en inteligencia de un grupo de niños de altas capacidades.

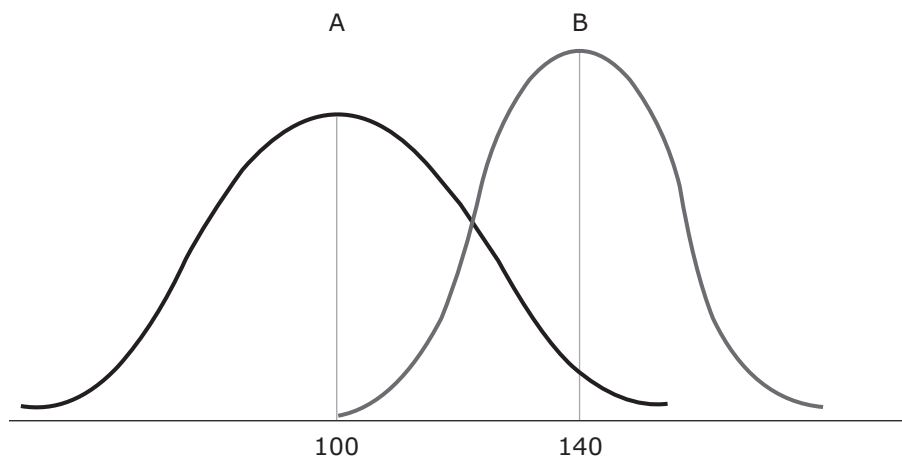


Figura 1.8. Ejemplo de dos distribuciones con tendencia central y variabilidad diferentes.

En este caso, además de una tendencia central distinta (el grupo B presenta, en líneas generales, un nivel mayor de inteligencia que el grupo A) podemos apreciar que las puntuaciones en inteligencia del grupo de estudiantes con altas capacidades están más próximas entre sí que las del otro grupo. Por tanto, el grupo A presenta una mayor variabilidad en inteligencia que el grupo B.

C) FORMA

Para estudiar la forma de una variable se analiza su asimetría y su curtosis.

La **asimetría** se refiere al grado en que los datos se reparten equilibradamente por encima y por debajo de la tendencia central. Una distribución será **simétrica** cuando al dividirla en dos partes iguales, las dos mitades se superponen. Una distribución tiene **asimetría positiva** cuando la mayor concentración de puntuaciones se produce en la parte baja de la escala y **asimetría negativa** cuando la mayor parte de las puntuaciones se sitúan en la parte alta de la escala.

En la Figura 1.9 se ha representado la puntuación obtenida por un grupo de alumnos en un examen muy difícil (A), en un examen de dificultad intermedia (B) y en un examen muy fácil (C).

Como se puede ver en el gráfico, el conjunto de puntuaciones presenta una distribución asimétrica positiva si la mayoría de las puntuaciones ob-

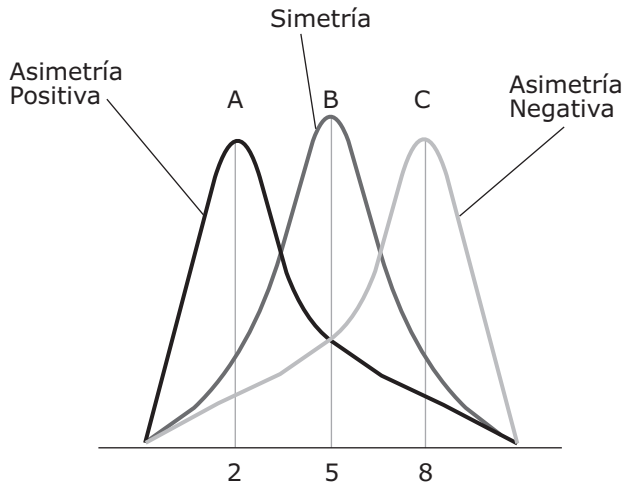


Figura 1.9. Ejemplo de tres distribuciones con distinta asimetría. La distribución A es asimétrica positiva, la distribución B es simétrica y la distribución C es asimétrica negativa.

tenidas son bajas (caso del examen A que es difícil), es simétrica cuando hay un número similar de puntuaciones a ambos lados del centro de la distribución (caso del examen B) y la distribución es asimétrica negativa si la mayoría de las puntuaciones son altas (caso del examen C que es fácil).

La **curtosis** se refiere al grado de apuntamiento de los datos (ver Figura 1.10). Si la distribución de frecuencias es muy apuntada se llama leptocúrtica (A), y si es muy aplastada se denomina platicúrtica (C). Si su grado de apuntamiento es intermedio se denomina mesocúrtica (B).

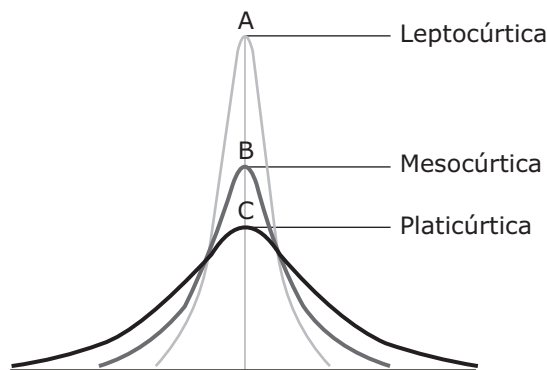


Figura 1.10. Ejemplo de tres distribuciones con distinta curtosis.

1.7. RESUMEN

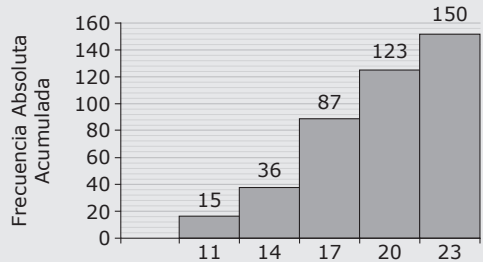
En este capítulo se ha tratado el papel que juega el Análisis de datos dentro del método general de la ciencia y algunos conceptos importantes relacionados con el análisis de datos. Posteriormente, se ha abordado el concepto de variable, su notación y clasificación, además de tratar el problema de la medición y los distintos tipos de escala: nominal, ordinal, de intervalo y de razón. También se ha tratado la organización y tabulación de los datos de variables cualitativas, cuasicuantitativas y cuantitativas, mediante la confección de una distribución de frecuencias. Además, se han presentado algunas formas de representar gráficamente una distribución de frecuencias, de modo que su visión aporte una información de carácter general acerca de la variable objeto de estudio. Por último, hemos adelantado de manera intuitiva los aspectos más relevantes que se deben analizar en toda distribución de frecuencias: la tendencia central, la variabilidad y la forma de la distribución (asimetría y curtosis), que serán objeto de estudio en los próximos temas.

1.8. EJERCICIOS

- 1.1. El *número de aciertos* en un examen es una variable: A) nominal; B) ordinal; C) de razón.
- 1.2. La variable *número de caras obtenidas al lanzar al aire dos monedas* es: A) dicotómica; B) discreta; C) continua.
- 1.3. ¿En qué escala de medida el origen no es arbitrario? A) En la escala nominal; B) En la escala de intervalo; C) En la escala de razón.
- 1.4. ¿Cuál es el nivel de medida de un ítem cuyas opciones de respuesta son: 1 = totalmente en desacuerdo, 2 = en desacuerdo, 3 = de acuerdo y 4 = totalmente de acuerdo? A) Nominal; B) Ordinal; C) De intervalo.
- 1.5. Se han asignado los valores 1, 2 y 3 a pacientes con un problema de claustrofobia muy leve, moderado y alto, respectivamente. ¿Qué nivel de medida tiene la variable *grado de claustrofobia*? A) Nominal; B) Ordinal; C) De razón.

- 1.6.** Las variables dicotómicas: A) solo admiten dos valores posibles; B) admiten como mínimo dos valores posibles; C) admiten dos o más valores siempre y cuando se trate de una variable nominal.
- 1.7.** El Centro de Investigaciones Sociológicas (CIS) realiza de manera regular una encuesta a los ciudadanos españoles mayores de edad. En una de ellas, preguntó a 1600 ciudadanos sobre el principal problema que existe actualmente en España, encontrando que la mayoría de los encuestados (el 52,5%) opinaron que el paro era el principal problema. ¿Cuál es la población objeto de estudio? A) Los 1600 ciudadanos; B) La población española; C) La población española mayor de edad.
- 1.8.** Continuando con el ejercicio anterior, 52,5% es el valor de: A) un parámetro; B) un estadístico; C) una muestra.
- 1.9.** La variable *flexibilidad psicológica*, recogida en la Gráfica 1, es: A) politómica; B) cuasicuantitativa; C) cuantitativa.
- 1.10.** ¿Cuál es la amplitud de los intervalos en los que está agrupada la variable *flexibilidad psicológica*? A) 2; B) 3; C) 4.
- 1.11.** Los límites exactos del primer intervalo de la variable flexibilidad psicológica son: A) 10–12; B) 9,5–12,5; C) 10,5–12,5.
- 1.12.** Con los datos de la Gráfica 1, la frecuencia relativa del tercer intervalo de puntuaciones es: A) 0,34; B) 0,76; C) 0,58.
- 1.13.** Si queremos construir un intervalo para el valor 18,56 de una variable, ¿cuáles son los límites exactos de dicho intervalo? A) 18,55–18,56; B) 18,555–18,565; C) 18,565–18,565.
- 1.14.** En un experimento de atención visual focalizada se ha utilizado como variable dependiente el *tiempo de reacción* en milisegundos a un determinado estímulo visual presentado en la pantalla de un ordenador. Los tiempos de reacción obtenidos han sido:
520, 487, 458, 399, 458, 465, 502, 389, 444, 478, 415, 501, 388, 466, 438, 474, 458, 468, 479, 511, 458, 499, 487, 468, 423, 415,

Gráfica 1. Puntuaciones en un test de flexibilidad psicológica (X) de una muestra de 150 personas. En el eje horizontal se muestran los puntos medios de los intervalos.



429, 473, 426, 409, 450, 410, 439, 490, 480, 417, 432, 491, 451, 382, 458, 510, 390, 433, 487, 429, 389, 477, 466, 520.

¿Qué nivel de medida tiene la variable *tiempo de reacción*? A) Ordinal; B) De intervalo; C) De razón.

1.15. La distribución de frecuencias de la variable *tiempo de reacción* del ejercicio anterior es:

A)		B)		C) cualquiera de las dos anteriores
<i>X</i>	<i>n_i</i>	<i>X</i>	<i>n_i</i>	
381-400	6	400 o menos	6	
401-420	5	401-425	6	
421-440	8	426-450	9	
441-460	8	451-475	13	
461-480	11	476-500	10	
481-500	6	más de 500	6	
501-520	6			

1.16. La amplitud de los intervalos de la distribución de frecuencias A del ejercicio anterior es: A) 19; B) 20; C) 25.

1.17. Según los datos del ejercicio 1.15, ¿Qué porcentaje de sujetos tardó 450,5 milisegundos o menos? A) 42%; B) 54%; C) 68%.

1.18. ¿Cuáles son los límites exactos del primer intervalo de la distribución de frecuencias de la alternativa A del ejercicio 1.15? A) 380,5 – 400,5; B) 380 – 401; C) 381,5 – 400,5.

1.19. Atendiendo a la distribución de frecuencias de la alternativa A del Ejercicio 1.15., el punto medio del primer intervalo es: A) 390; B) 390,5; C) 391.

1.20. ¿Qué gráfico representaría de manera apropiada los valores de la variable *tiempo de reacción* del Ejercicio 1.15? A) Diagrama de barras; B) Histograma; C) Diagrama de sectores.

1.9. SOLUCIONES A LOS EJERCICIOS

1.1. Solución: C

Es una variable de razón, ya que se dispone de una unidad constante de medida y el cero es absoluto.

1.2. Solución: B

Se trata de una variable discreta, ya que puede adoptar los valores 0, 1 y 2 pero no podría adoptar un valor intermedio entre ellos (al lanzar al aire dos monedas nunca se podría sacar cara y media, por ejemplo).

1.3. Solución: C

En la escala de razón el origen de la escala no es arbitrario, sino que representa un origen real que corresponde a la ausencia (valor cero) de la característica que se está midiendo.

1.4. Solución: B

El nivel de medida es ordinal, porque los números asignados a las opciones de respuesta solo nos permiten diferenciarlas y ordenarlas. Si una persona escoge la opción 4, solo podemos afirmar que está más de acuerdo con la cuestión planteada que otra persona que ha escogido la opción 3, pero no podemos saber cuánto más de acuerdo está.

1.5. Solución: B

El nivel de medida es ordinal, ya que podemos diferenciar entre tres niveles de claustrofobia y ordenarlos en función de su gravedad, pero no podemos precisar la distancia entre un nivel y otro.

1.6. Solución: A

Una variable dicotómica se define como aquella que solo puede presentar dos categorías o valores.

1.7. Solución: C

A) es el tamaño muestral y B) incluye a toda la población española, cuando en el estudio solo interesan los mayores de edad.

1.8. Solución: B

Es el valor de un estadístico, ya que 52,5 es un porcentaje obtenido sobre los 1600 encuestados que forman parte de la muestra.

1.9. Solución: C

La variable puntuaciones en un test de flexibilidad psicológica está en una escala de intervalo, ya que hay una unidad de medición común y constante pero el cero es arbitrario. Todas las puntuaciones que provienen de test psicológicos se consideran de intervalo. Por tanto, se trata de una variable cuantitativa.

1.10. Solución: B

La diferencia entre cada dos puntos medios consecutivos es 3, por lo que la amplitud de los intervalos es 3.

1.11. Solución: B

Por la distancia entre los puntos medios del histograma de la Gráfica, se sabe que los intervalos son de amplitud 3. El primer punto medio es 11, por lo que para que la amplitud sea 3 sus límites aparentes serán 10-12 y los límites exactos 9,5-12,5. Los límites aparentes y exactos de la variable flexibilidad psicológica son:

X Límites aparentes	X Límites exactos
10-12	9,5-12,5
13-15	12,5-15,5
16-18	15,5-18,5
19-21	18,5-21,5
22-24	21,5-24,5

1.12. Solución: A

El Gráfico 1 se basa en las frecuencias absolutas acumuladas, para calcular las frecuencias absolutas hay que restar la frecuencia acumulada anterior. Así, la frecuencia absoluta del intervalo 10-12 será 15 (no hay frecuencia acumulada anterior), la frecuencia absoluta del intervalo 13-15 será $36-15 = 21$, y así sucesivamente. Para calcular la frecuencia relativa del tercer intervalo se divide su frecuencia absoluta por el total de observaciones, $51/150 = 0,34$. En la Tabla se muestran todas las frecuencias absolutas y relativas de la variable.

X	n_a	n_i	p_i
10-12	15	15	0,10
13-15	36	21	0,14
16-18	87	51	0,34
19-21	123	36	0,24
22-24	150	27	0,18

1.13. Solución: B

El valor 18,56 tiene dos decimales, por lo que la unidad de medida de la variable es 0,01 habrá que sumar y restar la mitad de esta cantidad que es 0,005 para calcular los límites exactos. Así,

$$\text{Límites exactos} = 18,56 \pm 0,005 = \begin{cases} 18,555 \\ 18,565 \end{cases}$$

1.14. Solución: C

De razón, porque el cero representa la ausencia total de la característica medida (del tiempo).

1.15. Solución: C

Ambas reflejan adecuadamente los datos del ejercicio 1.14, diferenciándose únicamente en las decisiones tomadas respecto al número y amplitud de los intervalos.

1.16. Solución: B

La amplitud es la diferencia entre el límite superior exacto y el límite inferior exacto, por tanto $400,5 - 380,5 = 20$.

1.17. Solución: A

Si se utilizara la Tabla A el porcentaje de sujetos que tarda 450,5 milisegundos o menos estaría en el intervalo 441-460, que es un intervalo que incluye valores superiores a 450,5 milisegundos, por lo que la frecuencia absoluta de este intervalo puede incluir sujetos con un tiempo de reacción superior. En este Tema no se ha estudiado aún el cálculo de los percentiles que resolvería este problema, por lo que hay que utilizar la Tabla B.

Para obtener el porcentaje de sujetos que tardó 450,5 milisegundos o menos hay que calcular el porcentaje acumulado del intervalo 426-450. Para facilitar este cálculo, se añaden además las columnas correspondientes a las frecuencias acumuladas (absolutas y relativas).

<i>X</i>	<i>n_i</i>	<i>n_a</i>	<i>p_a</i>	<i>P_a</i>
400 o menos	6	6	0,12	12
401-425	6	12	0,24	24
426-450	9	21	0,42	42
451-475	13	34	0,68	68
476-500	10	44	0,88	88
más de 500	6	50	1	100

El 42% de sujetos tardó 450,5 milisegundos o menos.

1.18. Solución: A

Dado que los valores de la variable no tienen decimales, basta con restar y sumar 0,5 a los límites aparentes para obtener los límites exactos. Así, $381 - 0,5 = 380,5$ y $400 + 0,5 = 400,5$.

1.19. Solución: B

El Punto medio del intervalo es la suma de los límites exactos o de los límites aparentes, dividido entre dos:

Con los límites aparentes:

$$PM = \frac{LIA + LSA}{2} = \frac{381 + 400}{2} = 390,5$$

Con los límites exactos:

$$PM = \frac{LIE + LSE}{2} = \frac{380,5 + 400,5}{2} = 390,5$$

1.20. Solución: B

El histograma representa adecuadamente los valores de esta variable, ya que es cuantitativa. El diagrama de barras (opción A) no se puede utilizar en distribuciones de frecuencias agrupadas en intervalos y el diagrama de sectores (opción C) no se utiliza en variables cuantitativas.